Tina Memo No. 2015-006
Internal, Leverhulme preliminary work

Numerical Estimation of Parameter Covariances.

Neil A Thacker, Ashley Seepujak, Paul D Tar, Georgios Krokos.

Last updated
20 / 05 / 2015

Centre for Imaging Sciences,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Numerical Estimation of Parameter Covariances.

Neil A Thacker, Ashley Seepujak, Paul D Tar, Georgios Krokos. 20/05/2015

Abstract

The conventional method for estimation of parameter covariances, following a model fitting procedure, involves an analytical estimate of the second derivatives. This makes use of the minimum variance (or Kramer-Rao) bound. We present an approach based upon direct inspection of the behaviour of the cost function around the minima. The new approach is not only more convenient, but also provides a better approximation to the average behaviour of the cost function, whilst supporting useful tests of cost function behaviour. A test for asymmetry can indicate poor cost function construction and poor parameter specification. Elimination of these problems can lead to better optimisation and to more appropriate covariances. In general therefore, numerical estimation of covariances can result in both better software and algorithm design methodology.

1 The Motivation for a Numerical Estimation of Covariances

The purpose of many software algorithms is the estimation of parameters and the best algorithms are designed using statistical principles. Generally, we can relate the information available in data s_1, to the information we desire s_2, using the probability $P(s_2|s_1)$ or something monotonically related to it. The most common approach is to use Likelihood; many algorithms are based around the optimisation of a Likelihood-based cost function, using one of a number of optimisation routines.

A meaningful solution which minimises the cost function represents the optimal solution. A necessary condition for locating the optima of the cost function is that both the cost function and the feasible region are convex [9] in the optimisation problem. Most methods for locating the optimum in fact erroneously locate local optima, as opposed to the global optima. Discontinuous cost functions are particularly susceptible to this problem, where there may be several local optima. It is therefore always important to try to ensure that cost functions do not have numerical problems, which generate unnecessary discontinuities.

Although a cost function can be optimised using brute-force methods (testing every possible combination of parameter, and subsequently choosing the optimal set), generally it is much more efficient to use a search process. Search processes are more efficient when we can assume some properties of the underlying cost function. General cost functions can exhibit a range of behaviours, generating a consequent range of problems for optimisation algorithms. In order of difficulty, valid cost functions (i.e. those based appropriately on probability theory) can be : smooth , differentially discontinuous, noisy or even stochastic (i.e. non-deterministic). An approximately quadratic function might be considered the simplest and can be optimised using a number of efficient methods, generally involving the need to compute the derivatives of the cost function with respect to each parameter. One category of approach is referred to as 'quasi-Newton', and includes methods which iterate towards a solution utilising a series of matrix inverses or linear searches. Specific algorithms include methods such as Levenberg-Marquadt [1] and conjugate gradient [2,3].

Though there are slight differences in numerical stability, most common optimisation schemes are well-enough understood to be used to locate a reliable solution when the cost function is smooth and continuous. Provided that any choice of algorithm ultimately converges on the same optimal solution, then its statistical characteristics must be the same, irrespective of optimisation method. Therefore, the properties of the parametric solution (for example, estimation error and correlations) are dictated by the underlying probability theory, and not by the algorithm used to find it, despite often strong contrary opinions of developers (evaluations often show indistinguishable perfomance for competing approaches [4,5]). Choice of algorithm therefore, normally depend upon other issues, for example, robustness to local optima, computational efficiency, ease of application and auxiliary factors, such as the provision of parameter covariances. Parameter covariances are particularly important for scientific applications, where we need to use the parameters in hypothesis tests, or to combine them with other such estimates.

Quasi-Newton approaches are often selected due to good efficiency, but also provide easy access to parameter covariances (the required second derivatives being computed as part of the optimisation algorithm). To achieve this efficiency, however, they rely on a cost function which is quadratic in the region of the minimum. Nevertheless, in the context of basic research, we often do not know the behaviour of a novel Likelihood function; the requirement of an approximate quadratic behaviour of the cost function is something which one cannot simply expect. We

will generally be interested in proving that there is sufficient information in a data-set to estimate a number of parameters. Under these circumstances, it is often sensible to start by using a slower, but more robust, search-based approach, such as the downhill simplex algorithm of Nealder and Meade [6]. It is more robust, specifically, because it does not assume anything regarding the general behaviour of the cost function, and even works well with numerically-unstable, discontinuous or noisy cost functions, having quite good tolerance to local optima.

Simplex is more convenient than most other approaches for getting something working quickly, as it requires only the construction of the cost function, which is used to assess specific parameter choices. Importantly, the derivatives are not needed, and this gives a significant advantage when developing software. In a research context, computing numerous derivatives correctly, and with appropriate numerical stability for a novel form of a Likelihood function, can take many weeks of effort, even for a programmer with numerical analysis experience. If we want to ensure that the method has been implemented properly, and will work reliably on large numbers of datasets, new software generally needs to be tested. Analytic derivative calculations should be numerically tested using finite differences in order to identify, not only mathematical/programming errors, but any numerical imprecision arising due to rounding error. Furthermore, for many optimisation tasks, computational efficiency (particularly on modern machines) is not a huge issue, and the simplex solution is good enough. It is therefore most unfortunate that having implemented a cost function, and having been able to get parameter estimates working quickly (often in a few hours), we then find that we still need the derivatives of the cost function in order to obtain the parameter covariances needed for science.

