Linear Inverse Problems

- Problems of the form $d = Gm$

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

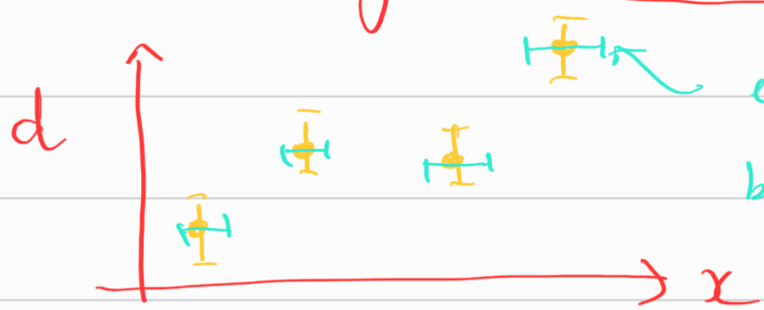
$$\Rightarrow G = \begin{array}{cc} & \xleftrightarrow{\quad M \quad} \\ \updownarrow N & \begin{bmatrix} \quad & \quad & \quad \\ & & \\ & & \end{bmatrix} \end{array}$$

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

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

$\rightarrow$ assuming the model is perfect



errors may exist in x too but we assume $\varepsilon_x = 0$ or $\varepsilon_x \ll \varepsilon_d$

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

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

we want to minimize ε,

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

$\hat{d} = G\hat{m}$

$d = G m$ 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 ϕ w.r.t $\hat{m}$,

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

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

scalar

$$\frac{\partial\, \hat{m}^T G^T G \hat{m}}{\partial \hat{m}} = \frac{\partial}{\partial \hat{m}}\left[\hat{m}^T (G^T G \hat{m})\right] + \frac{\partial}{\partial \hat{m}}\left[(\hat{m}^T G^T G)\hat{m}\right]$$
$$= (G^T G \hat{m})^T + (\hat{m}^T G^T G)$$
$$= 2\hat{m}^T G^T G$$

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

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1N} \\ A_{21} & \dots & & \cdot \\ A_{31} & \dots & & \cdot \\ \vdots & & & \\ A_{N1} & \dots & & \end{bmatrix}$$

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

Proof :

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

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

$$\begin{bmatrix} A_{2j}\, x_j \\ \vdots \\ A_{Nj}\, x_j \end{bmatrix} \quad A_{21} x_1 + A_{22} x_2 + \cdots$$

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

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

$$\frac{\partial \alpha}{\partial x_1} = \frac{\partial}{\partial x_1} \left[\begin{array}{l} x_1 (A_{11} x_1 + A_{12} x_2 + A_{13} x_3 + \cdots) \\ + x_2 (A_{21} x_1 + A_{22} x_2 + A_{23} x_3 + \cdots) \\ + x_3 (A_{31} x_1 + A_{32} x_2 + A_{33} x_3 + \cdots) \\ \cdots\cdots \end{array} \right]$$

$$= (A_{1j} x_j) + x_1 A_{11}$$
$$+ A_{21} x_2 + A_{31} x_3 + \cdots\; A_{N1} x_N$$
$$= A_{1j} x_j + A_{j1} x_j$$
$$= x_j (A_{1j} + A_{j1})$$

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

this shows that $\dfrac{\partial \alpha}{\partial x_i} = x^T (A_{ij} + A_{ji})$

$$\begin{bmatrix} \vdots \\ \vdots \end{bmatrix}$$

$\partial \alpha$

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

Coming back to ϕ

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

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

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

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

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

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

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

our assumptions about the data & model were $d = Gm + \varepsilon \rightarrow N(0, \sigma^2)$

$$\Rightarrow E(\varepsilon) = 0, \quad cov(\varepsilon) = \sigma^2 I$$

what about $\hat{m}$?

