Paul Avery
CBX 91 − 74
October 18, 1991

Applied Fitting Theory III
Non-Optimal Least Squares Fitting and Multiple Scattering

1. The Problem

In the first paper in this series, I gave a general overview of least squares fitting theory and showed that if the dependence of the measurements on the parameters can be linearized, the parameters which minimize the χ^2 function can be obtained rather easily, even in the presence of constraints.

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

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

where $\mathbf{\Delta y} = \mathbf{y} - \mathbf{f}_A(\alpha_A)$, $A_{li} = \partial f_l(\alpha)/\partial\alpha_i|_{\alpha_A}$ 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 $\mathbf{V}_\eta = \mathbf{V}_A = (\mathbf{A}^t \mathbf{V}_y^{-1} \mathbf{A})^{-1}$. The Gauss-Markov theorem states that the parameters obtained by this procedure are both *unbiased* and have *minimum* uncertainties, i.e., they are the best parameters that can be determined by any method. So far, so good.

Under normal circumstances the measurment errors σ_l are independent of each other, i.e., the weight matrix $\mathbf{V}_y^{-1}$ can be written in diagonal form:

$$\mathbf{V}_y^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_n^2 \end{pmatrix},$$

which is obtained by inversion of the diagonal covariance matrix $V_{y\,ij} = \langle \delta y_i \delta y_j \rangle = \sigma_i^2 \delta_{ij}$. Thus it is a trivial matter to fit an arbitrary number of measurements.

There are situations, however, where the data measurements y_l are not independent of each other, so that movement of a particular measurement will cause the simultaneous adjustment of other measurements. An example of this is multiple scattering, in which a single scattering event at a particular point affects the drift distance measurements at all subsequent points. In this case the deviation of the measurements from the original track will have an independent component due to random measurement errors and a correlated component due to the presence of the multiple scattering event. The correlations among the data points can be computed and put into the measurement covariance matrix $\mathbf{V}_y$. The problem now is obtaining $\mathbf{V}_y^{-1}$: the inversion of the covariance matrix can be prohibitively time consuming when the number of measurements n is large. In the CLEO central detector, for instance, a track can have as many as 71 measurements. Inverting a 71×71 matrix at least once per track to account for multiple scattering is just not very practical, unless one has a lot of CPU time to waste.

2. Non-Optimal Least Squares Fitting

Several attempts have been made in the literature to account for multiple scattering effects in track fitting. The methods employed range from the mundane to the elegant, but a common goal links all of them: finding a set of track parameters whose errors are as close as possible to the optimal values calculated in Section 1, but without the high cost in CPU time which that technique entails. I call this the "optimal" approach.

For most tracks inclusion of multiple scattering effects in the fitting algorithm does not improve the parameters of the fit significantly, except for very slow tracks. A large amount of programming and CPU time is invested in making marginal improvements in the track parameters. In fact, what is many times desired is not the *best* possible parameters, but parameters whose errors are *well understood.* Knowledge of these errors is important when they are used as input to kinematic fitting or when a sensitive lifetime measurement depends critically on uncertainties in the vertex resolution. In both cases, lack of understanding of the parameter errors and their correlations can give erroneous physics results, especially when those results are based on a quoted number of standard deviations.

I start by noting that the procedure whereby one looks for parameters which minimize the χ^2 function is merely one possible estimation scheme — although it is the optimal one — and that others are possible. Assume, for instance, that an approximation for $\mathbf{V}_y^{-1}$ exists which we call $\tilde{\mathbf{V}}_y^{-1}$. $\tilde{\mathbf{V}}_y^{-1}$ will almost certainly be diagonal, but this assumption is not necessary in what is to follow. We define a modified form of the χ^2 function

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

The parameters obtained by minimizing $\tilde{\chi}^2$ are given by $\eta = \tilde{\mathbf{V}}_A \mathbf{A}^t \tilde{\mathbf{V}}_y^{-1} \mathbf{\Delta y}$, where $\tilde{\mathbf{V}}_A = (\mathbf{A}^t \tilde{\mathbf{V}}_y^{-1} \mathbf{A})^{-1}$. The covariance matrix must be calculated from the definition

$$\mathbf{V}_\eta \equiv \langle \delta\eta\delta\eta^t \rangle = \tilde{\mathbf{V}}_A \mathbf{A}^t \tilde{\mathbf{V}}_y^{-1} \langle \delta\mathbf{y}\delta\mathbf{y}^t \rangle \tilde{\mathbf{V}}_y^{-1} \mathbf{A} \tilde{\mathbf{V}}_A = \tilde{\mathbf{V}}_A \mathbf{A}^t \tilde{\mathbf{V}}_y^{-1} \mathbf{V}_y \tilde{\mathbf{V}}_y^{-1} \mathbf{A} \tilde{\mathbf{V}}_A.$$

This equation collapses to $\mathbf{V}_\eta = \mathbf{V}_A$ when $\tilde{\mathbf{V}}_y^{-1} = \mathbf{V}_y^{-1}$.

