Das gedämpfte Newton-Verfahren

Das gedämpfte NT-Verfahren

- Die gedaempfte NT - Iteration

$$F'\left(x^k\right)\Delta x^k = -F\left(x^k\right),\ x^{k+1} = x^k + \lambda_k \Delta_k,\ \lambda_k \in \left]0,1\right]$$

- Ziel

Schwierigkeiten abzuheben, wenn die Jacobi - Matrizen schlecht - konditioniert sind.

Allgemeine Niveaufunktion

- Erinnerung:

  Allgemeine Niveaufunktion (general level function)

  $$T(x|A) := \frac{1}{2}\|AF(x)\|_2^2$$

  A beliebig, nicht singulär (n×n)- Matrix

- Lokaler Abstieg

  Die Newton-Richtung geht „abwärts“ bzgl. allen solchen Niveaufunktionen:

Allgemeine Niveaufunktion

Lemma 3.11

Sei $F \in c^1(D)$ und sei Δx die gewöhnliche Newton-Korrektur (die Iterationsindex k wird weggelassen), dann für alle $A \in Gl(n)$ gilt:

$$\Delta x^T \, grad \; T(x|A) = -2T(x|A) < 0.$$

Beweis :

$$\Delta x^T \, gradT \; (x|A) = \Delta x^T \cdot \left((AF'(x))^T \, AF(x) \right)$$

$$= \underbrace{\Delta x^T F'(x)^T}_{-F(x)^T} A^T AF(x)$$

$$= -2\,T(x|A) < 0$$

Allgemeine Niveaufunktion

Erinnerung :

Niveaumenge

$$G(z) := \{x \in D \subseteq \Re^n | T(x) \leq T(z)\}$$

$$G(z|A) := \{x \in D \subseteq \Re^n | AT(x) \leq AT(z)\}$$

Satz 3.12

$F \in C^1(D)$ mit $D \subset \Re^n$ konvex und $F'(x)$ nicht singulaer fuer alle $x \in D$. Fuer einen gegebenen Schritt $x^k \in D$ sei $G(x^k|A) \subset D$ fuer $A \in GL(D)$. Fuer $x, y \in D$ nehmen wir an, dass

$$\left\| F'(x)^{-1}(F'(y) - F'(x))(y - x) \right\| \le \omega \|y - x\|^2$$

Dann mit der Notation

$$h_k := \|\Delta x^k\| \omega, \quad \bar{h}_k := h_k cond\left(AF'(x^k)\right)$$

erhaelt man

$$\left\| AF\left(x + \lambda \Delta x^k\right)\right\| \le t_k(\lambda|A) \left\| AF\left(x^k\right)\right\|$$

fuer $\lambda \in \left[0, \min\left(1, 2/\bar{h}_k\right)\right]$, wobei

$$t_k(\lambda|A) := 1 - \lambda + \frac{1}{2}\lambda^2 \bar{h}_k$$

Die optimale Auswahl des Daempfungsfaktors bzgl. dieser Annahme nist

$$\bar{\lambda}_k(A) := \min\left(1, 1/\overline{h_k}\right)$$

Beweis

$$\left\| AF(x+\lambda\Delta x) \right\| = \left\| A\left(F(x+\lambda\Delta x) - F(x) - F'(x)\Delta x\right) \right\|$$

$$= \left\| A\left(\overbrace{\int_{\delta=0}^{\lambda} F'(x+\delta\Delta x)\Delta d\delta}^{\text{Hauptsatz d. Intergralr echnung}} - F'(x)\Delta x \right) \right\|$$

$$= \left\| A\left(\int_{\delta=0}^{\lambda} \left(F'(x+\delta\Delta x) - F'(x)\right)\Delta x d\delta - (1-\lambda)F'(x)\Delta x \right) \right\|$$

$$\leq (1-\lambda)\left\| AF(x) \right\| + O\left(\lambda^2\right) \text{ fuer } \lambda \in [0,1]$$

$$O\left(\lambda^2\right) = \left\| AF'(x)\left(\int_{\delta=0}^{\lambda} \underbrace{F'(x)^{-1}\left(F'(x+\delta\Delta x) - F'(x)\right)\Delta x}_{\leq \omega\delta \left\|\Delta x\right\|^2} d\delta \right) \right\|$$

$$\leq \left\| AF'(x) \right\| \omega \frac{1}{2}\lambda^2 \left\| \Delta x \right\|^2 = \frac{1}{2}\lambda^2 \left\| AF'(x) \right\| \cdot \underbrace{\omega \left\| \Delta x \right\|}_{h_k} \cdot \left\| \Delta x \right\|$$

