Technische Universität Ilmenau
Fakultät für Mathematik
und Naturwissenschaften
Institut für Mathematik
http://www.mathematik.tu-ilmenau.de/Math-Net/index_de.html

Postfach 10 05 65
D - 98684 Ilmenau
Germany
Tel.: 03677/69 3267
Fax: 03677/69 3272
Telex: 33 84 23 tuil d.
email: werner.neundorf@tu-ilmenau.de

Preprint No. M 19/04

Abstiegsverfahren

Teil I

Werner Neundorf

Oktober 2004

‡MSC (2000): 65F10, 65G50, 65-05, 65Y20

Zusammenfassung

Die Entwicklung moderner numerischer Algorithmen hat zu einem hohen Bedarf an effizienten, robusten iterativen Gleichungssystemlösern geführt. So entstand eine Vielzahl von Verfahren, die man zur Gruppe der Projektionsmethoden und Krylov-Unterraum-Methoden zählt.
Gegenstand der Betrachtungen in dieser mehrteiligen Arbeit sind grundlegende Abstiegsverfahren als Vertreter dieser Algorithmengruppe, die auf Minimierungsaufgaben nach Umformulierung eines regulären linearen Gleichungssystems führen.
Unter bestimmten Voraussetzungen an die Matrix des Gleichungssystems werden geeignete Funktionale konstruiert und damit der Weg des Abstiegs illustriert.

Das Verhalten der Abstiegsverfahren ist in den normalen gutartigen Fällen hinreichend bekannt und untersucht worden. Hier soll zunächst eine Gegenüberstellung zu anderen Situationen gemacht werden, wo man unter veränderten Voraussetzungen arbeitet, und damit der typische Charakter der Minimierungsaufgabe nicht mehr vorhanden ist. Dabei ist teilweise noch mit zufrieden stellenden Ergebnissen zu rechnen, es können aber auch starke Abweichungen vom Normalfall auftreten. Diese Darstellungen findet der Leser in den Teilen I und II.
Des Weiteren betrachten wir in einem Teil III die Abstiegsverfahren als polynomiale Iterationsverfahren und untersuchen den Einfluss von Rundungsfehlern bei der Implementation der Verfahren im Zusammenhang mit ihrer (eventuell) theoretischen Endlichkeit sowie mit der Notation des Formelapparates und seiner numerischen Auswertung (Fehlerverhalten und Fehlererinnerung).

Praktische Rechnungen mit kleindimensionierten Beispielen, wo man dies auch gut illustrieren kann, demonstrieren die unterschiedlichen Situationen.

Eileen Caddy

Spuren auf dem Weg zum Licht

Hört auf zu versuchen,
alles gedanklich außuarbeiten,
das bringt euch nirgendwohin.
Lebt auf der Intuition und Inspiration,
und laßt euer ganzes Leben
eine Offenbarung sein.

Inhaltsverzeichnis

Kapitel 1

Funktionale und Abstiegsszenarien

Wir betrachten im $\mathbb{R}^n$ das lineare Gleichungssystem (LGS)

$$Ax \;=\; b \tag{1.1}$$

mit der regulären Matrix $A \in \mathbb{R}^{n,n} = (a_{ij})_{i,j=1}^n$ und der exakten Lösung $x^* \in \mathbb{R}^n$.
Die Überführung in eine Optimierungsproblem ist verknüpft mit der Definition geeigneter zu minimierende Funktionale, deren eindeutige Minimumstelle eben diese Lösung x^* sein sollte.
Unsere Vorstellungen im dreidimensionalen Raum sagen uns natürlich schon, dass Figuren wie eine Rotationsparaboloid mit kreis- oder ellipsenförmigen Querschnitt (Höhenlinien) eine eindeutige Minimumstelle besitzen. Aber auch andere Figuren, wie eine flache Schale oder die Bananenschale, besitzen ein Minimum. Die Oberflächenstruktur eines Geländeabschnitts hat möglicherweise ein globales Minimum, aber auch zahlreiche lokale Minima.
Als zu minimierende Funktionale nehmen wir die folgenden drei Funktionen. Ihre Entstehung, Bedeutung und Eigenschaften werden später noch eingehender untersucht. Zunächst wollen wir sie in unterschiedlichen Beispielen anwenden.
Die Funktionale leiten wir aus dem LGS ab.
Ihre einfachste direkt mit (1.1) verknüpfte Form ist

$$\begin{aligned} Q(x) &= \frac{1}{2}x^T Ax - x^T b \\ &= \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}x_i x_j - \sum_{i=1}^{n} b_i x_i. \end{aligned} \tag{1.2}$$

Es gilt $Q(x^*) = \frac{1}{2}x^{*T}(Ax^*) - x^{*T}b = \frac{1}{2}x^{*T}b - x^{*T}b = -\frac{1}{2}x^{*T}b$.
Transformiert man das LGS (1.1) durch Multiplikation mit A^T, ohne dabei die Lösung zu verändern, auf das so genannte Normalgleichungssystem (Gaußsche Normalgleichungen)

$$A^T Ax \;=\; A^T b, \;\; \text{d. h.} \;\; Bx \;=\; c, \tag{1.3}$$

mit der symmetrischen und positiv definiten Koeffizientenmatrix (spd) $B = A^T A$, so folgt sofort das Funktional

$$\begin{aligned} R(x) &= \frac{1}{2}x^T Bx - x^T c \\ &= \frac{1}{2}x^T A^T Ax - x^T A^T b \;=\; \frac{1}{2}(Ax)^T Ax - (Ax)^T b \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j,k=1}^{n} a_{ij}a_{ik}x_j x_k - \sum_{i=1}^{n}\sum_{j=1}^{n} b_i a_{ij} x_j. \end{aligned} \tag{1.4}$$

An der Lösung ist $R(x^*) = \frac{1}{2}(Ax^*)^T Ax^* - (Ax^*)^T b = -\frac{1}{2}b^T b \leq 0$.
Die Matrix B hat im Allgemeinen eine schlechtere Kondition als A, so dass sich dies auf das Iterationsverfahren, dazu gehören die Abstiegsverfahren, ungünstig auswirken wird und somit dann mehr Schritte erforderlich sind.

Bei Implementierungen wird die Matrix B wegen des Aufwands von $2n^3$ Operationen nicht explizit ermittelt, sondern im Verfahren hat man anstelle einer Matrix-Vektor-Multiplikation in der Hauptschleife nun zwei.
Eine dritte mit $R(x)$ verwandte Form ist

$$f(x) \;=\; \sum_{i=1}^{n} [f_i(x)]^2 \;\geq\; 0, \tag{1.5}$$

wobei $f_i(x)$ die i-te Zeile des LGS (1.1) darstellt, also

$$f_i(x) \;=\; (b - Ax)_i \;=\; b_i - \sum_{j=1}^{n} a_{ij}x_j.$$

Man rechnet einfach nach, dass die Beziehungen

$$f(x) \;=\; 2R(x) + b^T b \tag{1.6}$$

und somit $f(x^*) = 2R(x^*) + b^T b = 0$ gelten. Die Funktionale $f(x)$ und $R(x)$ weisen deshalb keine qualitativen Unterschiede auf.

Im allgemeinen Fall können die Summanden $f_i(x)$ auch nichtlineare Terme in x sein, so dass man nicht unbedingt von einem LGS wie hier ausgehen muss.

Nun berechnen wir die genannten Funktionale für einige LGS im $\mathbb{R}^2$, wo wir sie auch grafisch darstellen können. Wir diskutieren die zugehörige Abstiegssituation und ihre eventuellen Besonderheiten. Dabei charakterisieren wir im Rahmen des Iterationsprozesses für einen Schritt die Wahl von möglichen Abstiegsrichtungen bzw. generell Suchrichtungen hin zur Lösung x^*. Der Grenzvektor x^* ist unter den für Abstiegsverfahren üblichen Voraussetzungen die Minimumstelle des Funktionals, anders jedoch bei abweichenden Bedingungen.
Die Abbildungen zu den folgenden Beispielen sind in Maple erzeugt worden.

Beispiel 1.1
Sei $Ax = b,\ A = I$ Einheitsmatrix und $b = 0$.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Die Funktionale sind

$$Q(x) \ = \ \frac{1}{2}(x_1^2 + x_2^2), \quad R(x) \ = \ \frac{1}{2}(x_1^2 + x_2^2), \quad f(x) \ = \ x_1^2 + x_2^2.$$

Am gemeinsamen eindeutigen Minimum bei $x^* = 0$ ist $Q(x^*) = R(x^*) = f(x^*) = 0$.

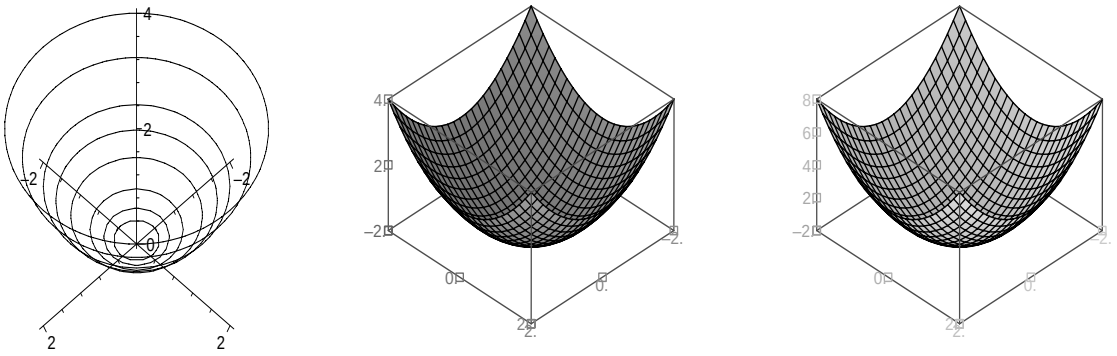


Abb. 1.1 Dateien *abst_10.ps*, *abst_11.ps*
3D-Portraits von $Q(x) = R(x) = \frac{1}{2}(x_1^2 + x_2^2)$ und $f(x) = x_1^2 + x_2^2$ (v.l.n.r.)

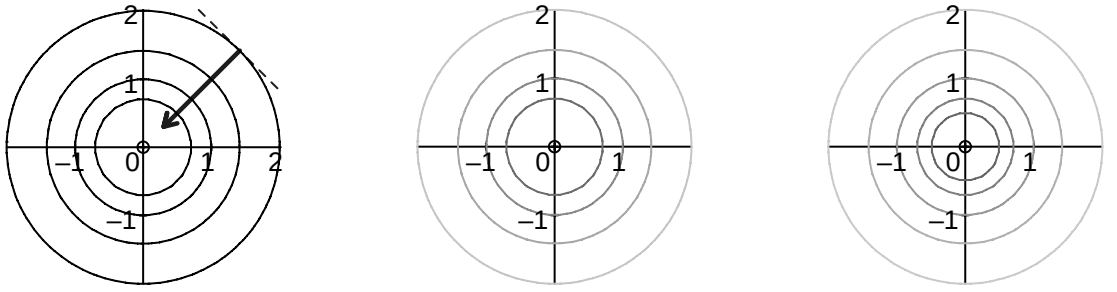


Abb. 1.2 Datei *abst_12.ps*
Höhenlinienbild von $Q(x),\ R(x)$ mit `contours=[0,0.25,0.5,1,2]` und $f(x)$ mit `contours=[0,0.25,0.5,1,2,4]`

Als Suchrichtung und Abstiegsrichtung in einem Schritt bietet sich die Richtung des steilsten Abstiegs an. Sie ist der negative Gradient des Funktionals

$$-\nabla Q(x) \ = \ -\Big(\frac{\partial Q(x)}{\partial x_1}, \frac{\partial Q(x)}{\partial x_2}, ..., \frac{\partial Q(x)}{\partial x_n}\Big)^T \tag{1.7}$$

und orthogonal zur Höhenlinie $Q(x)$=const, genauer gesagt: $\nabla Q(x) \perp$ zur Tangente an die Höhenlinie. Im Höhenlinienbild von $Q(x)$ in Abbildung 1.2 (linke Figur) ist eine solche Abstiegsrichtung eingetragen. Im Fall kreisförmiger Konturen kann ein Schritt in dieser Richtung genau zur Minimumstelle führen.

Beispiel 1.2
Sei $Ax = b,\ A = A^T > 0$, Diagonalmatrix und $b = 0$.

$$\begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Die Funktionale sind

$$\begin{aligned} Q(x) &= \frac{1}{2}x_1^2 + \frac{5}{2}x_2^2, \\ R(x) &= \frac{1}{2}x_1^2 + \frac{25}{2}x_2^2, \\ f(x) &= x_1^2 + 25x_2^2. \end{aligned}$$

Am gemeinsamen eindeutigen Minimum an der Stelle $x^* = 0$ gilt $Q(x^*) = R(x^*) = f(x^*) = 0$.

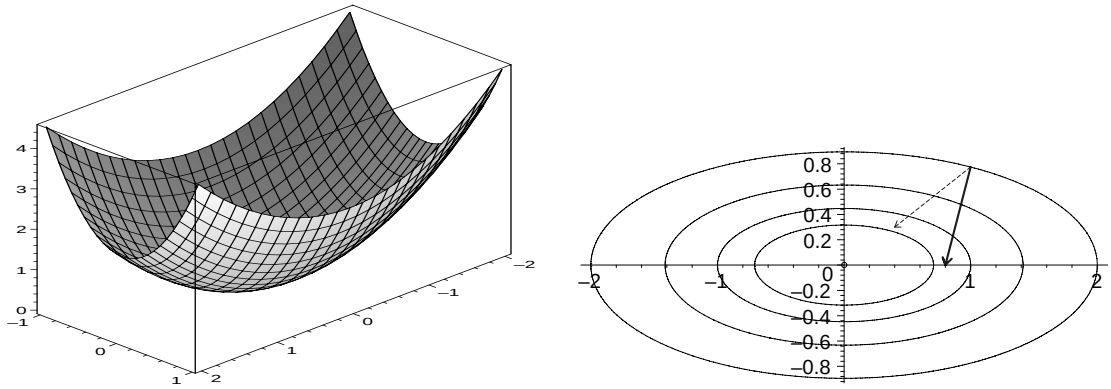


Abb. 1.3 Dateien *abst_211.ps*, *abst_212.ps*
3D- und Höhenlinienbild von $Q(x) = \frac{1}{2}x_1^2 + \frac{5}{2}x_2^2$ mit `contours=[0,0.25,0.5,1,2]`

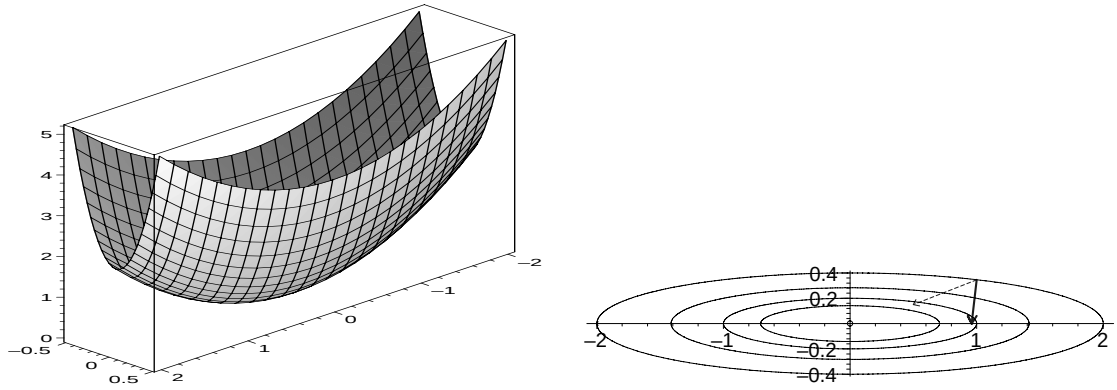


Abb. 1.4 Dateien *abst_221.ps*, *abst_222.ps*
3D- und Höhenlinienbild von $R(x) = \frac{1}{2}x_1^2 + \frac{25}{2}x_2^2$ mit `contours=[0,0.25,0.5,1,2]`

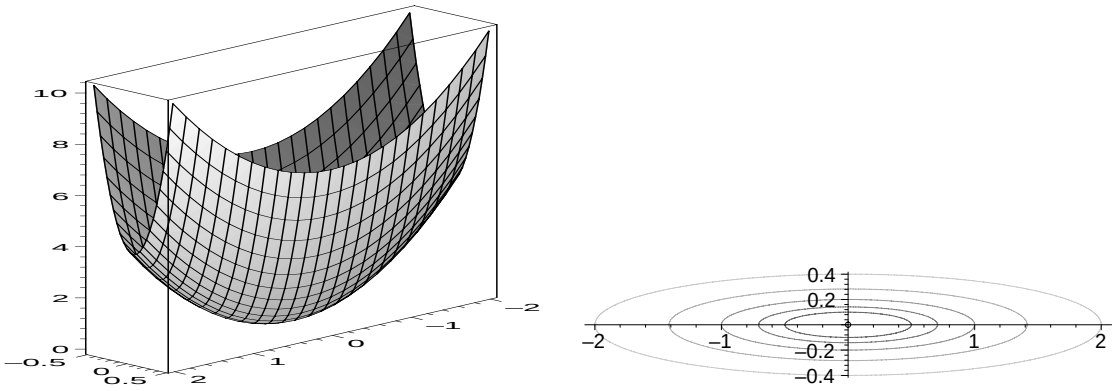


Abb. 1.5 Dateien *abst_231.ps*, *abst_232.ps*
3D- und Höhenlinienbild von $f(x) = x_1^2 + 25x_2^2$ mit `contours=[0,0.25,0.5,1,2,4]`

Als Suchrichtung und Abstiegsrichtung in einem Schritt bietet sich erneut die Richtung des steilsten Abstiegs an.
Je lang gestreckter die Konturen (Ellipsen) sind, desto mehr Suchschritte sind für die Annäherung an x^* vorzusehen. Damit ist das Funktional $R(x)$ gegenüber $Q(x)$ im Nachteil.
Nicht orthogonal zu den Höhenlinien $Q(x)$=const bzw. $R(x)$=const scheint es jedoch günstigere Suchrichtungen zu geben, die als gestrichelte Vektoren in den Abbildungen 1.3 und 1.4 eingezeichnet sind.

Beispiel 1.3
Sei $Ax = b,\ A = A^T > 0$.

$$\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad x^* = \begin{pmatrix} 1/5 \\ 3/5 \end{pmatrix}.$$

Die Funktionale sind

$$\begin{aligned} Q(x) &= x_1^2 + x_1x_2 + \frac{3}{2}x_2^2 - x_1 - 2x_2, \quad Q(x^*) = -\frac{7}{10}, \\ R(x) &= \frac{5}{2}x_1^2 + 5x_1x_2 + 5x_2^2 - 4x_1 - 7x_2, \quad R(x^*) = -\frac{5}{2}, \\ f(x) &= 5x_1^2 + 10x_1x_2 + 10x_2^2 - 8x_1 - 14x_2 + 5. \end{aligned}$$

Das gemeinsame eindeutige Minimum ist an der Stelle $x^* = (\frac{1}{5}, \frac{3}{5})^T$.

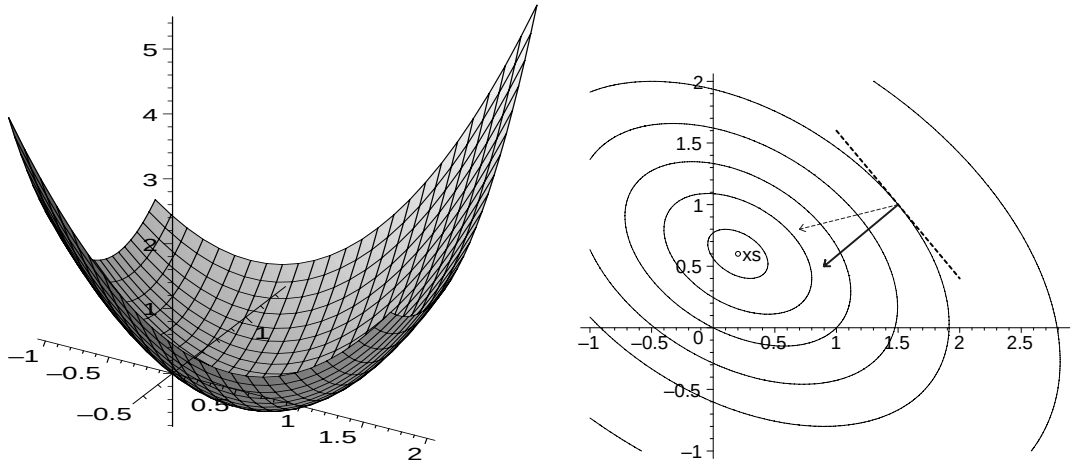


Abb. 1.6 Dateien *abst_311.ps*, *abst_312.ps*
3D- und Höhenlinienbild von $Q(x) = x_1^2 + x_1x_2 + \frac{3}{2}x_2^2 - x_1 - 2x_2$ mit
`contours=[5,1.75,0.7,0,-0.4,-0.65,-0.7]`

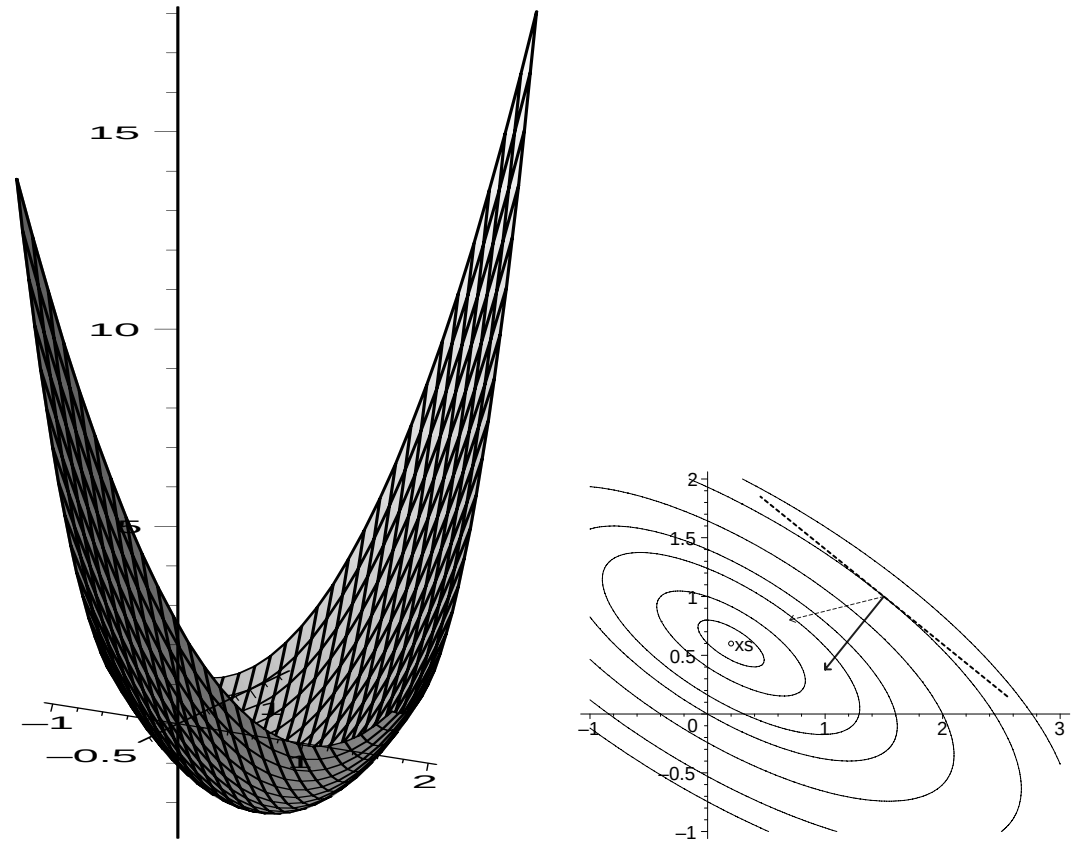


Abb. 1.7 Dateien *abst_321.ps*, *abst_322.ps*
3D- und Höhenlinienbild von $R(x) = \frac{5}{2}x_1^2 + 5x_1x_2 + 5x_2^2 - 4x_1 - 7x_2$ mit
`contours=[8,5.125,2,0,-1,-2,-2.4,-2.5]`

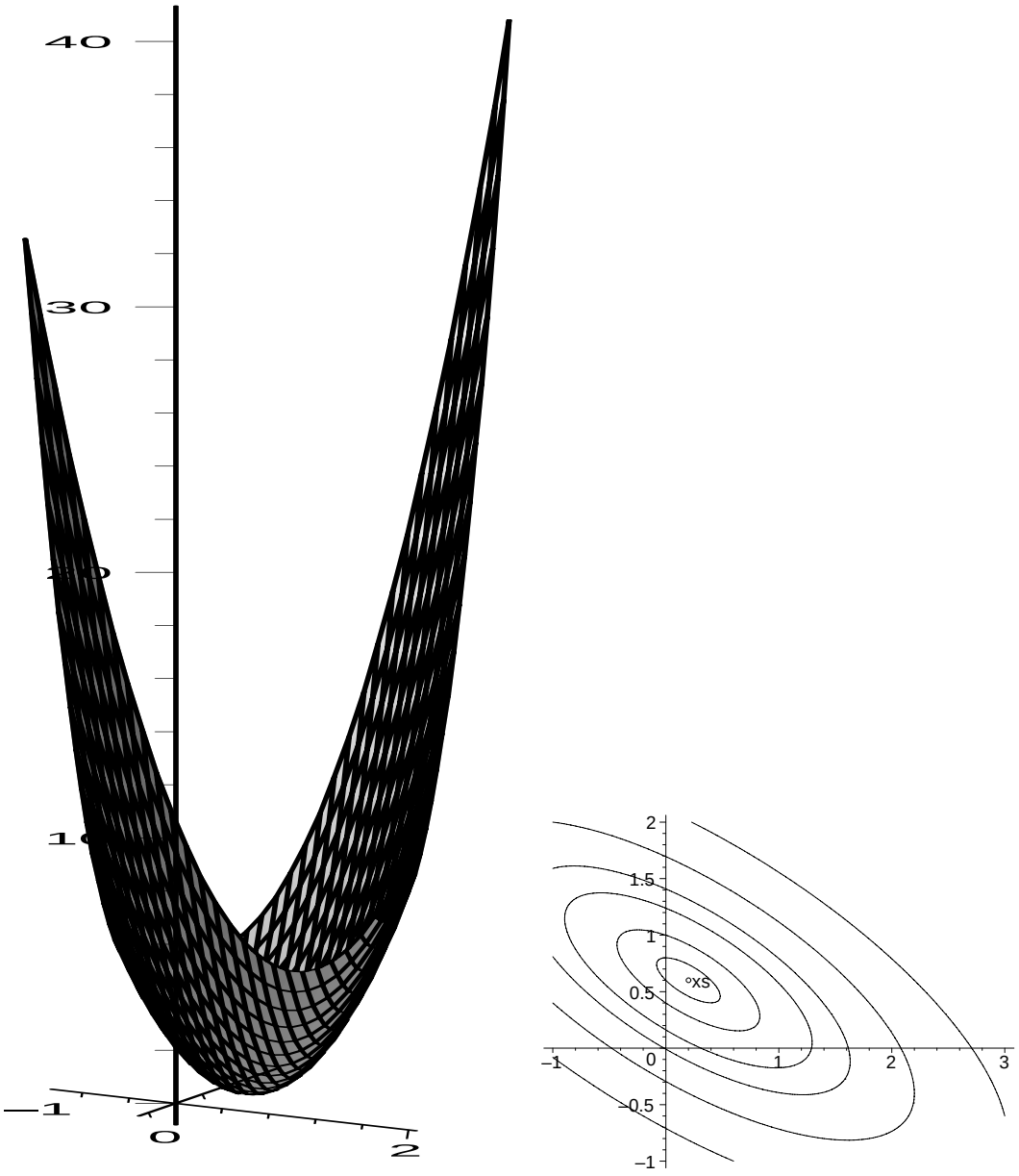


Abb. 1.8 Dateien *abst_331.ps*, *abst_332.ps*
3D- und Höhenlinienbild von $f(x) = 5x_1^2 + 10x_1x_2 + 10x_2^2 - 8x_1 - 14x_2 + 5$ mit
`contours=[20,10,5.125,3,1,0.2,0]`

Abstiegsrichtungen ergeben sich sowohl aus der Richtung des steilsten Abstiegs als auch durch "benachbarte" und eventuell günstigere Richtungen.
Die Kondition des LGS verschlechtert sich durch die Multiplikation mit A^T, was sich auch auf das zugehörige Funktional $R(x)$ auswirkt.
Mit Ausnahme der (künstlichen) Wahl der Suchrichtung entlang einer Tangente an eine Höhenlinie des Funktionals besteht im Fall $A = A^T > 0$ nicht die Gefahr, dass das Verfahren sich nicht schrittweise zur Minimumstelle hinbewegt.

Beispiel 1.4
Sei $Ax = b$, A diagonal (damit symmetrisch) und indefinit.

$$\begin{pmatrix} 1 & 0 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Die Funktionale sind

$$\begin{aligned} Q(x) &= \frac{1}{2}x_1^2 - 2x_2^2, \quad Q(x^*) = 0, \\ R(x) &= \frac{1}{2}x_1^2 + 8x_2^2, \quad R(x^*) = 0, \\ f(x) &= x_1^2 + 16x_2^2. \end{aligned}$$

Die Stelle x^* ist das eindeutige Minimum des Funktionals $R(x)$, aber für das Funktional $Q(x)$ ein Sattelpunkt, durch den die Null-Höhenlinien (zwei sich schneidende Geraden) gehen. Damit entsteht eine neue Situation.

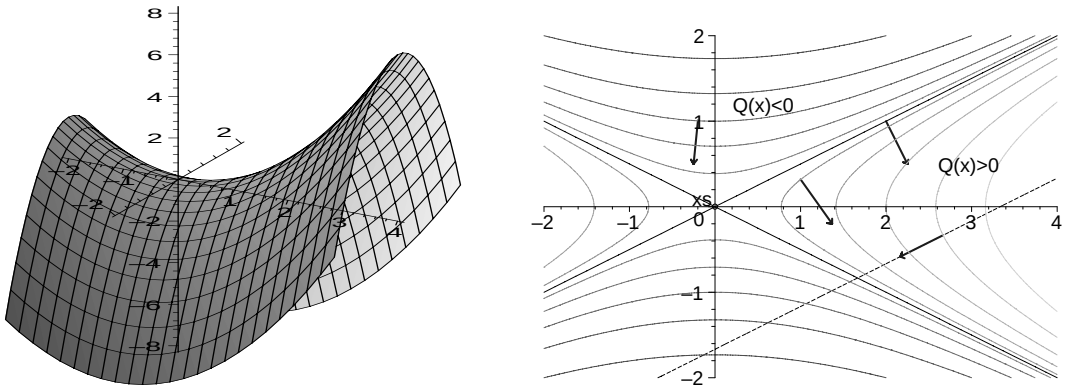


Abb. 1.9 Dateien *abst_411.ps*, *abst_412.ps*
3D- und Höhenlinienbild von $Q(x) = \frac{1}{2}x_1^2 - 2x_2^2$ mit
`contours=[-6,-3.5,-2,-1,-0.3,0.3,1,2,3.333,5]`

Der Sattelpunkt von $Q(x)$ ist die gesuchte Lösung x^*. Will man aber zu diesem gelangen, kommen Abstiegs- und Anstiegsrichtungen in Betracht. Einige sind im Höhenlinienbild der Abbildung 1.9 eingezeichnet worden. Dabei besteht durchaus die Gefahr, dass bei bestimmten Suchrichtungen das Verfahren abbricht oder divergiert. Ersteres merkt man daran, dass in Suchrichtung das Funktional $Q(x)$ einen streng monotonen Verlauf hat und sonst übliche Berechnungsvorschriften dann nicht ausführbar sind. Auf der gestrichelten Geraden in der rechten Figur der genannten Abbildung liegen solche problematischen Suchrichtungen.

Die Wahl von unterschiedlichen Abstiegsrichtungen ist generell nur noch für das Funktional $R(x)$ möglich.

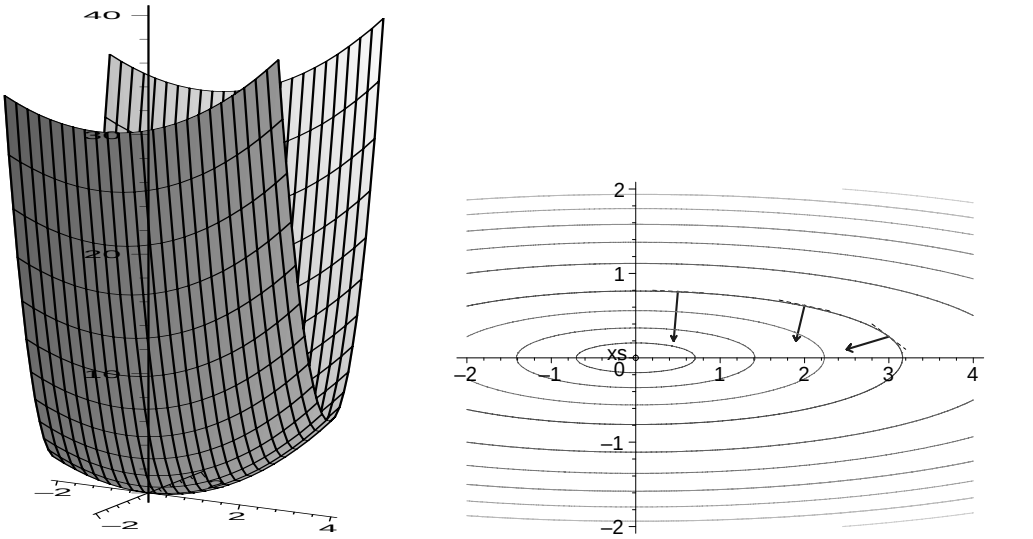


Abb. 1.10 Dateien *abst_421.ps*, *abst_422.ps*
3D- und Höhenlinienbild von $R(x) = \frac{1}{2}x_1^2 + 8x_2^2$ mit
`contours=[0.25,1,2.5,5,10,20,30]`

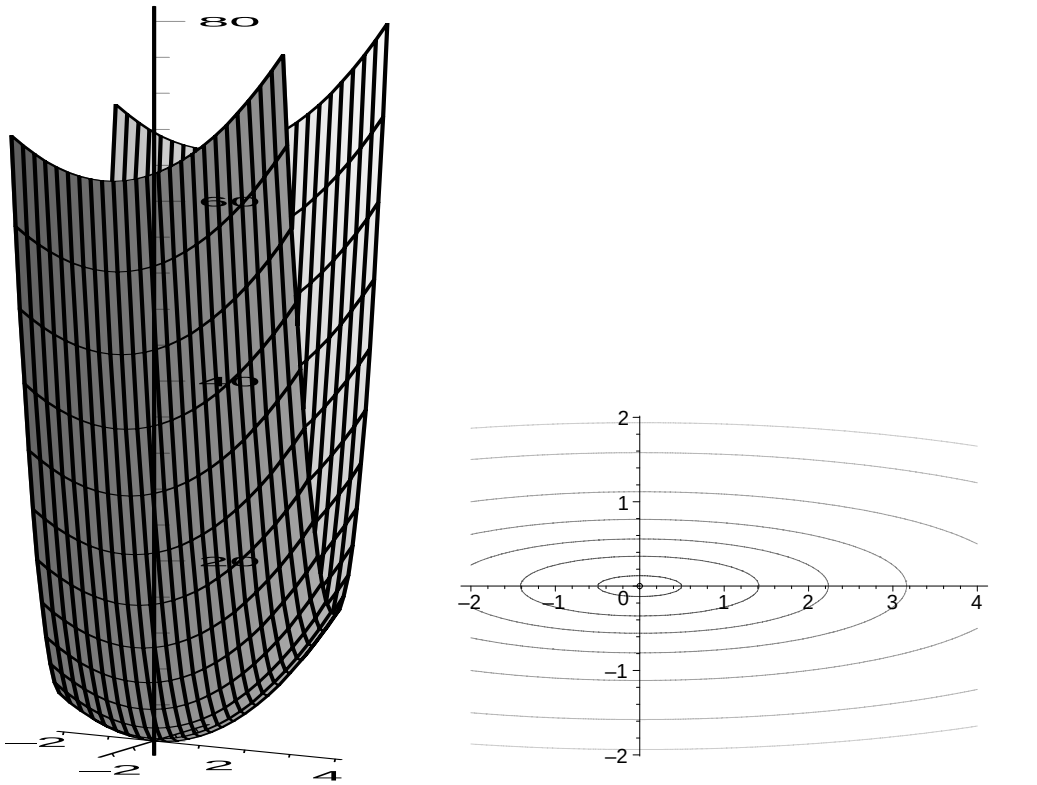


Abb. 1.11 Dateien *abst_431.ps*, *abst_432.ps*
3D- und Höhenlinienbild von $f(x) = x_1^2 + 16x_2^2$ mit
`contours=[0.25,2,5,10,20,40,60]`

Beispiel 1.5
Sei $Ax = b,\ A = A^T$ und indefinit.

$$\begin{pmatrix} 2 & 1 \\ 1 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \quad x^* = \begin{pmatrix} 10/7 \\ 1/7 \end{pmatrix}.$$

Die Funktionale sind

$$\begin{aligned} Q(x) &= x_1^2 + x_1x_2 - \frac{3}{2}x_2^2 - 3x_1 - x_2,\ Q(x^*) = -\frac{31}{14}, \\ R(x) &= \frac{5}{2}x_1^2 - x_1x_2 + 5x_2^2 - 7x_1,\ R(x^*) = -5, \\ f(x) &= 5x_1^2 - 2x_1x_2 + 10x_2^2 - 14x_1 + 10. \end{aligned}$$

Die Stelle x^* ist das eindeutige Minimum des Funktionals $R(x)$, aber für das Funktional $Q(x)$ wiederum ein Sattelpunkt.

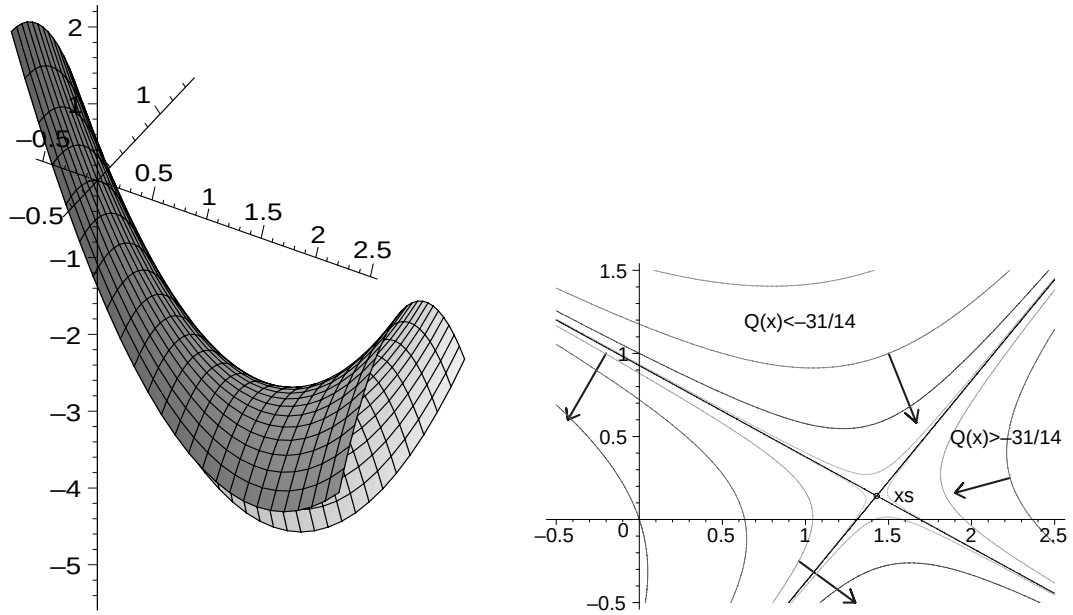


Abb. 1.12 Dateien *abst_511.ps*, *abst_512.ps*
3D- und Höhenlinienbild von $Q(x) = x_1^2 + x_1x_2 - \frac{3}{2}x_2^2 - 3x_1 - x_2$ mit
`contours=[-2.5,0,-1.5,-3.25,-5,-2.0417,-2.2431,-2.2095]`

Es treffen die gleichen Bemerkungen wie in Beispiel 1.4 bezüglich Minimum und Sattelpunkt sowie Suchrichtungen zu.

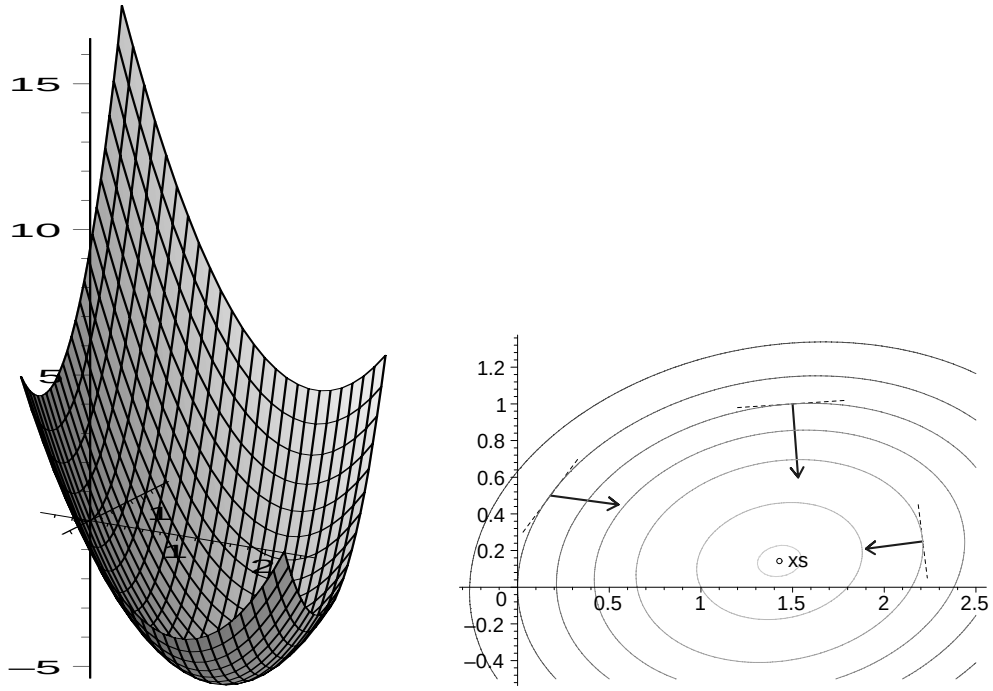


Abb. 1.13 Dateien *abst_521.ps*, *abst_522.ps*
3D- und Höhenlinienbild von $R(x) = \frac{5}{2}x_1^2 - x_1x_2 + 5x_2^2 - 7x_1$ mit
`contours=[2,0,-1.375,-2.5,-3.5,-4.5,-4.9641]`

Nun kommen wir zum nicht symmetrischen Fall, wo sich die Eigenschaften der Funktionale $Q(x)$ und $R(x)$ in Bezug auf die Lösung unterscheiden.
Selbst wenn das Funktional $Q(x)$ eine eindeutige Minimumstelle besitzt, zu welcher natürlich Abstiegsverfahren tendieren können, ist dies nicht die Lösung des LGS.
Dazu nehmen wir den Gradienten $\nabla Q(x)$ des Funktionals, der an der Minimumstelle verschwindet.
Es gilt

$$\nabla Q(x) \;=\; \frac{1}{2}(A + A^T)x - b \tag{1.8}$$

und aus $\nabla Q(x) = 0$ folgt die Lösung

$$z \;=\; \left[\frac{1}{2}(A + A^T)\right]^{-1} b, \tag{1.9}$$

die im Allgemeinen nicht mit x^* übereinstimmt. Wir greifen diesen Sachverhalt später noch einmal bei der Behandlung der Verfahren auf.
Das Funktional $R(x)$ bereitet, abgesehen von der schlechteren Kondition der Matrix A^TA, keine weiteren Probleme.

Beispiel 1.6
Sei $A \neq A^T,\ A > 0$.

$$\begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \quad x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Die Matrix A ist positiv definit, denn für $x \neq 0$ ist

$$x^T Ax = 2x_1^2 + x_1x_2 + 3x_2^2 = \left(x_1 + \frac{1}{2}x_2\right)^2 + x_1^2 + \frac{11}{4}x_2^2 > 0.$$

Damit hat $Q(x)$ ein eindeutiges Minimum und seine Stelle ist $z = (\frac{30}{23}, \frac{18}{23})^T$.
Die Funktionale sind

$$\begin{aligned} Q(x) &= x_1^2 + \frac{1}{2}x_1x_2 + \frac{3}{2}x_2^2 - 3x_1 - 3x_2,\ Q(z) = -\frac{72}{23} = -3.130...,\ Q(x^*) = -3, \\ R(x) &= 2x_1^2 + 2x_1x_2 + 5x_2^2 - 6x_1 - 12x_2,\ R(x^*) = -9,\ R(z) = -8.710..., \\ f(x) &= 4x_1^2 + 4x_1x_2 + 10x_2^2 - 12x_1 - 24x_2 + 18. \end{aligned}$$

Die Stelle x^* ist das eindeutige Minimum des Funktionals $R(x)$, aber für das Funktional $Q(x)$ haben wir ein Minimum bei $z \neq x^*$.

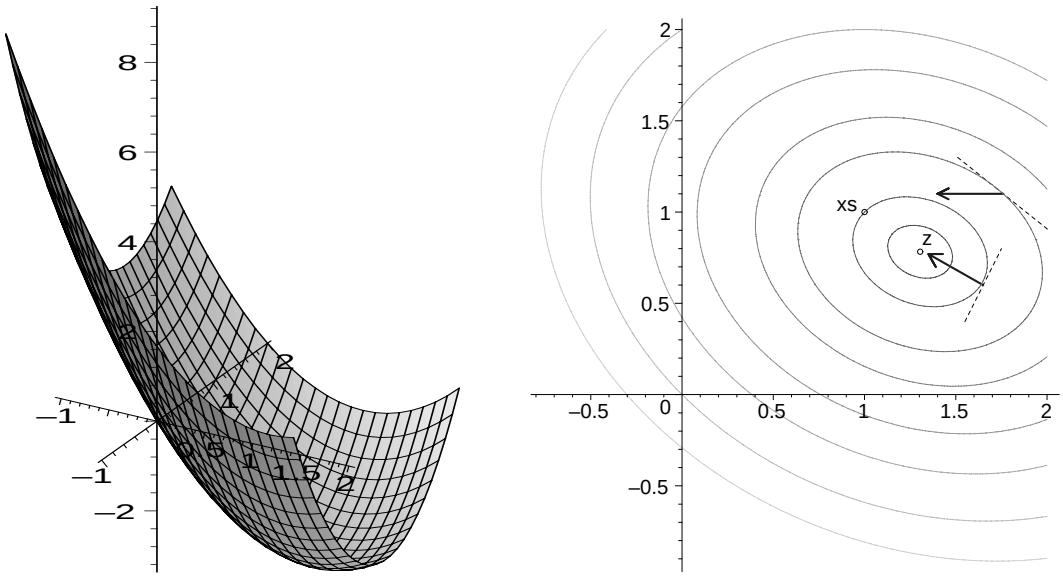


Abb. 1.14 Dateien *abst_611.ps*, *abst_612.ps*
3D- und Höhenlinienbild von $Q(x) = x_1^2 + \frac{1}{2}x_1x_2 + \frac{3}{2}x_2^2 - 3x_1 - 3x_2$ mit
`contours=[-3.1,-3,-2.7,-2.35,-1.7,-1,0,1]`

Wie verhalten sich Suchrichtungen, die aus dem Funktional $Q(x)$ abgeleitet werden? Wenn diese sich zu sehr dem Abstiegsverhalten von $Q(x)$ anpassen, wird das Verfahren möglicherweise zum Punkt z führen. Das nutzt uns jedoch wenig.
Geht der Prozess jedoch "wegwärts", so bleibt die Frage, ob sich die Folge der Iterierten nun der Lösung x^* nähert oder auch nicht.

Als Suchrichtung für die Minimumstelle von $R(x)$ bieten sich unterschiedliche Abstiegsrichtungen an.
Alles das ist ein Grund, die Situation und das Verhalten verschiedener Abstiegsverfahren im Weiteren genauer zu untersuchen.

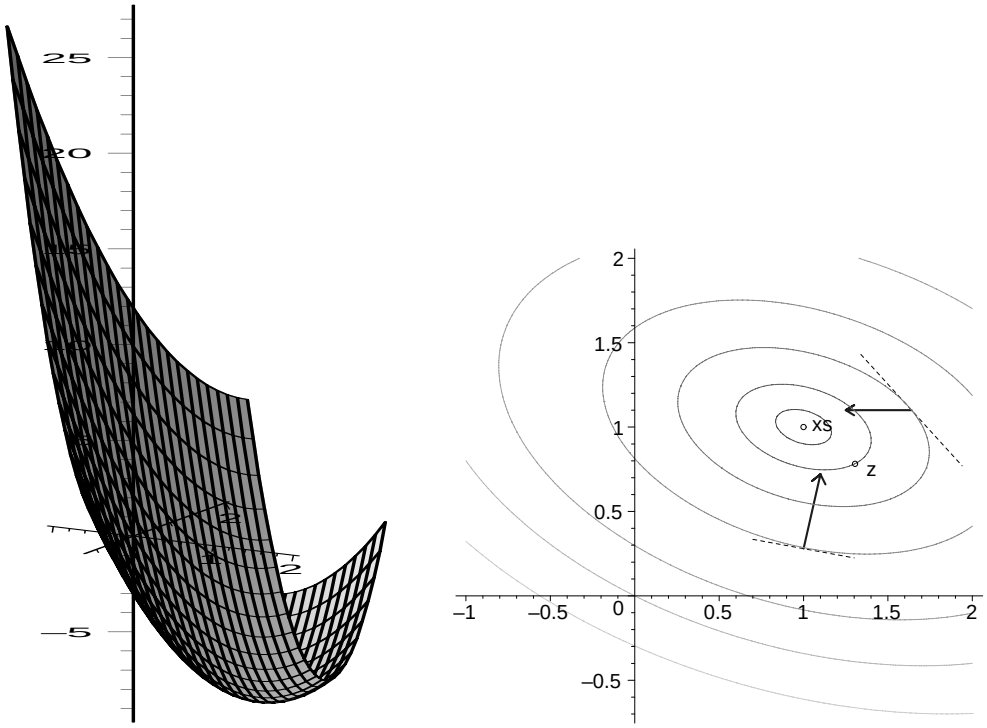


Abb. 1.15 Dateien *abst_621.ps*, *abst_622.ps*
3D- und Höhenlinienbild von $R(x) = 2x_1^2 + 2x_1x_2 + 5x_2^2 - 6x_1 - 12x_2$ mit
`contours=[-9,-8.95,-8.71,-8,-6.45,-3.13,0,4]`

Machen wir die Situation noch etwas komplizierter, indem wir eine nicht symmetrische und indefinite Matrix betrachten.

Beispiel 1.7
Sei $A = (a_{ij}) \neq A^T$ und A indefinit.

$$\begin{pmatrix} 0.780 & 0.563 \\ 0.913 & 0.659 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.217 \\ 0.254 \end{pmatrix} = b, \quad x^* = \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{1.10}$$

Der Wert der Determinante ist $\det(A) = 10^{-6}$ und die Inverse lautet

$$A^{-1} = \begin{pmatrix} 659\,000 & -563\,000 \\ -913\,000 & 780\,000 \end{pmatrix}.$$

Die Eigenwerte der Matrix A sind

$$\lambda_{1,2}(A) = \frac{1\,439 \pm \sqrt{2\,070\,717}}{2\,000} = 1.438\,999\,305\,072..., \; 0.000\,000\,694\,927... > 0.$$

Obwohl beide Eigenwerte positiv sind, ist die Matrix indefinit. Für Vektoren mit positiven Komponenten hat die quadratische Form $x^T Ax$ nur positive Werte. Aber es gilt zum Beispiel für $x = (1, -0.9)^T \neq 0$ die Beziehung

$$x^T Ax \;=\; 0.780x_1^2 + 1.476x_1x_2 + 0.659x_2^2 \;=\; -0.014\,61 < 0.$$

Somit ist die Matrix A indefinit.
An der Determinante und den Eigenwerten der Matrix erkennt man ihre schlechte Kondition. Sie wird noch schlechter für die symmetrische Matrix

$$B = A^T A = \begin{pmatrix} 1.441\,969 & 1.040\,807 \\ 1.040\,807 & 0.751\,250 \end{pmatrix} \quad \text{mit} \quad \det(B) = 10^{-12}, \tag{1.11}$$

$$\lambda_{1,2}(B) = 2.193\,218\,999\,999\,544..., \; 4.559\,508\,193\,209... \cdot 10^{-13} > 0$$

und den zugehörigen orthogonalen Eigenvektoren

$$v^{(1)} = (0.810\,843\,355\,393\,196..., 0.585\,263\,233\,950\,328...)^T, \quad v^{(2)} = (-v_2^{(1)}, v_1^{(1)})^T. \tag{1.12}$$

Das Funktional $R(x)$ besitzt sein eindeutiges Minimum bei $x^* = (1, -1)^T$ und dort ist $R(x^*) = -0.055\,802\,5$.
Das Funktional $Q(x)$ hat einen Sattelpunkt bei

$$z = \Big(\frac{44\,449}{30\,624}, -\frac{6\,329}{5\,104}\Big)^T = (1.451\,443\,312\,434\,691, -1.240\,007\,836\,990\,595)^T$$

(Dezimalzahlen sind gerundet), an welchem

$$Q(z) = -\frac{37}{61\,248\,000} = -0.604\,101\,358\,411\,702\,9 \cdot 10^{-6}$$

ist. Außerdem sind

$$Q(x^*) = 0.018\,5 > Q(z) \;\text{ und }\; R(z) = 0.952\,851\,707\,750\,301 \cdot 10^{-6} > R(x^*).$$

Die Funktionale sind

$$\begin{aligned}
Q(x) &= 0.5(0.780x_1 + 0.913x_2)x_1 + 0.5(0.563x_1 + 0.659x_2)x_2 - 0.217x_1 - 0.254x_2 \\
&= 0.390x_1^2 + 0.738x_1x_2 + 0.3295x_2^2 - 0.217x_1 - 0.254x_2, \\
R(x) &= 0.5(1.441\,969x_1 + 1.040\,807x_2)x_1 + 0.5(1.040\,807x_1 + 0.751\,250x_2)x_2 \\
&\quad -0.401\,162x_1 - 0.289\,557x_2 \\
&= 0.720\,984\,5x_1^2 + 1.040\,807x_1x_2 + 0.375\,625x_2^2 - 0.401\,162x_1 - 0.289\,557x_2, \\
f(x) &= (1.441\,969x_1 + 1.040\,807x_2)x_1 + (1.040\,807x_1 + 0.751\,250x_2)x_2 \\
&\quad -0.802\,324x_1 - 0.579\,114x_2 + 0.111\,605.
\end{aligned}$$

Das Minimum von $R(x)$ liegt in einem sehr lang gestreckten Tal auf einer Talkurve in einer fast ebenen Talsohle. Seine Höhenlinien sind in der Abbildung 1.17 fast parallel. Unweit von $x^* =$ `xs` befindet sich der Sattelpunkt z von $Q(x)$. Am Sattelpunkt sind beide Werte $Q(z)$ und $R(z)$ sehr nahe Null, so dass im betrachteten Bereich beide Funktionale fast ein gemeinsames Stück der Nullhöhenlinie besitzen. Das alles macht die Darstellung vieler Sachverhalte in einer Grafik unübersichtlich.

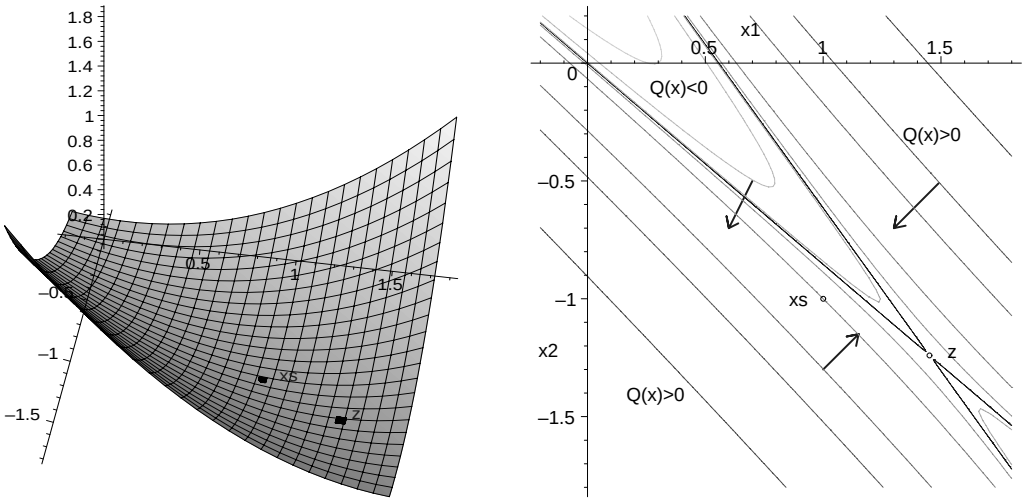


Abb. 1.16 Dateien *abst_711.ps*, *abst_712.ps*
3D- und Höhenlinienbild von
$Q(x) = 0.390x_1^2 + 0.738x_1x_2 + 0.3295x_2^2 - 0.217x_1 - 0.254x_2$ mit
`contours=[0.5,0.2,0.1,0.0185,0.003,-0.0000006041,-0.001,-0.01,-0.03]`

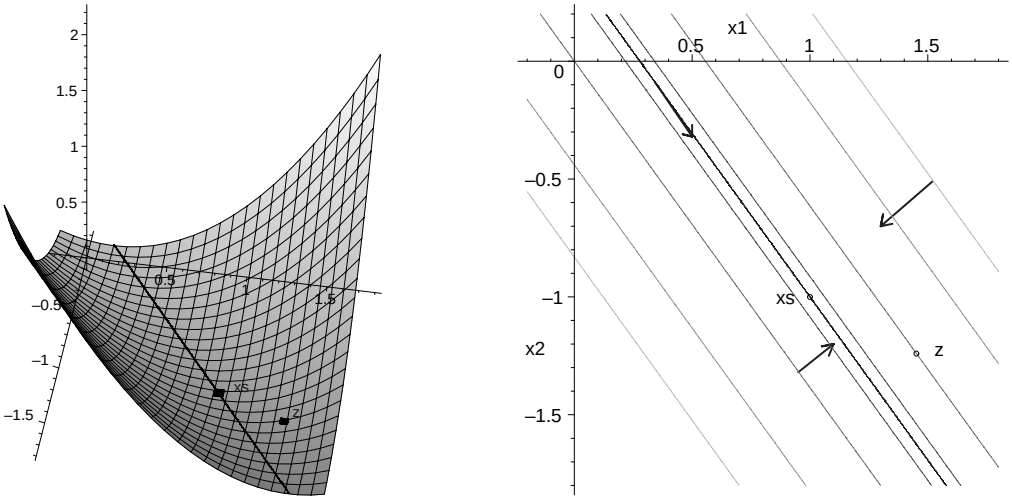


Abb. 1.17 Dateien *abst_721.ps*, *abst_722.ps*
3D-Bild mit Talkurve zur Tallinie $x_2 = v_2^{(2)}/v_1^{(2)}\,(x_1-1)-1$ und Höhenlinienbild von
$R(x) = 0.720\,984\,5x_1^2 + 1.040\,807x_1x_2 + 0.375\,625x_2^2 - 0.401\,162x_1 - 0.289\,557x_2$ mit
`contours=[-0.0558025,-0.0558,-0.053,0,0.2,0.5]`

Gleichzeitig ist x^* der Schnittpunkt zweier Geraden

$$\begin{aligned} x_2 &= g_1(x_1) \\ &= \frac{b_1 - a_{11}x_1}{a_{12}} = \frac{1}{0.563}(0.217 - 0.780x_1), \quad (x_1^*, x_2^*) = (1, -1) \\ &= -\frac{a_{11}}{a_{12}}(x_1 - 1) - 1 \end{aligned}$$

und

$$\begin{aligned} x_2 &= g_2(x_1) \\ &= \frac{b_2 - a_{21}x_1}{a_{22}} = \frac{1}{0.659}(0.254 - 0.913x_1) \\ &= -\frac{a_{21}}{a_{22}}(x_1 - 1) - 1 \end{aligned} \tag{1.13}$$

entsprechend den beiden Gleichungen des LGS (1.10). In einer grafischen Darstellung sind die Geraden jedoch kaum zu unterscheiden, damit auch nicht von der dazwischen liegenden Tallinie

$$x_2 = t(x_1) = \frac{v_2^{(2)}}{v_1^{(2)}}(x_1 - 1) - 1. \tag{1.14}$$

Analog kann man zum Normalgleichungssystem $Bx = c,\ B = (b_{ij}),\ b_{12} = b_{21}$, zwei Geradengleichungen aufstellen.

$$\begin{aligned} x_2 &= h_1(x_1) \\ &= \frac{c_1 - b_{11}x_1}{b_{12}} = \frac{1}{1.040\,807}(0.401\,162 - 1.441\,969x_1) \\ &= -\frac{b_{11}}{b_{12}}(x_1 - 1) - 1, \\ x_2 &= h_2(x_1) \\ &= \frac{c_2 - b_{21}x_1}{b_{22}} = \frac{1}{0.751\,250}(0.289\,557 - 1.040\,807x_1) \\ &= -\frac{b_{21}}{b_{22}}(x_1 - 1) - 1. \end{aligned} \tag{1.15}$$

Die Geraden $g_i(x_1)$ "schließen" die Geraden $h_i(x_1)$ ein und die Tallinie $t(x_1)$ "liegt in der Mitte".
Im Höhenlinienbild der quadratischen Form $x^TBx = x^TA^TAx$ können wegen der geringen Auflösung keine Ellipsen gezeichnet werden. Man erkennt nur gegenüber liegende fast parallele Ellipsenabschnitte.

