Appendix A

Constrained Fit Details

This appendix discusses the mathematical details of the constrained kinematic fit.

A.1 Basic Fitting Algorithm

The problem to be solved is this: given a set of measurements $\mathbf{x}^m$ and their error matrix $G_{ij}^{-1} = \langle \delta x_i^m \delta x_j^m \rangle$, what is the smallest change one can make to $\mathbf{x}$ to make them satisfy a given set of constraints $F_i(\mathbf{x}) = 0$?

That is, minimize the quantity $\chi^2 = (\mathbf{x} - \mathbf{x}^m)^T \mathbf{G}(\mathbf{x} - \mathbf{x}^m)$, subject to the constraints $\mathbf{F}(\mathbf{x}) = 0$. The algorithm, which is based on SQUAW kinematic fitting program [148, 149], is developed in this section.

A.1.1 Definitions

Divide the set of variables into 'well-measured' and 'poorly measured' sets. Let $\mathbf{x}$ and $\mathbf{y}$ be column vectors of the well-measured and poorly-measured variables, respectively. (Whole vectors and matrices are denoted by **boldface** symbols, and their elements by subscripted plain symbols.) Denote the measured values of the variables by $\mathbf{x}^m$, $\mathbf{y}^m$, the 'true' values of the variables by $\mathbf{x}^t$, $\mathbf{y}^t$, their values after iteration ν by $\mathbf{x}^\nu$, $\mathbf{y}^\nu$, and their values at the end of the fit by $\mathbf{x}^f$, $\mathbf{y}^f$. The error matrices are given by

$$\begin{aligned} G_{ij}^{-1} &\equiv \langle \delta x_i^m \delta x_j^m \rangle \\ Y_{ij}^{-1} &\equiv \langle \delta y_i^m \delta y_j^m \rangle, \end{aligned} \tag{A.1}$$

where it is tacitly assumed that $\langle \delta x_i^m \delta y_j^m \rangle = 0$. (Here, $\delta \mathbf{x}^m = \mathbf{x}^m - \mathbf{x}^t$.)

The fitting algorithm will be designed to reference only the arrays $\mathbf{G}^{-1}$ and $\mathbf{Y}$, not their inverses. This permits one to specify that a well-measured variable is constant by setting the appropriate diagonal element in $\mathbf{G}^{-1}$ to zero. One can also specify that a poorly measured variable is completely unknown by setting the appropriate element in $\mathbf{Y}$ to zero.

It is convenient to define the displacement vectors

$$\begin{aligned} \mathbf{c} &= \mathbf{x} - \mathbf{x}^m \\ \mathbf{d} &= \mathbf{y} - \mathbf{y}^m. \end{aligned} \tag{A.2}$$

The constraints which must be satisfied are given by the row vector

$$\mathbf{F}(\mathbf{x}^\nu, \mathbf{y}^\nu) = \mathbf{F}^\nu = 0 \tag{A.3}$$

with gradients

$$\begin{aligned} B_{\mathrm{x}i\lambda}(\mathbf{x}^\nu, \mathbf{y}^\nu) &= B^\nu_{\mathrm{x}i\lambda} = \frac{\partial F^\nu_\lambda}{\partial x_i} \\ B_{\mathrm{y}i\lambda}(\mathbf{x}^\nu, \mathbf{y}^\nu) &= B^\nu_{\mathrm{y}i\lambda} = \frac{\partial F^\nu_\lambda}{\partial y_i}. \end{aligned} \tag{A.4}$$

A column vector of Lagrange multipliers $\boldsymbol{\alpha}^\nu$ will also be needed.

The quantity which should be minimized subject to the constraints $\mathbf{F} = 0$ is

$$\chi^2 = \mathbf{c}^T\mathbf{G}\mathbf{c} + \mathbf{d}^T\mathbf{Y}\mathbf{d}. \tag{A.5}$$

A.1.2 Fit

In order to minimize the χ^2 subject to the constraints $\mathbf{F} = 0$, use the method of Lagrange multipliers and minimize the quantity

$$M = \mathbf{c}^T\mathbf{G}\mathbf{c} + \mathbf{d}^T\mathbf{Y}\mathbf{d} + 2\mathbf{F}\boldsymbol{\alpha}. \tag{A.6}$$

At the minimum, it must be the case that

$$0 = \frac{\partial M}{\partial \boldsymbol{\alpha}} = \mathbf{F} \tag{A.7}$$
$$0 = \frac{\partial M}{\partial \mathbf{x}} = \mathbf{G}\mathbf{c} + \mathbf{B}_{\mathrm{x}}\boldsymbol{\alpha} \tag{A.8}$$
$$0 = \frac{\partial M}{\partial \mathbf{y}} = \mathbf{Y}\mathbf{d} + \mathbf{B}_{\mathrm{y}}\boldsymbol{\alpha}. \tag{A.9}$$

