

## # Linear Inverse Problems

- Problems of the form  $d = G m$

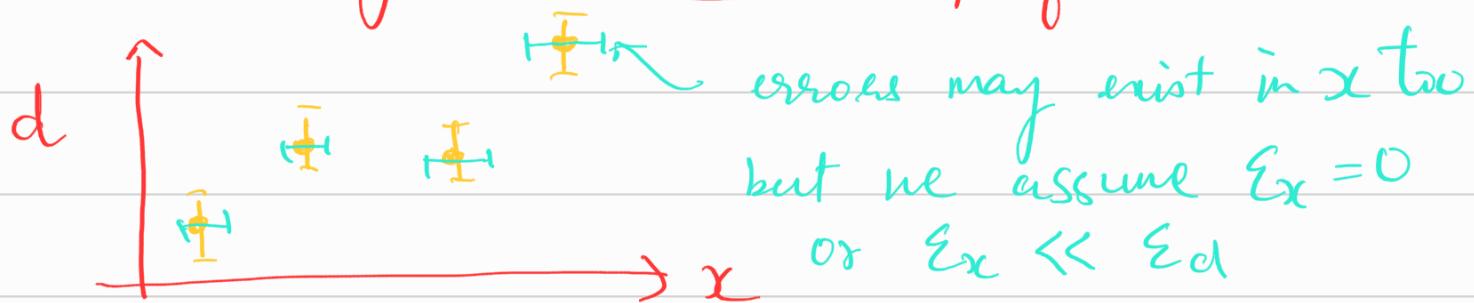
$$d \rightarrow \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_N \end{bmatrix} \quad m \rightarrow \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_M \end{bmatrix}$$

$$\Rightarrow G = \begin{bmatrix} & \leftarrow M \rightarrow \\ \uparrow N & \\ \downarrow & \end{bmatrix}$$

If there is a true set of model parameters,  $\rightarrow m$ , the best estimate of this is  $\hat{m}$   
 we want  $\hat{m} \sim m$ , however

$$d = G m + \varepsilon \leftarrow \text{observation errors}$$

→ assuming the model is perfect



From our best model,  $\hat{d} = G \hat{m}$

$$\Rightarrow d - \hat{d} = \varepsilon$$

We want to minimize  $\varepsilon$ ,

$$\phi = (d - \hat{d})^T (d - \hat{d}) = \sum_{i=1}^N \underbrace{(d_i - \hat{d}_i)^2}_{\hat{d} = \sum \hat{m}_i}$$

$d - Gm$

Least-squares

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

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

$$\phi = d^T d - d^T G \hat{m} - \hat{m}^T G^T d + \hat{m}^T G^T G \hat{m}$$

to minimize  $\phi$  w.r.t  $\hat{m}$ ,

$$\frac{\partial \phi}{\partial \hat{m}} = 0.$$

$$\phi = \underbrace{d^T d}_{\text{scalar}} - 2 \underbrace{d^T G \hat{m}}_{\text{scalar}} + \underbrace{\hat{m}^T G^T G \hat{m}}$$

$$\begin{aligned} \frac{\partial \hat{m}^T G^T G \hat{m}}{\partial \hat{m}} &= \frac{\partial [\hat{m}^T (G^T G \hat{m})]}{\partial \hat{m}} + \frac{\partial [(\hat{m}^T G^T G) \hat{m}]}{\partial \hat{m}} \\ &= (G^T G \hat{m})^T + (\hat{m}^T G^T G) \\ &= 2 \hat{m}^T G^T G \end{aligned}$$

consider  $\alpha = x^T A x$ ,  $\frac{\partial \alpha}{\partial x} = \frac{\partial}{\partial x} (x^T A x)$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1N} \\ A_{21} & \ddots & \ddots & \vdots \\ A_{31} & \ddots & \ddots & \vdots \\ \vdots & & & \ddots \\ A_{N1} & \ddots & \ddots & \ddots \end{bmatrix} \quad \begin{aligned} \frac{\partial \alpha}{\partial x} &= \frac{\partial}{\partial x} (x^T A x) \\ &= x^T (A^T + A) \end{aligned}$$

