Inverse Problems in Geophysics
Part 5: TSVD & Regularization

2. MGPY+MGIN

Thomas Günther

thomas.guenther@geophysik.tu-freiberg.de

TU BERGAKADEMIE FREIBERG
TUBAF
Die Ressourcenuniversität. Seit 1765.

Recap problem types/SVD

- over-determined problems: least-squares solution
- under-determined problems: minimum-norm solution
- rank determines problem type
- singular value decomposition (SVD) as fundamental tool
- for all cases: the pseudo-inverse (LS & MN special cases)
- data/model resolution determined by data/model eigenvectors

Inverse problem types - classification scheme

The rank r determines the type of the inverse problem

Even-determined

$M = N = r$, $\mathbf{R}^M$=$\mathbf{I}$, $\mathbf{R}^D$=$\mathbf{I}$

Over-determined

$N > r = M$, $\mathbf{R}^M$=$\mathbf{I}$, $\mathbf{R}^D \neq \mathbf{I}$

Under-determined

$N = r < M$, $\mathbf{R}^M \neq \mathbf{I}$, $\mathbf{R}^D = \mathbf{I}$

Mixed-determined

$r < N$, $r < M$, $\mathbf{R}^M \neq \mathbf{I}$, $\mathbf{R}^D \neq \mathbf{I}$

The Singular Value Decomposition

$\mathbf{G}$ consists of model & data eigenvectors, weighted by singular values

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

- eigenvalues in $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_i) \in \mathbb{R}^{r \times r}$
- orthonormal data eigenvectors $\mathbf{U}_r \in \mathbb{R}^{N \times r}$ with $\mathbf{U}_r^T \mathbf{U}_r = \mathbf{I}$
- orthonormal model eigenvectors $\mathbf{V}_r \in \mathbb{R}^{M \times r}$ with $\mathbf{V}_r^T \mathbf{V}_r = \mathbf{I}$
- generalized inverse $\mathbf{G}^\dagger = \mathbf{V}_r \boldsymbol{\Sigma}_r^{-1} \mathbf{U}_r^T = \sum \frac{\mathbf{u}_i^T \mathbf{d}}{\sigma_i} \mathbf{v}_i$

The problem of small eigenvalues

Generalized inverse $\mathbf{G}^\dagger = \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^T$ to invert noisy data $\mathbf{Gm} + \mathbf{n}$

$$\mathbf{G}^\dagger(\mathbf{Gm} + \mathbf{n}) = \mathbf{G}^\dagger \mathbf{Gm} + \mathbf{G}^\dagger \mathbf{n} = \mathbf{R}^M + \sum_{i=1}^{r} \frac{\mathbf{u}_i^T \mathbf{n}}{\sigma_i} \mathbf{v}_i$$

Problem

Small eigenvalues amplify noise in the data!

Moderate Example

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1.1 & 0 & 0 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0.5 & 1 \end{pmatrix}$$

$$\Rightarrow \sigma_i = [2.05, 1.5, 0.5, 0.05]$$

```
m = pinv(G) * (d + n)
```

$$\mathbf{m} = [10, 11, 12, 13], \epsilon = 0.2$$

$$\Rightarrow \mathbf{d} = [21.0, 22.1, 18.5, 19.0]$$

$$\mathbf{m}^1 = [4.45, 16.24, 11.62, 13.37]$$
$$\mathbf{m}^2 = [9.92, 10.92, 12.36, 12.59]$$
$$\mathbf{m}^3 = [15.11, 6.2, 11.66, 13.41]$$
$$\mathbf{m}^4 = [4.11, 16.76, 11.74, 13.05]$$
$$\mathbf{m}^5 = [4.4, 16.48, 11.55, 13.19]$$

Even a small noise level let the solution explode.

Solution: drop small singular values

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1.1 & 0 & 0 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0.5 & 1 \end{pmatrix}$$

$$\mathbf{m} = [10, 11, 12, 13], \epsilon = 0.2$$

$$\Rightarrow \mathbf{d} = [21.0, 22.1, 18.5, 19.0]$$

$$\Rightarrow \sigma_i = [2.05, 1.5, 0.5, 0.05]$$

$\Rightarrow$ `m=pinv(G, rtol=0.1)*(d+n)`

$$\begin{aligned} \mathbf{m}^1 &= [10.36, 10.89, 12.07, 13.2] \\ \mathbf{m}^2 &= [10.29, 10.82, 11.67, 13.08] \\ \mathbf{m}^3 &= [10.32, 10.84, 11.97, 13.16] \\ \mathbf{m}^4 &= [10.25, 10.78, 11.81, 13.04] \\ \mathbf{m}^3 &= [10.29, 10.82, 11.71, 13.13] \end{aligned}$$

The solution is much less noise-depending!

