XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6–11, 2009, Lisbon, Portugal

ADJUSTMENT OF A NETWORK OF FUNDAMENTAL CONSTANTS

A B Forbes

National Physical Laboratory, Teddington, UK, alistair.forbes@npl.co.uk

Abstract – This paper considers the analysis of observational data associated with the fundamental constants, in particular looking at issues of consistency and sensitivity.

Keywords: least squares adjustment, uncertainty

1. INTRODUCTION

An important interaction between metrology and the physical sciences in general is in the determination of the fundamental constants. From the metrology perspective, the goal of relating all the main units of measurement to fundamental quantities is within reach, although there is still debate about when and how the SI will be reformulated along those lines [1,2]. Irrespective of how these changes will be implemented, there is the issue of how the body of knowledge concerning the fundamental constants is assembled, analysed and the results disseminated to the science community. Most notable in this regard is the work of CODATA [3,4,5].

The body of knowledge of a network of fundamental constants comprises measurement data and physical theory relating the values of subsets of the fundamental constants to each other. Often a measurement does not directly provide an estimate of a constant but instead that of a quantity, related through theory, in a linear or nonlinear way, to a number of constants. A particular constant may figure in several experiments associated with different aspects of the physical theory. Each such experiment provides information about the constant, and an adjustment of the value of that constant should take into account all relevant experimental evidence.

Currently, exercises such as those performed by CODATA provide a simultaneous adjustment through solving a large nonlinear least squares problem involving a comprehensive set of measurement data to provide estimated values of the constants and an associated uncertainty matrix giving variances and covariances associated with the estimates.

In implementing such an adjustment, the following questions arise:

- How do we assess if the data and physical theory are consistent with each other?
- If there is inconsistency, is there enough information (e.g., through multiple experiments) to indicate where it arises?
- By how much can the input data be changed and still have consistency?
- How can the adjustment model be used to predict the likely impact of a proposed experiment or judge which of two competing experiments should be performed?
- How important is a given theoretical relationship and does the observational data support the theory?

This paper addresses some of these questions, using methods that require no special techniques beyond those used in an adjustment process but often ignored in an initial analysis of a set of observational data. In section 2, the main analysis methods associated with least squares adjustment applied to a network of fundamental constants are described, including those pertaining to changing which constants are to be considered exact. In section 3, these techniques are applied to a very simple but relevant network of constants. Our concluding remarks are given in section 4.

2. LEAST SQUARES ADJUSTMENT

It is assumed that the observational data takes the form of a vector $\mathbf{y} = (y_1,\dots,y_m)^{\mathrm{T}}$ and that each measured value y_i is an estimate of some function $\phi_i(\boldsymbol{\alpha})$ of a subset of the constants $\boldsymbol{\alpha} = (\alpha_1,\dots,\alpha_n)^{\mathrm{T}}$, $m \geq n$, and has an associated standard uncertainty $u(y_i)$. Assuming statistical independence of the measurements, appropriate estimates $\mathbf{a} = (a_1,\dots,a_n)^{\mathrm{T}}$ of the constants are found by minimising

$$F(\boldsymbol{\alpha}) = \sum_{i=1}^{m} w_i^2 (y_i - \phi_i(\boldsymbol{\alpha}))^2, \quad w_i = 1/u(y_i), \qquad (1)$$

with respect to $\boldsymbol{\alpha}$.

If $f_i = f_i(\boldsymbol{\alpha}) = w_i\,(y_i - \phi_i(\boldsymbol{\alpha}))$ and J is the Jacobian matrix with $J_{ij} = \partial f_i / \partial \alpha_j$ evaluated at the solution $\mathbf{a}$, then the variance matrix associated with the fitted parameters is given by $V_{\mathbf{a}} = \left(J^T J\right)^{-1}$. For the case in which there is correlation associated with the observations, the diagonal weighting matrix is replaced by the inverse of the variance matrix associated with the observations. We do not consider this case further here, but the analysis discussed below applies to this case also.