What is required is a numerical approach for estimating parameter covariances, which is also consistent with the way that derivatives are tested, so that implementations of analytical expressions can be evaluated more efficiently, or avoided altogether. However, if we are developing a numerical method for the estimation of parameter covariances based entirely upon the cost function, then there is another issue which we need to take into account.

1.1 Equal Variance Parameter Spaces (Homoscedasticity)

When building parametric models, we have considerable choice in how we define the parameters. In particular, any parameter can be equivalently represented mathematically by any monotonic mapping function $a' = f(a)$; for example $x = ax^b$ is mathematically identical to $x = ax^{\frac{1}{2c}}$. We would normally choose a parameter definition consistent with a familiar quantity. For example, in physics, we would not normally use the fourth root of mass ($m^{\frac{1}{4}}$), rather than simply mass (m). If we look in conventional statistical texts, for information regarding the selection of model parameters for use in maximum Likelihood, the general view is that such changes in parameter definitions are irrelevant.

Likelihood, and therefore the location of the maximum, is invariant to re-parametrisation. However, whilst this is true of the optimal parameter values, it is not quite true of the behaviour of the cost function around this optima, and therefore the parameter covariances. Parameter covariances are nothing more than a second-order (quadratic) approximation to the the cost function. A poor choice of parameter definition (with non-quadratic minima) will lead to sub-optimal approximation. Typically, poor descriptions of estimation uncertainty and poor statistical efficiency will result when we use these parameters and covariances for scientific tasks.

For a pure quadratic cost function, the second derivative of the function is invariant, i.e. the Fisher information and therefore the expected minimum variance bound (MVB) are uniform (homoscedasticity). The cost function will be symmetrical around the optimal value. For non-quadratic cost functions, the second derivative varies across parameter space; consequently the Fisher information will vary, as will the MVB (heteroscedasticity). Furthermore, provided that the dominant variation from quadratic behaviour is due to the next third order term in the Taylor expansion, then the function will not be symmetrical around the optima.

Although the MVB is generally defined in terms of the expectation of the second derivative (taking into account the uncertainties in the measurement), analytic computational approaches for determination of parameter covariances are based only on the second derivatives of the cost function at the optima. This calculation is exact for quadratic cost functions but provides no information regarding higher-order behaviour. In contrast, a finite difference based numerical approach provides the opportunity to test for asymmetry directly, thereby allowing us to evaluate our choice of parametrisation, and to modify it, if required. However, as for the MVB, this calculation needs to take appropriate account of the expected range of uncertainty. Provided that the general asymmetric behaviour is consistent between data-sets, non-linear mappings of the parameters can be used in the same way as is often done for measured data (for example, Anscombe transformations [7,8]), in order to improve general statistical efficiency. This will be discussed further below.

1.2 Numerical Calculation of the Covariance Matrix

For many uses, errors do not have to be estimated precisely, but statistical efficiency is improved if they are. The parameter error covariance is related to the second derivative of the corresponding log Likelihood via the MVB. The common procedure of using the second derivative evaluated at the optima, is unreliable for noisy (stochastic) and discontinuous cost functions, or in circumstances where poor selection of fit parameters leads to non-quadratic behaviour. However, direct inspection of the behaviour of the original cost function can tell us about both parameter uncertainty and its suitability. In the method described below, parameters are adjusted to step around the cost function minimum, in order to establish the quadratic behaviour, over a range similar to one standard deviation of the parameter error (this is the range over which we prefer the covariance to be accurate).

The use of Likelihood involves construction of the probability of observing the data $\mathbf{d}$ given the assumed model and parameters $\mathbf{x}$, specifically $P(\mathbf{d}|\mathbf{x})$. Consider calculation of a suitable log Likelihood function, for example a χ^2 value (i.e. $-2\ log[P(\mathbf{d}|\mathbf{x})]$). We can approximate the original function to second-order using the difference vector $\mathbf{\Delta_x}$ and its transpose, along with the inverse covariance matrix C^{-1},

$$\chi^2 \ = \ \chi_0^2 + \mathbf{\Delta_x}^T C^{-1} \mathbf{\Delta_x} \ + \ h.o.t. \tag{1}$$

The difference between the χ^2 value and the value about χ_0 is given by,

$$\mathbf{\Delta_x}^T C^{-1} \mathbf{\Delta_x} \ \approx \ \chi^2 - \chi_0^2 \tag{2}$$

In terms of matrix elements, the above can be used to evaluate the expected change in the original cost function, due to a change in the estimated parameters. For example, for a 2D system ($\mathbf{x} \ = \ (x, y)$), we can define the parameter change Δx and write,

$$\begin{pmatrix} \Delta x \\ 0 \end{pmatrix}^T \begin{bmatrix} \alpha_{xx} & \alpha_{xy} \\ \alpha_{xy} & \alpha_{yy} \end{bmatrix} \begin{pmatrix} \Delta x \\ 0 \end{pmatrix} \ = \ \Delta x^2 \alpha_{xx} \tag{3}$$

