Vertexing and Kinematic Fitting, Part I: Savior of CLEO or Agent of Evil?

Paul Avery
University of Florida
Sept. 27, 1996

avery@phys.ufl.edu
http://www.phys.ufl.edu/~avery

Overview of plan

- 1st of several lectures on kinematic fitting
- Focus in this lecture on theory
- Plan of future lectures
  - Lecture 2: Introduction to KNLIB
  - Lecture 3: Vertex fitting with KNLIB
  - Lecture 4: Fitting decay chains with KNLIB
- References
  - **KNLIB**
    http://www.phys.ufl.edu/~avery/knlib/knlib.html or
    http://w4.lns.cornell.edu/~pra/knlib/knlib.html
  - Writeups on different aspects of fitting theory and constraints
    http://www.phys.ufl.edu/~avery/fitting.html

What is Kinematic Fitting?

Kinematic fitting is a mathematical procedure in which one uses the physical laws governing a particle interaction or decay to improve the measurements of the process.

For example, the fact that the tracks coming from $D^0 \to K^-\pi^+\pi^+\pi^-$ decay must come from a common space point can be used to improve the 4-momentum and positions of the daughter particles, thus improving the mass resolution of the D^0.

Physical information is supplied via *constraints*. Each constraint is expressed in the form of an equation expressing some physical condition that the process must satisfy. In the example above, each track contributes 2 constraints (r–ϕ and z) to the vertex requirement, giving 8 constraints in all.

Vertexing is only one example. We can require instead that the invariant mass of the particles be equal to 1.8654. This is known as a *mass constraint*. We will discuss mass constraints later.

Implementation of Constraints

Constraints are generally implemented through a least-squares procedure. Each constraint equation is linearized and added, via the Lagrange multiplier technique, to the χ^2 equation of the tracks using the covariance matrices of the tracks. Each track contributes a 5 parameter "measurement", and the 5×5 track covariance matrix is the generalization of the σ^2 for a single measurement.

One then minimizes the χ^2 simultaneously with the constraint conditions. The constraints "pull" the tracks away from their unconstrained values, and the resulting χ^2 one obtains with n constraints is distributed like a standard χ^2 with n degrees of freedom, if gaussian errors apply. A histogram of fits to, say, 10,000 decays would clearly show this distribution. Of course, since track errors are only approximately gaussian, the actual distribution will have more events in the tail than predicted by theory. Still, knowledge of the distribution allows one to define reasonable χ^2 cuts.

For example, in the vertexing example $D^0 \to K^-\pi^+\pi^+\pi^-$, there are a total of 8 constraints, but 3 unknown parameters must be determined (the D^0 vertex). The total number of degrees of freedom is thus 2*4 – 3 = 5.

Trivial Example

Let's work out all the least squares machinery for a simple example. Suppose we have two measurements, x_1 and x_2 with (independent) errors 0.1. Now we impose the condition that we want the two variables to sum to 6. Why 6? I don't know; just humor me for now.

Without the constraint condition, the total χ^2 of the measurements could be written

$$\chi^2 = \frac{\left(x_1 - x_{10}\right)^2}{\sigma_1^2} + \frac{\left(x_2 - x_{20}\right)^2}{\sigma_2^2}$$

where x_{10} and x_{20} are the initial measurements of x_1 and x_2, and $\sigma_1 = \sigma_2 = 0.1$. Since there is no reason yet for the measurements to stray from their initial values, $\chi^2 = 0$ initially.

