

On Regularized Inversion with Minimization of a Model Parameter

Joshua D Carmichael, Ken C. Creager, Erin C. Pettit

Last update 11/2007

Many problems in linear inversion theory can be expressed as least-squares problems. In practice these problems involve a set of observations expressible as a data vector and a mathematical model based on the physics and geometry of the problem. The forward problem is to compute the data \mathbf{d} from the data kernel \mathbf{G} , whose elements depend upon mathematical model, given a parameter vector \mathbf{m} . The inverse problem is to estimate the parameter vector \mathbf{m} given the data. In the presence of noise, the best model is found when the 2-norm of the misfit between predicted data and observed data is minimized. Problems often arise in applications where the misfit is minimized subject to a constraint on the model. These problems usually can be recast as least-square problems in which the weighted sum of two scalar functions is minimized. These typically involve a trade-off parameter that weights the relative importance of minimizing each scalar in the sum. In this work we consider the less standard problem of minimizing a single model parameter along with the misfit between the predicted data and observed data.

This study will focus on problems of the form:

$$\min \|\mathbf{G}\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 (m_i - c_i)$$

where α is the trade-off parameter, m_i is the i^{th} model parameter, and c_i is the i^{th} component of some vector \mathbf{c} . This problem is equivalent to:

$$\min \left\| \begin{bmatrix} \mathbf{G} \\ \alpha \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T \end{bmatrix} \mathbf{m} - \begin{bmatrix} \mathbf{d} \\ \alpha \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T \mathbf{c} \end{bmatrix} \right\|_2^2$$

The least-square solution to this minimization problem can be calculated from the normal equations:

$$\left[\mathbf{G}^T \mathbf{G} + \alpha^2 \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T \right] \mathbf{m} = \mathbf{G}^T \mathbf{d} + \alpha^2 \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T \mathbf{c}$$

The data kernel \mathbf{G} may be factored into a product of three matrices using the SVD so that $\mathbf{G} = \hat{\mathbf{U}} \hat{\mathbf{S}} \hat{\mathbf{V}}^T$, where $\hat{\mathbf{U}} \in R^{m \times m}$, $\hat{\mathbf{S}} \in R^{m \times n}$ and $\hat{\mathbf{V}} \in R^{n \times n}$. Both $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ are orthogonal matrices, and $\hat{\mathbf{S}}$ is diagonal with non-negative elements. If only the first p diagonal elements (singular values) of $\hat{\mathbf{S}}$ are nonzero, \mathbf{G} is rank deficient and can be written more compactly using only the non-zero singular so that $\mathbf{G} = \mathbf{U} \mathbf{S} \mathbf{V}^T$, where $\mathbf{U} \in R^{m \times p}$, $\mathbf{S} \in R^{p \times p}$ and $\mathbf{V} \in R^{n \times p}$. Considering for the moment data kernels \mathbf{G} with full rank, one benefit of this factorization follows from the fact that any model \mathbf{m} can be written as a linear combination of the columns of \mathbf{V} so for some coefficients β_j :

$$\mathbf{m} = \sum_{j=1}^n \beta_j \mathbf{V}_j$$

The normal equations are now:

$$\left[\mathbf{V}^T \mathbf{S} \mathbf{V}^T + \alpha^2 \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T \right] \sum_{j=1}^n \beta_j \mathbf{V}_j = \mathbf{V}^T \mathbf{U}^T \mathbf{d} + \alpha^2 \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^T \mathbf{c}$$

The product on the left can be evaluated using the orthogonality of \mathbf{V} 's columns and the unit vector $\hat{\mathbf{e}}_i$ can also be re-written as a linear combination of the columns of \mathbf{V} by making a change of basis. This effectively redefines the problem as finding the coefficients $\beta_j, j = 1, 2, \dots, n$, for \mathbf{V} 's columns:

$$\sum_{j=1}^n \left(s_j^2 \beta_j + \alpha^2 \mathbf{V}_j^T \hat{\mathbf{e}}_i \sum_{k=1}^n \beta_k V_{ik} \right) \mathbf{V}_j = \sum_{j=1}^n \left(s_j \mathbf{U}_j^T \mathbf{d} + \alpha^2 c_i \mathbf{V}_j^T \hat{\mathbf{e}}_i \right) \mathbf{V}_j$$

To solve for each β_j we again make use of the orthogonality of \mathbf{V} 's columns and multiply the above expression by \mathbf{V}_1^T then \mathbf{V}_2^T , and so on. This allows us to construct a system of linear equations expressible in matrix form as:

$$\left[\mathbf{S}^T \mathbf{S} + \alpha^2 \mathbf{V}_{,i}^T \mathbf{V}_{,i} \right] \hat{\mathbf{a}} = \mathbf{S}^T \mathbf{U}^T \mathbf{d} + \alpha^2 c_i \mathbf{V}_{,i}^T$$

where $\hat{\mathbf{a}}$ is the coefficient vector that needs to be determined and $\mathbf{V}_{,i}$ is the i^{th} row of \mathbf{V} . Since we are assuming \mathbf{G} has full rank, $\mathbf{S}^T \mathbf{S}$ must have full rank. Multiplying through both sides by $(\mathbf{S}^T \mathbf{S})^{-1}$ gives:

$$\left[\mathbf{I} + \alpha^2 (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{V}_{,i}^T \mathbf{V}_{,i} \right] \hat{\mathbf{a}} = (\mathbf{S}^T \mathbf{S})^{-1} \left[\mathbf{S}^T \mathbf{U}^T \mathbf{d} + \alpha^2 c_i \mathbf{V}_{,i}^T \right]$$