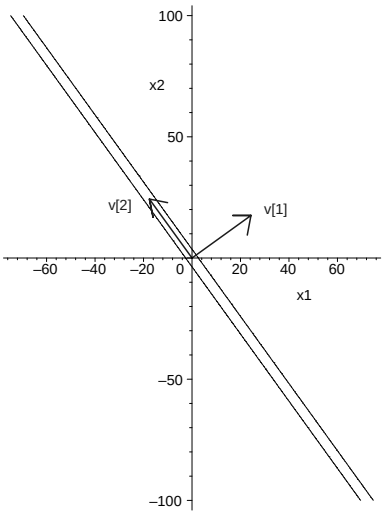


Abb. 1.18
Datei *abst_79.ps*
Höhenlinie $x^T Bx = 10$
und skalierte EV (1.12) von B
$cv^{(1)},\ cv^{(2)}$ mit $c = 30$

Wir betrachten auch das Funktional

$$\begin{aligned} f(x) &= 2R(x) + b^T b \\ &= [r_1(x)]^2 + [r_2(x)]^2 \\ &= (0.780x_1 + 0.563x_2 - 0.217)^2 + (0.913x_1 + 0.659x_2 - 0.254)^2 \\ &= 0.720\,9845x_1^2 + 1.040\,807x_1x_2 + 0.375\,625x_2^2 - 0.401\,162x_1 - 0.289\,557x_2 \\ &\quad -0.802\,324x_1 - 0.579\,114x_2 + 0.111\,605. \end{aligned}$$

Wir wissen, dass die Lösung $x^* = (1, -1)^T$ seine Minimumstelle ist.
Diese liegt wie bei $R(x)$ ebenfalls in einem flachen lang gestreckten Tal.
Wir berechnen einige Funktionswerte von $f(x)$ entlang der Geraden $g_1(x_1),\ g_2(x_1)$ und $t(x_1)$ als Hinweis auf die flache Talsohle.
Es gilt

$$\begin{aligned} f(x_1, x_2)|_{x_2=g_1(x_1)} &= [r_1(x_1, g_1(x_1))]^2 + [r_2(x_1, g_1(x_1))]^2 \\ &= 0 + [b_2 - (a_{21}x_1 + a_{22}g_1(x_1))]^2, \quad b_2 = a_{21} - a_{22} \\ &= \frac{[\det(A)]^2}{a_{12}^2}(x_1 - 1)^2 \ = \ \frac{10^{-12}}{a_{12}^2}(x_1 - 1)^2, \end{aligned}$$

$$\begin{aligned} f(x_1, x_2)|_{x_2=g_2(x_1)} &= [r_1(x_1, g_2(x_1))]^2 + [r_2(x_1, g_2(x_1))]^2 \\ &= [b_1 - (a_{11}x_1 + a_{12}g_1(x_1))]^2 + 0, \quad b_1 = a_{11} - a_{12} \\ &= \frac{[\det(A)]^2}{a_{22}^2}(x_1 - 1)^2 \ = \ \frac{10^{-12}}{a_{22}^2}(x_1 - 1)^2. \end{aligned}$$

Somit ist

$$f(1, -1) = 0, \quad f(0.341, -0.087) = 10^{-12}, \quad f(0.999, -1.001) = 4.274 \cdot 10^{-6}.$$

x_1	$f(x_1, g_1(x_1))$	$f(x_1, t(x_1))$	$f(x_1, g_2(x_1))$
1.010	$3.886 \cdot 10^{-16}$	$1.665 \cdot 10^{-16}$	$3.331 \cdot 10^{-16}$
1.005	0	0	0
1.001	0	0	0
1	0	0	0
0.999	0	0	0
0.341	$1.370 \cdot 10^{-12}$	$5.781 \cdot 10^{-13}$	$1.000 \cdot 10^{-12}$
0.278	$1.665 \cdot 10^{-12}$	$6.938 \cdot 10^{-13}$	$1.200 \cdot 10^{-12}$
0.100	$2.555 \cdot 10^{-12}$	$1.078 \cdot 10^{-12}$	$1.865 \cdot 10^{-12}$
0	$3.155 \cdot 10^{-12}$	$1.331 \cdot 10^{-12}$	$2.303 \cdot 10^{-12}$
-0.100	$3.817 \cdot 10^{-12}$	$1.611 \cdot 10^{-12}$	$2.786 \cdot 10^{-12}$
-1	$1.262 \cdot 10^{-11}$	$5.324 \cdot 10^{-12}$	$9.211 \cdot 10^{-12}$
-10	$3.817 \cdot 10^{-10}$	$1.611 \cdot 10^{-10}$	$2.786 \cdot 10^{-10}$
-10^2	$3.218 \cdot 10^{-8}$	$1.358 \cdot 10^{-8}$	$2.349 \cdot 10^{-8}$
-10^3	$3.161 \cdot 10^{-6}$	$1.334 \cdot 10^{-6}$	$2.307 \cdot 10^{-6}$
-10^6	$3.155 \cdot 10^{0}$	$1.331 \cdot 10^{0}$	$2.303 \cdot 10^{0}$
-10^9	$3.155 \cdot 10^{6}$	$1.331 \cdot 10^{6}$	$2.303 \cdot 10^{6}$
-10^{12}	$3.155 \cdot 10^{12}$	$1.331 \cdot 10^{12}$	$2.303 \cdot 10^{12}$
-10^{15}	$3.155 \cdot 10^{18}$	$1.331 \cdot 10^{18}$	$2.303 \cdot 10^{18}$

Tab. 1.1 Funktional $f(x_1, x_2)$ für ausgewählte Punkte entlang der drei Geraden $g_1(x_1),\ t(x_1),\ g_2(x_1)$, Rechnung in der Gleitpunktarithmetik *double* (16 Dezimalst.)

Einige Berechnungen zu $f(x)$ in Maple

```
> A:=matrix(2,2,[[0.780, 0.563],
                 [0.913, 0.659]]):
  b:=vector(2,[0.217,0.254]):

> r:=evalm(A&*[x1,x2]-b);
  r[1]; r[2];
  evalm(transpose(r)&*r);
  f:=(x1,x2)->(0.780*x1+0.563*x2-0.217)^2+(0.913*x1+0.659*x2-0.254)^2;

  f1:=r[1]^2+r[2]^2;
```

$$r := [0.780\,x1 + 0.563\,x2 - 0.217,\ 0.913\,x1 + 0.659\,x2 - 0.254]$$

$$0.780\,x1 + 0.563\,x2 - 0.217$$

$$0.913\,x1 + 0.659\,x2 - 0.254$$

$$(0.780\,x1 + 0.563\,x2 - 0.217)^2 + (0.913\,x1 + 0.659\,x2 - 0.254)^2$$

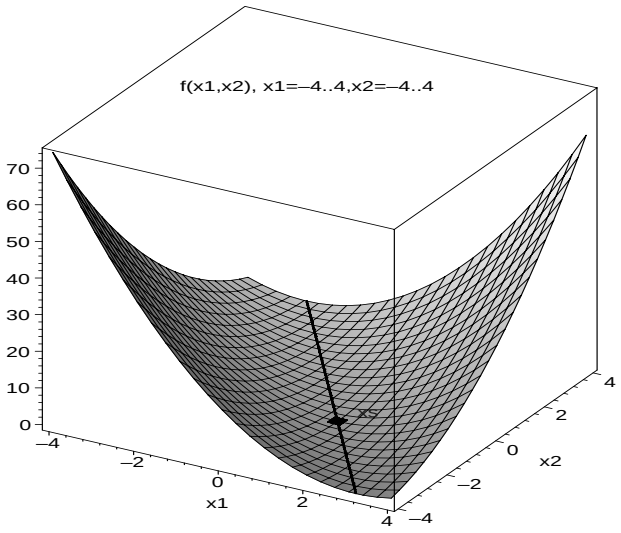
$$f := (x1,\ x2) \rightarrow (0.780\,x1 + 0.563\,x2 - 0.217)^2 + (0.913\,x1 + 0.659\,x2 - 0.254)^2$$

$$f1 := (0.780\,x1 + 0.563\,x2 - 0.217)^2 + (0.913\,x1 + 0.659\,x2 - 0.254)^2$$

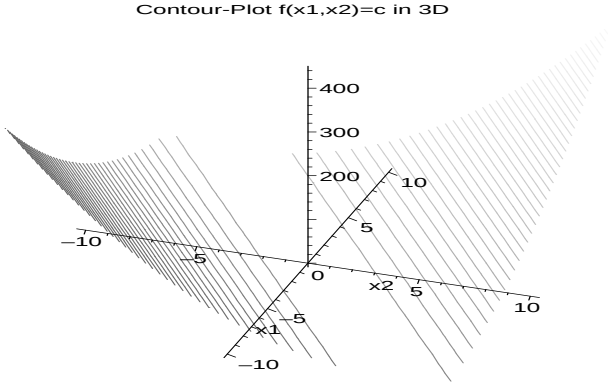
```
>  f(1,-1), f(0.341,-0.087), f(0.999,-1.001);
```

$$0.,\quad 0.1\,10^{-11},\quad 0.4274833\,10^{-5}$$

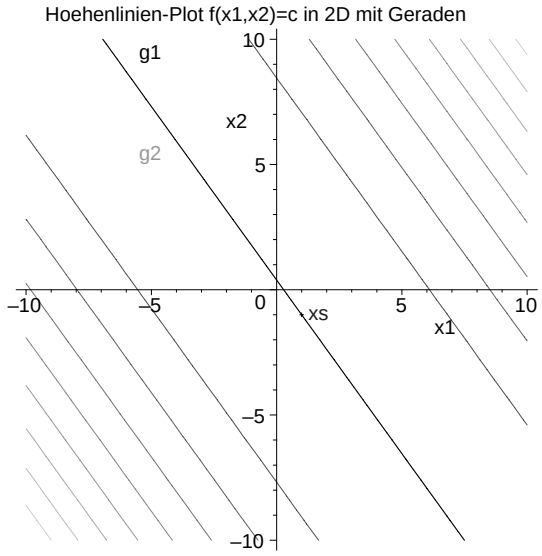
```
> p1:=plot3d(f(x1,x2),x1=-4..4,x2=-4..4,axes=boxed,
            title=‘f(x1,x2), x1=-4..4,x2=-4..4‘):
  # p1:=plot3d(f1,x1=-4..4,x2=-4..4):   # auch moeglich
  p2:=plot3d(0.06,0.85..1.15,-1.15..-0.85,color=blue,symbol=point):
  p3:=textplot3d([[1.35,-0.2,2,‘xs‘]],color=blue):
  p4:=plot3d(f(x1,x2),x1=-4..4,x2=-4..4,style=contour,contours=[0.0001],
             color=black,thickness=3,numpoints=100000):
  plots[display]([p1,p2,p3,p4],orientation=[300,60]);

> pfad:=‘C:/D/Neundorf/Verschie/Publikat/Abstieg/‘:
  dateiname:=‘abst_74‘:
  file:=cat(pfad,dateiname,‘.ps‘):
  interface(plotdevice=ps,plotoutput=file,
            plotoptions=‘portrait,noborder‘);
  plots[display]([p1,p2,p3,p4],orientation=[300,60]);
  interface(plotdevice=win);
```



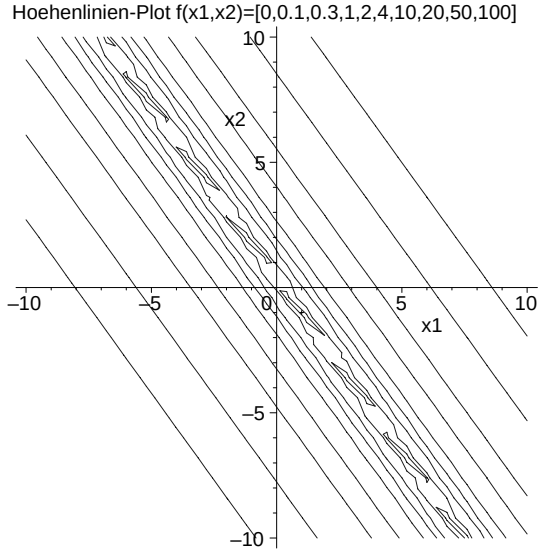

```
> p1:=plot3d(f(x1,x2), x1=-10..10,x2=-10..10,style=contour,
          contours=30,orientation=[290,45],
          axes=normal,title=‘Contour-Plot f(x1,x2)=c in 3D‘):
  plots[display](p1);
```



```
> p1:=contourplot(f(x1,x2),x1=-10..10,x2=-10..10,scaling=constrained):
  p2:=plot([[1,-1]],style=point,color=black):
  p3:=implicitplot(r[1],x1=-10..10,x2=-10..10,color=black,thickness=2):
  p3t:=textplot([-5,9.5,‘g1‘],color=black):
  p4:=implicitplot(r[2],x1=-10..10,x2=-10..10,color=green,thickness=2):
  p4t:=textplot([-5,5.5,‘g2‘],color=green):
  p5t:=textplot([1.55,-0.9,‘xs‘],color=blue):
  plots[display]([p1,p2,p3,p3t,p4,p4t,p5t],
        title=‘Hoehenlinien-Plot f(x1,x2)=c in 2D mit Geraden‘);
```

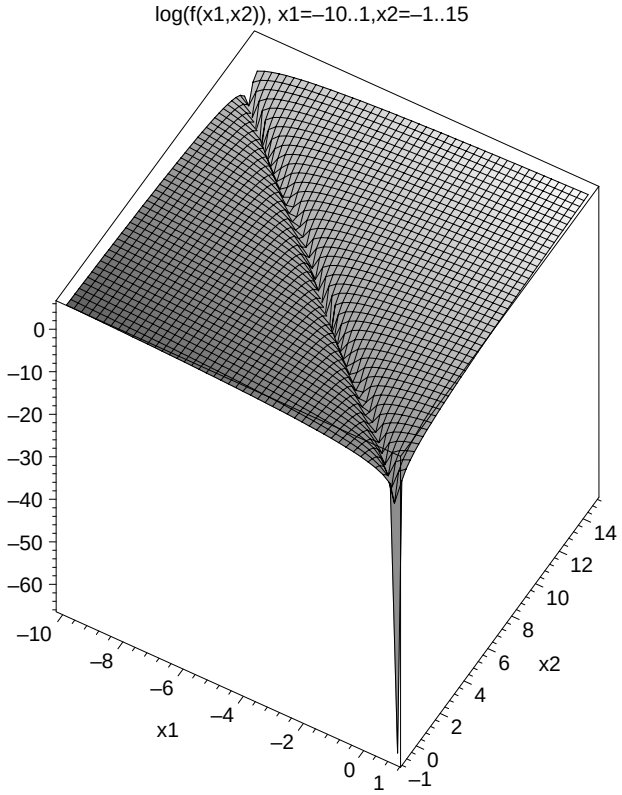



```
> p1:=contourplot(f(x1,x2),x1=-10..10,x2=-10..10,scaling=constrained,
      color=blue,contours=[0,0.1,0.3,1,2,4,10,20,50,100],
      title=‘Hoehenlinien-Plot f(x1,x2)=[0,0.1,0.3,1,2,4,10,20,50,100]‘):
  p2:=plot([[1,-1]],style=point,color=black):
  plots[display]([p1,p2]);
```



Wie flach die Talsohle des Funktionals ist, erkennt man deutlich bei seinem logarithmischen Plot.

```
> p1:=plot3d(log(f(x1,x2)),x1=-10..1,x2=-1..15,axes=boxed,
        orientation=[300,45],numpoints=2000,
        title=‘log(f(x1,x2)), x1=-10..1,x2=-1..15‘):
  plots[display](p1);
```



Wir sehen in den letzten Abbildungen, dass eine hohe Genauigkeit und feine grafische Auflösung in der Talsohle des Funktionals allgemein, besonders aber in der Nähe der Minimumstelle und des Sattelpunkts notwendig sind, um alle Merkmale der Situation zu erfassen und darzustellen.

Beispiel 1.8

Als letztes Beispiel in diesem Abschnitt betrachten wir noch ein Funktional $f(x)$ als Summe zweier nichtlinearer Terme. Es entsteht somit nicht aus einem LGS.
Sei

$$f(x) \;=\; [f_1(x)]^2 + [f_2(x)]^2 \;=\; [10(x_2 - x_1^2)]^2 + (1 - x_1)^2. \tag{1.16}$$

Das eindeutige globale Minimum ist an der Stelle $x^* = (1, 1)^T$.
Die Auswahl der Verfahren zur Bestimmung von x^* reicht vom allgemeinen Iterationsverfahren (Picard-Iteration) über Newton-Verfahren bis zu Abstiegsverfahren sowie ihren Modifikationen.

Es ist zu erwarten, dass bei Abstiegsverfahren bezüglich der Konvergenz zahlreiche Probleme auftreten werden, wenn es zum Beispiel zusätzliche lokale Minima gibt, wenn der Weg zum globalen Minimum durch ein sehr flaches lang gestrecktes, eventuell noch gekrümmtes und mit steilen Wänden umgebenes Tal (Gebiet) verläuft. Die vorliegende Funktion hat die zuletzt genannte Eigenschaft.

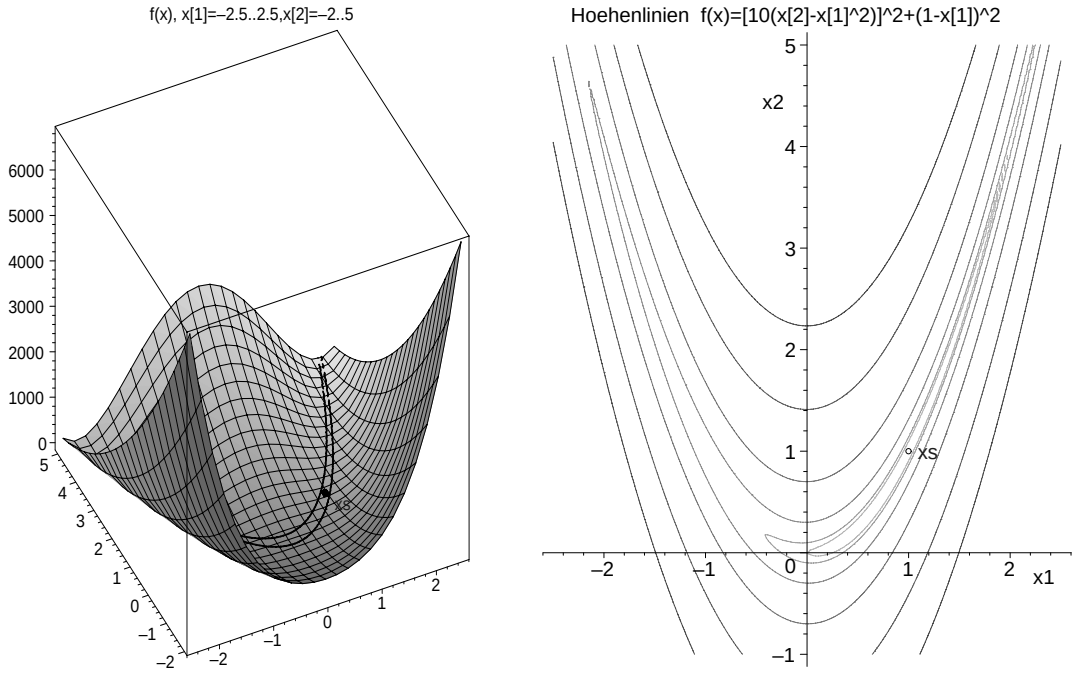


Abb. 1.19 Dateien *abst_81.ps*, *abst_82.ps*
3D-Bild mit Tal und Höhenlinie $f(x) = 4$ sowie Höhenlinienbild von $f(x) = [10(x_2 - x_1^2)]^2 + (1 - x_1)^2$ mit `contours=[500,200,50,10,2,1,0]`

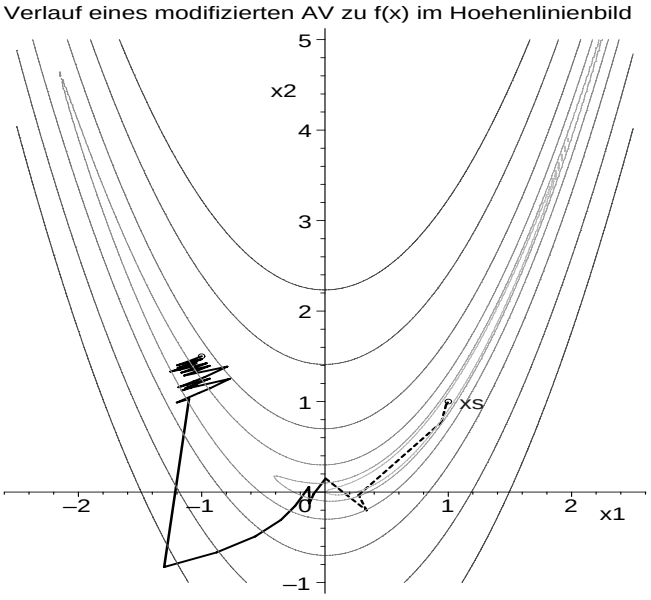


Abb. 1.20
Datei *abst_83.ps*
Verlauf eines modifizierten Abstiegsverfahrens zu $f(x) = [10(x_2 - x_1^2)]^2 + (1 - x_1)^2$ im Höhenlinienbild mit $x^{(0)} = (-1, 1.5)^T$ und den Iterierten $x^{(1)}, x^{(2)}, ..., x^{(43)}$, $x^{(46)}, x^{(50)}, x^{(100)}, x^{(200)}, x^{(299)}$

Kapitel 2

Grundlagen der Abstiegsverfahren

Gegeben sei das reguläre LGS (1.1) mit der exakten Lösung $x^* = A^{-1}b \in \mathbb{R}^n$. Damit gilt $b - Ax^* = 0$. Ein Näherungsverfahren zu seiner Lösung liefert im Allgemeinen auch nur eine Näherungslösung $\tilde{x}$.
Somit entsteht ein Lösungsfehler

$$e \;=\; e(\tilde{x}) \;=\; x^* - \tilde{x} \tag{2.1}$$

und ein nicht verschwindender Rest bzw. Residuum (Residuenvektor)

$$r \;=\; r(\tilde{x}) \;=\; b - A\tilde{x}. \tag{2.2}$$

Zwischen Fehler und Residuum erkennt man den Zusammenhang

$$\begin{aligned} r &= b - A\tilde{x} = A(A^{-1}b - \tilde{x}) = A(x^* - \tilde{x}) = Ae, \\ e &= A^{-1}r. \end{aligned} \tag{2.3}$$

Somit kann nur bei Kenntnis spezieller zusätzlicher Eigenschaften der Koeffizientenmatrix A – eine davon ist die Kondition – aus der einen Größe auf die andere geschlossen werden.
Ein kleines Residuum im Sinne der Norm des Vektors bedeutet also nicht automatisch einen kleinen Lösungsfehler. Dazu braucht man nur das Beispiel 1.7 mit der erweiterten Koeffizientenmatrix zu nehmen.

$$\left(\begin{array}{cc|c} 0.780 & 0.563 & 0.217 \\ 0.913 & 0.659 & 0.254 \end{array} \right).$$

Um einfach festzustellen, ob ein Vektor x Lösung des Systems ist, prüft man, ob das Residuum $r = b - Ax$ einen Nullvektor liefert. Nehmen wir also zwei Kandidaten für die Lösung und zwar

$$\begin{aligned} \bar{x} &= (0.341,\ -0.087)^T, \\ \hat{x} &= (0.999,\ -1.001)^T, \end{aligned}$$

und berechnen dafür das Residuum. Es gilt entsprechend

$$\begin{aligned} \bar{r} &= (0.000\,001,\ 0)^T, \\ \hat{r} &= (0.001\,343,\ 0.001\,572)^T. \end{aligned}$$

Die "Güte" des Fehlers könnte den Betrachter dazu verleiten, $\bar{x}$ als den besseren Vorschlag zu akzeptieren. Das ist jedoch ein Trugschluss bei Kenntnis der exakten Lösung $x^* = (1, -1)^T$.
Ähnlich ist es in der anderen Richtung, wo aus kleinen Fehlern nicht unbedingt kleine Residua folgen müssen. Wir brauchen nur im eindimensionalen Fall zu bleiben und die Lösung von $f(x_1) = a_{11}x_1 - b_1 = 0$ mit $|a_{11}| \gg 1$ zu suchen. Falls $\tilde{x}$ nur geringfügig von der exakten Lösung $x^* = b_1/a_{11}$ abweicht, wird $r = b_1 - a_{11}\tilde{x}$ betragsmäßig sehr groß.
Der Grund ist, dass die Matrix A eine schlechte Kondition hat. Kennzeichen dafür sind unter anderem:

- Die Matrix A ist fast singulär.
- Die Determinante von A ist nahe Null.
- Wenn die Matrix A Elemente der Größenordnung $\mathcal{O}(1)$ besitzt, dann hat die inverse Matrix A^{-1} betragsmäßig große Elemente, oder umgekehrt.
- Das Spektrum der Eigenwerte von A ist sehr "breit". Es liegen Größenordnungen zwischen dem betragsmäßig kleinsten ($\neq 0$) und größten Eigenwert.
- Eine einfache geometrische Interpretation ist die Charakterisierung der Lösung als Schnittpunkt zweier Geraden, die sich dann in einem extrem spitzen Winkel schneiden.

Zur Kondition gehört ihre wertemäßige Berechnung als Konditionszahl mit Hilfe der inversen Matrix und der Norm

$$\kappa(A) \ = \ \mathrm{cond}_s(A) = \|A\|_s \|A^{-1}\|_s, \tag{2.4}$$

wobei durch den Index die Wahl einer konkreten Norm $\|A\|_s$ gemeint ist.
Ein schlechter Konditionswert charakterisiert im Groben die Reduzierung der gültigen Mantissenstellen der in Rechnungen benutzten Gleitpunktarithmetik.

Eine wichtige Rolle bez. der Matrixeigenschaften und bei Abschätzungen spielen weiterhin die Eigenwerte $\lambda(A)$ gemäß $Ax = \lambda x,\ x \neq 0$, die sich auch im Zusammenhang mit den Nullstellen des charakteristischen Polynoms $p_n(\lambda) = \det(A - \lambda I)$ ergeben, das Spektrum

$$\sigma(A) \ = \ \{\lambda\} \ = \ \{\lambda_1, \lambda_2, ..., \lambda_n\}, \tag{2.5}$$

der Spektralradius

$$\rho(A) \ = \ \max_{\lambda \in \sigma(A)} |\lambda| \ = \ \max_{i=1,2,...,n} |\lambda_i| \ \leq \ \|A\| \tag{2.6}$$

und die Singulärwerte $\sigma_i(A)$ der Matrix (siehe [17]).

2.1 Ansätze für quadratische Formen

Die Idee zur Wahl von quadratischen Formen und ihre Einbindung in ein Optimierungsproblem für die Lösung von (1.1) kann man ganz einfach im eindimensionalen (skalaren) Fall nachvollziehen.
Für die Nullstellenaufgabe

$$f(x) \;=\; ax + b \;=\; 0, \quad a \neq 0,$$

mit der Nullstelle $x^* = -\frac{b}{a}$ ergibt sich sofort die daraus abgeleitete und zu minimierende Funktion

$$g(x) \;=\; [f(x)]^2.$$

Sie hat mindestens dieselben Glattheitseigenschaften wie $f(x)$ und an der Stelle x^* ihr eindeutiges absolutes Minimum.
Man schreibt

$$x^* \;=\; \arg\min_{x \in \mathbb{R}} g(x).$$

Bezüglich der Minimumstelle ändert sich nichts, wenn wir die Funktion $g(x)$ strecken und vertikal verschieben, also mit $c_1 g(x) + c_2,\ c_1 > 0$, arbeiten.

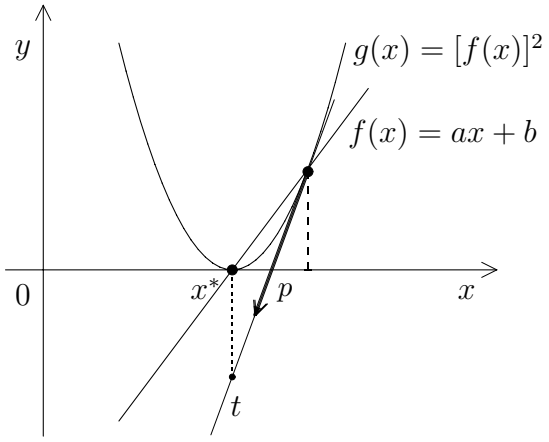


Abb. 2.1 Datei *abst1.pic*
Nullstellensituation $f(x) = 0$ und zugehörige Minimierungsaufgabe $\min\limits_x [f(x)]^2$ im $\mathbb{R}^1$

Wenn man entlang der Tangente t an die Kurve $g(x)$ in Richtung des steilsten Abstiegs (Vektor p) geht, "fährt" man sozusagen unter dem Minimum entlang und kann mittels einer Strahlenminimierung genau die Stelle darunter treffen.
Im mehrdimensionalen Fall werden im Allgemeinen mehrere aufeinander folgende Strahlenminimierungen in Richtung der Minimumstelle erforderlich sein.

Wir betrachten einige quadratische oder bilineare Formen für das LGS (1.1) bzw. für die Koeffizientenmatrix A, die daraus abgeleiteten Funktionale sowie erste Eigenschaften dieser.
Dabei sei $x, y, ... \in \mathbb{R}^n$. Weiterhin verwenden wir sowohl die Notation mit dem Skalarprodukt $(x, y) = x^T y \in \mathbb{R}$ als auch die Matrix-Vektor-Darstellung mit der entsprechenden Transposition der Felder.

(a) Bilinearform

$$\begin{aligned} (x, Ay) \;=\; x^T Ay \;&=\; (x^T Ay)^T \;=\; y^T A^T x \;=\; (y, A^T x) \;=\; (A^T x, y), \\ x^T Ay \;&\neq\; x^T A^T y, \text{ falls } \; A \neq A^T \text{ und } x \neq y. \end{aligned} \tag{2.7}$$

(b) Quadratische Form

$$(Ax, x) = (x, Ax) = x^T Ax = (x^T Ax)^T = x^T A^T x = (x, A^T x) = (A^T x, x). \tag{2.8}$$

Falls $A = A^T > 0$ ist, dann definiert man die energetische Norm (A-Norm)

$$x^T Ax \;=\; (Ax, x) \;=\; (x, x)_A \;=\; \|x\|_A^2 \tag{2.9}$$

(siehe [18]), und es gilt $x^T Ax > 0$ für $x \neq 0$.

(c) Quadratische Form

$$\begin{aligned} x^T Ax \;&=\; x^T A^T x \;\text{ für alle }\; A, x \;\text{ gemäß (2.8)} \\ &=\; \frac{1}{2} x^T Ax + \frac{1}{2} x^T A^T x \\ &=\; \frac{1}{2} x^T (A + A^T) x \\ &=\; \frac{1}{2} x^T \widetilde{A} x, \end{aligned} \tag{2.10}$$

wobei $\widetilde{A} = A + A^T = \widetilde{A}^T$.

(d) Quadratische Form

$$\begin{aligned} (Ax, Ax) \;&=\; (Ax)^T Ax \;=\; \|Ax\|_2^2, \\ (Ax, Ax) \;&=\; (x, Ax)_A \;=\; (Ax, x)_A, \text{ falls } \; A = A^T > 0. \end{aligned} \tag{2.11}$$

(e) Funktional $f(x)$ gemäß Formel (1.5)

$$\begin{aligned} f_i(x) \;&=\; (b - Ax)_i \;=\; b_i - \sum_{j=1}^{n} a_{ij} x_j, \; i = 1, 2, ..., n, \\ f(x) \;&=\; \sum_{i=1}^{n} [f_i(x)]^2 \;\geq\; 0, \\ f(x^*) \;&=\; \min_{x \in \mathbb{R}^n} f(x) \;=\; 0. \end{aligned} \tag{2.12}$$

In den Beispielen 1.7 und 1.8 haben wir dieses Funktional bereits verwendet.

(f) Quadratische Form $\delta_Q(x)$ und Funktional $Q(x)$ auf der Basis von A

$$\begin{aligned}
0 \le \delta_Q(x) &= \|e(x)\|_A^2 = \|x^* - x\|_A^2, \quad A = A^T > 0, \\
\delta_Q(x) &= (x^* - x)^T A(x^* - x) \quad \text{(allgemein)} \\
&= x^T Ax - x^T Ax^* - x^{*T} Ax + x^{*T} Ax^* \\
&= x^T Ax - x^T Ax^* - (x^{*T} Ax)^T + x^{*T} b \\
&= x^T Ax - x^T Ax^* - x^T A^T x^* + x^{*T} b \\
&= x^T Ax - x^T (A + A^T) x^* + x^{*T} b \qquad (2.13) \\
&= x^T Ax - 2x^T Ax^* + x^{*T} b, \quad \text{falls } A = A^T \\
&= x^T Ax - 2x^T b + x^{*T} b \\
&= 2\Big(\underbrace{\frac{1}{2} x^T Ax - x^T b}_{= Q(x)}\Big) + x^{*T} b,
\end{aligned}$$

$$Q(x) = \frac{1}{2} x^T Ax - x^T b. \qquad (2.14)$$

Falls die Matrix nicht symmetrisch ist, kommt man nur bis zum Funktional

$$Q_{ns}(x) = \frac{1}{2}\Big(x^T Ax - x^T (A + A^T) x^*\Big). \qquad (2.15)$$

Außerdem gilt $\delta_Q(x) \ge 0$ natürlich nur im Fall $A = A^T > 0$.

(g) Quadratische Form $\delta_R(x)$ und Funktional $R(x)$ auf der Basis von $A^T A$

$$\begin{aligned}
0 \le \delta_R(x) &= \|r(x)\|_2^2 = r(x)^T r(x), \quad r(x) = b - Ax \\
&= \|b - Ax\|_2^2 = (b - Ax)^T (b - Ax) \\
&= (Ax)^T Ax - (Ax)^T b - b^T Ax + b^T b \\
&= x^T A^T Ax - x^T A^T b - (b^T Ax)^T + b^T b \\
&= x^T A^T Ax - x^T A^T b - x^T A^T b + b^T b \\
&= x^T A^T Ax - 2x^T A^T b + b^T b \qquad (2.16) \\
&= x^T A^2 x - 2x^T Ab + b^T b, \quad \text{falls } A = A^T \\
&= (Ax, x)_A - 2(x, b)_A + (b, b), \quad \text{falls } A = A^T > 0 \\
&= 2\Big(\underbrace{\frac{1}{2}(Ax, x)_A - (x, b)_A}_{= R_s(x)}\Big) + (b, b),
\end{aligned}$$

$$R_s(x) = \frac{1}{2}(Ax, x)_A - (x, b)_A = \frac{1}{2} x^T A^2 x - (Ax)^T b. \qquad (2.17)$$

Wegen $b = Ax^*,\ e(x) = x^* - x$ gilt auch

$$\|r(x)\|_2^2 = \|A(x^* - x)\|_2^2 = [A^T A(x^* - x)]^T(x^* - x) = \|e(x)\|_{A^T A}^2. \tag{2.18}$$

Falls die Matrix nicht symmetrisch ist, und dafür ist dieser Zugang ja vorgesehen, denn er beinhaltet die Symmetrisierung, so kommt man zum Funktional

$$R(x) \ = \ \frac{1}{2}(Ax)^T Ax - (Ax)^T b \ = \ \frac{1}{2}x^T A^T Ax - x^T A^T b. \tag{2.19}$$

Im Fall $A = A^T$ ist eine zusätzliche Symmetrisierung von A und damit die Erzeugung von A^2 eigentlich überflüssig und man sollte mit $Q(x)$ operieren.

Es gibt jedoch diverse Beispiele, wo die Matrix A^2 im Vergleich zu A in ihren Elementen ausgeglichener ist, wie in

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}, \quad A^2 = \begin{pmatrix} 5 & 5 \\ 5 & 10 \end{pmatrix} = 5 \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Von weiterem Interesse sind also die Betrachtungen und Formeln in den Punkten (f) und (g).

Wir haben zum LGS $Ax = b,\ A = A^T > 0$,

$$e(x^*) = 0 \ \text{ und das Ziel } \ \|e(x)\|_A \ \to \ \min_x,$$

und zum LGS $A^T Ax = A^T b$

$$r(x^*) = 0 \ \text{ und das Ziel } \ \|r(x)\|_2 = \|e(x)\|_{A^T A} \ \to \ \min_x,$$

sowie nach Transformation die Gleichungen

$$\|e(x)\|_A^2 \ = \ 2\Big(\underbrace{\frac{1}{2}x^T Ax - x^T b}_{= Q(x) \to \min\limits_x} \Big) + \underbrace{x^{*T} b}_{= \text{const}} \tag{2.20}$$

und

$$\|r(x)\|_2^2 \ = \ 2\Big(\underbrace{\frac{1}{2}(Ax)^T Ax - (Ax)^T b}_{= R(x) \to \min\limits_x} \Big) + \underbrace{b^T b}_{= \text{const}} . \tag{2.21}$$

mit den neuen zu minimierenden Funktionalen $Q(x)$ und $R(x)$.

Zum Funktional $R(x)$ kommt man ebenfalls über das spd Normalgleichungssystem $A^T Ax = A^T b$, indem man genau dafür das Funktional $Q(x)$ nimmt, wie aus dem Vergleich der Formeln (2.14) und (2.19) zu sehen ist.

Wir nutzen die Idee aus Abschnitt (f) für das Normalgleichungssystem in zwei Varianten.

(1) Normalgleichungssystem $A^TAx = A^Tb,\ \ A^TA = (A^TA)^T > 0$

$$\begin{aligned} 0 &\leq \|e(x)\|^2_{A^TA} \\ &= \|x^* - x\|^2_{A^TA} \\ &= (x^* - x)^T A^T A(x^* - x) \\ &= (A(x^* - x))^T A(x^* - x) \ = \ \|A(x^* - x)\|_2^2, \ \ Ax^* = b \\ &= (b - Ax)^T(b - Ax) \ = \ \|b - Ax\|_2^2 \\ &= \|r(x)\|_2^2. \end{aligned} \tag{2.22}$$

Die Minimierung des Fehlers von x in der Norm $\|\cdot\|_{A^TA}$ entspricht der Minimierung seines Residuums in der Norm $\|\cdot\|_2$.

(2) Modifiziertes Normalgleichungssystem $AA^Ty = b,\ x = A^Ty,\ AA^T = (AA^T)^T > 0$

$$\begin{aligned} 0 &\leq \|e(y)\|^2_{AA^T} \\ &= \|y^* - y\|^2_{AA^T} \\ &= (y^* - y)^T AA^T(y^* - y) \\ &= (A^T(y^* - y))^T A^T(y^* - y) \ = \ \|A^T(y^* - y)\|_2^2 \\ &= (A^Ty^* - A^Ty)^T(A^Ty^* - A^Ty) \ = \ \|A^Ty^* - A^Ty\|_2^2, \ \ A^Ty^* = x^* \\ &= (x^* - x)^T(x^* - x) \\ &= \|e(x)\|_2^2. \end{aligned} \tag{2.23}$$

Die Minimierung des Fehlers von y in der Norm $\|\cdot\|_{AA^T}$ entspricht der Minimierung des Fehlers von x in der Norm $\|\cdot\|_2$ bei Anwendung auf das modifizierte Normalgleichungssystem.

Als Maß für den Fehler der Iterierten in einem Abstiegsverfahren eignet sich natürlich vorrangig die Residuumnorm $\|r(x)\|_2$. In der Formel erscheint erstens nicht die exakte Lösung x^* und zweitens ist die bequemere euklidische Norm anstelle der A-Norm auszuwerten. Nur für akademische Kontrollrechnungen, wo die exakte Lösung bekannt ist, bietet sich die Berechnung beider Fehlerterme $\|r(x)\|_2$ und $\|e(x)\|_A$ an.

Kapitel 3

Abstiegsverfahren – 1

3.1 Die quadratische Form $Q(x)$

Die Minimierung des Funktionals $\|e(x)\|_A$ geht konform mit der Minimierung des Funktionals und der quadratischen Form

$$Q(x) \;=\; \frac{1}{2}x^T Ax - x^T b, \quad A = A^T > 0. \tag{3.1}$$

Sein Gradient als Richtung des steilsten Anstiegs ist

$$\text{grad } Q(x) \;=\; \nabla Q(x) \;=\; Ax - b, \tag{3.2}$$

für die Hesse-Matrix gilt

$$\nabla^2 Q(x) \;=\; A \;=\; A^T > 0, \tag{3.3}$$

so dass $Q(x)$ konvex ist.
Wir bemerken, dass im Fall $A \neq A^T$ der Gradient $\nabla Q(x) = \frac{1}{2}(A + A^T)x - b$ ist.
Aus der Beziehung (2.20) erhält man

$$\begin{aligned}
0 \leq \|\varepsilon(x)\|_A^2 &= 2Q(x) + x^{*T} b = 2[Q(x) - (-\tfrac{1}{2}x^{*T} b)] = 2[Q(x) - Q(x^*)], \\
Q(x) &= \frac{1}{2}\Big(\|e(x)\|_A^2 - x^{*T} b\Big), \\
Q(x^*) &= \min_{x \in \mathbb{R}^n} Q(x) \;=\; -\frac{1}{2}x^{*T} b \;=\; -\frac{1}{2}x^{*T} A x^* \leq 0, \\
Q(x^* + \Delta x) \;-\; Q(x^*) \;&=\; \frac{1}{2}\|\Delta x\|_A^2 > 0, \quad \Delta x \neq 0.
\end{aligned} \tag{3.4}$$

Durch die Normäquivalenz (siehe [17]) kann man aus Abschätzungen für $\|e(x)\|_A$ auch zu solchen in der euklidischen Norm gelangen.
Die notwendige Bedingung am Minimum (Extremum) $\nabla Q(x) = 0$ entspricht der Lösung des LGS.

Der Zusammenhang zwischen Minimumstelle und Lösung soll als Satz formuliert werden.

Satz 3.1 *Unter der Voraussetzung $A = A^T > 0$ ist die einzige Minimumstelle x^* von $Q(x)$ zugleich die eindeutige Lösung des LGS $Ax = b$.*

Beweis.
(1) Für die eindeutige Lösbarkeit des LGS betrachten wir die Inverse A^{-1}.
Diese existiert. Wäre das nicht so, hätte das homogene LGS $Ax = 0$ eine nichttriviale Lösung x. Somit wäre auch $x^T Ax = 0,\ x \neq 0$, was der positiven Definitheit von A widerspricht.
(2) Für eine spd Matrix A existiert eine orthogonale Matrix U ($U^T U = I$), so dass die Matrix $T = U^T AU$ eine Diagonalmatrix ist mit den reellen positiven Diagonalelementen t_{ii} als Eigenwerte von A (siehe Hauptachsentheorem Satz 2.31 bzw. spezieller Fall des Satzes von SCHUR 2.15 in [18]).
(3) Zu zeigen ist nun der Zusammenhang mit der Extremwerteigenschaft von $Q(x)$, das heißt

$$Q(x^*) \ < \ Q(x) \ \text{ für alle } \ x \in \mathbb{R}^n \backslash \{x^*\}.$$

Betrachten wir zunächst den einfachen Fall, dass A eine Diagonalmatrix der Form $A = \text{diag}(d_1, d_2, ..., d_n)$ ist. Aus der spd-Eigenschaft folgt sofort für ihre Diagonalelemente $d_i > 0,\ i = 1, 2, ..., n$.
Das Funktional ist gegeben durch

$$Q(x) \ = \ \sum_{i=1}^{n} \Big(\frac{1}{2} d_i x_i^2 - x_i b_i \Big).$$

Das Extremum von $Q(x)$ erhält man durch die Bedingung

$$\nabla Q(x) \ = \ \Big(\frac{\partial Q(x)}{\partial x_1}, \frac{\partial Q(x)}{\partial x_2}, ..., \frac{\partial Q(x)}{\partial x_n} \Big)^T \ = \ 0.$$

Wegen

$$\frac{\partial Q(x)}{\partial x_j} \ = \ \frac{\partial}{\partial x_j} \Big(\frac{1}{2} d_j x_j^2 - x_j b_j \Big) \ = \ d_j x_j - b_j \ = \ (Ax - b)_j \ = \ 0, \quad j = 1, 2, ..., n,$$

ist es eindeutig und führt auf die Forderung $Ax - b = 0$, also auf die Lösung des LGS. Das Extremum ist wegen der Positivität von A ein Minimum.
Betrachten wir nun den allgemeinen Fall. Nach Teil (2) ist A ähnlich zu einer Diagonalmatrix $D = \text{diag}(d_1, d_2, ..., d_n)$ mit einer orthogonalen Transformationsmatrix U ($U^T U = UU^T = I$), d. h. $D = U^T AU = U^{-1} AU$ oder $A = UDU^T$.
Es gilt $d_i > 0\ \forall i$. Außerdem ist

$$Q(x) \ = \ \frac{1}{2} x^T UDU^T x - x^T UU^T b \ = \ \frac{1}{2} (U^T x)^T D (U^T x) - (U^T x)^T U^T b.$$

Wir definieren $z = U^T x,\ c = U^T b$ und

$$Q_D(z) \ = \ \frac{1}{2} z^T D z - z^T c.$$

Offenbar liegt das Minimum von $Q(x)$ genau dann in x, wenn das Minimum von $Q_D(z)$ in z angenommen wird. Das Minimum von $Q_D(z)$ wird gemäß Diagonalfall in dem Wert z erreicht, welcher Lösung des LGS $Dz = c$ ist.
Dann ist aber x genau die Lösung von $Ax = b$. □

Die Minimaleigenschaft des Funktionals $Q(x)$ war schon in der letzten Beziehung von (3.4) zu erkennen.

Wie konstruiert man nun eine Folge von Näherungen zum Minimum von $Q(x)$?
Welche Form haben die Iterationsverfahren zu seiner Bestimmung?
Ausgehend von einem Startvektor $x^{(0)}$ als ersten Näherungswert zu x^* sucht man weitere Iterierte $x^{(1)},\ x^{(2)}, ...$ eines Abstiegsverfahrens mit der Bedingung

$$Q(x^{(m+1)}) \ < \ Q(x^{(m)}),$$

also $\|e^{(m+1)}\|_A < \|e^{(m)}\|_A,\ e^{(m)} = e(x^{(m)})$.
Dazu betrachtet man eine Folge "eindimensionaler Probleme" mit einer Strahlenminimierung.

Definition 3.1 Strahlenminimierung und Abstiegsverfahren (AV)
Gegeben sei die Iterierte $x^{(m)}$ des AV mit $Q(x^{(m)})$.
Mit einem Richtungsvektor $p^{(m)}$ ermittelt man diejenige Stelle $x^{(m+1)}$ auf der Geraden

$$g(\alpha) \ = \ x^{(m)} + \alpha p^{(m)}, \quad \alpha \in \mathbb{R}, \tag{3.5}$$

in welcher die Funktion

$$f(\alpha) \ = \ Q(x^{(m)} + \alpha p^{(m)}) \tag{3.6}$$

ihr Minimum annimmt.
$p^{(m)}$ ist eine lokal günstige Suchrichtung, der optimale Wert $\alpha = \alpha_{min} = \alpha_m$ heißt Schrittzahl und die neue Iterierte im AV ist

$$x^{(m+1)} \ = \ x^{(m)} + \alpha_m p^{(m)}. \tag{3.7}$$

Wir notieren die Minimierungsaufgabe im m-ten Schritt in der kompakten Form

$$\boxed{\begin{aligned} Q(x^{(m+1)}) \ &= \ \min_{\alpha \in \mathbb{R}} \ Q(x^{(m)} + \alpha p^{(m)}) \\ x^{(m+1)} \ &= \ \arg\min_{\alpha \in \mathbb{R}} \ Q(x^{(m)} + \alpha p^{(m)}) \end{aligned}} \tag{3.8}$$

je nachdem, worauf wir den Schwerpunkt der Betrachtung legen.

Die Bestimmung des Parameters α und somit von $x^{(m+1)}$ in der Strahlenminimierung folgt aus dem folgenden Lemma.

Lemma 3.2 *Gegeben seien zwei Vektoren x, p $(p \neq 0)$.*
Das Minimum der Funktion $f(\alpha) = Q(x+\alpha p)$ auf der Geraden $g(\alpha) = x+\alpha p$, $\alpha \in \mathbb{R}$, wird angenommen für

$$\alpha \;=\; \alpha_{min} \;=\; \frac{p^T r}{p^T A p}. \tag{3.9}$$

Beweis.
Dazu brauchen wir nur die skalare Funktion $f(\alpha)$ und bezüglich α die notwendige Bedingung an ihrem Minimum auszuwerten.
Es gilt

$$\begin{aligned} f(\alpha) &= Q(x+\alpha p) \;=\; \frac{1}{2}(x+\alpha p)^T A(x+\alpha p) - (x+\alpha p)^T b \\ &= Q(x) + \alpha p^T(Ax-b) + \frac{1}{2}\alpha^2 p^T A p, \\ 0 \;=\; f'(\alpha) &= p^T(Ax-b) + \alpha p^T A p, \quad r = b - Ax, \\ \alpha &= \frac{p^T r}{p^T A p} = \frac{(p,r)}{(p,Ap)} = \frac{(r,p)}{(Ap,p)}, \\ f''(\alpha) &= p^T A p \;>\; 0, \quad p \neq 0. \end{aligned} \tag{3.10}$$

□

Für die Schrittzahl haben wir verschiedene Darstellungen angegeben.
Wenn man also von $x^{(m)}$ in Richtung $p^{(m)}$ geht und zur Stelle $x^{(m+1)}$ kommt, so muss diese nicht unbedingt den geringsten Abstand von der Lösung x^* (im Sinne des üblichen euklidischen Abstands) haben. Es wäre durchaus denkbar, dass eine andere Stelle $x'^{(m+1)}$ näher an x^* liegt. In der nachfolgenden Abbildung ist z. B. das Lot von x^* auf die Gerade g gezeichnet worden, um $x'^{(m+1)}$ zu erhalten.
Der Abstand zwischen $x^{(m+1)}$ und x^* ist eventuell minimal im Sinne einer anderen Norm. Leider kennt man x^* nicht, um sein Lot (Projektion) auf die Gerade g zu fällen.

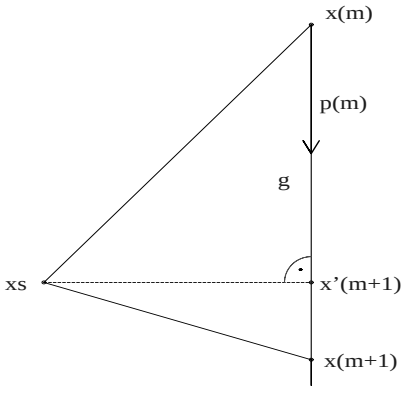


Abb. 3.1
Datei *abst_90.ps*
Geometrische Deutung der Abstiegssituation in der Draufsicht,
`xs`$=x^*$, `z(m)`$=z^{(m)}$

Wünschenswert in Bezug auf die Wahl von $x^{(m+1)}$ ist also ein möglichst kleiner Abstand (euklidische Norm) $\|x^* - x^{(m+1)}\|_2$.

Anders ausgedrückt heißt dies, dass

$$x^* - x^{(m+1)} \perp p^{(m)} \quad \text{bzw.} \quad (x^* - x^{(m+1)}, p^{(m)}) = 0$$

ist. Damit ist der Vektor $x^{(m+1)} - x^{(m)}$ die Projektion von $x^* - x^{(m)}$ auf die Gerade g bzw. den Vektor $p^{(m)}$. Auf diesem Weg haben wir den Begriff der Projektionsmethode eingeführt.

In dem von uns betrachteten AV liegt jedoch das Augenmerk auf der Berechnung von Funktionswerten $Q(x)$ entlang eines Strahls, auf der Minimierung des Funktionals $Q(x)$ in Strahlenrichtung sowie damit wegen (3.4) auf der Minimierung von $\|e^{(m+1)}\|_A = \|x^* - x^{(m+1)}\|_A$. In [11] wurde gezeigt, dass bei $A = A^T > 0$ auch die Normwerte $\|e^{(m+1)}\|_2$ kontinuierlich kleiner werden.

In der folgenden Abbildung soll zumindest in einer etwas anderen Situation angedeutet werden, dass trotz einer Verkleinerung von $Q(x)$ bzw. $\|e(x)\|_A$ die Ungleichung $\|x^* - x^{(m+1)}\|_2 > \|x^* - x^{(m)}\|_2$ auftreten kann. Dabei stammt das Höhenlinienbild natürlich nicht von einer spd Matrix.

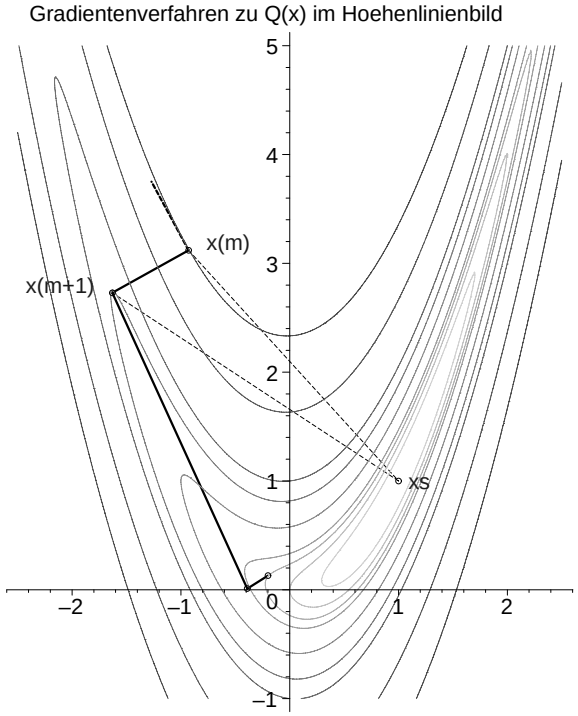


Abb. 3.2
Datei *abst_93.ps*
Abstiegssituation für $Q(x)$,
aber $\|x^* - x^{(m+1)}\|_2 > \|x^* - x^{(m)}\|_2$

Die Wahl der Suchrichtung p zum Funktional $Q(x)$ charakterisiert sowohl Algorithmengruppen als auch einzelne Methoden.

In unsere Betrachtungen beziehen wir folgende Verfahren ein.

– Gradientenverfahren, auch als Methode des steilsten Abstiegs bezeichnet,
– Abstiegsverfahren mit linear unabhängigen Richtungen,
– Abstiegsverfahren mit konjugierten (A-orthogonalen) Richtungen,
– Verfahren der konjugierten Gradienten.

3.2 Das Gradientenverfahren

Das (klassische) Gradientenverfahren (GV) ist ein AV der Gestalt (3.7) und verwendet für das Funktional $Q(x)$ gemäß Formel (3.1) als Suchrichtung $p(x)$ die Richtung des steilsten Abstiegs des Funktionals (deepest descent method), die gleich dem negativen Gradient $-\nabla Q(x)$ ist und damit dem Residuum $r(x)$ entspricht.

$$p(x) \;=\; r(x) \;=\; -\nabla Q(x) \;=\; b - Ax. \tag{3.11}$$

Eine Normierung von $p(x)$ wäre denkbar, wird aber praktisch nicht vorgenommen.
Die Durchführung der Strahlenminimierung mit der notwendigen Bedingung beim Minimum erlaubt die Berechnung der Schrittzahl (Suchschritt, Vielfache) α im GV.

$$\alpha \;=\; \alpha_m \;=\; \frac{r^{(m)T} r^{(m)}}{r^{(m)T} A r^{(m)}} \;=\; \frac{(r^{(m)}, r^{(m)})}{(Ar^{(m)}, r^{(m)})} \;=\; \frac{\parallel r^{(m)} \parallel_2^2}{\parallel r^{(m)} \parallel_A^2} \;\geq\; 0. \tag{3.12}$$

Der Iterationsschritt als Aufdatierungsformel im GV ist

$$x^{(m+1)} \;=\; x^{(m)} + \alpha_m r^{(m)}. \tag{3.13}$$

Damit ergibt sich aus allen Schritten die monoton fallende Folge von Funktionswerten

$$Q(x^{(0)}) \geq Q(x^{(1)}) \geq ... \geq Q(x^{(m)}) \geq Q(x^{(m+1)}) \geq \;...\; \geq\; Q(x^*).$$

Der wesentliche Aufwand in einem Schritt des Zyklus entsteht bei der Berechnung der Schrittzahl durch eine Matrix-Vektor-Multiplikation $Ar^{(m)}$ sowohl für α_m als auch $r^{(m+1)}$, ansonsten sind höchstens Skalarprodukte auszuwerten.
Das Residuum selber kann man rekursiv bestimmen. Ausgehend von seiner expliziten Formel $r^{(m)} = b - Ax^{(m)}$ bieten sich zwei Versionen der rekursiven Ermittlung an. Auf Unterschiede bei der Implementierung dieser Versionen der Residuumsberechnung wird später eingegangen.
Für die rekursive Darstellung führen wir die Kürzel

$$v \;=\; Ax, \quad w \;=\; Ap, \quad t \;=\; Ar \tag{3.14}$$

ein. Wegen $p = r$ ist im GV $w = Ar$ und die Größe t wird nicht benötigt.

$$\begin{aligned}
r^{(m+1)} &= b - Ax^{(m+1)} \;=\; b - A(x^{(m)} + \alpha_m r^{(m)}) \\
&= b - (Ax^{(m)} + \alpha_m A r^{(m)}) \\
&= b - (v^{(m)} + \alpha_m w^{(m)}), \quad w^{(m)} = Ar^{(m)} \\
r^{(m+1)} &= b - v^{(m+1)}, \quad \text{wobei} \quad v^{(m+1)} = v^{(m)} + \alpha_m w^{(m)} \\
&\qquad\qquad\qquad\qquad \text{(1. rekursive Berechnung)}, \\
&= b - Ax^{(m)} - \alpha_m A r^{(m)} \\
&= r^{(m)} - \alpha_m A r^{(m)} \\
r^{(m+1)} &= r^{(m)} - \alpha_m w^{(m)} \quad \text{(2. rekursive Berechnung)}.
\end{aligned} \tag{3.15}$$

Mit den Abkürzungen notieren wir auch das Funktional

$$\begin{aligned} Q(x) &= \frac{1}{2}x^TAx - x^Tb \;=\; -\frac{1}{2}x^T(2b - Ax) \;=\; -\frac{1}{2}x^T(b - Ax + b) \\ &= -\frac{1}{2}x^Ts, \quad s = r + b = 2b - Ax = 2b - v \\ &= \frac{1}{2}(\|e(x)\|_A^2 - x^{*T}b), \quad e(x) = x^* - x, \\ Q(x^{(m)}) &= -\frac{1}{2}x^{(m)T}s^{(m)} \;=\; -\frac{1}{2}x^{(m)T}(2b - v^{(m)}). \end{aligned}$$

Die Funktionswerte $Q(x^{(m)})$ bzw. $\|e(x^{(m)})\|_A$ sind abnehmend und $Q(x^*) = -\frac{1}{2}x^{*T}b$. Den Verlauf von $Q(x^{(m)})$ kann man kontrollieren, genauso ist unter Beachtung der Beziehung (2.21) eine Kontrolle der Größen $\|r(x^{(m)})\|_2^2 = 2R(x^{(m)}) + b^Tb$ möglich.

Bezüglich der Abbruchbedingungen für den Iterationsprozess eignen sich bei vorgegebener Toleranz $\varepsilon > 0$ beispielsweise folgende Kriterien:

– Iterationsanzahl begrenzen, $m \le m_{max}$,
– $\|r^{(m)}\|_2 < \varepsilon$,
– $\|r^{(0)}\|_2^2 < \varepsilon$ bzw. $\dfrac{r^{(m)T}r^{(m)}}{r^{(0)T}r^{(0)}} = \dfrac{\|r^{(m)}\|_2^2}{\|r^{(0)}\|_2^2} < \varepsilon$ bei $m > 0$,
– $\|x^{(m+1)} - x^{(m)}\|_2 < \varepsilon$,
– $|Q(x^{(m+1)}) - Q(x^{(m)})| < \varepsilon$.

Meistens wird der relative Fehler mit dem Residuumnormquadrat verwendet.

Neben der Wahl geeigneter Abbruchbedingungen ist im Fall beliebiger Matrizen noch ein zusätzlicher Test auf die Durchführbarkeit der Berechnung von

$$\alpha_m = \frac{r^{(m)T}r^{(m)}}{r^{(m)T}Ar^{(m)}} = \frac{\|r^{(m)}\|_2^2}{r^{(m)T}Ar^{(m)}}$$

mit Nenner $\neq 0$ zu machen.

Weiterhin kann sich eine Kontrolle zur Divergenz gemäß $\|x^{(m+1)} - x^{(m)}\|_2^2 > 10^k \gg 1$ als sinnvoll erweisen.

Die Such- und Abstiegsrichtungen $r^{(m)}$ erfüllen eine Orthogonalitätsbedingung.
Es gilt für aufeinander folgende Richtungen $r^{(m+1)} \perp r^{(m)}$, denn mit (3.15) haben wir

$$\begin{aligned} (r^{(m+1)}, r^{(m)}) &= (r^{(m)}, r^{(m)}) - \alpha_m(Ar^{(m)}, r^{(m)}) \\ &= (r^{(m)}, r^{(m)}) - \frac{(r^{(m)}, r^{(m)})}{(Ar^{(m)}, r^{(m)})}\,(Ar^{(m)}, r^{(m)}) \\ &= 0. \end{aligned}$$

Die Orthogonalität ist aber nicht transitiv, so dass aus $r^{(m+1)} \perp r^{(m)}$ und $r^{(m)} \perp r^{(m-1)}$ nicht notwendigerweise $r^{(m+1)} \perp r^{(m-1)}$ folgt.

Varianten der Realisierung des GV als (unendliches) Iterationsverfahren

Version 1 GV mit Indizierung

$x^{(0)}$ Startvektor,

$r^{(0)} = b - Ax^{(0)}$ Anfangsresiduum, Abstiegsrichtung, negativer Gradient,

ε Toleranz für den Test auf Abbruch der Iteration.

m = 0,1,2,...

$$\begin{aligned} &\text{falls } \| r^{(m)} \| < \varepsilon, \quad \text{dann break} \\ \alpha_m &= \frac{r^{(m)T} r^{(m)}}{r^{(m)T} w^{(m)}}, \quad \text{wobei } \; w^{(m)} = Ar^{(m)} \\ x^{(m+1)} &= x^{(m)} + \alpha_m r^{(m)} \\ r^{(m+1)} &= r^{(m)} - \alpha_m w^{(m)} \quad \text{oder } \; r^{(m+1)} = b - Ax^{(m+1)} \end{aligned} \tag{3.16}$$

end m

Näherungslösung $x^* = x^{(m)}$

Version 2 GV mit Test auf relativen Fehler

K maximale Iterationsanzahl,

$x^{(0)}$ Startvektor,

$v^{(0)} = Ax^{(0)}$ Hilfsvektor,

$r^{(0)} = b - v^{(0)}$ Anfangsresiduum, Abstiegsrichtung, negativer Gradient,

$\gamma[0..K]$ Vektor der Normquadrate der euklidischen Norm, $\gamma_m = \| r^{(m)} \|_2^2$,

ε Toleranz für den Test auf Abbruch der Iteration,

falls $\gamma_0 < \varepsilon$, dann $x^* = x^{(0)}$ und Stopp.

m = 0,1,...,K

$$\begin{aligned} \alpha_m &= \frac{\gamma_m}{r^{(m)T} w^{(m)}}, \quad \text{wobei } \; w^{(m)} = Ar^{(m)} \\ x^{(m+1)} &= x^{(m)} + \alpha_m r^{(m)} \\ &\left\{ \begin{array}{lcl} v^{(m+1)} &=& v^{(m)} + \alpha_m w^{(m)} \\ r^{(m+1)} &=& b - v^{(m+1)} \end{array} \right\} \quad \text{oder } \; r^{(m+1)} = b - Ax^{(m+1)} \\ \gamma_{m+1} &= \| r^{(m+1)} \|_2^2, \quad \text{falls } \; \frac{\gamma_{m+1}}{\gamma_0} < \varepsilon, \quad \text{dann break} \end{aligned} \tag{3.17}$$

end m

Näherungslösung $x^* = x^{(m+1)}$

Bemerkung 3.1 Das Produkt $w^{(m)} = Ar^{(m)}$ tritt in den Formeln für α_m und $r^{(m+1)}$ auf und ist nur einmal zu berechnen.

