Paul Avery
CBX 91 – 73
October 18, 1991

Applied Fitting Theory II
Determining Systematic Effects by Fitting

1. Introduction

This is the second paper in my thrilling series on fitting theory. In this note I will concentrate on determining systematic contributions to measurement error. These systematic effects can degrade the resolution of the parameters beyond what is reported in the statistical error. Examples of effects that are commonly found in track fitting are misalignment of detector positions and orientations, nonuniformity of the magnetic field, noncircularity of a drift chamber layer, unaccounted bowing in the endplates (which affects stereo angles), extra wire sag, twisting of the endplates relative to each other, and so forth.

I present here a method which allows one to determine these systematic effects provided that one can parametrize the effect of any given parameter on the measurement. The heart of the technique lies in performing many thousands of fits (constrained or unconstrained), each of which depends on parameters local to that fit together with the unknown parameters describing the systematic effect. One naively would expect this procedure to be hopeless because the χ^2 minimization would lead to thousands of equations to be solved simultaneously. It turns out, however, that one can remove the parameters describing the many distinct fits and only determine the small number of remaining systematic unknowns. Better still, one can accumulate statistics steadily, without any intermediate matrix inversion, in order to reduce the errors in these parameters.

Note: I make no claim to be the only person ever to have invented this algorithm. My treatment is very general, however, since it permits the calculation of the full covariance matrix of the systematic parameters. I also include the interesting case where the individual fits were performed with constraints.

2. Finding Systematic Effects Using Unconstrained Fits

In the first paper in the series (CBX 91−72), I gave a general overview of fitting theory using matrix notation. I showed that if the dependence of the measurements on the parameters can be linearized, then the solution which is best from the point of view of minimum variance, can be obtained rather easily, even in the presence of constraints, by minimizing the χ^2 function,

To recapitulate the argument, suppose we want to fit a set of n measurements $\mathbf{y}$ to a set of m parameters α through the relation $y_l = f_l(\alpha)$ for $1 \le l \le n$. If the $f_l(\alpha)$ are nonlinear we can expand them about an approximate solution $\alpha = \alpha_A$: $y_l = f_l(\alpha_A) + (\partial f_l/\partial\alpha_i)(\alpha_i - \alpha_{A\,i})$. This linearization permits us to define the χ^2 statistic as

$$\begin{aligned}\chi^2 &= (\mathbf{y} - \mathbf{f}(\alpha_A) - \mathbf{A}(\alpha - \alpha_A))^t \mathbf{V}_y^{-1} \ (\mathbf{y} - \mathbf{f}(\alpha_A) - \mathbf{A}(\alpha_A - \alpha)) \\ &\equiv (\mathbf{\Delta y} - \mathbf{A}\eta)^t \mathbf{V}_y^{-1} \ (\mathbf{\Delta y} - \mathbf{A}\eta)\end{aligned}$$

where $\mathbf{\Delta y} = \mathbf{y} - \mathbf{f}(\alpha_A)$, $A_{li} = \partial f_l(\alpha)/\partial\alpha_i|_{\alpha_A}$ is a constant matrix, $\eta = \alpha - \alpha_A$ is the new vector of unknowns, and $\mathbf{V}_y^{-1}$ is the inverse of the covariance matrix of the measurements. Since the χ^2 measures how much the measurements "miss" the function, the solution we want is that which minimizes χ^2, i.e., $\partial\chi^2/\partial\eta_i = \mathbf{0}$. The solution was found to be $\eta = \mathbf{V}_A \mathbf{A}^t \mathbf{V}_y^{-1} \mathbf{\Delta y}$ with covariance matrix $\mathbf{V}_\eta$, where $\mathbf{V}_\eta = \mathbf{V}_A = (\mathbf{A}^t \mathbf{V}_y^{-1} \mathbf{A})^{-1}$.

Now suppose we have many hundreds, or even thousands, of such fits and that each fit l depends on parameters η_l which are local to it as well as some unknown parameters $\mathbf{v}$ (s of them) which affect all fits (note: when I use the word "fit" I refer to any complete fit, for example a single track or even a set of tracks within an event). Our goal is to find $\mathbf{v}$. For this case the χ^2 can be written as the sum

$$\chi^2 = \sum_l (\mathbf{\Delta y}_l - \mathbf{A}_l \eta_l - \mathbf{B}_l \mathbf{v})^t \mathbf{V}_{y\,l}^{-1} (\mathbf{\Delta y}_l - \mathbf{A}_l \eta_l - \mathbf{B}_l \mathbf{v}),$$

