NPL Report CMSC 01/00

Report to the National Measurement System Policy Unit, Department of Trade and Industry

From the Software Support for Metrology Programme

Simulated Instruments and Uncertainty Estimation

By
A B Forbes and P M Harris

March 2000

Simulated Instruments and Uncertainty Estimation

A B Forbes and P M Harris
Centre for Mathematics and Scientific Computing

March 2000

ABSTRACT

Many metrology data analysis problems can be posed in terms of estimation: given measurement data X and a mathematical model of the experimental system described in terms of parameters $\mathbf{a}$, find the values $\mathbf{a}^*$ of $\mathbf{a}$ that best explain the data. This is usually achieved by formulating an error function and determining the values of the parameters that minimises this function. While $\mathbf{a}$ may describe the parameters of interest in the experiment, there are usually other parameters $\mathbf{b}$, often treated as constants, that define the configuration of the system. In general, the prior information $\mathbf{b}$ is not known exactly and its associated uncertainty will need to be taken into account in deriving an uncertainty statement for the estimates $\mathbf{a}^*$. In recent years, a number of laboratories have developed approaches based on numerical simulation to estimate uncertainties associated with co-ordinate measuring machine (CMM) measurements. A Virtual CMM attempts to model and simulate all the main sources of measurement error and, in particular, how the uncertainty and drift in the calibration parameters contribute to the uncertainty in assessing the quality of manufactured parts. This report discusses how the Virtual CMM concept can be generalised to models of other measuring instruments and provides a general methodology to take into account prior calibration information in uncertainty estimation. This report is a deliverable of the *Validation of Simulated Instruments* project within the NMS Software Support for Metrology Programme, 1998–2001.

1361–407X

Authorised by Dr Dave Rayner,
Head of the Centre for Mathematics and Scientific Computing

National Physical Laboratory,
Queens Road, Teddington, Middlesex, United Kingdom TW11 0LW

Contents

1 Introduction

Many measuring instruments undergo periodic calibration, a process that usually involves using the instrument to measure a number of calibrated artefacts (traceable to national standards) and, from these measurements and the calibration information, deriving a calibrated model of the instrument response. This response model relates the output from the instrument to units traceable to standards, for example, converting a voltage to a temperature measurement. The response y is modelled as a function $y = \phi(\mathbf{x}, \mathbf{b})$ of the variables $\mathbf{x}$ and calibration parameters $\mathbf{b} = (b_1, \ldots, b_n)$, for example, the coefficients of a polynomial representing a calibration curve. These calibration parameters, determined from measurements, will have an associated uncertainty. Often, when the instrument is re-calibrated, there is a significant change in the calibration parameters as behaviour of the instrument drifts with time (due to wear, etc.). Both these effects need to be taken into account when evaluating the uncertainty associated with the measurements produced by the instrument.

In recent years, a number of laboratories, including NPL, PTB and IMGC, have developed approaches based on numerical simulation to estimate uncertainties associated with co-ordinate measuring machine (CMM) measurements. A Virtual CMM attempts to model and simulate all the main sources of measurement error and, in particular, how the uncertainty and drift in the calibration parameters contribute to the uncertainty in assessing the quality of manufactured parts, such as the circularity of a cylindrical shaft. This report discusses how the Virtual CMM concept can be generalised to models of other measuring instruments and provides a general methodology to take into account prior calibration information in uncertainty estimation.

2 The Virtual CMM

2.1 Geometric tolerance assessment in co-ordinate metrology

The co-ordinate measuring machine (CMM) is an important instrument in dimensional metrology and industrial inspection. It can be used for a wide range of measurement tasks that can be defined in terms of the co-ordinates of points relating to a workpiece surface. In geometric tolerance assessment using CMMs, the CMM is used to gather data points $X = \{\mathbf{x}_i = (x_i, y_i, z_i)^{\mathrm{T}}\}_1^m$ related to the surface of the workpiece $\mathcal{W}$ so that (x_i, y_i, z_i) represents the Cartesian co-ordinates of a point in 3-dimensional space in a fixed frame of reference. Tolerance assessment software [10, 3, 1] is then

applied to the data points X to determine whether the workpiece has been manufactured to meet its design specification. Usually, the ideal workpiece is described as a mathematical surface $W(\mathbf{a})$ specified by parameters $\mathbf{a}$. The values of $\mathbf{a} = \mathbf{a}(X)$ which produce the best-fit surface to X are calculated and can then be used to test whether the part is in tolerance. Thus, the execution of a measurement task using a CMM is the two stage process

$$\mathcal{W} \longrightarrow X \longrightarrow \mathbf{a}(X).$$

For example, $\mathcal{W}$ may be a nominally cylindrical artefact, $\mathbf{a} = \mathbf{a}(X)$ the estimate of the parameters (specifying a point $\mathbf{x}_0(\mathbf{a})$ on the cylinder axis, the axis direction $\mathbf{n}(\mathbf{a})$ and radius $r(\mathbf{a})$) of the best-fit substitute cylinder to the data X). The estimated diameter $d = 2r(\mathbf{a})$ can then be compared with the dimensional tolerance on the diameter.

In order for the CMM to be fully integrated into an accredited measurement system, the measurements performed using a CMM have to be traceable to the metre. This implicitly requires that an uncertainty statement, derived according to accepted procedures such as the ISO *Guide to the Expression of Uncertainty of Measurement* [15], can be given for each measurement task. For example, given an estimate of a workpiece diameter $d = d(\mathbf{a})$, what uncertainty can we associate with d? These estimates of uncertainty are also precisely what is required by a user in order to determine whether a particular CMM will measure the parts to the accuracy required.

However, a major difficulty with task-specific measurement uncertainties is that the raw CMM measurements $X = \{\mathbf{x}_i\}$ alone do not provide information from which uncertainties can readily be deduced. Even for relatively simple tasks such as measuring the diameter of a cylindrical workpiece, there are a large number of factors that need to be taken into account in producing such an uncertainty statement. The *Virtual CMM* (VCMM) has been developed [21, 5, 7] as a tool with which to estimate these task-specific uncertainties. The VCMM attempts to simulate all the main components of a CMM's behaviour, including the geometrical and probing errors, etc. Using the VCMM in numerical simulations, it is possible in principle to analyse its behaviour for a particular measurement task and from this determine the estimates of associated uncertainties.

2.2 Mathematical models of CMM behaviour

For a conventional CMM, the probe (usually a small sphere that makes contact with the artefact) is carried on a system of three nominally orthogonal axes on which are mounted linear scales. By measuring how far the probe has moved along each axis, it is possible to estimate the nominal position of

the probe. However, due to manufacturing imperfection, the motion along each axis will have scale, straightness and rotational errors so that there is a systematic and repeatable departure of the actual probe location from that deduced from the scale readings. We model this as

$$\mathbf{x}^* = \mathbf{x} + \mathbf{e}(\mathbf{x}, \mathbf{b}, \mathbf{p}) + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{x}^*$ is the true location of the probe, and $\mathbf{x} = (x, y, z)^{\mathrm{T}}$ are the scale values. The function

$$\mathbf{e}(\mathbf{x}, \mathbf{b}, \mathbf{p}) = \mathbf{e}_0(\mathbf{x}, \mathbf{b}) + R(\mathbf{x}, \mathbf{b})\mathbf{p} \tag{2}$$

describes the systematic error in terms of a positional term $\mathbf{e}_0$, a rotation matrix R specified by three rotation angles and a probe offset $\mathbf{p}$ and $\boldsymbol{\epsilon}$ represents random, non-systematic error. The components $\mathbf{e}_0$ and R are each a function of the scale values $\mathbf{x}$ and model parameters $\mathbf{b}$, and different types of behaviour can be modelled by choosing different functions for $\mathbf{e}_0$ and R.

In a completely general model, the six component functions are represented by empirical functions of the scale readings $\mathbf{x}$ such as multivariate polynomials or tensor product splines. We term this an *empirical model*. In the *kinematic model* of CMM behaviour (e.g., [22]), it is assumed that the probe location is built up from the behaviour along each axis. For example, if

$$\mathbf{x}_x = (x, 0, 0)^{\mathrm{T}} + \mathbf{e}_{0,x}(x, \mathbf{b})$$

and $R_x(x, \mathbf{b})$ describe the three positional and three rotational terms of the error behaviour along the x-axis as a function of the x-scale value, with y- and z-motion described similarly, the overall motion can be described by

$$\mathbf{x}^* = \mathbf{x}_x(x, \mathbf{b}) + R_x(x, \mathbf{b}) \left\{ \mathbf{x}_y(y, \mathbf{b}) + R_y(y, \mathbf{b}) \left[\mathbf{x}_z(z, \mathbf{b}) + R_z(z, \mathbf{b})\mathbf{p} \right] \right\} + \boldsymbol{\epsilon}, \tag{3}$$

or similar, the exact representation depending on the CMM architecture. The 18 error terms are each a function of a single variable and can be represented using polynomials or splines, for example. Often the kinematic model is used in a linearised form with all higher order terms ignored.

The systematic errors described above concern only the geometric behaviour of the CMM. In practice, the probe mechanism has to be regarded as another measuring system that exhibits significant systematic departure from nominal behaviour and appropriate models included. For our purposes here, we regard the model simply as

$$\mathbf{x}^* = \mathbf{x} + \mathbf{e}(\mathbf{x}, \mathbf{b}) + \boldsymbol{\epsilon}.$$

and refer to $\mathbf{e}(\mathbf{x}, \mathbf{b})$ as the *parametric errors*.

2.3 Estimating the error model parameters

To determine the parametric errors of a CMM, one can design experiments to measure each of the error functions individually and then amalgamate the results to determine the composite error. A second approach is to use the CMM to measure one or more calibrated, dimensionally stable, artefacts in different locations and orientations and determine the errors from these measurements and calibration information.

