Inverse Problems in Geophysics
Part 3: Resolution matrices and Under-determined problems

2. MGPY+MGIN

Thomas Günther

thomas.guenther@geophysik.tu-freiberg.de

TU Bergakademie Freiberg
TUBAF
Die Ressourcenuniversität.
Seit 1765.

Recap

- minimization of objective function (L2 norm of error-weighted misfit)
- least-squares solution for over/even-determined problems (= `G\d`)
- forward operator: $\mathbf{f} = \mathbf{G}\mathbf{d}$ $(N \times M)$
- least-squares inverse operator $\mathbf{m}_{LS} = \mathbf{G}^{\dagger}\mathbf{d}$
  with $\mathbf{G}^{\dagger} = (\mathbf{G}^{T}\mathbf{G})^{-1}\mathbf{G}^{T}$ $(M \times M$ x $M \times N$=$M \times N)$
- error weighting by multiplying $\mathbf{G}$ and $\mathbf{d}$ with $\mathrm{diag}(1/\mathbf{e})$

Resolution analysis

Model resolution

$$\mathbf{d} = \mathbf{G}\mathbf{m}^{\text{true}} + \mathbf{n}$$

Matrix inversion with inverse operator $\mathbf{G}\dagger$:

$$\mathbf{m}^{\text{est}} = \mathbf{G}^{\dagger}\mathbf{d} = \mathbf{G}^{\dagger}\mathbf{G}\mathbf{m}^{\text{true}} + \mathbf{G}^{\dagger}\mathbf{n} = \mathbf{R}^{M}\mathbf{m}^{\text{true}} + \mathbf{G}^{\dagger}\mathbf{n}$$

with the model resolution matrix $\mathbf{R}^{M} = \mathbf{G}^{\dagger}\mathbf{G}$

$\Rightarrow$ How is the true model ($\mathbf{m}^{\text{true}}$) reflected in the estimated ($\mathbf{m}^{\text{est}}$)?

Data resolution

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

How are the data explained by the model?

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

with the data resolution (information density) matrix:

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

Diagonal of R^D: information content of individual data

Resolution of least-squares solution

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

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

Important

Over-determined problems have a perfect model resolution. The data are correlated and share the total information content (rank).

$$\Rightarrow \mathbf{R}^D = \mathbf{G}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T$$

Wrap-up overdetermined problems ($N > M$)

- objective function Φ as squared data misfit
- weighting of individual data (unitless, more flexible & objective)
- broad minimum of the objective function
- grid search to plot $\phi \Rightarrow$ only nice for M=2
- least squares method yields
- resolution matrices for model (perfect) and data (distributed)

Under-determined problems

- specifically $N < M$ (too less data for the model pameters), but, more general, if the rank $r < M$

Note

The rank of a matrix is the number of independent columns and rows, i.e. the degree of non-degenerateness.

Example of a mixed-determined problem

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- N<M: clearly under-determined,
- parameter m_3 is perfectly determined (=d_2)
- any combination of m_1 and $m_2 = d_1 - m_1$ fulfils the first equation there is no unique solution (model ambiguity)

Changing the system

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

adding data 1+2

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

does not help, because they are linear dependent ($rk(\mathbf{G}) = 2$).

Minimum-norm solution

Analog to the least-squares inverse $(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T$ there is an inverse for under-determined problems:

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

that is called minimum-norm solution.

Observation from Notebook

The backslash operator (`G\d`) yields the minimum-norm solution.

Derivation

From the models fulfilling $\mathbf{Gm} = \mathbf{d}$ we are looking for the “smallest”

$$\min \Phi = \mathbf{m}^{\mathbf{T}}\mathbf{m} \quad \text{with} \quad \mathbf{Gm} = \mathbf{d}$$

We use the method of Lagrangian parameters (λ_i)

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

Derivation (2)

The derivations vanish

$$\frac{\partial \Phi}{\partial \lambda} = \mathbf{Gm} - \mathbf{d} = \mathbf{0}$$

$$\frac{\partial \Phi}{\partial m} = 2\mathbf{m}^{\mathbf{T}} + \lambda^{\mathbf{T}}\mathbf{G} = \mathbf{0}$$

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

Derivation (3)

$$\Rightarrow \mathbf{d} = \mathbf{Gm} = \mathbf{G}(-\frac{1}{2}\mathbf{G^T}\lambda) = -\frac{1}{2}\mathbf{GG^T}\lambda$$

from which we can determine the Langrangian parameters

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

Replacing them we obtain the minimum-norm solution

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

Resolution of the minimum norm inverse

$$\mathbf{G}^{\dagger} = \mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{G}^{\mathrm{T}})^{-1}$$

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

$$\mathbf{R}^{D} = \mathbf{G}\mathbf{G}^{\dagger} = \mathbf{G}\mathbf{G}^{\mathrm{T}}(\mathbf{G}\mathbf{G}^{\mathrm{T}})^{-1} = \mathbf{I}$$

Note

For underdetermined problems, all data are independent and equally important. Model parameters are insufficiently resolved.