The constraint is imposed using the Lagrange multiplier method, e.g.

$$\chi^2 = \frac{\left(x_1 - x_{10}\right)^2}{\sigma_1^2} + \frac{\left(x_2 - x_{20}\right)^2}{\sigma_2^2} + 2\lambda(x_1 + x_2 - 6)$$

where λ is a lagrange multiplier which must be determined (the factor of 2 is inserted to simplify the algebra). We minimize the χ^2 by setting the partial derivatives wrt x_1, x_2 *and also* λ to 0. This yields, using $\sigma_1 = \sigma_2 = \sigma$

$$0 = \left(x_1 - x_{10}\right) / \sigma^2 + \lambda$$

$$0 = \left(x_2 - x_{20}\right) / \sigma^2 + \lambda$$

$$0 = x_1 + x_2 - 6$$

Using the first two equations to eliminate λ, we solve for x_1 and x_2:

$$x_1 = 3 + 0.5\left(x_{10} - x_{20}\right)$$

$$x_2 = 3 - 0.5\left(x_{10} - x_{20}\right)$$

Error Analysis

The solution is only half the story, because what we're *really* interested in is the error of the updated parameters. From the above discussion, we fully expect that the constraint will reduce the errors.

We calculate the errors for x_1 and x_2 directly from the definition of standard deviation, by averaging over all possible measurements. For example, for x_1 we get the variance:

$$\begin{aligned}\sigma_{x_1}^2 &= \langle \delta x_1 \delta x_1 \rangle = 0.5^2 \left\langle \left(\delta x_{10} + \delta x_{20} \right)^2 \right\rangle \\ &= 0.5^2 \left\langle \delta x_{10}^2 + 2\delta x_{10} \delta x_{20} + \delta x_{20}^2 \right\rangle \\ &= 0.5\sigma^2\end{aligned}$$

where I used the independence of the initial measurements and $\left\langle \delta x_{10}^2 \right\rangle = \left\langle \delta x_{20}^2 \right\rangle = \sigma^2$. The same result holds for x_2. Thus the errors are

$$\begin{aligned}\sigma_{x_1} &= 0.071 \\ \sigma_{x_2} &= 0.071\end{aligned}$$

which are substantially smaller than before.

However, the fit has introduced a correlation between the updated parameters which was not originally present. We define the *covariance* of x_1 and x_2 as

$$V_{x_1 x_2} = \langle \delta x_1 \delta x_2 \rangle$$

Plugging in the expresssions for x_1 and x_2 yields

$$\begin{aligned} V_{x_1 x_2} &= -0.5^2 \left\langle \left(\delta x_{10} - \delta x_{20} \right)^2 \right\rangle \\ &= -0.5\sigma^2 \end{aligned}$$

The familiar correlation coefficient $r_{x_1 x_2}$ is more commonly used to express the variation of one parameter with another. It is defined as

$$r_{x_1 x_2} = \frac{V_{x_1 x_2}}{\sigma_{x_1} \sigma_{x_2}}$$

Our simple constraint leads to $r_{x_1 x_2} = -1$, i.e., every fluctuation of x_1 *upward* is matched by an equal fluctuation of x_2 *downward*. Other kinds of constraints lead to different correlations.

The Covariance Matrix

The error information for more than one variable is more elegantly expressed in terms of the “covariance matrix”. For example, let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

The covariance matrix of the two variables wrt one another is $V_{x\,ij} \equiv \left\langle \delta x_i \delta x_j \right\rangle$, or in matrix form

$$\mathbf{V_x} \equiv \left\langle \delta\mathbf{x}\delta\mathbf{x}^T \right\rangle = \left\langle \begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix} \begin{pmatrix} \delta x_1 & \delta x_2 \end{pmatrix} \right\rangle$$

$$= \begin{pmatrix} \left\langle \delta x_1 \delta x_1 \right\rangle & \left\langle \delta x_1 \delta x_2 \right\rangle \\ \left\langle \delta x_2 \delta x_1 \right\rangle & \left\langle \delta x_2 \delta x_2 \right\rangle \end{pmatrix}$$

It is clear from the definition that $\mathbf{V_x}$ is symmetric ($V_{x\,ij} = V_{x\,ji}$) and the diagonal elements are just the squares of the standard deviations ($V_{x\,ii} = \sigma_i^2$).

