4c. Newton-artige Verfahren

Ausgleichsprobleme

Optimierung SoSe 2020

Dr. Alexey Agaltsov

Plan

- Lineares Ausgleichsproblem
- Nichtlineares Ausgleichsproblem
- Konvergenz des Gauß-Newton-Verfahrens

Lineares Ausgleichsproblem

Minimiere $f(x) = \frac{1}{2}\|Ax - y\|_2^2$ über $x \in \mathbb{R}^n$

$A \in \mathbb{R}^{m \times n}, \; y \in \mathbb{R}^m$

a_i i-te Zeile von A

$y = (y_1, \dots, y_m)$

Minimiere $f(x) = \frac{1}{2}\sum_{i=1}^{m}\left(a_i^T x - y_i\right)^2$ über $x \in \mathbb{R}^n$

- x_* ist eine optimale Lösung $\Leftrightarrow$ Ax_* ist die Projektion von y auf im A
- Der Optimalwert ist gleich $0 \Leftrightarrow y \in \operatorname{im} A$

Existenz von Lösungen

$$f(x) = \frac{1}{2}||Ax - b||_2^2$$
$$= \frac{1}{2}x^T A^T Ax - x^T A^T b + \frac{1}{2}b^T b$$

- f ist quadratisch und f ist beschränkt nach unten $(f \geq 0)$
- Es gibt eine optimale Lösung x_*

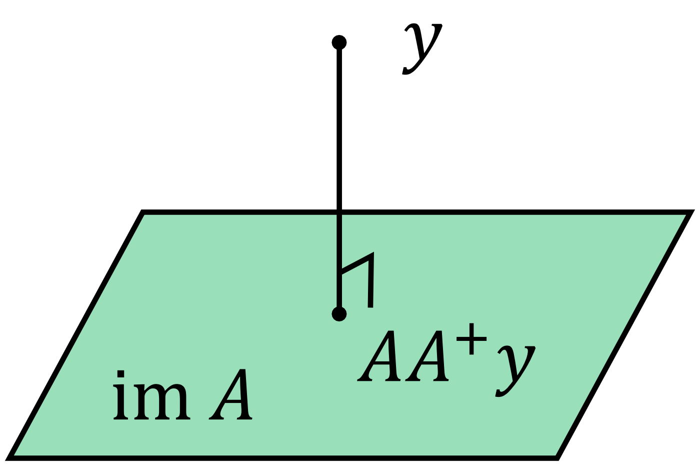
Aufgabe 2.21

- x_* ist eine optimale Lösung $\Leftrightarrow$ x_* erfüllt die Normalgleichung

$$A^T Ax_* = A^T b \qquad \text{(Aufgabe 2.21)}$$

Die Pseudo-Inverse

$$A^T A x_* = A^T y$$



- Sei $\ker A = \{0\}$, sodass $A^T A \succ 0$

$$x_* = \underbrace{(A^T A)^{-1} A^T}_{A^+} y$$

- A^+ heißt die (Moore-Penrose) Pseudo-Inverse von A

$$A^+ y = \mathrm{argmin}_x \{\|Ax - y\|_2\}$$

Statistische Interpretation

- Lineares Messmodel:

a_i i-te Zeile von A

$y = (y_1, \dots, y_m)$

$\varepsilon = (\varepsilon_1, \dots, \varepsilon_m)$

$$\underbrace{y_i}_{\text{Messwerte}} = \overbrace{a_i}^{\text{bekannt}}{}^T \overbrace{x}^{\text{unbekannte Parameterwerte}} + \underbrace{\varepsilon_i}_{\text{Messfehler}} \qquad i = 1, \dots, m$$

$$y = Ax + \varepsilon$$

- Bestimme x

Distribution der Messfehler

- Seien ε_i unabhängige Zufallsvariablen von $N(0, \sigma^2)$:

$$„\underbrace{\mathbb{P}(\varepsilon_1 = t_1, \dots, \varepsilon_m = t_m)}_{\textit{Wahrscheinlichkeitsdichtefunktion}}“ = \frac{1}{(2\pi\sigma^2)^{m/2}} \prod_{i=1}^{m} \exp\left(-\frac{t_i^2}{2\sigma^2}\right)$$

$$= \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{\|t\|_2^2}{2\sigma^2}\right)$$

$$t = (t_1, \dots, t_m)$$

- Der Vektor ε hat die mehrdimensionale Normalverteilung $N(0, \Sigma)$

$$\Sigma = \mathrm{diag}(\sigma^2, \cdots, \sigma^2)$$

Maximum-Likelihood-Methode

- Für welches x ist die Realisierung $\varepsilon = y - Ax$ von $N(0, \Sigma)$ möglichst wahrscheinlich?

