Inverse Problems in Geophysics
Part 6: Regularization

2. MGPY+MGIN

Thomas Günther
thomas.guenther@geophysik.tu-freiberg.de

TUBAF
Die Ressourcenuniversität.
Seit 1765.

Recap

- singular value decomposition (SVD) as fundamental tool
- general solution: the pseudo-inverse (LS & MN special cases)
- small singular values amplify noise in the model
- for ill-conditioned (wide SV range) matrices: TSVD
    - truncate singular values at some point
- regularization: adding equations to make solution unique

Truncated singular value (TSVD) method

- Look at the singular value spectrum
- Choose a maximum number p or `rtol` and compute (e.g. by `pinv`)

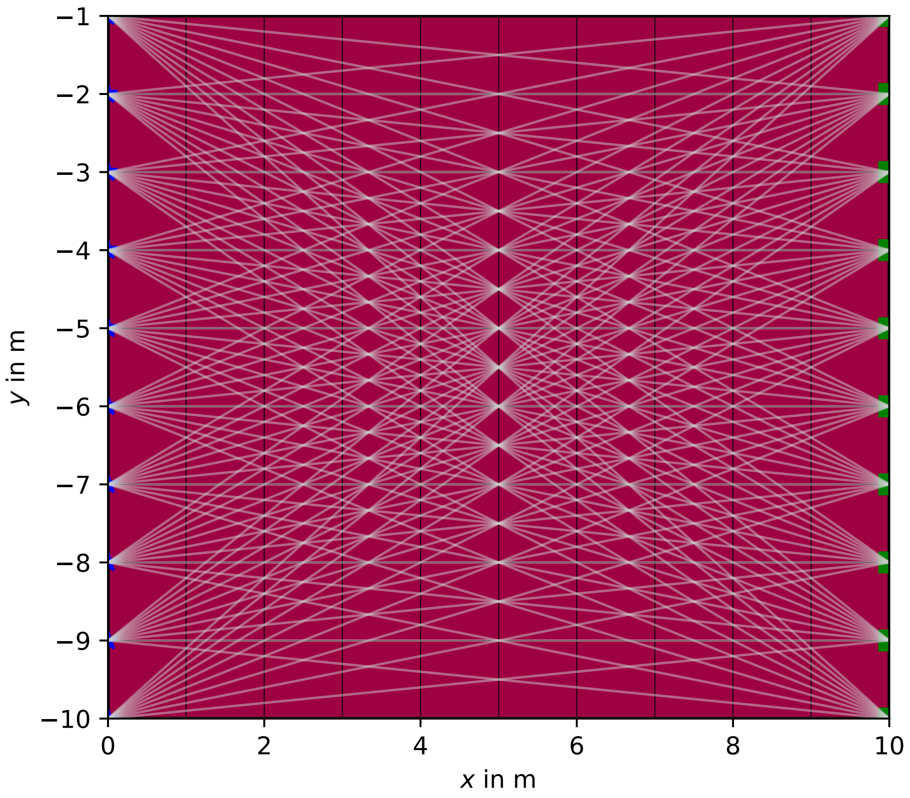
$$\mathbf{G}_p^\dagger = \mathbf{V}_p \mathbf{\Sigma}_p^{-1} \mathbf{U}_p^T$$

- How to choose p? Trade-off between resolution and artifacts.

Discrepancy principle (free after Occam)

Look at data fit and choose p such that the data can be fitted within the error.