The initial and final covariance matrices are then

$$\mathbf{V}_{\mathbf{x}_0} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \qquad \mathbf{V_x} = \begin{pmatrix} \dfrac{\sigma^2}{2} & -\dfrac{\sigma^2}{2} \\ -\dfrac{\sigma^2}{2} & \dfrac{\sigma^2}{2} \end{pmatrix}$$

General Constrained Fits

Kinematic fitting involving tracks is more complicated for several reasons:

1. There are generally several constraints
2. The constraints are generally non-linear
3. The initial tracks are defined by 5 parameters apiece, each governed by a 5×5 covariance matrix with off-diagonal terms.

Non-linearity is not a problem since we expand about a point close to the final answer anyway. We do this in the following way. Suppose that there are m variables α and r constraints $\mathbf{H}(\alpha) = \mathbf{0}$. The constraints can be expanded to first order about the point α_A, e.g.

$$\mathbf{0} = \frac{\partial \mathbf{H}(\alpha_A)}{\partial \alpha_i}(\alpha_i - \alpha_{iA}) + \mathbf{H}(\alpha_A)$$

$$\mathbf{0} = \mathbf{D}\Delta\alpha + \mathbf{d}$$

where $\Delta\alpha = \alpha - \alpha_A$ and $\mathbf{D}$ is a r row by m column matrix of partial derivatives:

$$\mathbf{D} = \begin{pmatrix} \frac{\partial H_1}{\partial \alpha_1} & \frac{\partial H_1}{\partial \alpha_2} & \frac{\partial H_1}{\partial \alpha_3} & \cdots & \frac{\partial H_1}{\partial \alpha_m} \\ \frac{\partial H_2}{\partial \alpha_1} & \frac{\partial H_2}{\partial \alpha_2} & \frac{\partial H_2}{\partial \alpha_3} & \cdots & \frac{\partial H_2}{\partial \alpha_m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial H_r}{\partial \alpha_1} & \frac{\partial H_r}{\partial \alpha_2} & \frac{\partial H_r}{\partial \alpha_3} & & \frac{\partial H_r}{\partial \alpha_m} \end{pmatrix} \qquad \mathbf{d} = \begin{pmatrix} H_1(\alpha_A) \\ H_2(\alpha_A) \\ \vdots \\ H_m(\alpha_A) \end{pmatrix}$$

For example, for our simple example of two variables satisfying the constraint $x_1 + x_2 - 6 = 0$ expanded about the point $\mathbf{x}_A = (3,3)$, we get $\mathbf{D} = (1 \quad 1)$ and $\mathbf{d} = 0$. The constraint equation becomes

$$\mathbf{D}\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix} = 0$$

where $\Delta x_1 = x_1 - 3$ and $\Delta x_2 = x_2 - 3$.

The complete χ^2 equation for a set of m parameters $\boldsymbol{\alpha}$ with initial covariance matrix $\mathbf{V}_{\alpha_0}$ and r constraint equations $\mathbf{H}(\alpha) = \mathbf{0}$ can be written compactly in matrix form:

$$\chi^2 = (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \mathbf{V}_{\alpha_0}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + 2\boldsymbol{\lambda}^T (\mathbf{D}\Delta\boldsymbol{\alpha} + \mathbf{d})$$

where $\boldsymbol{\alpha}_0$ are the unconstrained parameters, $\Delta\boldsymbol{\alpha} = \boldsymbol{\alpha} - \boldsymbol{\alpha}_A$ as before and

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_r \end{pmatrix}$$

The reason I use m parameters rather than 5 is that there are typically several tracks involved, i.e., for n tracks $m = 5n$.

The first term in the χ^2 expression is the general form for a set of m correlated variables. When the variables are uncorrelated, it collapses to the familiar expression

$$\sum_i \frac{\left(\alpha_i - \alpha_{i0}\right)^2}{\sigma_i^2}$$

The second term is the sum of the products of each of r Lagrange multipliers λ_i by its corresponding constraint.