Maximiere „$\mathbb{P}(\varepsilon = y - Ax)$" über $x \in \mathbb{R}^n$

Minimiere $-\log($„$\mathbb{P}(\varepsilon = y - Ax)$"$)$ über $x \in \mathbb{R}^n$

„$\mathbb{P}$" $= \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{\|t\|_2^2}{2\sigma^2}\right)$

Minimiere $\frac{1}{2\sigma^2}\|Ax - y\|_2^2$ über $x \in \mathbb{R}^n$

Beispiel: Lineare Regression

- $R_i \in \mathbb{R}$ Rendite des Investmentfonds PRSVX für die Woche $i = 1, \dots, m$

  Von 15/01/2010 bis 27/12/2013

- $R_i^M \in \mathbb{R}$ die Marktrendite für die Woche i

- Gibt es einen Zusammenhang zwischen R_i und R_i^M?

Lineare Regression

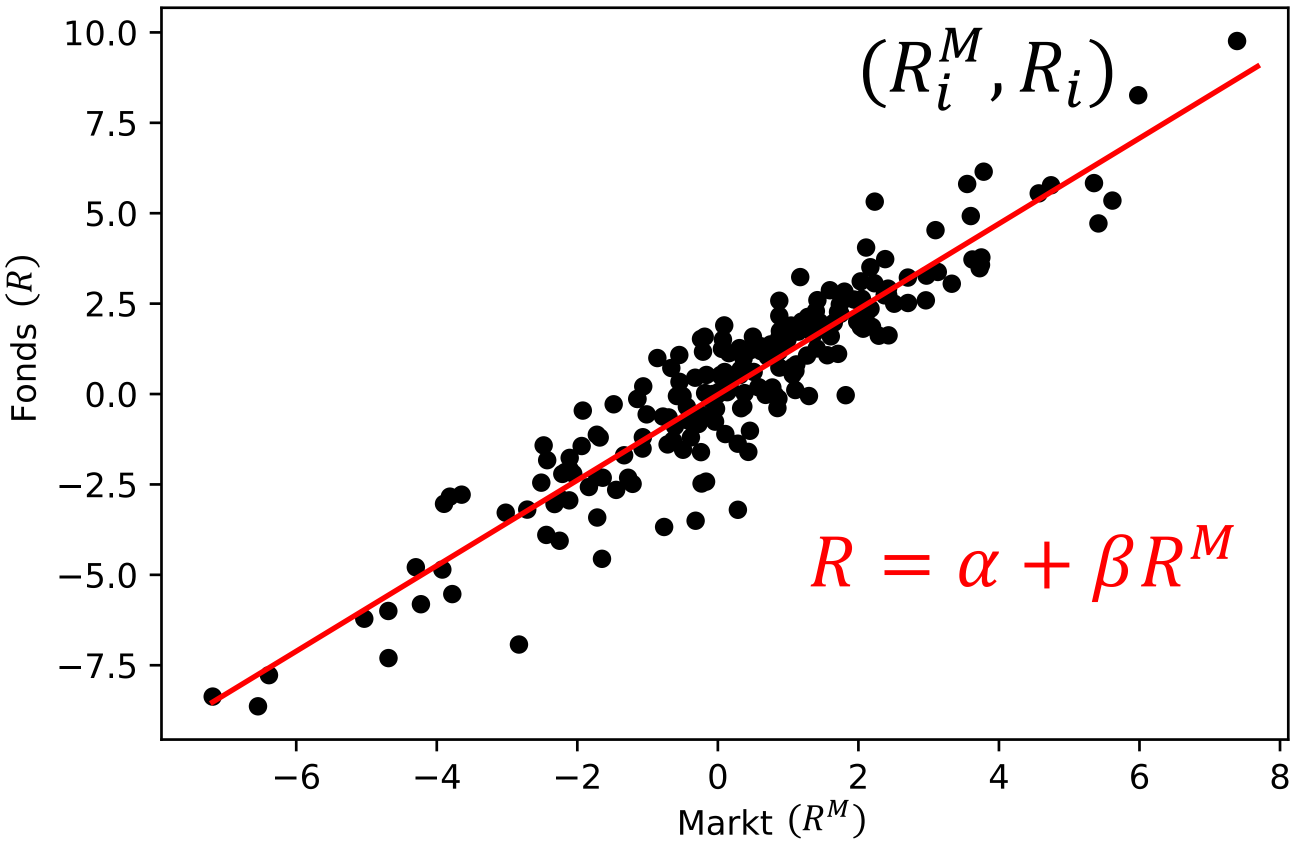
$$R_i = \alpha + \beta R_i^M + \underbrace{\varepsilon_i}_{\text{Rauschen}}$$