Truncated singular value (TSVD) method

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

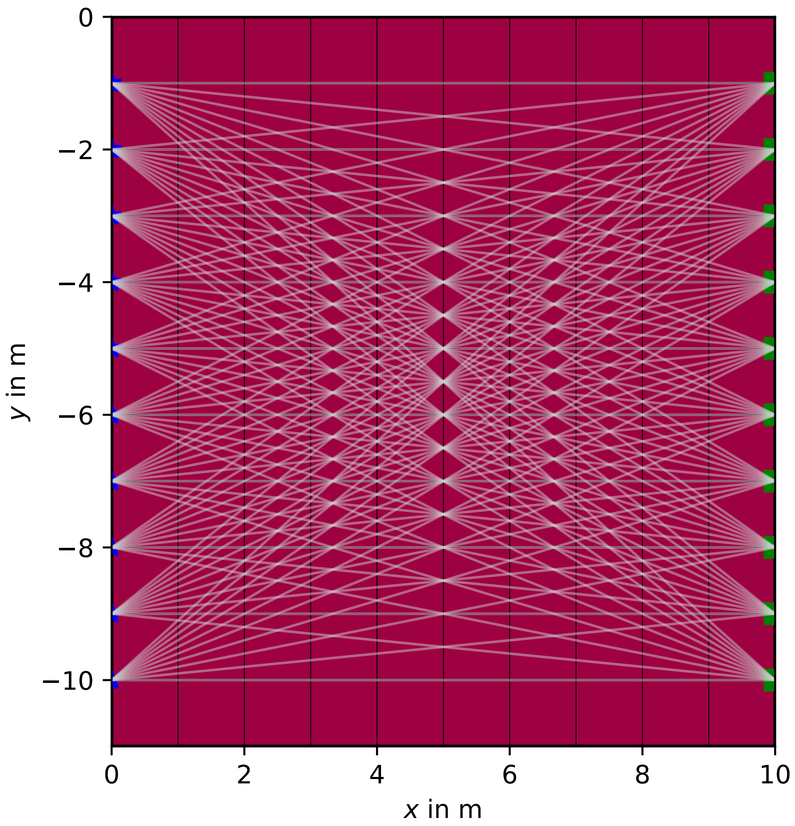
$$\mathbf{G}_p^{\dagger} = \mathbf{V}_p \mathbf{\Sigma}_p^{-1} \mathbf{U}_p^T$$

- 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.