ISBN 978-963-88410-0-1

2.1 Observation equations for fundamental constants

The observation equations for adjustment of a network of fundamental constants generally have the form

$$y_i = \alpha_1^{c_{i1}} \alpha_2^{c_{i2}} \cdots \alpha_n^{c_{in}} + e_i, \quad e_i \in \mathrm{N}(0, u^2(y_i)),$$

where the indices c_{ij} are integer or half integer powers, most of which are zero. It is usually convenient to reparametrize the quantities, replacing α_j by $\alpha_{j,0}(1+\alpha_j)$, where $\alpha_{j,0}$ is the current or nominal value of a quantity and α_j now represents the relative adjustment. Similarly, the observations can be transformed so that y_i is replaced by $(y_i - \phi_i(\boldsymbol{\alpha}_o))/\phi_i(\boldsymbol{\alpha}_o)$, the relative difference between the observed value y_i and the (current) model prediction $\phi_i(\boldsymbol{\alpha}_o)$. The associated standard uncertainty is essentially the relative standard uncertainty associated with the observation. With these transformations, to first order the observation equations can be written as

$$y_i = \mathbf{c}_i^{\mathrm{T}} \boldsymbol{\alpha} + e_i, \quad e_i \in \mathrm{N}(0, u^2(y_i)),$$

with $\mathbf{c}_i^{\mathrm{T}} = (c_{i1}, \ldots, c_{in})$. We make one further weighting transformation, replacing y_i by $y_i / u(y_i)$ and $\mathbf{c}_i$ by $\mathbf{c}_i / u(y_i)$, so that the final form of the observation equations (to first order) can be written as

$$\mathbf{y} \in \mathrm{N}(C\boldsymbol{\alpha}, I). \tag{2}$$

The (transformed) observation matrix C will be sparse with only a few nonzero elements in each row. In each row, the nonzero elements will be integer multiples of a common factor. Treating the observations as nonlinear, the associated Jacobian matrix will be very similar to C. For the rest of the paper, we will only be concerned with the linearised equations. If the relative standard uncertainties associated with the observations are of the order of one part in 10^6, then second order effects will only become apparent at the one part in 10^{12} level. With this assumption, the analysis for the linear case will provide a good approximation to the nonlinear case.

2.2 Analysis of fitted results

The analysis of the fitted results of an adjustment exercise follows from the standard theory for least squares adjustment. The least squares estimate $\mathbf{a}$ of $\boldsymbol{\alpha}$ is given by

$$\mathbf{a} = S\mathbf{y}, \quad S = \left(C^{\mathrm{T}}C\right)^{-1} C^{\mathrm{T}}.$$

The $n \times m$ matrix S is the sensitivity matrix of $\mathbf{a}$ with respect to $\mathbf{y}$, and specifies the least squares solution as a linear combination of the observations. The matrix S is the pseudo-inverse of C, also known as the Moore-Penrose inverse [6], since $SC = I$. If C has QR factorisation [6]

$$C = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ \mathbf{0} \end{bmatrix},$$

where Q is orthogonal and R is upper-triangular, then $S = R_1^{-1} Q_1^{\mathrm{T}}$.

The variance matrix associated with $\mathbf{y}$ (in our case, the identity matrix since the observations and observation matrix already incorporate weights) is propagated through to that $V_{\mathbf{a}}$ associated with $\mathbf{a}$:

$$V_{\mathbf{a}} = SIS^{\mathrm{T}} = \left(C^{\mathrm{T}}C\right)^{-1} = \left(R_1^{\mathrm{T}} R_1\right)^{-1}.$$

The variance associated with the jth parameter is the sum of the squares of elements of the jth row of S so that S specifies the uncertainty contribution from each of the observations to each of the fitted parameters.

A measure of the total variation associated with the fitted parameters is given by the trace($V_{\mathbf{a}}$), the sum of the diagonal elements of $V_{\mathbf{a}}$, also the sum of the eigenvalues of $V_{\mathbf{a}}$. It is easy to check that

$$\mathrm{trace}(V_{\mathbf{a}}) = \sum_{i=1}^{m} \sum_{j=1}^{n} s_{ji}^2,$$

so that the sum of squares of the elements in the ith column of S is the contribution of the ith observation to the total variation in the fitted parameters.

