

Inverse Probleme in der Geophysik
Vorlesung (Vertretung K. Spitzer)
TU Bergakademie Freiberg, SS 2020

Thomas Günther (LIAG Hannover)
(Thomas.Guenther@extern.tu-freiberg.de)

8. Juni 2020

Teil 1: Einführung und Motivation

Angewandte Geophysik

Messung und Rückschluss auf Struktur & Parameter des Untergrunds

- direkte Verwendung sehr selten (Punktmessungen): Bohrlochgeophysik, flache Magnetik, Bodensensoren, Eigenpotential
- ansonsten: Messung = \sum Effekte des Untergrundes + Fehler
- Modellbildung (Vereinfachung) und Rekonstruktion

Meist verwendet man fertige Programme zur Auswertung, die man oft nicht durchschaut.

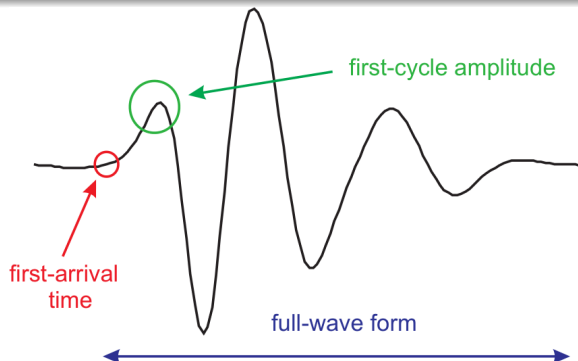
Ziel der Veranstaltung

- Verständnis für Prozess der Inversion, um Ergebnisse einzuschätzen
- zielgerichtete Beeinflussung der (meist mehrdeutigen) Ergebnisse

Daten in der Geophysik = diskrete Anzahl von Messwerten

Arten von Daten

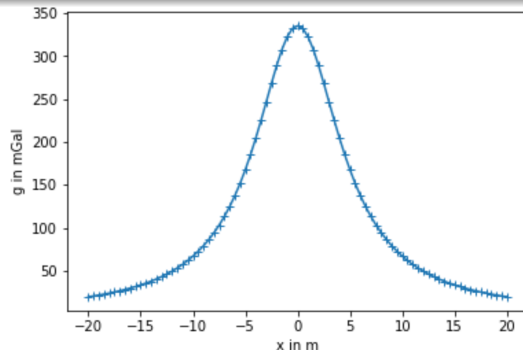
- Seismologie: Zeitreihe von von Beschleunigungswerten
- Gravimetrie: Schwerewerte an diskreten Positionen
- Laufzeittomographie: Laufzeiten zwischen Sender und Empfänger
- Geoelektrik: Ströme und Spannungen f. A-B/M-N Kombinationen



Daten in der Geophysik = diskrete Anzahl von Messwerten

Arten von Daten

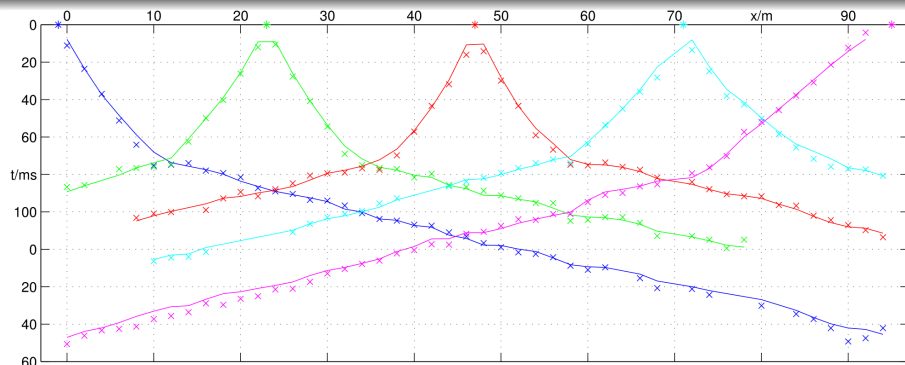
- Seismologie: Zeitreihe von von Beschleunigungswerten
- Gravimetrie: Schwerewerte an diskreten Positionen
- Laufzeittomographie: Laufzeiten zwischen Sender und Empfänger
- Geoelektrik: Ströme und Spannungen f. A-B/M-N Kombinationen



Daten in der Geophysik = diskrete Anzahl von Messwerten

Arten von Daten

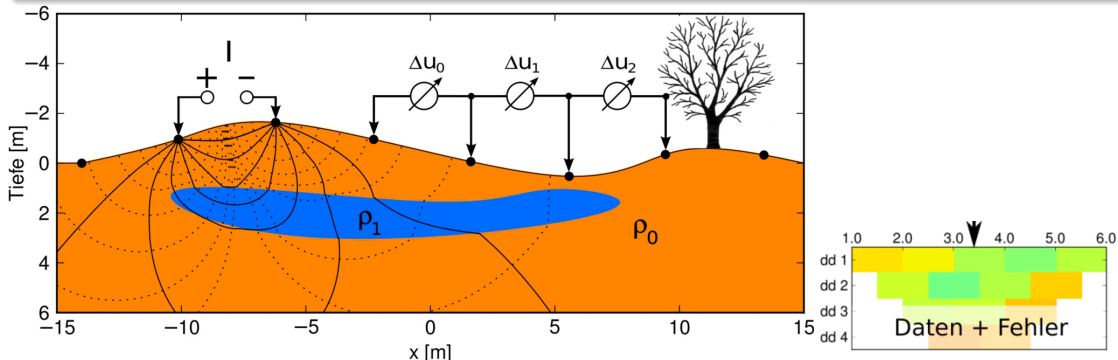
- Seismologie: Zeitreihe von von Beschleunigungswerten
- Gravimetrie: Schwerewerte an diskreten Positionen
- Laufzeittomographie: Laufzeiten zwischen Sender und Empfänger
- Geoelektrik: Ströme und Spannungen f. A-B/M-N Kombinationen



Daten in der Geophysik = diskrete Anzahl von Messwerten

Arten von Daten

- Seismologie: Zeitreihe von von Beschleunigungswerten
- Gravimetrie: Schwerewerte an diskreten Positionen
- Laufzeittomographie: Laufzeiten zwischen Sender und Empfänger
- Geoelektrik: Ströme und Spannungen f. A-B/M-N Kombinationen



Daten in der Geophysik = diskrete Anzahl von Messwerten

Arten von Daten

- Seismologie: Zeitreihe von von Beschleunigungswerten
 - Gravimetrie: Schwerewerte an diskreten Positionen
 - Laufzeittomographie: Laufzeiten zwischen Sender und Empfänger
 - Geoelektrik: Ströme und Spannungen f. A-B/M-N Kombinationen
-
- kann diskretisierte Funktion von Zeit, Ort, oder Frequenz sein (und so geplottet werden)
 - kann mehreren Positionen (Tx-Rx, AB-MN) zugeordnet werden
(gesamter Untergrund nimmt Einfluss – Plotten von Pseudosektion, Crossplots etc.)
 - werden durch ein Modell **m** und Noise **n** verursacht: **d = f(m) + n**

Modell in der Geophysik

Beschreibung von Parametern im Untergrund (räumlich, zeitlich) durch endliche Anzahl an Freiheitsgraden

Unzusammenhängende Parameter

- Seismologie: Herdflächenlösung (Erdbebenposition, Spannung, Winkel, ...)
- Gravimetrie: Dichtekontrast, Tiefe, Durchmesser eines Störkörpers
- Spektroskopie (z.B. SIP): Parameter einer Funktion (z.B. Cole-Cole)

Parameter als Funktion des Ortes (oder/und der Zeit)

- Refraktion: Tiefe des Refraktors (plus Geschwindigkeiten)
- Verteilung von Dichte, Geschwindigkeit oder Leitfähigkeit $p(\vec{r})$

Occams Rasiermesser - Ein grundlegendes Prinzip

William v. Occam, Schottland 14. Jh.:

Pluralitas non est ponenda sine neccesitate!

Eine Mehrheit darf nie ohne Not zugrunde gelegt werden.

Übertragung auf inverse Probleme

Wähle aus allen möglichen Modellen, welche die Daten (im Rahmen der Messfehler) erklären können, das einfachste aus!

Daten und Modell

Daten

Einzelwerte in Vektor $\mathbf{d} = [d_1, d_2, \dots, d_N]$, ggf. Fehlerwerte $\mathbf{e} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N]$

Modell

Verteilung eines (oder mehrerer) Parameter $p(x, y, z)$

oft diskretisiert: $p_{ijk} \Rightarrow \mathbf{m} = [m_1, m_2, \dots, m_M]$

allgemeiner: $p = \sum m_i p_i(x, y, z)$ mit Basisfunktionen p_i

oder: Strukturparameter (vorgegeben oder flexibel),

z.B. 3-Schichtmodell: $\mathbf{m} = [p_1, p_2, p_3, h_1, h_2]$

Inverses Problem

Bestimme ein Modell **m**, das die Daten **d** im Rahmen des Fehlers erklärt:

$$\mathbf{d} = \mathbf{f}(\mathbf{m}) + \mathbf{n}$$

Vorwärtsantwort (ideale Messung) **f**, Noise **n**

Lineares Problem

f(m) ist linear bezüglich der Modellparameter m_i
⇒ kann als Matrix-Vektorgleichung geschrieben werden

$$\mathbf{d} = \mathbf{Gm} + \mathbf{n}$$

Gravimetrie, Magnetik, MRS, VSP, Tomographie mit geraden Strahlen, Regression

Korrekt gestellte Probleme

Korrekt gestelltes Problem

Definition nach Hadamard:

- 1 Es existiert eine Lösung.
- 2 Sie ist eindeutig.
- 3 Die Lösung hängt stabil von den Eingangsdaten ab, d.h. kleine Variationen führen zu kleinen Änderungen.

Schlecht gestellte Probleme

- 1 Kein Modell kann die Daten perfekt anpassen.
- 2 Innerhalb eines Fehlers können viele Modelle die Daten fitten.
- 3 Kleine Änderungen in den Daten führen zu großen Modelländerungen.

Wie lösen wir das inverse Problem?

Vorwärtsmodellierung

- gezielt ausprobieren und variieren
- bestimmtes Raster an Lösungen absuchen (grid search)
- intelligent suchen (Genetische Algorithmen etc.)

Matrix-basierte Minimierung

- strahlenbasierte Rekonstruktion (ART, SIRT)
- Gradientenverfahren (steepest descent)
- Newton-Verfahren (Gauss-Newton)
- Mischung von Verfahren, Filterung, Dekonvolution

Ziel

Minimierung des Residuums $\mathbf{d} - \mathbf{f}(\mathbf{m})$

Wie invertieren wir nun **G**?

Problem

- Matrix **G** ist meist nicht invertierbar
- im Allgemeinen nicht einmal quadratisch

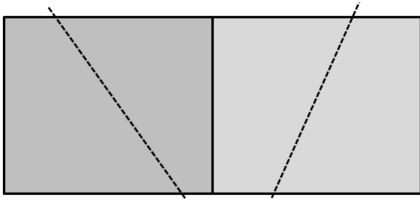
Verschiedene Aufgabentypen

Anzahl unabhängiger Messungen N , Anzahl Modellparameter M

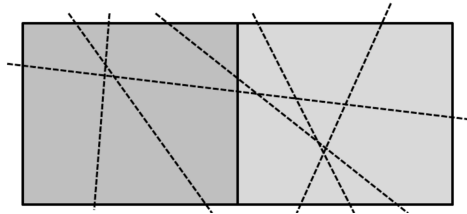
- $N > M$: überbestimmtes Problem \Rightarrow Ausgleichsrechnung, Lösung im Sinne kleinster Quadrate
- $N < M$: Unterbestimmtes Problem \Rightarrow Zusätzliche Forderungen an Lösung führen zu Eindeutigkeit
- In vielen Fällen: sowohl über- als auch unterbestimmte Parameter gleichzeitig

Über- und Unterbestimmtheit (Menke, 2012)

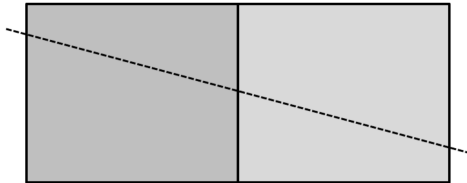
Even-determined



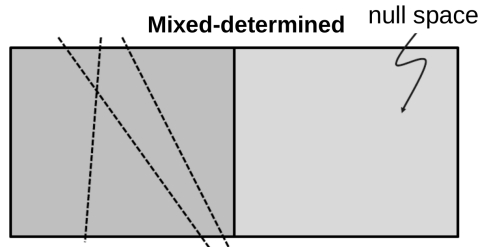
Over-determined



Under-determined



Mixed-determined



Beispiel überbestimmtes Problem

$$m_1 - m_2 = -1$$

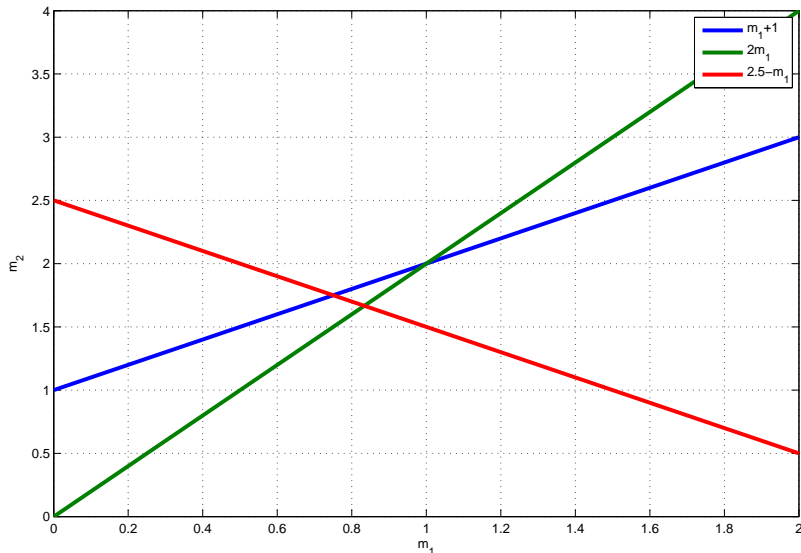
(1)

$$2m_1 - m_2 = 0$$

(2)

$$m_1 + m_2 = 2.5$$

(3)



Es gibt mehr unabhängige Gleichungen als Unbekannte.

