

# Crashkurs

## Ableitung, Gradient, Abstieg, Minimierung, cg – Verfahren

Stand: Februar 2020

© Ingenieurbüro Dr. Tilman Hasse

---

**Ingenieurbüro Dr. Tilman Hasse**

Eibenweg 5 • 76337 Waldbronn • Telefon: 07243 / 572113 • FAX: 07243 / 572114  
[tilman.hasse@t-online.de](mailto:tilman.hasse@t-online.de)

# Inhaltsverzeichnis

1	Der „1D – Fall“: Funktion einer Variablen.....	3
2	Etwas lineare Algebra.....	4
2.1	Eine Lösung für $Ax = b$ , wenn $x$ , $b$ gegeben sind.....	5
3	Richtungsableitung, Differenzierbarkeit.....	6
4	Das allgemeine CG – Verfahren.....	8
4.1	Berechnung der Schrittweite $\gamma$ .....	8
4.2	Berechnung einer Suchrichtung $p^+$ .....	9
4.3	Der CG – Algorithmus.....	11
4.4	Genauigkeitsfragen – Abbruchkriterium.....	12
4.5	Beweis zum Beispiel (4.2).....	12
5	Literatur.....	14

# 1 Der „1D – Fall“: Funktion einer Variablen

Auf der Definitionsmenge  $D \subset \mathbb{R}$ , einer konvexen Teilmenge der reellen Zahlen, sei die reellwertige Funktion  $f: x \rightarrow f(x)$  mit  $x \in D$  und  $f(x) \in \mathbb{R}$  definiert. Dann wird für  $x_0 \in D$  die Ableitung nach  $x$  definiert durch:

$$(df/dx)(x_0) = f'(x_0) = \lim (f(x_0+h) - f(x_0)) / h \text{ für } h \rightarrow 0 \quad (1.1a)$$

(1.1a) gilt genau dann, falls in  $x_0$  die folgende **lineare Approximation der Funktion  $f$**  existiert:

$$f(x) = f(x_0) + f'(x_0) \cdot (x - x_0) + R(x), \text{ wobei für } \varepsilon > 0 \text{ eine Konstante } K > 0 \text{ existiert, so dass } |R(x)| < K \cdot (|x - x_0|^2) \text{ für } |x - x_0| < \varepsilon \text{ gilt.} \quad (1.1b)$$

$f'(x_0)$  ist also die **lokale Steigung** im Punkt  $x_0$ .

$$\text{Wichtiges Beispiel: } (x^n)' = n \cdot x^{n-1} \quad (1.2)$$

Siehe Buch S. 287ff, insbesondere auch Abbildung 10.4

Lösung der **Nullstellenaufgabe**  $f(x) = 0$  mit dem Newton – Verfahren auf S. 294

**Differenzierungsregeln:** Linearität des Differenzierens, Produktregel, Ableitung von Verkettungen, Umkehrfunktionen S.295 bis S.297

**Wichtige Bemerkung:** Für  $s \in \mathbb{R}$ ,  $s > 0$  gilt:

$$\lim (f(x_0+h \cdot s) - f(x_0)) / h = s \cdot \lim (f(x_0+h) - f(x_0)) / h \text{ für } h \rightarrow 0 \quad (1.3)$$

**Minimierungsproblem  $f(x^*) = \text{Min } f(x)$ ,  $x \in D$ ,  $f$  hinreichend glatt**

I) Polynom 2. Grades:  $p(x) = s \cdot x^2 + a \cdot x + b$ , o.B.d.A.  $s > 0$ ; Lösung für:

$$p'(x^*) = 0 = 2 \cdot s \cdot x + a \rightarrow x^* = -a / (2 \cdot s) \quad (1.4)$$

II) Zusammenhang zwischen Nullstellenaufgabe

$$f(x^*) = 0 \rightarrow f^2(x^*) = \text{Min } f^2(x), \text{ lokales Minimum für lokale Nullstelle}$$

III) Numerische Berechnung: Es existiere das lokale Minimum  $(x^*, f(x^*))$ ; bis  $x_i$ ,  $f_i = f(x_i)$  sei alles berechnet. Gesucht:  $x_{i+1}$ ,  $f_{i+1} = f(x_{i+1}) < f_i$ ; Idee: man nähere  $f$  in  $(x_i, f_i)$  durch ein Polynom 2. Grades an.

A) Taylorreihe:  $p_i(x) = f_i + f'(x_i) \cdot (x - x_i) + 0,5 \cdot f''(x_i) \cdot (x - x_i)^2$ ; es müssen dazu also die 1. und 2. Ableitung verfügbar sein. Es ergibt sich  $x_{i+1}$  als Nullstelle gemäß (1.4); Taylorreihe siehe Buch S.316

B)  $(x_i, f_i = f(x_i))$ ,  $(x_{i-1}, f_{i-1} = f(x_{i-1}))$ ,  $(x_{i-2}, f_{i-2} = f(x_{i-2}))$  seien berechnet. Dann berechne man das Polynom 2. Grades durch diese 3 Punkte:  $p_i(x) = A_0 + A_1 \cdot (x - x_i) + A_2 \cdot (x - x_i) \cdot (x - x_{i-1})$ . Man bestimme  $A_0$ ,  $A_1$  und  $A_2$ . Es ergibt sich  $x_{i+1}$  als Nullstelle gemäß (1.4) (1.5)

## 2 Etwas lineare Algebra

Wir haben die Menge der **Vektoren** als  $n$  – Tupel  $(x_1, x_2, \dots, x_n)^T = \underline{x} \in \mathbb{R}^n$ , die traditionell als Spalten geschrieben werden, daher das hochgestellte T (wie transponiert). Es sind **Addition** gemäß  $\underline{x} + \underline{y} = (x_1+y_1, \dots, x_n+y_n)^T$  und **skalare Multiplikation** gemäß  $a \cdot \underline{x} = (a \cdot x_1, \dots, a \cdot x_n)^T$  komponentenweise definiert und es gelten die entsprechenden Rechenregeln.