Notice what has been accomplished here: I have produced an estimate of the fitting parameters and their associated covariance matrix *without inverting the measurement covariance matrix* $\mathbf{V}_y$. Furthermore, the estimate is *unbiased* (the parameters converge to the true values if the experiment is repeated indefinitely) and the covariance matrix — and hence the errors — are known *exactly.* The price paid for this estimation scheme is that the parameter errors are no longer the best possible; that would require the minimization of the original χ^2 function. However, if $\tilde{\mathbf{V}}_y^{-1}$ is a reasonable representation of $\mathbf{V}_y^{-1}$ then the "non-optimal" estimate may in fact be close to optimal.

3. Non-Optimal Fitting with Constraints

The effect of constraints can easily be taken into account by the standard Lagrange Multiplier method where the "initial" unconstrained values and their covariance matrix are obtained by the non-optimal fitting procedure. This procedure works exactly the same as if the parameters were obtained by the normal optimal least squares technique. The reason is because constraints act only to move the fitted parameters away from their unconstrained values and the relative motion of each of the parameters is governed by the covariance matrix of the fit, which has absorbed all the information regarding details of the measurement errors and correlations. This point was discussed in Section 5 of the first paper in this series, CBX 91–72.

4. Simple Application to Multiple Scattering

Particles moving through a detector suffer innumerable collisions which alter the trajectory of the particle by a stochastic process. In general, a particle traversing material of length Δx and a radiation length L_R will be deflected in any particular plane by an random angle θ whose r.m.s. value can be calculated from

$$\langle \theta^2 \rangle = \left(\frac{0.0141}{p\beta} \right)^2 \frac{\Delta x}{L_R} = H \Delta x,$$

where p is the momentum in GeV/c and β is the velocity in units of c. Note that this is the width in each of two planes perpendicular to each other and lying along the flight path. I am only accounting for standard multiple scattering effects that can be described by Gaussian statistics; single scattering theory is too complicated to incorporate in least squares fitting, at least by me.

Now consider a particle moving in a solenoidal detector (having a magnetic field in the z direction) from a point near the center and assume that the drift distance to a wire is measured in every layer. If the particle has sufficient momentum so that there is not much bending, we can equate this situation to a simple geometry in which n planar drift chamber are arrayed along the x axis which is defined to be the direction of the particle in the $r - \phi$ plane (i.e., x is really a measurement of the radius r and y measures the drift distance). The covariance matrix of the measurements can be represented by the sum $\mathbf{V}_y = \mathbf{V}_{y0} + \mathbf{V}_{yD} + \mathbf{V}_{yC}$, where $\mathbf{V}_{y0} = \mathrm{diag}(\sigma_i^2)$ is the component due to independent measurement errors, $\mathbf{V}_{yD}$ is due to multiple scattering at discrete surfaces and $\mathbf{V}_{yC}$ is a result of continuous multiple scattering in gas.

Let's consider discrete scatterers first. If material of thickness t is present at $x = x_S$, the track will be deflected stochastically from its nominal flight path by an angle θ_s, causing the drift distance measurements at plane i to be changed by an amount $\delta y_i = \theta_s(x_i - x_S)$. The covariance matrix element $(V_{yD})_{ij}$ can then be calculated from its definition, viz.

$$(V_{yD})_{ij} = \langle \delta y_i \delta y_j \rangle = \langle \theta_s^2 \rangle (x_i - x_S)(x_j - x_S) \equiv Ht(x_i - x_S)(x_j - x_S),$$

and is 0 if $x_i \leq x_S$ or $x_j \leq x_S$. Note that t should reflect the total path length of the track through the scatterer; this is expecially important for tracks having significant components in the z direction.

