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

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

20. April 2020

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.

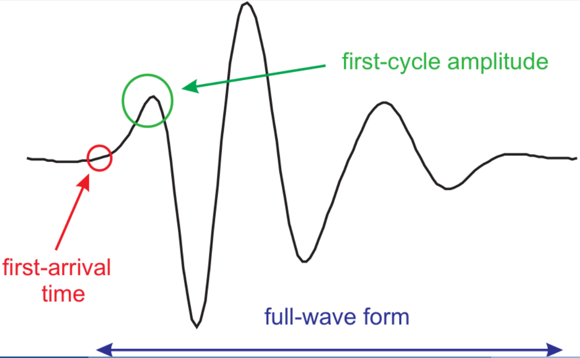
Ziel der Veranstaltung

- 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

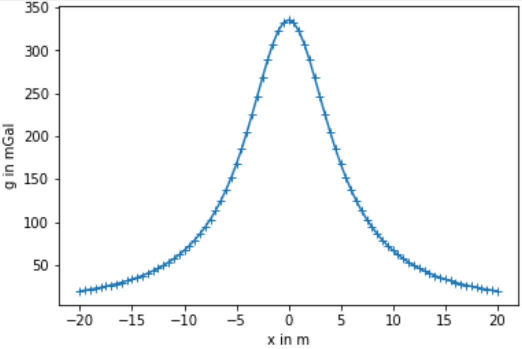
- Seismologie: Zeitreihe von von Beschleunigungswerten
- Gravimetrie: Schwerewerte an diskreten Positionen
- Laufzeittomographie: Laufzeiten zwischen Sender und Empfänger
- Geoelektrik: Ströme und Spannungen f. A-B/M-N Kombinationen



Daten in der Geophysik = diskrete Anzahl von Messwerten

Arten von Daten

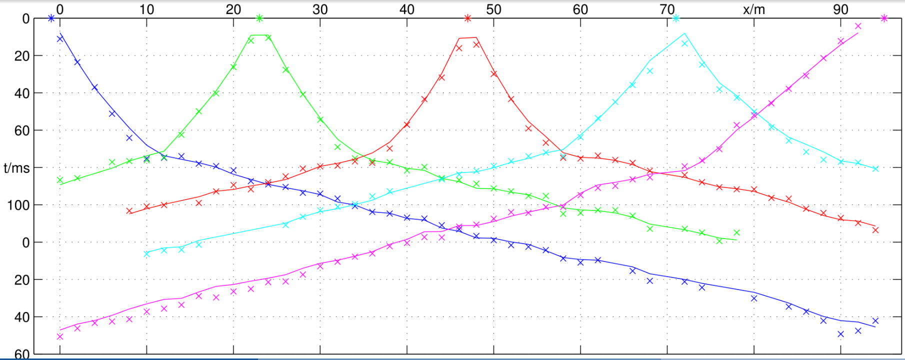
- Seismologie: Zeitreihe von von Beschleunigungswerten
- Gravimetrie: Schwerewerte an diskreten Positionen
- Laufzeittomographie: Laufzeiten zwischen Sender und Empfänger
- Geoelektrik: Ströme und Spannungen f. A-B/M-N Kombinationen



Daten in der Geophysik = diskrete Anzahl von Messwerten

Arten von Daten

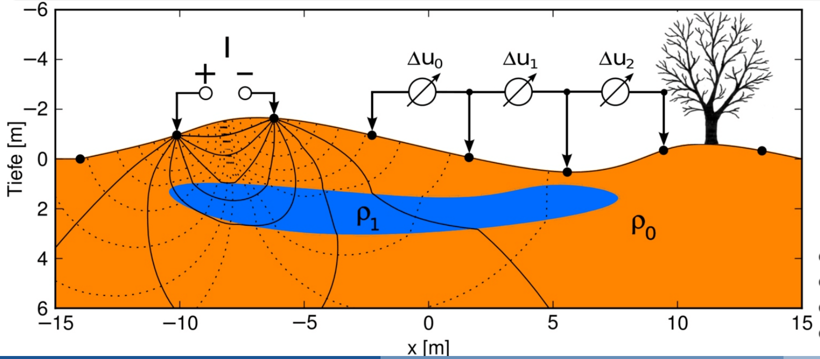
- Seismologie: Zeitreihe von von Beschleunigungswerten
- Gravimetrie: Schwerewerte an diskreten Positionen
- Laufzeittomographie: Laufzeiten zwischen Sender und Empfänger
- Geoelektrik: Ströme und Spannungen f. A-B/M-N Kombinationen

