6. ERROR ANALYSIS IN LINEAR INVERSION

One question that is fundamental to geophysical data analysis is, **how representative of the real geophysical system is our reconstructed least squares model or how accurate is our solution to the given problem ?** Recall that our initial assumption was that the experimental data contain errors (which is why we cannot fit them exactly). **One may therefore be interested in how the experimental errors translate into errors in the model estimates.** The answers obviously come from statistics. Inverse theory not only provides us with estimate of the relevant parameters but also furnishes a plethora of related information that enable us to gauge the "goodness" of the least squares solution to the inverse problem. Some of such "auxiliary parameters" are described following a discussion of how to incorporate available observational errors directly in the inversion process.

6.1 Elaborate Treatment of Observational Errors in Inversion

It may be appreciated from our treatment of inversion so far that the process involves finding a solution that minimizes a suitably chosen quantity- the squared distance between our solution and the given data. Our desire should be that the solution be both numerically and statistically stable. Statistical stability is important because of the differening uncertainties associated with our observations. If observational errors are available, we can incorporate them directly in the problem formulation to obtain a more acceptble **weighted solution.** Thus, assuming that the n standard data errors, σ_i are Gaussian with zero mean and are statistically independent, we can define an $n \times n$ diagonal weighting matrix, W as

$$W = diag[\,\frac{1}{\sigma_1}\,,\frac{1}{\sigma_2}\,,\frac{1}{\sigma_3}\,,\,\cdots\,,\frac{1}{\sigma_n}\,] = \sigma^{-1}I.$$

We can then re-state the constrained inversion problem of Section 5.2, for example, as:

minimize

$$q = (Wd - WGm)^{\mathrm{T}}(Wd - WGm) + \beta^2(m^{\mathrm{T}}D^{\mathrm{T}}Dm)$$

where the weighting (or scaling) of each datum by its associated observational error ensures that undue importance is not given to poorly estimated (i.e., noisy) data. This is a somewhat more robust formulation and the solution is identically,

$$m_s = ((WG)^T WG + \beta^2 H)^{-1}(WG)^T Wd \quad (6.1a)$$

or simply

$$m_s = ((WG)^T WG + \beta^2 I)^{-1}(WG)^T Wd \quad (6.1b)$$

if $D = I$. Notice that β is just a single number in the above formulation. If we are interested in retaining specified values of the parameters, i.e., biased estimation *proper* (Section 5.1), a recommended elegant approach is to define a diagonal matrix of undetermined multipliers β and re-state the problem as *minimize*

$$q = (Wd - WGm)^T(Wd - WGm) + \beta(Dm-h)^T\beta^T(Dm-h).$$

The solution to this problem is

$$m_b = ((WG)^T WG + D^T BD)^{-1}[(WG)^T Wd + BD^T h] \quad (6.2)$$

where $B = \beta^T\beta$. The diagonal elements of β are assigned a constant positive value for those parameters specified *a priori* and the rest are set to nought. The somewhat identical role of the matrices W and β is obvious and we shall examine the statistical implications later. Notice that if we define another matrix $E = W^T W$, then eq. 6.2 may be expressed in another enlightening form

$$m_b = (G^T EG + D^T BD)^{-1}[G^T Ed + D^T Bh] \quad (6.3)$$

which in every respect is equivalent to the so-called **Bayesian Estimator** (see e.g., Jackson and Matsu'ura, 1985; Duijndam,1988).

6.2 Assessing the Quality of a Solution

6.2.1 Goodness-of-fit

This is the commonly used model acceptance criterion. Assuming that our data d_i are normally distributed about their expected values and with known uncertainties σ_i (the experimental errors), we can assess the fit beween the observed and predicted data by calculating the statistical parameter q defined by

$$q = \sum_{i=1}^{n} \frac{(d_i^{obs} - G_{ij}m_j)^2}{\sigma^2} \qquad j = 1, p$$

or simply

$$q = \sum_{i=1}^{n} \| Wd^{obs} - WGm \|^2 \quad \text{for a weighted solution.}$$

For n independent observations and p independent parameters, q is distributed as χ^2 (Chi-square) with $(n-p)$ degrees of freedom.