Write (A.7) in terms of $\mathbf{x}^{\nu+1}$, $\mathbf{y}^{\nu+1}$ and linearize around $\mathbf{x}^{\nu}$, $\mathbf{y}^{\nu}$:

$$\begin{aligned} 0 &= \mathbf{F}^{\nu} + (\mathbf{x}^{\nu+1} - \mathbf{x}^{\nu})^T\mathbf{B}_{\mathrm{x}}^{\nu} + (\mathbf{y}^{\nu+1} - \mathbf{y}^{\nu})^T\mathbf{B}_{\mathrm{y}}^{\nu} \\ &= \mathbf{F}^{\nu} + (\mathbf{c}^{\nu+1} - \mathbf{c}^{\nu})^T\mathbf{B}_{\mathrm{x}}^{\nu} + (\mathbf{d}^{\nu+1} - \mathbf{d}^{\nu})^T\mathbf{B}_{\mathrm{y}}^{\nu}. \end{aligned} \tag{A.10}$$

Make the definitions

$$\mathbf{E} \equiv \mathbf{G}^{-1}\mathbf{B}_{\mathrm{x}} \tag{A.11}$$

$$\mathbf{H} \equiv \mathbf{E}^T\mathbf{B}_{\mathrm{x}} = \mathbf{B}_{\mathrm{x}}^T\mathbf{G}^{-1}\mathbf{B}_{\mathrm{x}} \tag{A.12}$$

$$\mathbf{r} \equiv \mathbf{c}^T\mathbf{B}_{\mathrm{x}} + \mathbf{d}^T\mathbf{B}_{\mathrm{y}} - \mathbf{F}. \tag{A.13}$$

Note that $\mathbf{H}$ is symmetric, $\mathbf{H}^T = \mathbf{H}$.

Then (A.10) yields

$$\mathbf{r}^\nu = \mathbf{c}^{\nu+1T}\mathbf{B}_\mathbf{x}^\nu + \mathbf{d}^{\nu+1T}\mathbf{B}_\mathbf{y}^\nu. \tag{A.14}$$

From (A.8) one gets

$$\mathbf{c}^{\nu+1} = -\mathbf{E}^\nu \boldsymbol{\alpha}^{\nu+1} \tag{A.15}$$

so

$$\mathbf{r}^\nu = -\boldsymbol{\alpha}^{\nu+1T}\mathbf{H}^\nu + \mathbf{d}^{\nu+1T}\mathbf{B}_\mathbf{y}^\nu \tag{A.16}$$

or

$$\mathbf{r}^{\nu T} = -\mathbf{H}^\nu \boldsymbol{\alpha}^{\nu+1} + \mathbf{B}_\mathbf{y}^{\nu T}\mathbf{d}^{\nu+1}. \tag{A.17}$$

Combining this with (A.9) yields

$$\begin{pmatrix} -\mathbf{H}^\nu & \mathbf{B}_\mathbf{y}^{\nu T} \\ \mathbf{B}_\mathbf{y}^\nu & \mathbf{Y} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}^{\nu+1} \\ \mathbf{d}^{\nu+1} \end{pmatrix} = \begin{pmatrix} \mathbf{r}^{\nu T} \\ 0 \end{pmatrix}. \tag{A.18}$$

Thus, for each iteration, given $\mathbf{x}^\nu$ and $\mathbf{y}^\nu$, evaluate $\mathbf{F}^\nu$, $\mathbf{B}_\mathbf{x}^\nu$, and $\mathbf{B}_\mathbf{y}^\nu$. Then calculate $\mathbf{E}^\nu$, $\mathbf{H}^\nu$, and $\mathbf{r}^\nu$, and find $\boldsymbol{\alpha}^{\nu+1}$ and $\mathbf{d}^{\nu+1}$ by solving (A.18). Finally, $\mathbf{c}^{\nu+1}$ follows from (A.15).

A.1.3 χ^2 Calculation

Define

$$\begin{pmatrix} \mathbf{W} & \mathbf{V}^T \\ \mathbf{V} & \mathbf{U} \end{pmatrix} \equiv \begin{pmatrix} -\mathbf{H} & \mathbf{B}_\mathbf{y}^T \\ \mathbf{B}_\mathbf{y} & \mathbf{Y} \end{pmatrix}^{-1}. \tag{A.19}$$

Note that since both $\mathbf{H}$ and $\mathbf{Y}$ are symmetric, $\mathbf{W}$ and $\mathbf{U}$ are also symmetric.

