Properties of Generalized Least Squares

Bill Menke, February 23, 2014

1. Definition of the inverse problem. We are given a forward problem $\mathbf{d} = \mathbf{G}\mathbf{m}$, where a known $N \times M$ data kernel $\mathbf{G}$ links model parameters $\mathbf{m}$ to data $\mathbf{d}$ The generalized inverse $\mathbf{G}^{-g}$ turns this equation around, linking data $\mathbf{d}$ to model parameters $\mathbf{m}$ through $\mathbf{m} = \mathbf{G}^{-g}\mathbf{d}$. The generalized inverse is an $M \times N$ matrix that is a function of the data kernel. However, at this point, neither the method by which it has been obtained nor its functional form has been specified. We note for future reference that the simple least squares solution, obtained by minimizing the prediction error

$$\Phi_{SLS} = [\mathbf{d} - \mathbf{G}\mathbf{m}]^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}[\mathbf{d} - \mathbf{G}\mathbf{m}]$$

is

$$\mathbf{m} = [\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{d}$$

has so has generalized inverse $\mathbf{G}^{-g} = [\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}$. Here $\mathbf{C}_{\mathrm{d}}^{-1}$ is the prior covariance of the data. One of the limitations of simple least squares is that this generalized inverse exists only when the observations are sufficient to uniquely specify a solution. Otherwise the matrix $[\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{G}]^{\mathbf{-1}}$ does not exist.

2. Definition of model resolution. The model resolution matrix $\mathbf{R}^{\mathrm{G}} = \mathbf{G}^{-g}\mathbf{G}$ can be obtained by using the fact that an asserted (or "true") model predicts the data, $\mathbf{d}^{\mathrm{pre}} = \mathbf{G}\mathbf{m}^{\mathrm{true}}$ and that those data can then be inverted for estimated model parameters,$\mathbf{m}^{\mathrm{est}} = \mathbf{G}^{-g}\mathbf{d}^{\mathrm{pre}}$**:**

$$\mathbf{m}^{\mathrm{est}} = \mathbf{G}^{-g}\mathbf{G}\mathbf{m}^{\mathrm{true}} = \mathbf{R}^{\mathrm{G}}\,\mathbf{m}^{\mathrm{true}}$$

The resolution matrix $\mathbf{R}^{\mathrm{G}}$ indicates that the estimated model parameters only equal the true model parameters in the special case where $\mathbf{R}^{\mathrm{G}} = \mathbf{I}$. In typical cases, the estimated model parameters are linear combinations ("weighted averages") of the true model parameters. In general, $\mathbf{R}^{\mathrm{G}}$ has no special symmetry; it is neither symmetric nor anti-symmetric. We note for future reference that the simple least squares solution, when it exists, has perfect resolution

$$\mathbf{R}^{\mathrm{G}} = \mathbf{G}^{-g}\mathbf{G} = [\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{G} = \mathbf{I}$$

3. Meaning of the k-th row of the resolution matrix. The k-th estimated model parameter satisfies:

$$m_k^{\mathrm{est}} = \sum_{\mathrm{i}} R_{ki}^{G} m_i^{\mathrm{true}}$$

and so can be interpreted as being equal to a linear combination of the true model parameters, where the coefficients are given by the elements of the k-th row of the resolution matrix.

Colloquially, we might speak of the estimated model parameters as *weighted averages* of the true model parameters. However, strictly speaking, they are only thru weighted averages when the elements of the row sum to unity,

$$\sum_{\mathrm{i}} R_{ki}^{G} = [1]_k$$

which is, in general, not the case.

