Estimating Measurement Uncertainty

Oscar Chinellato, Erwin Achermann*
Institute of Scientific Computing
ETH Zürich

28th June 2000

Abstract

Most analytical methods are relative and hence require calibration. Calibration measurements are normally performed with reference materials or calibration standards. Typically least squares methods have been employed in such a manner as to ignore uncertainties associated with the calibration standards. However different authors [3, 4, 11] have shown that by taking into account the uncertainties associated with the calibration standards, better approximations to the model are possible.

We show a complete, numerically stable and fast way to compute the *Maximum Likelihood (fitting of a) Functional Relationship (MLFR)*, and we show how the model described in [11] is a unnecessarily restricted case of our fully general model.

1 Calibration and Measurement

Since measurement instruments obey physical laws, they can be modelled by mathematical functions. These functions describe only the qualitative behaviour unless all free parameters $\mathbf{p}$ are quantified. Calibration is the process of quantifying these parameters. For example, consider the measurement of a weight using a spring balance.

Here we denote weights with $x \in \mathbb{R}$ and stretches of the spring with $y \in \mathbb{R}$. Let x_{cs} be a known weight of a calibration standard and y_{cs} be the stretch of the spring balance for this calibration standard. For a given calibrated balance $\mathbf{p}$ the following holds true: $y_{\text{cs}} = f(x_{\text{cs}}, \mathbf{p})$. This f is called the *calibration function*. The art of calibrating consists of adjusting the parameters $\mathbf{p}$ (e.g. spring constant) using several calibrating pairs x_i and y_i such that $y_i \approx f(x_i, \mathbf{p})$ holds for all i as well as possible.

For an unknown mass $\hat{x}$, *weighing with a calibrated spring balance* requires the measurement of a spring stretch $\hat{y}$ and subsequent application of the *measurement function* $g(\hat{y}, \mathbf{p})$. In other words, $\hat{x} \approx g(\hat{y}, \mathbf{p})$. Note that g is the inverse function of f.

In the case that the pairs (x_i, y_i) were exact this would be a standard problem. But in practice every measurement is subject to uncertainties. Therefore, not only the calibrating pairs (x_i, y_i) but also their respective uncertainties (α_i, β_i) must be known and taken into consideration. Throughout this paper the error random variables $e_i^{(x)}$ respectively $e_i^{(y)}$ are treated as $e_i^{(x)} \sim \mathcal{N}(0, \alpha_i)$ respectively $e_i^{(y)} \sim \mathcal{N}(0, \beta_i)$. This fully conforms with the "Guide to the Expression of Uncertainty in Measurement" (GUM) [10].

With respect to our spring balance example, this implies that the real calibration weights are not x_i but $x_i + e_i^{(x)}$. Analogously the correct values for the stretch are no longer y_i but $y_i + e_i^{(y)}$.

*{achermann | chinellato}@inf.ethz.ch

Note that the uncertainties of x and y are uncorrelated. That is, the error of the calibration standard is independent of the error of the reading.

2 Calibration Models

Before we start considering calibration models, it is necessary to agree on a number of the assumptions.

The measurement tool is expected to work according to the function $f(\cdot,\cdot)$ in use. This means that there exists a set of parameters which describes the apparatus correctly. In other words, the model is assumed to be exact.

Every error *random vector* (rv) $\mathbf{u}$ is assumed to be normally distributed and is denoted by $\mathbf{u} \sim \mathcal{N}(\mu, \mathbf{C})$ where $\mu \in \mathbb{R}^n$ is the *mean vector* and $\mathbf{C} \in \mathbb{R}^{n\times n}$ is the *covariance matrix* of the rv $\mathbf{u} \in \mathbb{R}^n$. The well known *probability density function* (pdf) is

$$\mathrm{pdf}(\mathbf{u}) = \mathrm{err}\,(\mathbf{u} - \mu, \mathbf{C}) = (2\pi)^{-\frac{n}{2}} \frac{1}{\sqrt{\det(\mathbf{C})}} \mathrm{e}^{-\frac{1}{2}(\mathbf{u}-\mu)^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{u}-\mu)}.$$

A crucial point for the further development of calibration models is the exact specification of the calibration process. A reasonable assumption for one calibration measurement is the following (the process is schematically depicted in figure 1):

- Choose the i-th calibration standard x_i (whose real but unknown value ξ_i differs from x_i by $e_i^{(x)}$).
- Let $\mathbf{p}$ be the real but unknown set of parameters that describes the tool exactly. Here η_i shall denote the true but (again) unknown value $f(\xi_i, \mathbf{p})$.
- Finally, the reading mechanism of the tool generally imposes an additional error $e_i^{(y)}$. This leads to $y_i = \eta_i + e_i^{(y)}$.

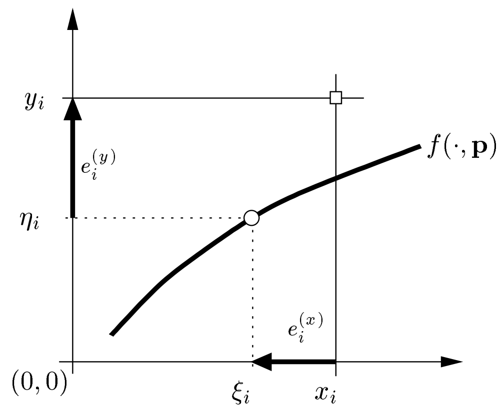


Figure 1: Process of a calibration measurement

2.1 Berkson Model

The *Berkson Model* [5, 6] is a statistical model that handles exactly the described process depicted in Figure 1. It behaves as follows

$$\begin{aligned} \xi_i &= x_i + e_i^{(x)} \\ y_i &= f(\xi_i, \mathbf{p}) + e_i^{(y)}, \end{aligned} \tag{1}$$

where x_i is a known and constant value and the errors $e_i^{(x)}$ and $e_i^{(y)}$ are independent and normally distributed with means 0 and variances α_i^2 and respectively β_i^2.

A *Taylor series expansion* of $f(\xi_i, \mathbf{p})$ around x_i yields an expression that reveals the connection between ξ_i and y_i

$$y_i \approx f(x_i, \mathbf{p}) + \frac{\partial}{\partial x} f(x_i, \mathbf{p}) \cdot e_i^{(x)} + e_i^{(y)} = f^{(i)} + f_x^{(i)} e_i^{(x)} + e_i^{(y)}, \tag{2}$$

where $f^{(i)} = f(x_i, \mathbf{p})$ and $f_x^{(i)} = \frac{\partial}{\partial x} f(x_i, \mathbf{p})$. Equations (1) and (2) make the following relations obvious:

$$\begin{array}{rclcl} \mathrm{Var}\,(\xi_i) & = & \mathrm{Var}\left(x_i + e_i^{(x)}\right) & = & \alpha_i^2 \\ \mathrm{Var}\,(y_i) & \approx & \mathrm{Var}\left(f^{(i)} + f_x^{(i)} e_i^{(x)} + e_i^{(y)}\right) & = & {f_x^{(i)}}^2 \alpha_i^2 + \beta_i^2 \\ \mathrm{Cov}\,(\xi_i, y_i) & \approx & \mathrm{E}\left[e_i^{(x)}\left(f_x^{(i)} e_i^{(x)} + e_i^{(y)}\right)\right] & = & f_x^{(i)} \alpha_i^2 . \end{array}$$