$$y = (R_1, \dots, R_m)^T$$

$$x = (\alpha, \beta)^T$$

$$A = \begin{bmatrix} 1 & R_1^M \\ \vdots & \vdots \\ 1 & R_m^M \end{bmatrix}$$

$$y = Ax + \varepsilon$$



Wöchentliche Renditen

Aufgabe 4.6. Lineare Regression

Minimiere $f(\alpha,\beta) = \frac{1}{2}\sum_{i=1}^{m}\left(R_i - \alpha - \beta R_i^M\right)^2$ über $\alpha,\beta \in \mathbb{R}$

- Wir nehmen an, es gibt $i \neq j$ mit $R_i^M \neq R_j^M$
- Zeigen Sie, dass es eine eindeutige optimale Lösung α_*, β_* gibt, sodass:

$$\beta_* = \frac{\sum_{i=1}^{m}(R_i^M - \overline{R^M})(R_i - \bar{R})}{\sum_{i=1}^{m}\left(R_i^M - \overline{R^M}\right)^2} \quad \alpha_* = \bar{R} - \beta_* \overline{R^M}$$

$$\bar{R} = (R_1 + \cdots + R_m)/m, \quad \overline{R^M} = (R_1^M + \cdots + R_m^M)/m$$

Verzerrung-Varianz-Dilemma

$$\text{Minimiere } f(x) = \tfrac{1}{2}||Ax - y||_2^2 \text{ über } x \in \mathbb{R}^n$$

$$\ker A = \{0\},\ x_* = (A^T A)^{-1} A^T y$$

- Seien die Spalten von A fast linear abhängig, so ist $(A^T A)^{-1}$ groß

Die Messfehler in y werden verstärkt

x_* *spiegelt das Rauschen wider (Überanpassung)*

- Finde einen geeigneten Ausgleich zwischen Empfindlichkeit auf Rauschen (*Varianz*) und Modellfehler (*Verzerrung*)

Falsche Beziehung zwischen x und y (Unteranpassung)

Tikhonov-Regularisierung

$$\text{Minimiere } \|Ax - y\|_2^2 + \alpha\|x\|_2^2 \text{ über } x \in \mathbb{R}^n$$

$$A \in \mathbb{R}^{m\times n}, y \in \mathbb{R}^m, \alpha > 0$$

$$\text{Minimiere } \left\|\begin{bmatrix} A \\ \sqrt{\alpha} I \end{bmatrix} x - \begin{bmatrix} y \\ 0 \end{bmatrix}\right\|_2^2 \text{ über } x \in \mathbb{R}^n$$

$$x_* = (A^T A + \alpha I)^{-1} A^T y$$

- Die Kondition $\kappa(A^T A + \alpha I)$ fällt mit α
- Ist $\kappa(A^T A + \alpha I) = N$, so wird der Fehler in $A^T y \approx$ um Faktor N verstärkt

Aufgabe 4.7. Kondition fällt mit α

- Sei $A \in \mathbb{R}^{m\times n}$ mit $\ker A = \{0\}$.
- Singulärwerte von $A \in \mathbb{R}^{m\times n}$ sind die Eigenwerte von $(A^T A)^{1/2}$