4. **Meaning of the k-th column of the resolution matrix.** The k-th column of the resolution matrix specifies how each of the estimated model parameters is influenced by the k-th true model parameter. This can be seen by setting $\mathbf{m}^{\text{true}} = \mathbf{s}^{(k)}$ with $s_i^{(k)} = \delta_{ik}$; that is, the all the true model parameters are zero except the k-th, which is unity. Denoting the set of estimated model parameters associated with $\mathbf{s}^{(k)}$ as $\mathbf{m}^{\text{est(k)}}$, we have:

$$\mathbf{m}^{\text{est(k)}} = \mathbf{R}^{\mathrm{G}}\,\mathbf{s}^{(k)} \quad \text{or} \quad m_j^{\text{est(k)}} = \sum_{\mathrm{i}} R_{ji}^{G}\delta_{ik} = R_{jk}^{G}$$

Thus the k-th column of the resolution matrix is a "point-spread function"; that is, a single true model parameter spreads out into many estimated model parameters.

5. The generalized least squares solution. Generalized least squares supplements the observations with prior information, represented by the linear equation $\mathbf{h} = \mathbf{Hm}$. Both the matrix $\mathbf{H}$ and vector $\mathbf{h}$ are assumed to be known. The solution is obtained by minimizing the sum of the prediction error and the error in prior information, $\varepsilon^2[\mathbf{h} - \mathbf{Hm}]^{\mathrm{T}}[\mathbf{h} - \mathbf{Hm}]$:

$$\Phi_{GLS} = [\mathbf{d} - \mathbf{Gm}]^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}[\mathbf{d} - \mathbf{Gm}] + \varepsilon^2[\mathbf{h} - \mathbf{Hm}]^{\mathrm{T}}[\mathbf{h} - \mathbf{Hm}]$$

The parameter ε^2 is a known constant that specifies the strength of the prior information relative to the data. It can be interpreted as the reciprocal of the variance of the prior information. By defining

$$\boldsymbol{F} = \begin{bmatrix} \mathbf{C}_{\mathrm{d}}^{-1/2}\mathbf{G} \\ \varepsilon\mathbf{H} \end{bmatrix} \quad \text{and} \quad \mathbf{f} = \begin{bmatrix} \mathbf{C}_{\mathrm{d}}^{-1/2}\mathbf{d} \\ \varepsilon\mathbf{h} \end{bmatrix}$$

we can write $\Phi = [\mathbf{f} - \mathbf{Fm}]^{\mathrm{T}}\mathbf{C}_{\mathrm{f}}^{-1}[\mathbf{f} - \mathbf{Fm}]$, with $\mathbf{C}_{\mathrm{f}}^{-1} = \mathbf{I}$, which is in the form of a simple least squares minimization problem. The solution is given by the simple least squares formula:

$$\mathbf{m}^{\text{est}} = [\mathbf{F}^{\mathrm{T}}\mathbf{F}]^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{f}^{\text{obs}} \quad \textbf{with} \quad \mathbf{f}^{\text{obs}} = \begin{bmatrix} \mathbf{C}_{\mathrm{d}}^{-1/2}\mathbf{d}^{\text{obs}} \\ \varepsilon\mathbf{h}^{\text{pri}} \end{bmatrix}$$

or

$$\mathbf{m}^{\text{est}} = \mathbf{G}^{-g}\mathbf{d}^{\text{obs}} + \mathbf{H}^{-g}\mathbf{h}^{\text{pri}}$$

$$\text{with } \mathbf{G}^{-g} = \mathbf{A}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1} \quad \text{and} \quad \mathbf{H}^{-g} = \varepsilon^2\mathbf{A}^{-1}\mathbf{H}^{\mathrm{T}} \quad \text{and} \quad \mathbf{A} = \mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{G} + \varepsilon^2\mathbf{H}^{\mathrm{T}}\mathbf{H}$$

Here $\mathbf{d}^{\text{obs}}$ denotes the observed values of the data and $\mathbf{h}^{\text{pri}}$ the prior values of the information (that is, the values that are asserted). The combined vector $\mathbf{f}^{\text{obs}}$ includes both observations and prior information, but we simplify its superscript to "obs". The presumption in generalized least squares is that the addition of prior information to the problem is sufficient to eliminate any non-uniqueness that would have been present had only observations been used. Thus, the inverse of $\mathbf{A}$ is presumed to exist. Note that since $\mathbf{A}$ is symmetric, its inverse $\mathbf{A}^{-1}$ will also be symmetric.