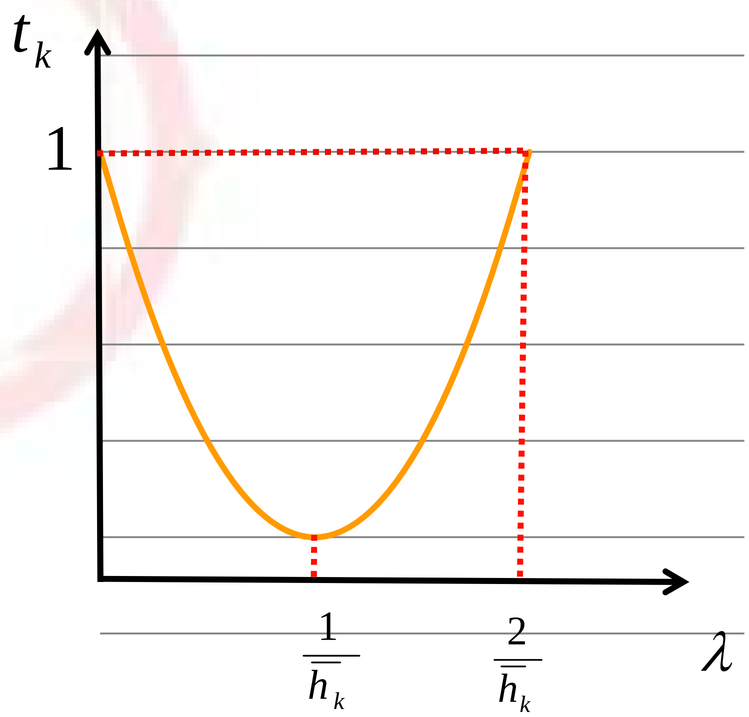
$$= \frac{1}{2}\lambda^2 \left\| AF'(x) \right\| h_k \cdot \left\| \left(AF'(x)\right)^{-1} AF(x) \right\|$$

$$\leq \frac{1}{2}\lambda^2 h_k \underbrace{\left\| AF'(x) \right\| \cdot \left\| \left(AF'(x)\right)^{-1} \right\|}_{cond\left(AF'(x)\right)} \cdot \left\| AF(x) \right\| = \frac{1}{2}\lambda^2 \overline{h}_k \left\| AF(x) \right\|$$

$$\Rightarrow \left\| AF\left(x+\lambda \Delta x^{k}\right)\right\| \leq t_{k}\left(\lambda | A\right)\left\| AF\left(x^{k}\right)\right\|$$

wobei $t_{k}\left(\lambda | A\right)=1-\lambda+\frac{1}{2}\lambda^{2}\overline{h}_{k}$

$$0 \leq t_{k} \leq 1 \Rightarrow 0 \leq \lambda \leq \frac{2}{\overline{\overline{h}}_{k}}$$



Globale Konvergenz

Aus dem vorigen Satz können wir jetzt die globale Konvergenz ableiten:

Satz 3.13

Notation und Annahmen wie in Sazt 3.12, weiter sei D_0 die zusammenhængende Komponente von $G(x^0|A)$ in x^0 und sei $D_0 \in D$ kompakt. Ausserdem wird angenommen, dass die Jacobi-Matrix $F'(x)$ nicht singulaer ist fuer alle $x \in D_0$. Dann die gedaempfte Newton-Iteration $(k = 0, 1, \ldots)$ mit Daempfungsfaktor im Intervall

$$\lambda_k \in \left[\varepsilon, 2\overline{\lambda}_k(A) - \varepsilon\right]$$

und genuegende kleine $\varepsilon > 0$, welche von D_0 abhaengt, konvergiert gegen den Loesungspunkt x^*.

Beweisskizze (per Induktion)

die Parabel $t_k(\lambda|A)$ hat die globale obere Schranke :

$$t_k(\lambda|A) \leq 1 - \frac{1}{2}\varepsilon,\ \ 0 < \varepsilon \leq \frac{1}{\overline{h}_k},\ \ (3.39)$$

dies indukziert strenge Reduktion der allgemeinen Niveaufunktion $T(x|A)$ in jedem Schritt k. Wir bleiben noch zu zeigen, ob eine globale ε existiert. Mit anderen Worten, ob $\overline{h}_k$ beschraenkt ist.

Wegen der Kompaktheit von D_0 gilt :

$$\max_{x \in D_0} \left\| F'(x)^{-1} F(x) \right\| \cdot cond_2(AF'(x)) < \infty$$

Das garantiert die Existenz von einer globalen ε.