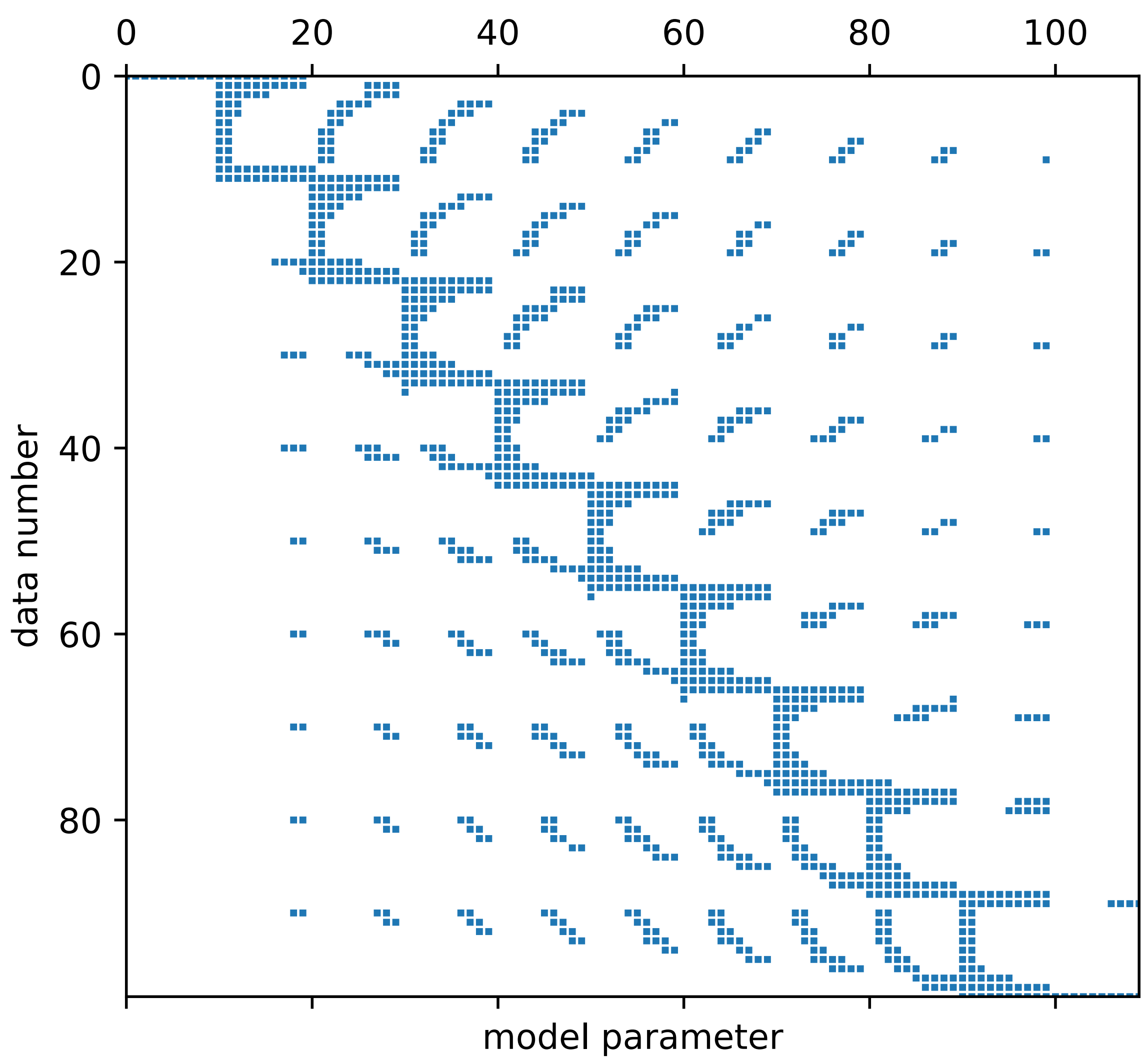
Geophysical example



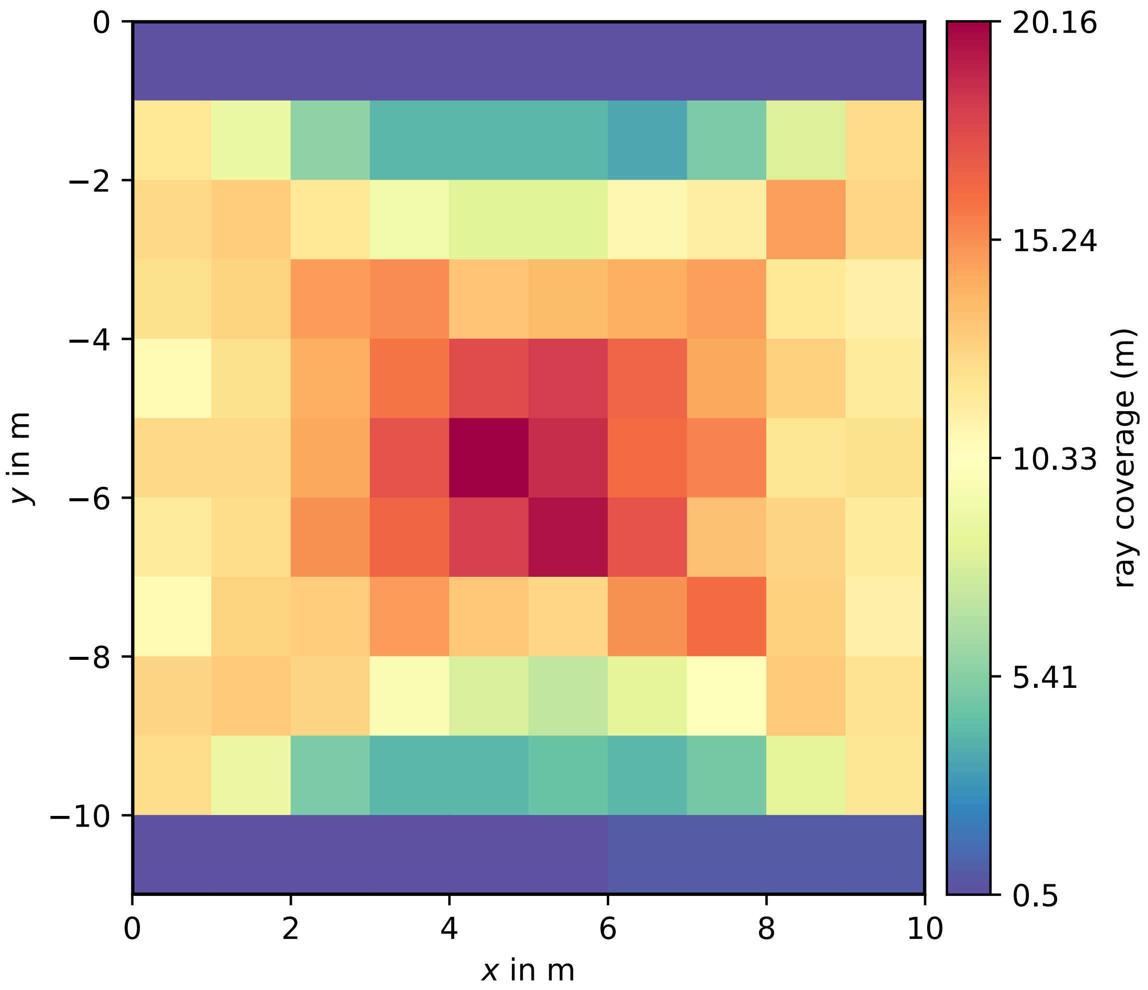
Seismic crosshole tomography

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

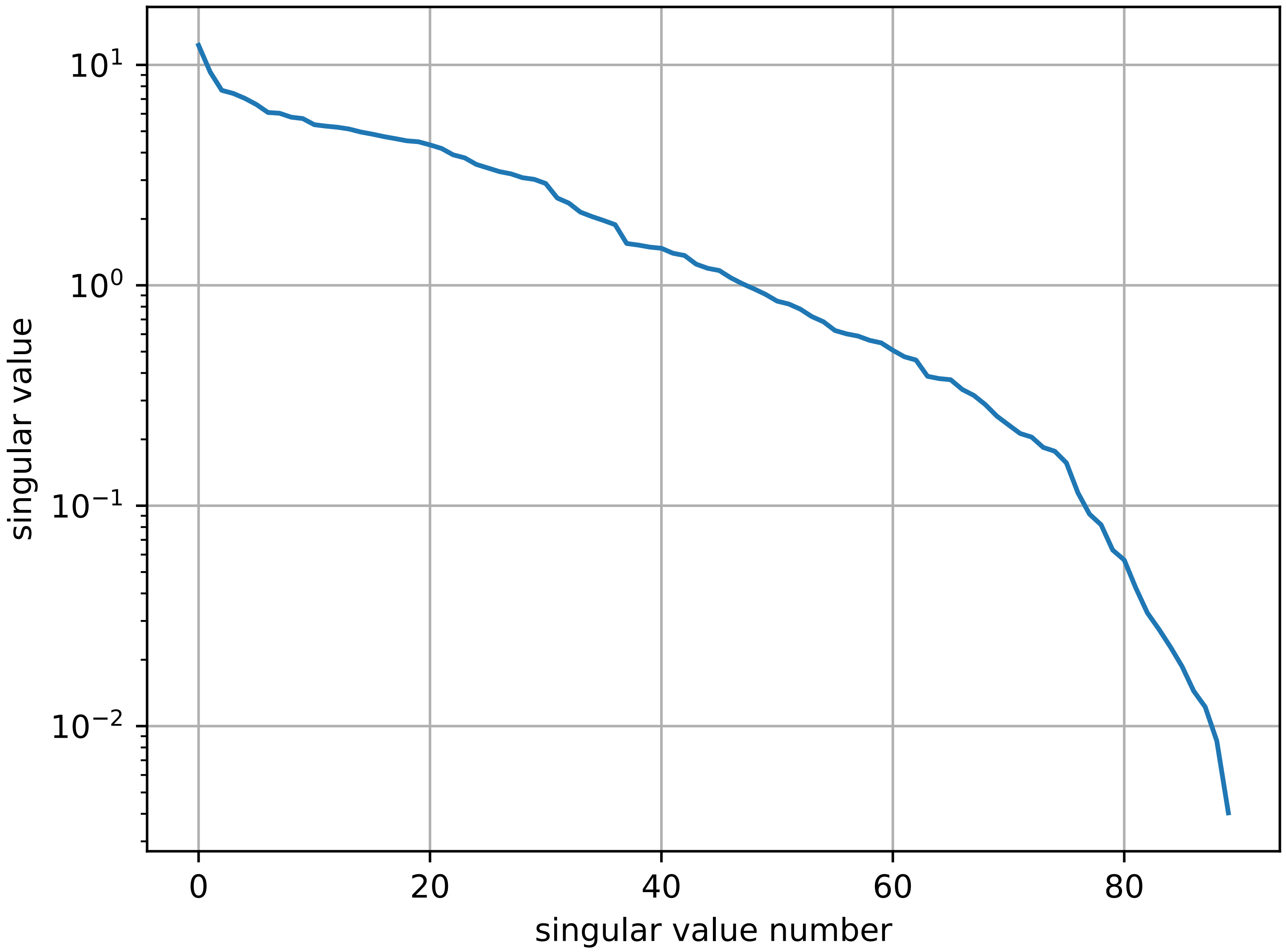
Way matrix and coverage



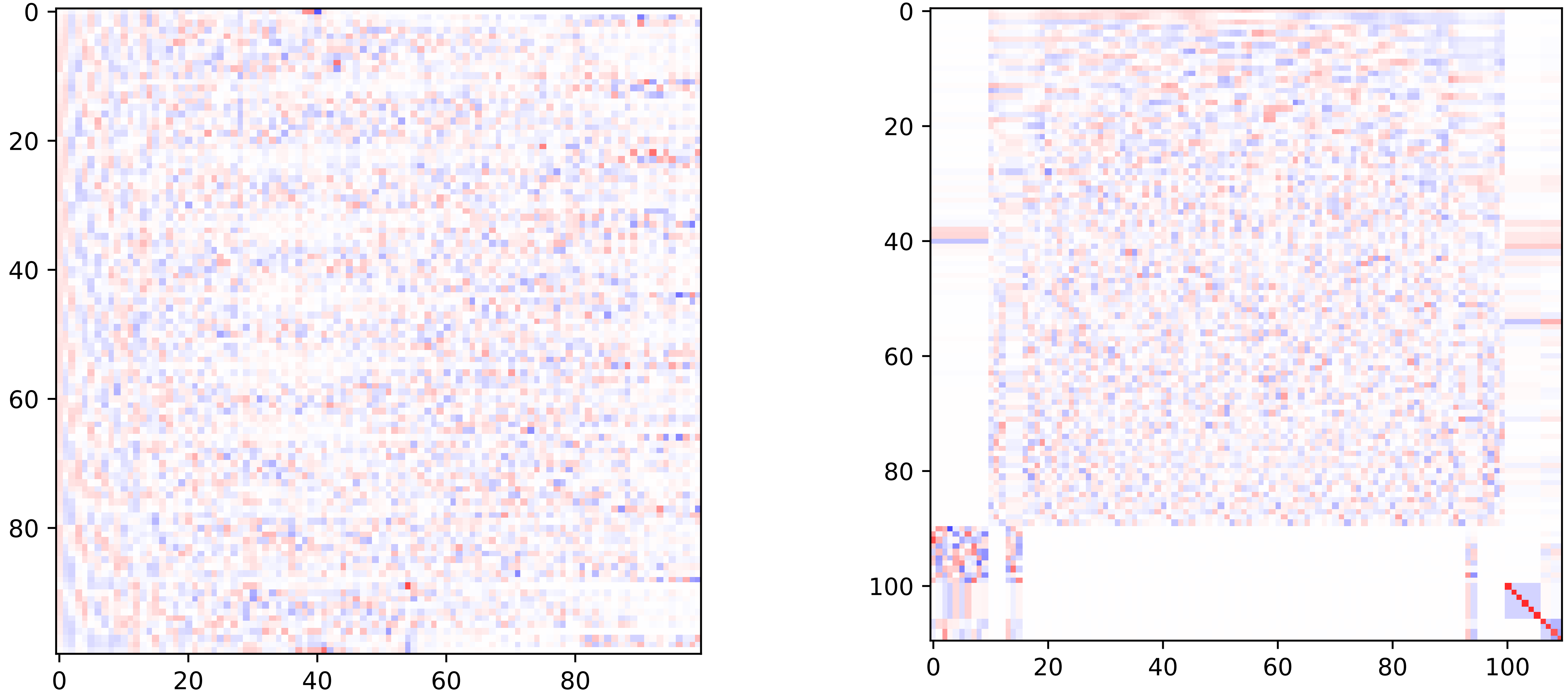
Coverage: data-sum of Jacobian



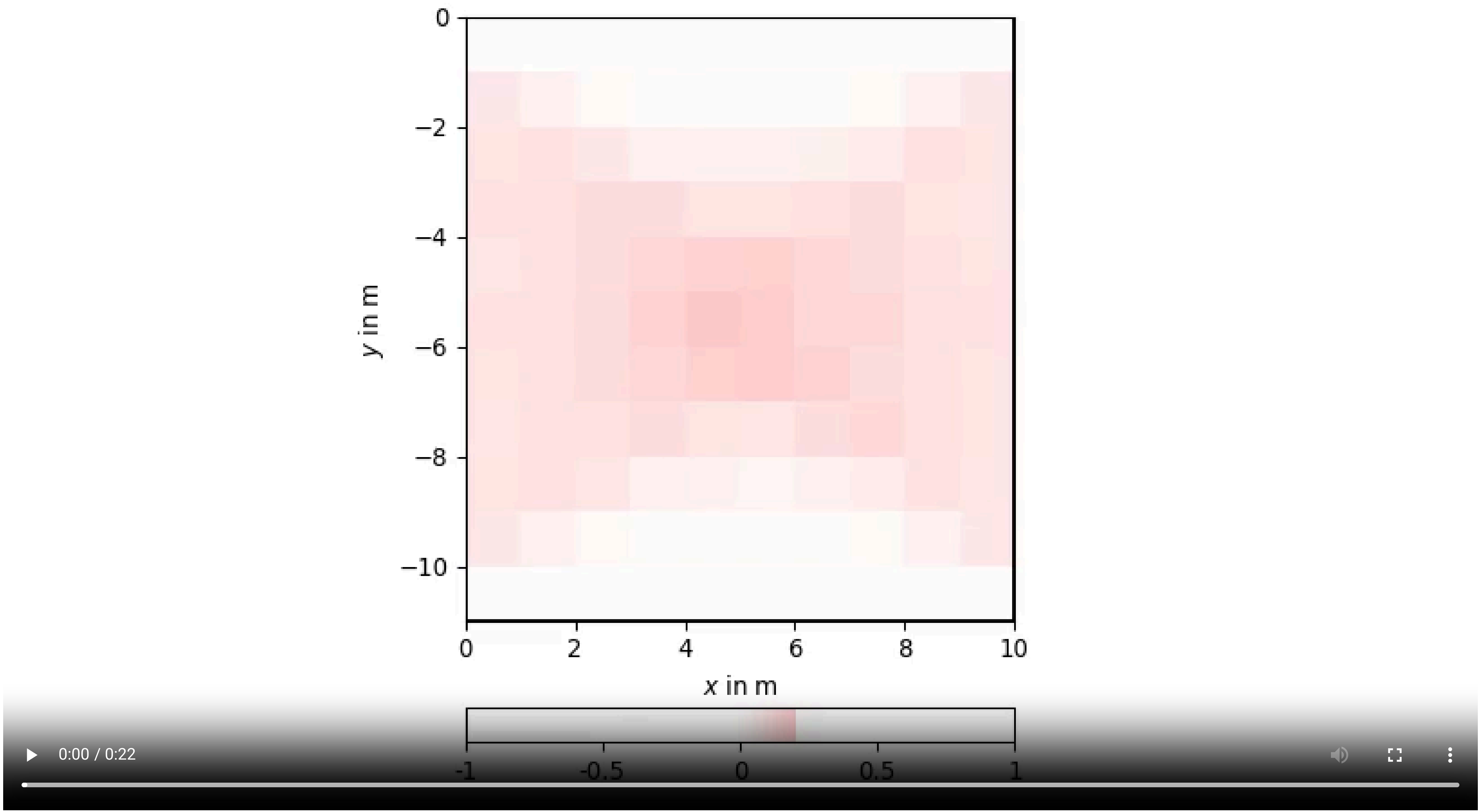
