ECE295, Data Assimilation and Inverse Problems, Spring 2015

1 April, Intro; Linear discrete Inverse problems (Aster Ch 1 and 2) Slides

8 April, SVD (Aster ch 2 and 3) Slides

15 April, Regularization (ch 4)

22 April, Sparse methods (ch 7.2-7.3)

29 April, more on Sparse

6 May, Bayesian methods and Monte Carlo methods (ch 11)

13 May, Introduction to sequential Bayesian methods, Kalman Filter (KF)

20 May, Ensemple Kalman Filer (EnKF)

27 May, EnKF, Particle Filter,

3 June, Markov Chain Monte Carlo

Homework:

Just email the code to me (I dont need anything else).

Call the files LastName_ExXX.

Homework is **due 8am on Wednesday.**

8 April: Hw 1: Download the matlab codes for the book (cd_5.3) from this website

15 April: SVD analysis:
SVD homework. You can also try replacing the matrix in the Shaw problem with the beamforming sensing matrix. The sensing matrix is available here .

22 April

Late April: Beamforming

May: Ice-flow from GPS

Recap: Linear discrete inverse problems

- The Least squares solution minimizes the prediction error.

$$\boxed{\phi(\mathbf{m}_{LS}) = (\mathbf{d} - G\mathbf{m}_{LS})^T C_d^{-1} (\mathbf{d} - G\mathbf{m}_{LS})}$$

$$\boxed{\mathbf{m}_{LS} = (G^T C_d^{-1} G)^{-1} G^T C_d^{-1} \mathbf{d} = G^{-g}\mathbf{d}}$$

- Goodness of fit criteria tells us whether the least squares model adequately fits the data, given the level of noise.

$\phi(\mathbf{m}_{LS}) \rightarrow \chi^2_{N-M}$ Chi-square with *N-M* degrees of freedom

- The covariance matrix describes how noise propagates from the data to the estimated model

$$\boxed{C_M = (G^T C_d^{-1} G)^{-1}}$$

$\Delta^2 \rightarrow \chi^2_M$

$$\boxed{(\mathbf{m} - \mathbf{m}_{LS})^T C_M^{-1} (\mathbf{m} - \mathbf{m}_{LS}) < \Delta^2}$$

Chi-square with *M* degrees of freedom

Gives confidence intervals

- The resolution matrix describes how the estimated model relates to the true model

Recap: Goodness of fit and model covariance

- Once a best fit solution has been obtained we test goodness of fit with a chi-square test (assuming Gaussian statistics)

$$\left[\chi^2_{obs} = \sum_{i=1}^{N} \frac{(d_i - \sum_{j=1}^{M} G_{i,j} m_j)^2}{\sigma_i^2}\right]$$

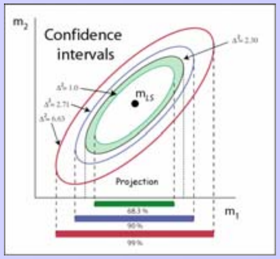
- If the model passes the goodness of fit test we may proceed to evaluating model covariance (if not then your data errors are probably too small)

- Evaluate model covariance matrix

$$C_M = (G^T C_d^{-1} G)^{-1}$$

- Plot model or projections of it onto chosen subsets of parameters

$$\Delta'^2 = (m - m_{LS})'^T C'^{-1}_M (m - m_{LS})'$$



- Calculate confidence intervals using projected equation

$$\sigma_{M,i} = \Delta \times \sigma_{i,i}$$

Where Δ^2 follows a χ^2_1 distribution

Recap: Linear discrete inverse problems

- The Least squares solution minimizes the prediction error.

$$\boxed{\phi(\boldsymbol{m}_{LS}) = (\boldsymbol{d} - G\boldsymbol{m}_{LS})^T C_d^{-1} (\boldsymbol{d} - G\boldsymbol{m}_{LS})}$$

$$\boxed{\boldsymbol{m}_{LS} = (G^T C_d^{-1} G)^{-1} G^T C_d^{-1} \boldsymbol{d} = G^{-g} \boldsymbol{d}}$$

- Goodness of fit criteria tells us whether the least squares model adequately fits the data, given the level of noise.

$$\phi(\boldsymbol{m}_{LS}) \rightarrow \chi^2_{N-M}$$

Chi-square with *N-M* degrees of freedom

- The covariance matrix describes how noise propagates from the data to the estimated model

$$\boxed{C_M = (G^T C_d^{-1} G)^{-1}}$$

$$\Delta^2 \rightarrow \chi^2_M$$

$$\boxed{(\boldsymbol{m} - \boldsymbol{m}_{LS})^T C_M^{-1} (\boldsymbol{m} - \boldsymbol{m}_{LS}) < \Delta^2}$$

