Adding Linear Prior Information to the Natural Solution of an Inverse Problem

Bill Menke, April 2, 2026

Introduction. This note explores interaction of data and prior information, in the context of the so-called natural solution (SVD solution) to an inverse problem. The key attribute of this solution is that the data are given precedence over the prior information, in the sense that the solution, first and foremost, minimizes the prediction error, and minimizes the error in prior information only to the extent that it does not increase the prediction error. This asymmetry contrasts to the completely symmetrical Generalized Least Squares, and alternative solution that gives the two types of error equitable weight.

The natural solution to the inverse problem $\mathbf{Gm} = \mathbf{d}^{obs}$ (with covariance $\sigma_d^2\mathbf{I}$) is:

$$\mathbf{m}_N = \mathbf{G}_N^{-g}\mathbf{d}^{obs} \quad \text{with} \quad \mathbf{G}_N^{-g} \equiv \mathbf{V}_p\mathbf{\Lambda}_p^{-1}\mathbf{U}_p^T \tag{1}$$

Here, $\mathbf{G}_N^{-g}$ is the natural generalized inverse. The matrix $\mathbf{G}$ has singular value decomposition $\mathbf{G} = \mathbf{U}_p\mathbf{\Lambda}_p\mathbf{V}_p^T$ where $\mathbf{\Lambda}_p$ contains no zero singular values. A corresponding set of matrices, $\mathbf{U}_0, \mathbf{\Lambda}_p, \mathbf{V}_0$ are associated with the zero singular values, with the columns of $\mathbf{V}_0$ being termed the null vectors of the problem. The natural solution minimizes the prediction error $E \equiv \mathbf{e}^T\mathbf{e}$ with $\mathbf{e} \equiv \mathbf{d}^{obs} - \mathbf{Gm}_N$ and it also minimizes the solution length $L_m \equiv \mathbf{m}_N^T\mathbf{m}_N$. Formulas involving the natural generalized inverse often contain the matrix products:

$$\mathbf{R}_G \equiv \mathbf{G}_N^{-g}\mathbf{G} = \mathbf{V}_p\mathbf{V}_p^T \quad \text{and} \quad \mathbf{N}_G \equiv \mathbf{G}\mathbf{G}_N^{-g} = \mathbf{U}_p\mathbf{U}_p^T \tag{2}$$

which are called the model resolution matrix and data resolution matrix, respectively. Any solution:

$$\mathbf{m} = \mathbf{m}_N + \mathbf{V}_o\boldsymbol{\alpha} \tag{3}$$

has minimum prediction error, irrespective of the choice of $\boldsymbol{\alpha}$, but does not, in general, minimize L_m. The minimum error property is derived by inserting Eqn. (3) into $\mathbf{Gm} = \mathbf{d}^{obs}$ and applying the rule $\mathbf{V}_0^T\mathbf{V}_p = 0$.

Inclusion of Prior Information. Now suppose that we want to impose prior information of the form $\mathbf{Hm} = \mathbf{h}$ (with covariance $\sigma_h^2\mathbf{I}$). An equation for $\boldsymbol{\alpha}$ is given by:

$$\mathbf{Hm} = \mathbf{H}[\mathbf{m}_N + \mathbf{V}_0\boldsymbol{\alpha}] = \mathbf{Hm}_N + \mathbf{HV}_0\boldsymbol{\alpha} = \mathbf{h}$$

$$\mathbf{X}\boldsymbol{\alpha} = \mathbf{x} \quad \text{with} \quad \mathbf{X} \equiv [\mathbf{HV}_0] \ \text{and} \ \mathbf{x} \equiv [\mathbf{h} - \mathbf{Hm}_N] \tag{4}$$

The solution that minimizes the error in prior information $L_m \equiv \mathbf{l}^T\mathbf{l}$ where $\mathbf{l} \equiv \mathbf{h} - \mathbf{Hm}$ can be found by solving Eqn. (4) in the least squares sense. In practice, the use of damped least squares (say, with damping factor ε^2) is preferable, to avoid numerical problems in cases where $\det(\mathbf{X})$ is close to zero. The $\boldsymbol{\alpha}$ that leads to the best fit of the prior information without increasing the prediction error is:

$$\boldsymbol{\alpha} = \mathbf{X}^{-g}\mathbf{x} = \mathbf{X}^{-g}\mathbf{h} - \mathbf{X}^{-g}\mathbf{H}\mathbf{G}_N^{-g}\mathbf{d}^{obs}$$

$$\text{with} \quad \mathbf{X}^{-g} \equiv \mathbf{A}^{-1}\mathbf{V}_0^T\mathbf{H}^T \quad \text{and} \quad \mathbf{A} \equiv [\mathbf{V}_0^T\mathbf{H}^T\mathbf{H}\mathbf{V}_0 + \varepsilon^2\mathbf{I}] \tag{5}$$