Version 3 **GV ohne Indizierung**

K maximale Iterationsanzahl,
x Startvektor,
$v = Ax$ Hilfsvektor,
$r = b - v$ Anfangsresiduum, Abstiegsrichtung, negativer Gradient,
$\gamma[0..K]$ Vektor der Normquadrate der euklidischen Norm von r,
$\gamma_0 = r^T r = \| r \|_2^2$,
ε Toleranz für den Test auf Abbruch der Iteration,
falls $\gamma_0 < \varepsilon$, dann $x^* = x$ und Stopp.

m = 1,2,...,K

$$
\begin{aligned}
w &= Ar \\
\alpha &= \frac{\gamma_{m-1}}{d}, \quad \text{wobei} \quad d = r^T w = w^T r \\
x &= x + \alpha r \\
&\left\{ \begin{array}{lcl} v &=& v + \alpha w \\ r &=& b - v \end{array} \right\} \quad \text{oder} \quad r = b - Ax \quad \text{oder} \quad r = r - \alpha w \\
\gamma_m &= \| r \|_2^2 \\
&\text{falls} \quad \frac{\gamma_m}{\gamma_0} < \varepsilon, \quad \text{dann break}
\end{aligned}
\tag{3.18}
$$

end m

Näherungslösung $x^* = x$

Wenn A nur regulär ist, dann bildet man das Normalgleichungssystem $A^T Ax = Bx = c = A^T b$ mit $B = B^T > 0$, auf welches das GV angewendet werden kann. Von der Kondition der Koeffizientenmatrix her ist aber die Symmetrisierung des LGS nicht unbedingt der günstigste Weg, auch wenn sie die Matrix in gewisser Weise skaliert. Daraus ergibt sich das zu minimierende Funktional

$$
\begin{aligned}
R(x) &= \frac{1}{2} x^T Bx - x^T c = -\frac{1}{2}(Ax)^T (2b - Ax) \\
&= -\frac{1}{2}(Ax)^T s, \quad s = r + b = 2b - Ax = 2b - v, \\
&= \frac{1}{2}(\|r(x)\|_2^2 - b^T b), \quad r(x) = b - Ax.
\end{aligned}
$$

Die Matrix B hat im Allgemeinen eine schlechtere Kondition, so dass meistens mehr Gradientenschritte erforderlich sind.

Bei Implementierungen wird die Matrix $B = A^T A$ wegen des Aufwands von $2n^3$ Operationen nicht explizit ermittelt, sondern im GV hat man anstelle einer Matrix-Vektor-Multiplikation in der Hauptschleife nun zwei.

Version 4 **GV mit Symmetrisierung für beliebiges** A

K maximale Iterationsanzahl,
x Startvektor,
$v = A^T(Ax)$, $c = A^T b$ Hilfsvektoren,
$\tilde{r} = c - v$ Anfangsresiduum, Abstiegsrichtung, negativer Gradient,
$\gamma[0..K]$ Vektor der Normquadrate der euklidischen Norm von $\tilde{r}$,
$\gamma_0 = \tilde{r}^T\tilde{r} = \| \tilde{r} \|_2^2$,
ε Toleranz für den Test auf Abbruch der Iteration,
falls $\gamma_0 < \varepsilon$, dann $x^* = x$ und Stopp.

m = 1,2,...,K

$$w = A^T(A\tilde{r})$$
$$\alpha = \frac{\gamma_{m-1}}{d}, \quad \text{wobei} \quad d = \tilde{r}^T w = w^T\tilde{r}$$
$$x = x + \alpha\tilde{r}$$
$$\left\{ \begin{array}{lcl} v & = & v + \alpha w \\ \tilde{r} & = & c - v \end{array} \right\} \quad \text{oder} \quad \tilde{r} = c - A^T(Ax) \quad \text{oder} \quad \tilde{r} = \tilde{r} - \alpha w$$
$$\gamma_m = \| \tilde{r} \|_2^2, \quad \text{falls} \quad \frac{\gamma_m}{\gamma_0} < \varepsilon, \quad \text{dann break}$$

end m

Näherungslösung $x^* = x$

Da LGS mit Koeffizientenmatrizen gleich oder nahe der Einheitsmatrix der ideale Fall sind, betrachten wir noch kurz das GV mit Vorkonditierung (preconditioned gradient method). Mit einer einfachen regulären Vorkonditionierungsmatrix C (preconditioner) transformieren wir das LGS durch beidseitige Multiplikation gemäß

$$C^{-1}Ax = Bx = c = C^{-1}b. \tag{3.19}$$

Im GV selbst soll der Mehraufwand, der nun durch die zusätzliche Lösung eines Gleichungssystems entstehen wird, von der Größenordnung nicht den einer Matrix-Vektor-Multiplikation übersteigen. Das wäre dann der Fall, wenn z. B. die Matrix C in Produktform $C = LL^T > 0$ mit der Dreiecksmatrix L vorliegt und die Lösung des Gleichungssystems auf gestaffelte Systeme führt.

Die Lösung des Gleichungssystems $Cs = r$ erfolgt dann in zwei Schritten mit den Dreieckssystemen

$$Lz \;=\; r \;\text{ und }\; L^T s \;=\; z.$$

Version 5 **Vorkonditioniertes GV mit Indizierung**

$x^{(0)}$ Startvektor,
$r^{(0)} = b - Ax^{(0)}$ Anfangsresiduum, negativer Gradient,
ε Toleranz für Test auf Abbruch der Iteration,
C Vorkonditionierungsmatrix.

m = 0,1,2,...

falls $\| r^{(m)} \| < \varepsilon$, dann break

$$\begin{aligned} Cs^{(m)} &= r^{(m)} \\ w^{(m)} &= As^{(m)} \\ \alpha_m &= \frac{r^{(m)T} s^{(m)}}{w^{(m)T} s^{(m)}} \\ x^{(m+1)} &= x^{(m)} + \alpha_m s^{(m)} \\ r^{(m+1)} &= r^{(m)} - \alpha_m w^{(m)} \end{aligned}$$

end m

Näherungslösung $x^* = x^{(m)}$

Wie man zu den leicht modifizierten Darstellungen in Version 5 im Vergleich mit Version 1 gelangt, soll kurz erläutert werden.
Ausgangspunkt ist einfach das Funktional $Q(x + \alpha s)$ zum LGS (1.1) und die neue Suchrichtung

$$s \;=\; C^{-1} r, \quad r = b - Ax,$$

in der Strahlenminimierung.
Damit ergibt sich wie in Lemma 3.2 die Schrittzahl

$$\alpha \;=\; \frac{s^T r}{s^T A s} \;=\; \frac{r^T s}{(As)^T s},$$

so dass wir mit $w = As$ und $Cs = r$ auf die Beziehungen in Version 5 kommen.

Eine Implementierung des GV soll sich auf die Versionen 2 (3.17) und 3 (3.18) stützen. Dieser Algorithmus liefert in Maple die folgenden Kommandos.

```
# GV, Versionen 2,3
x:=evalm(x0):
v:=evalm(A&*x):
r:=evalm(b-v):
gamma0:=evalm(transpose(r)&*r):

if gamma0<etol then RETURN(x,0); end if;
gamman:=gamma0:

for k from 1 by 1 to maxiter do
 w:=evalm(A&*r);
 d:=evalm(transpose(r)&*w);

 alpha:=gamman/d;
 x:=evalm(x+alpha*r);
 v:=evalm(v+alpha*w);
 r:=evalm(b-v);
 gamman:=evalm(transpose(r)&*r);

 if gamman/gamma0<etol then break; end if;   # relativer Fehler
end do:

if k>maxiter then k:=k-1; end if;
[x,k];
```

Ergänzt um die zusätzliche Abfrage des Nenners bei der Berechnung der Schrittzahl α, die Bestimmung von Werten des Funktionals sowie einige Ergebnisfelder und Zwischenausgaben entsteht dann die erweiterte Maple-Prozedur in Anlehnung an diese Versionen des GV, die wir in den Beispielrechnungen verwenden.

```
> gv:=proc(n::posint,A::matrix,b::vector,x0::vector,
          maxiter::posint,etol::numeric,aus::name)

 local k,i,m,v,r,w,d,gamma0,gamman,x,alpha,Q,fh,fh1,fh2,xv1,rv1,xx;
 global xv,rv,lr;

x:=evalm(x0):
v:=evalm(A&*x):
r:=evalm(b-v):  lr:=evalm(r):
gamma0:=evalm(transpose(r)&*r):
Q:=0.5*evalm(transpose(x)&*A&*x)-evalm(transpose(x)&*b);

fh2:=`%+.16e`;  # Ausgabeformate einstellen
fh1:=`%+.16e`;
fh :=fh1;
for m from 2 to n do
 fh:=cat(fh,` `,fh1);
end do;
```

```
xx:=matrix(n,0,[]):
xv1:=concat(xx,x);
xv:=evalm(xv1);
rv1:=concat(xx,r);
rv:=evalm(rv1);

if aus=ja then
 printf(`\n`):
 printf(`Startvektor            x = [`||fh||`]\n`,seq(x[i],i=1..n));
 printf(`Residuum        r = b-Ax = [`||fh||`]\n`,seq(r[i],i=1..n));
 printf(`Funktionswert       Q(x) = `||fh1||`\n`,Q);
 printf(`Anfangsfehlerquadrat r'r = `||fh2||`\n\n`,gamma0);
end if;

if gamma0<etol then RETURN(x,0); end if;
gamman:=gamma0:

for k from 1 by 1 to maxiter do
 w:=evalm(A&*r);
 d:=evalm(transpose(r)&*w);
 if d=0 then
  lprint(`Abbruch wegen Nenner r'Ar=0`):
  RETURN(x,k-1);
 end if;

 alpha:=gamman/d;
 x:=evalm(x+alpha*r);
 v:=evalm(v+alpha*w);
 r:=evalm(b-v);   lr:=evalm(r);
 gamman:=evalm(transpose(r)&*r);

 Q:=0.5*evalm(transpose(x)&*A&*x)-evalm(transpose(x)&*b);
 xv1:=concat(xv,x);
 xv:=evalm(xv1);
 rv1:=concat(rv,r);
 rv:=evalm(rv1);

 if aus=ja then
  printf(`Schritt k = %g\n`,k);
  printf(`Suchrichtung    p=r=b-Ax = [`||fh||`]\n`,seq(r[i],i=1..n));
  printf(`Suchschritt        alpha = `||fh1||`\n`,alpha);
  printf(`Iterationsvektor       x = [`||fh||`]\n\n`,seq(x[i],i=1..n));
  printf(`Residuum        r = b-Ax = [`||fh||`]\n`,seq(r[i],i=1..n));
  printf(`Funktionswert       Q(x) = `||fh1||`\n`,Q);
  printf(`Fehlernormquadrat    r'r = `||fh2||`\n\n`,gamman);
 end if;

 if gamman/gamma0<etol then break; end if;   # relativer Fehler
end do:

if k>maxiter then k:=k-1; end if;
[x,k];

end:
```

3.2.1 Zur Konvergenz des Gradientenverfahrens

Um Konvergenzaussagen zum GV zu machen, sind einige vorbereitende Überlegungen zu Iterationsverfahren anzustellen.
Das LGS (1.1) kann man zunächst als Nullstellengleichung $0 = b - Ax$ notieren und daraus die Gestalt einer Fixpunktgleichung

$$x \;=\; x + b - Ax \;=\; (I - A)x + b, \quad I \text{ Einheitsmatrix}, \tag{3.20}$$

ableiten. Damit ergeben sich mehrere äquivalente Formen der Iterationsformel

$$\begin{aligned} x^{(m+1)} &= (I - A)x^{(m)} + b \\ &= x^{(m)} - I^{-1}(Ax^{(m)} - b) \\ &= x^{(m)} + r^{(m)}, \quad r^{(m)} = r(x^{(m)}) = b - Ax^{(m)} \quad \text{(Residualform)}, \end{aligned} \tag{3.21}$$

mit

$$\begin{aligned} H &= I - A \quad \text{als Iterationsmatrix}, \\ W &= I \qquad \text{als Wichtung, Skalierungsmatrix}. \end{aligned}$$

Diese Umformungen sind jedoch trivial, so dass wir es als naives Iterationsverfahren bezeichnen wollen.
Eleganter und gebräuchlicher ist das Gesamtschrittverfahren (GSV, Jacobi-Iterationsverfahren), das unter der Voraussetzung $a_{ii} \neq 0,\ i = 1, 2, ..., n$, in der Matrix-Vektor-Form

$$\begin{aligned} x^{(m+1)} &= (I - D^{-1}A)x^{(m)} + D^{-1}b, \quad D = \operatorname{diag}(A) = \operatorname{diag}(a_{11}, a_{22}, ..., a_{nn}) \\ &= x^{(m)} - D^{-1}(Ax^{(m)} - b) \\ &= x^{(m)} + D^{-1}r^{(m)} \end{aligned} \tag{3.22}$$

notiert werden kann, wobei

$$\begin{aligned} H = J &= I - D^{-1}A \quad \text{Iterationsmatrix}, \\ W &= D \qquad\quad \text{Wichtung, Skalierungsmatrix}, \\ & D^{-1}r^{(m)} \quad \text{gewichtetes Residuum (weighted residual)}, \\ & \qquad\quad \text{Verbesserung (update) für } x^{(m)}. \end{aligned}$$

Auf diese Art und Weise können weitere Basisverfahren definiert werden.
Das allgemeine Iterationsverfahren

$$x^{(m+1)} \;=\; Hx^{(m)} + c \tag{3.23}$$

konvergiert für jede beliebige Startnäherung $x^{(0)}$ gegen die Lösung x^*, wenn $\|H\| < 1$ gilt. Hinreichendes und notwendiges Konvergenzkriterium ist $\rho(H) < 1$. Je kleiner die Norm oder der Spektralradius der Iterationsmatrix H sind, desto schneller konvergiert das Verfahren. Dies kann man mit einer günstigen Wahl der Skalierungsmatrix eventuell bewirken.

Hier wollen wir jedoch die Wichtung möglichst einfach wählen.
Somit liegt nahe, mit der parameterabhängigen Matrix $W = \frac{1}{\omega} I,\ \omega > 0$, zu arbeiten. Den Parameter ω nennt man Relaxationsparameter. Die Vorgehensweise beschreibt die so genannte Richardson-Iteration (auch semiiterative Richardson-Methode oder Richardson-Relaxation genannt) mit festem Parameter (RF)

$$\begin{aligned} x^{(m+1)} &= x^{(m)} - \Big(\frac{1}{\omega} I\Big)^{-1} (Ax^{(m)} - b) \\ &= (I - \omega A) x^{(m)} + \omega b \\ &= x^{(m)} + \omega r^{(m)}, \quad r^{(m)} = b - Ax^{(m)}, \end{aligned} \tag{3.24}$$

wobei

$$\begin{aligned} H &= I - \omega A \quad \text{Iterationsmatrix,} \\ W &= \frac{1}{\omega} I \qquad \text{Wichtung, Skalierungsmatrix.} \end{aligned}$$

Bei geeigneter Wahl dieses Relaxationsparameters wird das Iterationsverfahren konvergieren und möglicherweise besonders schnell.
Unter der Voraussetzung $A = A^T > 0$ untersuchen wir die symmetrische Iterationsmatrix $H = I - \omega A$, speziell das Verhalten ihrer Eigenwerte.
Wir gehen von den reellen Eigenwerten der spd Matrix A aus, die der Bedingung (Ungleichungskette) $0 < \lambda = \lambda_1(A) \le \lambda_2(A) \le ... \le \lambda_n(A) = \Lambda = \rho(A)$ genügen. Weiter ist die Spektralnorm $\|A\|_2 = \rho(A)$.
Die Eigenwerte von $H = H(\omega)$ sind dann

$$\lambda(H(\omega)) \ = \ 1 - \omega\, \lambda(A). \tag{3.25}$$

Für ihr Verhalten in Abhängigkeit vom Parameter $\omega \ge 0$ braucht man nur die beiden "äußeren" Eigenwerte zu betrachten. Es gilt

$$\lambda_n(H(\omega)) = 1 - \omega\lambda_n(A) \le \lambda_i(H(\omega)) \le \lambda_1(H(\omega)) = 1 - \omega\lambda_1(A) \le 1,\ i = 2, 3, ..., n-1.$$

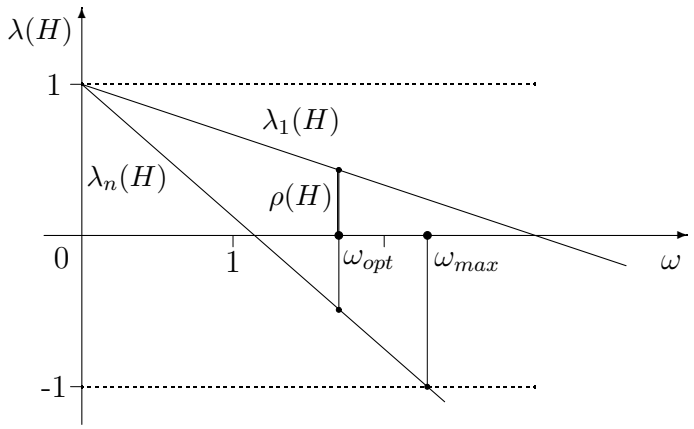


Abb. 3.3 Datei *abst2.pic*
Verlauf der Funktionen
$\lambda_i(H) = 1 - \omega\lambda_i(A),\ i = 1, n,$
$H = H(\omega)$,
in Abhängigkeit
vom Parameter $\omega \ge 0$

Die Abschätzung der Eigenwerte $\lambda(H(\omega))$ hat sowohl das Ziel, den Spektralradius $\rho(H(\omega)) < 1$ zu machen, als auch noch möglichst klein durch eine günstige Wahl des Parameters ω.
Dazu genügt es das Verhalten des Eigenwerts $\lambda_n(H(\omega)) = 1 - \omega\lambda_n(A)$ als Geradengleichung in Abhängigkeit von $\omega \geq 0$ bei abfallenden Verlauf der Geraden für den ersten Grenzfall $\lambda_n(H(\omega)) = -1$ zu untersuchen. Daraus erhält man

$$\omega_{max} \;=\; \frac{2}{\lambda_n(A)} \;=\; \frac{2}{\Lambda} \;=\; \frac{2}{\rho(A)} \tag{3.26}$$

und die Folgerung $\rho(H(\omega)) < 1$ für $\omega \in (0, \omega_{max})$.
Der Spektralradius $\rho(H(\omega))$ wir jedoch am kleinsten im zweiten Grenzfall $-\lambda_n(H(\omega)) = \lambda_1(H(\omega))$, was auf den optimalen Parameter

$$\omega_{opt} \;=\; \frac{2}{\lambda_1(A) + \lambda_n(A)} \;=\; \frac{2}{\lambda + \Lambda} \;<\; \omega_{max} \tag{3.27}$$

führt.
Ist $\lambda_1(A) \approx 0$, so liegt ω_{opt} nahe bei ω_{max}.
Spezielle Werte von ω sind z. B. $\omega = \frac{1}{\lambda_n(A)} = \frac{1}{\rho(A)} < \omega_{opt}$ oder $\omega = \frac{1}{\|A\|} \leq \frac{1}{\|A\|_2} = \frac{1}{\rho(A)}$.
Bei optimaler Situation der Konvergenz beträgt der Spektralradius der Iterationsmatrix

$$\eta \;=\; \rho(H(\omega_{opt})) = 1 - \omega_{opt}\,\lambda_1(A) = \frac{\Lambda - \lambda}{\Lambda + \lambda} = \frac{\dfrac{\Lambda}{\lambda} - 1}{\dfrac{\Lambda}{\lambda} + 1} = \frac{\kappa(A) - 1}{\kappa(A) + 1} < 1. \tag{3.28}$$

Das Ergebnis fassen wir als Satz zusammen.

Satz 3.3 *Falls $A = A^T > 0$ gilt, ist die Richardson-Iteration (3.24) für einen bestimmten Parameterbereich $\omega \in (0, \omega_{max})$, $\omega_{max} > 0$, konvergent. Unter diesen gibt es einen optimalen Wert $\omega = \omega_{opt}$.*

Unter den gleichen Voraussetzungen arbeiten das GV (3.13) mit variablem Parameter

$$x^{(m+1)} \;=\; x^{(m)} + \alpha_m r^{(m)}$$

und die Richardson-Iteration (3.24) mit festem Parameter

$$x^{(m+1)} \;=\; x^{(m)} + \omega_{opt} r^{(m)} \;=\; (I - \omega_{opt} A)x^{(m)} + \omega_{opt} b \;=\; Hx^{(m)} + c \tag{3.29}$$

mit derselben Struktur der Iterationsformel sowie jeweils optimaler Wahl der Schrittzahl in der Abstiegsrichtung $r^{(m)} = b - Ax^{(m)}$. Somit kann das GV als spezielle Richardson-Iteration mit variablem Parameter interpretiert werden und es entstehen vergleichbare Iterationsfolgen. Damit übertragen sich Konvergenzaussagen, Fehleranalyse und Abschätzungen zur Richardson-Iteration (3.29) auf das GV.

Satz 3.4 *Für den Fehlervektor $e^{(m)}$ des GV (3.13) mit (3.11), (3.12) gilt bei dem Faktor η aus Formel (3.28) die Abschätzung*

$$\|e^{(m)}\|_A \;\leq\; \eta^m \,\|e^{(0)}\|_A. \tag{3.30}$$

Beweis.
Da der Startvektor $x^{(0)} \in \mathbb{R}^n$ beliebig ist, reicht der Nachweis der Abschätzung (3.30) für $m = 1$. Sei also $x^{(1)} = x^{(0)} + \alpha_0 r^{(0)}$.
Wir haben im RF mit optimalem Gewichtsparameter ω_{opt} die Größen $H = H(\omega_{opt})$, $\|H\|_2 = \rho(H) = \eta$, $x^{(1)}_{RF} = x^{(0)} + \omega_{opt} r^{(0)}$ und den Fehlervektor

$$x^* - x^{(1)}_{RF} \;=\; e^{(1)}_{RF} \;=\; He^{(0)}.$$

Zudem gilt für beliebiges $q \in \mathbb{R}$ die Gleichheit $HA^q = A^q H$ und es folgt mit

$$\tilde{e}^{(0)} \;=\; A^{1/2} e^{(0)}, \quad \tilde{e}^{(1)}_{RF} \;=\; A^{1/2} e^{(1)}_{RF}$$

die Beziehung

$$\tilde{e}^{(1)}_{RF} \;=\; A^{1/2} e^{(1)}_{RF} \;=\; A^{1/2} H e^{(0)} \;=\; H A^{1/2} e^{(0)} \;=\; H\tilde{e}^{(0)}.$$

Hiermit erhält man

$$\begin{aligned}
\|e^{(1)}_{RF}\|_A \;&=\; (Ae^{(1)}_{RF}, e^{(1)}_{RF})^{1/2} \;=\; (A^{1/2}e^{(1)}_{RF}, A^{1/2}e^{(1)}_{RF})^{1/2} \\
&=\; \|\tilde{e}^{(1)}_{RF}\|_2 \;=\; \|H\tilde{e}^{(0)}\|_2 \;\leq \\
&\leq\; \|H\|_2 \,\|\tilde{e}^{(0)}\|_2 \;=\; \|H\|_2 \,\|A^{1/2}e^{(0)}\|_2 \\
&=\; \|H\|_2 \,\|e^{(0)}\|_A \;=\; \eta \,\|e^{(0)}\|_A.
\end{aligned}$$

Wegen

$$x^{(1)} \;=\; \underset{x \in x^{(0)} + \mathrm{span}\{r^{(0)}\}}{\arg\min}\; Q(x) \;=\; \underset{x \in x^{(0)} + \mathrm{span}\{r^{(0)}\}}{\arg\min} \left(\frac{1}{2}\|x^* - x\|_A^2 - \frac{1}{2} x^{*T} b \right)$$

im GV folgt

$$\|x^* - x^{(1)}\|_A \;\leq\; \|x^* - x^{(1)}_{RF}\|_A \;=\; \|e^{(1)}_{RF}\|_A$$

und

$$\|e^{(1)}\|_A \;\leq\; \eta \,\|e^{(0)}\|_A.$$

Das GV ist also lokal besser als das optimale RF, so dass für den m-ten Fehlervektor $e^{(m)}$ die Schranke η^m auftritt. □

Aus (3.4) und der Abschätzung (3.30) ergeben sich auch

$$\begin{aligned} \|e^{(m)}\|_A^2 &\leq \eta^{2m}\|e^{(0)}\|_A^2, \\ Q(x^{(m)}) - Q(x^*) &\leq \eta^{2m}\,[Q(x^{(0)}) - Q(x^*)], \\ Q(x^{(m)}) &\leq \eta^2\,Q(x^{(m-1)}) + (1-\eta^2)Q(x^*) \leq \eta^2\,Q(x^{(m-1)}) \leq \eta^{2m}\,Q(x^{(0)}), \end{aligned}$$

wobei die letzte Ungleichung nur für $(Q(x^*) =)\, 0 \leq Q(x^{(m)}) \leq Q(x^{(0)})$ brauchbar ist. Damit erkennt man jedoch allgemein die lineare Konvergenz des GV.

Wir wollen noch einige Bemerkungen zum Iterationsfehler von (3.29) $e_{RF}^{(m)} = x^* - x_{RF}^{(m)}$ machen, woraus man auch auf das Verhalten des GV dann schließen kann. Er genügt mit $H = H(\omega_{opt})$ der Beziehung

$$e_{RF}^{(m)} \;=\; x^* - x_{RF}^{(m)} \;=\; Hx^* + c - (Hx_{RF}^{(m-1)} + c) \;=\; He_{RF}^{(m-1)} \;=\; H^m e^{(0)}. \qquad (3.31)$$

Die symmetrische Matrix H hat n linear unabhängige Eigenvektoren $u_1, u_2, ..., u_n$ (Basis in $\mathbb{R}^n$) mit den Eigenwerten $\lambda_1(H), \lambda_2(H), ..., \lambda_n(H)$, wobei $1 > \eta = \rho(H) = \lambda_1(H) = |\lambda_1(H)| \geq |\lambda_i(H)|,\ i = 2, 3, ..., n$.
Der Ausgangsfehler lässt sich in der Basis $\{u_i\}$ darstellen, und die weiteren Fehlervektoren genügen den folgenden Formeln.

$$\begin{aligned} e^{(0)} &= \sum_{i=1}^{n} c_i u_i, \\ e_{RF}^{(1)} &= He^{(0)} \;=\; H\sum_{i=1}^{n} c_i u_i \;=\; \sum_{i=1}^{n} c_i \lambda_i(H) u_i, \\ e_{RF}^{(m)} &= He_{RF}^{(m-1)} \;=\; H^m e^{(0)} \;=\; \sum_{i=1}^{n} c_i [\lambda_i(H)]^m u_i. \end{aligned} \qquad (3.32)$$

Die Konvergenz $e_{RF}^{(m)} \to 0$ bei $m \to \infty$ und beliebigem Startvektor folgt aus den Ungleichungen $|\lambda_i(H)| < 1$ für alle i bzw. aus $\eta < 1$. Für großes m dominiert dann in (3.32) der Fehlerterm $c_1[\lambda_1(H)]^m u_1$.
Für beliebige kompatible Normen folgt aus (3.32) die Abschätzung

$$\|e_{RF}^{(m)}\| \;=\; \|H^m e^{(0)}\| \;\leq\; \|H^m\|\,\|e^{(0)}\| \;\leq\; \|H\|^m\,\|e^{(0)}\|. \qquad (3.33)$$

Bezüglich der Norm folgt ebenfalls die Konvergenz unter Verwendung des folgenden Lemmas.

Lemma 3.5 *Zu einer gegebenen Matrix C mit ihrem Spektralradius $\rho(C)$ gibt es für ein beliebiges $\varepsilon > 0$ eine Matrixnorm $\|\cdot\|$ mit $\|C\| \leq \rho(C) + \varepsilon$*

Beweis. Satz 2.52 in [18] zu Matrixnorm und Spektralradius.

Das heißt insbesondere, wenn wir von $\eta = \rho(H) < 1$ ausgehen, dann gilt auch für die Norm die Ungleichung $\|H\| < 1$, weil ja ε beliebig klein sein kann.

Damit ist wegen (3.33) und $\lim\limits_{m\to\infty} \|H\|^m = 0$ die Beziehung

$$\lim_{m\to\infty} \|e_{RF}^{(m)}\| \;=\; 0$$

für alle $e^{(0)}$ erfüllt. Mit den Normeigenschaften bedeutet das

$$\lim_{m\to\infty} x_{RF}^{(m)} \;=\; x^*.$$

Für die symmetrische Iterationsmatrix H kann die in Lemma 3.5 benötigte Norm leicht angegeben werden. Es ist nämlich die Spektralnorm $\|\cdot\|_2$, denn

$$\|H\|_2 = \sqrt{\rho(H^T H)} = \sqrt{\rho(H^2)} = \sqrt{\rho(H)^2} = \eta < \eta + \varepsilon.$$

Für $A = A^T > 0$ möchten wir noch eine Fehlerabschätzung in der A-Norm $\|\cdot\|_A$ erhalten, weil die Minimierung des Funktionals $Q(x)$ auch die von $\|e(x)\|_A$ bedeutet. Dazu brauchen wir noch eine weitere Hilfsaussage zur Matrix $A^{1/2}$.

Lemma 3.6 *Sei A symmetrisch und positiv definit. Dann gibt es eine eindeutige Matrix F, ebenfalls symmetrisch und positiv definit, mit $F^2 = A$.*

Beweis. Satz 2.18 in [18] allgemein für A hermitesch.

Damit gelingt uns der Nachweis der Verträglichkeit mit der A-Norm.
Wegen

$$\begin{aligned} e_{RF}^{(m+1)} &= H e_{RF}^{(m)}, \\ \|e_{RF}^{(m+1)}\|_2 &\le \|H\|_2 \, \|e_{RF}^{(m)}\|_2 \;=\; \eta \, \|e_{RF}^{(m)}\|_2 \end{aligned} \tag{3.34}$$

und $\|A^{1/2}x\|_2^2 = (A^{1/2}x, A^{1/2}x) = (Ax, x) = (x, x)_A = \|x\|_A^2$ erhält man für die Vektoren $\tilde{e}_{RF}^{(m)} = A^{1/2} e_{RF}^{(m)}$ die Beziehungen

$$\begin{aligned} \tilde{e}_{RF}^{(m+1)} &= A^{1/2} e_{RF}^{(m+1)} \;=\; A^{1/2} H e_{RF}^{(m)} \;=\; A^{1/2}(I - \omega_{opt} A) e_{RF}^{(m)} \\ &= (A^{1/2} - \omega_{opt} A^{1/2} A) e_{RF}^{(m)} \;=\; (A^{1/2} - \omega_{opt} A A^{1/2}) e_{RF}^{(m)} \\ &= (I - \omega_{opt} A) A^{1/2} e_{RF}^{(m)} \;=\; H \tilde{e}_{RF}^{(m)}, \end{aligned}$$

$$\|\tilde{e}_{RF}^{(m+1)}\|_2 \;\le\; \eta \, \|\tilde{e}_{RF}^{(m)}\|_2,$$

und somit in verschiedenen Varianten die Fehlerschätzungen bzw. Abschätzungen

$$\begin{aligned} \|A^{1/2} e_{RF}^{(m+1)}\|_2 &\le \eta \, \|A^{1/2} e_{RF}^{(m)}\|_2, \\ \|e_{RF}^{(m+1)}\|_A &\le \eta \, \|e_{RF}^{(m)}\|_A, \\ \|e_{RF}^{(m)}\|_A &\le \eta^m \, \|e^{(0)}\|_A, \\ \|e_{RF}^{(m)}\|_A^2 &\le \eta^{2m} \, \|e^{(0)}\|_A^2. \end{aligned}$$

Die Abschätzung $\|e_{RF}^{(m+1)}\|_2 \leq \eta \, \|e_{RF}^{(m)}\|_2$ aus (3.34) ist scharf.
Man kann also eine Situation konstruieren, wo die Gleichheit zutrifft. Sei

v_1 Eigenvektor von A zu Λ mit $\|v_1\|_2 = \lambda$,
v_2 Eigenvektor von A zu λ mit $\|v_2\|_2 = \Lambda$

und

$$e^{(0)} \;=\; x^* - x^{(0)} \;=\; v_1 \pm v_2.$$

Dann erhält man mit (3.28) für die weiteren Fehler

$$\begin{aligned}
e_{RF}^{(1)} &= He^{(0)} \;=\; (I - \omega_{opt} A)e^{(0)} \;=\; \Big(I - \frac{2}{\Lambda + \lambda} A\Big)e^{(0)} \\
&= \Big(I - \frac{2}{\Lambda + \lambda} A\Big)(v_1 \pm v_2) \\
&= \Big(I - \frac{2}{\Lambda + \lambda} A\Big)v_1 \pm \Big(I - \frac{2}{\Lambda + \lambda} A\Big)v_2 \\
&= \Big(1 - \frac{2}{\Lambda + \lambda}\Lambda\Big)v_1 \pm \Big(1 - \frac{2}{\Lambda + \lambda}\lambda\Big)v_2 \\
&= \frac{-\Lambda + \lambda}{\Lambda + \lambda} v_1 \pm \frac{\Lambda - \lambda}{\Lambda + \lambda} v_2 \\
&= -\eta v_1 \pm \eta v_2, \\
e_{RF}^{(2)} &= He_{RF}^{(1)} \;=\; -\eta H v_1 \pm \eta H v_2 \\
&= -\eta(-\eta v_1) \pm \eta\eta v_2 \;=\; \eta^2(v_1 \pm v_2) \;=\; \eta^2 e^{(0)}, \\
e_{RF}^{(2k)} &= \eta^{2k} \, e^{(0)}, \\
\|e_{RF}^{(2k)}\|_2 &= \eta^{2k} \, \|e^{(0)}\|_2.
\end{aligned}$$

Das trifft auch auf die A-Norm zu. Wir haben sowohl in der euklidischen Norm als auch in der A-Norm die Fehlerabschätzung

$$\|e_{RF}^{(m)}\| \;\leq\; \eta^m \, \|e^{(0)}\| \text{ mit dem Konvergenzfaktor } \eta = \rho(H) = \frac{\Lambda - \lambda}{\Lambda + \lambda} < 1. \quad (3.35)$$

Bemerkung 3.2 Einige Bemerkungen zur Konvergenzgeschwindigkeit und Konvergenzrate.

(1) Mit der Ungleichung $\|e_{RF}^{(m)}\| \leq \|H^m\| \, \|e^{(0)}\|$ und der Matrixnormdefinition erhält man

$$\frac{\|e_{RF}^{(m)}\|}{\|e^{(0)}\|} \;\leq\; \|H^m\| \;=\; \sup_{x \neq 0} \frac{\|H^m x\|}{\|x\|}.$$

$\|H^m\|$ ist somit ein Maß für die Verringerung des Fehlers nach m Iterationen.
Wegen $\lim\limits_{m\to\infty} \|H^m\| = 0$ gibt es einen Wert m, so dass eine gegebene Toleranz $\varepsilon > 0$ erreicht wird, d. h.

$$\begin{aligned} \|H^m\| &\leq \varepsilon, \\ m\ln(\|H^m\|) &\leq m\ln(\varepsilon), \\ m &\geq \frac{-\ln(\varepsilon)}{-m^{-1}\ln(\|H^m\|)}. \end{aligned}$$

Die Größe im Nenner

$$R_m(H) \;=\; -\frac{1}{m}\ln(\|H^m\|) \;=\; -\ln(\sqrt[m]{\|H^m\|})$$

heißt mittlere Konvergenzrate (average rate of convergence), die näherungsweise durch

$$\widetilde{R}_m(H) \;=\; -\ln(\|H\|), \quad \|H\| < 1,$$

abgeschätzt werden kann.
Auf die Schätzung für m mit $\widetilde{R}_m(H)$ kommt man auch über $\|H^m\| \leq \|H\|^m \leq \varepsilon$ und der Umstellung nach m.

(2) Wenn man die allgemein gültige Beziehung

$$\eta \;=\; \rho(H) \;=\; \lim_{m\to\infty} \sqrt[m]{\|H^m\|} \tag{3.36}$$

(siehe Satz 2.52 in [18]) verwendet, kommt man im Grenzübergang auf die asymptotische Konvergenzrate, auch einfach Konvergenzrate genannt (asymptotic average rate of convergence).

$$R(H) \;=\; \lim_{m\to\infty} R_m(H) \;=\; -\ln(\rho(H)) \;>\; 0, \tag{3.37}$$

die mit fallendem Spektralradius wächst.
Die Anzahl der erforderlichen Iterationen zur Erreichung der Genauigkeit ε beträgt dann bei Rundung auf die nächst größere ganze Zahl

$$m \;=\; \Big[\frac{-\ln(\varepsilon)}{R(H)} + 1\Big]. \tag{3.38}$$

Je kleiner ε ist, desto mehr Iterationen sind auszuführen. Je größer die Konvergenzrate ist, desto weniger Iterationen werden gebraucht.
Im Sonderfall $\eta^m = 10^{-k}$ ist $m \approx k/(-\log_{10}(\eta))$ und m wächst linear mit der geforderten Genauigkeit.

(3) In der Abschätzung (3.35) mit dem Spektralradius kann man auch die Matrixkondition einbeziehen.
Es gilt in der Spektralnorm mit (2.4) und

$$\kappa \;=\; \kappa(A) \;=\; \mathrm{cond}_2(A) \;=\; \frac{\max\lambda(A)}{\min\lambda(A)} \;=\; \frac{\Lambda}{\lambda} \tag{3.39}$$

die Formel

$$\eta \;=\; \frac{\Lambda-\lambda}{\Lambda+\lambda} \;=\; \frac{\frac{\Lambda}{\lambda}-1}{\frac{\Lambda}{\lambda}+1} \;=\; \frac{\kappa-1}{\kappa+1} < 1. \tag{3.40}$$

(4) Mit der Beziehung (3.40) drückt man die Iterationsanzahl m in Abhängigkeit von der Kondition aus. Unter Verwendung von

$$-\ln\Big(\frac{z-1}{z+1}\Big) \;=\; \ln\Big(\frac{z+1}{z-1}\Big) \;=\; 2\Big(\frac{1}{z}+\frac{1}{3z^3}+\frac{1}{5z^5}+...\Big),\quad |z|>1,$$

erhält man ähnlich zu (3.38) die Schätzung

$$m \;\approx\; \frac{-\ln(\varepsilon)}{-\ln(\eta)} \;=\; \frac{-\ln(\varepsilon)}{-\ln\Big(\frac{\kappa-1}{\kappa+1}\Big)} \;\approx\; \frac{\kappa}{2}\ln\Big(\frac{1}{\varepsilon}\Big), \tag{3.41}$$

die für κ nahe Eins natürlich zu grob ausfallen kann.
Aber man sieht in den Beziehungen, dass eine Verschlechterung der Kondition von A mehr Iterationen erforderlich macht.

Als eine wichtige Erkenntnis aus den Bemerkungen nehmen wir mit, dass in einem AV nicht nur die Bestimmung einer optimalen Schrittzahl α in der Suchrichtung wichtig ist, sondern auch eine Verbesserung der Kondition der Matrix des LGS anzustreben ist.
Die Verbesserung der Kondition haben wir im vorkonditionierten GV (Version 5) durch den Zugang (3.19) und somit durch eine sich ergebende Veränderung der Suchrichtung beschrieben.
Für die Vorkonditionierungsmatrix C fordert man im Allgemeinen die Bedingung der Spektraläquivalenz

$$0 \;<\; \gamma C \;\le\; A \;\le\; \Gamma C, \quad \text{wobei } 0<\gamma\le\Gamma. \tag{3.42}$$

Das führt für die Matrix $B = C^{-1}A$ auf die Konditionszahl

$$\kappa' \;=\; \kappa(B) \;=\; \frac{\Gamma}{\gamma}$$

und wegen

$$\kappa' \;=\; \frac{\Gamma}{\gamma} \;\le\; \frac{\Lambda}{\lambda} \;=\; \kappa$$

ist dann

$$\eta' \;=\; \frac{\kappa'-1}{\kappa'+1} \;\leq\; \frac{\kappa-1}{\kappa+1} \;=\; \eta \;<\; 1.$$

Damit erhält man in der Fehlerschätzung

$$\|e^{(m)}\|_A \;\leq\; {\eta'}^m \, \|e^{(0)}\|_A \tag{3.43}$$

einen kleineren Konvergenzfaktor.
Bezüglich detaillierter Konvergenzaussagen verweisen wir auf [8].

Zusammenfassung und Bewertung der Eigenschaften des GV
in der Skala $\{+, \pm, -\}$ unter der Voraussetzung $A = A^T > 0$

(3.44)

- \+ Das Abstiegsszenario und die Minimierungsaufgabe sind dem Problem angepasst.
- \+ Einfache Implementation des Algorithmus.
- $\pm$ Hauptaufwand in einem Iterationsschritt ist eine Matrix-Vektor-Multiplikation.
- $\pm$ Der Lösungsfehler $\|e^{(m)}\|_A = \|x^* - x^{(m)}\|_A$ wird ständig verkleinert, aber nicht unbedingt $\|r^{(m)}\|_2 = \|b - Ax^{(m)}\|_2$.
  Für ein Abbruchkriterium kann man natürlich die Größe $\|e^{(m)}\|_A$ wegen der darin enthaltenen unbekannten Lösung x^* nicht verwenden.
- $\pm$ Aufeinander folgende Residua $r^{(m+1)} \perp r^{(m)}$ sind orthogonal und damit linear unabhängig.
- $-$ Konvergenzverhalten
  Das GV ist i. Allg. ein unendliches Iterationsverfahren.
  Die Orthogonalität aufeinander folgender Residua $r^{(m+1)} \perp r^{(m)}$ bedeutet nicht unbedingt die lineare Unabhängigkeit von $r^{(m+1)}$ und $r^{(m-1)}$.
  Wünschenswert wäre z. B., dass das System der Suchrichtungen nur aus linear unabhängigen Vektoren besteht.

3.2.2 Krylov-Unterraum und Suchrichtungen

Aus linear unabhängigen Vektoren $x_i \in \mathbb{R}^n,\ i = 1, 2, ...,$ kann man $k \leq n$ Vektoren auswählen, die dann eine Basis $\mathrm{X} = \{x_1, x_2, ..., x_k\}$ des linearen Unterraums

$$\mathcal{X} \ = \ [\mathrm{X}] \ = \ \mathrm{span}\,\mathrm{X} \ = \ \mathrm{span}\{x_1, x_2, ..., x_k\} \subset \mathbb{R}^n, \tag{3.45}$$

bilden. Wir notieren hierbei verschiedene gebräuchliche Bezeichnungen des Unterraums. Seine Dimension $k \leq n$ kann durch einen zusätzlichen Index gekennzeichnet werden. Nimmt man die Vektoren x_i als Spalten einer Matrix, so erhält man

$$X \ = \ X(n, k) \ = \ (x_1, x_2, ..., x_k) \ = \ [x_1, x_2, ..., x_k]. \tag{3.46}$$

Definition 3.2 Krylov-Unterraum und Krylov-Matrix

Ein spezielles mittels der Matrix A und einem Vektor $x \neq 0$ generiertes Vektorsystem ist $\{x, Ax, ..., A^m x\},\ m \geq 0$.
Sind seine ersten k Vektoren linear unabhängig, so bezeichnet man den durch sie aufgespannten Raum als Krylov-Unterraum oder Krylov-Teilraum

$$\mathcal{K}_k \ = \ \mathcal{K}_k(A, x) \ = \ \mathrm{span}\{x, Ax, ..., A^{k-1}x\}. \tag{3.47}$$

Damit ergibt sich die so genannte Krylov-Matrix

$$K^{(k)} \ = \ (x, Ax, ..., A^{k-1}x). \tag{3.48}$$

Ist x ein Eigenvektor der Matrix A, d. h. $Ax = \lambda x,\ x \neq 0$, so erhält man wegen $A^m x = \lambda^m x,\ m = 1, 2, 3, ...,$ die einfache Situation mit den Krylov-Unterräumen $\mathcal{K}_1 = \mathcal{K}_2 = \mathcal{K}_3 = ... = \mathrm{span}\{x\}$.
Für eine singuläre Matrix macht die Konstruktion von $\mathcal{K}_k$ wenig Sinn.

Beispiel 3.1 Krylov-Unterräume für eine singuläre Matrix A

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

(1) $x = (1, 0, 0)^T$ Eigenvektor (EV) zu $\lambda = 0$

$\mathcal{K}_1 = \mathrm{span}\{x\}, \quad Ax = 0 = 0 \cdot x$

(2) $x = (0, 1, 0)^T$ kein EV

$\mathcal{K}_2 = \mathrm{span}\{x, Ax\}, \quad Ax = (1, 0, 0)^T,\ A^2 x = 0$

(3) $x = (0, 0, 1)^T$ kein EV

$\mathcal{K}_3 = \mathrm{span}\{x, Ax, A^2 x\}, \quad Ax = (1, 1, 0)^T,\ A^2 x = (1, 0, 0)^T,\ A^3 x = 0$

Für reguläre Matrizen ist zu prüfen, wie groß maximal die Dimension von $\mathcal{K}_k$ werden kann.

Beispiel 3.2 Krylov-Unterräume für eine reguläre Matrix A

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

(1) $x = (1,0,0)^T$ EV

$\mathcal{K}_1 = \text{span}\{x\}, \quad Ax = x = 1 \cdot x$

(2) $x = (1,1,0)^T$ kein EV

$\mathcal{K}_2 = \text{span}\{x, Ax\}, \quad Ax = (2,1,0)^T, \; A^m x = (c_1, c_2, 0)^T, \; m \geq 2, \; \mathcal{K}_3 = \mathcal{K}_2$

(3) $x = (0,0,1)^T$ kein EV

$\mathcal{K}_3 = \text{span}\{x, Ax, A^2x\}, \quad Ax = (1,1,1)^T, \; A^2x = (3,2,1)^T$

Die Kontrolle, ob bei Hinzunahme des nächsten Vektors $A^m x$ die Dimension des Vektorraums um Eins wächst, erfolgt in Anwendungen durch die Transformation des Vektorsystems auf eine orthogonale Basis $\{q_1, q_2, ..., q_m, q_{m+1}\}$, $q_i^T q_j = \delta_{ij}$, unter Verwendung z. B. des Orthogonalisierungsverfahrens von Gram-Schmidt. Kann der Vektor q_{m+1} nicht erzeugt werden, ist der maximale Krylov-Unterraum $\mathcal{K}_m(A, x)$.
Die Such- und Abstiegsrichtungen $r^{(m)}$ im GV (3.13) sind orthogonal, was zwei aufeinander folgende Richtungen betrifft. Das genügt jedoch nicht für ihre lineare Unabhängigkeit insgesamt. Es gilt jedoch der folgende Satz.

Satz 3.7 *Solange im GV die nicht verschwindenden Suchrichtungen $r^{(m)}$ linear unabhängig sind, gilt*

$$\text{span}\{r^{(0)}, r^{(1)}, ..., r^{(k)}\} \;=\; \text{span}\{r^{(0)}, Ar^{(0)}, ..., A^{k-1}r^{(0)}\} \;=\; \mathcal{K}_k(A, r^{(0)}). \quad (3.49)$$

Beweis.
Wir zeigen zunächst die Behauptung für $k = 1, 2$ unter Verwendung der rekursiven Berechnungsformeln der Suchrichtungen (3.15).
(1) $k = 1$
Wir haben $r^{(1)} \perp r^{(0)}$ und $\dim(\text{span}\{r^{(0)}, r^{(1)}\}) = 2$.
Aus $r^{(1)} = r^{(0)} - \alpha_0 Ar^{(0)}$ folgen unmittelbar die beiden Beziehungen mit reellen Koeffizienten

$$\begin{aligned} r^{(1)} &= \delta_0 r^{(0)} + \delta_1 Ar^{(0)} \in \text{span}\{r^{(0)}, Ar^{(0)}\}, \\ Ar^{(0)} &= \gamma_0 r^{(0)} + \gamma_1 r^{(1)} \in \text{span}\{r^{(0)}, r^{(1)}\}. \end{aligned}$$

(2) $k = 2$
Wir haben $r^{(2)} \perp r^{(1)} \perp r^{(0)}$, $r^{(2)} \not\parallel r^{(0)}$ und $\dim(\text{span}\{r^{(0)}, r^{(1)}, r^{(2)}\}) = 3$.
Aus $r^{(2)} = r^{(1)} - \alpha_1 A r^{(1)}$ und Teil (1) folgen die Beziehungen

$$\begin{aligned}
r^{(2)} &= r^{(1)} - \alpha_1 A(r^{(0)} - \alpha_0 A r^{(0)}) \\
&= r^{(1)} - \alpha_1 A r^{(0)} + \alpha_0\alpha_1 A^2 r^{(0)} \\
&= r^{(0)} - \alpha_0 A r^{(0)} - \alpha_1 A r^{(0)} + \alpha_0\alpha_1 A^2 r^{(0)} \\
&= r^{(0)} - (\alpha_0 + \alpha_1) A r^{(0)} + \alpha_0\alpha_1 A^2 r^{(0)} \\
&= \delta_0 r^{(0)} + \delta_1 A r^{(0)} + \delta_2 A^2 r^{(0)} \ \in \text{span}\{r^{(0)}, Ar^{(0)}, A^2 r^{(0)}\}, \\
A^2 r^{(0)} &= \frac{1}{\alpha_0\alpha_1}(\alpha_1 A r^{(0)} - r^{(1)} + r^{(2)}) \\
&= \frac{1}{\alpha_0\alpha_1}\Big(\alpha_1 \frac{1}{\alpha_0}(r^{(0)} - r^{(1)}) - r^{(1)} + r^{(2)}\Big) \\
&= \frac{1}{\alpha_0^2} r^{(0)} - \Big(\frac{1}{\alpha_0^2} + \frac{1}{\alpha_0\alpha_1}\Big) r^{(1)} + \frac{1}{\alpha_0\alpha_1} r^{(2)} \\
&= \gamma_0 r^{(0)} + \gamma_1 r^{(1)} + \gamma_2 r^{(2)} \ \in \text{span}\{r^{(0)}, r^{(1)}, r^{(2)}\}.
\end{aligned}$$

Diese Vorgehensweise ist auf beliebige k übertragbar. □

In der praktischen Durchführung des GV kann es sein, dass

$$r^{(m+1)} \ \in \ \text{span}\{r^{(0)}, r^{(1)}, ..., r^{(m)}\},$$

das bedeutet auch $r^{(m+1)} \not\perp \text{span}\{r^{(0)}, r^{(1)}, ..., r^{(m)}\}$, obwohl $r^{(m+1)} \perp r^{(m)}$.
Die Dimension des Unterraums der Suchrichtungen muss also nicht ständig wachsen, was natürlich konform geht mit der Unendlichkeit des GV.
Wenn $\dim(\text{span}\{r^{(0)}, r^{(1)}, ..., r^{(n-1)}\}) = n$ ist, müssen die n-te Suchrichtung $r^{(n)}$ und weitere im Vektorraum $\text{span}\{r^{(0)}, r^{(1)}, ..., r^{(n-1)}\} = \mathbb{R}^n$ liegen.
Aus $x^{(m+1)} = x^{(m)} + \alpha_m r^{(m)}$, $m = 0, 1, ...$, ergibt sich

$$x^{(m+1)} \ = \ x^{(0)} + \alpha_0 r^{(0)} + \alpha_1 r^{(1)} + ... + \alpha_m r^{(m)}$$

und somit

$$x^{(m+1)} \ \in \ x^{(0)} + \mathcal{R}_{m+1}, \quad \mathcal{R}_{m+1} = \text{span}\{r^{(0)}, r^{(1)}, ..., r^{(m)}\}. \tag{3.50}$$

Die m-te Iterierte $x^{(m)}$ im GV wird als dasjenige Element aus dem Unterraum $x^{(0)} + \text{span}\{r^{(0)}, r^{(1)}, ..., r^{(m-1)}\}$ bestimmt, für welches das Funktional $Q(x)$ seinen minimalen Wert annimmt.

3.2.3 Beispiele zum Gradientenverfahren

In einigen Beispielen illustrieren wir das GV für das LGS (1.1) unter der üblichen Voraussetzung $A = A^T > 0$.

Beispiel 3.3
Sei A eine Diagonalmatrix und $b = 0$.

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x^{(0)} = \begin{pmatrix} 9/2 \\ 3 \end{pmatrix}.$$

Die Funktionale sind

$$\begin{aligned} Q(x) &= \frac{1}{2}x_1^2 + x_2^2, \\ R(x) &= \frac{1}{2}x_1^2 + 2x_2^2. \end{aligned}$$

Am gemeinsamen eindeutigen Minimum an der Stelle $x^* = 0$ gilt $Q(x^*) = R(x^*) = 0$.

Als Suchrichtung und Abstiegsrichtung in einem Schritt nehmen wir die Richtung des steilsten Abstiegs $r(x) = -\nabla Q(x)$, die orthogonal zu den Höhenlinien $Q(x)$=const sind.
Im GV werden unendlich viele Schritte ausgeführt. Der Iterationsverlauf ist wie eine rechtwinklige "Zick-Zack"-Kurve.

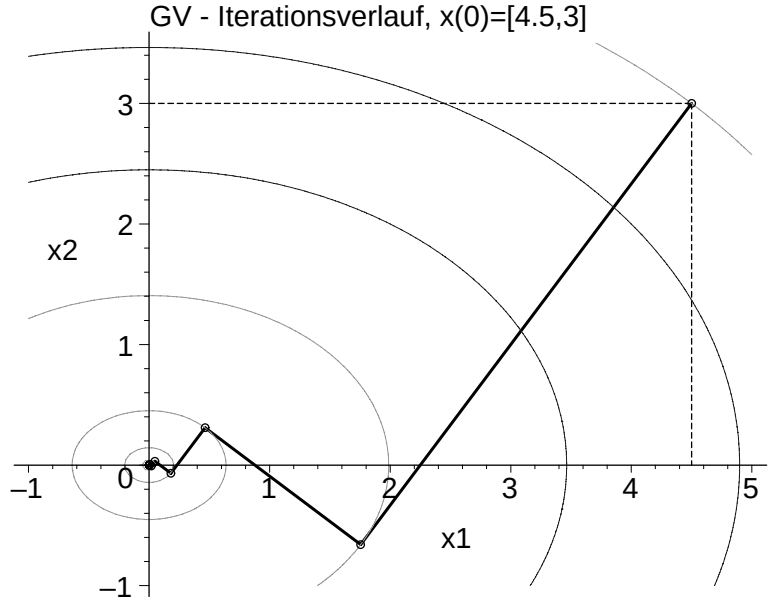


Abb. 3.4 Datei *abst*101.*ps*
Höhenlinienbild mit Iterationsverlauf des GV zu $Q(x) = \frac{1}{2}x_1^2 + x_2^2$ mit
`contours=[19.125,12,6,1.976,0.204,0.021,0.0022,0]`

Wir erhalten die Beziehungen

$$r^{(0)} = \Big(-\frac{9}{2}, -6\Big)^T \parallel (3,4)^T, \quad r^{(1)} = \Big(-\frac{72}{41}, \frac{54}{41}\Big)^T \parallel (-4,3)^T,$$

$$\alpha_0 = \frac{25}{41} = 0.609\,756..., \quad \alpha_1 = \frac{25}{34} = 0.735\,294...,$$

$$r^{(0)} \perp r^{(1)} \perp r^{(2)} \perp r^{(3)} \perp ...,$$

$$r^{(0)} \parallel r^{(2)} \parallel r^{(4)} \parallel ..., \quad r^{(1)} \parallel r^{(3)} \parallel r^{(5)} \parallel ...,$$

$$\mathcal{K}_2(A, r^{(0)}) = \text{span}\{r^{(0)}, r^{(1)}\} = \text{span}\{r^{(0)}, Ar^{(0)}\} = \text{span}\left\{\begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} -4 \\ 3 \end{pmatrix}\right\} = \mathbb{R}^2,$$

$$x^* \;=\; 0 \cdot r^{(0)} + 0 \cdot r^{(1)} \;\in\; \text{span}\{r^{(0)}, r^{(1)}\}.$$

Ergebnisse aus Berechnungen mit Maple

```
Startvektor            x = [+4.5000000000000000e+00 +3.0000000000000000e+00]
Residuum/SR p = r = b-Ax = [-4.5000000000000000e+00 -6.0000000000000000e+00]
Funktionswert       Q(x) =  +1.9125000000000000e+01
Anfangsfehlerquadrat r'r =   5.6250000000000000e+01
```

k	Schrittzahl alpha	Iterationsvektor x Residuum/neue Suchrichtung p=r=b-Ax	Funktionswert Q(x) Fehlernormquadrat r'r
1	6.0975609756097560e-01	[+1.7560975609756100e+00,-6.5853658536585400e-01] [-1.7560975609756100e+00,+1.3170731707317070e+00]	1.9756097560975620e+00 4.8185603807257580e+00
2	7.3529411764705890e-01	[+4.6484935437589700e-01,+3.0989956958393070e-01] [-4.6484935437589700e-01,-6.1979913916786200e-01]	2.0408020436014960e-01 6.0023589517691110e-01
3	6.0975609756097590e-01	[+1.8140462609791090e-01,-6.8026734786717000e-02] [-1.8140462609791090e-01,+1.3605346957343350e-01]	2.1081455830603760e-02 5.1418184952692060e-02
4	7.3529411764705830e-01	[+4.8018871614153000e-02,+3.2012581076101700e-02] [-4.8018871614153000e-02,-6.4025162152203900e-02]	2.1777113627022390e-03 6.4050334197125120e-03
5	6.0975609756097620e-01	[+1.8739071849425530e-02,-7.0271519435348600e-03] [-1.8739071849425530e-02,+1.4054303887069220e-02]	2.2495727132649200e-04 5.4867627152802140e-04
6	7.3529411764705750e-01	[+4.9603425483773700e-03,+3.3068950322513100e-03] [-4.9603425483773700e-03,-6.6137900645031300e-03]	2.3238053852949840e-05 6.8347217214563220e-05
7	6.0975609756097660e-01	[+1.9357434335131150e-03,-7.2590378756767800e-04] [-1.9357434335131150e-03,+1.4518075751348470e-03]	2.4004876289996720e-06 5.8548478756080670e-06
8	7.3529411764705690e-01	[+5.1240267357700500e-04,+3.4160178238441300e-04] [-5.1240267357700500e-04,-6.8320356476933400e-04]	2.4797002767263920e-07 7.2932361080238830e-07
9	6.0975609756097700e-01	[+1.9996201895687930e-04,-7.4985757109084300e-05] [-1.9996201895687930e-04,+1.4997151421766060e-04]	2.5615268281878260e-08 6.2476264102049340e-08
10	7.3529411764705700e-01	[+5.2931122665056600e-05 +3.5287415109783500e-05] [-5.2931122665056600e-05 -7.0574830220075000e-05]	2.6460535384218120e-09 7.7825104071756800e-09

Beispiel 3.4
Sei A eine Tridiagonalmatrix.

$$\begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ 2 \end{pmatrix}, \quad x^* = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad x^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Die Funktionale sind

$$\begin{aligned} Q(x) &= 2x_1^2 - x_1x_2 + 2x_2^2 - x_2x_3 + +2x_3^2 - 2x_1 - 6x_2 - 2x_3, \\ R(x) &= \frac{17}{2}x_1^2 - 8x_1x_2 + x_1x_3 + 9x_2^2 - 8x_2x_3 + \frac{17}{2}x_3^2 - 2x_1 - 20x_2 - 2x_3. \end{aligned}$$

Am gemeinsamen eindeutigen Minimum an der Stelle x^* gilt $Q(x^*) = -8$ und $R(x^*) = -22$.

Als Suchrichtung und Abstiegsrichtung in einem Schritt nehmen wir die Richtung des steilsten Abstiegs $r(x) = -\nabla Q(x)$, die orthogonal zu den Höhenlinien $Q(x)$=const sind.

Die ersten Schritte des GV sind wie folgt.

$x^{(0)} = (0,0,0)^T$ Startvektor,

$r^{(0)} = b - Ax^{(0)} = b = 2(1,3,1)^T$ Anfangsresiduum, Abstiegsrichtung.

S1
$$\begin{aligned} \alpha_0 &= \frac{\|r^{(0)}\|_2^2}{\|r^{(0)}\|_A^2} = \frac{11}{32} = 0.343\,75, \\ x^{(1)} &= x^{(0)} + \alpha_0 r^{(0)} = \frac{11}{16}(1,3,1)^T, \\ r^{(1)} &= b - Ax^{(1)} = r^{(0)} - \alpha_0 Ar^{(0)} = \frac{7}{16}(3,-2,3)^T. \end{aligned}$$

S2
$$\begin{aligned} \alpha_1 &= \frac{11}{56} = 0.196\,428..., \\ x^{(2)} &= \frac{121}{128}(1,2,1)^T, \quad r^{(2)} = \frac{7}{64}(1,3,1)^T. \end{aligned}$$

S3
$$\begin{aligned} \alpha_2 &= \frac{11}{32}, \\ x^{(3)} &= \frac{11}{2\,048}(183,373,183)^T, \quad r^{(3)} = \frac{49}{2\,048}(3,-2,3)^T. \end{aligned}$$

S4
$$\begin{aligned} \alpha_3 &= \frac{11}{56}, \\ x^{(4)} &= \frac{16\,335}{16\,384}(1,2,1)^T, \quad r^{(4)} = \frac{49}{8\,192}(1,3,1)^T. \end{aligned}$$

Startvektor $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, x_3^{(0)})^T$ und Iterierte $x^{(k)},\ k = 1, 2, ..., 10$

```
 k   [ x(k)[1],              x(k)[2],              x(k)[3]              ]
-------------------------------------------------------------------------
 0   [ 0,                   0,                    0                    ]
 1   [ 0.6875000000000000, 2.062500000000000, 0.6875000000000000]
 2   [ 0.9453125000000000, 1.890625000000000, 0.9453125000000000]
 3   [ 0.9829101562500000, 2.003417968750000, 0.9829101562500000]
 4   [ 0.9970092773437500, 1.994018554687500, 0.9970092773437500]
 5   [ 0.9990653991699219, 2.000186920166016, 0.9990653991699219]
 6   [ 0.9998364448547363, 1.999672889709473, 0.9998364448547363]
 7   [ 0.9999488890171049, 2.000010222196579, 0.9999488890171049]
 8   [ 0.9999910555779933, 1.999982111155987, 0.9999910555779933]
 9   [ 0.9999972048681227, 2.000000559026376, 0.9999972048681227]
10   [ 0.9999995108519213, 1.999999021703844, 0.9999995108519213]
```

Dazu erhalten wir die Beziehungen

$r^{(0)} \perp r^{(1)} \perp r^{(2)} \perp r^{(3)} \perp ...,$

$r^{(0)} \parallel r^{(2)} \parallel r^{(4)} \parallel ...,\quad r^{(1)} \parallel r^{(3)} \parallel r^{(5)} \parallel ...,$

$$\mathcal{K}_2(A, r^{(0)}) = \text{span}\{r^{(0)}, r^{(1)}\} = \text{span}\{r^{(0)}, Ar^{(0)}\} = \text{span}\left\{\begin{pmatrix}1\\3\\1\end{pmatrix}, \begin{pmatrix}3\\-2\\3\end{pmatrix}\right\} \subset \mathbb{R}^3,$$

$$x^* \ = \ \frac{28}{77}r^{(0)} + \frac{16}{77}r^{(1)} \ \in \ \text{span}\{r^{(0)}, r^{(1)}\}.$$

Das GV konvergiert nur langsam.
Bei den Abstiegsrichtungen gibt es keine und braucht man auch keine dritte linear unabhängige Richtung.

Beispiel 3.5
Sei $A = A^T > 0$.

$$\begin{pmatrix}2 & 1\\1 & 3\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}1\\2\end{pmatrix}, \quad x^* = \begin{pmatrix}1/5\\3/5\end{pmatrix}, \quad x^{(0)} = \begin{pmatrix}3/2\\1\end{pmatrix}.$$

Die Funktionale sind

$$\begin{aligned} Q(x) &= x_1^2 + x_1x_2 + \frac{3}{2}x_2^2 - x_1 - 2x_2, \\ R(x) &= \frac{5}{2}x_1^2 + 5x_1x_2 + 5x_2^2 - 4x_1 - 7x_2. \end{aligned}$$

Am gemeinsamen eindeutigen Minimum an der Stelle x^* gilt $Q(x^*) = -\frac{7}{10}$ und $R(x^*) = -\frac{5}{2}$.