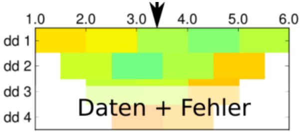


Daten in der Geophysik = diskrete Anzahl von Messwerten

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

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

Modell

Verteilung eines (oder mehrerer) Parameter $p(x, y, z)$
oft diskretisiert: $p_{ijk} \Rightarrow \mathbf{m} = [m_1, m_2, \ldots, m_M]$
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

$\mathbf{f}(\mathbf{m})$ ist linear bezüglich der Modellparameter m_i
$\Rightarrow$ kann als Matrix-Vektorgleichung geschrieben werden

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{n}$$

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

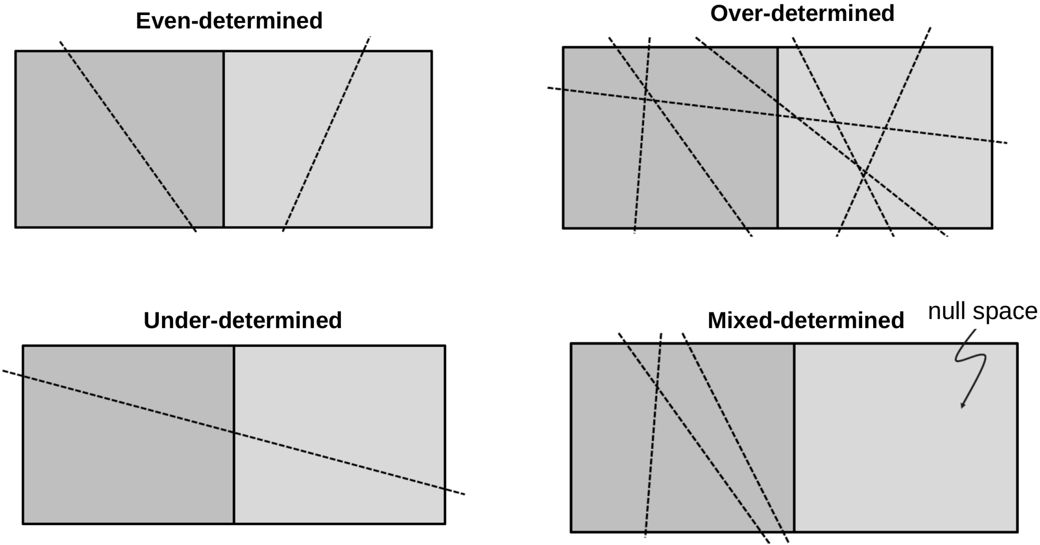
- Matrix **G** ist meist nicht invertierbar
- im Allgemeinen nicht einmal quadratisch

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)