Beispiel überbestimmtes Problem

$$m_1 - m_2 = -1$$

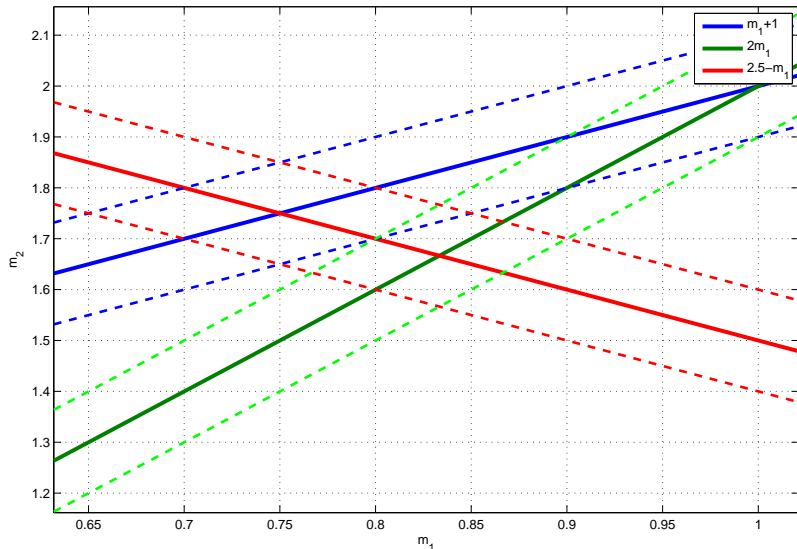
(1)

$$2m_1 - m_2 = 0$$

(2)

$$m_1 + m_2 = 2.5$$

(3)



Es gibt mehr unabhängige Gleichungen als Unbekannte.

Teil 2: Methode der kleinsten Quadrate

Was bisher geschah

Inversion = Rekonstruktion Modell aus Daten

- Übersicht Inversion in der Angewandten Geophysik
- Daten sind (teils zuordenbare) Zahlen mit Fehlern (Vektoren **d**, **e**)
- Modell abstrahiert Untergrund auf wenige Freiheitsgrade (Modell-Vektor **m**)
(nach Occams Razor: möglichst einfache Beschreibung)
- Lineares Inversionsproblem **Gm = d**
- Korrekt gestelltes Problem: Existenz, Eindeutigkeit, Stetigkeit
- Aufgabentypen: überbestimmt, unterbestimmt
(meist sowohl über- als auch unterbestimmte Anteile)
- Einfaches Matrix-Beispiel (3 Gleichungen für 2 Unbekannte)
Lösung = Kompromiss zwischen allen incl. Fehlerbereiche

Teil 2: Methode der kleinsten Quadrate

Was bisher geschah

Inversion = Rekonstruktion Modell aus Daten

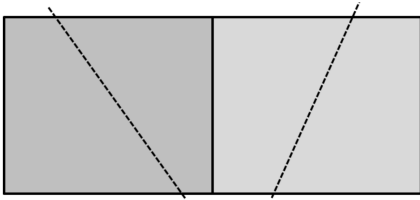
- Übersicht Inversion in der Angewandten Geophysik
- Daten sind (teils zuordenbare) Zahlen mit Fehlern (Vektoren **d**, **e**)
- Modell abstrahiert Untergrund auf wenige Freiheitsgrade (Modell-Vektor **m**)
(nach Occams Razor: möglichst einfache Beschreibung)
- Lineares Inversionsproblem $\mathbf{Gm} = \mathbf{d} = \mathbf{Gm}^{true} + \mathbf{n}$
- Korrekt gestelltes Problem: Existenz, Eindeutigkeit, Stetigkeit
- Aufgabentypen: überbestimmt, unterbestimmt
(meist sowohl über- als auch unterbestimmte Anteile)
- Einfaches Matrix-Beispiel (3 Gleichungen für 2 Unbekannte)
Lösung = Kompromiss zwischen allen incl. Fehlerbereiche

Inhalt der heutigen Veranstaltung

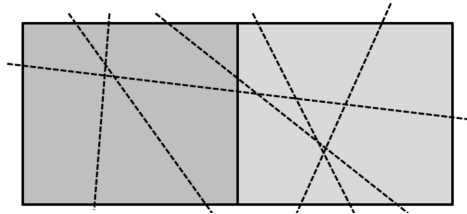
- Zusammenfassung und Fragen
- Troubleshooting Julia und Jupyter Notebooks (live)
- Die Methode der kleinsten Quadrate (pdf)
- Berücksichtigung von Daten-Fehlern
- Fortsetzung minimalistisches Matrix-Problem
- Auflösungsmatrizen: Modellauflösung, Dateninformation
- Übungsbeispiel lineare Regression
- Eigenwertzerlegung, Singulärwertzerlegung

Über- und Unterbestimmtheit (Menke, 2012)

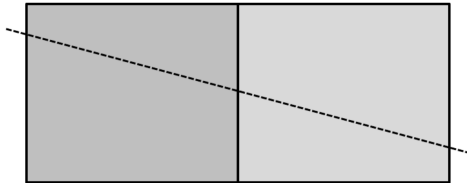
Even-determined



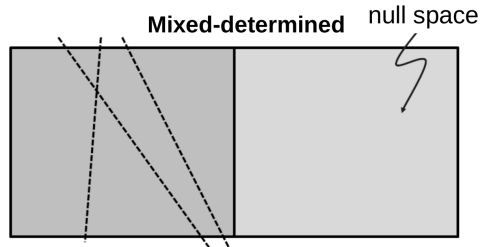
Over-determined



Under-determined



Mixed-determined

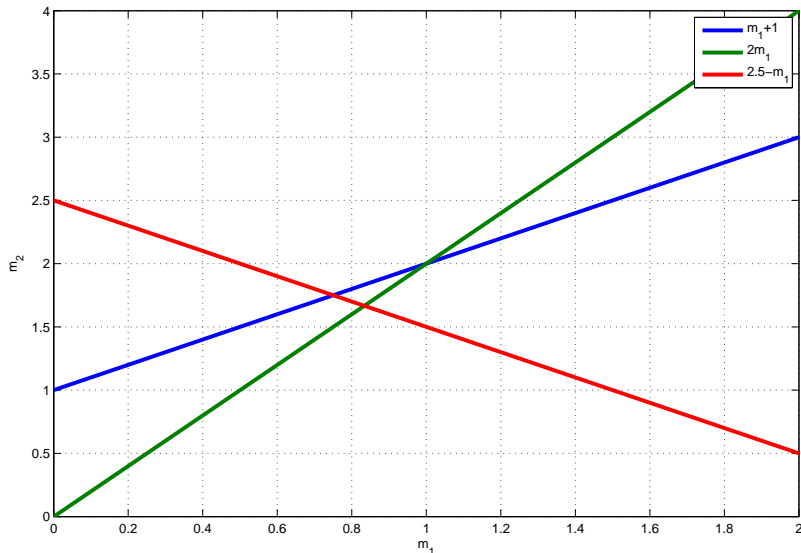


Beispiel überbestimmtes Problem

$$m_1 - m_2 = -1 \quad (4)$$

$$2m_1 - m_2 = 0 \quad (5)$$

$$m_1 + m_2 = 2.5 \quad (6)$$



Es gibt mehr unabhängige Gleichungen als Unbekannte.

Beispiel überbestimmtes Problem

$$m_1 - m_2 = -1$$

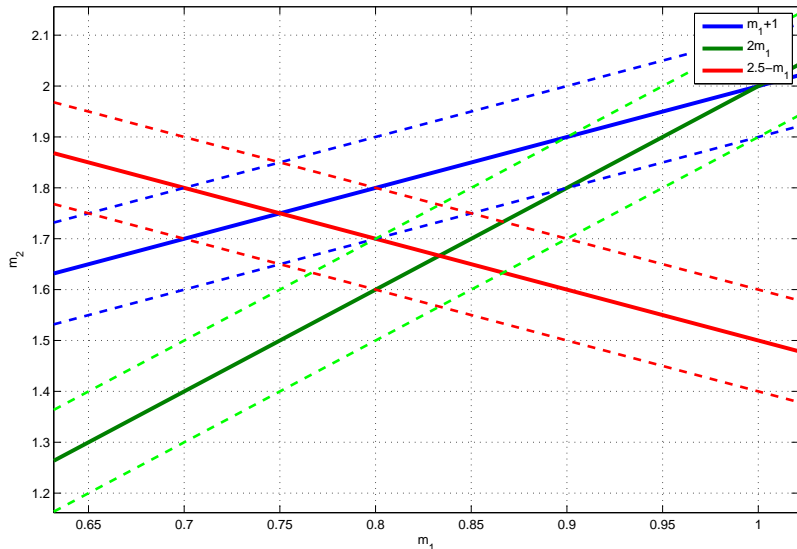
(4)

$$2m_1 - m_2 = 0$$

(5)

$$m_1 + m_2 = 2.5$$

(6)



Es gibt mehr unabhängige Gleichungen als Unbekannte.

Die Methode der kleinsten Quadrate

Ausgangspunkt ist die Minimierung des Residuums $\mathbf{d} - \mathbf{Gm}$, im Sinne der kleinsten Quadrate

$$\Phi = \|\mathbf{d} - \mathbf{Gm}\|_2^2 = (\mathbf{d} - \mathbf{Gm})^T (\mathbf{d} - \mathbf{Gm}) = (\mathbf{Gm} - \mathbf{d})^T (\mathbf{Gm} - \mathbf{d}) \quad (7)$$

Die Funktion Φ wird auch Zielfunktion (objective function) genannt.

Bedingung für ein Extremum ist das Verschwinden der Ableitungen nach allen freien Parametern.

$$\frac{\partial \Phi}{\partial m} = \frac{\partial}{\partial m} (\mathbf{Gm} - \mathbf{d})^T (\mathbf{Gm} - \mathbf{d}) + (\mathbf{Gm} - \mathbf{d})^T \frac{\partial}{\partial m} (\mathbf{Gm} - \mathbf{d}) = 0 \quad (8)$$

$$\mathbf{G}^T \mathbf{Gm} - \mathbf{G}^T \mathbf{d} + \mathbf{G}^T \mathbf{Gm} - \mathbf{G}^T \mathbf{d} = 0 \quad (9)$$

$$\mathbf{G}^T \mathbf{Gm} = \mathbf{G}^T \mathbf{d} \quad \Rightarrow \quad \mathbf{m} = \mathbf{G}^\dagger \mathbf{d} \quad \text{mit} \quad \mathbf{G}^\dagger = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \quad (10)$$

\mathbf{G}^\dagger wird auch Pseudo-Inverse (Moore-Penrose-Inverse) von \mathbf{G} genannt

$$\Phi = (\mathbf{d} - \mathbf{Gm})^T (\mathbf{d} - \mathbf{Gm}) = \sum_i \left[(d_i - \sum_j G_{ij} m_j) (d_i - \sum_k G_{ik} m_k) \right]$$

$$\Phi = \sum_i \left[d_i d_i - d_i \sum_k G_{ik} m_k - d_i \sum_j G_{ij} m_j + \sum_j G_{ij} m_j \sum_k G_{ik} m_k \right]$$

$$\Phi = \sum_i d_i d_i - 2 \sum_j m_j \sum_i d_i G_{ij} + \sum_i \sum_j \sum_k m_j G_{ij} G_{ik} m_k$$

$$\Phi = \sum_i d_i d_i - 2 \sum_j m_j \sum_i d_i G_{ij} + \sum_j \sum_k m_j m_k \sum_i G_{ij} G_{ik}$$

$$\partial \Phi = \partial m_q = \sum_i \sum_k (\delta_{iq} m_k + m_j \delta_{ik}) \sum_j G_{ij} G_{ik} - 2 \sum_j \delta_{iq} \sum_i G_{ij} d_i = 0$$

$$0 = 2 \sum_k \sum_i G_{iq} G_{ik} - 2 \sum_i G_{iq} d_i = 2 \mathbf{G}^T \mathbf{G} - 2 \mathbf{G}^T \mathbf{d}$$

Herleitung (2)

Wir stören unser Modell \mathbf{m} durch eine Änderung $t\delta\mathbf{m}$

$$\Phi(t) = (\mathbf{d} - \mathbf{G}(\mathbf{m} + t\delta\mathbf{m}))^T (\mathbf{d} - \mathbf{G}(\mathbf{m} + t\delta\mathbf{m}))$$

$$\Phi(t) = (\mathbf{m} + t\delta\mathbf{m})^T \mathbf{G}^T \mathbf{G} (\mathbf{m} + t\delta\mathbf{m}) - 2(\mathbf{m} + t\delta\mathbf{m})^T \mathbf{G}^T \mathbf{d} + \mathbf{d}^T \mathbf{d}$$

$$\Phi(t) = t^2(\delta\mathbf{m}^T \mathbf{G}^T \mathbf{G} \delta\mathbf{m}) + 2t(\delta\mathbf{m}^T \mathbf{G}^T \mathbf{G} \mathbf{m} - \delta\mathbf{m}^T \mathbf{G}^T \mathbf{d}) + (\mathbf{m}^T \mathbf{G}^T \mathbf{G} \mathbf{m} + \mathbf{d}^T \mathbf{d} - 2\mathbf{m}^T \mathbf{G}^T \mathbf{d})$$

$\Phi(t)$ hat ein Minimum bei $t = 0$, also muss $\partial\Phi/\partial t$ verschwinden:

$$\partial\Phi(t=0)/\partial t = 2(\delta\mathbf{m}^T \mathbf{G}^T \mathbf{G} \mathbf{m} - \delta\mathbf{m}^T \mathbf{G}^T \mathbf{d}) = 2\delta\mathbf{m}^T (\mathbf{G}^T \mathbf{G} \mathbf{m} - \mathbf{G}^T \mathbf{d}) = 0$$

Das das für jedes $\delta\mathbf{m}$ gilt, muss $\mathbf{G}^T \mathbf{G} \mathbf{m} = \mathbf{G}^T \mathbf{d}$ sein

Die Methode der kleinsten Quadrate

Daraus folgen die Normalgleichungen

$$\mathbf{G}^T(\mathbf{d} - \mathbf{Gm}) = 0 = \mathbf{G}^T\mathbf{Gm} - \mathbf{G}^T\mathbf{d}$$

mit der (nun eindeutigen) Least Squares Lösung

$$\mathbf{m} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d} = \mathbf{G}^\dagger\mathbf{d} \quad \text{mit} \quad \mathbf{G}^\dagger = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T$$

Maß für die Anpassung ist die (normalisierte) Residuumsnorm

$$\|\mathbf{d} - \mathbf{f}(\mathbf{m})\| = \sqrt{1/N \sum (d_i - f_i(\mathbf{m}))^2}$$

auch bezeichnet als RMS (root mean square)

Gewichtete Minimierung

Was passiert bei verschiedener Genauigkeit der Daten?