where I have assumed that the parameters $\mathbf{v}$ are small enough that their effect on the measurements $\mathbf{y}_l$ can be expressed linearly as $\delta\mathbf{y}_l = \mathbf{B}_l \mathbf{v}$, with $\mathbf{B}_l$ a $n \times s$ matrix. Minimizing this total χ^2 function with respect to the variables η and $\mathbf{v}$ we obtain the set of equations

$$
\begin{aligned}
\eta_l : \qquad & \mathbf{A}_l^t \mathbf{V}_y l^{-1} (\mathbf{\Delta y}_l - \mathbf{A}_l \eta_l - \mathbf{B}_l \mathbf{v}) = \mathbf{0}, \\
\mathbf{v} : \qquad & \sum_l \mathbf{B}_l^t \mathbf{V}_y l^{-1} (\mathbf{\Delta y}_l - \mathbf{A}_l \eta_l - \mathbf{B}_l \mathbf{v}) = \mathbf{0}.
\end{aligned}
$$

Although the number of equations is huge, straightforward algebra gives a surprisingly simple solution for $\mathbf{v}$ and its covariance matrix $\mathbf{V_v}$ in terms of two matrices, $\mathbf{s}$ and $\mathbf{S}$:

$$
\begin{aligned}
\mathbf{v} &= \mathbf{S}^{-1}\mathbf{s} \\
\mathbf{V_v} &= \mathbf{S}^{-1}
\end{aligned}
$$

where $\mathbf{s}$ and $\mathbf{S}$ are defined as

$$
\begin{aligned}
\mathbf{s} &= \sum_l \mathbf{B}_l^t \mathbf{V}_{y\,l}^{-1} (\mathbf{\Delta y}_l - \mathbf{A}_l \eta_{l0}) \equiv \sum_l \mathbf{B}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{r}_l \\
\mathbf{S} &= \sum_l (\mathbf{B}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{B}_l - \mathbf{U}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{U}_l) \\
\eta_{l0} &= \mathbf{V}_{A\,l} \mathbf{A}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{\Delta y}_l \\
\eta_l &= \eta_{l0} - \mathbf{V}_{A\,l} \mathbf{U}_l \mathbf{v}.
\end{aligned}
$$

The auxiliary matrices $\mathbf{V}_{A\,l}$ and $\mathbf{U}_l$ are defined as $\mathbf{V}_{A\,l} = (\mathbf{A}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{A}_l)^{-1}$ and $\mathbf{U}_l = \mathbf{A}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{B}_l$. Note that the equation for $\mathbf{s}$ involves the term $\mathbf{r}_l = \mathbf{\Delta y}_l - \mathbf{A}_l \eta_{l0}$, which is the just the vector of residuals for each individual fit l, ignoring $\mathbf{v}$.

The end result is that the parameters $\mathbf{v}$ can be obtained by performing a large number of individual fits (ignoring the effect of $\mathbf{v}$) and then accumulating the results of each fit in two matrices, $\mathbf{s}$ and $\mathbf{S}$. Recall that the word "accumulating" refers to a sum over complete fits. In high energy physics language this normally means a sum over individual tracks, but it could even be a sum over events if all the tracks within an event were fit together. The sequence of steps that should be followed is shown below.

1. For each fit in which the data values $\mathbf{y}_l$ and coefficients $\mathbf{A}_l$ are given, compute $\mathbf{\Delta y} = \mathbf{y} - \mathbf{f}(\alpha_A)$, and $\mathbf{V}_{A\,l}$, solve for η_{0l} as shown above and compute the residuals $\mathbf{r}_l$. This step assumes nothing about the effect of the $\mathbf{v}$ parameters.

2. Compute $\mathbf{B}_l$ and $\mathbf{U}_l$ and accumulate over some sample of fits the matrices $\mathbf{s} = \sum_l \mathbf{B}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{r}_l$ and $\mathbf{S} = \sum_l (\mathbf{B}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{B}_l - \mathbf{U}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{U}_l)$.

3. After accumulating the above matrices for awhile, occasionally invert $\mathbf{S}$ to get $\mathbf{V}_\mathbf{v}$ (remember that $\mathbf{V}_\mathbf{v} = \mathbf{S}^{-1}$) to see if the errors on $\mathbf{v}$ are small enough.

4. If the errors on $\mathbf{v}$ are not small enough, go to step 1 and accumulate more data. If enough data has been taken, solve for $\mathbf{v} = \mathbf{S}^{-1}\mathbf{s}$.