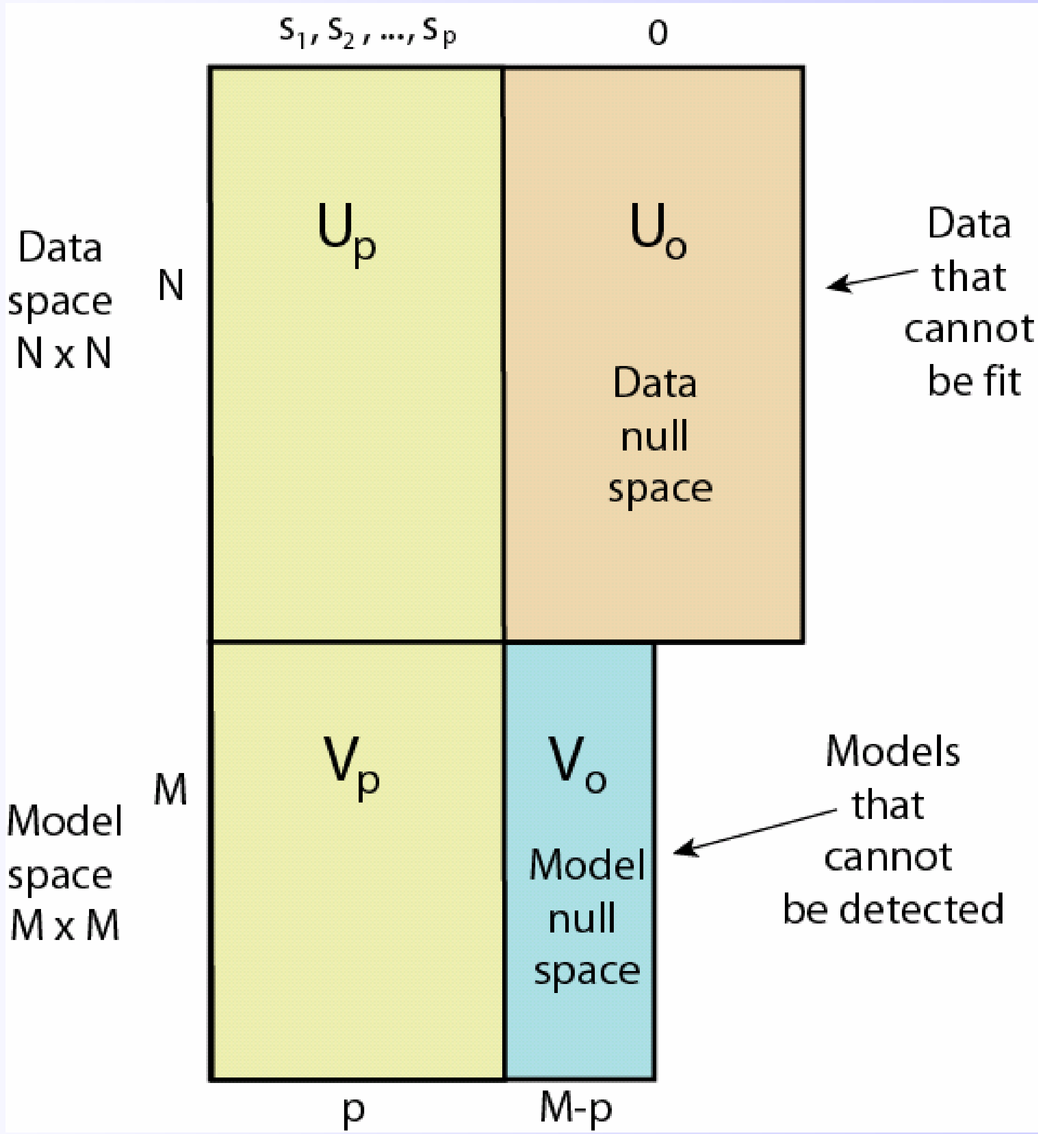
Chi-square with *M* degrees of freedom

Gives confidence intervals

- The resolution matrix describes how the estimated model relates to the true model

SVD

Data and model null spaces



Moore-Penrose inverse and data null space

$$\boldsymbol{d} = G\boldsymbol{m} \qquad G = U_p S_p V_p^T$$

The Moore-Penrose pseudo or generalized inverse of G is written

It is the unique matrix that satisfies four special properties

$$\boxed{G^\dagger = V_p S_p^{-1} U_p^T}$$

$$\boxed{GG^\dagger G = G} \qquad \boxed{(GG^\dagger)^T = GG^\dagger}$$

$$\boxed{G^\dagger G G^\dagger = G^\dagger} \qquad \boxed{(G^\dagger G)^T = G^\dagger G}$$