Als Suchrichtung und Abstiegsrichtung in einem Schritt nehmen wir die Richtung des steilsten Abstiegs $r(x) = -\nabla Q(x)$, die orthogonal zu den Höhenlinien $Q(x)$=const sind, aber $r(x) \not\perp R(x)$=const.

Die Eigenwerte von A sind

$$\lambda_{1,2} \;=\; \frac{5 \pm \sqrt{5}}{2} = 3.618\,033\,988...,\; 1.381\,966\,011...,$$

woraus die Kondition

$$\kappa \;=\; \frac{\lambda_1}{\lambda_2} = \frac{5+\sqrt{5}}{5-\sqrt{5}} = 2.618\,033...$$

und der Konvergenzfaktor

$$\eta \;=\; \frac{\kappa - 1}{\kappa + 1} = 0.447\,213... < 1$$

folgen.
So wie die Folge der Funktionalwerte $Q(x^{(m)})$ streng monoton fallend gegen $-\frac{7}{10}$ strebt, so tun dies auch die Fehler $e^{(m)} = e(x^{(m)}) = x^* - x^{(m)}$ in der A-Norm, aber natürlich gegen Null.
Man prüft leicht nach, dass die Fehlerschätzung

$$\|e^{(m)}\|_A \;\le\; \eta^m \, \|e^{(0)}\|_A$$

erfüllt ist. So haben wir für $m = 10$ die Abschätzung

$$\|e^{(10)}\|_A = 0.000\,008\,648... \le 0.000\,708\,350... = \eta^{10} \, \|e^{(0)}\|_A.$$

Der erste Schritt des GV ist wie folgt.

$x^{(0)} = (\frac{3}{2}, 1)^T$ Startvektor,
$r^{(0)} = b - Ax^{(0)} = (-3, -\frac{5}{2})^T$ Anfangsresiduum, Abstiegsrichtung,
$Ar^{(0)} = -\frac{1}{2}(17, 21)^T$.

S1

$$\begin{aligned}
\alpha_0 &= \frac{\|r^{(0)}\|_2^2}{\|r^{(0)}\|_A^2} = \frac{61}{207} = 0.294\,685... \\
x^{(1)} &= x^{(0)} + \alpha_0 r^{(0)} = \frac{1}{414}(255, 109)^T, \\
r^{(1)} &= b - Ax^{(1)} = r^{(0)} - \alpha_0 A r^{(0)} = \frac{41}{414}(-5, 6)^T.
\end{aligned}$$

Iterationsverlauf mit verschiedenen Fehlern

m	[x(m)[1],	x(m)[2]]	Q(x(m))	||xs-x(m)||_A^2	||xs-x(m)||_2^2	||r(m)||_2^2
0	[1.5000000000,	1.0000000000]	1.750000000000	4.900000000000	1.850000000000	15.250000000000
1	[0.6159420290,	0.2632850242]	-0.496980676329	0.406038647343	0.286384746435	0.598269504539
2	[0.3077245391,	0.6331460120]	-0.683176797639	0.033646404722	0.012703234436	0.104715851430
3	[0.2344670487,	0.5720981034]	-0.698605944830	0.002788110339	0.001966493283	0.004108085282
4	[0.2089265972,	0.6027466453]	-0.699884481576	0.000231036847	0.000087228197	0.000719043249
5	[0.2028561130,	0.5976879085]	-0.699990427562	0.000019144875	0.000013503149	0.000028208633
6	[0.2007397027,	0.6002276008]	-0.699999206780	0.000001586441	0.000000598962	0.000004937392
7	[0.2002366719,	0.5998084085]	-0.699999934270	0.000000131460	0.000000092721	0.000000193698
8	[0.2000612955,	0.6000188602]	-0.699999994553	0.000000010893	0.000000004113	0.000000033903
9	[0.2000196118,	0.5999841238]	-0.699999999549	0.000000000903	0.000000000637	0.000000001330
10	[0.2000050793,	0.6000015628]	-0.699999999963	0.000000000075	0.000000000028	0.000000000233

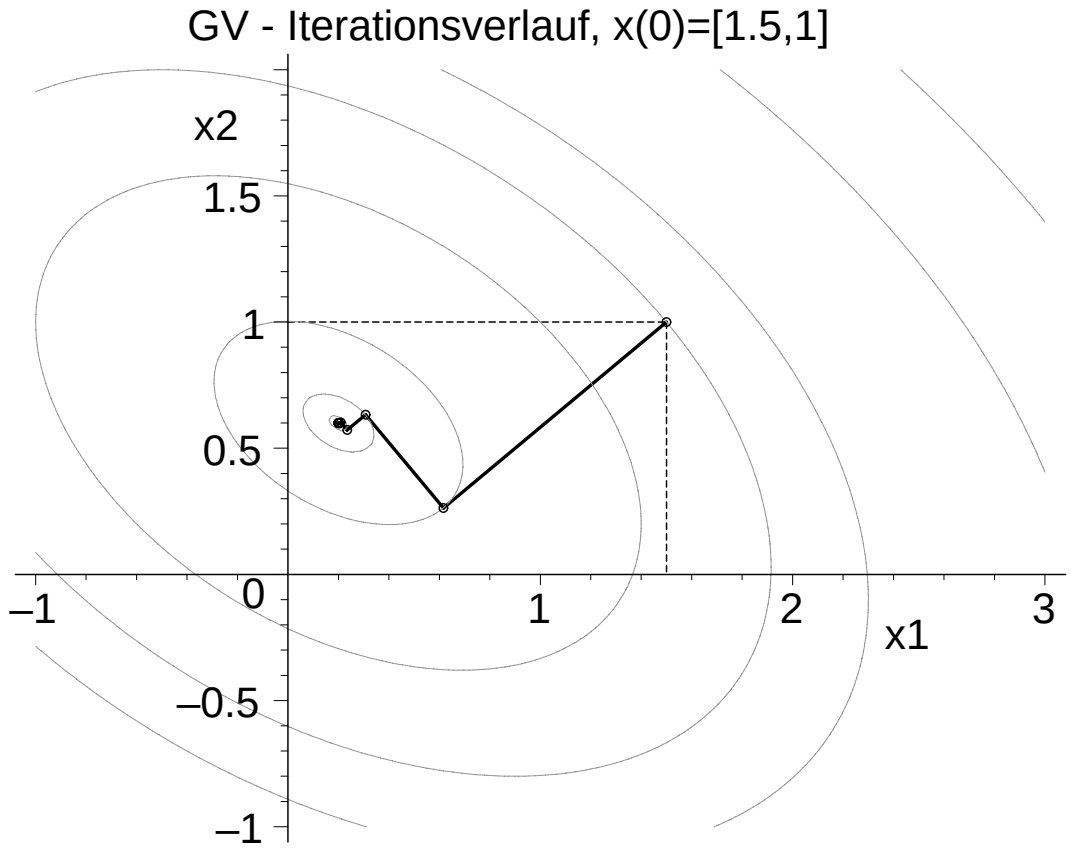


Abb. 3.5 Datei *gv_001.ps*
Höhenlinienbild mit Iterationsverlauf des GV zu $Q(x) = x_1^2 + x_1x_2 + \frac{3}{2}x_2^2 - x_1 - 2x_2$, `contours=[1.75,0.5,-0.497,-0.6832,-0.6986,-0.6999,-0.7]` und `contours=3`

Das GV für $A^TAx = A^Tb$ ist mit dem Funktional $R(x)$ verknüpft.

Es bringt eine Verschlechterung der Kondition des Systems durch Multiplikation mit A^T, aber für beliebiges reguläres A ist die Matrix $B = A^TA$ spd.
Im Allgemeinen ergeben sich dadurch mehr Iterationsschritte, aber nicht generell.

Wir betrachten also das LGS

$$Bx = \begin{pmatrix} 5 & 5 \\ 5 & 10 \end{pmatrix} x = \begin{pmatrix} 4 \\ 7 \end{pmatrix} = c, \quad B = B^T > 0.$$

Die Richtung $\hat{r}(x) = A^T r(x) = A^T(b - Ax)$ ist orthogonal zu $R(x)$=const.
Die Eigenwerte von B sind

$$\mu_{1,2} \;=\; \frac{15 \pm 5\sqrt{5}}{2} = \lambda_{1,2}^2 = 13.090\,169\,943..., \;\; 1.909\,830\,056...,$$

woraus die Kondition

$$\kappa' \;=\; \frac{\mu_1}{\mu_2} = \frac{15 + 5\sqrt{5}}{15 - 5\sqrt{5}} = 6.854\,101...$$

und der Konvergenzfaktor

$$\eta' \;=\; \frac{\kappa' - 1}{\kappa' + 1} = 0.745\,355... < 1$$

folgen.
So wie die Folge der Funktionalwerte $R(x^{(m)})$ streng monoton fallend gegen $-\frac{5}{2}$ strebt, so tun dies auch die Residua $r^{(m)} = r(x^{(m)}) = b - Ax^{(m)}$ in der euklidischen Norm sowie wegen (2.18) auch $\|e^{(m)}\|_{A^TA}$, aber gegen Null.
Man prüft auch leicht nach, dass die Fehlerschätzung

$$\|e^{(m)}\|_{A^TA} \;\leq\; \eta'^m \, \|e^{(0)}\|_{A^TA}$$

erfüllt ist. Auch die Norm des Residuums $\hat{r}(x) = A^T r(x) = A^T(b - Ax)$ verkleinert sich stetig.

```
Iterationsverlauf mit verschiedenen Fehlern
```

m	[x(m)[1], x(m)[2]]	R(x(m))	||r(m)||_2^2 = ||xs-x(m)||_B^2	||xs-x(m)||_2^2	||rd(m)||_2^2
0	[1.5000000000, 1.0000000000]	5.125000000000	15.250000000000	1.850000000000	182.500000000000
1	[0.8416445623, 0.1867374005]	-1.942639257294	1.114721485411	0.582493720493	2.158479268833
2	[0.2950254381, 0.6292385963]	-2.459258885573	0.081482228855	0.009884729402	0.975115197772
3	[0.2469019659, 0.5697919542]	-2.497021967502	0.005956064996	0.003112320435	0.011532964050
4	[0.2069460261, 0.6021372388]	-2.499782316275	0.000435367450	0.000052815068	0.005210135063
5	[0.2034283691, 0.5977918979]	-2.499984088084	0.000031823833	0.000016629430	0.000061621745
6	[0.2005077301, 0.6001562247]	-2.499998836895	0.000002326211	0.000000282196	0.000027838257
7	[0.2002506018, 0.5998385955]	-2.499999914981	0.000000170038	0.000000088853	0.000000329251
8	[0.2000371133, 0.6000114195]	-2.499999993785	0.000000012429	0.000000001508	0.000000148743
9	[0.2000183181, 0.5999882019]	-2.499999999546	0.000000000909	0.000000000475	0.000000001759
10	[0.2000027129, 0.6000008347]	-2.499999999967	0.000000000066	0.000000000008	0.000000000795

Im Iterationsverlauf und Konvergenzverhalten des GV erkennt man hier bezüglich des LGS und des Normalgleichungssystems keine wesentlichen Unterschiede. Das liegt mit an der kleinen Dimension des Problems und der Ausgeglichenheit der Elemente der Matrix $B = A^TA$.

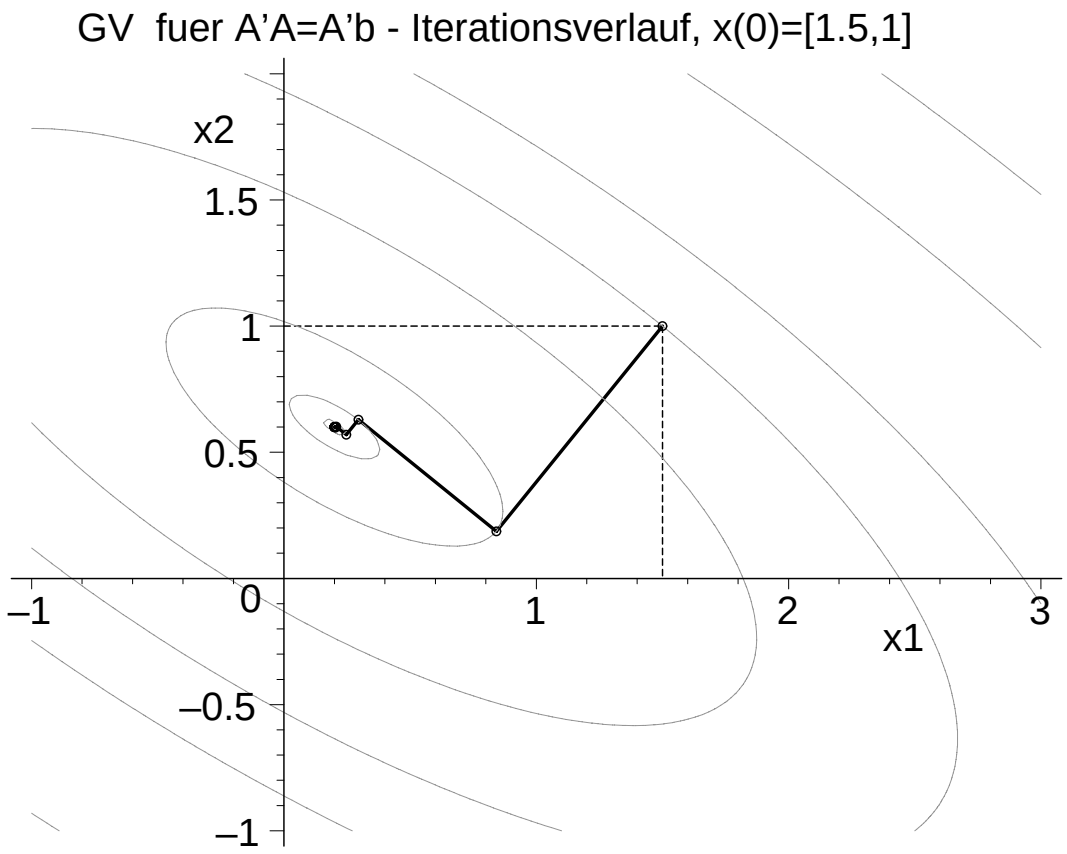


Abb. 3.6 Datei *gv_0011.ps*
Höhenlinienbild mit Iterationsverlauf des GV zu $R(x) = \frac{5}{2}x_1^2 + 5x_1x_2 + 5x_2^2 - 4x_1 - 7x_2$, `contours=[5.125,1,-1.9426,-2.4593,-2.4970,-2.4998,-2.5]` und `contours=3`

Beispiel 3.6

Gegeben sei das LGS aus Beispiel 1.7 mit der regulären Matrix A.
Wir nehmen hier Bezug auf einige Eigenschaften des Systems.

$$\begin{pmatrix} 0.780 & 0.563 \\ 0.913 & 0.659 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.217 \\ 0.254 \end{pmatrix}, \quad x^* = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Um das GV zu verwenden, muss man es symmetrisieren.

$$Bx = A^TAx = \begin{pmatrix} 1.441\,969 & 1.040\,807 \\ 1.040\,807 & 0.751\,250 \end{pmatrix} x = \begin{pmatrix} 0.401\,162 \\ 0.289\,557 \end{pmatrix} = A^Tb = c.$$

Die Funktionale sind

$$\begin{aligned} Q(x) &= 0.5(0.780x_1 + 0.913x_2)x_1 + 0.5(0.563x_1 + 0.659x_2)x_2 - 0.217x_1 - 0.254x_2 \\ &= 0.390x_1^2 + 0.738x_1x_2 + 0.3295x_2^2 - 0.217x_1 - 0.254x_2, \\ R(x) &= 0.5(1.441\,969x_1 + 1.040\,807x_2)x_1 + 0.5(1.040\,807x_1 + 0.751\,250x_2)x_2 \\ &\quad -0.401\,162x_1 - 0.289\,557x_2 \\ &= 0.720\,9845x_1^2 + 1.040\,807x_1x_2 + 0.375\,625x_2^2 - 0.401\,162x_1 - 0.289\,557x_2. \end{aligned}$$

Das Minimum von $R(x)$ ist bei der Lösung $x^* = (1, -1)^T$ und $R(x^*) = -0.055\,802\,5$ sowie $Q(x^*) = 0.018\,5$.
Aber $Q(x)$ hat die Oberfläche eines Sattels mit dem Sattelpunkt bei

$$z = \Big(\frac{44\,449}{30\,624}, -\frac{6\,329}{5\,104}\Big)^T = (1.451\,443\,312\,434\,691\,7, -1.240\,007\,836\,990\,595\,6)^T,$$

$Q(z) = -\frac{37}{61\,248\,000} = -0.604\,101\,358\,411\,702\,9\,10^{-6}$, und keine Minimumstelle.
Die Eigenwerte von B sind

$$\mu_{1,2} \;=\; 2.193\,218\,999\,999\,544..., \; 4.559\,508\,193\,209... \cdot 10^{-13} > 0$$

woraus sich die sehr schlechte Kondition

$$\kappa' \;=\; 0.481\,020\,958\,195\,900 \cdot 10^{13}$$

und der Konvergenzfaktor

$$\eta' \;=\; \frac{\kappa' - 1}{\kappa' + 1} = 0.999\,999\,999\,999\,584... < 1,$$

aber nahe der Eins, ergeben.
Wir testen den Iterationsverlauf des GV mit ausgewählten Startvektoren

$$x^{(0)} = \begin{pmatrix} 1.2 \\ -1.2 \end{pmatrix}, \begin{pmatrix} 0.999 \\ -1.001 \end{pmatrix}, \begin{pmatrix} 0.341 \\ -0.087 \end{pmatrix}, \begin{pmatrix} 0.991\,891\,566\,446\,068\,04 \\ -1.005\,852\,632\,339\,509\,328 \end{pmatrix}.$$

Der Vektor $(0.341, -0.087)^T$ liegt nahe der Tallinie (1.14) $t(x_1) = \frac{v_2^{(2)}}{v_1^{(2)}}(x_1 - 1) - 1$ von $R(x)$, auf der die Werte $R(x)$ nur minimal größer sind als $R(x^*)$ bzw. die Werte $\|r(x)\|_2$ oder $\|\hat{r}(x)\|_2$, $\hat{r}(x) = A^T r(x) = c - Bx$, ganz nahe bei der Null liegen.
Der Vektor $x^* - (0.991\,891\,566\,446\,068\,04, -1.005\,852\,632\,339\,509\,328)^T$ ist orthogonal zur Tallinie $t(x_1)$.

Welche Startsituation findet man vor?

$x^{(0)}$	$r(x^{(0)})$	$\|r(x^{(0)})\|_2^2$	$A^T r(x^{(0)})$	$R(x^{(0)})$
$[1.2, -1.2]$	$[-0.043\,4, -0.050\,8]$	$0.004\,464\,20$	$[-0.080\,232\,4, -0.057\,911\,4]$	$-0.053\,570\,400$
$[0.999, -1.001]$	$[\ 0.001\,343,\ 0.001\,572]$	$0.427 \cdot 10^{-5}$	$[0.002\,482\,776,\ 0.001\,792\,057]$	$-0.055\,800\,362\,583\,5$
$[0.341, -0.087]$	$[10^{-6},\ 0]$	10^{-12}	$[0.780 \cdot 10^{-6},\ 0.563 \cdot 10^{-6}]$	$-0.055\,802\,499\,999\,5$
$[0.991\,891..., -1.005\,852...]$	$[0.009\,619...,\ 0.011\,259...]$	$0.000\,219...$	$[0.017\,783...,\ 0.012\,836...]$	$-0.055\,692\,839...$
$x^* = (1, -1)^T$				$R(x^*) = -0.055\,802\,5$

Tab. 3.1 Startvektor $x^{(0)} = (x_1^{(0)}, x_2^{(0)})^T$, dazu Anfangsresidua und Funktionale

Wir werden den Iterationsverlauf und das Verhalten des GV mit dem Startvektor $x^{(0)} = (1.2, -1.2)^T$ etwas ausführlicher besprechen, denn ähnliche Aspekte bemerken wir auch bei den anderen Startvektoren.
Auf Grund der schlechten Kondition der Matrix $B = A^T A$ der Größenordnung 10^{12} muss das GV mit einer entsprechend starken Gleitpunktarithmetik arbeiten, um relevante Ergebnisse zu erzielen. Rechnet man mit t Dezimalstellen in der Mantisse, so ist wegen der Matrixkondition ein Verlust von ungefähr 12 Dezimalstellen, wenn nicht sogar einige mehr, zu erwarten. Von den angezeigten t Stellen sind also die letzten 12 und ev. mehr Stellen nicht verwertbar. Wenn also die Iterierten des GV in die Nähe von x^* kommen und mit dem Grenzvektor auf ca. $t - 12$ übereinstimmen, ist dann keine signifikante Verbesserung mehr durch weitere Iterationen zu erwarten. Im Gegenteil, es können sogar "zwischenzeitlich" Verschlechterungen auftreten.

Im CAS Maple, das standardmäßig mit dem Gleitpunktformat `Digits:=10` arbeitet, ist also die Genauigkeit deutlich zu erhöhen.

Ergebnisse aus Berechnungen (erste 3 Iterationen) mit Maple mit `Digits:=24`.

```
Startvektor              x=[+1.2000000000000000e+00 -1.2000000000000000e+00]
Residuum/SR  p = rd = c-Bx=[-8.0232400000000000e-02 -5.7911400000000000e-02]
Funktionswert         R(x)= -5.3570400000000000e-02
Anfangsfehlerquadrat rd'rd=  9.7909682597200000e-03

Schritt k = 1
Suchschritt          alpha= +4.5595081932091957e-01
Iterationsvektor         x=[+1.1634179714839163e+00 -1.2264047502780215e+00]
Residuum/neue SR rd = c-Bx=[-7.4510557990000000e-14 +1.0322943138800000e-13]
Funktionswert         R(x)= -5.5802499999982226e-02
Fehlernormquadrat    rd'rd=  1.6208138756670952e-26
Schritt k = 2
Suchschritt          alpha= +2.1932189999663202e+12
Iterationsvektor         x=[+1.0000000000021559e+00 -1.0000000000021403e+00]
Residuum/neue SR rd = c-Bx=[-8.8118966549200000e-13 -6.3603889693100000e-13]
Funktionswert         R(x)= -5.5802500000000000e-02
Fehlernormquadrat    rd'rd=  1.1810407049791061e-24
Schritt k = 3
Suchschritt          alpha= +4.5595081932091957e-01
Iterationsvektor         x=[+1.0000000000017542e+00 -1.0000000000024303e+00]
Residuum/neue SR rd = c-Bx=[-1.0000000000000000e-24 +2.0000000000000000e-24]
Funktionswert         R(x)= -5.5802500000000000e-02
Fehlernormquadrat    rd'rd=  5.0000000000000000e-48
```

Die Ergebnisse sind mit 16 Nachkommastellen angezeigt. Durch den Genauigkeitsverlust sind nur ca. 12 anfängliche Stellen genau. Mit wenigen Iterationsschritten ist man in diesem Bereich. In der folgenden Übersicht nehmen wir noch die nächsten Iterationen dazu.

```
Iterationsverlauf mit verschiedenen Fehlern

m  [     x(m)[1],          x(m)[2]    ]     R(x(m))       ||r(m)||_2^2=  ||xs-x(m)||_2^2 ||rd(m)||_2^2
                                                          ||xs-x(m)||_B^2
-------------------------------------------------------------------------------------------------------
0  [1.20000000000000,-1.20000000000000] -0.05357040000000   4.464200e-03   8.000000e-02   9.790968e-03
1  [1.16341797148392,-1.22640475027802] -0.05580249999998   3.554800e-14   7.796454e-02   1.620814e-26
2  [1.00000000000216,-1.00000000000214] -0.05580250000000   5.384965e-25   9.228861e-24   1.181041e-24
3  [1.00000000000175,-1.00000000000243] -0.05580250000000   4.095957e-36   8.983333e-24   5.000000e-48
4  [1.00000000000175,-1.00000000000243] -0.05580250000000   4.095958e-36   8.983333e-24   1.800000e-46
5  [1.00000000000175,-1.00000000000243] -0.05580250000000   4.095960e-36   8.983333e-24   5.000000e-48
6  [1.00000000000175,-1.00000000000243] -0.05580250000000   4.095957e-36   8.983333e-24   1.800000e-46
```

Im Iterationsverlauf und Konvergenzverhalten des GV erkennt man hier, dass zwischen der 3. und 4. Iteration kaum noch Unterschiede sind. Somit braucht man nicht mehr als vier Iterationen auszuführen.

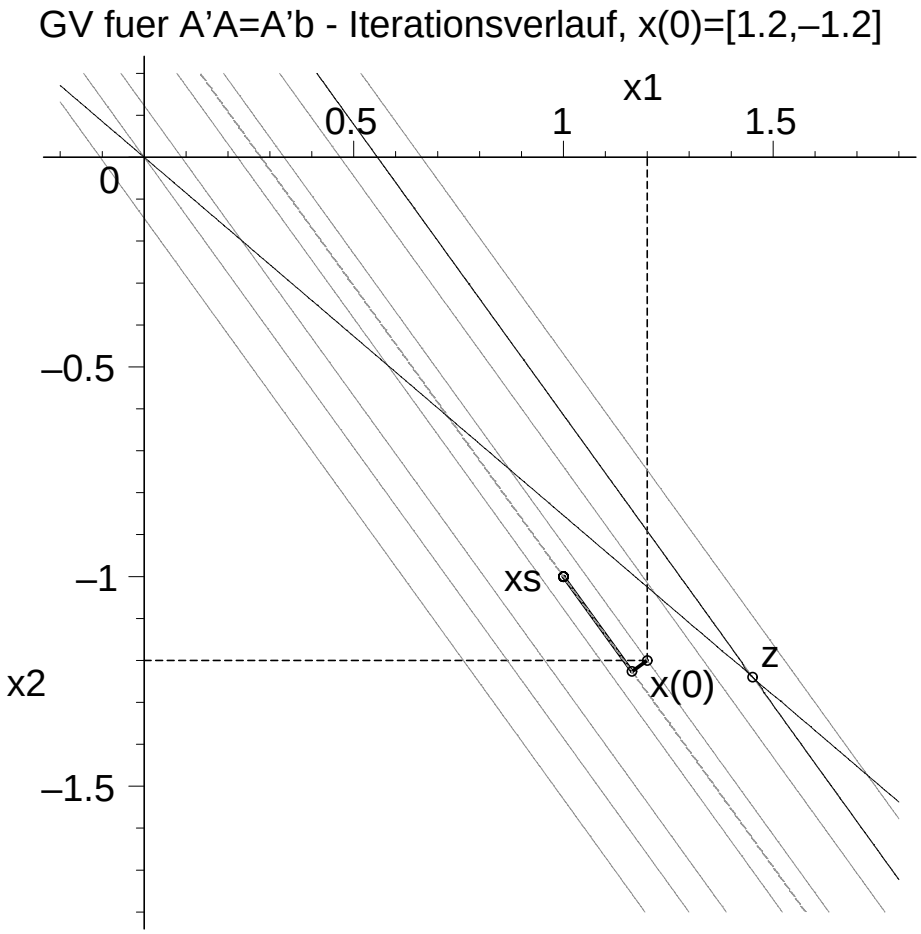


Abb. 3.7 Datei *abst*102.*ps*
Höhenlinienbild mit Iterationsverlauf des GV mit $x^{(0)} = (1.2, -1.2)^T$ zu
$R(x) = 0.720\,9845x_1^2 + 1.040\,807x_1x_2 + 0.375\,625x_2^2 - 0.401\,162x_1 - 0.289\,557x_2$
mit `contours=[0.05,0,-0.03,-0.05357,-0.0558,-0.558025]` sowie
zu $Q(x)$ `contours=[-0.000000604]`, Sattelpunkt z und Höhenlinien $Q(x) = Q(z)$

Eine weitere Rechnung mit Maple macht den Einfluss des Gleitpunktformats mit `Digits:=k` deutlich. Wenn die Größenangabe k mit dem Vielfachen einer Bytelänge korrespondiert, liegt meistens eine günstigere Situation vor.
Die Angaben beziehen sich auf die Iterierte $x^{(4)}$, bis auf `Digits:=10,11,12`, wo die Iteration nur bis $x^{(1)}$ geht. Es ist `rd(m)`$= \hat{r}^{(m)} = A^T r^{(m)} = c - Bx^{(m)}$.

```
 k                x(4)                                                R(x(4))          rd(4)'rd(4)
-------------------------------------------------------------------------------------------------------
10-12 [1.16341797148,-1.22640475028]                          -5.58024999940000e-2          0

13    [1.163417971480,-1.226404750274]                        -5.58024999997000e-2  7.20000000000000e-25
16    [1.163403311543425,-1.226384441360354]                  -5.58024999999826e-2  3.49588444849201e-18
20    [1.0002885805872075026,-1.0003998092376051912]          -5.58025000000000e-2  8.76383150044784e-23
21    [1.00000267784244182094,-1.00000370997290789266]        -5.58025000000000e-2  3.72231382425185e-28
22    [1.000000000015091736707,-1.000000000020908599277]                      "     1.00000000000000e-44
23    [1.0000000000917736240901,-1.00000000001271462633864]                   "     1.25800000000000e-42

24    [1.000000000000175416320873,-1.000000000000243027666794]                "     1.80000000000000e-46

25    [1.0000000000001536092273749,-1.0000000000002128153865046]              "     7.10770000000000e-45
26    [1.0000000000013383836164967,-1.0000000000018542416459392]              "     1.46551680400000e-43
27    [1.0000000000130795540772203,-1.0000000000181208537818450]              "     3.96363229563600e-41
28    [1.0000000000129950186271784 9,-1.0000000000018003734962759 16]         "     6.21875341211672e-39
29    [1.0000000000012501382045546179,-1.0000000000017319838229222509]        "     1.86429119749795e-37
30    [1.0000000000001106279212990 6489,-1.0000000000001532681235483 0612]    "     3.84545937986319e-37
31    [1.0000000000000012048519905714 75,-1.0000000000000016692951288719 88]  "     4.57346124818143e-39
32    [1.0000000000000000740065234317793,-1.0000000000000001025180632499014] "     2.81104120905338e-40
33    [1.0000000000000000002076446052997,-1.0000000000000000002883705214140] "     7.90614155818790e-43
34    [1.0000000000000000000008933432720,-1.0000000000000000000011902072492] "     3.71135023798440e-45
35    [1.0000000000000000000000022003177,-0.9999999999999999999999989856456] "     2.71960832567827e-47
36    [1.0000000000000000000000144232654,-1.0000000000000000000000147549151] "     4.50261495085612e-47

40    [1.0000000000000000000000012681931119283061,                            "     3.93664865153553e-47
      -1.0000000000000000000000012681986133793689]
```

Im GV mit den anderen Startvektoren machen wir nun 4 und mehr Iterationen mit `Digits:=24`.

Ergebnisse aus Berechnungen mit Maple für den Startvektor $x^{(0)} = (0.999, -1.001)^T$

```
Startvektor              x = [+9.9900000000000000e-01 -1.0010000000000000e+00]
Residuum/SR  p = rd = c-Bx = [+2.4827760000000000e-03 +1.7920570000000000e-03]
Funktionswert         R(x) =  -5.5800362583500000e-02
Anfangsfehlerquadrat rd'rd =   9.3756449574250000e-06
```

```
 k   Schrittzahl  alpha     Iterationsvektor  x                                Funktionswert  R(x)
                            Residuum/neue Suchrichtung  p=rd=c-Bx             Fehlernormquadrat rd'rd
-------------------------------------------------------------------------------------------------------

 1  4.5595081932091957e-01  [+1.0001320237513903e+00,-1.0001829101425802e+00] -5.5802500000000000e-02
                            [-6.0196338000000000e-17,+8.3398029000000000e-17]  1.0578830349695085e-32

 2  2.1929403522476651e+12  [+1.0000000167727326e+00,-1.0000000232394882e+00] -5.5802500000000000e-02
                            [+2.0615646946770000e-12,+1.4880285140750000e-12]  6.4642778490389247e-24

 3  4.5595081932091958e-01  [+1.0000000167736725e+00,-1.0000000232388097e+00] -5.5802500000000000e-02
                            [-7.6480000000000000e-21,+1.0596000000000000e-20]  1.7076712000000000e-40

 4  5.4621008188331627e+09  [+1.0000000167318984e+00,-1.0000000231809333e+00] -5.5802500000000000e-02
                            [-1.1583007480000000e-15,-8.3604040400000000e-16]  2.0406241799378427e-30
```

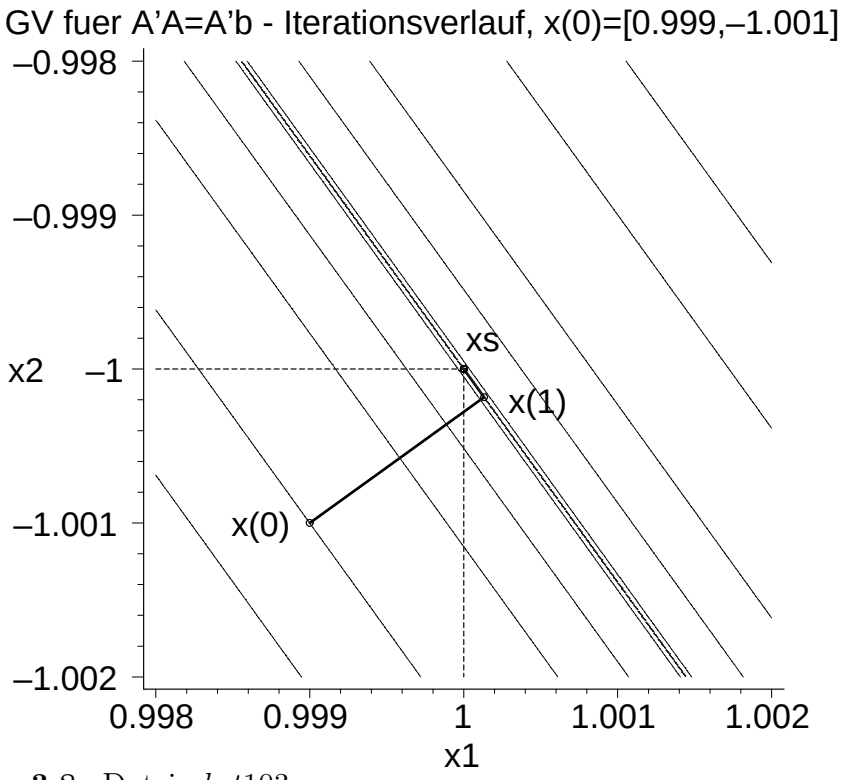


Abb. 3.8 Datei *abst*103.*ps*
Höhenlinienbild mit Iterationsverlauf des GV mit $x^{(0)} = (0.999, -1.001)^T$ zu $R(x) = 0.720\,9845x_1^2 + 1.040\,807x_1x_2 + 0.375\,625x_2^2 - 0.401\,162x_1 - 0.289\,557x_2$ mit `contours=[-0.055798,-0.05580036,-0.055802,-0.0558024,-0.055802499, -0.05580249999,-0.0558025]`

Ergebnisse aus Berechnungen mit Turbo Pascal im Format *extended* (19-20 Dezimalstellen), was mit Maple und `Digits:=19..20` vergleichbar ist. Mehr als 5..6 genaue Dezimalziffern der Iterierten sind trotz sehr vieler Schritte nicht zu erzielen.

```
Startvektor              x = [+9.9900000000000000e-01 -1.0010000000000000e+00]
Residuum/SR  p = rd = c-Bx = [+2.4827759999999999e-03 +1.7920570000000000e-03]
Funktionswert         R(x) =  -5.5800362583499999e-02
Anfangsfehlerquadrat rd'rd =   9.3756449574250004e-06

   k   Schrittzahl  alpha    Iterationsvektor  x                              Funktionswert  R(x)
                             Residuum/neue Suchrichtung  p=rd=c-Bx            Fehlernormquadrat rd'rd
-------------------------------------------------------------------------------------------------------
    1 4.5595081932091957e-01 [+1.0001320237513903e+00,-1.0001829101425802e+00] -5.5802499999999999e-02
                             [-6.0254535735881909e-17,+8.3483567281383841e-17]  1.0600115082772013e-32
    2 5.8553341988633887e+08 [+1.0001319884703459e+00,-1.0001828612601616e+00] -5.5802499999999999e-02
                             [+2.9912992384733411e-12,-2.1589799446917329e-12]  1.3609065535672309e-23
  100 4.5595082279735614e-01 [+1.0001318155843438e+00,-1.0001826217409574e+00] -5.5802499999999999e-02
                             [-6.0227430681569771e-17,+8.3456462227071704e-17]  1.0592324493961938e-32
 1000 4.5595082279735614e-01 [+1.0001301947908607e+00,-1.0001803762391899e+00] -5.5802499999999997e-02
                              unveraendert                                      unveraendert
10000 4.5595082279735614e-01 [+1.0001139868560290e+00,-1.0001579212215151e+00] -5.5802499999999969e-02
60000 4.5595082279735614e-01 [+1.0000239427736309e+00,-1.0000331711233213e+00] -5.5802499999999969e-02

70000 4.5595082279735614e-01 [+1.0000059339571513e+00,-1.0000082211036826e+00] -5.5802499999999785e-02
80000 4.5595082279735614e-01 [+9.9998792514067167e-01,-9.9998327108404382e-01] -5.5802499999999754e-02
```

Ergebnisse aus Berechnungen mit Maple für den Startvektor $x^{(0)} = (0.341, -0.087)^T$

Hier verläuft die Iteration von einer Seite kommend zickzackförmig entlang der Tallinie. Die Funktionswerte $R(x^{(m)})$ und damit die Größen $\|r(x^{(m)})\|_2$ werden langsam kleiner. Die Iterierten $x^{(2m)}$ und $x^{(2m+1)}$ liegen sehr nahe beieinander, aber $\|\hat{r}(x^{(2m+1)})\|_2 \approx \|\hat{r}(x^{(2m)})\|_2^2$.

Von $x^{(2m+1)}$ zu $x^{(2m+2)}$ ist ein sprungförmiger Zuwachs.

Die Höhenlinien $R(x) = R(x^{(m)}),\ m = 0, 1, ...,$ an den Iterierten $x^{(m)}$ sind extrem lang gestreckte Ellipsen. Man kann diese wie auch den Zick-Zack-Verlauf der Iteration im Bereich $[0, 1] \times [-1, 0]$ grafisch nicht mehr auflösen.

```
Startvektor              x = [+3.4100000000000000e-01 -8.7000000000000000e-02]
Residuum/SR  p = rd = c-Bx = [+7.8000000000000000e-07 +5.6300000000000000e-07]
Funktionswert         R(x) =  -5.5802499999500000e-02
Anfangsfehlerquadrat rd'rd =   9.2536900000000000e-13
```

k	Schrittzahl	alpha	Iterationsvektor x Residuum/neue Suchrichtung p=rd=c-Bx	Funktionswert R(x) Fehlernormquadrat rd'rd
1	4.5595081932104944e-01	[+3.4100035564163907e-01,-8.6999743299688722e-02] [+3.0047120560800000e-13,-4.1628337544300000e-13]	-5.5802499999710961e-02 2.6357479406974271e-25
2	9.2536831401682559e+11	[+6.1904688858571698e-01,-4.7221518858661085e-01] [+4.5090020694879444e-07,+3.2545745706665156e-07]	-5.5802499999832913e-02 3.0923355298675699e-13
3	4.5595081932104944e-01	[+6.1904709417403577e-01,-4.7221504019401665e-01] [+1.7369566110300000e-13,-2.4064407754900000e-13]	-5.5802499999903411e-02 8.8079754745417353e-26
4	9.2537077648614870e+11	[+7.7977998296119382e-01,-6.9490003709232776e-01] [+2.6065538469065704e-07,+1.8813971997578711e-07]	-5.5802499999944164e-02 1.0333778380080199e-13
5	4.5595081932104944e-01	[+7.7978010180723003e-01,-6.9489995130986829e-01] [+1.0040936876900000e-13,-1.3911067076300000e-13]	-5.5802499999967723e-02 2.9433820056720816e-26
6	9.2537204880815851e+11	[+8.7269612510453348e-01,-8.2362907772490280e-01] [+1.5067876386208253e-07,+1.0875915904439018e-07]	-5.5802499999981341e-02 3.4532644555048186e-14
7	4.5595081932104944e-01	[+8.7269619380663932e-01,-8.2362902813607512e-01] [+5.8044231818000000e-14,-8.0416520103000000e-14]	-5.5802499999989214e-02 9.8359495528179267e-27
8	9.2537007162094862e+11	[+9.2640858876124492e-01,-8.9804406910329568e-01] [+8.7103730289064610e-08,+6.2871025836921303e-08]	-5.5802499999993765e-02 1.1539825720056938e-14
9	4.5595081932104944e-01	[+9.2640862847626212e-01,-8.9804404043719994e-01] [+3.3554021331000000e-14,-4.6486921204000000e-14]	-5.5802499999996396e-02 3.2869061905081078e-27
10	9.2538092236035269e+11	[+9.5745887968444184e-01,-9.4106215045865050e-01] [+5.0352721728414277e-08,+3.6344336324939120e-08]	-5.5802499999997916e-02 3.8563073683594125e-15
20	9.2575999861985026e+11	[+9.9725556695314744e-01,-9.9619777333683666e-01] [+3.2495299036777120e-09,+2.3454940211824430e-09]	-5.5802499999999991e-02 1.6060786798298267e-17
21	4.5595081932104934e-01	[+9.9725556843477326e-01,-9.9619777226740674e-01] [+1.2513248960000000e-15,-1.7336295170000000e-15]	-5.5802499999999995e-02 4.5712852975630641e-30
50	1.5220152382301831e+12	[+9.9999971880877108e-01,-9.9999961042878408e-01] [+5.8663684026700000e-13,+4.2343166894500000e-13]	-5.5802500000000000e-02 5.2343716062399775e-25
100	7.0094230891129941e+07	[+9.9999997250723875e-01,-9.9999996191060431e-01] [-2.1535254600000000e-16,-1.5546695700000000e-16]	-5.5802500000000000e-02 7.0546693787521965e-32

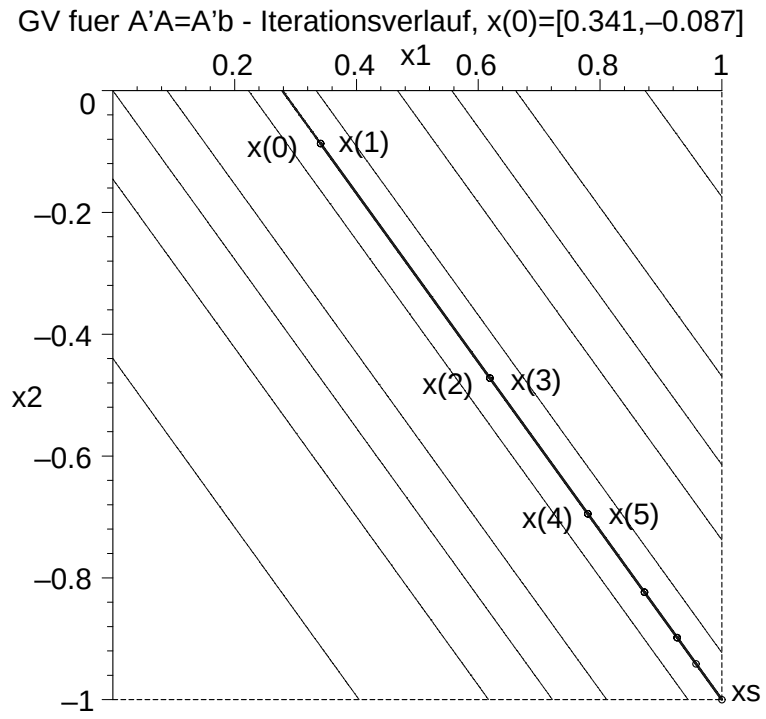


Abb. 3.9 Datei *abst*104.*ps*
Höhenlinienbild mit Iterationsverlauf des GV mit $x^{(0)} = (0.341, -0.087)^T$ zu $R(x) = 0.720\,9845x_1^2 + 1.040\,807x_1x_2 + 0.375\,625x_2^2 - 0.401\,162x_1 - 0.289\,557x_2$ mit `contours=[0.2,0.05,0,-0.03,-0.05357,-0.055801,-0.0558025]`

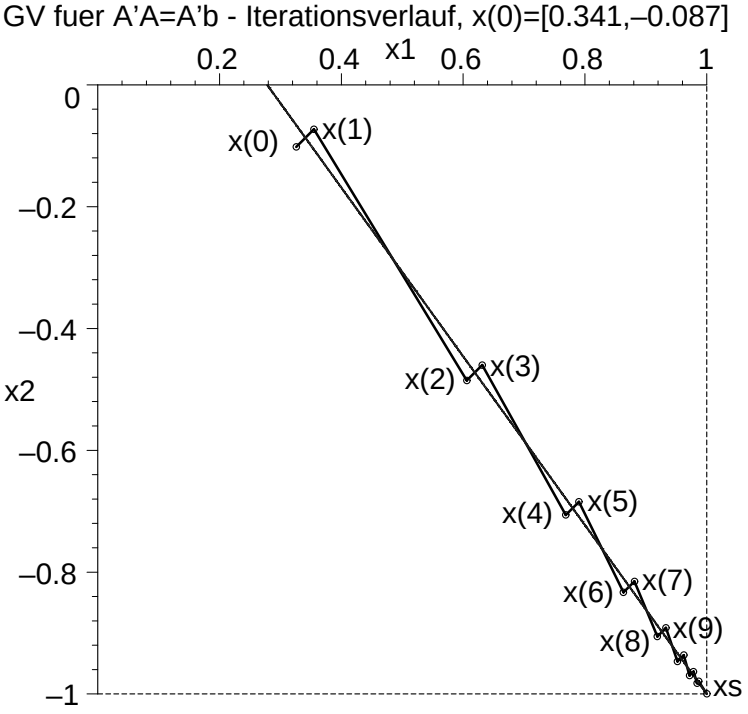


Abb. 3.10
Datei *abst*1041.*ps*
Höhenlinienbild mit Iterationsverlauf des GV mit $x^{(0)} = (0.341, -0.087)^T$ zu $R(x)$ mit `contours=[-0.055801]`, bei künstlicher Verstärkung des Zick-Zack-Verlaufs

Ergebnisse aus Berechnungen mit Maple für den Startvektor
$x^{(0)} = (0.991\,891\,566\,446\,068\,04, -1.005\,852\,632\,339\,509\,328)^T$

Der Vektor $x^* - x^{(0)}$ ist orthogonal zur Tallinie $t(x_1)$ mit der Genauigkeitsordnung $\mathcal{O}(10^{-17})$.

Somit muss der erste Gradientenschritt bis auf geringe Ungenauigkeiten in den letzten Dezimalstellen in das Minimum führen, also $x^{(1)} \approx x^*$. Weitere Iterationen werden sich dann um die Minimumstelle bewegen.

Die Höhenlinien $R(x) = R(x^{(m)}),\ m = 0, 1, ...,$ an den Iterierten $x^{(m)}$ sind extrem lang gestreckte Ellipsen. Man kann diese wie auch den Verlauf der Iteration im Bereich $[0.99, 1.01] \times [-1.01, -0.99]$ grafisch nur schwer auflösen.

```
Startvektor              x = [+9.9189156644606804e-01 -1.0058526323395093e+00]
Residuum/SR  p = rd = c-Bx = [+1.7783570530717400e-02 +1.2836104447023644e-02]
Funktionswert         R(x) =  -5.5692839050000023e-02
Anfangsfehlerquadrat rd'rd =   4.8102095819590051e-04

 k   Schrittzahl  alpha     Iterationsvektor  x                              Funktionswert  R(x)
                            Residuum/neue Suchrichtung  p=rd=c-Bx            Fehlernormquadrat rd'rd

-------------------------------------------------------------------------------------------------

 1  4.5595081932091957e-01  [+1.0000000000000000e+00,-1.0000000000000000e+00] -5.5802500000000000e-02
                            [+1.0000000000000000e-24,-2.0000000000000000e-24]  5.0000000000000000e-48

 2  1.7621704300753152e+01  [+1.0000000000000000e+00,-1.0000000000000000e+00] -5.5802500000000000e-02
                            [+1.2000000000000000e-23,+6.0000000000000000e-24]  1.8000000000000000e-46
```

Bei einer Ausgabe mit 16 Nachkommastellen werden die Iterierten $x^{(m)},\ m = 1, 2, ...,$ gerundet und wie x^* angezeigt.

Betrachtet man sich die 24-stelligen Zahlen aus der Rechnung mit `Digits:=24`, so sieht man die Stagnation des Iterationsprozesses nach der ersten Iteration.

```
 m   [ x(m)[1],                        x(m)[2]                      ]
-------------------------------------------------------------------

 0   [ 0.99189156644606804,          -1.005852632339509328        ]
 1   [ 0.99999999999999998743683, -0.99999999999999998259454]
 2   [ 0.99999999999999998743701, -0.99999999999999998259489]
 3   [ 0.99999999999999998743707, -0.99999999999999998259486]
 4   [ 0.99999999999999998743725, -0.99999999999999998259521]
 5   [ 0.99999999999999998743731, -0.99999999999999998259518]
 6   [ 0.99999999999999998743749, -0.99999999999999998259553]
 7   [ 0.99999999999999998743755, -0.99999999999999998259550]
 8   [ 0.99999999999999998743773, -0.99999999999999998259585]
 9   [ 0.99999999999999998743779, -0.99999999999999998259582]
10   [ 0.99999999999999998743797, -0.99999999999999998259617]
```

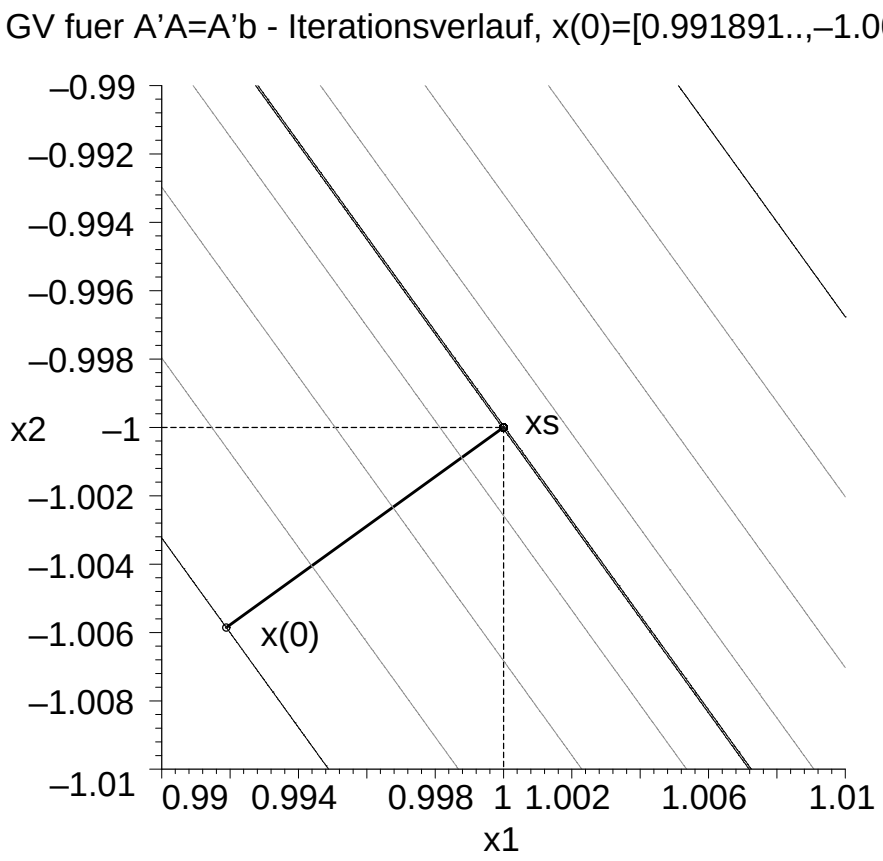


Abb. 3.11 Datei *abst*105.*ps*
Höhenlinienbild mit Iterationsverlauf des GV mit
$x^{(0)} = (0.991\,891\,566\,446\,068\,04, -1.005\,852\,632\,339\,509\,328)^T$ zu
$R(x) = 0.720\,9845x_1^2 + 1.040\,807x_1x_2 + 0.375\,625x_2^2 - 0.401\,162x_1 - 0.289\,557x_2$
mit `contours=[-0.05569283905,-0.05575,-0.055785,-0.05580,-0.055802499]`

Ergebnisse aus Berechnungen mit Turbo Pascal in Format *extended* (19-20 Dezimalstellen), was mit Maple und `Digits:=19..20` vergleichbar ist.
Mehr als 17..18 anfängliche genaue Dezimalziffern der Iterierten sind durch weitere Schritte nicht zu erzielen.

```
Startvektor              x = [+9.9189156644606804e-01 -1.0058526323395093e+00]
Residuum/SR  p = rd = c-Bx = [+1.7783570530717400e-02 +1.2836104447023644e-02]
Funktionswert         R(x) =  -5.5692839050000023e-02
Anfangsfehlerquadrat rd'rd =   4.8102095819590051e-04

  k  Schrittzahl  alpha     Iterationsvektor  x                                Funktionswert  R(x)
                            Residuum/neue Suchrichtung  p=rd=c-Bx             Fehlernormquadrat rd'rd

-------------------------------------------------------------------------------------------------

  1  4.5595081932091957e-01 [+9.9999999999999999e-01,-9.9999999999999999e-01] -5.5802500000000000e-02
                            [-5.4210108624275222e-20,+8.1315162936412833e-20]  9.5508916004310860e-39

  2  3.2959789057350032e+02 [+9.9999999999999998e-01,-9.9999999999999997e-01] -5.5802499999999999e-02
                            [-2.1955093992831465e-18,-1.4636729328554310e-18]  6.9625999767142617e-36
```

```
  3  4.5658243481715015e-01 [+9.9999999999999998e-01,-9.9999999999999997e-01] -5.5802499999999999e-02
                            [-5.4210108624275222e-20,+8.1315162936412833e-20]  9.5508916004310860e-39

  4  3.2959789057350032e+02 [+9.9999999999999996e-01,-9.9999999999999995e-01] -5.5802499999999999e-02
                            [-2.1955093992831465e-18,-1.4636729328554310e-18]  6.9625999767142617e-36

  5  4.5658243481715015e-01 [+9.9999999999999996e-01,-9.9999999999999995e-01] -5.5802499999999999e-02
                            [-5.4210108624275222e-20,+8.1315162936412833e-20]  9.5508916004310860e-39

                            [+9.9999999999999980e-01,-9.9999999999999972e-01] (x, Maple, Digits:=20)
 10  3.2959789057350032e+02 [+9.9999999999999991e-01,-9.9999999999999987e-01] -5.5802499999999999e-02
                            [-2.1955093992831465e-18,-1.4636729328554310e-18]  6.9625999767142617e-36

 11  4.5658243481715015e-01 [+9.9999999999999990e-01,-9.9999999999999987e-01] -5.5802499999999999e-02
                            [-5.4210108624275222e-20,+8.1315162936412833e-20]  9.5508916004310860e-39

 20  3.2959789057350032e+02 [+9.9999999999999981e-01,-9.9999999999999974e-01] -5.5802499999999999e-02
                            [-2.1955093992831465e-18,-1.4636729328554310e-18]  6.9625999767142617e-36

 50  3.2959789057350032e+02 [+9.9999999999999911e-01,-9.9999999999999876e-01] -5.5802499999999999e-02
                             unveraendert                                       unveraendert

                            [+9.9999999999999800e-01,-9.9999999999999724e-01] (x, Maple, Digits:=20)
100  3.2959789057350032e+02 [+9.9999999999999863e-01,-9.9999999999999811e-01] -5.5802499999999999e-02
                             unveraendert                                       unveraendert

                            [+9.9999999999998006e-01,-9.9999999999997238e-01] (x, Maple, Digits:=20)
1000 3.2959789057350032e+02 [+9.9999999999999057e-01,-9.9999999999998693e-01] -5.5802499999999999e-02
                             unveraendert                                       unveraendert

  10000                     [+9.9999999999990567e-01,-9.9999999999986935e-01]
  20000                     [+9.9999999999986473e-01,-9.9999999999973871e-01]
  30000                     [+9.9999999999971702e-01,-9.9999999999960806e-01]
  40000                     [+9.9999999999962270e-01,-9.9999999999947741e-01]
  50000                     [+9.9999999999952837e-01,-9.9999999999934677e-01]
  60000                     [+9.9999999999943405e-01,-9.9999999999921612e-01]

 100000                     [+9.9999999999905674e-01,-9.9999999999869353e-01]
 200000                     [+9.9999999999811349e-01,-9.9999999999738707e-01]
1000000                     [+9.9999999999056744e-01,-9.9999999998693536e-01]
2000000                     [+9.9999999998113488e-01,-9.9999999997387073e-01]
```

Der erste Gradientenschritt ist der beste und da sollte man stoppen.
Bei weiteren Schritten gehen die Iterierten wegen der sehr schlechten Kondition langsam von der Lösung weg, aber die Fehler und Funktionalwerte ändern dabei nicht ihre Größenordnung.
Es besteht eher die Gefahr wie beim Startvektor $x^{(0)} = (0.999, -1.001)^T$, dass mit wachsender Schrittzahl nur die Hälfte oder weniger der Nachkommastellen als genaue Dezimalziffern in den Iterierten erhalten bleiben.
Bei den Rechnungen in Turbo Pascal wurde der Startvektor mit einer `readln()`-Anweisung eingegeben. Macht man im Programm jedoch Ergibtanweisungen gemäß `x[1]:=0.99189156644606804; x[2]:=-1.005852632339509328`, so werden damit ihre internen Werte geringfügig genauer. Das wirkt sich auf die weiteren Iterierten aus. So erhält man $x^{(2000000)} = (0.99999999999918685, -0.99999999999886159)^T$.

In den Formeln (3.15) haben wir schon einen Hinweis auf gewisse Unterschiede in den Berechnungen der Residua gemacht, das spiegelt sich ebenfalls in den Ergebnissen wider.

3.3 Abstiegsverfahren mit linear unabhängigen Richtungen

Man kann natürlich von vornherein die linear unabhängigen Suchrichtungen

$$p^{(0)}, p^{(1)}, ..., p^{(m)}, ..., \text{ wobei } p^{(m)} \in \mathbb{R}^n \backslash \{0\}, \tag{3.51}$$

wählen.
Das AV ist wie in Definition 3.1 und hat im m-ten Schritt die Iterationsformel

$$x^{(m+1)} = x^{(m)} + \alpha_m p^{(m)} \tag{3.52}$$

mit der optimalen Wahl der Schrittzahl

$$\alpha_m = \frac{p^{(m)T} r^{(m)}}{p^{(m)T} A p^{(m)}}$$

in der Suchrichtung $p^{(m)}$.
Daraus ergibt sich

$$\begin{aligned} x^{(m+1)} &= x^{(0)} + \sum_{k=0}^{m} \alpha_k p^{(k)}, \\ x^{(m+1)} &\in x^{(0)} + \text{span}\{p^{(0)}, p^{(1)}, ..., p^{(m)}\}. \end{aligned} \tag{3.53}$$

Man sucht also von $x^{(m)}$ aus in Richtung des Vektors $p^{(m)}$ und findet die optimale Lösung $x^{(m+1)}$ aus dem Vektorrraum

$$x^{(0)} + \mathcal{P}_{m+1}, \quad \mathcal{P}_{m+1} = \text{span}\{p^{(0)}, p^{(1)}, ..., p^{(m)}\}. \tag{3.54}$$

Als Ergebnis wünscht man sich, dass bei exakter Rechnung spätestens für $m = n$ die Lösung x^* erreicht wird, und damit $x^{(m)} = A^{-1}b$, d. h. $r^{(m)} = 0$ gelten. Dies wird uns unter den gegebenen Voraussetzungen im Allgemeinen noch nicht gelingen.
Das Residuum genügt den Beziehungen

$$\begin{aligned} r^{(m)} &= b - Ax^{(m)}, \\ r^{(m+1)} &= r^{(m)} - \alpha_m A p^{(m)} \\ &= r^{(0)} - \sum_{k=0}^{m} \alpha_k A p^{(k)}. \end{aligned} \tag{3.55}$$

Welche Eigenschaften haben nun die aufeinander folgenden Residua $r^{(0)}, r^{(1)}, r^{(2)}, ...$?

$$
\begin{aligned}
r^{(0)} &= b - Ax^{(0)}, \\
r^{(1)} &= b - Ax^{(1)}, \\
& \quad r^{(1)T}p^{(0)} = (r^{(0)} - \alpha_0 Ap^{(0)})^T p^{(0)} \\
& \qquad = r^{(0)T}p^{(0)} - \frac{p^{(0)T}r^{(0)}}{p^{(0)T}Ap^{(0)}}\, p^{(0)T}Ap^{(0)} \\
& \qquad = r^{(0)T}p^{(0)} - p^{(0)T}r^{(0)} \\
& \qquad = 0, \\
r^{(1)} &\perp p^{(0)}, \quad r^{(1)} \perp \mathcal{P}_1, \\
r^{(2)} &= b - Ax^{(2)}, \\
& \quad r^{(2)T}p^{(1)} = (r^{(1)} - \alpha_1 Ap^{(1)})^T p^{(1)} \\
& \qquad = r^{(1)T}p^{(1)} - \frac{p^{(1)T}r^{(1)}}{p^{(1)T}Ap^{(1)}}\, p^{(1)T}Ap^{(1)} \\
& \qquad = 0, \\
r^{(2)} &\perp p^{(1)}, \\
r^{(2)} &\not\perp \mathcal{P}_2 = \operatorname{span}\{p^{(0)}, p^{(1)}\}, \text{ denn} \\
& \quad r^{(2)T}p^{(0)} = (r^{(1)} - \alpha_1 Ap^{(1)})^T p^{(0)} \\
& \qquad = \underbrace{r^{(1)T}p^{(0)}}_{=0} - \alpha_1 \underbrace{p^{(1)T}Ap^{(0)}}_{\text{i. Allg.} \neq 0}, \\
r^{(3)} &= b - Ax^{(3)}, \\
& \quad r^{(3)T}p^{(2)} = (r^{(2)} - \alpha_2 Ap^{(2)})^T p^{(2)} \\
& \qquad = r^{(2)T}p^{(2)} - \frac{p^{(2)T}r^{(2)}}{p^{(2)T}Ap^{(2)}}\, p^{(2)T}Ap^{(2)} \\
& \qquad = 0, \\
r^{(3)} &\perp p^{(2)}, \\
r^{(3)} &\not\perp \mathcal{P}_3 = \operatorname{span}\{p^{(0)}, p^{(1)}, p^{(2)}\}, \text{ denn i. Allg. wie oben} \\
& \quad r^{(3)T}p^{(0)} \neq 0, \quad r^{(3)T}p^{(1)} \neq 0.
\end{aligned}
$$

Damit erkennen wir zwar, dass $r^{(n)} \perp p^{(n-1)}$ ist, aber leider gilt

$$r^{(n)} \not\perp \mathcal{P}_n = \text{span}\{p^{(0)}, p^{(1)}, ..., p^{(n-1)}\} \; = \; \mathbb{R}^n$$

und damit ist das Verfahren **nicht** nach endlich vielen Schritten beendet.
Wir sehen jedoch die erforderliche Bedingung, um in jedem Schritt $r^{(m)} \perp \mathcal{P}_m$ zu erhalten, nämlich die A-Orthogonalität von $\{p^{(m)}\}$, d. h.

$$p^{(m)T} A p^{(k)} \; = \; 0, \quad k = 0, 1, ..., m-1. \tag{3.56}$$

Beispiel 3.7
Gegeben sei das triviale LGS $Ax = b$ mit $A = I$ und $b = (0, 0, 1)^T$, so dass $x^* = b$ ist.
Als Startvektor wählen wir $x^{(0)} = (0, 0, 0)^T$.
Die linear unabhängigen, aber nicht A-orthogonalen Suchrichtungen sind

$$p^{(0)} = (1, 0, 0)^T, \quad p^{(1)} = (1, 1, 0)^T, \quad p^{(2)} = (1, 1, 1)^T.$$

Es gilt

$$\begin{aligned} x^* &= \gamma_0 p^{(0)} + \gamma_1 p^{(1)} + \gamma_2 p^{(2)} \; = \; 0 \cdot p^{(0)} - 1 \cdot p^{(1)} + 1 \cdot p^{(2)}, \\ x^* &\in x^{(0)} + \mathcal{P}_3, \quad \mathcal{P}_3 = \text{span}\{p^{(0)}, p^{(1)}, p^{(2)}\}. \end{aligned}$$

Was macht das AV in den einzelnen Schritten?