$$E(\hat{m}) = E((G^T G)^{-1} G^T \underset{\downarrow \atop (Gm+\varepsilon)}{d})$$

$$= E((G^T G)^{-1} G^T Gm + (G^T G)^{-1} G^T \varepsilon)$$

$$= \cancel{(G^T G)^{-1} G^T G}\, E(m) + (G^T G)^{-1} G^T \cancel{E(\varepsilon)}^{\,0}$$

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

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

$$\left(\quad cov(m) = 0, \quad cov(\varepsilon) = \sigma^2 I \right.$$

$$cov(\hat{m}) = cov\left((G^TG)^{-1}G^T\varepsilon\right)$$

recall $cov(Am) = A\,cov(m)\,A^T$

$$= (G^TG)^{-1}G^T\left[\sigma^2 I\right]\left((G^TG)^{-1}G^T\right)^T$$

$$= (G^TG)^{-1}G^T\left[\sigma^2 I\right] G\,(G^TG)^{-T}$$

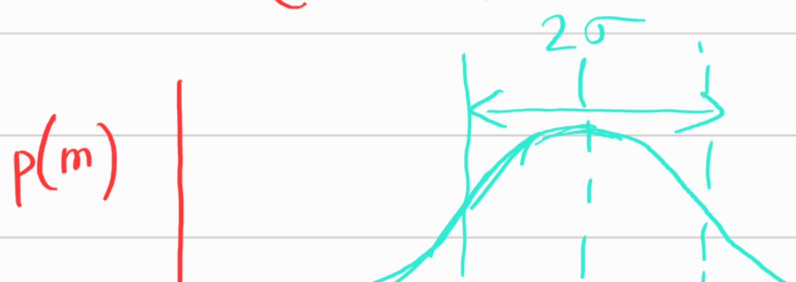
$$cov(\hat{m}) = \sigma^2 (G^TG)^{-T} = \sigma^2 (G^TG)^{-1}$$

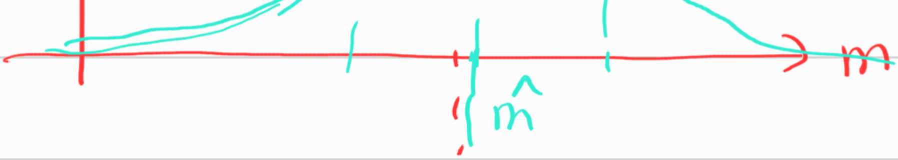
This means we can fully describe $\hat{m}$ using a Gaussian probability distribution,

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

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

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

$$\Rightarrow \boxed{p(m|d) \propto p(d|m)\,p(m)}$$

$$p(\text{ball}|\text{red}) = \frac{3}{5}$$

Bayes'

$$p(\text{ball}|\text{red}) = \frac{p(\text{red}|\text{ball})\,p(\text{ball})}{p(\text{red})} = \frac{\left(\frac{3}{\cancel{7}} \times \frac{\cancel{7}}{12}\right)}{\left(\frac{5}{12}\right)} = \frac{3}{5}.$$

$p(\text{red})$

$$= \left(p(\text{red}|\text{ball})\,p(\text{ball}) + p(\text{red}|\text{square})\,p(\text{square})\right)$$

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

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

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

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

$$\underbrace{\log(p(d|\hat{m},\sigma^2)) = \frac{-1}{2\sigma^2}(d-G\hat{m})^T C_d^{-1}(d-Gm)}_{\text{maximize this function}}$$

⇒ <u>same as minimizing</u> $\phi = \underbrace{(d-G\hat{m})^T C_d^{-1}(d-G\hat{m})}_{\text{least-squares}}$