Eine **quadratische Matrix**  $A \in \mathbb{R}^n \times \mathbb{R}^n$  ist ein quadratisches Zahlenschemata der Form  $(A_{i,j})_{i,j=1}^n$ , wobei der erste Index, hier  $i$ , der Zeilenindex ist und der zweite Index, hier  $j$ , der Spaltenindex ist. Auch für quadratische Matrizen sind die **Addition** und die **skalare Multiplikation** komponentenweise definiert und es gelten die entsprechenden Rechenregeln. Mit der Matrix  $A$  kann eine Abbildung  $\mathbb{R}^n \xrightarrow{A} \mathbb{R}^n$  definiert werden durch:  $(A\underline{x})_i = (\sum_j A_{i,j} \cdot x_j)$ . Die Abbildung durch zwei Matrizen  $A, B \in \mathbb{R}^n \times \mathbb{R}^n$  hintereinander  $\mathbb{R}^n \xrightarrow{A} \mathbb{R}^n \xrightarrow{B} \mathbb{R}^n$ , führt zur **Matrixmultiplikation**  $(BA)_{i,k} = \sum_j B_{i,j} \cdot A_{j,k}$ , nämlich  $B(A\underline{x}) = (BA)\underline{x}$ . Die Matrix  $E$ , die definiert ist durch  $E_{i,i} = 1$  und  $E_{i,j} = 0$  für  $i \neq j$ , heißt Einheitsmatrix. Eine Matrix  $A^{-1}$  mit  $AA^{-1} = E$  heißt inverse Matrix zu  $A$ . Die **transponierte Matrix**  $A^T$  ist definiert durch:  $A^T = (A_{j,i})$  für  $A = (A_{i,j})$ . Für die Matrixmultiplikation gelten die folgenden Rechenregeln:

- $AE = A = EA$  und  $E\underline{x} = \underline{x}$
- $(AB)C = A(BC)$
- $A(B + C) = AB + AC$  und  $(A + B)C = AC + BC$
- Die Matrixmultiplikation ist i.A. nicht kommutativ, also meist  $AB \neq BA$
- $(A^{-1})^{-1} = A$ ,  $(AB)^{-1} = B^{-1}A^{-1}$   $E^{-1} = E$
- $(A + B)^T = A^T + B^T$   $(aA)^T = aA^T$   $(A^T)^T = A$
- $(AB)^T = B^T A^T$  man beachte die Reihenfolge! (2.1)

Das **Standardskalarprodukt oder kanonische Skalarprodukt** zweier Vektoren  $\underline{x}, \underline{y} \in \mathbb{R}^n$  ist definiert durch  $\langle \underline{y}, \underline{x} \rangle = \underline{y} \cdot \underline{x} = \underline{y}^T \underline{x} = \underline{x}^T \underline{y} = \sum_j y_j \cdot x_j$  (Zeilenvektor x Spaltenvektor) (2.2)

**Achtung:** „ $\cdot$ “ hat hier genau zwei Bedeutungen: Multiplikation zwischen Zahlen und Skalarprodukt. Die Matrixmultiplikation wird hier ohne Zwischenzeichen definiert !!!!! Siehe Seite 677.

Eine symmetrische Matrix  $S = S^T$  heißt **positiv definit**, falls  $\underline{x}^T S \underline{x} > 0$  für alle  $\underline{x} \in \mathbb{R}^n$  mit  $\underline{x} \neq \underline{0}$ . Insbesondere die Einheitsmatrix  $E$  ist positiv definit. Mit einer positiv definiten Matrix  $S$  kann das **Skalarprodukt**  $\langle \underline{y}, \underline{x} \rangle$  definiert werden durch  $\langle \underline{y}, \underline{x} \rangle = \underline{y}^T S \underline{x}$  für  $\underline{x}, \underline{y} \in \mathbb{R}^n$ . Zwei Vektoren  $\underline{x}, \underline{y} \in \mathbb{R}^n$  mit  $\underline{x} \neq \underline{0}$  und  $\underline{y} \neq \underline{0}$  heißen **orthogonal** bezüglich  $S$ , falls  $\underline{y}^T S \underline{x} = 0$  gilt. Der **Betrag bzw. die Norm** eines Vektors sei definiert durch:  $\|\underline{x}\| = \sqrt{\underline{x}^T S \underline{x}}$ . Für  $S=E$  erhält man natürlich das Standardskalarprodukt. Es gilt der wichtige Satz:

Zu jedem durch eine positiv definite Matrix  $S$  definierten Skalarprodukt gibt es eine Orthonormalbasis von  $n$  (Basis –) Vektoren  $\underline{e}_i$ ,  $i = 1, \dots, n$  für die gilt:  $\underline{e}_i^T S \underline{e}_j = 0$  für  $i \neq j$  und  $\underline{e}_i^T S \underline{e}_i = 1$  und diese Basis kann als Standardbasis gemäß (2.4) verwendet werden.

Siehe Seite 679 (2.3)

Die **Standardbasis zum Standardskalarprodukt** wird gebildet durch die  $n$  Vektoren  $\underline{e}_i$ ,  $i = 0, \dots, n-1$ , für deren Komponenten gilt:  $e_{j,i} = 0$  für  $i \neq j$  und  $e_{i,i} = 1$ . Bezüglich des Standardskalarprodukts bilden diese  $n$  Vektoren  $\underline{e}_i$  eine Orthonormalbasis, d.h.,  $\underline{e}_i^T \underline{e}_j = 0$  für  $i \neq j$  und  $\underline{e}_i^T \underline{e}_i = 1$  (2.4)

Für zwei Vektoren  $\underline{x}, \underline{y} \in \mathbb{R}^n$  gilt die **Cauchy – Schwarz'sche Ungleichung**:

$$|\underline{x}^T \underline{y}| \leq \|\underline{x}\| \cdot \|\underline{y}\| \quad \text{Siehe Seite 681} \quad (2.5)$$

## **2.1 Eine Lösung für $A\underline{x} = \underline{b}$ , wenn $\underline{x}, \underline{b}$ gegeben sind.**

Diese Aufgabenstellung spielt in der allgemeinen Theorie über Approximation eine gewisse wichtige Rolle.

**(2.1.1) Problem:** Gegeben  $\underline{x}, \underline{b} \in \mathbb{R}^n$ ,  $\underline{b}^T \underline{x} \neq 0$ . Gesucht:  $A \in \mathbb{R}^n \times \mathbb{R}^n$  mit  $A\underline{x} = \underline{b}$

Lösung:

- $A = c \cdot \underline{b}\underline{b}^T$  mit  $c = 1 / \underline{b}^T \underline{x}$
- Anwendung:  $\underline{E} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\underline{E}_0 = \underline{E}(\underline{x}_0)$ ,  $\underline{E}_1 = \underline{E}(\underline{x}_1)$ , lineare Näherung durch:
- $\underline{G}(\underline{x}) = \underline{E}_0 + A(\underline{x} - \underline{x}_0)$ , also  $\underline{E}_1 - \underline{E}_0 = A(\underline{x}_1 - \underline{x}_0)$