Wichtung des Datenmisfits durch individuellen Datenfehler ε_i :

$$\sum \left(\frac{d_i - f_i(\mathbf{m})}{\varepsilon_i} \right)^2 \rightarrow \min$$

(Ersetzung d_i durch $\hat{d}_i = d_i/\varepsilon_i$) führt zu

$$\mathbf{m} = (\hat{\mathbf{G}}^T \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^T \hat{\mathbf{d}}$$

mit $\hat{\mathbf{G}} = \text{diag}(1/\varepsilon_i) \cdot \mathbf{G}$

zugehöriges Fehlermaß: fehlergewichteter Misfit (idealerweise im Mittel 1)

$$\chi^2 = \frac{1}{N} \sum \left(\frac{d_i - f_i(\mathbf{m})}{\varepsilon_i} \right)^2$$

Rauschen und Fehler

- Fehler (immer da) werden mit invertiert
- Least-Squares-Inversion = Gauss-Verteilung des Residuums
- Modellvariation durch Wiederholung: Fehleranalyse
- je größer Daten-Fehler desto größer Modell-Variation
- auch abhängig von Gutartigkeit des Problems
- ungleiches Rauschen \Rightarrow systematische Verzerrung
- Wichtung der Daten mit reziprokem Fehler
 \Rightarrow gewichtete Normalgleichungen

$$\mathbf{m} = (\hat{\mathbf{G}}^T \hat{\mathbf{G}})^{-1} \hat{\mathbf{G}}^T \hat{\mathbf{d}} \text{ mit } \hat{\mathbf{G}} = \text{diag}(1/\varepsilon_i) \cdot \mathbf{G}$$

- Maß für Anpassung: χ^2 (fehlergewichtetes Quadratmittel)

Auflösungsmatrizen

Modell-Auflösung

$$\mathbf{d} = \mathbf{G}\mathbf{m}^{\text{true}} + \mathbf{n}$$

Matrix-Inversion mit inversem Operator \mathbf{G}^\dagger :

$$\mathbf{m}^{\text{est}} = \mathbf{G}^\dagger \mathbf{d} = \mathbf{G}^\dagger \mathbf{G} \mathbf{m}^{\text{true}} + \mathbf{G}^\dagger \mathbf{n} = \mathbf{R}^M \mathbf{m}^{\text{true}} + \mathbf{G}^\dagger \mathbf{n}$$

mit der Modell-Auflösungsmatrix $\mathbf{R}^M = \mathbf{G}^\dagger \mathbf{G}$

⇒ Wie spiegelt sich die Wahrheit (\mathbf{m}^{true}) im Ergebnis (\mathbf{m}^{est}) wider?

Diagonale von \mathbf{R}^M : Auflösung der Modellparameter, Nebendiagonale: Verzerrung

Überbestimmte Probleme

$$\mathbf{G}^\dagger = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \Rightarrow \mathbf{R}^M = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{G} = \mathbf{I} \Rightarrow \text{perfekte Modellauflösung}$$

Auflösungsmatrizen

Daten-Informationsdichtematrix

$$\mathbf{m}^{\text{est}} = \mathbf{G}^\dagger \mathbf{d}^{\text{obs}}$$

Wie werden die Daten durch das Modell erklärt?

$$\mathbf{d}^{\text{est}} = \mathbf{G} \mathbf{m}^{\text{est}} = \mathbf{G} \mathbf{G}^\dagger \mathbf{d}^{\text{obs}} = \mathbf{R}^D \mathbf{d}^{\text{obs}}$$

mit der Daten-Auflösungsmatrix (Informationsdichtematrix):

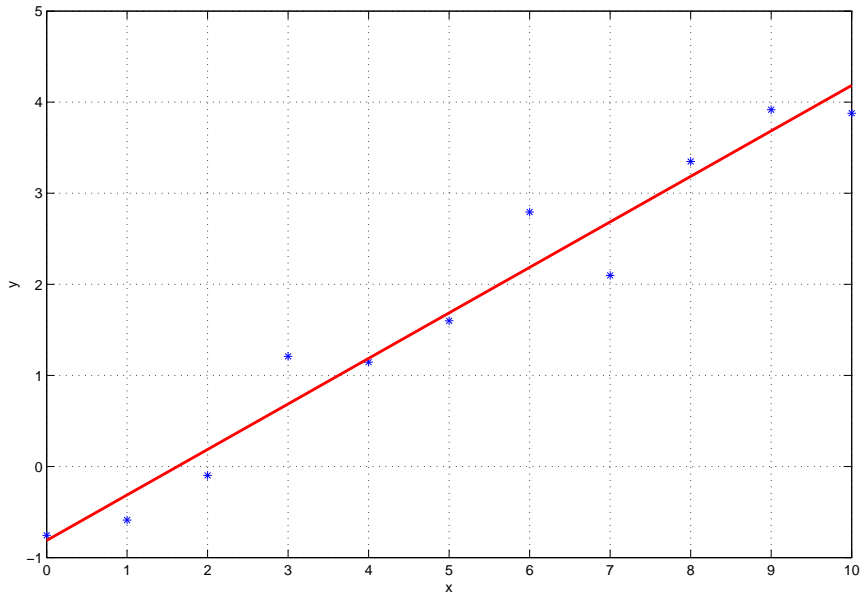
$$\mathbf{R}^D = \mathbf{G} \mathbf{G}^\dagger$$

Diagonale von \mathbf{R}^D : Informationsgehalt der Daten, Nebendiagonale: Korrelation

Überbestimmte Probleme

$$\mathbf{G}^\dagger = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \quad \Rightarrow \quad \mathbf{R}^D = \mathbf{G} (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$$

Lineare Regression(1)



Lineare Regression(2)

Die Daten: y_i Das Modell: a, b Der Vorwärtsoperator: Abbildung von (a, b) auf $a + bx$ durch Matrix-Vektor-Produkt.

- 1 Wie muss diese aussehen? (Überlegung 1, danach 2 Werte)
- 2 Stellen Sie \mathbf{G} auf und lösen Sie die Normalgleichungen

$$\mathbf{G}^T(\mathbf{d} - \mathbf{Gm}) = 0 \text{ bzw. } \mathbf{G}^T\mathbf{Gm} = \mathbf{G}^T\mathbf{d}$$

- 3 Testen Sie mit idealen Daten (graphischer Vergleich)!
- 4 Verrauschen Sie die Daten und variieren Sie die Fehler.
- 5 Berechnen Sie die Fehlerquadratsumme!
- 6 Wiederholen Sie (neue Verrauschung) & plotten Sie die alle Ergebnisse zusammen! Wie verteilen sie sich?
- 7 Erhöhen Sie den Polynomgrad schrittweise!

Daten-Auflösung Überbestimmte Probleme

Berechnen Sie für die beiden Beispiel-Probleme (3 Geraden, Lineare Regression) die Datenauf Lösungsmatrix und stellen Sie diese dar

Problem mit Unterbestimmung

2 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{Gm} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{d}$$

Zerlegung in zwei Unterprobleme:

- Parameter 3 ist überbestimmt
- Parameter 1+2 unterbestimmt

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Problem mit Unterbestimmung

2 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{Gm} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{d}$$

Zerlegung in zwei Unterprobleme:

- Parameter 3 ist überbestimmt
- Parameter 1+2 unterbestimmt

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Problem mit Unterbestimmung

2 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{Gm} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{d}$$

Zerlegung in zwei Unterprobleme:

- Parameter 3 ist überbestimmt
- Parameter 1+2 unterbestimmt

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Teil 3: Auflösungsanalyse

Was bisher geschah

Überbestimmte Probleme

Lösung: Methode der kleinsten Quadrate

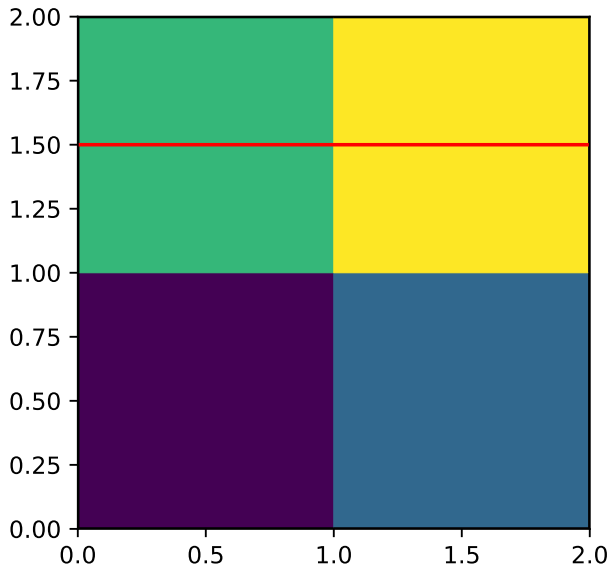
- Einfaches Matrix-Beispiel (3 Gleichungen für 2 Unbekannte)
Lösung = Kompromiss zwischen allen incl. Fehlerbereiche
- Generalisierte Inverse durch Lösung von Normalengleichung
- Einfluss von Fehlerwerten durch Wichtung
- Einführung von Auflösungsmatrizen
Überbestimmt = perfekte Modellauflösung, Dateninformation noch anschauen!

IJulia Notebooks unter <https://github.com/halbmy/IJulia> verfügbar

Inhalt der heutigen Veranstaltung

- Zusammenfassung und Fragen
- Kurzer Test Problemtypen
- gemeinsame Diskussion und Abschluss Matrix-Problem (JNB live)
- Wiederholung Auflösungsmatrizen: Modellauflösung, Dateninformation (pdf)
- Übungsbeispiel lineare Regression (JNB live+selbst)
- Normen und Robuste Inversion (JNB live)
- Unterbestimmte und gemischt bestimmte Probleme (pdf+JNB)
Lösung der kleinsten Modell-Norm (JNB)
- Singulärwertzerlegung (pdf)
 - ▶ Matrix-Kompression mit SVD
 - ▶ Generalisierte Inverse
 - ▶ Auflösungsmatrizen
 - ▶ Anwendung auf bisherige Probleme (JNB)

Test Problemtypen mit Strahlentomographie



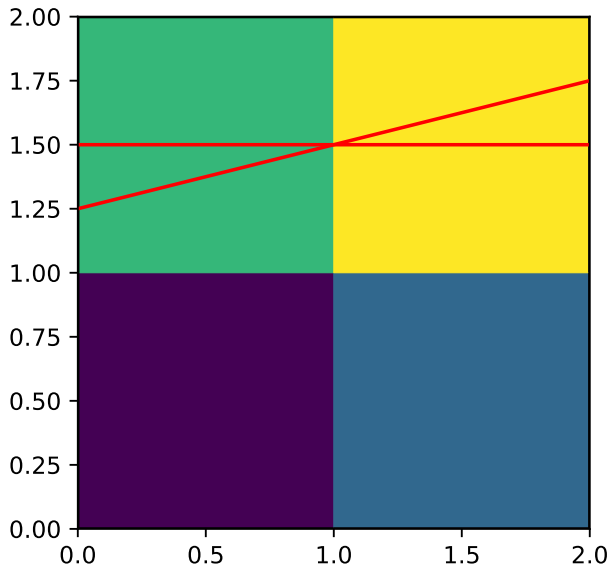
Boxen = Zellen (Modell)
Rote Linien = Strahlen (Daten)
Matrix enthält Laufwege der Zellen für
einzelne Strahlen

$$d_i = \int_l s_j dl = \sum_j w_{ij} s_j$$

Aufgabe 1

- ☐ A Gleichbestimmt
- ☐ B Überbestimmt
- ☐ C Unterbestimmt
- ☐ D Gemischt bestimmt

Test Problemtypen mit Strahlentomographie



Boxen = Zellen (Modell)

Rote Linien = Strahlen (Daten)

