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

Teil 2: Methode der kleinsten Quadrate und Auflösungsmatritzen

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

27. April 2020

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 $\mathbf{Gm} = \mathbf{d}$
- Korrekt gestelltes Problemes: 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

Was bisher geschah

Inversion = Rekonstruktion Modell aus Daten

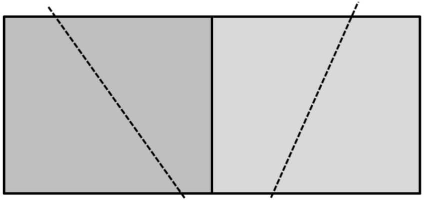
- Übersicht Inversion in der Angewandten Geophysik
- Daten sind (teils zuordenbare) Zahlen mit Fehlern (Vektoren **d**, **e**)
- Modell abstrahiert Untergrund auf wenige Freiheitsgrade (Modell-Vektor **m**)
  (nach Occams Razor: möglichst einfache Beschreibung)
- Lineares Inversionsproblem $\mathbf{Gm} = \mathbf{d} = \mathbf{Gm}^{true} + \mathbf{n}$
- Korrekt gestelltes Problemes: 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

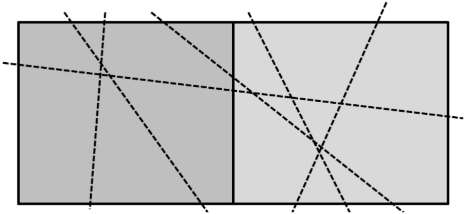
- Zusammenfassung und Fragen
- Troubleshooting Julia und Jupyter Notebooks (live)
- Die Methode der kleinsten Quadrate (pdf)
- Berücksichtigung von Daten-Fehlern
- Fortsetzung minimalistisches Matrix-Problem
- Auflösungsmatritzen: Modellauflösung, Dateninformation
- Übungsbeispiel lineare Regression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

Über- und Unterbestimmtheit (Menke, 2012)

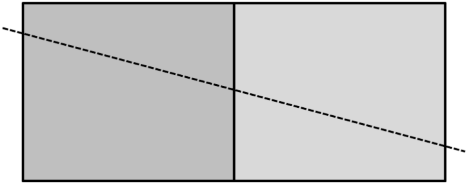
Even-determined



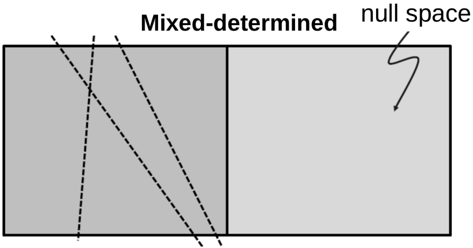
Over-determined



Under-determined

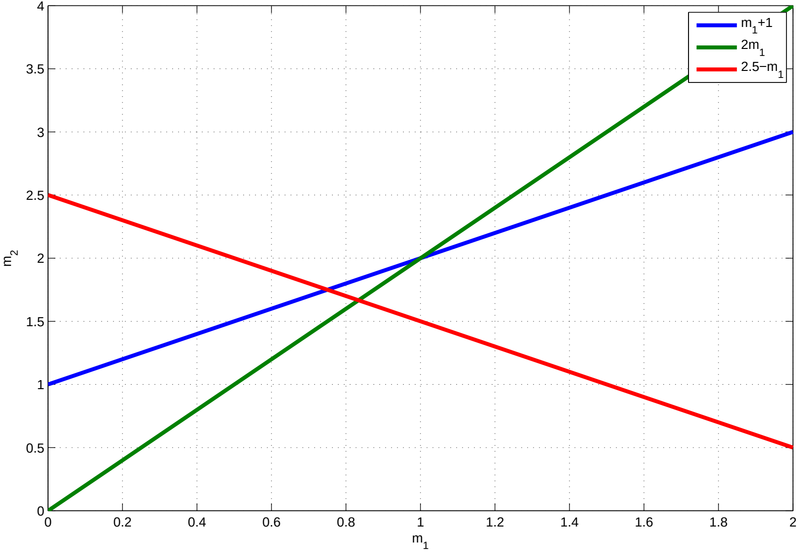


Mixed-determined



Beispiel überbestimmtes Problem

$$m_1 - m_2 = -1 \qquad (1)$$
$$2m_1 - m_2 = 0 \qquad (2)$$
$$m_1 + m_2 = 2.5 \qquad (3)$$



