Joint Inversions by Forcing Separate Inversions to Have Equal Model Parameters

Bill Menke, December 16, 2018

This idea arose from a conversation that I had with Zhongmin Tao and Aibing Li at the AGU.

Consider an inversion that involves M model parameters $\mathbf{m}_\mathrm{b}$ that are constrained by two sets of data equations:

$$\mathbf{g}_v(\mathbf{m}_\mathrm{b}) = \mathbf{d}_v \quad \text{with} \quad \sigma_{d_v}^2$$

$$\mathbf{g}_h(\mathbf{m}_\mathrm{b}) = \mathbf{d}_h \quad \text{with} \quad \sigma_{d_h}^2$$

A common problem with such an inversion is that no reasonable $\mathbf{m}_b$ may satisfy both data equations, because: (a) the theories $\mathbf{g}_v(\mathbf{m})$ and $\mathbf{g}_h(\mathbf{m})$ may not be quite right; or (b) the noise in the data $\mathbf{d}_v$ and $\mathbf{d}_h$ might be mischaracterized. In such as case it may be advantageous to consider each data equation to depend on a *different* set of model parameters, labeled v and h:

$$\mathbf{g}_v(\mathbf{m}_v) = \mathbf{d}_v \quad \text{with} \quad \sigma_{d_v}^2$$

$$\mathbf{g}_h(\mathbf{m}_h) = \mathbf{d}_h \quad \text{with} \quad \sigma_{d_h}^2$$

The overall vector model parameters $\mathbf{m} = [\mathbf{m}_v, \mathbf{m}_h]^\mathrm{T}$ is of length $2M$. The notion that $\mathbf{m}_v = \mathbf{m}_h$ can then be implemented as prior information with variance σ_c^2. The quality of the assumption $\mathbf{m}_v = \mathbf{m}_h$ can then be evaluated with a standard squeezing test.

The linearized generalized least squares formulation is derived as follows:

(A) The $N_v + N_h$ data equations are linearized about trial solutions $\mathbf{m}_v^0$ and $\mathbf{m}_h^0$ with unknown corrections $\Delta\mathbf{m}_v$ and $\Delta\mathbf{m}_h$ via matrices $[\mathbf{G}_v]_{ij} = \partial g_{vi}/\partial m_{vj}\big|_{\mathbf{m}_v^0}$ and $[\mathbf{G}_h]_{ij} = \partial g_{hi}/\partial m_{hj}\big|_{\mathbf{m}_h^0}$ of partial derivatives:

$$\mathbf{g}_v(\mathbf{m}_v^0) + \mathbf{G}_v\Delta\mathbf{m}_v = \mathbf{d}_v \quad \text{with} \quad \sigma_{d_v}^2$$

$$\mathbf{g}_h\left(\mathbf{m}_h^0\right) + \mathbf{G}_h\Delta\mathbf{m}_h = \mathbf{d}_h \quad \text{with} \quad \sigma_{d_h}^2$$

(B) The $2M$ regularization equations for closeness to a base (prior) model $\mathbf{m}^\mathrm{b}$ are:

$$\mathbf{m}_v^0 + \Delta\mathbf{m}_v = \mathbf{m}_v^b \quad \text{with} \quad \sigma_{b_v}^2$$

$$\mathbf{m}_h^0 + \Delta\mathbf{m}_h = \mathbf{m}_h^b \quad \text{with} \quad \sigma_{b_h}^2$$

(C) The $2K$ regularization equations for the smoothness of the solution, with first (or second) derivative operator $\mathbf{D}$, are:

$$\mathbf{D}_v(\mathbf{m}_v^0 + \Delta\mathbf{m}_v) = 0 \quad \text{with} \quad \sigma_{s_v}^2$$

$$\mathbf{D}_h(\mathbf{m}_h^0 + \Delta\mathbf{m}_h) = 0 \quad \text{with} \quad \sigma_{s_h}^2$$

(D) The M regularization equations for the v and h models being close to one another are:

$$(\mathbf{m}_v^0 + \Delta\mathbf{m}_v) - (\mathbf{m}_h^0 + \Delta\mathbf{m}_h) = 0 \quad \text{with} \quad \sigma_c^2$$