Notice that you will not be able to compute the corrected values of η_l during the accumulation since they depend on the value of $\mathbf{v}$. Usually this is not important since what is most often desired is $\mathbf{v}$ and its covariance matrix $\mathbf{V}_\mathbf{v}$. Of course, if you want the best parameters η_l for each individual fit l, you can always go back and reanalyze the data using the now known values of $\mathbf{v}$ via the correction $\eta_l = \eta_{l0} - \mathbf{V}_{A\,l}\mathbf{U}_l\mathbf{v}$.

3. Finding Systematic Effects Using Constrained Fits

It is fairly easy to extend the results of the previous section to the case where the parameters $\mathbf{v}$ describing the systematic effects must be determined using fits which were performed with constraints. This situation occasionally arises in high energy physics when individual tracks carry insufficient information to determine a systematic parameter but a set of tracks satisfying a global constraint does. For example, if the lower and upper hemispheres of a drift chamber were displaced relative to one another, the shift would be in principle unobservable in individual tracks because each track would lie completely within a single hemisphere. However, the shift would have an effect if one constrained the dimuons in $e^+e^- \to \mu^+\mu^-$ events to meet in the middle.

Let the r constraint equations for the l^{th} fit be written $\mathbf{H}_l(\alpha_l) = \mathbf{0}$. Expanding about α_A as before, we obtain the linearized equations $\mathbf{H}_l(\alpha_{A\,l}) + (\alpha_l - \alpha_{A\,l})\partial\mathbf{H}_l(\alpha_{A\,l})/\partial\alpha_l \equiv \mathbf{D}_l\eta_l + \mathbf{d}_l = \mathbf{0}$, using obvious notation, where $\mathbf{D}_l$ is a $r \times m$ matrix and $\mathbf{d}_l$ is a vector of length r. Note that the number of constraints r can vary from fit to fit. The full χ^2 equation can now be written, using the method of Lagrange multipliers, as

$$\chi^2 = \sum_l (\mathbf{\Delta y}_l - \mathbf{A}_l\eta_l - \mathbf{B}_l\mathbf{v})^t \mathbf{V}_{y\,l}^{-1} (\mathbf{\Delta y}_l - \mathbf{A}_l\eta_l - \mathbf{B}_l\mathbf{v}) + \sum_l 2\lambda_l^t(\mathbf{D}_l\eta_l + \mathbf{d}_l),$$

which, when minimized with respect to the parameters η_l, $\mathbf{v}$ and λ_l, yields the equations

$$\begin{aligned}
\eta_l: &\quad -\mathbf{A}_l^t\mathbf{V}_{y\,l}^{-1}(\mathbf{\Delta y}_l - \mathbf{A}_l\eta_l - \mathbf{B}_l\mathbf{v}) + \mathbf{D}_l^t\lambda_l = \mathbf{0} \\
\mathbf{v}: &\quad \sum_l \mathbf{B}_l^t\mathbf{V}_{y\,l}^{-1}(\mathbf{\Delta y}_l - \mathbf{A}_l\eta_l - \mathbf{B}_l\mathbf{v}) = \mathbf{0} \\
\lambda_l: &\quad \mathbf{D}_l\eta_l + \mathbf{d}_l = \mathbf{0}
\end{aligned}$$

Again, straightforward matrix algebra yields the solution for $\mathbf{v}$ and its covariance matrix $\mathbf{V_v}$ in terms of the accumulating matrices $\mathbf{s}$ and $\mathbf{S}$:

$$\begin{aligned}
\mathbf{v} &= \mathbf{S}^{-1}\mathbf{s}, \\
\mathbf{V_v} &= \mathbf{S}^{-1},
\end{aligned}$$

where $\mathbf{s}$ and $\mathbf{S}$ are given by

$$\begin{aligned}
\mathbf{s} &= \sum_l \mathbf{B}_l^t\mathbf{V}_{y\,l}^{-1}(\mathbf{\Delta y}_l - \mathbf{A}_l\eta'_{l0}) \equiv \sum_l \mathbf{B}_l^t\mathbf{V}_{y\,l}^{-1}\mathbf{r}_l \\
\mathbf{S} &= \sum_l (\mathbf{B}_l^t\mathbf{V}_{y\,l}^{-1}\mathbf{B}_l - \mathbf{U}_l^t\mathbf{V}_{A\,l}\mathbf{U}_l + \mathbf{W}_l^t\mathbf{V}_{D\,l}\mathbf{W}_l) \\
\eta'_{l0} &= \eta_{l0} - \mathbf{V}_{A\,l}\mathbf{D}_l^t\lambda_{l0} \\
\eta_{l0} &= \mathbf{V}_{A\,l}\mathbf{A}_l^t\mathbf{V}_{y\,l}^{-1}\mathbf{\Delta y}_l \\
\lambda_{l0} &= \mathbf{V}_{D\,l}(\mathbf{D}_l\eta_{l0} + \mathbf{d}_l) \\
\lambda_l &= \lambda_{l0} - \mathbf{V}_{D\,l}\mathbf{W}_l\mathbf{v} \\
\eta_l &= \eta_{l0} - \mathbf{V}_{A\,l}\mathbf{D}_l^t\lambda_l - \mathbf{V}_{A\,l}\mathbf{U}_l\mathbf{v}.
\end{aligned}$$

