Degrees of Freedom of the Natural Solution to an Inverse Problem

William Menke, December 22, 2018
inspired by a question from Zach Eilon
(typo corrected 01/28/19)

The general issue is how to compute the number of degrees of freedom of the natural solution to a linear inverse problem; that is, the solution constructed via the singular value decomposition (SVD).

In this discussion, I will use the "full version" of SVD, where the data kernel $\mathbf{G}$ is represented as:

$$\mathbf{G} = \mathbf{U}_{N\times N}\begin{bmatrix}\mathbf{\Sigma}_{M\times M}\\ \mathbf{0}_{(N-M)\times M}\end{bmatrix}\mathbf{V}^{\mathrm{T}}_{M\times M}$$

Here $\mathbf{U}$ is an $N\times N$ orthogonal matrix, $\mathbf{V}$ an $M\times M$ orthogonal matrix, and $\mathbf{\Sigma}$ is an $M\times M$ diagonal matrix of non-negative singular values Σ_{ii}, sorted by decreasing size. The SVD of the data equation $\mathbf{d}=\mathbf{Gm}$ can be interpreted as a set of rotations in the data and model spaces that bring the equation into diagonal form:

$$\mathbf{d}' = \begin{bmatrix}\mathbf{d}'_M\\ \mathbf{d}'_{N-M}\end{bmatrix} = \mathbf{\Sigma}\mathbf{m}' = \begin{bmatrix}\mathbf{\Sigma}_{M\times M}\\ \mathbf{0}_{(N-M)\times M}\end{bmatrix}\mathbf{m}' \quad \text{with} \quad \mathbf{d}' = \mathbf{U}^{\mathrm{T}}\mathbf{d} \quad \text{and} \quad \mathbf{m}' = \mathbf{V}^{\mathrm{T}}\mathbf{m}$$

Presuming that $\mathbf{\Sigma}_{M\times M}$ is invertible, the top system can be solved exactly as $\mathbf{m}' = \mathbf{\Sigma}^{-1}_{M\times M}\mathbf{d}'_M$, in which case $\mathbf{m} = \mathbf{V}\mathbf{\Sigma}^{-1}_{M\times M}\mathbf{U}^{\mathrm{T}}\mathbf{d}$. The bottom system cannot be solved at all and represents the linear combinations of the data that cannot be fit. The number of degrees of freedom is clearly $\nu = N - M$, equal to the number of data that cannot be fit.

When $\mathbf{\Sigma}_{M\times M}$ is not invertible because $M-p$ singular values are identically zero, the so-called "natural solution" is to keep only the first p rows of the top system (those with the non-zero Σ_{ii}s) and supplement them with $M-p$ "prior" equations of the form $0 = m'_i$:

$$\begin{bmatrix}\mathbf{d}'_p\\ \mathbf{0}_{M-p}\\ \mathbf{d}'_{N-M}\end{bmatrix} = \begin{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_{p\times p} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{(M-p)\times(M-p)}\end{bmatrix}\\ \mathbf{0}_{(N-M)\times M}\end{bmatrix}\mathbf{m}'$$

Thus, $\mathbf{m}'_p = \mathbf{\Sigma}^{-1}\mathbf{d}'_p$ and $\mathbf{m}'_{M-p} = \mathbf{0}$. More generally, any set of any equations of $M-p$ the form $h'_i = \sum_{j=1}^{M-p} H'_{ij}m'_{j+p}$ can be used in place of $0 = m'_i$ to represent other types of prior information, where $\mathbf{H}'$ is an invertible matrix. Then the equation becomes:

$$\begin{bmatrix} \mathbf{d}'_p \\ \boldsymbol{h}'_{M-p} \\ \mathbf{d}'_{N-M} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{p\times p} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}'_{(M-p)\times(M-p)} \end{bmatrix} \\ \mathbf{0}_{(N-M)\times M} \end{bmatrix} \begin{bmatrix} \mathbf{m}'_{\boldsymbol{p}} \\ \mathbf{m}'_{M-p} \end{bmatrix}$$

Thus, $\mathbf{m}'_p = \boldsymbol{\Sigma}^{-1}\mathbf{d}'_p$ and $\mathbf{m}'_{M-p} = \mathbf{H}'^{-1}\mathbf{h}'_{M-p}$.