Proof:

$$\alpha = [x_1, x_2, \dots, x_N] \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1N} \\ A_{21} & & & \\ \vdots & & & \\ A_{N1} & & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

$$= [x_1, x_2, \dots, x_N] \{ A_{ij} x_j \} = A_{11} x_1 + A_{12} x_2 + \dots$$

$$A_{1j} x_j + A_{2j} x_j + \dots + A_{nj} x_j$$

$$\alpha = x_1 (A_{1j} x_j) + x_2 (A_{2j} x_j) + x_3 (A_{3j} x_j) \dots$$

$$\alpha = \sum_{i=1}^N x_i \sum_{j=1}^N A_{ij} x_j$$

$$\frac{\partial \alpha}{\partial x_i} = \frac{\partial}{\partial x_i} \left[ x_1 (A_{11} x_1 + A_{12} x_2 + A_{13} x_3 + \dots) + x_2 (A_{21} x_1 + A_{22} x_2 + A_{23} x_3 + \dots) + x_3 (A_{31} x_1 + A_{32} x_2 + A_{33} x_3 + \dots) + \dots - - - - - \right]$$

$$= (A_{1j} x_j) + x_1 A_{11} + A_{21} x_2 + A_{31} x_3 + \dots + A_{N1} x_N$$

$$= A_{1j} x_j + A_{j1} x_j$$

$$= x_j (A_{1j} + A_{j1})$$

$$\frac{\partial \alpha}{\partial x_2} = x_1 (A_{2j} + A_{j2}) \dots \text{and so on, } [\dots]$$

This shows that  $\frac{\partial \alpha}{\partial x_i} = x^T \begin{pmatrix} A_{1j} & A_{j1} \\ \vdots & \vdots \end{pmatrix}$

$$\frac{\partial \underline{x}}{\partial \underline{x}} = \underline{x}^T (A + A^T)$$

coming back to  $\phi$

$$\frac{\partial \phi}{\partial \hat{m}} = -2 \frac{\partial (d^T G \hat{m})}{\partial \hat{m}} + \frac{\partial (\hat{m}^T G^T G \hat{m})}{\partial \hat{m}}$$

$$= -2 d^T G + (G^T G \hat{m})^T + \hat{m}^T G^T G$$

$$0 = -2 d^T G + 2 \hat{m}^T G^T G$$

$$\Rightarrow d^T G = \hat{m}^T G^T G$$

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

This gives us an analytical closed-form expression for the least-squares solution

our assumptions about the data & model

$$\text{were } d = G m + \varepsilon \rightarrow N(0, \sigma^2)$$

$$\Rightarrow E(\varepsilon) = 0, \text{cov}(\varepsilon) = \sigma^2 I$$

what about  $\hat{m}$ ?

$$\begin{aligned} E(\hat{m}) &= E((G^T G)^{-1} G^T d) \\ &= E((G^T G)^{-1} G^T G m + (G^T G)^{-1} G^T \varepsilon) \\ &= (G^T G)^{-1} G^T E(m) + (G^T G)^{-1} G^T E(\varepsilon) \end{aligned}$$

$$E(\hat{m}) = E(m)$$

$$\text{cov}(\hat{m}) = \text{cov}\left((G^T G)^{-1} G^T (Gm + \epsilon)\right)$$

$$\left\{ \begin{array}{l} \text{cov}(m) = 0, \text{cov}(\epsilon) = \sigma^2 I \end{array} \right.$$

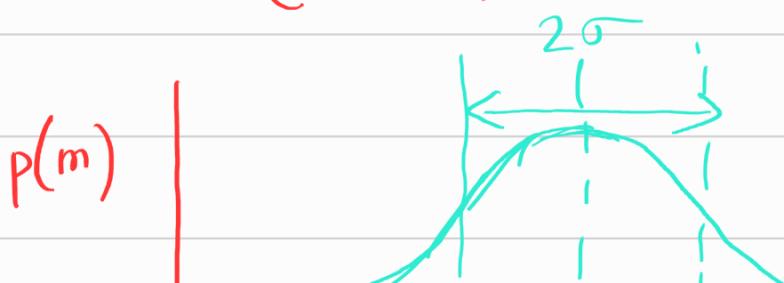
$$\begin{aligned} \text{cov}(\hat{m}) &= \text{cov}\left((G^T G)^{-1} G^T \epsilon\right) && \text{recall } \text{cov}(Am) \\ &= (G^T G)^{-1} G^T [\sigma^2 I] (G^T G)^{-1 T} && = A \text{cov}(m) A^+ \\ &= (G^T G)^{-1} G^T [\sigma^2 I] G (G^T G)^{-T} \\ \text{cov}(\hat{m}) &= \sigma^2 (G^T G)^{-T} &= \sigma^2 (G^T G)^{-1} \end{aligned}$$