If the difference between the χ^2 value and the value about χ_0 is unity, then,

$$\chi^2 - \chi_0^2 \ = \ 1 \Rightarrow 1 \ = \ \Delta x^2 \alpha_{xx} \tag{4}$$

i.e.,

$$\Delta x^2 \ = \ \frac{1}{\alpha_{xx}} \tag{5}$$

Similarly,

$$\Delta y^2 \ = \ \frac{1}{\alpha_{yy}} \tag{6}$$

Eqns. (5) and (6) thus provide the means for calculating the diagonal terms of the inverse covariance matrix. The values of Δx and Δy are adjusted until they generate a change in 1 of the χ^2 (see the code detailed in Appendix A). As we have two possible signs for Δx, we can make two independent estimates and use these to test the symmetry of the minimum ($(\Delta x_+ - \Delta x_-)/(\Delta x_+ + \Delta x_-)$). Now, suppose that we wish to calculate the non-diagonal terms of the inverse covariance matrix, i.e. α_{xy}. Let,

$$z = \chi^2 \left(\Delta x, \Delta y\right) - \chi_0^2 \tag{7}$$

Expanding (7) gives,

$$\chi^2 \left(\Delta x, \Delta y\right) - \chi_0^2 \ = \ \Delta x^2 \alpha_{xx} + \Delta x \Delta y \alpha_{xy} + \Delta x \Delta y \alpha_{xy} + \Delta y^2 \alpha_{yy} \tag{8}$$

From (5) and (6), $\Delta x^2 \alpha_{xx} = 1$ and $\Delta y^2 \alpha_{yy} = 1$. Substituting 1 into (8) for both $\Delta x^2 \alpha_{xx}$ and $\Delta y^2 \alpha_{yy}$, gives,

$$\chi^2 \left(\Delta x, \Delta y\right) - \chi_0^2 \ = \ 2 + 2 \Delta x \Delta y \ \alpha_{xy} \tag{9}$$

Substituting (9) into (7) for $\chi^2 \left(\Delta x, \Delta y\right) - \chi_0^2$, gives,

$$z \ = \ 2 + 2 \Delta x \Delta y \ \alpha_{xy} \tag{10}$$

Thus,

$$\alpha_{xy} \ = \ \frac{z - 2}{2 \Delta x \Delta y} \tag{11}$$

Although this process is derived specifically for a 2D system, we can equally apply these calculations to elements of a covariance for any number of parameters, so this can be seen as a general approach to numerical estimation of the inverse covariance (see code detailed in Appendix B).

In order to test this approach, we have taken three current projects within our Group which require the estimation of parameter covariances, and investigated the reliability of the information determined. In particular, we are interested in the level of agreement between numerical and analytic approaches, and the reliability and nature of asymmetries seen in Δx.

2 Tests of Numerical Covariances

2.1 Covariances for Error Models

In previous work, we used a power-law model to approximate general noise models. The expected error (standard deviation) on measurement x is modelled as

$$\epsilon_x \;=\; ax^b$$

For the case of mass spectra, we wish to confirm that data in the signal peaks have Poisson-like measurement noise. A Bland-Altman plot is a scatter-plot of individual values of the data vector, versus the fitted residual of that data-point. A noise data vector ϵ_x exhibiting a Poisson-like characteristic will possess an error function with a power-law function s.t. $b \approx \frac{1}{2}$. That is, the statistical error is proportional to the square root of the value of that data-point. A typical Bland-Altman plot for several hundred repeat measurements of peaks from mass spectra ($(x_1[j]+x_2[j])/2$ vs $R_{12[j]} = x_1[j]-x_2[j]$) are shown in Fig. 1. Data-points with value less than the observed level of background noise (in this case 50), are purposely excluded from the population sample space, since these do not possess the same noise characteristics as the signal which we are interested in (see [10] for details).

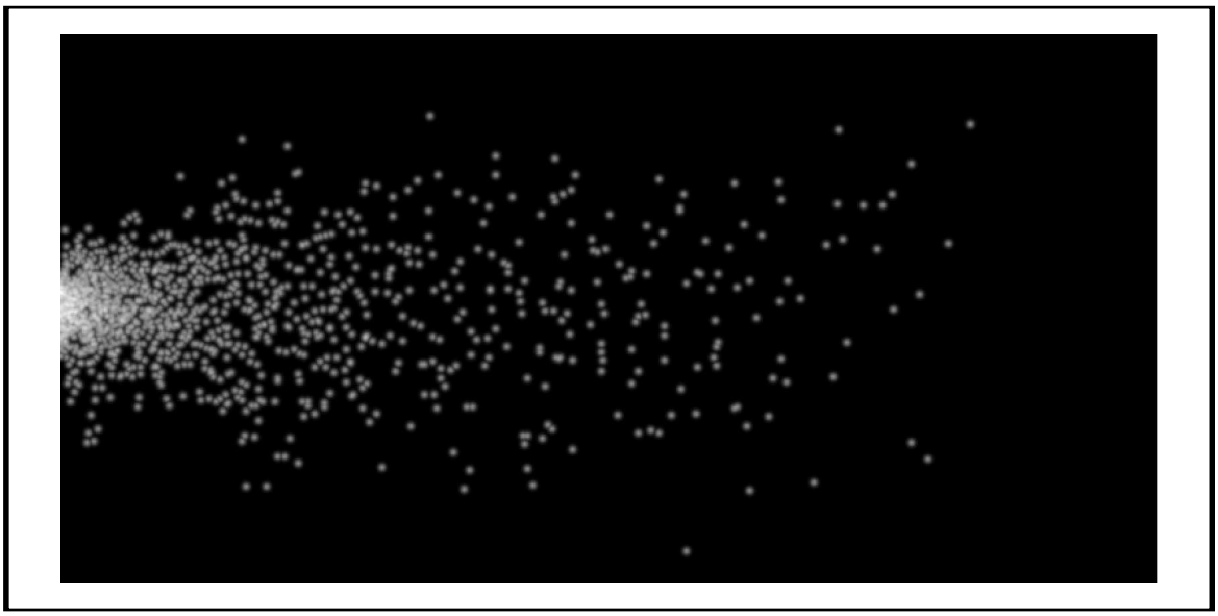


Figure 1: *A Bland-Altman plot of repeat measurement differences against the average value of the data-point. Final values of the fitting parameters were a=9.482800 ± 1.678087, and b=0.465157 ± 0.027671. The value of the index b is equal to that of a square-root function (to within 3 s.d.), representing a test for Poisson-like data.*