As an example of the second approach, suppose a ball plate is placed in a number of positions within the working volume of a CMM and the scale readings corresponding to measurements of the spheres are recorded. Let $\mathbf{x}_i$ be the scale readings corresponding to an estimate of the centre of the jth sphere in the kth position of the plate using the lth probe. This measurement information is described by three model equations of the form

$$\mathbf{x}_i + \mathbf{e}(\mathbf{x}_i, \mathbf{b}) = T(\mathbf{y}_j, \mathbf{t}_k) + \mathbf{f}_i, \tag{4}$$

where $\mathbf{e}$ describes the kinematic errors, $\mathbf{y}_j$ is the location of the jth sphere for the ball plate in a fixed frame of reference, $\mathbf{t}_k$ are parameters describing the kth transformation $\hat{\mathbf{y}} = T(\mathbf{y}, \mathbf{t})$ and $\mathbf{f}_i = \mathbf{f}_i(\mathbf{a})$ are parameter-dependent deviations of the model values from the data. If the model validly represents the data, the values of $\mathbf{f}_i$ corresponding to the best estimates (below) of $\mathbf{b}$ are estimates of the measurement errors corresponding to the transformed sphere centre co-ordinates. These errors can be expected to behave as random variables. Calibration information associated with the ball plate can be in the form of estimates d_{jl} of the inter-sphere distances and modelled as

$$\|\mathbf{y}_j - \mathbf{y}_l\| = d_{jl} + f_{jl}, \tag{5}$$

where the f_{jl} again represent deviations of the model values from the data. In theory, best estimates of the parametric errors defined by $\mathbf{b}$ can be determined by minimising

$$\sum \alpha_i^2 \mathbf{f}_i^{\mathrm{T}} \mathbf{f}_i + \sum \beta_{jl}^2 f_{jl}^2 \tag{6}$$

with respect to the parameters $\mathbf{b}$, $\{\mathbf{y}_j\}$ and $\{\mathbf{t}_k\}$ subject to the constraints (4–5). Here, α_i and β_{jl} are weights chosen to reflect the relative uncertainty in the measurement and calibration information. In practice, a key issue is to design measurement and calibration strategies that provide enough information to allow all the system parameters to be determined (system identifiability) with sufficiently small uncertainties.

The same approach can be adopted for the multiple measurement of other calibrated artefacts such as step gauges, ring gauges (cylinders) and spheres. See, for example, [5, 6], and the references therein. Once estimates of the

parametric errors have been determined, the CMM can either be re-adjusted to eliminate these geometrical errors or software correction can be applied to the estimates of the probe locations in the form

$$\hat{\mathbf{x}}(\mathbf{b}) = \mathbf{x} + \mathbf{e}(\mathbf{x}, \mathbf{b}).$$

2.4 Uncertainty in the assessment parameters

The key issue is that the estimates of the parameters $\mathbf{b}$ representing the CMM's parametric errors are determined from experiment and the uncertainty in these parameters will propagate to uncertainties in the measured points $\hat{X}(\mathbf{b}) = \{\mathbf{x}_i(\mathbf{b})\}$ and hence to any parameter $c = c(\hat{X})$ determined from $\hat{X}$. The aim of the Virtual CMM is to quantify the uncertainty in c due to the uncertainty in $\mathbf{b}$ and other relevant factors. The following approach, based on Monte Carlo simulation, is generally adopted:

1. Given a nominal shape $W = W(\mathbf{a}_0)$ for the workpiece and a measurement strategy specifying the number m and distribution of measurement points, generate points $\{\mathbf{x}_i^*\}_1^m$ exactly on the $W(\mathbf{a}_0)$.

2. Perform Monte Carlo simulations, $q = 1, \ldots, N$:

   a) Generate error parameters $\mathbf{b}_q$ according to some pre-assigned statistical model.

   b) Calculate scale readings $\mathbf{x}_{i,q}$ such that

   $$\hat{\mathbf{x}}_{i,q} = \mathbf{x}_{i,q} + \mathbf{e}(\mathbf{x}_{i,q}, \mathbf{b}_q) = \mathbf{x}_i^*, \quad i = 1, \ldots, m.$$

   c) Add random error to the scale readings to produce data sets

   $$\check{X}_q = \{\check{\mathbf{x}}_{i,q}\}_1^m$$

   with

   $$\check{\mathbf{x}}_{i,q} = \hat{\mathbf{x}}_{i,q} + \boldsymbol{\epsilon}_{i,q}.$$

   d) Determine assessment parameters $\mathbf{c}_q = \mathbf{c}_q(\check{X}_q)$.

3. Evaluate statistics associated with the sample $C = \{\mathbf{c}_q\}_1^N$.

If the parameters $\mathbf{b}_q$ are generated from some statistical distribution describing the uncertainty in the estimates of $\mathbf{b}$, then the covariance matrix of $\{\mathbf{c}_q\}$ will reflect that uncertainty.

2.5 The main components of the virtual CMM concept

If we look at the virtual CMM at a more general level, ignoring the particular details (which are many, complex and significant), the main components of the virtual CMM concept are:

- An instrument/sensor that outputs readings $X = \{\mathbf{x}_i(\mathbf{b})\}$ depending on calibration parameters $\mathbf{b}$.
- Assessment parameters $\mathbf{c}(X)$ depending on the readings X and hence on $\mathbf{b}$.
- Monte Carlo simulation to assess how uncertainty in $\mathbf{b}$ propagates through to uncertainty in $\mathbf{c}$.

3 Least squares incorporation of prior calibration information

In this section we look at how calibration information can be incorporated in determining measurement uncertainty in a least squares analysis of data. Least squares methods are appropriate if the measurement errors can be modelled as belonging to (multi-)normal distributions [4].

3.1 Least squares estimation

Many data analysis problems that arise in metrology can be posed as linear or nonlinear least squares systems. In a 'standard' experiment a response y_i of a system is measured for a number of settings of the independent variables $\mathbf{x}_i$. The model of the system describes the output y as a function $\phi(\mathbf{x}, \mathbf{a})$ of the variables $\mathbf{x}$ and parameters $\mathbf{a}$ describing the state of the system. The aim of the experiment is to determine estimates of $\mathbf{a}$ that best explain the data $X = \{(\mathbf{x}_i, y_i)\}_1^m$. This is usually achieved by defining an error function $F(X, \mathbf{a})$ in terms of the data X and parameters $\mathbf{a}$ and minimising F with respect to $\mathbf{a}$. If the measurements are related to the model values according to $y_i = \phi(\mathbf{x}_i, \mathbf{a}) + \epsilon_i$, where $\epsilon_i \in N(0, \sigma)$ are drawn from a Gaussian distribution, then the appropriate error function is the sum-of-squares

$$F(X, \mathbf{a}) = \sum_{i=1}^{m} \; (y_i - \phi(\mathbf{x}_i, \mathbf{a}))^2. \tag{7}$$

If the function ϕ is linear in the parameters $\mathbf{a}$ then the minimum of F is the maximum likelihood estimate of $\mathbf{a}$ and provides unbiased and efficient

estimates [16, 19]. If ϕ is (mildly) nonlinear, then the estimates produced are still favourable with respect to bias and statistical efficiency.

Determining least squares estimates involves the solution of overdetermined systems of linear equations $A\mathbf{a} = \mathbf{y}$. Let $\mathbf{f}(\mathbf{a}) = A\mathbf{a} - \mathbf{y}$ and $F(\mathbf{a}) = \mathbf{f}^{\mathrm{T}}(\mathbf{a})\mathbf{f}(\mathbf{a})$ Assuming A is full rank, then the requirement that $\nabla_{\mathbf{a}} F = \mathbf{0}$ at the minimum of F yields the *normal equations*:

$$\mathbf{a} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{y}. \tag{8}$$

A stable method of solving this system (see [14], for example) is to find an orthogonal factorisation

$$A = QR,$$

where Q is an $m \times n$ orthogonal matrix and U is $n \times n$ upper triangular. The solution $\mathbf{a}$ can be found efficiently by solving the upper-triangular system

$$R\mathbf{a} = Q^{\mathrm{T}}\mathbf{f}.$$

Nonlinear least squares systems are generally tackled using a variant of the Gauss-Newton algorithm to solve a sequence of linear least squares systems. Suppose we want to minimise

$$F(\mathbf{a}) = \mathbf{f}^{\mathrm{T}}(\mathbf{a})\mathbf{f}(\mathbf{a}) = \sum_{i}^{m} f_i^2(\mathbf{a})$$

with respect to $\mathbf{a} = (a_1, \ldots, a_n)^{\mathrm{T}}$. Given an estimate of the solution $\mathbf{a}$, let J be the *Jacobian matrix* of partial derivatives $J = \nabla_{\mathbf{a}^{\mathrm{T}}}\mathbf{f}$. An updated estimate is given by $\mathbf{a} + \mathbf{p}$ where $\mathbf{p}$ (known as the *Gauss-Newton step*) solves the matrix equation

$$J\mathbf{p} = -\mathbf{f}$$

in the least squares sense.

3.1.1 Covariance matrix of fitted parameters

Recall [15] that if $\mathbf{y} = (y_1, \ldots, y_m)$ has associated covariance matrix $V_{\mathbf{y}}$ and $\mathbf{a}(\mathbf{y}) = (a_1(\mathbf{y}), \ldots, a_n(\mathbf{y}))$ are n functions of $\mathbf{y}$, then the covariance matrix of $\mathbf{a}$ is estimated by

$$V_{\mathbf{a}} = \left(\nabla_{\mathbf{y}^{\mathrm{T}}}\mathbf{a}\right)^{\mathrm{T}} V_{\mathbf{y}} \left(\nabla_{\mathbf{y}^{\mathrm{T}}}\mathbf{a}\right).$$

Applying this to the normal equations (8), we see that if $V_{\mathbf{y}} = \sigma^2 I$ (as in the case where the measurements $\mathbf{y}$ are subject to Gaussian noise), then

$$V_{\mathbf{a}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\sigma^2 I A(A^{\mathrm{T}}A)^{-1} = \sigma^2(A^{\mathrm{T}}A)^{-1}.$$

For nonlinear systems,

$$V_{\mathbf{a}} = \sigma^2 (J^{\mathrm{T}} J)^{-1},$$

where J is the Jacobian matrix evaluated at the solution $\mathbf{a}^*$.

