Backus-Gilbert Resolution Computed from Ensembles

Bill Menke, October 31, 2018

Suppose the model parameter vector $\mathbf{m}$ (of length $M$) and its covariance $\mathbf{C}_m$ are estimated from an ensemble of solutions. Writing the eigenvalue expansion of the covariance matrix as $\mathbf{C}_m = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}$, we consider the linear transformation $\mathbf{d} = \mathbf{Gm}$ with $\mathbf{G} \equiv \mathbf{V}^{\mathrm{T}}$. The covariance matrix $\mathbf{C}_d$ of $\mathbf{d}$ is:

$$\mathbf{C}_d = \mathbf{G}\mathbf{C}_m\mathbf{G}^{\mathrm{T}} = \mathbf{V}^{\mathrm{T}}\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}} = \boldsymbol{\Lambda}$$

That is, the $\mathbf{d}$'s are uncorrelated with covariance $\boldsymbol{\Lambda}$. Now suppose that we discard all eigenvectors $\mathbf{v}^{(i)}$ for which $\Lambda_{ii} > \sigma^2$, where $\sigma$ is a threshold variance, and create an eigenvalue matrix $\mathbf{V}_P^{\mathrm{T}}$ of the remaining, say $P$, eigenvectors. The linear combinations $\mathbf{d}_P = \mathbf{G}_P\mathbf{m}$ with $\mathbf{G}_P \equiv \mathbf{V}_P^{\mathrm{T}}$ all have variance above the threshold.

Now suppose that we define an average $a = \mathbf{a}^T\mathbf{d}_P$, with the averaging vector $\mathbf{a}$ chosen to have unit length; that is $\mathbf{a}^T\mathbf{a} = 1$. Then the variance $\sigma_a^2$ of $a$ can be no larger than the threshold $\sigma^2$:

$$\sigma_a^2 = \mathbf{a}^T\mathbf{C}_g\mathbf{a} = \sum_{i=1}^{P} a_i^2\Lambda_{ii} \le \max_j \Lambda_{jj} \sum_{i=1}^{P} a_i^2 = \sigma^2$$

The Backus-Gilbert (1967) method allows one to determine an averaging vector $\mathbf{a}^{(k)}$ associated with a linear combination of model that is "most-localized" around a given model parameter $m_k$. The goal is to make the resolution vector $\mathbf{r}^{(k)}$ in:

$$a^{(k)} = \left[\mathbf{a}^{(k)}\right]^{\mathrm{T}} \mathbf{g}_P = \left\{\left[\mathbf{a}^{(k)}\right]^{\mathrm{T}}\mathbf{G}_P\right\}\mathbf{m} \equiv \left[\mathbf{r}^{(k)}\right]^{\mathrm{T}}\mathbf{m}$$

as close to a "unit spike" as possible, so that $a^{(k)}$ can be interpreted as the average of just a few model parameters, centered about $m_k$. The solution minimizes a measure $J_k$ of the spread of $\mathbf{r}^{(k)}$:

$$\text{minimize} \quad J_k = \sum_i w(l,k)\, r_l^{(k)} r_l^{(k)} \quad \text{where} \quad \mathbf{r}^{(k)} \equiv \left[\mathbf{a}^{(k)}\right]^T\mathbf{G}_P \quad \text{and} \quad \sum_{\boldsymbol{i}} a_i^{(k)} = 1$$

Here $w(l,k)$ is a penalty function that quantifies the distance between model parameters $m_l$ and $m_k$ and that satisfied $w(l,l) = 0$ and $w(l,k) > 0$ for $l \neq k$. Backus and Gilbert (1967) showed that the solution to this problem is:

$$a_l^{(k)} = \frac{\sum_i u_i \left[\left\{\mathbf{S}^{(k)}\right\}^{-1}\right]_{il}}{\sum_i \sum_i u_i \left[\left\{\mathbf{S}^{(k)}\right\}^{-1}\right]_{ij} u_j} \quad \text{with} \quad u_j = \sum_i [\mathbf{G}_P]_{jk}$$

$$\text{and} \quad \left[\mathbf{S}^{(k)}\right]_{ij} = \sum_l w(l,k)\, [\mathbf{G}_P]_{il}[\mathbf{G}_P]_{jl}$$

The function $J_k(\sigma)$ defines how spread of resolution and variance trade off. Note that the Backus-Gilbert constraint on the size of the elements of $\mathbf{a}^{(k)}$ are different than the one discussed in the context of variance, so the upper limit on variance is, at best, only approximate.

Backus, G. and F. Gilbert (1967). Numerical application of a formalism for geophysical inverse problems, Geophys. J. R. Astr. Soc. 13, 247-276.