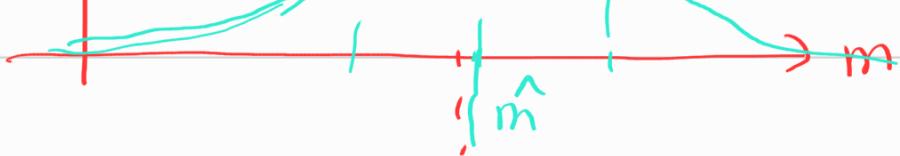
This means we can fully describe  $\hat{m}$  using a Gaussian probability distribution,

$$\hat{m} \sim N\left((G^T G)^{-1} G^T d, \underbrace{\sigma^2 (G^T G)^{-1}}_{\sigma^2 C_m}\right)$$

## # Gaussian probability distributions

$$p(m) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \left| C_m^{-1} \right|^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} (m - \hat{m})^T C_m^{-1} (m - \hat{m})\right)$$





## # Bayes' Theorem

$$P(m_i | d) = \frac{P(d|m_i) p(m_i)}{\sum_{i=1}^M P(d|m_i) p(m_i)}$$

$$\Rightarrow P(m|d) \propto \frac{P(d|m) p(m)}{P(d)}$$

  $P(\text{ball} | \text{red}) = \frac{3}{5}$  Bayes'

$$P(\text{ball} | \text{red}) = \frac{P(\text{red} | \text{ball}) p(\text{ball})}{p(\text{red})}$$

$$p(\text{red}) = \left( \frac{3}{7} \times \frac{7}{12} \right) = \frac{3}{5}$$

$$= \frac{P(\text{red} | \text{ball}) p(\text{ball})}{P(\text{red} | \text{square}) p(\text{square})} = \frac{(5/12)}{(5/12)}$$

$$= \frac{3}{7} \times \frac{7}{12} + \frac{2}{8} \times \frac{5}{12} = \frac{5}{12}$$

$$\text{for } d = G \hat{m} + \varepsilon \leftarrow N(0, \sigma^2 C_d)$$

$$P(d | \hat{m}, \sigma^2) \propto \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma^2} (d - G \hat{m})^T C_d^{-1} (d - G \hat{m}) \right)$$

To find the best  $\hat{m}$ , we need to maximize  $p(d|\hat{m}, \sigma^2)$  ← likelihood function

$$\log(p(d|\hat{m}, \sigma^2)) = -\frac{1}{2\sigma^2} (d - G\hat{m})^T (G^T G)^{-1} (d - G\hat{m})$$

maximize this function

⇒ same as minimizing  $\phi = (d - G\hat{m})^T (G^T G)^{-1} (d - G\hat{m})$

least-squares.

The least-squares minimization method is identical to maximizing the log-likelihood function which in a Bayesian framework is the same as maximizing the posterior distribution with a uniform prior.

$$p(m|d) \propto p(d|m) p(m)$$

↑  
posterior

↑  
likelihood

uniform i.e

$$p(m) \propto 1 \quad \forall m \in \mathbb{R}^M$$

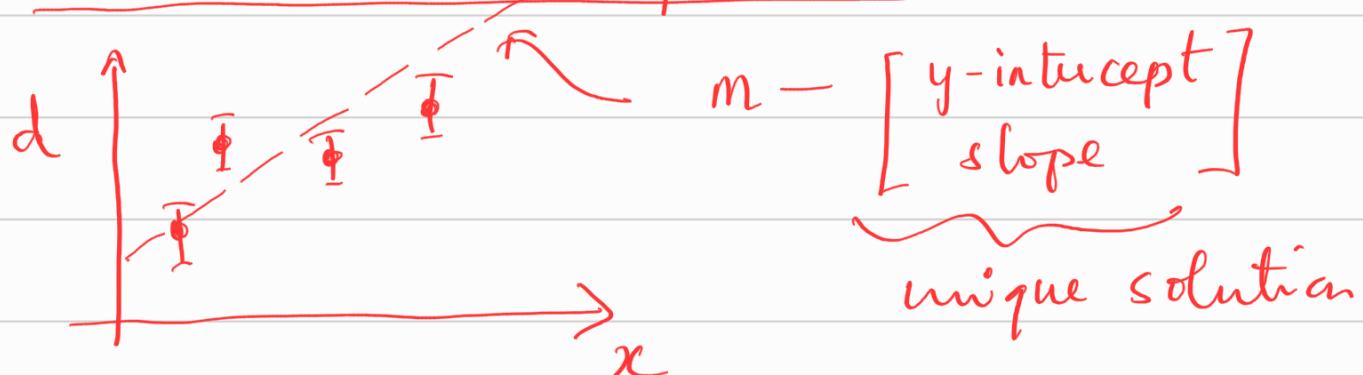
$$\xrightarrow{N(\hat{m}, \sigma^2(m))}$$