Setting

$$\hat{\mathbf{y}} = C\mathbf{a} = CS\mathbf{y} = Q_1 R_1 R_1^{-1} Q_1^{\mathrm{T}} \mathbf{y} = Q_1 Q_1^{\mathrm{T}} \mathbf{y},$$

the fitted model values of the relative adjustment, the associated variance matrix $V_{\hat{\mathbf{y}}}$ is given by

$$V_{\hat{\mathbf{y}}} = CV_{\mathbf{a}}C^{\mathrm{T}} = Q_1 R_1 \left(R_1^{\mathrm{T}} R\right)^{-1} R_1^{\mathrm{T}} Q_1^{\mathrm{T}} = Q_1 Q_1^{\mathrm{T}}.$$

Thus, the standard uncertainty $u(\hat{y}_i)$ associated with the ith model prediction is the norm of the ith row of the matrix Q_1. Since the norm of the complete ith row of the orthogonal matrix $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ is 1, we have $0 \le u(\hat{y}_i) \le 1 = u(y_i)$.

The residual vector $\mathbf{r}$ is given by

$$\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - C\mathbf{a} = (I - CS)\mathbf{y} = (I - Q_1 Q_1^{\mathrm{T}})\mathbf{y} = Q_2 Q_2^{\mathrm{T}} \mathbf{y},$$

and the associated variance matrix is $V_{\mathbf{r}} = Q_2 Q_2^{\mathrm{T}}$. It follows that

$$u^2(\hat{y}_i) + u^2(r_i) = 1, \quad 0 \le u(\hat{y}_i), u(r_i) \le 1.$$

If $u(r_i) = 0$, it means that the ith model prediction must match the ith observation exactly, in other words, the ith model prediction is determined by the ith observation. Conversely, if $u(r_i) = 1$, it means that the ith observation plays no part in determining the ith model prediction; it is determined using other information.

2.3 The chi squared test

If $\mathbf{r} \in \mathrm{N}\left(\mathbf{0}, Q_2 Q_2^{\mathrm{T}}\right)$, then $Q_2^{\mathrm{T}} \mathbf{r} \in \mathrm{N}\left(\mathbf{0}, I_{m-n}\right)$, and

$$\mathbf{r}^{\mathrm{T}} \mathbf{r} = \left(Q_2^{\mathrm{T}} \mathbf{r}\right)^{\mathrm{T}} Q_2^{\mathrm{T}} \mathbf{r}$$

is a draw from a χ^2_{m-n} distribution with $m-n$ degrees of freedom. The observed value can be compared with quantiles of χ^2_{m-n} as a test of model-data consistency. Note that

$$\sum_{i=1}^{m} u^2(r_i) = \text{trace}(Q_2 Q_2^{\mathrm{T}}) = m - n,$$

the sum of the eigenvalues of $Q_2 Q_2^{\mathrm{T}}$. The expected value of a draw from a χ^2_{m-n} distribution is $m-n$ and the $u(r_i)$ specify how the sum of squares is expected to be partitioned across the observations.

2.4 Fixed values for some quantities

In order to define the SI, some quantities are to be regarded as fixed, with no associated uncertainty. For example, in the current SI, the magnetic constant μ_0 is treated as exact. We assume that such exactness constraints can be written as $D^{\mathrm{T}}\boldsymbol{\alpha} = \mathbf{z}$, where D is a $n \times p$ constraint matrix and $\mathbf{z}$ a p-vector of constrained values. If D has QR factorisation

$$D = UT = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} T_1 \\ \mathbf{0} \end{bmatrix}, \tag{3}$$

and $\boldsymbol{\alpha}_o$ satisfies the constraints, then the constrained adjustment problem can be written as an unconstrained problem

$$\min_{\tilde{\boldsymbol{\alpha}}} \left\| y - CU_2 \tilde{\boldsymbol{\alpha}} \right\|_2^2,$$

involving the $n-p$ parameter vector $\tilde{\boldsymbol{\alpha}}$. The unconstrained parameters are given by $\boldsymbol{\alpha} = U_2 \tilde{\boldsymbol{\alpha}}$. The analysis above is easily modified to account for linear constraints.