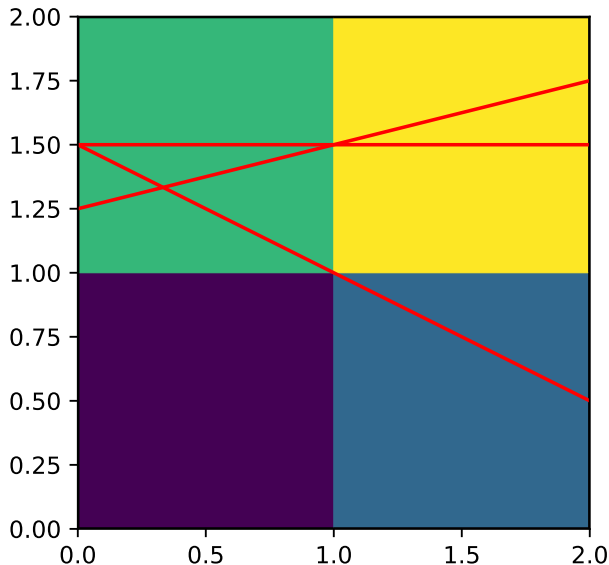
Matrix enthält Laufwege der Zellen für
einzelne Strahlen

$$d_i = \int_l s_j dl = \sum_j w_{ij} s_j$$

Aufgabe 2

- ☐ A Gleichbestimmt
- ☐ B Überbestimmt
- ☐ C Unterbestimmt
- ☐ D Gemischt bestimmt

Test Problemtypen mit Strahlentomographie



Boxen = Zellen (Modell)

Rote Linien = Strahlen (Daten)

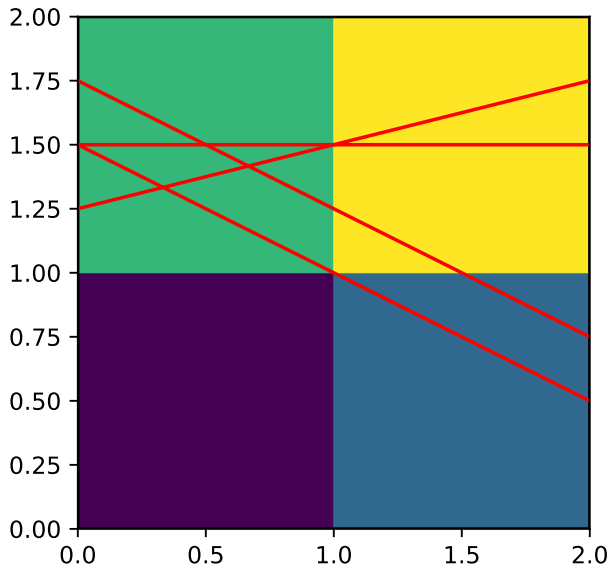
Matrix enthält Laufwege der Zellen für
einzelne Strahlen

$$d_i = \int_l s_j dl = \sum_j w_{ij} s_j$$

Aufgabe 3

- ☐ A Gleichbestimmt
- ☐ B Überbestimmt
- ☐ C Unterbestimmt
- ☐ D Gemischt bestimmt

Test Problemtypen mit Strahlentomographie



Boxen = Zellen (Modell)

Rote Linien = Strahlen (Daten)

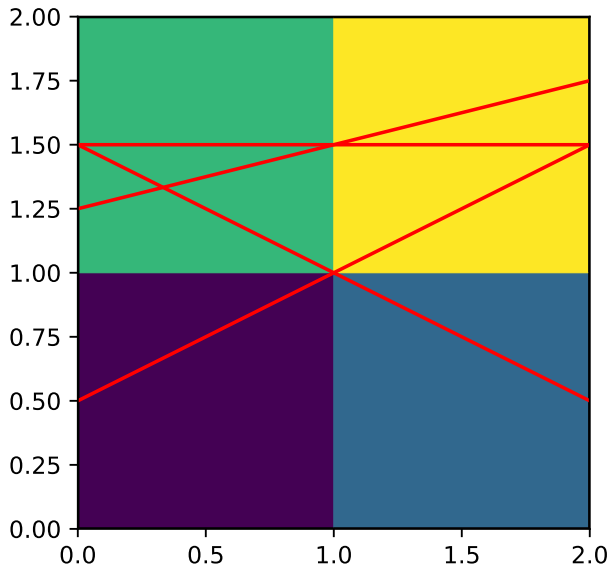
Matrix enthält Laufwege der Zellen für
einzelne Strahlen

$$d_i = \int_l s_j dl = \sum_j w_{ij} s_j$$

Aufgabe 4

- ☐ A Gleichbestimmt
- ☐ B Überbestimmt
- ☐ C Unterbestimmt
- ☐ D Gemischt bestimmt

Test Problemtypen mit Strahlentomographie



Boxen = Zellen (Modell)

Rote Linien = Strahlen (Daten)

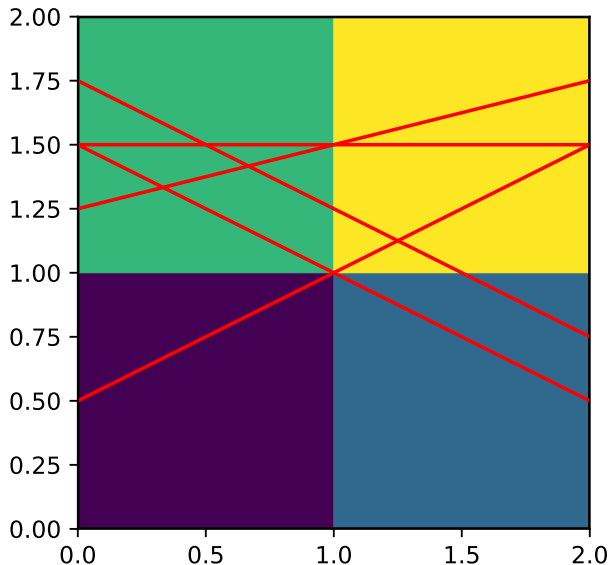
Matrix enthält Laufwege der Zellen für
einzelne Strahlen

$$d_i = \int_l s_j dl = \sum_j w_{ij} s_j$$

Aufgabe 5

- ☐ A Gleichbestimmt
- ☐ B Überbestimmt
- ☐ C Unterbestimmt
- ☐ D Gemischt bestimmt

Test Problemtypen mit Strahlentomographie



Boxen = Zellen (Modell)

Rote Linien = Strahlen (Daten)

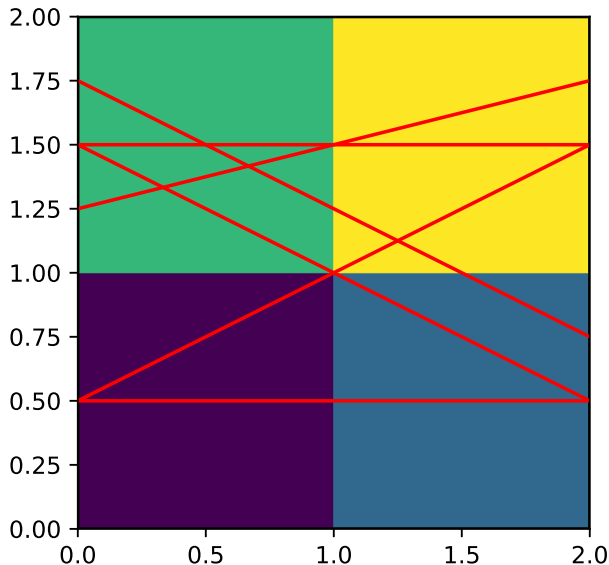
Matrix enthält Laufwege der Zellen für
einzelne Strahlen

$$d_i = \int_l s_j dl = \sum_j w_{ij} s_j$$

Aufgabe 6

- ☐ A Gleichbestimmt
- ☐ B Überbestimmt
- ☐ C Unterbestimmt
- ☐ D Gemischt bestimmt

Test Problemtypen mit Strahlentomographie



Boxen = Zellen (Modell)

Rote Linien = Strahlen (Daten)

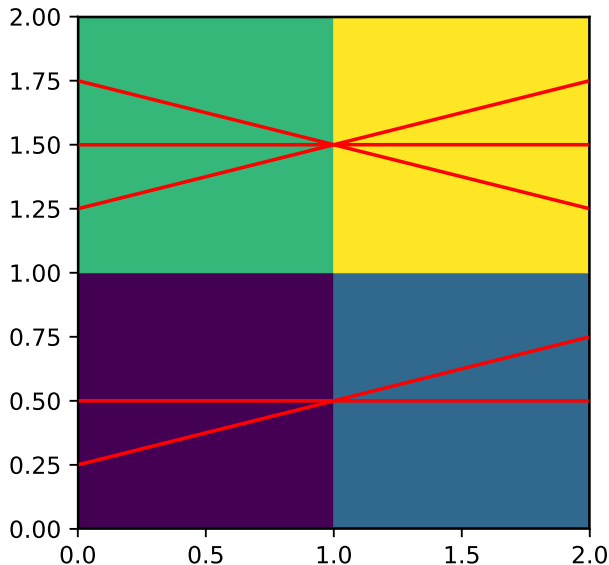
Matrix enthält Laufwege der Zellen für
einzelne Strahlen

$$d_i = \int_l s_j dl = \sum_j w_{ij} s_j$$

Aufgabe 7

- ☐ A Gleichbestimmt
- ☐ B Überbestimmt
- ☐ C Unterbestimmt
- ☐ D Gemischt bestimmt

Test Problemtypen mit Strahlentomographie



Boxen = Zellen (Modell)

Rote Linien = Strahlen (Daten)

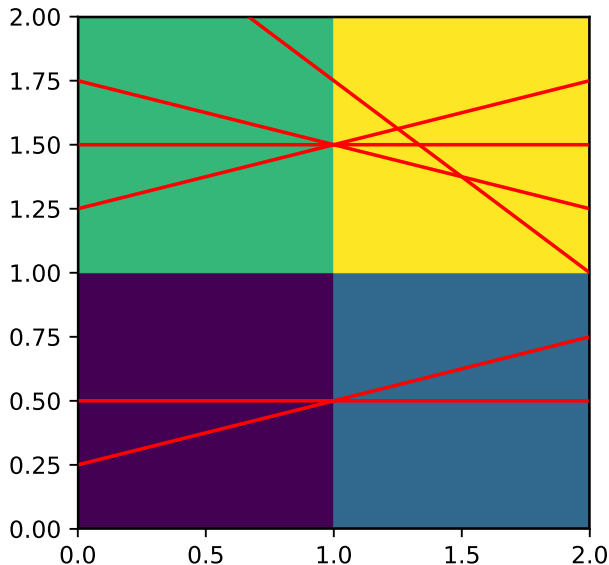
Matrix enthält Laufwege der Zellen für
einzelne Strahlen

$$d_i = \int_l s_j dl = \sum_j w_{ij} s_j$$

Aufgabe 8

- ☐ A Gleichbestimmt
- ☐ B Überbestimmt
- ☐ C Unterbestimmt
- ☐ D Gemischt bestimmt

Test Problemtypen mit Strahlentomographie



Boxen = Zellen (Modell)

Rote Linien = Strahlen (Daten)

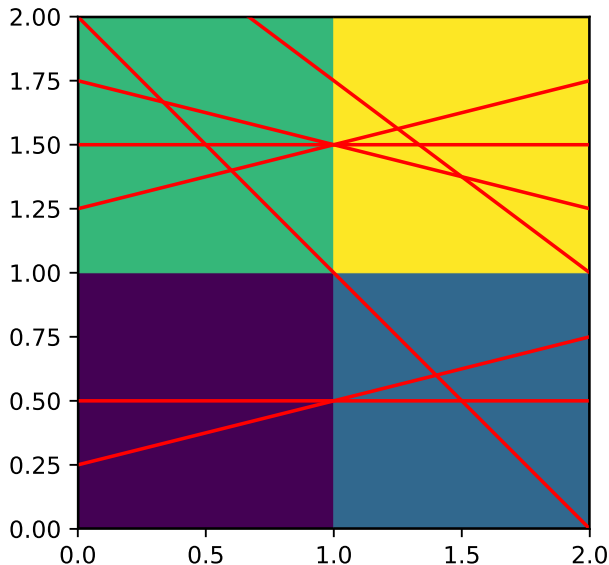
Matrix enthält Laufwege der Zellen für
einzelne Strahlen

$$d_i = \int_l s_j dl = \sum_j w_{ij} s_j$$

Aufgabe 9

- ☐ A Gleichbestimmt
- ☐ B Überbestimmt
- ☐ C Unterbestimmt
- ☐ D Gemischt bestimmt

Test Problemtypen mit Strahlentomographie



Boxen = Zellen (Modell)
Rote Linien = Strahlen (Daten)
Matrix enthält Laufwege der Zellen für
einzelne Strahlen

$$d_i = \int_l s_j dl = \sum_j w_{ij} s_j$$

Aufgabe 10

- ☐ A Gleichbestimmt
- ☐ B Überbestimmt
- ☐ C Unterbestimmt
- ☐ D Gemischt bestimmt

Auflösungsmatrizen

Modell-Auflösung

$$\mathbf{d} = \mathbf{G}\mathbf{m}^{\text{true}} + \mathbf{n}$$

Matrix-Inversion mit inversem Operator \mathbf{G}^\dagger :

$$\mathbf{m}^{\text{est}} = \mathbf{G}^\dagger \mathbf{d} = \mathbf{G}^\dagger \mathbf{G} \mathbf{m}^{\text{true}} + \mathbf{G}^\dagger \mathbf{n} = \mathbf{R}^M \mathbf{m}^{\text{true}} + \mathbf{G}^\dagger \mathbf{n}$$

mit der Modell-Auflösungsmatrix $\mathbf{R}^M = \mathbf{G}^\dagger \mathbf{G}$

⇒ Wie spiegelt sich die Wahrheit (\mathbf{m}^{true}) im Ergebnis (\mathbf{m}^{est}) wider?