### 3 Richtungsableitung, Differenzierbarkeit

$D \subset \mathbb{R}^n$  sei eine konvexe Teilmenge des  $\mathbb{R}^n$ , in dem das Standardskalarprodukt mit der Standardbasis definiert ist. Gegeben sei die Funktion  $f: D \rightarrow \mathbb{R}$ ,  $\underline{x} \rightarrow f(\underline{x})$  und eine Richtung  $\underline{e} \in \mathbb{R}^n$  mit  $\|\underline{e}\| = 1$ . Dann wird für den Punkt  $\underline{x}^0 \in D$  die **Richtungsableitung** in Richtung  $\underline{e}$  definiert durch:

$$(\partial f / \partial \underline{e})(\underline{x}^0) = \lim_{h \rightarrow 0} (f(\underline{x}^0 + h \cdot \underline{e}) - f(\underline{x}^0)) / h \quad (3.1)$$

Die Ableitung wird also auf den 1D – Fall zurückgeführt. (Buch S.796)

Sei nun durch einen beliebigen Vektor  $\underline{p} \neq \underline{0}$  eine Richtung vorgegeben, so bildet man den Einheitsvektor  $\underline{e}_p = (1/\|\underline{p}\|) \cdot \underline{p}$  und  $(\partial f / \partial \underline{e}_p)(\underline{x}^0)$  ist dann **die Steigung in Richtung  $\underline{p}$** .

#### Definition:

(a) Die **partiellen Ableitungen in  $\underline{x}_0$**  werden definiert durch:

$$f_{x_i}(\underline{x}^0) = (\partial f / \partial x_i)(\underline{x}^0) = (\partial f / \partial \underline{e}_i)(\underline{x}^0), \quad i = 1, \dots, n$$

(b) Existieren die partiellen Ableitungen in  $\underline{x}^0$ , so wird der **Gradient**  $(\nabla f)$  bzw.  $(\text{grad } f)$  in  $\underline{x}^0$  definiert durch:  $(\nabla f)(\underline{x}^0) = (\text{grad } f)(\underline{x}^0) = (f_{x_1}(\underline{x}^0), f_{x_2}(\underline{x}^0), \dots, f_{x_n}(\underline{x}^0))^T$  (3.2)

#### Ein wichtiges Beispiel: Die Linearform :

Es seien  $\underline{a} \in \mathbb{R}^n$  und  $b \in \mathbb{R}$ ,  $\underline{a} \neq \underline{0}$  gegeben; damit wird die Funktion  $g$ , eine sog. Linearform, definiert durch  $g(\underline{x}) = \underline{a}^T \underline{x} + b$  (3.3)

Gesucht ist die Richtungsableitung für eine beliebige Richtung  $\underline{e}$  mit  $\|\underline{e}\| = 1$ , die Richtung mit der größten Steigung und die Lösungsmenge für  $g(\underline{x}) = 0$ . Wir erhalten:

$$(\partial g / \partial \underline{e})(\underline{x}^0) = \lim_{h \rightarrow 0} (\underline{a}^T(\underline{x}^0 + h \cdot \underline{e}) - \underline{a}^T \underline{x}^0) / h = \underline{a}^T \underline{e} \quad \text{und daher: } (\nabla g)(\underline{x}^0) = (\text{grad } g)(\underline{x}^0) = \underline{a}$$

Mit der Cauchy – Schwarz'sche Ungleichung (2.5) folgt:  $|\underline{a}^T \underline{e}| \leq \|\underline{a}\|$ ; das Gleichheitszeichen gilt für  $\underline{e} = (1/\|\underline{a}\|) \cdot \underline{a}$ , also zeigt  $\underline{a}$  und damit der Gradient in die **Richtung des steilsten Abstiegs**  $\|\underline{a}\|$ .

$$(3.4)$$

Um eine Lösung für  $g(\underline{x}) = 0$  zu erhalten, wählen wir sinnvollerweise die Richtung des steilsten Abstiegs; zu lösen also:  $g(u \cdot \underline{a}) = 0$ , also:  $u = -b / \|\underline{a}\|^2$ . Damit ergibt sich die **Lösungsmenge** zu:  $L = \{ \underline{v} \in \mathbb{R}^n \mid (-b / \|\underline{a}\|^2) \cdot \underline{a} + \underline{v} \text{ mit } \underline{a} \cdot \underline{v} = \underline{a}^T \underline{v} = 0 \}$

#### Definition: Differenzierbarkeit einer Funktion $f$ im Punkt $\underline{x}_0$

$D \subset \mathbb{R}^n$  sei eine konvexe Teilmenge des  $\mathbb{R}^n$ , auf der die Funktion  $f: D \rightarrow \mathbb{R}$ ,  $\underline{x} \rightarrow f(\underline{x})$  definiert sei. Weiter sei  $\underline{x}^0 \in D$ . Dann heißt  $f$  in  $\underline{x}^0$  differenzierbar, falls der Gradient  $(\text{grad } f)(\underline{x}^0)$  gemäß (3.2) existiert und damit folgendes in einer  $\varepsilon$  – Umgebung gilt:

$$f(\underline{x}) = f(\underline{x}^0) + (\text{grad } f)(\underline{x}^0) \cdot (\underline{x} - \underline{x}^0) + R(\underline{x}) \quad \text{und es existieren Konstanten } \varepsilon, K \in \mathbb{R}, \varepsilon > 0 \text{ und } K > 0, \text{ so dass für das Restglied gilt: } R(\underline{x}) < K \cdot \|\underline{x} - \underline{x}^0\|^2 \quad \text{für } \|\underline{x} - \underline{x}^0\| < \varepsilon \quad (3.5)$$

**Bemerkung I:** Wegen (3.4) und (3.5) gilt also allgemein: Der Gradient zeigt die Richtung des steilsten Anstiegs an.

**Bemerkung II:** Es gilt:  $(\partial f / \partial \underline{e})(\underline{x}^0) = (\text{grad } f)(\underline{x}^0) \cdot \underline{e}$

**Beispiel: Die allgemeine quadratische Funktion**  $f(\underline{x}) = \underline{x}^T \underline{A} \underline{x} + \underline{b}^T \underline{x} + c$  mit einer beliebigen  $n \times n$  – Matrix  $A$ . Wir bilden die Ableitung in Richtung  $\underline{e} \in \mathbb{R}^n$  mit  $\|\underline{e}\| = 1$  für  $\underline{x}^0$ :  
 $\lim [(\underline{x}^0 + h \cdot \underline{e})^T \underline{A} (\underline{x}^0 + h \cdot \underline{e}) - \underline{x}^{0T} \underline{A} \underline{x}^0] / h = \underline{e}^T (\underline{A}^T + \underline{A}) \underline{x}^0$  und mit (3.4) ergibt sich:

$(\partial f / \partial \underline{e})(\underline{x}^0) = \underline{e}^T (\underline{A}^T + \underline{A}) \underline{x}^0 + \underline{b}^T \underline{e}$  und damit:

$$(\text{grad } f)(\underline{x}^0) = (\underline{A}^T + \underline{A}) \underline{x}^0 + \underline{b} \tag{3.6}$$

Analog zum eindimensionalen Fall wird eine **stationäre Stelle** in  $\underline{x}^0$  definiert durch:

$$(\text{grad } f)(\underline{x}^0) = 0$$

Dies ist eine notwendige Bedingung dafür, dass  $f$  im Punkt  $\underline{x}^0$  ein Extremum annimmt. Hinreichend ist in diesem Fall die Definitheit der Hesse-Matrix von  $f$ : ist sie positiv definit, liegt ein lokales Minimum vor; ist sie negativ definit, handelt es sich um ein lokales Maximum; ist sie indefinit, liegt kein Extrempunkt, sondern ein Sattelpunkt vor. Wenn sie nur semidefinit ist, ist keine Entscheidung anhand der Hesse-Matrix möglich.

**Definition: Hesse – Matrix**

Sei  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  eine zweimal stetig differenzierbare Funktion. Dann ist die Hesse-Matrix von  $f$  am Punkt  $\underline{x} \in D$  in  $D$  definiert durch:

$$H_f(\underline{x}) := \left( \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}) \right)_{i,j=1,\dots,n} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\underline{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\underline{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\underline{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\underline{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\underline{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\underline{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\underline{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\underline{x}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\underline{x}) \end{pmatrix}$$

Mit  $\partial^2 f / (\partial x_i \partial x_j)$  werden die zweiten partiellen Ableitungen bezeichnet. Die Hesse-Matrix ist bei stetigen zweiten Ableitungen wegen der Vertauschbarkeit der Differenziationsreihenfolge symmetrisch! Siehe Seite 823 (3.7)

**Taylor-Entwicklung (3.8)**

Die Taylor-Entwicklung einer zweimal stetig differenzierbaren Funktion  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  um eine Entwicklungsstelle  $\underline{x}^0 \in D$  ergibt sich zu:

$$f(\underline{x}) = f(\underline{x}^0) + (\underline{x} - \underline{x}^0)^T (\nabla f)(\underline{x}^0) + 0,5 \cdot (\underline{x} - \underline{x}^0)^T H_f(\underline{x}^0) (\underline{x} - \underline{x}^0) + R(\underline{x}) \text{ bzw.}$$

$$(\nabla f)(\underline{x}) = (\nabla f)(\underline{x}^0) + H_f(\underline{x}^0) (\underline{x} - \underline{x}^0) + \underline{R}(\underline{x})$$

Die Terme zweiter Ordnung dieser Entwicklung sind also durch die quadratische Form gegeben, deren Matrix die an der Entwicklungsstelle ausgewertete Hesse-Matrix ist. (3.8)

**Für Beispiel (3.6) gilt:**  $H_f \equiv \underline{A}^T + \underline{A}$

## 4 Das allgemeine CG – Verfahren

Das CG-Verfahren (von engl. conjugate gradients oder auch Verfahren der konjugierten Gradienten) ist eine effiziente numerische Methode zur Lösung von Minimierungsproblemen.

**Problemstellung:**  $D \subset \mathbb{R}^n$  sei eine konvexe Teilmenge des  $\mathbb{R}^n$  und gegeben sei eine zweimal stetig differenzierbare Funktion  $f: D \rightarrow \mathbb{R}$ ,  $\underline{x} \rightarrow f(\underline{x})$ . **Aufgabe:** Gesucht ist  $\underline{x}^* \in D$  mit:  $f(\underline{x}^*) = \text{Min } f(\underline{x})$  auf  $D$ , wobei  $\underline{x}^*$  nicht auf dem Rand von  $D$  liege. Hier Voraussetzung: Es existiert die Minimumstelle  $\underline{x}^*$ , also  $(\nabla f)(\underline{x}^*) = \underline{0}$ . (4.1)

**Beispiel, quadratisches Polynom:** Sei  $S$  eine positiv definite Matrix, also  $\underline{x}^T S \underline{x} > 0$  für  $\underline{x} \neq \underline{0}$ , und  $\underline{a} \in \mathbb{R}^n$  und  $b \in \mathbb{R}$ . Damit definieren wir die Funktion  $f_s(\underline{x}) = \underline{x}^T S \underline{x} + \underline{a}^T \underline{x} + b$ . Wir erhalten:  $(\nabla f_s)(\underline{x}) = 2 \cdot S \underline{x} + \underline{a}$  und damit ergibt sich  $\underline{x}^* = -0,5 \cdot S^{-1} \underline{a}$ . (4.2)

Wir setzen weiter voraus, dass für  $n = 1$  ein Minimierungsverfahren existiert, das wir **M1 – Verfahren** nennen wollen (siehe (1.5) B).

Die **prinzipielle Vorgehensweise** entspricht einem Abstiegsverfahren:

Wir gehen von einer bereits berechneten Näherung  $\underline{x}$  und einer bereits berechneten Suchrichtung  $\underline{p}$  aus:

Schritt 1: Zu  $\underline{p}$  wird eine sogenannte Schrittweite  $\gamma$  berechnet mit:

$$\underline{x}^+ = \underline{x} + \gamma \cdot \underline{p} \quad \text{mit } f(\underline{x}^+) < f(\underline{x})$$

Schritt 2: Anschließend wird eine neue Suchrichtung  $\underline{p}^+$  berechnet.