2.1.1 Fitting of ξ and $\mathbf{p}$ (XiP-fit)

As can be seen in Section 2.1, the pdf of (ξ_i, y_i) can be written as *bivariate normal distribution*

$$\mathrm{pdf}(\mathbf{z}_i) = \mathrm{err}\,(\mathbf{z}_i, \mathbf{C}_i) = \frac{1}{2\pi} \frac{1}{\sqrt{\det\left(\mathbf{C}_i\right)}}\, \mathrm{e}^{-\frac{1}{2} \mathbf{z}_i{}^{\mathrm{T}} \mathbf{C}_i^{-1} \mathbf{z}_i},$$

where $\mathbf{z}_i$ and $\mathbf{C}_i$ are defined as follows:

$$\mathbf{z}_i = \begin{pmatrix} \xi_i - x_i \\ y_i - f^{(i)} \end{pmatrix} \quad \text{and} \quad \mathbf{C}_i = \begin{pmatrix} \alpha_i^2 & f_x^{(i)} \alpha_i^2 \\ f_x^{(i)} \alpha_i^2 & {f_x^{(i)}}^2 \alpha_i^2 + \beta_i^2 \end{pmatrix}.$$

We generalize to all n measurements, and allow additional correlations amongst the errors $e_i^{(x)}$ respectively $e_i^{(y)}$ denoted by $\mathbf{C}_x$ respectively $\mathbf{C}_y$. This leads to the

$$\mathrm{pdf}(\xi, \mathbf{y}) = \mathrm{err}\,(\mathbf{z}, \mathbf{C}) = (2\pi)^{-n} \frac{1}{\sqrt{\det\left(\mathbf{C}\right)}}\, \mathrm{e}^{-\frac{1}{2} \mathbf{z}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{z}},$$

where $\mathbf{z}$ and $\mathbf{C}$ now are defined as

$$\mathbf{z} = \begin{pmatrix} \xi - \mathbf{x} \\ \mathbf{y} - \mathbf{f} \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} \mathbf{C}_x & \mathbf{C}_x \mathbf{F}_x \\ \mathbf{F}_x \mathbf{C}_x & \mathbf{F}_x \mathbf{C}_x \mathbf{F}_x + \mathbf{C}_y \end{pmatrix}.$$

Now, analogously,

$$\mathbf{f} = \left(f^{(1)}, \ldots, f^{(n)}\right)^{\mathrm{T}} \quad \text{and} \quad \mathbf{F}_x = \mathrm{diag}\left(f_x^{(1)}, \ldots, f_x^{(n)}\right).$$

Thus we know the value of the pdf at $(\xi, \mathbf{p})$ given $\mathbf{x}$ and $\mathbf{y}$. The maximum likelihood principle aims at maximizing this value. Hence

$$\max_{\xi, \mathbf{p}} \left((2\pi)^{-n} \frac{1}{\sqrt{\det\left(\mathbf{C}\right)}}\, \mathrm{e}^{-\frac{1}{2} \mathbf{z}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{z}} \right) \tag{3}$$

must be solved for $\mathbf{p}$ and ξ. See Section 3 for computational details.

2.1.2 Fitting of $\mathbf{p}$ (P-fit)

In the previous section we fitted for both parameters $\mathbf{p}$ and ξ. Hence we found the most probable pairing of $\mathbf{p}$ and ξ, i.e. an estimation for the true shape of f and simultaneously the true values of the calibration standards $\mathbf{x}$. However, it is often the case that one is not interested in an estimation of the calibration standard values, but rather one is interested in the most probable shape of the calibration function f regardless of any estimation for $\mathbf{x}$.

Consequently, we are looking for the pdf of $\mathbf{y}$ only. This is computed by the *law of total probability* [13]. We show the derivation of it for one calibration pair (x, y). We use $f = f(x, \mathbf{p})$ and $f_x = \frac{\partial}{\partial x} f(x, \mathbf{p})$ as short forms.

$$\begin{aligned}
\mathrm{pdf}(y) &= \int_{-\infty}^{\infty} \mathrm{pdf}\left(y \middle| \xi\right) \mathrm{pdf}\left(\xi\right) \, d\xi \\
&= \int_{-\infty}^{\infty} \mathrm{err}\left(y - f(\xi, \mathbf{p}), \beta\right) \mathrm{err}\left(\xi - x, \alpha\right) \, d\xi \\
&\approx \int_{-\infty}^{\infty} \mathrm{err}\left(y - f - f_x(\xi - x), \beta\right) \mathrm{err}\left(\xi - x, \alpha\right) \, d\xi \\
&= \frac{1}{2\pi} \frac{1}{\alpha\beta} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\frac{(y-f-f_x(\xi-x))^2}{\beta^2}} \mathrm{e}^{-\frac{1}{2}\frac{(\xi-x)^2}{\alpha^2}} \, d\xi \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\alpha^2 f_x^2 + \beta^2}} \mathrm{e}^{-\frac{1}{2}\frac{(y-f)^2}{\alpha^2 f_x^2 + \beta^2}} \\
\Longrightarrow \quad \mathrm{pdf}(y) &\approx \mathrm{err}\left(y - f, \sqrt{\alpha^2 f_x^2 + \beta^2}\right)
\end{aligned}$$

The generalization to the n-dimensional case is now straight forward:

$$\mathrm{pdf}(\mathbf{y}) \approx \mathrm{err}\left(\mathbf{z}, \mathbf{C}\right) = (2\pi)^{-\frac{n}{2}} \frac{1}{\sqrt{\det\left(\mathbf{C}\right)}} \mathrm{e}^{-\frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{z}}$$

with

$$\mathbf{z} = \mathbf{y} - \mathbf{f} \qquad \text{and} \qquad \mathbf{C} = \mathbf{F}_x \mathbf{C}_x \mathbf{F}_x + \mathbf{C}_y.$$

Hence we know the value of the pdf at $\mathbf{p}$ given $\mathbf{x}$ and $\mathbf{y}$. Again using the principle of maximum likelihood, we are left with the maximization problem

$$\max_{\mathbf{p}} \left((2\pi)^{-\frac{n}{2}} \frac{1}{\sqrt{\det\left(\mathbf{C}\right)}} \mathrm{e}^{-\frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{z}} \right). \tag{4}$$

Superficially, this looks equivalent to equation (3), but recall that $\mathbf{C}$ and $\mathbf{z}$ are defined differently. See Section 3 for the computation of $\mathbf{p}$.

2.2 Model proposed in ISO 6143

The method proposed by ISO 6143 [11] is again a maximum likelihood fit:

$$\max_{\xi, \mathbf{p}} \left(K \, \mathrm{e}^{-\frac{1}{2}(\xi - \mathbf{x})^{\mathrm{T}} \mathbf{C}_x^{-1} (\xi - \mathbf{x}) - \frac{1}{2}(\mathbf{f}(\xi, \mathbf{p}) - \mathbf{y})^{\mathrm{T}} \mathbf{C}_y^{-1} (\mathbf{f}(\xi, \mathbf{p}) - \mathbf{y})} \right) \tag{5}$$

with $K = (2\pi)^{-n} (\det\left(\mathbf{C}_x\right) \det\left(\mathbf{C}_y\right))^{-1/2}$ constant. Comparing this formula with equation (3), one can see that it is transformed to (5) by assuming that $\mathbf{C}_x$ and $\mathbf{C}_y$ are both diagonal matrices. This is not always the case [11, p. 22].

However, using this assumption (5), leads to the minimization problem

$$\min_{\xi, \mathbf{p}} \left(\sum_{i=1}^{n} \left(\frac{(\xi_i - x_i)^2}{\alpha_i^2} + \frac{(f(\xi_i, \mathbf{p}) - y_i)^2}{\beta_i^2} \right) \right).$$

This is the minimization problem described in [11, p. 19]. It belongs to the class of *weighted orthogonal distance regression* (Weighted ODR) problems that are thoroughly studied in [2] and [9].