The matrix $\mathbf{I} + \alpha^2 (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{V}_{,i}^T \mathbf{V}_{,i}$ is a special type of matrix known as a rank-one perturbation of the identity. Given full rank, its inverse has a particularly simple form which we will denote as \mathbf{F}_i . In this case, it is given by:

$$\mathbf{F}_i = \mathbf{I} - \alpha^2 \kappa_i (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{V}_{,i}^T \mathbf{V}_{,i}$$

where the scalar κ_i is a function of α and is defined by:

$$\kappa_i = \frac{1}{1 + \alpha^2 \sum_k^p \frac{V_{ik}^2}{s_k^2}}$$

Now $\hat{\mathbf{a}}$ may be computed directly. It follows that \mathbf{m} can be written as the product:

$$\mathbf{m} = \mathbf{V} \mathbf{F}_i (\mathbf{S}^T \mathbf{S})^{-1} [\mathbf{S}^T \mathbf{U}^T \mathbf{d} + \alpha^2 c_i \mathbf{V}_i^T]$$

The model may be expressed explicitly in terms of \mathbf{V} 's columns as well:

$$\mathbf{m} = \sum_j^p \frac{\mathbf{U}_j^T \mathbf{d}}{s_j} \mathbf{V}_j + \left[\alpha^2 \left(c_i - \kappa_i \sum_k^p \frac{\mathbf{U}_k^T \mathbf{d}}{s_k} V_{ik} \right) - \alpha^4 \left(\kappa_i c_i \sum_k^p \frac{V_{ik}^2}{s_k^2} \right) \right] \sum_j^p \frac{V_{ij}}{s_j^2} \mathbf{V}_j$$

To aid interpretation, we make the following observations: the un-regularized model solution $\mathbf{m}(0)$ corresponding to $\alpha = 0$ satisfies $\min \|\mathbf{G}\mathbf{m}(0) - \mathbf{d}\|_2^2$ and is expressible as

$$\mathbf{m}(0) = \sum_j^p \frac{\mathbf{U}_j^T \mathbf{d}}{s_j} \mathbf{V}_j$$

In addition, the scalar function of α in the square brackets has coefficients with statistical aspects related to the un-regularized problem. The covariance of $\mathbf{m}(0)$ is

$$\text{cov} \mathbf{m}(0) = \sum_j^p \frac{\mathbf{V}_j \mathbf{V}_j^T}{s_j^2}$$

The i^{th} diagonal entry in the above matrix is just

$$\text{var} m_i(0) = \sum_k^p \frac{V_{ik}^2}{s_k^2}$$

which provides a measure of uncertainty in the estimates of the i^{th} model parameter. The inverted model can be written now as

$$\mathbf{m} = \mathbf{m}(0) + \alpha^2 \left(\frac{c_i - m_i(0)}{1 + \alpha^2 \text{var} m_i(0)} \right) \sum_j^p \frac{V_{ij}}{s_j^2} \mathbf{V}_j$$

so that the un-regularized solution is explicitly decoupled from the regularized one. It is also clear that a vanishing variance of $m_i(0)$ results in only the difference $c_i - m_i(0)$ being weighted by the trade-off parameter during inversion.

The column weights in the first sum are identical to those obtained without any regularization. The regularized solution is completely decoupled from the un-regularized one and can be computed simultaneously. It too can be separated into two terms, one dependent upon \mathbf{c} and one independent of \mathbf{c} .

Resolution and Covariance

The model estimate obtained above is of the form:

$$\mathbf{m} = \mathbf{G}^{-g} \mathbf{d} + \mathbf{p}$$

where matrix \mathbf{G}^{-g} and vector \mathbf{p} are respectively:

$$\mathbf{G}^{-g} = \mathbf{V} \mathbf{F}_i (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{U}^T$$

$$\mathbf{p} = \alpha^2 c_i \mathbf{V} \mathbf{F}_i (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{V}_i^T$$

\mathbf{G}^{-g} may be interpreted as a generalized inverse. Assuming normally and equally distributed errors in the data with variance σ^2 , the covariance, model resolution, and data resolution matrices are respectively:

$$\text{cov} \mathbf{m} = \sigma^2 \mathbf{V} \mathbf{F}_i (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{F}_i^T \mathbf{V}^T$$

$$\mathbf{R}_m = \mathbf{V} \mathbf{F}_i \mathbf{V}^T$$

$$\mathbf{R}_d = \mathbf{U} \mathbf{F}_i (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{U}^T$$

The resolution quality and the size of the covariance matrix elements are each dependent upon the matrix \mathbf{F}_i and it is natural to interpret the columns as filter factors weighting the columns of \mathbf{U} and \mathbf{V} that are used to build those matrices.