August 19, 2000
CERN-TH/2000-249

Multivariate Fitting and the Error Matrix in Global Analysis of Data

J. Pumplin,[a] D. R. Stump,[a] and W. K. Tung[a,b]

[a] Department of Physics and Astronomy
Michigan State University
East Lansing, MI 48824 USA

[b] Theory Division
CERN
CH-1211 Geneva 23, Switzerland

When a large body of data from diverse experiments is analyzed using a theoretical model with many parameters, the standard error matrix method and the general tools for evaluating the error may become inadequate. We present an iterative method that significantly improves the reliability, and hence the applicability, of the error matrix calculation. Also, to obtain more accurate estimates of the uncertainties on predictions of physical observables, we present a Lagrange multiplier method that explores the entire parameter space and avoids the linear approximations assumed in conventional error propagation calculations. These methods are illustrated by an example involving the global analysis of parton distribution functions.

1 Introduction

The subject of this paper is a problem that arises when a large body of data from diverse experiments is analyzed according to a theoretical model that has many adjustable parameters. Consider a generic data fitting problem based on experimental measurements $\{D_I,\, I = 1, \ldots, N\}$ with errors $\{\sigma_I\}$. The data are to be compared to theoretical values $\{T_I\}$ obtained from a theoretical model with unknown parameters $\{a_i,\, i = 1, \ldots, n\}$. A common technique for comparing data with theory is to compute the χ^2 function defined by

$$\chi^2 = \sum_{I=1}^{N} \left(\frac{D_I - T_I}{\sigma_I} \right)^2 , \tag{1}$$

or generalizations of that formula in the more general case that correlations between the errors are known in terms of a set of correlation matrices. The physics objectives are (i) to find the best estimates of the parameters $\{a_i\}$ and their uncertainties, and (ii) to predict the values and uncertainties of other physical observables $\{X_\alpha,\, \alpha = 1, \ldots\}$ that are functions of the $\{a_i\}$.

If the errors are randomly distributed, and the correlations well determined, then standard statistical methods of χ^2 minimization [1, 2] apply, and established fitting tools like the program MINUIT [3] can be employed. However, real problems are often more complex. This is particularly so in a *global analysis* where the large number of data points $\{D_I\}$ do not come from a uniform set of measurements, but instead consist of a collection of results from many experiments on a variety of physical processes with diverse characteristics and errors. The difficulties are compounded if there are unquantified theoretical uncertainties, if the number of theoretical parameters n is large, or if the best parametrization cannot be uniquely defined *a priori*. All of these difficulties arise, for example, in the global analysis of hadronic parton distribution functions (PDFs) [4, 5, 6], which originally motivated this investigation. But the problem is clearly more general than that application.

Of the many possible issues that confront a global analysis of data, we shall address two, for which we have been able to significantly improve on the traditional treatment. The improvements allow a much more reliable determination of the uncertainties of $\{a_i\}$ and $\{X_\alpha\}$ in complex systems for which conventional methods may fail. To define these problems, we assume the system can be described by a global fitting function $\chi^2_{\rm global}$, or χ^2 for short, that characterizes the goodness-of-fit for a given set of theory parameters $\{a_i\}$. This χ^2 distills all available information on the theory and on the global experimental data sets, including their errors and correlations. One finds the minimum value χ^2_0 of χ^2, and the best estimates of the theory parameters are the values $\{a_i^0\}$ at that minimum. The dependence of $\chi^2(a_i)$ on $\{a_i\}$ near the minimum provides information on the uncertainties in the $\{a_i\}$. These are usually characterized by the error matrix and its inverse, the Hessian matrix H_{ij}, where one assumes that χ^2 can be well-approximated by a quadratic expansion in $\{a_i\}$ around $\{a_i^0\}$. Once the Hessian is known, one can estimate not only the uncertainties of $\{a_i\}$, but also the uncertainty in the theoretical prediction for any physical quantity X, provided the dependence of X on $\{a_i\}$ can be approximated by a linear expansion around $\{a_i^0\}$, and thus characterized by its gradient function at $\{a_i^0\}$ (cf. Sec. 2).

The first problem we address is a technical one that is important in practice: If the uncertainties on $\{a_i\}$ are very disparate for different directions in the n-dimensional parameter

space, *i.e.,* if the eigenvalues of H_{ij} span many orders of magnitude, how can one calculate the matrix H_{ij} with sufficient accuracy that reliable predictions are obtained in all directions? To solve this problem we have developed an iterative procedure that adapts the step sizes used in the numerical calculation of the Hessian to the magnitudes of the uncertainties in all directions of parameter space. We will demonstrate the effectiveness of this procedure in our specific application, where a standard tool fails to yield reliable results in some directions.

The second problem we address concerns the reliability of estimating ΔX, the uncertainty in the theoretical prediction for some physical variable X that is a function of the $\{a_i\}$. Can one obtain a robust estimate of ΔX that takes into account the variation of χ^2 over the entire parameter space $\{a_i\}$, without assuming the quadratic approximation to χ^2 and the linear approximation to X that are part of the error matrix approach? We solve this problem by using Lagrange's method of the undetermined multiplier to construct constrained fits that derive the dependence of χ^2 on X. Because this method is more robust, it can also be used to check the reliability of the Hessian method.