A log-Likelihood cost function [1] can be minimised, in order to fit reproducibility data for J pairs of samples $x[j]$ to an error model,

$$-logL \;= \sum_{j=1}^{J} \left[\frac{1}{2} \left(\frac{R_{12}\,[j]}{\epsilon_{R_{12[j]}}} \right)^2 + log\left(\sqrt{2\pi}\epsilon_{R_{12[j]}} \right) \right] \tag{12}$$

$R_{12[j]}/\epsilon_{R_{12}}$ represents a Pull distribution with unit standard deviation. Behaviour of the cost function around the optimal parameters is shown in Figure 2; the asymmetry is $\Delta_- \;=\; 0.423500$ and $\Delta_+ \;=\; 0.398967$ about the

minimum. This observed skewing of the cost function can be removed by fitting instead with the mathematically equivalent function,

$$\epsilon_x \;=\; a x^{\frac{1}{2c}}$$

For this model, the asymmetry is reduced to $\Delta_- \;=\; 0.013870$ and $\Delta_+ \;=\; 0.013760$, i.e. the asymmetry becomes negligible. This is therefore a better paramterisation of the model.

Figure 2: *Plot of the cost function. The skew of the quadratic function is typical of a cost function in heteroscedastic space. The minimum of the quadratic function represents the converged value of the cost function.*

2.2 Covariances for Linear Poisson Model weighting factors: Q

Linear Poisson models (LPM) describe histogram data as linear combinations of conditional probability functions, $P(X|k)$, and weighting factors $\mathbf{Q}_k$, thus,

$$\mathbf{H}_X \approx \mathbf{M}_X = \sum_k P(X|k)\mathbf{Q}_k \tag{13}$$

$\mathbf{H}$ represents a histogram; X a histogram bin; $\mathbf{M}$ the model of the histogram; and $\mathbf{Q}$ the quantity of component k in the model. LPM parameters are estimated using expectation maximisation (EM), which implicitly maximises the following extended maximum Likelihood (EML),

$$\ln \mathcal{L} = \sum_X \ln \left[\sum_k P(X|k)\mathbf{Q}_k \right] \mathbf{H}_X - \sum_k \mathbf{Q}_k \tag{14}$$

The statistical repeatability with which the $\mathbf{Q}$ terms are computed can be estimated using several methods. Three methods used herein include error propagation applied to the EM algorithm; using the MVB to analytically inspect the EML cost function shape; and numerically inspecting the EML cost function shape (see Section 2.1) [11,12].

As a test of the numerical covariance estimation process, Monte-Carlo data were generated for histograms comprised of two overlapping Gaussian distributions. An LPM analysis was then applied to build an ICA model of the generators and used to estimated their quantities.

For the first method (error propagation), the EM algorithm was applied to estimate LPM weighting parameters before error propagation was applied to the EM algorithm, in order to give statistical error estimates. For the second method (error propagation using the MVB), the EM algorithm was applied to estimate parameters, then the MVB was applied to the EML function, thus giving statistical error estimates. For the final method (numerical errors), the EML function ($-2 \ln \mathcal{L}$) was optimised directly using the Simplex algorithm, then the numerical method was applied to inspect the shape of the EML function. Table 1 summarises the results. Note the close agreement between all error estimation methods, typically to within 1%.

Error Prop σ	**MBV σ**	**Numerical σ**	**Error Prop σ**	**MBV σ**	**Numerical σ**
88.371	91.021	87.846	89.332	89.793	88.700
90.598	89.518	91.202	91.131	88.288	91.877
140.508	140.010	140.454	140.873	140.417	140.644
153.258	152.818	153.082	155.193	154.228	154.620
160.302	159.651	160.185	161.131	160.144	160.444
169.832	169.712	169.557	172.618	173.587	172.641
173.764	174.690	173.792	175.974	171.542	171.946
185.043	184.383	184.259	185.713	185.708	185.604
196.759	196.065	196.740	196.908	195.611	196.748
197.147	195.484	196.318	197.169	196.741	197.081
197.395	196.346	197.019	197.443	198.073	197.448
201.735	201.863	201.427	202.780	200.128	200.656
204.156	204.759	203.942	204.358	204.627	204.279
211.870	212.292	211.857	212.216	212.353	212.115
212.330	212.402	211.962	212.531	212.221	212.393
212.816	212.618	212.647	212.925	212.513	212.622
218.891	218.974	218.658	220.245	218.965	218.993
224.632	224.207	224.313	224.705	224.435	224.556
230.952	230.070	230.829	231.630	230.322	229.104
233.463	233.740	233.439	233.981	234.170	233.908
234.020	233.610	233.932	234.660	234.594	234.560
234.838	233.157	234.511	234.981	234.310	234.066
234.996	234.954	234.888	235.395	234.478	234.200
236.175	235.822	235.636	236.402	235.589	235.645
253.400	253.266	253.088	253.404	253.229	253.373
266.851	266.435	266.584	267.014	266.267	266.747
275.996	274.892	275.500	281.637	275.017	274.595