3 Computations

In this section we show how to solve the maximization problems (3) and (4). For quantifying the measurement uncertainty later, when using the calibrated tool, we need the covariances of the parameters $\mathbf{p}$, i.e. the covariance matrix $\mathbf{C_p}$. Its use will be demonstrated in Section 4. We explain the computation of the covariance matrix $\mathbf{C_p}$ at the end of each subsection.

Notation We use the following symbols, notations and abbreviations for readability. For matrices, $\mathbf{M}^{(i)}$ denotes the i-th column vector. Additionally, ∂_k abbreviates the element-wise derivative with respect to $\mathbf{p}_k$.

$$\begin{aligned}
\mathbf{J}_{\mathbf{p}}^{\mathrm{T}} &= \left(\tfrac{\partial}{\partial \mathbf{p}} f^{(1)}, \ldots, \tfrac{\partial}{\partial \mathbf{p}} f^{(n)} \right) \\
\mathbf{J}_{x\mathbf{p}}^{\mathrm{T}} &= \left(\tfrac{\partial}{\partial \mathbf{p}} f_x^{(1)}, \ldots, \tfrac{\partial}{\partial \mathbf{p}} f_x^{(n)} \right) \\
\mathbf{J}_{\mathbf{p}\mathbf{p}_k} &= \partial_k \mathbf{J}_{\mathbf{p}} \\
\mathbf{J}_{x\mathbf{p}\mathbf{p}_k} &= \partial_k \mathbf{J}_{x\mathbf{p}} \\
\mathbf{F}_{x\mathbf{p}_k} &= \partial_k \mathbf{F}_x
\end{aligned}$$

3.1 Computing ξ and p

Consider the maximization problem (3). This is equivalent (after taking the logarithm) to the following minimization problem:

$$\min_{\xi,\mathbf{p}} \left(n \log(2\pi) + \frac{1}{2} \log(\det(\mathbf{C})) + \frac{1}{2} \mathbf{z}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{z} \right).$$

Now we can drop constant terms. Note that $\det(\mathbf{C})$ is also constant. This can be seen by applying a *Cholesky factorization*[1] on $\mathbf{C}$.

$$\mathbf{C} = \mathbf{R}^{\mathrm{T}}\mathbf{R} = \begin{pmatrix} \mathbf{R}_x^{\mathrm{T}} & \mathbf{0} \\ \mathbf{F}_x \mathbf{R}_x^{\mathrm{T}} & \mathbf{R}_y^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \mathbf{R}_x & \mathbf{R}_x \mathbf{F}_x \\ \mathbf{0} & \mathbf{R}_y \end{pmatrix},$$

where $\mathbf{C}_x = \mathbf{R}_x^{\mathrm{T}}\mathbf{R}_x$ and $\mathbf{C}_y = \mathbf{R}_y^{\mathrm{T}}\mathbf{R}_y$. Note, that $\det(\mathbf{C}) = \det(\mathbf{R}_x^{\mathrm{T}})\det(\mathbf{R}_x)\det(\mathbf{R}_y^{\mathrm{T}})\det(\mathbf{R}_y) = \det(\mathbf{R}_x^{\mathrm{T}}\mathbf{R}_x)\det(\mathbf{R}_y^{\mathrm{T}}\mathbf{R}_y) = \det(\mathbf{C}_x)\det(\mathbf{C}_y)$ and since it is independent from ξ and $\mathbf{p}$, it is constant. Lucky us! We are left with the following minimization problem

$$\min_{\xi,\mathbf{p}} \left(\mathbf{z}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{z} \right).$$

Solving this is equivalent to solving a *nonlinear least squares problem*:

$$\begin{aligned}
\min_{\xi,\mathbf{p}} \left(\mathbf{z}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{z} \right) &= \min_{\xi,\mathbf{p}} \left(\mathbf{z}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{R}^{-\mathrm{T}} \mathbf{z} \right) \\
&= \min_{\xi,\mathbf{p}} \left(\mathbf{r}^{\mathrm{T}} \mathbf{r} \right) \\
&= \min_{\xi,\mathbf{p}} \|\mathbf{r}\|^2 = \min_{\xi,\mathbf{p}} \mathcal{S}(\xi, \mathbf{p}).
\end{aligned}$$

There are different standard solvers for such problems, e.g. the *Gauss-Newton* (or the further developed *Levenberg-Marquardt* method), the *Newton* method, the *spectral decomposition* method, etc. [14, 1, 12, 8, 7]. The following condition holds for the computed minimum

$$\mathbf{J}^{\mathrm{T}}\mathbf{r} = \mathbf{0}, \tag{6}$$

where $\mathbf{J}$ denotes the *jacobian* of $\mathbf{r}$ with respect to ξ and $\mathbf{p}$. Its structure is shown in equation (8). The statement above can be verified as follows: we define $\mathbf{w} = (\xi, \mathbf{p})^{\mathrm{T}}$. Linearizing the function $\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \mathbf{J}\Delta\mathbf{w}$ we get $\mathcal{S}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathcal{S}(\mathbf{w}) + 2\Delta\mathbf{w}^{\mathrm{T}}\mathbf{J}^{\mathrm{T}}\mathbf{r} + \mathcal{O}(\|\Delta\mathbf{w}\|^2)$. Thus equation (6) is obviously true in a minimum.

[1] There exists a Cholesky factorization for any symmetric positive definite matrix. This is the case for all covariance matrices by definition.

3.1.1 The Computation of $\mathbf{C_p}$

For a good estimation of the measurement error we will need the covariance matrix $\mathbf{C_p}$ of $\mathbf{p}$. We therefore rewrite equation (6) as

$$\mathbf{h}(\mathbf{w}, \mathbf{y}) = \mathbf{J}^{\mathrm{T}}\mathbf{r} = \mathbf{0}. \tag{7}$$

The function $\mathbf{h}$ is an implicit description of the maximum likelihood criterion. Thus, $\mathbf{w}$ is determined by $\mathbf{y}$ only (and of course $\mathbf{x}$ which is constant for a given set of calibration standards). Since $\mathbf{y}$ is a rv with a known covariance matrix, we are now able to compute the covariance matrix of $\mathbf{w}$ and thus of ξ and $\mathbf{p}$.

We define $\bar{\mathbf{w}} = \left(\bar{\xi}, \bar{\mathbf{p}}_{\mathrm{est}}\right)^{\mathrm{T}} = \mathrm{E}\left[\mathbf{w}\right]$ and $\bar{\mathbf{y}} = \mathrm{E}\left[\mathbf{y}\right]$ as the means of the respective rvs[2]. Then

$$\mathbf{w} = \bar{\mathbf{w}} + \Delta\mathbf{w} \quad \text{and} \quad \mathbf{y} = \bar{\mathbf{y}} + \Delta\mathbf{y}.$$

By definition, we may write

$$\mathbf{h}(\bar{\mathbf{w}}, \bar{\mathbf{y}}) = \mathbf{0}$$

and after linearizing $\mathbf{h}$ and inserting the equations above we get

$$\begin{aligned}
\mathbf{0} &\approx \frac{\partial}{\partial \mathbf{w}}\mathbf{h}(\mathbf{w}, \mathbf{y})\Delta\mathbf{w} + \frac{\partial}{\partial \mathbf{y}}\mathbf{h}(\mathbf{w}, \mathbf{y})\Delta\mathbf{y} \\
&= \mathbf{H_w}\Delta\mathbf{w} + \mathbf{H_y}\Delta\mathbf{y}
\end{aligned}$$

Solving for $\Delta\mathbf{w}$ and rearranging the expressions we get[3]

$$\begin{aligned}
\Delta\mathbf{w} &\approx -\mathbf{H_w}^{-1}\mathbf{H_y}\Delta\mathbf{y} \\
\Longrightarrow \quad \mathrm{Cov}\left(\mathbf{w}, \mathbf{w}\right) &= \mathrm{E}\left[\Delta\mathbf{w}\Delta\mathbf{w}^{\mathrm{T}}\right] \\
&\approx \mathrm{E}\left[\mathbf{H_w}^{-1}\mathbf{H_y}\Delta\mathbf{y}\Delta\mathbf{y}^{\mathrm{T}}\mathbf{H_y}^{\mathrm{T}}\mathbf{H_w}^{-\mathrm{T}}\right] \\
&= \mathbf{H_w}^{-1}\mathbf{H_y}\left(\mathbf{F}_x\mathbf{C}_x\mathbf{F}_x + \mathbf{C}_y\right)\mathbf{H_y}^{\mathrm{T}}\mathbf{H_w}^{-\mathrm{T}}
\end{aligned}$$

The covariance matrix $\mathbf{C_p}$ corresponds to the lower right $m \times m$ block of $\mathrm{Cov}\left(\mathbf{w}, \mathbf{w}\right)$.