Diagonale von \mathbf{R}^M : Auflösung der Modellparameter, Nebendiagonale: Verzerrung

Überbestimmte Probleme

$$\mathbf{G}^\dagger = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \Rightarrow \mathbf{R}^M = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{G} = \mathbf{I} \Rightarrow \text{perfekte Modellauflösung}$$

Auflösungsmatrizen

Daten-Informationsdichtematrix

$$\mathbf{m}^{\text{est}} = \mathbf{G}^\dagger \mathbf{d}^{\text{obs}}$$

Wie werden die Daten durch das Modell erklärt?

$$\mathbf{d}^{\text{est}} = \mathbf{G} \mathbf{m}^{\text{est}} = \mathbf{G} \mathbf{G}^\dagger \mathbf{d}^{\text{obs}} = \mathbf{R}^D \mathbf{d}^{\text{obs}}$$

mit der Daten-Auflösungsmatrix (Informationsdichtematrix):

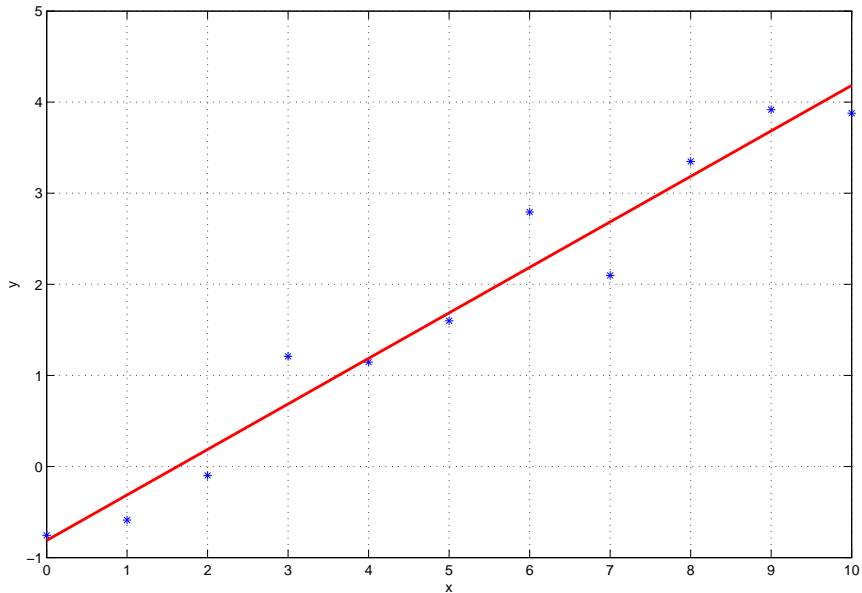
$$\mathbf{R}^D = \mathbf{G} \mathbf{G}^\dagger$$

Diagonale von \mathbf{R}^D : Informationsgehalt der Daten, Nebendiagonale: Korrelation

Überbestimmte Probleme

$$\mathbf{G}^\dagger = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \quad \Rightarrow \quad \mathbf{R}^D = \mathbf{G} (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$$

Lineare Regression(1)



Lineare Regression(2)

Die Daten: y_i Das Modell: a, b Der Vorwärtsoperator: Abbildung von (a, b) auf $a + bx$ durch Matrix-Vektor-Produkt.

- 1 Wie muss diese aussehen? (Überlegung 1, danach 2 Werte)
- 2 Stellen Sie \mathbf{G} auf und lösen Sie die Normalengleichungen

$$\mathbf{G}^T(\mathbf{d} - \mathbf{Gm}) = 0 \text{ bzw. } \mathbf{G}^T\mathbf{Gm} = \mathbf{G}^T\mathbf{d}$$

- 3 Testen Sie mit idealen Daten (graphischer Vergleich)!
- 4 Verrauschen Sie die Daten und variieren Sie die Fehler.
- 5 Berechnen Sie die Fehlerquadratsumme!
- 6 Wiederholen Sie (neue Verrauschung) & plotten Sie die alle Ergebnisse zusammen! Wie verteilen sie sich?
- 7 Erhöhen Sie den Polynomgrad schrittweise!

Unterbestimmte Probleme

2 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{G}\mathbf{m} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{d}$$

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow
Regularisierung

Analog zur Inversen $(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$ für überbestimmte Probleme gibt es eine Inverse für unterbestimmte Probleme

$$\mathbf{m} = \mathbf{G}^T (\mathbf{G} \mathbf{G}^T)^{-1} \mathbf{d}$$

Sie wird auch als Minimum-Norm-Lösung bezeichnet. \Rightarrow Jupyter Notebook

Teil 4: Singulärwertzerlegung

Was bisher geschah

Überbestimmte Probleme

Lösung: Methode der kleinsten Quadrate

Kompromiss zwischen allen incl. Fehlerbereiche

- Einfaches Matrix-Beispiel (3 Gleichungen für 2 Unbekannte)
- Lineare Regression
- Generalisierte Inverse durch Lösung von Normalengleichung
- Einfluss von Fehlerwerten durch Wichtung

Auflösungsmatrizen

- Kombination zwischen Vorwärts- und inversem Operator
- Modellauflösung: $\mathbf{G}^\dagger \mathbf{G} \Rightarrow$ perfekt (**I**) für überbestimmte Probleme
- Dateninformation: $\mathbf{G}\mathbf{G}^\dagger \Rightarrow$ Wichtigkeit und Korrelation der Messungen

Inhalt der heutigen Veranstaltung

- Zusammenfassung und Fragen
- gemeinsame Diskussion Lineare Regression (JNB live)
- Exkurs Normen und Robuste Inversion (JNB live)
- Troubleshooting Aufgabe Problemtypen (Strahlentomographie) bos 18.5.
- Unterbestimmte Probleme und Minimum-Norm Lösung (pdf, JNB)
- Singulärwertzerlegung
 - ▶ Theorie (pdf)
 - ▶ Generalisierte Inverse und Auflösungsmatrizen
 - ▶ Beispiel Bild-Kompression mit SVD (JNB)
 - ▶ Anwendung auf bisherige Probleme (JNB)
- Regularisierung: Einführung

Offene Aufgaben

- Messwerte am Rand wichtiger, benachbarte korreliert
- Tests mit verschiedenen Fehlerwerten und Fehlerwichtung
- Anzahl Messwerte und Statistik
- Art des Rauschens
- Erweiterung auf beliebige Ordnung \Rightarrow Test quadratisch
- Was machen wir mit Ausreißern?

Demonstration robuste Verfahren und Normen

Unterbestimmte Probleme

2 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{G}\mathbf{m} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \mathbf{d}$$

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig. Wir können zu einer beliebigen Lösung ein Vielfaches des Vektors $[1, -1, 0]$ addieren, und fitten weiterhin die Daten.

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Analog zur Inversen $(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$ für überbestimmte Probleme gibt es eine Inverse für unterbestimmte Probleme

$$\mathbf{m} = \mathbf{G}^T (\mathbf{G} \mathbf{G}^T)^{-1} \mathbf{d}$$

Sie wird auch als Minimum-Norm-Lösung bezeichnet.

Herleitung Minimum-Norm-Lösung

Das inverse Problem kann exakt gelöst werden: $\mathbf{Gm} = \mathbf{d}$.

Wir suchen unter allen Lösungen die "kleinste":

$$\min \Phi = \mathbf{m}^T \mathbf{m} \text{ mit } \mathbf{Gm} = \mathbf{d}$$

Dazu wenden wir die Methode der Lagrange-Parameter ($\boldsymbol{\lambda} = [\lambda_i]$) an:

$$\Phi = \mathbf{m}^T \mathbf{m} + \boldsymbol{\lambda}^T (\mathbf{Gm} - \mathbf{d}) \rightarrow \min$$

Die Ableitungen verschwinden:

$$\frac{\partial \Phi}{\partial \boldsymbol{\lambda}} = \mathbf{Gm} - \mathbf{d} = 0$$

$$\frac{\partial \Phi}{\partial \mathbf{m}} = 2\mathbf{m}^T + \boldsymbol{\lambda}^T \mathbf{G} = 0$$

$$\Rightarrow \mathbf{m} = -\frac{1}{2}(\boldsymbol{\lambda}^T \mathbf{G})^T = -\frac{1}{2}\mathbf{G}^T \boldsymbol{\lambda}$$

Herleitung Minimum-Norm-Lösung (2)

$$\mathbf{m} = -\frac{1}{2}(\boldsymbol{\lambda}^T \mathbf{G})^T = -\frac{1}{2}\mathbf{G}^T \boldsymbol{\lambda}$$

Die Daten ergeben sich mit

$$\mathbf{d} = \mathbf{Gm} = \mathbf{G}\left(-\frac{1}{2}\mathbf{G}^T \boldsymbol{\lambda}\right) = -\frac{1}{2}\mathbf{GG}^T \boldsymbol{\lambda}$$

woraus wir die Lagrange-Parameter bestimmen können:

$$\boldsymbol{\lambda} = -2(\mathbf{GG}^T)^{-1} \mathbf{d}$$

Durch Einsetzen erhalten wir die Lösung von $\mathbf{Gm} = \mathbf{d}$ mit der kleinsten Norm:

$$\mathbf{m}_{MN} = \mathbf{G}^T (\mathbf{GG}^T)^{-1} \mathbf{d}$$

im Gegensatz zur Least-Squares Lösung $\mathbf{m}_{LS} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{d} \Rightarrow$ Jupyter Notebook

Problem mit Unterbestimmung

2 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{Gm} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \mathbf{d}$$

Zerlegung in zwei Unterprobleme:

- Parameter 3 ist überbestimmt
- Parameter 1+2 unterbestimmt

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Problem mit Unterbestimmung

3 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{Gm} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{d}$$

Zerlegung in zwei Unterprobleme:

- Parameter 3 ist überbestimmt
- Parameter 1+2 unterbestimmt

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Problem mit Unterbestimmung

3 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{Gm} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{d}$$

Zerlegung in zwei Unterprobleme:

- Parameter 3 ist überbestimmt
- Parameter 1+2 unterbestimmt

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Problem mit Unterbestimmung

3 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{Gm} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{d}$$

Zerlegung in zwei Unterprobleme:

- Parameter 3 ist überbestimmt
- Parameter 1+2 unterbestimmt

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Regularisierung

Wie können wir die Inversion regulär machen?

Zusätzliche Gleichungen im Modellraum

- A-priori-Wissen über eine Unbekannte (Modellreduktion)
ungefähre Schätzung (Referenzmodell) $m_1 = d_3$
- Beziehung zwischen Unbekannten (z.B. Summe zweier Mächtigkeiten) $m_1 + m_2 = d_3$
- Differenz/Glattheit soll klein sein $m_1 - m_2 = d_3$

Zusammen mit Daten im Sinne kleinster Quadrate zu lösen:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Aber: Erste 2 Gleichungen im Datenraum, letzte im Modellraum (Einheiten, Rauschen etc.)

Regularisierung

Wie können wir die Inversion regulär machen?

Zusätzliche Gleichungen im Modellraum

- A-priori-Wissen über eine Unbekannte (Modellreduktion)
ungefähre Schätzung (Referenzmodell) $m_1 = d_3$
- Beziehung zwischen Unbekannten (z.B. Summe zweier Mächtigkeiten) $m_1 + m_2 = d_3$
- Differenz/Glattheit soll klein sein $m_1 - m_2 = d_3$

Zusammen mit Daten im Sinne kleinster Quadrate zu lösen:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Aber: Erste 2 Gleichungen im Datenraum, letzte im Modellraum (Einheiten, Rauschen etc.)

Regularisierung

Wie können wir die Inversion regulär machen?

Zusätzliche Gleichungen im Modellraum

- A-priori-Wissen über eine Unbekannte (Modellreduktion)
ungefähre Schätzung (Referenzmodell) $m_1 = d_3$
- Beziehung zwischen Unbekannten (z.B. Summe zweier Mächtigkeiten) $m_1 + m_2 = d_3$
- Differenz/Glattheit soll klein sein $m_1 - m_2 = d_3$

Zusammen mit Daten im Sinne kleinster Quadrate zu lösen:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$$

Aber: Erste 2 Gleichungen im Datenraum, letzte im Modellraum (Einheiten, Rauschen etc.)

Eigenwertzerlegung

Die (quadratische) Matrix **A** projiziert einen Vektor in eine andere Richtung. Besondere Vektoren sind Eigenvektoren, die ihre Richtung beibehalten:

$$\mathbf{Ax} = \lambda \mathbf{x}$$

Die Lösung der Gleichung

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = 0$$

führt zur Bestimmung der Eigenwerte über das charakteristische Polynom $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$

Verschiedene Eigenwerte korrespondieren mit linear unabhängigen Eigenvektoren. Für symmetrische Matrizen existiert eine Faktorisierung mit den Eigenvektoren in \mathbf{Q} (als Spalten) und den Eigenwerten in $\mathbf{\Lambda} = \text{diag}(\lambda_i)$

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$$

Singulärwertzerlegung

Wir machen aus unserer rechteckigen Matrix \mathbf{G} eine quadratische+symmetrische Matrix

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{G} \\ \mathbf{G}^T & 0 \end{bmatrix}$$

Diese besitzt eine Eigenwertzerlegung der Form

$$\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i \quad \text{mit} \quad \mathbf{x}_i = \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix}$$

Damit erhalten wir zwei gekoppelte Eigenwertprobleme für \mathbf{G} und \mathbf{G}^T :

$$\mathbf{G}\mathbf{v} = \lambda\mathbf{u} \quad \text{und} \quad \mathbf{G}^T\mathbf{u} = \lambda\mathbf{v} \quad \text{bzw.}$$

sowie durch Multiplikation mit \mathbf{G}^T bzw. \mathbf{G} erhalten wir

$$\mathbf{G}^T\mathbf{G}\mathbf{v} = \lambda\mathbf{G}^T\mathbf{u} = \lambda^2\mathbf{v} \quad \text{und} \quad \mathbf{G}\mathbf{G}^T\mathbf{u} = \lambda\mathbf{G}\mathbf{v} = \lambda^2\mathbf{u}$$

Singulärwertzerlegung als 2 Eigenwertprobleme

Das Eigenwertproblem für die Modell-Eigenvektoren \mathbf{v}

$$\mathbf{G}^T \mathbf{G} \mathbf{v} = \lambda^2 \mathbf{v}$$

Das Eigenwertproblem für die Daten-Eigenvektoren \mathbf{u} :

$$\mathbf{G} \mathbf{G}^T \mathbf{u} = \lambda^2 \mathbf{u}$$

Die Matrix \mathbf{G} wird aufgespannt durch alle Eigenvektoren:

$$\mathbf{G} = \mathbf{U} \mathbf{S} \mathbf{V} \quad \text{mit} \quad \mathbf{S} = \text{diag}(\lambda_i)$$

$\mathbf{U} \in R^{N \times N}$ enthält die Daten-Eigenvektoren, $\mathbf{V} \in R^{N \times N}$ die Modell-Eigenvektoren

Beide Matrizen sind orthonormal mit $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ und $\mathbf{V}^T \mathbf{V} = \mathbf{I}$.