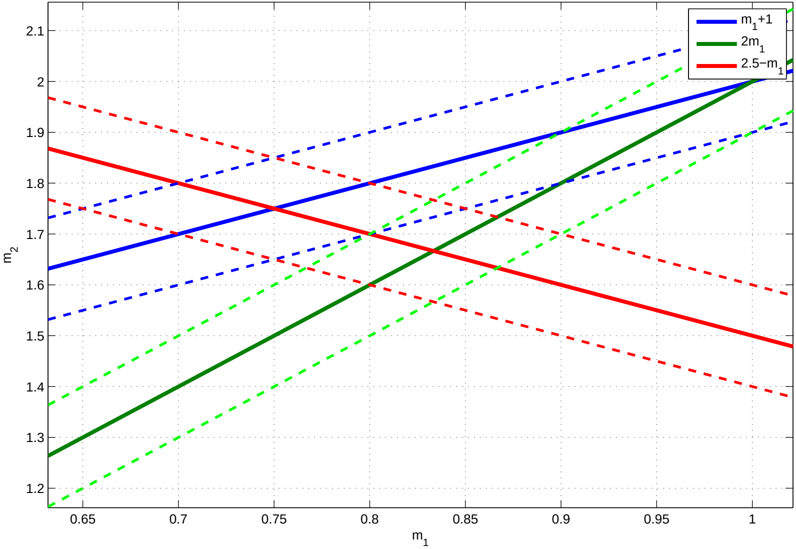
Es gibt mehr unabhängige Gleichungen als Unbekannte.

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

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

Die Funktion Φ wird auch Zielfunktion (objective function) genannt.
Bedingung für ein Extremum ist das Verschwinden der Ableitungen nach allen freien Parametern.

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

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

$$\mathbf{G}^T\mathbf{Gm} = \mathbf{G}^T\mathbf{d} \qquad \Rightarrow \qquad \mathbf{m} = \mathbf{G}^\dagger\mathbf{d} \quad \text{mit} \quad \mathbf{G}^\dagger = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T \tag{7}$$

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

Herleitung

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

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

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

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

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

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

Herleitung (2)

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

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

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

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

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

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

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:

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

(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 (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

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

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

Auflösungsmatritzen

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

Überbestimmte Probleme

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

Auflösungsmatritzen

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):

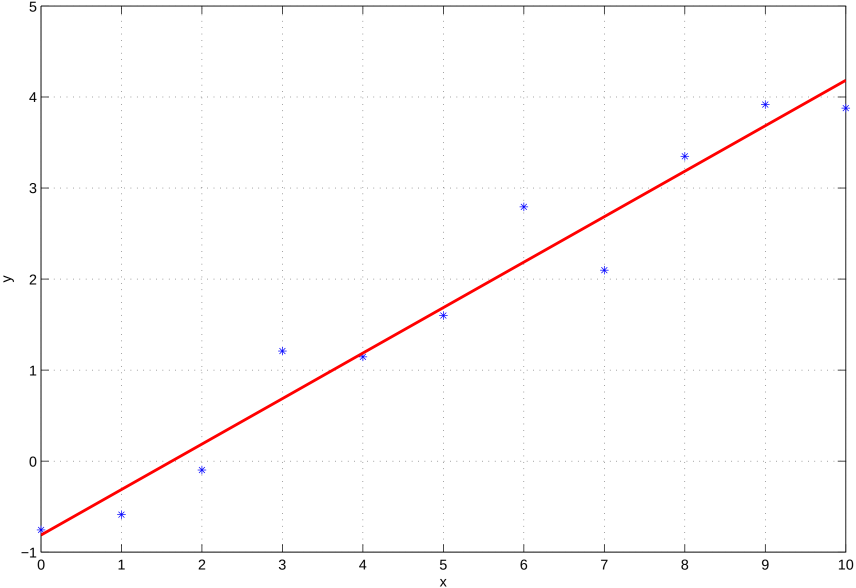
$$\mathbf{R}^{D} = \mathbf{G}\mathbf{G}^{\dagger}$$

Diagonale von 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 **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!

