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

Teil 4: Singulärwertzerlegung

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

11. Mai 2020

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

Auflösungsmatritzen

- Kombination zwischen Vorwärts- und inversem Operator
- Modellauflösung: $\mathbf{G}^\dagger\mathbf{G} \Rightarrow$ perfekt ($\mathbf{I}$) 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 Auflösungsmatritzen
  - Beispiel Bild-Kompression mit SVD (JNB)
  - Anwendung auf bisherige Probleme (JNB)
- Regularisierung: Einführung

Rückschau Lineare Regression

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)

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

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:

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

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

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

woraus wir die Lagrange-Parameter bestimmen können:

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

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

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

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 Matritzen existiert eine Faktorisierung mit den Eigenvektoren in Q (als Spalten) und den Eigenwerten in $\Lambda = \mathrm{diag}(\lambda_i)$

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

Singulärwertzerlegung

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

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

Singulärwertzerlegung als 2 Eigenwertprobleme

Das Eigenwertproblem für die Modell-Eigenvektoren **v**

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

Das Eigenwertproblem für die Daten-Eigenvektoren **u**:

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

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

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

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

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

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

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

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

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

Multiplikation mit Λ_r^{-1} von links

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

Multiplikation mit $\mathbf{V}_r$ von links

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

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

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

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

Auflösungsmatritzen und SVD

Durch Einsetzen der verallgemeinerten Inversen

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

ergibt sich für die Modellauflösung

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

sowie für die Informationsdichtematrix

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

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

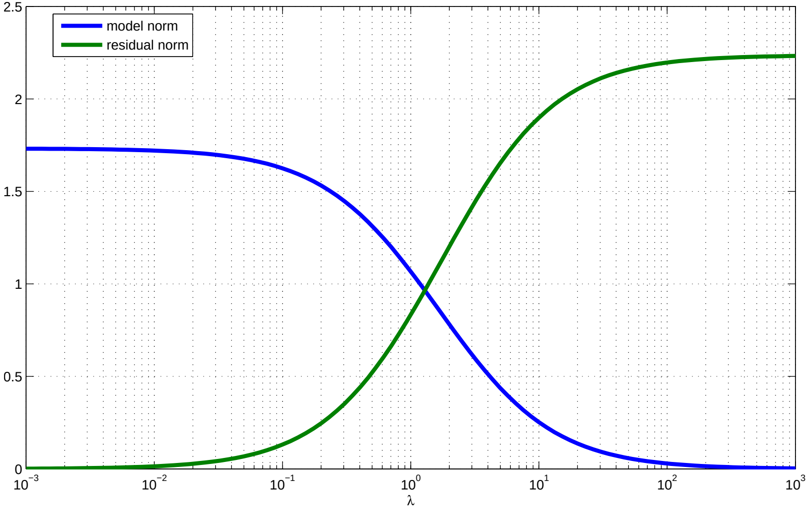
William v. Occam, Schottland 14. Jh.:

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