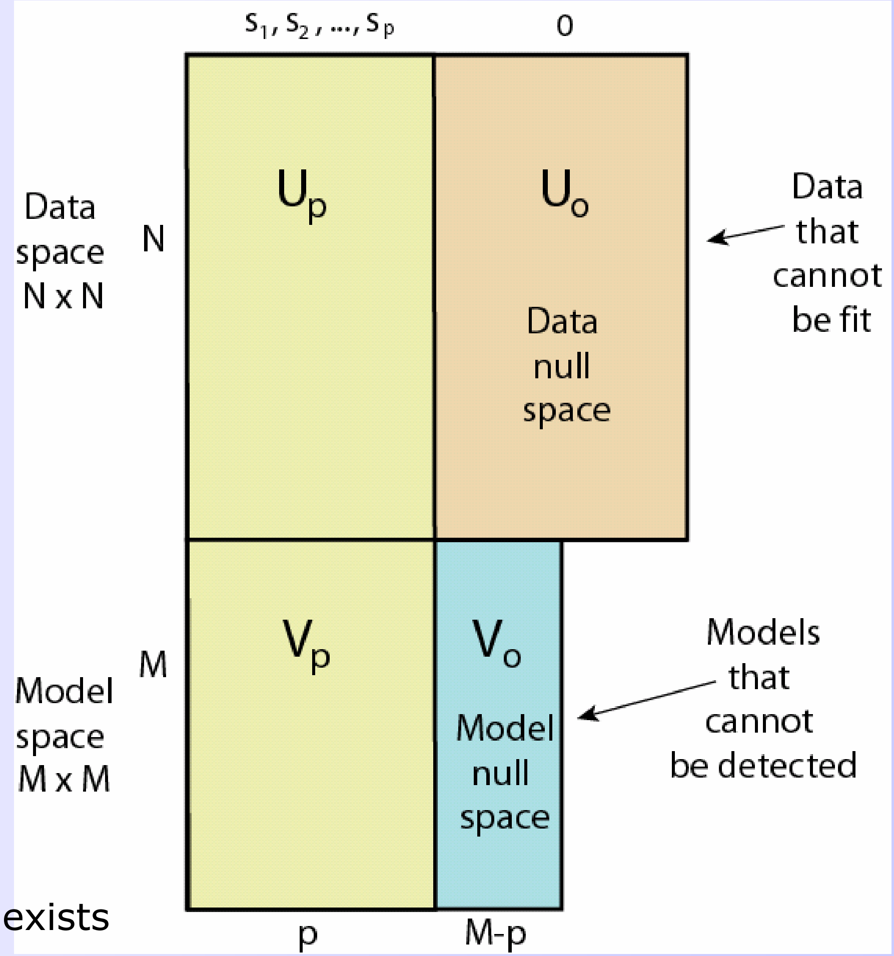
Even when G has zero eigenvalues the Moore-Penrose inverse always exists and has desirable properties

$$\boxed{\boldsymbol{m}^\dagger = G^\dagger \boldsymbol{d}_{obs} = V_p S_p^{-1} U_p^T \boldsymbol{d}_{obs}}$$

Properties of the Moore-Penrose inverse

U_p and V_p always exist and there are FOUR possible situations to consider

- Case 1:
  No model or data null space
  U_o and V_o do not exist
- Case 2:
  Only a model null space exists
  U_o does not exist, V_o exists
- Case 3:
  Only a data null space exists
  U_o exists, V_o does not exist
- Case 4:
  Both data and model null space exists
  U_o and V_o exists



Covariance and Resolution of the pseudo inverse

How does data noise propagate into the model ?

What is the model covariance matrix for the generalized inverse ?

$$C_M = G^\dagger C_d (G^\dagger)^T$$

$$G^\dagger = V_p S_p^{-1} U_p^T$$

For the case $C_d = \sigma^2 I$

$$C_M = \sigma^2 G^\dagger (G^\dagger)^T$$

$$= \sigma^2 V_p S_p^{-2} V_p^T$$

Recall that S_p is a diagonal matrix of singular ordered values

$$S_p = diag[s_1, s_2, \ldots, s_p]$$

$$\Rightarrow C_M = \sigma^2 \sum_{i=1}^{p} \frac{\boldsymbol{v}_i \boldsymbol{v}_i^T}{s_i^2}$$

As the number of singular values, p, increases the variance of the model parameters increases !

Covariance and Resolution of the pseudo inverse

How is the estimated model related to the true model ?

Model resolution matrix

$$m^{\dagger} = R m_{true}$$

$$\begin{aligned} R &= G^{\dagger} G \\ &= V_p S_p^{-1} U_p^T U_p S_p V_p^T \\ &= V_p V_p^T \end{aligned}$$

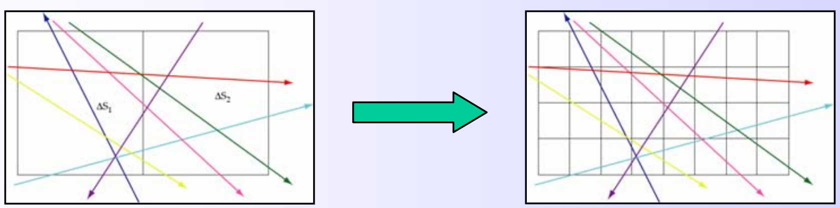
$$G^{\dagger} = V_p S_p^{-1} U_p^T$$

As p increases the model null space decreases

$$p \rightarrow M : \quad V_p^T \rightarrow V_p^{-1}, \quad R \rightarrow I$$

As the number of singular values, p, increases the resolution of the model parameters increases !

We see the trade-off between variance and resolution

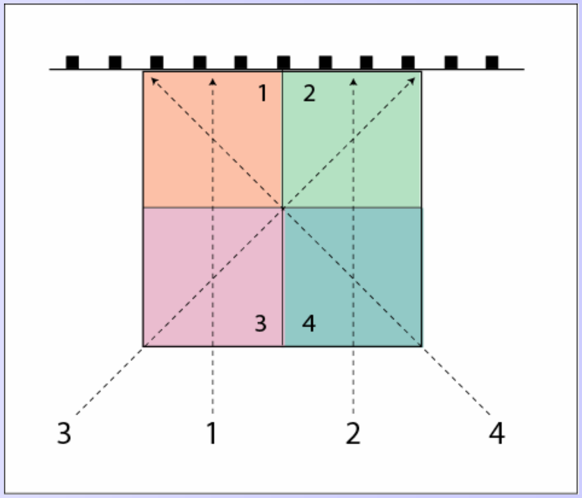


Worked example: tomography

Using rays 1- 4

$$\delta \boldsymbol{d} = G \delta \boldsymbol{m}$$

$$G = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & \sqrt{2} & \sqrt{2} & 0 \\ \sqrt{2} & 0 & 0 & \sqrt{2} \end{bmatrix}$$



$$G^T G = \begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & 2 & 1 \\ 1 & 2 & 3 & 0 \\ 2 & 1 & 0 & 3 \end{bmatrix}$$