Continuous scattering is slightly more complicated. Assume the presence of gas in the region $x_{G1} \leq x \leq x_{G2}$, where $L = x_{G2} - x_{G1}$. If we divide up the path into small regions Δx the total scattering angle and deviation up to a point x in the gas volume is

$$\begin{aligned} \theta &= \sum_{l=1}^{N} \delta\theta_l \\ \delta y &= \sum_{l=1}^{N} \theta_l \Delta x = \sum_{l=1}^{N} (N - l)\delta\theta_l \Delta x, \end{aligned}$$

where $\Delta x = (x - x_{G1})/N$. Thus the covariance matrix element for points x_i and x_j (where $x_j \geq x_i$) in the gas volume is

$$\begin{aligned} (V_{yG})_{ij} &= \sum_{l=1}^{N_i} \sum_{l'=1}^{N_j} (N_i - l)(N_j - l')\langle \delta\theta_l \delta\theta_{l'} \rangle \Delta x^2 = \int_{x_{G1}}^{x_i} H(x_i - x)(x_j - x)dx \\ &= H\frac{1}{2}(x_i - x_{G1})^2 (x_j - x_{G1} - \frac{1}{3}(x_i - x_{G1})), \end{aligned}$$

where I have used $\langle \delta\theta_l \delta\theta_{l'} \rangle = H \Delta x \delta_{ll'}$.

If x_i and x_j both are beyond the gas region the above equations have to be modified to

$$\theta = \sum_{l=1}^{N_G} \delta\theta_l$$
$$\delta y = \sum_{l=1}^{N_G} \theta_l \Delta x + \theta(x - x_{G2}) = \sum_{l=1}^{N_G} (N_G - l)\delta\theta_l \Delta x + \theta(x - x_{G2}),$$

where $N_G = L/\Delta x$. In this case the covariance matrix is

$$(V_{yD})_{ij} = HL\left\{\frac{1}{3}L^2 + \frac{1}{2}L[(x_i - x_{G2}) + (x_j - x_{G2})] + (x_j - x_{G2})(x_i - x_{G2})\right\}.$$

If x_j is beyond the gas region and x_i is inside it the covariance matrix becomes

$$(V_{yD})_{ij} = H\frac{1}{2}(x_i - x_{G1})^2(L - \frac{1}{3}(x_i - x_{G1}) + (x_j - x_{G2})).$$

x_i and x_j are interchanged in the above equation if x_i is beyond the gas region and x_j is inside of it.

5. Summary of Multiple Scattering Formulas

A summary of the contributions to the multiple scattering matrix $\mathbf{V}_y = \mathbf{V}_{y0} + \mathbf{V}_{yD} + \mathbf{V}_{yC}$ is shown below (remember that x_j and x_i are defined such that $x_j \geq x_i$).

1. Measurement error:

$$(V_{y0})_{ij} = \sigma_i^2 \delta_{ij}$$

where σ_i is the measurement error of layer i.

2. Scattering from discrete material (thin) at $x = x_S$ and thickness t (remember that $H = (0.0141/p\beta)^2/L_R$ is the square of the r.m.s. scattering angle per unit length in a plane

and L_R is the radiation length):

$$(V_{yD})_{ij} = Ht(x_i - x_S)(x_j - x_S) \quad \text{for } x_i, x_j \geq x_S$$
$$= 0 \quad \text{otherwise}$$

3. Scattering in a gas region $x_{G1} \leq x \leq x_{G2}$, $L = x_{G1} - x_{G2}$:

$$(V_{yG})_{ij} = H\frac{1}{2}(x_i - x_{G1})^2(x_j - x_{G1} - \frac{1}{3}(x_i - x_{G1}))$$
$$\text{for } x_{G1} \leq x_i, x_j \leq x_{G2}$$
$$(V_{yD})_{ij} = H\frac{1}{2}(x_i - x_{G1})^2(L - \frac{1}{3}(x_i - x_{G1}) + (x_j - x_{G2}))$$
$$\text{for } x_{G1} \leq x_i \leq x_{G2},\ x_j \geq x_{G2}$$
$$(V_{yD})_{ij} = H\frac{1}{2}(x_j - x_{G1})^2(L - \frac{1}{3}(x_j - x_{G1}) + (x_i - x_{G2}))$$
$$\text{for } x_{G1} \leq x_j \leq x_{G2},\ x_i \geq x_{G2}$$
$$(V_{yD})_{ij} = HL\left\{\frac{1}{3}L2 + \frac{1}{2}L[(x_i - x_{G2}) + (x_j - x_{G2})] + (x_j - x_{G2})(x_i - x_{G2})\right\}$$
$$\text{for } x_i, x_j \geq x_{G2}$$
$$= 0 \quad \text{otherwise}$$