  Ist $A \in \mathbb{S}^n_>$ so stimmen die Singulärwerte mit Eigenwerten überein

- Die Kondition von A ist definiert durch:

$$\kappa(A) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)}$$

max. Singulärwert von A

min. Singulärwert von A

- Beweisen Sie die folgende Formel:

$$\kappa(A^T A + \alpha I) = \frac{\sigma_{max}^2(A) + \alpha}{\sigma_{min}^2(A) + \alpha} \qquad \alpha \geq 0$$

Plan

- Lineares Ausgleichsproblem
- Nichtlineares Ausgleichsproblem
- Konvergenz des Gauß-Newton-Verfahrens

Beispiel: Radioaktiver Zerfall

- Zwei radioaktive Substanzen werden in den Boden eingebracht
- y_i Zerfälle pro Minute werden zum Zeitpunkt t_i gemessen, $i = 1, \dots, m$
- Bestimme die Anfangsaktivitäten A_1, A_2 und die Halbwertszeiten T_1, T_2

$$y_i = A_1 2^{-t_i/T_1} + A_2 2^{-t_i/T_2} + \underbrace{\varepsilon_i}_{\text{Messfehler}}, \qquad i = 1, \dots, m$$

Methode der kleinsten Quadrate

$$y_i = A_1 2^{-t_i/T_1} + A_2 2^{-t_i/T_2} + \varepsilon_i, \quad i = 1, \dots, m$$

Bestimme A_1, A_2, T_1, T_2

- Wähle A_1, A_2, T_1, T_2, sodass $\sum_{i=1}^{m} \varepsilon_i^2$ minimal ist

$$F_i(A_1, A_2, T_1, T_2) := A_1 2^{-t_i/T_1} + A_2 2^{-t_i/T_2}$$

Minimiere $\frac{1}{2} \sum_{i=1}^{m} (F_i(A_1, A_2, T_1, T_2) - y_i)^2$

über $A_1, A_2, T_1, T_2 \in \mathbb{R}$

Nichtlineares Ausgleichsproblem

$$\text{Minimiere } f(x) = \frac{1}{2}\|F(x) - y\|_2^2 \text{ über } x \in \mathbb{R}^n$$

$$F \in C^1(\mathbb{R}^n, \mathbb{R}^m),\ y \in \mathbb{R}^m$$

- Sei x_* eine optimale Lösung
- Das Gauß-Newton-Verfahren versucht x_* näherungsweise zu bestimmen
- Die Konvergenz ist lokal quadratisch, falls das Problem kompatibel ist:

$$F(x_*) = y$$

Konvergenzgeschwindigkeit fällt mit $\|F(x_*) - y\|_2$

Gauß-Newton-Verfahren

Zwei Herleitungen:

- Wende das Newton-Verfahren mit einer Approximation an die Hesse-Matrix von f an
- Linearisiere F neben der aktuellen Näherungslösung und löse das resultierende quadratische Program für eine neue Näherungslösung

Ableitungen

$$f(x) = \frac{1}{2}\sum_{i=1}^{m}(F_i(x) - y_i)^2$$

$$\frac{\partial f}{\partial x_i} = \sum_{k=1}^{m}\frac{\partial F_k}{\partial x_i}(F_k(x) - y_k)$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \sum_{k=1}^{m}\left[\frac{\partial^2 F_k}{\partial x_i \partial x_j}(F_k(x) - y_k) + \frac{\partial F_k}{\partial x_i}\frac{\partial F_k}{\partial x_j}\right] \approx \frac{\partial F_k}{\partial x_i}\frac{\partial F_k}{\partial x_j}$$

Ableitungen in Vektorform

$$DF = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix}$$

$$f(x) = \frac{1}{2}(F(x) - y)^T(F(x) - y)$$

$$\nabla f(x) = DF(x)^T(F(x) - y)$$

$$\nabla^2 f(x) \approx DF(x)^T DF(x)$$

Erste Herleitung

$$\text{Minimiere } f(x) = \tfrac{1}{2}\|F(x) - y\|_2^2 \text{ über } x \in \mathbb{R}^n$$

$$F \in C^2(\mathbb{R}^n, \mathbb{R}^m),\; y \in \mathbb{R}^m$$

Newton-Verfahren

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

aktuelle Approximation

$$\nabla f(x_k) = DF(x_k)^T (F(x_k) - y)$$

$$\nabla^2 f(x_k) \approx DF(x_k)^T DF(x_k) \succ 0$$

$$x_{k+1} = x_k - \underbrace{\left(DF(x_k)^T DF(x_k)\right)^{-1} DF(x_k)^T}_{DF(x_k)^+} (F(x_k) - y)$$