Eigenwertzerlegung

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

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

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 Matritzen existiert eine Faktorisierung mit den EV in Q und den EW in $\Lambda = \mathrm{diag}\lambda_i$

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

Singulärwertzerlegung

Wir machen aus unserer rechteckigen Matrix **G** eine symmetrische Matrx

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

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 **G** und $\mathbf{G}^T$:

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

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

Singulärwertzerlegung

Wir erhalten zwei Eigenwertprobleme für die quadratischen Matritzen

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

Mit den Modell-Eigenvektoren **v** und den Daten-Eigenvektoren **v**:

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

Die Matrix **G** wird aufgespannt durch alle Eigenvektoren:

$$\mathbf{G} = \mathbf{USV} \quad \text{mit} \quad \mathbf{S} = \mathrm{diag}(\lambda_i)$$

$\mathbf{U} \in R^{N\times N}$ enthält die Daten-Eigenvektoren, $\mathbf{V} \in R^{N\times N}$ die Modell-Eigenvektoren
Beide Matritzen sind orthonormal mit $\mathbf{U}^{-1} = \mathbf{U}^T$ und $\mathbf{V}^{-1} = \mathbf{V}^T$. Zu Null ($\lambda_i = 0$) gehörende Vektoren spannen den Modell/Daten-Nullraum auf.

Die verallgemeinerte Inverse

Der Rang der nichtverschwindenen Singulärwerte sei r.
Durch ausschließliche Betrachtung des Daten- und Modellvektorraums, also der r nichtverschwindenen Singulärwerte, erhalten wir einen für unser Problem äquivalenten Operator:

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

Da **U** und **V** orthogonal sind, existiert eine verallgemeinerte Inverse

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

Diese existiert und ist im überbestimmten Fall ($r >= N >= M$) identisch mit der Lösung der Normalengleichung.
Für schlecht gestellte Probleme werden Singulärwerte nicht exakt Null, sondern sehr klein. Dann kann der Rang künstlich verkleinert werden (Pseudorang).

Auflösungsmatritzen und SVD

Durch Einsetzen der verallgemeinerten Inversen ergibt sich

$$\mathbf{m}^{est} = \mathbf{G}^{\dagger}\mathbf{G} = \mathbf{V}_r\Lambda_r^{-1}\mathbf{U}_r^T\mathbf{U}_r\Lambda_r\mathbf{V}_r\mathbf{m}^{true} = \mathbf{V}_r^T\mathbf{V}_r\mathbf{m}^{true}$$

sowie für die Informationsdichtematrix

$$\mathbf{d}^{est} = \mathbf{G}\mathbf{G}^{\dagger} = \mathbf{U}_r\Lambda_r\mathbf{V}_r^T\mathbf{V}_r\Lambda_r^{-1}\mathbf{U}_r^T\mathbf{d} = \mathbf{U}_r^T\mathbf{U}_r\mathbf{d}$$

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):

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

(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

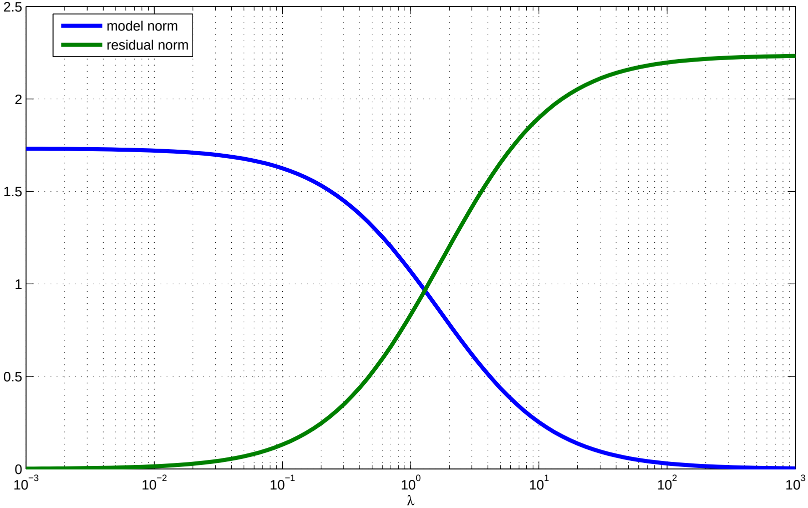
William v. Occam, Schottland 14. Jh.:

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