(4.3)

### 4.1 Berechnung der Schrittweite $\gamma$

Wir haben bereits berechnet:

- Den Wert  $f_0 := f(\underline{x})$  zum Näherungsvektor  $\underline{x}$
- Den Gradienten  $\underline{\nabla} := (\nabla f)(\underline{x})$
- Die Suchrichtung  $\underline{p}$  mit  $\underline{e}_p = (1/\|\underline{p}\|) \cdot \underline{p}$

Wir betrachten nun die Funktion  $f_p(z) = f(\underline{x} + z \cdot \underline{e}_p)$ ; dann gilt:

- $(df_p / dz)(0) = (\partial f / \partial \underline{e}_p)(\underline{x}) = \underline{\nabla}^T \underline{e}_p = f'_{p0}$  und damit lineare Näherung:
- $f_p(z) \cong f_0 + z \cdot f'_{p0}$

Um eine robuste Näherung für das Minimum in Richtung  $\underline{p}$  zu erhalten, benötigen wir aber ein quadratisches Polynom. Wir berechnen dazu einen zusätzlichen Funktionswert  $f_1$ :

(1)  $f_1 := f_p(\Delta z) = f(\underline{x} + \Delta z \cdot \underline{e}_p)$ ,  $\Delta z$  so groß wie möglich, so klein wie nötig! Damit ergibt sich das gewünschte Polynom  $g$  zu:

(2)  $f_p(z) \cong g(z) = f_0 + z \cdot f'_{p0} + 0,5 \cdot A \cdot z^2$  mit  $A = 2 \cdot (f_1 - f_0 - f'_{p0} \cdot \Delta z) / \Delta z^2$



$$(3) \quad g'(z) = f'_{p0} + A \cdot z \quad \text{fürs Minimum gilt:}$$

$$(4) \quad 0 = f'_{p0} + A \cdot z_{\text{Min}}, \text{ also } z_{\text{Min}} = -f'_{p0} / A$$

$$(5) \quad \text{Im Allgemeinen gilt nicht: } (df_p / dz)(z_{\text{Min}}) = 0 \quad (4.1.1)$$

Betrachten wir dazu unser **Beispiel (4.2)**: mit  $\nabla := (\nabla f_S)(\underline{x}) = 2 \cdot S\underline{x} + \underline{a}$

$$(1) \quad f_{Sp}(z) = f_S(\underline{x} + z \cdot \underline{e}_p) = (\underline{x} + z \cdot \underline{e}_p)^t S (\underline{x} + z \cdot \underline{e}_p) + \underline{a}^t (\underline{x} + z \cdot \underline{e}_p) + b, \text{ also:}$$

$$(2) \quad f_{Sp}(z) = g(z) = f_S(\underline{x}) + z \cdot (2 \cdot \underline{x}^t S \underline{e}_p + \underline{a}^t \underline{e}_p) + z^2 \cdot \underline{e}_p^t S \underline{e}_p = f_S(\underline{x}) + z \cdot \nabla^t \underline{e}_p + z^2 \cdot \underline{e}_p^t S \underline{e}_p \quad \text{und} \\ \text{daher } (df_{Sp} / dz)(0) = \nabla^t \underline{e}_p + 2 \cdot z \cdot \underline{e}_p^t S \underline{e}_p$$

$$(3) \quad (df_{Sp} / dz)(0) = (\partial f_S / \partial \underline{e}_p)(\underline{x}) = \nabla^t \underline{e}_p = f'_{p0} = 2 \cdot \underline{x}^t S \underline{e}_p + \underline{a}^t \underline{e}_p \quad \text{und}$$

$$(4) \quad A = 2 \cdot \underline{e}_p^t S \underline{e}_p, \text{ also } z_{\text{Min}} = -f'_{p0} / A = -(2 \cdot \underline{x}^t S \underline{e}_p + \underline{a}^t \underline{e}_p) / (2 \cdot \underline{e}_p^t S \underline{e}_p) = -(\nabla^t \underline{e}_p) / (2 \cdot \underline{e}_p^t S \underline{e}_p) = -\|\underline{p}\| \cdot (\nabla^t \underline{p}) / (2 \cdot \underline{p}^t S \underline{p})$$

$$(5) \quad \text{speziell gilt: } (df_{Sp} / dz)(z_{\text{Min}}) = 0 = \nabla^t \underline{e}_p - 2 \cdot (\nabla^t \underline{e}_p / (2 \cdot \underline{e}_p^t S \underline{e}_p)) \cdot \underline{e}_p^t S \underline{e}_p \quad (4.1.2)$$

**Man beachte**, dass in (4.1.2)(2)  $f_{Sp}(z) = g(z)$  steht, während im allgemeinen Fall nur  $f_p(z) \cong g(z)$  (siehe (4.1.1)(2)) gilt, siehe auch: (4.1.1)(5) und (4.1.2)(5)

**(4.1.3) Es erscheint so ziemlich sinnvoll**, als Schrittweite  $\gamma = z_{\text{Min}} / \|\underline{p}\|$  zu wählen, also  $\underline{x}^+ = \underline{x} + \gamma \cdot \underline{p} = \underline{x} + z_{\text{Min}} \cdot \underline{e}_p$  mit  $f(\underline{x}^+) < f(\underline{x})$ ; für **Beispiel (4.2)**:  $\gamma = -(\nabla^t \underline{p}) / (2 \cdot \underline{p}^t S \underline{p})$

Dazu noch folgende Definition:

**Definition:**  $\gamma$  heißt **perfekte Schrittweite** falls gilt:

$$df(\underline{x} + \gamma \cdot \underline{p}) / d\gamma = 0 \quad (4.1.4)$$

**Satz (4.1.5)**

(1) Mit einem Minimierungsverfahren 1D kann man immer die perfekte Schrittweite berechnen. Dann gilt:

$$(2) \quad df(\underline{x} + \gamma \cdot \underline{p}) / d\gamma = 0 = (\nabla f)(\underline{x} + \gamma \cdot \underline{p}) \cdot \underline{p} = 0, \text{ d.h., mit } \nabla^+ := (\nabla f)(\underline{x} + \gamma \cdot \underline{p}) \text{ gilt } \nabla^+ \cdot \underline{p} = 0$$

(3) **Für unsere Wahl von  $\gamma$  gemäß (4.1.3) gilt für das Beispiel(4.2) wegen (4.1.2)(5) also:  $\nabla^+ \cdot \underline{p} = 0$**

Wir haben nun also die Schrittweite  $\gamma$  gemäß (4.1.3) gewählt und erhalten einen verbesserten Näherungswert für das Minimum:  $\underline{x}^+ = \underline{x} + \gamma \cdot \underline{p} = \underline{x} + z_{\text{Min}} \cdot \underline{e}_p$  mit  $f(\underline{x}^+) < f(\underline{x})$  und berechnen nun  $\nabla^+ := (\nabla f)(\underline{x} + \gamma \cdot \underline{p})$ . Was damit berechnet werden muss, ist eine neue Suchrichtung  $\underline{p}^+$ . Dazu der folgende Abschnitt:

