Linear Model : Where were we last time?
Did I do non linear models yet? Yes, I think so?

1) start with simple approach

$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

$\vec{x} = \begin{pmatrix} r_e \\ \tau \end{pmatrix}$ @ 800 nm e.g.

2) Write down cost function

$$J(x) = \chi^2(\vec{x}) = (\vec{y} - F(\vec{x}))^T S_y^{-1} (\vec{y} - F(\vec{x})) + (x - x_A)^T S_A^{-1} (x - x_A)$$

3) Gauss-Newton Minimization
Blind man down the hill

$$\vec{x}_{i+1} = \vec{x}_i + (K^T S_y^{-1} K + S_A^{-1})^{-1} (K^T S_y^{-1} (\vec{y} - \vec{F}(\vec{x})) + S_A^{-1} (\vec{x}_A - \vec{x}))$$

4) Stop iterating. Assume/hope this is global minimum. Typically χ^2 is ~~true~~ quadratic near here.

5) $\hat{x}$ is the "optimal estimate".

$$\vec{x}_A + \hat{S}\,(K^T S_y^{-1}\vec{y} + S_A^{-1}\vec{x}_A)$$

$$= G\vec{y} + (I-A)\vec{x}_A$$

$$\vec{x}_A + \hat{S}K^T S_y^{-1}(\vec{y} - K\vec{x}_A)$$

$$= \vec{x}_A + G\vec{y} \cancel{A} - A\vec{x}_A$$

$$\hat{x} = (I-A)\vec{x}_A + G\vec{y}$$

Notice this is a direct expression of weighted avg of prior & measurement.

G is a (n,m) matrix.

The "gain matrix"

$$S_A = \begin{bmatrix} 10^2\mu m^2 & 0 \\ 0 & \delta b^2 \end{bmatrix}$$

$$S_y = \begin{bmatrix} 2.5\cdot10^{-5} & 0 \\ 0 & 2.5\cdot10^{-5} \end{bmatrix}$$

$K = \cancel{b}\,(m,n)$: $\begin{matrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{matrix}$

$$K = 10^{-3}\cdot \begin{bmatrix} -1.8 & 4.6 \\ -16.7 & 0 \end{bmatrix} \begin{pmatrix} r_e \\ \tau \end{pmatrix}$$

$y \sim K\vec{x}$

So

$y_1 \sim -1.8r_e + 4.6\tau$

$y_2 \sim -16.7r_e + 0\tau$

No dependence on τ. implies direct constraint on r_e from y_2!

Continuing.

$$\hat{x} \sim (I-A)\vec{x}_A + G\vec{y}$$

$$\begin{pmatrix} R_e \\ \tau \end{pmatrix} \approx \underbrace{\begin{pmatrix} -.003 & -60 \\ 219 & -23 \end{pmatrix}}_{G} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

Re taken mostly from y_2 (basically all)

or

$R_e \sim -60y_2$

$\tau \sim 219y_1 - 23y_2$

What did this buy us?

Key is $\hat{S}$: This is our "posterior error".

$$\hat{S} = \begin{bmatrix} 0.09 & .033 \\ .033 & 1.15 \end{bmatrix} \rightarrow \begin{bmatrix} \sigma_{x} = 0.3 & \rho_{12} = 10.5\% \\ \rho_{12} = 10.5\% & \sigma_{T} = 1.1 \end{bmatrix}$$

Next: ~~Write~~

In Linear case:

$$\hat{x} = A\vec{x}_{true} + (I - A)\vec{x}_{Ap} + G\vec{\varepsilon}$$

where does this come from?

$$\hat{x} = \hat{S}\left(K^T S_y^{-1}\vec{y} + S_A^{-1}\vec{x}_A\right)$$

let $\vec{y} = K\vec{x}_{true} + \vec{\varepsilon}$

then

$$\hat{x} = \underbrace{\hat{S}K^T S_y^{-1} K}_{A=GK}\vec{x}_{true} + \hat{S}S_A^{-1}\vec{x}_A + \underbrace{\hat{S}K^T S_y^{-1}}_{\equiv G}\vec{\varepsilon}$$

$$\hat{x} = A\vec{x}_{true} + \hat{S}S_A^{-1}\vec{x}_A + G\vec{\varepsilon}$$

{only true in strictly linear case!}

$$\hat{S}S_A^{-1} + A \cancel{= \hat{S}S_A^{-1} + \hat{S}S^{-1}}$$

$$\cancel{= \hat{S}S_A^{-1} + \hat{S}(K^T S_y^{-1} K + S_A^{-1})}$$

$$\cancel{= \hat{S}}$$

$$= \hat{S}\left[S_A^{-1} + K^T S_y^{-1} K\right] = I$$

Show in more steps!