Multiplying out (A.19) yields

$$\begin{aligned} -\mathbf{HW} + \mathbf{B}_{\mathbf{y}}^T\mathbf{V} &= 1 \qquad & \mathbf{B}_{\mathrm{y}}\mathbf{W} + \mathbf{YV} &= 0 \\ -\mathbf{HV}^T + \mathbf{B}_{\mathbf{y}}^T\mathbf{U} &= 0 \qquad & \mathbf{B}_{\mathrm{y}}\mathbf{V}^T + \mathbf{YU} &= 1. \end{aligned} \tag{A.20}$$

From these relations, one can derive the identities

$$\begin{aligned} \mathbf{WHW} + \mathbf{V}^T\mathbf{YV} &= (\mathbf{V}^T\mathbf{B}_{\mathrm{y}} - 1)\mathbf{W} + \mathbf{V}^T(-\mathbf{B}_{\mathrm{y}}\mathbf{W}) \\ &= -\mathbf{W}, \end{aligned} \tag{A.21}$$

$$\begin{aligned} \mathbf{VHV}^T + \mathbf{UYU} &= (\mathbf{UB}_{\mathrm{y}})\mathbf{V}^T + \mathbf{U}(1 - \mathbf{B}_{\mathrm{y}}\mathbf{V}^T) \\ &= \mathbf{U}, \end{aligned} \tag{A.22}$$

and

$$\begin{aligned} \mathbf{V}^T\mathbf{YU} + \mathbf{WHV}^T &= \mathbf{V}^T(1 - \mathbf{B}_{\mathrm{y}}\mathbf{V}^T) + (\mathbf{V}^T\mathbf{B}_{\mathrm{y}} - 1)\mathbf{V}^T \\ &= 0. \end{aligned} \tag{A.23}$$

Now, combining (A.19) with (A.18) yields

$$\begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{d} \end{pmatrix}^{\nu+1} = \begin{pmatrix} \mathbf{W} & \mathbf{V}^T \\ \mathbf{V} & \mathbf{U} \end{pmatrix}^{\nu} \begin{pmatrix} \mathbf{r}^T \\ 0 \end{pmatrix}^{\nu}, \tag{A.24}$$

or, multiplying it out,

$$\boldsymbol{\alpha}^{\nu+1} = \mathbf{W}^{\nu}\mathbf{r}^{\nu T} \tag{A.25}$$

and

$$\mathbf{d}^{\nu+1} = \mathbf{V}^{\nu}\mathbf{r}^{\nu T}. \tag{A.26}$$

Also, from (A.15),

$$\mathbf{c}^{\nu+1} = -\mathbf{E}^{\nu}\mathbf{W}^{\nu}\mathbf{r}^{\nu T}. \tag{A.27}$$

Pulling these all together and plugging them into the χ^2 definition (A.5), the result is (suppressing ν superscripts)

$$\begin{aligned}
\chi^2 &= \mathbf{c}^T\mathbf{G}\mathbf{c} + \mathbf{d}^T\mathbf{Y}\mathbf{d} \\
&= (-\mathbf{r}\mathbf{W}\mathbf{E}^T)\mathbf{G}(-\mathbf{E}\mathbf{W}\mathbf{r}^T) + (\mathbf{r}\mathbf{V}^T)\mathbf{Y}(\mathbf{V}\mathbf{r}^T) \\
&= \mathbf{r}(\mathbf{W}\mathbf{H}\mathbf{W} + \mathbf{V}^T\mathbf{Y}\mathbf{V})\mathbf{r}^T \\
&= \; -\mathbf{r}\mathbf{W}\mathbf{r}^T \\
&= -\,\mathbf{r}\boldsymbol{\alpha}.
\end{aligned}$$

Thus,

$$\left(\chi^2\right)^{\nu+1} = -\mathbf{r}^{\nu}\boldsymbol{\alpha}^{\nu+1}. \tag{A.28}$$

A.1.4 Convergence

The fit is deemed to have converged if the following two criteria are satisfied for two iterations in a row:

$$\sum_{\lambda} |F_{\lambda}^{\nu+1}| < \epsilon_2 \tag{A.29}$$

$$|(\chi^2)^{\nu+1} - (\chi^2)^{\nu}| < \epsilon_3 \tag{A.30}$$

That is, when the constraints are satisfied and the χ^2 has stopped changing.

A.1.5 Error Propagation

The final error matrices are defined by

$$
\begin{aligned}
Q_{ij} &\equiv \langle \delta x_i^f \delta x_j^f \rangle \\
R_{ij} &\equiv \langle \delta y_i^f \delta y_j^f \rangle \\
S_{ij} &\equiv \langle \delta x_i^f \delta y_j^f \rangle .
\end{aligned} \tag{A.31}
$$

Noting that $\mathbf{B}_{\mathrm{x}}$, $\mathbf{B}_{\mathrm{y}}$, and $\mathbf{E}$ are independent of $\mathbf{x}^m$ and $\mathbf{y}^m$, differentiate (A.13) with respect to $\mathbf{x}^m$ and $\mathbf{y}^m$ to get

$$
\frac{\partial \mathbf{r}}{\partial \mathbf{x}^m} = -\mathbf{B}_{\mathrm{x}} \qquad \frac{\partial \mathbf{r}}{\partial \mathbf{y}^m} = -\mathbf{B}_{\mathrm{y}}. \tag{A.32}
$$