## 4.2 Berechnung einer Suchrichtung $\underline{p}^+$

Zusammenfassend haben wir bereits berechnet, **hier** aus gutem Grund  $\underline{x}^0$  statt  $\underline{x}$ :

$$\triangleright \quad \nabla := (\nabla f)(\underline{x}^0)$$

$$\triangleright \quad \underline{x}^+ = \underline{x}^0 + \gamma \cdot \underline{p}$$

$$\triangleright \quad \nabla^+ := (\nabla f)(\underline{x}^+)$$

Wir setzen als Näherung an:

$$(1) g(\underline{x}) = f(\underline{x}^+) + (\underline{x} - \underline{x}^+)^t S^+ (\underline{x} - \underline{x}^+) + \nabla^{+t} (\underline{x} - \underline{x}^+)$$

$$(2) (\nabla g)(\underline{x}) = 2 \cdot S^+ (\underline{x} - \underline{x}^+) + \nabla^+, \quad (\nabla g)(\underline{x}^0) = \nabla = 2 \cdot S^+ (-\gamma \cdot \underline{p}) + \nabla^+, \text{ also:}$$

$$(3) (\nabla^+ - \nabla) = 2 \cdot S^+ (\gamma \cdot \underline{p}) \rightarrow S^+ = c \cdot (\nabla^+ - \nabla) (\nabla^+ - \nabla)^t \text{ mit } c = 1 / (2 \cdot \gamma \cdot (\nabla^+ - \nabla)^t \underline{p}), \text{ siehe hierzu Abschnitt 2.1 und } - \text{ bezogen auf (4.2): Satz (4.5.2)}$$

$$(4) \text{ Da wir an } c \text{ oft nicht interessiert sind: } S^{++} = (\nabla^+ - \nabla) (\nabla^+ - \nabla)^t \quad (4.2.1)$$

Da sich für den allgemeinen nichtlinearen Fall die Funktion  $f(\underline{x})$  in der Nähe einer Lösung in etwa wie ein quadratisches Polynom verhält, kann man als Näherung  $g(\underline{x})$  verwenden. Die Vorgehensweise zur Berechnung einer neuen Suchrichtung  $\underline{p}^+$  wird nun wie folgt begründet.  $\nabla^+$  zeigt ja bereits die Richtung des steilsten Anstiegs an, die aber meist für eine neue Suchrichtung ungünstig ist; daher:

$$\textbf{Ansatz:} \quad \underline{p}^+ = \nabla^+ + \beta \cdot \underline{p} \quad (4.2.2)$$

Weiter verwenden wir eine grobe Näherung  $S^+$  für die Funktionalmatrix:

$$\text{➤ } S^+ = c \cdot (\nabla^+ - \nabla) (\nabla^+ - \nabla)^t \text{ es gilt für } c > 0:$$

$$\text{➤ } \underline{x}^t S^+ \underline{x} \geq 0 \text{ für alle } \underline{x} \text{ (positiv semidefinit)} \quad (4.2.3)$$

$$\text{und fordern: } \underline{p}^t S^+ \underline{p}^+ = 0 = \underline{p}^t S^{++} \underline{p}^+ \quad (4.2.4)$$

also:

$$\text{➤ } \underline{p}^t S^{++} \nabla^+ + \beta \cdot \underline{p}^t S^{++} \underline{p} = 0 \rightarrow$$

$$\text{➤ } \beta = - (\nabla^{+t} - \nabla^t) \nabla^+ / ((\nabla^{+t} - \nabla^t) \underline{p})$$

$$\text{➤ Es gilt im Allgemeinen nicht: } \nabla^{+t} \underline{p} = 0 \text{ (siehe (4.1.1)(5))} \quad (4.2.5)$$

Damit ergibt sich als **neue Suchrichtung**:  $\underline{p}^+ = \nabla^+ + \beta \cdot \underline{p}$

Bemerkenswert ist die Ableitung mit dem Ansatz (4.2.2) für unser **Beispiel (4.2)** wie folgt:

$$(1) \beta \text{ ergibt sich aus der Forderung: } \underline{p}^t S \underline{p}^+ = 0 \text{ (vergleichbar (4.2.4), wobei ja } \mathbf{2 \cdot S} \text{ die Hesse-Matrix ist.}$$

$$(2) \underline{p}^t S (\nabla^+ + \beta \cdot \underline{p}) = 0, \text{ also } \beta = - (\underline{p}^t S \nabla^+) / (\underline{p}^t S \underline{p})$$

$$(3) \text{ wegen } \nabla^+ := (\nabla f_s)(\underline{x} + \gamma \cdot \underline{p}) = 2 \cdot S(\underline{x} + \gamma \cdot \underline{p}) + \underline{a} = \nabla + 2 \cdot \gamma \cdot S \underline{p} \text{ und damit: } \underline{p}^t S = (\nabla^{+t} - \nabla^t) / (2 \cdot \gamma); \text{ eingesetzt ergibt sich:}$$

$$(4) \beta = - (\nabla^{+t} - \nabla^t) \nabla^+ / ((\nabla^{+t} - \nabla^t) \underline{p}) \text{ und bei Wahl von } \gamma \text{ gemäß (4.1.3) gilt gemäß (4.1.5)(2): } \nabla^+ \cdot \underline{p} = 0 \text{ und daher:}$$

$$(5) \beta = - (\nabla^{+t} - \nabla^t) \nabla^+ / ((\nabla^{+t} - \nabla^t) \underline{p}) = (\nabla^{+t} - \nabla^t) \nabla^+ / \nabla^t \underline{p} \quad (4.2.6)$$

Formal und inhaltlich stimmt (4.2.6)(4) mit (4.2.5) überein!!

Für unser **Beispiel (4.2)** gilt also:

### Satz (4.2.7)

- (1)  $\gamma = -(2 \cdot \underline{x}^t \underline{S} \underline{p} + \underline{a}^t \underline{p}) / (2 \cdot \underline{p}^t \underline{S} \underline{p}) = -(\nabla^t \underline{p}) / (2 \cdot \underline{p}^t \underline{S} \underline{p})$
- (2)  $\underline{x}^+ = \underline{x}^0 + \gamma \cdot \underline{p}$
- (3)  $\nabla^+ = 2 \cdot \underline{S}(\underline{x}^0 + \gamma \cdot \underline{p}) + \underline{a} = \nabla + 2 \cdot \gamma \cdot \underline{S} \underline{p}$
- (4)  $\underline{p}^+ = \nabla^+ + \beta \cdot \underline{p}$
- (5)  $\nabla^{+t} \underline{p}^+ = \nabla^{+t}(\nabla^+ + \beta \cdot \underline{p}) = \nabla^{+t} \nabla^+ + \beta \cdot \nabla^{+t} \underline{p} = \nabla^{+t} \nabla^+ + \beta \cdot (\nabla + 2 \cdot \gamma \cdot \underline{S} \underline{p})^t \underline{p} = \nabla^{+t} \nabla^+$
- (6)  $\underline{p}^t \underline{S}(\nabla^+ + \beta \cdot \underline{p}) = 0 = \underline{p}^t \underline{S} \underline{p}^+ \quad (\beta \text{ wurde ja so bestimmt!})$  daher:
- (7)  $\underline{p}^{+t} \underline{S} \underline{p}^+ = (\nabla^+ + \beta \cdot \underline{p})^t \underline{S} \underline{p}^+ = \nabla^{+t} \underline{S} \underline{p}^+$  wegen (6)
- (8)  $\nabla^t \nabla^+ = \nabla^t (\nabla + 2 \cdot \gamma \cdot \underline{S} \underline{p}) = \nabla^t \nabla + 2 \cdot \gamma \cdot \nabla^t \underline{S} \underline{p} = \nabla^t \nabla - (\nabla^t \underline{p} / \underline{p}^t \underline{S} \underline{p}) \cdot \nabla^t \underline{S} \underline{p} = 0$  wegen (5) und (7)
- (9)  $\nabla^+ \cdot \underline{p} = 0$ , siehe Satz (4.1.5)