so $\hat{S}S_A^{-1} + A = I$

$$\hat{S}S_A^{-1} = I - A$$

or

$$A = I - \hat{S}S_A^{-1}$$

Alternatively

$$\hat{x} = \vec{x}_{true} + \hat{S}S_A^{-1}(\vec{x}_A - \vec{x}_{true}) + G\vec{\varepsilon}$$

- G tells us how the retrieval "inflates" noise.
- Noise on y_1 leads mostly to noise on I.

  Noise on y_2 " " " " " Re.

"degrees of freedom for signal"

- χ_m^2 should be of the order $m-n$.

  $\chi_a^2 \sim n$

  χ^2 of order m. (though really this isn't true!)

  $$d_s = \mathcal{E}\{(\hat{x}-x_a)^T S_a^{-1} (\hat{x}-\vec{x}_a)\}$$

  CAN show this is the <u>trace</u> of A.

Trace:

- sum of its main diagonal $A_{11} + A_{22} + \ldots + A_{nn}$.
- Equal to sum of its eigenvalues.
- Invariant to a basis change.

$d_{noise} = m - d_s$ or $d_s + d_n = m$

$\hat{x}$: best estimate $\quad E\{\vec{x}_{true}\}$?

$\hat{S}$: covariance of best estimate

$E\{(\hat{\vec{x}}-\vec{x}_{true})(\hat{x}-\vec{x}_{true})^T\}$ what 1σ, 2σ Limits Look like.

$P(\vec{x}|\vec{y})$ can typically be represented as a gaussian near $x=\hat{x}$.

$\vec{x}_a, \hat{S}_a$: Prior Knowledge

$\vec{y}$: the measurement

$\vec{F}(\vec{x})$: the forward model. $\quad \vec{y} = \vec{F}(\vec{x}) + \vec{\varepsilon}$

$\vec{\varepsilon}$: inst noise + measurement error.

χ^2 : $P(x|y) = e^{-\frac{1}{2}\chi^2(\vec{x})}$

At the solution

G : the gain matrix $\quad (n,m)$

A : the Avg. Kernel Matrix $\quad (n,n)$

$d_s = tr(A)$ "deg of freedom for signal"

$d_n = m - d_s$ " " " " " noise"

example of G use :

- Suppose inst tells you they recalibrated inst and $\vec{y} \to \vec{y} + \delta\vec{y}$. Can we estimate change in $\hat{x}$ w/o repeating minimization?

Repeat last step in minimization

old $\vec{y}$: $\hat{x} = \vec{x}_i + G(\vec{y}-\vec{F}(\vec{x}_i)) - \hat{S}S_a^{-1}(\vec{x}_i - \vec{x}_a)$

new $\vec{y}$: $\hat{x} = \vec{x}_i + G(\vec{y}+\delta\vec{y}-\vec{F}(\vec{x}_i)) - \hat{S}S_a^{-1}(\vec{x}_i - \vec{x}_a)$

$\boxed{\text{new } \vec{y} = \text{old } \vec{y} + G\,\delta\vec{y}}$ this why its called the "gain matrix"

Error Analysis (for Chris)

$$\vec{y} = f(\vec{x}, \vec{b}) + \vec{\varepsilon}$$

true forward model

or

$$\vec{y} = \vec{F}(\vec{x}, \vec{b}) + \vec{\varepsilon}_m + \vec{\varepsilon}_i$$

FM error

inst noise

$$\hat{x} - \vec{x}_{true} = A\vec{x}_{true} - I\vec{x}_t + (I-A)x_a + \vec{G}\vec{\varepsilon}$$

$$= (A - I)\vec{x}_t + (I-A)\vec{x}_a + G\vec{\varepsilon}$$

$$= \underbrace{(A-I)(\vec{x}_t - \vec{x}_a)}_{\text{"smoothing" error}} + \underbrace{G\vec{\varepsilon}}_{\text{inst noise}} \qquad 3.16$$ missing two terms.

- random.

- it is directly dependence of ~~the~~ $\hat{x}$ on prior.
- depends on $A - I$: (ideal A is I) and how far away prior is from truth.
- Can easily be systematic.