Alternative Herleitung

Minimiere $f(x) = \frac{1}{2}\|F(x) - y\|_2^2$

$F \in C^1(\mathbb{R}^n, \mathbb{R}^m),\ y \in \mathbb{R}^m$

linearisiere F

Minimiere $\frac{1}{2}\|F(x_k) + DF(x_k)(x - x_k) - y\|_2^2$

aktuelle Approximation

lineares Ausgleichsproblem

$$DF(x_k)^T DF(x_k)(x_{k+1} - x_k) = -DF(x_k)^T(F(x_k) - y)$$

$\ker DF(x_k) = \{0\}$

$$x_{k+1} = x_k - DF(x_k)^+(F(x_k) - y)$$

Gauß-Newton-Verfahren

1. *Initialisierung:* Startwert $x_0 \in \mathbb{R}^n$, Toleranzwert $\epsilon > 0$
2. **for** $k = 0,1,2,\dots$ **do**:
3. **if** $\|\nabla f(x_k)\|_2 < \epsilon$ **then** break
4. $x_{k+1} = x_k - DF(x_k)^+(F(x_k) - y)$
5. **end for**

$\|DF(x_k)^T(F(x_k) - y)\|_2$

Plan

- Lineares Ausgleichsproblem
- Nichtlineares Ausgleichsproblem
- Konvergenz des Gauß-Newton-Verfahrens

Satz 4.8. Lokale Konvergenz

Sei $F \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ und $x_* \in \operatorname{Argmin}_x \|F(x) - y\|_2^2$. Angenommen:

- $\operatorname{rang}\big(DF(x)\big) = n \;\; \forall x \in B_\delta(x_*) \longrightarrow DF(x)^T DF(x) \succ 0$
- $\left\|DF(x)^+\big(DF(x) - DF(y)\big)\right\|_2 \leq \omega\|x - y\|_2 \; \forall x, y \in B_\delta(x_*)$
- $\|DF(x)^+(F(x_*) - y)\|_2 \leq \kappa\|x - x_*\|_2$ mit $\kappa \in [0,1) \;\; \forall x \in B_\delta(x_*)$

  ist $F(x_*) = y$, so gilt es mit $\kappa = 0$

Ist $\|x_0 - x_*\|_2 < \min\left\{\frac{2}{\omega}(1 - \kappa), \delta\right\}$, so gilt:

$$\|x_{k+1} - x_*\|_2 \leq \frac{\omega}{2}\|x_k - x_*\|_2^2 + \kappa\|x_k - x_*\|_2$$

Beweis

- Ohne Beschränkung der Allgemeinheit: $x_* = 0, y = 0$

$$x_{k+1} = x_k - DF(x_k)^+ F(x_k)$$

$$\pm DF(x_k)^+ F(0)$$

$$= x_k - DF(x_k)^+ \big(F(x_k) - F(0)\big) - DF(x_k)^+ F(0)$$

$$x_k = DF(x_k)^+ DF(x_k) x_k$$

$$= -DF(x_k)^+ (F(x_k) - F(0) - DF(x_k) x_k) - DF(x_k)^+ F(0)$$

Taylor-Formel für x_k und 0

$$= -\int_0^1 DF(x_k)^+ \big(DF(t x_k) - DF(x_k)\big) x_k \, dt - DF(x_k)^+ F(0)$$

Beweis

$$\|DF(x_k)^+F(0)\|_2 \le \kappa\|x_k\|_2$$

Annahme

$$x_{k+1} = -\int_0^1 \underbrace{DF(x_k)^+\big(DF(tx_k) - DF(x_k)\big)}\, x_k\, dt \;-\; \overbrace{DF(x_k)^+F(0)}$$