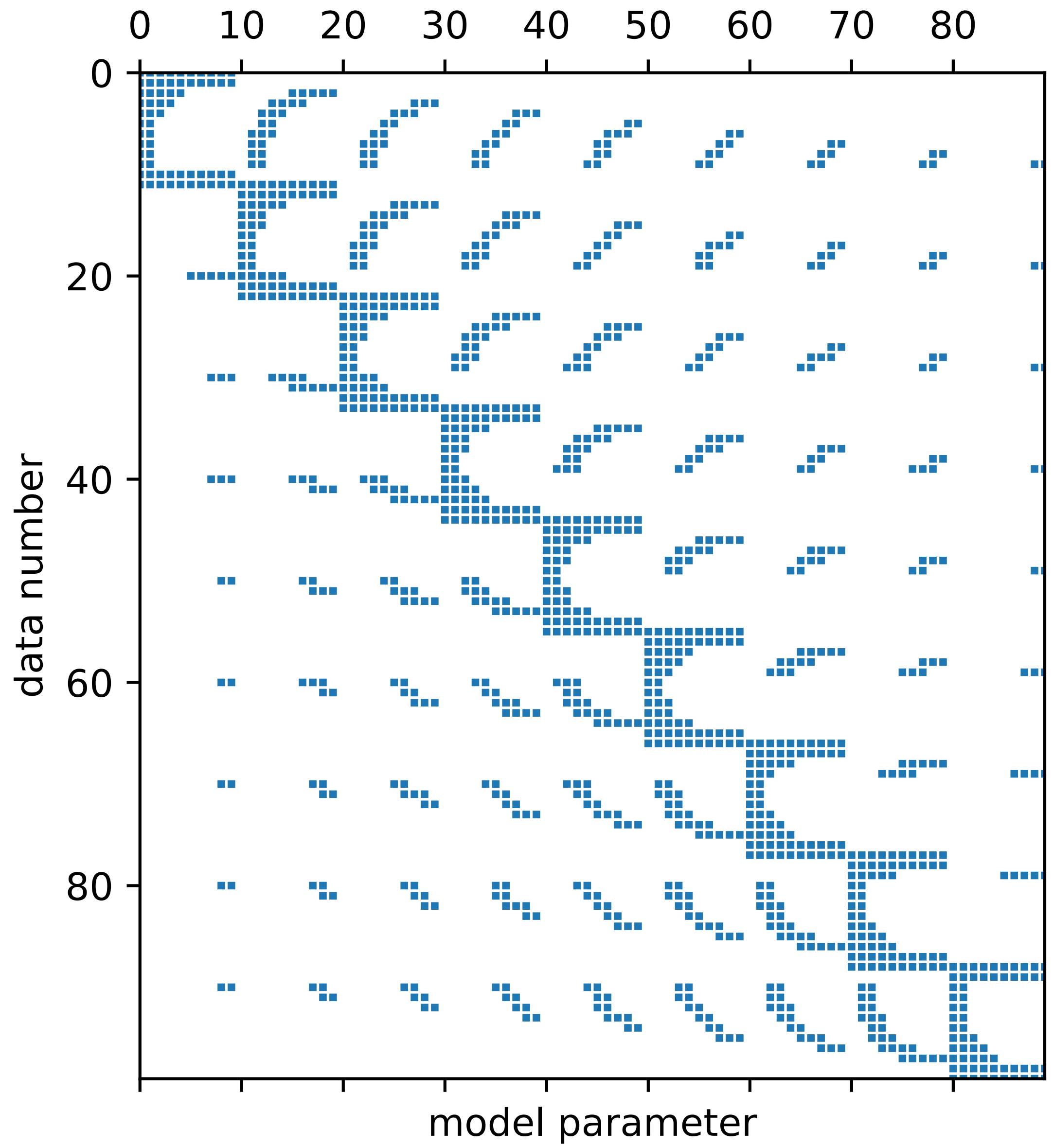
Geophysical example



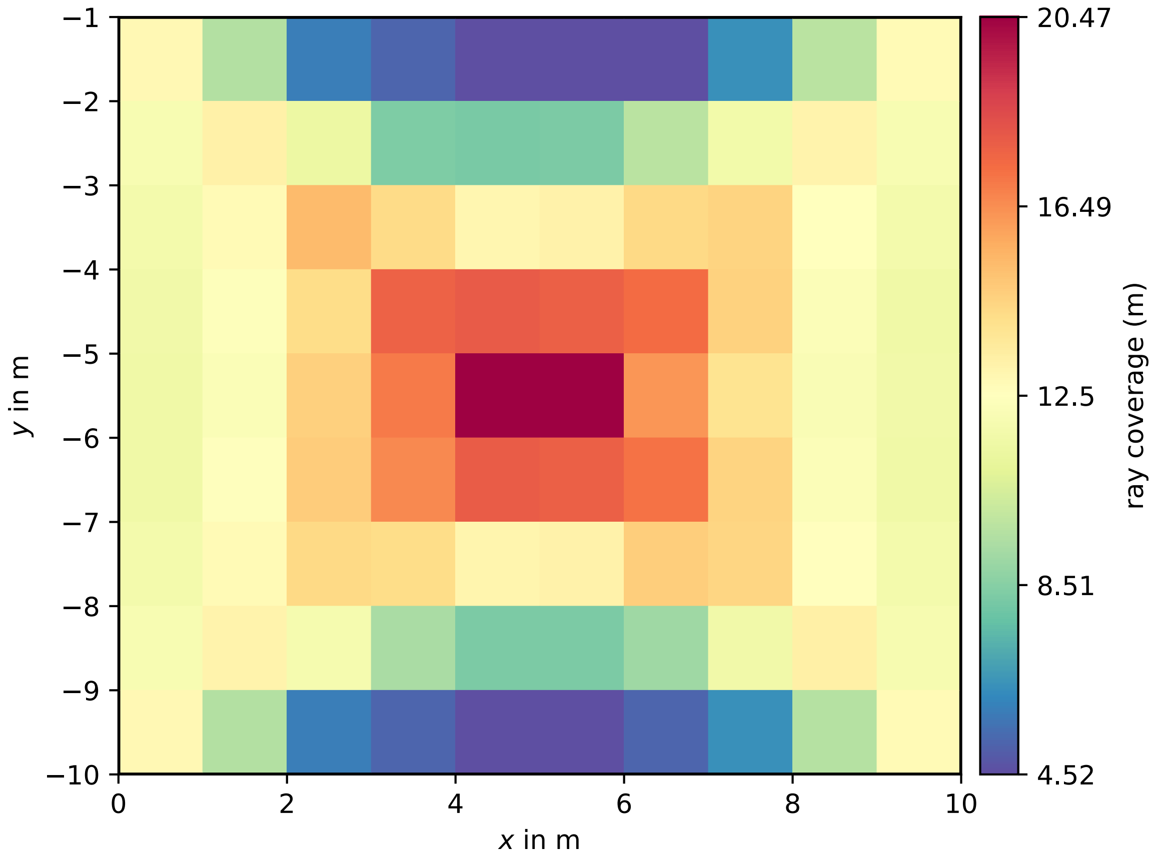
Seismic crosshole tomography (NEW)

- grid with 1m spacing (~~11~~9x10 cells)
- two boreholes: shots left, geophones right, fully connected (10x10 data)
- straight ray paths (x-ray, small contrasts)