Robustness of inversion

Error weighting

Unweighted residual norm (root-mean square RMS)

$$\|\mathbf{d} - \mathbf{f}(\mathbf{m})\| = \sqrt{1/N \sum (d_i - f_i(\mathbf{m}))^2}$$

Weighting by individual error ϵ_i (chi-square value):

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

In case of exact error estimates: $\chi^2 = 1$

General L_p norm

The L_2 norm is the rooted sum of squared elements of a vector $\mathbf{v}$:

$$\Phi^P = \|\mathbf{v}\|_2 = \sqrt{\sum_i v_i^2}$$

More generally, the L_P norm is computed by

$$\Phi^P = \|\mathbf{v}\|_P = \left(\sum_i |v_i|^P\right)^{1/P}$$

The L_1 norm

The L_1 norm

$$\mathbf{\Phi}^1 = \|\mathbf{v}\|_1 = \sum_i |v_i|$$

is much less prone to outliers as the misfit is not squared

Error weighting

replace d_i by $\hat{d}_i = d_i/\epsilon_i$ leads to

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

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

Derivation (1)

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

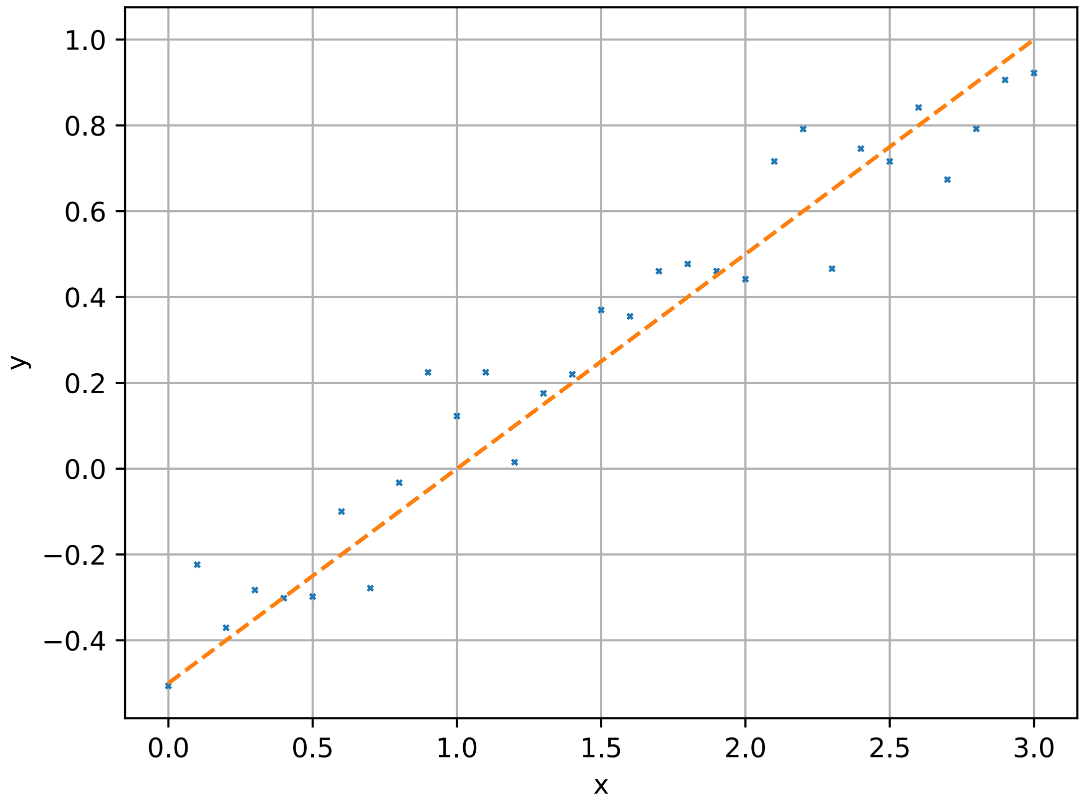
$$\Phi = (\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d})$$