6. Resolution of generalized least squares. Generalized least squares does not distinguish the weighted data equation $\mathbf{C}_{\mathrm{d}}^{-1/2}\{\mathbf{d} = \mathbf{Gm}\}$ from the weighted prior information equation $\varepsilon\{\mathbf{Hm} = \mathbf{h}\}$; the latter is simply appended to the bottom of the former to create the combined equation $\mathbf{Fm} = \mathbf{f}$. Consequently, in analogy to the simple least squares case, we can define a generalized inverse $\mathbf{F}^{-g}$ and a resolution matrix $\mathbf{R}^{\mathrm{F}}$ as:

$$\mathbf{f}^{\text{pre}} = \mathbf{F}\mathbf{m}^{\text{true}} \quad \text{and} \quad \mathbf{m}^{\text{est}} = \mathbf{F}^{-g}\mathbf{f}^{\text{pre}} \quad \text{and} \quad \mathbf{F}^{-g} = [\mathbf{F}^{\mathrm{T}}\mathbf{F}]^{-1}\mathbf{F}^{\mathrm{T}}$$

$$\text{so} \quad \mathbf{m}^{\text{est}} = \mathbf{R}^{\mathrm{F}}\mathbf{m}^{\text{true}} \quad \text{with} \quad \mathbf{R}^{\mathrm{F}} = \mathbf{F}^{-g}\mathbf{F}$$

However, when defined in this way, the resolution of generalized least squares is perfect, since

$$\mathbf{R}^{\mathrm{F}} = \mathbf{F}^{-g}\mathbf{F} = [\mathbf{F}^{\mathrm{T}}\mathbf{F}]^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{F} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

In general, the estimated model parameters depend upon both $\mathbf{d}$ and $\mathbf{h}$; that is $\mathbf{m} = \mathbf{G}^{-g}\mathbf{d} + \mathbf{H}^{-g}\mathbf{h}$. Consider, however, the special case of $\mathbf{h} = 0$. (This case commonly arises in practice, e.g. for the prior information of smoothness). The estimated model parameters depend only upon $\mathbf{d}$; that is $\mathbf{m} = \mathbf{G}^{-g}\mathbf{d} + 0$. We can use the forward equation to predict data associated with true model parameters, $\mathbf{d}^{\text{pre}} = \mathbf{G}\mathbf{m}^{\text{true}}$, and then invert these predictions back to estimated model parameters, $\mathbf{m}^{\text{est}} = \mathbf{G}^{-g}\mathbf{d}^{\text{pre}} + 0$. Hence, we obtain the usual formula for resolution:

$$\mathbf{m}^{\text{est}} = \mathbf{R}^{\mathrm{G}}\mathbf{m}^{\text{true}} \quad \text{with} \quad \mathbf{R}^{\mathrm{G}} = \mathbf{G}^{-g}\mathbf{G}$$

Superficially, we seemed to have achieved contradictory results, as the two resolution matrices have radically different properties:

$$\mathbf{R}^{\mathrm{F}} = \mathbf{I} \quad \text{and} \quad \mathbf{R}^{\mathrm{G}} \neq \mathbf{I}$$

However, one step in the derivations is critically different. In the case of $\mathbf{R}^{\mathrm{G}}$, we asserted that $\mathbf{h}^{\text{pre}} = 0$, even though an arbitrary $\mathbf{m}^{\text{true}}$ predicts $\mathbf{h}^{\text{pre}} = \mathbf{H}\mathbf{m}^{\text{true}} \neq 0$. In the case of $\mathbf{R}^{\mathrm{F}}$, we made no such assertion; the $\mathbf{h}^{\text{pre}}$ imbedded in $\mathbf{f}^{\text{pre}}$ arises from $\mathbf{h}^{\text{pre}} = \mathbf{H}\mathbf{m}^{\text{true}}$ and is not equal to zero.