This has eigenvalues 0, 2, 4, 6.

$$V_p = \begin{bmatrix} 0.5 & -0.5 & -0.5 \\ 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 \\ 0.5 & -0.5 & 0.5 \end{bmatrix} \qquad V_o = \begin{bmatrix} 0.5 \\ 0.5 \\ -0.5 \\ -0.5 \end{bmatrix} \qquad G \boldsymbol{v}_o = \boldsymbol{0}$$

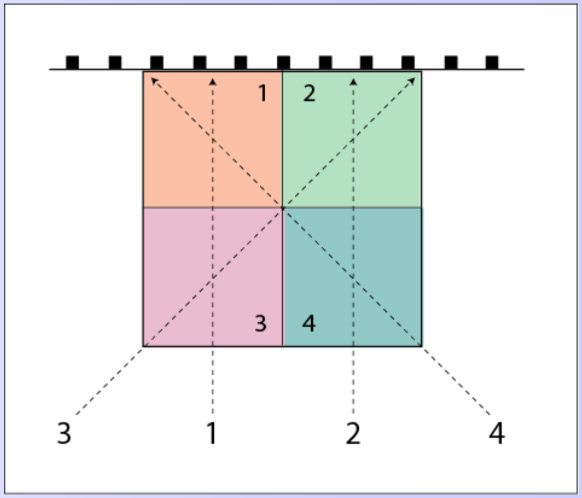
$s_1^2 = 6$ $s_2^2 = 4$ $s_3^2 = 2$ $s_4^2 = 0$

Worked example: tomography

Using rays 1- 4

$$\boldsymbol{\delta d = G\delta m}$$

$$G = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & \sqrt{2} & \sqrt{2} & 0 \\ \sqrt{2} & 0 & 0 & \sqrt{2} \end{bmatrix}$$



$$G^T G = \begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & 2 & 1 \\ 1 & 2 & 3 & 0 \\ 2 & 1 & 0 & 3 \end{bmatrix}$$

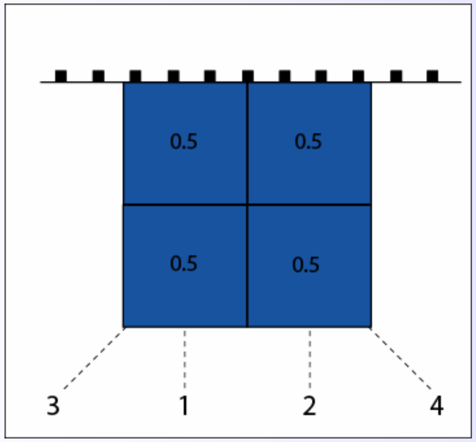
This has eigenvalues 0, 2, 4, 6.

$$V_p = \begin{bmatrix} 0.5 & -0.5 & -0.5 \\ 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 \\ 0.5 & -0.5 & 0.5 \end{bmatrix} \qquad V_o = \begin{bmatrix} 0.5 \\ 0.5 \\ -0.5 \\ -0.5 \end{bmatrix} \qquad \boldsymbol{Gv_o = 0}$$

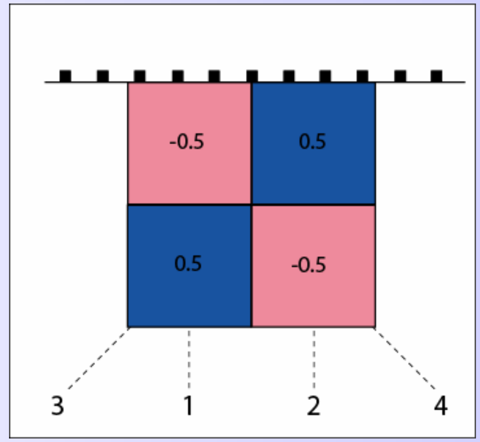
$s_1^2 = 6$ $\quad s_2^2 = 4$ $\quad s_3^2 = 2$ $\quad s_4^2 = 0$