Finally we must show how to compute the matrices $\mathbf{H_w}$ and $\mathbf{H_y}$. We start by expanding the expression (7) using the variables introduced throughout Section 2 and 3.1. On the one hand, we have

$$\begin{aligned}
\mathbf{H_y} &= \frac{\partial}{\partial \mathbf{y}}\mathbf{h}(\mathbf{w}, \mathbf{y}) \\
&= \left(\frac{\partial}{\partial \mathbf{y}}\mathbf{J}^{\mathrm{T}}\right)\mathbf{r} + \mathbf{J}^{\mathrm{T}}\left(\frac{\partial}{\partial \mathbf{y}}\mathbf{r}\right) \\
&= \mathbf{0}\cdot\mathbf{r} + \mathbf{J}^{\mathrm{T}}\mathbf{R}^{-\mathrm{T}}\begin{pmatrix}\mathbf{0}\\ \mathbf{I}\end{pmatrix} \\
&= \begin{pmatrix}\mathbf{R}_x^{-\mathrm{T}} & \mathbf{0}\\ -\mathbf{R}_y^{-\mathrm{T}}\mathbf{F}_x & -\mathbf{R}_y^{-\mathrm{T}}\left(\mathrm{diag}\left(\xi - \mathbf{x}\right)\mathbf{J}_{x\mathbf{p}} - \mathbf{J}_{\mathbf{p}}\right)\end{pmatrix}^{\mathrm{T}}\mathbf{R}^{-\mathrm{T}}\begin{pmatrix}\mathbf{0}\\ \mathbf{I}\end{pmatrix} \\
&= \begin{pmatrix}-\mathbf{F}_x\mathbf{C}_y^{-1}\\ -\left(\mathbf{J}_{x\mathbf{p}}{}^{\mathrm{T}}\mathrm{diag}\left(\xi - \mathbf{x}\right) - \mathbf{J}_{\mathbf{p}}{}^{\mathrm{T}}\right)\mathbf{C}_y^{-1}\end{pmatrix}.
\end{aligned} \tag{8}$$

On the other hand, for the i-th column $\mathbf{H_w}^{(i)}$, for $1 \le i \le n$, i.e. the derivatives with respect to ξ_i, we have

$$\begin{aligned}
\mathbf{H_w}^{(i)} &= \left(\frac{\partial}{\partial \xi_i}\mathbf{J}^{\mathrm{T}}\right)\mathbf{r} + \mathbf{J}^{\mathrm{T}}\left(\frac{\partial}{\partial \xi_i}\mathbf{r}\right) \\
&= \begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & -\mathbf{J}_{x\mathbf{p}}{}^{\mathrm{T}}\mathrm{diag}\left(\mathbf{e}_i\right)\mathbf{R}_y^{-1}\end{pmatrix}\mathbf{r} + \mathbf{J}^{\mathrm{T}}\mathbf{R}^{-\mathrm{T}}\begin{pmatrix}\mathbf{e}_i\\ \mathbf{0}\end{pmatrix},
\end{aligned}$$

[2] Note that the mean of several estimated parameter vectors $\bar{\mathbf{p}}_{\mathrm{est}}$ is not necessarily the true and unknown parameter vector.

[3] Notice that the inverse of $\mathbf{H_w}$ exists if f is a well behaving function.

and for $1 \le k \le m$, i.e. the derivatives with respect to $\mathbf{p}_k$, we have

$$\begin{aligned}
\mathbf{H}_{\mathbf{w}}^{(n+k)} &= \left(\partial_k \mathbf{J}^{\mathrm{T}}\right)\mathbf{r} + \mathbf{J}^{\mathrm{T}}\left(\partial_k \mathbf{r}\right) \\
&= \begin{pmatrix} \mathbf{0} & -\mathbf{F}_{x\mathbf{p}_i}\mathbf{R}_y^{-1} \\ \mathbf{0} & -\left(\mathbf{J}_{x\mathbf{pp}_i}^{\mathrm{T}}\operatorname{diag}(\xi - \mathbf{x}) - \mathbf{J}_{\mathbf{pp}_i}^{\mathrm{T}}\right)\mathbf{R}_y^{-1} \end{pmatrix}\mathbf{r} \\
&\quad + \mathbf{J}^{\mathrm{T}}\left(\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ -\mathbf{R}_y^{-\mathrm{T}}\mathbf{F}_{x\mathbf{p}_i} & \mathbf{0} \end{pmatrix}\mathbf{z} + \mathbf{R}^{-\mathrm{T}}\begin{pmatrix} \mathbf{0} \\ -\mathbf{J}_{\mathbf{p}}^{(i)} \end{pmatrix}\right),
\end{aligned}$$

where $\mathbf{e}_i$ is the i-th unity vector.

3.2 Computing p only

Consider the maximization problem (4). This is equivalent to the following minimization problem (after taking logarithms again):

$$\min_{\mathbf{p}}\left(\frac{1}{2}n\log(2\pi) + \frac{1}{2}\log(\det(\mathbf{C})) + \frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{z}\right).$$

Note that the matrix $\mathbf{C}$ used here is different from the one in Section 3.1. $\mathbf{R}$ is again the upper triangular matrix resulting from the Cholesky factorization of the new $\mathbf{C}$. It is readily seen that $\det(\mathbf{C})$ is no longer constant in $\mathbf{p}$ and cannot be neglected. The only constant term we can drop is $\frac{1}{2}n\log(2\pi)$. Thus, we are not so lucky anymore! We are now left with the following general minimization problem

$$\min_{\mathbf{p}}\left(\log(\det(\mathbf{C})) + \mathbf{z}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{z}\right). \tag{9}$$

Solving (9) is no longer equivalent to solving a nonlinear least squares problem, as can be seen by the following derivation

$$\begin{aligned}
& \min_{\mathbf{p}}\left(\log(\det(\mathbf{C})) + \mathbf{z}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{z}\right) \\
= & \min_{\mathbf{p}}\left(\log(\det(\mathbf{R}^{\mathrm{T}}\mathbf{R})) + \mathbf{z}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{R}^{-\mathrm{T}}\mathbf{z}\right) \\
= & \min_{\mathbf{p}}\left(2\log(\det(\mathbf{R})) + \mathbf{r}^{\mathrm{T}}\mathbf{r}\right) = \min_{\mathbf{p}}\mathcal{S}(\mathbf{p})
\end{aligned} \tag{10}$$

Nonetheless the standard techniques mentioned in section 3.1 can be applied to this problem. But the effort to solve problem (10) is larger. The construction of the jacobian is more difficult (see Section 3.2.1). In the minimum the following condition holds:

$$\mathbf{q} + \mathbf{J}^{\mathrm{T}}\mathbf{r} = \mathbf{0} \tag{11}$$

Here $\mathbf{q} = \frac{\partial}{\partial \mathbf{p}}2\log(\det(\mathbf{R}))$ and $\mathbf{J}$ denotes the jacobian of $\mathbf{r}$ with respect to $\mathbf{p}$. Again, as in equation (6), this means that the first derivative with respect to $\mathbf{p}$ is zero. The computation of these entities is shown in the following section.