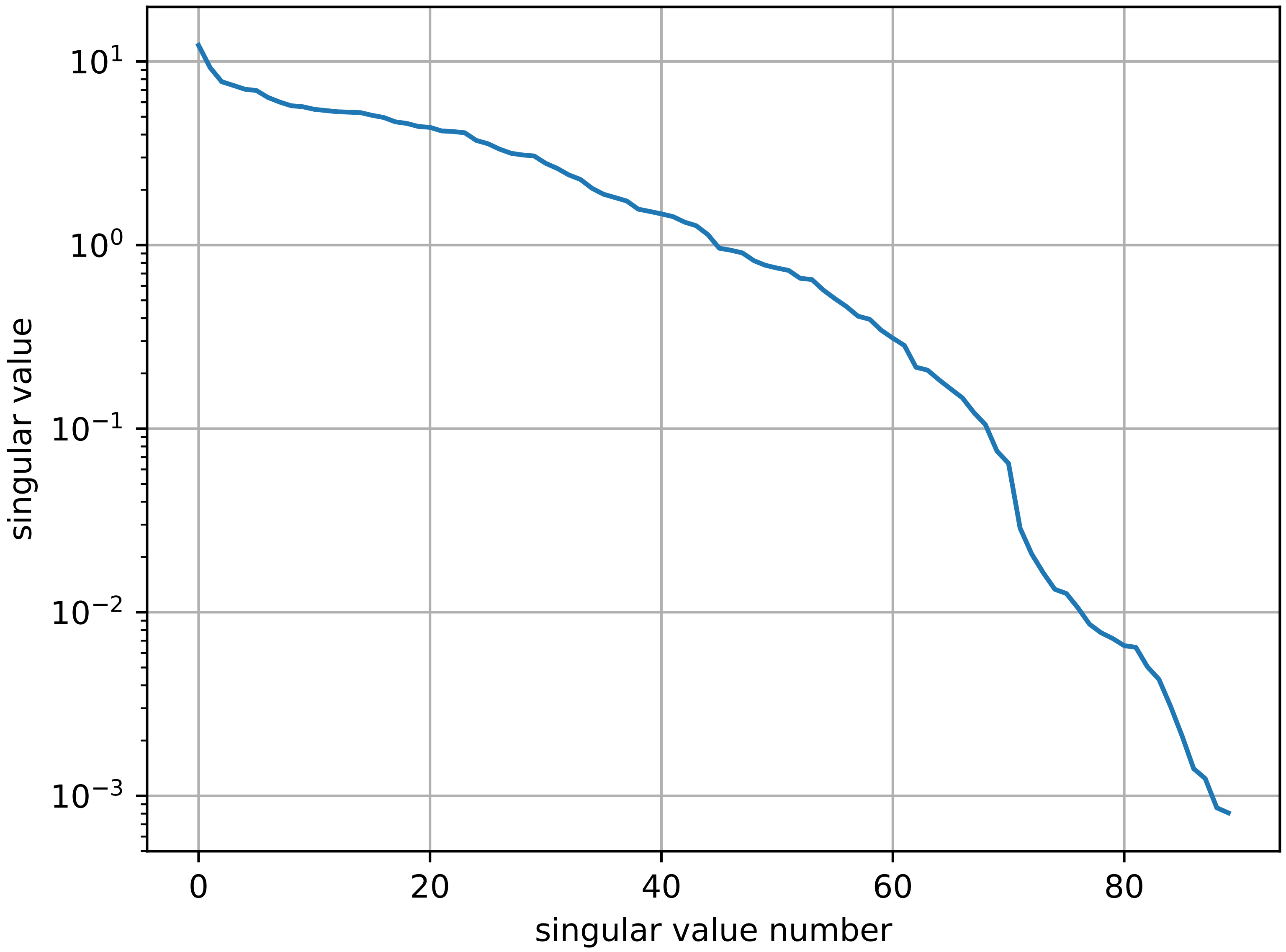
Way matrix and coverage



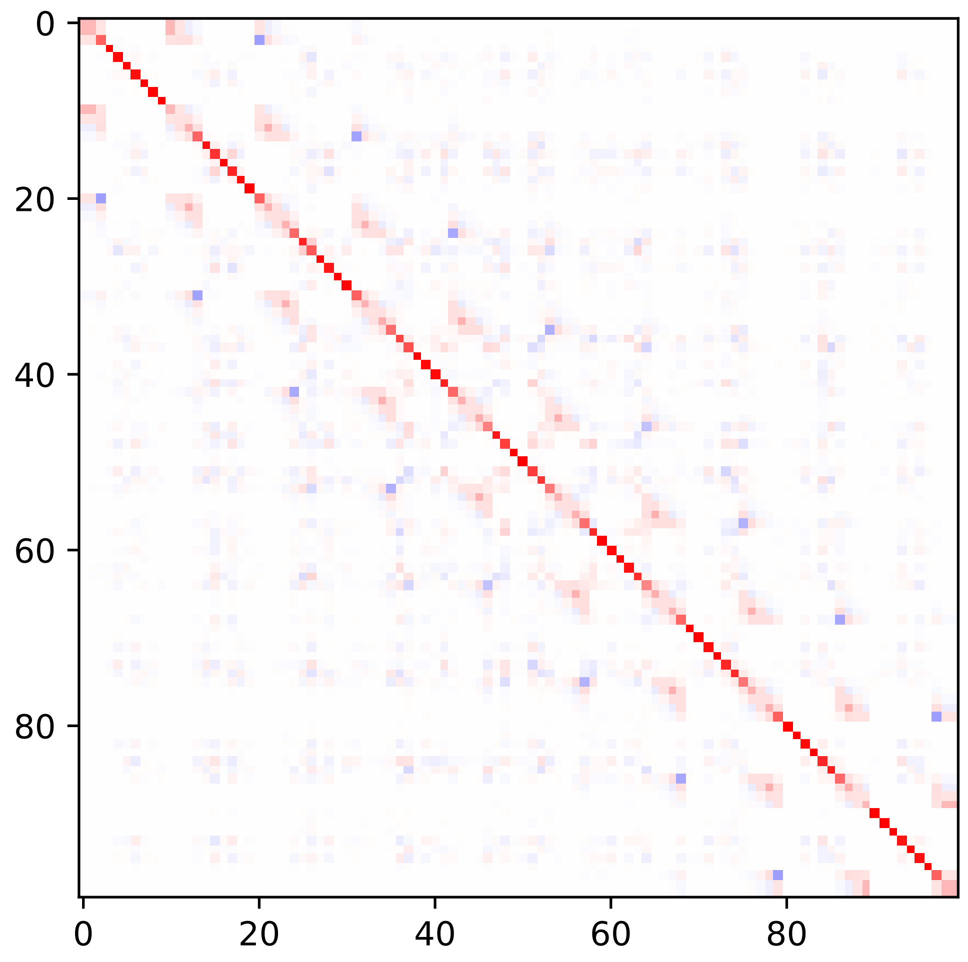
Coverage: data-sum of Jacobian



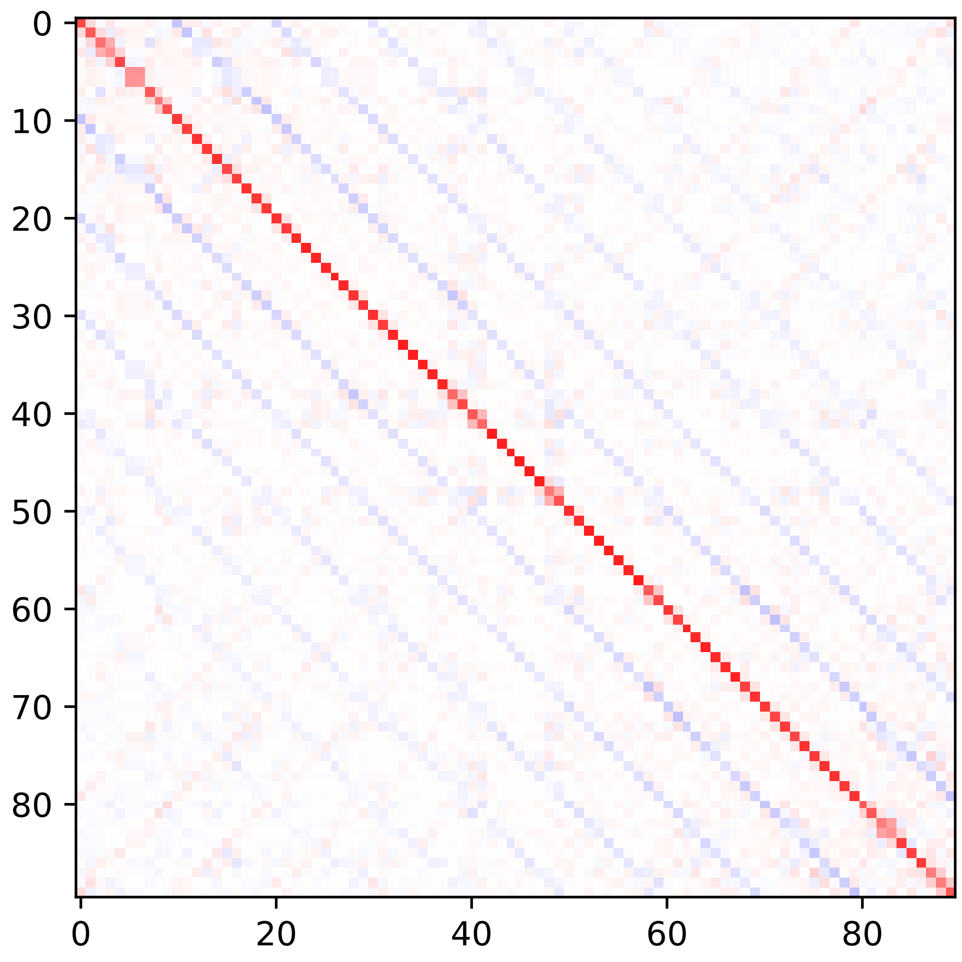
Singular value spectrum



Resolution matrices

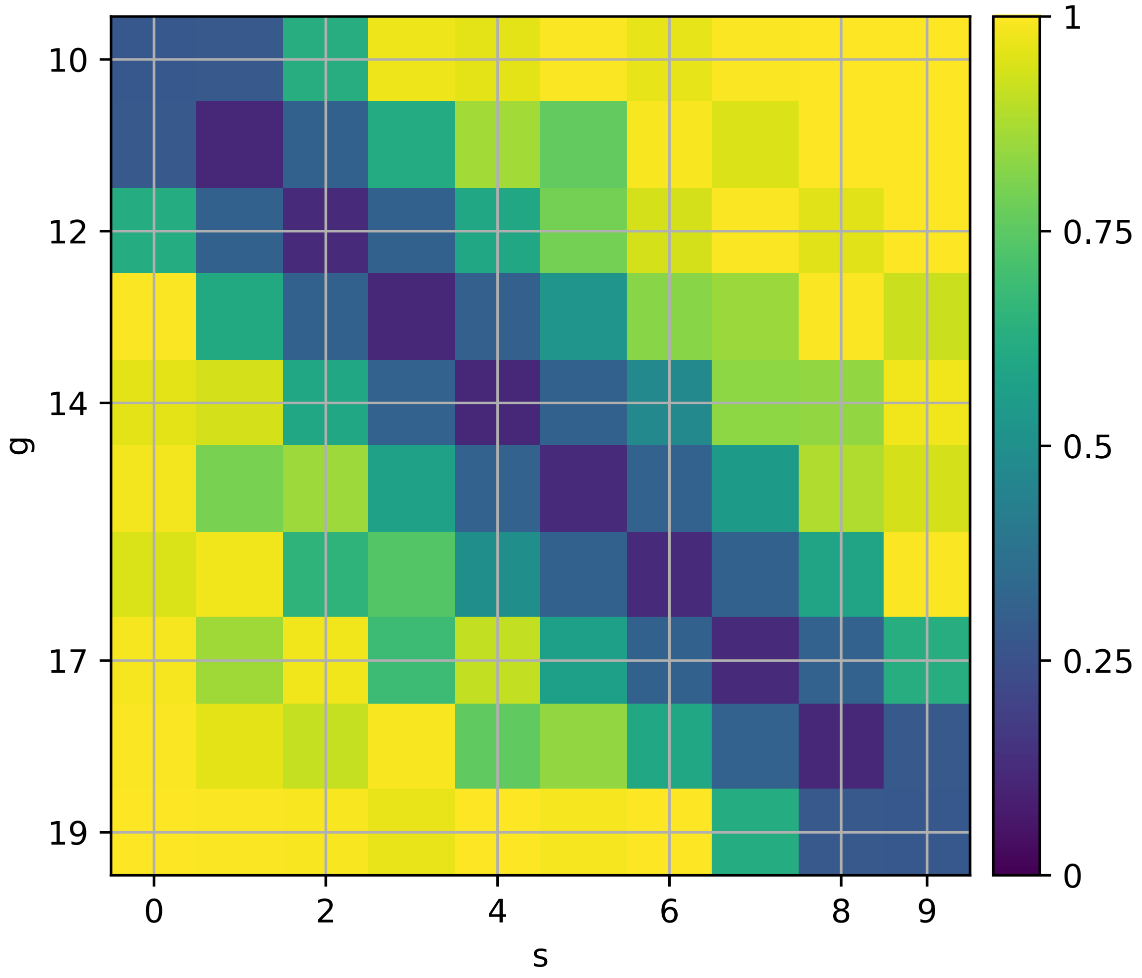


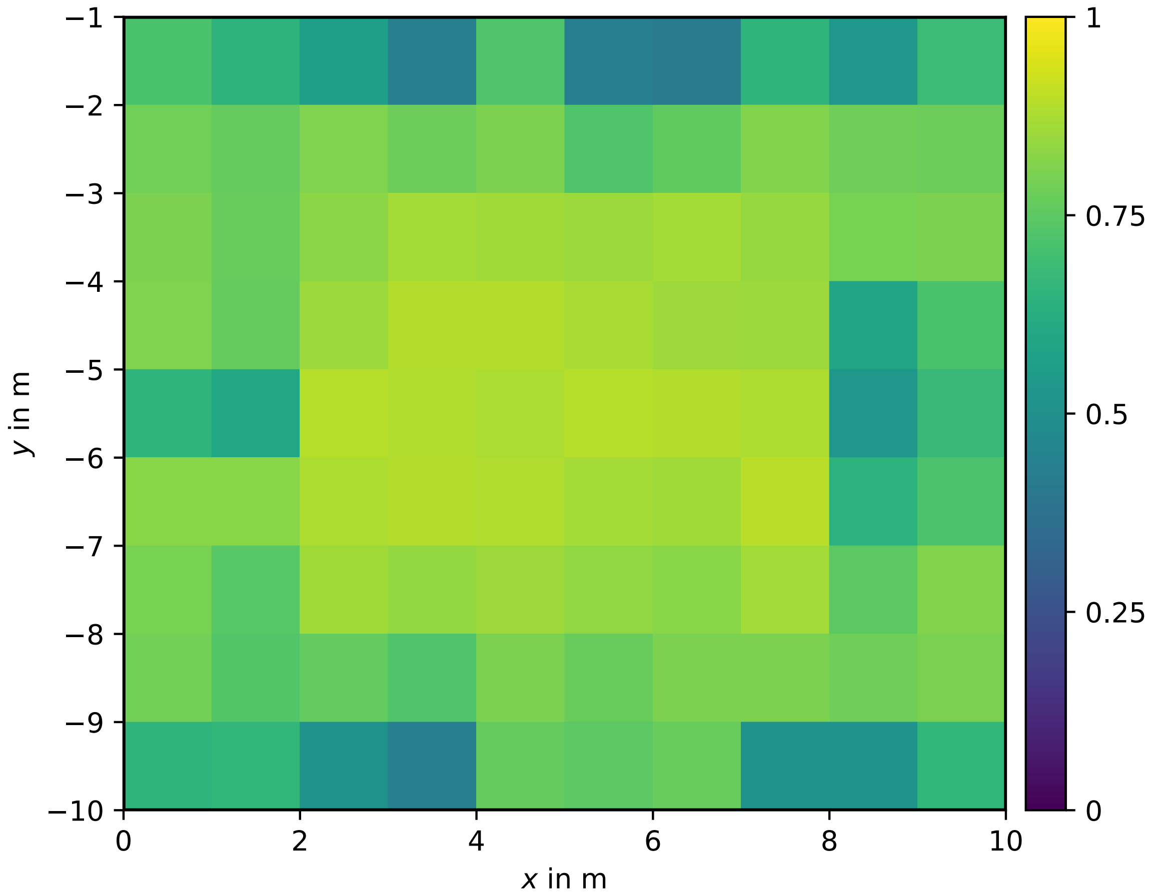
Data importance matrix



Model resolution matrix

Data importance and model resolution



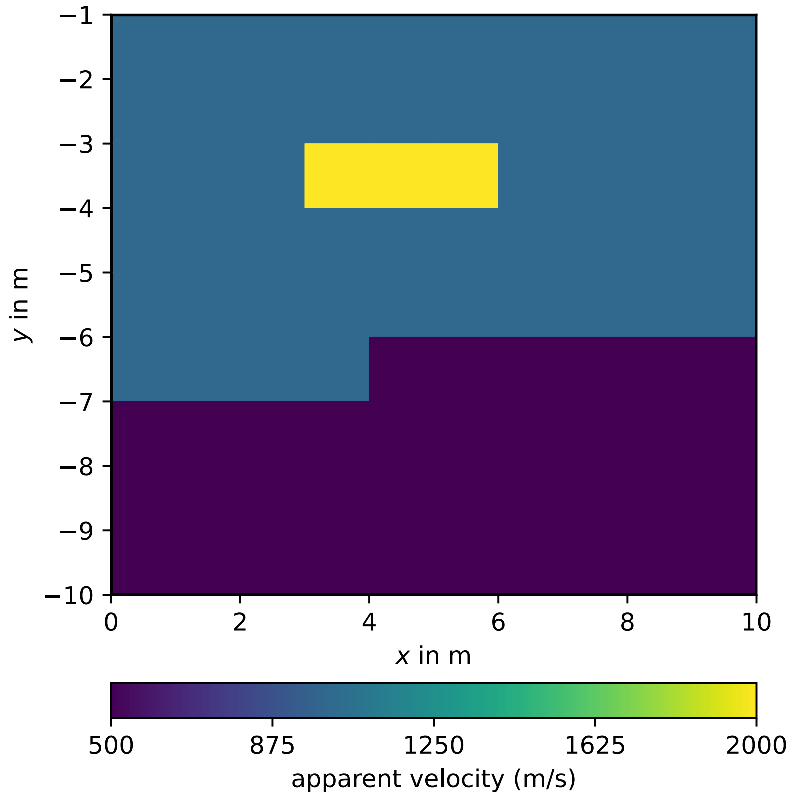


Model resolution

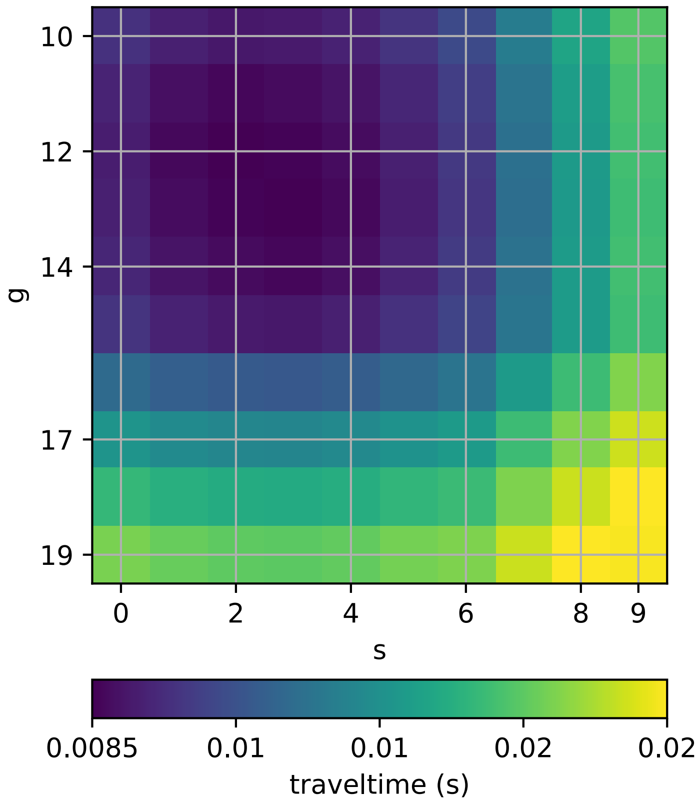
Data importance

A synthetic model

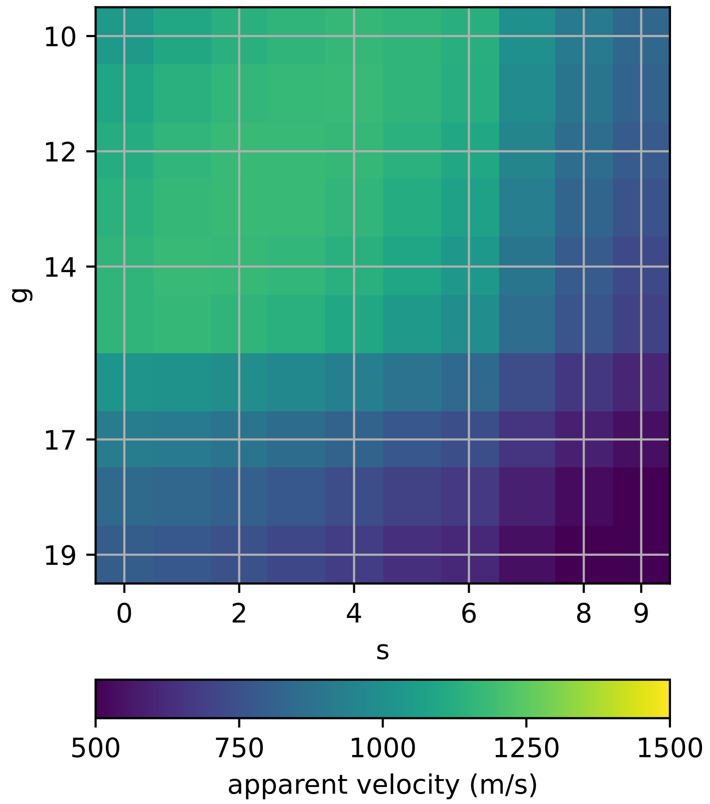
Synthetic model



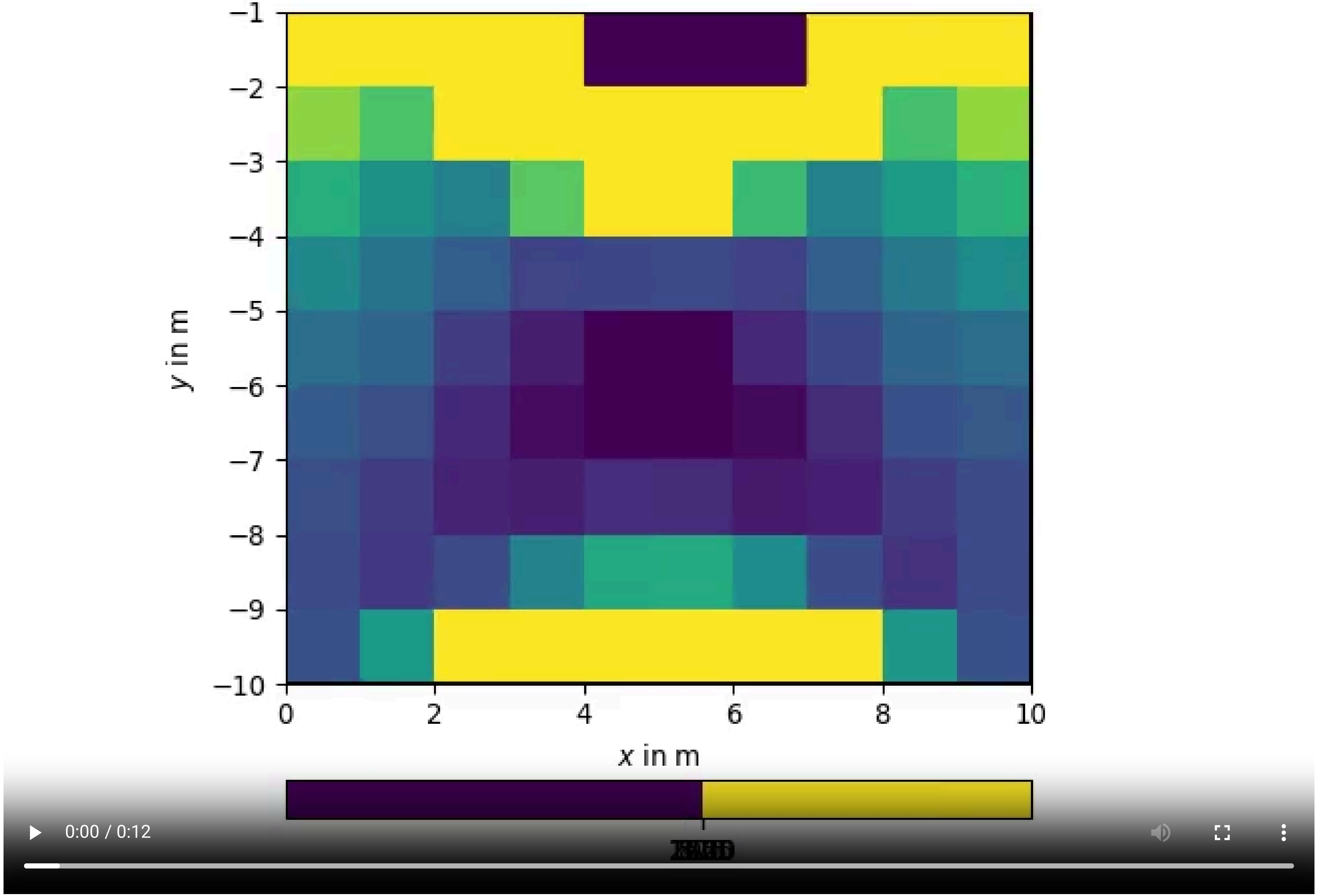
Traveltimes



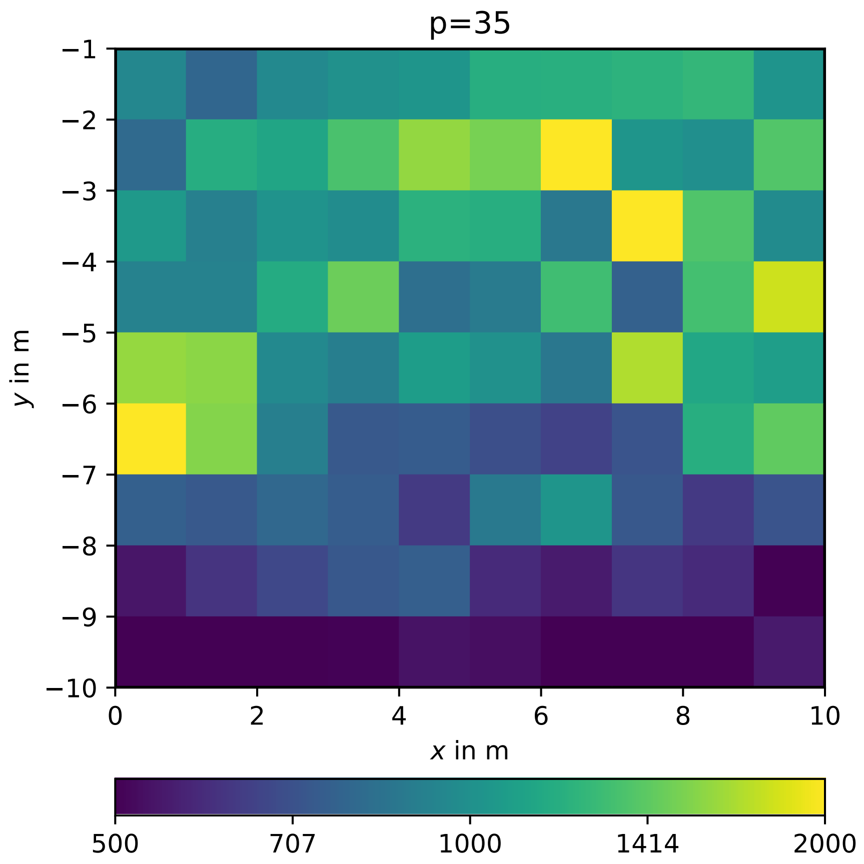
Apparent velocity



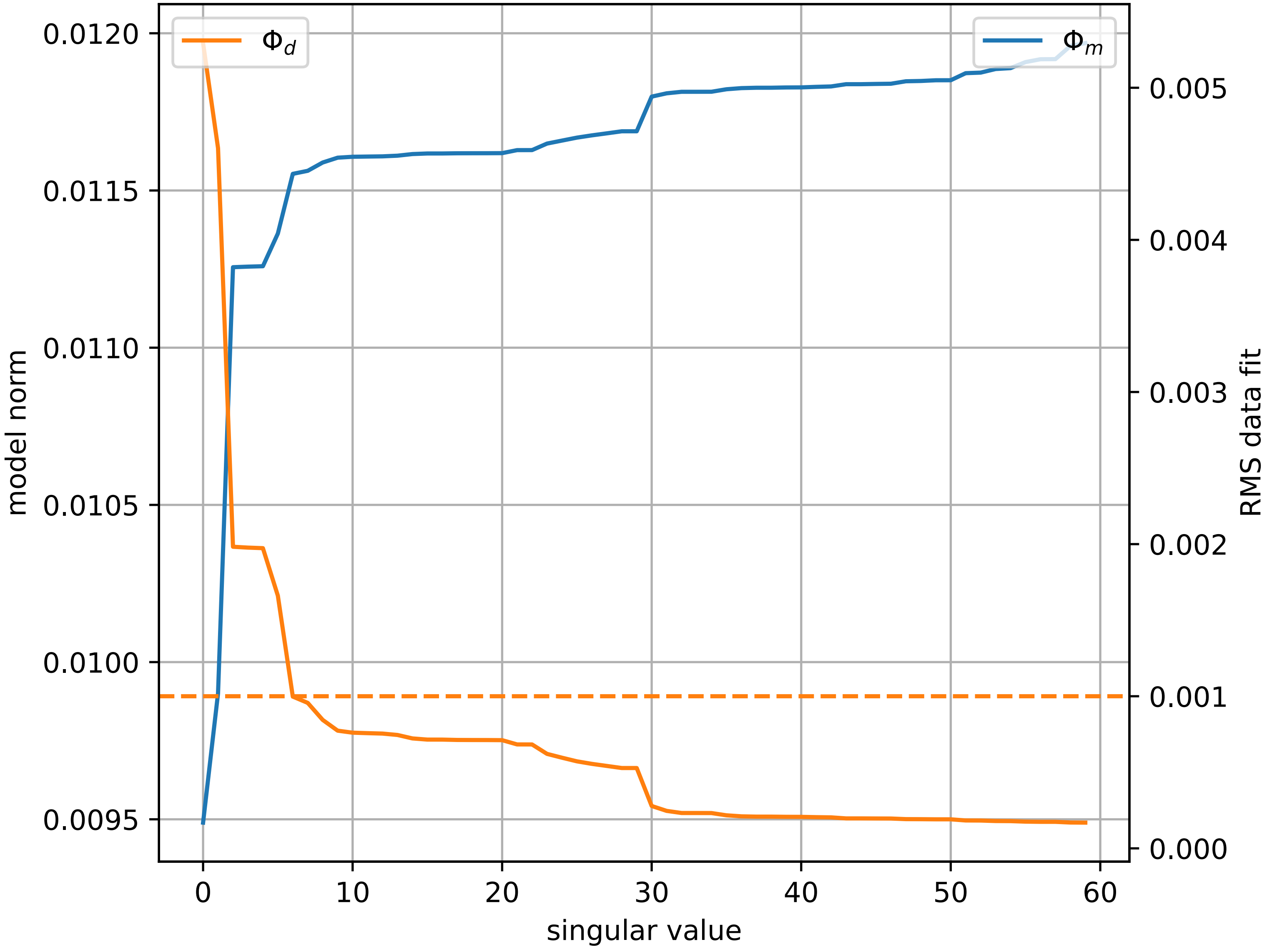