That the $\mathbf{R}^{\mathrm{G}}$ version is the more appropriate choice can be understood from the following scenario: Suppose that the model $\mathbf{m}$ represents a discrete version of a continuous function $m(x)$ and that that one in trying to find an $\mathbf{m}^{\mathrm{est}}$ that approximately satisfies $\mathbf{Gm} = \mathbf{d}^{\mathrm{obs}}$ but is smooth. Smoothness is the opposite of roughness, and the roughness of a function can be quantified by the r.m.s. value of its second derivative. Thus, we take $\mathbf{H}$ to be the second derivative operator (i.e. with rows like $[0 \quad \cdots \quad 0 \quad 1 \quad -2 \quad 1 \quad 0 \quad \cdots \quad 0]$) and $\mathbf{h}^{\mathrm{pri}} = 0$, which leads to the minimization of $[\mathbf{Hm}]^{\mathrm{T}}[\mathbf{Hm}]$, a quantity proportional to the r.m.s. average of the second derivative. Now suppose that the true solution is the spike $\mathbf{m}^{\mathrm{true}} = \mathbf{s}^{(\mathrm{k})}$ (that is, zero except for the k-th element, which is unity). We want to know how this spike spreads out during the inversion process, presuming that an experiment produced the data $\mathbf{d}^{\mathrm{pre}} = \mathbf{Gs}^{(\mathrm{k})}$ that this model predicts. What values should one use for $\mathbf{h}$ in such an inversion? The model predicts $\mathbf{h}^{\mathrm{pre}} = \mathbf{Hs}^{(\mathrm{k})}$, but these are the actual values of the second derivative. To use them in the inversion would be to assert that second derivatives are known, which is stronger prior information than merely asserting that their r.m.s. average is small. One should, therefore, use $\mathbf{h}^{\mathrm{pri}} = 0$, which leads to a solution that is a column of $\mathbf{R}^{\mathrm{G}}$, not $\mathbf{R}^{\mathrm{F}}$.

So far, our discussion has been limited to the special case of $\mathbf{h}^{\mathrm{pri}} = 0$. We now relax that condition, but require that the prior information is complete, in the sense that $[\mathbf{H}^{\mathrm{T}}\mathbf{H}]^{-1}$ exists. Then we can solve the prior information equation by least squares, and determine a set of model parameters $\mathbf{m}^{\mathrm{H}}$:

$$\mathbf{m}^{\mathrm{H}} = [\mathbf{H}^{\mathrm{T}}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{h}^{\mathrm{pri}}$$

We now use $\mathbf{m}^{\mathrm{H}}$ as a reference model, defining the deviation of a given model from it as $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}^{\mathrm{H}}$. The generalized least squares solution can be rewritten in terms of this deviation:

$$\Delta\mathbf{m}^{\mathrm{est}} = \mathbf{m}^{\mathbf{est}} - \mathbf{m}^{\mathrm{H}} = \mathbf{A}^{-1}\left(\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{d}^{\mathrm{obs}} + \varepsilon^2\mathbf{H}^{\mathrm{T}}\mathbf{h}^{\mathrm{pri}}\right) - \mathbf{m}^{\mathrm{H}}$$

$$= \mathbf{A}^{-1}\left(\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{d}^{\mathrm{obs}} + \varepsilon^2\mathbf{H}^{\mathrm{T}}\mathbf{h}^{\mathrm{pri}}\right) - \mathbf{A}^{-1}\mathbf{A}\mathbf{m}^{\mathrm{H}}$$

$$= \mathbf{A}^{-1}\left(\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{d}^{\mathrm{obs}} + \varepsilon^2\mathbf{H}^{\mathrm{T}}\mathbf{h}^{\mathrm{pri}} - \mathbf{A}\mathbf{m}^{\mathrm{H}}\right)$$

$$= \mathbf{A}^{-1}\left(\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{d}^{\mathrm{obs}} + \varepsilon^2\mathbf{H}^{\mathrm{T}}\mathbf{h}^{\mathrm{pri}} - \mathbf{G}^{\mathbf{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{G}\mathbf{m}^{\mathrm{H}} - \varepsilon^2\mathbf{H}^{\mathbf{T}}\mathbf{H}[\mathbf{H}^{\mathrm{T}}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{h}^{\mathrm{pri}}\right)$$

$$= \mathbf{G}^{-\mathbf{g}}\left(\mathbf{d}^{\mathrm{obs}} - \mathbf{G}\mathbf{m}^{\mathrm{H}}\right)$$