3.2.1 Computing of $\mathbf{C}_{\mathbf{p}}$

In the first step we derive an expression for

$$\begin{aligned}
\mathbf{q} &= \frac{\partial}{\partial \mathbf{p}}2\log(\det(\mathbf{R})) \\
&= \frac{\partial}{\partial \mathbf{p}}2\log\left(\prod_{i=1}^{n}\mathbf{R}_{i,i}\right) \\
&= 2\frac{\partial}{\partial \mathbf{p}}\sum_{i=1}^{n}\log(\mathbf{R}_{i,i}).
\end{aligned}$$

For the k-th entry of $\mathbf{q}$, we get

$$\begin{aligned}\mathbf{q}_k &= 2\frac{\partial}{\partial \mathbf{p}_k}\sum_{i=1}^{n}\log\left(\mathbf{R}_{i,i}\right) \\ &= 2\sum_{i=1}^{n}\frac{\frac{\partial}{\partial \mathbf{p}_k}\mathbf{R}_{i,i}}{\mathbf{R}_{i,i}} = 2\sum_{i=1}^{n}\frac{\partial_k\mathbf{R}_{i,i}}{\mathbf{R}_{i,i}}.\end{aligned}$$

It remains to show how $\partial_k\mathbf{R}_{i,i}$ can be computed. We begin by writing

$$\begin{aligned}\partial_k\mathbf{C} &= \left(\partial_k\mathbf{F}_x\right)\mathbf{C}_x\mathbf{F}_x + \mathbf{F}_x\mathbf{C}_x\left(\partial_k\mathbf{F}_x\right) \\ &= \mathbf{F}_{x\mathbf{p}_k}\mathbf{C}_x\mathbf{F}_x + \mathbf{F}_x\mathbf{C}_x\mathbf{F}_{x\mathbf{p}_k}.\end{aligned}$$

With $\partial_k\mathbf{C}$ and with the derivation rules applied on $\mathbf{C} = \mathbf{R}^{\mathrm{T}}\mathbf{R}$ we get

$$\partial_k\mathbf{C} = \left(\partial_k\mathbf{R}^{\mathrm{T}}\right)\mathbf{R} + \mathbf{R}^{\mathrm{T}}\left(\partial_k\mathbf{R}\right).$$

This can be easily solved for $\partial_k\mathbf{R}$. The diagonal elements of the latter are the wanted $\frac{\partial}{\partial \mathbf{p}_k}\mathbf{R}_{i,i}$. As in Section 3.1.1 we introduce again an implicit function

$$\mathbf{h}(\mathbf{p},\mathbf{y}) = \mathbf{q} + \mathbf{J}^{\mathrm{T}}\mathbf{r} = \mathbf{0}. \tag{12}$$

Due to the tricky structure of $\mathbf{C}$ we cannot write down an explicit expression for $\mathbf{J}$. However, its computation is still feasible in a column-wise way. We write the k-th column of $\mathbf{J}$ by $\mathbf{J}^{(k)}$, for $1 \leq k \leq m$. Then we have

$$\begin{aligned}\mathbf{J}^{(k)} &= \frac{\partial}{\partial \mathbf{p}_k}\mathbf{r} \\ &= \left(\partial_k\mathbf{R}^{-\mathrm{T}}\right)\mathbf{r} + \mathbf{R}^{-\mathrm{T}}\left(\partial_k\mathbf{r}\right) \\ &= \left(\partial_k\mathbf{R}^{-\mathrm{T}}\right)\mathbf{r} - \mathbf{R}^{-\mathrm{T}}\mathbf{J}_{\mathbf{p}}^{(k)}.\end{aligned}$$

It remains to compute $\partial_k\mathbf{R}^{-\mathrm{T}}$. We have $\mathbf{I} = \mathbf{R}^{-1}\mathbf{R}$. Thus, after taking derivatives on both sides:

$$\begin{aligned}\mathbf{0} &= \left(\partial_k\mathbf{R}^{-1}\right)\mathbf{R} + \mathbf{R}^{-1}\left(\partial_k\mathbf{R}\right) \\ \Longrightarrow \quad \partial_k\mathbf{R}^{-1} &= -\mathbf{R}^{-1}\left(\partial_k\mathbf{R}\right)\mathbf{R}^{-1}\end{aligned}$$

Again, as in section 3.1.1, we have to compute $\mathbf{H}_{\mathbf{p}}$ and $\mathbf{H}_{\mathbf{y}}$. Let $\mathbf{H}_{\mathbf{y}}^{(i)}$ denote the i-th column of $\mathbf{H}_{\mathbf{y}}$. For $1 \leq i \leq n$, we get

$$\begin{aligned}\mathbf{H}_{\mathbf{y}}^{(i)} &= \frac{\partial}{\partial y_i}\left(\mathbf{q} + \mathbf{J}^{\mathrm{T}}\mathbf{r}\right) \\ &= \mathbf{0} + \left(\frac{\partial}{\partial y_i}\mathbf{J}^{\mathrm{T}}\right)\mathbf{r} + \mathbf{J}^{\mathrm{T}}\left(\frac{\partial}{\partial y_i}\mathbf{r}\right) \\ &= \left(\frac{\partial}{\partial y_i}\left(\mathbf{J}^{(1)},\ldots,\mathbf{J}^{(m)}\right)^{\mathrm{T}}\right)\mathbf{r} + \mathbf{J}^{\mathrm{T}}\mathbf{R}^{-\mathrm{T}}\mathbf{e}_i \\ &= \left(\left(\partial_1\mathbf{R}^{-\mathrm{T}}\right)\mathbf{R}^{-\mathrm{T}}\mathbf{e}_i,\ldots,\left(\partial_m\mathbf{R}^{-\mathrm{T}}\right)\mathbf{R}^{-\mathrm{T}}\mathbf{e}_i\right)^{\mathrm{T}}\mathbf{r} + \mathbf{J}^{\mathrm{T}}\mathbf{R}^{-\mathrm{T}}\mathbf{e}_i.\end{aligned}$$

The derivation for $\mathbf{H}_{\mathbf{p}}$ is exceedingly odd. Nevertheless here it comes: First we have

$$\partial_l\mathbf{q}_k = 2\sum_{i=1}^{n}\frac{\left(\partial_l\partial_k\mathbf{R}_{i,i}\right)\mathbf{R}_{i,i} - \left(\partial_k\mathbf{R}_{i,i}\right)\left(\partial_l\mathbf{R}_{i,i}\right)}{\mathbf{R}_{i,i}^2},$$

$(\partial_l\partial_k\mathbf{R}_{i,i})$ can be found by solving

$$\partial_l\partial_k\mathbf{C} = \left(\partial_l\partial_k\mathbf{R}^{\mathrm{T}}\right)\mathbf{R} + \left(\partial_l\mathbf{R}^{\mathrm{T}}\right)\left(\partial_k\mathbf{R}\right) + \left(\partial_k\mathbf{R}^{\mathrm{T}}\right)\left(\partial_l\mathbf{R}\right) + \mathbf{R}^{\mathrm{T}}\left(\partial_l\partial_k\mathbf{R}\right).$$