Wann immer wir $G(x^k|A) \subseteq D_0$ haben, (3.39) garantiert, dass

$$G(x^{k+1}(\lambda)|A) \subset G(x^k|A) \subseteq D_0$$

Daher schliessen wir mit Induktion ab, dass

$$\lim_{k \to \infty} x^k = x^*.$$

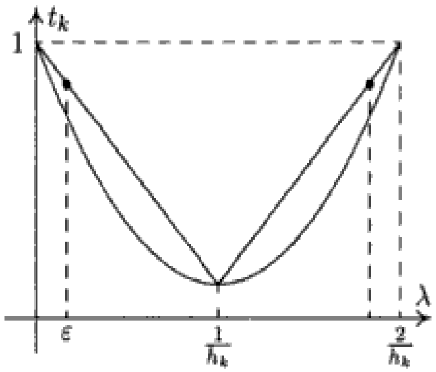


Fig. 3.5. Local reduction parabola t_k together with polygonal upper bounds.

Algorithmische Einschränkung der Residuum-Monotonie

Trotz bewiesener Konvergenzeigenschaft das gedämpfte Newton-Verfahren, das mit dem traditionellen Residuum-Monotonietest

$$T\left(x^{k+1} \middle| I\right) \leq T\left(x^{k} \middle| I\right)$$

ausgestattet, kann bei praktischer Rechnung scheitern.

Warum?

$$\overline{\lambda}_k(I) = \left(h_k \, cond_2\left(F'\left(x^k\right)\right)\right)^{-1} < \lambda_{\min} << 1$$

$$\Rightarrow \quad x^{k+1} \approx x^k$$

Iteration beleibt „stehen“

Natürliche Niveaufunktion

$$cond_2\left(AF'\left(x^k\right)\right) \geq 1 = cond_2(I)$$

=> $A_k := F'\left(x^k\right)^{-1}$ „lokal optimal“ in Allgemeiner Niveaufunktion

$T\left(x \middle| F'\left(x^k\right)^{-1}\right)$ „natürliche Niveaufunktion“

Natürliche Niveaufunktion

1) Extremale Eigenschaften

Für $A \in GL(n)$ gilt:

Reduktionsfaktor:

$$t_k(\lambda | A_k) = 1 - \lambda + \frac{1}{2}\lambda^2 h_k \leq t_k(\lambda | A) \quad \text{und}$$

lokal optimaler Daempfungsfaktor:

$$\overline{\lambda}_k(A_k) = \min\left(1, \frac{1}{h_k}\right) \geq \overline{\lambda}_k(A)$$

Natürliche Niveaufunktion

2) Steilste Abstiegseigenschaft

Die steilste Abstiegsrichtung für $T(x|A)$ in x^k ist $-\,gradT\left(x^k|A\right) = -\left(AF'\left(x^k\right)\right)^T AF\left(x^k\right)$.

Mit $A = A_k := F'\left(x^k\right)^{-1}$ führt dies zu

$$\Delta x^k = -\,gradT\left(x^k|A_k\right).$$

=> das gedämpfte NT-Verfahren ist ein Gradientenverfahren für die natürliche Niveaufunktion $T(x|A_k)$.

Natürliche Niveaufunktion

3) Fusionseigenschaft

$$h_k \leq 1 \Rightarrow \overline{\lambda}_k \left(A_k\right) = 1$$

die quadratische Konvergenz ist asymptotisch erreicht

Natürliche Niveaufunktion

4) Asymptotische Distanzfunktion

Fuer $F \in C^2(D)$, man hat

$$T\left(x \middle| F'(x^*)^{-1}\right) = \frac{1}{2}\left\|x - x^*\right\|_2^2 + O\left(\left\|x - x^*\right\|^3\right)$$

$\Rightarrow$ Fuer $x^k \to x^*$, das natuerliche Monotonie Kriterium geht asymptotisch in die Distanzfunktion ueber :

$$\left\|x^{k+1} - x^*\right\|_2 \stackrel{\cdot}{\leq} \left\|x^k - x^*\right\|_2$$

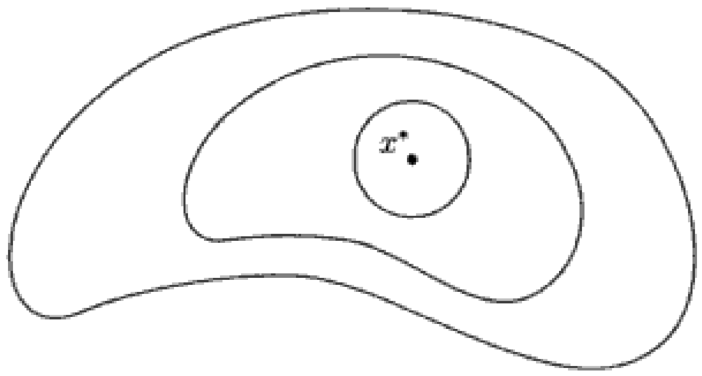


Fig. 3.7. Natural level sets: asymptotic distance spheres.