Thus, using (A.27), (A.26), and (A.20),

$$
\frac{\partial \mathbf{x}}{\partial \mathbf{x}^m} = \frac{\partial (\mathbf{x}^m + \mathbf{c})}{\partial \mathbf{x}^m} = 1 - \frac{\partial}{\partial \mathbf{x}^m}(\mathbf{EWr}^T) = 1 + \mathbf{B}_{\mathrm{x}}\mathbf{WE}^T \tag{A.33}
$$

$$
\frac{\partial \mathbf{x}}{\partial \mathbf{y}^m} = \frac{\partial (\mathbf{x}^m + \mathbf{c})}{\partial \mathbf{y}^m} = -\frac{\partial}{\partial \mathbf{y}^m}(\mathbf{EWr}^T) = \mathbf{B}_{\mathrm{y}}\mathbf{WE}^T = -\mathbf{YVE}^T \tag{A.34}
$$

$$
\frac{\partial \mathbf{y}}{\partial \mathbf{x}^m} = \frac{\partial (\mathbf{y}^m + \mathbf{d})}{\partial \mathbf{x}^m} = \frac{\partial}{\partial \mathbf{x}^m}(\mathbf{Vr}^T) = -\mathbf{B}_{\mathrm{x}}\mathbf{V}^T \tag{A.35}
$$

$$
\frac{\partial \mathbf{y}}{\partial \mathbf{y}^m} = \frac{\partial (\mathbf{y}^m + \mathbf{d})}{\partial \mathbf{y}^m} = 1 + \frac{\partial}{\partial \mathbf{y}^m}(\mathbf{Vr}^T) = 1 - \mathbf{B}_{\mathrm{y}}\mathbf{V}^T = \mathbf{YU}. \tag{A.36}
$$

(Note that since $\frac{\partial F_\lambda}{\partial x_i} = B_{i\lambda}$, the form $\mathbf{M} = \frac{\partial}{\partial \mathrm{x}}(\mathbf{AF})$ can be written as $M_{ij} = \frac{\partial}{\partial x_i}(A_{jk}F_k) = A_{jk}B_{ik}$, or $\mathbf{M} = \mathbf{BA}^T$.)

So, to evaluate $\mathbf{Q} = \langle \delta \mathbf{x} \delta \mathbf{x}^T \rangle$, note that

$$
\delta \mathbf{x} = \left(\frac{\partial \mathbf{x}}{\partial \mathbf{x}^m} \right)^T \delta \mathbf{x}^m + \left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}^m} \right)^T \delta \mathbf{y}^m, \tag{A.37}
$$

and similarly for $\mathbf{y}$. So, using (A.21), one gets

$$\begin{aligned}
\mathbf{Q} &= \left(\frac{\partial \mathbf{x}}{\partial \mathbf{x}^m}\right)^T \langle \delta\mathbf{x}^m \delta\mathbf{x}^{mT} \rangle \left(\frac{\partial \mathbf{x}}{\partial \mathbf{x}^m}\right) + \left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}^m}\right)^T \langle \delta\mathbf{y}^m \delta\mathbf{y}^{mT} \rangle \left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}^m}\right) \\
&= (1 + \mathbf{E}\mathbf{W}\mathbf{B}_{\mathrm{x}}^T)\mathbf{G}^{-1}(1 + \mathbf{B}_{\mathrm{x}}\mathbf{W}\mathbf{E}^T) + (\mathbf{E}\mathbf{V}^T\mathbf{Y})\mathbf{Y}^{-1}(\mathbf{Y}\mathbf{V}\mathbf{E}^T) \\
&= \mathbf{G}^{-1} + \mathbf{E}(2\mathbf{W} + \mathbf{W}\mathbf{H}\mathbf{W} + \mathbf{V}^T\mathbf{Y}\mathbf{V})\mathbf{E}^T \\
&= \mathbf{G}^{-1} + \mathbf{E}\mathbf{W}\mathbf{E}^T.
\end{aligned} \tag{A.38}$$