The auxiliary matrices are given by $\mathbf{V}_{A\,l} = (\mathbf{A}_l^t\mathbf{V}_{y\,l}^{-1}\mathbf{A}_l)^{-1}$, $\mathbf{U}_l = \mathbf{A}_l^t\mathbf{V}_{y\,l}^{-1}\mathbf{B}_l$, $\mathbf{V}_{D\,l} = (\mathbf{D}_l\mathbf{V}_{A\,l}\mathbf{D}_l^t)^{-1}$ and $\mathbf{W}_l = \mathbf{D}_l\mathbf{V}_{A\,l}\mathbf{U}_l$. Note that η_{l0} and η'_{l0} are, respectively, the unconstrained and constrained solutions for the individual fits obtained without knowledge of the systematic parameters $\mathbf{v}$.

The solution should be obtained using the following sequence of steps.

1. For each fit in which the data values $\mathbf{y}_l$ and coefficients $\mathbf{A}_l$ are given, compute $\mathbf{\Delta y} = \mathbf{y} - \mathbf{f}(\alpha_A)$ and $\mathbf{V}_{A\,l}$ and solve for the unconstrained parameters η_{0l} as shown above. The effects of $\mathbf{v}$ are ignored here.

2. Determine $\mathbf{V}_{D\,l}$, solve for λ_{l0} and the constrained parameters η'_{l0} and compute the residuals $\mathbf{r}_l$. The effects of $\mathbf{v}$ still play no role here.

3. Compute $\mathbf{B}_l$, $\mathbf{U}_l$, $\mathbf{W}_l$ and accumulate over some sample of fits the two matrices $\mathbf{s} = \sum_l \mathbf{B}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{r}_l$ and $\mathbf{S} = \sum_l (\mathbf{B}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{B}_l - \mathbf{U}_l^t \mathbf{V}_{y\,l}^{-1} \mathbf{U}_l + \mathbf{W}_l^t \mathbf{V}_{D\,l} \mathbf{W}_l)$.

4. After accumulating the above matrices for awhile, occasionally invert $\mathbf{S}$ to get $\mathbf{V}_{\mathbf{v}}$ (remember that $\mathbf{V}_{\mathbf{v}} = \mathbf{S}^{-1}$) to see if the errors on $\mathbf{v}$ are small enough.

5. If the errors on $\mathbf{v}$ are not small enough, go to step 1 and accumulate more data. If enough data has been taken, solve for $\mathbf{v} = \mathbf{S}^{-1}\mathbf{s}$.

As mentioned in the last section, you will not be able to compute the corrected values of η_l during the accumulation since they depend on the value of $\mathbf{v}$. However, after determining $\mathbf{v}$ you can go back and reanalyze the data.

4. Tracking example

This example is derived from a real application in the CLEO central tracking detector. The central detector consisted at one time of three separate pieces: (1) a 3 layer microvertex chamber consisting of straws; (2) a 10 layer gaseous vertex detector; and (3) a 51 layer drift chamber. The large drift chamber defined the coordinate system and the other two chambers had no z information available on the anode wires. Since the accuracy of chamber placement was far inferior to the internal machining of each chamber, 10 positioning parameters, 5 for each chamber, had to be determined:

1. Offset in x

2. Offset in y

3. Rotation in the $x - z$ plane

4. Rotation in the $y - z$ plane

5. Rotation in the $x - y$ plane

The z offset could not be determined because there was no z information on the anode wires of the inner chambers.