Worked example: Eigenvectors

$S_1^2=6$

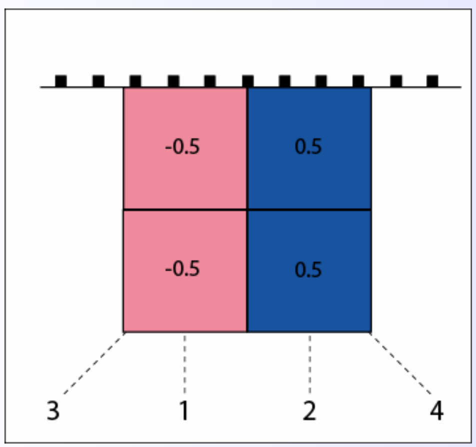


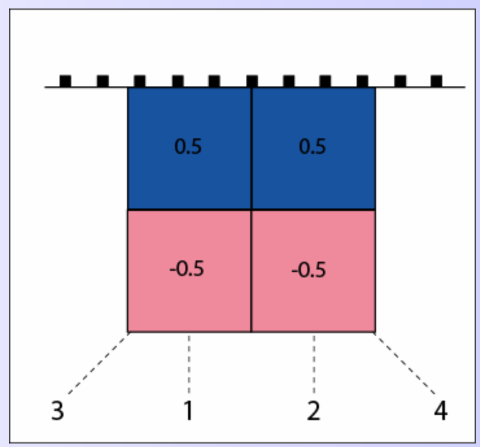
$S_2^2=4$



$$V_p = \begin{bmatrix} 0.5 & -0.5 & -0.5 \\ 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 \\ 0.5 & -0.5 & 0.5 \end{bmatrix}$$

$S_3^2=2$





$$V_o = \begin{bmatrix} 0.5 \\ 0.5 \\ -0.5 \\ -0.5 \end{bmatrix}$$

Worked example: tomography

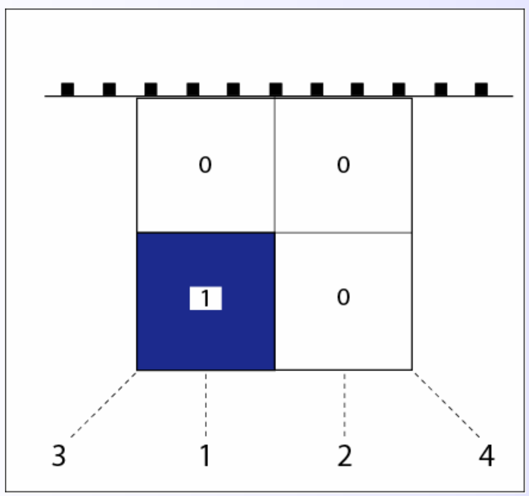
Using eigenvalues all non zero eigenvalues s_1, s_2 and s_3 the resolution matrix becomes

$$\delta \boldsymbol{m} = R \delta \boldsymbol{m}_{true} = V_p V_p^T \delta \boldsymbol{m}_{true}$$

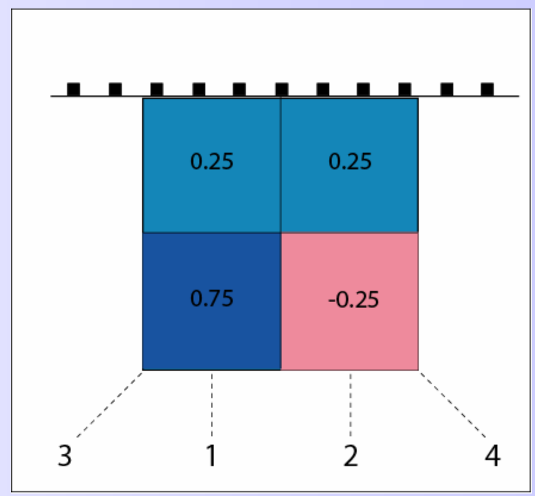
$$V_p = \begin{bmatrix} 0.5 & -0.5 & -0.5 \\ 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 \\ 0.5 & -0.5 & 0.5 \end{bmatrix}$$

$$R = \begin{bmatrix} 0.75 & -0.25 & 0.25 & 0.25 \\ -0.25 & 0.75 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.75 & -0.25 \\ 0.25 & 0.25 & -0.25 & 0.75 \end{bmatrix}$$

Input model



Recovered model



Worked example: tomography

Using eigenvalues s_1, s_2 and s_3 the model covariance becomes

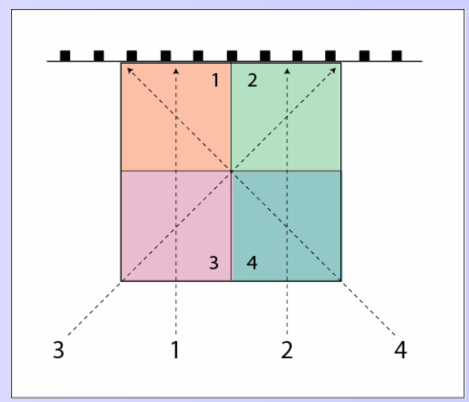
$$\Rightarrow C_M = \sigma^2 \sum_{i=1}^{p} \frac{\boldsymbol{v}_i \boldsymbol{v}_i^T}{s_i^2}$$

$$\boldsymbol{V_p} = \begin{bmatrix} -0.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & 0.5 \\ 0.5 & -0.5 & 0.5 \end{bmatrix}$$

$s_1^2 = 2$ $s_2^2 = 4$ $s_3^2 = 6$

$$C_M = \frac{\sigma^2}{4}\left\{\frac{1}{2}\begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} + \frac{1}{6}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}\right\}$$

$$C_M = \frac{\sigma^2}{48}\begin{bmatrix} 11 & -7 & 5 & -1 \\ -7 & 11 & -1 & 5 \\ 5 & -1 & 11 & -7 \\ -1 & 5 & -7 & 11 \end{bmatrix}$$