Also, using (A.23),

$$\begin{aligned}
\mathbf{S} &= \langle \delta\mathbf{x} \delta\mathbf{y} \rangle \\
&= \left(\frac{\partial \mathbf{x}}{\partial \mathbf{x}^m}\right)^T \langle \delta\mathbf{x}^m \delta\mathbf{x}^{mT} \rangle \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}^m}\right) + \left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}^m}\right)^T \langle \delta\mathbf{y}^m \delta\mathbf{y}^{mT} \rangle \left(\frac{\partial \mathbf{y}}{\partial \mathbf{y}^m}\right) \\
&= (1 + \mathbf{E}\mathbf{W}\mathbf{B}_{\mathrm{x}}^T)\mathbf{G}^{-1}(-\mathbf{B}_{\mathrm{x}}\mathbf{V}^T) + (-\mathbf{E}\mathbf{V}^T\mathbf{Y})\mathbf{Y}^{-1}(\mathbf{Y}\mathbf{U}) \\
&= -\mathbf{E}(\mathbf{V}^T + \mathbf{W}\mathbf{H}\mathbf{V}^T + \mathbf{V}^T\mathbf{Y}\mathbf{U}) \\
&= -\mathbf{E}\mathbf{V}^T.
\end{aligned} \tag{A.39}$$

Finally, using (A.22),

$$\begin{aligned}
\mathbf{R} &= \langle \delta\mathbf{y} \delta\mathbf{y} \rangle \\
&= \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}^m}\right)^T \langle \delta\mathbf{x}^m \delta\mathbf{x}^{mT} \rangle \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}^m}\right) + \left(\frac{\partial \mathbf{y}}{\partial \mathbf{y}^m}\right)^T \langle \delta\mathbf{y}^m \delta\mathbf{y}^{mT} \rangle \left(\frac{\partial \mathbf{y}}{\partial \mathbf{y}^m}\right) \\
&= (\mathbf{V}\mathbf{B}_{\mathrm{x}}^T)\mathbf{G}^{-1}(\mathbf{B}_{\mathrm{x}}\mathbf{V}^T) + (\mathbf{U}\mathbf{Y})\mathbf{Y}^{-1}(\mathbf{Y}\mathbf{U}) \\
&= \mathbf{V}\mathbf{H}\mathbf{V}^T + \mathbf{U}\mathbf{Y}\mathbf{U} \\
&= \mathbf{U}.
\end{aligned} \tag{A.40}$$

A.1.6 Pull Functions

Define the pull functions

$$\xi_i = \frac{c_i}{\sqrt{\langle c_i^2 \rangle}}, \qquad \phi_i = \frac{d_i}{\sqrt{\langle d_i^2 \rangle}}. \tag{A.41}$$

These quantities should be normally distributed with means of zero and widths of one.

For the moment, ignore the y's. Then, since $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^t$ and $\delta\mathbf{x}^m = \mathbf{x}^m - \mathbf{x}^t$, it follows that

$$\delta\mathbf{x} = \delta\mathbf{x}^m + \mathbf{c}. \tag{A.42}$$

Thus,

$$\langle \delta\mathbf{x}^m \delta\mathbf{x}^{mT} \rangle = \langle \delta\mathbf{x}\delta\mathbf{x}^T \rangle + \langle \mathbf{c}\mathbf{c}^T \rangle - 2\langle \mathbf{c}\delta\mathbf{x}^T \rangle, \tag{A.43}$$

or

$$\langle \mathbf{c}\mathbf{c}^T \rangle = \mathbf{G}^{-1} - \mathbf{Q} + 2\langle \mathbf{c}\delta\mathbf{x}^T \rangle. \tag{A.44}$$

But,

$$\langle \mathbf{c}\delta\mathbf{x}^T \rangle = \langle \delta\mathbf{x}\delta\mathbf{x}^T \rangle - \langle \delta\mathbf{x}^m \delta\mathbf{x}^T \rangle = \mathbf{Q} - \mathbf{G}^{-1}\left(\frac{\partial\mathbf{x}}{\partial\mathbf{x}^m}\right). \tag{A.45}$$

Distinguishing again between well-measured and poorly measured variables,

$$\begin{aligned}
\langle \mathbf{c}\delta\mathbf{x}^T \rangle \rightarrow & \begin{pmatrix} \langle \mathbf{c}\delta\mathbf{x}^T \rangle & \langle \mathbf{d}\delta\mathbf{x}^T \rangle \\ \langle \mathbf{c}\delta\mathbf{y}^T \rangle & \langle \mathbf{d}\delta\mathbf{y}^T \rangle \end{pmatrix} \\
= & \begin{pmatrix} \mathbf{Q} - \mathbf{G}^{-1}\left(\frac{\partial\mathbf{x}}{\partial\mathbf{x}^m}\right) & \mathbf{S}^T - \mathbf{Y}^{-1}\left(\frac{\partial\mathbf{x}}{\partial\mathbf{y}^m}\right) \\ \mathbf{S} - \mathbf{G}^{-1}\left(\frac{\partial\mathbf{y}}{\partial\mathbf{x}^m}\right) & \mathbf{R} - \mathbf{Y}^{-1}\left(\frac{\partial\mathbf{y}}{\partial\mathbf{y}^m}\right) \end{pmatrix} \\
= & \begin{pmatrix} \mathbf{Q} - (\mathbf{G}^{-1} + \mathbf{E}\mathbf{W}\mathbf{E}^T) & \mathbf{S}^T + \mathbf{V}\mathbf{E}^T \\ \mathbf{S} + \mathbf{E}\mathbf{V}^T & \mathbf{R} - \mathbf{U} \end{pmatrix} \\
= & \; 0.
\end{aligned} \tag{A.46}$$