2.5 Continuity constraints on redefinition

A possible consequence of a redefinition of the SI is that some quantities regarded as fixed become free and vice versa. In the context of the previous section, the constraint equations change. In going from one definition to another, there could be a change in the values of the constants. There is some scope for controlling that change by the choice of the constrained values. If there are p constraints, up to p linear combinations of the solution parameters can be specified. The problem can be posed as: determine $\mathbf{z}$ such that the solution $\mathbf{a}$ of

$$\min_{\boldsymbol{\alpha}} \left\| \mathbf{y} - C\boldsymbol{\alpha} \right\|_2^2 \quad \text{subject to} \quad D^{\mathrm{T}}\boldsymbol{\alpha} = \mathbf{z},$$

satisfies $F^{\mathrm{T}}\boldsymbol{\alpha} = \mathbf{g}$, where F and $\mathbf{g}$ are a pre-assigned $n \times p$ matrix and p-vector, respectively. Let U_2 be defined as in (3), and

$$CU_2 = \begin{bmatrix} P_1 & P_2 \end{bmatrix} \begin{bmatrix} H_1 \\ \mathbf{0} \end{bmatrix},$$

be the QR factorisation of CU_2. Any $\boldsymbol{\alpha}$ satisfying the constraint can be written as $\boldsymbol{\alpha} = U_1 T_1^{-\mathrm{T}} \mathbf{z} + U_2 \tilde{\boldsymbol{\alpha}}$ for some $\tilde{\boldsymbol{\alpha}}$. It can be checked that the required $\mathbf{z}$ satisfies

$$\left(F^{\mathrm{T}} - F^{\mathrm{T}} U_2 \tilde{S} C \right) U_1 T_1^{-\mathrm{T}} \mathbf{z} = \mathbf{g} - F^{\mathrm{T}} U_2 \tilde{S} \mathbf{y}, \quad \tilde{S} = H_1^{-1} P_1^{\mathrm{T}}.$$

Similarly, $\mathbf{z}$ can be chosen so that the solution of the constrained optimisation problem minimises $\left\| F^{\mathrm{T}}\boldsymbol{\alpha} - \mathbf{g} \right\|_2^2$ and can be used to minimise some measure of the size of the relative adjustment on implementing a change in the SI.

3. EXAMPLE CALCULATIONS

Analysis methods described above are illustrated in the following simple example involving six observations, five parameters and two constraints. Although the example is extremely simple, it is motivated by a study of issues associated with the redefinition of the SI [7]. The unweighted observation equations have the form

$$\begin{aligned} y_1 &= -\alpha_3 + \alpha_4, \\ y_2 &= -\alpha_1 + \alpha_2 + \alpha_4/2 + \alpha_5/2, \\ y_3 &= -\alpha_1 + \alpha_4, \\ y_4 &= \alpha_1 - 2\alpha_2 - \alpha_5, \\ y_5 &= \alpha_1 - \alpha_3, \\ y_6 &= \alpha_1 - 2\alpha_2 - \alpha_5, \end{aligned}$$

with relative uncertainties of

$$(10.0, 10.0, 1.0, 1.0, 0.1, 0.1) \times 10^{-8}.$$

In the first scenario, the fourth and fifth parameters are considered exact. In the second, the first and second parameters are treated as exact.