If a prior estimate of σ is not available, then σ can be estimated by

$$\hat{\sigma} = \|\mathbf{f}\|/(m-n)^{1/2},$$

where $\mathbf{f}$ is the vector $(f_1, \ldots, f_m)^{\mathrm{T}}$ of residuals evaluated at the solution.

The covariance matrix provides an important summary of what information about $\mathbf{a}$ can be derived from the data. The jth diagonal term is the variance of the fitted parameter a_j, while the (j,k)th element is the covariance of a_j with a_k. If $V_{\mathbf{a}}$ has eigenvalue decomposition [14]

$$V_{\mathbf{a}} = P \Lambda P^{\mathrm{T}}, \tag{9}$$

then the eigenvector $\mathbf{p}_{\max}$ ($\mathbf{p}_{\min}$) corresponding to the largest (smallest) eigenvalue $\lambda_{\max}$ ($\lambda_{\min}$) defines the linear combination $\mathbf{p}_{\max}^{\mathrm{T}}\mathbf{a}$ ($\mathbf{p}_{\min}^{\mathrm{T}}\mathbf{a}$) of the parameters $\mathbf{a}$ with maximal (minimal) variance $\lambda_{\max}$ ($\lambda_{\min}$).

The covariance matrix $V_{\mathbf{a}}$ defines the uncertainty ellipsoids about the solution $\mathbf{a}^*$:

$$(\mathbf{a} - \mathbf{a}^*)^{\mathrm{T}} V_{\mathbf{a}}^{-1} (\mathbf{a} - \mathbf{a}^*) = c^2. \tag{10}$$

If $\mathbf{a} \in N_n(\mathbf{a}^*, V_{\mathbf{a}})$ has multinormal distribution about $\mathbf{a}^*$ with covariance $V_{\mathbf{a}}$, then the probability that $\mathbf{a}$ lies within the ellispoid (10) is $\chi_n^2(U \leq c^2)$, where χ_n^2 is the the χ^2-distribution with n degrees of freedom. Using the eigenvalue decomposition (9), let $\hat{\mathbf{a}} = P^{\mathrm{T}}(\mathbf{a} - \mathbf{a}^*)$, then (10) becomes

$$\sum_j \frac{\hat{a}_j^2}{\lambda_j} \;=\; c^2.$$

The correlation matrix $C_{\mathbf{a}}$ defined from $V_{\mathbf{a}}$ by

$$C_{\mathbf{a}}(i,j) = \frac{V_{\mathbf{a}}(i,j)}{(V_{\mathbf{a}}(i,i) V_{\mathbf{a}}(j,j))^{1/2}},$$

presents a normalised version of the covariance information.

Importantly, even though the measurements y_i have uncorrelated errors, there will in general be correlation in the errors associated with the fitted parameters. If these estimates are fed into a subsequent data analysis, this correlation will have to be taken into account.

3.1.2 The Singular Value Decomposition of a matrix

See, for example, [13, 14].

Any $m \times n$ matrix A can be factored as the product

$$A = USV^{\mathrm{T}},$$

where U is an $m \times n$ orthogonal matrix, S is an $n \times n$ diagonal matrix with diagonal entries $s_1 \geq s_2 \geq \ldots \geq s_n$, and V is an $n \times n$ orthogonal matrix. This is known as the Singular Value Decomposition (SVD). (An orthogonal matrix U is such that its column vectors have unit norm and are orthogonal to each other: $\mathbf{u}_j^{\mathrm{T}}\mathbf{u}_k = 1$, $j = k$, 0 otherwise.) The columns of U (V) are the *left (right) singular vectors* and the s_j are known as the *singular values*. The vectors $\mathbf{u}_j$ form an orthogonal basis for the subspace spanned by the columns of A. The singular values are such that

$$\begin{aligned}
s_1 &= \|A\mathbf{v}_1\| = \max_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|, \\
s_k &= \|A\mathbf{v}_k\| = \|A\mathbf{v}_k\| = \max_{\mathbf{v}}\{\|A\mathbf{v}\| : \|\mathbf{v}\| = 1, \quad \mathbf{v}^{\mathrm{T}}\mathbf{v}_q = 0, q = 1, \ldots, k-1\}, \\
s_n &= \|A\mathbf{v}_n\| = \min_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|.
\end{aligned}$$

The SVD shows that A maps the orthonormal vectors $\mathbf{v}_j$ onto the vectors $s_j\mathbf{u}_j$. If A has singular values all equal to one then it is an orthogonal matrix, and conversely. A is full rank if and only if $s_n > 0$. The ratio $\kappa = s_1/s_n$ of the largest singular value to the smallest of a matrix is known as the *condition* [14, 20] of the matrix.

If $A = USV^{\mathrm{T}}$ then the eigenvalue decomposition of $A^{\mathrm{T}}A$ is given by

$$A^{\mathrm{T}}A = VS^2V^{\mathrm{T}},$$

showing that the eigenvalues $\{\lambda_j\}$ of $A^{\mathrm{T}}A$ are the squares of the singular values of A: $\lambda_j = s_j^2$ and the eigenvectors of $A^{\mathrm{T}}A$ are precisely the right singular vectors of A.

The singular values have a geometrical interpretation. The matrix A maps the unit sphere $\{\mathbf{x} : \|\mathbf{x}\| = 1\}$ in $\mathcal{R}^n$ into a hyper-ellipsoid in $\mathcal{R}^m$. The singular values are the lengths of the semi-axes of the ellipsoid. The condition number is the ratio of the largest semi-axis to the smallest. An ill-conditioned matrix is one which maps the sphere into a long thin ellipsoid. Orthogonal matrices map the unit sphere to a unit sphere.

3.1.3 Quadratic approximation to a sum of squares

Let $F(\mathbf{a}) = \frac{1}{2}\sum_{i=1}^{m} f_i^2(\mathbf{a})$, J be the Jacobian matrix $J_{ij} = \frac{\partial f_i}{\partial a_j}$ with QR factorisation $J = QR$ (and singular value decomposition $J = USV^{\mathrm{T}}$), $\mathbf{g}$ the

gradient $\mathbf{g} = \nabla_{\mathbf{a}} F = J^{\mathrm{T}}\mathbf{f}$, H the *Hessian* matrix

$$H = \left[\frac{\partial^2 F}{\partial a_j \partial a_k} \right]_{jk} = J^{\mathrm{T}}J + \sum_i f_i \frac{\partial^2 F}{\partial a_j \partial a_k},$$

and $\mathbf{a}^*$ a minimum of $F(\mathbf{a})$. We assume that at the minimum f_i are small or approximately linear so that H is approximated by

$$H \doteq J^{\mathrm{T}}J.$$

Then function $F(\mathbf{a})$ has Taylor expansion about $\mathbf{a}^*$ of the form

$$\begin{aligned} F(\mathbf{a}) &= F(\mathbf{a}^*) + \left[\mathbf{g}^{\mathrm{T}}(\mathbf{a}-\mathbf{a}^*) = 0\right] + \frac{1}{2}(\mathbf{a}-\mathbf{a}^*)^{\mathrm{T}} H (\mathbf{a}-\mathbf{a}^*) + \ldots \\ &\doteq F(\mathbf{a}^*) + \frac{1}{2}(\mathbf{a}-\mathbf{a}^*)^{\mathrm{T}} J^{\mathrm{T}} J (\mathbf{a}-\mathbf{a}^*) \\ &= F(\mathbf{a}^*) + \frac{1}{2}(\mathbf{a}-\mathbf{a}^*)^{\mathrm{T}} R^{\mathrm{T}} R (\mathbf{a}-\mathbf{a}^*). \end{aligned}$$

The quadratic approximation to $F(\mathbf{a})$ at $\mathbf{a}^*$ is defined essentially by the triangular factor R of J. Since

$$V_{\mathbf{a}} = \sigma^2 (J^{\mathrm{T}}J)^{-1} = \sigma^2 (R^{\mathrm{T}}R)^{-1} \ (= \sigma^2 V S^{-2} V^{\mathrm{T}}),$$

the covariance matrix $V_{\mathbf{a}}$ also stores this information.

3.2 Numerical simulation and experimental design

Two of the main uses of numerical simulation in discrete modelling are analysing system identifiability and effectiveness. Both can be performed from an analysis of the Jacobian matrix associated with a model of the measurement system.

3.2.1 System identifiability

A system is identifiable using a given measurement strategy if all the system parameters $\mathbf{a}$ can be determined from the resulting measurement data. For nonlinear least squares problems of the form

$$\min_{\mathbf{a}} \sum_i f^2(\mathbf{x}_i, \mathbf{a}),$$

system identifiability can be examined using the following approach. Given software to calculate $\mathbf{f}$ and the associated Jacobian matrix J and nominal values $\mathbf{a}_0$ of the parameters:

1 Generate a dense sample of data points $\{\mathbf{x}_i\}$ satisfying the model equations exactly so that $f(\mathbf{x}_i, \mathbf{a}_0) = 0$.

2 Calculate the associated Jacobian matrix J.

3 Determine the SVD of $J = USV^{\mathrm{T}}$ and singular values $s_1, \ldots, s_n$.

4 If J is full rank ($s_n > 0$) then the system is identifiable.

5 If J is not full rank, so that $s_k = \ldots = s_n = 0$, examine the corresponding columns $\mathbf{v}_k, \ldots, \mathbf{v}_n$ of V to determine which combinations of parameters are not determined from the data.