Zu Null ($\lambda_i = 0$) gehörende Vektoren spannen den Modell/Daten-Nullraum auf.

Rang und Reduktion der Matrix mit SVD

Der Rang der nichtverschwindenden (>0) Singulärwerte sei r . Die dazu gehörigen Vektoren spannen den Modell- und Datenraum auf, die zu $\lambda_i = 0$ gehörenden die Null-Räume. Durch ausschließliche Betrachtung des Daten- und Modellvektorraums, also der r nichtverschwindenden Singulärwerte, erhalten wir einen für das Vorwärts-Problem äquivalenten (abgeschnittenen) Operator:

$$\mathbf{G}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{V}_r^T$$

Für schlecht gestellte (großskalige) Probleme werden Singulärwerte nicht exakt Null, sondern sehr klein und damit die Inverse instabil.

Dann kann der Rang künstlich verkleinert werden (Pseudorang).

Die verallgemeinerte Inverse

$$\mathbf{G}_r \mathbf{m} = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{V}_r^T = \mathbf{d}$$

Multiplikation mit \mathbf{U}_r^T von links

$$\mathbf{U}^T \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{V}_r^T \mathbf{m} = \mathbf{\Lambda}_r \mathbf{V}_r^T \mathbf{m} = \mathbf{U}^T \mathbf{d}$$

Multiplikation mit $\mathbf{\Lambda}_r^{-1}$ von links

$$\mathbf{\Lambda}_r^{-1} \mathbf{\Lambda}_r \mathbf{V}_r^T \mathbf{m} = \mathbf{V}_r^T \mathbf{m} = \mathbf{\Lambda}_r^{-1} \mathbf{U}^T \mathbf{d}$$

Multiplikation mit \mathbf{V}_r von links

$$\mathbf{V}_r \mathbf{V}_r^T \mathbf{m} = \mathbf{m} = \mathbf{V}_r \mathbf{\Lambda}_r^{-1} \mathbf{U}^T \mathbf{d}$$

Die SVD definiert eine verallgemeinerte (Pseudo) Inverse (Moore-Penrose-Inverse)

$$\mathbf{G}^\dagger = \mathbf{V}_r \mathbf{\Lambda}_r^{-1} \mathbf{U}_r^T$$

Diese entspricht im überbestimmten Fall ($N \geq M = r$) den Least-Squares.

Auflösungsmatrizen und SVD

Durch Einsetzen der verallgemeinerten Inversen

$$\mathbf{G}^\dagger = \mathbf{V}_r \mathbf{\Lambda}_r^{-1} \mathbf{U}_r^T$$

ergibt sich für die Modellauflösung

$$\mathbf{R}^M == \mathbf{G}^\dagger \mathbf{G} = \mathbf{V}_r \mathbf{\Lambda}_r^{-1} \mathbf{U}_r^T \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{V}_r^T = \mathbf{V}_r \mathbf{V}_r^T$$

sowie für die Informationsdichtematrix

$$\mathbf{R}^D = \mathbf{G} \mathbf{G}^\dagger = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{V}_r^T \mathbf{V}_r \mathbf{\Lambda}_r^{-1} \mathbf{U}_r^T = \mathbf{U}_r \mathbf{U}_r^T$$

Teil 5: Regularisierung

Was bisher geschah

Überbestimmte Probleme

- Methode der kleinsten Quadrate
- Beispiele: Matrixproblem, Lineare Regression
- perfekte Modellauflösung, korrelierte Datenauflösung

Unterbestimmte Probleme

- Keine eindeutige Lösung: zusätzliche Forderung
- Minimum-Norm-Lösung
- Auflösungsmatrizen noch anschauen

Singulärwertzerlegung

- klärt Problemtyp über Bestimmung des Rangs \Rightarrow Test-Aufgabe
- verallgemeinerte Inverse für alle Probleme, LS und MN-Lösung Spezialfälle

Inhalt der heutigen Veranstaltung

- Zusammenfassung
- Singulärwertzerlegung und Nullräume
- Auflösungsmatrizen unterbestimmte Probleme (JNB live)
- Rückschau Aufgabe Problemtypen (Strahlentomographie)
- Regularisierung
 - ▶ Einführung und Occams Prinzip (pdf)
 - ▶ gedämpfte Inverse, Glattheits-Nebenbedingungen (pdf)
 - ▶ Übung an realistischer Strahlentomographie (JNB)

Singulärwertzerlegung

Holzhammer der Inversion und Analyse der Anatomie

$$\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad \text{mit} \quad \mathbf{S} = \text{diag}(\lambda_i), \mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$$

Datenraum+Nullraum $\mathbf{U} = (\mathbf{U}_r, \mathbf{U}_0)$, Modellraum+Nullraum $\mathbf{V} = (\mathbf{V}_r, \mathbf{V}_0)$, Reduktion

$$\mathbf{G}_r = \mathbf{U}_r\mathbf{S}_r\mathbf{V}_r^T$$

Verallgemeinerte Inverse:

$$\mathbf{G}^\dagger = \mathbf{V}_r\mathbf{S}_r^{-1}\mathbf{U}_r^T$$

Modellauflösungsmatrix und Informationsdichtematrix

$$\mathbf{R}^M = \mathbf{V}_r\mathbf{V}_r^T \quad \text{und} \quad \mathbf{R}^D = \mathbf{U}_r\mathbf{U}_r^T$$

Fälle inverser Probleme

Gleichbestimmt $M = N = r$

Es existieren weder Modell- noch Daten-Nullraum: $M = N = r \Rightarrow$ reguläre Inverse, exakte Datenanpassung und Auflösung: $\mathbf{R}^M = \mathbf{R}^D = \mathbf{I}$,

Überbestimmt: $r = M < N$

Es existiert ein Datennulraum: Jedes \mathbf{d}_0 , das aus \mathbf{U}_0 aufgespannt wird

$$\mathbf{m}^0 = \mathbf{G}^\dagger \mathbf{d}^0 = \mathbf{V}_r \mathbf{S}_r^{-1} \mathbf{U}_r^T \mathbf{d}^0 = 0$$

keine exakte Datenanpassung, Least-Squares Lösung mit $\mathbf{R}^D \neq \mathbf{I}$, $\mathbf{R}^M = \mathbf{I}$

$$\mathbf{m}_{LS} = (\mathbf{G}^T \mathbf{G}) \mathbf{G}^T \mathbf{d} = (\mathbf{V}_r \mathbf{S}_r^2 \mathbf{V}_r^T)^{-1} (\mathbf{U}_r \mathbf{S}_r \mathbf{V}_r)^T \mathbf{d} = \mathbf{V}_r \mathbf{S}^{-1} \mathbf{U}_r^T \mathbf{d} = \mathbf{G}^\dagger \mathbf{d}$$

Fälle inverser Probleme

Unterbestimmt: $r = N < M$

Es existiert ein Modell-Nullraum: Jedes \mathbf{m}_0 das aus \mathbf{V}_0 aufgespannt wird

$$\mathbf{d}^0 = \mathbf{G}\mathbf{m}^0 = \mathbf{U}_r \mathbf{S}_r \mathbf{V}_r^T \mathbf{m}^0 = 0$$

keine eindeutige Lösung, $\mathbf{R}^M \neq \mathbf{I}$, $\mathbf{R}^D = \mathbf{I}$, Minimum-Norm-Lösung

$$\mathbf{m}_{MN} = \mathbf{G}^T (\mathbf{G}\mathbf{G}^T)^{-1} \mathbf{d} = \mathbf{V}_r \mathbf{S}_r \mathbf{U}_r^T \mathbf{U}_r \mathbf{S}_r^{-2} \mathbf{U}_r^T \mathbf{d} = \mathbf{V}_r \mathbf{S}^{-1} \mathbf{U}_r^T \mathbf{d} = \mathbf{G}^\dagger \mathbf{d}$$

Gleichzeitig über- und unterbestimmt: $r < M$, $r < N$

Es existieren sowohl Modell- als auch Daten-Nullräume, $\mathbf{R}^D \neq \mathbf{I}$ und $\mathbf{R}^M \neq \mathbf{I}$
weder Least-Squares noch Minimum-Norm anwendbar, weil $\mathbf{G}^T \mathbf{G}$ und $\mathbf{G}\mathbf{G}^T$ singulär

Über-/Unter-Bestimmtheit beschreibt unabhängige Informationen (r) verglichen mit M und N !

Auswertung Problemtypen Strahlentomographie

Was haben wir aus den Problemtypen gelernt?

- formell oft gleich oder überbestimmt ($M \geq N$), aber
- Daten oft redundant
trotz $N > M$ Unterbestimmtheit eines Teils des Modells
- Rang der Matrix ausschlaggebend für Problemtyp
- LS-Lösung oder MN-Lösung nur selten anwendbar oder unsinnig
- SVD-Inverse ($\text{pinv}(G)$) immer anwendbar (Holzhammer)
- unbestimmte Parameter 0, unterbestimmte Parameter gemittelt
- Lösungen oft weit weg von synthetischem Modell
-

Problem kleiner Singulärwerte

Inversion mit verallgemeinerter Inversen (auch LS, MN!)

$$\mathbf{m} = \mathbf{G}^\dagger \mathbf{d} = \mathbf{V} \mathbf{S}^{-1} \mathbf{U}^T \mathbf{d} = \sum \frac{\mathbf{U}_i^T \mathbf{d}}{s_i} \mathbf{V}_i$$

kleine Singulärwerte führen zu großen Faktoren und haben starken Einfluss auf die Lösung
Rauschen kann sich verstärken und die Lösung instabil machen

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1.1 & 0 & 0 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0.5 & 1 \end{bmatrix} \Rightarrow \mathbf{s} = \begin{bmatrix} 2.05 \\ 1.5 \\ 1.0 \\ 0.05 \end{bmatrix}$$

Lösung: Abgeschnittene SVD-Inverse mit `pinv(G, rtol=0.05)` \Rightarrow Jupyter Notebook

Problem mit Unterbestimmung

3 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{Gm} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{d}$$

Zerlegung in zwei Unterprobleme:

- Parameter 3 ist überbestimmt
- Parameter 1+2 unterbestimmt

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Problem mit Unterbestimmung

3 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{Gm} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{d}$$

Zerlegung in zwei Unterprobleme:

- Parameter 3 ist überbestimmt
- Parameter 1+2 unterbestimmt

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Problem mit Unterbestimmung

3 Messungen (z.B. Strahlen), 3 Parameter (Zellen)

$$\mathbf{Gm} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \mathbf{d}$$

Zerlegung in zwei Unterprobleme:

- Parameter 3 ist überbestimmt
- Parameter 1+2 unterbestimmt

Alle Lösungen mit $m_1 = d_1 - m_2$ sind gleichrichtig

Für eindeutige (reguläre) Lösungen müssen zusätzliche Bedingungen gestellt werden \Rightarrow

Regularisierung

Occams Prinzip

William v. Occam, Schottland 14. Jh.:

Pluralitas non est ponenda sine neccesitate!

Eine Mehrheit darf nie ohne Not zugrunde gelegt werden.

(Wähle aus allen möglichen Lösungen die einfachste)

Doch wie können wir einfach mathematisch definieren?

- wenige Modellzellen (z.B. Schichten)
- große Glattheit
- möglichst geringe Kontraste
- möglichst wenige Kontraste
- Höchste Wahrscheinlichkeit (Bayes)
- Maximum der Entropie/Informationsgehalt

Regularisierung

Wie können wir die Inversion regulär machen?

Zusätzliche Gleichungen im Modellraum

- A-priori-Wissen über eine Unbekannte (Modellreduktion)
ungefähre Schätzung (Referenzmodell) $m_1 = d_3$
- Beziehung zwischen Unbekannten (z.B. Summe zweier Mächtigkeiten) $m_1 + m_2 = d_3$
- Differenz/Glattheit soll klein sein $m_1 - m_2 = d_3$

Zusammen mit Daten im Sinne kleinster Quadrate zu lösen:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Aber: Erste 2 Gleichungen leben im Datenraum, letzte im Modellraum (Einheiten, Rauschen etc.)

Regularisierung

Wie können wir die Inversion regulär machen?

Zusätzliche Gleichungen im Modellraum

- A-priori-Wissen über eine Unbekannte (Modellreduktion)
ungefähre Schätzung (Referenzmodell) $m_1 = d_3$
- Beziehung zwischen Unbekannten (z.B. Summe zweier Mächtigkeiten) $m_1 + m_2 = d_3$
- Differenz/Glattheit soll klein sein $m_1 - m_2 = d_3$

Zusammen mit Daten im Sinne kleinster Quadrate zu lösen:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Aber: Erste 2 Gleichungen leben im Datenraum, letzte im Modellraum (Einheiten, Rauschen etc.)