Thus, the deviation of the model from $\mathbf{m}^{\mathrm{H}}$ depends only on the deviation of the data from that predicted by $\mathbf{m}^{\mathrm{H}}$:

$$\Delta\mathbf{m} = \mathbf{G}^{-\mathbf{g}}\Delta\mathbf{d} \quad \text{with} \quad \Delta\mathbf{m} = \mathbf{m} - \mathbf{m}^{\mathrm{H}} \quad \text{and} \quad \Delta\mathbf{d} = (\mathbf{d} - \mathbf{G}\mathbf{m}^{\mathrm{H}})$$

and furthermore

$$\mathbf{G}\Delta\mathbf{m} = \Delta\mathbf{d} \quad \text{since} \quad \mathbf{G}\Delta\mathbf{m} = \mathbf{G}(\mathbf{m} - \mathbf{m}^{\mathrm{H}}) = \mathbf{Gm} - \mathbf{Gm}^{\mathrm{H}} = \mathbf{d} - \mathbf{Gm}^{\mathrm{H}} = \Delta\mathbf{d}$$

Once again, we can combine $\Delta\mathbf{d}^{\mathrm{pre}} = \mathbf{G}\Delta\mathbf{m}^{\mathrm{true}}$ with $\Delta\mathbf{m}^{\mathrm{est}} = \mathbf{G}^{-\mathbf{g}}\Delta\mathbf{d}^{\mathrm{pre}}$ into the usual statement about resolution,

$$\Delta\mathbf{m}^{\mathrm{est}} = \mathbf{R}^{\mathrm{G}}\Delta\mathbf{m}^{\mathrm{true}} \quad \text{with} \quad \mathbf{R}^{\mathrm{G}} = \mathbf{G}^{-\mathbf{g}}\mathbf{G}$$

In this case, too, $\mathbf{R}^{\mathrm{G}}$ is the correct choice for quantifying resolution. However, the quantity being resolved is the deviation of the model from a reference model, and not the model itself.

The general case can be approached by adding additional – but very weak - prior information, enough that $\mathbf{m}^{\mathrm{H}}$ is uniquely specified. For instance, the choice:

$$\varepsilon\{\mathbf{Hm} = \mathbf{h}\} \rightarrow \varepsilon\left\{\begin{bmatrix}\mathbf{H} \\ (\varepsilon'/\varepsilon)\mathbf{I}\end{bmatrix}\mathbf{m} = \varepsilon\begin{bmatrix}\mathbf{h} \\ 0\end{bmatrix}\right\}$$

with $\varepsilon'/\varepsilon \ll 1$ forces any linear combinations of model parameters that is not resolved by $\mathbf{Hm} = \mathbf{h}$ to zero while having negligible effect on the others. It leads to a "damped least squared" estimate:

$$\mathbf{m}^{\mathrm{H}} = [\mathbf{H}^{\mathrm{T}}\mathbf{H} + (\varepsilon'/\varepsilon)^2\mathbf{I}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{h}^{\mathrm{pri}}$$

This modification is limited to the definitions of $\mathbf{H}$ and $\mathbf{h}$ and so does not affect other aspects of the derivation, implying that $\mathbf{R}^{\mathrm{G}}$ is the correct choice in the general case, too.

7. Calculating the generalized least squares solution. In practice, the quantity $[\mathbf{F}^{\mathrm{T}}\mathbf{F}]^{-1}$ is not needed when computing an estimate of the model parameters from data; instead, once solves the linear system:

$$[\mathbf{F}^{\mathrm{T}}\mathbf{F}]\,\mathbf{m}^{\mathrm{est}} = \mathbf{F}^{\mathbf{T}}\mathbf{f}^{\mathrm{obs}}$$

Furthermore, when an iterative linear equation solver such as biconjugate gradients is used, the matrices $\mathbf{G}^{\mathbf{T}}\mathbf{C}_{\mathrm{d}}^{-1/2}\mathbf{G}$ and $\mathbf{H}^{\mathbf{T}}\mathbf{H}$ need never to be explicitly calculated, since the solver can be configured to use $\mathbf{G}$, $\mathbf{C}_{\mathrm{d}}^{-1/2}$ and $\mathbf{H}$ directly. This technique can lead to substantial efficiencies in speed and memory requirements, especially when the matrices are very large but sparse.