3.2.2 System effectiveness

Given an identifiable system, it is possible to use the Jacobian evaluation software in numerical simulations of an experiment to determine the likely uncertainties that will be obtained for a given measurement strategy. In this situation, exact measurement data $\mathbf{x}_i^*$ are generated according to a specified measurement strategy and from this, the Jacobian matrix J. Given an estimate of the standard deviation σ of the measurement errors, the covariance matrix is given by

$$V_{\mathbf{a}} = \sigma^2 (J^{\mathrm{T}} J)^{-1}.$$

The variances of the fitted parameters are estimated by the diagonal elements of the covariance matrix. Often we are not so much concerned with the variances of the fitted parameters but with variances and covariances of functions of the fitted parameters. If $h = \mathbf{h}^{\mathrm{T}}\mathbf{a}$ and $g = \mathbf{g}^{\mathrm{T}}\mathbf{a}$, then the covariance of h with g, $cov(h, g)$, is estimated by

$$cov(h, g) = \mathbf{g}^{\mathrm{T}} V_{\mathbf{a}} \mathbf{h},$$

and the variance of g is given by $var(g) = cov(g, g)$.

By changing the measurement strategy and monitoring the effect on the variances of the parameters of interest it is often possible to improve the experimental efficiency. Importantly, this can be achieved by using the same data analysis modules required to determine estimates of the parameters and their uncertainties from actual measurement data. In other words, with very little additional effort the model solving tools can be used to improve experimental strategy.

3.3 Generalisations of least squares methods

3.3.1 Covariance information

The least squares objective function is appropriate if the errors associated with the observation equations are uncorrelated. If these errors are correlated and V is the associated (full rank) covariance matrix then it is appropriate to minimise the Gauss-Markov objective function[16, 19]

$$\min_{\mathbf{a}} \ \mathbf{f}^{\mathrm{T}} V^{-1} \mathbf{f}. \tag{11}$$

Let $V^{-1} = LL^{\mathrm{T}}$ be the Cholesky factorisation [14] of V^{-1}. Then (11) can be written as $\min_{\mathbf{a}} \ \tilde{\mathbf{f}}^{\mathrm{T}} \tilde{\mathbf{f}}$, with $\tilde{\mathbf{f}}(\mathbf{a}) = L^{\mathrm{T}} \mathbf{f}(\mathbf{a})$. In this way, (11) can be posed as a standard least squares problem.

3.3.2 Prior information and updating strategies

Suppose $\mathbf{a}_0$ are prior estimates of a linear combination $A_0\mathbf{a}$ of the parameters with associated covariance matrix V_0, assumed to be full rank. This prior information can be included in the optimisation process by augmenting the objective function to

$$\left(A_0\mathbf{a} - \mathbf{a}_0\right)^{\mathrm{T}} V_0^{-1} \left(A_0\mathbf{a} - \mathbf{a}_0\right) \quad + \quad \sum_{i=1}^{m} f_i^2(\mathbf{a}). \tag{12}$$

As above, the Cholesky factorisation of V_0^{-1} can be used to transform the optimisation into a standard least squares problem.

Updating strategies are an example of this. Suppose in an experiment two sets of data $X_k = \{\mathbf{x}_i, i \in I_k\}$, $k = 1, 2$, have been gathered with the system in the same state (i.e., the system is specified by the same parameters $\mathbf{a}^*$ when both sets of data were gathered) and that the measurements are subject to Gaussian noise with $\sigma = 1$. Estimates of the system parameters can be determined from a least squares analysis $\mathbf{a} = \mathcal{A}(X)$ of the form

$$\min_{\mathbf{a}} \sum_i f^2(\mathbf{x}_i, \mathbf{a}).$$

Let $\mathbf{a}_k = \mathcal{A}(X_k)$ be the least squares estimates of the parameters determined from data set X_k, and $V_{\mathbf{a}_k}$ the corresponding estimates of covariance matrices of the fitted parameters, $k = 1, 2$. Note if $J_k = Q_k R_k$ are the Jacobian matrices at the solutions, then

$$V_{\mathbf{a}_k} = (R_k^{\mathrm{T}} R_k)^{-1},$$

X1		X2	
95.5453	-29.5590	-95.5413	-29.5758
97.1401	-23.7578	-97.1252	-23.7730
98.3851	-17.9093	-98.3849	-17.9062
99.2844	-11.9654	-99.2757	-11.9680
99.8131	-6.0000	-99.8161	-6.0015
100.0170	-0.0014	-99.9924	0.0000
99.8206	5.9829	-99.8160	6.0125
99.2988	11.9585	-99.2943	11.9797
98.3870	17.9128	-98.3806	17.9056
97.1425	23.7698	-97.1225	23.7610
95.5192	29.5440	-95.5264	29.5513

Table 1: Data points for arc 1 and arc 2.

(since $\sigma = 1$). Assuming that f_i are small at the solution, up to a constant,

$$F_k(\mathbf{a}) = \sum_{i \in I_k} f_i^2(\mathbf{x}_i, \mathbf{a}) \doteq (\mathbf{a} - \mathbf{a}_k)^{\mathrm{T}} R_k^{\mathrm{T}} R_k (\mathbf{a} - \mathbf{a}_k) = (\mathbf{a} - \mathbf{a}_k)^{\mathrm{T}} V_{\mathbf{a}_k}^{-1} (\mathbf{a} - \mathbf{a}_k).$$

This means that, for example, given $\mathbf{a}_1$, $V_{\mathbf{a}_1}$ and data set X_2, an updated estimate of the parameters can be found by solving

$$\min_{\mathbf{a}} \left\{ (\mathbf{a} - \mathbf{a}_1)^{\mathrm{T}} V_{\mathbf{a}_1}^{-1} (\mathbf{a} - \mathbf{a}_1) + \sum_{i \in I_2} f_i^2(\mathbf{x}_i, \mathbf{a}) \right\}.$$

3.4 Example: circle fitting I

We illustrate some of the concepts discussed above using a circle fitting example. Suppose we wish to find the centre co-ordinates (a_1, a_2) and radius a_3 of a circle from data points $\mathbf{x}_i = (x_i, y_i)$. Estimates of the circle centre can be found by solving

$$\min_{\mathbf{a}} \sum_i d^2(\mathbf{x}_i, \mathbf{a}),$$

where $d(\mathbf{x}, \mathbf{a}) = \{(x - a_1)^2 + (y - a_2)^2\}^{1/2} - a_3$ is the distance from $\mathbf{x}$ to the circle specified by parameters $\mathbf{a}$.

We determine circle fits to datasets $X_1 = \{\mathbf{x}_i, i \in I_1\}$ and $X_2 = \{\mathbf{x}_i, i \in I_2\}$ (see Table 1) representing points nominally on a circle centred at the origin with radius 100, but perturbed by Gaussian noise with $\sigma = 0.01$. X_1 represents an arc on the right near $(100, 0)$ and X_2 an arc on the left near $(-100, 0)$. These data points appear in Figure 1.

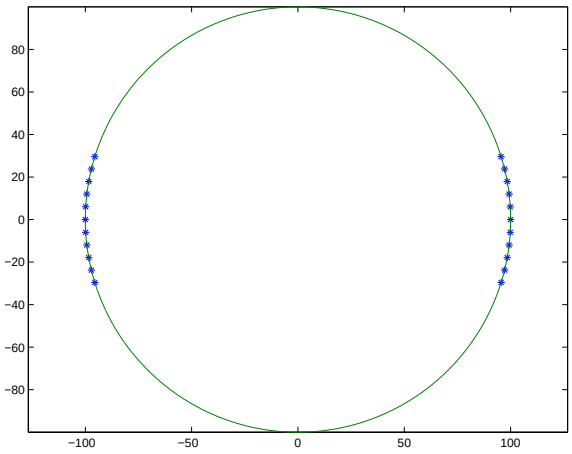

Figure 1: Data points in data sets X_1 (right arc) and X_2 (left arc).

We first consider minimising three objective functions

$$F_1(\mathbf{a}) = \sum_{i \in I_1} d^2(\mathbf{x}_i, \mathbf{a}), \quad F_2(\mathbf{a}) = \sum_{i \in I_2} d^2(\mathbf{x}_i, \mathbf{a}), \quad F_3(\mathbf{a}) = \sum_{i \in I_1 \cup I_2} d^2(\mathbf{x}_i, \mathbf{a}). \tag{13}$$

The best-fit parameters $\mathbf{a}_k$ and their standard uncertainties $u(\mathbf{a}_k)$, $k = 1, 2, 3$ for each of the objective functions are given in Table 2 and the covariance matrices V_k, $k = 1, 2$ are given in Table 3. For $k = 1, 2$ the arc nature of the data sets results in large uncertainties in the first co-ordinate of the circle centre and the radius. For $k = 3$, the data includes both right and left arcs and the uncertainties are much reduced. The covariance matrices also reflect the arc nature of the data. The eigenvalue decomposition of $V_1 = P_1 \Lambda_1 P_1^{\mathrm{T}}$ is given in Table 4. The first column of P_1 specifies the linear combination of $\mathbf{a}$ with maximal variation while the second column that with minimal variation. From this, we see that the combination of parameters $a_1 + a_3$ is well determined from the data but that $a_1 - a_3$ is poorly determined. For data set X_2, these roles are reversed.

