# The Least Square Solution and its Covariance are Invariant under a Linear Transformation

Bill Menke, December 3, 2020

Let $\mathbf{d}$ be a vector of $N$ data with covariance $\mathbf{C}_d$. Suppose a vector $\mathbf{m}$ of $M < N$ model parameters is related to the data through a linear equation $\mathbf{Gm} = \mathbf{d}$. The least squares solution and its covariance are:

$$\mathbf{m}(\mathbf{d}) = [\,\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\,\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{d} \quad \text{and} \quad \mathbf{C}_{\mathbf{m}(\mathbf{d})} = [\,\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\,\mathbf{G}]^{-1}$$

Now, suppose a second set of data $\mathbf{d}'$ are related to the first by the linear transformation $\mathbf{d}' = \mathbf{Dd}$, where $\mathbf{D}$ is an invertible matrix with inverse $\mathbf{D}^{-1}$. The first difference operator (with top-row boundary condition) and integration operator $\mathbf{D}^{-1} = \mathbf{L}$ (with top-row integration constant) are one such pair or matrices:

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & \cdots & \cdots & 1 \end{bmatrix}$$

The model parameters satisfy $\mathbf{DGm} = \mathbf{Dd} = \mathbf{d}'$, or $\mathbf{G}'\mathbf{m} = \mathbf{d}'$ with $\mathbf{G}' = \mathbf{DG}$. By the standard rules of error propagation, the covariance of $\mathbf{d}'$ is $\mathbf{C}_{d'} = \mathbf{DC}_d\mathbf{D}^{\mathrm{T}}$. The least squares solution and its covariance is using $\mathbf{d}'$ are:

$$\mathbf{m}(\mathbf{d}') = [\,\mathbf{G}'^{\mathrm{T}}\mathbf{C}_{d'}^{-1}\,\mathbf{G}']^{-1}\mathbf{G}'^{\mathrm{T}}\mathbf{C}_{d'}^{-1}\mathbf{d}'$$

$$= [\,\{\mathbf{G}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\}\ \{\,[\mathbf{D}^{T}]^{-1}\mathbf{C}_d^{-1}\mathbf{D}^{-1}\}\ \{\mathbf{DG}\}]^{-1}\,\{\,\mathbf{G}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\}\,\{\,[\mathbf{D}^{T}]^{-1}\mathbf{C}_d^{-1}\mathbf{D}^{-1}\}\,\{\mathbf{Dd}\}$$

$$= [\,\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\,\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{d} = \mathbf{m}(\mathbf{d})$$

$$\mathbf{C}_{\mathbf{m}(\mathbf{d}')} = [\,\mathbf{G}'^{\mathrm{T}}\mathbf{C}_{d'}^{-1}\,\mathbf{G}']^{-1}$$

$$= [\,\{\mathbf{G}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\}\ \{\,[\mathbf{D}^{T}]^{-1}\mathbf{C}_d^{-1}\mathbf{D}^{-1}\}\ \{\mathbf{DG}\}]^{-1} = [\,\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\,\mathbf{G}]^{-1} = \mathbf{C}_{\mathbf{m}(\mathbf{d})}$$

Here, we have used the rules $[\mathbf{A}^{-1}]^{\mathrm{T}} = [\mathbf{A}^{\mathrm{T}}]^{-1}$, $[\mathbf{AB}]^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$ and $[\mathbf{ABA}^{\mathrm{T}}]^{-1} = [\mathbf{A}^{\mathrm{T}}]^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$. Hence, the least squares solution and its covariance is invariant under an invertible linear transformation.

The invariance implies that that there is no special value in working with difference data $\mathbf{d}' = \mathbf{Dd}$, as contrasted to the original data $\mathbf{d}$, unless their actual covariance is smaller than that implied by standard error propagation; that is, $\|\mathbf{C}_{d'}\| < \|\mathbf{DC}_d\mathbf{D}^{\mathrm{T}}\|$. Thus, an “improved” solution can only result when the observational method for determining $\mathbf{d}'$ is inherently more accurate for than the one for determining $\mathbf{d}$.

In seismology, a reduction in variance occurs when differential travel time data are used in preference to absolute arrival times, because cross-correlation allows the time difference between two seismic arrivals to be determined much more accurately than the start time of either individual arrival.