The quantity that is measured by all the drift chambers is the drift distance, which is the distance of closest approach of the track to the wire. Let ϕ_k be the azimuthal position at which the track crosses the radial position of the wire and α_k be the crossing angle in the $r-\phi$ plane (a perfectly radial track would have $\alpha_k = 0$). Let v_{1-5} be the positioning parameters for the first chamber and v_{6-10} be the parameters for the second chamber. If we call δd_k the change in the drift distance for layer k due to v_i, then we can write the contribution to the drift distance as

$$\delta d_k = \sum_i B_{ki} v_i.$$

For $1 \leq k \leq 3$ (the first chamber), we can determine B_{ki} to be

$$\begin{aligned}
B_{k1} &= \sin\phi_k \cos\alpha_k \\
B_{k2} &= -\cos\phi_k \cos\alpha_k \\
B_{k3} &= z_k \sin\phi_k \cos\alpha_k \\
B_{k4} &= -z_k \cos\phi_k \cos\alpha_k \\
B_{k5} &= -r_k \cos\alpha_k \\
B_{ki} &= 0 \qquad i \geq 6,
\end{aligned}$$

where r_k and z_k are, respectively, the radius of the layer and z position of the hit (as determined by an initial, approximate fit). For $4 \leq k \leq 13$ (the second chamber), B_{ki} is

$$\begin{aligned}
B_{k6} &= \sin\phi_k \cos\alpha_k \\
B_{k7} &= -\cos\phi_k \cos\alpha_k \\
B_{k8} &= z_k \sin\phi_k \cos\alpha_k \\
B_{k9} &= -z_k \cos\phi_k \cos\alpha_k \\
B_{k10} &= -r_k \cos\alpha_k \\
B_{ki} &= 0 \qquad i \leq 5,
\end{aligned}$$

For $14 \leq k \leq 64$ (the third chamber), $B_{ki} = 0$.

The procedure for finding these 10 constants is straightforward. For each track in the data sample, we perform a fit using all three chambers to determine the track parameters η_0

and their covariance matrix $\mathbf{V}_A$. Each track is then traced through the detector to compute the values of r_k (the residual), z_k, r_k, ϕ_k and α_k for each layer. Finally, we calculate the matrix B_{ki} and accumulate the two matrices $\mathbf{s}$ and $\mathbf{S}$ over the data sample. After sufficient data has been gathered, we compute the 10 constants $\mathbf{v} = \mathbf{S}^{-1}\mathbf{s}$ and their covariance matrix $\mathbf{V}_\mathbf{v} = \mathbf{S}^{-1}$. Note that this approach requires only a single pass over the data. By comparison, the conventional approach in which highly selected data (i.e., tracks lying close to the x or y axes) are used to isolate each individual parameter requires several iterations to converge properly.

5. **Limitations of This Algorithm**

I should comment before closing this paper that determining drift chamber positioning constants with this χ^2 technique can lead to problems if you try to push it too far in terms of accuracy, especially if the constants are highly correlated. The reason is that once you get to the few micron level, many new effects arise that can partially mimic the ones you are trying to determine. These effects can systematically skew the fit depending on the precise set of tracks you pick. In CLEO, for example, there is a correction to the drift measurement due to pulse height saturation that is a function of θ, the dip angle. Unfortunately, θ is highly correlated with the z value of the track at any layer, causing problems in determining the tilts of chambers since the effect of tilt on drift distance is linear in z.

Another example from CLEO is illustrative. Layer by layer measurements of drift distance have an $E \times B$ correction that has the net effect of making the time–distance $(t - d)$ relation asymmetric, i.e., left is different from right. This effect interferes with the ability to precisely determine the $x - y$ rotation angle of one chamber relative to another. If it were symmetric, then the fact that there would be roughly an equal number of tracks on the left or right of any cell would cause any uncertainty in the $t - d$ relation to cancel when determining the rotation angle. A possible solution is to avoid using the innermost or outermost layers (since these have the largest asymmetry) and only use drift distances well within the cell, so that $E \times B$ effects are minimized.

Avoidance of systematic bias can be tricky. For example, if you plan to use Bhabha events $(e^+e^- \to e^+e^-)$ to measure chamber positions and orientations, be aware of the fact that

Bhabha events are primarily forward scattered, meaning that positive charged tracks tend to have a particular sign of p_z and negative tracks another. This can cause great problems in determing chamber positions and orientations if there are systematic effects which are charge dependent. Suppose, for example, that the drift chamber endplates are slightly rotated with respect to one another. This causes the curvature of tracks to be slightly shifted systematically, resulting in a change of momentum. If positive tracks along the $+z$ direction get an increase in momentum, then they are reduced in momentum along the $-z$ direction. The opposite behavior holds for negative tracks. Now, if an equal number of positive and negative tracks are used in each direction (for instance by using $e^+e^- \to \mu^+\mu^-$ events) then the effect cancels. It does not cancel for Bhabha events because of the aforementioned asymmetry.