We consider other ways to combine the two sets of data using objective functions

$$\begin{aligned} F_4(\mathbf{a}) &= (\mathbf{a} - \mathbf{a}_1)^{\mathrm{T}} V_1^{-1} (\mathbf{a} - \mathbf{a}_1) + \sum_{i \in I_2} d^2(\mathbf{x}_i, \mathbf{a}), \\ F_5(\mathbf{a}) &= (\mathbf{a} - \mathbf{a}_2)^{\mathrm{T}} V_2^{-1} (\mathbf{a} - \mathbf{a}_2) + \sum_{i \in I_1} d^2(\mathbf{x}_i, \mathbf{a}), \\ F_6(\mathbf{a}) &= (\mathbf{a} - \mathbf{a}_1)^{\mathrm{T}} V_1^{-1} (\mathbf{a} - \mathbf{a}_1) + (\mathbf{a} - \mathbf{a}_2)^{\mathrm{T}} V_2^{-1} (\mathbf{a} - \mathbf{a}_2). \end{aligned}$$

a1	u(a1)	a2	u(a2)	a3	u(a3)
1.3251e-01	1.9060e-01	4.3867e-02	1.9116e-01	3.0318e-03	2.1706e-03
-1.2846e-02	1.6043e-02	-1.5502e-02	1.6065e-02	-1.4182e-02	1.1357e-02
9.9874e+01	1.8721e-01	1.0004e+02	1.8776e-01	1.0000e+02	2.1320e-03

Table 2: Solution parameters $\mathbf{a}_k$ (centre co-ordinates and radius, columns 1, 3 and 5) along with their standard uncertainties $u(\mathbf{a}_k)$ (columns 2, 4 and 6) associated with minimising the objective functions F_k, $k = 1, 2, 3$ in (13) to circle data.

V1			V2		
3.6330e-02	4.7417e-06	-3.5678e-02	3.6541e-02	-2.6030e-06	3.5887e-02
4.7417e-06	2.5737e-04	-4.6839e-06	-2.6030e-06	2.5808e-04	-2.5931e-06
-3.5678e-02	-4.6839e-06	3.5047e-02	3.5887e-02	-2.5931e-06	3.5254e-02

Table 3: Covariance matrices V_1 and V_2 associated with the best-fit circle parameters to data sets X_1 and X_2 corresponding to objective functions F_1 and F_2.

P1			L1		
7.1344e-01	7.0072e-01	-1.1963e-04	7.1373e-02	0	0
9.3721e-05	7.5297e-05	1.0000e+00	0	4.6275e-06	0
-7.0072e-01	7.1344e-01	1.1953e-05	0	0	2.5737e-04

Table 4: Eigenvalue decomposition $V_1 = P_1 \Lambda_1 P_1^{\mathrm{T}}$ of the covariance matrix for the best-fit circle parameters to data set X_1.

a4-a3	u(a4)	a5-a3	u(a5)	a6-a3	u(a6)
-1.4469e-06	2.1706e-03	1.4711e-07	2.1706e-03	-1.2997e-06	2.1706e-03
1.9358e-06	1.1350e-02	4.5483e-07	1.1359e-02	2.3908e-06	1.1352e-02
-1.4849e-06	2.1320e-03	-1.5107e-07	2.1320e-03	-1.6360e-06	2.1320e-03

Table 5: Difference between solutions for F_4, F_5 and F_6 and that determined for F_3 (columns 1, 3 and 5), along with estimates of the standard uncertainties $u(\mathbf{a})$ for the parameters (columns 2, 4 and 6).

A comparison of the solution estimates $\mathbf{a}_k$, $k = 4, 5, 6$, with $\mathbf{a}_3$ along with their standard uncertainties are given in Table 5. The table shows that the four sets of results associated with F_3 to F_6 are close relative to the noise in the data of 0.01. These results demonstrate that the estimate of the circle parameters and the corresponding covariance matrix together retain sufficient information from the data sets X_k, $k = 1, 2$.

If, instead of using the covariance matrices, we use only the variances of the fitting parameters and ignore the covariances, we find that significant information is lost. Let

$$\begin{aligned}
\tilde{F}_4(\mathbf{a}) &= (\mathbf{a}-\mathbf{a}_1)^{\mathrm{T}} D_1^{-1} (\mathbf{a}-\mathbf{a}_1) + \sum_{i \in I_2} d^2(\mathbf{x}_i, \mathbf{a}), \\
\tilde{F}_5(\mathbf{a}) &= (\mathbf{a}-\mathbf{a}_2)^{\mathrm{T}} D_2^{-1} (\mathbf{a}-\mathbf{a}_2) + \sum_{i \in I_1} d^2(\mathbf{x}_i, \mathbf{a}), \\
\tilde{F}_6(\mathbf{a}) &= (\mathbf{a}-\mathbf{a}_1)^{\mathrm{T}} D_1^{-1} (\mathbf{a}-\mathbf{a}_1) + (\mathbf{a}-\mathbf{a}_2)^{\mathrm{T}} D_2^{-1} (\mathbf{a}-\mathbf{a}_2),
\end{aligned}$$

where D_k is the diagonal matrix with $D_k(j,j) = V_k(j,j)$. A comparison of the solution estimates $\tilde{\mathbf{a}}_k$, $k = 4, 5, 6$ associated with $\tilde{F}_k$, $k = 4, 5, 6$, with $\mathbf{a}_3$ along with their standard uncertainties are given in Table 6. In contrast with Table 5, Table 6 shows that the estimates associated with $\tilde{F}_k$, $k = 4, 5, 6$ differ significantly from those associated with F_3 for the first and third circle parameters. The standard uncertainties $u(\tilde{\mathbf{a}}_k)$ are also much larger. This emphasizes the fact the variances alone are not sufficient to provide all the information associated with the fits.

Suppose now that the radius of the circle has been estimated to be $a_{0,3} = 1$ with standard uncertainty σ_{a_3}. This prior information can be incorporated with data X_1 for example by minimising

$$\hat{F}_1(\mathbf{a}) = \frac{1}{\sigma_{a_3}^2}(a_3 - a_{0,3})^2 \; + \; \sum_{i \in I_1} d^2(\mathbf{x}_i, \mathbf{a}). \tag{14}$$

Table 7 shows the covariance matrices for the fitted parameters for the cases $\sigma = 0.001, 0.0001$. These can be compared with V_1 in Table 3. The prior

a4t-a3	u(a4t)	a5t-a3	u(a5t)	a6t-a3	u(a6t)
1.3569e-02	1.1014e-01	4.3245e-02	1.1033e-01	8.5285e-02	1.3497e-01
1.0120e-05	1.1350e-02	2.4270e-06	1.1357e-02	9.3875e-06	1.1352e-02
1.3311e-02	1.0818e-01	-4.2471e-02	1.0837e-01	-4.3774e-02	1.3257e-01

Table 6: Difference between solutions for $\tilde{F}_k$, $k = 4, 5, 6$, and that determined for F_3 (columns 1, 3 and 5), along with estimates of the standard uncertainties $u(\tilde{\mathbf{a}}_k)$ for the parameters (column 2, 4 and 6).

sigma = 0.001			sigma = 0.0001		
1.0459e-05	-2.6472e-08	-1.0179e-06	9.4333e-06	-2.6607e-08	-1.0180e-08
-2.6472e-08	2.5801e-04	-1.3362e-10	-2.6607e-08	2.5801e-04	-1.3363e-12
-1.0179e-06	-1.3362e-10	9.9997e-07	-1.0180e-08	-1.3363e-12	1.0000e-08

Table 7: Covariance matrices associated with the fitted parameters determined by solving (14) with $\sigma_{a_3} = 0.001, 0.0001$.

information much reduces the uncertainty in the x-co-ordinate a_1 and the radius a_3.

4 Calibration information

We now consider systems in which the system output y depends on the system parameters $\mathbf{a}$ and also on calibration information specified by parameters $\mathbf{b}$. For example, $\mathbf{b}$ may be the coefficients defining the calibration curve associated with a sensor. We assume that the calibration information is given in terms of estimates $\mathbf{b}_0$ of $\mathbf{b}$ along with the corresponding covariance matrix $V_{\mathbf{b}}$.

Consider firstly the linear system of observation equations with

$$\mathbf{y} = A\mathbf{a} + B\mathbf{b} + \boldsymbol{\epsilon}, \quad \epsilon_i \in N(0, \sigma^2).$$

We wish to determine best estimates of $\mathbf{a}$ taking into account the prior calibration information $\{\mathbf{b}_0, V_{\mathbf{b}}\}$. There are three possible approaches: i) regard $\mathbf{b} = \mathbf{b}_0$ as accurately known constants, ii) include $\mathbf{b}$ in the estimation process as additional parameters and solve

$$\min_{\mathbf{a},\mathbf{b}} \left\{ (A\mathbf{a} + B\mathbf{b} - \mathbf{y})^{\mathrm{T}} \frac{1}{\sigma^2} I (A\mathbf{a} + B\mathbf{b} - \mathbf{y}) \;+\; (\mathbf{b} - \mathbf{b}_0)^{\mathrm{T}} V_{\mathbf{b}}^{-1} (\mathbf{b} - \mathbf{b}_0) \right\}, \tag{15}$$

as considered in section 3.3, or iii) regard $\mathbf{b} = \mathbf{b}_0$ as constant but contributing to the covariance matrix $V_{\mathbf{y}}$ of $\mathbf{y}$ according to

$$V_{\mathbf{y}} = \sigma^2 I + B V_{\mathbf{b}} B^{\mathrm{T}},$$

and solve

$$\min_{\mathbf{a}} \ (A\mathbf{a} + B\mathbf{b}_0 - \mathbf{y})^{\mathrm{T}} V_{\mathbf{y}}^{-1} (A\mathbf{a} + B\mathbf{b}_0 - \mathbf{y}). \tag{16}$$

In fact, (15) and (16) are equivalent in the sense that they produce the same estimate of $\mathbf{a}$ and its covariance matrix $V_{\mathbf{a}}$. It is sufficient to consider the case where $\sigma = 1$, $\mathbf{b}_0 = \mathbf{0}$ and $V_{\mathbf{b}} = I$, so that the function in (15) becomes

$$(A\mathbf{a} + B\mathbf{b} - \mathbf{y})^{\mathrm{T}} (A\mathbf{a} + B\mathbf{b} - \mathbf{y}) \ + \ \mathbf{b}^{\mathrm{T}}\mathbf{b}, \tag{17}$$

that in (16) becomes

$$\min_{\mathbf{a}} \ (A\mathbf{a} - \mathbf{y})^{\mathrm{T}} V_{\mathbf{y}}^{-1} (A\mathbf{a} - \mathbf{y}), \tag{18}$$

and

$$V_{\mathbf{y}}^{-1} = \left(I + BB^{\mathrm{T}}\right)^{-1} = I - B\left(I + B^{\mathrm{T}}B\right)^{-1} B^{\mathrm{T}}. \tag{19}$$

With

$$\mathbf{b}^* = -\left(I + B^{\mathrm{T}}B\right)^{-1} B^{\mathrm{T}}(A\mathbf{a} - \mathbf{y}),$$

$\mathbf{b}^*$ minimises (17) with $\mathbf{a}$ regarded as fixed and

$$A\mathbf{a} + B\mathbf{b}^* - \mathbf{y} = V_{\mathbf{y}}^{-1} (A\mathbf{a} - \mathbf{y}),$$

so that (17) can be written as

$$(A\mathbf{a} - \mathbf{y})^{\mathrm{T}} \left\{V_{\mathbf{y}}^{-2} + B(I + B^{\mathrm{T}}B)^{-2}B^{\mathrm{T}}\right\} (A\mathbf{a} - \mathbf{y}).$$

Using (19), it can be shown that the middle term above is $V_{\mathbf{y}}^{-1}$, as required in (18).