Thus,

$$\langle \mathbf{c}\mathbf{c}^T \rangle = \mathbf{G}^{-1} - \mathbf{Q}, \tag{A.47}$$

so the pull is

$$\xi_i = \frac{c_i}{\sqrt{(G^{-1})_{ii} - Q_{ii}}}. \tag{A.48}$$

Similarly, for poorly measured variables (if the off-diagonal elements are small),

$$\frac{1}{\langle \mathbf{d}\mathbf{d}^T \rangle_{ii}} = \frac{1}{(Y^{-1})_{ii} - R_{ii}} \approx \frac{Y_{ii}}{1 - Y_{ii} R_{ii}}, \tag{A.49}$$

so

$$\phi_i = \frac{d_i}{\sqrt{\langle \mathbf{d}\mathbf{d}^T \rangle_{ii}}} = d_i \sqrt{\frac{Y_{ii}}{1 - Y_{ii} R_{ii}}}. \tag{A.50}$$

A.1.7 Cut Steps

The constraint sum $\sum_\lambda |F_\lambda|$ should get smaller with each step. If it does not, then see if the χ^2 is changing. If it is not, then the fit is near a solution, and a directed step should be tried (see Section A.1.8). Otherwise, the step is 'cut' by a fraction ϵ and retried. That is, instead of stepping to $\mathbf{x}^{\nu+1}$, step to $\mathbf{x}'$, defined by

$$\mathbf{x}' = \epsilon(\mathbf{x}^{\nu+1} - \mathbf{x}^\nu) + \mathbf{x}^\nu, \tag{A.51}$$

or

$$\mathbf{c}' = \epsilon \mathbf{c}^{\nu+1} + (1 - \epsilon)\mathbf{c}^\nu. \tag{A.52}$$

Also,

$$\mathbf{d}' = \epsilon \mathbf{d}^{\nu+1} + (1 - \epsilon)\mathbf{d}^\nu. \tag{A.53}$$

Now to find the χ^2. At $\mathbf{x}'$, $\mathbf{y}'$, the χ^2 is

$$
\begin{aligned}
\left(\chi^2\right)' &= \mathbf{c}'^T\mathbf{G}\mathbf{c}' + \mathbf{d}'^T\mathbf{Y}\mathbf{d}' \\
&= (\epsilon\mathbf{c}^{\nu+1T} + (1-\epsilon)\mathbf{c}^{\nu T})\mathbf{G}(\epsilon\mathbf{c}^{\nu+1} + (1-\epsilon)\mathbf{c}^{\nu}) \\
&\quad + (\epsilon\mathbf{d}^{\nu+1T} + (1-\epsilon)\mathbf{d}^{\nu T})\mathbf{Y}(\epsilon\mathbf{d}^{\nu+1} + (1-\epsilon)\mathbf{d}^{\nu}) \\
&= \epsilon^2\mathbf{c}^{\nu+1T}\mathbf{G}\mathbf{c}^{\nu+1} + (1-\epsilon)^2\mathbf{c}^{\nu T}\mathbf{G}\mathbf{c}^{\nu} + 2\epsilon(1-\epsilon)\mathbf{c}^{\nu+1T}\mathbf{G}\mathbf{c}^{\nu} \\
&\quad + \epsilon^2\mathbf{d}^{\nu+1T}\mathbf{Y}\mathbf{d}^{\nu+1} + (1-\epsilon)^2\mathbf{d}^{\nu T}\mathbf{Y}\mathbf{d}^{\nu} + 2\epsilon(1-\epsilon)\mathbf{d}^{\nu+1T}\mathbf{Y}\mathbf{d}^{\nu}.
\end{aligned}
\tag{A.54}
$$

Define

$$
\psi \equiv \mathbf{c}^{\nu+1T}\mathbf{G}\mathbf{c}^{\nu} + \mathbf{d}^{\nu+1T}\mathbf{Y}\mathbf{d}^{\nu}. \tag{A.55}
$$

Then,

$$
\left(\chi^2\right)' = \epsilon^2\left(\chi^2\right)^{\nu+1} + (1-\epsilon)^2\left(\chi^2\right)^{\nu} + 2\epsilon(1-\epsilon)\psi. \tag{A.56}
$$