For the first scenario, labelled A, the 3×6 sensitivity matrix is

$$S_{\mathrm{A}} = \begin{bmatrix} -0.10 & -0.05 & -0.99 & 0.00 & 0.00 & 0.00 \\ -0.05 & -0.02 & -0.49 & 0.00 & 0.00 & -0.05 \\ -0.10 & -0.05 & -0.99 & 0.00 & -0.10 & 0.00 \end{bmatrix} \times 10^{-8}.$$

The rows correspond to parameters 1, 2 and 3, respectively. The sensitivity matrix immediately tells us that the main contribution to the uncertainties associated with the fitted parameters is that associated with the third observation. If uncertainty associated with y_3 could be halved, then the uncertainties associated with all the fitted parameters would also be halved. By the same argument, reducing the uncertainty associated with any or all of the other observations will have only a marginal effect on the uncertainties associated with the fitted parameters.

The vector of standard uncertainties $u(\mathbf{r})$ associated with the vector $\mathbf{r}$ of residuals is given by

$$u(\mathbf{r}) = \begin{pmatrix} 1.00 & 1.00 & 0.11 & 1.00 & 0.01 & 0.10 \end{pmatrix}^{\mathrm{T}},$$

showing that only observations 3, 5 and 6 are influential; all the other observations are essentially ignored. These uncertainties also provide information about the ability to detect discrepant data. Discrepant values for observations 1 2 and 4 will be detected through large evaluated residuals. Discrepant values associated with observations 3, 5 and 6 will not be detected, as the fit will accommodate them.

In scenario B, the sensitivity matrix is

$$S_{\mathrm{B}} = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & -0.10 & 0.00 \\ 0.10 & 0.05 & 0.99 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & -0.01 & 0.00 & -0.10 \end{bmatrix} \times 10^{-8}.$$

Here, the rows correspond to parameters 3, 4 and 5. In this case, the uncertainties associated with parameters α_3, α_4, and α_5 are dominated by the uncertainties associated with observations 5, 3 and 6, respectively. The uncertainties associated with the residuals are the same as for scenario A; only observations 3, 5 and 6 are influential. The fact that the sensitivity matrices are different points to the different roles of the observations in the two scenarios. Note that an observation can be influential without contributing significantly to the sensitivities. In can be argued that scenario B is better in that the sum of the variances, $\mathrm{trace}(V_{\mathbf{a}})$, is less than half that for scenario A.

We now consider a revised set of observation equations, encoding doubt about the physics associated with observations 2, 3 and 4, and examining the potential value of a new experiment aimed at eliminating that doubt. The revised set of observation equations is

$$
\begin{aligned}
y_1 &= -\alpha_3 + \alpha_4, \\
y_2 &= -\alpha_1 + \alpha_2 + \alpha_4/2 + \alpha_5/2 + \alpha_6, \\
y_3 &= -\alpha_1 + \alpha_4 + 2\alpha_6 + \alpha_7, \\
y_4 &= \alpha_1 - 2\alpha_2 - \alpha_5 + \alpha_7, \\
y_5 &= \alpha_1 - \alpha_3, \\
y_6 &= \alpha_1 - 2\alpha_2 - \alpha_5, \\
y_7 &= \alpha_6 + \alpha_7,
\end{aligned}
$$

involving a sixth and seventh parameter and one new (proposed) observation.

S_{A}	y_1	y_2	y_3	y_4	y_5	y_6	y_7
α_1	-8.01	0.00	-0.20	-0.20	0.08	0.02	3.98
α_2	-4.00	0.00	-0.20	-0.10	0.04	-0.04	1.99
α_3	-8.01	0.00	-0.20	-0.20	-0.02	0.02	3.98
α_6	-4.00	0.05	0.40	-0.60	0.04	0.06	1.99
α_7	0.02	-0.05	0.00	1.00	0.00	-0.10	0.04

S_{A}	y_1	y_2	y_3	y_4	y_5	y_6	y_7
α_1	-0.57	0.00	-0.94	-0.94	0.01	0.09	1.89
α_2	-0.28	0.00	-0.47	-0.47	0.00	0.00	0.94
α_3	-0.57	0.00	-0.94	-0.94	-0.09	0.09	1.89
α_6	-0.28	0.05	0.03	-0.97	0.00	0.10	0.94
α_7	0.10	-0.05	-0.01	0.99	0.00	-0.10	0.02