Singular value spectrum



Data and model eigenvectors

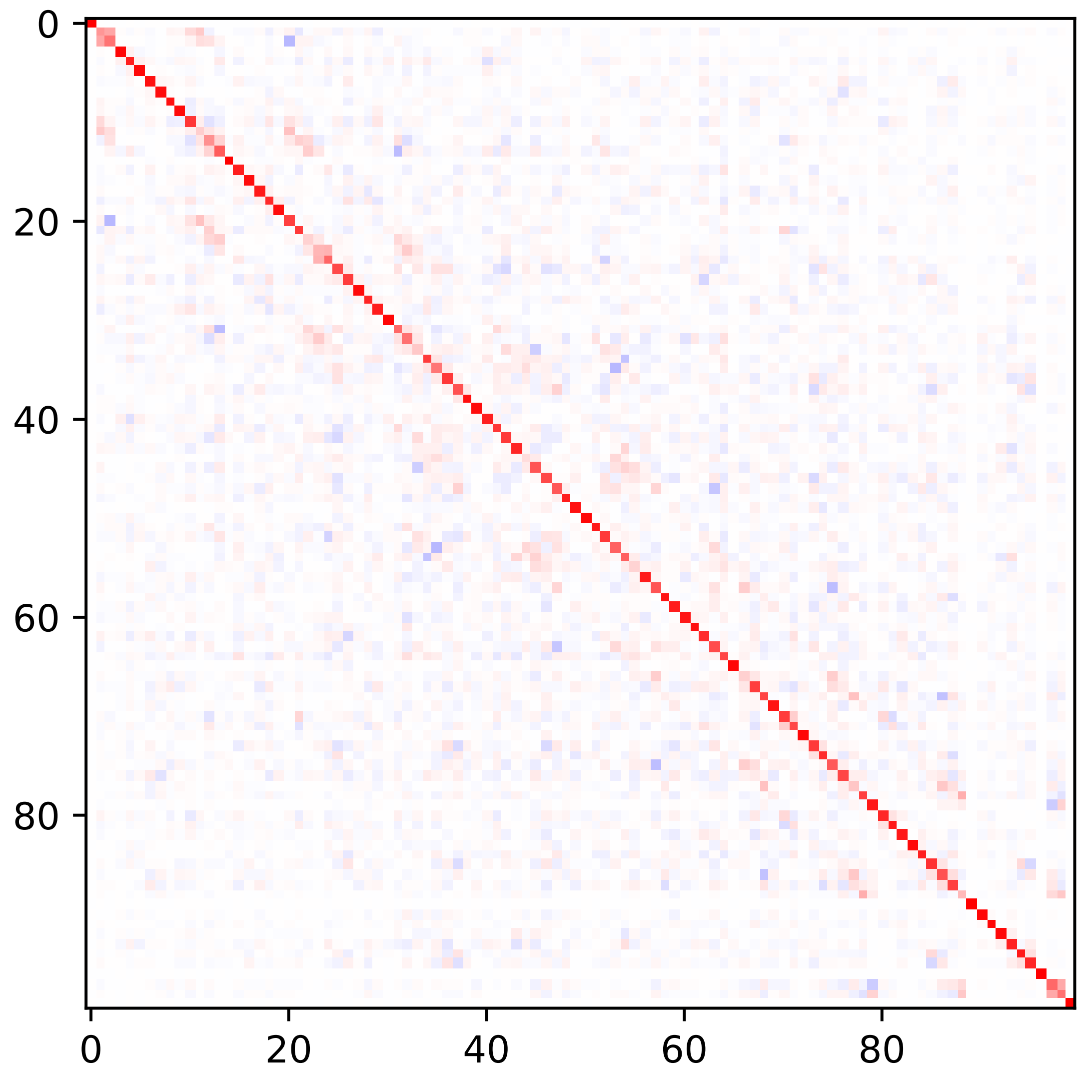


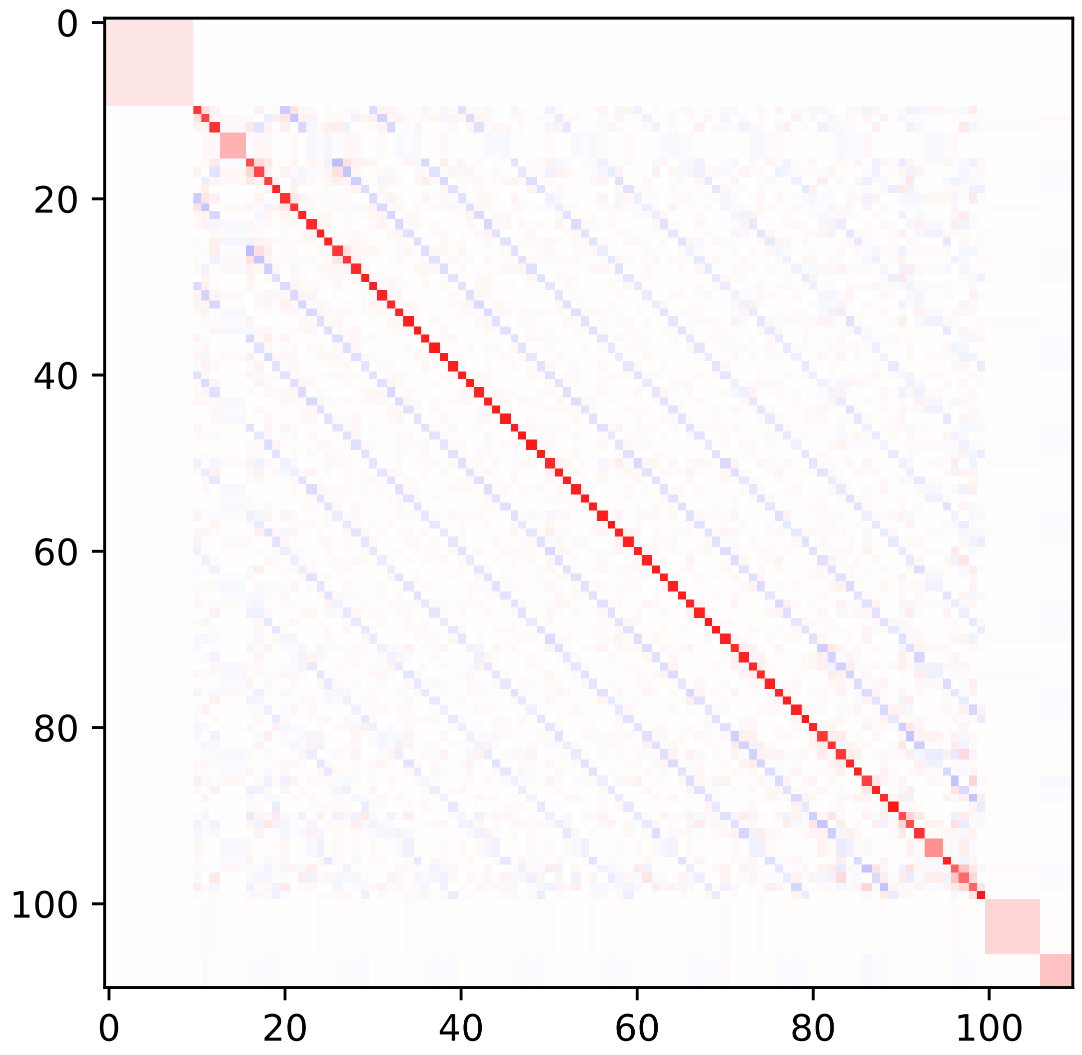
Model eigenvectors



y in m
0
−2
−4
−6
−8
−10
x in m
0
2
4
6
8
10
-1
-0.5
0
0.5
1
0:00 / 0:22

Data and model resolution matrix





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?