Regularisierung

Wie können wir die Inversion regulär machen?

Zusätzliche Gleichungen im Modellraum

- A-priori-Wissen über eine Unbekannte (Modellreduktion)
ungefähre Schätzung (Referenzmodell) $m_1 = d_3$
- Beziehung zwischen Unbekannten (z.B. Summe zweier Mächtigkeiten) $m_1 + m_2 = d_3$
- Differenz/Glattheit soll klein sein $m_1 - m_2 = d_3$

Zusammen mit Daten im Sinne kleinster Quadrate zu lösen:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$$

Aber: Erste 2 Gleichungen leben im Datenraum, letzte im Modellraum (Einheiten, Rauschen etc.)

Regularisierung

Wie können wir die Inversion regulär machen?

Zusätzliche Gleichungen im Modellraum

- A-priori-Wissen über eine Unbekannte (Modellreduktion)
- Beziehung zwischen mehreren Unbekannten (z.B. Summe zweier Mächtigkeiten, Differenz/Glattheit)
- ungefähre Schätzung (Referenzmodell)

Zusammen mit Daten im Sinne kleinster Quadrate zu lösen:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Aber: Erste 2 Gleichungen leben im Datenraum, letzte im Modellraum (Einheiten, Rauschen etc.)

Regularisierung

Wie können wir die Inversion regulär machen?

Zusätzliche Gleichungen im Modellraum

- A-priori-Wissen über eine Unbekannte (Modellreduktion)
- Beziehung zwischen mehreren Unbekannten (z.B. Summe zweier Mächtigkeiten, Differenz/Glattheit)
- ungefähre Schätzung (Referenzmodell)

Zusammen mit Daten im Sinne kleinster Quadrate zu lösen:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Aber: Erste 2 Gleichungen leben im Datenraum, letzte im Modellraum (Einheiten, Rauschen etc.)

Regularisierung

Wie können wir die Inversion regulär machen?

Zusätzliche Gleichungen im Modellraum

- A-priori-Wissen über eine Unbekannte (Modellreduktion)
- Beziehung zwischen mehreren Unbekannten (z.B. Summe zweier Mächtigkeiten, Differenz/Glattheit)
- ungefähre Schätzung (Referenzmodell)

Zusammen mit Daten im Sinne kleinster Quadrate zu lösen:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$$

Aber: Erste 2 Gleichungen leben im Datenraum, letzte im Modellraum (Einheiten, Rauschen etc.)

Regularisierung (2)

Minimierung einer gewichteten Summe (Residuum + Constraints):

$$\|\mathbf{G}\mathbf{m} - \mathbf{d}\|^2 + \lambda^2 \|\mathbf{W}\mathbf{m}\|^2 \rightarrow \min$$

λ ist ein Wichtungsfaktor mit Einheit $[\lambda] = [\text{Daten}]/[\text{Modell}]$, führt zu

$$(\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{W}^T \mathbf{W}) \mathbf{m} = \mathbf{G}^T \mathbf{d}$$

Ist identisch zum inversen Problem

$$\begin{bmatrix} \mathbf{G} \\ \lambda \mathbf{W} \end{bmatrix} \mathbf{m} - \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} \rightarrow \min$$

Regularisierung (2)

Minimierung einer gewichteten Summe (Residuum + Constraints):

$$\|\mathbf{G}\mathbf{m} - \mathbf{d}\|^2 + \lambda^2 \|\mathbf{W}\mathbf{m}\|^2 \rightarrow \min$$

λ ist ein Wichtungsfaktor mit Einheit $[\lambda] = [\text{Daten}]/[\text{Modell}]$, führt zu

$$(\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{W}^T \mathbf{W}) \mathbf{m} = \mathbf{G}^T \mathbf{d}$$

- Einfachster Fall: \mathbf{W} ist Einheitsmatrix \mathbf{I} :
gedämpfte Normalengleichungen \Rightarrow kleinstes Modell
- Weiterer häufiger Fall: \mathbf{W} ist diskrete Ableitungsmatrix:
smoothness constraints \Rightarrow glattestes Modell:

Glattheits-Nebenbedingungen (Smoothness Constraints)

Wir minimieren die Rauigkeit, d.h. Gradienten oder Krümmung im Modell.

Beispiel Rauigkeitsoperator 1. Ableitung für 1D-Modell

$$\begin{bmatrix} -1 & 1 & 0 & \dots & \\ 0 & -1 & 1 & 0 & \dots \\ & \vdots & \ddots & \ddots & \vdots \\ & \dots & 0 & -1 & 1 \end{bmatrix}$$

Beispiel Rauigkeitsoperator 2. Ableitung für 1D-Modell

$$\begin{bmatrix} -1 & 2 & -1 & 0 & \dots & \\ 0 & -1 & 2 & -1 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \dots & \dots & 0 & -1 & 2 & -1 \end{bmatrix}$$

Glattheits-Nebenbedingungen (Smoothness Constraints)

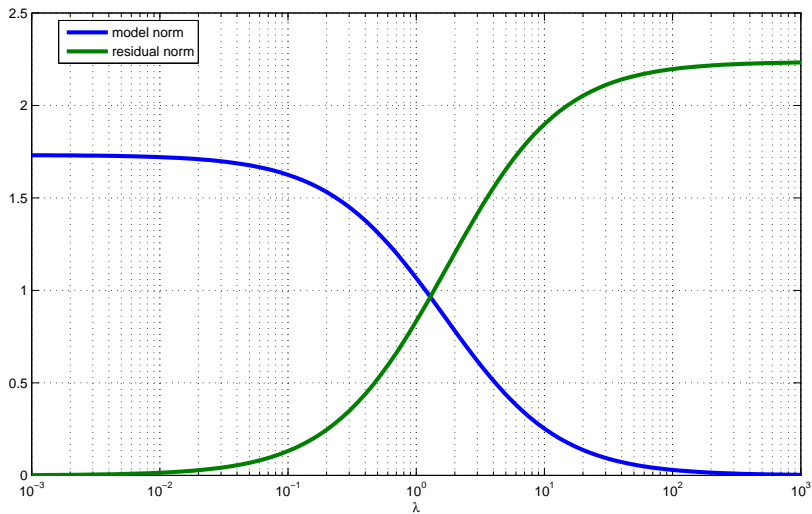
Wir minimieren die Rauigkeit, d.h. Gradienten oder Krümmung im Modell.

Beispiel Rauigkeitsoperator 1. Ableitung für 2D-Modell

$$\begin{bmatrix} -1 & 1 & 0 & \dots & \dots \\ 0 & -1 & 1 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \dots & \dots & 0 & -1 & 1 \\ -1 & 0 & \dots & 1 & 0 \\ \vdots & -1 & \dots & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{W}_x \\ \mathbf{W}_y \end{bmatrix}$$

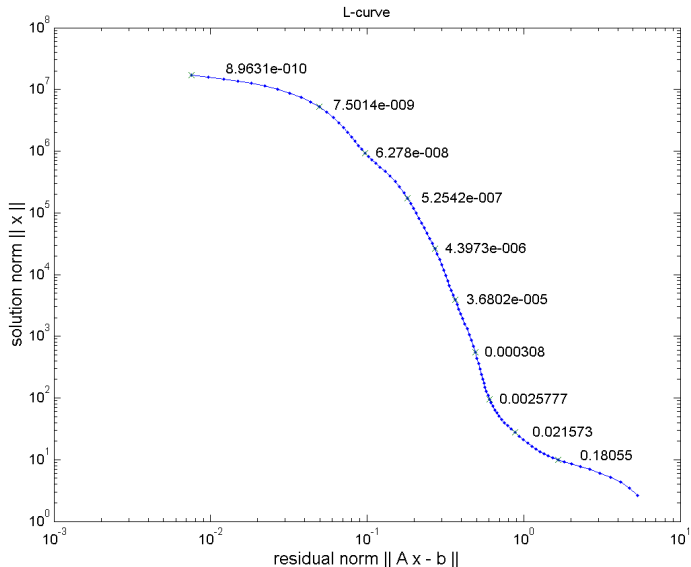
Alternativ: $\|\mathbf{Gm} - \mathbf{d}\|^2 + \lambda_x \|\mathbf{W}_x \mathbf{m}\| + \lambda_y \|\mathbf{W}_y \mathbf{m}\| \rightarrow \min$

Wahl des Regularisierungsparameters λ



Kompromiss zwischen Datenanpassung und Modellnorm

Wahl des Regularisierungsparameters λ



Wahl des Regularisierungsparameters λ

Das Diskrepanzprinzip

Wähle λ so, dass die Daten im Rahmen ihrer Fehler angepasst werden ($\chi^2 = 1$):

$$\min \|\mathbf{Wm}\|_2^2 \quad \text{subject to} \quad \|\hat{\mathbf{Gm}} - \hat{\mathbf{d}}\|_2^2 = N$$

Auflösung für regularisierte Inversion

generalisierte Inverse:

$$\mathbf{G}^\dagger = (\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{W}^T \mathbf{W})^{-1} \mathbf{G}^T$$

Modell-Auflösung:

$$\mathbf{R}^M = \mathbf{G}^\dagger \mathbf{G} = (\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{W}^T \mathbf{W})^{-1} \mathbf{G}^T \mathbf{G}$$

nähert sich Einheitsmatrix \mathbf{I} für $\lambda \rightarrow 0$

Daten-Auflösung:

$$\mathbf{R}^D = \mathbf{G} \mathbf{G}^\dagger = \mathbf{G} (\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{W}^T \mathbf{W})^{-1} \mathbf{G}^T$$

Referenzmodell-Inversion

Oft macht kleinstes Modell wenig Sinn.

Dann invertiert man oft Modelländerungen $\Delta \mathbf{m} = \mathbf{m} - \mathbf{m}^R$

$$\mathbf{G}\Delta \mathbf{m} = \Delta \mathbf{d} = \mathbf{d} - \mathbf{G}\mathbf{m}^R$$

und verwendet die gedämpften Normalengleichungen (Abstand zu Referenzmodell wird minimiert)
Dadurch werden smoothness constraints bewusst vermieden (z.B. bei Timelapse-Inversion sehr kleiner Änderungen)

Teil 6: Strahlentomographie

Was bisher geschah

Singulärwertzerlegung

- Unterteilung in Modell- bzw. Daten-Raum und Nullräume
- klärt Problemtyp über Bestimmung des Rangs \Rightarrow Test-Aufgabe
- verallgemeinerte Inverse für alle Probleme, LS und MN-Lösung Spezialfälle
- Modell-Konstruktion aus Modellvektoren (Wichtung Projektion auf Datenraum)
- kleine Singulärwerte haben großen Einfluss (Verstärkung Rauschen)
 \Rightarrow Reduktion Pseudorang bzw. rtol für svd und pinv

Regularisierung

- zusätzliche Gleichungen im Modellraum (Achtung Wichtung)
- gewichtete Minimierung von Residuum und Modellnorm
- Auswahl des Regularisierungsparameters (L-Kurve, Diskrepanzprinzip)

Inhalt der heutigen Veranstaltung

- Grundlagen (pdf)
 - ▶ Wiederholung Regularisierung
 - ▶ Zusammenhang gedämpfte Least-Squares zu SVD
 - ▶ Die Kovarianz-Matrix
- kurzer Überblick pyGIMLi (live) und Berechnung Wegmatrix
- Strahlentomographie (live JNB)
- Übung Regularisierung am Beispiel (selbst JNB)
- bis nächste Veranstaltung: Ausgabe der Belegaufgabe 1

Regularisierung

Minimierung einer gewichteten Summe (Residuum + Constraints):

$$\|\mathbf{G}\mathbf{m} - \mathbf{d}\|^2 + \lambda^2 \|\mathbf{W}\mathbf{m}\|^2 \rightarrow \min$$

λ ist ein Wichtungsfaktor mit Einheit $[\lambda] = [\text{Daten}]/[\text{Modell}]$, führt zu

$$(\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{W}^T \mathbf{W}) \mathbf{m} = \mathbf{G}^T \mathbf{d}$$

Ist identisch zum inversen Problem

$$\begin{bmatrix} \mathbf{G} \\ \lambda \mathbf{W} \end{bmatrix} \mathbf{m} - \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} \rightarrow \min$$

Regularisierung

Minimierung einer gewichteten Summe (Residuum + Constraints):

$$\|\mathbf{G}\mathbf{m} - \mathbf{d}\|^2 + \lambda^2 \|\mathbf{W}\mathbf{m}\|^2 \rightarrow \min$$

λ ist ein Wichtungsfaktor mit Einheit $[\lambda] = [\text{Daten}]/[\text{Modell}]$, führt zu

$$(\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{W}^T \mathbf{W}) \mathbf{m} = \mathbf{G}^T \mathbf{d}$$

- Einfachster Fall: \mathbf{W} ist Einheitsmatrix \mathbf{I} :
gedämpfte Normalengleichungen \Rightarrow kleinstes Modell
- Weiterer häufiger Fall: \mathbf{W} ist diskrete Ableitungsmatrix:
smoothness constraints \Rightarrow glattestes Modell:

Glattheits-Nebenbedingungen (Smoothness Constraints)

Wir minimieren die Rauhigkeit, d.h. Gradienten oder Krümmung im Modell.