Inversion with TSVD


y in m
x in m
−1
−2
−3
−4
−5
−6
−7
−8
−9
−10
0
2
4
6
8
10
0:00 / 0:12

Inversion with TSVD



Choosing p: Data and model norm



Regularization

- making under-determined and ill-posed problems unique (regular)
- make the model less sensible to small changes in the data
- adding our assumptions or knowledge (valid ranges, prior data, geostatistical behaviour)

> **Occams razor**
>
> Of all possible models, choose the simplest! How to define simple?

Minimum norm

All model parameters are expected to be (similarly) small

$$\tilde{\mathbf{G}}\mathbf{m} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{m} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{pmatrix}$$

or close to some prior knowledge (d_3, d_4, d_5)

Smoothness constraints

Gradient (roughness) between neighboring model parameters

$$\tilde{\mathbf{G}}\mathbf{m} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ 0 \\ 0 \end{pmatrix}$$

Regularization scheme

Splitting into original matrix & data and constraints

$$\tilde{\mathbf{G}} = \begin{bmatrix} \mathbf{G} \\ \mathbf{C} \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{d}} = \begin{bmatrix} \mathbf{d} \\ \mathbf{c} \end{bmatrix}$$

Damping (minimum norm)

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

Smoothness constraints

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

Solving the regularized problem

Splitting into original matrix & data and constraints

