

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

Teil 6: Strahlentomographie

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

25. Mai 2020

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 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)

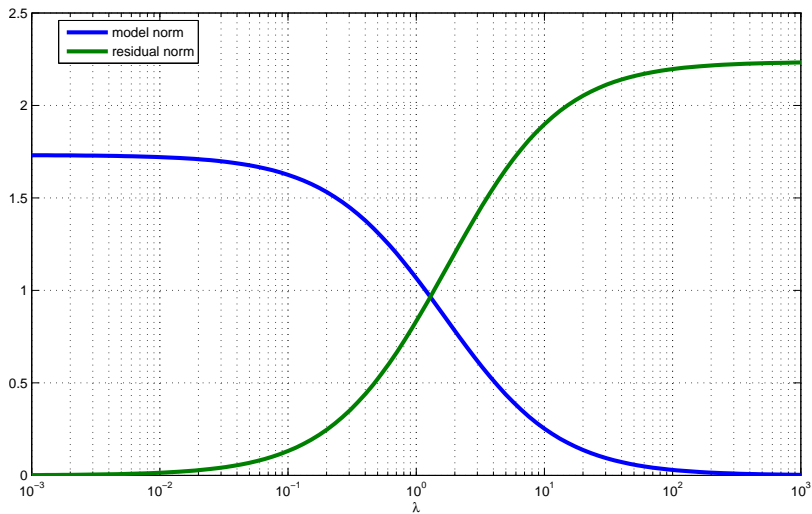
Wir minimieren die Rauigkeit, d.h. Gradienten oder Krümmung im Modell.

Beispiel Rauigkeitsoperator 1. Ableitung für 2D-Modell

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

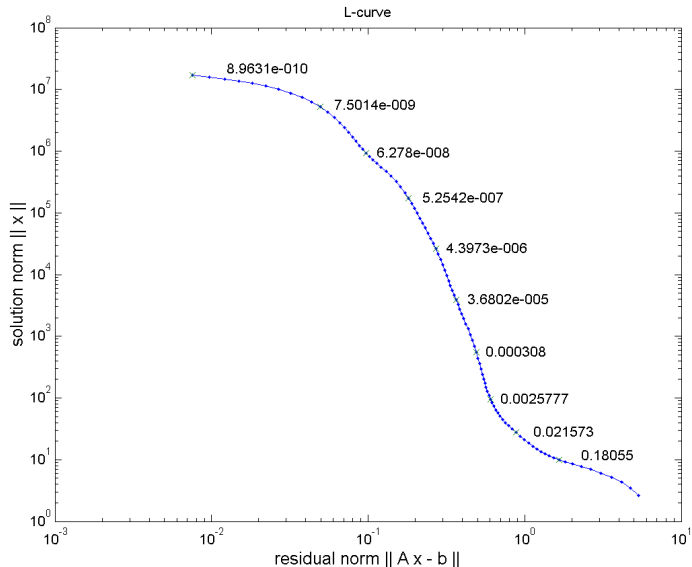
Alternativ: $\|\mathbf{Gm} - \mathbf{d}\|^2 + \lambda_x \|\mathbf{W}_x \mathbf{m}\| + \lambda_y \|\mathbf{W}_y \mathbf{m}\| \rightarrow \min$

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 ($\chi^2 = 1$):

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

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