Beispiel Rauigkeitsoperator 1. Ableitung für 1D-Modell

$$\begin{bmatrix} -1 & 1 & 0 & \dots & \\ 0 & -1 & 1 & 0 & \dots \\ & \vdots & \ddots & \ddots & \vdots \\ & \dots & 0 & -1 & 1 \end{bmatrix}$$

Beispiel Rauigkeitsoperator 2. Ableitung für 1D-Modell

$$\begin{bmatrix} -1 & 2 & -1 & 0 & \dots & \\ 0 & -1 & 2 & -1 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \dots & \dots & 0 & -1 & 2 & -1 \end{bmatrix}$$

Glattheits-Nebenbedingungen (Smoothness Constraints)

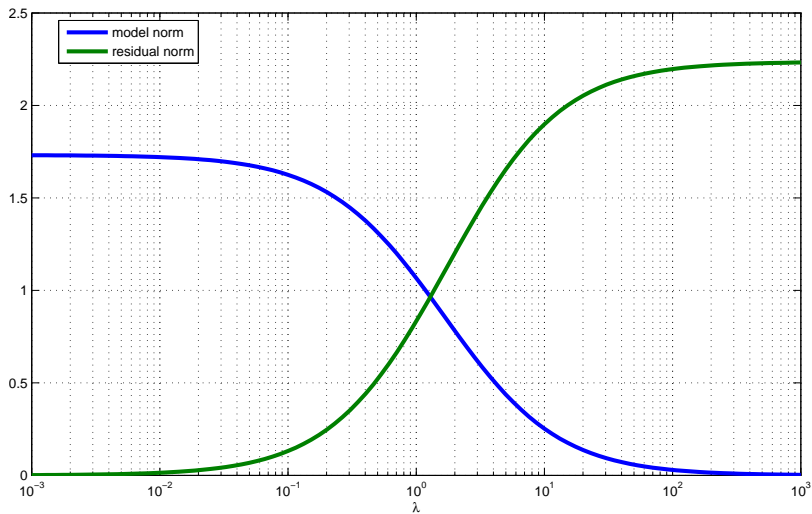
Wir minimieren die Rauigkeit, d.h. Gradienten oder Krümmung im Modell.

Beispiel Rauigkeitsoperator 1. Ableitung für 2D-Modell

$$\begin{bmatrix} -1 & 1 & 0 & \dots & \dots \\ 0 & -1 & 1 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \dots & \dots & 0 & -1 & 1 \\ -1 & 0 & \dots & 1 & 0 \\ \vdots & -1 & \dots & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{W}_x \\ \mathbf{W}_y \end{bmatrix}$$

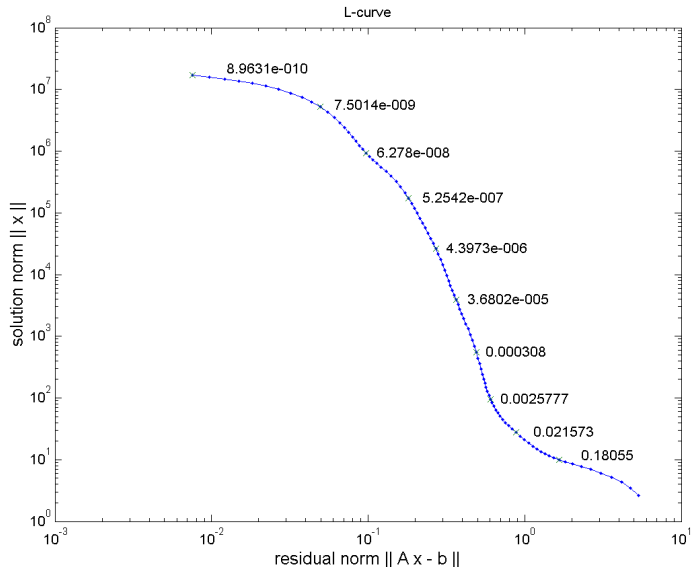
Alternativ: $\|\mathbf{Gm} - \mathbf{d}\|^2 + \lambda_x \|\mathbf{W}_x \mathbf{m}\| + \lambda_y \|\mathbf{W}_y \mathbf{m}\| \rightarrow \min$

Wahl des Regularisierungsparameters λ



Kompromiss zwischen Datenanpassung und Modellnorm

Wahl des Regularisierungsparameters λ



Wahl des Regularisierungsparameters λ

Das Diskrepanzprinzip

Wähle λ so, dass die Daten im Rahmen ihrer Fehler angepasst werden ($\chi^2 = 1$):

$$\min \|\mathbf{Wm}\|_2^2 \quad \text{subject to} \quad \|\hat{\mathbf{Gm}} - \hat{\mathbf{d}}\|_2^2 = N$$

z.B. Such das glatteste Modell (größte λ) das die Daten (gerade noch so) anpassen kann.

Beziehung zwischen gedämpften Normalgleichungen und SVD

$$\mathbf{m} = (\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{I})^{-1} \mathbf{G}^T \mathbf{d}$$

$$\mathbf{m} = (\mathbf{V} \mathbf{S} \mathbf{U}^T \mathbf{U} \mathbf{S} \mathbf{V}^T + \lambda^2 \mathbf{I})^{-1} \mathbf{V} \mathbf{S} \mathbf{U}^T \mathbf{d}$$

$$\mathbf{m} = (\mathbf{V} \cdot \text{diag}(s_i^2 + \lambda^2) \mathbf{V}^T)^{-1} \mathbf{V} \mathbf{S} \mathbf{U}^T \mathbf{d}$$

$$\mathbf{m} = \sum_i \frac{s_i}{s_i^2 + \lambda^2} \mathbf{U}_i^T \mathbf{d} \mathbf{V}_i = \sum_i \frac{s_i^2}{s_i^2 + \lambda^2} \frac{\mathbf{U}_i^T \mathbf{d}}{s_i} \mathbf{V}_i$$

Die Filterfaktoren $f_i = \frac{s_i^2}{s_i^2 + \lambda^2}$ sorgen für ein geringeres Gewicht $g_i = \frac{s_i}{s_i^2 + \lambda^2}$ kleiner Singulärwerte.
Extremfälle: $\lambda_i \gg s_i \Rightarrow f_i/g_i \rightarrow 0$, $\lambda_i \ll s_i \Rightarrow f_i = 1$, $g_i = 1/s_i$

Modellauflösungsmatrix und Informationsdichtematrix für SVD:

$$\mathbf{R}^M = \mathbf{V}_r \mathbf{V}_r^T \quad \text{und} \quad \mathbf{R}^D = \mathbf{U}_r \mathbf{U}_r^T$$

Beachte: $\mathbf{V}^T \mathbf{V} = \mathbf{I}$ und $\mathbf{V}_r^T \mathbf{V}_r = \mathbf{I}$ aber nicht anders herum!

Mit wachsendem Rang r geht \mathbf{R}^M gegen \mathbf{I} .

$$\mathbf{R}^M - \mathbf{I} = \mathbf{V}_r \mathbf{V}_r^T - \mathbf{V} \mathbf{V}^T = -\mathbf{V}_0 \mathbf{V}_0^T$$

Modellauflösungsmatrix und Informationsdichtematrix für SVD:

$$\mathbf{R}^M = \mathbf{V} \cdot \text{diag}(f_i) \cdot \mathbf{V}^T \quad \text{und} \quad \mathbf{R}^D = \mathbf{U} \cdot \text{diag}(f_i) \cdot \mathbf{U}^T$$

Auflösung für regularisierte Inversion

generalisierte Inverse:

$$\mathbf{G}^\dagger = (\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{W}^T \mathbf{W})^{-1} \mathbf{G}^T$$

Modell-Auflösung:

$$\mathbf{R}^M = \mathbf{G}^\dagger \mathbf{G} = (\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{W}^T \mathbf{W})^{-1} \mathbf{G}^T \mathbf{G} = \mathbf{V} \cdot \text{diag}(f_i) \cdot \mathbf{V}^T$$

nähert sich Einheitsmatrix \mathbf{I} für $\lambda \rightarrow 0$

Daten-Auflösung:

$$\mathbf{R}^D = \mathbf{G} \mathbf{G}^\dagger = \mathbf{G} (\mathbf{G}^T \mathbf{G} + \lambda^2 \mathbf{W}^T \mathbf{W})^{-1} \mathbf{G}^T = \mathbf{U} \cdot \text{diag}(f_i) \cdot \mathbf{U}^T$$

Auflösungsradius

Die Hauptdiagonalen-Elemente definieren die Rekonstruierbarkeit der Parameterzellen

Beispiel

Ein Wert von 0.25 bedeutet, dass nur der Mittelwert über 2x2 Zellen bestimmt werden kann.
(umgekehrt proportional zur Zellgröße des auflösbaren Bereichs)

Ableitung eines äquivalenten Radius

$$R_{ii}^M = \frac{A_{Zelle}}{A_{Bereich}} = \frac{A}{\pi r_{res}^2} \quad \Rightarrow \quad r_{res} = \sqrt{\frac{A}{\pi R_{ii}^M}}$$

Damit erhalten wir ein geometrisches Maß für die Auflösbarkeit kleiner Anomalien oder Grenzen

Parameter-Kovarianz

Theorem

Sei \mathbf{x} ein multivariabler, normalverteilter Zufallsvektor mit dem Erwartungswert μ und der Kovarianz \mathbf{C} und sei $\mathbf{y} = \mathbf{Ax}$. Dann ist \mathbf{y} ebenfalls ein multivariabler, normalverteilter Zufallsvektor mit dem Erwartungswert $E(\mathbf{y}) = \mathbf{A}\mu$ und der Kovarianz $\text{cov}(\mathbf{y}) = \mathbf{ACA}^T$

Inverse Probleme

$$E(\mathbf{m}) = E(\mathbf{G}^\dagger \mathbf{d}) = \mathbf{G}^\dagger E(\mathbf{d}) = \mathbf{R}^M \mathbf{m}^{true}$$
$$\text{cov}(\mathbf{m}) = \mathbf{G}^\dagger \cdot \text{cov}(\mathbf{d})(\mathbf{G}^\dagger)^T$$

Beispiel Least-Squares mit einheitlicher Datenvarianz σ

$$\text{cov}(\mathbf{m}) = \sigma^2 (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{G} (\mathbf{G}^T \mathbf{G})^{-1} = \sigma^2 (\mathbf{G}^T \mathbf{G})^{-1}$$

Trade-off zwischen Kovarianz und Bias

Kovarianz

Bei Verwendung fehlergewichteter Daten ($\text{cov}=\mathbf{I}$) ergibt sich

$$\text{cov}(\mathbf{m}) = \mathbf{G}^\dagger (\mathbf{G}^\dagger)^T$$

$$\text{cov}(\mathbf{m}) = \mathbf{V}_r \mathbf{S}^{-2} \mathbf{V}_r^T = \sum_i \frac{V_i V_i^T}{s_i^2}$$

Bias = systematische Abweichung

$$E(\mathbf{m}^{true}) - \mathbf{m}^{true} = \mathbf{R}^M \mathbf{m}^{true} - \mathbf{m}^{true} = (\mathbf{R}^M - \mathbf{I}) \mathbf{m}^{true}$$

kleiner Bias \Rightarrow Modellbestimmtheit steigt, aber Kovarianz (Unsicherheit) auch

Inverse Probleme in der Geophysik
Vorlesung (Vertretung K. Spitzer)
TU Bergakademie Freiberg, SS 2020

Teil 7: Abschluss Lineare Probleme mit Strahlentomographie

Thomas Günther (LIAG Hannover)
(Thomas.Guenther@extern.tu-freiberg.de)

8. Juni 2020

Abschluss Lineare Probleme

Inhalt der heutigen Veranstaltung

- Termindiskussion
- Zusammenfassung Lineare Probleme
- Fortsetzung Strahlentomographie, Fragensammlung?
 - ▶ Vorstellung meines Notebooks
 - ▶ einige Umstellungen und Vereinfachungen
 - ▶ An-Lösung der Aufgaben Teile A bis C
 - ▶ Fragestunde?
 - ▶ individuelles Üben in Breakout Rooms
 - ▶ Erreichung Kenntnisstand für eigenständige Lösungen
- Ausblick Belegaufgaben Teil 1: Strahlentomographie mit JNB
 - ▶ Skizzierung der zugrunde liegenden Gleichungen
 - ▶ Inversion synthetischer Daten mit realistischen Größen
 - ▶ Parameter-Studien mit verschiedenen Verfahren

Lineare Inversionsprobleme

Haupt-Werkzeug: SVD

- Analyse des Problemtyps und Verständnis für Aufgabe
⇒ Unterteilung in Modell- bzw. Daten-Raum und Nullräume
- verallgemeinerte Inverse für alle Probleme (LS und MN-Lösung Spezialfälle)
- Modell-Konstruktion aus Modellvektoren (Wichtung Projektion auf Datenraum)
- kleine Singulärwerte haben großen Einfluss (Verstärkung Rauschen)

Regularisierung

- TSVD-Regularisierung durch Reduktion des Rangs (pinv, svd)
- Explizite Regularisierung (gewichtete Minimierung von Residuum und Modellnorm)
 - ▶ Einheitsmatrix: Gedämpfte Normalgleichungen (im Modellraum!)
 - ▶ Ableitungsmatrizen: Smoothness constraints (im Null-Raum!)
- Auswahl des Regularisierungsparameters (L-Kurve, Diskrepanzprinzip)

Lektionen Lineare Inversionsprobleme

Was sollten wir gelernt haben (oder heute zu Ende lernen)

- Überbestimmte und unterbestimmte Probleme weniger Parameter einfach
- L2-Norm-Minimierung ist Verbunden mit Gauss-schem Rauschen
- Inversions-Ergebnisse stehen und fallen mit Stärke des Rauschens
- Wichtung mit Datenfehlern hilft, Robuste Inversion durch Re-Wichtung
- Diskrepanzprinzip bei synthetischen Rechnungen, Fehleranalyse sonst
- mehrdimensionale Probleme sind oft gemischt (und schlecht) gestellt
- kleine Singulärwerte machen Probleme \Rightarrow Regularisierung nötig
(Abschneiden oder Dämpfung von Modellvektoren, Nebenbedingungen)
- Auflösungsmatrizen zeigen Grenzen des Machbaren und Wichtung der Daten