## 4.3 Der CG – Algorithmus

**Problemstellung(4.1):**  $D \subset \mathbb{R}^n$  sei eine konvexe Teilmenge des  $\mathbb{R}^n$  und gegeben sei eine zweimal stetig differenzierbare Funktion  $f: D \rightarrow \mathbb{R}$ ,  $\underline{x} \rightarrow f(\underline{x})$ . **Aufgabe:** Gesucht ist  $\underline{x}^* \in D$  mit:  $f(\underline{x}^*) = \text{Min } f(\underline{x})$  auf  $D$ , wobei  $\underline{x}^*$  nicht auf dem Rand von  $D$  liege. Hier Voraussetzung: Es existiert die Minimumstelle  $\underline{x}^*$ , also  $(\nabla f)(\underline{x}^*) = \underline{0}$ .

**Beispiel (4.2), quadratisches Polynom:** Sei  $\underline{S}$  eine positiv definite Matrix, also  $\underline{x}^t \underline{S} \underline{x} > 0$  für  $\underline{x} \neq \underline{0}$ , und  $\underline{a} \in \mathbb{R}^n$  und  $b \in \mathbb{R}$ . Damit definieren wir die Funktion  $f_s(\underline{x}) = \underline{x}^t \underline{S} \underline{x} + \underline{a}^t \underline{x} + b$ . Wir erhalten:  $(\nabla f_s)(\underline{x}) = 2 \cdot \underline{S} \underline{x} + \underline{a}$  und damit ergibt sich  $\underline{x}^* = -0,5 \cdot \underline{S}^{-1} \underline{a}$

Wir geben nun einen Algorithmus an, der sich der Lösung  $\underline{x}^*$  iterativ nähert.

**Start:** Wähle einen geeigneten Startvektor  $\underline{x}^0$  und berechne  $\nabla^0 := (\nabla f)(\underline{x}^0)$

- (4.2):  $\nabla^0 := (\nabla f_s)(\underline{x}^0) = 2 \cdot \underline{S} \underline{x}^0 + \underline{a}$

Setze  $\underline{p}^0 = \nabla^0$  und  $k = 0$  und wähle  $\varepsilon > 0$  hinreichend klein, z.B.  $\varepsilon = 1e-7$

**Schritt 1:** Falls  $\nabla^k = 0$  oder  $\nabla^k < \varepsilon$ : stoppe

**Schritt 2:** Berechne gemäß (4.1.3) die Schrittweite  $\gamma^k = z_{\text{Min}}^k / \|\underline{p}^k\|$

- (4.2): Es gilt dann:  $\gamma^k = -(2 \cdot \underline{x}^{kt} \underline{S} \underline{p}^k + \underline{a}^t \underline{p}^k) / (2 \cdot \underline{p}^{kt} \underline{S} \underline{p}^k) = -\nabla^{kt} \underline{p}^k / (2 \cdot \underline{p}^{kt} \underline{S} \underline{p}^k)$

Setze:  $\underline{x}^{k+1} = \underline{x}^k + \gamma^k \cdot \underline{p}^k$

**Schritt 3:** Berechne  $\nabla^{k+1} = (\nabla f)(\underline{x}^{k+1})$

- (4.2)  $\nabla^{k+1} = 2 \cdot \underline{S}(\underline{x}^k + \gamma^k \cdot \underline{p}^k) + \underline{a} = \nabla^k + 2 \cdot \gamma^k \cdot \underline{S} \underline{p}^k$

**Schritt 4:** Berechne  $\beta^k = -(\nabla^{k+1} - \nabla^k)^t \nabla^{k+1} / (\nabla^{k+1} - \nabla^k)^t \underline{p}^k$

Setze:  $\underline{p}^{k+1} = \nabla^{k+1} + \beta^k \cdot \underline{p}^k$

**Schritt 5:** Setze  $k = k + 1$  und gehe nach Schritt 1

**Bemerkungen zu Schritt 2:** Wird  $\gamma^k$  wie beschrieben bestimmt, so hat man im Beispiel (4.2) schöne Konvergenzeigenschaften, wie in Abschnitt 4.5 gezeigt wird. Die Bestimmung von  $\gamma^k$  kann aber auch modifiziert werden, wobei zur Bestimmung ja auch ein  $\Delta z$  erforderlich ist.

#### 4.4 Genauigkeitsfragen – Abbruchkriterium

Grundgedanke: Sei für  $\underline{x}_0$   $f_0 := f(\underline{x}_0)$  das tatsächliche Minimum genau bekannt. Wir betrachten die beliebige Suchrichtung  $\underline{p}$  mit  $\underline{e}_p = (1/\|\underline{p}\|) \cdot \underline{p}$  und dazu  $f_p(z) = f(\underline{x}_0 + z \cdot \underline{e}_p)$ . Dann gilt:

- $f_p(z)$  verhält sich in der Nähe von 0 fast genau wie ein quadratisches Polynom. (4.3.1)

Daher ist es sinnvoll, einen vernünftigen Polynomansatz  $h(z)$  zu untersuchen, nämlich:

- (1) Mit  $q > 0$  sei  $h(0) = -q$ ,  $h'(0) = 0$ ,  $h(q) = 0$ , also:
- (2)  $h(z) = -q + (1/q) \cdot z^2$ ,  $h'(z) = (2/q) \cdot z$
- (3) für ein hinreichend kleines  $\varepsilon > 0$  fordern wir:  $\Delta z > 0$  sei Lösung, falls  $h(\Delta z) - h(0) \leq \varepsilon^2$
- (4)  $\leftrightarrow \Delta z \leq \sqrt{(q) \cdot \varepsilon}$  oder grenzwertig:  $\Delta z = \sqrt{(q) \cdot \varepsilon}$  (4.3.2)