$$\tilde{\mathbf{G}} = \begin{bmatrix} \mathbf{G} \\ \mathbf{C} \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{d}} = \begin{bmatrix} \mathbf{d} \\ \mathbf{c} \end{bmatrix}$$

now over-determined $\Rightarrow$ (constrained) least-squares solution

$$\tilde{\mathbf{G}}^T\tilde{\mathbf{G}} = \mathbf{G}^T\mathbf{G} + \mathbf{C}^T\mathbf{C} \quad \& \quad \tilde{\mathbf{G}}^T\tilde{\mathbf{d}} = \mathbf{G}^T\mathbf{d} + \mathbf{C}^T\mathbf{c}$$

$$\Rightarrow \mathbf{m} = (\mathbf{G}^T\mathbf{G} + \mathbf{C}^T\mathbf{C})^{-1}(\mathbf{G}^T\mathbf{d} + \mathbf{C}^T\mathbf{c})$$

Weighting data vs. constraints

$$\Phi = \|\tilde{\mathbf{G}}\mathbf{m} - \tilde{\mathbf{d}}\|_2^2 = \|\mathbf{G}\mathbf{m} - \mathbf{d}\|^2 + \|\mathbf{C}\mathbf{m} - \mathbf{c}\|^2 = \rightarrow \min$$