Table 1: *The table shows estimated parameter errors. From left to right, these are computed using error propagation (analytically), the MVB (analytically) and numerically, respectively. Note the close agreement between methods, typically to within less than 1% of each other.*

2.3 Estimating Parameter Covariances for Pharmokinetic Models

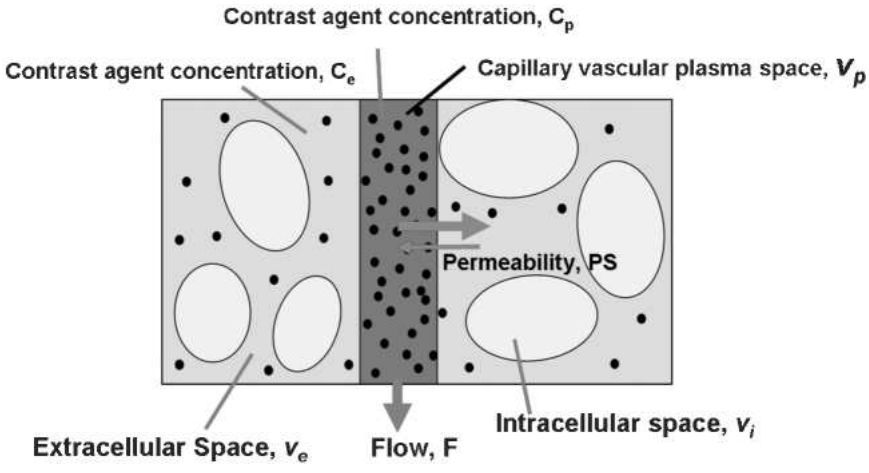


Figure 3: *Distribution of the contrast agent (black dots) within a voxel. The blood plasma, which flows across the arterial vessel, contains the greatest concentration of the contrast agent (C_p), whilst a fraction is transferred into the EES (C_e) according to the contrast agent permeability. The contrast agent does not enter the intracellular space v_i [15].*

Dynamic contrast-enhanced MRI (DCE-MRI) is increasingly being used in the study of tumour physiology and in patient follow-up, as it is easily performed in most MR scanners. The quantification of the physiological parameters

is performed using kinetic modelling. In tumours where the blood brain barrier is disrupted, the contrast agent leaks into the extravascular extracellular space (EES) (see Figure 3). This space is characterised by the fractional volume v_e. By solving the differential equation for a one compartment model, one obtains the Tofts model [13],

$$C_t(t) = K^{trans} \cdot C_p(t) \otimes e^{-\frac{K^{trans}}{v_e}} \tag{15}$$

The term C_t represents the concentration of the substance in tissue, and C_p the concentration in the plasma of the blood. However, the recorded signal from the scanner is due to both the amount of the contrast agent which has been diffused in the EES, and (more importantly), the component which remained in the blood vessels. As a result, an additional term describing the signal due to the blood volume is added to the original Tofts model,

$$C_t(t) = K^{trans} \cdot C_p(t \ + \ \tau) \otimes e^{-\frac{K^{trans}}{v_e}} + v_p \cdot C_p(t \ + \ \tau) \tag{16}$$

The term v_p represents the volume of blood plasma per unit volume of tissue. Equation (16) is known as the extended Tofts model [14]. The bolus delay (τ), i.e. the time from the beginning of the acquisition until the arrival of the contrast agent in the region of interest, is usually included as an additional parameter during the kinetic analysis by linear interpolation of a measured signal $C_p(t) \ = \ C(t)$ in an artery. Linear interpolation results in continuity of the cost function, but introduces differential discontinuities as the interpolation switches between data-points. Calculation of contrast concentration $C(t)$ is performed using a non-linear mapping, based upon a theory of MR image formation and requires absolute calibration of T_{10}. A weighted least-squared cost function (χ^2), can be constructed of the form,

$$\chi^2 \ = \ \sum_t \frac{(C(t) \ - \ C_t(t))^2}{var(C(t))} \tag{17}$$

The term $var(C(t))$ is estimated numerically, based upon noise propagated from the original image data though the non-linear calculation of concentration.