For nonlinear systems, the equivalent of (15) is

$$\min_{\mathbf{a},\mathbf{b}} \ \left\{(\mathbf{b} - \mathbf{b}_0)^{\mathrm{T}} V_{\mathbf{b}}^{-1} (\mathbf{b} - \mathbf{b}_0) \ + \ \sum_i f_i^2(\mathbf{x}_i, \mathbf{a}, \mathbf{b})\right\}. \tag{20}$$

If $B = \nabla_{\mathbf{b}}^{\mathrm{T}} \mathbf{f}$ evaluated at $\mathbf{b} = \mathbf{b}_0$ is (essentially) independent of $\mathbf{a}$, then it is also possible to consider an approach corresponding to (16).

4.1 Prior information and self-calibration

Although the two approaches to incorporating the prior information are equivalent (at least in the linear case) as far as estimates of $\mathbf{a}$ and its covariance matrix $V_{\mathbf{a}}$ are concerned, they are different in their behaviour with respect to the calibration parameters $\mathbf{b}$. In (15), updated estimates of $\mathbf{b}$ and $V_{\mathbf{b}}$ can be determined from the analysis of the data. If the data measurement information represented by A and $\mathbf{y}$ say nothing about the calibration parameters, then we expect these updated estimates to be very close to their prior values. At the other extreme, in a self-calibration experiment, the measurement data is sufficient to determine accurate estimates of $\mathbf{a}$ and estimates of $\mathbf{b}$ without any prior information. Often, the position is somewhere in between these extremes: without some prior information, the estimates of $\mathbf{a}$ would not be accurate enough; after analysing the measurement data, the estimates of $\mathbf{b}$ are significantly improved (in terms of reduced variances, etc.).

In an analysis according to (16) no additional information about $\mathbf{b}$ is available directly. Issues in self-calibration are discussed in [12]. (The two methods of analysis differ also in their computational amd memory requirements.)

4.2 Example: circle fitting II

In this section, we combine our analysis of least squares estimation with prior information with the concept of the Virtual CMM, illustrated in the simpler case of a 2-dimensional CMM. Following from section 3.4, suppose now that the measurements of the circle are gathered using a CMM with scale and orthogonality errors so that the true co-ordinates $\mathbf{x}^* = \mathbf{x}^*(\mathbf{x}, \mathbf{b})$ are related to the measured co-ordinates $\mathbf{x}$ by

$$x^* = x(1 + b_1) + y(1 + b_2) \sin b_3, \quad y^* = y(1 + b_2) \cos b_3.$$

We suppose that prior information in the form of estimates $\mathbf{b}_0$ of $\mathbf{b}$ and associated covariance matrix $V_{\mathbf{b}}$ is available. In practice, these estimates can be determined from (multiple) measurements of a calibrated artefact. For example, if we measure a circular artefact for which an accurate estimate $a_{0,3}$ of the radius a_3 is known with standard uncertainty σ_{a_3}, it is possible to determine estimates of the circle parameters $\mathbf{a}$ and geometric errors $\mathbf{b}$ from measurement data $X = \{\mathbf{x}_i\}$ by solving

$$\min_{\mathbf{a},\mathbf{b}} \; F_7(\mathbf{a}, \mathbf{b}) \;=\; \left\{ \frac{1}{\sigma_{a_3}^2}(a_3 - a_{0,3})^2 \;+\; \frac{1}{\sigma^2} \sum_i \; d_i^2(\mathbf{x}^*(\mathbf{x}_i, \mathbf{b}), \mathbf{a}) \right\}. \qquad (21)$$

Once an estimate of machine errors has been determined, we assume that the system has been 'error corrected' so that in effect $\mathbf{b}_0 = \mathbf{0}$.

3.8182e-03	1.1490e-02	-2.9374e-10	5.2095e-05	1.1551e-04	4.6401e-05
1.1490e-02	5.4638e-01	-1.2784e-09	2.3748e-03	5.4796e-03	1.1945e-04
-2.9374e-10	-1.2784e-09	1.0000e-06	9.9929e-09	9.9873e-09	-7.2290e-21
5.2095e-05	2.3748e-03	9.9929e-09	1.0334e-05	2.3815e-05	5.4307e-07
1.1551e-04	5.4796e-03	9.9873e-09	2.3815e-05	5.4957e-05	1.2019e-06
4.6401e-05	1.1945e-04	-3.2946e-21	5.4307e-07	1.2019e-06	5.6874e-07

Table 8: Covariance matrix associated with F_7 for circle parameters $\mathbf{a}$ and machine error parameters $\mathbf{b}$ determined from 23 measurements representing a $\pi/2$-arc of a calibrated circular artefact.

Table 8 shows the covariance matrix V_7 of the fitted parameters $\mathbf{a}$ and $\mathbf{b}$ associated with minimising F_7 in (21) for 23 data points representing measurements of a calibrated circular artefact centred at the $(0, 0)$ and of nominal radius 100. The standard uncertainties are taken to be $\sigma = 0.01$ and $\sigma_{a3} = 0.001$ and the data points are distributed in a $\pi/2$-arc centred on $(0, 100)$. The lower-right 3×3 submatrix of V_7 is $V_{\mathbf{b}}$.

Having determined estimates of the calibration parameters along with their covariance matrix, we wish to find estimates of the best-fit circle parameters and their uncertainties, taking into account the uncertainty in the calibration information $\mathbf{b}$. Corresponding to (15) we can solve

$$\min_{\mathbf{a},\mathbf{b}} \; F_8(\mathbf{a}, \mathbf{b}) \;=\; (\mathbf{b} - \mathbf{b}_0)^{\mathrm{T}} V_{\mathbf{b}}^{-1} (\mathbf{b} - \mathbf{b}_0) \;+\; \frac{1}{\sigma^2} \sum_i \; d_i^2(\mathbf{x}^*(\mathbf{x}_i, \mathbf{b}), \mathbf{a}). \quad (22)$$

Alternatively, given good estimates of the solution parameters $\mathbf{a}$, we can set $B = \nabla_{\mathbf{b}^{\mathrm{T}}} \mathbf{d}$ and $V_{\mathbf{d}} = \sigma^2 I + B V_{\mathbf{b}} B^{\mathrm{T}}$ and, corresponding to (16), solve instead

$$\min_{\mathbf{a}} \; F_9(\mathbf{a}) \;=\; \mathbf{d}^{\mathrm{T}} V_{\mathbf{d}}^{-1} \mathbf{d}, \quad (23)$$

where $d_i = d_i(\mathbf{a}) = d(\mathbf{x}^*(\mathbf{x}_i, \mathbf{b}_0), \mathbf{a})$.

The estimates V_8 and V_9 of covariance matrices of the fitted parameters $\mathbf{a}$ associated with (22) and (23) are given in Table 9 and show good agreement. (Recall that for linear problems, the agreement would be exact.) These covariance matrices reflect all the sources of uncertainty in the calibration chain, including the uncertainty in the radius of the circular artefact used to estimate the calibration parameters $\mathbf{b}$ of the CMM.

We can compare these covariance estimates with those from a Monte Carlo simulation. Given data points $X^* = \{(x_i^*, y_i^*)\}$ lying exactly on a circle and covariance matrix $V_{\mathbf{b}}$, we generate parameters $\mathbf{b}_q$ such that $\mathbf{b}_q \in N(\mathbf{0}, V_{\mathbf{b}})$ and data sets $X_q = \{(x_{q,i}, y_{q,i})\}$ with

$$x_{q,i} \;=\; (x_i^* - y_i^* \tan b_{q,3})/(1 + b_{q,1}) + \epsilon_{q,i},$$

V8			V9		
7.2882e-01	-1.1343e-02	-9.9203e-01	7.2373e-01	-1.1389e-02	-9.8602e-01
-1.1343e-02	5.7216e-03	1.6564e-02	-1.1389e-02	5.7107e-03	1.6640e-02
-9.9203e-01	1.6564e-02	1.3586e+00	-9.8602e-01	1.6640e-02	1.3516e+00

Table 9: Covariance matrices associated with minimising F_8 and F_9 to determine the best-fit circle to data gathered by a CMM with calibration parameters $\mathbf{b}$.