Inserting Eqn. (5) into Eqn. (3) leads to the solution:

$$\mathbf{m} = \mathbf{G}_N^{-g}\mathbf{d}^{obs} + \mathbf{V}_0\left[\mathbf{X}^{-g}\mathbf{h} - \mathbf{X}^{-g}\mathbf{H}\mathbf{G}_N^{-g}\mathbf{d}^{obs}\right]$$

$$= [\mathbf{I} - \mathbf{Y}]\mathbf{G}_N^{-g}\mathbf{d}^{obs} + [\mathbf{B}\mathbf{H}^T]\mathbf{h}$$

$$\text{with} \quad \mathbf{Y} \equiv \mathbf{V}_0\mathbf{X}_N^{-g}\mathbf{H} = \mathbf{B}\mathbf{H}^T\mathbf{H} \quad \text{and} \quad \mathbf{B} \equiv \mathbf{V}_0\mathbf{A}^{-1}\mathbf{V}_0^T \tag{6}$$

With the assumption that the data and prior information are uncorrelated, the usual rules of error propagation can be used to calculate the posterior covariance of the solution::

$$[\operatorname{cov}\mathbf{m}] = \sigma_d^2[\mathbf{I} - \mathbf{Y}]\left[\mathbf{V}_p\boldsymbol{\Lambda}_p^{-2}\mathbf{V}_p^T\right][\mathbf{I} - \mathbf{Y}]^T + \sigma_h^2\mathbf{B} \tag{7}$$

Here, we have used the approximation**:**

$$\{\mathbf{B}\mathbf{H}^T\}\{\mathbf{B}\mathbf{H}^T\}^T = \{\mathbf{V}_0\mathbf{A}^{-1}\mathbf{V}_0^T\mathbf{H}^T\}\{\mathbf{V}_0\mathbf{A}^{-1}\mathbf{V}_0^T\mathbf{H}^T\}^T = \mathbf{V}_0\mathbf{A}^{-1}[\mathbf{V}_0^T\mathbf{H}^T\mathbf{H}\mathbf{V}_0]\mathbf{A}^{-1}\mathbf{V}_0^T \approx \mathbf{V}_0\mathbf{A}^{-1}\mathbf{V}_0^T = \mathbf{B}$$

$$\mathbf{G}_N^{-g}\mathbf{G}_N^{-gT} = \left\{\mathbf{V}_p\boldsymbol{\Lambda}_p^{-1}\mathbf{U}_p^T\right\}\left\{\mathbf{V}_p\boldsymbol{\Lambda}_p^{-1}\mathbf{U}_p^T\right\}^T = \mathbf{V}_p\boldsymbol{\Lambda}_p^{-2}\mathbf{V}_p^T \tag{8}$$

The quantity $[\operatorname{cov}\mathbf{m}_N] = \sigma_d^2\mathbf{V}_p\boldsymbol{\Lambda}_p^{-2}\mathbf{V}_p^T$ is the covariance of the natural solution $\mathbf{m}_N$.

A Special Case. Suppose the case of a minimum length solution $\mathbf{m}_{ML}$, with $\mathbf{H} = \mathbf{I}$ and $\mathbf{h} = 0$ and $\sigma_m^2 \equiv \sigma_h^2$. Then:

$$\mathbf{A} = [\mathbf{V}_0^T\mathbf{H}^T\mathbf{H}\mathbf{V}_0 + \varepsilon^2\mathbf{I}] \approx \mathbf{I}$$

$$\mathbf{X}_N^{-g} = \mathbf{A}^{-1}\mathbf{V}_0^T\mathbf{H}^T = \mathbf{V}_0^T$$

$$\mathbf{Y} \equiv \mathbf{V}_0\mathbf{X}_N^{-g}\mathbf{H} = \mathbf{V}_0\mathbf{V}_0^T$$

$$\mathbf{V}_0\mathbf{A}^{-1}\mathbf{V}_0^T = \mathbf{V}_0\mathbf{V}_0^T = \mathbf{I} - \mathbf{V}_p\mathbf{V}_p^T = \mathbf{I} - \mathbf{R}_G$$