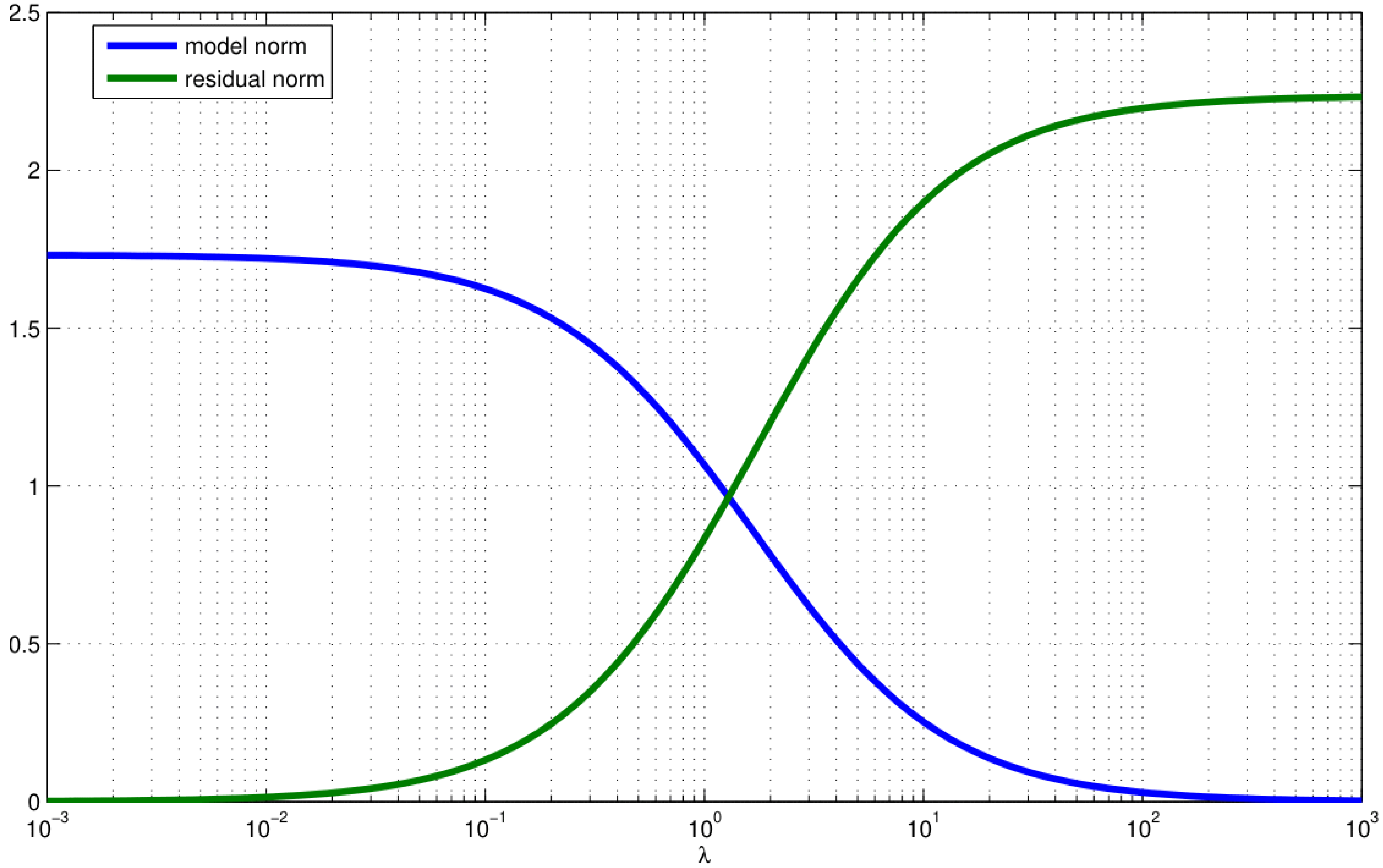
$\mathbf{d}$ and $\mathbf{c}$ have different magnitudes & units, data maybe too weak or too strong $\Rightarrow$ weighting by regularization parameter λ:

$$\Phi = \|\mathbf{G}\mathbf{m} - \mathbf{d}\|^2 + \lambda\|\mathbf{C}\mathbf{m} - \mathbf{c}\|^2 = \Phi_d + \lambda\Phi_m \rightarrow \min$$

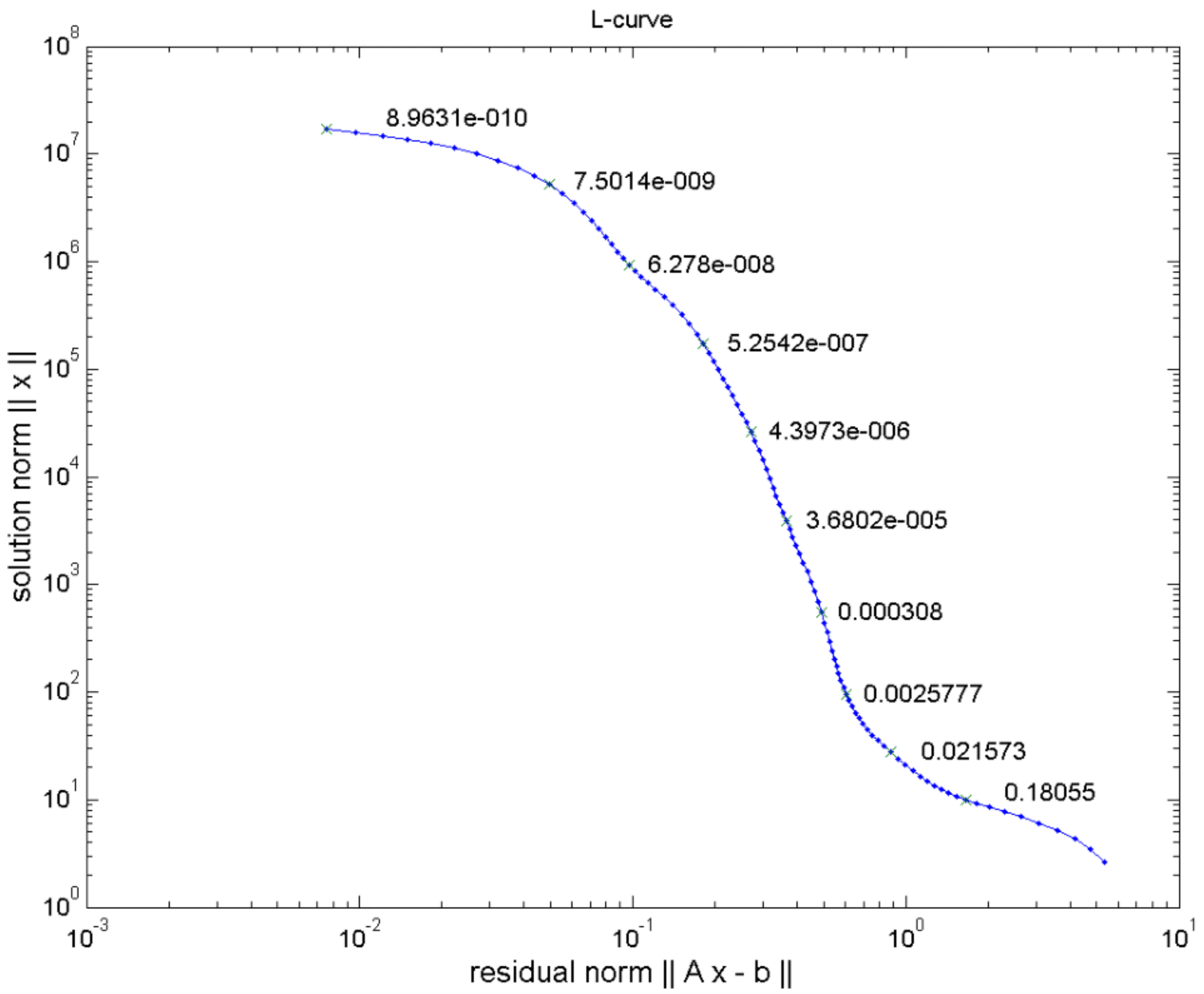
λ..regularization strength, Φ_d/Φ_m..data/model objective function

$$\Rightarrow \mathbf{m} = (\mathbf{G}^T\mathbf{G} + \lambda\mathbf{C}^T\mathbf{C})^{-1}(\mathbf{G}^T\mathbf{d} + \lambda\mathbf{C}^T\mathbf{c})$$

Model and data norms



The L-curve



Data vs. model norm for wide range of λ

- low data residual achieved by high norm (oscillating model)
- low model norm cannot fit the data (large misfit)
- optimum somewhere “at the corner” (not always a corner)

Choice of regularization strength

Always have a look at your data fit and model plausibility.

- use different values and look at models (and misfit)
- try to determine the corner of the L-curve (maximum curvature)
- start large λ, decrease & stop when data misfit show no systematics

Discrepancy principle

Choose the highest λ value that is able to fit the data (χ^2=1)!