Sei nun  $\delta$  die bequem erreichbare Auflösung, also bei double  $1e-14$  mit einer gewissen Rechenreserve. Für  $\Delta z = \sqrt{(q) \cdot \varepsilon}$  haben wir:

- (1)  $h(\Delta z) = -q + \varepsilon^2$  bzw..  $h(\Delta z)/q = -1 + \varepsilon^2/q$ , also:
- (2)  $\varepsilon^2/q \geq \delta$ ; grenzwertig setzen wir:  $\varepsilon^2 = \delta \cdot q$ ; mit  $\Delta z = \sqrt{(q) \cdot \varepsilon}$  haben wir:
- (3)  $\Delta z = q \cdot \sqrt{(\delta)}$ ; eingesetzt in (4.3.2) ergibt sich:
- (4)  $h(\Delta z) = q \cdot (-1 + \delta)$  und
- (5)  $h'(\Delta z) = (2/q) \cdot q \cdot \sqrt{(\delta)} = 2 \cdot \sqrt{(\delta)}$  (4.3.2)

Bemerkenswert ist, dass in (4.3.2)(5) sich  $q$  wegekürzt, was sofort zum allgemein gültigen Abbruchkriterium führt:

#### Abbruchkriterium nach (4.2.1)(8):

- Falls  $\|\nabla^+\| < 2 \cdot \sqrt{(\delta)}$  Abbruch der Iteration erforderlich (4.3.3)

#### 4.5 Beweis zum Beispiel (4.2)

Für unser Beispiel (4.2) liefert das CG – Verfahren, in exakter Arithmetik, nach spätestens  $n$  Schritten die exakte Lösung. Dies beweisen wir in diesem Abschnitt!

##### (4.5.1) Was wir schon gezeigt haben:

- (1)  $(\nabla^{k+1})^t \nabla^k = 0$  siehe (4.2.7)(8)

(2)  $(\nabla^{k+1})^t \underline{p}^k = 0$  siehe Satz (4.1.5), also wegen perfekter Schrittweite

(3)  $(\underline{p}^{k+1})^t S \underline{p}^k = 0 \leftarrow \beta^k$  wurde so konstruiert

(4)  $(\nabla^{k+1})^t \underline{p}^{k+1} = (\nabla^{k+1})^t \nabla^{k+1} > 0$ , andernfalls bereits Lösung, siehe (4.2.7)(5)

(5)  $(\nabla^+ - \nabla) \cdot (1 / (2 \cdot \gamma)) = S \underline{p}$ , siehe (4.2.7)(3)

Wir beweisen nun den folgenden Satz:

### Satz (4.5.2)

Mit exakter Arithmetik wurde berechnet gemäß CG – Algorithmus 4.3 berechnet:

➤  $\nabla^i \neq 0$ ,  $i = 0, 1, \dots, k \leq n$ ,  $\nabla^i = 0 \rightarrow$  wir haben bereits die Lösung

Dann gilt für  $k = 1, 2, \dots$ :

I.  $(\nabla^k)^t \underline{p}^k > 0$

II.  $(\nabla^k)^t \underline{p}^j = 0$  für  $j = 0, 1, \dots, k-1$

III.  $(\nabla^k)^t \nabla^j = 0$  für  $j = 0, 1, \dots, k-1$

IV.  $(\underline{p}^k)^t S \underline{p}^j = 0$  für  $j = 0, 1, \dots, k-1$

Vollständige Induktion nach  $k$

### Schritt $k = 1$ :

Wir haben  $\underline{p}^0 = \nabla^0 = 2 \cdot S \underline{x}^0 + \underline{a}$  mit  $\nabla^0 \neq 0$  (andernfalls schon Lösung!), also  $(\nabla^0)^t \underline{p}^0 > 0$ .

$(\nabla^1)^t \underline{p}^1 > 0$  wegen (4.5.1)(4)

$(\nabla^1)^t \underline{p}^0 = 0$  wegen (4.5.1)(2)

$(\nabla^1)^t \nabla^0 = 0$  wegen (4.5.1)(1)

$(\underline{p}^1)^t S \underline{p}^0 = 0$  wegen (4.5.1)(3)

### Nun Induktion von $k$ auf $k+1$ , also bis $k$ alles bewiesen:

I.) folgt direkt aus (4.5.1)(4)

II.)  $(\nabla^{k+1})^t \underline{p}^k = 0$  wegen (4.5.1)(2);  $(\nabla^{k+1})^t \underline{p}^j = (\nabla^k + 2 \cdot \gamma^k \cdot S \underline{p}^k)^t \underline{p}^j = 0$  für  $j \leq k-1$  wegen Induktionsvoraussetzung

III.)  $(\nabla^{k+1})^t \nabla^k = 0$  wegen (4.5.1)(1);  $(\nabla^{k+1})^t \nabla^j = (\nabla^k + 2 \cdot \gamma^k \cdot S \underline{p}^k)^t \nabla^j = (\nabla^k)^t \nabla^j + 2 \cdot \gamma^k \cdot (\underline{p}^k)^t S \nabla^j = (\nabla^k)^t \nabla^j + 2 \cdot \gamma^k \cdot (\underline{p}^k)^t S (\underline{p}^j - \beta^{j-1} \cdot \underline{p}^{j-1}) = 0$  für  $j \leq k-1$  wegen Induktionsvoraussetzung

IV.)  $(\underline{p}^{k+1})^t S \underline{p}^k = 0$  wegen (4.5.1)(3);  $(\underline{p}^{k+1})^t S \underline{p}^j = (\nabla^{k+1} + \beta^k \cdot \underline{p}^k)^t S \underline{p}^j = (\nabla^{k+1})^t S \underline{p}^j$ ; wegen (4.5.1)(5) gilt:  $(\nabla^{k+1})^t S \underline{p}^j = (\nabla^{k+1})^t (\nabla^{j+1} - \nabla^j) \cdot (1 / (2 \cdot \gamma^j)) = 0$  wegen III.)

**q.e.d.**

**Korollar:** Spätestens für  $k = n$  gilt  $\nabla^k = \underline{0}$ , da nur  $n$  Vektoren linear unabhängig sein können.

## 5 Literatur

/1/ bzw. „Buch“: Arens, T., Hettlich, F., Karpfinger, C., Kockelkorn, U., Lichtenegger, K., Stachel, H.: Mathematik (2013)