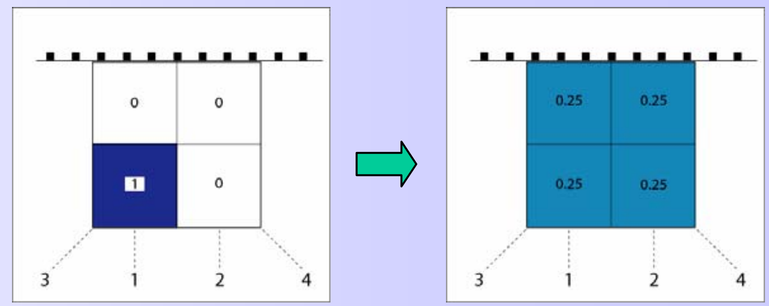
Worked example: tomography

Repeat using only one singular value s_3 =6

$$V_p = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$$

Model resolution matrix

$$R = V_p V_p^T = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

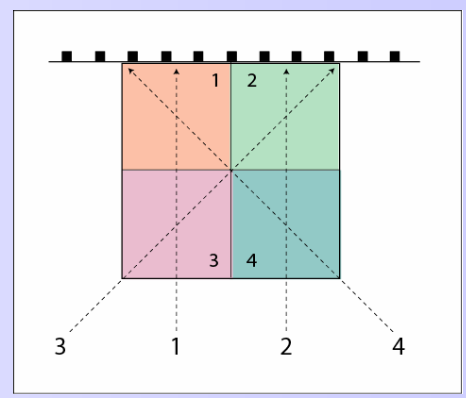


Input Output

Model covariance matrix

$$C_M = \sigma^2 \sum_{i=1}^{p} \frac{\boldsymbol{v}_i \boldsymbol{v}_i^T}{s_i^2}$$

$$= \frac{\sigma^2}{24} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$



Recap: Singular value decomposition

- There may exist a model null space -> models that can not be constrained by the data.
- There may exist a data null space -> data that can not be fit by any model.
- The general linear discrete inverse problem may be simultaneously under and over determined (mix-determined).
- Singular value decomposition is a framework for dealing with ill-posed problems.
- The Pseudo inverse is constructed using SVD and provides a unique model with desirable properties.
  - Fits the data in a least squares sense
  - Gives a minimum length model (no component in the null space)
- Model Resolution and Covariance can be traded off by choosing the number of eigenvalues to use in reconstruction.

Regularization

Regularization

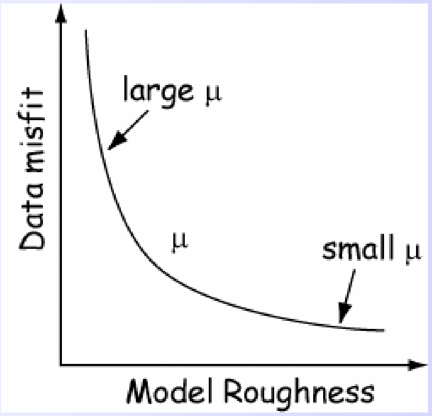
The idea behind SVD is to limit the degree of freedom in the model and fit the data to an acceptable level. Retain only those features necessary to fit the data.

A general framework for solving non-unique inverse problems is to introduce regularization. Regularization makes a non-unique problem become a unique problem. How does it do this ?

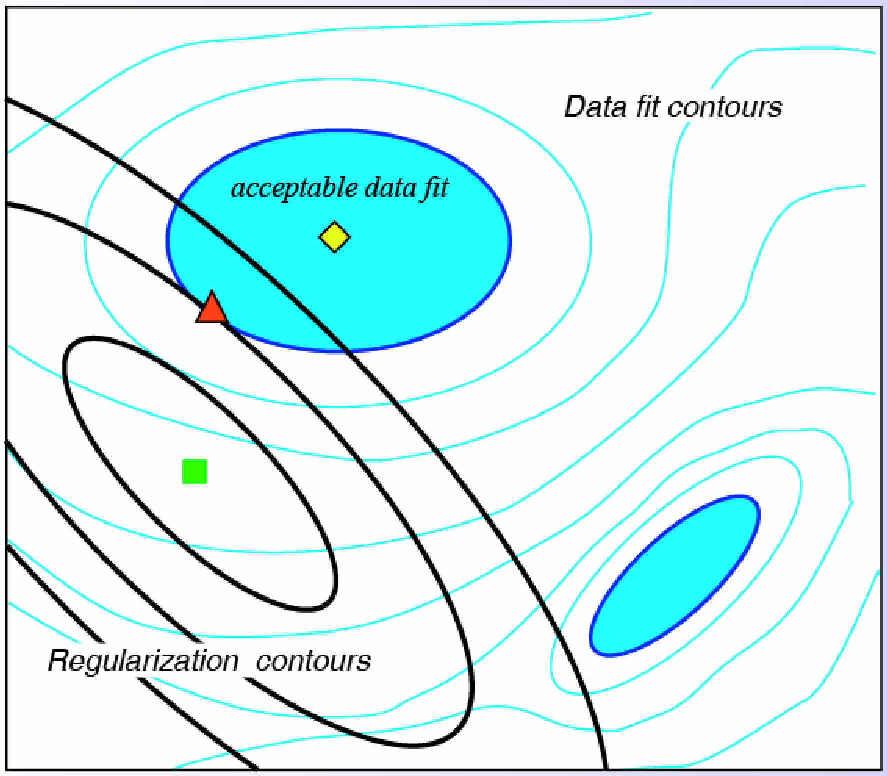
We want to minimize a combination of data misfit and some property of the model that measures extravagant behaviour, e.g.