8. Calculating the k-th row or column of $\mathbf{A}^{-1}$. We will first discuss how to calculate the k-th row of the damped least squares generalized inverse $\mathbf{G}^{-g} = \mathbf{A}^{-1}\mathbf{G}^{\mathbf{T}}$ without having to compute the others. Note that the equation

$$\mathbf{A}\,\mathbf{A}^{-1} = \mathbf{I} \quad \text{or} \quad \sum_k A_{ik}A_{kj}^{-1} = \delta_{ij}$$

can be read as a sequence of vector equations:

$$\mathbf{A}\,\mathbf{v}^{(\mathrm{j})} = \mathbf{s}^{(\mathrm{j})} \quad \text{with} \quad [\mathbf{v}^{(\mathrm{j})}]_{\mathbf{i}} = [\mathbf{A}^{-1}]_{\mathbf{ij}} \quad \text{and} \quad [\mathbf{s}^{(\mathrm{j})}]_{\mathbf{i}} = \delta_{ij}$$

That is, $\mathbf{v}^{(j)}$ is the j-th column of $\mathbf{A^{-1}}$ and $\mathbf{s}^{(j)}$ is the corresponding column of the identity matrix. Hence we can solve for the j-th column of $\mathbf{A^{-1}}$ by solving the system $\mathbf{A}\,\mathbf{v}^{(j)} = \mathbf{s}^{(j)}$. Furthermore, when an a iterative solver such as biconjugate gradients is used to solve this system, the matrices $\mathbf{G^T G}$ and $\mathbf{H^T H}$ need never be explicitly calculated; the solver uses $\mathbf{G}$ and $\mathbf{H}$ directly. Finally, note that since $\mathbf{A^{-1}}$ is symmetric, its j-th row is $\mathbf{v}^{(j)\mathrm{T}}$.

9. Calculating the k-th row of the generalized inverse. Now notice that:

$$\mathbf{G}^{-g} = \mathbf{A^{-1}}\,\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1} \quad \text{or} \quad [\mathbf{G}^{-g}]_{kj} = \sum_i A_{ki}^{-1}[\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}]_{ij}$$

Hence, the k-th row of the generalized inverse $\mathbf{G}^{-\mathrm{g}}$ it's the k-th row of $\mathbf{A^{-1}}$ dotted into $\left[\mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\right]$.

10. Calculating the k-th row of the resolution matrix. The resolution matrix is formed from the generalized inverse and data kernel through

$$\mathbf{R} = \mathbf{G}^{-g}\mathbf{G} \quad \textbf{or} \quad R_{kj} = \sum_{\mathrm{i}} {G^{-g}}_{ki} G_{ij}$$

Thus, the k-th row of the resolution matrix is the k-th row of the generalized inverse dotted into the data kernel. We can construct the k-th row of the resolution matrix after having calculated the k-th row of the generalized inverse.

10. Calculating the k-th column of the resolution matrix. The Let us define the k-th column of the resolution matrix as the vector $\mathbf{r}^{(\mathrm{k})}$; that is:

$$r_i^{(k)} = R_{ik}$$

Then notice that the definition $\mathbf{R} = \mathbf{G}^{-g}\mathbf{G}$ can be written as

$$\mathbf{R} = \mathbf{G}^{-g}\mathbf{G}\,\mathbf{I} \quad \text{or} \quad r_i^{(k)} = R_{ik} = \sum_p G_{ip}^{-g} \sum_q G_{pq}\delta_{qk} = \sum_p G_{pq}^{-g}\, d_q^{(k)}$$