Gaussian

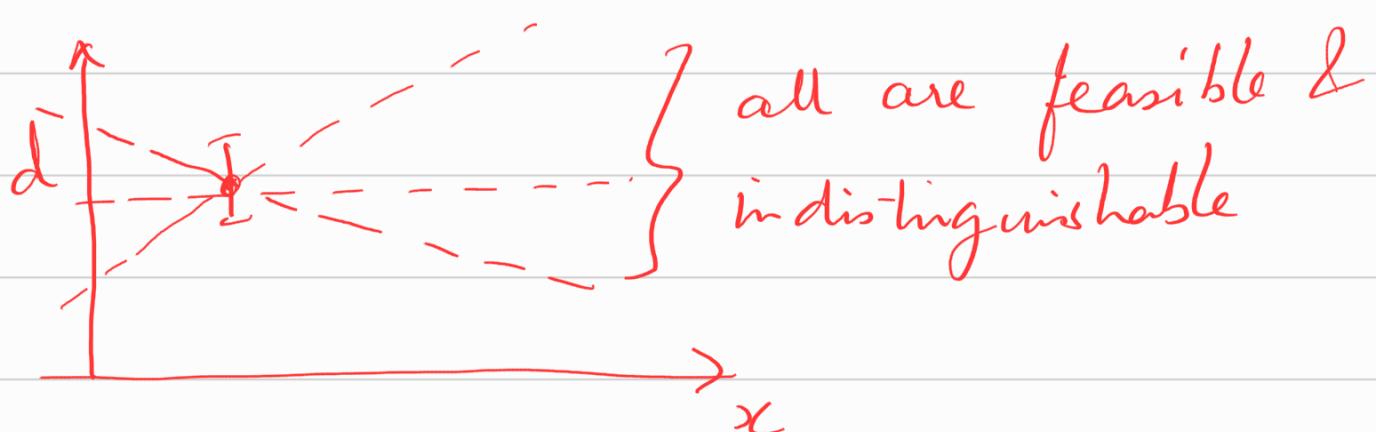
product is also Gaussian

7) Fitting data & resolution of issues

### Over-determined problems



### Under-determined problems



### Mix-determined problems

- Some parameters can be resolved while others cannot be.

In terms of the design matrix, we can define a sensitivity kernel

$$S = \text{diag}(\mathbf{G}^T \mathbf{G})$$

$m_1 \ m_2$  square matrix  $[N_m \times N_m]$

$$G = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ \vdots \end{bmatrix} \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 0 \end{bmatrix}$$

$$G^T = \begin{bmatrix} d_1, d_2, \dots & m_1 \\ d_1, d_2, \dots & m_2 \\ \vdots & \vdots \\ d_1, d_2, \dots & m_n \end{bmatrix} \quad \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ \vdots \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ \vdots \end{bmatrix}$$

$$G^T G = \begin{bmatrix} (d_1^2 + d_2^2 + \dots) & m_1 \\ (d_1^2 + d_2^2 + \dots) & m_2 \\ \vdots & \vdots \\ (d_1^2 + d_2^2 + \dots) & m_n \end{bmatrix}$$

tells us the strength of the signal of a given model parameter  $m_i$  due to the arrangement of the data.

if  $G \rightarrow$  geodetic network, then  $G^T G$  contains the sensitivity of each fault patch to the station configuration.

A patch  $m_i$  located far away from the network will have lower  $S_{ii}$  than a patch located close to stations.

→ Because we like linear-inverse methods, we are forced to use non-uniform priors to condition mixed-determined problems.