Thus, the generalized least squares equations are:

$$\mathbf{F}\begin{bmatrix}\Delta\mathbf{m}_v \\ \Delta\mathbf{m}_h\end{bmatrix} = \begin{bmatrix} \sigma_{d_v}^{-2}\mathbf{G}_v & 0 \\ 0 & \sigma_{d_h}^{-2}\mathbf{G}_h \\ \sigma_{b_v}^{-2}\mathbf{I} & 0 \\ 0 & \sigma_{b_h}^{-2}\mathbf{I} \\ \sigma_{b_v}^{-2}\mathbf{D}_v & 0 \\ 0 & \sigma_{b_h}^{-2}\mathbf{D}_h \\ \sigma_c^{-2}\mathbf{I} & -\sigma_c^{-2}\mathbf{I} \end{bmatrix}\begin{bmatrix}\Delta\mathbf{m}_v \\ \Delta\mathbf{m}_h\end{bmatrix} = \begin{bmatrix} \sigma_{d_v}^{-2}(\mathbf{d}_v - \mathbf{g}_v(\mathbf{m}_v^0)) \\ \sigma_{d_h}^{-2}\left(\mathbf{d}_h - \mathbf{g}_h(\mathbf{m}_h^0)\right) \\ \sigma_{b_v}^{-2}(\mathbf{m}_v^b - \mathbf{m}_v^0) \\ \sigma_{b_h}^{-2}(\mathbf{m}_h^b - \mathbf{m}_h^0) \\ -\mathbf{D}_v\mathbf{m}_v^0 \\ -\mathbf{D}_h\mathbf{m}_h^0 \\ \sigma_c^{-2}(-\mathbf{m}_v^0 + \mathbf{m}_h^0) \end{bmatrix} = \mathbf{f}$$

and the generalized least squares solution is:

$$\begin{bmatrix}\Delta\mathbf{m}_v \\ \Delta\mathbf{m}_h\end{bmatrix} = [\mathbf{F}^{\mathrm{T}}\mathbf{F}]^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{f}$$

This solution must be iterated, $\mathbf{m}_v^0 \rightarrow \mathbf{m}_v^0 + \Delta\mathbf{m}_v$ with $\mathbf{g}_v(\mathbf{m}_v^0)$, $\mathbf{g}_h(\mathbf{m}_h^0)$, $\mathbf{G}_v$ and $\mathbf{G}_h$ being recomputed at the start of each iteration.

The variance σ_c^2 controls the degree to which the v and h solutions are forced to equal one another. They become more and more similar as $\sigma_c^2 \rightarrow 0$. The proposition that two data equations are incompatible, and require the different model parameters, can be tested by generating a series of solutions, each for a different σ_c and examining the behavior of the prediction error

$$E = \sigma_{d_v}^{-2}\|\mathbf{e}_v\|_2^2 + \sigma_{d_v}^{-2}\|\mathbf{e}_v\|_2^2 \quad \text{with} \quad \mathbf{e}_v = \mathbf{d}_v - \mathbf{g}_v(\mathbf{m}_v) \;\text{ and }\; \mathbf{e}_h = \mathbf{d}_h - \mathbf{g}_h(\mathbf{m}_h)$$

now viewed as $E(\sigma_c)$. Is the error significantly smaller for large σ_c than for small σ_c?

The overall problem has $2M$ unknowns and $N = (N_v + N_h + 3M + 2K)$ constraints, so the total number of degrees of freedom are $\nu = N - 2M$. The data equations have approximately $\nu_E = \nu(N_v + N_h)/N$ degrees of freedom. The prediction error E is chi-squared distributed with approximately ν_E degrees of freedom, and has mean ν_E and variance $2\nu_E$. An F-test can be used to test against the Null Hypothesis that the difference between $E(\sigma_c^{small})$ and $E(\sigma_c^{large})$ is due to random variation (as contrasted to the data requiring two different models). Only when the Null Hypothesis can be rejected to greater than 95% confidence can the data be said to require two different models.