A careful look at the equation shows it is identical to that given for the simple example:

$$\chi^2 = \left(x_1 - x_{10},\ x_2 - x_{20}\right)\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{pmatrix}\begin{pmatrix} x_1 - x_{10} \\ x_2 - x_{20} \end{pmatrix} + 2\lambda\left[1(x_1 - 3) + 1(x_2 - 3) + 0\right]$$

$$= \frac{\left(x_1 - x_{10}\right)^2}{\sigma^2} + \frac{\left(x_2 - x_{20}\right)^2}{\sigma^2} + 2\lambda(x_1 + x_2 - 6)$$

The solution is obtained by minimizing the χ^2. We set to zero the partial derivatives of the χ^2 wrt each of the α_i variables and λ_i Lagrange multipliers, giving a total of $m + r$ equations, enough to solve for all the α_i and λ_i unknowns.

The solution is demonstrated in my first fitting note, CBX 91–72. Without going into details, the answer is

$$
\begin{aligned}
\boldsymbol{\alpha} &= \boldsymbol{\alpha}_0 - \mathbf{V}_{\alpha_0}\mathbf{D}^T\boldsymbol{\lambda} \\
\boldsymbol{\lambda} &= \mathbf{V}_D(\mathbf{D}\Delta\boldsymbol{\alpha}_0 + \mathbf{d}) \\
\mathbf{V}_D &= \left(\mathbf{D}\mathbf{V}_{\alpha_0}\mathbf{D}^T\right)^{-1} \\
\mathbf{V}_\alpha &= \mathbf{V}_{\alpha_0} - \mathbf{V}_{\alpha_0}\mathbf{D}^T\mathbf{V}_D\mathbf{D}\mathbf{V}_{\alpha_0} \\
\chi^2 &= \boldsymbol{\lambda}^T\mathbf{V}_D^{-1}\boldsymbol{\lambda} = \boldsymbol{\lambda}^T(\mathbf{D}\Delta\boldsymbol{\alpha}_0 + \mathbf{d}) = (\mathbf{D}\Delta\boldsymbol{\alpha}_0 + \mathbf{d})^T\mathbf{V}_D(\mathbf{D}\Delta\boldsymbol{\alpha}_0 + \mathbf{d})
\end{aligned}
$$

The last equation shows that $\mathbf{V}_D$ is the covariance matrix for the Lagrange multipliers $\boldsymbol{\lambda}$ and that $\mathbf{V}_D^{-1}$ is the covariance matrix of the initial constraints $\mathbf{D}\Delta\boldsymbol{\alpha}_0 + \mathbf{d}$. Thus the number of standard deviations constraint i is from being satisfied by the unconstrained parameters is

$$
N_i = \frac{D_{ij}\Delta\alpha_{0j} + d_i}{\sqrt{V_{D\,ii}^{-1}}} = \frac{D_{ij}\left(\alpha_{0j} - \alpha_{Aj}\right) + d_i}{\sqrt{D_{ij}V_{\alpha_0\,jk}D_{ki}}}
$$

The following points should be noted about the solution.

1. The solution requires the inverse of only a single matrix, the $r \times r$ matrix $\mathbf{D}\mathbf{V}_{\alpha_0}\mathbf{D}^T$ which is used to obtain $\mathbf{V}_D$ (r is the number of constraints).
2. It can be shown that the new covariance matrix $\mathbf{V}_\alpha$ has diagonal elements smaller than the initial covariance matrix $\mathbf{V}_{\alpha_0}$. Thus the constraints are doing their job.
3. The χ^2 does *not* require the evaluation of $\mathbf{V}_{\alpha_0}^{-1}$, although the formal definition uses that matrix. This is a great simplification and permits the use track representations with non-invertible covariance matrices (such as that used in **KNLIB**).
4. The χ^2 can be written as a sum of r terms, one per constraint. It's then possible to look at each of these terms separately in order to get more discriminating power than what's available from the overall χ^2.