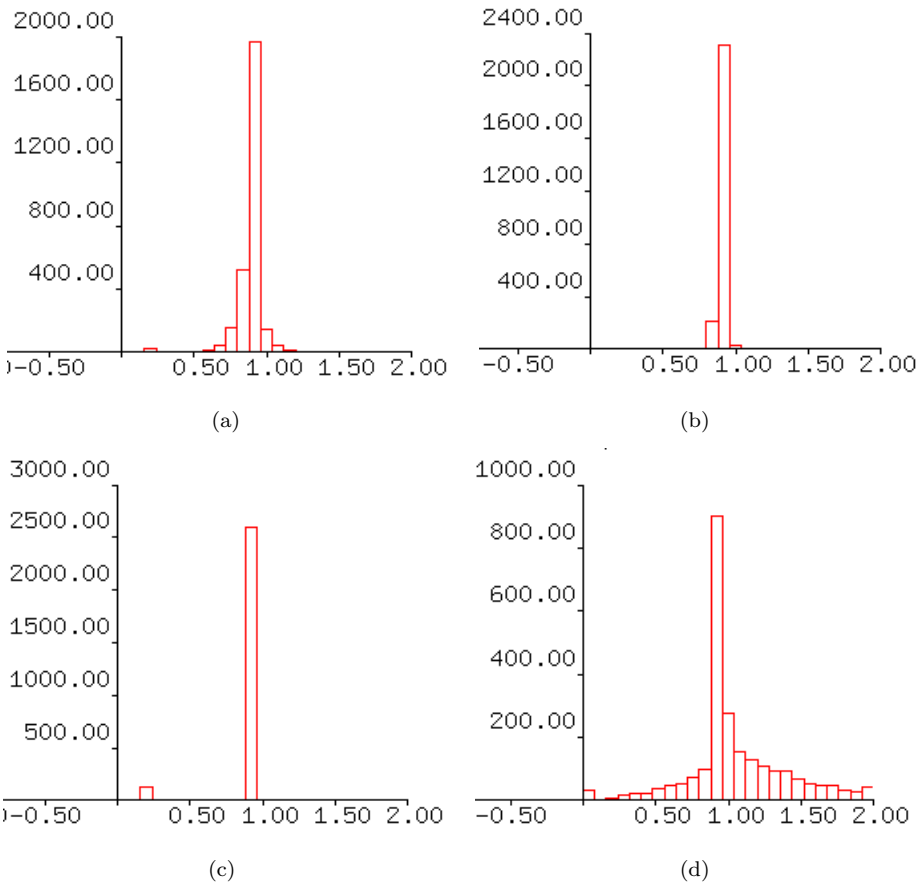


Figure 4: *The ratio of error estimated using the analytical approach to the error estimated with the numerical approach for* ***(a)*** K^{trans}, ***(b)*** v_e, ***(c)*** v_p, *and* ***(d)*** *bolus delay. Analysis was performed on a patient with grade IV glioblastoma.*

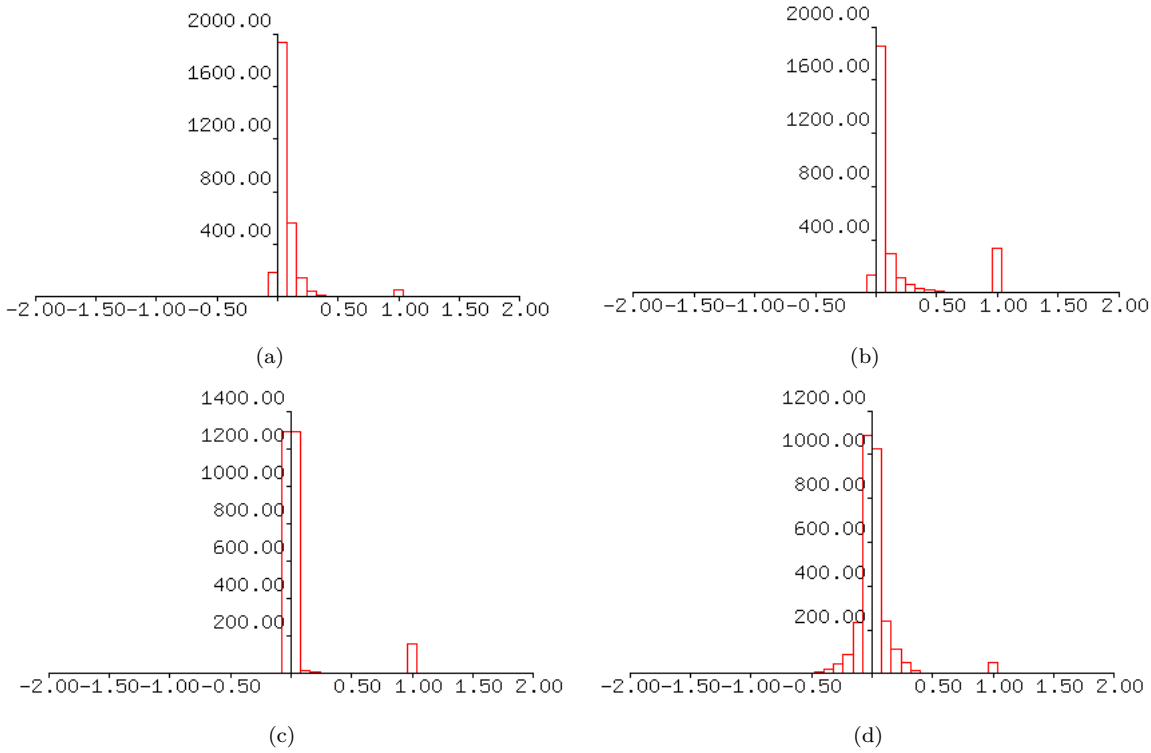


Figure 5: *The estimated asymmetries using the numerical method for* ***(a)*** K^{trans}, ***(b)*** v_e, ***(c)*** v_p, *and* ***(d)*** *bolus delay. Analysis was performed on a patient with grade IV glioblastoma.*

Correlations derived from the minimum variance bound, can be used as an estimate of the precision of physiological parameters when applying kinetic modelling. This can be performed either with;

i) the conventional analytical method using equation (16) to calculate the Jacobian matrices by assuming the cost function has quadratic behaviour;

ii) the numerical method which estimates the approximate quadratic-behaviour of the cost function (Eqn. (17)) using two symmetrical points from the curve.