S_{A}	y_1	y_2	y_3	y_4	y_5	y_6	y_7
α_1	-0.20	0.00	-0.98	-0.98	0.00	0.10	0.20
α_2	-0.10	0.00	-0.49	-0.49	0.00	0.00	0.10
α_3	-0.20	0.00	-0.98	-0.98	-0.10	0.10	0.20
α_6	-0.10	0.05	0.01	-0.99	0.00	0.10	0.10
α_7	0.10	-0.05	-0.01	0.99	0.00	-0.10	0.00

Table 1. Sensitivity matrix for scenario A with $u(y_7) = 10^{-7}$, 10^{-8} and 10^{-9}, top, middle and bottom, respectively.

We calculate, in Table 1, the sensitivity matrix in scenario A, for parameters 1, 2, 3, 6 and 6 (with the fourth and fifth parameter treated as exact) for $u(y_7) = 10^{-7}, 10^{-8}$ and 10^{-9}. The Table shows that increasing the accuracy associated with observation seven means that the main contribution to the sensitivity of the parameters is moved from observation one to the more accurate observations three and four. The calculations show that observation seven is only taken into account only if its standard uncertainty is of the order of one part in 10^7. Additionally, increasing its accuracy beyond one part in 10^8 has only a marginal effect on the standard uncertainties standard uncertainty associated with the fitted parameters.

CONCLUSIONS

We have reviewed some of the analysis techniques that can be applied to least squares adjustment. Often, the focus of attention is on the values of the fitted parameters and the associated variance matrix. However, the sensitivity matrix contains valuable information about the dependence of the fitted parameters on the observations. In the first example considered in section 3, the sensitivity matrix immediately pointed to the relative importance of the various observations. In scenario A, we saw that in order to improve the uncertainties associated with the fitted parameters, only the third observation is important; improving the uncertainties associated with any of the other observations would have only a marginal effect. In this way the sensitivity matrix can give a strong guide to how measurement resources should be applied in order to have maximal impact. In the second example, the sensitivity matrix was used to show the potential impact of a new proposed experiment. In a similar vein, the uncertainties associated with the residuals give valuable information about the ability to detect discrepant data.

ACKNOWLEDGMENTS

The work described in this paper was supported by the National Measurement System Directorate of the UK Department for Innovation, Universities and Skills as part of the NMS Software Support for Metrology programme. The work was carried out as part of a study on the redefinition of the SI [7] and the author gratefully acknowledges the contribution of colleagues Martin Milton and Jonathan Williams in the discussions leading to this paper. Ian Smith made valuable comments on an earlier draft.

REFERENCES

[1] I M Mills, P J Mohr, T J Quinn, B N Taylor, and E R Williams: "Redefinition of the kilogram, ampere, kelvin and mole; a proposed approach to implementing CIPM recommendation 1 (CI-2005)", *Metrologia*, **43**, pp 227-246, 2006

[2] M J T Milton, J M Williams and S J Bennett, "Modernizing the SI: towards an improved, accessible and enduring system", *Metrologia*, **44**, pp 356-364, 2007

[3] P J Mohr and B N Taylor, "CODATA recommended values of the fundamental physical constants: 1998", *Rev Mod Phys*, **72**, pp 351-495, 2000

[4] P J Mohr and B N Taylor, "CODATA recommended values of the fundamental physical constants: 2002", *Rev Mod Phys*, **77**, pp 1-107, 2005

[5] P J Mohr and B N Taylor, "CODATA recommended values of the fundamental physical constants: 2006", *Rev Mod Phys*, **80**, pp 633-730, 2008

[6] G H Golub and C F Van Loan, *Matrix Computations*, John Hopkins, Baltimore, 3rd edition, 1996

[7] M J T Milton, J M Williams and A B Forbes, "The quantum metrology triangle and the re-definition of the SI ampere and kilogram; analysis of a reduced set of observational equations" http://arxiv.org/abs/0905.3635