Section 2 summarizes the standard definitions of the error matrix and establishes our notation. Section 3 describes our iterative method for calculating the Hessian, and demonstrates the improvement over a standard method in a concrete example. Section 4 introduces the Lagrange multiplier method and compares the results from it against the Hessian approach in the same application. Section 5 concludes.

2 Error Matrix and Hessian

We first review the well-known connection between the error matrix and the Hessian matrix of second derivatives. We emphasize the eigenvector representations of those matrices, which we use extensively for subsequent developments.

The basic assumption of the error matrix approach is that χ^2 can be approximated by a quadratic expansion in the fit parameters $\{a_i\}$ near the global minimum. This assumption will be true if the variation of the theory values T_I with $\{a_i\}$ is approximately linear near the minimum. Defining $y_i = a_i - a_i^0$ as the displacement of parameter a_i from its value a_i^0 at the minimum, we have

$$\chi^2 = \chi_0^2 + \tfrac{1}{2}\sum_{i,j} H_{ij}\, y_i\, y_j\,, \tag{2}$$

$$H_{ij} = \left(\frac{\partial^2 \chi^2}{\partial y_i\, \partial y_j}\right), \tag{3}$$

where the derivatives are evaluated at the minimum point $y_i = 0$, and H_{ij} are the elements of the *Hessian matrix*. There are no linear terms in y_i in Eq. 2, because the first derivatives of χ^2 are zero at the minimum.

Being a symmetric matrix, H_{ij} has a complete set of n orthonormal eigenvectors $V_i^{(k)} \equiv v_{ik}$ with eigenvalues ϵ_k :

$$\sum_j H_{ij}\, v_{jk} = \epsilon_k\, v_{ik} \tag{4}$$

$$\sum_i v_{ij}\, v_{ik} = \delta_{jk}\,. \tag{5}$$

The eigenvectors provide a natural basis to express arbitrary variations around the minimum, so the $\{y_i\}$ can be replaced by a new set of parameters $\{z_i\}$, defined by

$$y_i = \sum_j v_{ij} \sqrt{\tfrac{2}{\epsilon_j}}\, z_j \,. \tag{6}$$

These parameters have the simple property that

$$\Delta\chi^2 = \chi^2 - \chi_0^2 \,=\, \sum_i z_i^2 \,. \tag{7}$$

In other words, *the surfaces of constant χ^2 are spheres in $\{z_i\}$ space,* with $\Delta\chi^2$ *the squared length of the distance from the minimum in z-space.*

The orthonormality of v_{ij} can be used to invert the transformation (6),

$$z_i = \sqrt{\frac{\epsilon_i}{2}} \sum_j y_j\, v_{ji} \,. \tag{8}$$

The Hessian and its inverse are easily expressed in terms of the eigenvalues and eigenvector components:

$$H_{ij} \;=\; \sum_k \epsilon_k\, v_{ik}\, v_{jk} \tag{9}$$

$$(H^{-1})_{ij} \;=\; \sum_k \frac{1}{\epsilon_k}\, v_{ik}\, v_{jk} \,. \tag{10}$$

Now consider any *physical quantity* X that can be calculated according to the theory as a function of the parameters $\{a_i\}$. The best estimate of X is the value at the minimum $X_0 = X(a_i^0)$. In the neighborhood of the minimum, assuming the first term of the Taylor-series expansion of X gives an adequate approximation, the deviation of X from its best estimate is given by

$$\Delta X = X - X_0 \,\cong\, \sum_i \frac{\partial X}{\partial y_i}\, y_i = \sum_i X_i\, z_i \tag{11}$$

where $X_i \equiv \frac{\partial X}{\partial z_i}$ are the components of the z-gradient evaluated at the global minimum, *i.e.,* at the origin in z-space.

Since χ^2 increases uniformly in all directions in z-space, the gradient vector X_i gives the direction in which the physical observable X varies fastest with increasing χ^2. The maximum deviation in X for a given increase in χ^2 is therefore obtained by the dot product of the gradient vector X_i and a displacement vector Z_i in the same direction with length $\sqrt{\Delta\chi^2}$, *i.e.,* $Z_i = X_i\sqrt{\Delta\chi^2/\Sigma_j X_j^2}$. For the square of the deviation, we therefore obtain

$$(\Delta X)^2 = (X\cdot Z)^2 = \Delta\chi^2 \sum_i X_i^2 \,. \tag{12}$$

The traditional formula for the error estimate $(\Delta X)^2$ in terms of the original coordinates $\{y_i\}$ can be derived by substituting $X_i = \frac{\partial X}{\partial z_i} = \sum_j \frac{\partial X}{\partial y_j}\,\frac{\partial y_j}{\partial z_i}$ in (12) and using (6) and (10).