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 → $N(\hat{m},\sigma^2 C_m)$

likelihood ↓ Gaussian

uniform i.e. $p(m) \propto 1 \; \forall m \in R^M$

$\underbrace{\text{Gaussian} \quad p(m)}_{\text{product is also Gaussian}}$

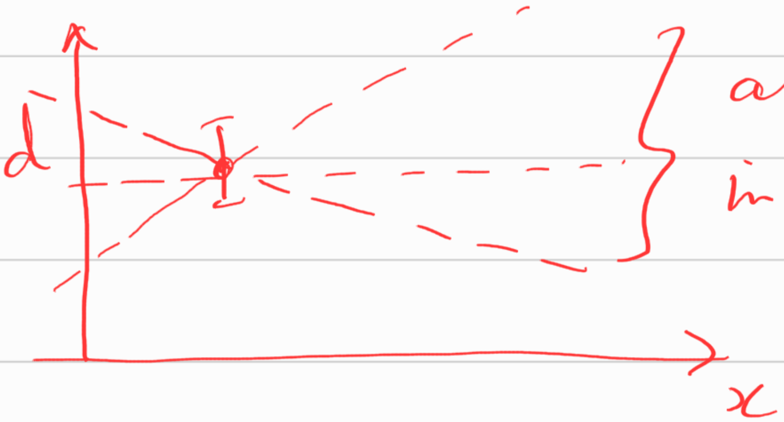
Filtering data & resolution of inverse methods

Over-determined problems

m — $\begin{bmatrix} \text{y-intercept} \\ \text{slope} \end{bmatrix}$

unique solution

Under-determined problems



all are feasible & indistinguishable

Mix-determined problems

— some parameters can be resolved while others cannot be.

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

$$S = \mathrm{diag}(\underbrace{G^T G}_{\text{square matrix } [N_m \times N_M]})$$

$m_1 \quad m_2$

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

$$G^T = \begin{array}{l} \\ m_1 \\ m_2 \\ \\ \end{array} \begin{bmatrix} d_1 & d_2 & \cdots \\ d_1 & d_2 & \cdots \\ \vdots & & \end{bmatrix} \qquad \begin{array}{ccc} m_1 & m_2 & \\ \end{array} \begin{bmatrix} d_1 & d_1 & \cdots \\ d_2 & d_2 & \\ d_3 & d_3 & \\ \vdots & \vdots & \end{bmatrix}$$

$$G^T G = \begin{bmatrix} \overbrace{(d_1^2 + d_2^2 + \cdots)}^{m_1} & & \\ & \overbrace{(d_1^2 + d_2^2 + \cdots)}^{m_2} & \\ & & \ddots \end{bmatrix}$$

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 $\underline{m}$

$$p(m|d) = \frac{p(d|m)\, p(m)}{\int_M p(d|m)\, p(m)\, dm}$$

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

$p(m) \rightarrow$ we can say $\underline{m \text{ is smooth}}$

$$\Rightarrow \nabla^2 m \sim 0 \rightarrow \text{Laplacian}$$

$$\exp\left(-\frac{1}{2\alpha^2} (Lm)^T I (Lm)\right)$$

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

Combing $p(d|m)\,p(m)$, we get.

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

maximizing $p(m|d)$ is same as minimizing ϕ

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

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

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

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

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

for an arbitrary prior $Am \sim m_0$

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

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

Non-linear Inverse problems

→ linearized version

$$d = G(m)$$

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

$$\underbrace{G(m)}_{d} = \underbrace{G(m_0)}_{d_0} + \underbrace{\left.\frac{\partial G}{\partial m}\right]_{m_0}}_{\text{Jacobian}} (m - m_0) + \varepsilon$$

$$\Rightarrow \overbrace{\begin{bmatrix} d_1 - d_1^0 \\ d_2 - d_2^0 \\ d_3 - d_3^0 \\ \vdots \end{bmatrix}}^{\text{perturbed obs}} = \overbrace{\begin{bmatrix} \left(\frac{\partial G}{\partial m_1}\right)_{m_0} & \left(\frac{\partial G}{\partial m_2}\right)_{m_0} & \cdots \end{bmatrix}}^{\text{Jacobian}} \underbrace{\begin{bmatrix} m_1 - m_1^0 \\ m_2 - m_2^0 \\ \vdots \end{bmatrix}}_{\text{model perturbation}}$$

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 m_0 is relatively close to $\hat{m}$ $\Rightarrow$ 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 $(m - m_0)$

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

The uncertainties in $\underline{m}$ are taken from the uncertainties in $C_{(m-m_0)}$ at the final iteration & is an underestimate.