Next suppose

$$\vec{y} = F(\vec{x}, \vec{b}) + \vec{\varepsilon}$$

and $\vec{b} \neq \vec{b}_{true}$

then

$$\vec{y} \approx \vec{F}(\vec{x}, \vec{b}_{true}) + \frac{\partial \vec{F}}{\partial \vec{b}}(\hat{b} - \vec{b}_{true}) + \vec{\varepsilon}$$

we know how this goes! changes $\hat{x}$ by the gain matrix, so

$$\hat{x} - \vec{x}_{true} = (A - I)(\vec{x}_{true} - \vec{x}_a) + \underbrace{G\hat{K}_b(\vec{b}_{true} - \hat{b})}_{\text{errors in unfit parameters}} + G\vec{\varepsilon}$$

- Requires $\hat{K}_b$.

- Examples : Frankenberg Aerosol DOF paper (suggest to milk this)
- L'ecuyer channel selection paper.

How do we include errors due to unfit parameters?

- Will they affect $\hat{x}$ or just $\hat{S}$?

Answer: Typically included in Forward Model error.

$$\vec{y} = \vec{F}(\vec{x}, \hat{b}) + \frac{\partial \vec{F}}{\partial \vec{b}} (\vec{b}_{true} - \hat{b}) + \vec{\varepsilon}$$

$$= \vec{F}(\vec{x}, \hat{b}) + K_b \Delta \vec{b} + \vec{\varepsilon}$$

question : what are the statistics of $\Delta \vec{b}$?

we Assume $\mathcal{E}\{\Delta \vec{b}\} = 0$. {otherwise we'd make a different $\hat{b}$ guess!}

but

$$\mathcal{E}\{\Delta b \, \Delta b^T\} = S_b$$

covariance matrix!

example:

surface Albedo $\sigma = 0.02$

$\sigma^2 = 0.0004$ {seems small but is transformed like K_b!}

How is S_y transformed?

$$\mathcal{E}\{(y - F(x))(y - F(x))^T\} = \mathcal{E}\{\varepsilon \varepsilon^T\} + \mathcal{E}((K_b \Delta \hat{b})(K_b \Delta b)^T\}$$

$$S_y = S_\varepsilon + \mathcal{E}\{K_b \Delta b \, \Delta b^T K_b^T\} \; 2\mathcal{E}\{\vec{\varepsilon} (K_b \Delta \vec{b})^T\}$$

$$= S_\varepsilon + K_b \mathcal{E}\{\Delta b \, \Delta b^T\} K_b^T + \alpha \, \mathcal{E}\underbrace{\{\vec{\varepsilon} \Delta \vec{b}^T\}}_{\text{assume uncorrelated}}$$

$$\boxed{S_y = S_\varepsilon + K_b S_b K_b^T}$$

increases our "meas. error covariance matrix"

~~ZH~~

Note: B/c $\chi^2 = \chi^2_m + \chi^2_A$

AND $\chi^2_m = (y-F(x))^T S_y^{-1} (y-F(x))$

this will change the χ^2.

It will therefore move the minimum more towards the prior. Depending on x_0, it will even change channel's relative weights. This is <u>very</u> powerful!

Albedo error in Re/T.

Next Topic: Opposite of inverse model: the ~~*~~ Forward model problem?

suppose general function

$\vec{z} = h(\hat{\vec{x}})$ A "Derived Quantity"

Very common to see this.

$(c, d)\begin{pmatrix} S_{11}c + S_{12}d \\ S_{12}c + S_{22}d \end{pmatrix} = \cancel{S_{11}c^2 + S_{12}cd + S_{12}cd + S_{22}d^2 + S_{12}c^2 +}$

$S_{11}c^2 + S_{12}cd + S_{12}cd + S_{22}d^2$

$= S_{11}c^2 + S_{22}d^2 + 2S_{12}cd$

Example

$\hat{x}_{CO_2} = \vec{h}^T \hat{u}_{CO_2}$

"Weighting Function"

Atmos. profile

RWP → RR. etc.

If errors in $\hat{x}$ are relatively small, Again Assume $h(\vec{x})$ is linear in neighborhood encompassing $\hat{x}$ errors.

Then Find $K_z = \dfrac{\partial \vec{h}}{\partial \vec{x}}$ (Matrix or vector)

(z scalar then K = row vector)

$\hat{S}_z = K_z \hat{S} K_z^T$

ex

$z = cx_1 + dx_2$ then

$S_z = (c, d)\begin{pmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{pmatrix}\begin{pmatrix} c \\ d \end{pmatrix} = \hat{S}_{11}c^2 + \hat{S}_{22}d^2 + 2\hat{S}_{12}cd$