In geophysical inversion, we reject or accept a solution to the problem being considered based on the value of q. The expected value of q is n (fron Chi-square statistics) but in practice a model with $\boxed{n-p < q \lesssim n+\sqrt{2n}}$ is acceptable. However, if $q \ll n$, the model is said to over-fit the observed data and if $q \gg n$, the model under-fits the data. Over-fitting may lead to solutions that contain artifacts of the computer !

If experimental errors are not available, an unbiased estimate of σ^2 is given by

$$\sigma^2 = \frac{(d^T d - m^T G^T G m)}{n-p} \equiv \frac{\text{sum of squares of residuals}}{n-p}.$$

Another measure of goodness of a solution is the Root Mean Square (rms) error given by

$$\text{rms} = \frac{1}{n} \sum_{i=1}^{n} \frac{(d_i^{obs} - G_{ij} m_j)^2}{\sigma^2}$$

or simply

$$\text{rms} = \frac{1}{n} \sum_{i=1}^{n} \| Wd^{obs} - WGm \|^2 \quad \text{for a weighted solution.}$$

Obviously one should not aim for solutions with $1.0 \ll rms \gg 1.0$ Note that the number of degrees of freedom for a constrained solution with l independent constraint equations is $\{n-p+l\}$ and we should substitute this quantity for $n-p$ in the above formulations where appropriate.

6.2.2 Parameter Resolution Matrix

For a linear system, we can assess the quality of the model derived from a given data set by examining the parameter resolution matrix (Jackson,1972). We will show how it is derived for the unbiased least squares solution and then apply the same strategy to the cases where we have constrained the solution process.

Case 1: Unconstrained solution

Recall that the unconstrained least squares solution is $m = \{(G^T G)^{-1} G^T\} d$ which in terms of SVD is simply $m = \{V\Lambda^{-1} U^T\} d$. Let us denote the quantity in braces as H.

The matrix H is the generalized inverse used in the estimation and post-multiplying it by the applicable design matrix (or data kernel) gives the resolution matrix R, i.e.,

$$R = HG = (G^{T}G)^{-1}G^{T} \cdot G = I \tag{6.3a}$$

In terms of the SVD of G,

$$R = \{V\Lambda^{-1}U^{T}\}\{U\Lambda V^{T}\} = VV^{T}.$$

R is of dimension $p \times r$ (where r is the number of non-zero eigenvalues of the problem). If $R = I$ (identity matrix), *i.e.*, $r=p$, then each model parameter is uniquely determined The deviation of the rows of R from those of the identity matrix, I, is generally assumed to be a measure of the lack of resolution for the corresponding model parameters.

Case 2 : The Marquardt damped solution

For the damped least squares solution given by eq. 5.10, we have that

$$R = (G^{T}G + \beta I)^{-1}G^{T} \cdot G \ = I + (\frac{G^{T}G}{\beta I}) \tag{6.3b}$$

which, in terms of the SVD of G, is simply

$$R = V\Lambda_{D}^{-2}V^{T} \cdot V\Lambda^{2}V^{T} = \frac{V\Lambda^{2}V^{T}}{(\Lambda + \beta I)^{2}} \ .$$

It is obvious that the damped solution obtained by simply adding a constant bias to the eigenvalues of a problem does not have a perfect resolution.