Regularization basics: adding equations

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

- we only measure the mean value of m_1 and m_2
- difference between m_1 and m_2 should be small

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

Regularization basics: adding equations

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

- we only measure the mean value of m_1 and m_2
- size of subvector $[m_1, m_2]$ should be small

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

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

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

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

$$\tilde{\mathbf{G}}^T \tilde{\mathbf{G}} = \mathbf{G}^T \mathbf{G} + \mathbf{C}^T \mathbf{C}$$

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

Weighting data vs. constraints

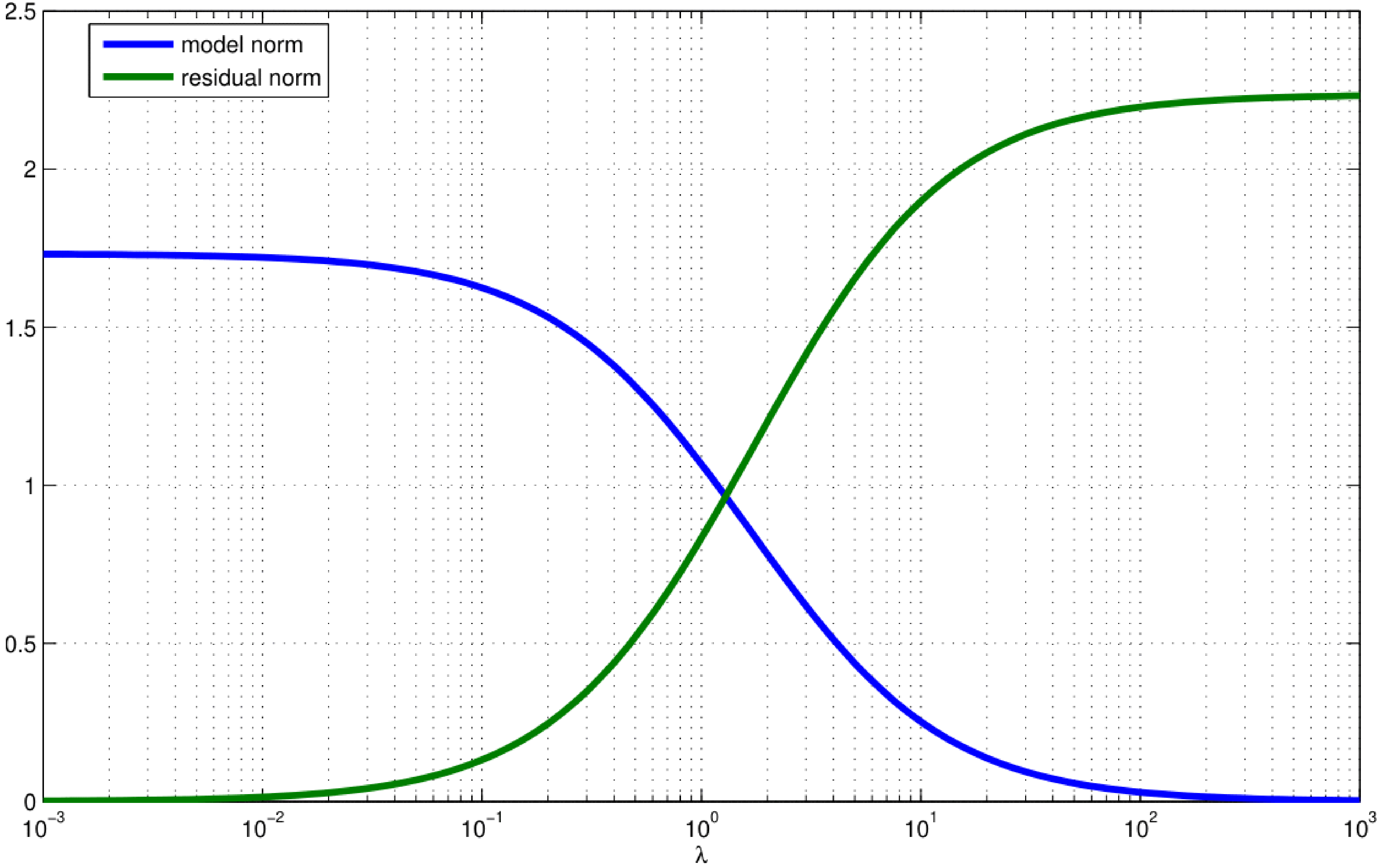
$\mathbf{d}$ and $\mathbf{c}$ may have completely different magnitudes and physical units, data maybe too weak or too strong $\Rightarrow$ weighting of constraints by regularization parameter λ:

$$\Phi = \|\mathbf{Gm} - \mathbf{d}\|^2 + \lambda\|\mathbf{Cm} - \mathbf{c}\|^2 = \Phi_d + \lambda\Phi_m \rightarrow \min$$

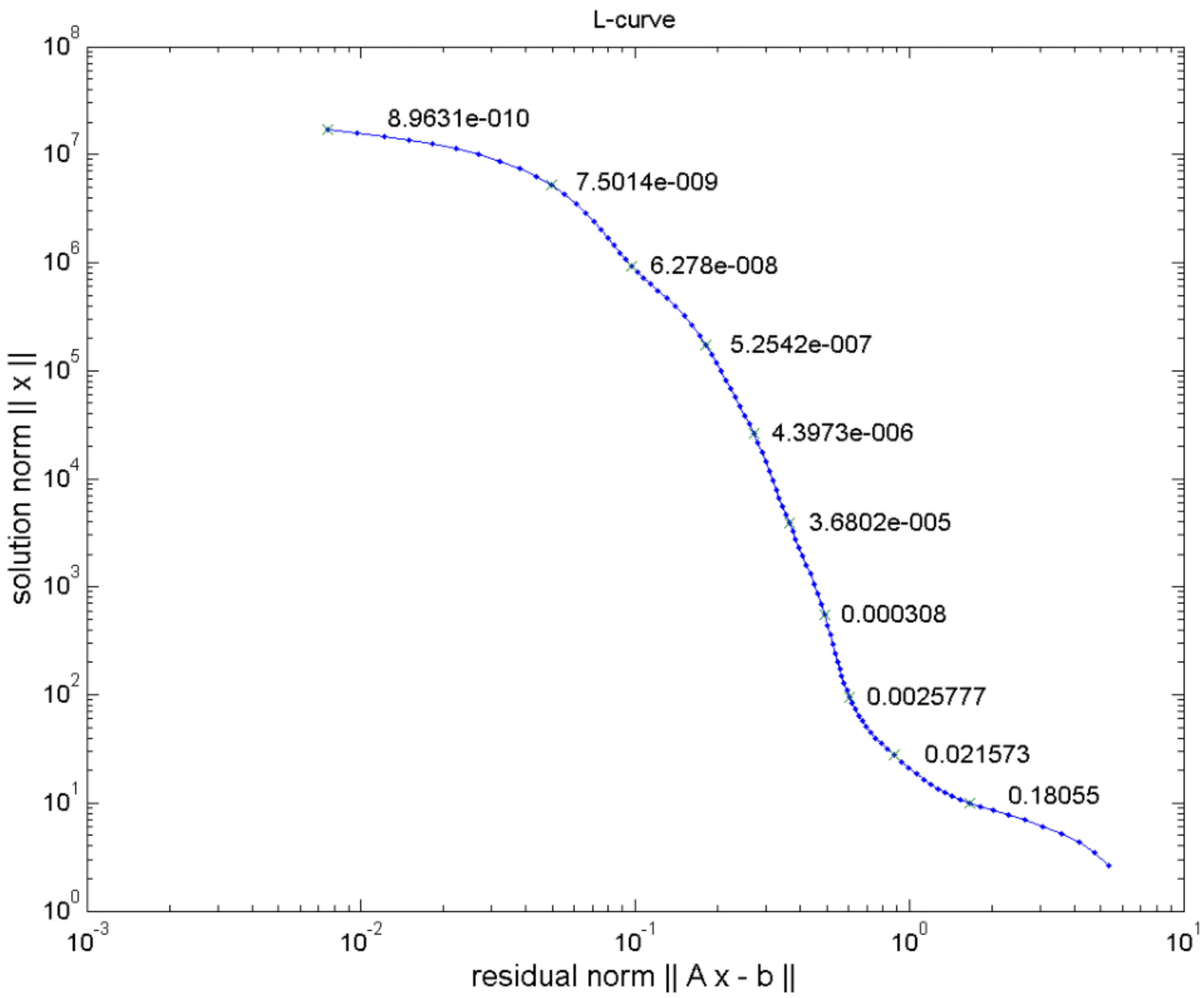
λ..regularization strength, Φ_d/Φ_m..data/model objective function

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

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 $\lambda \to 0$ and deviates if λ grows

Wrap up

- SVD provides a general tool, BUT:
- ill-conditioned inverse problems (SV spectrum) tend to amplify noise
- truncated SVD is a method so suppress this
- regularization can (also) be done to make solution unique
- different strategies
- choice of regularization strength is vital