Now, using (A.20), (A.27), (A.26), and (A.25),

$$
\begin{aligned}
\psi &= \mathbf{c}^{\nu+1T}\mathbf{G}\mathbf{c}^{\nu} + \mathbf{d}^{\nu+1T}\mathbf{Y}\mathbf{d}^{\nu} \\
&= (-\mathbf{r}\mathbf{W}\mathbf{E}^T)^{\nu}\mathbf{G}\mathbf{c}^{\nu} + (\mathbf{r}\mathbf{V}^T)^{\nu}\mathbf{Y}\mathbf{d}^{\nu} \\
&= (-\mathbf{r}\mathbf{W}\mathbf{B}_{\mathbf{x}}^T)^{\nu}\mathbf{c}^{\nu} + (-\mathbf{r}\mathbf{W}\mathbf{B}_{\mathbf{y}}^T)^{\nu}\mathbf{d}^{\nu} \\
&= -\mathbf{r}^{\nu}\mathbf{W}^{\nu}(\mathbf{B}_{\mathbf{x}}^T\mathbf{c} + \mathbf{B}_{\mathbf{y}}^T\mathbf{d})^{\nu} \\
&= -(\mathbf{r} + \mathbf{F})^{\nu}\boldsymbol{\alpha}^{\nu+1}.
\end{aligned}
\tag{A.57}
$$

Since the step may have to be cut multiple times, note that

$$
\begin{aligned}
\psi' &= \mathbf{c}'^T\mathbf{G}\mathbf{c}^{\nu} + \mathbf{d}'^T\mathbf{Y}\mathbf{d}^{\nu} \\
&= (\epsilon\mathbf{c}^{\nu+1T} + (1-\epsilon)\mathbf{c}^{\nu T})\mathbf{G}\mathbf{c}^{\nu} + (\epsilon\mathbf{d}^{\nu+1T} + (1-\epsilon)\mathbf{d}^{\nu T})\mathbf{Y}\mathbf{d}^{\nu} \\
&= \epsilon\psi + (1-\epsilon)\left(\chi^2\right)^{\nu}.
\end{aligned}
\tag{A.58}
$$

A.1.8 Directed Steps

For a directed step, try to take the smallest step which satisfies the constraints; i.e., minimize

$$N = \boldsymbol{\gamma}^T \mathbf{G} \boldsymbol{\gamma} + \boldsymbol{\delta}^T \mathbf{Y} \boldsymbol{\delta} + 2\mathbf{F}\boldsymbol{\beta} \tag{A.59}$$

by stepping to $\mathbf{x}'$, $\mathbf{y}'$, where

$$\boldsymbol{\gamma} = \mathbf{x}' - \mathbf{x}^\nu, \qquad \boldsymbol{\delta} = \mathbf{y}' - \mathbf{y}^\nu. \tag{A.60}$$

After linearizing, this gives

$$0 = \mathbf{F}^\nu + \boldsymbol{\gamma}^T \mathbf{B}_\mathrm{x}^\nu + \boldsymbol{\delta}^T \mathbf{B}_\mathrm{y}^\nu \tag{A.61}$$

$$0 = \mathbf{G}\boldsymbol{\gamma} + \mathbf{B}_\mathrm{x}^\nu \boldsymbol{\beta} \quad \longrightarrow \quad \boldsymbol{\gamma} = -\mathbf{G}^{-1}\mathbf{B}_\mathrm{x}^\nu \boldsymbol{\beta} \tag{A.62}$$

$$0 = \mathbf{Y}\boldsymbol{\delta} + \mathbf{B}_\mathrm{y}^\nu \boldsymbol{\beta} \quad \longrightarrow \quad \boldsymbol{\delta} = -\mathbf{Y}^{-1}\mathbf{B}_\mathrm{y}^\nu \boldsymbol{\beta}. \tag{A.63}$$

Putting together (A.61) and (A.62) yields

$$-\mathbf{H}^\nu \boldsymbol{\beta} + \mathbf{B}_\mathrm{y}^{\nu T} \boldsymbol{\delta} = -\mathbf{F}^{\nu T}. \tag{A.64}$$

Combining this with (A.63),

$$\begin{pmatrix} -\mathbf{H} & \mathbf{B}_\mathrm{y}^T \\ \mathbf{B}_\mathrm{y} & \mathbf{Y} \end{pmatrix}^\nu \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{pmatrix} = \begin{pmatrix} -\mathbf{F}^T \\ 0 \end{pmatrix}^\nu. \tag{A.65}$$

This can be solved to find $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, and thence $\boldsymbol{\gamma}$.