## Priors, data-likelihood & posteriors

→ Regularization → prior information on m

$$p(m|d) = \underbrace{p(d|m) p(m)}_{\text{Posterior}}$$

$$\int_M p(d|m) p(m) dm$$

$$p(d|m) \propto \exp\left(-\frac{(d - Gm)^T C_a^{-1} (d - Gm)}{2\sigma^2}\right)$$

$p(m) \rightarrow$  we can say  $m$  is smooth  
 $\Rightarrow \nabla^2 m \sim 0 \rightarrow$  Laplacian  
 $\exp\left(-\frac{1}{2\alpha^2} (Lm)^T I (Lm)\right)$

$p(m) \rightarrow$  we already have a guess for  $m$  as  $m_0$

$$\Rightarrow \exp\left(-\frac{1}{2\alpha^2} (m - m_0)^T C_{m_0}^{-1} (m - m_0)\right)$$

If  $m_0 \sim 0$ , then  $p(m)$  is a minimum length solution  $\Rightarrow \exp\left(-\frac{1}{2\sigma^2} (Im)^T C_m^{-1} (Im)\right)$

Combining  $p(d|m) p(m)$ , we get.

$$P(m|d) \propto \exp\left(-\frac{1}{2\sigma^2} (d-Gm)^T C_d^{-1} (d-Gm) - \frac{1}{2\lambda^2} (m-m_0)^T C_{m_0}^{-1} (m-m_0)\right)$$

Maximizing  $p(m|d)$  is same as minimizing  $\phi$

$$\phi = (d-Gm)^T C_d^{-1} (d-Gm) + \lambda^2 (m-m_0)^T C_{m_0}^{-1} (m-m_0)$$

$\lambda^2 = \left(\frac{\sigma^2}{\lambda^2}\right)$   $\leftarrow$  relative weight of prior & likelihood

To find  $\hat{m}$ ,  $\frac{\partial \phi}{\partial m} = 0$

$$\boxed{\hat{m} = (G^T C_d^{-1} G + \lambda^2 C_{m_0}^{-1})^{-1} G^T C_d^{-1} d}$$

$$\boxed{C_{\hat{m}} = \hat{\sigma}^2 (G^T C_d^{-1} G + \lambda^2 C_{m_0}^{-1})^{-1}}$$

for an arbitrary prior  $\underline{A m \sim m_0}$

$$\hat{m} = \left( G^T G_d^{-1} G + \lambda^2 A^T (C_{m_0}^{-1} A) \right)^{-1} G^T G_d^{-1} d$$

$$C_m = \hat{\sigma}^2 \left[ G^T G_d^{-1} G + \lambda^2 (A^T C_{m_0}^{-1} A) \right]$$

## # Non-linear Inverse problems

→ linearized version

$$d = G(m)$$

we can use the Taylor expansion of  
 $G(m)$  about some point  $m_0$ ,

$$G(m) = G(m_0) + \underbrace{\frac{\partial G}{\partial m}}_{\text{in}} \Big|_{m_0} (m - m_0) + \epsilon$$

perturbed obs

$$\Rightarrow \begin{bmatrix} d_1 - d_1^0 \\ d_2 - d_2^0 \\ d_3 - d_3^0 \\ \vdots \end{bmatrix} = \begin{bmatrix} \left( \frac{\partial G}{\partial m_1} \right)_{m_0} & \left( \frac{\partial G}{\partial m_2} \right)_{m_0} & \cdots \\ & & \ddots \end{bmatrix} \begin{bmatrix} m_1 - m_1^0 \\ m_2 - m_2^0 \\ \vdots \\ m_n - m_n^0 \end{bmatrix}$$

model perturbation

\ predictions for a guessed value  
of  $\underline{m}^0$

We can estimate  $[\underline{m} - \underline{m}_0]$  using  
matrix inversion & the uncertainties in  
 $C_{(\underline{m}-\underline{m}_0)}$ .

- This method works iteratively when  
 $\underline{m}_0$  is relatively close to  $\hat{\underline{m}} \Rightarrow$  the  
initial guess must be good for  
convergence of  $(\underline{m} - \underline{m}_0)$

- each iteration will have a new  $\underline{m}_0$   
based on the correction factor  $(\underline{m} - \underline{m}_0)$

$$\Rightarrow \boxed{\underline{m}_0^k = \underline{m}_0^{k-1} + [\hat{\underline{m}} - \underline{m}_0]_0^k}$$

The uncertainties in  $\underline{m}$  are taken  
from the uncertainties in  $C_{(\underline{m}-\underline{m}_0)}$  at  
the final iteration & is  
an underestimate -