Case 3: Inversion with *a priori* data

The constrained inversion formula represented by eq. 5.2, may be written as

$$\hat{m}_c = \{(G^{T}G + \beta^{2}D^{T}D)^{-1}G^{T}\} \cdot d + \{(G^{T}G + \beta^{2}D^{T}D)^{-1}\beta D^{T}\} \cdot \beta h$$

and hence the parameter resolution matrix may be calculated as

$$R = (G^{T}G + \beta^{2}D^{T}D)^{-1}G^{T} \cdot G + (G^{T}G + \beta^{2}D^{T}D)^{-1}\beta D^{T} \cdot \beta D$$

$$= (G^T G + \beta^2 D^T D)^{-1}(G^T G + \beta^2 D^T D) = I \qquad (6.3c)$$

which in terms of SVD of the augmented (or partitioned) design matrix G_*, defined as

$$G_* = \begin{bmatrix} G \\ \beta D \end{bmatrix}$$

is simply $V_* V_*^T$. Thus the constrained solution incorporating *a priori* parameter estimates has perfect resolution. It is may be noted that the constrained solution obtained by augmenting the matrix G with the identity matrix D and the data vector d with the null vector h (i.e., eq. 5.8) also has perfect resolution since the null vector is some kind of fictitious *a priori* data eventhough the inversion formula is mathematically equivalent to the popular Marquardt formula represented by eq. 5.10.

It is remarked here that the resolution matrix is only an experimental design guide. It is always an identity matrix for cases 1 and 3. It has nothing to do with the actual field observations and its utility in parameter error analysis has often been over-stated in the geophysical inversion literature. A perfect resolution does not imply in every respect an accurate or reliable model. To illustrate, the calculated values of R for the refraction seismology delay-time problem of Examples 4.4.3 and 5.1.2 are shown in Tables 6.2.2.1 and 6.2.2.2 respectively.

col.1	col.2	col.3	col.4	col.5	col.6
.9999997	7.14512E-08	1.82836E-08	1.75506E-08	2.93858E-08	-1.18523E-08
7.14512E-08	1.00000	3.67795E-08	3.40229E-08	2.22605E-08	-7.54078E-08
1.82836E-08	3.67795E-08	1.00000	-1.38533E-08	1.32156E-07	-9.01755E-09
1.75506E-08	3.40229E-08	-1.3853E-08	1.000000	1.12674E-07	1.25385E-08
2.93858E-08	2.22605E-08	1.32156E-07	1.12674E-07	.9999999	-9.54640E-09
-1.1852E-08	-7.5408E-08	-9.0176E-09	-1.25385E-08	-9.54640E-09	.9999999

Table 6.2.2.1. Resolution matrix, R for an unconstrained six-parameter problem. The i^{th} column elements relate to the i^{th} parameter. The degree of correlation between the various parameters can be gleaned from the rows of the matrix. Notice that R is delta-like suggesting perfect resolution.

col.1	col.2	col.3	col.4	col.5	col.6
1.000000	4.10187E-08	-2.0286E-09	-7.94157E-09	-4.34869E-09	-6.02772E-09
4.10187E-08	1.000000	-1.6353E-08	-5.59144E-09	-9.41843E-09	-9.81052E-08
-2.0286E-09	-1.6353E-08	1.000000	-3.89008E-08	-1.34107E-07	1.57051E-08
-7.9416E-09	-5.5914E-09	-3.8901E-08	1.000000	-9.6212E-08	4.16797E-09
-4.3487E-09	-9.4184E-09	-1.3411E-07	-9.6212E-08	1.000000	1.50810E-08
-6.0277E-09	-9.8105E-08	1.57051E-08	4.16797E-09	1.50810E-08	1.000000

Table 6.2.2.2. Resolution matrix, $\boldsymbol{R}$ for a constrained six-parameter problem. Notice in this case that $\boldsymbol{R}$ is also delta-like suggesting perfect resolution.