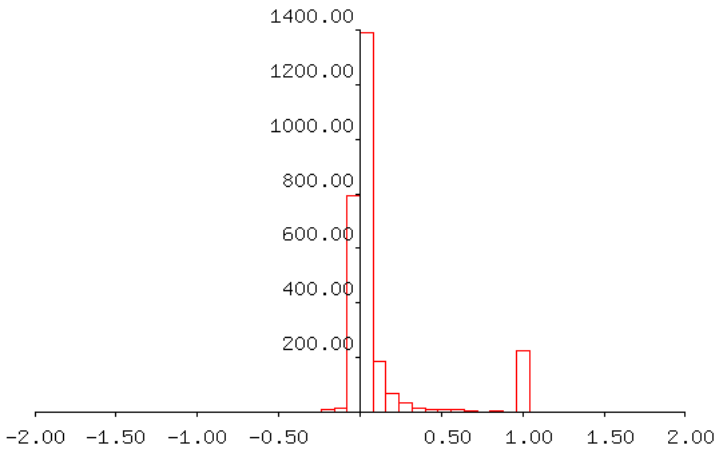
For the present work, a region of interest which encompassed one slice of the entire brain, resulting in 2500 individual fits of time course data, was selected. The error ratio and asymmetry across parameters are detailed in Figures 4 and 5, with the mean values listed in Table 2. As a demonstration of the effects of parameter mapping, the asymmetry for v_e^2 is shown in Figure 6. It is observed that the mean asymmetry of v_e is reduced from 0.172 to 0.11.

Parameter	**Error ratio**	**Asymmetry**
K_{trans}	0.88 ± 0.11	0.077 ± 0.145
v_e	0.91 ± 0.08	0.172 ± 0.317
v_p	0.87 ± 0.16	0.058 ± 0.232
τ	1.04 ± 0.33	0.017 ± 0.17

Table 2: *Comparison of the error ratio and asymmetry across parameters utilised for the analytical and numerical approaches.*

3 Discussion

The initial example shows how transformation of parameters can be used to remove asymmetries in a cost function (Figure 2) and so improve the statistical efficiency of covariance estimates; transformation from heteroscedastic to homoscedastic space.



(a)

Figure 6: *Asymmetry of the v_e parameter following the mapping to v_e^2.*

The LPM results demonstrate that for well-behaved cost functions, solutions obtained using the numerical approach closely match those obtained using analytical methods, to better than 1%. This level of agreement is similar to differences seen between analytic approaches and considered more than good enough for any subsequent use of the covariances.

In contrast, analytic approaches generate a much poorer approximations to the numerical approach for problems like the pharmokinetic model detailed above. Estimated parameter covariances may differ by up to a factor of two (Figure 4(d)). These problems are reflected in the asymmetry distributions (Figure 5), making it possible to assess the quality of these calculations even without knowing the analytic result. Figure 5(d) shows that the cost function is not quadratic as a function of bolus delay. When assessing the quality of a cost function using the asymmetry distribution a sensible limit would seem to be around ± 0.1. In this case the process of linear interpolation used to model bolus delay results in differential discontinuities. Under these circumstances the numerical estimate based on the local behaviour of the cost function should provide a better estimate of covariance than that based upon the second derivative computed at one point.

When the degree of observed asymetry is variable, parameter transformation can adjust the average behaviour but does not remove asymmetry in all cases (Figure 6). Parameter transformation is therefore not a quick fix for poorly behaved cost functions. Under these circumstances better numerical appraches should perhaps be considered for the calculation of the cost fucntion. For example, for pharmokinetic data, replacing the noisy estimate of $C_p(t)$ with a smoothly fitted curve would eliminate the need for linear interpolation and resulting differential discontinuities.

4 Conclusion

Developing the software for calculation of parameter covariances can be difficult and time-consuming for relatively straightforward functions, even for highly-experienced coders. Moreover, obtaining analytical solutions is not always possible for more complicated functions. In those instances where analytical solutions can be obtained, a numerical approach serves as a useful comparison, allowing the more-difficult to obtain analytical solution to be tested. Intuitively, better agreement would be expected for well-behaved cost functions. The distribution of asymetries for multiple fits allows the numerical behaviour of the cost function to be assessed. Poorly behaving cost functions might then be improved via better numerical implementation in order to gain more reliability and statistical efficiency.

For some cost functions, the second derivative at the local minimum may not give a good estimate of parameter error. Utilising numerical methods computed using a suitable distance over the change in the cost function, this issue can be avoided. Additionally, parameter definitions can cause problems with constructing a representative covariance. If parameters are poorly chosen, the cost function is likely to be skewed. Inspecting the asymmetry of estimated inverse covariances on either side of the optima allows us to assess this problem. Performing a parameter transformation can then provide a better representation of error.