Globale Konvergenz

- $A_k = F'(x_k)^{-1}$ variiert in jedem Schritt

$\Rightarrow$ Satz 3.13 ist nicht anwendbar

- affin - kovariante gloabale Konvergenz

=> fixe Wahl $A = F'(x^*)^{-1}$

Satz 3.14

Sei $F : D \to \Re^n$ eine stetig differenzierbare Abbildung mit $D \subseteq \Re^n$ offen und konvex. Angenommen, dass $x^0, x^* \in D$ und x^* die eindeutige Loesung von F in D ist, die Jacobi-Matrix $F'(x^*)$ nicht singulaer ist. Weiter nehmen wir an, dass

(1) $F'(x)$ ist nicht singulaer fuer alle $x \in D$,

(2) die Zusammenhaengende Komponente D_0 von

$$G\left(x^0 \middle| F'(x^*)^{-1}\right) \text{ in } x^0 \text{ kompakt und in } D \text{ enthalten ist,}$$

(3) die folgende affin-kovariante Lipschitzbedingung ist erfuellt:

$$\left\| F'(x^*)^{-1}(F'(y) - F'(x))(y - x) \right\| \le \omega_* \|y - x\|^2 \text{ fuer } y, x \in D,$$

(4) fuer jeden Schritt $x^k \in D$ sei $h_k^* := \omega_* \|\Delta x^k\| \cdot \left\| F'(x^k)^{-1} F'(x^*) \right\| < \infty.$

Als das lokal optimale Daempfungsfaktor erhalten wir

$$\lambda_k^* := \min(1, 1/h_k^*).$$

Dann konvergiert jede gedaempfte NT-Iteration mit iterativem Daempfungsfaktor λ_k gegen x^*, wobei

$$\lambda_k \in [\varepsilon, 2\lambda_k^* - \varepsilon] \text{ fuer } 0 < \varepsilon < 1/h_k^*.$$

Globale Konvergenz

Man kann zeigen, dass die lokal optimale Dämpfungsfaktoren bzgl. der lokal definierten natürlichen Niveaufunktion außerordentlich sind unter allen möglichen globalen optimalen Dämpfungsfaktoren bzgl. jeder globalen definierten affin-kovarianten Niveaufunktion.

Theoretische optimale Dämpfungsstrategie für das exakte NT-Verfahren

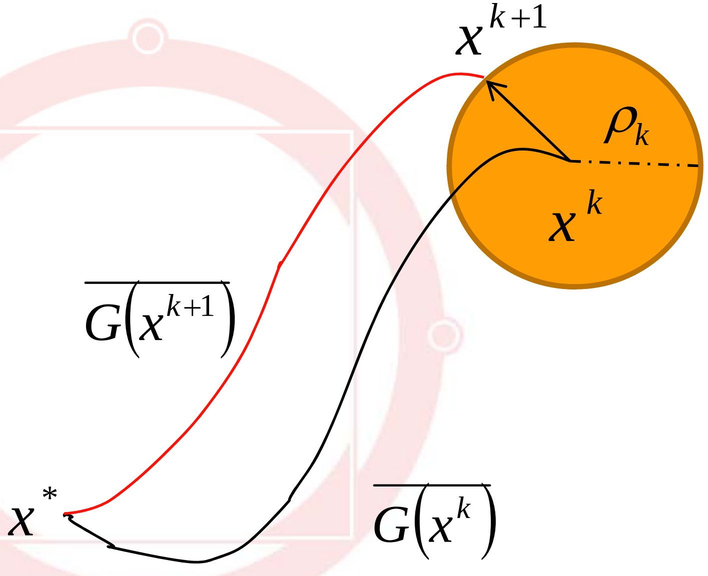
Unsere theoretische Konvergenzanalysis zeigt, dass die globale affin-kovariante Lipschitzkonstante ω durch ihr mehr lokaleres Duplikat $\omega_k = \omega(x^k)$ ersetzt werden muss:

$$\left\| F'(x^k)^{-1}\left(F'(x) - F'(x^k)\right)(x - x^k)\right\| \le \omega_k \left\|x - x^k\right\|^2 \text{ fuer } x, x^k \in D_0.$$