It is obvious that even the unconstrained problem which we know not to be determined by the practical data has a perfect resolution. The calculated $\boldsymbol{R}$ for an optimal damped solution (i.e., augmented singular values) for the seismic delay-time problem is given in Table 6.2.2.3 for the sake of completeness. Notice in this case that $\boldsymbol{R}$ is not an identity

col.1	col.2	col.3	col.4	col.5	col.6
7.99918E-01	-2.00064E-01	1.99958E-01	1.99957E-01	1.99939E-01	1.65436E-05
-2.00064E-01	7.99907E-01	1.99955E-01	1.99954E-01	1.99935E-01	1.78145E-05
1.99958E-01	1.99955E-01	7.99956E-01	-2.00016E-01	-2.00027E-01	9.78324E-06
1.99957E-01	1.99954E-01	-2.00016E-01	7.99955E-01	-2.00027E-01	9.94825E-06
1.99939E-01	1.99935E-01	-2.00027E-01	-2.00027E-01	7.99929E-01	1.44205E-05
1.65436E-05	1.78145E-05	9.78324E-06	9.94825E-06	1.44205E-05	9.99994E-01

Table 6.2.2.3 Resolution matrix for an optimal damped solution for $\beta = 0.00001$. Notice that the diagonal elements corresponding to the first five parameters have values of about 0.8 while the sixth parameter has a diagonal element of 1.0 so that $\boldsymbol{R} \neq \boldsymbol{I}$. The velocity parameter is thus the only parameter that may have been well determined in this inversion.

matrix which is in accord with our derivation in Section 6.2.2. The information provided by $\boldsymbol{R}$ here is that the slowness (and by implication the velocity) parameter may have been well determined. As we saw for the unconstrained inversion problem, a delta-like

R does not guarantee a meaningful solution. A cautious use of R would thus seem appropriate when dealing with practical data. For instance, it may be said that providing that the generalized inverse used in the calculations exists, then the true solution may be found if $R = I$; a lack of perfect resolution would therefore suggest that the true solution may not be found.

6.3 Errors/Bounds on the Parameter Estimates

An important aspect of geophysical data analysis (or interpretation) is the determination of bounds (or confidence limits) on the various model parameters that are consistent with the experimental data and their associated errors.

6.3.1 Parameter Covariance matrix

The simplest form of model error estimation is the determination of the limits of the parameters from the Covariance matrix, $Cov(m)$. The Covariance matrix depends on the covariance of the experimental errors and the way in which we map the data errors into parameter errors (Menke,1984). For illustration, let us express the estimated least squares solution as

$$m^{est} = (G^T G)^{-1} G^T d = Ld$$

where L is the generalized inverse used in the inversion. The above expression shows that m^{est} is a linear transformation of d. The mathematical expectation (E) value of m^{est} is

$$E(\ m^{est}) = E(Ld)\ = LE(d)$$

If the experimental data are uncorrelated and of equal variance σ^2, then (by Law of propagation of errors)

$$Cov(m^{est}) = L[Cov(d)]L^T$$

$$Cov(m^{est}) = \{(G^T G)^{-1} G^T\}\ [\sigma^2 I]\ \{(G^T G)^{-1} G^T\}^T = \sigma^2 (G^T G)^{-1} \qquad (6.4a)$$

since $\{(G^T G)^{-1} G^T\}^T = G(G^T G)^{-1}$.

For the Marquardt-type damped least squares solution,

$$m^{est} = (G^TG + \beta I)^{-1}G^Td = Ld$$

and the covariance matrix is given by

$$Cov(m^{est}) = (G^TG + \beta I)^{-1}G^T[\sigma^2 I]G(G^TG + \beta I)^{-1}$$

$$= \sigma^2(G^TG + \beta I)^{-1}G^TG(G^TG + \beta I)^{-1}\,. \tag{6.4b}$$

Using a similar argument, the covariance matrix for the solution incorporating *a priori* information is identically

$$Cov(m^{est}) = (G^TG + \beta^2D^TD)^{-1}G^T[\sigma^2 I]G\,(G^TG+\beta^2D^TD)^{-1} + (G^TG+\beta^2D^TD)^{-1}\beta D^T[\beta^2 I]^{-1}\beta D\,(G^TG+\beta^2D^TD)^{-1}$$

giving

$$Cov(m^{est}) = (G^TG+\beta^2D^TD)^{-1}\{\sigma^2G^TG + D^TD\}\,(G^TG+\beta^2D^TD)^{-1}. \tag{6.4c}$$