The result is

$$(\Delta X)^2 = 2\,\Delta\chi^2 \sum_{i,j} \frac{\partial X}{\partial y_i}\,(H^{-1})_{ij}\,\frac{\partial X}{\partial y_j}\ . \tag{13}$$

This standard result can of course also be derived directly by minimizing Eq. (2) with respect to $\{a_i\}$, subject to a constraint on X.

Equations (12) and (13) are equivalent in principle, provided the assumptions of linear approximation for X and quadratic approximation for χ^2 are valid. In practical applications, the numerical accuracy of the two calculations can differ considerably if these conditions are not well met over the relevant region of parameter space. To calculate the error estimate ΔX, we prefer to use Eq. (12), using derivatives X_i calculated by finite differences of X at the points $z_i = \pm\sqrt{\Delta\chi^2}$ (with $z_j = 0$ for $j \neq i$). This is generally more accurate, because it estimates the necessary derivatives over the relevant region of parameter space, thereby reducing the effect of higher order terms and numerical noise.

The value of $\Delta\chi^2$ that corresponds to a given confidence level is well defined for ideal experiments. But in a real-world global analysis, the uncertainties σ_I in Eq. (1) are not accurately known because they include subjective estimates of experimental and theoretical systematic errors. The relation between $\Delta\chi^2$ and confidence level therefore requires further analysis.[1]

In a complex problem such as a global analysis, the region of applicability of the approximations is generally unknown beforehand. A situation of particular concern is when the various eigenvalues $\{\epsilon_i\}$ have very different orders of magnitude—signaling that the function χ^2 varies slowly in some directions of a_i space, and rapidly in others. The iterative method described in the next section attempts to accommodate this situation.

3 Iterative Procedure

In practical applications, the Hessian matrix H_{ij} is calculated using finite differences to estimate the second derivatives in (3). A balance must be maintained in choosing the step sizes for this, since higher-order terms will contribute if the intervals are too large, while numerical noise will dominate if the intervals are too small. This noise problem may arise more often than is generally realized, since the theory values $\{T_I\}$ that enter the χ^2 calculation are not always the ideally smooth functions of the fit parameters that one would associate with analytic formulas. For in complex theoretical models, the $\{T_I\}$ are often computed from multiple integrals, and will have small discontinuities as functions of $\{a_i\}$ induced by adaptive integration methods. These numerical errors forbid the use of a very small step size in the finite difference calculations of derivatives. Furthermore, as noted above, the eigenvalues of H_{ij} may span a wide range, so good accuracy is needed especially to get the smaller ones right.

[1] In the extreme case where fully correlated systematic errors are completely left out, it can be shown that the relevant $\Delta\chi^2$ could be as large as order N, the number of data points.

The Procedure

We want to evaluate H_{ij} by sampling the values of χ^2 in a region of parameter space where Eq. (2) is a good approximation. In principle, the parameters $\{z_i\}$ are the natural choice for exploring this space; but, of course, they are not known in advance. We therefore adopt the following iterative procedure:

1. Define a new set of coordinates $\{\xi_i\}$ by

$$y_i = \sum_j u_{ij}\, t_j\, \xi_j \tag{14}$$

where u_{ij} is an orthogonal matrix and $\{t_i\}$ are scale factors. In the first iteration, these are chosen as $u_{ij} = \delta_{ij}$ and $t_i = 1$, so that $\xi_i = y_i$. This makes the first round of iteration similar to the usual procedure of taking derivatives with respect to a_i. As the iteration progresses, u_{ij}, t_i, and ξ_i converge to v_{ij}, $\sqrt{2/\epsilon_i}$, and z_i.

2. Calculate the effective second derivative matrix Φ_{ij} defined by

$$\chi^2 = \chi_0^2 + \tfrac{1}{2}\sum_{ij} \Phi_{ij}\, \xi_i\, \xi_j \tag{15}$$

$$\Phi_{ij} = \frac{\partial^2 \chi^2}{\partial \xi_i\, \partial \xi_j} \tag{16}$$

using finite differences of the ξ_i. The step size in ξ_i is chosen to make the increase in χ^2, due to the diagonal element $\frac{1}{2}\frac{\partial^2\chi^2}{\partial\xi_i^2}{\xi_i}^2$, equal to a certain value $\delta\chi^2$. The choice of $\delta\chi^2$ is determined by the particular physics application at hand. Naively, one might expect $\delta\chi^2 \simeq 1$ to be the "right choice". That would indeed be appropriate for a χ^2 function obeying ideal statistical requirements. But when the input to $\chi^2_{\rm global}$ is imperfect, a reasonable choice of $\delta\chi^2$ must be based on a physics judgement of the appropriate range of that particular χ^2 function. We therefore leave the choice of $\delta\chi^2$ open in this general discussion.[2] In any case, if the final results are to be trustworthy, they must not be very sensitive to that choice.