$x^{(0)} = (0, 0, 0)^T$	Startvektor,
$r^{(0)} = b - Ax^{(0)} = b = (0, 0, 1)^T$	Anfangsresiduum.

S1

$$\begin{aligned} \alpha_0 &= \frac{p^{(0)T} r^{(0)}}{p^{(0)T} A p^{(0)}} \; = \; \frac{p^{(0)T} r^{(0)}}{p^{(0)T} p^{(0)}} \; = \; 0, \quad p^{(0)} = (1, 0, 0)^T \text{ 1. Abstiegsrichtung}, \\ x^{(1)} &= x^{(0)} + \alpha_0 p^{(0)} \; = \; x^{(0)} \; = \; (0, 0, 0)^T, \\ r^{(1)} &= b - Ax^{(1)} \; = \; b \; = \; (0, 0, 1)^T. \end{aligned}$$

S2

$$\begin{aligned} \alpha_1 &= \frac{p^{(1)T} r^{(1)}}{p^{(1)T} A p^{(1)}} \; = \; 0, \quad p^{(1)} = (1, 1, 0)^T \text{ 2. Abstiegsrichtung}, \\ x^{(2)} &= x^{(1)} \; = \; (0, 0, 0)^T, \\ r^{(2)} &= b \; = \; (0, 0, 1)^T. \end{aligned}$$

S3

$$\begin{aligned} \alpha_2 &= \frac{p^{(2)T} r^{(2)}}{p^{(2)T} A p^{(2)}} \; = \; \frac{1}{3}, \quad p^{(2)} = (1, 1, 1)^T \text{ 3. Abstiegsrichtung}, \\ x^{(3)} &= x^{(2)} + \alpha_2 p^{(2)} \; = \; \tfrac{1}{3}(1, 1, 1)^T, \\ r^{(3)} &= b - Ax^{(3)} \; = \; \tfrac{1}{3}(-1, -1, 2)^T \; \neq \; 0. \end{aligned}$$

Die Orthogonalitätsbeziehungen sind

$$
\begin{aligned}
r^{(1)} &\perp p^{(0)}, \quad r^{(1)} \perp \mathcal{P}_1, \\
r^{(2)} &\perp p^{(1)}, \quad \text{aber auch } r^{(2)} \perp p^{(0)}, \\
r^{(3)} &\perp p^{(2)}, \quad \text{aber } r^{(3)} \not\perp p^{(0)}, p^{(1)}, \quad r^{(3)} \not\perp \mathbb{R}^3.
\end{aligned}
$$

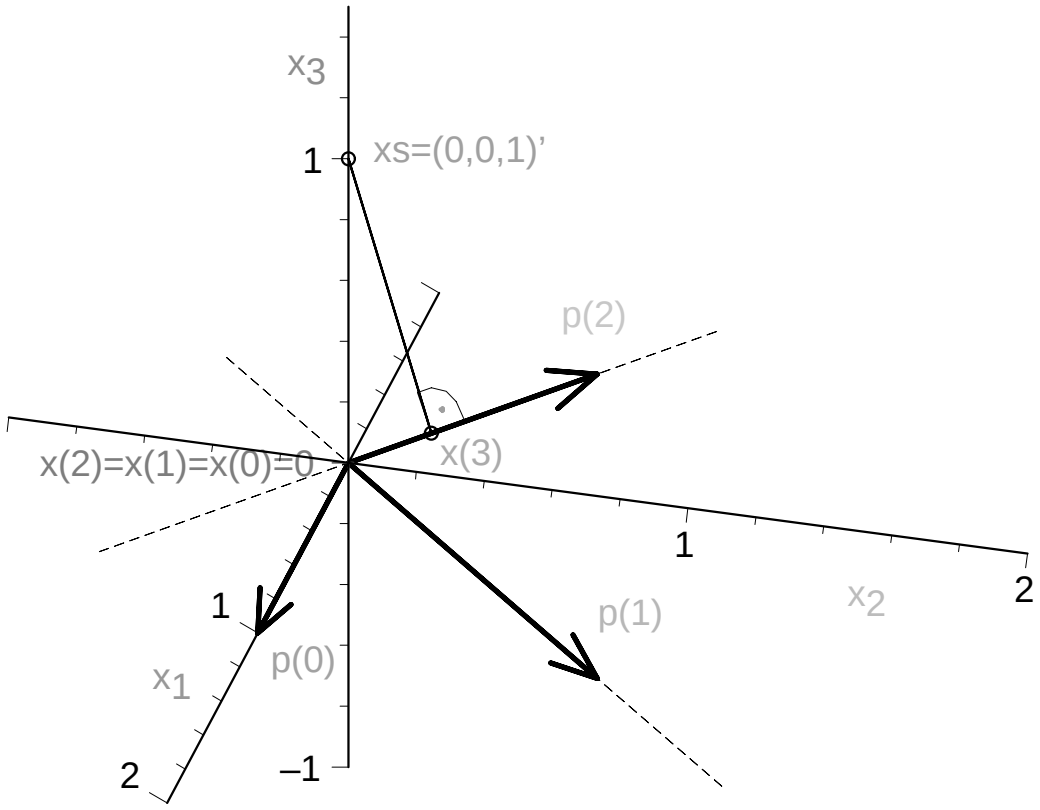


Abb. 3.12 Datei *abst*106.*ps*
Iterationsverlauf (3 Iterationen) des AV mit linear unabhängigen Suchrichtungen $p^{(0)}, p^{(1)}, p^{(2)}$ zu $Q(x) = \frac{1}{2}x^T Ax - x^T b = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) - x_3$

Der Iterationsverlauf zeigt am Anfang in den Richtungen $p^{(0)}$ und $p^{(1)}$ keinen Fortschritt, so dass $x^{(2)} = x^{(1)} = x^{(0)}$ ist.
Erst mit der Suchrichtung $p^{(2)}$ kommt etwas "Bewegung" in den Ablauf, aber es gilt $x^{(3)} = \frac{1}{3}(1,1,1)^T \neq x^*$.

Wir haben wegen $A = I$ und $b = x^*$ die Beziehungen

$$
\begin{aligned}
e^{(3)} &= x^* - x^{(3)} = \frac{1}{3}(-1,-1,2)^T, \\
\|e^{(m)}\|_A^2 &= \|e^{(m)}\|_I^2 = \|e^{(m)}\|_2^2, \quad \|e^{(3)}\|_A^2 = \frac{2}{3},
\end{aligned}
$$

$$
\begin{aligned}
r^{(m)} &= b - Ax^{(m)} = Ax^* - Ax^{(m)} = Ae^{(m)} = e^{(m)},\\
r^{(3)} &= e^{(3)} = \frac{1}{3}(-1,-1,2)^T \perp p^{(2)} = (1,1,1)^T,\\
0 &= (r^{(3)}, p^{(2)}) = (x^* - x^{(3)}, p^{(2)}),\\
\|r^{(3)}\|_2^2 &= \|e^{(3)}\|_2^2 = \frac{2}{3},\\
Q(x) &= \frac{1}{2}x^TAx - x^Tb = \frac{1}{2}x^Tx - x^Tx^* = \frac{1}{2}\|x\|_2^2 - x_3,\\
&= \frac{1}{2}\Big(\|e(x)\|_A^2 - x^{*T}b\Big) = \frac{1}{2}\Big(\|e(x)\|_2^2 - x^{*T}b\Big) = \frac{1}{2}\Big(\|r(x)\|_2^2 - x^{*T}b\Big),\\
Q(x^*) &= \frac{1}{2}\Big(\|e(x^*)\|_A^2 - x^{*T}b\Big) = -\frac{1}{2}x^{*T}b = -\frac{1}{2},\\
Q(x^{(3)}) &= -\frac{1}{6},
\end{aligned}
$$

sowie

$$
\min_\alpha \|x^* - (x^{(2)} + \alpha p^{(2)})\|_2 = \|x^* - x^{(3)}\|_2 = \sqrt{\frac{2}{3}}.
$$

Die Iterierte $x^{(3)}$ ist das Lot von x^* auf die Gerade $f(\alpha) = x^{(2)} + \alpha p^{(2)}$.
Wir wiederholen die Abstiegsschritte bei zyklischer Anwendung der Suchrichtungen. Dabei erkennen wir die relativ langsame lineare Konvergenz.

```
Iterationsverlauf mit verschiedenen Fehlern

  m   [ x(m)[1],   x(m)[2],   x(m)[3]  ]        Q(x(m))          ||xs-x(m)||_A^2
      [ r(m)[1],   r(m)[2],   r(m)[3]  ]                         =||r(m)||_2^2
-------------------------------------------------------------------------------------
Start
  0   [ 0,         0,         0        ]   0                     1
      [ 0,         0,         1        ]
-------------------------------------------------------------------------------------
  1   [ 0,         0,         0        ]   0                     1
      [ 0,         0,         1        ]

  2   [ 0,         0,         0        ]   0                     1
      [ 0,         0,         1        ]

  3   [ 1/3,       1/3,       1/3      ]   -1/6=-0.166667        2/3=0.666667
      [-1/3,      -1/3,       2/3      ]

  4   [ 0,         1/3,       1/3      ]   -2/9=-0.222222        5/9=0.555556
      [ 0,        -1/3,       2/3      ]

  5   [-1/6,       1/6,       1/3      ]   -1/4=-0.25            1/2=0.5
      [ 1/3,      -1/6,       2/3      ]

  6   [ 1/18,      7/18,      5/9      ]   -35/108=-0.324074     19/54=0.351852
      [-1/18,     -7/18,      4/9      ]
```

```
 7   [ 0,         7/18,      5/9      ]   -211/648=-0.325617       113/324=0.348765
     [ 0,        -7/18,      4/9      ]

 8   [-7/36,      7/36,      5/9      ]   -157/432=-0.363426       59/216=0.273148
     [ 7/36,     -7/36,      4/9      ]

 9   [-5/108,     37/108,    19/27    ]   -1541/3888=-0.396348     403/1944=0.207305
     [ 5/108,    -37/108,    8/27     ]

10   [ 0,         37/108,    19/27    ]   -9271/23328=-0.397419    2393/11664=0.205161
     [ 0,        -37/108,    8/27     ]

11   [-37/216,    37/216,    19/27    ]   -6637/15552=-0.426762    1139/7776=0.146476
     [ 37/216,   -37/216,    8/27     ]

12   [-47/648,    175/648,   65/81    ]   -61781/139968=-0.441394  8203/69984=0.117212
     [ 47/648,   -175/648,   16/81    ]

21   [-0.042903,  0.101430,  0.941472 ]   -0.492223                0.015554
     [ 0.042903, -0.101430,  0.058528 ]

30   [-0.015388,  0.032730,  0.982658 ]   -0.499196                0.001609
     [ 0.015388, -0.032730,  0.017342 ]

60   [-0.000299,  0.000600,  0.999699 ]   -0.499999730             0.000000539
     [ 0.000299, -0.000600,  0.000301 ]
```

Auf Grund der Besonderheit der Suchrichtungen unterliegen die Iterierten und Residua in ihren Komponenten einer gewissen Regelmäßigkeit und Wiederholung.

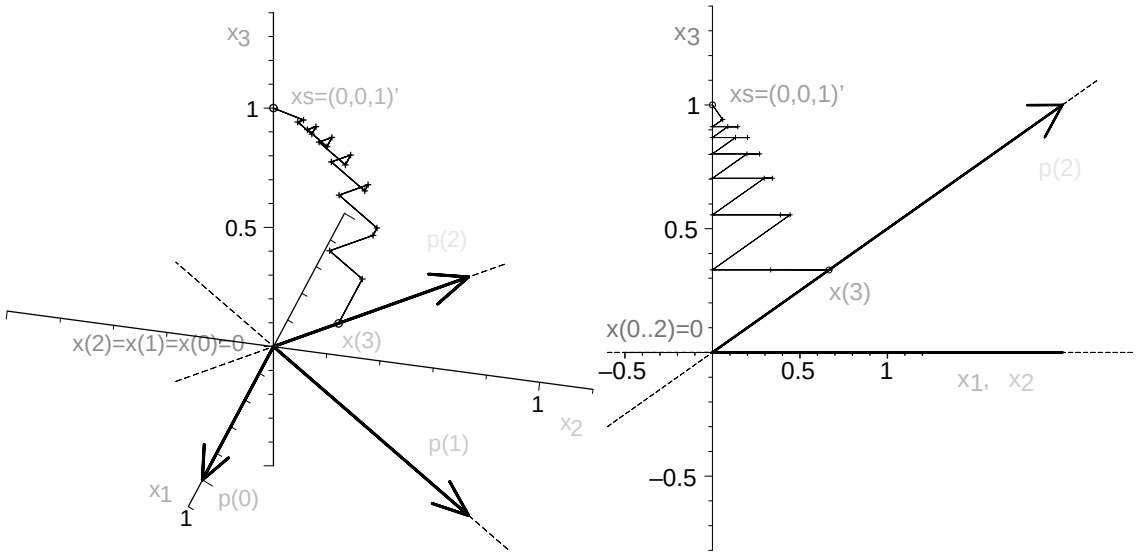


Abb. 3.13 Dateien *abst*1061.*ps*, *abst*1062.*ps*
Iterationsverlauf (21 Iterationen) des AV mit linear unabhängigen Suchrichtungen $p^{(0)}, p^{(1)}, p^{(2)}$ (bei zyklischer Verwendung) zu $Q(x) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) - x_3$,
Maple `orientation=[15,60]`, `orientation=[315,90]` (Blick parallel zur Ebene (x_1, x_2))

3.4 Optimalität im Abstiegsverfahren

Wir gewinnen aus dem AV mit linear unabhängigen Suchrichtungen die folgende Erkenntnis.

Man braucht im AV nicht nur eine lokale Optimalität bezüglich der Suche in einer bestimmten Richtung (Strahlenminimierung), sondern auch noch eine Optimalität bezüglich eines Unterraums (Orthogonalität).

Definition 3.3 Optimalität zum Funktional $Q(x)$
Der Vektor $x \in \mathbb{R}^n$ heißt optimale Lösung von $\min_x Q(x)$

(a) bezüglich der Suchrichtung $p \in \mathbb{R}^n \backslash \{0\}$, falls

$$Q(x) \;\leq\; Q(x+\alpha p) \quad \forall\, \alpha \in \mathbb{R}, \tag{3.57}$$

(b) bezüglich des Unterraums $U \subset \mathbb{R}^n$, falls

$$Q(x) \;\leq\; Q(x+\Delta x) \quad \forall\, \Delta x \in U. \tag{3.58}$$

Wann ist x optimal bezüglich $U \subset \mathbb{R}^n$?
Dazu betrachten wir wie in (3.10) aus der Strahlenminimierung die Funktion

$$\begin{aligned} f(\alpha) &= Q(x+\alpha\Delta x) \\ &= \frac{1}{2}(x+\alpha\Delta x)^T A(x+\alpha\Delta x) - (x+\alpha\Delta x)^T b \\ &= Q(x) + \alpha(Ax-b,\Delta x) + \frac{1}{2}\alpha^2(A\Delta x,\Delta x), \\ f'(\alpha) &= (Ax-b,\Delta x) + \alpha(A\Delta x,\Delta x). \end{aligned}$$

Der Vektor x ist optimal bezüglich U mit $0 \neq \Delta x \in U$,

$$\begin{aligned} &\Leftrightarrow \quad f'(0) = 0 \\ &\Leftrightarrow \quad (Ax-b,\Delta x) = 0 \\ &\Leftrightarrow \quad r(x) \perp U. \end{aligned}$$

Wann erhält man die Optimalität bei gegebenen Suchrichtungen $p^{(j)}$?
Sei

$$x^{(m)} \;\in\; x^{(0)} + \mathcal{P}_m, \quad \mathcal{P}_m = \operatorname{span}\{p^{(0)}, p^{(1)}, ..., p^{(m-1)}\}.$$

Gilt, dass $x^{(m)}$ optimal bez. $\mathcal{P}_m$ ist, d. h. $r^{(m)} \perp \mathcal{P}_m$, und

$$x^{(m+1)} \;=\; x^{(m)} + \alpha_m p^{(m)}, \quad \alpha_m \;=\; \frac{p^{(m)T} r^{(m)}}{p^{(m)T} A p^{(m)}},$$

dann haben wir für beliebiges $\xi \in \mathcal{P}_m$ die Beziehung

$$\begin{aligned}
(r^{(m+1)}, \xi) &= (b - Ax^{(m+1)}, \xi) \\
&= (b - A(x^{(m)} + \alpha_m p^{(m)}), \xi) \\
&= \underbrace{(b - Ax^{(m)}, \xi)}_{=0} - \alpha_m (Ap^{(m)}, \xi) \\
&= -\alpha_m (Ap^{(m)}, \xi).
\end{aligned}$$

Es gilt $(r^{(m+1)}, \xi) = 0$ für alle $\xi \in \mathcal{P}_m$, falls

$$(r^{(m+1)}, p^{(j)}) \;=\; -\alpha_m (Ap^{(m)}, p^{(j)}) \;=\; 0, \quad j = 0, 1, ..., m-1,$$

ist, d. h. wir brauchen die Bedingung

$$0 \;=\; (Ap^{(m)}, p^{(j)}) \;=\; p^{(j)T} A p^{(m)} \;=\; p^{(m)T} A p^{(j)}, \quad j = 0, 1, ..., m-1. \tag{3.59}$$

Das bedeutet aber die A-Orthogonalität der Suchrichtungen $p^{(j)}$.
Weiterhin berechnen wir unter Verwendung von α_m das Skalarprodukt

$$\begin{aligned}
(r^{(m+1)}, p^{(m)}) &= (r^{(m)}, p^{(m)}) - \frac{p^{(m)T} r^{(m)}}{p^{(m)T} A p^{(m)}} (Ap^{(m)}, p^{(m)}) \\
&= (r^{(m)}, p^{(m)}) - (p^{(m)}, r^{(m)}) \\
&= 0,
\end{aligned}$$

womit dann

$$r^{(m+1)} \;\perp\; \mathcal{P}_{m+1} \;=\; \text{span}\{\mathcal{P}_m, p^{(m)}\}, \quad \mathcal{P}_0 = \emptyset, \tag{3.60}$$

und $x^{(m+1)}$ optimal bezüglich $\mathcal{P}_{m+1}$ ist.
Falls sich die Anzahl der A-orthogonalen Richtungen bis auf n erstreckt, also $p^{(0)}, p^{(1)}, ..., p^{(n-1)}$ vorliegen, erhält man

$$r^{(n)} \;\perp\; \mathcal{P}_n \;=\; \text{span}\{p^{(0)}, p^{(1)}, ..., p^{(n-1)}\} \;=\; \mathbb{R}^n, \tag{3.61}$$

was für das Residuum $r^{(n)} = 0$ bedeutet.
Damit ist die (theoretische) Endlichkeit des AV gegeben.

3.5 Abstiegsverfahren mit konjugierten Richtungen

Mit dem Hinweis auf den vorherigen Abschnitt nehmen wir als Suchrichtungen die A-orthogonalen bzw. konjugierten Vektoren

$$p^{(0)}, p^{(1)}, ..., p^{(m)}, \quad m \leq n-1, \quad p^{(m)} \in \mathbb{R}^n \backslash \{0\}. \tag{3.62}$$

Ihre A-Orthogonalität heißt

$$(Ap^{(i)}, p^{(j)}) \;=\; p^{(i)T} A p^{(j)} \;=\; 0 \quad \forall\, i \neq j. \tag{3.63}$$

Satz 3.8 *Ein System von* A*-orthogonalen Vektoren ist linear unabhängig.*

Beweis. Seien die Vektoren $z_i,\ i = 1, 2, ..., k$, A-orthogonal und

$$\sum_{i=1}^{k} c_i z_i = 0, \quad c_i \in \mathbb{R}.$$

Daraus folgt für beliebiges $j \in \{1, 2, ..., k\}$

$$0 \;=\; z_j^T A \sum_{i=1}^{k} c_i z_i \;=\; \sum_{i=1}^{k} c_i z_j^T A z_i \;=\; c_j z_j^T A z_j, \quad z_j^T A z_j > 0$$

und damit $c_j = 0$, was die lineare Unabhängigkeit der Vektoren z_i bedeutet. □

Natürlich gilt die Behauptung des Satzes nicht in umgekehrter Richtung.

Realisierung des AV als (endliches) Iterationsverfahren mit Indizierung

$x^{(0)}$	Startvektor,
$r^{(0)} = b - Ax^{(0)}$	Anfangsresiduum,
ε	Toleranz für den Test auf Abbruch der Iteration,
$p^{(0)}, p^{(1)}, ..., p^{(m)}, ...$	A-orthogonale Suchrichtungen.

m = 0,1,2,...≤n–1

$$\begin{aligned}
&\text{falls } \| r^{(m)} \| < \varepsilon, \quad \text{dann break} \\
\alpha_m \;&=\; \frac{p^{(m)T} r^{(m)}}{p^{(m)T} w^{(m)}}, \quad \text{wobei} \quad w^{(m)} = Ap^{(m)} \\
x^{(m+1)} \;&=\; x^{(m)} + \alpha_m p^{(m)} \\
r^{(m+1)} \;&=\; r^{(m)} - \alpha_m w^{(m)} \quad \text{oder} \quad r^{(m+1)} = b - Ax^{(m+1)}
\end{aligned} \tag{3.64}$$

end m

Näherungslösung $x^* = x^{(m)}$

Bemerkung 3.3 Im Abbruchkriterium kann man auch den relativen Fehler der Residua $\|r^{(m)}\|/\|r^{(0)}\|,\ m \geq 1$, einbeziehen.

Das erste Argument, welches für die Wahl zueinander konjugierter Richtungen spricht, ist, dass das Verfahren (3.64) im Gegensatz zu den AV des Abschnitts 3.3 bereits nach endlich vielen Schritten gegen die Lösung konvergiert.
Dazu formulieren wir den folgenden Satz, dessen Beweis sich an die Betrachtungen im Abschnitt 3.4 anlehnt.

Satz 3.9 *Sind die Richtungen $p^{(0)}, p^{(1)}, ..., p^{(n-1)}$ zueinander konjugiert gewählt, so konvergiert das AV (3.64) für beliebige Startvektoren $x^{(0)}$ nach höchstens n Schritten gegen die Lösung von (1.1).*

Beweis. Wir untersuchen die Komponenten der Residua $r^{(k)} = b - Ax^{(k)}$ in Richtung der Vektoren $p^{(j)}$.
(1) $j < k$
Wegen $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$ ist

$$(r^{(k+1)}, p^{(j)}) = (b - A(x^{(k)} + \alpha_k p^{(k)}), p^{(j)}) = (r^{(k)}, p^{(j)}) - \alpha_k \underbrace{(Ap^{(k)}, p^{(j)})}_{=0} = (r^{(k)}, p^{(j)}).$$

Das bedeutet, dass sich im Verlauf der Iteration für $k > j$ diejenige Komponente des Residuums, welche in Richtung $p^{(j)}$ zeigt, nicht mehr ändert.

(2) $j = k$
Wegen

$$\alpha_k \ = \ \frac{(r^{(k)}, p^{(k)})}{(Ap^{(k)}, p^{(k)})}$$

ist

$$(r^{(k+1)}, p^{(k)}) \ = \ (r^{(k)}, p^{(k)}) - \alpha_k (Ap^{(k)}, p^{(k)}) \ = \ 0.$$

Damit gilt für $k = 0, 1, ..., n-1$ die Beziehung

$$(r^{(n)}, p^{(k)}) \overset{(1)}{=} (r^{(n-1)}, p^{(k)}) \overset{(1)}{=} ... \overset{(1)}{=} (r^{(k+1)}, p^{(k)}) \overset{(2)}{=} 0.$$

Da $p^{(0)}, p^{(1)}, ..., p^{(n-1)}$ eine Basis des $\mathbb{R}^n$ bilden, muss gelten $r^{(n)} = b - Ax^{(n)} = 0$. $\square$

Natürlich gilt nicht nur die Orthogonalität von $r^{(n)}$ zu allen konjugierten Suchrichtungen, sondern auch für die Zwischenschritte jeweils $(r^{(k+1)}, p^{(j)}) = 0,\ j = 0, 1, ..., k$.

Das zweite Argument für die Wahl konjugierter Richtungen besteht darin, dass solche Richtungen ohne allzu großen Rechenzeit- und Speicheraufwand bestimmt werden können. Betrachten wir jedoch zunächst zwei Beispiele, wie man es nicht machen sollte.

Beispiel 3.8

(a) Da die Matrix A spd ist, gibt es einen Satz $v_1, v_2, ..., v_n$ von paarweise senkrecht aufeinander stehenden Eigenvektoren, d. h. $v_i^T v_j = 0$ für $i \neq j$. Damit sind diese Vektoren auch A-orthogonal, denn

$$(v_i, v_j)_A \;=\; v_i^T A v_j \;=\; \lambda_j v_i^T v_j \;=\; 0.$$

Da die Bestimmung der Eigenvektoren in der Regel sehr aufwändig ist, ist diese Wahl praktisch unbrauchbar.

(b) Gegeben seien beliebige linear unabhängige Vektoren $y_1, y_2, ..., y_n$.
Mit Hilfe des Gram-Schmidt-Verfahrens können hieraus n zueinander konjugierte Richtungen $p^{(0)}, p^{(1)}, ..., p^{(n-1)}$ konstruiert werden. Man überzeugt sich jedoch leicht, dass für diese Konstruktion alle Richtungsvektoren abgespeichert werden müssen. Darüber hinaus ist der Rechenaufwand sehr groß.

Berechnungen in Maple

Prozedur für die A-Orthogonalisierung von n linear unabhängigen Vektoren nach Gram-Schmidt
– Vektoren als Spaltenvektoren in Matrix `Y(n,n)` $\Rightarrow$ `Q` mit A-orth. Spalten
– Test auf Durchführbarkeit mit gegebener Toleranz `etol`
– Matrix $A(n,n) = A^T > 0$
– Ergebnis ist `[j,Q,R]`, Durchführbarkeit bis j-te Spalte von `Y`

```
> AOrth_GS:=proc(n::posint, Y::matrix, A::matrix, etol::numeric)
    local Q,R,p,ps,y,nen,r,w,s,i,j;

  # Initialisierungen
  R:=array(sparse,1..n,1..n);      # R:=array(identity,1..n,1..n);
  for i from 1 by 1 to n do
   R[i,i]:=1;
  end do;
  Q:=matrix(n,0,[]);
  y:=vector(n,[]);
  p:=evalm(y);
  nen:=evalm(y);

  # Schritt j=1
  y:=evalm(col(Y,1));
  p:=evalm(y);
  w:=evalm(A&*p);
  nen[1]:=evalm(transpose(w)&*p);
  Q:=concat(Q,p);
  if nen[1]<etol then RETURN(1,Q,R); end if;

  # Schritte j=2,3,...,n
  for j from 2 by 1 to n do
   y:=evalm(col(Y,j));
   ps:=evalm(y);
```

```
 for i from 1 by 1 to j-1 do
  p:=evalm(col(Q,i));
  w:=evalm(A&*p);
  s:=evalm(transpose(w)&*y);
  r:=s/nen[i];
  R[i,j]:=r;
  ps:=evalm(ps-r*p);
 end do;
 p:=evalm(ps);
 w:=evalm(A&*p);
 s:=evalm(transpose(w)&*p);
 nen[j]:=s;
 Q:=concat(Q,p);
 if s<etol then break end if;
end do:
if j>n then j:=j-1; end if;
[j,Q,R];
end:
```

Matrizen

```
> n:=4:
  fehler:=1E-6:
  A:=matrix(n,n,[[ 2,-1, 0, 0],
                 [-1, 2,-1, 0],
                 [ 0,-1, 2,-1],
                 [ 0, 0,-1, 2]]):
  Y:=matrix(n,n,[[ 1, 1, 1, 1],
                 [ 0, 2, 2, 2],
                 [ 1, 0, 3, 3],
                 [ 1, 0, 0, 4]]):
  rank(Y);
```

$$4$$

A-Orthogonalisierung der linear unabhängigen Spalten von Y
Q enthält A-orthogonale Vektoren und R obere Dreiecksmatrix, so dass zusätzlich $Y = QR$ ist.

```
> erg:=AOrth_GS(n,Y,A,fehler);
  erg[1];
  Q:=evalm(erg[2]);
  R:=evalm(erg[3]);
```

$$erg := [4, Q, R]$$

$$4$$

$$Q := \begin{bmatrix} 1 & \frac{3}{2} & \frac{3}{5} & \frac{-12}{13} \\ 0 & 2 & \frac{9}{5} & \frac{16}{13} \\ 1 & \frac{1}{2} & \frac{27}{10} & \frac{24}{13} \\ 1 & \frac{1}{2} & \frac{-3}{10} & \frac{32}{13} \end{bmatrix}$$

$$R := \begin{bmatrix} 1 & \frac{-1}{2} & \frac{1}{4} & \frac{5}{4} \\ 0 & 1 & \frac{1}{10} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{-5}{39} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Kontrolle

```
> evalm(transpose(Q)&*A&*Q); # A-Orthogonalitaet
  evalm(transpose(Q)&*Q);    # keine Orthogonalitaet
  evalm(Q&*R):               # Y=QR
```

$$\begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & \frac{117}{10} & 0 \\ 0 & 0 & 0 & \frac{160}{13} \end{bmatrix}$$

$$\begin{bmatrix} 3 & \frac{5}{2} & 3 & \frac{44}{13} \\ \frac{5}{2} & \frac{27}{4} & \frac{57}{10} & \frac{42}{13} \\ 3 & \frac{57}{10} & \frac{549}{50} & \frac{384}{65} \\ \frac{44}{13} & \frac{42}{13} & \frac{384}{65} & \frac{2000}{169} \end{bmatrix}$$

Will man die Suchrichtungen $p^{(j)}$ während der Durchführung des AV erst konstruieren, dann wählt man dazu einen geeigneten Ansatz.
Man gibt sich eine Startsuchrichtung $p^{(0)}$ vor und erzeugt rekursiv aus dem Vektorsystem $\{p^{(0)}, p^{(2)}, ..., p^{(m-1)}\}$ unter Einbeziehung des nicht verschwindenden Residuums $r^{(m)} \perp \mathcal{P}_m$ den nächsten Vektor $p^{(m)}$.
Der Ansatz ist

$$p^{(m)} \;=\; r^{(m)} + \sum_{j=0}^{m-1} \beta_j p^{(j)}. \tag{3.65}$$

Die reellen Koeffizienten β_j lassen sich einfach aus der A-Orthogonalität von $p^{(j)}$ bestimmen. Es gilt

$$\begin{aligned} p^{(k)T} A p^{(m)} &= p^{(k)T} A r^{(m)} + \sum_{j=0}^{m-1} \beta_j \, p^{(k)T} A p^{(j)}, \quad k = 0, 1, ..., m-1, \\ 0 &= p^{(k)T} A r^{(m)} + \beta_k \, p^{(k)T} A p^{(k)}, \end{aligned}$$

somit

$$\beta_k \;=\; -\frac{p^{(k)T} A r^{(m)}}{p^{(k)T} A p^{(k)}}, \quad k = 0, 1, ..., m-1, \tag{3.66}$$

und

$$p^{(m)} = r^{(m)} + \sum_{j=0}^{m-1} \left(-\frac{p^{(k)T} A r^{(m)}}{p^{(k)T} A p^{(k)}} \right) p^{(j)} = r^{(m)} - \sum_{j=0}^{m-1} \frac{(Ap^{(k)}, r^{(m)})}{(Ap^{(k)}, p^{(k)})} p^{(j)}. \tag{3.67}$$

Wie schon bemerkt, sind bei dieser Vorgehensweise alle bisherigen Suchrichtungen zu speichern.

Zusammenfassung und Bewertung der Eigenschaften des AV
in der Skala $\{+, \pm, -\}$ unter der Voraussetzung $A = A^T > 0$

(3.68)

- $+$ Bei exakter Rechnung ist spätestens $x^{(n)} = A^{-1}b$, damit erhält man die genannte Optimalität. Bei $r^{(m)} = 0$, $m \leq n-1$, haben wir ein vorzeitiges Ende des AV.
- $+$ Die Residua $r^{(m)}$ sind orthogonal zu $\mathcal{P}_m$, aber untereinander nicht zwingend linear unabhängig.
- $+$ Einfache Implementation des Algorithmus.
- $\pm$ Das Abstiegsszenario und die Minimierungsaufgabe sind dem Problem angepasst, nicht aber die vorgegebenen Suchrichtungen.
- $\pm$ Hauptaufwand in einem Iterationsschritt ist eine Matrix-Vektor-Multiplikation.
- $\pm$ Der Lösungsfehler $\|e^{(m)}\|_A = \|x^* - x^{(m)}\|_A$ wird ständig verkleinert, aber nicht unbedingt $\|r^{(m)}\|_2 = \|b - Ax^{(m)}\|_2$.
  Für ein Abbruchkriterium kann man natürlich die Größe $\|e^{(m)}\|_A$ wegen der darin enthaltenen unbekannten Lösung x^* nicht verwenden.
- $\pm$ Eine ungeschickte Reihenfolge der Suchrichtungen verursacht einen ungünstigen Fehlerverlauf, so dass dann eventuell n Iterationsschritte zu machen sind.
  Wünschenswert wäre z. B., dass das System der Suchrichtungen erst im Iterationsprozess auf einfache Weise generiert wird und dabei von zusätzlichen Informationen aus den Residua lebt.
- $-$ Bei numerischen Rechnungen kann der Fall $r^{(n)} \neq 0$ eintreten, so dass eine Fortsetzung des AV sinnvoll ist, möglichst verbunden mit genaueren Rechnungen (verbesserte Gleitpunktarithmetik).
- $-$ Speicherbedarf zum Merken aller Suchrichtungen.

Beispiel 3.9

Wir nehmen das LGS aus Beispiel 3.7 mit $A = I$, $b = (0,0,1)^T = x^*$ und dem Startvektor $x^{(0)} = (0,0,0)^T$.

Wir testen vier Varianten der Wahl der konjugierten Suchrichtungen $p^{(j)}$:

(1) $p^{(j)},\ j = 1,2,3$, gegeben,

(2) 1. Berechnung von $p^{(j)},\ j \geq 1,\ p^{(0)}$ gegeben,

(3) 2. Berechnung von $p^{(j)},\ j \geq 1,\ p^{(0)} = r^{(0)} \neq 0$,

(4) 3. Berechnung von $p^{(j)},\ j \geq 1,\ p^{(0)}$ gegeben.

(1) Die gegebenen A-orthogonalen und damit linear unabhängigen Suchrichtungen sind der Reihe nach die Einheitsvektoren $p^{(0)} = e_1,\ p^{(1)} = e_2,\ p^{(2)} = e_3$.

Ablauf des Iterationsprozesses:

$$
\begin{aligned}
x^{(0)} &= (0,0,0)^T, \\
r^{(0)} &= b - Ax^{(0)} = (0,0,1)^T, \\
\alpha_0 &= \frac{p^{(0)T} r^{(0)}}{p^{(0)T} A p^{(0)}} = 0, \quad p^{(0)} = (1,0,0)^T, \\
x^{(1)} &= x^{(0)} + \alpha_0 p^{(0)} = (0,0,0)^T, \\
r^{(1)} &= b - Ax^{(1)} = (0,0,1)^T, \\
\alpha_1 &= \frac{p^{(1)T} r^{(1)}}{p^{(1)T} A p^{(1)}} = 0, \quad p^{(1)} = (0,1,0)^T, \\
x^{(2)} &= x^{(1)} + \alpha_1 p^{(1)} = (0,0,0)^T, \\
r^{(2)} &= b - Ax^{(2)} = (0,0,0)^T, \\
\alpha_2 &= \frac{p^{(2)T} r^{(2)}}{p^{(2)T} A p^{(2)}} = 1, \quad p^{(2)} = (0,0,1)^T, \\
x^{(3)} &= x^{(2)} + \alpha_2 p^{(2)} = (0,0,1)^T, \\
r^{(3)} &= b - Ax^{(3)} = (0,0,0)^T \quad \text{und Ende.}
\end{aligned}
$$

Wir bemerken, dass wir die Suchrichtungen gerade in der ungünstigsten Reihenfolge ausgewählt haben. Es gelten allgemein bei $A = I$ und $x^{(0)} = x^* - A^{-1}p^{(n-1)}$ die Beziehungen $r^{(0)} = p^{(n-1)}$, $x^{(0)} = x^{(1)} = ... = x^{(n-1)}$ und $x^{(n)} = x^*$.

(2) Wahl von $p^{(0)} = (1,0,0)^T \neq r^{(0)}$ und der Formel (3.67).

$$
\begin{aligned}
x^{(0)} &= (0,0,0)^T, \quad r^{(0)} = (0,0,1)^T, \quad p^{(0)} = (1,0,0)^T, \quad \alpha_0 = 0, \\
x^{(1)} &= (0,0,0)^T, \quad r^{(1)} = (0,0,1)^T, \\
p^{(1)} &= r^{(1)} - \frac{(Ap^{(0)}, r^{(1)})}{(Ap^{(0)}, p^{(0)})}\, p^{(0)} = r^{(1)} = (0,0,1)^T,
\end{aligned}
$$

$$\begin{aligned}
\alpha_1 &= \frac{p^{(1)T} r^{(1)}}{p^{(1)T} A p^{(1)}} = 1, \\
x^{(2)} &= x^{(1)} + \alpha_1 p^{(1)} = (0,0,1)^T, \\
r^{(2)} &= b - Ax^{(2)} = (0,0,0)^T \quad \text{und “vorzeitiges“ Ende.}
\end{aligned}$$

Hier haben wir einen günstigen Ablauf erhalten.

(3) Wahl von $p^{(0)} = r^{(0)} = (0,0,1)^T$ und der Formel (3.67).

$$\begin{aligned}
x^{(0)} &= (0,0,0)^T, \quad r^{(0)} = (0,0,1)^T, \quad p^{(0)} = (0,0,1)^T, \\
\alpha_0 &= 1, \\
x^{(1)} &= (0,0,1)^T, \\
r^{(1)} &= (0,0,0)^T \quad \text{und “vorzeitiges“ Ende.}
\end{aligned}$$

Hier haben wir mit der Wahl der ersten Suchrichtung als Richtung des steilsten Abstiegs einen besonders günstigen Ablauf erhalten.

(4) Wahl von $p^{(0)} = (1,0,0)^T \neq r^{(0)}$, aber etwas andere Berechnung der weiteren A-orthogonalen Suchrichtungen.

$$\begin{aligned}
x^{(0)} &= (0,0,0)^T, \quad r^{(0)} = (0,0,1)^T, \quad p^{(0)} = (1,0,0)^T, \quad \|p^{(0)}\|_2 = 1, \\
p^{(1)} &= (\alpha,\beta,0)^T \quad \text{(Ansatz)}, \\
&\quad 0 = (Ap^{(0)}, p^{(1)}) = \alpha \quad \text{(A-Orthogonalität)}, \\
&\quad 1 = \|p^{(1)}\|_2 = |\beta|, \ \beta = 1, \\
p^{(1)} &= (0,1,0)^T, \quad \|p^{(1)}\|_2 = 1, \\
p^{(2)} &= (\alpha,\beta,\gamma)^T \quad \text{(Ansatz)}, \\
&\quad 0 = (Ap^{(0)}, p^{(2)}) = \alpha, \\
&\quad 0 = (Ap^{(1)}, p^{(2)}) = \beta, \\
&\quad 1 = \|p^{(2)}\|_2 = |\gamma|, \ \gamma = 1, \\
p^{(2)} &= (0,0,1)^T, \quad \|p^{(2)}\|_2 = 1.
\end{aligned}$$

Es entsteht eine Situation wie im Fall (1).

Ausgehend von den Eigenschaften des AV in (3.68) und der Formel (3.67) ist es unser Ziel, den Aufwand bei der Bestimmung uns Speicherung der Suchrichtungen $p^{(j)}$ durch entprechend kürzere Berechnungsformeln (verkürzte Rekursion) zu verringern.

3.6 Verfahren der konjugierten Gradienten

Das Verfahren der konjugierten Gradienten oder Richtungen (CG), wobei das Kürzel von *conjugate gradient method* bzw. *conjugate gradient acceleration* herrührt, benutzt auch die Voraussetzung $A = A^T > 0$.

Es vereinigt alle bisher genannten positiven Eigenschaften von AV.

– Abstiegsszenario und die Minimierungsaufgabe sind dem Problem angepasst.
– Die Residua $r^{(m)}$ sind orthogonal zum Unterraum $\mathcal{P}_m = \text{span}\{p^{(0)}, p^{(2)}, ..., p^{(m-1)}\}$ von A-orthogonalen Suchrichtungen $p^{(j)}$.
– Bei exakter Rechnung ist spätestens $r^{(n)} \perp \mathcal{P}_n = \mathbb{R}^n$ und damit $r^{(n)} = 0$ bzw. $x^{(n)} = A^{-1}b$. Das bedeutet Optimalität bez. des Unterraums.
– Die theoretische Schrittzahl übersteigt nicht n, praktisch werden es jedoch durch Rundungsfehler meist einige Schritte mehr. Dann erfolgt ein so genannter Restart des Verfahrens. Am einfachsten ist, mit den Formeln weiter zu iterieren. Man kann aber auch neu starten mit $x^{(0)} := x^{(n)}$ und $p^{(0)} := r^{(n)} \neq 0$ oder wie üblich mit $x^{(0)} := x^{(n)}$ und $p^{(0)} := r^{(0)} = b - Ax^{(0)}$.
– Einfache Implementation des Algorithmus.
– Hauptaufwand in einem Iterationsschritt ist eine Matrix-Vektor-Multiplikation.
– Die Suchrichtungen $p^{(j)}$ werden rekursiv im Iterationsverlauf erzeugt.

Dazu kommen noch einige weitere nützliche Eigenschaften, auf die bereits hingewiesen wurde und die sich als vorteilhaft erweisen werden.

Das betrifft die folgenden Aspekte.

– Das System der Suchrichtungen $p^{(j)}$ wird auf einfache Weise durch eine verkürzte Rekursion generiert.
– Nur die letzten zwei konjugierten Suchrichtungen sind zu merken.
– Es gilt $r^{(m)} = 0$ gdw. $p^{(m)} = 0$.
– Die Residua $r^{(j)}$ sind zueinander orthogonal und es gibt weitere Bedingungen der Orthogonalität bzw. A-Orthogonalität, so zum Beispiel die Bedingung $(Ar^{(m)}, p^{(j)}) = 0,\ j = 0, 1, ..., m-2$.
– Nicht nur der Lösungsfehler $\|e^{(m)}\|_A = \|x^* - x^{(m)}\|_A$ wird ständig verkleinert, sondern auch die Fehlernorm $\|e^{(m)}\|_2$ (siehe [11]).
– Anwendung für LGS mit großen schwach besetzten Matrizen.

Der entscheidende Aspekt im CG für das LGS ist, dass bei der Berechnung der Suchrichtung $p(x)$ (Abstiegsrichtung, Relaxationsrichtung, Projektionsvektor) der Gradient grad $Q(x) = \nabla Q(x) = Ax - b$ (Richtung des steilsten Anstiegs) als Abstiegsrichtung und Residuum $r(x) = -\text{grad } Q(x)$ einbezogen wird.
Der aktuelle Richtungsvektor $p(x)$ berechnet sich bei Anwendung der Eigenschaft der A-Orthogonalität aus der Linearkombination von $r(x)$ und vorheriger Suchrichtung, wobei die erste Suchrichtung gleich $r(x)$ ist.

3.6.1 Beschreibung als (endliches) Iterationsverfahren

Die Durchführung der Strahlenminimierung mit der notwendigen Bedingung beim Minimum des Funktionals

$$Q(x) \;=\; \frac{1}{2}x^T Ax - x^T b \;=\; \frac{1}{2}\Big(\|e(x)\|_A^2 - x^{*T} b\Big) \tag{3.69}$$

erlaubt die Berechnung des Parameters α im CG und liefert die monoton fallende Folge von Funktionswerten.

$$\min_{\alpha}\; Q(x^{(m)} + \alpha p^{(m)}) \;\Rightarrow\; \frac{\partial Q}{\partial \alpha} = 0,$$

$$\alpha \;=\; \alpha_m \;=\; \frac{p^{(m)T} r^{(m)}}{p^{(m)T} A p^{(m)}}, \tag{3.70}$$

$$Q(x^{(0)}) \geq Q(x^{(1)}) \geq ... \geq Q(x^{(m)}) \geq Q(x^{(m+1)}) \geq ... \geq Q(x^*).$$

Die A-Orthogonalität der Suchrichtungen

$$p^{(m)T} A p^{(k)} \;=\; 0, \quad k = 0, 1, ..., m-1,$$

angewandt auf die verkürzte Rekursion

$$p^{(m)} \;=\; LK(p^{(m-1)}, r^{(m)}) \;=\; r^{(m)} + \beta p^{(m-1)},\; m \geq 1,\; p^{(0)} = r^{(0)}, \tag{3.71}$$

die natürlich noch gezeigt werden muss, liefert sofort

$$\beta \;=\; \beta_{m-1} \;=\; -\frac{p^{(m-1)T} A r^{(m)}}{p^{(m-1)T} A p^{(m-1)}}. \tag{3.72}$$

Wir fassen alle Schritte und Formeln zusammen, lassen aber mögliche Abbruchkriterien zunächst weg.
Die darin enthaltenen weiter führenden Berechnungsvorschriften für α_m und β_m bedürfen noch einiger Erläuterungen.
Der Iterationsschritt

$$x^{(m+1)} \;=\; x^{(m)} + \alpha_m p^{(m)}$$

selber entspricht dem Richardson-Verfahren mit Parameter und Konditionierungsmatrix, die in den Suchrichtungen enthalten ist (vergl. Formel (3.24)).

Formeln zum CG als (endliches) Iterationsverfahren mit Indizierung

$x^{(0)}$ gegebener Startvektor,
$r^{(0)} = b - Ax^{(0)}$ Anfangsresiduum,
$p^{(0)} = r^{(0)}$ 1. Suchrichtung.

m = 0,1,...≤n–1

$$
\begin{aligned}
x^{(m+1)} &= x^{(m)} + \alpha_m p^{(m)}, \quad \text{wobei} \\
\alpha_m &= \frac{p^{(m)T} r^{(m)}}{p^{(m)T} A p^{(m)}} \\
&= \frac{(r^{(m)}, p^{(m)})}{(w^{(m)}, p^{(m)})}, \quad w^{(m)} = Ap^{(m)} \\
&= \frac{\| r^{(m)} \|_2^2}{\| p^{(m)} \|_A^2} > 0, \\
r^{(m+1)} &= b - Ax^{(m+1)} \quad \text{(direkte Berechnung)}, \\
r^{(m+1)} &= b - v^{(m+1)}, \quad \text{wobei} \\
v^{(m+1)} &= Ax^{(m+1)} \\
&= A(x^{(m)} + \alpha_m p^{(m)}) \\
&= Ax^{(m)} + \alpha_m Ap^{(m)} \\
&= v^{(m)} + \alpha_m w^{(m)} \\
&\qquad \text{(1. rekursive Berechnung)}, \\
r^{(m+1)} &= b - A(x^{(m)} + \alpha_m p^{(m)}) \\
&= b - Ax^{(m)} - \alpha_m Ap^{(m)} \\
&= r^{(m)} - \alpha_m w^{(m)}, \quad \text{(2. rekursive Berechnung)}, \\
p^{(m+1)} &= r^{(m+1)} + \beta_m p^{(m)}, \quad \text{wobei} \\
\beta_m &= -\frac{p^{(m)T} A r^{(m+1)}}{p^{(m)T} A p^{(m)}} \\
&= -\frac{(w^{(m)}, r^{(m+1)})}{(w^{(m)}, p^{(m)})}, \quad w^{(m)} = Ap^{(m)} \\
&= \frac{\| r^{(m+1)} \|_2^2}{\| r^{(m)} \|_2^2} > 0.
\end{aligned}
\tag{3.73}
$$

end m

3.6.2 Wichtige Eigenschaften der Verfahrensgrößen

Die folgenden Betrachtungen und Umformungen dienen der Erkenntnis weiterer Zusammenhänge und Merkmale des CG.

(a) Darstellung des Residuums

$$r^{(m)} = r^{(m-1)} - \alpha_{m-1} A p^{(m-1)}, \quad \alpha_{m-1} = \frac{p^{(m-1)T} r^{(m-1)}}{p^{(m-1)T} A p^{(m-1)}}, \tag{3.74}$$

$$r^{(m+1)} = r^{(0)} - \sum_{k=0}^{m} \alpha_k A p^{(k)}. \tag{3.75}$$

(b) Vereinfachung von Formeln durch Skalarproduktbildung und A-Orthogonalität

$$p^{(m+1)T} r^{(m+1)} = p^{(m+1)T} \Big(r^{(0)} - \sum_{k=0}^{m} \alpha_k A p^{(k)} \Big) = p^{(m+1)T} r^{(0)},$$

$$p^{(m)T} r^{(m)} = p^{(m)T} r^{(0)}, \tag{3.76}$$

$$\alpha_m = \frac{p^{(m)T} r^{(m)}}{p^{(m)T} A p^{(m)}} = \frac{p^{(m)T} r^{(0)}}{p^{(m)T} A p^{(m)}}, \tag{3.77}$$

$$p^{(m)} = r^{(m)} + \sum_{j=0}^{m-1} \beta_j p^{(j)}, \quad r^{(m)} = p^{(m)} - \sum_{j=0}^{m-1} \beta_j p^{(j)},$$

$$\begin{aligned} r^{(m)T} A p^{(k)} &= \Big(p^{(m)} - \sum_{j=0}^{m-1} \beta_j p^{(j)} \Big)^T A p^{(k)}, \quad k \geq m \\ &= p^{(m)T} A p^{(k)} - \sum_{j=0}^{m-1} \beta_j \underbrace{p^{(j)T} A p^{(k)}}_{=0} \\ &= p^{(m)T} A p^{(k)}, \end{aligned}$$

$$r^{(m)T} A p^{(k)} = \begin{cases} 0, & \text{falls } k > m, \\ p^{(m)T} A p^{(m)}, & \text{falls } k = m. \end{cases} \tag{3.78}$$

Satz 3.10 *Nicht verschwindende Residua sind zueinander orthogonal.*

Beweis. Wegen $r^{(0)} = p^{(0)}$ und $r^{(1)} = r^{(0)} - \alpha_0 A p^{(0)} = p^{(0)} - \alpha_0 A p^{(0)}$ folgt

$$\begin{aligned} r^{(0)T} r^{(1)} &= p^{(0)T} (p^{(0)} - \alpha_0 A p^{(0)}) \\ &= p^{(0)T} p^{(0)} - \alpha_0 p^{(0)T} A p^{(0)} \\ &= p^{(0)T} p^{(0)} - \frac{p^{(0)T} r^{(0)}}{p^{(0)T} A p^{(0)}} p^{(0)T} A p^{(0)} \\ &= 0. \end{aligned}$$

Mittels vollständiger Induktion und (3.74) zeigen wir die Orthogonalität des Vektorsystems $\{r^{(0)}, r^{(1)}, ..., r^{(k)}\}$, falls $\{r^{(0)}, r^{(1)}, ..., r^{(k-1)}\}$ orthogonal sind.

$$\begin{aligned} r^{(j)T} r^{(k)} &= r^{(j)T}(r^{(k-1)} - \alpha_{k-1} A p^{(k-1)}), \quad j = 0, 1, ..., k-1 \\ &= r^{(j)T} r^{(k-1)} - \alpha_{k-1} r^{(j)T} A p^{(k-1)}, \end{aligned}$$

$j = 0, 1, ..., k-2$

$$r^{(j)T} r^{(k)} = 0 - \alpha_{k-1}\, r^{(j)T} A p^{(k-1)} = 0, \quad k-1 > j \text{ mit (3.77), (3.78),}$$

$j = k-1$

$$\begin{aligned} r^{(j)T} r^{(k)} &= r^{(k-1)T} r^{(k-1)} - \alpha_{k-1} r^{(k-1)T} A p^{(k-1)} \text{ mit (3.78)} \\ &= r^{(k-1)T} r^{(k-1)} - \frac{p^{(k-1)T} r^{(k-1)}}{p^{(k-1)T} A p^{(k-1)}}\, p^{(k-1)T} A p^{(k-1)} \\ &= r^{(k-1)T} r^{(k-1)} - p^{(k-1)T} r^{(k-1)} \text{ mit (3.65)} \\ &= r^{(k-1)T} r^{(k-1)} - \left(r^{(k-1)} + \sum_{j=0}^{k-2} \beta_j p^{(j)}\right)^T r^{(k-1)}, \ \beta_j = \beta_j^{(k-1)} \\ &= -\sum_{j=0}^{k-2} \beta_j^{(k-1)}\, p^{(j)T} r^{(k-1)} \text{ mit (3.65)} \\ &= -\sum_{j=0}^{k-2} \beta_j^{(k-1)} \left(r^{(j)} + \sum_{l=0}^{j-1} \beta_l^{(j)} p^{(l)}\right)^T r^{(k-1)} \\ &= -\sum_{j=0}^{k-2} \beta_j^{(k-1)} \underbrace{r^{(j)T} r^{(k-1)}}_{=0} - \sum_{j=0}^{k-2} \sum_{l=0}^{j-1} \beta_j^{(k-1)} \beta_l^{(j)}\, p^{(l)T} r^{(k-1)} \\ &= -\sum_{j=0}^{k-2} \sum_{l=0}^{j-1} \beta_j^{(k-1)} \beta_l^{(j)}\, p^{(l)T} r^{(k-1)} \text{ mit (3.65)} \\ &= -\sum_{j=0}^{k-2} \sum_{l=0}^{j-1} \sum_{s=0}^{l-1} \beta_j^{(k-1)} \beta_l^{(j)} \beta_s^{(l)}\, p^{(s)T} r^{(k-1)} \text{ mit (3.65)} \\ &= ... \\ &= -\sum_{l_0=0}^{k-2} \sum_{l_1=0}^{l_0-1} \sum_{l_2=0}^{l_1-1} ... \sum_{l_t=0}^{0} \beta_{l_0}^{(k-1)} \beta_{l_1}^{(l_0)} \beta_{l_2}^{(l_1)} \cdot ... \cdot \beta_{l_t}^{(l_{t-1})}\, p^{(l_t)\,T} r^{(k-1)}, \\ &\quad p^{(l_t)} = p^{(0)} = r^{(0)}, \ p^{(0)T} r^{(k-1)} = r^{(0)T} r^{(k-1)} = 0, \\ &= 0, \end{aligned}$$

somit

$$r^{(j)T} r^{(k)} = 0 \text{ für } j \neq k. \tag{3.79}$$

Nicht verschwindende orthogonale Vektoren sind linear unabhängig. □

(c) Nachweis der verkürzten Rekursion für die Suchrichtungen

$$
\begin{aligned}
p^{(m)} &= r^{(m)} + \sum_{j=0}^{m-1} \beta_j p^{(j)}, \\
p^{(m)T} A p^{(k)} &= r^{(m)T} A p^{(k)} + \sum_{j=0}^{m-1} \beta_j \, p^{(j)T} A p^{(k)}, \quad k = 0, 1, ..., m-1, \\
0 &= r^{(m)T} A p^{(k)} + \beta_k \, p^{(k)T} A p^{(k)}, \\
\beta_k &= -\frac{r^{(m)T} A p^{(k)}}{p^{(k)T} A p^{(k)}}, \quad k = 0, 1, ..., m-1, \qquad (3.80) \\
-r^{(m)T} A p^{(k)} &= r^{(m)T} \frac{1}{\alpha_k} (r^{(k+1)} - r^{(k)}) \ \text{ mit (3.74)} \\
&= \frac{1}{\alpha_k} (r^{(m)T} r^{(k+1)} - r^{(m)T} r^{(k)}), \\
-r^{(m)T} A p^{(k)} &= \begin{cases} 0, & \text{falls } k = 0, 1, ..., m-2, \\ \dfrac{1}{\alpha_{m-1}} \, r^{(m)T} r^{(m)}, & \text{falls } k = m-1. \end{cases} \qquad (3.81)
\end{aligned}
$$

Damit ist

$$
\beta_k = \begin{cases} 0, & \text{falls } k = 0, 1, ..., m-2, \\ \dfrac{\frac{1}{\alpha_{m-1}} r^{(m)T} r^{(m)}}{p^{(m-1)T} A p^{(m-1)}}, & \text{falls } k = m-1, \end{cases}
$$

und

$$
\begin{aligned}
p^{(m)} &= r^{(m)} + \beta_{m-1} p^{(m-1)}, \quad \text{wobei} \qquad (3.82) \\
\beta_{m-1} &= -\frac{r^{(m)T} A p^{(m-1)}}{p^{(m-1)T} A p^{(m-1)}} \\
&= \frac{r^{(m)T} r^{(m)}}{\alpha_{m-1} \, p^{(m-1)T} A p^{(m-1)}} \\
&= \frac{r^{(m)T} r^{(m)}}{\dfrac{r^{(m-1)T} r^{(m-1)}}{p^{(m-1)T} A p^{(m-1)}} \, p^{(m-1)T} A p^{(m-1)}} \ \text{ mit (3.74)}, \\
\beta_{m-1} &= \frac{r^{(m)T} r^{(m)}}{r^{(m-1)T} r^{(m-1)}} = \frac{\|r^{(m)}\|_2^2}{\|r^{(m-1)}\|_2^2} > 0 \ \text{ bei } r^{(m)} \neq 0. \qquad (3.83)
\end{aligned}
$$

(d) Orthogonalität von $p^{(k)}$ und $r^{(m)}$ mit (3.82)

$$\begin{aligned}
p^{(k)T}r^{(m)} &= (r^{(k)} + \beta_{k-1}p^{(k-1)})^T r^{(m)},\ k = 0, 1, ..., m-1, m, \\
&= r^{(k)^T} r^{(m)} + \beta_{k-1} p^{(k-1)T} r^{(m)} \\
&= r^{(k)^T} r^{(m)} + \beta_{k-1}(r^{(k-1)} + \beta_{k-2}p^{(k-2)})^T r^{(m)} \\
&= r^{(k)^T} r^{(m)} + \beta_{k-1} r^{(k-1)T} r^{(m)} + \beta_{k-1}\beta_{k-2}\, p^{(k-2)T} r^{(m)} \\
&= ... \\
&= r^{(k)^T} r^{(m)} + \beta_{k-1}\, r^{(k-1)T} r^{(m)} \\
&\qquad + \beta_{k-1}\beta_{k-2}\, r^{(k-2)T} r^{(m)} \\
&\qquad + ... \\
&\qquad + \beta_{k-1}\beta_{k-2} \cdot ... \cdot \beta_0\, p^{(0)T} r^{(m)},\quad p^{(0)} = r^{(0)},
\end{aligned}$$

$$p^{(k)T}r^{(m)} = \begin{cases} 0, & \text{falls } k = 0, 1, ..., m-1, \\ r^{(m)T}r^{(m)}, & \text{falls } k = m, \end{cases} \tag{3.84}$$

$$p^{(m)T}r^{(m)} = r^{(m)T}r^{(m)} = p^{(m)T}r^{(0)} \text{ mit (3.76).} \tag{3.85}$$

Der Nachweis der Orthogonalität kann auch unter Verwendung von (3.74) durch vollständige Induktion wie folgt gemacht werden.

$$\begin{aligned}
r^{(1)T}p^{(0)} &= (r^{(0)} - \alpha_0 A p^{(0)})^T p^{(0)} \\
&= r^{(0)T}p^{(0)} - \alpha_0 p^{(0)T} A p^{(0)} \\
&= r^{(0)T}p^{(0)} - \frac{p^{(0)T}r^{(0)}}{p^{(0)T}Ap^{(0)}}\, p^{(0)T} A p^{(0)} \\
&= 0.
\end{aligned}$$

Sei $r^{(m)T}p^{(k)} = 0$ für $m \geq 1,\ k = 0, 1, ..., m-1$. Dann ist

$$\begin{aligned}
r^{(m+1)T}p^{(k)} &= (r^{(m)} - \alpha_m A p^{(m)})^T p^{(k)},\quad k \leq m-1 \\
&= \underbrace{r^{(m)T}p^{(k)}}_{=0 \text{ nach Vor.}} - \alpha_m \underbrace{p^{(m)T}Ap^{(k)}}_{=0 \text{ wegen A-Orth.}} = 0,
\end{aligned}$$

$$\begin{aligned}
r^{(m+1)T}p^{(m)} &= (r^{(m)} - \alpha_m A p^{(m)})^T p^{(m)},\quad k = m \\
&= r^{(m)T}p^{(m)} - \frac{p^{(m)T}r^{(m)}}{p^{(m)T}Ap^{(m)}}\, p^{(m)T} A p^{(m)} = 0.
\end{aligned}$$

Mit (3.85) ist die Schrittzahl

$$\alpha_m = \frac{p^{(m)T}r^{(m)}}{p^{(m)T}Ap^{(m)}} = \frac{r^{(m)T}r^{(m)}}{p^{(m)T}Ap^{(m)}} = \frac{(r^{(m)}, r^{(m)})}{(p^{(m)}, p^{(m)})_A} = \frac{\|r^{(m)}\|_2^2}{\|p^{(m)}\|_A^2} > 0. \tag{3.86}$$

(e) A-Orthogonalität von $p^{(k)}$ und $r^{(m)}$ mit (3.78), (3.81)

$$r^{(m)T}Ap^{(k)} = \begin{cases} 0, & \text{falls } k > m, \\ p^{(m)T}Ap^{(m)}, & \text{falls } k = m, \\ -\dfrac{1}{\alpha_{m-1}}\, r^{(m)T}r^{(m)} = -\beta_{m-1}\, p^{(m-1)T}Ap^{(m-1)}, & \text{falls } k = m-1, \\ 0, & \text{falls } k < m-1. \end{cases}$$

(f) Wichtige Vektoren sind

$$r^{(0)},\ Ar^{(0)},\ A^2r^{(0)}, ...,$$

weil, solange die Vektoren $p^{(k)} \neq 0$ sind, diese lineare Unterräume aufspannen und mit (3.74) und (3.82) die folgenden Beziehungen gelten.

$$\begin{aligned}
r^{(0)}, \quad & p^{(0)} = r^{(0)}, \ \ \text{span}\{p^{(0)}\} = \text{span}\{r^{(0)}\}, \\
Ar^{(0)} &= Ap^{(0)} = \frac{1}{\alpha_0}(r^{(0)} - r^{(1)}) = \frac{1}{\alpha_0}(p^{(0)} - p^{(1)} + \beta_0 p^{(0)}) \\
&= \frac{1}{\alpha_0}(1+\beta_0)p^{(0)} - \frac{1}{\alpha_0}p^{(1)}, \\
r^{(1)} &= r^{(0)} - \alpha_0 Ap^{(0)} \\
&= (I - \alpha_0 A)r^{(0)}, \\
p^{(1)} &= r^{(1)} + \beta_0 p^{(0)} \\
&= (I - \alpha_0 A)r^{(0)} + \beta_0 r^{(0)} \\
&= (1+\beta_0)Ir^{(0)} - \alpha_0 Ar^{(0)}, \\
& \quad \text{span}\{p^{(0)}, p^{(1)}\} = \text{span}\{r^{(0)}, r^{(1)}\} = \text{span}\{r^{(0)}, Ar^{(0)}\}, \\
Ar^{(0)} &= \gamma_0 p^{(0)} + \gamma_1 p^{(1)}, \\
A^2r^{(0)} &= \gamma_0 Ap^{(0)} + \gamma_1 Ap^{(1)} \\
&= \gamma_0 \frac{1}{\alpha_0}(r^{(0)} - r^{(1)}) + \gamma_1 \frac{1}{\alpha_1}(r^{(1)} - r^{(2)}) \\
&= \gamma_0 \frac{1}{\alpha_0}(p^{(0)} - p^{(1)} + \beta_0 p^{(0)}) + \gamma_1 \frac{1}{\alpha_1}(p^{(1)} - \beta_0 p^{(0)} - p^{(2)} + \beta_1 p^{(1)}) \\
&= \delta_0 p^{(0)} + \delta_1 p^{(1)} + \delta_2 p^{(2)}, \\
r^{(2)} &= r^{(1)} - \alpha_1 Ap^{(1)} \\
&= (I - \alpha_0 A)r^{(0)} - \alpha_1 A[(1+\beta_0)Ir^{(0)} - \alpha_0 Ar^{(0)}] \\
r^{(2)} &= [I - (\alpha_0 + \alpha_1 + \alpha_1\beta_0)A + \alpha_0\alpha_1 A^2]\, r^{(0)}, \qquad (3.87)
\end{aligned}$$

$$
\begin{aligned}
p^{(2)} &= r^{(2)} + \beta_1 p^{(1)} \\
&= [(1+\beta_1+\beta_0\beta_1)I - (\alpha_0+\alpha_1+\alpha_1\beta_0+\alpha_0\beta_1)A + \alpha_0\alpha_1 A^2]\, r^{(0)}, \\
&\quad \operatorname{span}\{p^{(0)}, p^{(1)}, p^{(2)}\} = \operatorname{span}\{r^{(0)}, r^{(1)}, r^{(2)}\} \;=\; \operatorname{span}\{r^{(0)}, Ar^{(0)}, A^2 r^{(0)}\}, \\
&\quad \text{usw.}
\end{aligned}
$$

Daraus erkennen wir die Entstehung der linearen Unterräume sowie die polynomiale Darstellung der Residua und Suchrichtungen.