Wir haben dann die folgende theoretische optimale Daempfungsstrategie fuer das exakte NT-Verfahren:

$$x^{k+1} = x^k + \overline{\lambda}_k \Delta x_k,\ \overline{\lambda}_k := \min(1, 1/h_k),\ h_k = \omega_k \left\|\Delta x^k\right\|$$

Geometrische Interpretation



$$\left\| x^{k+1} - x^k \right\| = \overline{\lambda}_k \left\| \Delta x^k \right\| \le \rho_k := 1/\omega_k$$

Interpretation durch Jacobi-Information

In Form von der relativen Veraenderung

der Jacobi - Matrix koennen wir schreiben :

$$\left\| F'\left(x^{k+1}\right) - F'\left(x^{k}\right) \right\| \Big/ \left\| F'\left(x^{k}\right) \right\|$$

$$= \underbrace{\left\| F'\left(x^{k}\right)^{-1} \left(F'\left(x^{k+1}\right) - F'\left(x^{k}\right) \right) \overline{\lambda}_k \Delta x^k \right\|}_{\leq \omega_k \left\| \overline{\lambda}_k \Delta x^k \right\|^2} \Big/ \left\| \overline{\lambda}_k \Delta x^k \right\|$$

$$\leq \omega_k \left\| \overline{\lambda}_k \Delta x^k \right\| = \overline{\lambda}_k h_k \leq 1$$

=> Jacobi - Information gueltig entlang der NT - Richtung bis zum Rand des "Vertrauenkreis" (Trust region ball).

Ausserhalb des "Vertrauenkreis" muss ein neuer NT - Pfad konstruiert werden.

Verhalten beim kritischen Punkt

singulaer Jacobi - Matrix $F'(\hat{x})$

$\Rightarrow$ NT - Pfad & gedaempfte NT - Iteration mit optimaler. Daempfung werden "angelockt".

In Abb., 2 Loesungspunkte(x^* postiv, x^{**} negativ) sind getrennt durch die Mannigfaltigkeit mit singulaerer J - Matrix $F'(\hat{x})$.

Residuum Monotonietest : $T(\hat{x}|I)$ Lokalminimum $\Rightarrow$ Iteration "terminiert"

Natuerliche Niveaufunktion : $T\left(\hat{x}\middle|F'(\hat{x})^{-1}\right)$ Lokalmaximum

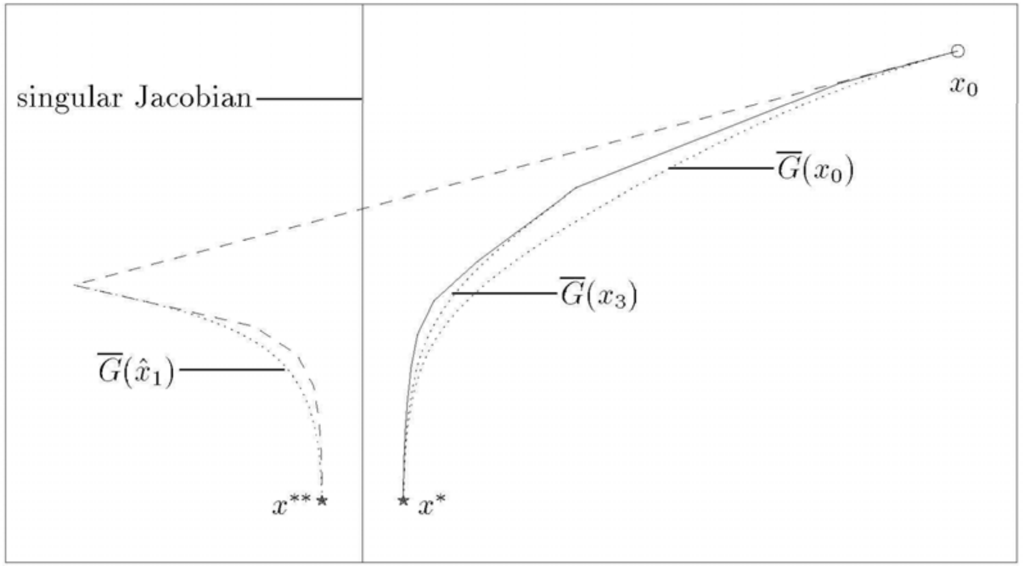


Figure 1.1 Newton path and Newton iterates