Note, however, that if the elaborate treatment of data errors described in Section 6.1 was effected (see eq. 6.2 & 6.3), then

$$Cov(m^{est}) = (G^TEG + D^TBD)^{-1}(WG)^T\{E\,[\sigma^2 I]\}\,WG(G^TEG + D^TBD)^{-1} + (G^TEG + D^TBD)^{-1}(\beta D)^T\,\{B[B]^{-1}\}\,\beta D\,(G^TEG + D^TBD)^{-1}$$

or simply,

$$Cov(m^{est}) = (G^TEG + D^TBD)^{-1}\{G^TEG + D^TBD\}\,(G^TEG + D^TBD)^{-1}$$

$$= (G^TEG + D^TBD)^{-1}. \tag{6.4d}$$

where the symbols are as previously defined and the covariance matrices of the actual and *a priori* data in this standardized framework are identity matrices.

Having derived working expressions for the covariance matrices, let us try and see what they represent. $Cov(m)$ is a parameter-by-parameter matrix whose i^{th} diagonal element is the statistical variance of the i^{th} parameter m_i, and whose off-diagonal elements, the covariances, indicate the correlations between the model parameters. Large off-diagonal elements Cov_{ij} mean that the i^{th} and j^{th} model parameters are highly correlated. The square roots of the diagonal elements of $Cov(m)$ are generally referred to as the standard deviations of the least squares parameter estimates and may be used to estimate the bounds of the model parameters. Notice that eq. 6.4c reduces to eq.6.4a when β=0 but yields smaller values when $\beta \neq 0$. Note also that the variances and covariances of parameters constrained to be equal to fixed values in biased estimation are effectively zero while the corresponding quantities for the free (i.e., unconstrained) parameters in the problem is reduced.

6.3.2 Extreme parameter sets : Extremal inversion.

We may elect to determine a solution with the maximum tolerable sum of squared residuals. One method of extremal inversion is the **Most Squares Method** of Jackson (1976) in which a value is determined for each parameter which is maximum (or minimum) under the constraint that the misfit of the observed and calculated data is equal to some specified value. The essential feature of the method is that it produces a class of models which defines a zone within which the true solution, if it exists, may be found. The mathematical formulation of this extremal inversion problem is straightforward. We state the problem as follows:

Given an optimal least squares solution to an inverse problem, m with residuals q_{LS}, find (on account of the observational uncertainties) other solutions which fit the data to a specified threshold residual q_T; or equivalently, **extremize the linear objective function $m^T b$ under the constraint**

$$|d - Gm|^2 = q_T \tag{6.5}$$

where the function b is a vector of zeros with the k^{th} element (to be maximized) set equal to unity; i.e., $b^T = [0,...,0,1^k,0,...,0]$ and m is the vector of model parameters. This search for extreme solutions can be effected simply by minimizing the function

$$L = m^T b + \frac{1}{2\mu}\left\{(d-Gm)^T(d-Gm) - q_T\right\} \tag{6.6}$$

where we have introduced the Lagrange multiplier $1/2\mu$. At the extremum, we have that

$$\frac{\partial}{\partial m}\left(m^T b + \frac{1}{2\mu}\left\{m^T G^T G m - 2m^T G^T d + d^T d - q_T\right\}\right) = 0 \tag{6.7}$$

or

$$G^T G m = G^T d - \mu b$$

from which we obtain the most squares solution

$$m_{ms} = [G^T G]^{-1}[G^T d - \mu b] \tag{6.8}$$

Thus the value of $\mu = 0$ corresponds to the unconstrained least squares solution. However, since eq. 6.8 must satisfy eq.6.5, we have that