Annahme

$$\left\|DF(x_k)^+\big(DF(tx_k) - DF(x_k)\big)\right\|_2 \le \omega(1-t)\|x_k\|_2$$

$$\begin{aligned}\|x_{k+1}\|_2 &\le \int_0^1 \omega(1-t)\|x_k\|_2^2 dt + \kappa\|x_k\|_2 \\ &= \tfrac{1}{2}\omega\|x_k\|_2^2 + \kappa\|x_k\|_2\end{aligned}$$

$\square$

Lemma 4.9. Konvergenzgeschwindigkeit

$$\|DF(x)^+(F(x_*) - y)\|_2 \leq \kappa \|x - x_*\|_2 \text{ mit } \kappa \in [0,1) \;\; \forall x \in B_\delta(x_*)$$

$$\Longrightarrow \|x_{k+1} - x_*\|_2 \leq \tfrac{\omega}{2}\|x_k - x_*\|_2^2 + \kappa\|x_k - x_*\|_2$$

Die Bedingung ist erfüllt mit $\kappa = \varepsilon L$, falls:

$$\|DF(x)^+ - DF(x_*)^+\|_2 \leq L\|x - x_*\|_2 \;\text{ für } x \in B_\delta(x_*)$$

$$\|F(x_*) - y\|_2 = \varepsilon < \frac{1}{L} \;\; \textit{Rauschpegel in Messwerten}$$

- Quadratische Konvergenz für exakte Daten: $\varepsilon = 0$
- Die Konvergenzgeschwindigkeit fällt mit steigendem ε

Beweis

Behauptung: $\|DF(x)^{+}(F(x_*) - y)\|_2 \leq \kappa \|x - x_*\|_2$ mit $\kappa \in [0,1)$

$$f(x) = \tfrac{1}{2}\|F(x) - y\|_2^2$$

$$\nabla f(x_*) = DF(x_*)^T (F(x_*) - y) = 0$$

$$DF(x_*)^{+}(F(x_*) - y) = 0$$

$$\|DF(x)^{+}(F(x_*) - y)\|_2 = \|(DF(x)^{+} - DF(x_*)^{+})(F(x_*) - y)\|_2$$

$$\leq \underbrace{\|DF(x)^{+} - DF(x_*)^{+}\|_2}_{\leq L\|x - x_*\|_2} \underbrace{\|F(x_*) - y\|_2}_{\varepsilon}$$

$$\leq \underbrace{L\varepsilon}_{\kappa < 1} \|x - x_*\|_2 \qquad \square$$

Stabilität

$$\text{Minimiere } \; f(x) = \frac{1}{2}||F(x) - y||_2^2 \; \text{ über } x \in \mathbb{R}^n$$
$$F \in C^1(\mathbb{R}^n, \mathbb{R}^m), \; y \in \mathbb{R}^m$$

- Oft sind sind Matrizen $DF(x)^T DF(x)$ schlecht konditioniert, d.h.

$$\kappa(DF(x)^T DF(x)) \; \textit{ist groß}$$

- Der Fehler in Messwerten y wird beim Gauß-Newton-Schritt verstärkt :

$$x_{k+1} = x_k - \left(DF(x_k)^T DF(x_k)\right)^{-1} DF(x_k)^T (F(x_k) - y)$$

- Die Iterationen akkumulieren die Fehler

Levenberg-Marquardt-Verfahren

Minimiere $f(x) = \frac{1}{2}\|F(x) - y\|_2^2$ über $x \in \mathbb{R}^n$

$F \in C^1(\mathbb{R}^n, \mathbb{R}^m),\ y \in \mathbb{R}^m$

linearisiere F

Minimiere $\|F(x_k) + DF(x_k)(x - x_k) - y\|_2^2$

aktuelle Näherungslösung

Anpassungsparameter > 0

Minimiere $\|F(x_k) + DF(x_k)(x - x_k) - y\|_2^2 + \alpha_k \|x - x_k\|_2^2$

Ausgleich zwischen Empfindlichkeit auf Messfehler und Modellqualität

$$x_{k+1} = x_k - (DF(x_k)^T DF(x_k) + \alpha_k)^{-1}\, DF(x_k)^T (y - F(x_k))$$

Zusammenfassung

- Lineares Ausgleichsproblem
- Nichtlineares Ausgleichsproblem
- Konvergenz des Gauß-Newton-Verfahrens

Nächstes Video

5a. Lineare Nebenbedingungen: Gleichungen