Vm			Vz		
7.9376e-01	-1.3884e-02	-1.0810e+00	8.3025e-01	-1.2381e-02	-1.1152e+00
-1.3884e-02	5.7201e-03	1.9787e-02	-1.2381e-02	6.5362e-03	1.8113e-02
-1.0810e+00	1.9787e-02	1.4806e+00	-1.1152e+00	1.8113e-02	1.5125e+00

Table 10: Covariance matrices derived from 1000 Monte Carlo simulations of circle fits to data gathered by a CMM with calibration parameters $\mathbf{b}$.

$$y_{q,i} \quad = \quad y_i^*/((1 + b_{q,2})(\cos b_{q,3})) + \delta_{q,i},$$

so that X_q represents the measurements of the circle using a CMM with parametric errors $\mathbf{b}_q$. We determine estimates $\mathbf{a}_q$ of the best-fit circle parameters determined from X_q and calculate the covariance matrix V_m of $\{\mathbf{a}_q\}$.

Table 10 shows the estimate of V_m determined from 1000 Monte Carlo simulations. If we compare this with V_8 and V_9 in Table 9, we see that there is reasonable agreement. Since only 1000 Monte Carlo have been performed, we can only expect agreement at somewhere between 1 and 2 figures. If we increase the number of Monte Carlo simulations and reduce the noise in the data, the agreement is closer. This is to be expected, since all components of the system fall within the framework of multinormal statistics. We move outside this framework if, rather than finding the least squares best-fit circle to X_q, we instead adopt the Chebyshev fitting criterion [2]

$$\min_{\mathbf{a}} \max_{i} d(\mathbf{x}_i, \mathbf{a}),$$

to determine estimates $\tilde{\mathbf{a}}_q$ of the circle parameters corresponding to X_q. The covariance matrix V_z determined from 1000 Monte Carlo simulations of Chebyshev fits is also given in Table 10. It can be seen that V_z is reasonably close to V_m (and hence V_8 and V_9) showing that in this non-normal case the estimate of the covariance of the fitted parameters derived from multinormal statistics is a fair guide.

The circle fitting examples considered above are reasonably straightforward but have many of the features that can arise in metrology data analysis problems: nonlinearity, a calibration chain with calibration information passed to the next element in the chain, correlation amongst parameters, etc.

4.3 Example: determination of the volume ratio from multiple expansions

The same approach based on multinormal statistical methods can be applied to other metrology data analysis problems where uncertainties of parameters associated with subcomponents need to be taken into account. In this example, we consider the determination of the volume ratio of two vessels from a multiple expansion experiment.

4.3.1 Generation of low pressures

In vacuum metrology, low pressures are generated by expanding a gas in a small volume at a relatively high pressure into a larger evacuated volume. The pressure in the combined volume is a function of the ratio of the volumes. Knowing the inlet pressure and the volume ratio, the pressure in the combined volume can be estimated from (variations of) the ideal gas law; see [8, 9, 18, 17].

The volume ratio can be determined from multiple expansion measurements in which gas is repeatedly expanded from the small volume into the combined volume and the inlet pressures and the rising pressures in the combined volume measured at each expansion. The idea of the overall approach is that relatively accurate pressure measurement at high pressures can be used to determine the volume ratio. Accurate low pressures can then be generated to test the accuracy of low pressure measurement devices.

The ideal gas law can be written as

$$pV = nRT, \tag{24}$$

where p is the pressure, V is the volume, n is the number of moles of the gas, T is the absolute temperature and R is the gas constant. Let v and V represent two volumes connected by a valve. Suppose that with the valve closed, v and V have quantities of the same gas at pressures p and P. Assuming that the temperature is constant throughout, when the valve is opened the pressure Q in the combined volume is given by

$$\frac{pv}{RT} + \frac{PV}{RT} = \frac{Q(v+V)}{RT},$$

or, in terms of the volume ratio $a = (v + V)/v$,

$$\frac{p}{RT} + \frac{P(a-1)}{RT} = \frac{Qa}{RT}. \tag{25}$$

This equation states that the number of moles in the combined volume is the sum of the numbers of moles in the two separated volumes, i.e., the conservation of molecules. In practice, this equation has to be modified to take into account factors such as variations in temperature and the non-ideal behaviour of the gas. The ideal gas law is a first order approximation to the virial equation of state:

$$pV = n(RT + Bp + Cp^2 + \ldots). \tag{26}$$

The coefficient B is the second virial coefficient, C the third, etc. Taking into account temperature and the second virial coefficient, the counterpart of (25) is

$$\frac{p}{Rs + Bp} + \frac{P(a-1)}{RS + BP} = \frac{Qa}{RT + BQ}, \tag{27}$$

where s and S are the temperatures in volumes v and V before expansion and T is the temperature in the combined volume after expansion.

4.3.2 Sensors

The measurements of pressure and temperature will be subject to measurement error. We can model the behaviour of a sensor as

$$r^* = \mathcal{B}(r, \mathbf{b}) + \epsilon, \tag{28}$$

where r^* is the true value of the measurand, r is the value output by the sensor, $\mathcal{B}$ is a function describing the systematic error behaviour of the sensor in terms of parameters $\mathbf{b}$ and ϵ represent uncorrelated random errors sampled from a normal distribution with mean zero. For example, the inlet pressure p could be modelled as

$$p^* = \mathcal{B}(p, b_1) = p + b_1 + \epsilon,$$

where b_1 represents some unknown offset to take into account calibration error or long term drift in the instrument behaviour since calibration.

The sensor parameters $\mathbf{b}_k$ will be assumed to be drawn from a statistical distribution such as the normal or rectangular distributions. We wish to take into account the uncertainty in these calibration parameters in determining the volume ratio a from measurements.

4.3.3 Model equations

Suppose that a multiple expansion experiment is carried out to determine the volume ratio of a pair of volumes. We suppose that only four sensors are involved:

p pressure in the small volume.

t temperature in the small volume.

P pressure in the large volume.

T temperature in the large volume.

The sensors P and T take readings before and after each expansion. We can also regard the measured value of the second virial coefficient B of the gas as the output of a fifth pseudo-sensor with known behaviour (in other words, there is an estimate of the standard uncertainty $u(B)$ of B).

For the multiple expansion experiment, the model equations take the form

$$\frac{p_i^*}{Rt_i^* + B^* p_i^*} + \frac{P_{i-1}^*(a-1)}{RT_{i-1}^* + B^* P_{i-1}^*} = \frac{P_i^* a}{RT_i^* + B^* P_i^*}, \tag{29}$$

where p_i^*, etc., represent the true values of the sensor measurands. To simplify the example, we will assume that all the temperature measurements and the virial coefficient estimate are accurate so that (with appropriate corrections applied) equation (29) can be simplified to

$$p_i^* + P_{i-1}^*(a-1) = P_i^* a. \tag{30}$$

Our model for the sensor behaviour is

$$p_i^* = p_i + b_1 + \epsilon_i, \quad P_i^* = P_i(1+b_3) + b_2 + \delta_i, \tag{31}$$

where b_1 and b_2 represent unknown offsets and b_3 a scale error associated with the pressure measurements for the combined volume.

4.3.4 Estimators

Linear least squares estimator (LLSE). The parameter a appears linearly in equation (30) (and also in (29)) and an approximate estimate of a can be easily obtained by solving the equations

$$(p_i - P_{i-1})\,\nu = (P_i - P_{i-1})\,, \quad i = 2, \ldots, m, \tag{32}$$

in the least squares sense for $\nu = 1/a$. This estimator does not take into account any systematic behaviour of the sensors nor reflect accurately the measurement errors. However, it can be used to provide an initial estimate for other methods.

Simple generalised regression estimator (SGRE). In order to reflect more properly error structure in the data, we can consider the equations

$$(p_i + \epsilon_i) + (P_{i-1} + \delta_{i-1})(a-1) = (P_i + \delta_i)a, \quad i = 2, \ldots, m. \tag{33}$$

This ignores the systematic error behaviour of the sensors but models the random errors correctly. The terms ϵ_i appears linearly in equation (33) so that $\epsilon_i = \epsilon_i(a, \boldsymbol{\delta})$ can be written in terms of the parameters a and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)^{\mathrm{T}}$. An estimate of a can be determined by solving the nonlinear least squares optimisation problem

$$\min_{a,\boldsymbol{\delta}} \left\{ \boldsymbol{\delta}^{\mathrm{T}} D_{\boldsymbol{\delta}} \boldsymbol{\delta} + \boldsymbol{\epsilon}^{\mathrm{T}} D_{\boldsymbol{\epsilon}} \boldsymbol{\epsilon} \right\}. \tag{34}$$

Here, the weighting diagonal matrices $D_{\boldsymbol{\delta}}$ and $D_{\boldsymbol{\epsilon}}$ can be chosen to reflect the relative measurement accuracies.

Comprehensive generalised regression estimator (CGRE). The simple generalised regression estimator can be developed to take into account the systematic error behaviour. Equation (30) can be re-arranged to write $\epsilon_i = \epsilon_i(a, \mathbf{b}, \boldsymbol{\delta})$ as a function of a, $\mathbf{b}$ and $\boldsymbol{\delta}$. If it is assumed that the calibration parameters are normally distributed with mean zero and standard deviation σ_k, i.e., $b_k \in N(0, \sigma_k^2)$, $k = 1, 2, 3$, then estimates of the volume ratio are determined by solving

$$\min_{a,\mathbf{b},\boldsymbol{\delta}} \left\{ \mathbf{b}^{\mathrm{T}} D_{\mathbf{b}} \mathbf{b} + \boldsymbol{\delta}^{\mathrm{T}} D_{\boldsymbol{\delta}} \boldsymbol{\delta} + \boldsymbol{\epsilon}^{\mathrm{T}} D_{\boldsymbol{\epsilon}} \boldsymbol{\epsilon} \right\}, \tag{35}$$

where $D_{\mathbf{b}}$ is the 3×3 diagonal weighting matrix with $D(k,k) = 1/\sigma_k^2$.