$$\phi(\boldsymbol{m}) = (\boldsymbol{d} - g(\boldsymbol{m}))^T C_d^{-1} (\boldsymbol{d} - g(\boldsymbol{m})) + \mu (\boldsymbol{m} - \boldsymbol{m}_o)^T C_m^{-1} (\boldsymbol{m} - \boldsymbol{m}_o)$$

Inverse problem $\rightarrow$ nonlinear optimization problem



Solutions to inverse problems



Optimal data fitting model

Extremal model

Data acceptable models

The general framework is optimization

$$\phi(\boldsymbol{m}) = (\boldsymbol{d} - g(\boldsymbol{m})^T C_d^{-1}(\boldsymbol{d} - g(\boldsymbol{m}) + \mu(\boldsymbol{m} - \boldsymbol{m}_o)^T C_M^{-1}(\boldsymbol{m} - \boldsymbol{m}_o)$$

Damped least squares

We seek the model that minimizes

$$\phi(\boldsymbol{d},\boldsymbol{m}) = (\boldsymbol{d}-G\boldsymbol{m})^T C_d^{-1}(\boldsymbol{d}-G\boldsymbol{m})+\mu(\boldsymbol{m}-\boldsymbol{m}_o)^T C_M^{-1}(\boldsymbol{m}-\boldsymbol{m}_o)$$

After some algebra we get

$$\boldsymbol{m}_{DLS} = (G^T C_D^{-1} G+\mu C_M^{-1})^{-1}(G^T C_D^{-1}\boldsymbol{d}+\mu C_M^{-1}\boldsymbol{m}_o)$$

This is the damped least squares solution. A special case is to minimise a weighted sum of the data misfit and the model norm.

$$\text{min} \quad ||(\boldsymbol{d} - G\boldsymbol{m})||_2^2 + \mu||\boldsymbol{m}||_2^2$$

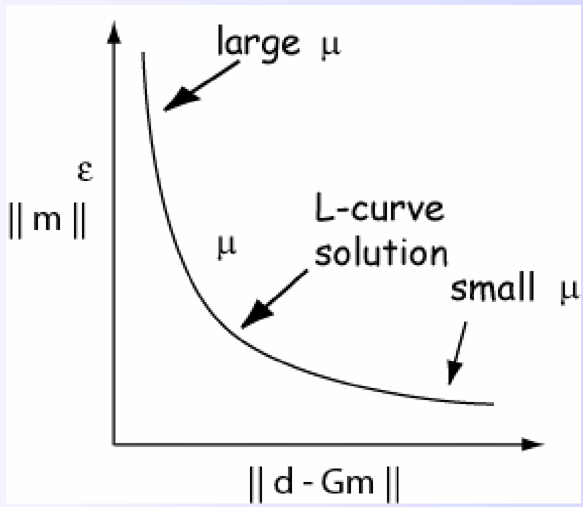
Normal equations become

$$(G^T G + \mu I)\boldsymbol{m} = G^T\boldsymbol{d}$$

this system of linear equations will have a unique solution

which is called the Tikhonov solution

L-curve solution



Choose μ that is nearest the elbow of the curve

With this value of μ we can construct a particular Tikhonov solution

$$(G^T G + \mu I)\boldsymbol{m} = G^T \boldsymbol{d}$$

Perform repeat solutions of the normal equations for different μ and select the one which lies near the elbow of the trade-off curve

Elbow is estimated visually -> potentially subjective if elbow is not clear

See example 5.1 of Aster et al. (2005)

Tikhonov: Discrepancy principle

Discrepancy principle gives us a value of δ and consequently a value for μ

If the N data have Gaussian errors with known variance $N(0,\sigma^2)$ then we would expect that on average that each residual $(d_i - G_{I,j}\, m_j)$ would be approximately σ . Hence we have

$$||\boldsymbol{d} - G\boldsymbol{m}||^2 = N\sigma^2$$

c.f. $\chi^2_N(50\%)$

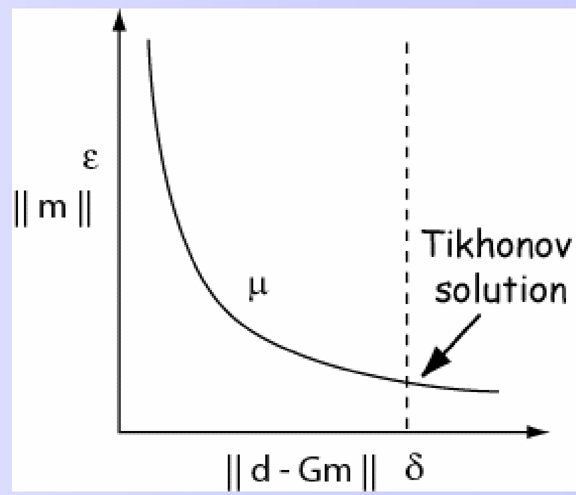
$$\Rightarrow \delta = \sigma\sqrt{N}$$

We get an expected value of the norm of the prediction error from the number of data and the standard deviation of the data