We calculate each off-diagonal second derivative by evaluating χ^2 at the four corners of a rectangle $(+\delta_i, +\delta_j)$, $(-\delta_i, -\delta_j)$, $(+\delta_i, -\delta_j)$, $(-\delta_i, +\delta_j)$. This is a modification of the technique used in MINUIT [3]. For the sake of efficiency, the MINUIT subroutine HESSE estimates off-diagonal elements using only one of those corners, together with values at $(\delta_i, 0)$ and $(0, \delta_j)$ that are already known from calculating the diagonal elements of the Hessian. Our method is slower by a factor of 4, but is more accurate because it fully or partly cancels some of the contributions from higher derivatives. The first derivatives $\partial\chi^2/\partial\xi_i$ are also calculated at this stage of the iteration and are used to refine the estimate of the location of the minimum.

3. Compute the Hessian according to Φ_{ij},

$$H_{ij} = \sum_{m,n} \frac{\Phi_{mn}\, u_{im}\, u_{jn}}{t_m\, t_n}\,. \tag{17}$$

[2] Cf. discussion in the following subsection on a sample problem.

4. Find the normalized eigenvectors of the Hessian, as defined by Eqs. (4) and (5).

5. Replace u_{ij} by v_{ij}, t_j by $\sqrt{2/\epsilon_j}$, and go back to step 1. The steps are repeated typically 10–20 times, until the changes become small and Φ_{ij} converges to $2\,\delta_{ij}$.

This iterative procedure improves the estimate of the Hessian matrix, and hence of the error matrix, because in the later iterations, it calculates the Hessian based on points that sample the region where $\Delta\chi^2$ has the magnitude of physical interest.

Results from a Sample Application

As an example, we apply the above procedure to the application that motivated this study—the global analysis of PDFs [7]—and compare the results with those obtained from MINUIT. The experimental input for this problem consists of $N \approx 1300$ data points from 15 different experimental data sets involving four distinct physical processes. All the potential complexities mentioned earlier are present in this system. The theory is the quark parton model, based on next-to-leading order perturbative Quantum Chromodynamics (QCD). The model contains $n = 16$ physical parameters a_i that characterize the quark and gluon distributions in the proton at some low momentum scale Q_0. From a calculational point of view, the theoretical model consists of the numerical integration of an integro-differential equation and multiple convolution integrals that are evaluated mostly by adaptive algorithms. The fitting function $\chi^2_{\rm global}$ in this case combines the published statistical and systematic errors of the data points in quadrature. The only correlated errors incorporated are the published overall normalization uncertainties of the individual experiments. The fitting program is the same as that used to generate the CTEQ parton distributions [4, 6]. The global χ^2 minimum for this system is given by the CTEQ5M1 set of PDFs, for which $\chi^2_0 \approx 1200$. For details, see Ref. [6].

We find that the eigenvalues $\{\epsilon_i\}$ of the Hessian for this system range over 5 orders of magnitude. Because $\chi^2_{\rm global}$ does not contain full information on correlated systematic errors, it represents an "imperfect" χ^2 function as discussed in the previous section. From an independent detailed study of the uncertainties associated with this system [8], we estimate that an appropriate choice of $\delta\chi^2$ for the iterative calculation is around 10. (It is natural that this value is larger than 1 because adding the statistical and systematic errors in quadrature inflates the denominator in Eq. (1), also cf. footnote 1.)

The error matrix approach relies on a quadratic approximation to χ^2 in the neighborhood of the minimum. To see how good that approximation is, we plot χ^2 as a function of distance along a particular direction in $\{a_i\}$ space, as shown in Fig. 1. The direction chosen is a typical one—specifically it is the direction of the eigenvector with median eigenvalue. The dotted curve in Fig. 1 is the exact χ^2 and the solid curve is the quadratic approximation (2). The approximation is seen to provide a rather good description of the function. Even at points where χ^2 has increased by 50, the quadratic approximation reproduces the increase to an accuracy of 20%.

To correctly measure the curvature of the quadratic approximation, it is important to fit points that are displaced by an appropriate distance. This can be seen from Fig. 2, which shows the difference between the two curves in Fig. 1 in the central region. The difference displays a small cubic contribution to χ^2. It also reveals contributions that vary erratically

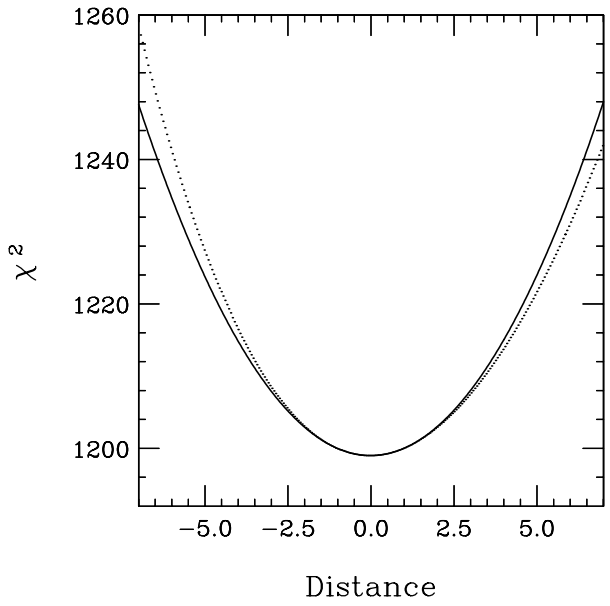

Figure 1: Variation of χ^2 with distance along a typical direction in parameter space. The dotted curve is the exact χ^2 and the solid curve is the quadratic approximation based on the Hessian. The quadratic form is seen to be a rather good approximation over the range shown.