4.3.5 Comparison of estimators

For each of these least-squares estimators, an estimate of the standard uncertainty $\hat{u}_q$ of the volume ratio can be determined using the approaches outlined in section 3.1.1. These estimates can be compared with the actual variation in the parameter estimates determined from Monte Carlo simulations. Given the true volume ratio a^*, and exact inlet pressures $p_i^* = 95,000$ Pa, we generate exact pressures P_i^* for the combined volume according to equation (30). For $q = 1, \ldots, N$, we generate random values for the calibration parameters $\mathbf{b}_q$ and random errors $(\delta_{q,i}, \epsilon_{q,i})$ and pressures $p_{q,i}$ and $P_{q,i}$ such that

$$p_i^* = p_{q,i} + b_1 + \epsilon_{q,i}, \quad P_i^* = P_{q,i}(1 + b_3) + b_2 + \delta_{q,i}. \tag{36}$$

	$\bar{a} - a^*$	s	$\bar{u}$
LLSE	-0.0092	0.1828	0.0700
SGRE	-0.0136	0.2622	0.0277
CGRE	0.0003	0.0229	0.0217

Table 11: Mean bias $\bar{a} - a^*$, standard deviation s of the estimates $\{a_q\}$ and mean estimate of the standard uncertainty of the volume ratio for each estimator in 1000 Monte Carlo simulations. The calibration parameters $\mathbf{b}_q$ are normally distributed.

For each estimator, we determine estimates of the volume ratio a_q from the qth set of pressures. For each estimate, we give the mean bias $\bar{a} - a^*$, the standard deviation s of the estimates $\{a_q\}$ and the mean $\bar{u}$ of estimates of the standard uncertainties $\hat{u}_q$. Table 11 gives these statistics for an experiment involving 1000 Monte Carlo simulations of an experiment involving 30 expansions and a volume ratio of 10. It is assumed that δ_i, ϵ_i, b_1, b_2 and b_3 are drawn from normal distributions with mean zero and standard deviation 100.0, 100.0, 100.0, 100.0 and 0.01, respectively.

The following points can be noted. Firstly, the variation in the estimates provided by the comprehensive estimator CGRE is smaller by approximately an order of magnitude than those for the other estimators. In other words, CGRE makes much more efficient use of the data. Secondly, only the estimates of the standard uncertainty ($\bar{u}$) provided by CGRE accord with the actual variation in the estimates (as measured by s). We emphasize that there is no error in the way the uncertainties associated with LLSE and SGRE have been calculated. These estimates are incorrect because they each assume a statistical model for the measurement errors that does not hold.

Table 12 gives the corresponding results for the case where the calibration parameters b_1, b_2 and b_3 are drawn from rectangular distributions with mean zero and standard deviation 100.0, 100.0 and 0.01, respectively. These results show a similar pattern to Table 11. Even though the model does not fall completely within multinormal statistics, the estimate of the uncertainty provided by CGRE is in agreement with the actual spread of estimates.

4.4 Discussion

If the statistical models for the measurement errors and the calibration information are essentially linear and multinormal, then least squares methods are appropriate. Their success does depend on being able to weight

	$\bar{a} - a^*$	s	$\bar{u}$
LLSE	0.0003	0.1906	0.0706
SGRE	-0.0008	0.2724	0.0276
CGRE	-0.0002	0.0217	0.0217

Table 12: Mean bias $\bar{a} - a^*$, standard deviation s of the estimates $\{a_q\}$ and mean estimate of the standard uncertainty of the volume ratio for each estimator in 1000 Monte Carlo simulations. The calibration parameters $\mathbf{b}_q$ are uniformly distributed.

appropriately the prior information relative to the current measurement information. Often this information is available since the behaviour of the measurement system components is reasonable well understood. A main challenge of data fusion in metrology is to determine parameter estimates when the prior covariance information is unavailable or unreliable.

The incorporation of prior information can be achieved by modifying standard least squares regression methods to solve the augmented least squares problems that arise. At NPL, we have developed a suite of software modules that convert these augmented systems to standard form so that existing regression software can be applied without change [11]. The software also caters for linear constraints on the optimisation parameters and sparsity structure in the matrix equations.

For systems in which there is a significant non-normal component or that are highly nonlinear, the application of multinormal statistics based on linearisations could give misleading results. Monte Carlo simulations provide another method of analysing the behaviour of a measurement system and the effectiveness of data analysis methods. They can be used to verify whether or not multinormal statistics apply, providing valuable validation of standard approaches where they do and offering an alternative where they do not.

The Monte Carlo approach can require significant computational resources, and it is not unusual to perform 10,000 or more simulations. For large, complicated systems, this may not be practical for everyday calculations (but feasible for an exploratory study). The Virtual CMM is a reasonably complex system with hundreds of parameters used to describe CMM behaviour. The estimation of reliable task-specific uncertainties from Monte Carlo simulations is a sizeable task. It may be that the use of multinormal statistics are sufficient alone to provide reliable uncertainty estimates. It is possible that Monte Carlo simulations on a small subspace of parameters used in conjunction with multinormal statistics could provide a more effective approach.

5 Summary and concluding remarks

In the presence of normally distributed uncertainty in measurement data, least squares parameter estimation methods are appropriate.

There are standard approaches to calculating the covariance matrix of the fitted parameters in a least squares estimation process. In general, the parameter estimates will have correlated uncertainties.

The parameter estimates along with their covariance matrix provide a summary of the information held by the data (in terms of the model).

There are consistent approaches to incorporating the uncertainties associated with parameters representing prior calibration information into the estimation process.

Software for incorporating covariances and prior information can be easily constructed from standard least squares optimisation software modules.

Monte Carlo approaches can be used to validate the uncertainty estimation for multinormal systems and provide estimates for non-normal situations.

Acknowledgement. We thank our colleague Dr I M Smith for comments on a draft of this report.

References

[1] G. T. Anthony, H. M. Anthony, B. Bittner, B. P. Butler, M. G. Cox, R. Drieschner, R. Elligsen, A. B. Forbes, H. Groß, S. A. Hannaby, P. M. Harris, and J. Kok. Chebyshev best-fit geometric elements. Technical Report DITC 221/93, National Physical Laboratory, Teddington, 1993.

[2] G. T. Anthony, H. M. Anthony, B. Bittner, B. P. Butler, M. G. Cox, R. Drieschner, R. Elligsen, A. B. Forbes, H. Groß, S. A. Hannaby, P. M. Harris, and J. Kok. Reference software for finding Chebyshev best-fit geometric elements. *Precision Engineering*, 19:28 – 36, 1996.

[3] B. P. Butler, A. B. Forbes, and P. M. Harris. Algorithms for geometric tolerance assessment. Technical Report DITC 228/94, National Physical Laboratory, Teddington, 1994.

[4] M. G. Cox, A. B. Forbes, and P. M. Harris. Software Support for Metrology Best Practice Guide 4: Modelling Discrete Data. Technical report, National Physical Laboratory, Teddington, 2000.

[5] M. G. Cox, A. B. Forbes, P. M. Harris, and G. N. Peggs. Determining CMM behaviour from measurements of standard artefacts. Technical Report CISE 15/98, National Physical Laboratory, Teddington, March 1998.

[6] M. G. Cox, A. B. Forbes, P. M. Harris, and G. N. Peggs. Experimental design in determining the parametric errors of CMMs. In V. Chiles and D. Jenkinson, editors, *Laser Metrology and Machine Performance IV*, pages 13–22, Southampton, 1999. WIT Press.

[7] M. G. Cox, A. B. Forbes, and G. N. Peggs. The virtual co-ordinate measuring machine. *Measurements for the future*, (Issue 4), Spring 1999. National Physical Laboratory.

[8] J. H. Dymond and E. B. Smith. *The virial coefficients of gases*, volume 2 of *Oxford Science Research Papers*. Clarendon Press, Oxford, 1969.

[9] K. W. T. Elliott and P. B. Clapham. The accurate measurement of the volume ratios of vacuum vessels. Technical Report MOM 28, National Physical Laboratory, Teddington, 1978.

[10] A. B. Forbes. Least-squares best-fit geometric elements. Technical Report DITC 140/89, National Physical Laboratory, Teddington, 1989.

[11] A. B. Forbes, P. M. Fossati, and I. M. Smith. General purpose least squares solvers for metrology applications. Technical report, National Physical Laboratory, Teddington. *To appear.*

[12] A. B Forbes and I. M. Smith. Self-calibration and error separation techniques in metrology. *Extended presentation at* Advanced Mathematical and Computational Tools in Metrology V*, Lisbon, May 2000.*

[13] G. H. Golub. The singular value decomposition with applications. In P. Ciarlini, M. G. Cox, F. Pavese, and D. Richter, editors, *Advanced Mathematical Tools in Metrology, II*, pages 48–55, Singapore, 1996. World Scientific.

[14] G. H. Golub and C. F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, third edition, 1996.

[15] International Organization for Standardization, Geneva. *ISO Guide to the expression of uncertainty in measurement*, 1993.

[16] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.

[17] K. F. Poulter. The calibration of vacuum gauges. *Journal of Physics E: Scientific Instruments*, 10:112–125, 1977.

[18] F. J. Redgrave, A. B Forbes, and P. M. Harris. A discussion of methods for the estimation of expansion ratios determined by multiple expansions. *Vacuum*, 53:159–162, 1999.

[19] J. A. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, Belmont, CA, second edition, 1995.

[20] J. R. Rice. A theory of condition. *SIAM J. Num. Anal.*, 3:287–310, 1966.

[21] E. Trapet and F. Waldele. The virtual CMM concept. In P. Ciarlini, M. G. Cox, F. Pavese, and D. Richter, editors, *Advanced Mathematical Tools in Metrology, II*, pages 238–247, Singapore, 1996. World Scientific.

[22] G. Zhang, R. Ouyang, B. Lu, R. Hocken, R. Veale, and A. Donmez. A displacement method for machine geometry calibration. *Annals of the CIRP*, 37:515–518, 1998.