$$\text{where } d_q^{(k)} = \sum_q G_{pq}\delta_{qk} = \sum_q G_{pq} s_p^{(k)}$$

As before, $\mathbf{s}^{(\mathrm{k})}$ is the k-th column of the identity matrix. The quantity $\mathbf{d}^{(\mathrm{k})} = \mathbf{G}\mathbf{s}^{(\mathrm{k})}$ is the data predicted by a set of model parameters $\mathbf{m}^{\mathrm{true}(\mathrm{k})} = \mathbf{s}^{(\mathrm{k})}$ that are all zero, except for the k-th, which is unity. Thus, the two step process

$$\mathbf{d}^{(k)} = \mathbf{G}\mathbf{s}^{(k)} \quad \text{and} \quad \mathbf{r}^{(k)} = \mathbf{G}^{-g}\mathbf{d}^{(k)}$$

forms the k-th column of the resolution matrix. In practice, the equivalent linear system $\mathbf{Ar}^{(k)} = \mathbf{G}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{d}^{(k)}$ is solved instead of the equation containing the generalized inverse.

11. Symmetric resolution in the special case of convolutions. Let us consider the special case where $\mathbf{m}$ represents the discrete version of a continuous function $m(x)$ and where $\mathbf{G}$ and $\mathbf{H}$ represent convolutions. That is, $\mathbf{Gm}$ is the discrete version of $\mathrm{G(t)} * \mathrm{m(t)}$, where $*$ is the convolution operator. Furthermore, let us assume that the data are uncorrelated and with uniform variance; that is $\mathbf{C}_{\mathrm{d}}^{-1} = \sigma_d^2\mathbf{I}$. Convolutions commute; that is $\mathrm{G(t)} * \mathrm{H(t)} = \mathrm{H(t)} * \mathrm{G(t)}$. Consequently, the corresponding matrices will commute as well (except possibly for "edge effects"); that is $\mathbf{GH} = \mathbf{HG}$. Furthermore, the transform of a convolution matrix is itself a convolution – namely, the original convolution backward in time; that is, $\mathbf{H}^{\mathrm{T}} \to \mathrm{H}(-\mathrm{t})$. These properties imply that the resolution matrix:

$$\mathbf{R}^{\mathrm{G}} = \mathbf{G}^{-g}\mathbf{G} = \mathbf{A}^{-1}\mathbf{G}^{\mathrm{T}}\sigma_{\mathrm{d}}^{-2}\mathbf{G}$$

is symmetric, since

$$\mathbf{R}^{\mathrm{G}} \stackrel{?}{=} \mathbf{R}^{\mathrm{GT}}$$

$$\mathbf{A}^{-1}\mathbf{G}^{\mathrm{T}}\sigma_{\mathrm{d}}^{-2}\mathbf{G} \stackrel{?}{=} \mathbf{G}^{\mathrm{T}}\sigma_{\mathrm{d}}^{-2}\mathbf{G}\mathbf{A}^{-1}$$

$$\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{A} \stackrel{?}{=} \mathbf{A}\mathbf{G}^{\mathrm{T}}\mathbf{G}$$

$$\mathbf{G}^{\mathrm{T}}\mathbf{G}\left[\sigma_{\mathrm{d}}^2\mathbf{G}^{\mathrm{T}}\mathbf{G} + \varepsilon^2\mathbf{H}^{\mathrm{T}}\mathbf{H}\right] \stackrel{?}{=} \left[\sigma_{\mathrm{d}}^2\mathbf{G}^{\mathrm{T}}\mathbf{G} + \varepsilon^2\mathbf{H}^{\mathrm{T}}\mathbf{H}\right]\mathbf{G}^{\mathrm{T}}\mathbf{G}$$

$$\varepsilon^2\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{H}^{\mathrm{T}}\mathbf{H} \stackrel{?}{=} \varepsilon^2\mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{G}^{\mathrm{T}}\mathbf{G}$$

$$\mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{G}^{\mathrm{T}}\mathbf{G} = \mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{G}^{\mathrm{T}}\mathbf{G}$$