with a magnitude on the order of 0.03 . These fluctuations come from the noise associated with switching of intervals in the adaptive integration routines. Because the fluctuations are small, they do not affect our results in principle. But they do require care in estimating the derivatives. In particular, they would make finite-difference estimates based on small intervals extremely unreliable. The iterative method avoids this problem by choosing a suitable scale for each eigenvector direction when evaluating the Hessian.

Figures 1 and 2 show the behavior of χ^2 along a single, albeit typical, direction in the 16 dimensional parameter space. Fig. 3 shows a complementary test of the iterative method for all possible directions. We have chosen 1000 directions at random in $\{z_i\}$ space. We displace the parameters away from the minimum in each of these directions by a distance that makes $\Delta\chi^2 = 5$. We then compute the value of $\Delta\chi^2$ predicted by the quadratic approximation (2), using the Hessian calculated by the iterative method and, for comparison, by the routine HESSE within the MINUIT package. The results are displayed in Fig. 3 as histograms, with $\Delta\chi^2$ on the horizontal axis and number of counts on the vertical axis. If χ^2 were quadratic in $\{a_i\}$, then a perfect computational method would yield a delta function at $\Delta\chi^2 = 5$. From Fig. 3 we see that:

- For the *solid histogram*—the result of the iterative procedure—the quadratic approximation is close to the exact result in all directions, and hence Eq. (2) is a pretty good representation of χ^2. Quantitatively, the middle 68% of the distribution is contained in the region 5.4 ± 0.6 .
- For the *dotted histogram*—based on the general purpose program MINUIT—the distribution is also spread around the expected value of 5, but it is very broadly distributed. This estimate of the Hessian is therefore unsatisfactory, because we might be interested in a

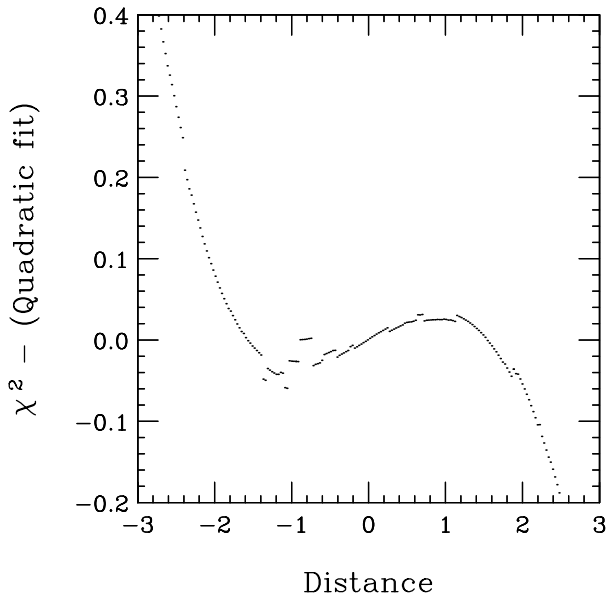

Figure 2: Difference between χ^2 and its quadratic approximation (2) as shown in Fig. 1. A cubic contribution can be seen, along with a noticeable amount of numerical noise.

quantity whose gradient direction is one for which the Hessian computed by MINUIT is widely off the mark. A major source of this problem is the numerical noise visible in Fig. 2: MINUIT uses a small step size to calculate the derivatives, and gets misled by the small-scale discontinuities in χ^2. For some directions, $\Delta\chi^2$ even becomes negative because the errors in one or more of the small eigenvalues are big enough to allow their calculated values to become negative. (Within MINUIT, this circumstance is handled by issuing a warning message and adding a constant to all the eigenvalues, which in the context of Fig. 3 corresponds to shifting the dotted distribution to the right.)

Figure 4 shows the results of a similar study, in which the 1000 random directions are chosen only from the subspace of $\{z_i\}$ that is spanned by the 10 directions with the largest eigenvalues ϵ_i. These larger eigenvalues correspond to directions in which χ^2 rises most rapidly, or in other words, directions in which the parameters are most strongly constrained by data. Because the distance moved away from the minimum in $\{a_i\}$ space is smaller in this case, the quadratic approximation is generally better, so it is not surprising that the histograms are more sharply peaked than in Fig. 3. But the advantage of the iterative method remains apparent.

Comment

Information from the iteratively-improved Hessian provides a useful tool for refining the choice of functional forms used to parametrize a continuous degree of freedom in the theoretical model.