(g) Lineare Unterräume von Vektoren im CG

Definition 3.4
Als lineare Unterräume der Vektoren des CG definiert man für $m = 1, 2, ...$

$$
\begin{aligned}
\mathcal{R}_m &= \operatorname{span}\{r^{(0)}, r^{(1)}, ..., r^{(m-1)}\}, \quad r^{(k)} \neq 0, \\
\mathcal{P}_m &= \operatorname{span}\{p^{(0)}, p^{(1)}, ..., p^{(m-1)}\}, \quad p^{(k)} \neq 0, \\
\mathcal{K}_m(A, r^{(0)}) &= \operatorname{span}\{r^{(0)}, Ar^{(0)}, ..., A^{m-1} r^{(0)}\} \quad \textit{(Krylov-Unterraum).}
\end{aligned} \tag{3.88}
$$

Offensichtlich gilt für die Krylov-Unterräume die Kette der Enthaltungen

$$
\mathcal{K}_1(A, r^{(0)}) \;\subseteq\; \mathcal{K}_2(A, r^{(0)}) \;\subseteq\; ... \;\subseteq\; \mathcal{K}_m(A, r^{(0)}) \;\subseteq\; \mathcal{K}_{m+1}(A, r^{(0)}) \;\subseteq\; ... \tag{3.89}
$$

Aber es lässt sich noch eine Reihe weiterer Beziehungen zu den Vektoren des CG zeigen.

Lemma 3.11 *Für die Iterierten* $x^{(m)}$, $m = 0, 1, ...$, *die Residua* $r^{(m)}$ *und die Suchrichtungen* $p^{(m)}$ *gilt*

(a)

$$
\begin{aligned}
r^{(m)} &\in \mathcal{K}_{m+1}(A, r^{(0)}), \\
p^{(m)} &\in \mathcal{K}_{m+1}(A, r^{(0)}), \\
x^{(m+1)} &\in x^{(0)} + \mathcal{K}_{m+1}(A, r^{(0)}).
\end{aligned} \tag{3.90}
$$

(b) Es ist

$$
\mathcal{K}_{m+1}(A, r^{(0)}) \;=\; \mathcal{R}_{m+1} \;=\; \mathcal{P}_{m+1}. \tag{3.91}
$$

(c) Weiterhin haben wir

$$
r^{(m+1)} \;\perp\; \mathcal{K}_{m+1}(A, r^{(0)}). \tag{3.92}
$$

Beweis. Vollständige Induktion bezüglich m.
Für alle Behauptungen ist in den bisher gemachten Betrachtungen im Teil (f) und in den Formeln (3.79), (3.84) der Induktionsanfang für $m = 0$ zu finden.
Die Behauptungen seien also richtig für ein $m \in \mathbb{N}$.

Zu (a): Aus den Iterationsvorschriften folgt mit $\mathcal{K}_{m+1} = \mathcal{K}_{m+1}(A, r^{(0)})$

$$\begin{aligned} r^{(m+1)} &= \underbrace{r^{(m)}}_{\in \mathcal{K}_{m+1}} - \alpha_m \underbrace{Ap^{(m)}}_{\in \mathcal{K}_{m+2}} \in \mathcal{K}_{m+2}, \\ p^{(m+1)} &= \underbrace{r^{(m+1)}}_{\in \mathcal{K}_{m+2}} + \beta_m \underbrace{p^{(m)}}_{\in \mathcal{K}_{m+1}} \in \mathcal{K}_{m+2}, \\ x^{(m+2)} &= \underbrace{x^{(m+1)}}_{\in x^{(0)} + \mathcal{K}_{m+1}} + \alpha_{m+1} \underbrace{p^{(m+1)}}_{\in \mathcal{K}_{m+2}} \in x^{(0)} + \mathcal{K}_{m+2}. \end{aligned}$$

Zu (b): Nach (a) ist

$$\mathcal{P}_{m+2} = \operatorname{span}\{p^{(0)}, ..., p^{(m)}, p^{(m+1)}\} \subseteq \mathcal{K}_{m+2}.$$

Da die Vektoren $p^{(k)}$ linear unabhängig sind, ist die Dimension von $\mathcal{P}_{m+2}$ gleich $m+2$. Aber $\dim(\mathcal{K}_{m+2}) \leq m+2$, also müssen beide Unterräume gleich sein.
Da $r^{(m+2)} \in \mathcal{R}_{m+2}$ wegen $r^{(m+2)} = p^{(m+2)} - \beta_{m+1} p^{(m+1)}$ auch zu $\mathcal{P}_{m+2}$ gehört und umgekehrt $p^{(m+2)} \in \mathcal{P}_{m+2}$ wegen $p^{(m+2)} = r^{(m+2)} + \beta_{m+1} p^{(m+1)}$ und $\mathcal{P}_{m+1} = \mathcal{R}_{m+1}$ auch zu $\mathcal{R}_{m+2}$ gehört, sind die Unterräume $\mathcal{P}_{m+2}$ und $\mathcal{R}_{m+2}$ identisch.
Zu (c): Wegen der Gleichheit der Unterräume gemäß (b) zeigt man

$$r^{(m+1)} \perp \mathcal{R}_{m+1} \quad \text{oder} \quad r^{(m+1)} \perp \mathcal{P}_{m+1}.$$

Ersteres ist in (3.79) gemacht worden.
Der Nachweis von $r^{(m+1)T} p^{(k)} = p^{(k)T} r^{(m+1)} = 0,\ k = 0, 1, ..., m$, wurde schon in (3.84) geführt, damit gilt insbesondere auch

$$r^{(m+1)T} p^{(m)} = p^{(m)T} r^{(m+1)} = 0.$$

□

Aus der Orthogonalität (3.92) folgt

$$\begin{aligned} b - Ax^{(m)} = Ax^* - Ax^{(m)} &= A(x^* - x^{(m)}) \perp \mathcal{K}_m(A, r^{(0)}), \\ (A(x^* - x^{(m)}), A^k r^{(0)}) &= 0, \quad k = 0, 1, ..., m-1, \\ (x^* - x^{(m)}, A^k r^{(0)})_A &= 0. \end{aligned} \tag{3.93}$$

Damit kann das CG wie folgt geometrisch interpretiert werden.
Die Krylov-Unterräume bilden, wenn kein vorzeitiges Ende eintritt, eine aufsteigende Folge von Unterräumen des $\mathbb{R}^n$.

$$\mathcal{K}_1(A, r^{(0)}) \subset \mathcal{K}_2(A, r^{(0)}) \subset ... \subset \mathcal{K}_n(A, r^{(0)}).$$

Die m-te Iterierte $x^{(m)}$ liegt im affinen Unterraum $x^{(0)} + \mathcal{K}_m$ und ist so gewählt, dass das Residuum $r^{(m)}$ senkrecht auf $\mathcal{K}_m$ steht.

Man sieht mit (3.93) leicht, dass durch diese Bedingung der Abstand der Iterierten $x^{(m)}$ zur gesuchten Lösung x^* bezüglich der Norm $\|\cdot\|_A$ minimiert wird.
Falls $\mathcal{K}_n(A, r^{(0)}) = \mathbb{R}^n$ ist, gilt $x^{(n)} \in x^{(0)} + \mathcal{K}_n(A, r^{(0)})$ und $r^{(n)} \perp \mathbb{R}^n$, was noch einmal die Endlichkeit des CG unterstreicht.

(h) Es gilt $r^{(m)} = 0$ gdw. $p^{(m)} = 0$.
Diese Aussage folgt aus dem gemeinsamen gleichmäßigen Anwachsen der Dimension der genannten Unterräume.
Der Nachweis kann aber auch einfach erfolgen. Sei $\dim(\mathcal{R}_{m-1}) = m - 1$.
$\Rightarrow$

$$r^{(m)} = 0 \;\rightarrow\; \beta_{m-1} = \frac{\|r^{(m)}\|_2^2}{\|r^{(m-1)}\|_2^2} = 0 \;\rightarrow\; p^{(m)} = r^{(m)} + \beta_{m-1} p^{(m-1)} = 0.$$

$\Leftarrow$

$$p^{(m)} = 0 \;\rightarrow\; 0 = p^{(m)T} r^{(m)} = r^{(m)T} r^{(m)} \text{ mit (3.85)} \;\rightarrow\; r^{(m)} = 0.$$

Aus den gemachten Erläuterungen zum CG ergibt sich eine Verallgemeinerung und folgende Definition, welche auf beliebige reguläre, nicht notwendig spd Matrizen anwendbar ist.
Die Iterierten $x^{(m)}$ bestimmt man mittels Suchrichtungen $p^{(m)}$ aus einem ersten linearen Unterraum ("Suchraum", search subspace) und fordert dabei, dass die Residua $r^{(m)}$ orthogonal zu einem zweiten Unterraum ("Testunterraum", subspace of constraints) sind.

Definition 3.5 Projektionsmethode
Es seien K_m und L_m m-dimensionale Unterräume von $\mathbb{R}^n$.
Ein Verfahren zur Berechnung einer Näherungslösung $x^{(m)} \in x^{(0)} + K_m$ unter der **Galerkin-Bedingung**

$$r^{(m)} = b - Ax^{(m)} \perp L_m, \quad \textit{das ist} \;\; r^{(m)T} w = (r^{(m)}, w) = 0 \;\forall\; w \in L_m,$$

heißt Projektionsmethode. (Die Orthogonalitätsbedingung bezieht sich auf das euklidische Skalarprodukt.)
Im Fall $K_m \neq L_m$ ist es eine schiefe Projektionsmethode.
Für $K_m = L_m$ heißt die orthogonale Projektionsmethode auch **Galerkin-Verfahren**.
Für $K_m = \mathcal{K}_m$ heißt das Projektionsverfahren auch **Krylov-Unterraum-Methode**.

Das CG ist eine (orthogonale) Krylov-Unterraum-Methode mit $L_m = K_m = \mathcal{K}_m = \mathcal{K}_m(A, r^{(0)})$ und

$$\begin{aligned} x^{(m)} &= x^{(0)} + \sum_{k=0}^{m-1} \alpha_k p^{(k)} = x^{(0)} + \delta^{(m)} \;\in\; x^{(0)} + \mathcal{P}_m = x^{(0)} + \mathcal{K}_m(A, r^{(0)}), \\ r^{(m)} &= b - Ax^{(m)} = r^{(0)} - \sum_{k=0}^{m-1} \alpha_k A p^{(k)} = r^{(0)} - A\delta^{(m)} \;\perp\; \mathcal{K}_m(A, r^{(0)}). \end{aligned} \tag{3.94}$$

Zudem gilt

$$x^{(m)} = \underset{x \in x^{(0)}+K_m}{\arg\min} Q(x), \quad K_m = \mathcal{K}_m(A, r^{(0)}). \tag{3.95}$$

Damit haben wir den Iterierten eine neue geometrische Interpretation gegeben, welche sich verallgemeinern lässt.

Untersuchen wir die anderen AV im Sinne von Projektionsmethoden.

- GV mit Richtungen des steilsten Abstiegs $r^{(m)}$

  $$\begin{aligned} x^{(m)} &= x^{(m-1)} + \alpha_{m-1} r^{(m-1)} \in x^{(0)} + K, \quad K = \operatorname{span}\{r^{(m-1)}\}, \\ x^{(m)} &\in x^{(0)} + K_m = x^{(0)} + \mathcal{R}_m = x^{(0)} + \operatorname{span}\{r^{(0)}, ..., r^{(m-1)}\}, \\ r^{(m)} &\perp L_m = \operatorname{span}\{r^{(m-1)}\} = L, \\ r^{(m)} &\perp r^{(m-1)}, \quad r^{(m)} \not\perp K_m. \end{aligned}$$

  Das GV stellt jedoch in jedem Schritt eine orthogonale Projektionsmethode mit $K = L = \operatorname{span}\{r^{(m-1)}\}$ dar.

- AV mit linear unabhängigen Richtungen $p^{(m)}$

  $$\begin{aligned} x^{(m)} &= x^{(m-1)} + \alpha_{m-1} p^{(m-1)}, \\ x^{(m)} &\in x^{(0)} + K_m = x^{(0)} + \mathcal{P}_m = x^{(0)} + \operatorname{span}\{p^{(0)}, ..., p^{(m-1)}\}, \\ r^{(m)} &\perp L_m = \operatorname{span}\{r^{(m-1)}\}, \\ r^{(m)} &\perp p^{(m-1)}, \quad r^{(m)} \not\perp K_m. \end{aligned}$$

- AV mit konjugierten Richtungen $p^{(m)}$

  $$\begin{aligned} x^{(m)} &= x^{(m-1)} + \alpha_{m-1} p^{(m-1)}, \\ x^{(m)} &\in x^{(0)} + K_m = x^{(0)} + \mathcal{P}_m = x^{(0)} + \operatorname{span}\{p^{(0)}, ..., p^{(m-1)}\}, \\ r^{(m)} &\perp L_m = K_m = \mathcal{P}_m = \operatorname{span}\{p^{(0)}, ..., p^{(m-1)}\}, \\ &\quad (r^{(m)}, p^{(k)}) = (r^{(m-1)}, p^{(k)}) = ... = (r^{(k+1)}, p^{(k)}) = 0, \ m > k, \\ r^{(n)} &\perp \mathcal{P}_n = \mathbb{R}^n \rightarrow r^{(n)} = 0. \end{aligned}$$

Betrachtet man die AV in Bezug auf die Wahl der Unterräume, so erkennt man dort Vorteile, wo die Unterräume nicht unabhängig voneinander definiert bzw. konstruiert werden.
Bei kleindimensionalen Beispielen kann man dabei schnell an Grenzen stoßen. Besonders unterliegt dann die Möglichkeit der Wahl des nächsten Unterraums L_m gewissen Einschränkungen und Dimensionsbegrenzungen, gerade wenn man an die Endlichkeit des AV denkt.

Wir illustrieren die Situation der Projektionsmethoden an einfachen Beispielen aus dem $\mathbb{R}^2$ mit A-orthogonalen Richtungsvektoren.

Orthogonale Projektionsmethode

Dabei spannt der Richtungsvektor $p^{(0)}$ den Raum K_1 auf und es ist i. Allg. $p^{(0)} \neq r^{(0)}$.
Ausgehend von $x^{(0)}$ suchen wir im Unterraum K_1 die nächste Iterierte $x^{(1)}$, so dass $r^{(1)} = b - Ax^{(1)} = A(x^* - x^{(1)}) \perp K_1 = \text{span}\{p^{(0)}\}$ ist.
Damit gilt $x^* - x^{(1)} \perp p^{(0)}$ nur für $A = I$.
Die Iterierte $x^{(1)}$ ist aber nichts anderes als die Lösung aus der Strahlenminimierung für $Q(x)$, denn aus

$$x^{(1)} \;=\; x^{(0)} + \alpha_0 p^{(0)}, \quad \alpha_0 = \frac{p^{(0)T} r^{(0)}}{p^{(0)T} A p^{(0)}}, \quad r^{(1)} \;=\; r^{(0)} - \alpha_0 A p^{(0)}$$

folgt

$$r^{(1)T} p^{(0)} \;=\; (r^{(0)} - \alpha_0 A p^{(0)})^T p^{(0)} \;=\; r^{(0)T} p^{(0)} - \frac{p^{(0)T} r^{(0)}}{p^{(0)T} A p^{(0)}}\, p^{(0)T} A p^{(0)} \;=\; 0.$$

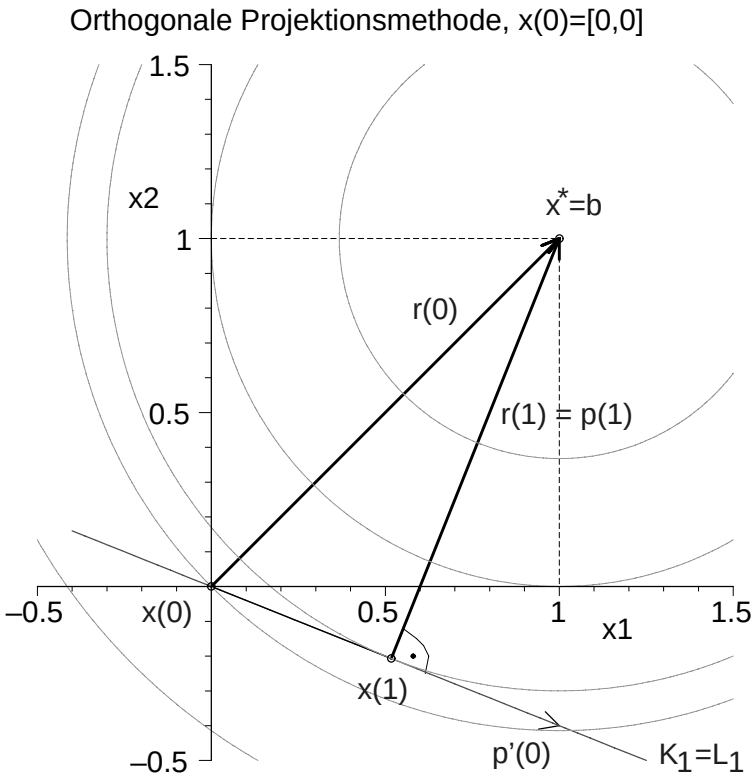


Abb. 3.14 Datei *abst_101.ps*
Orthogonale Projektionsmethode und erste Iterierte $x^{(m)}$ bei $Ax = Ix = b = (1,1)^T$ mit `contours=[0.5,0,-0.1552,-0.5,-0.8,-1]`, $x^* = b,\ Q(x^*) = -1,$
$x^{(0)} = 0,\ r^{(0)} = b,\ p^{'(0)} \| p^{(0)} = (5,-2)^T \neq r^{(0)},\ x^{(1)} = \frac{3}{29}(5,-2)^T,\ Q(x^{(1)}) = -\frac{9}{58},$
$r^{(1)} = x^* - x^{(1)} = \frac{7}{29}(2,5)^T \perp p^{(0)},\ \beta_0 = 0,\ p^{(1)} = r^{(1)},\ (Ap^{(1)}, p^{(0)}) = 0,\ x^{(2)} = x^*$

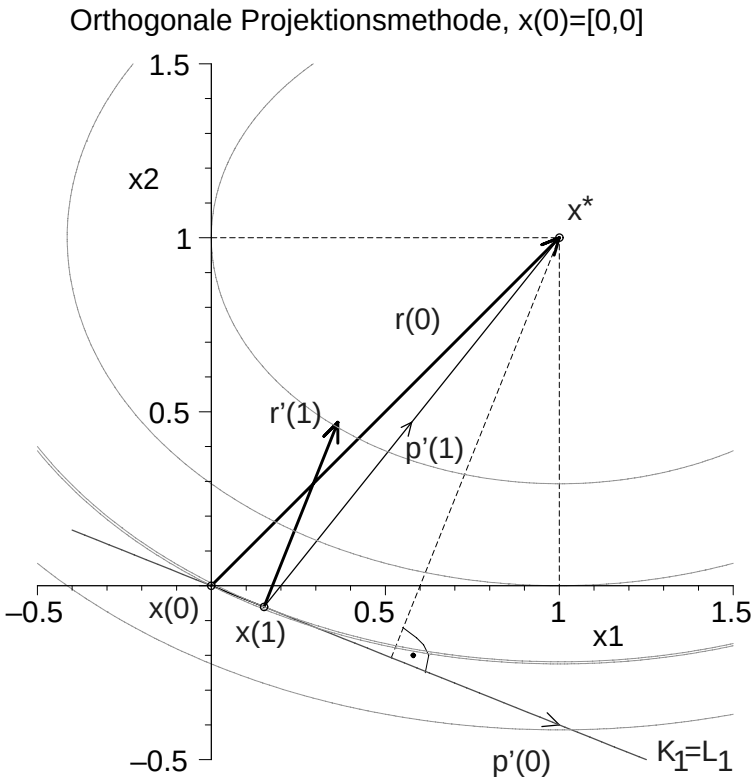


Abb. 3.15 Datei *abst_102.ps*
Orth. Projektionsmethode und erste Iterierte $x^{(m)}$ bei $Ax = \mathrm{diag}(1,2)x = b = (1,2)^T$ mit `contours=[0.5,0,-0.01515,-0.5,-1,-1.5]`, $x^* = (1,1)^T,\ Q(x^*) = -\frac{3}{2}$,
$x^{(0)} = 0,\ r^{(0)} = b,\ p'^{(0)} \| p^{(0)} = (5,-2)^T \neq r^{(0)},\ x^{(1)} = \frac{1}{33}(5,-2)^T,\ Q(x^{(1)}) = -\frac{1}{66}$,
$r^{(1)} = A(x^* - x^{(1)}) = \frac{14}{33}(2,5)^T \perp p^{(0)},\ r'^{(1)} \| r^{(1)}$,
$\beta_0 = \frac{140}{1089},\ p'^{(1)} \| p^{(1)} = \frac{406}{1089}(4,5)^T,\ (Ap^{(1)}, p^{(0)}) = 0,\ x^{(2)} = x^*$

In beiden Beispielen aus dem $\mathbb{R}^2$ haben wir $K_2 = \mathrm{span}\{p^{(0)}, p^{(1)}\}$ und $\dim(K_2) = 2$. Da $r^{(2)} \perp K_2$ ist, muss $r^{(2)} = 0$ sein, d. h. das AV ist nach 2 Schritten am Minimum.

Schiefe Projektionsmethode

Dabei spannt der Richtungsvektor $p^{(0)}$ den Raum K_1 auf, aber

$$r^{(1)} = b - Ax^{(1)} = A(x^* - x^{(1)}) \perp L_1 \neq K_1.$$

Hier und im weiteren Verlauf unterscheidet sich die Situation von der orthogonalen Projektionsmethode. Wenn nun das Residuum $r^{(2)}$ orthogonal zum zweidimensionalen Unterraum $L_2 = \mathrm{span}\{l^{(0)}, l^{(1)}\}$ sein soll, dann kann nur $r^{(2)} = 0$ gelten, d. h. $x^{(2)} = x^*$. Aber die Wahl von L_2 ist im Allgemeinen unabhängig von der Auswahl der Suchrichtungen $p^{(k)}$, die den Verlauf der Iterierten gemäß

$$x^{(m)} \ \in \ x^{(0)} + K_m \ = \ x^{(0)} + \mathrm{span}\{p^{(0)}, p^{(1)}, ..., p^{(m-1)}\}$$

wesentlich bestimmen.

So muss bei angenommener A-Orthogonalität der Suchrichtungen die Iterierte $x^{(2)}$ nicht unbedingt schon zu x^* führen. Im Gegenteil, es kann passieren, dass die Forderung

$$r^{(2)} \perp L_2 = \text{span}\{l^{(0)}, l^{(1)}\}$$

nicht erfüllbar ist.
Denkbar wäre dann, mit solchen "reduzierten" Unterräumen wie $L_1 = \text{span}\{l^{(0)}\}$, $L_2 = \text{span}\{l^{(1)}\}$ usw. zu arbeiten.
Weiterhin wird mit der Forderung der Orthogonalität der Residua zu den Unterräumen L_m die lokale Optimalität verhindert und die Strahlenminimierung wird nicht zum lokalen Minimum in der Suchrichtung führen.
In den folgenden zwei Beispielen finden wir diese Situation vor.

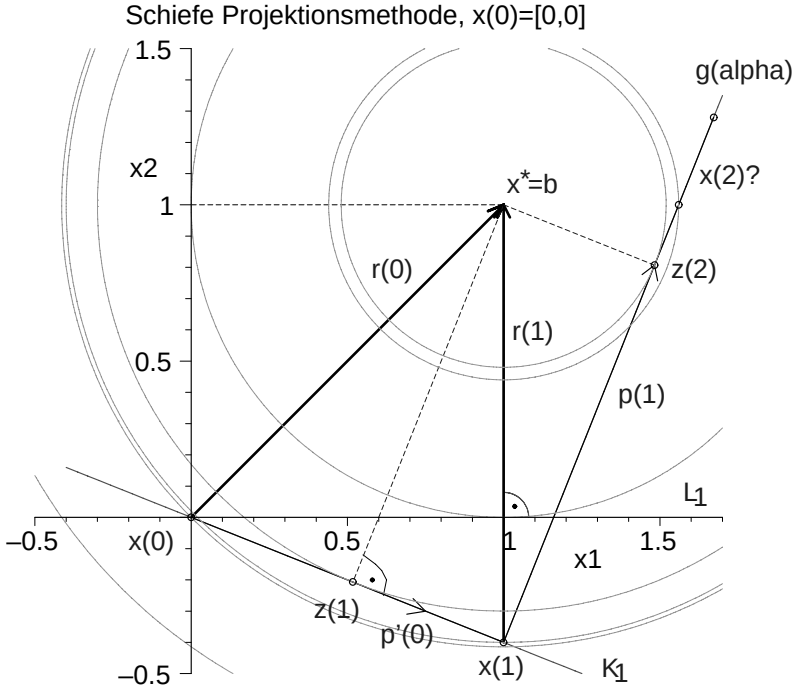


Abb. 3.16 Datei *abst_103.ps*
Schiefe Projektionsmethode und erste Iterierte $x^{(m)}$ bei $Ax = Ix = b = (1,1)^T$
mit `contours=[0.5,0,-0.02,-0.1552,-0.5,-0.8432,-0.8648,-1]`,
$x^* = b,\ Q(x^*) = -1,\ x^{(0)} = 0,\ r^{(0)} = b,\ p'^{(0)} \| p^{(0)} = (5,-2)^T \neq r^{(0)},\ K_1 = \text{span}\{p^{(0)}\}$,
$L_1 = \text{span}\{l^{(0)}\},\ l^{(0)} = (1,0)^T,\ x^{(1)} = (1,-\frac{2}{5})^T,\ r^{(1)} = x^* - x^{(1)} = (0,\frac{7}{5})^T \perp l^{(0)}$,
$Q(x^{(1)}) = -\frac{1}{50} = -0.02 > Q(z^{(1)}) = -\frac{9}{58} = -0.155...,\ z^{(1)} = \frac{3}{29}(5,-2)$ von orth. PM,
$\beta_0 = \frac{14}{145},\ p^{(1)} = r^{(1)} + \beta_0 p^{(0)} = \frac{7}{29}(2,5)^T,\ (Ap^{(1)}, p^{(0)}) = 0,\ K_2 = \text{span}\{p^{(0)}, p^{(1)}\}$,
$\min\limits_{\alpha} Q(x^{(1)} + \alpha p^{(1)})$ führt auf $z^{(2)} = \frac{1}{145}(215,117)^T$ mit $Q(z^{(2)}) = -\frac{627}{725} = -0.864...$,
$x^{(2)}$ ist ebenfalls auf der Geraden $g(\alpha) = x^{(1)} + \alpha p^{(1)}$ zu suchen.

Die Wahl des nächsten Unterraums L_2, so dass $r^{(2)} \perp L_2$ gilt, ist problematisch. Wenn wir $L_2 = L_1$ nehmen, wird $x^{(2)} = x^{(1)}$ werden.
Da das Residuum $r^{(2)} = b - Ax^{(2)} = x^* - x^{(2)}$ mit einer Iterierten $x^{(2)}$ auf der Geraden $g(\alpha) = x^{(1)} + \alpha p^{(1)}$ nicht Null werden kann, macht die Wahl eines Unterraums $L_2 = \text{span}\{l^{(0)}, l^{(1)}\}$ mit $l^{(1)}$ linear unabhängig von $l^{(0)}$ keinen Sinn. Wir führen die Iteration jedoch fort mit $L_2 = \text{span}\{l^{(1)}\},\ l^{(1)} = (0,1)^T$. Anschließend fordern wir einfach abwechselnd $r^{(2k-1)} \perp L_1$ und $r^{(2k)} \perp L_2$. Auch die Dimension des Unterraums K_m kann nicht größer als 2 werden, so dass es zu einer "Wiederholung" der konjugierten Suchrichtungen mit $p^{(2)} \| p^{(0)},\ p^{(3)} \| p^{(1)}$ usw. kommt und $K_m = K_2,\ m > 3$, ist.
Damit wird der Iterationsprozess nicht endlich sein.

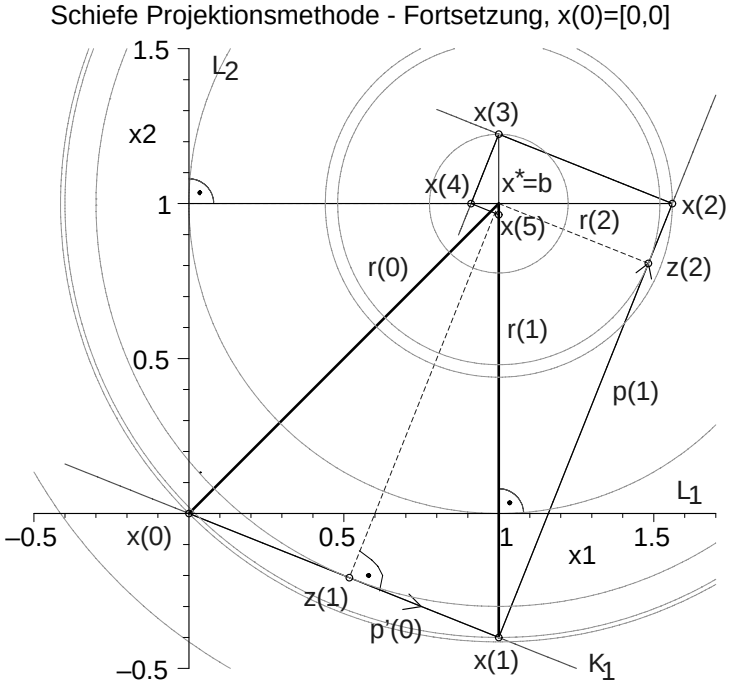


Abb. 3.17 Datei *abst_104.ps*
Fortsetzung der schiefen Projektionsmethode und erste Iterierte $x^{(m)}$ bei $Ax = b$ mit `contours=[0.5,0,-0.02,-0.1552,-0.5,-0.8432,-0.8648,-0.9749,-1]`,

$x^* = b,\ Q(x^*) = -1,\quad x^{(0)} = 0,\ r^{(0)} = b,\quad l^{(0)} = (1,0)^T,\ l^{(1)} = (0,1)^T,$
$L_1 = \text{span}\{l^{(0)}\},\ K_1 = \text{span}\{p^{(0)}\},\ x^{(1)} = (1, -\frac{2}{5})^T,\ r^{(1)} = (0, \frac{7}{5})^T \perp l^{(0)},$
$L_2 = \text{span}\{l^{(1)}\},\ K_2 = \text{span}\{p^{(0)}, p^{(1)}\},\ x^{(2)} = (\frac{39}{25}, 1)^T,\ r^{(2)} = (-\frac{14}{25}, 0)^T \perp l^{(1)},$
$L_3 = L_1,\ K_3 = K_2,\ x^{(3)} = (1, \frac{153}{125})^T,\ r^{(3)} \perp l^{(0)},$
$L_4 = L_2,\ K_4 = K_2,\ x^{(4)} = (\frac{569}{625}, 1)^T,\ r^{(4)} \perp l^{(1)},$
$L_5 = L_1,\ K_5 = K_2,\ x^{(5)} = (1, \frac{3013}{3125})^T,\ r^{(5)} \perp l^{(0)}$

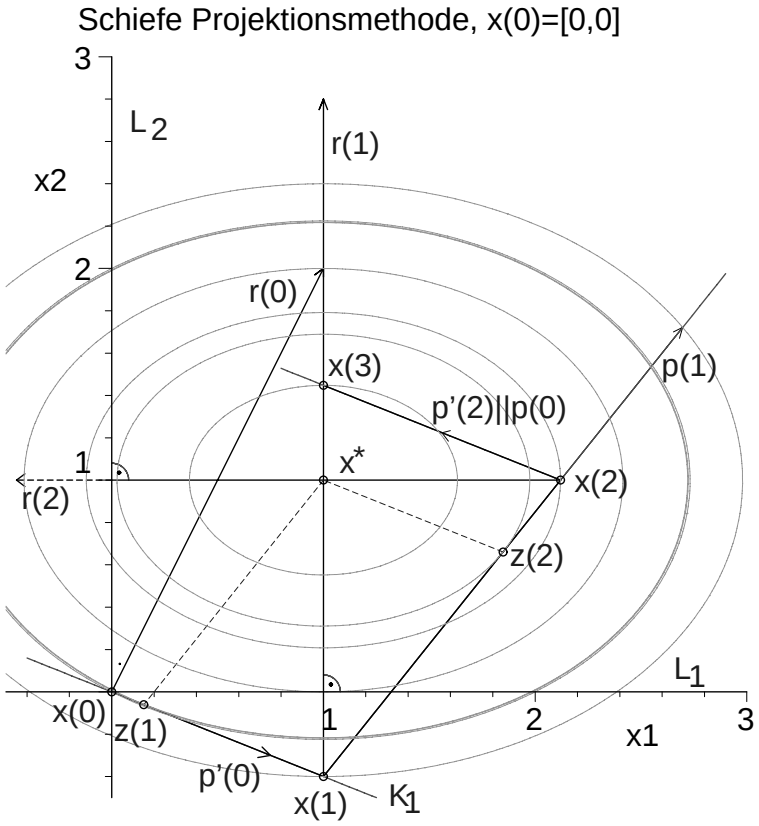


Abb. 3.18 Datei *abst_105.ps*
Schiefe PM und erste Iterierte $x^{(m)}$ bei $Ax = \mathrm{diag}(1,2)\,x = b = (1,2)^T$
mit `contours=[0.46,0,-0.01515,-0.5,-0.8728,-1.0248,-1.2993]`,

$x^* = b,\ Q(x^*) = -\frac{3}{2},\ \ x^{(0)} = 0,\ r^{(0)} = b,$
$l^{(0)} = (1,0)^T,\ l^{(1)} = (0,1)^T,\ p^{(0)} = (5,-2)^T,\ p^{(1)} = \frac{14}{33}(4,5)^T,$
$L_1 = \mathrm{span}\{l^{(0)}\},\ K_1 = \mathrm{span}\{p^{(0)}\},\ x^{(1)} = (1,-\frac{2}{5})^T,\ r^{(1)} = (0,\frac{14}{5})^T \perp l^{(0)},$
$L_2 = \mathrm{span}\{l^{(1)}\},\ K_2 = \mathrm{span}\{p^{(0)}, p^{(1)}\},\ x^{(2)} = (\frac{53}{25},1)^T,\ r^{(2)} = (-\frac{28}{25},0)^T \perp l^{(1)},$
$L_3 = L_1,\ K_3 = K_2,\ x^{(3)} = (1,\frac{181}{125})^T,\ r^{(3)} \perp l^{(0)},$
$L_4 = L_2,\ K_4 = K_2,\ x^{(4)} = (\frac{401}{625},1)^T,\ r^{(4)} \perp l^{(1)},$

$z^{(1)},\ z^{(2)}$ von der Strahlenminimierung:
$Q(x^{(1)}) = \frac{23}{50} = 0.46 > Q(z^{(1)}) = -\frac{1}{66} = -0.015...,\ z^{(1)} = \frac{1}{33}(5,-2),$
$Q(x^{(2)}) = -\frac{1091}{1250} = -0.872... > Q(z^{(2)}) = -\frac{1691}{1650} = -1.024...,\ z^{(2)} = \frac{1}{165}(305,109),$
$Q(x^{(3)}) = -\frac{40603}{31250} = -1.299...$

Das CG als orthogonale Krylov-Unterraum-Methode mit $L_m = K_m = \mathcal{K}_m(A, r^{(0)})$ enthält neben $p^{(0)} = r^{(0)}$ und der Orthogonalität $r^{(m)} = b - Ax^{(m)} = Ae^{(m)} \perp \mathcal{K}_m(A, r^{(0)})$ zusätzlich die Minimierung des Funktionals $Q(x)$ in der Suchrichtung.

Eigenschaften der Suchrichtung $p^{(m)},\ m = 0, 1, ...$

$r^{(0)} = b - Ax^{(0)} = p^{(0)}$, 1. Abstiegsrichtung = steilster Abstieg,
$r^{(0)} \perp$ zur Tangente an die Niveaulinie von $Q(x)$ im Punkt $x^{(0)}$,
$p^{(0)}$ liegt auf der Tangente an eine Niveaulinie im Punkt $x^{(1)}$,
$x^{(1)} = x^{(0)} + \alpha_0 p^{(0)}$,
$r^{(1)} \perp r^{(0)}$,
$p^{(1)}$ wie $r^{(1)}$ geht durch $x^{(1)}$,
$p^{(1)} = r^{(1)} + \beta_0 p^{(0)}$ liegt in der Ebene aufgespannt durch $r^{(1)}$ und $p^{(0)}$,
$p^{(1)}$ und $p^{(0)}$ sind konjugiert in Bezug auf $Q(x) = Q(x^{(1)}) = \text{const}$,
also $(Ap^{(1)}, p^{(0)}) = 0$ bzw. $Ap^{(1)} \perp p^{(0)},\ p^{(1)} \not\perp p^{(0)}$.

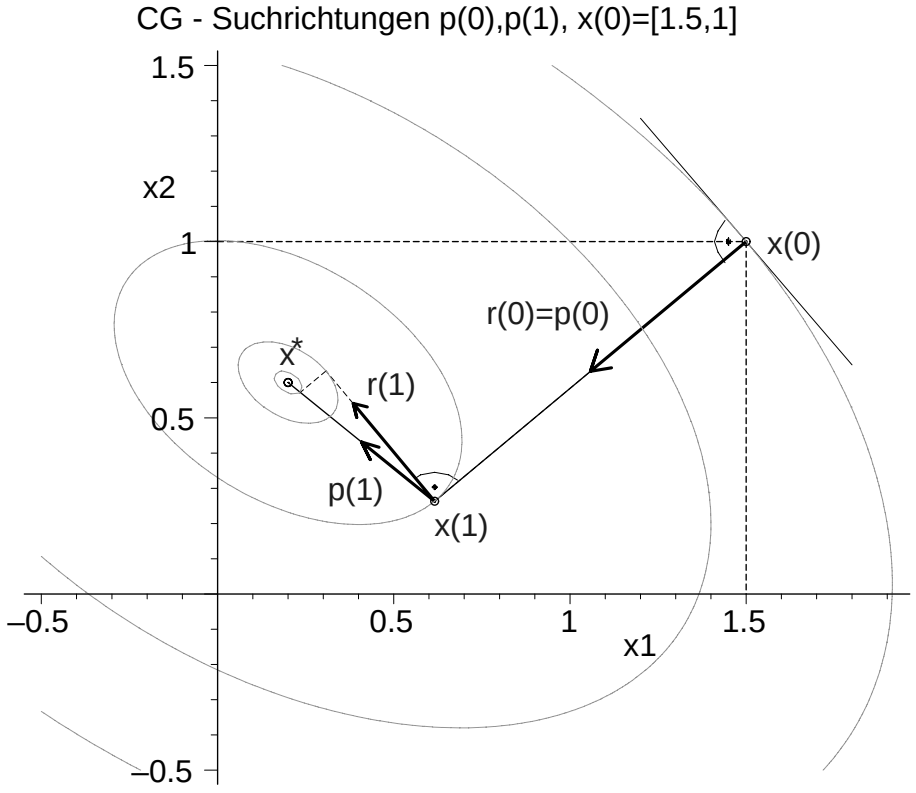


Abb. 3.19 Datei *abst_100.ps*
CG mit den ersten Suchrichtungen $p^{(m)}$ zu $Q(x) = x_1^2 + x_1x_2 + \frac{3}{2}x_2^2 - x_1 - 2x_2$
mit `contours=[1.75,0.5,-0.497,-0.6832,-0.6986,-0.7]`
(eingetragene Vektoren sind skaliert)

3.6.3 Varianten der Realisierung des Verfahrens

Auf die theoretische Endlichkeit des CG, aber eventuelle Fortsetzung bei einer numerischen Rechnung wegen $r^{(n)} \approx 0$ haben wir schon verwiesen.

Version 1 CG mit Indizierung

$x^{(0)}$ Startvektor,
$r^{(0)} = b - Ax^{(0)}$ Anfangsresiduum, negativer Gradient,
$p^{(0)} = r^{(0)}$ 1. Abstiegsrichtung,
ε Toleranz für den Test auf Abbruch der Iteration,
falls $\gamma_0 = \|r^{(0)}\|_2^2 < \varepsilon$, dann $x^* = x^{(0)}$ und Stopp.

m = 0,1,...,n–1,...,K

$$
\begin{aligned}
\alpha_m &= \frac{p^{(m)T} r^{(m)}}{p^{(m)T} w^{(m)}} = \frac{\| r^{(m)} \|_2^2}{\| p^{(m)} \|_A^2}, \quad \text{wobei } w^{(m)} = Ap^{(m)} \\
x^{(m+1)} &= x^{(m)} + \alpha_m p^{(m)} \\
r^{(m+1)} &= r^{(m)} - \alpha_m w^{(m)} \\
\gamma_{m+1} &= \| r^{(m+1)} \|_2^2, \quad \text{falls } \frac{\gamma_{m+1}}{\gamma_0} < \varepsilon, \quad \text{dann break} \\
\beta_m &= -\frac{r^{(m+1)T} Ap^{(m)}}{p^{(m)T} Ap^{(m)}} = -\frac{r^{(m+1)T} w^{(m)}}{p^{(m)T} w^{(m)}} = \frac{\| r^{(m+1)} \|_2^2}{\| r^{(m)} \|_2^2} \\
p^{(m+1)} &= r^{(m+1)} + \beta_m p^{(m)}
\end{aligned}
\tag{3.96}
$$

end m

Näherungslösung $x^* = x^{(m+1)}$

Bemerkung 3.4

(1) Das Produkt $w^{(m)} = Ap^{(m)}$ tritt in den Formeln für α_m und $r^{(m+1)}$ auf und ist nur einmal zu berechnen. Man kann es auch für β_m benutzen oder alternativ dafür die Norm $\|r^{(m)}\|_2^2$ aus zwei aufeinander folgenden Iterationen.

(2) Beim CG wird die m-te Iterierte $x^{(m)}$ als dasjenige Element aus dem Unterraum $x^{(0)} + \mathcal{K}_m(A, r^{(0)})$ bestimmt, für welches die quadratische Form $Q(x)$ ihren minimalen Wert annimmt. Die Existenz dieses Minimums ist für spd Matrizen garantiert.

(3) Die noch folgende vorkonditionierte Variante des CG verwendet einen anderen Teilraum für die Konstruktion der Iterierten, aber das gleiche zu minimierende Funktional.

Version 2 **CG ohne Indizierung, mit Test auf den relativen Fehler**

K maximale Iterationsanzahl,
x Startvektor,
$r = b - Ax$ Anfangsresiduum, negativer Gradient,
$p = r$ 1. Abstiegsrichtung,
$\gamma[0..K]$ Vektor der Normquadrate der euklidischen Norm von r, $\gamma_0 = \|r\|_2^2$,
ε Toleranz für den Test auf Abbruch der Iteration,
falls $\gamma_0 < \varepsilon$, dann $x^* = x$ und Stopp.

m = 0,1,...,n–1,...,K

$$
\begin{aligned}
w &= Ap \\
d &= w^T p = p^T w \\
h &= \gamma_{m-1}/d, \quad \text{Hilfsgröße für } \alpha \text{ und } \beta \\
x &= x + hp \\
r &= r - hw \\
\gamma_m &= \| r \|_2^2, \quad \text{falls } \gamma_m/\gamma_0 < \varepsilon, \quad \text{dann break} \\
h &= \gamma_m/\gamma_{m-1} \\
p &= r + hp
\end{aligned}
\tag{3.97}
$$

end m

Näherungslösung $x^* = x$

Wenn A nicht symmetrisch ist, dann bildet man das Normalgleichungssystem $A^T Ax = Bx = c = A^T b$ mit $B = B^T > 0$, auf welches das CG angewendet werden kann. Von der Kondition der Koeffizientenmatrix her ist aber die Symmetrisierung des LGS nicht unbedingt der günstigste Weg, auch wenn sie die Matrix in gewisser Weise skaliert.
Daraus ergibt sich das zu minimierende Funktional

$$
\begin{aligned}
Q(x) &= \frac{1}{2} x^T Bx - x^T c \;=\; -\frac{1}{2}(Ax)^T(2b - Ax) \\
&= -\frac{1}{2}(Ax)^T s, \quad s = r + b = 2b - Ax = 2b - v, \\
&= \frac{1}{2}(\|r(x)\|_2^2 - b^T b),
\end{aligned}
$$

welches somit wegen (2.21) auf $R(x)$ führt.

Man minimiert also die A^TA-Norm des Fehlers von $x^{(m)}$, was der Norm $\|r(x)\|_2$ entspricht.
Die Matrix B hat im Allgemeinen eine schlechtere Kondition, was zu einer Verlangsamung der Konvergenz führen kann.

Bei Implementierungen wird die Matrix $B = A^TA$ wegen des Aufwands von $2n^3$ Operationen nicht explizit ermittelt, sondern im CG hat man anstelle einer Matrix-Vektor-Multiplikation in der Hauptschleife nun zwei.

Version 3 CG mit Symmetrisierung für beliebiges A mit Indizierung

$x^{(0)}$ Startvektor,
$r^{(0)} = b - Ax^{(0)}$ Anfangsresiduum, negativer Gradient,
$s^{(0)} = A^Tr^{(0)}$ $(s = A^Tr)$,
$p^{(0)} = s^{(0)}$ 1. Abstiegsrichtung,
ε Toleranz für den Test auf Abbruch der Iteration,
falls $\gamma_0 = \|r^{(0)}\|_2^2 < \varepsilon$, dann $x^* = x^{(0)}$ und Stopp.

m = 0,1,...,n–1,...,K

$$
\begin{aligned}
w^{(m)} &= Ap^{(m)} \\
\alpha_m &= \frac{s^{(m)T}s^{(m)}}{w^{(m)T}w^{(m)}} = \frac{\| s^{(m)} \|_2^2}{\| w^{(m)} \|_2^2} \\
x^{(m+1)} &= x^{(m)} + \alpha_m p^{(m)} \\
r^{(m+1)} &= r^{(m)} - \alpha_m w^{(m)} \\
\gamma_{m+1} &= \| r^{(m+1)} \|_2^2, \quad \text{falls} \quad \frac{\gamma_{m+1}}{\gamma_0} < \varepsilon, \quad \text{dann break} \\
s^{(m+1)} &= s^{(m)} - \alpha_m A^T w^{(m)} \quad \text{oder} \quad s^{(m+1)} = A^T r^{(m+1)}) \\
\beta_m &= \frac{\| s^{(m+1)} \|_2^2}{\| s^{(m)} \|_2^2} \\
p^{(m+1)} &= s^{(m+1)} + \beta_m p^{(m)}
\end{aligned}
\tag{3.98}
$$

end m

Näherungslösung $x^* = x^{(m+1)}$

Bemerkung 3.5

(1) Als Produkte sind $Ap^{(m)}$ und A^Tz zu berechnen.

(2) Da das CG auf das Normalgleichungssystem angewendet wird, spricht man auch vom CGNR-Verfahren (conjugate gradient normal residual method).

Wir betrachten noch kurz analog zum vorkonditionierten GV das CG mit Vorkonditierung (PCG, preconditioned conjugate gradient method) für das LGS.
Als wichtige Voraussetzung für die Durchführbarkeit erweist sich, dass eine spd Vorkonditionierungsmatrix C (preconditioner) genommen wird. Zusätzlich sollte sie möglichst noch von einfacher faktorisierter Form, so dass der Mehraufwand, der nun durch die zusätzliche Lösung eines Gleichungssystems entstehen wird, von der Größenordnung nicht den einer Matrix-Vektor-Multiplikation übersteigt.

Version 4 Vorkonditioniertes CG mit Indizierung

$x^{(0)}$ Startvektor,
$r^{(0)} = b - Ax^{(0)}$ Anfangsresiduum, negativer Gradient,
C Vorkonditionierungsmatrix,
$Cs^{(0)} = r^{(0)}$ $(Cs = r)$,
$p^{(0)} = s^{(0)}$ 1. Abstiegsrichtung,
ε Toleranz für den Test auf Abbruch der Iteration,
falls $\gamma_0 = \|r^{(0)}\|_2^2 < \varepsilon$, dann $x^* = x^{(0)}$ und Stopp.

m = 0,1,2,...,n–1,...,K

$$
\begin{aligned}
w^{(m)} &= Ap^{(m)} \\
\alpha_m &= \frac{r^{(m)T} s^{(m)}}{p^{(m)T} w^{(m)}} \\
x^{(m+1)} &= x^{(m)} + \alpha_m p^{(m)} \\
r^{(m+1)} &= r^{(m)} - \alpha_m w^{(m)} \\
\gamma_{m+1} &= \| r^{(m+1)} \|_2^2, \quad \text{falls } \frac{\gamma_{m+1}}{\gamma_0} < \varepsilon, \quad \text{dann break} \\
Cs^{(m+1)} &= r^{(m+1)} \\
\beta_m &= \frac{s^{(m+1)T} r^{(m+1)}}{s^{(m)T} r^{(m)}} \\
p^{(m+1)} &= s^{(m+1)} + \beta_m p^{(m)}
\end{aligned}
\tag{3.99}
$$

end m

Näherungslösung $x^* = x^{(m+1)}$

Eine Implementierung des CG soll sich auf die Versionen 1 (3.96) und 2 (3.97) stützen. Ergänzt um die zusätzliche Abfrage des Nenners bei der Berechnung der Schrittzahl α, die Bestimmung von Werten des Funktionals sowie einige Ergebnisfelder und Zwischenausgaben entsteht dann die erweiterte Maple-Prozedur in Anlehnung an diese Versionen des CG, die wir in den Beispielrechnungen verwenden.

Dieser Algorithmus liefert in Maple die folgenden Kommandos.

```
> cg:=proc(n::posint,A::matrix,b::vector,x0::vector,
          maxiter::posint,etol::numeric,aus::name)

   local i,k,m,p,r,v,w,d,x,xx,gamma0,gamman,gammaa,alpha,beta,
         Q,fh,fh1,fh2,xv1,rv1,pv1;
   global xv,rv,pv,lp,lr;

  x:=evalm(x0):
  v:=evalm(A&*x);
  r:=evalm(b-v): lr:=evalm(r):
  p:=evalm(r);   lp:=evalm(p):
  gamma0:=evalm(transpose(r)&*r):
  Q:=0.5*evalm(transpose(x)&*A&*x)-evalm(transpose(x)&*b);

  fh2:=`%+.16e`;   # Ausgabeformate einstellen
  fh1:=`%+.16e`;
  fh :=fh1;
  for m from 2 to n do
   fh:=cat(fh,` `,fh1);
  end do;

  xx:=matrix(n,0,[]):
  xv1:=concat(xx,x);
  xv:=evalm(xv1);
  rv1:=concat(xx,r);
  rv:=evalm(rv1);
  pv1:=concat(xx,p);
  pv:=evalm(pv1);

  if aus=ja then
   printf(`\n`):
   printf(`Startvektor            x = [`||fh||`]\n`,seq(x[i],i=1..n));
   printf(`Residuum        r = b-Ax = [`||fh||`]\n`,seq(r[i],i=1..n));
   printf(`Funktionswert       Q(x) = `||fh1||`\n`,Q);
   printf(`Anfangsfehlerquadrat r'r = `||fh2||`\n\n`,gamma0);
  end if;

  if gamma0<etol then RETURN(x,0); end if;
  gammaa:=gamma0:

  for k from 1 by 1 to maxiter do
   w:=evalm(A&*p);
   d:=evalm(transpose(p)&*w);
   if d=0 then
    lprint(`Abbruch wegen Nenner p'Ap=0`):
```

```
 RETURN(x,k-1);
end if;

alpha:=gammaa/d;
x:=evalm(x+alpha*p);
v:=evalm(v+alpha*w);
r:=evalm(b-v);  lr:=evalm(r):
gamman:=evalm(transpose(r)&*r);
Q:=0.5*evalm(transpose(x)&*A&*x)-evalm(transpose(x)&*b);

xv1:=concat(xv,x); xv:=evalm(xv1);
rv1:=concat(rv,r); rv:=evalm(rv1);

if aus=ja then
 printf(‘Schritt k = %g\n‘,k);
 printf(‘Suchrichtung           p = [‘||fh||‘]\n‘,seq(p[i],i=1..n));
 printf(‘Suchschritt        alpha = ‘||fh1||‘\n‘,alpha);
 printf(‘Iterationsvektor       x = [‘||fh||‘]\n\n‘,seq(x[i],i=1..n));
 printf(‘Residuum        r = b-Ax = [‘||fh||‘]\n‘,seq(r[i],i=1..n));
 printf(‘Funktionswert       Q(x) = ‘||fh1||‘\n‘,Q);
 printf(‘Fehlernormquadrat    r’r = ‘||fh2||‘\n‘,gamman);
end if;

beta:=gamman/gammaa;
p:=evalm(r+beta*p); lp:=evalm(p);
gammaa:=gamman;

pv1:=concat(pv,p);
pv:=evalm(pv1);

if gamman/gamma0<etol then break; end if;   # relativer Fehler

if aus=ja then
 printf(‘Schritt            beta = ‘||fh1||‘\n‘,beta);
 printf(‘neue Suchrichtung     p = [‘||fh||‘]\n\n‘,seq(p[i],i=1..n));
end if;

end do:

if k>maxiter then k:=k-1; end if;
[x,k];

end:
```

3.6.4 Zur Konvergenz des Verfahrens

Bezüglich Konvergenzaussagen verweisen wir auf [8], [10] und [15].

Unter der Voraussetzung wie beim GV (siehe Abschnitt 3.2.1) erhalten wird für den Fehler $e^{(m)} = x^* - x^{(m)}$ verschiedene Abschätzungen.

Für ihren Nachweis benötigt man zwei Dinge.

Einmal sind es einige Eigenschaften von Tschebyscheff-Polynomen im Zusammenhang mit anderen Formeln. Die meisten sind in [17] aufgeführt, einige davon auch bewiesen.

Tschebyscheff-Polynome 1. Art

$$
\begin{aligned}
T_m(x) &= \cos(m \arccos(x)), \quad T_m : [-1,1] \to [-1,1], \quad m = 0,1,..., \\
&\quad \text{(explizite Darstellung)}, \\
T_m(x) &= 2xT_{m-1}(x) - T_{m-2}(x), \quad m = 2,3,..., \; T_0(x) = 1, \; T_1(x) = x, \\
&\quad \text{(rekursive Form, Nachweis mittels Cosinus-Additionstheorem)}, \\
&= 2^{m-1}x^m - a_2x^{m-2} + ... \; = \; 2^{m-1}(x-x_0)(x-x_1)\cdot ... \cdot (x-x_{m-1}), \\
&\quad \text{Hauptkoeffizient } a_0 \text{ von } T_m(x) \text{ bei } x^m \text{ ist } 2^{m-1}, \\
|T_m(x)| &\leq 1, \quad \|T_m\|_\infty = 1 \text{ für } |x| \leq 1, \\
T_m(x) &\text{ hat die } m{+}1 \text{ Extremalstellen } \bar{x}_i = -\cos\left(\tfrac{i}{m}\pi\right) \in [-1,1], \quad i = 0,...,m, \\
&\quad T_m(\bar{x}_i) = (-1)^{m-i} = \pm 1, \quad i = 0,1,...,m, \quad T_m(1) = 1, \\
T_m(x) &= \cosh(m \operatorname{arcosh}(x)), \quad |x| \geq 1, \quad \cosh(z) = \tfrac{1}{2}(e^z + e^{-z}), \\
&\quad \text{(Nachweis mittels Rekursionsformel)}.
\end{aligned}
$$

Die Fortsetzung der Tschebyscheff-Polynome auf $\mathbb{R}$ ist

$$
T_m(x) = \begin{cases} \cos(m \arccos(x)), & \text{falls } x \in [-1,1], \\ \cosh(m \operatorname{arcosh}(x)), & \text{falls } x \geq 1, \\ (-1)^m \cosh(m \operatorname{arcosh}(-x)), & \text{falls } x \leq -1. \end{cases}
$$

Damit gilt auf $\mathbb{R}$ für gerades m die Beziehung $T_m(-x) = T_m(x)$ und für ungeraden Index $T_m(-x) = -T_m(x)$. Mit $T_1(x) = x$ und

$$
x = \frac{1}{2}\Big(x \pm \sqrt{x^2-1} + \frac{1}{x \pm \sqrt{x^2-1}}\Big) \tag{3.100}
$$

hat man

$$
T_1(x) = \frac{1}{2}\Big(x \pm \sqrt{x^2-1} + \frac{1}{x \pm \sqrt{x^2-1}}\Big), \quad x \geq 1.
$$

Sehr nützlich ist die Formel

$$T_m\Big(\frac{1}{2}\Big(x+\frac{1}{x}\Big)\Big) \;=\; \frac{1}{2}\Big(x^m+\frac{1}{x^m}\Big) \;=\; \frac{1+x^{2m}}{2x^m}, \tag{3.101}$$

die mit vollständiger Induktion gezeigt werden kann.
Folglich gilt

$$\begin{aligned} T_m(x) &= T_m\Big(\frac{1}{2}\Big(x\pm\sqrt{x^2-1}+\frac{1}{x\pm\sqrt{x^2-1}}\Big)\Big), \quad x\in\mathbb{C} \\ &= \frac{1}{2}[(x-\sqrt{x^2-1})^m+(x-\sqrt{x^2-1})^{-m}] \\ T_m(x) &= \frac{1}{2}[(x+\sqrt{x^2-1})^m+(x+\sqrt{x^2-1})^{-m}], \qquad (3.102) \\ &\quad \text{(Nachweis auch mit Substitution } x=\cos(\xi)), \\ &= x^m-\binom{m}{2}x^{m-2}(x^2-1)+\binom{m}{4}x^{m-4}(x^2-1)^2-... \\ &\qquad \pm\begin{cases} \binom{m}{m}(x^2-1)^{m/2}, & \text{falls } m \text{ gerade,} \\ \binom{m}{m-1}x(x^2-1)^{(m-1)/2}, & \text{falls } m \text{ ungerade.} \end{cases} \end{aligned}$$

Die Formel (3.102) benutzen wir später für reelles Argument $x\geq 1$.
Weiterhin hat man die Beziehungen

$$T_m(x) \;=\; \frac{1}{2}\Big[(x+\sqrt{x^2-1})^m+\Big(\frac{1}{x+\sqrt{x^2-1}}\Big)^m\Big] \;\geq\; x^m \geq \frac{1}{2}x^m \;\text{ für } x\geq 1, \tag{3.103}$$

$$|T_m(x)| \;\geq\; |x|^m \;\text{ für } |x|\geq 1, \tag{3.104}$$

$$\frac{x+1}{x-1} \;=\; \frac{1}{2}\Big(\frac{\sqrt{x}+1}{\sqrt{x}-1}+\frac{\sqrt{x}-1}{\sqrt{x}+1}\Big) \;\text{ für } x>1, \tag{3.105}$$

und mit den Größen

$$1 \;>\; \eta \;=\; \frac{x-1}{x+1} \;>\; \frac{\sqrt{x}-1}{\sqrt{x}+1} \;=\; c \tag{3.106}$$

die Formeln

$$\begin{aligned} \eta &= \frac{2c}{1+c^2}, \quad c \;=\; \frac{1-\sqrt{1-\eta^2}}{\eta}, \\ 1 \;<\; \frac{1}{\eta} &= \frac{1}{2}\Big(c+\frac{1}{c}\Big) \;=\; \frac{1+c^2}{2c}, \\ 1 \;<\; \frac{1}{c} &= \frac{1+\sqrt{1-\eta^2}}{\eta} \;=\; \frac{\eta}{1-\sqrt{1-\eta^2}} \;=\; \frac{1}{\eta}+\sqrt{\Big(\frac{1}{\eta}\Big)^2-1}\,. \end{aligned}$$

Besonders nützlich ist die **Minimax-Eigenschaft** der Tschebyscheff-Polynome auf $[-1,1]$ im Vergleich mit anderen Polynomen desselben Grades.
Ihre geringe Abweichung von der x-Achse entsteht durch ihr oszillierendes Verhalten sowie die Verteilung der Nullstellen und damit der Extremalstellen im Intervall $[-1,1]$.

Lemma 3.12 *Jedes Polynom $p_m(x)$ m-ten Grades mit führendem Koeffizient $a_0 \neq 0$ nimmt in $[-1,1]$ einen Wert vom Betrag $\geq |a_0|/2^{m-1}$ an.*
Insbesondere sind die Tschebyscheff-Polynome $T_m(x)$ unter allen Polynomen m-ten Grades mit führendem Koeffizienten 2^{m-1} minimal bezüglich der Maximumnorm der Funktion $\|f\|_\infty = \max\limits_{x\in[-1,1]} |f(x)|$.

Beweis. [17]

Lemma 3.13 *$p_m(x)$ ein beliebiges Polynom n-ten Grades mit dem Koeffizienten 1 bei x^m. Dann gilt*

$$\max_{-1\leq x\leq 1} |2^{1-m}\, T_m(x)| \;\leq\; \max_{-1\leq x\leq 1} |p_m(x)|. \tag{3.107}$$

Lemma 3.14 *Sei $[a,b]$, $a<b$, ein Intervall. Die Aufgabe, die Funktion $\max\limits_{a\leq x\leq b} |p_m(x)|$ unter allen Polynomen m-ten Grades zu minimieren, hat die eindeutige Lösung*

$$p_m(x) \;=\; T_m\Big(\frac{2x-(a+b)}{b-a}\Big). \tag{3.108}$$

Es gilt $\|p_m(x)\|_\infty = \max\limits_{a\leq x\leq b} |p_m(x)| = 1$.