Further we need

$$
\begin{aligned}
\partial_l \mathbf{J}^{\mathrm{T}}\mathbf{r} &= \left(\partial_l \mathbf{J}\right)^{\mathrm{T}}\mathbf{r} + \mathbf{J}^{\mathrm{T}}\left(\partial_l \mathbf{r}\right) \\
&= \left(\partial_l \mathbf{J}^{(1)}, \ldots, \partial_l \mathbf{J}^{(m)}\right)^{\mathrm{T}}\mathbf{r} + \mathbf{J}^{\mathrm{T}}\left(\partial_l \mathbf{r}\right) \\
\partial_l \mathbf{r} &= \left(\partial_l \mathbf{R}^{-\mathrm{T}}\right)\mathbf{z} - \mathbf{R}^{-\mathrm{T}}\mathbf{J}_{\mathbf{p}}^{(l)} \\
\partial_l \mathbf{J}^{(k)} &= \left(\partial_l \partial_k \mathbf{R}^{-\mathrm{T}}\right)\mathbf{r} + \left(\partial_k \mathbf{R}^{-\mathrm{T}}\right)\left(\partial_l \mathbf{r}\right) - \left(\partial_l \mathbf{R}^{-\mathrm{T}}\right)\mathbf{J}_{\mathbf{p}}^{(k)} - \mathbf{R}^{-\mathrm{T}}\left(\partial_l \mathbf{J}_{\mathbf{p}}^{(k)}\right).
\end{aligned}
$$

Putting things together yields

$$\mathbf{H}_{\mathbf{p}}^{(l)} = \partial_l \mathbf{q} + \partial_l \mathbf{J}^{\mathrm{T}}\mathbf{r},$$

where $\partial_l \mathbf{q}$ is of course the vector with elements $\partial_l \mathbf{q}_k$, for $k = 1, \ldots, m$.

4 Measurement

The first goal of calibration is to compute a good set of parameters to quantify the measurement tool. An equally important goal is to estimate the measurement uncertainty when using the calibrated device.

Assume that the measuring tool is calibrated. Thus we have a set of estimated parameters $\hat{\mathbf{p}}$ together with an estimated covariance matrix $\hat{\mathbf{C}}_{\mathbf{p}} = \mathrm{Cov}\left(\hat{\mathbf{p}}, \hat{\mathbf{p}}\right)$.

Let $\mathbf{y} \in \mathbb{R}^s$ be a measurement vector whose elements are readings taken from the measurement tool. The readings are subject to errors, of the form $\mathbf{y} - \bar{\mathbf{y}} = \mathcal{N}(\mathbf{0}, \mathbf{E})$, where $\bar{\mathbf{y}}$ is the correct but unknown reading vector.

For $1 \le i \le s$ compute $\hat{x}_i$ by applying $\hat{x}_i = g(y_i, \hat{\mathbf{p}})$ or by solving $y_i = f(\hat{x}_i, \hat{\mathbf{p}})$. For arbitrary (but still well behaved) functions $f(x, \hat{\mathbf{p}})$ we propose the use of a standard solver such as *bisection*, *regula falsi*, *Newton method* etc.

The computation of the measurement uncertainty is performed as follows. We represent the exact physics with $\bar{y}_i = f(\bar{x}_i, \mathbf{p}_{\mathrm{ex}})$, where these are the exact and never known values. The given values are $\bar{x}_i + \Delta x_i = \hat{x}_i$, $\bar{y}_i + \Delta y_i = y_i$ and $\mathbf{p}_{\mathrm{ex}} + \Delta \mathbf{p} = \hat{\mathbf{p}}$. Substituting and linearizing these equations leads to

$$\Delta y_i \approx f_x^{(i)} \Delta x_i + \nabla_{\mathbf{p}} f^{(i)\,\mathrm{T}} \Delta \mathbf{p},$$

where $f^{(i)} = f(\hat{x}_i, \hat{\mathbf{p}})$ and $f_x^{(i)} = f_x(\hat{x}_i, \hat{\mathbf{p}})$. Extending this approximation to s readings $\mathbf{y}$ and solving for $\Delta \mathbf{x}$, we get

$$\Delta \mathbf{x} \approx \mathbf{F}_x^{-1} \Delta \mathbf{y} - \mathbf{F}_x^{-1} \mathbf{J}_{\mathbf{p}} \Delta \mathbf{p}$$

and furthermore

$$
\begin{aligned}
\mathbf{C}_x &= \mathrm{E}\left[\Delta \mathbf{x} \Delta \mathbf{x}^{\mathrm{T}}\right] \\
&\approx \quad \mathbf{F}_x^{-1}\,\mathrm{Cov}\left(\mathbf{y}, \mathbf{y}\right)\mathbf{F}_x^{-1} \\
&\quad + \mathbf{F}_x^{-1}\mathbf{J}_{\mathbf{p}}\,\mathrm{Cov}\left(\hat{\mathbf{p}}, \hat{\mathbf{p}}\right)\mathbf{J}_{\mathbf{p}}{}^{\mathrm{T}}\mathbf{F}_x^{-1} \\
&\quad - 2\mathbf{F}_x^{-1}\,\mathrm{Cov}\left(\mathbf{y}, \hat{\mathbf{p}}\right)\mathbf{J}_{\mathbf{p}}{}^{\mathrm{T}}\mathbf{F}_x^{-1}.
\end{aligned}
$$

There are two thing to note. First, $\mathrm{Cov}\left(\mathbf{y}, \hat{\mathbf{p}}\right)$ can safely be assumed to be $\mathbf{0}$, since the uncertainties of the reading for the s measurements are independent of $\hat{\mathbf{p}}$ found during calibration. Second, it is obvious that no $f_x^{(i)}$ is not allowed to be 0. In practice this does not present a problem, since nobody would use a measurement device in a range where the slope of the measurement function is too large.

So we end up with the calculation for

$$\mathbf{C}_x \approx \mathbf{F}_x^{-1}\left(\mathrm{Cov}\left(\mathbf{y}, \mathbf{y}\right) + \mathbf{J}_{\mathbf{p}}\,\mathrm{Cov}\left(\hat{\mathbf{p}}, \hat{\mathbf{p}}\right)\mathbf{J}_{\mathbf{p}}{}^{\mathrm{T}}\right)\mathbf{F}_x^{-1}.$$

5 Numerical Comparisons and Simulations

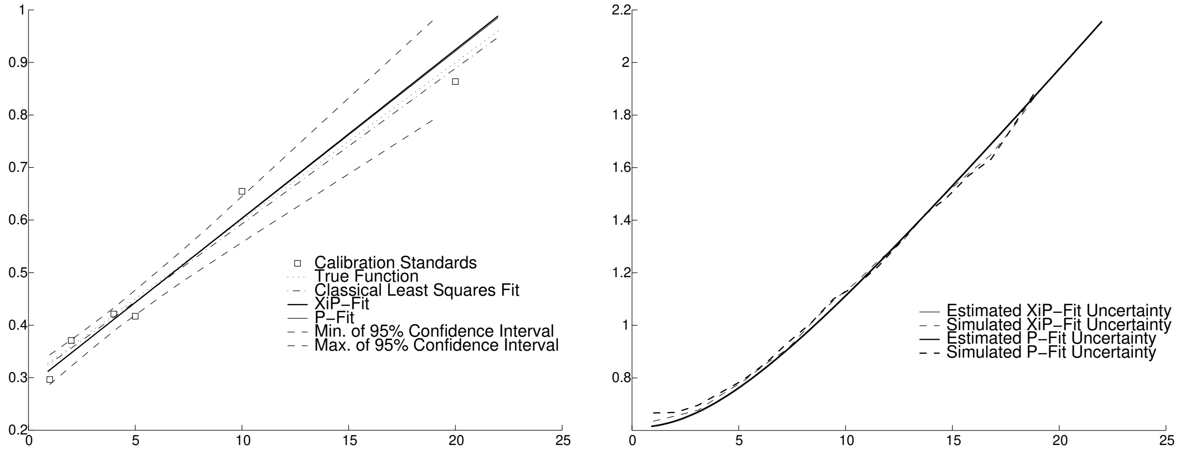
We show four simulations with different calibration functions. The comparison with the predicted values is quite encouraging. The left hand side figure shows the fitted functions using XIP-FIT (thick line), P-FIT (thin line) and the classical least squares fit (dash-dotted line). The two dashed lines build an approximated 95%–confidence region for the XIP-FIT.