The length squared of the displacement vector in the space of fit parameters is

$$\sum_i (a_i - a_i^0)^2 = \sum_i y_i^2 = \sum_i \frac{2}{\epsilon_i} z_i^2 \tag{18}$$

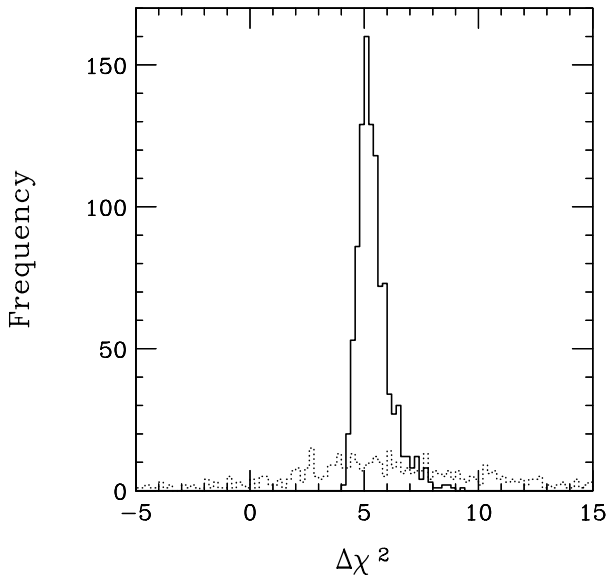

Figure 3: Frequency distribution of $\Delta\chi^2$ according to the Hessian approximation (2) for displacements in random directions for which the true value is $\Delta\chi^2 = 5.0$. *Solid histogram*: Hessian calculated by iterative method of Section 3; *Dotted histogram*: Hessian calculated by MINUIT.

while $\Delta\chi^2 = \sum_i z_i^2$ by Eq. (7). Hence the directions in which the parameters are well determined (the steep directions) correspond to eigenvectors of the Hessian with large eigenvalues, while the shallow directions in which they are weakly determined correspond to small eigenvalues.

The extreme values of any particular a_i are

$$a_i = a_i^0 \pm \Delta a_i \tag{19}$$

where

$$(\Delta a_i)^2 = 2\,\Delta\chi^2 \sum_j \frac{v_{ij}^2}{\epsilon_j}\,. \tag{20}$$

Equation (20) can be used to see if each parameter is appropriately well constrained. Furthermore, the individual terms in the sum show the contributions to Δa_i from the various eigenvectors, so if a parametrization leads to a poorly defined minimum because it allows too much freedom—which is indicated by a failure of the iteration to converge for the smallest eigenvalues of the Hessian—it is easy to see which of the parameters are most responsible for the overly-shallow directions.

4 Lagrange Multiplier Method

The Hessian, *via* its inverse which is the error matrix, provides a general way to propagate the uncertainties of experimental and theoretical input to the fit parameters $\{a_i\}$, and then

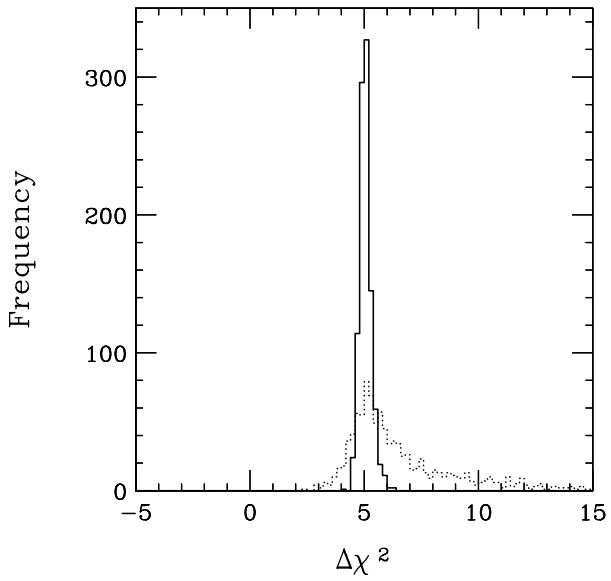

Figure 4: Same as Fig. 3, except that the displacements are restricted to the parameter subspace spanned by the 10 "steepest directions".

on to measurable physical quantities $\{X_\alpha(a_i)\}$ by Eqs. (12) or (13). But it requires assuming that χ^2 and X can be treated as quadratic and linear functions of $\{a_i\}$ respectively. In this section we describe a different approach, based on the mathematical method of the *Lagrange undetermined multiplier*, which avoids those assumptions.

The Procedure

Let X_0 be the value of X at the χ^2 minimum, which is the best estimate of X. For a fixed value of λ, called the Lagrange multiplier, one performs a new minimization with respect to the fit parameters $\{a_i\}$, this time on the quantity

$$M = \chi^2 + \lambda(X - X_0)\,, \tag{21}$$

to obtain a pair of values $(\chi^2(\lambda), X(\lambda))$. (The constant term $-\lambda X_0$ here is not necessary, because it does not affect the minimization; but it makes the minimum value of M easier to interpret.) At this new minimum, $\chi^2(\lambda)$ is the lowest possible χ^2 for the corresponding value $X(\lambda)$ of the physical variable X. Thus one achieves a *constrained fit* in which χ^2 is minimized for a particular value of X.