$$\begin{aligned} q_T &= d^T d - d^T G m - m^T G^T d + m^T G^T G m \\ &= d^T d - d^T G(G^T G)^{-1}(G^T d - \mu b) - (d^T G - \mu b^T)(G^T G)^{-1} G^T d \\ &\quad + (d^T G - \mu b^T)(G^T G)^{-1} G^T G (G^T G)^{-1}(G^T d - \mu b) \end{aligned}$$

which simplifies to

$$q_T = d^T d - d^T G[G^T G]^{-1} G^T d + \mu^2 b^T [G^T G]^{-1} b$$

so that

$$\mu = \left(\frac{q_T - d^T d + d^T G[G^T G]^{-1} G^T d}{b^T [G^T G]^{-1} b} \right)^{1/2} = \pm \left(\frac{q_T - q_{LS}}{b^T [G^T G]^{-1} b} \right)^{1/2} \tag{6.9}$$

where $q_{LS} = (d^T d - d^T G[G^T G]^{-1} G^T d)$ is the sum of squared residuals of the optimal least squares solution. Since there are two solutions for μ for each model parameter, there will be $2p$ solutions (for the p parameters) whenever $q_T > q_{LS}$. Note that setting all the elements of the parameter projection vector b to unity (i.e., $b = [1, 1, \ldots, 1]^T$) will yield two (i.e., *plus* and *minus*) solutions that may be interpreted as the upper and lower solution envelopes of our least squares solution. If the errors on the data are assumed univariant and uncorrelated, it is expected that q_T will have a value close to the number of the data, i.e., $q_T \approx n$.

We can compare directly the most squares and least sqaures solutions. Equation 6.8 can be written as

$$m_{ms} = (G^T G)^{-1} G^T d - \mu (G^T G)^{-1} b$$

$$= m_{LS} \pm \left(\frac{q_{\mathrm{T}} - q_{LS}}{b^{\mathrm{T}}(G^{\mathrm{T}}G)^{-1}b} \right)^{1/2} (G^{\mathrm{T}}G)^{-1}b \tag{6.10}$$

where m_{LS} is the least squares solution. The most squares solution envelope may thus be interpreted as the confidence limits of the least squares solution.

As we saw in Section 4 (Ex. 4.4.3), the unconstrained least squares solution process is unstable in certain situations. The same problem may plague the most squares method (Eq. 6.10) if the matrix $G^{\mathrm{T}}G$ is ill-conditioned. Meju (1994d) suggests that a bound could be placed on the size of the solutions as in the conventional inversion employing smoothness constraints. The constrained problem is defined simply as:

Given an optimal least squares solution to an inverse problem, m with residuals q_{LS}, find (on account of the observational uncertainties) the smoothest solutions (as gauged by the measure $m^{\mathrm{T}}D^{\mathrm{T}}Dm$) which fit the data to a specified threshold residual q_{T}.

Note in this case that there are two types of data under consideration: the actual experimental data and the extraneous constraining data. The threshold residual is therefore conveniently defined as

$$q_{\mathrm{T}} = |\, d - Gm \,|^2 + \beta^2 m^{\mathrm{T}}D^{\mathrm{T}}Dm \tag{6.11}$$

and the objective function is stated as:

$$\mathcal{L} = m^{\mathrm{T}}b + \frac{1}{2\mu}\left\{(d-Gm)^{\mathrm{T}}(d-Gm) + \beta^2 m^{\mathrm{T}}D^{\mathrm{T}}Dm - q_{\mathrm{T}}\right\} \tag{6.12}$$

so that

$$\frac{\partial}{\partial m}\left(m^{\mathrm{T}}b + \frac{1}{2\mu}\left\{m^{\mathrm{T}}G^{\mathrm{T}}Gm - 2m^{\mathrm{T}}G^{\mathrm{T}}d + d^{\mathrm{T}}d + \beta^2 m^{\mathrm{T}}D^{\mathrm{T}}Dm - q_{\mathrm{T}}\right\}\right) = 0$$