The right hand side figure shows the simulated measurement uncertainty for the XIP-FIT and the P-FIT (dashed lines) compared to the estimated uncertainties (solid lines).

The calibration data was generated randomly according to figure 1 using the true parameters $\mathbf{p}_{\text{ex}}$. The covariances in x and y are known. The uncertainty in y was chosen to be equal at calibration and at measurement.

5.1 Linear Fit

The calibration function is $f(x, \mathbf{p}) = p_1 + p_2 x$.



(a) (b)

Figure 2: **(a)** Fitting a linear function with different methods **(b)** Comparison of estimated and simulated measurement uncertainty

The calibration data for figure 2 is

x	1.0000	2.0000	4.0000	5.0000	10.0000	20.0000
y	0.2960	0.3712	0.4222	0.4201	0.6402	0.8675

The covariance in x is given by the following matrix:

$$\mathbf{C}_x = \begin{pmatrix} 0.0012 & 0.0016 & 0.0016 & & & \\ 0.0016 & 0.0032 & 0.0032 & & & \\ 0.0016 & 0.0032 & 0.0064 & & & \\ & & & 0.0100 & & \\ & & & & 0.0400 & \\ & & & & & 0.1600 \end{pmatrix}$$

The covariance in y is $\mathbf{C}_y = \operatorname{diag}(\texttt{0.0003, 0.0003, 0.0004, 0.0005, 0.0009, 0.0020})$. Finally the fitted parameters $\mathbf{p}$ are as follows[4]:

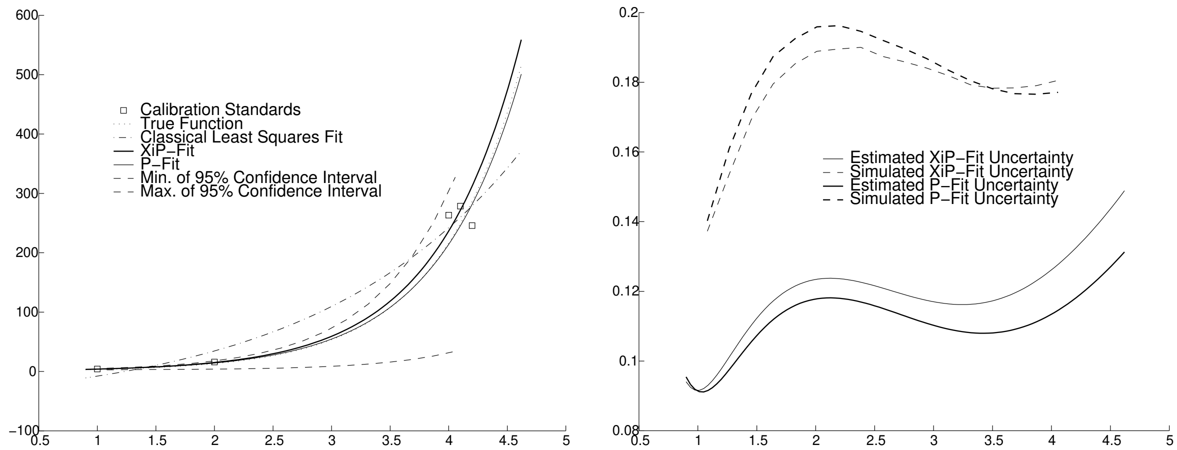
[4] The ISO-Method is implemented in `B-Least`, a software that BAM provides.

	$\mathbf{p}_1$	$\mathbf{p}_2$
Exact	0.3	0.03
Classic LS	0.2956	0.029
XiP-FIT	0.28567	0.031282
P-FIT	0.28577	0.031256
ISO-Method	0.28577	0.031272

Remarks The match of estimated and simulated measurement uncertainty is quite nice. Note that the measurement uncertainty[5] is chosen to be 2%.

5.2 Exponential Fit

The calibration function is $f(x, \mathbf{p}) = p_1 + p_2 e^{\mathbf{p}_3 x}$.



(a) (b)

Figure 3: **(a)** Fitting an exponential function with different methods **(b)** Comparison of estimated and simulated measurement uncertainty

The data for figure 3 is

x	1.0000	2.0000	4.0000	4.1000	4.2000
y	4.1682	15.9362	263.3923	278.6612	245.8233

The covariance in x is given by the following matrix:

$$\mathbf{C}_x = \begin{pmatrix} 0.0075 & 0.0100 & 0.0100 & & \\ 0.0100 & 0.0200 & 0.0200 & & \\ 0.0100 & 0.0200 & 0.0400 & & \\ & & & 0.0420 & \\ & & & & 0.0441 \end{pmatrix}$$

The covariance in y is $\mathbf{C}_y = \mathrm{diag}\,(\texttt{0.0280, 0.4393, 117.1169, 154.9422, 204.9871})$. Finally the fitted parameters $\mathbf{p}$ are as follows:

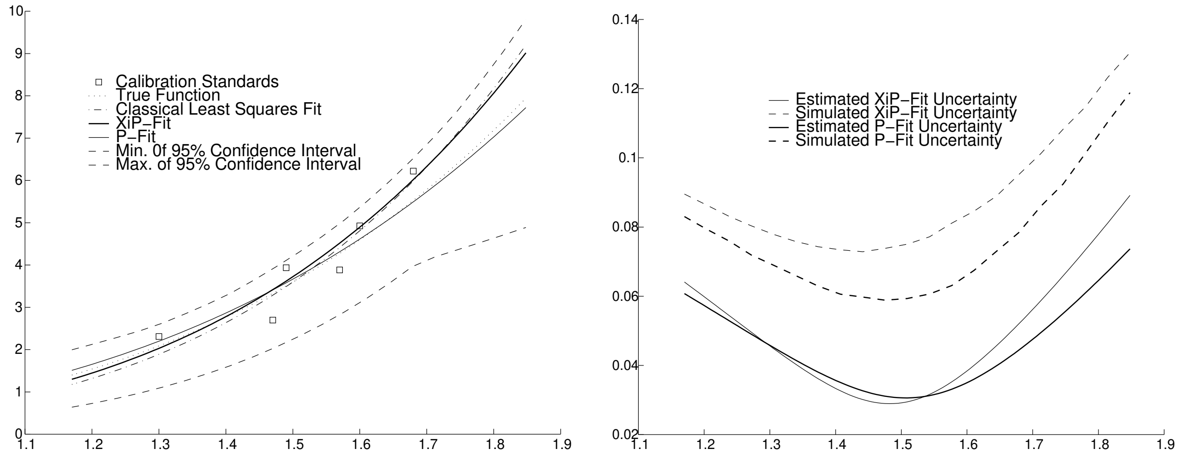
[5]The uncertainty induced by the reading mechanism

	$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$
Exact	0.1	0.8	1.4
Classic LS	-62.9821	31.2164	0.5699
XIP-FIT	0.39654	0.91511	1.3883
P-FIT	0.47711	0.88251	1.3724
ISO-Method	-0.060929	1.1143	1.3338

Remarks The uncertainties in the calibration and the measurement data are 5%. This is quite considerable. The errors induced by linearization become now obvious. The simulated uncertainties are thus bigger but qualitatively still acceptable. Note also that the classical least squares fit is way out of bounds. The ISO-Method, despite the different numbers, would be just crowding the picture between the XIP-FIT and the P-FIT. There was a warning of the `B-Least` software about a division by zero, but the output seems still reasonable.