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

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

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

Linear regression



- x - measuring positions
- y - measurements (data)
- $\mathbf{y} = a \cdot \mathbf{x} + b + \mathbf{n}$
- model: slope & intersection $\mathbf{m} = (a, b)^T$
- How is $\mathbf{G}$ looking like? $\mathbf{Gm} = a \cdot \mathbf{x} + b$

The singular value decomposition

Any matrix $\mathbf{A}$ can be decomposed into model and data eigenvectors, weighted by singular values:

$$\mathbf{A} = \sum_{i=1} r\sigma_i \mathbf{u}_i \cdot \mathbf{v}_i^T = \mathbf{U\Sigma V}^T$$

with the eigenvalues in $\Sigma = \mathrm{diag}(\sigma_i)$, a set of orthogonal data eigenvectors $\mathbf{U} \in R^{N\times N}$ with $\mathbf{U}^{-1} = \mathbf{U}^T$, and model eigenvectors $\mathbf{V} \in \mathbb{R}^{M\times M}$ with $\mathbf{V}^{-1} = \mathbf{V}^T$.

Eigenvectors and singular values

The eigenvectors associated with zero singular values ($\sigma_i = 0$) span the null spaces of data and model. The number of non-zero eigenvalues p defines the rank of the matrix which can be abbreviated by

$$\mathbf{A}_p = \sum_{i=1}^{p} \sigma_i \mathbf{u} \cdot \mathbf{v}^T = \mathbf{U}_p \mathbf{\Sigma}_p \mathbf{V}_p^T$$

where $\mathbf{B}_p$ holds the first p columns of the matrix $\mathbf{B}$.

A matrix can be approximated by choosing the rank by hand.

Forward operator in terms of SVD

The forward problem $\mathbf{Gm} = \mathbf{d}$ can be written in terms of SVD

$$\mathbf{U}_p \mathbf{\Sigma}_p \mathbf{V}_p^T \mathbf{m} = \mathbf{d}$$

and rearranged by using $\mathbf{U_p^T U_p = V_p^T V_p = I}$ to

$$\mathbf{\Sigma}_p \mathbf{V}_p^T \mathbf{m} = \mathbf{U}_p^T \mathbf{d}$$

$$\mathbf{V}_p^T \mathbf{m} = \mathbf{\Sigma}_p^{-1} \mathbf{U}_p^T \mathbf{d}$$

$$\mathbf{m} = \mathbf{V}_p \mathbf{\Sigma}_p^{-1} \mathbf{U}_p^T \mathbf{d}$$

Inverse operator in terms of SVD

The generalized inverse $\mathbf{V}_p \mathbf{\Sigma}_p^{-1} \mathbf{U}_p^T$ is called the pseudo-inverse and equals

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

i.e., the least-squares inverse for overdetermined problems.

By further truncating the number of nonzero singular values one can obtain a regularized solution for ill-posed (badly conditioned under-determined or mixed-determined) problems.

Derivation

The quadratic matrix $\mathbf{A}$ projects a vector in another direction. Special vectors are eigenvectors who keep their direction

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

The solution of the equation $\mathbf{A} - \lambda\mathbf{I} = 0$ leads to eigenvalues λ over the characteristic polynome $\det(\mathbf{A} - \lambda\mathbf{I})$ that correspond to eigenvectors in the matrix $\mathbf{Q}$ by

$$\mathbf{A} = \mathbf{Q}\mathrm{diag}(\lambda_i)\mathbf{Q}^T = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T$$

Appendix

The generalized inverse

The matrix

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

is called pseudo-inverse (Moore-Penrose inverse)

The solution $\mathbf{m}_{LS} = \mathbf{G}^{\dagger}\mathbf{d}$ is the least-squares solution

Derivation (2)

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

$$\Phi = \sum_i \Bigg[d_i d_i - d_i \sum_k G_{ik}m_k - d_i \sum_j G_{ij}m_j + \sum_j G_{ij}m_j \sum_k G_{ik}m$$

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

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

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

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

Derivation by change $t\delta\mathbf{m}$

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

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

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

$\Phi(t)$ has a minimum at $t = 0$, therefore $\partial\Phi/\partial t$ must be 0

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

As this holds for every $\delta\mathbf{m}$, we obtain $\mathbf{G}^T\mathbf{G}\mathbf{m} = \mathbf{G}^T\mathbf{d}$