Damped normal equations and SVD

$$\mathbf{m} = (\mathbf{G}^T\mathbf{G} + \lambda\mathbf{I})^{-1}\mathbf{G}^T\mathbf{d}$$

$$\mathbf{m} = (\mathbf{V\Sigma U}^T\mathbf{U\Sigma V}^T + \lambda\mathbf{I})^{-1}\mathbf{V\Sigma U}^T\mathbf{d}$$

$$\mathbf{m} = (\mathbf{V}\mathrm{diag}(s_i^2 + \lambda)\mathbf{V}^T + \lambda\mathbf{I})^{-1}\mathbf{V\Sigma U}^T\mathbf{d}$$

$$\mathbf{m} = \sum_i^r \frac{s_i}{s_i^2 + \lambda}\mathbf{u}_i^T\mathbf{d} \cdot v_i^T = \sum_i^r \frac{s_i^2}{s_i^2 + \lambda}\frac{\mathbf{u}_i^T\mathbf{d}}{s_i}v_i^T$$

Small singular values are damped in inversion, large unchanged

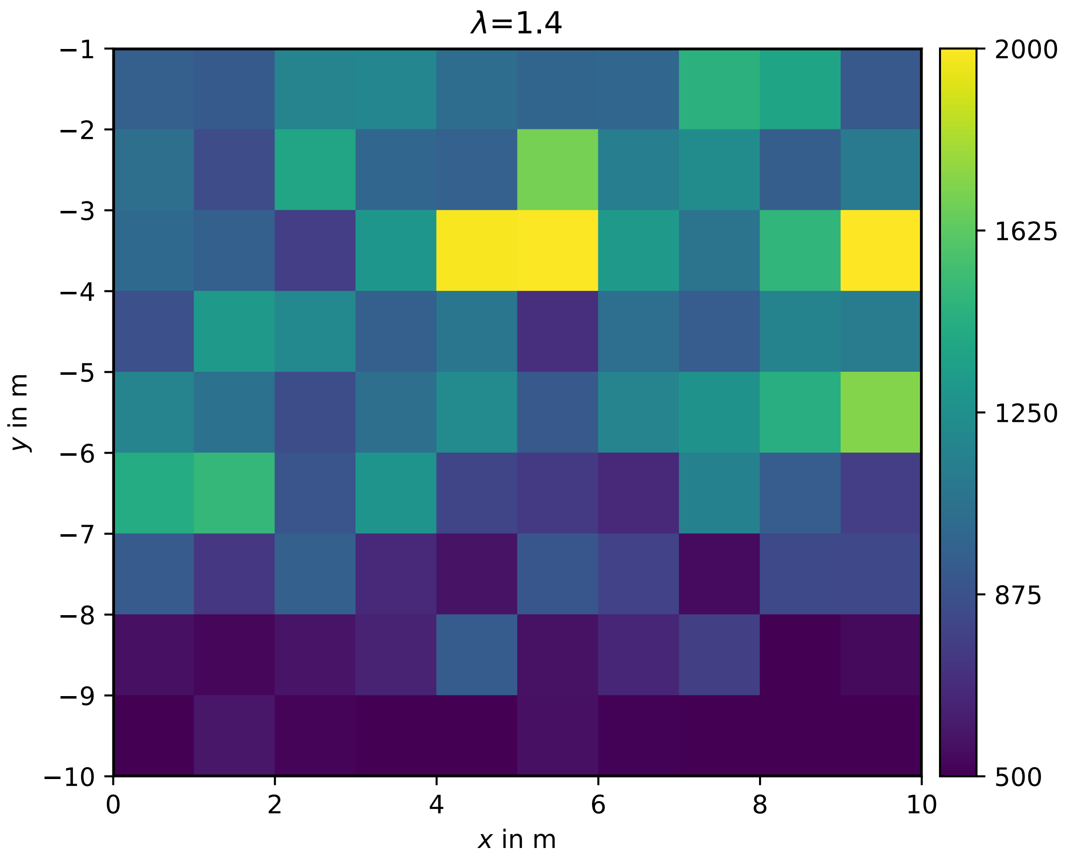
Resolution of regularized inverse problems

For $c = 0$ we have $\mathbf{G}^\dagger = (\mathbf{G}^T\mathbf{G} + \lambda\mathbf{C}^T\mathbf{C})^{-1}\mathbf{G}^T$

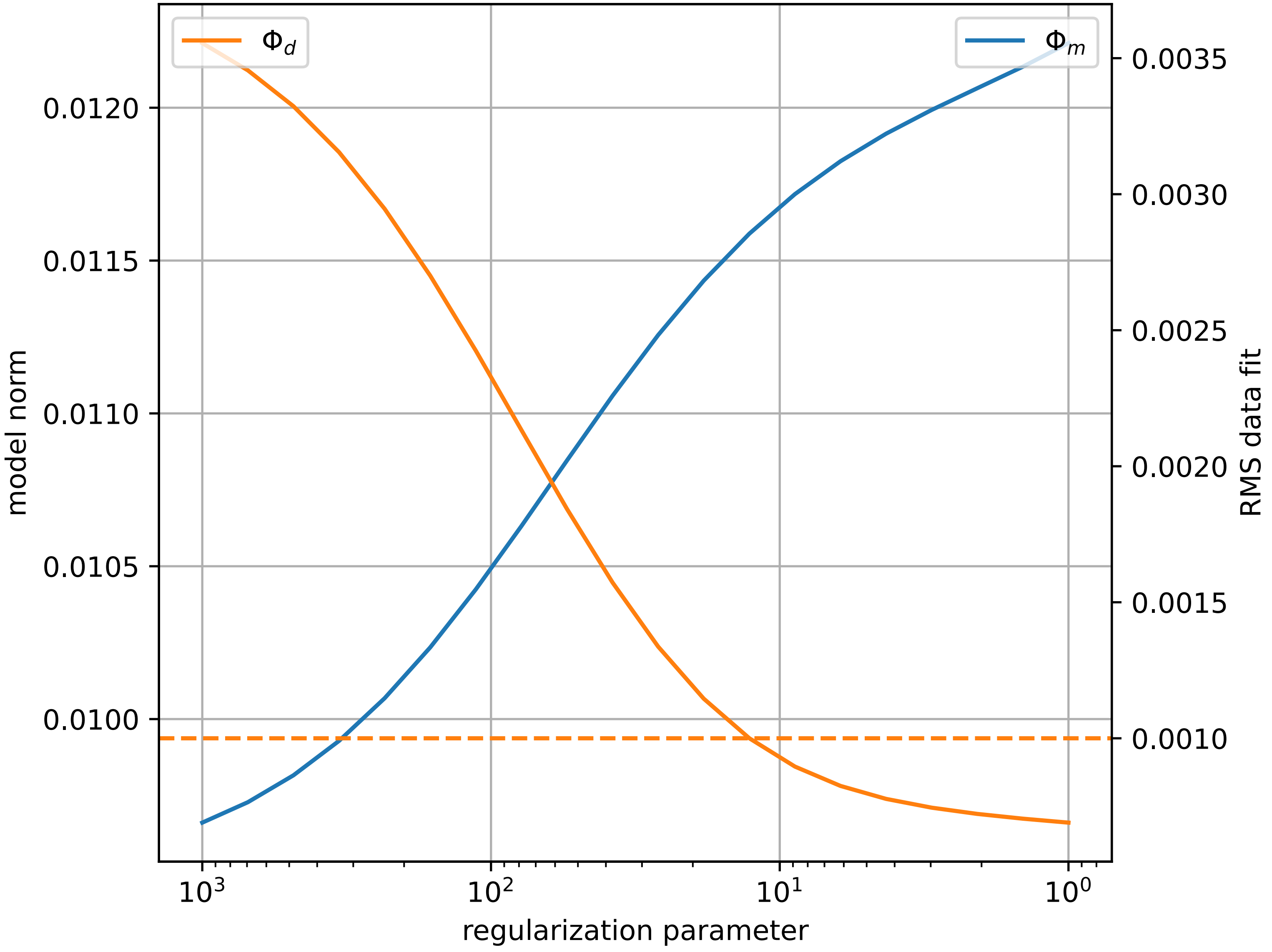
$$\Rightarrow \mathbf{R}^M = \mathbf{G}^\dagger\mathbf{G} = (\mathbf{G}^T\mathbf{G} + \lambda\mathbf{C}^T\mathbf{C})^{-1}\mathbf{G}^T\mathbf{G}$$

approaches $\mathbf{I}$ for $\lambda \rightarrow 0$ and deviates if λ grows

Inversion with damping



Choosing λ: Data and model norm



Wrap up

- SVD provides a general tool, BUT:
  - can amplify noise for ill-conditioned problems (SV spectrum)
  - truncated SVD (limiting p) can suppress this
- explicit regularization to make solution unique
  - different strategies: smoothness, minimum norm
- choice of regularization strength (λ, p) is vital