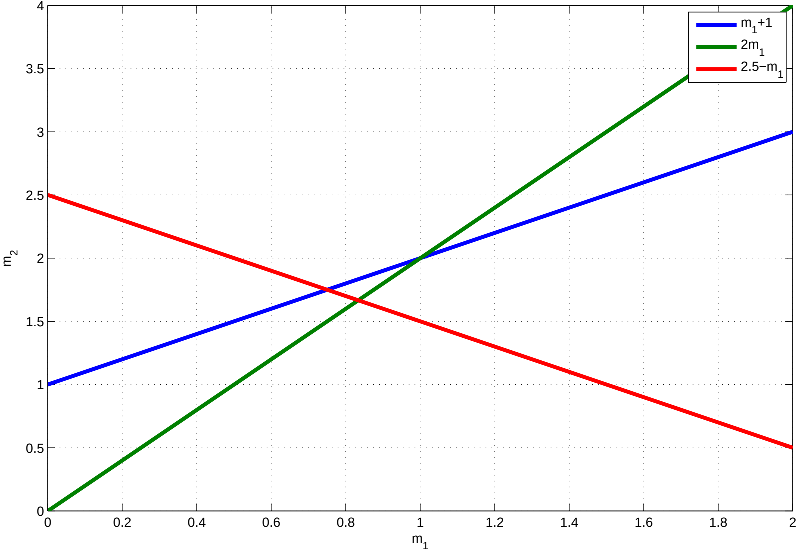


Beispiel überbestimmtes Problem

$$m_1 - m_2 = -1 \tag{1}$$

$$2m_1 - m_2 = 0 \tag{2}$$

$$m_1 + m_2 = 2.5 \tag{3}$$



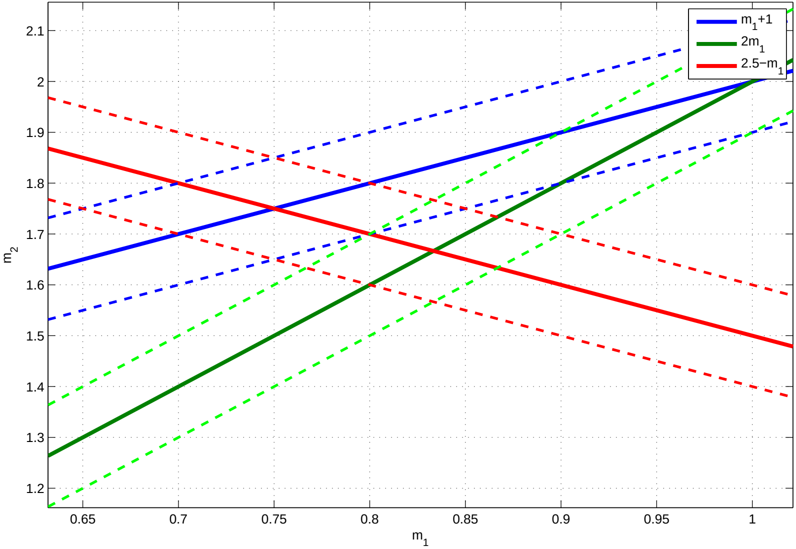
Es gibt mehr unabhängige Gleichungen als Unbekannte.

Beispiel überbestimmtes Problem

$$m_1 - m_2 = -1 \qquad (1)$$

$$2m_1 - m_2 = 0 \qquad (2)$$

$$m_1 + m_2 = 2.5 \qquad (3)$$



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

$$\Phi = \|\mathbf{d} - \mathbf{Gm}\|_2^2 = (\mathbf{d} - \mathbf{Gm})^T(\mathbf{d} - \mathbf{Gm}) = (\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d}) \tag{4}$$

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

$$\frac{\Phi}{m} = \frac{\partial}{\partial m}(\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d}) + (\mathbf{Gm} - \mathbf{d})^T\frac{\partial}{\partial m}(\mathbf{Gm} - \mathbf{d}) = 0 \tag{5}$$

$$\mathbf{G}^T\mathbf{Gm} - \mathbf{G}^T\mathbf{d} + \mathbf{G}^T\mathbf{Gm} - \mathbf{G}^T\mathbf{d} = 0 \tag{6}$$

$$\mathbf{G}^T\mathbf{Gm} = \mathbf{G}^T\mathbf{d} \tag{7}$$

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

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

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:

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

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

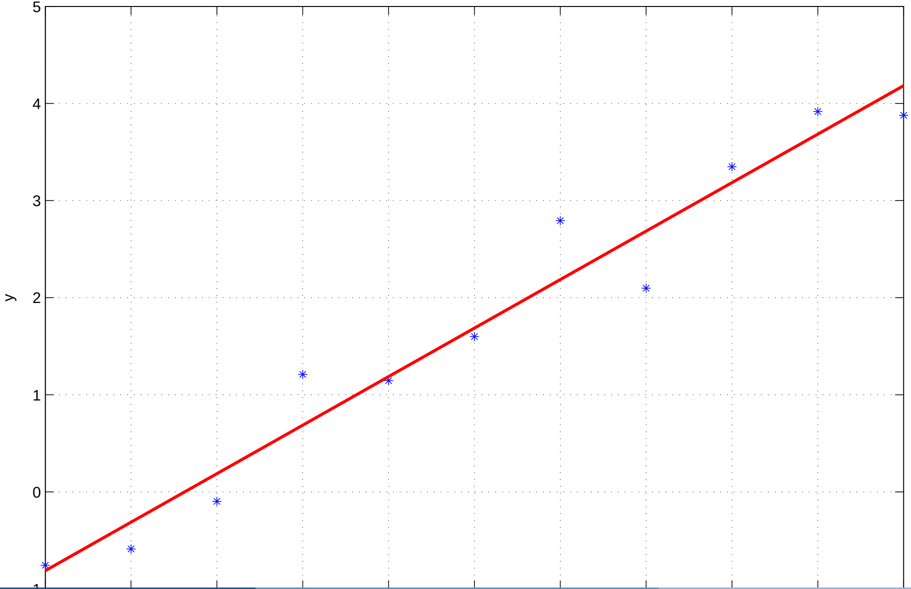
zugehöriges Fehlermaß: fehlergewichteter Misfit (ideal 1)

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

Aufgaben Ausgleichsrechnung

- Bestimmen Sie die Lösung mit der Ausgleichsmethode und das RMS-Fehlermaß.
- Verwenden Sie alternativ die gewichtete Methode mit konstanten Fehlern und geben Sie das χ^2-Fehlermaß an.
- Wie verändert sich die Lösung, wenn Sie das Fehlermodell variieren?
- Variieren Sie die rechten Seiten (Verschiebung der Geraden) oder Koeffizienten.