$$\text{because } \mathbf{I} = \mathbf{V}\mathbf{V}^T = [\mathbf{V}_p \quad \mathbf{V}_0]\begin{bmatrix}\mathbf{V}_p^T \\ \mathbf{V}_0^T\end{bmatrix} = \mathbf{V}_p\mathbf{V}_p^T + \mathbf{V}_0\mathbf{V}_0^T$$

$$\mathbf{m}_{ML} = [\mathbf{I} - \mathbf{V}_0\mathbf{V}_0^T]\mathbf{G}_N^{-g}\mathbf{d}^{obs} = [\mathbf{I} - \mathbf{V}_0\mathbf{V}_0^T]\mathbf{V}_p\mathbf{\Lambda}_p^{-1}\mathbf{U}_p^T\mathbf{d}^{obs} = \mathbf{V}_p\mathbf{\Lambda}_p^{-1}\mathbf{U}_p^T\mathbf{d}^{obs} = \mathbf{G}_N^{-g}\mathbf{d}^{obs} = \mathbf{m}_N$$

$$[\text{cov}\,\mathbf{m}_{ML}] = \sigma_d^2\mathbf{G}_N^{-g}\mathbf{G}_N^{-gT} + \sigma_m^2\mathbf{V}_0\mathbf{V}_0^T = \sigma_d^2\mathbf{V}_p\mathbf{\Lambda}_p^{-2}\mathbf{V}_p^T + \sigma_m^2(\mathbf{I} - \mathbf{R}_G)$$

(9)

Thus, the covariance of the minimum length solution contains a "hidden" contribution that depends on the accuracy of the prior information. In other words, the minimum length solution should be written

$$\mathbf{m}_{ML} = \mathbf{m}_N + 0$$

(10)

to emphasize that the covariance of the zero matters.

An example. This example is based on the gravity problem in Figure 9.8 of Menke (2024). The model is a square object described by 20×20 grid of density pixels (that is, $M = 200$ model parameters). The data are noisy gravity observations on a line above the object. The problem in the book compares the true solution to the natural solution for $p = 4$ (Fig. 2A,B).

I now add linear prior information of the form $m_{ij} - 2m_{kl} = 1$ where (i, j) and (k, l) are corresponding pixels in the top and bottom half of the model, respectively. The constrained solution (Fig. 2C) has the same prediction error as the natural solution ($E = 15{,}787$ in both cases) and lower error in prior information ($L = 0.00$ compared to 443.0). The data and prior parts of the covariance for a model parameter in the top half of the model is shown in Fig. 3.

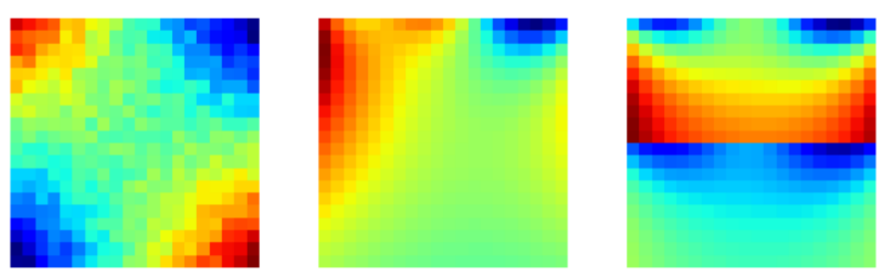

Fig. 2. (left) True solution, (middle) natural solution for $p = 4$, (right) Solution approximately satisfying the prior information.

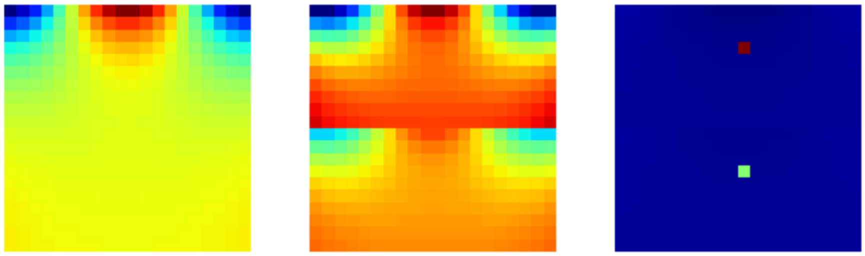

Fig. 3. Covariance for the model parameter at row 4, column 11 of the 2D model. (left) Natural solutions. (middle) Part of the covariance associated with the data in the constrained solution. (Right) Part associated with the prior information.

Reference

Menke, W. (2024), Geophysical Data Analysis and Inverse Theory with MATLAB(R) and Python, Fifth Edition, Academic Press, Elsevier (Amsterdam), 400pp.