or

$$[G^{\mathrm{T}}G + \beta^2 D^{\mathrm{T}}D]m \;=\; G^{\mathrm{T}}d - \mu b \tag{6.13}$$

from which we obtain the **damped most squares solution**

$$m_{dms} = [G^{\mathrm{T}}G + \beta^2 H]^{-1}[G^{\mathrm{T}}d - \mu b] \tag{6.14}$$

where $H = D^TD$. Now, equation 6.14 must satisfy eq. (6.11). Therefore,

$$q_T = d^Td - d^TGm - m^TG^Td + m^TG^TGm + \beta^2 m^THm$$

$$= d^Td - d^TGm - m^TG^Td + m^T\{G^TG + \beta^2H\}m$$

$$= d^Td - d^TG(G^TG+\beta^2H)^{-1}(G^Td-\mu b) - (d^TG-\mu b^T)(G^TG+\beta^2H)^{-1}G^Td$$
$$+ (d^TG-\mu b^T)(G^TG+\beta^2H)^{-1}\{G^TG+\beta^2H\}(G^TG+\beta^2H)^{-1}(G^Td-\mu b)$$

which simplifies to

$$q_T = d^Td - d^TG(G^TG+\beta^2H)^{-1}G^Td + \mu b^T(G^TG+\beta^2H)^{-1}\mu b$$

so that

$$\mu = \pm\left(\frac{q_T - q_{LS}}{b^T[G^TG + \beta^2H]^{-1}b}\right)^{1/2} \tag{6.15}$$

where $q_{LS}= d^Td-d^TG(G^TG+\beta^2H)^{-1}G^Td$. If $D = I$, then $H = I$. Equation (6.14) is a stable inversion formula and the operation is equivalent to adding a positive constant bias to the main diagonal of the matrix G^TG. β may be chosen to be a small number much less than unity ($\ll 1$) and the expected value of q_T is $n-l$, where there are l constraints in the problem.

Note that we can also formulate the problem to preserve any *a priori* parameter estimates as in Section 5.1. In this case the constrained most squares solution is

$$m_{cms} = [G^TG + \beta^2H]^{-1}[G^Td + \beta^2h - \mu b] \tag{6.16}$$

where $H = D^TD$ (cf. Eq. 5.2) and must satisfy the condition:

$$q_T = |d-Gm|^2 + |Dm-h|^2 \tag{6.17}$$

$$= d^Td + \beta^2h^Th - (d^TG+\beta^2h^TD)m - m^T(G^Td+\beta^2D^Th) + m^T(G^TG+\beta^2D^TD)m\ .$$

Using eq. (6.16) in the place of m in eq. (6.17), we have that

$$q_{\mathrm{T}} = \boldsymbol{d}^{\mathrm{T}}\boldsymbol{d} + \beta^2\boldsymbol{h}^{\mathrm{T}}\boldsymbol{h} - (\boldsymbol{d}^{\mathrm{T}}\boldsymbol{G} + \beta^2\boldsymbol{h}^{\mathrm{T}}\boldsymbol{D})(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\beta^2\boldsymbol{H})^{-1}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}+\beta^2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{h}) + \mu\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\beta^2\boldsymbol{H})^{-1}\mu\boldsymbol{b}$$

giving

$$\mu = \pm \left(\frac{q_{\mathrm{T}} - q_{LS}}{\boldsymbol{b}^{\mathrm{T}}[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G} + \beta^2\boldsymbol{H}]^{-1}\boldsymbol{b}} \right)^{1/2} \tag{6.18}$$

where $q_{LS}= \boldsymbol{d}^{\mathrm{T}}\boldsymbol{d} + \beta^2\boldsymbol{h}^{\mathrm{T}}\boldsymbol{h} - (\boldsymbol{d}^{\mathrm{T}}\boldsymbol{G}+\beta^2\boldsymbol{h}^{\mathrm{T}}\boldsymbol{D})(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\beta^2\boldsymbol{H})^{-1}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{d}+\beta^2\boldsymbol{D}^{\mathrm{T}}\boldsymbol{h})$.
The threshold residual may be set to a value close to $n-l$ for a problem with l constraints.