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 **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!

Wiederholung 1. Veranstaltung

- Lineare Probleme: Vorwärtsoperator $\mathbf{Gm}$
- Daten: Modellantwort plus Fehler $\mathbf{d} = \mathbf{Gm} + \mathbf{n}$
- Überbestimmte Probleme (M>N) $\Rightarrow$ Ausgleichsrechung
  $\Rightarrow$ Minimierung des Residuums $\|\mathbf{d} - \mathbf{Gm}\| \rightarrow \min$
- Least Squares Lösung durch Normalgleichungen:

$$\mathbf{G}^T\mathbf{Gm} = \mathbf{G}^T\mathbf{d}$$

- Matlab denkt mit: m = G \ d
- Maß für Anpassung: Root Mean Square (RMS)

$$\sqrt{1/N \sum (\mathbf{d} - \mathbf{Gm})_i^2} = \|\mathbf{d} - \mathbf{Gm}\| / \sqrt{N}$$

- 3-Geraden-Problem, Lineare Regression

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 ⇒ systematische Verzerrung
- Wichtung der Daten mit reziprokem Fehler
  ⇒ gewichtete Normalgleichungen

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

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

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 $\mathbf{R}^{M} = \mathbf{G}^{\dagger}\mathbf{G}$
$\Rightarrow$ Wie spiegelt sich die Wahrheit ($\mathbf{m}^{\text{true}}$) im Ergebnis ($\mathbf{m}^{\text{est}}$) wider?
Überbestimmte Probleme: $\mathbf{G}\dagger = (\mathbf{G}^{T}\mathbf{G})^{-1}\mathbf{G}^{T}$
$\Rightarrow$ perfekte Modellauflösung

$$\mathbf{R}^{M} = (\mathbf{G}^{T}\mathbf{G})^{-1}\mathbf{G}^{T}\mathbf{G} = \mathbf{I}$$

Daten-Auflösung

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

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 R^D: Informationsgehalt der einzelnen Daten
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(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 **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!

Daten-Auflösung Überbestimmte Probleme

Berechnen Sie für die beiden Beispiel-Probleme (3 Geraden, Lineare Regression) die Datenauflösungsmatrix 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

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 zwei 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 zwei 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 zwei Gleichungen leben im Datenraum, letzte im Modellraum (Einheiten, Rauschen etc.)

Regularisierung (2)

Minimierung einer gewichteten Summe (Residuum + Constraints):

$$\|\mathbf{Gm} - \mathbf{d}\|^2 + \lambda^2 \|\mathbf{Wm}\|^2 \to \min$$

(λ-Wichtungsfaktor mit Einheit [λ]=[Daten]/[Modell])
führt zu

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

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

Occams Prinzip

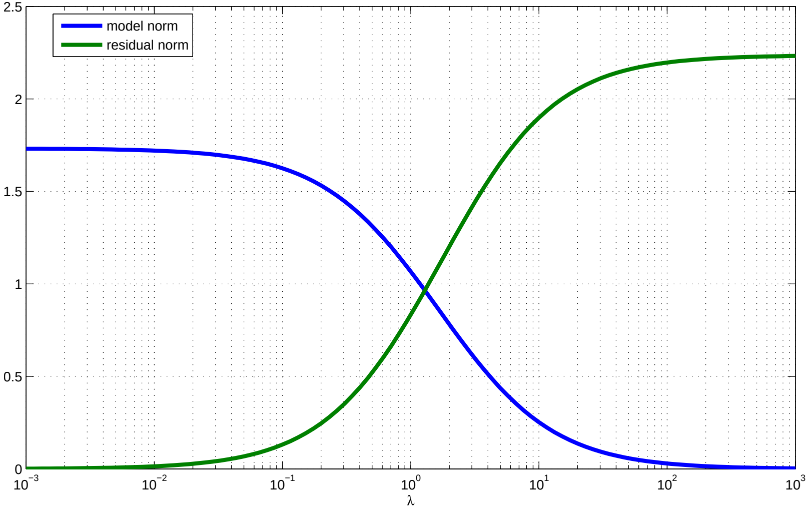
William v. Occam, Schottland 14. Jh.:

Pluralitas non est ponenda sine neccesitate!
Eine Mehrheit darf nie ohne Not zugrunde gelegt werden.
(Wähle aus allen möglichen Lösungen die einfachste)

Doch wie kännen wir einfach mathematisch definieren?

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

Wahl des Regularisierungsparameters



Kompromiss zwischen Datenanpassung und Modellnorm