By repeating the minimization for many values of λ, one maps out the parametrically-defined curve $(\chi^2(\lambda), X(\lambda))$. Since λ is just the parameter for this curve, its value is of no particular physical significance. The relevant range for λ can be found by trial-and-error; or it can be estimated using the Hessian approximation, which predicts that $\lambda \approx -2\,\Delta\chi^2/\Delta X$. According to that approximation, M goes down by the same amount that χ^2 goes up.

One way to understand the Lagrange Multiplier method is first to imagine that the quantity X is simply one of the fitting parameters, say a_1. The variation of χ^2 with a_1 could be mapped out by minimizing χ^2 with respect to $\{a_2, \ldots, a_n\}$ for various fixed values of

a_1. (That operation is indeed so useful that MINUIT already provides a procedure MINOS to carry it out.) In the more general case that X is a function of all $\{a_i\}$, one wants to similarly minimize χ^2 for fixed values of X. That is exactly what the Lagrange Multiplier method does, since including the undetermined multiplier term in (21) renders the $\{a_i\}$ independent in spite of the constraint on X.

A more phenomenological way to understand the Lagrange Multiplier method is to imagine that X has just been measured, with result $X_{\rm new} \pm \sigma_{\rm new}$. To decide whether this hypothetical new measurement is consistent with the old body of data, one would add a term $((X_{\rm new} - X)/\sigma_{\rm new})^2$ to $\chi^2_{\rm global}$ of Eq. (1) and redo the minimization. The added contribution to χ^2 consists of a constant, a linear term in X, and a quadratic term in X. This is equivalent to Eq. (21), because a constraint on X^2 is equivalent to a constraint on X itself.

The essential feature of the Lagrange Multiplier method is that, for a given $\Delta\chi^2$, it finds the largest range of X allowed by the global data sets and the theoretical model, independent of any approximations. The full parameter space $\{a_i\}$ is explored by the minimization procedure, not just the immediate neighborhood of the original χ^2 minimum as in the Hessian method, and no approximations based on a small deviation from the original minimum are made.

The only drawback to the Lagrange Multiplier method is that it can be slow computationally, since it requires a separate series of minimizations for each observable X that is of interest.

Example

We now look at an example of the Lagrange Multiplier method from our application—the uncertainty of parton distribution functions. For the physical quantity X we consider the cross section σ_W for $W^\pm$ production in $p\overline{p}$ collisions at the energy $\sqrt{s} = 1.8\,\text{TeV}$ of the Tevatron collider at Fermilab. We would like to estimate σ_W, and the uncertainty on that estimate, based on the global analysis of parton distributions.

The points in Fig. 5 show $\chi^2_{\rm global}$ as a function of $\sigma_W B$ in nanobarns, where $B = 0.106$ is the branching ratio for $W \to e\,\nu$. These points are obtained by the Lagrange Multiplier method using $\lambda = 0$, ± 1000, ± 2000, ± 3000, ± 4000. They are thus discrete cases of $\chi^2_{\rm global}$ versus σ_W without approximations.

The smooth curve in Fig. 5 is the parabola given in Eq. (13), using the Hessian computed by the iterative method and treating σ_W in the linear approximation. The comparison between this curve and the discrete points from the Lagrange Multiplier calculation tests the quality of the quadratic and linear approximations and the reliability of the iterative calculation of H_{ij}. For this application, we conclude that the improved Hessian method works very well, since the difference between the points and the curve is very small, and indicates only a small cubic contribution. If the two results did not agree, the correct result would be the one given by the Lagrange Multiplier method.

To estimate ΔX, the uncertainty of X consistent with the global analysis of existing data, one needs to specify what range of $\Delta\chi^2_{\rm global}$ is allowed. As discussed earlier, the acceptable limit of $\chi^2_{\rm global}$ depends on the nature of the original definition of this fitting function, in the context of the specific system under study. For the case of σ_W, Ref. [8] estimates $\Delta\chi^2_{\rm global} \approx 100$, which translates into $\Delta\sigma_W/\sigma_W \approx \pm 3\%$ according to Fig. 5.