Addendum, 12/04/20: Effect of error in the theory.

In some cases, error in the theory can also be important. That is, $\mathbf{Gm}$ is not a good predictor of $\mathbf{d}$ because the theory embodied in $\mathbf{G}$ is not an accurate one. Tarantola and Valette (1982) have shown that this error can be accounted by adding to the covariance of the data $\mathbf{C}_d$ the covariance of the theory $\mathbf{C}_G$; that is, all instances of $\mathbf{C}_d$ in the least squares formula are replaced by $\mathbf{C}_d + \mathbf{C}_G$.

Suppose now that the data $d_i$ are associated with an auxiliary variable $x_n = n\,\Delta x$ that increases linearly with index $n$. We will assume that observational errors are uniform and uncorrelated; that is, $\mathbf{C}_d = \sigma_d^2 \mathbf{I}$. In contrast, errors in the theory might be highly correlated for small distance intervals but less correlated for large ones. In this case, a two-sized declining exponential might suffice to model the covariance of the theory:

$$[\mathbf{C}_G]_{ij} = \gamma^2 \exp\{-b|i-j|\}$$

We now transform to the primed data $\mathbf{d}' = \mathbf{Dd}$, where $\mathbf{D}$ is the first difference operator, stated above. The transformed covariance of the data is $\mathbf{C}_{d'} = \mathbf{D}\mathbf{C}_d\mathbf{D}^{\mathrm{T}}$. It is easy to show that:

$$[\mathbf{C}_{d'}]_{11} = \sigma_d^2 \quad \text{and for } \; i > 1 \quad [\mathbf{C}_{d'}]_{ii} = 2\sigma_d^2 \;\; \text{and} \;\; [\mathbf{C}_{d'}]_{i,i+1} = -\sigma_d^2$$

$$\text{and for } j > i+1 \quad [\mathbf{C}_{d'}]_{ij} = 0$$

The transformed covariance of the theory is $\mathbf{C}_{G'} = \mathbf{D}\mathbf{C}_G\mathbf{D}^{\mathrm{T}}$. It is east to show that $[\mathbf{C}_G]_{11} = \gamma^2$. Because both $\mathbf{D}$ and $\mathbf{C}_G$ are Toeplitz, they asymptotically commute, so $\mathbf{C}_{G'} \approx \mathbf{D}\mathbf{D}^{\mathrm{T}}\mathbf{C}_G$. Furthermore $-\mathbf{D}\mathbf{D}^{\mathrm{T}}$ is asymptotically a second difference operator. Consequently, for $i > 1$ $[\mathbf{C}_{G'}]_{ii} \approx 2b\gamma^2$ and for $i > 1$ and $j > i$ $[\mathbf{C}_{G'}]_{ij} \approx -b^2[\mathbf{C}_G]_{ij}$. For highly correlated error, $b \ll 1$ and both $2b\gamma^2$ and $b^2\gamma^2$ can be considered negligible compared to $\gamma^2$:

$$[\mathbf{C}_{G'}]_{11} = \gamma^2 \quad \text{and for } \; i > 1 \text{ and } j > 1 \quad [\mathbf{C}_{G'}]_{ij} \approx 0$$

Consequently, the total covariance is:

$$[\mathbf{C}_{d'} + \mathbf{C}_{G'}]_{11} = \sigma_d^2 + \gamma^2$$

$$[\mathbf{C}_{d'} + \mathbf{C}_{G'}]_{ij} \approx [\mathbf{C}_{d'}]_{ij} \;\; \text{for } \; i > 1 \quad \text{and} \quad j > i$$

Thus, the effect of the error in the theory is only to increase the variance of $d'_1 = d_1$. It has negligible effect on the variance of $d'_i = d_i - d_{i-1}$ or on any of their covariances.

Because of the invariance demonstrated previously, the same results are obtained irrespective of whether the least squared problem is solved using $\mathbf{d}$ or $\mathbf{d}'$, as long as the error in the theory is properly accounted for in both cases. However, since the covariance is diagonal in the primed coordinate system, one might say colloquially that its use has "removed" the effect of the poor theory.