$$(G^T G + \mu I)\boldsymbol{m} = G^T\boldsymbol{d}$$

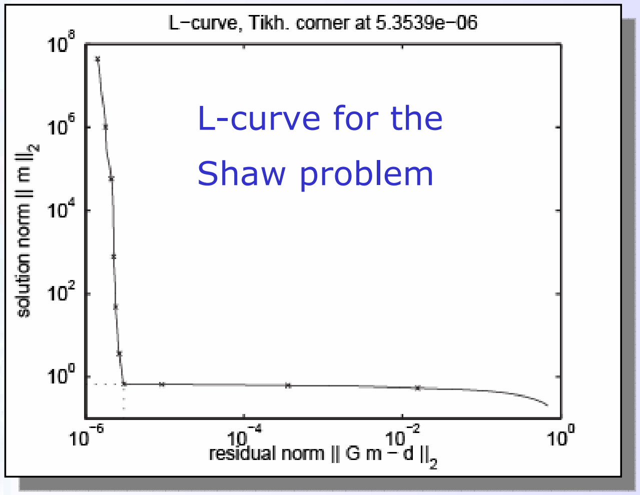
Perform repeated solutions until

$$||\boldsymbol{d} - G\boldsymbol{m}|| = \delta$$



Tikhonov regularization: Shaw problem

$$(G^T G + \mu I)\boldsymbol{m} = G^T \boldsymbol{d}$$



Fortunately a nice clear elbow

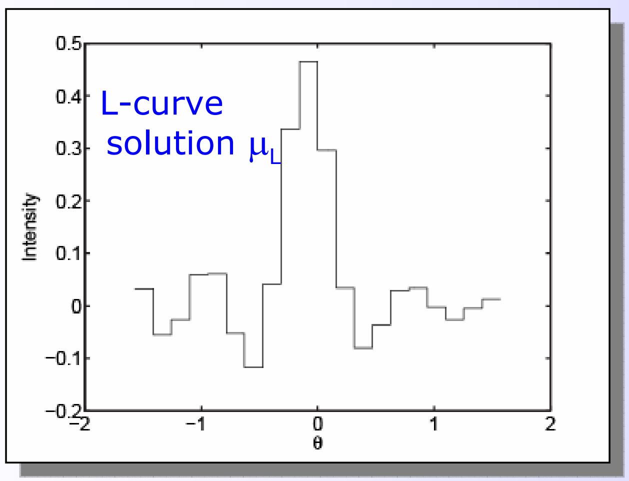
is visible in this log-log plot, Trade off parameter μ

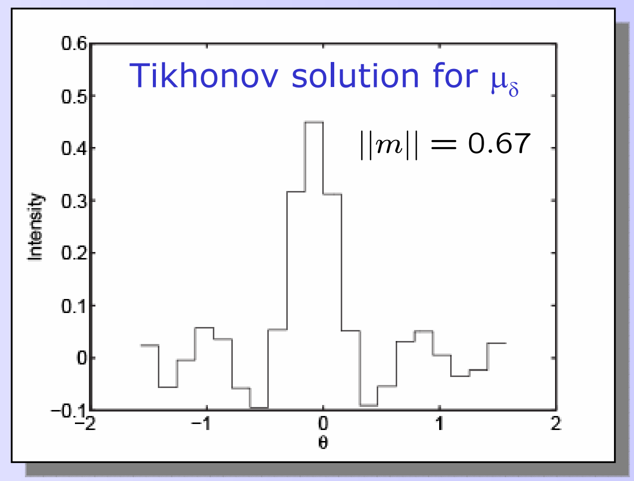
is easily picked. $N[0,(10^{-6})^2]$

$$\sqrt{\mu_L} = 5.35 \times 10^{-6}$$

$$\delta = 20 \times 10^{-12} = 4.47 \times 10^{-6}$$

$$\sqrt{\mu_\delta} = 4.29 \times 10^{-6}$$





SVD and Tikhonov regularization

Tkihonov solution

$$\boxed{G = U_p S_p V_p^T}$$

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

It can be shown... that this can be written as

Look !

Page 91-93 of

Aster et al (2005)

$$\boldsymbol{m} = \sum_{i=1}^{M} \frac{s_i^2}{s_i^2 + \mu} \left(\frac{\boldsymbol{u}_i \cdot \boldsymbol{d}}{s_i} \right) \boldsymbol{v}_i$$

$$\boldsymbol{m}^\dagger = \sum_{i=1}^{p} \left(\frac{\boldsymbol{u}_i \cdot \boldsymbol{d}}{s_i} \right) \boldsymbol{v}_i$$

SVD

Now we get very different behaviour

Let

$$f_i = \frac{s_i^2}{s_i^2 + \mu}$$

Filter factors

As singular values $s_i \to 0$ the solution is not highly sensitive to noise

Because $f_i \to 0$ rather than ∞.

If $s_i >> \sqrt{\mu}$ then $f_i \to 1$

If $s_i << \sqrt{\mu}$ then $f_i \to 0$

SVD

$s_i > s_p \quad f_i = 1$

$s_i < s_p \quad f_i = 0$

μ filters out (damps) the unstable influence of the small eigenvalues