6.4 Example of inversion and detailed error analysis: a recommended practice

As an illustration of a good inversion practice, let us analyse a simple linear problem (see also, Meju 1994c).

Problem: Given the following $\{x,y\}$ data pairs (Jackson,1976) for a straight-line inverse problem:

x	y
-1.000000	-1.124600
-8.000000E-01	7.080000E-02
-6.000000E-01	-9.942000E-01
-4.000000E-01	-7.038000E-01
-2.000000E-01	9.637000E-01
0.000000E+00	5.810000E-02
2.000000E-01	-7.820000E-02
4.000000E-01	-1.069000E-01
6.000000E-01	-9.231000E-01
8.000000E-01	-7.819000E-01
1.000000	-4.250000E-02

determine the least solution for the slope and intercept, compute the residuals, the resolution and covariance matrices and the bounding models.

Solution: There are two model parameters in this problem: m_1 (intercept) and m_2 (slope). Using the procedure outlined for fitting straight lines to x,y data pairs, we will obtain the following unconstrained least squares estimates:

$$\boxed{m_1 = \text{-3.329636E-01}, \quad m_2 = \text{1.074954E-01}}.$$

The computed data misfit, $q = |\boldsymbol{d}-\boldsymbol{Gm}|^2 = 3.898074$.

The resolution information is simply:

[1.000000 -1.594859E-15]$\longleftarrow$row$_1$ of $\boldsymbol{R}$

[-1.594859E-15 1.000000]$\longleftarrow$row$_2$

To calculate $\boldsymbol{Cov}(\boldsymbol{m})$, we will assume that the experimental errors in the data are Gaussian, statistically independent, of zero mean, and of unit variance.

The resulting covariance information is simply:

[9.090909E-02 -3.624680E-16]$\longleftarrow$row$_1$ of $\boldsymbol{Cov}(\boldsymbol{m})$

[-1.449872E-16 2.272727E-01]$\longleftarrow$row$_2$

The square roots of the diagonal elements of $\boldsymbol{Cov}(\boldsymbol{m})$ are the standard deviations of the estimates, viz:

$\boxed{\sigma_{m_1} = \pm 3.015113\text{E-}01, \quad \sigma_{m_2} = \pm 4.767313\text{E-}01}$.

The most-squares extreme parameter sets are:

(a) plus solutions (μ positive)

[4.705472E-01 1.074954E-01]$\longleftarrow m_1$ extremized using $\boldsymbol{b} = [1,0]^T$

[-3.329636E-01 1.377958]$\longleftarrow m_2$ extremized using $\boldsymbol{b} = [0,1]^T$

(b) minus solutions (μ negative)

[-1.136474 1.074954E-01]$\longleftarrow m_1$ extremized using $\boldsymbol{b} = [1,0]^T$

[-3.329636E-01 -1.162967]$\longleftarrow m_2$ extremized using $\boldsymbol{b} = [0,1]^T$

The most-squares solution envelopes are:

(a) plus solution

[9.653091E-02 1.181232]$\longleftarrow$upper envelope: $\boldsymbol{m}$ extremized using $\boldsymbol{b} = [1,1]^T$

(b) minus solution

[-7.624581E-01 -9.662411E-01]$\longleftarrow$lower envelope

The threshold residual, q_T was set to 11 (i.e., n) in these calculations. Notice that the range of parameters provided by the most-squares method for the data are greater than that indicated by the standard deviations of the least squares estimates. All the above results were obtained using the inversion program SVDINV (which incorporates the subroutine MOSTSQ) and are in good agreement with those given in Jackson(1976). This concludes our discussion of linear inversion theory and practice.