5 Appendix A: Calculation of the Diagonal Inverse Covariance

The following C function was written to calculate a numerical estimate of the diagonal inverse covariance matrix elements.

```
double diagerror(int i, int n, double *a, double (*fitfunc), void *data)

/* calculate diagonal parameter covariance element */
{
        int j=0;
        double sig, sig2, f, olda, val1, val2;
        char temp[256];
        olda = a[i];

/* start with a small step in the parameter */
        sig = a[i]/10000.0;
        f = 2.0;
        val1 = fitfunc(n, a, data);

/* try to estimate error on parameter to around 1 percent */
while ((!(f<1.01) || !(f>0.99)) && j <= 10)
        {
             j++;
             a[i] = olda + sig;
             val2 = fitfunc(n, a, data);
             f = fabs(val1-val2);
             if (f < 100.0)
                  sig=sig/sqrt(f);
             else
                  sig = sig/2.0;
        }
        a[i] = olda;

/* now check with negative value to within 10 percent */

        a[i] = olda - sig;
        val2 = fitfunc(n, a, data);
        f = fabs(val1-val2);
        sig2=sig/sqrt(f);

        if (fabs((sig-sig2)/(sig+sig2)) >= 0.1)
        {
            sprintf(temp, "  Asymmetric covariance structure in diagerror()
            %f %f \n", sig, sig2);
            Print(temp);
        }
        else
        {
            sig = (sig+sig2)/2.0;
        }
        a[i] = olda;
        return (sig);
}
```

6 Appendix B: Calculation of Off-Diagonal Inverse Covariance

The following C function was written to calculate a numerical estimate of the off-diagonal inverse covariance matrix elements.

```
double offinvdiagerror (int i, int j, double delta_i2, double delta_j2,
        int n, double *a , double (*fitfunc), void *data)

/* calculate inverse off-diagonal parameter covariance */
{
        double z1,z;
        double mint;
        double alpha_ij, alpha_ij2;
        double oldi, oldj;
        char temp[256];

        oldi = a[i];
        oldj = a[j];

        mint = fitfunc(n, a, data);
        a[i] +=delta_i2;
        a[j] +=delta_j2;
        z1 = fitfunc(n, a, data);
        z=z1-mint;

        alpha_ij=2.0*(delta_i2*delta_j2);
        alpha_ij=(z-2.0)/alpha_ij;

/* test for symmetry here */

        a[i] = oldi - delta_i2;
        a[j] = oldj - delta_j2;
        z1 = fitfunc(n, a, data);
        z=z1-mint;

        alpha_ij2=2.0*(delta_i2*delta_j2);
        alpha_ij2=(z-2.0)/alpha_ij2;

        if (fabs((alpha_ij2-alpha_ij)/(alpha_ij+alpha_ij2)) >= 0.2)
        {
            sprintf(temp, "  Asymmetric covariance structure in offinvdiagerror()
            %f %f \n", alpha_ij, alpha_ij2);
            Print(temp);
        }
        alpha_ij = (alpha_ij+alpha_ij2)/2.0;

/* restore old values */
        a[i] = oldi;
        a[j] = oldj;

        return(alpha_ij);
}
```


Acknowledgements

Funding from Leverhulme is gratefully acknowledged.

References

[1] D. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM J. Appl. Math.,* 11 431 (1963).

[2] J. Reid, "On the Method of Conjugate Gradients for the Solution of Large Sparse Systems of Linear Equations," *In Large Sparse Sets of Linear Equations: Proceedings of the Oxford conference of the Institute of Mathematics and Its Applications,* held in April, 1970 (Ed. J. Reid). London: Academic Press, 231 (1971).

[3] J. Dongarra, I. Duff, D. Sorensen and H. van der Vorst, "Solving Linear Systems on Vector and Shared Memory Computers," Philadelphia, PA: SIAM, (1991).

[4] R. B. Fisher, A. W. Fitzgibbon, M. Waite, E. Trucco and M. Orr, "Recognition of complex 3-d objects from range data," *Proc. 7th International Conference on Image Analysis and Processing,* 569 (1993).

[5] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. Goldgof, K. Bowyer, D. Eggert, A. W. Fitzgibbon and R. B. Fisher, "An Experimental Comparison of Range Segmentation Algorithms," *IEEE Trans. Pat. Anal. and Mach. Intel.,* 18 673 (1996).

[6] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal,* 7 308 (1965).

[7] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika* 35 246 (1948).

[8] S. K. Bar-Lev and P. Enis, "On the classical choice of variance stabilizing transformations and an application for a Poisson variate," *Biometrika,* 75 803 (1988).

[9] H. Ragheb, N. A. Thacker, R. Pathak, D. M. Morris and A. Jackson, "Multi-site Liver Tumour ADC Reproducibility at 1.5 T," Tina Memo 2014-007 (2014).

[10] A. Seepujak, N. A. Thacker and P. D. Tar, "Background Subtraction and Noise Modelling for Spectroscopy Data," Tina Memo 2015-005 (2015).

[11] P. D. Tar and N. A. Thacker, "Linear Poisson Models: A Pattern Recognition Solution to the Histogram Composition Problem," Tina Memo 2013-006 (2013).

[12] W. H. Press et al., "Numerical Recipies in C," Cambridge University Press (1991).

[13] P. S. Tofts et al., "Estimating Kinetic Parameters From Dynamic Contrast-Enhanced T1-Weighted MRI of a Diffusible Tracer: Standardized Quantities and Symbols," *Journal of Magnetic Resonance Imaging,* 223 (1999).

[14] S. S. Kety and C. F. Schmidt, "The Determination of Cerebral Blood Flow in Man by the Use of Nitrous Oxide in Low Concentrations," *American Journal of Physiology,* 53 (1945).

[15] A. Jackson, "Imaging microvascular structure with contrast enhanced MRI," *The British Journal of Radiology,* S159 (2003).