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

Teil 5: Regularisierung

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

18. Mai 2020

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
- Auflösungsmatritzen 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
- Auflösungsmatritzen 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

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

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 Datennullraum: 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 $R^D \neq \mathbf{I}$, $\mathbf{R}^M = \mathbf{I}$

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

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, $R^M \neq \mathbf{I}$, $\mathbf{R}^D = \mathbf{I}$, 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 (`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

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

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 kännen 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):

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

λ ist ein Wichtungsfaktor mit Einheit [λ]=[Daten]/[Modell], führt zu

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

Ist identisch zum inversen Problem

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

Regularisierung (2)

Minimierung einer gewichteten Summe (Residuum + Constraints):

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

λ ist ein 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:

Glattheits-Nebenbedingungen (Smoothness Constraints)

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

Beispiel Rauhigkeitsoperator 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 Rauhigkeitsoperator 2. Ableitung für 1D-Modell

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

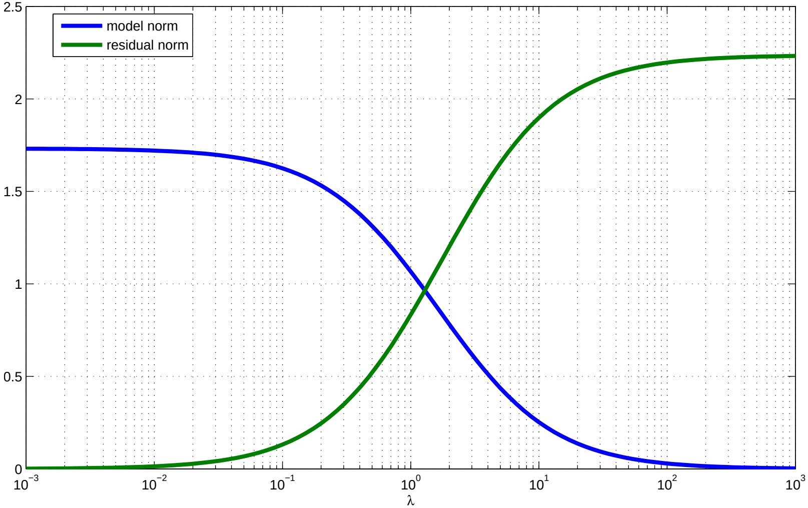
Glattheits-Nebenbedingungen (Smoothness Constraints)

Wir minimieren die Rauhigkeit, d.h. Gradienten oder Krümmung im Modell.
Beispiel Rauhigkeitsoperator 1. Ableitung für 2D-Modell

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

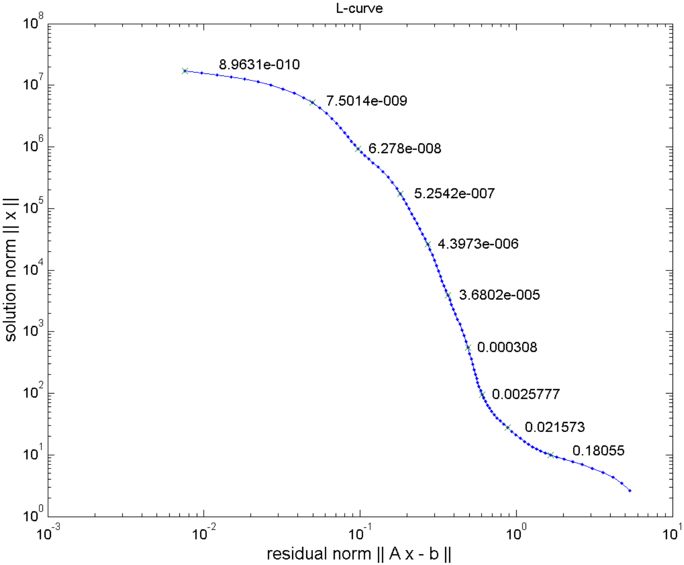
Alternativ: $\|\mathbf{Gm} - \mathbf{d}\|^2 + \lambda_x \|\mathbf{W}_x \mathbf{m}\| + \lambda_y \|\mathbf{W}_y \mathbf{m}\| \to \min$

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 ($\chi^2 = 1$):

$$\min \|\mathbf{Wm}\|_2^2 \qquad \text{subject to} \qquad \|\hat{\mathbf{G}}\mathbf{m} - \hat{\mathbf{d}}\|_2^2 = N$$

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 **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)