Lemma 3.15 *Sei $[a,b]$, $a<b$, ein Intervall. Die Aufgabe, die Funktion $\max\limits_{a\leq x\leq b} |p_m(x)|$ unter allen Polynomen mit höchstens m-tem Grad und $p_m(1)=1$ zu minimieren, hat die eindeutige Lösung*

$$p_m(x) \;=\; \frac{T_m\Big(\dfrac{2x-(a+b)}{b-a}\Big)}{T_m\Big(\dfrac{2-(a+b)}{b-a}\Big)}. \tag{3.109}$$

Das minimierende Polynom hat den Grad m und führt auf das Minimum

$$\max_{a\leq x\leq b} |p_m(x)| \;=\; \frac{1}{\Big|T_m\Big(\dfrac{2-(a+b)}{b-a}\Big)\Big|}. \tag{3.110}$$

Beweis. [8] für den Fall $-\infty < a < b < 1$.

Bemerkung 3.6

(1) Wegen $t = \frac{2x-(a+b)}{b-a} \in [-1,1]$ ist die Funktion im Zähler von (3.109) beschränkt durch $|T_m(t)| \le 1$, nimmt aber für einige Argumente die Werte ± 1 an.

(2) Die Betrachtung kann man auch auf das Intervall $[a,b]$ mit $1 < a < b < \infty$ anwenden. Im Fall $-\infty < a < b < 1$ ist $t_0 = \frac{2-(a+b)}{b-a} > 1$, während bei $1 < a < b < \infty$ die Ungleichung $t_0 < -1$ gilt. In beiden Situationen erhält man $|t_0| \ge 1$ sowie mit (3.104) die Abschätzungen $|T_m(t_0)| \ge |t_0|^m$ und

$$\max_{a\le x\le b} |p_m(x)| \;=\; \frac{1}{C_m}, \quad C_m = \Big|T_m\Big(\frac{2-(a+b)}{b-a}\Big)\Big| \ge \Big|\frac{2-(a+b)}{b-a}\Big|^m.$$

Damit kann $\max\limits_{a\le x\le b} |p_m(x)|$ durch die Schranke $|\frac{b-a}{2-(a+b)}|^m < 1$ abgeschätzt werden.

(3) Für $a \le 1 \le b$ ist $t_0 \in [-1,1]$ und der Wert C_m kann nahe Null sein, so dass die Schranke $\frac{1}{C_m}$ nicht viel nutzt.

(4) Wir wollen den Wert $\frac{1}{C_m}$ in der Formel (3.110) nun genauer untersuchen.
Ohne Beschränkung der Allgemeinheit betrachten wir den Fall $t_0 > 1$ und ermitteln die Größe (3.103)

$$T_m(t_0) \;=\; \frac{1}{2}\Big[(t_0+\sqrt{t_0^2-1}\,)^m + \Big(\frac{1}{t_0+\sqrt{t_0^2-1}}\Big)^m\Big], \quad t_0 = \frac{(1-a)+(1-b)}{b-a}.$$

Man rechnet nach, dass

$$\begin{aligned}
t_0^2-1 &= \frac{(1-a)^2+(1-b)^2+2(1-a)(1-b)}{(b-a)^2} - \frac{[(1-a)-(1-b)]^2}{(b-a)^2} \\
&= \frac{4(1-a)(1-b)}{(b-a)^2}, \\
t_0+\sqrt{t_0^2-1} &= \frac{(1-a)+(1-b)}{b-a} + \frac{2\sqrt{1-a}\,\sqrt{1-b}}{b-a} \\
&= \frac{(\sqrt{1-a}+\sqrt{1-b})^2}{b-a} \;=\; \frac{(\sqrt{1-a}+\sqrt{1-b})^2}{(1-a)-(1-b)} \\
&= \frac{(\sqrt{1-a}+\sqrt{1-b})^2}{(\sqrt{1-a}-\sqrt{1-b})(\sqrt{1-a}+\sqrt{1-b})} \\
&= \frac{\sqrt{1-a}+\sqrt{1-b}}{\sqrt{1-a}-\sqrt{1-b}} \;=\; \frac{\sqrt{\frac{1-a}{1-b}}+1}{\sqrt{\frac{1-a}{1-b}}-1},
\end{aligned}$$

$$
\begin{aligned}
t_0 + \sqrt{t_0^2 - 1} &= \frac{1}{d}, \quad d = \frac{\sqrt{\tau} - 1}{\sqrt{\tau} + 1}, \ \tau = \frac{1 - a}{1 - b}, \\
\frac{1}{C_m} &= \frac{1}{T_m(t_0)} = \frac{1}{T_m\Big(\frac{1}{d}\Big)} \\
&= \frac{1}{\frac{1}{2}\Big[\Big(\frac{1}{d}\Big)^m + d^m\Big]} \\
&= \frac{2d^m}{1 + d^{2m}}.
\end{aligned}
$$

Man kann daher die Größe $\tau = \frac{1-a}{1-b}$ als Konditionszahl einer Matrix $A = A^T > 0$ interpretieren, denn mit ihren Eigenwerten z. B. $b = \lambda_{max} \approx 1$ und $a = \lambda_{min} \approx 0$ erhält man

$$
\kappa(A) = \frac{\lambda_{max}}{\lambda_{min}} \approx \frac{1 - \lambda_{min}}{1 - \lambda_{max}} = \frac{1 - a}{1 - b}.
$$

Der Faktor $d = \frac{\sqrt{\tau}-1}{\sqrt{\tau}+1}$ taucht in der Fehlerschätzung zur Iteration auf und spielt dieselbe Rolle wie die Größen c bzw. η in den Abstiegsverfahren (vergleiche Kap. 3.2.1 und Formel (3.106)).
In der Konvergenzuntersuchung zum CG ist zu erwarten, dass auf der Basis der Matrixkondition die Formeln zu d und $\frac{1}{C_m}$ auftreten werden.

Ausgehend von Lemma 3.15 verschieben wir die Betrachtung der transformierten Tschebyscheff-Polynome vom Intervall $[a, b],\ -\infty < a < b < 1$ auf das Intervall mit $0 < a < b < \infty$.

Lemma 3.16 *Sei $[a, b],\ 0 < a < b < \infty$, ein Intervall. Die Aufgabe, die Funktion $\max\limits_{a \le x \le b} |p_m(x)|$ unter allen Polynomen mit höchstens m-tem Grad und $p_m(0) = 1$ zu minimieren, hat die eindeutige Lösung*

$$
p_m(x) = \frac{T_m\Big(\frac{b + a - 2x}{b - a}\Big)}{T_m\Big(\frac{b + a}{b - a}\Big)}. \tag{3.111}
$$

Das minimierende Polynom hat den Grad m und führt auf das Minimum

$$
\max_{a \le x \le b} |p_m(x)| = \frac{1}{T_m\Big(\frac{b + a}{b - a}\Big)}. \tag{3.112}
$$

Beweis.
(1) Wegen $t_0 = \frac{b+a}{b-a} > 1$ ist die Konstante $C_m = T_m(t_0) \neq 0$.
Das transformierte Polynom $p_m(x)$ ist per Konstruktion genau wie das Tschebyscheff-Polynom T_m vom Grad m und es gilt $p_m(0) = 1$.
Für $x \in [a, b]$ ist $t = \frac{b+a-2x}{b-a} \in [-1, 1]$. Dort ist $|T_m| \leq 1$ und T_m nimmt diese Grenze auch an, so dass (3.112) folgt.
(2) Es bleibt zu zeigen, dass für jedes andere Polynom m-ten Grades das Maximum in (3.112) größer wird.
Sei $q_m(x)$ ein solches Polynom mit $q_m(0) = 1$ und $\max\limits_{a\leq x\leq b} |q_m(x)| \leq \frac{1}{C_m}$. An den $m+1$ verschiedenen Extremalstellen $\bar{x}_i$ von T_m mit den wechselnden Funktionswerten ± 1 hat das Polynom p_m die Werte

$$p_m(\xi_i) \;=\; \pm\frac{(-1)^i}{C_m}, \quad \xi_i = \frac{1}{2}[b + a - (b - a)\bar{x}_i], \;\; i = 0, 1, ..., m.$$

Aus $|q_m(\xi_i)| \leq \frac{1}{C_m} = |p_m(\xi_i)|$ schließt man für die Differenz $r = p_m - q_m$, dass sie an den $m+1$ Stellen $\xi_i \in [a, b]$ abwechselnd die Bedingung $r(\xi_i) \geq 0$ und $r(\xi_i) \leq 0$ erfüllt. Nach dem Zwischenwertsatz existiert in jedem Teilintervall $[\xi_i, \xi_{i+1}]$ eine Nullstelle von r. Falls die Nullstellen in $[\xi_{i-1}, \xi_i]$ und $[\xi_i, \xi_{i+1}]$ auf den gemeinsamen Punkt ξ_i fallen, wird sie als zweifache mitgezählt. Wegen $p_m(0) = q_m(0) = 1$ hat r in $1 \notin [a, b]$ eine $(m + 1)$-te Nullstelle, Als Polynom m-ten Grades muss r somit identisch Null sein. Daraus folgt $q_m = p_m$ und die Eindeutigkeit $q_m = p_m$. □.

Bemerkung 3.7
(1) Im Spezialfall $a = b$ definieren wir das einfache Polynom $p_m(x) = 1 - \frac{x}{b}$, womit $p_m(0) = 1$ und $\max\limits_{a\leq x\leq b} |p_m(x)| = 0$ gilt.
(2) Wir wollen die rechte Seite in der Formel (3.112) nun genauer untersuchen und daraus verschiedene Abschätzungen herleiten. Dazu verwenden wir die Beziehungen (3.103), (3.105), (3.101) und (3.106).
Interpretiert man die Größe $\tau = \frac{b}{a}$ als Konditionszahl einer Matrix $A = A^T > 0$ mit den Eigenwerten $b = \lambda_{max}$ und $a = \lambda_{min} > 0$ erhält man

$$t_0 \;=\; \frac{b+a}{b-a} \;=\; \frac{\tau+1}{\tau-1} \;=\; \frac{1}{\eta}, \quad \frac{1}{c} = \frac{\sqrt{\tau}+1}{\sqrt{\tau}-1}$$

und die Beziehungen

$$\frac{1}{T_m(t_0)} = \frac{1}{T_m\Big(\frac{1}{\eta}\Big)} \begin{cases} \overset{(3.103)}{\leq} \quad 2\eta^m & \text{(Abschätzung schlechter als GV)}, \\ \overset{(3.103)}{\leq} \quad \eta^m & \text{(Abschätzung wie GV)}, \\ = \dfrac{1}{T_m\Big(\frac{\tau+1}{\tau-1}\Big)} \overset{(3.105)}{=} \dfrac{1}{T_m\Big(\frac{1}{2}\Big(\frac{\sqrt{\tau}+1}{\sqrt{\tau}-1} + \frac{\sqrt{\tau}-1}{\sqrt{\tau}+1}\Big)\Big)} = \dfrac{1}{T_m\Big(\frac{1}{2}\Big(\frac{1}{c} + c\Big)\Big)}, \end{cases} \tag{3.113}$$

und weiter

$$\frac{1}{T_m\Big(\frac{1}{2}\Big(\frac{1}{c}+c\Big)\Big)} \overset{(3.101)}{=} \frac{1}{\frac{1}{2}\Big(\frac{1}{c^m}+c^m\Big)} \begin{cases} \leq \dfrac{2}{c^m} & \text{(zu grobe Abschätzung)}, \\ \leq 2c^m & \text{(gute Abschätzung)}, \\ = \dfrac{2c^m}{1+c^{2m}} & \text{(beste Formel)}. \end{cases} \tag{3.114}$$

Ein zweiter Sachverhalt ist die mögliche Darstellung von Iterierten des AV als Polynome in Potenzen von A. Dies haben wir im CG andeutungsweise in der Beziehung (3.87) für die Residua $r^{(m)}$ gesehen, z. B. $r^{(2)} = (I + a_1 A + a_2 A^2)\, r^{(0)}$.
Später kommen wir auf diese polynomiale Iterationsformel zurück. Hier wollen wir eine solche Beziehung voraussetzen und sie allgemein als

$$r^{(m)} \;=\; \Big(\sum_{i=0}^{m} a_i A^i\Big)\, r^{(0)} \;=\; p_m(A)\, r^{(0)} \tag{3.115}$$

notieren, wobei $p_m(A) \in \mathcal{P}_m^1$ und $p_m(0) = I$ ist.
Mit $\mathcal{P}_m^1$ bezeichnen wir dabei die Menge aller Polynome vom Höchstgrad m, die der Nebenbedingung $p(0) = I$ genügen.
Wegen $e^{(m)} = x^* - x^{(m)} = A^{-1}b - x^{(m)} = A^{-1}(b - Ax^{(m)}) = A^{-1}r^{(m)}$ gilt auch

$$e^{(m)} \;=\; A^{-1}r^{(m)} \;=\; A^{-1}p_m(A)r^{(0)} \;=\; A^{-1}p_m(A)A\,A^{-1}r^{(0)} \;=\; p_m(A)\, e^{(0)}. \tag{3.116}$$

Alle diese Informationen fließen in den folgenden Konvergenzsatz zum CG ein.

Satz 3.17 Konvergenz des CG
Sei die Matrix A spd und $x^{(m)}$ die m-te Iterierte des CG (3.73).
Der zugehörige Fehlervektor $e^{(m)} = x^* - x^{(m)} = A^{-1}b - x^{(m)}$ erfüllt mit der spektralen Konditionszahl (3.39) $\kappa = \mathrm{cond}_2(A) = \frac{\lambda_n}{\lambda_1} \in (1, \infty)$ und

$$0 < c = \frac{\sqrt{\lambda_n}-\sqrt{\lambda_1}}{\sqrt{\lambda_n}+\sqrt{\lambda_1}} = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \leq \eta = \frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1} = \frac{\kappa-1}{\kappa+1} \;=\; \frac{2c}{1+c^2} < 1 \tag{3.117}$$

die folgenden Abschätzungen

$$\begin{aligned} \|e^{(m)}\|_A \;&\leq\; 2\frac{(1-1/\kappa)^m}{(1+1/\sqrt{\kappa})^{2m}+(1-1/\sqrt{\kappa})^{2m}}\,\|e^{(0)}\|_A \;= \\ &=\; \frac{2c^m}{1+c^{2m}}\,\|e^{(0)}\|_A \;=\; \tilde{\eta}(m)\,\|e^{(0)}\|_A, \end{aligned} \tag{3.118}$$

$$\|e^{(m)}\|_A \;\leq\; \Big(\frac{\kappa-1}{\kappa+1}\Big)^m \|e^{(0)}\|_A \;=\; \eta^m\,\|e^{(0)}\|_A \;\leq\; 2\eta^m\,\|e^{(0)}\|_A, \tag{3.119}$$

$$\|e^{(m)}\|_A \;\leq\; 2\Big(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\Big)^m \|e^{(0)}\|_A \;=\; 2c^m\,\|e^{(0)}\|_A \;=\; \hat{\eta}(m)\,\|e^{(0)}\|_A. \tag{3.120}$$

Beweis.
Die Abschätzung (3.118) ist die beste von allen, ihren Nachweis findet man in [8]. Bei der groben Abschätzung (3.119) mit dem Faktor $2\eta^m$ aus [15] sind wir sogar schlechter als die analoge Ungleichung (3.35) zum GV. Aber als Ergänzung ist dort noch die zweite bessere Beziehung (3.120) für den Iterationsfehler gezeigt worden, die ohne Probleme auch für die Herleitung von (3.118) verwendet werden kann.
Wir nutzen die Beziehung (3.116) für den Lösungsfehler mit dem Polynom $p_m \in \mathcal{P}_m^1$ vom Grad m, das der Nebenbedingung $p_m(0) = I$ genügt.
Die Gleichung (3.95) liefert zusammen mit der Beziehung (3.69) zwischen Funktional $Q(x)$ und der Norm $\|e(x)\|_A$ die Aussage

$$\|x^* - x^{(m)}\|_A \;=\; \min_{x \in x^{(0)} + K_m} \|x^* - x\|_A, \quad K_m = \mathcal{K}_m(A, r^{(0)}),$$

wodurch

$$\|e^{(m)}\|_A \;=\; \|e(x^{(m)})\|_A \;=\; \min_{p \in \mathcal{P}_m^1} \|p(A)\, e^{(0)}\|_A \tag{3.121}$$

folgt. Als spd Matrix besitzt A reelle und positive Eigenwerte $0 < \lambda_1 \le \lambda_2 \le ... \le \lambda_n$ und die zugehörigen Eigenvektoren $v_1, v_2, ..., v_n$ können als Orthonormalbasis des $\mathbb{R}^n$ gewählt werden. Somit existiert eine Darstellung des Fehlervektors $e^{(0)}$ in der Form

$$e^{(0)} \;=\; \sum_{i=1}^{n} \alpha_i v_i, \quad \alpha_i \in \mathbb{R}.$$

Hieraus erhalten wir

$$\|e^{(0)}\|_A^2 \;=\; \left(\sum_{i=1}^{n} \alpha_i v_i, \sum_{i=1}^{n} \alpha_i \lambda_i v_i \right) \;=\; \sum_{i=1}^{n} \alpha_i^2 \lambda_i$$

und analog

$$\|p(A)e^{(0)}\|_A^2 \;=\; \sum_{i=1}^{n} [p(\lambda_i)]^2\, \alpha_i^2 \lambda_i.$$

Folglich ergibt sich unter Ausnutzung der Gleichung (3.121) die Ungleichung

$$\begin{aligned} \|e^{(m)}\|_A \;&=\; \min_{p \in \mathcal{P}_m^1} \left(\sum_{i=1}^{n} [p(\lambda_i)]^2\, \alpha_i^2 \lambda_i \right)^{1/2} \\ &\le\; \min_{p \in \mathcal{P}_m^1} \max_{j=1,2,...,n} |p(\lambda_j)| \left(\sum_{i=1}^{n} \alpha_i^2 \lambda_i \right)^{1/2} \\ &\le\; \min_{p \in \mathcal{P}_m^1} \max_{\lambda \in [\lambda_1, \lambda_n]} |p(\lambda)|\; \|e^{(0)}\|_A. \end{aligned} \tag{3.122}$$

Den weiteren Schritt zum Nachweis der Behauptungen haben wir unter Verwendung der Tschebyscheff-Polynome mit dem Lemma 3.16 und den Abschätzungen (3.113) und (3.114) schon vorbereitet.

In diesen Aussagen sei $a = \lambda_1$ und $b = \lambda_n$.

Wir betrachten zunächst den Fall $\lambda_1 \neq \lambda_n$.

Das gesuchte wohl definierte eindeutige Polynom aus $\mathcal{P}_m^1$ ist

$$p_m(\lambda) \;=\; \frac{T_m\Big(\dfrac{\lambda_n + \lambda_1 - 2\lambda}{\lambda_n - \lambda_1}\Big)}{T_m\Big(\dfrac{\lambda_n + \lambda_1}{\lambda_n - \lambda_1}\Big)}.$$

Die Abschätzung (3.122) führt mit den in (3.117) definierten Größen auf die folgenden Varianten.

$$\|e^{(m)}\|_A \;\leq\; \|e^{(0)}\|_A \cdot \begin{cases} 2\eta^m & \text{(Abschätzung schlechter als GV),} \\ \eta^m & \text{(Abschätzung wie GV),} \\ \dfrac{2}{c^m} & \text{(zu grobe Abschätzung),} \\ 2c^m & \text{(gute Abschätzung),} \\ \dfrac{2c^m}{1+c^{2m}} & \text{(beste Abschätzung).} \end{cases} \tag{3.123}$$

Für den Spezialfall $\lambda_1 = \lambda_n$ nutzen wir $p_m \in \mathcal{P}_m^1$ mit $p_m(\lambda) = 1 - \frac{\lambda}{\lambda_n}$ und erhalten dann mit $\eta = 0$ und $c = 0$

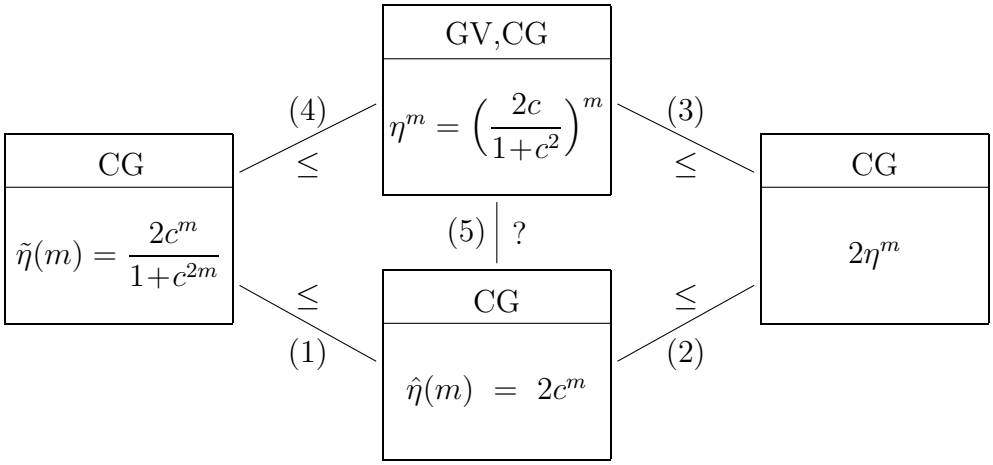
$$\|e^{(m)}\|_A \;\leq\; \max_{\lambda\in[\lambda_1,\lambda_n]} |p(\lambda)|\, \|e^{(m)}\|_A \;=\; 0 \cdot \|e^{(0)}\|_A \;=\; 0 \;=\; \|e^{(0)})\|_A \cdot \begin{cases} 2\eta^m, \\ \eta^m, \\ 2c^m, \\ \dfrac{2c^m}{1+c^{2m}}. \end{cases}$$

Die grobe Abschätzung mit der oberen Schranke $\frac{2}{c^m} = \frac{2}{0} =$ "∞" brauchen wir hier nicht zu berücksichtigen. □

Einige vergleichende Betrachtungen zur Konvergenzgeschwindigkeit von GV und CG

In der Fehlerabschätzung (3.35) des GV haben wir den Faktor η^m erhalten. Wir wollen ihn vergleichen mit denen in den Formeln (3.118) – (3.120) zum CG.

Dazu konstruieren wir das folgende Schema zu den Beziehungen zwischen den Faktoren der Fehlerabschätzungen.



Nachweis zu den Fehlerabschätzungen (1)–(5) im obigen Schema

(1) Die Gültigkeit der Ungleichung

$$\tilde{\eta}(m) \;=\; \frac{2c^m}{1+c^{2m}} \;\le\; 2c^m \;=\; \hat{\eta}(m)$$

für $c \in (0,1)$ ist leicht zu sehen.

(2) Wegen $c \le \eta$ gilt auch $\hat{\eta}(m) = 2c^m \le 2\eta^m$.

(3) Offensichtlich gilt $\eta^m \le 2\eta^m$.

(4) Für diese Ungleichung machen wir einige Illustrationen.

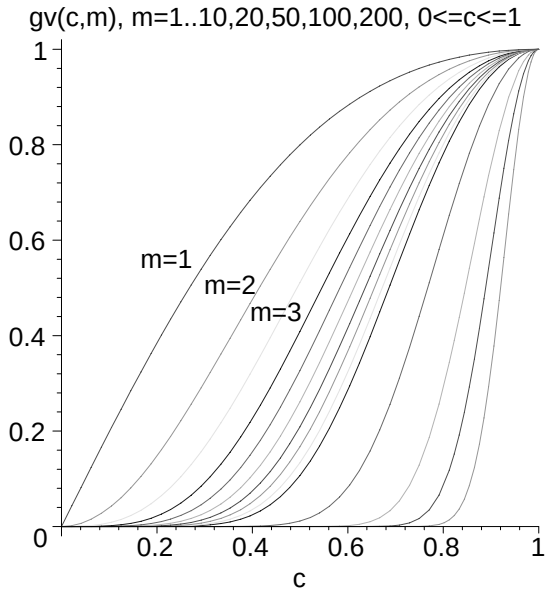


Abb. 3.20 Datei *abst_110.ps*
$gv(c,m) = \eta^m = (\frac{2c}{1+c^2})^m,\ c \in [0,1]$

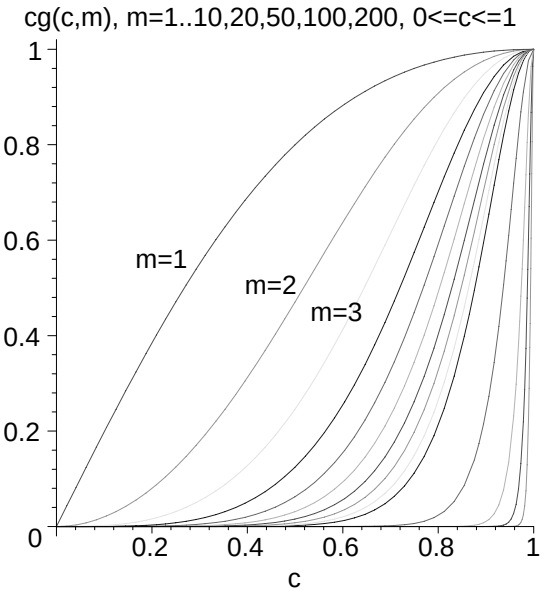


Abb. 3.21 Datei *abst_111.ps*
$cg(c,m) = \tilde{\eta}(m) = \frac{2c^m}{1+c^{2m}},\ c \in [0,1]$

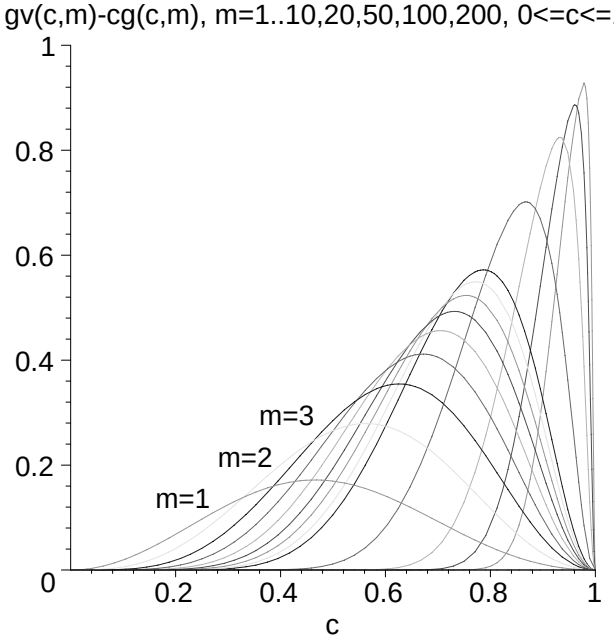


Abb. 3.22 Datei *abst_112.ps*
$\delta(c,m) = gv(c,m) - cg(c,m)$
$= \eta^m - \tilde{\eta}(m)$
$= (\frac{2c}{1+c^2})^m - \frac{2c^m}{1+c^{2m}} \geq 0,$
$c \in [0,1],$
$m = 1(1)10, 20, 50, 100, 200$

Dazu folgen die entsprechenden Rechnungen, speziell der Nachweis von $\delta(c,m) \geq 0$. Mit $m = (0,)1, 2, ...$ sei

$$\begin{aligned} gv(c,m) &= \eta^m = \Big(\frac{2c}{1+c^2}\Big)^m, \quad c = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} \leq \eta = \frac{\kappa-1}{\kappa+1} \in [0,1], \\ cg(c,m) &= \frac{2c^m}{1+c^{2m}}, \\ \delta(c,m) &= gv(c,m) - cg(c,m). \end{aligned} \tag{3.124}$$

Die Sonderfälle $\delta(c,0) \geq 0$ und $\delta(c,1) \geq 0$ sind sofort überprüfbar.
Bei $m = 2$ rechnet man

$$\begin{aligned} \delta(c,2) &= \Big(\frac{2c}{1+c^2}\Big)^2 - \frac{2c^2}{1+c^4} \\ &= 2c^2\,\frac{2(1+c^4) - (1+c^2)^2}{(1+c^2)^2(1+c^4)} \\ &= 2c^2\,\frac{c^4 - 2c^2 + 1}{(1+c^2)^2(1+c^4)} \\ &= 2c^2\,\frac{(1-c^2)^2}{(1+c^2)^2(1+c^4)} \\ &\geq 0. \end{aligned}$$

Für weitere Werte m ist eine analoge Vorgehensweise sehr beschwerlich.
Deshalb wiederholen wir den Schritt $m = 2$ unter Verwendung von (3.117) mit einer anderen Technik und Idee, um damit dann die Ungleichung allgemein zu zeigen.

$$
\begin{aligned}
\delta(c,2) &= \eta^2 - \tilde{\eta}(2) \\
&= \Big(\frac{\kappa-1}{\kappa+1}\Big)^2 - \frac{2\Big(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\Big)^2}{1+\Big(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\Big)^4} \\
&= \frac{(\kappa-1)^2}{(\kappa+1)^2} - \frac{2(\sqrt{\kappa}-1)^2(\sqrt{\kappa}+1)^2}{(\sqrt{\kappa}+1)^4+(\sqrt{\kappa}-1)^4} \\
&= (\kappa-1)^2\Big(\frac{1}{(\kappa+1)^2} - \frac{2}{(\sqrt{\kappa}+1)^4+(\sqrt{\kappa}-1)^4}\Big) \\
&= (\kappa-1)^2\Big(\frac{1}{\kappa^2+2\kappa+1} - \frac{1}{\kappa^2+6\kappa+1}\Big) \\
&= \frac{4\kappa(\kappa-1)^2}{(\kappa^2+2\kappa+1)(\kappa^2+6\kappa+1)} \\
&\geq 0,
\end{aligned}
$$

und allgemein

$$
\begin{aligned}
\delta(c,m) &= \eta^m - \tilde{\eta}(m) \\
&= \Big(\frac{\kappa-1}{\kappa+1}\Big)^m - \frac{2\Big(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\Big)^m}{1+\Big(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\Big)^{2m}} \\
&= \frac{(\kappa-1)^m}{(\kappa+1)^m} - \frac{2(\sqrt{\kappa}-1)^m(\sqrt{\kappa}+1)^{2m}}{(\sqrt{\kappa}+1)^m[(\sqrt{\kappa}+1)^{2m}+(\sqrt{\kappa}-1)^{2m}]} \\
&= (\kappa-1)^m\Big(\frac{1}{(\kappa+1)^m} - \frac{2}{(\sqrt{\kappa}+1)^{2m}+(\sqrt{\kappa}-1)^{2m}}\Big) \\
&= (\kappa-1)^m\Big(\frac{1}{\kappa^m+\binom{m}{1}\kappa^{m-1}+\binom{m}{2}\kappa^{m-2}+\ldots+1} - \\
&\qquad \frac{1}{\kappa^m+\binom{2m}{2}\kappa^{m-1}+\binom{2m}{4}\kappa^{m-2}+\ldots+1}\Big) \\
&\geq 0,
\end{aligned}
\tag{3.125}
$$

weil gemäß Pascalschem Dreieck und Rekursion gilt

$$
\binom{2m}{2k} = \binom{2m-1}{2k} + \binom{2m-1}{2k-1} = \ldots = \binom{m}{k} + \ldots \geq \binom{m}{k}.
$$

(5) Welche Beziehung gilt nun zwischen den Faktoren η^m und $\hat{\eta}(m)$?
Wünschenswert wäre für den Normalfall die Ungleichung $\hat{\eta}(m) = 2c^m \leq \eta^m$.
Dazu machen wir ebenfalls eine grafische Darstellung.

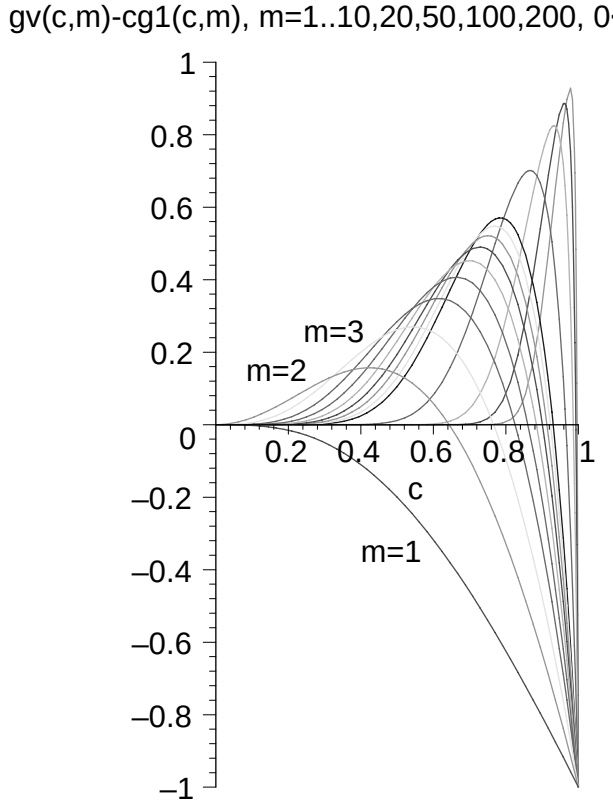


Abb. 3.23 Datei *abst_114.ps*
$gv(c,m) - cg1(c,m) =$
$= \eta^m - \hat{\eta}(m)$
$= (\frac{2c}{1+c^2})^m - 2c^m$, $c \in [0,1]$,
$m = 1(1)10, 20, 50, 100, 200$

Für die Iterationen $m = (0,)1, 2, ...$ und $c \in [0,1]$ ergeben sich folgende Situationen.
– Bei nur einer Iteration ist $\hat{\eta}(m)$ schlechter ($\geq$) als η^m, weil $\frac{2c}{1+c^2} \leq 2c$.
– Wenn nur wenigen Iterationen gemacht werden, dann ist $\hat{\eta}(m)$ besser ($\leq$) als η^m für hinreichend kleine Werte c.
– Bei 2 Iterationen ist $\hat{\eta}(m)$ besser als η^m für $\frac{4c^2}{(1+c^2)^2} \leq 2c^2$,
d. h. $c \leq \sqrt{\sqrt{2}-1} = 0.643...$
– Bei 3 Iterationen ist $\hat{\eta}(m)$ besser als der η^m für $\frac{8c^3}{(1+c^2)^3} \leq 2c^3$,
d. h. $c \leq \sqrt{\sqrt[3]{4}-1} = 0.766...$
– Bei vielen Iterationen kann $\hat{\eta}(m)$ besser werden als der η^m für Werte c nahe 1.
– Wenn c sehr nahe 1 ist, dann muss die Iterationsanzahl m gross werden, damit $\hat{\eta}(m)$ kleiner wird als η^m, d. h.

$$m \approx \frac{1}{1 - \dfrac{\ln(1+c^2)}{\ln(2)}}.$$

– Glücklicherweise hat man aber die theoretische Endlichkeit des CG.

Insgesamt gibt es bezüglich der Konvergenzfaktoren für $\kappa \approx 1$, also $c \approx 0$, nur geringe Unterschiede zwischen GV und CG, mit wachsendem m und κ wird das CG im Vergleich zu GV immer günstiger.

Kann man dieses Verhältnis etwas genauer beschreiben?
Man denkt dabei an den Vergleich der beiden Iterationsverfahren, nämlich des GSV und des Einzelschrittverfahrens (ESV, Gauß-Seidel-Iterationsverfahren), wo unter der Voraussetzung der Konvergenz und $A = A^T > 0$ der Konvergenzfaktor des ESV das Quadrat des vom GSV ist und damit zur Erreichung der gleichen Genauigkeit im ESV nur halb soviel Iterationen gebraucht werden.
Ganz so genau ist ein ähnlicher Vergleich zwischen CG und GV nicht möglich, weil das CG ja theoretisch nur endlich viele Iterationen braucht.
Deshalb machen wir folgende Vereinfachung. Für die praktisch relevanten Anwendungen auf LGS mit $\kappa \gg 1$ und damit $\eta \approx 1$ nehmen wir anstelle des Faktors $\eta^m = (\frac{2c}{1+c^2})^m$ des GV und $\tilde{\eta}(m) = \frac{2c^m}{1+c^{2m}}$ bzw. $\hat{\eta}(m) = 2c^m$ des CG einfach die Größen $(\frac{2c}{1+c^2})^m$ sowie c^m, d. h. $\eta = \frac{2c}{1+c^2} = \frac{\kappa-1}{\kappa+1}$ und $c = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$.
Wir prüfen dann bei $m \gg 1$, wieviel Iterationen k muss man ausführen für

$$\Big(\frac{2c}{1+c^2}\Big)^k \;\approx\; c. \tag{3.126}$$

Die notwendige Iterationsanzahl ist abhängig von c und beträgt

$$k \;=\; k(c) \;=\; \frac{\ln\Big(\frac{1}{c}\Big)}{\ln\Big(\frac{1+c^2}{2c}\Big)} \;=\; \frac{1}{1-\dfrac{\ln(2)-\ln(1+c^2)}{\ln(1/c)}}. \tag{3.127}$$

Für $c = 0$ definieren wir $k(c) = 1$.
Die Funktion $k(c)$ hat leider keinen annähernd konstanten Verlauf, sondern verhält sich in einem Bereich $c \in [0, c_0],\ c_0 < 1$, qualitativ wie $e^{k_0 c/(1-c)}$, also stark wachsend.

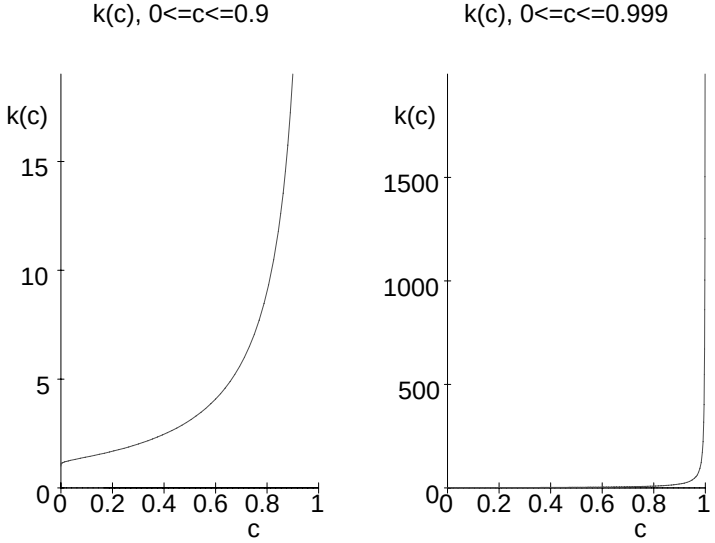


Abb. 3.24 Datei *abst_115.ps*
Funktion der Iterationsanzahl $k(c)$ für die Intervalle $c \in [0, 0.9]$ bzw. $c \in [0, 0.999]$

Das bedeutet für das GV, dass erst einmal k Iterationsschritte gebraucht werden bis $\eta^k = c$ ist. Wenn man insgesamt nur $m \geq k$ Schritte machen will, ist

$$\eta^m \;=\; (\eta^k)^{m/k} \;=\; c^{m/k},$$

so dass das CG nur $\frac{m}{k}$ Iterationen für die gleiche Fehlerordnung benötigen wird. Dies unterstreicht noch einmal alle Vorzüge des CG und sein deutlich besseres Konvergenzverhalten als die typisch lineare Konvergenz des GV. Ob das CG als Iterationsverfahren mit einer höheren Konvergenzordnung als Eins interpretiert werden kann, ist hier nicht zu erkennen.

Deshalb machen wir noch einen anderen Versuch, der etwas Aufschluss darüber bringen soll. Dazu vereinfachen wir die Fehlerabschätzung des CG

$$\|e^{(m)}\|_A \;\leq\; \tilde{\eta}(m)\,\|e^{(0)}\|_A, \quad \tilde{\eta}(m) = \frac{2}{1+c^{2m}}c^m,$$

zu einem Ansatz, in dem die Konvergenzordnung $q > 1$ enthalten ist.

$$\|e^{(m)}\|_A \;\leq\; d\,\|e^{(m-1)}\|_A^q, \quad d \approx c < 1.$$

Damit gilt

$$\|e^{(m)}\|_A \;\leq\; d^{(q^m-1)/(q-1)}\,\|e^{(0)}\|_A^{q^m}.$$

Nehmen wir an, dass $\|e^{(0)}\|_A \approx 1$ ist und wir bei gegebener Dimension n des Problems nach n Schritten auf den Fehler Null bzw. fast Null kommen wollen. Dann fordern wir einfach die Bedingung

$$d^{(q^n-1)/(q-1)} \;=\; 10^{-k} \;\approx\; 0, \quad k \gg 1.$$

Daraus ergibt sich in Abhängigkeit von q und $d \in (0,1)$ ein Polynom n-ten Grades

$$p_n(q,d) \;=\; q^n - aq + a - 1, \quad a = \frac{k}{\log_{10}(d^{-1})}.$$

Seine Nullstelle $q = 1$ ist nicht relevant. Die andere $q > 1$ ist in Abhängigkeit von der Wahl des vereinfachten Konvergenzfaktors d die gesuchte Konvergenzordnung. Wir werten das Polynom $p_n(q,d)$ für den Fall $n = 10,\ k = 16$ (wie Mantissenlänge im Gleitpunktformat *double*) und $d = 0.1(0.1)0.9$ aus.

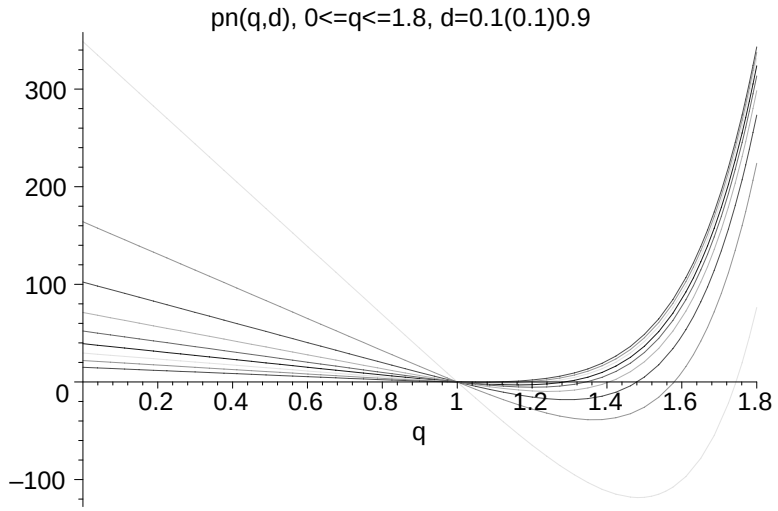


Abb. 3.25 Datei *abst_106.ps*

Polynome $p_n(q,d) = q^n - \dfrac{k}{\log_{10}(d^{-1})}(q-1) - 1,\ q \in [0,1.8],\ d = 0.1(0.1)0.9$
und ihre Nullstellen $q_1^* = 1$ und $q_2^* > 1$

d	Nullstelle q>1
0.1	1.100817
0.2	1.174475
0.3	1.233236
0.4	1.288256
0.5	1.344590
0.6	1.406642
0.7	1.480663
0.8	1.579561
0.9	1.744750

Es ist also im CG durch unsere Vorgehensweise keine einheitliche Konvergenzordnung größer als Eins zu finden.

3.6.5 Modellproblem mit Vergleich von Abstiegsverfahren

Als Modellproblem wird eine einfache eindimensionale Zweipunktrandwertaufgabe mit inhomogenen Randbedingungen gewählt. Das zur numerischen Behandlung verwendete Diskretisierungsverfahren führt auf ein LGS, das mit AV gelöst werden soll. Insbesondere sind dabei von Interesse die Fragen der Konvergenz im Zusammenhang mit dem Spektrum der spd Systemmatrix und ein kurzer Vergleich mit Iterationsverfahren.

- Zweipunktrandwertaufgabe mit inhomogenen Randbedingungen:

$$\begin{aligned} -U''(x) &= F(x), \quad x \in \Omega = (0,1) \subset \mathbb{R}, \\ U &= \varphi \text{ für } x \in \partial\Omega \quad \text{bzw. } U(0) = \varphi_0, \quad U(1) = \varphi_1. \end{aligned}$$

- Gitter: $\Omega_h = \{x \mid x = ih,\ i = 1(1)N,\ h = 1/N\}$, h Maschenweite.
- Gitterfunktion: $u_h = (u_1, u_2, ..., u_{N-1})^T$ mit $u_i \approx U_i = U(ih)$.
- Analog für rechte Seite: $f_h = (f_1, f_2, ..., f_{N-1})^T$ mit $f_i = F_i = F(ih)$, d. h. auf dem Gitter wird die rechte Seite exakt dargestellt.
- Approximation der Ableitungen (Operatoren) mittels Differenzenausdrücken:

$$U_{xx} \sim \frac{1}{h^2}(U_{i+1} - 2U_i + U_{i-1}) \text{ zentraler Differenzenquotient 2. Ordnung.}$$

- Diskretisierte Aufgabe als LGS:

$$\begin{aligned} -\frac{1}{h^2}(u_{i+1} - 2u_i + u_{i-1}) &= f_i, \quad i = 1, 2, ..., N-1, \\ u_0 = \varphi_0, \quad u_N &= \varphi_1. \end{aligned}$$

- Matrixschreibweise des LGS:

$$A_h u_h = b_h \quad \text{bzw.} \quad Au = b$$

mit

$$A_h = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}, \quad b_h = \begin{pmatrix} f_1 + \varphi_0/h^2 \\ f_2 \\ f_3 \\ \cdots \\ f_{N-2} \\ f_{N-1} + \varphi_1/h^2 \end{pmatrix},$$

$$A = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}, \quad b = h^2 \begin{pmatrix} f_1 + \varphi_0/h^2 \\ f_2 \\ f_3 \\ \cdots \\ f_{N-2} \\ f_{N-1} + \varphi_1/h^2 \end{pmatrix},$$

$A = \text{tridiag}(-1, 2, -1)$.

EW und EV der speziellen Tridiagonalmatrix

Für Tridiagonalmatrizen mit jeweils gleichen Elementen auf den Diagonalen und Nebendiagonalen kann man das Eigenwertproblem direkt lösen. Wir fassen die Ergebnisse zusammen.

EW und EV der Systemmatrix $A(n,n),\ n = N-1$

$A = \text{tridiag}(-1, 2, -1)$,

EW $\lambda_i = 2[1-\cos(i\pi/(n+1))] = 4\sin^2(i\pi/(2(n+1))),\quad i = 1,2,...,n,$

EV $v^{(i)} = (v_1^{(i)}, v_2^{(i)}, ..., v_n^{(i)})^T \quad \text{mit} \quad v_j^{(i)} = \sin(ij\pi/(n+1)).$

Wegen $h = 1/(n+1),\ \cos(x) = 1-2\sin^2(x/2),$ gelten die Beziehungen

$\lambda_i = 4\sin^2(i\pi h/2),$

$v^{(i)},\ v_j^{(i)} = \sin(ij\pi h),$

$0 < 2(1-\cos(\pi h)) = 4\sin^2(\pi h/2) = \lambda_{min} = \lambda_1 < \lambda_2 < .... < \lambda_n = \lambda_{max} =$

$= 4\sin^2(n\pi h/2) = 4[1-\sin^2(\pi h/2)] = 4\cos^2(\pi h/2) = 4 - \lambda_{min} < 4,$

$\lambda_1 \approx \pi^2 h^2, \qquad \lim_{h\to 0}\ \lambda_1 = 0,$

$\lambda_n \approx 4 - \pi^2 h^2, \quad \lim_{h\to 0}\ \lambda_n = 4, \quad \lambda_{\frac{n+1}{2}} = 2, \quad \text{falls } n \text{ ungerade.}$

Damit ist $A = A^T > 0$ mit dem Spektrum $\sigma(A) \in (0,4)$.

Vergleich der Eigenschaften der AV

Die Kenntnis der Eigenwerte der Matrix A ist ein gute Ausgangsbasis für Abschätzungen sowohl der Kondition der Koeffizientenmatrix A als auch der Eigenwerte und des Spektrums der Iterationsmatrizen vom GSV und ESV.

EW: $\lambda = \lambda(A) \in (0,4)$,

$\lambda_1 = \lambda_{min} = 4\sin^2(\pi h/2),\quad \lambda_n = \lambda_{max} = 4 - \lambda_{min} = 4\cos^2(\pi h/2).$

Betrachtet man den Spektralradius der Iterationsmatrix des GSV $\rho_{GSV} \approx 1 - \frac{\pi^2 h^2}{2}$ und des ESV $\rho_{ESV} \approx 1 - \pi^2 h^2$, so ist wegen $\rho_{GSV}^2 \approx \rho_{ESV}$ genau die Situation gegeben, dass bei gleicher Genauigkeit das ESV ungefähr die Hälfte der Iterationsschritte des GSV braucht.

Wie stehen hier das GV und CG zueinander?

Betrachtet man in der nachfolgenden Tabelle den Spektralradius der Iterationsmatrix des GV $1 - \frac{\pi^2 h^2}{2}$ und des CG $1 - \pi h$, so braucht man hier ungefähr n Schritte des GV für einen Schritt des CG.

Für diese Modellgleichung sind die Vorteile von CG gegenüber den anderen Iterationsverfahren deutlich erkennbar.

AV,IV	Iterationsmatrix H	EW $\mu = \mu(H)$ bzw. κ, η, c	$\varrho(H)$	
			$\max \lvert\mu(H)\rvert$	$\approx$
GV	$x^{(m+1)} = x^{(m)} + \alpha_m r^{(m)}$ $r^{(m+1)} = b - Ax^{(m+1)}$ $H_{GV} = I - \alpha_m A$	$1 - \alpha_m \lambda_i$ $\eta^m = \left(\frac{\kappa - 1}{\kappa + 1}\right)^m$ $\kappa = \operatorname{cond}(A)$ $= \frac{\lambda_{max}}{\lambda_{min}}$	$\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}$ $= \cos(\pi h)$ $= 1 - 2\sin^2(\frac{\pi h}{2})$	$1 - \frac{\pi^2 h^2}{2}$
CG	$x^{(m+1)} = x^{(m)} + \alpha_m p^{(m)}$ $p^{(m+1)} = r^{(m+1)} + \beta_m p^{(m)}$ $H_{CG} = I - \alpha_m W^{-1} A$	$c = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}$ $\tilde{\eta} = \frac{2c^m}{1 + c^{2m}}$ $= \frac{2}{1 + c^{2m}} c^m$ $\approx c^m$	$c = \frac{\kappa - 1}{\kappa + 1 + 2\sqrt{\kappa}}$ $= \frac{1 - \lambda_{min}/2}{1 + \sqrt{\lambda_{min} - \lambda_{min}^2/4}}$ $\approx \frac{1 - \lambda_{min}/2}{1 + \sqrt{\lambda_{min}}}$ $= \frac{1 - 2\sin^2(\pi h/2)}{1 + 2\sin(\pi h/2)}$ $\approx \frac{1 - \pi^2 h^2/2}{1 + \pi h}$	$1 - \pi h$
RF	$H_{RF} = I - \omega A$ $\omega_{opt} = \frac{2}{\lambda_{min} + \lambda_{max}} \approx \frac{1}{2}$	$1 - \omega \lambda_i$	$\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}$ $= \cos(\pi h)$	$1 - \frac{\pi^2 h^2}{2}$

Tab. 3.2 Abstiegsverfahren GV und CG sowie RF mit Iterationsmatrix H, EW $\mu(H)$, Spektralradius $\rho(H)$ ($\approx$ heißt gerundet) sowie EW $\lambda_i = \lambda_i(A)$

Die Fehlerabschätzungen zum CG selbst kann man benutzen, um Schranken für die Anzahl benötigter Iterationsschritte zu finden, unabhängig von seiner theoretischen Endlichkeit.
Mit der Abschätzung (3.120) wird wegen $\ln((1+x)/(1-x)) \approx 2x$ für kleines $|x|$ bei

$$m \;\geq\; \frac{1}{2}\sqrt{\kappa}\,\ln\left(\frac{2}{\varepsilon}\right) \tag{3.128}$$

Iterationen der relative Approximationsfehler (bez. der Norm $\|\cdot\|_A$) gegenüber dem Startfehler um mindestens den Faktor ε verringert, d. h. $\|e^{(m)}\|_A \leq \varepsilon\,\|e^{(0)}\|_A$.
Bezüglich des Einflusses der Kondition auf die Iterationsanzahl erhält man also ähnliche Aussagen wie beim GV im Abschnitt 3.2.1.

3.6.6 Beispiele zum Verfahren der konjugierten Gradienten

In den Beispielen, die wir zum GV im Abschnitt 3.2.3 betrachtet haben, illustrieren wir nun das CG für das LGS (1.1) unter der üblichen Voraussetzung $A = A^T > 0$.
Einführend machen wir einen Vergleich zwischen CG und GV.

Beispiel 3.10
Gegeben sei das LGS mit der spd Tridiagonalmatrix

$$A = A(7,7) = \text{tridiag}(-1, 2, -1) \quad \text{und} \quad b = (2, -7, 11, -13, 8, 2, 5)^T.$$

Die exakte Lösung des LGS ist $x^* = (1, 0, 6, 1, 9, 9, 7)^T$.
Für die Funktionale $Q(x)$ und $R(x)$ gilt $Q(x^*) = -90,\ R(x^*) = -218$.
Der Startvektor sei $x^{(0)} = 0$.

Bei exakter Rechnung mit dem CG werden genau $n = 7$ Schritte bis zu x^* ausgeführt, wobei im letzten Schritt die entscheidende Fehlerverbesserung von $x^{(6)} = (0.126, -1.141, 5.043, 0.541, 8.232, 8.540, 6.978)^T$ zu $x^{(7)} = x^*$ auftritt (Angaben in $x^{(6)}$ auf 3 Stellen nach dem Komma abgeschnitten).
Im Iterationsverlauf bilden wie erwartet die Funktionswerte $Q(x^{(m)})$ sowie die Normwerte $\|e^{(m)}\|_A = \|x^* - x^{(m)}\|_A$ und $\|e^{(m)}\|_2 = \|x^* - x^{(m)}\|_2$ monoton abnehmende Folgen. Hier gilt dies auch für die Norm $\|r^{(m)}\|_2$ der Residua.
Wenn bei numerischer Rechnung das CG auf Grund von Rundungsfehlern nicht mit dem n-ten Schritt endet, werden jedoch die weiteren Iterierten in der Nähe von $x^{(n)}$ bleiben, vorausgesetzt die Kondition der Matrix A ist moderat. Bei schlechter Kondition kann der Iterationsverlauf sich auch etwas "wegbewegen", insbesondere bei schwacher Gleitpunktarithmetik.
Die folgende Rechnung ist mit Maple bei `Digits:=16` gemacht worden.

```
CG - Iterationsverlauf mit verschiedenen Fehlern
```

m	[x(m)[1],x(m)[2],x(m)[3],x(m)[4],x(m)[5],x(m)[6],x(m)[7]]	Q(x(m))	|e(m)|_A	|e(m)|_2	|r(m)|_2
0	[0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000]	0.000	13.416	15.780	20.881
1	[0.583, -2.040, 3.206, -3.789, 2.332, 0.583, 1.457]	-63.535	7.275	13.458	5.681
2	[-0.393, -1.716, 2.812, -4.571, 2.999, 4.986, 4.261]	-78.425	4.811	10.281	3.949
3	[-0.015, -2.379, 2.055, -3.526, 4.873, 6.066, 6.247]	-83.707	3.548	8.303	2.395
4	[-0.144, -2.884, 2.571, -2.135, 6.503, 7.482, 5.933]	-86.287	2.725	6.395	1.838
5	[-0.696, -2.183, 3.527, -1.117, 7.646, 7.806, 6.266]	-87.658	2.164	4.695	1.618
6	[0.126, -1.142, 5.403, 0.542, 8.233, 8.540, 6.978]	-89.233	1.239	1.853	1.402
7	[1.000, -0.000, 6.000, 1.000, 9.000, 9.000, 7.000]	-90.000	0.000	0.000	0.000
8	[1.000, -0.000, 6.000, 1.000, 9.000, 9.000, 7.000]	-90.000	0.000	0.000	0.000
9	[1.000, -0.000, 6.000, 1.000, 9.000, 9.000, 7.000]	-90.000	0.000	0.000	0.000
10	[1.000, -0.000, 6.000, 1.000, 9.000, 9.000, 7.000]	-90.000	0.000	0.000	0.000

Die Iterierten $x^{(7)} \approx x^*, x^{(8)}, x^{(9)}, x^{(10)}$ liegen nahe beieinander.

$x^{(7)}$	$x^{(8)}$	$x^{(10)}$	x^*
9.9999999999999540e−01	9.9999999999999400e−01	9.9999999999999510e−01	1
−1.5000000000000000e−14	−1.2294871794871580e−14	−1.1606303132920410e−14	0
5.9999999999999840e+00	5.9999999999999810e+00	5.9999999999999810e+00	6
9.9999999999998120e−01	9.9999999999998660e−01	9.9999999999998730e−01	1
8.9999999999999900e+00	8.9999999999999860e+00	8.9999999999999910e+00	9
8.9999999999999850e+00	8.9999999999999890e+00	8.9999999999999940e+00	9
6.9999999999999940e+00	6.9999999999999920e+00	6.9999999999999950e+00	7

Tab. 3.3 Iterierte $x^{(7)}, x^{(8)}, x^{(10)}$ des CG bei fortlaufender Rechnung

Im CG kann man bei mehr als n Iterationen einen Restart machen.
Dies bedeutet, dass die Iterierte $x^{(n)}$ mit $r^{(n)} \neq 0$ als neuer Startvektor $x_r^{(0)}$ genommen wird und dann als nächstes vor einer neuen Schleife das Residuum $r_r^{(0)} = b - Ax_r^{(0)}$ ermittelt und die Anfangssuchrichtung $p_r^{(0)} = r_r^{(0)}$ definiert wird. Damit werden die erste und auch weitere Iterierte nach dem Restart geringfügig von den ursprünglichen Iterationsvektoren bei fortlaufender Rechnung (vergl. Tabelle 3.3) abweichen.

$x_r^{(0)} = x^{(7)}$	$x_r^{(1)}$	$x_r^{(3)}$	x^*
9.9999999999999540e−01	9.9999999999999370e−01	9.9999999999999490e−01	1
−1.5000000000000000e−14	−1.2398866608544030e−14	−7.0154479668353690e−15	0
5.9999999999999840e+00	5.9999999999999840e+00	5.9999999999999950e+00	6
9.9999999999998120e−01	9.9999999999998410e−01	9.9999999999998580e−01	1
8.9999999999999900e+00	8.9999999999999860e+00	8.9999999999999880e+00	9
8.9999999999999850e+00	8.9999999999999890e+00	8.9999999999999900e+00	9
6.9999999999999940e+00	6.9999999999999930e+00	6.9999999999999960e+00	7

Tab. 3.4 Iterierte $x_r^{(0)} = x^{(7)}, x_r^{(1)}, x_r^{(3)}$ des CG bei Restart nach n-tem Schritt

Interessanter sind Betrachtungen zu den Konvergenzfaktoren.
Dazu benötigen wir die Matrixeigenschaften bezüglich Eigenwerte, Kondition usw., also die Größen

$$\begin{aligned}
\lambda &= 0.152\,240..., \ \Lambda = 3.847\,759..., \ \kappa = \kappa(A) = \frac{\Lambda}{\lambda} = 25.274..., \\
c &= \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} = 0.668..., \quad \eta = \frac{\kappa-1}{\kappa+1} = \frac{2c}{1+c^2} = 0.923..., \quad \eta^n = 0.574..., \\
\tilde{\eta}(m) &= \frac{2c^m}{1+c^{2m}} = 0.118... \text{ für } m = 7.
\end{aligned}$$

Da die Systemmatrix aus einer Zweipunktrandwertaufgabe wie in Kap. 3.6.5 entsteht, so trifft die dortige Bemerkung zu, dass im GV ungefähr n Schritte gebraucht werden für einen Schritt des CG. Vergleichbar sind nämlich $\eta^n = \eta^7 = 0.574$ und $c = 0.688$.

Wir machen weitere Vergleiche zu den Konvergenzfaktoren aus der Fehlerabschätzung

$$\|e^{(m)}\|_A \;\leq\; \tilde{\eta}(m)\,\|e^{(0)}\|_A, \quad \tilde{\eta}(m) \leq \eta^m,$$

und erkennen, dass dabei die Faktoren $\tilde{\eta}(m)$ =`etas(m)` und η^m =`etaˆm` nur grobe obere Schranken für die berechneten Fehlerquotienten $\frac{\|e^{(m)}\|_A}{\|e^{(m-1)}\|_A}$ bzw. $\frac{\|e^{(m)}\|_A}{\|e^{(0)}\|_A}$ darstellen. Für $m > n$ liefern die Quotienten keine vernünftigen Werte mehr.

```
CG - Untersuchung zu Konvergenzfaktor/-ordnung
```

m	||e(m)||_A / ||e(m-1)||_A	||e(m)||_A / ||e(0)||_A	etas(m)	etaˆm
1	0.542	0.542	0.706	0.924
2	0.661	0.359	0.638	0.854
3	0.737	0.264	0.506	0.789
4	0.768	0.203	0.369	0.729
5	0.794	0.161	0.257	0.673
6	0.572	0.092	0.175	0.622
7	0.000	0.000	0.118	0.575
8	0.809	0.000	0.079	0.531
9	0.980	0.000	0.053	0.490
10	0.971	0.000	0.035	0.453

Das GV konvergiert linear und wesentlich langsamer als das CG.

```
 GV - Iterationsverlauf mit verschiedenen Fehlern
```

m	[x(m)[1],x(m)[2],x(m)[3],x(m)[4],x(m)[5],x(m)[6],x(m)[7]]	Q(x(m))	|e(m)|_A	|e(m)|_2	|r(m)|_2
0	[0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000]	0.000	13.416	15.780	20.881
1	[0.583, -2.040, 3.206, -3.789, 2.332, 0.583, 1.457]	-63.535	7.275	13.458	5.681
2	[-0.319, -1.390, 2.278, -3.703, 2.429, 4.039, 3.452]	-75.598	5.367	10.955	5.095
3	[0.116, -2.179, 2.749, -4.012, 3.642, 3.971, 4.197]	-80.125	4.444	10.152	2.662
4	[-0.176, -2.020, 2.259, -3.005, 4.122, 5.321, 4.607]	-82.646	3.835	8.863	2.984
5	[-0.056, -2.336, 2.784, -3.225, 4.869, 5.352, 5.006]	-84.251	3.391	8.313	1.755
6	[-0.222, -2.040, 2.688, -2.408, 5.158, 6.219, 5.258]	-85.391	3.036	7.333	2.144
7	[-0.074, -2.206, 3.120, -2.532, 5.706, 6.211, 5.516]	-86.234	2.744	6.901	1.323
8	[-0.118, -1.859, 3.137, -1.857, 5.908, 6.818, 5.652]	-86.898	2.491	6.099	1.696
9	[0.021, -1.956, 3.510, -1.946, 6.331, 6.790, 5.842]	-87.429	2.267	5.745	1.067
10	[0.022, -1.616, 3.569, -1.385, 6.470, 7.244, 5.923]	-87.865	2.067	5.080	1.392
20	[0.588, -0.677, 5.004, 0.041, 8.003, 8.321, 6.587]	-89.656	0.829	2.041	0.556
30	[0.834, -0.272, 5.600, 0.615, 8.600, 8.728, 6.834]	-89.945	0.333	0.820	0.223
40	[0.933, -0.109, 5.839, 0.845, 8.839, 8.891, 6.933]	-89.991	0.134	0.329	0.090
50	[0.973, -0.044, 5.935, 0.938, 8.935, 8.956, 6.973]	-89.999	0.054	0.132	0.036
60	[0.989, -0.018, 5.974, 0.975, 8.974, 8.982, 6.989]	-90.000	0.022	0.053	0.014
70	[0.996, -0.007, 5.990, 0.990, 8.990, 8.993, 6.996]	-90.000	0.009	0.021	0.006
80	[0.998, -0.003, 5.996, 0.996, 8.996, 8.997, 6.998]	-90.000	0.003	0.009	0.002
90	[0.999, -0.001, 5.998, 0.998, 8.998, 8.999, 6.999]	-90.000	0.001	0.003	0.001
100	[1.000, -0.000, 5.999, 0.999, 8.999, 9.000, 7.000]	-90.000	0.001	0.001	0.000

Betrachtet man im GV den Verlauf der Werte des Quotienten

$$q_m \;=\; \frac{\|e^{(m)}\|_A}{\|e^{(m-1)}\|_A},\; m = 1, 2, ...,$$

so sieht man sehr gut die lineare Konvergenz. Es ist

```
q(1)  q(2)  q(3)  q(4)  q(5)  q(6)  q(7)  q(8)  q(9)  q(10) q(11) q(12) q(13)
0.542 0.738 0.828 0.863 0.884 0.895 0.904 0.908 0.910 0.911 0.912 0.912 0.913
```

Beispiel 3.11
Sei

$$A(n,n) \;=\; \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 1 \\ -1 & 4 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 4 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 4 & -1 \\ 1 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}, \quad b \;=\; \begin{pmatrix} n \\ 4 \\ 6 \\ \cdots \\ 2(n-1) \\ n+2 \end{pmatrix}.$$

Damit ist $x^* = (1, 2, ..., n)^T$.