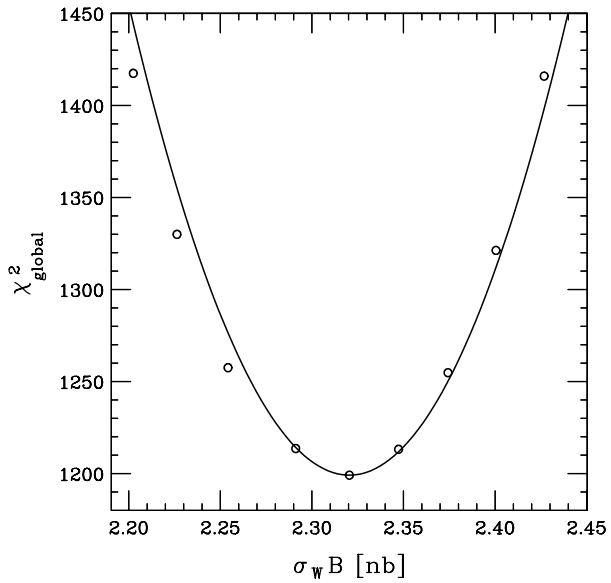

Figure 5: Minimum χ^2 as a function of the predicted cross section for $W^{\pm}$ production in $p\overline{p}$ collisions. *Parabolic curve* is the prediction of the iteratively improved Hessian method. *Points* are from the Lagrange multiplier method.

5 Conclusion

We have addressed some computational problems that arise in a global phenomenological analysis, in which a complex theory with many parameters confronts a large number of data points from diverse experiments.

The traditional error-matrix analysis is based on a quadratic approximation to the function χ^2 that measures the quality of the fit, in the neighborhood of the minimum that defines the best fit. The iterative method proposed in Sec. 3 improves the calculation of the Hessian matrix which expresses that quadratic approximation, for a complex system in which standard general-purpose programs may fall short.[3] The inverse of this improved version of the Hessian matrix is an improved version of the error matrix, which can be used to estimate the uncertainty of predictions using standard error matrix formulas.

The Lagrange Multiplier method proposed in Sec. 4 calculates the uncertainty on a given physical observable *directly*, without going through the error matrix. It thus avoids the assumption that the theoretical quantities can be approximated by linear functions of the search parameters, which is intrinsic to the Hessian approach.

For simplicity, we have discussed only the problem of obtaining error estimates on a single quantity X. It is straightforward to generalize our methods to find the region allowed simultaneously for two or more variables by a given $\Delta\chi^2$. For example, in the case of two variables $X^{(1)}$ and $X^{(2)}$, the allowed region according to the Hessian method is the interior of an ellipse. The Lagrange multiplier method is generalized for this case by adding *two* Lagrange multiplier terms, $\lambda_1 X^{(1)} + \lambda_2 X^{(2)}$, to χ^2.

[3]Our iterative procedure for calculating the Hessian is implemented as an extension to the widely used CERN fortran library routine MINUIT [3]. The code can be requested by e-mail from pumplin@pa.msu.edu.

Although the Lagrange Multiplier procedure is conceptually simple and straightforward to apply, it is slow computationally because it requires many full minimizations to map out χ^2 as a function of X, and this must be done separately for each quantity X whose error limits are of interest. In contrast, once the Hessian has been determined from the global analysis, it can be applied to any physical observable. One needs only to compute the gradient $\partial X/\partial a_i$ of the observable and substitute into Eq. (13); or better, to compute the gradient $X_i = \partial X/\partial z_i$ and substitute into Eq. (12). For computational efficiency, the iteratively calculated Hessian may therefore be the method of choice, provided its linear approximations are sufficiently accurate. Whether or not that is the case can be determined by comparing Hessian and Lagrange Multiplier results. We have used both methods in a detailed study of the uncertainties in the CTEQ5 parton distribution functions [7, 8].

Acknowledgments

The authors wish to thank R. Brock, D. Casey, J. Huston, and F. Olness for useful discussions on the uncertainties of parton distribution functions. This work was supported in part by NSF grant PHY-9802564.

References

[1] Among standard textbooks, see for instance, R. J. Barlow, *Statistics*, (John Wiley, Chichester, 1989).

[2] D.E. Soper and J.C. Collins, "Issues in the Determination of Parton Distribution Functions", CTEQ Note 94/01; hep-ph/9411214.

[3] F. James, M. Roos, Comput. Phys. Commun. **10**, 343 (1975); Minuit manual, http://wwwinfo.cern.ch/asdoc/minuit/.

[4] James Botts, et al., Phys. Lett. **B304**, 159 (1993); H. L. Lai, et al., Phys. Rev. **D51**, 4763 (1995); H. L. Lai, et al., Phys. Rev. **D55**, 1280 (1997).

[5] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Eur. Phys. J. **C4**, (1998) 463, hep-ph/9803445; and references cited therein.

[6] H. L. Lai, J. Huston, S. Kuhlmann, J. Morfin, F. Olness, J. F. Owens, J. Pumplin and W. K. Tung, Eur. Phys. J. **C12**, 375 (2000), hep-ph/9903282.

[7] R. Brock, D. Casey, J. Huston, J. Kalk, J. Pumplin, D. Stump, W.K. Tung, Contribution to the Proceedings of Fermilab RunII Workshop on *QCD and Gauge Boson Physics* (to be published), MSU-HEP-03101, June 2000, hep-ph/0006148.

[8] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, W.-K. Tung, "Study of Uncertainties of Parton Distribution Functions"; and "Uncertainties of Physical Observables due to Parton Distribution Functions", *manuscripts in preparation.*