5.3 Power Function Fit

The calibration function is $f(x, \mathbf{p}) = p_1 x^{\mathbf{p}_2}$.



(a) (b)

Figure 4: **(a)** Fitting a power function with different methods **(b)** Comparison of estimated and simulated measurement uncertainty

Some of the data for figure 4 is taken from NIST[6].

x	1.3000	1.4700	1.4900	1.5700	1.6000	1.6800
y	2.3093	2.6972	3.9351	3.8846	4.9255	6.2209

The covariance in x is $\mathbf{C}_x = \operatorname{diag}(\texttt{0.0027, 0.0035, 0.0036, 0.0039, 0.0041, 0.0045})$. The covariance in y is $\mathbf{C}_y = \operatorname{diag}(\texttt{0.0070, 0.0177, 0.0196, 0.0292, 0.0338, 0.0489})$. Finally the fitted parameters $\mathbf{p}$ are as follows:

	$\mathbf{p}_1$	$\mathbf{p}_2$
Exact	0.7700	3.8000
Classic LS	0.5801	4.5005
XIP-FIT	0.6676	4.2387
P-FIT	0.8618	3.5703

[6] `http://www.itl.nist.gov/div898/strd/index.html` → Nonlinear Regression → Dan Wood

Remarks The uncertainties in the calibration and the measurement data are 4%. Again the linearization induces a shift between the estimated and the simulated measurement uncertainty. Note that the P-FIT is visually better in the sense that it is closer to the exact function. This is the case for most of the experiments we have studied.

5.4 Sigmoid Function Fit

The calibration function is $f(x, \mathbf{p}) = \frac{p_1}{1+e^{p_2-p_3x}}$

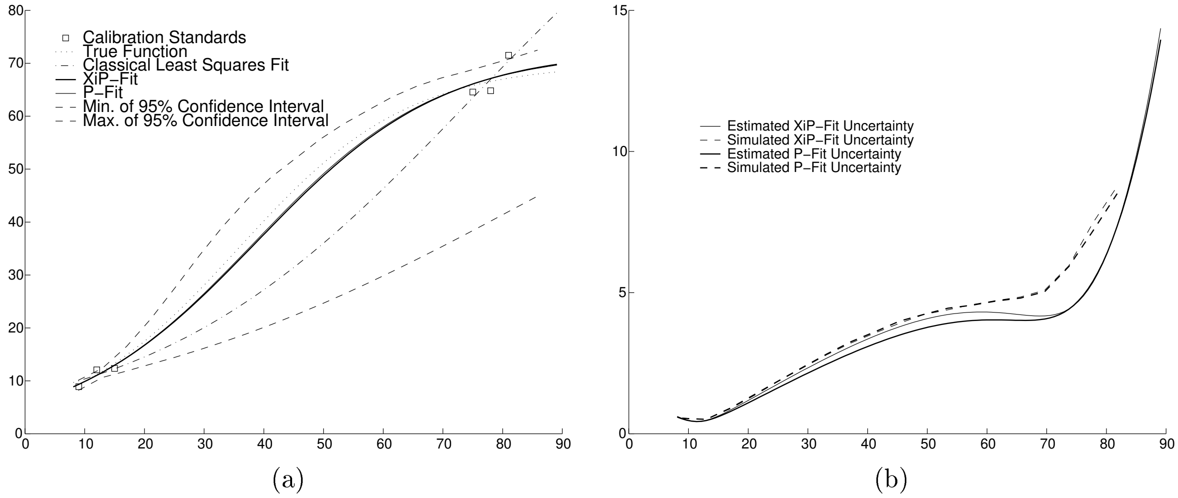


Figure 5: **(a)** Fitting a sigmoid function with different methods **(b)** Comparison of estimated and simulated measurement uncertainty

Some of the data for figure 5 is taken from NIST[7]. They assign it a "higher level of difficulty".

x	9	12	15	75	78	81
y	8.8916	12.0857	12.3492	64.5600	64.8196	71.5168

The covariance in x is $\mathbf{C}_x = \text{diag}\,(\texttt{0.0729, 0.1296, 0.2025, 5.0625, 5.4756, 5.9049})$. The covariance in y is $\mathbf{C}_y = \text{diag}\,(\texttt{0.0786, 0.1126, 0.1592, 3.8960, 3.9862, 4.0616})$. The fitted parameters $\mathbf{p}$ are as follows:

	$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$
Exact	70.0000	2.5000	0.0700
Classic LS	124.6656	2.7761	0.0375
XIP-FIT	72.5830	2.4889	0.0641
P-FIT	72.2162	2.4931	0.0650

Remarks The uncertainties in the calibration and the measurement data are 3%. As opposed to the other examples estimated and the simulated measurement uncertainty are very close together. This is due to the fact that the uncertainties in the calibration data are relatively small, which is the case for reasonable calibrations anyhow. Notice how the measurement uncertainty seems to explode as soon as the calibration function's slope is heading towards zero.

[7] `http://www.itl.nist.gov/div898/strd/index.html` → Nonlinear Regression → Rat42

6 Conclusion

Taking the of calibration data into account is important for cases with very strong correlation or with relatively large uncertainties. However, there is no sound reason to ignore it, since it does not seriously affect the runtime behaviour.

The P-FIT leads to slightly higher measurement uncertainties, but it recovers the "exact" parameters generally better (in our experiments). This is explainable with it's property of "integrating over all possible true x–values".

Classical least squares fitting is not suitable for calibration tasks.

References

[1] Åke Björck. *Numerical Methods for Least Squares Problems.* SIAM, Philadelphia, 1996. ISBN 0-89871-360-9.

[2] Paul T. Boggs, Richard H. Byrd, and Robert B. Schnabel. A stable and efficient algorithm for nonlinear orthogonal distance regression. *SIAM Journal Sci. and Stat. Computing*, 8(6):1052–1078, November 1987.

[3] Paul T. Boggs and Janet E. Rogers. The computation and use of the asymptotic covariance matrix for measurement error models. Technical Report IR 89-4102, NIST, 1990. U.S. Government Printing Office.

[4] Wolfram Bremser and Werner Hässelbarth. Controlling uncertainty in calibration. *Analytica Chimica Acta*, pages 61–69, 1997.

[5] R.J. Carroll, D. Ruppert, and L.A. Stefanski. *Measurement Error in Nonlinear Models.* Number 63 in Monographs on Statistics and Applied Probability. Chapman and Hall, London, 1995. ISBN 0-412-04721-7.

[6] Wayne A. Fuller. *Measurement Error Models.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., 1987. ISBN 0-471-86187-1.

[7] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Practical Optimization.* Academic Press Inc., London, 1981.

[8] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Numerical Linear Algebra and Optimization*, volume 1. Addison-Wesley, 1991.

[9] Sabine Van Huffel, editor. *Recent Advances in Total Least Squares Techniques and Errors-In-Variables Modelling*, Proceedings of the 2nd International Workshop. Society for Industrial and Applied Mathematics, SIAM, 1997. ISBN 0-89871-393-5.

[10] ISO, IEC, BIPM, IFCC, IUPAC, IUPAP, and OIML. *Guide to the Expression of Uncertainty in Measurement.* ISO Central Secretariat, 1st edition, 1995.

[11] ISO, IEC, BIPM, IFCC, IUPAC, IUPAP, and OIML. *ISO 6143*, chapter B.2, pages 18–31. ISO Central Secretariat, 2000.

[12] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in* C. Cambridge University Press, 2 edition, 1996.

[13] John A. Rice. *Mathematical Statistics and Data Analysis.* Duxbury Press, 2nd edition, 1995.

[14] H. R. Schwarz. *Numerische Mathematik.* B.G. Teubner, Stuttgart, 4th edition, 1997. ISBN 3-519-32960-3.