$n = 10$	$n = 100$
$\lambda_{min} = 0.666787139$	$\lambda_{min} = 0.666666666666658$
$\lambda_{max} = 5.891556604$	$\lambda_{max} = 5.999003177628156$
$\text{cond}_2(A) = 8.835738213$	$\text{cond}_2(A) = 8.998504766442351$
$\text{cond}_\infty(A) = 11.16494845$	$\text{cond}_\infty(A) = 11.19615242270663$

Tab. 3.5
Charakteristika der Matrizen $A(n,n)$

Wir berechnen den Fehler $\|e^{(m)}\|_\infty = \|x^* - x^{(m)}\|_\infty$ und vergleichen diesen zwischen GV und dem besseren CG für ausgewählte Schritte m (Rechnung mit 16 Mantissenstellen). Der Startvektor für GV und CG ist der Nullvektor.

```
                n=10                          n=100
  m      GV            CG            GV            CG
-------------------------------------------------------
  0   1.000e+01     1.000e+01     1.000e+02     1.000e+02
  1   4.202e+00     4.202e+00     4.936e+01     4.936e+01
  2   2.946e+00     2.894e+00     3.518e+01     3.502e+01
  3   2.223e+00     1.375e+00     2.670e+01     1.633e+01
  4   1.714e+00     3.142e-01     2.072e+01     3.531e+00
  5   1.332e+00     8.785e-02     1.606e+01     1.019e+00
  6   1.053e+00     1.100e-14     1.269e+01     2.753e-01
  7   8.221e-01     1.100e-14     9.837e+00     7.385e-02
  8   6.583e-01                   7.850e+00     1.980e-02
  9   5.130e-01                   6.079e+00     5.309e-03
```

10	4.134e-01	4.880e+00	1.424e-03
20	4.043e-02	4.683e-01	2.729e-09
30	3.949e-03	4.599e-02	3.000e-13
40	3.858e-04	4.558e-03	
50	3.769e-05	4.541e-04	
60	3.681e-06	4.539e-05	
70	3.596e-07	4.547e-06	
80	3.513e-08	4.562e-07	
90	3.432e-09	4.582e-08	
100	3.353e-10	4.607e-09	

Die weiteren Beispielen werden wir nicht in jedem Fall so detailliert beschreiben.

Beispiel 3.12

Sei A eine Diagonalmatrix und $b = 0$.

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad x^{(0)} = \begin{pmatrix} 9/2 \\ 3 \end{pmatrix}.$$

Die Funktionale sind $Q(x) = \frac{1}{2}x_1^2 + x_2^2$, $R(x) = \frac{1}{2}x_1^2 + 2x_2^2$.
Am gemeinsamen eindeutigen Minimum an der Stelle $x^* = 0$ gilt $Q(x^*) = R(x^*) = 0$.

Als Suchrichtung und Abstiegsrichtung in einem Schritt nehmen wir eine Richtung $p(x)$, die A-orthogonal zu den bisherigen ist.
Im CG werden (theoretisch) endlich viele Schritte ausgeführt.

Die Schritte des CG sind wie folgt.

$x^{(0)} = (\frac{9}{2}, 3)^T$	Startvektor,
$r^{(0)} = b - Ax^{(0)} = (-\frac{9}{2}, -6)^T$	Anfangsresiduum,
$p^{(0)} = r^{(0)}$	Abstiegsrichtung.

S1

$$\begin{aligned}
\alpha_0 &= \frac{\|r^{(0)}\|_2^2}{\|p^{(0)}\|_A^2} = \frac{\frac{225}{4}}{\frac{369}{4}} = \frac{25}{41} = 0.609..., \\
x^{(1)} &= x^{(0)} + \alpha_0 p^{(0)} = \frac{9}{41}(8, -3)^T, \\
r^{(1)} &= b - Ax^{(1)} = r^{(0)} - \alpha_0 A p^{(0)} = \frac{27}{41}(-4, 3)^T, \\
\beta_0 &= \frac{\|r^{(1)}\|_2^2}{\|r^{(0)}\|_2^2} = \frac{\frac{8\,100}{1\,681}}{\frac{225}{4}} = \frac{144}{1\,681} = 0.085..., \quad p^{(1)} = r^{(1)} + \beta_0 p^{(0)} = \frac{450}{1\,681}(-8, 3)^T.
\end{aligned}$$

S2

$$\begin{aligned}
\alpha_1 &= \frac{\|r^{(1)}\|_2^2}{\|p^{(1)}\|_A^2} = \frac{\frac{8\,100}{1\,681}}{\frac{405\,000}{68\,921}} = \frac{41}{50} = 0.82, \\
x^{(2)} &= x^{(1)} + \alpha_1 p^{(1)} = (0, 0)^T, \\
r^{(2)} &= b - Ax^{(2)} = (0, 0)^T \text{ und Stopp mit } x^{(2)} = x^*.
\end{aligned}$$

Der Iterationsverlauf ist wie eine “optimale“ “Zick-Zack“-Kurve.

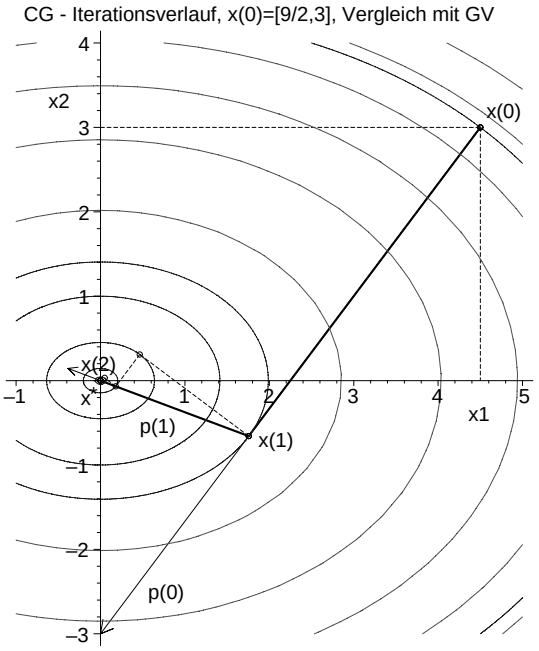


Abb. 3.26 Datei *abst*201.*ps*
Höhenlinienbild mit Iterationsverlauf des CG im Vergleich mit GV zu $Q(x) = \frac{1}{2}x_1^2 + x_2^2$ mit `contours=[19.125,1.9756, 1,0.2041,0.02108,0.0022]` und `contours=6`

Wir erhalten die Beziehungen

$$\begin{aligned}
K_1 &= \mathcal{K}_1(A, r^{(0)}) = \operatorname{span}\{r^{(0)}\} = \operatorname{span}\{p^{(0)}\},\\
r^{(1)} &\perp K_1,\ r^{(1)} \perp r^{(0)},\ \ (Ap^{(1)}, p^{(0)}) = 0,\\
K_2 &= \mathcal{K}_2(A, r^{(0)}) = \operatorname{span}\{r^{(0)}, Ar^{(0)}\} = \operatorname{span}\{r^{(0)}, r^{(1)}\} = \operatorname{span}\{p^{(0)}, p^{(1)}\} = \mathbb{R}^2,\\
r^{(2)} &\perp K_2,\ r^{(2)} \perp r^{(1)}, r^{(0)} \ \rightarrow\ r^{(2)} = 0,\ x^{(2)} = x^*.
\end{aligned}$$

Dies sind auch die Ergebnisse aus Berechnungen mit Maple in der Rationalarithmetik. Bei numerischen Rechnungen in Maple (`Digits:=16`) taucht in der Iterierten $x^{(2)}$ ein Fehler in der Größenordnung der Mantissengenauigkeit auf. Ist die Abbruchbedingung moderat, so endet das CG. Nimmt man jedoch eine extrem kleine Toleranz, z. B. $\varepsilon = 10^{-40}$, so wird der Iterationsprozess fortgesetzt.
Aber nach der n-ten Iteration ist auch ein Restart möglich, wobei sich die nächsten Iterierten vom ursprünglichen Iterationslauf ganz wenig unterscheiden werden.

Berechnungen mit Maple

```
Urspruengliche Iteration mit sehr kleiner Toleranz
x(0): [ 4.5,                    3                    ]
x(1): [ 1.756097560975610,     -0.658536585365854    ]
x(2): [ 0,                     -0.8e-15              ] r(2)=[0,0.1e-14 ]
x(3): [-0.2222222222222223e-30,-0.3000000000000000e-15] r(3)=[0.2e-30,0 ]
-------------------------------------------------------------------------
Restart mit x(2) --> x(0)
x(0): [ 0,                     -0.8e-15              ] r(0)=[0,0.16e-14]
x(1): [ 0,                      0                    ] r(1)=[0,0       ]
```

Beispiel 3.13
Sei A eine Tridiagonalmatrix.

$$\begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ 2 \end{pmatrix}, \quad x^* = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad x^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Die Funktionale sind

$$\begin{aligned} Q(x) &= 2x_1^2 - x_1x_2 + 2x_2^2 - x_2x_3 + +2x_3^2 - 2x_1 - 6x_2 - 2x_3, \\ R(x) &= \frac{17}{2}x_1^2 - 8x_1x_2 + x_1x_3 + 9x_2^2 - 8x_2x_3 + \frac{17}{2}x_3^2 - 2x_1 - 20x_2 - 2x_3. \end{aligned}$$

Am gemeinsamen eindeutigen Minimum an der Stelle x^* gilt $Q(x^*) = -8$ und $R(x^*) = -22$.

Die Schritte des CG sind wie folgt.

$x^{(0)} = (0,0,0)^T$	Startvektor,
$r^{(0)} = b - Ax^{(0)} = b = 2(1,3,1)^T$	Anfangsresiduum,
$p^{(0)} = r^{(0)}$	Abstiegsrichtung.

S1
$$\begin{aligned} \alpha_0 &= \frac{11}{32} = 0.343\,75, \\ x^{(1)} &= \frac{11}{16}(1,3,1)^T, \quad r^{(1)} = \frac{7}{16}(3,-2,3)^T, \\ \beta_0 &= \frac{49}{512} = 0.095\,703..., \quad p^{(1)} = \frac{77}{256}(5,-1,5)^T. \end{aligned}$$

S2
$$\begin{aligned} \alpha_1 &= \frac{16}{77} = 0.207\,792..., \\ x^{(2)} &= (1,2,1)^T, \\ r^{(2)} &= 0 \text{ und vorzeitiger Stopp mit } x^{(2)} = x^*, \\ & \beta_1 = 0, \quad p^{(2)} = r^{(2)} = 0. \end{aligned}$$

Bei numerischer Rechnung in Maple mit `Digits:=5,6,...,16,...` werden auch bei beliebig kleiner Toleranz nur 2 Iterationen ausgeführt.
Startvektor $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, x_3^{(0)})^T$ und Iterierte $x^{(k)},\ k = 1,2$

```
k  [ x(k)[1],              x(k)[2],              x(k)[3]              ]
-------------------------------------------------------------------
0  [ 0,                    0,                    0                    ]
1  [ 0.6875000000000000, 2.062500000000000, 0.6875000000000000]
2  [ 1.000000000000000, 2.000000000000000, 1.000000000000000]
```

Rechnet man ungenauer, ist kein vorzeitiges Ende zu erwarten, eher geht die Iterationsanzahl über n hinaus.
Numerische Berechnung in Maple mit `Digits:=4` und Ausgabe der Iterationsvektoren $x^{(i)}, r^{(i)}, p^{(i)}$, $i = 0, 1, ..., m$, als Spalten der folgenden Felder.

$$m = 4$$

$$xv = \begin{bmatrix} 0. & 0.6876 & 1.000 & 0.9997 & 0.9996 \\ 0. & 2.063 & 2.000 & 2.000 & 2.000 \\ 0. & 0.6876 & 1.000 & 0.9997 & 0.9996 \end{bmatrix}$$

$$rv = \begin{bmatrix} 2. & 1.312 & -0.001 & 0. & 0. \\ 6. & -0.876 & 0. & -0.001 & 0. \\ 2. & 1.312 & -0.001 & 0. & 0. \end{bmatrix}$$

$$pv = \begin{bmatrix} 2. & 1.503 & -0.0009993 & -0.0004996 & 0. \\ 6. & -0.3020 & -0.1435\,10^{-6} & -0.00100 & 0. \\ 2. & 1.503 & -0.0009993 & -0.0004996 & 0. \end{bmatrix}$$

Berechnung in Maple mit `Digits:=3`

$$m = 3$$

$$xv = \begin{bmatrix} 0. & 0.688 & 0.998 & 0.998 \\ 0. & 2.06 & 2.00 & 2.00 \\ 0. & 0.688 & 0.998 & 0.998 \end{bmatrix}$$

In beiden Rechnungen mit der ungenauen Gleitpunktarithmetik erhält man für $x^{(m)} \neq x^*$ das Residuum $r^{(m)} = 0$, so dass das Verfahren dann endet.
Zum CG erhalten wir u. a. die Beziehungen

$$r^{(1)} \perp r^{(0)},\ 0 = r^{(2)} \perp r^{(1)}, r^{(0)}, ...,\quad (Ap^{(1)}, p^{(0)}) = 0,$$

$$\mathcal{K}_2(A, r^{(0)}) = \mathrm{span}\{r^{(0)}, Ar^{(0)}\} = \mathrm{span}\left\{\begin{pmatrix}1\\3\\1\end{pmatrix}, \begin{pmatrix}1\\10\\1\end{pmatrix}\right\} = \mathrm{span}\{\tilde{r}^{(0)}, \widetilde{Ar^{(0)}}\}$$

$$= \mathrm{span}\{r^{(0)}, r^{(1)}\} = \mathrm{span}\left\{\begin{pmatrix}1\\3\\1\end{pmatrix}, \begin{pmatrix}3\\-2\\3\end{pmatrix}\right\} = \mathrm{span}\{\tilde{r}^{(0)}, \tilde{r}^{(1)}\}$$

$$= \mathrm{span}\{p^{(0)}, p^{(1)}\} = \mathrm{span}\left\{\begin{pmatrix}1\\3\\1\end{pmatrix}, \begin{pmatrix}5\\-1\\5\end{pmatrix}\right\} = \mathrm{span}\{\tilde{p}^{(0)}, \tilde{p}^{(1)}\} \subset \mathbb{R}^3,$$

$$\widetilde{Ar^{(0)}} = \tfrac{32}{11}\tilde{r}^{(0)} - \tfrac{7}{11}\tilde{r}^{(1)} = \tfrac{51}{16}\tilde{p}^{(0)} - \tfrac{7}{16}\tilde{p}^{(1)},$$

$$\tilde{r}^{(1)} = \tfrac{40}{11}\tilde{r}^{(0)} - \tfrac{7}{11}\widetilde{Ar^{(0)}} = -\tfrac{7}{16}\tilde{p}^{(0)} + \tfrac{11}{16}\tilde{p}^{(1)},$$

$$\tilde{p}^{(1)} = \tfrac{51}{7}\tilde{r}^{(0)} - \tfrac{16}{7}\widetilde{Ar^{(0)}} = \tfrac{7}{11}\tilde{r}^{(0)} + \tfrac{16}{11}\tilde{r}^{(1)},$$

$$x^* = x^{(2)} = x^{(0)} + \mathcal{K}_2(A, r^{(0)}) = \mathcal{K}_2(A, r^{(0)}).$$

Beispiel 3.14
Wir nehmen das LGS aus Beispiel 3.7 mit $A = I$, $b = (0,0,1)^T = x^*$ und dem Startvektor $x^{(0)} = (0,0,0)^T$.
Während wir im AV der konjugierten Richtungen verschiedene Varianten der Auswahl solcher Richtungen testen konnten (siehe Beispiel 3.9), gilt für das CG
$p^{(0)} = r^{(0)} = b = (0,0,1)^T$.
Der Ablauf des Iterationsprozesses des CG ist sehr kurz, denn

$$\begin{aligned}
\alpha_0 &= \frac{\|r^{(0)}\|_2^2}{\|p^{(0)}\|_A^2} = 1, \\
x^{(1)} &= x^{(0)} + \alpha_0 p^{(0)} = (0,0,1)^T, \\
r^{(1)} &= b - Ax^{(1)} = r^{(0)} - \alpha_0 A p^{(0)} = 0 \quad \text{und vorzeitiger Stopp mit } x^{(1)} = x^*, \\
&\quad \beta_0 = 0, \quad p^{(1)} = r^{(1)} = 0.
\end{aligned}$$

Dazu erhalten wir u. a. die Beziehungen

$$\begin{aligned}
p^{(1)} &= 0, \\
Ar^{(0)} &= r^{(0)} = (0,0,1)^T, \\
\mathcal{K}_2(A, r^{(0)}) &= \mathcal{K}_1(A, r^{(0)}).
\end{aligned}$$

Beispiel 3.15
Sei $A = A^T > 0$.

$$\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad x^* = \begin{pmatrix} 1/5 \\ 3/5 \end{pmatrix}, \quad x^{(0)} = \begin{pmatrix} 3/2 \\ 1 \end{pmatrix}.$$

Die Funktionale sind

$$Q(x) = x_1^2 + x_1 x_2 + \frac{3}{2}x_2^2 - x_1 - 2x_2, \quad R(x) = \frac{5}{2}x_1^2 + 5x_1 x_2 + 5x_2^2 - 4x_1 - 7x_2.$$

Am gemeinsamen eindeutigen Minimum an der Stelle x^* gilt $Q(x^*) = -\frac{7}{10}$ und $R(x^*) = -\frac{5}{2}$.
Wir testen das CG sowohl für $Ax = b$ als auch $A^T Ax = A^T b$.
Der Vorabschritt im CG ist wie folgt.

$x^{(0)} = (\frac{3}{2}, 1)^T$	Startvektor,
$r^{(0)} = b - Ax^{(0)} = (-3, -\frac{5}{2})^T$	Anfangsresiduum, $Ar^{(0)} = -\frac{1}{2}(17,21)^T$,
$p^{(0)} = r^{(0)}$	Abstiegsrichtung.

Bei exakter Rechnung haben wir zwei Schritte und die Iterationsvektoren $x^{(i)}, r^{(i)}, p^{(i)}$, $i = 0,1,2$, sind die Spalten der folgenden Felder.

$$m = 2$$

$$xv = \begin{bmatrix} \frac{3}{2} & \frac{85}{138} & \frac{1}{5} \\ 1 & \frac{109}{414} & \frac{3}{5} \end{bmatrix}$$

$$rv = \begin{bmatrix} -3 & \frac{-205}{414} & 0 \\ \frac{-5}{2} & \frac{41}{69} & 0 \end{bmatrix}$$

$$pv = \begin{bmatrix} -3 & \frac{-17507}{28566} & 0 \\ \frac{-5}{2} & \frac{42517}{85698} & 0 \end{bmatrix}$$

Bei numerischen Rechnungen in Maple (`Digits:=16`) taucht in der Iterierten $x^{(2)}$ ein Fehler in der Größenordnung der Mantissengenauigkeit auf. Ist die Abbruchbedingung moderat, so endet das CG spätestens beim n-ten Schritt. Nimmt man jedoch eine extrem kleine Toleranz, z. B. $\varepsilon = 10^{-40}$, so wird der Iterationsprozess fortgesetzt. Aber nach der n-ten Iteration ist auch ein Restart möglich, wobei sich die nächsten Iterierten vom ursprünglichen Iterationslauf ganz wenig unterscheiden werden.

Berechnungen mit Maple

```
Urspruengliche Iteration mit sehr kleiner Toleranz
x(0): [ 1.5,                1                   ]
x(1): [ 0.6159420289855071, 0.2632850241545892 ]
x(2): [ 0.2000000000000002, 0.6000000000000003 ] r(2)=[-0.1e-14,-0.1e-14]
x(3): [ 0.1999999999999999, 0.6000000000000000 ] r(3)=[0, 0]
-------------------------------------------------------------------------
Restart mit x(2) --> x(0)
x(0): [ 0.2000000000000002, 0.6000000000000003 ]
x(1): [ 0.1999999999999999  0.6000000000000000 ]
```

Numerische Berechnung in Maple mit `Digits:=16`
Man bemerke die kleine Ungenauigkeit bei $m = 2$ mit $Q(x^{(2)}) < Q(x^*)$.

```
Iterationsverlauf mit verschiedenen Fehlern

 m  [       x(m)[1],          x(m)[2]     ]        Q(x(m))          ||e(m)||_A
                                                                     ||e(m)||_2
                                                                     ||r(m)||_2
 ------------------------------------------------------------------------------
 0  [ 1.5000000000000000, 1.0000000000000000]   1.750000000000000   2.213594362117866
                                                                    1.360147050873544
                                                                    3.905124837953327
 1  [ 0.6159420289855071, 0.2632850241545892]  -0.496980676328503   0.637211618964214
                                                                    0.535149274908559
                                                                    0.773478832638098
 2  [ 0.2000000000000002, 0.6000000000000003]  -0.700000000000001   0.000000000000001
                                                                    0.000000000000000
                                                                    0.000000000000001
 3  [ 0.1999999999999999, 0.6000000000000000]  -0.700000000000000   0.000000000000000
                                                                    0.000000000000000
                                                                    0.000000000000000
```

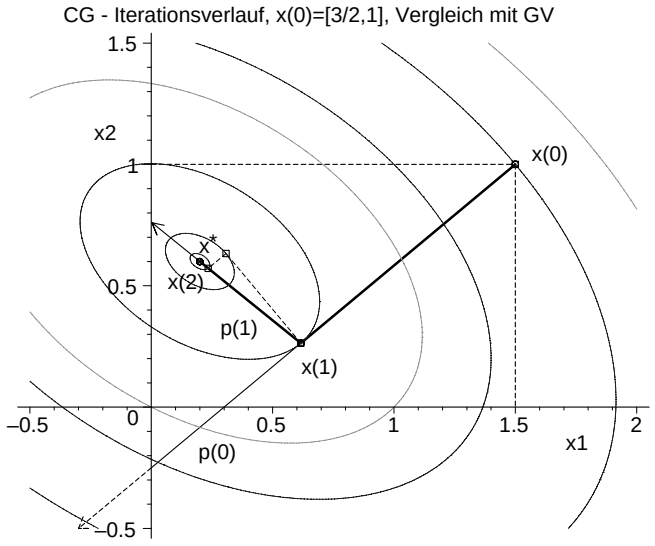


Abb. 3.27 Datei *abst*202.*ps*
Höhenlinienbild mit Iterationsverlauf des CG im Vergleich mit GV zu $Q(x) = x_1^2 + x_1x_2 + \frac{3}{2}x_2^2 - x_1 - 2x_2$ mit `contours=[3,1.75,0.5,0, -0.497,-0.6832,-0.6986,-0.7]`

Das CG für $A^TAx = A^Tb$ ist mit dem Funktional $R(x)$ verknüpft.

Es bringt eine Verschlechterung der Kondition des Systems durch Multiplikation mit A^T, aber für beliebiges reguläres A ist die Matrix $B = A^TA$ spd.
Im Allgemeinen ergeben sich dadurch eventuell mehr Iterationsschritte, aber nicht generell.
Wir betrachten also das LGS

$$Bx = \begin{pmatrix} 5 & 5 \\ 5 & 10 \end{pmatrix} x = \begin{pmatrix} 4 \\ 7 \end{pmatrix} = c, \quad B = B^T > 0.$$

So wie die Folge der Funktionalwerte $R(x^{(m)})$ streng monoton fallend gegen $-\frac{5}{2}$ strebt, so tun dies auch die Residua $r^{(m)} = r(x^{(m)}) = b - Ax^{(m)}$ in der euklidischen Norm sowie wegen (2.18) auch $\|e^{(m)}\|_{A^TA}$, aber gegen Null. Auch die Norm des Residuums $\hat{r}(x) = A^Tr(x) = A^T(b - Ax)$ verkleinert sich stetig.

Bei exakter Rechnung haben wir wieder zwei Schritte und die Iterationsvektoren $x^{(i)}, \hat{r}^{(i)}, p^{(i)}$, $i = 0, 1, 2$, sind die Spalten der folgenden Felder.

$$m = 2$$

$$xv = \begin{bmatrix} \frac{3}{2} & \frac{3173}{3770} & \frac{1}{5} \\ 1 & \frac{352}{1885} & \frac{3}{5} \end{bmatrix}$$

$$rv = \begin{bmatrix} \frac{-17}{2} & \frac{-861}{754} & 0 \\ \frac{-21}{2} & \frac{697}{754} & 0 \end{bmatrix}$$

$$pv = \begin{bmatrix} \frac{-17}{2} & \frac{-176587}{142129} & 0 \\ \frac{-21}{2} & \frac{113734}{142129} & 0 \end{bmatrix}$$

Numerische Berechnung in Maple mit `Digits:=16`

```
Iterationsverlauf mit verschiedenen Fehlern

 m  [       x(m)[1],          x(m)[2]     ]        Q(x(m))          ||e(m)||_A
                                                                     ||e(m)||_2
                                                                     ||rd(m)||_2
 ------------------------------------------------------------------------------
 0  [ 1.5000000000000000, 1.0000000000000000]   +5.125000000000000   3.905124837953327
                                                                     1.360147050873544
                                                                    13.509256086106300
 1  [ 0.8416445623342175, 0.1867374005305040]   -1.942639257294430   1.055803715380440
                                                                     0.763212762271016
                                                                     1.469176391327209
 2  [ 0.1999999999999995, 0.5999999999999990]   -2.500000000000000   0.000000000000004
                                                                     0.000000000000001
                                                                     0.000000000000012
 3  [ 0.2000000000000000, 0.5999999999999998]   -2.500000000000001   0.000000000000001
                                                                     0.000000000000000
                                                                     0.000000000000000
```

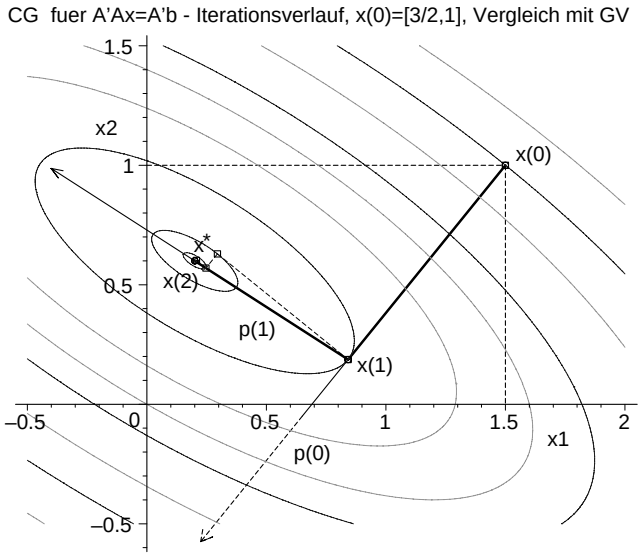


Abb. 3.28 Datei *abst*203.*ps*
Höhenlinienbild mit Iterationsverlauf des CG im Vergleich mit GV zu $R(x) = \frac{5}{2}x_1^2 + 5x_1x_2 + 5x_2^2 - 4x_1 - 7x_2$ mit `contours=[8,5.125,3,1,0, -1,-1.9426,-2.4593, -2.4970,-2.4998,-2.5]`

Beispiel 3.16

Gegeben sei das LGS aus den Beispielen 1.7, 3.6 mit der regulären Matrix A. Wir nehmen hier Bezug auf einige Eigenschaften des Systems.

$$\begin{pmatrix} 0.780 & 0.563 \\ 0.913 & 0.659 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0.217 \\ 0.254 \end{pmatrix}, \quad x^* = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Um das CG zu verwenden, muss man das LGS symmetrisieren.

$$Bx = A^T Ax = \begin{pmatrix} 1.441\,969 & 1.040\,807 \\ 1.040\,807 & 0.751\,250 \end{pmatrix} x = \begin{pmatrix} 0.401\,162 \\ 0.289\,557 \end{pmatrix} = A^T b = c.$$

Die Funktionale sind

$$\begin{aligned} Q(x) &= 0.5(0.780x_1 + 0.913x_2)x_1 + 0.5(0.563x_1 + 0.659x_2)x_2 - 0.217x_1 - 0.254x_2 \\ &= 0.390x_1^2 + 0.738x_1x_2 + 0.3295x_2^2 - 0.217x_1 - 0.254x_2, \\ R(x) &= 0.5(1.441\,969x_1 + 1.040\,807x_2)x_1 + 0.5(1.040\,807x_1 + 0.751\,250x_2)x_2 \\ &\quad -0.401\,162x_1 - 0.289\,557x_2 \\ &= 0.720\,984\,5x_1^2 + 1.040\,807x_1x_2 + 0.375\,625x_2^2 - 0.401\,162x_1 - 0.289\,557x_2. \end{aligned}$$

Das Minimum von $R(x)$ ist bei der Lösung $x^* = (1, -1)^T$ und $R(x^*) = -0.055\,802\,5$ sowie $Q(x^*) = 0.018\,5$.
Aber $Q(x)$ hat die Oberfläche eines Sattels mit dem Sattelpunkt bei

$$z = \Big(\frac{44\,449}{30\,624}, -\frac{6\,329}{5\,104}\Big)^T = (1.451\,443\,312\,434\,691\,7, -1.240\,007\,836\,990\,595\,6)^T,$$

$$Q(z) = -\frac{37}{61\,248\,000} = -0.604\,101\,358\,411\,702\,9\,10^{-6},$$

und keine Minimumstelle.

Die Matrix B hat eine sehr schlechte Kondition, ihre spektrale Konditionszahl ist ungefähr $0.5 \cdot 10^{13}$.
Wir testen den Iterationsverlauf des CG mit ausgewählten Startvektoren

$$x^{(0)} = \begin{pmatrix} 1.2 \\ -1.2 \end{pmatrix}, \begin{pmatrix} 0.999 \\ -1.001 \end{pmatrix}, \begin{pmatrix} 0.341 \\ -0.087 \end{pmatrix}, \begin{pmatrix} 0.991\,891\,566\,446\,068\,04 \\ -1.005\,852\,632\,339\,509\,328 \end{pmatrix}.$$

Der Vektor $(0.341, -0.087)^T$ liegt nahe der Tallinie (1.14) $t(x_1) = \frac{v_2^{(2)}}{v_1^{(2)}}(x_1 - 1) - 1$ von $R(x)$, auf der die Werte $R(x)$ nur minimal größer sind als $R(x^*)$ bzw. die Werte $\|r(x)\|_2$ oder $\|\hat{r}(x)\|_2$, $\hat{r}(x) = A^T r(x) = c - Bx$, ganz nahe bei der Null liegen.
Der Vektor $x^* - (0.991\,891\,566\,446\,068\,04, -1.005\,852\,632\,339\,509\,328)^T$ ist orthogonal zur Tallinie $t(x_1)$.

Welche Startsituation findet man vor?

$x^{(0)}$	$r(x^{(0)})$	$\|r(x^{(0)})\|_2$	$\hat{r}(x^{(0)}) = A^T r(x^{(0)})$	$R(x^{(0)})$
$[1.2, -1.2]$	$[-0.043\,4, -0.050\,8]$	$0.066\,814...$	$[-0.080\,232\,4, -0.057\,911\,4]$	$-0.053\,570\,400$
$[0.999, -1.001]$	$[\ 0.001\,343,\ 0.001\,572]$	$0.002\,067...$	$[0.002\,482\,776,\ 0.001\,792\,057]$	$-0.055\,800\,362\,583\,5$
$[0.341, -0.087]$	$[10^{-6},\ 0]$	10^{-6}	$[0.780 \cdot 10^{-6},\ 0.563 \cdot 10^{-6}]$	$-0.055\,802\,499\,999\,5$
$[0.991\,891..., -1.005\,852...]$	$[0.009\,619...,\ 0.011\,259...]$	$0.014\,809...$	$[0.017\,783...,\ 0.012\,836...]$	$-0.055\,692\,839...$
$x^* = (1, -1)^T$				$R(x^*) = -0.055\,802\,5$

Tab. 3.6 Startvektor $x^{(0)} = (x_1^{(0)}, x_2^{(0)})^T$, dazu Anfangsresidua und Funktionale

Wir werden den Iterationsverlauf und das Verhalten des CG mit dem Startvektor $x^{(0)} = (1.2, -1.2)^T$ etwas ausführlicher besprechen, denn ähnlich Aspekte bemerken wir auch bei den anderen Startvektoren.

Bei exakter Rechnung haben wir zwei Schritte und die Iterationsvektoren $x^{(i)}, \hat{r}^{(i)}, p^{(i)}$, $i = 0, 1, 2$, sind die Spalten der folgenden Felder.

$$m = 2$$

$$xv = \begin{bmatrix} \frac{6}{5} & \frac{1561433266058205653}{1342108600975648405} & 1 \\ \frac{-6}{5} & \frac{-3291936727251049773}{2684217201951296810} & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1.2 & 1.16341797148391625295701... & 1 \\ -1.2 & -1.22640475027802150139476... & -1 \end{bmatrix}$$

$$rv = \begin{bmatrix} \frac{-200581}{2500000} & \frac{-200002521483}{268421720195129681000000} & 0 \\ \frac{-289557}{5000000} & \frac{138545107739}{134210860097564840500000} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -0.0802324 & -0.74510557989721467354928 1\ 10^{-13} & 0 \\ -0.0579114 & 0.10322943138750796315419 0\ 10^{-12} & 0 \end{bmatrix}$$

$$pv = \begin{bmatrix} \frac{-200581}{2500000} & \frac{-838828450936342729913215801}{112578468550800764425340988136190251562 50} & 0 \\ \frac{-289557}{5000000} & \frac{148754064575549669487250588759}{144100439745024978464436464814323522000 0000} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -0.0802324 & -0.74510557989854285462134 0\ 10^{-13} & 0 \\ -0.0579114 & 0.10322943138741209561771 2\ 10^{-12} & 0 \end{bmatrix}$$

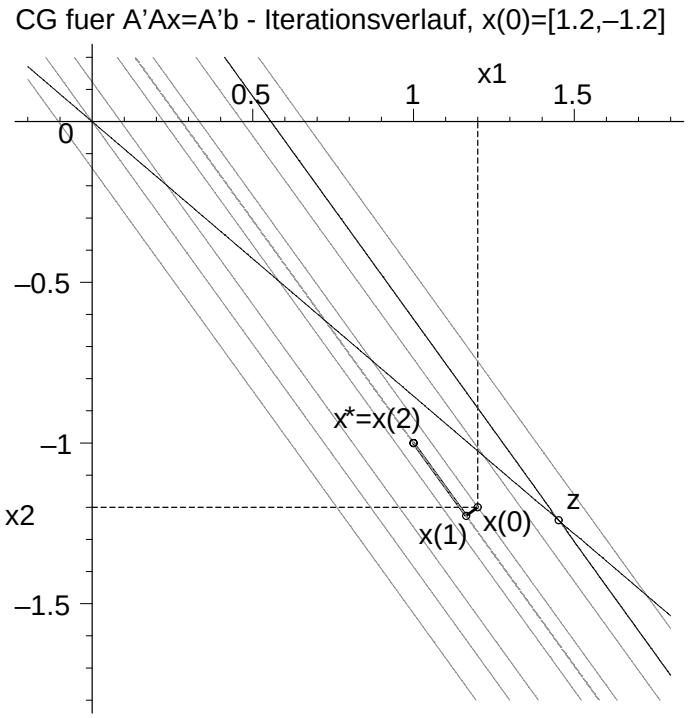


Abb. 3.29 Datei *abst*204.*ps* Höhenlinienbild mit Iterationsverlauf des CG mit $x^{(0)} = (1.2, -1.2)^T$ zu $R(x)$ mit `contours=[0.05,0,-0.03, -0.05357,-0.0558]` sowie zu $Q(x)$ mit `contours=[-0.000000604]`, Sattelpunkt z und Höhenlinien $Q(x) = Q(z)$

Bei numerischen Rechnungen muss das CG auf Grund der schlechten Kondition der Matrix $B = A^T A$ der Größenordnung 10^{12} mit einer entsprechend starken Gleitpunktarithmetik arbeiten, um relevante Ergebnisse zu erzielen. Rechnet man mit t Dezimalstellen in der Mantisse, so ist wegen der Matrixkondition ein Verlust von ungefähr 12 Dezimalstellen, wenn nicht sogar einige mehr, im Laufe der Rechnung zu erwarten. Von den angezeigten t Stellen sind also die letzten 12 und ev. mehr Stellen nicht verwertbar. Wenn die Iterierten des CG in die Nähe von x^* kommen und mit dem Grenzvektor auf ca. $t-12$ übereinstimmen, ist dann keine signifikante Verbesserung mehr durch weitere Iterationen zu erwarten. Im Gegenteil, es können sogar "zwischenzeitlich" Verschlechterungen auftreten. Dadurch wird das Ende des CG nach spätestens n Iterationen bei sehr kleiner Toleranz nicht erreicht. Es kann durch die schwache Gleitpunktarithmetik und die schlechte Matrixkondition der Fall eintreten, dass ein späterer Iterationsvektor $x^{(m)},\ m > n$, noch verschieden von x^*, plötzlich ein Residuum gleich Null hat.

Ergebnisse aus Berechnungen (9 Iterationen) mit Maple mit `Digits:=13`.

```
m  [     x(m)[1],          x(m)[2]     ]  [    rd(m)[1],    rd(m)[2] ]  [    p(m)[1],      p(m)[2]   ]
-------------------------------------------------------------------------------------------------------
0  [ 1.2,            -1.2             ]  [ -0.0802324,  -0.0579114  ]  [ -0.0802324,   -0.0579114   ]
1  [ 1.163417971484, -1.226404750278 ]  [ -0.1e-12,     0.1e-12    ]  [ -0.100...e-12, 0.999...e-13]
2  [ 1.163417971482, -1.226404750276 ]  [  0.6e-12,     0.6e-12    ]  [ -0.300...e-11, 0.419...e-11]
3  [ 1.163417969976, -1.226404748168 ]  [ -0.222e-10,  -0.159e-10  ]  [ -0.312...e-8,  0.433...e-8 ]
4  [ 1.163416386952, -1.226402555701 ]  [  0.7153e-19,  0.5164e-9  ]  [ -0.326...e-5,  0.452...e-5 ]
5  [ 1.161788718964, -1.224147507159 ]  [ -0.228119e-7,-0.164654e-7]  [ -0.332...e-2,  0.460...e-2 ]
6  [ 0.9835323939010,-0.9771853078956]  [  0.862542e-7, 0.622579e-7]  [ -0.047...e-1,  0.657...e-1 ]
7  [ 0.9579886624732,-0.9417960810587]  [  0.1009e-9,   0.728e-10  ]  [ -0.648...e-7,  0.900...e-7 ]
8  [ 0.9579886329638,-0.9417960400785]  [  0.1e-12,     0.1e-12    ]  [  0.162...e-13, 0.216...e-12]
9  [ 0.9579886329638,-0.9417960400784]  [  0.,          0.         ]  [  0.,           0.          ]
```

In Maple, das standardmäßig mit dem Gleitpunktformat `Digits:=10` arbeitet, ist also die Genauigkeit deutlich zu erhöhen.

Ergebnisse aus Berechnungen (erste 3 Iterationen) mit Maple mit `Digits:=24`.

```
Startvektor              x=[+1.2000000000000000e+00 -1.2000000000000000e+00]
Residuum/SR  p = rd = c-Bx=[-8.0232400000000000e-02 -5.7911400000000000e-02]
Funktionswert         Q(x)= -5.3570400000000000e-02
Anfangsfehlerquadrat rd'rd=  9.7909682597200000e-03

Schritt k = 1
Suchrichtung             p=[-8.0232400000000000e-02 -5.7911400000000000e-02]
Suchschritt          alpha= +4.5595081932091957e-01
Iterationsvektor         x=[+1.1634179714839163e+00 -1.2264047502780215e+00]
Residuum         rd = c-Bx=[-7.4510557990000000e-14 +1.0322943138800000e-13]
Funktionswert         R(x)= -5.5802499999982226e-02
Fehlernormquadrat    rd'rd=  1.6208138756670952e-26
Schritt               beta=  1.6554173526791179e-24
neue Suchrichtung        p=[-7.4510557990132818e-14 +1.0322943138790413e-13]
```

```
Schritt k = 2
Suchrichtung             p=[-7.4510557990132818e-14 +1.0322943138790413e-13]
Suchschritt          alpha= +2.1932190000053422e+12
Iterationsvektor         x=[+9.9999999999895709e-01 -9.9999999999832231e-01]
Residuum         rd = c-Bx=[-2.4230673356000000e-13 -1.7489595437700000e-13]
Funktionswert         R(x)= -5.5802500000000000e-02
Fehlernormquadrat    rd'rd=  8.9301147985958496e-26
Schritt               beta=  5.5096485368626236e+00
neue Suchrichtung        p=[-6.5283372037115295e-13 +3.9386193123052659e-13]

Schritt k = 3
Suchrichtung             p=[-6.5283372037115295e-13 +3.9386193123052659e-13]
Suchschritt          alpha= +4.5595081932144182e-01
Iterationsvektor         x=[+9.9999999999865943e-01 -9.9999999999814273e-01]
Residuum         rd = c-Bx=[+0.0000000000000000e-01 +0.0000000000000000e-01]
Funktionswert         R(x)= -5.5802500000000000e-02
Fehlernormquadrat    rd'rd=  0.0000000000000000e-01
```

Die Ergebnisse sind mit 16 Nachkommastellen angezeigt.
Während die Iterierte $x^{(1)}$ noch auf 24 Mantissenstellen genau erhalten wird und ihr Residuum $\hat{r}^{(1)} = c - Bx^{(1)}$ erst durch die Multiplikation mit B die Norm $\|\hat{r}^{(1)}\| \approx 10^{-13}$ bekommt, hat die nächste Iterierte $x^{(2)}$ durch den Matrixeinfluss nur noch ca. 12 Mantissenstellen genau. Das CG macht dadurch einen Schritt mehr als theoretisch notwendig, aber $x^{(3)} \approx x^*$ liefert trotzdem $\hat{r}^{(3)} = 0$.

Eine weitere Rechnung mit Maple macht den Einfluss des Gleitpunktformats mit `Digits:=k` deutlich. Wenn die Größenangabe k mit dem Vielfachen einer Bytelänge korrespondiert, liegt meistens eine günstige Situation vor.
Wir machen maximal 2 Iterationen.

Bei `Digits:=10,11,12` endet das CG mit der Iterierten $x^{(1)}$.

```
k      x(1)                                          ||e(1)||_2=||xs-x(1)||_2   R(x(1))
       rd(1)=c-Bx(1)                                     ||rd(1)||_2
       r(1)=b-Ax(1)                                      ||r(1)||_2

-------------------------------------------------------------------------------------------

10  [ 1.163417971,-1.226404750]                         0.2792213174       -0.05580250030
    [ 0, 0]                                             0
    [-0.1432e-6, 0.1222e-6]                             0.1882527025e-6

11  [ 1.1634179715,-1.2264047503]                       0.27922131789      -0.055802500025
    [ 0, 0]                                             0
    [-0.14335e-6, 0.12245e-6]                           0.188529...e-6

12  [ 1.16341797148,-1.22640475028]                     0.279221317868     -0.0558024999940
    [ 0, 0]                                             0
    [-0.143346e-6, 0.122475e-6]                         0.188542...e-6
```

Mit `Digits:=13,14,...` erkennt man im Iterationsverlauf und Konvergenzverhalten des CG, dass zwischen $x^{(2)}$ und $x^{(3)}$ kaum noch Unterschiede sind.
Die Kondition der Matrix B und die Wahl der Genauigkeit haben deutlichen Einfluss auf die Ergebnisse.

```
k      x(2)                                                ||e(2)||_2=||xs-x(2)||_2   R(x(2))
       rd(2)=c-Bx(2)                                           ||rd(2)||_2
       r(2)=b-Ax(2)                                            ||r(2)||_2

-------------------------------------------------------------------------------------------------------

13 [ 1.163417971482,-1.226404750276]                           0.2792213178660 -0.05580250000010
   [ 0.6e-12, 0.6e-12]                                         0.848528...e-12
   [-0.1433506e-6, 0.1224689e-6]                               0.188541...e-6

14 [ 1.1634179714690,-1.2264047502594]                         0.2792213178449 -0.055802499999935
   [ 0.202e-11, 0.161e-11]                                     0.258311...e-11
   [-0.14334978e-6, 0.12246974e-6]                             0.188541...e-6

15 [ 1.16341796951608,-1.22640474757552]                       0.2792213145257 -0.0558024999999800
   [ 0.24708e-10, 0.17991e-10]                                 0.305640...e-10
   [-0.143337524e-6, 0.122484088e-6]                           0.188541...e-6

16 [ 1.163410641168179,-1.226394596068840]                     0.2792087942314 -0.05580249999998215
   [ 0.15159824e-8, 0.10943865e-8]                             0.186972...e-8
   [-0.1425244227e-6, 0.1234228186e-6]                         0.188537...e-6

17 [ 1.1625297817689348,-1.2251742064009944]                   0.2777037147576 -0.055802499999982340
   [-0.1664596933e-7, -0.1201483265e-7]                        0.205291...e-7
   [-0.15157600929e-6, 0.11126321781e-6]                       0.188028...e-6

18 [ 1.16195559071665090,-1.22437873929677543]                 0.2767226626295 -0.0558024999999823810
   [ 0.21321160566e-7, 0.15389677270e-7]                       0.262951...e-7
   [-0.130534903135e-6, 0.134872272738e-6]                     0.187696...e-6

19 [ 1.073426518281103759,-1.101727674543438430]               0.1254590505101 -0.05580249999999201395
   [ 0.1126192476153e-6, 0.812881573601e-7]                    0.138891...e-6
   [-0.34913050959e-8, 0.1263334781934e-6]                     0.126381...e-6

20 [ 1.0002887216172102614,-1.0004000137678948943]             0.493326...e-3  -0.055802499999999968610
   [ 0.950807431784e-8, 0.686288727881e-8]                     0.117261...e-7
   [ 0.488990082160e-8, 0.623652976668e-8]                     0.792498...e-8

21 [ 1.00000268136619826112,-1.00000371573596083211]           0.458218...e-5  -0.0558024999999999997085
   [ 0.917062645403e-9, 0.661931861565e-9]                     0.113099...e-8
   [ 0.493711304804e-9, 0.582659175957e-9]                     0.763703...e-9

22 [ 1.000000000120819184363,-1.000000000171221117411]         0.209556...e-9  -0.0558025+0.4e-21
   [ 0.39906190925e-11, 0.28804116356e-11]                     0.492156...e-11
   [ 0.21585252993e-11, 0.25268010504e-11]                     0.332324...e-11

23 [ 1.0000000000775233080423,-1.0000000001027841519780]       0.128741...e-9  -0.0558025-0.5e-22
   [-0.480774210674e-11,-0.347020750015e-11]                   0.592931...e-11
   [-0.260070270938e-11,-0.304402408912e-11]                   0.400371...e-11

24 [ 0.99999999998957091850949,-0.99999999998322313402639]     0.197542...e-11 -0.0558025+0.5e-24
   [-0.242306733560e-12,-0.174895954377e-12]                   0.298832...e-12
   [-0.131069198054e-12,-0.153420327577e-12]                   0.201784...e-12

25 [ 1.00000000000097447929867,-1.00000000000087242747042]     0.130795...e-12 -0.0558025-0.4e-24
   [-0.497140321615e-13,-0.358833738256e-13]                   0.613115...e-13
   [-0.268917187117e-13,-0.314769896679e-13]                   0.414000...e-13
```

```
k      x(2)
                                                          ||e(2)||_2=||xs-x(2)||_2   R(x(2))
       rd(2)=c-Bx(2)                                      ||rd(2)||_2
       r(2)=b-Ax(2)                                       ||r(2)||_2

-------------------------------------------------------------------------------------------------------

26 [ 0.99999999999999496951332695,-0.99999999999998410470076396]
                                                          0.166723...e-13    -0.0558025-0.2e-25
   [-0.929013287453e-14,-0.670557780836e-14]              0.114573...e-13
   [-0.502527386491e-14,-0.588216786406e-14]              0.773648...e-14

27 [ 1.00000000000000176855795130,-1.00000000000000094932302494]
                                                          0.179153...e-14    -0.0558025
   [-0.156214369085e-14,-0.112754857313e-14]              0.192656...e-14
   [-0.845006338973e-15,-0.989089536102e-15]              0.130089...e-14

28 [ 0.99999999999999996685274922553,-0.99999999999999997747799858447]
                                                          0.403307...e-16    -0.0558025-0.5e-28
   [ 0.191332132472e-16, 0.138102707341e-16]              0.235966...e-16
   [ 0.103496880715e-16, 0.121144102426e-16]              0.159334...e-16

29 [ 1.00000000000000000008433263713,-1.00000000000000000161234090411]
                                                          0.161454...e-16    -0.0558025+0.4e-28
   [ 0.155653065096e-16, 0.112349710516e-16]              0.191964...e-16
   [ 0.841968472053e-17, 0.985536958108e-17]              0.129622...e-16

30 [ 1.00000000000000000009702106702,-0.99999999999999999394117634467]
                                                          0.613601...e-18    -0.0558025-0.1e-29
   [-0.770507978208e-18,-0.556149332805e-18]              0.950255...e-18
   [-0.416788204071e-18,-0.487856713075e-18]              0.641651...e-18

31 [ 0.99999999999999999988899177855117,-0.99999999999999999988839697146443]
                                                          0.159940...e-18    -0.0558025+0.2e-30
   [ 0.379700075332e-19, 0.274065875415e-19]              0.468277...e-19
   [ 0.205390766456e-19, 0.240411037784e-19]              0.316200...e-19

32 [ 0.99999999999999999997998719896002,-0.99999999999999999999687348460174]
                                                          0.202555...e-19     -0.0558025-0.5e-32
   [ 0.2560373959073e-19, 0.1848066871910e-19]            0.315766...e-19
   [ 0.1384975664196e-19, 0.1621131370205e-19]            0.213218...e-19

33 [ 0.99999999999999999999548607493079,-1.00000000000000000000000884486456]
                                                          0.451479...e-21     -0.0558025-0.5e-33
   [ 0.660099798756e-21, 0.476457185448e-21]              0.814090...e-21
   [ 0.357065814145e-21, 0.417950124564e-21]              0.549707...e-21

34 [ 1.00000000000000000000000005275538314,-0.99999999999999999999998990283915254]
                                                          0.101109...e-21     -0.0558025+0.15e-33
   [-0.112699119609e-21,-0.813457380729e-22]              0.138990...e-21
   [-0.609619354561e-22,-0.713568564655e-22]              0.938517...e-22

35 [ 0.99999999999999999999999853704616070,-1.00000000000000000000000027816275541]
                                                          0.314287...e-23     -0.0558025-0.1e-34
   [ 0.500467151442e-23, 0.361235029665e-23]              0.617218...e-23
   [ 0.270716030761e-23, 0.316876941343e-23]              0.416771...e-23

36 [ 1.00000000000000000000000099780322576,-1.00000000000000000000000011372446346]
                                                          0.100426...e-23     -0.0558025+0.4e-35
   [-0.132043610200e-23,-0.953085078817e-24]              0.162847...e-23
   [-0.714259643165e-24,-0.836049923699e-24]              0.109961...e-23

40 [ 0.99999999999999999999999999345649771297,-0.99999999999999999999999999488817027852]
                                                          0.830350...e-28     -0.0558025+0.25e-39
   [ 0.411509929242e-28, 0.297026090661e-28]              0.507508...e-28
   [ 0.222597165069e-28, 0.260552180160e-28]              0.342690...e-28
```

Im CG mit den anderen Startvektoren machen wir einige Iterationen mit `Digits:=24`. Auf grafische Darstellungen verzichten wir wegen $x^{(1)}$ wie beim GV und $x^{(2)} \approx x^*$. Man beachte Situationen, dass im Vergleich von $x^{(k)}$ mit $x^* = (1, -1)^T$ trotz kleinerem Wert $\|\hat{r}^{(k)}\|_2$ die Iterierte ungenauer sein kann.

Ergebnisse aus Berechnungen mit Maple für Startvektor $x^{(0)} = (0.999, -1.001)^T$

```
Startvektor              x = [ 9.9900000000000000e-01, -1.0010000000000000e+00]
Residuum         rd = c-Bx = [ 2.4827760000000000e-03,  1.7920570000000000e-03]
Funktionswert         R(x) =  -5.5800362583500000e-02
Anfangsfehlerquadrat rd'rd =   9.3756449574250000e-06

 k   Iterationsvektor  x                                          Funktionswert  R(x)
     Residuum  rd=c-Bx                                          Fehlernormquadrat  rd'rd
-------------------------------------------------------------------------------------

 1  [ 1.0001320237513903153 9812, -1.0001829101425802108 4407]  -5.5802500000000000e-02
    [-6.0196338e-17, 8.3398029e-17]                              1.0578830349695085e-32

 2  [ 1.0000000167725131832 3426, -1.0000000232391842229 5359]  -5.5802500000000000e-02
    [ 2.061551224528e-12, 1.488018791379e-12]                    6.4641933748499162e-24

 3  [ 1.0000000000000982359 3845, -1.0000000000001360993 7091]  -5.5802500000000000e-02
    [ 0, 0]                                                      0
```

Ergebnisse aus Berechnungen mit Maple für Startvektor $x^{(0)} = (0.341, -0.087)^T$

```
Startvektor              x = [ 3.4100000000000000e-01, -8.7000000000000000e-02]
Residuum         rd = c-Bx = [ 7.8000000000000000e-07,  5.6300000000000000e-07]
Funktionswert         R(x) =  -5.5802499999500000e-02
Anfangsfehlerquadrat rd'rd =   9.2536900000000000e-13

 k   Iterationsvektor  x                                            Funktionswert  R(x)
     Residuum  rd=c-Bx                                            Fehlernormquadrat  rd'rd
-----------------------------------------------------------------------------------------

 1  [ 0.341000355641639070418560, -0.086999743299688722249 1675]  -5.5802499999710961e-02
    [ 3.00471205608e-13, -4.16283375443e-13]                      2.6357479406974271e-25

 2  [ 0.999999999997844633251351, -0.999999999997672238245988 ]  -5.5802500000000000e-02
    [ 6.85221307274e-13,  4.94589781861e-13]                      7.1414729226360109e-25

 3  [ 0.999999999998528258016127, -0.999999999997960999189337 ]  -5.5802500000000000e-02
    [ 0, 0]                                                       0
```

Ergebnisse aus Berechnungen mit Maple für Startvektor
$x^{(0)} = (0.991\,891\,566\,446\,068\,04, -1.005\,852\,632\,339\,509\,328)^T$

Der Vektor $x^* - x^{(0)}$ ist orthogonal zur Tallinie $t(x_1)$ mit der Genauigkeitsordnung $\mathcal{O}(10^{-17})$.

Somit muss der erste CG-Schritt, der ein Gradientenschritt ist, bis auf Ungenauigkeiten in den letzten Dezimalstellen in das Minimum führen, also $x^{(1)} \approx x^*$. Weitere Iterationen werden sich dann um die Minimumstelle bewegen, wobei die Situation zwischenzeitlich sich verschlechtern kann.

```
Startvektor              x = [ 9.9189156644606804e-01, -1.0058526323395093e+00]
Residuum         rd = c-Bx = [ 1.7783570530717400e-02,  1.2836104447023644e-02]
Funktionswert         R(x) =  -5.5692839050000023e-02
Anfangsfehlerquadrat rd'rd =   4.8102095819590051e-04
```

k	Iterationsvektor x Residuum rd=c-Bx	Funktionswert R(x) Fehlernormquadrat rd'rd
1	[0.99999999999999998743683, -0.99999999999999998259454] [1e-24, -2e-24]	-5.5692839050000023e-02 5e-48
2	[0.99999999999999998743701, -0.99999999999999998259489] [1.2e-23, 6e-24]	-5.5802500000000000e-02 1.8e-46
3	[0.99999999999999998789555, -0.99999999999999998322538] [-4.86e-22, -3.53e-22]	-5.5802500000000000e-02 3.60805e-43
4	[1.00000000000000008997270, -1.00000000000000012467212] [2.1775e-20, 1.5715e-20]	-5.5802500000000000e-02 7.2111185e-40
5	[1.00000000000017093336465, -1.00000000000023681595209] [-9.12236e-19, -6.58450e-19]	-5.5802500000000000e-02 1.265730922196e-36
6	[1.00000000000268737388995, -1.00000000000372317841966] [3.20728e-19, 2.31500e-19]	-5.5802500000000000e-02 1.56458699984e-37
7	[1.00000000000270730622973, -1.00000000000375079303803] [-7.226e-21, -5.215e-21]	-5.5802500000000000e-02 7.9411301e-41
8	[1.00000000000270731626859, -1.00000000000375080695330] [1.61e-22, 1.17e-22]	-5.5802500000000000e-02 3.961e-44
9	[1.00000000000270731627367, -1.00000000000375080696018] [-4e-24, -2e-24]	-5.5802500000000000e-02 2e-47
10	[1.00000000000270731627367, -1.00000000000375080696018] [0, 1e-24]	-5.5802500000000000e-02 1e-48

Literaturverzeichnis

[1] Axelsson, O.: *Iterative Solution Methods*. Cambridge University Press Cambridge 1994, 1996.

[2] Beresin, I. S. und N. P. Shidkow: *Numerische Methoden*. Bd. 1,2. DVW Berlin 1970, 1971.

[3] Brezinski, C.: *Projection Methods for Systems of Equations*. Studies in Computational Mathematics. Elsevier Amsterdam 1997.

[4] Deuflhard, P. und H. Hohmann: *Numerische Mathematik*. 1: Eine algorithmisch orientierte Einführung. 3. überarbeitete und erweiterte Auflage, Lehrbuch. Walter de Gruyter Berlin 2002.

[5] Faddejew, D. K. und W. N. Faddejewa: *Numerische Methoden der linearen Algebra*. Math. für Naturwiss. und Technik, Bd. 10. DVW Berlin 1973.

[6] Fischer, B.: *Polynomial Based Iteration Methods for Symmetric Linear Systems*. Advances in Numerical Mathematics. Wiley-Teubner Stuttgart 1996.

[7] Greenbaum, A.: *Iterative Methods for Solving Linear Systems*. SIAM Philadelphia 1997.

[8] Hackbusch, W.: *Iterative Lösung großer schwachbesetzter Gleichungssysteme*. Leitfäden der angewandten Mathematik und Mechanik Band 69. B. G. Teubner Stuttgart 1991, 1993.

[9] Hämmerlin G. und K.-H. Hoffmann: *Numerische Mathematik*. Grundwissen Mathematik 7. Springer-Verlag Berlin 1991.

[10] Hermann, M.: *Numerische Mathematik*. R. Oldenbourg Verlag München 2001.

[11] Hestenes, J. R. und E. Stiefel: *Methods of conjugate gradients for solving linear systems*. Journ. Res. Nat. Bur. Stand. 49 (1952) 409-436.

[12] Kelley, C. T.: *Iterative Methods for Linear and Nonlinear Equations*. Frontiers in Applied Mathematics. SIAM Philadelphia 1995.

[13] Kiełbasiński, A. und H. Schwetlick: *Numerische lineare Algebra*. DVW Berlin 1988.

[14] Maess, G.: *Vorlesungen über Numerische Mathematik I, II*. Akademie-Verlag Berlin 1984, 1988.

[15] Meister, A.: *Numerik linearer Gleichungssysteme*. Ein Einführung in moderne Verfahren. Vieweg Braunschweig 1999.

[16] Meurant, G.: *Computer Solution of Large Linear Systems*. Studies in Mathematics and Its Applications, Vol 28. Elsevier Science B. V. 1999.

[17] Neundorf, W.: *Numerische Mathematik*. Vorlesungen, Übungen, Algorithmen und Programme. Shaker Verlag Aachen 2002.

[18] Neundorf, W.: *Grundlagen der numerischen linearen Algebra*. Preprint No. M 04/04 IfMath der TU Ilmenau, Februar 2004.

[19] Opfer, G.: *Numerische Mathematik für Anfänger*. Vieweg Studium Grundkurs Mathematik Wiesbaden 1993, 3. überarbeitete und erw. Auflage 2001.

[20] Plato, R.: *Numerische Mathematik kompakt*. Grundlagenwissen für Studium und Praxis. Vieweg Wiesbaden 2000.

[21] Quarteroni, A., R. Sacco und F. Saleri: *Numerische Mathematik*. Band 1, 2. Springer-Verlag Berlin 2002.

[22] Ralston, A.: *A First Course in Numerical Analysis*. McGraw-Hill New York 1965.

[23] Roos, H.-G. und H. Schwetlick: *Numerische Mathematik*. Das Grundwissen für jedermann. B. G. Teubner Stuttgart 1999.

[24] Saad, Y.: *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company Boston 1995.

[25] Schwarz, H. R.: *Numerische Mathematik*. B. G. Teubner Stuttgart 1988.

[26] Schwarz, H. R., H. Rutishauser und E. Stiefel: *Numerik symmetrischer Matrizen*. B. G. Teubner Stuttgart 1972.

[27] Stoer, J. und R. Bulirsch: *Einführung in die Numerische Mathematik II*. Heidelberger Taschenbücher 114. Springer-Verlag Berlin 1990.

[28] Trefethen, L. N. und D. Bau: *Numerical Linear Algebra*. SIAM Philadelphia 1997.

[29] Überhuber, C.: *Computer-Numerik 1,2*. Springer-Verlag Berlin 1995.

[30] Van der Vorst, H. A.: *Iterative Krylov Methods for Large Linear Systems*. Cambridge University Press Cambridge 2003.

[31] Kanzow, Ch.: *Numerik linearer Gleichungssysteme*. Direkte und iterative Verfahren. Springer-Verlag Berlin Heidelberg 2005.

[32] Watkins, David S.: *Fundamentals of Matrix Computations*. 2nd Ed. Pure and Applied Mathematics. A Wiley-Interscience Series of Texts, Monographs, and Tracts. A John Wiley & Sons, Inc., Publication, New York 2002.

Anschrift:

Dr. rer. nat. habil. Werner Neundorf
Technische Universität Ilmenau, Institut für Mathematik
PF 10 05 65
D - 98684 Ilmenau

E-mail : werner.neundorf@tu-ilmenau.de
Homepage : http://www.mathematik.tu-ilmenau.de/∼neundorf/index_de.html