Returning now to the natural solution, we consider the prediction error

$$E' = \sigma_d^{-2}\boldsymbol{e}'^{\mathrm{T}}\boldsymbol{e}' \quad \text{with} \quad \mathbf{e}' = \mathbf{d}'^{obs} - \mathbf{d}'$$

Here, the variance of the data is σ_d^2. The errors $\mathbf{e}'$ are:

$$\boldsymbol{e}' = \begin{bmatrix} \mathbf{d}'_p \\ \boldsymbol{d}'_{M-p} \\ \mathbf{d}'_{N-M} \end{bmatrix}^{obs} - \begin{bmatrix} \mathbf{d}'_{\boldsymbol{p}} \\ \boldsymbol{d}'_{M-p} \\ \mathbf{d}'_{N-M} \end{bmatrix} = \begin{bmatrix} \boldsymbol{d}'_p \\ \boldsymbol{d}'_{M-p} \\ \mathbf{d}'_{N-M} \end{bmatrix}^{obs} - \begin{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{p\times p} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{(M-p)\times(M-p)} \end{bmatrix} \\ \mathbf{0}_{(N-M)\times M} \end{bmatrix} \begin{bmatrix} \mathbf{m}'_{\boldsymbol{p}} \\ \boldsymbol{m}'_{M-p} \end{bmatrix} =$$

$$= \begin{bmatrix} \boldsymbol{d}'_p \\ \boldsymbol{d}'_{M-p} \\ \mathbf{d}'_{N-M} \end{bmatrix}^{obs} - \begin{bmatrix} \boldsymbol{\Sigma}_{p\times p}\boldsymbol{\Sigma}_{p\times p}^{-1}\boldsymbol{d}'_p \\ \mathbf{0}_{M-p} \\ \mathbf{0}_{N-M} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_p \\ \boldsymbol{d}'^{obs}_{M-p} \\ \boldsymbol{d}'^{obs}_{N-M} \end{bmatrix}$$

The number of elements in $\boldsymbol{e}'$ is N, but p of them are identically zero, so the number of degrees of freedom is $\nu = N - p$. The total error is in the form of a dot product, which is invariant under rotations, so $E = E'$. A rotation only mixes linear combinations of $\mathbf{e}$, so E also has $\nu = N - p$ degrees of freedom. Whether either E or E' are chi-squared distributed is unclear to me, since e′ does not appear to be guaranteed to have zero mean. Yet normal practice would be to consider it to be chi-squared distributed, with $\nu = N - p$ degrees of freedom.

Sometimes, one throws out rows with near-zero singular values as well as rows with zero singular values. This practice merely decreases p; all the points made above remain unchanged.

Consider the special case where the prior equations are $m'^0_i = m'_i$, which reduces to the natural solution when $m'^0_i = 0$. The solution is then:

$$\mathbf{m}' = \begin{bmatrix} \boldsymbol{\Sigma}_{p\times p}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{(M-p)\times(M-p)}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{d}'_M \\ \mathbf{m}'^0_{M-p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{p\times p}^{-1}\mathbf{d}'_M \\ \mathbf{M}_{(M-p)\times(M-p)}^{-1}\mathbf{m}'^0_{M-p} \end{bmatrix}$$

or

$$\mathbf{m} = \mathbf{V}_p\boldsymbol{\Sigma}_{p\times p}^{-1}\mathbf{U}_p^{\mathrm{T}}\mathbf{d} + \mathbf{V}_0\mathbf{M}_{(M-p)\times(M-p)}^{-1}\mathbf{m}'^0_{M-p}$$

Here $\mathbf{U} = [\mathbf{U}_p, \mathbf{U}_0]$ and $\mathbf{V} = [\mathbf{V}_p, \mathbf{V}_0]$ have been partitioned into two submatrices, the first of which has p columns. This form of the solution emphasizes that the natural solution has a “hidden” zero part:

$$\mathbf{m} = \mathbf{V}_p \mathbf{\Sigma}_{p\times p}^{-1} \mathbf{U}_p^{\mathrm{T}} \mathbf{d} + \mathbf{0}$$

Consequently, the covariance $\mathbf{C}_m$ of the solution $\mathbf{m}$ depends both upon the covariance of the data (say $\mathbf{C}_d$) and the covariance of the prior information (say $\mathbf{C}_0$). Furthermore, the so-called resolution matrix $\mathbf{R} = \mathbf{V}_p \mathbf{V}_p^{\mathrm{T}}$ is really a "deviatoric resolution"; that is, the resolution of deviations about a prior solution (see Menke (2018) for details).

Reference

W. Menke, Geophysical Data Analysis: Discrete Inverse Theory, 4th Edition, Elsevier, 2018.