Using (A.19), the result is

$$\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{pmatrix} = \begin{pmatrix} \mathbf{W} & \mathbf{V}^T \\ \mathbf{V} & \mathbf{U} \end{pmatrix}^\nu \begin{pmatrix} -\mathbf{F}^T \\ 0 \end{pmatrix}^\nu, \tag{A.66}$$

or

$$\boldsymbol{\beta} = -\mathbf{W}^{\nu}\mathbf{F}^{\nu T} \qquad \boldsymbol{\delta} = -\mathbf{V}^{\nu}\mathbf{F}^{\nu T}. \tag{A.67}$$

Then calculate the resulting χ^2 (using (A.20)):

$$\begin{aligned}
\left(\chi^2\right)' &= \mathbf{c}'^T\mathbf{G}\mathbf{c}' + \mathbf{d}'^T\mathbf{Y}\mathbf{d}' \\
&= (\boldsymbol{\gamma} + \mathbf{c}^{\nu})^T\mathbf{G}(\boldsymbol{\gamma} + \mathbf{c}^{\nu}) + (\boldsymbol{\delta} + \mathbf{d}^{\nu})^T\mathbf{Y}(\boldsymbol{\delta} + \mathbf{d}^{\nu}) \\
&= \boldsymbol{\gamma}^T\mathbf{G}\boldsymbol{\gamma} + \boldsymbol{\delta}^T\mathbf{Y}\boldsymbol{\delta} + 2\boldsymbol{\gamma}^T\mathbf{G}\mathbf{c}^{\nu} + 2\boldsymbol{\delta}^T\mathbf{Y}\mathbf{d}^{\nu} + \mathbf{c}^{\nu T}\mathbf{G}\mathbf{c}^{\nu} + \mathbf{d}^{\nu T}\mathbf{Y}\mathbf{d}^{\nu} \\
&= \boldsymbol{\beta}^T\mathbf{B}_{\mathrm{x}}^{\nu T}\mathbf{G}^{-1}\mathbf{G}\mathbf{G}^{-1}\mathbf{B}_{\mathrm{x}}^{\nu}\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{B}_{\mathrm{y}}^{\nu T}\mathbf{Y}^{-1}\mathbf{Y}\boldsymbol{\delta} \\
&\quad - 2\boldsymbol{\beta}^T\mathbf{B}_{\mathrm{x}}^{\nu T}\mathbf{G}^{-1}\mathbf{G}\mathbf{c}^{\nu} - 2\boldsymbol{\beta}^T\mathbf{B}_{\mathrm{y}}^{\nu T}\mathbf{Y}^{-1}\mathbf{Y}\mathbf{d}^{\nu} + \left(\chi^2\right)^{\nu} \qquad \text{(A.68)} \\
&= \boldsymbol{\beta}^T\mathbf{H}^{\nu}\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{B}_{\mathrm{y}}^{\nu T}\boldsymbol{\delta} - 2\boldsymbol{\beta}^T(\mathbf{B}_{\mathrm{x}}^{\nu T}\mathbf{c}^{\nu} + \mathbf{B}_{\mathrm{y}}^{\nu T}\mathbf{d}^{\nu}) + \left(\chi^2\right)^{\nu} \\
&= \boldsymbol{\beta}^T(-\mathbf{H}^{\nu}\mathbf{W}^{\nu}\mathbf{F}^{\nu T} + \mathbf{B}_{\mathrm{y}}^{\nu T}\mathbf{V}^{\nu}\mathbf{F}^{\nu T}) - 2\boldsymbol{\beta}^T(\mathbf{r} + \mathbf{F})^{\nu T} + \left(\chi^2\right)^{\nu} \\
&= \mathbf{F}^{\nu}\boldsymbol{\beta} - 2(\mathbf{r} + \mathbf{F})^{\nu}\boldsymbol{\beta} + \left(\chi^2\right)^{\nu} \\
&= \left(\chi^2\right)^{\nu} - 2\mathbf{r}^{\nu}\boldsymbol{\beta} - \mathbf{F}^{\nu}\boldsymbol{\beta}.
\end{aligned}$$

A.1.9 Summary of Fitting Algorithm

1. The inputs are the vectors of measured values $\mathbf{x}^m$, $\mathbf{y}^m$, the initial starting point for the fit $\mathbf{x}$, $\mathbf{y}$, and the error matrices $\mathbf{G}^{-1}$ and $\mathbf{Y}$. There's also a constraint function `constraint_fcn`, which takes a point $\mathbf{x}$, $\mathbf{y}$ and returns the values of the constraints $\mathbf{F}$ and their gradients $\mathbf{B}_{\mathrm{x}}$ and $\mathbf{B}_{\mathrm{y}}$.

2. Call `constraint_fcn` for the initial values of $\mathbf{x}$ and $\mathbf{y}$ to get the initial values for $\mathbf{F}$, $\mathbf{B}_{\mathrm{x}}$, and $\mathbf{B}_{\mathrm{y}}$. If the initial starting point is rejected by `constraint_fcn` as invalid, return an error status.

3. Initialize $\mathbf{c} \leftarrow \mathbf{x} - \mathbf{x}^m$ and $\mathbf{d} \leftarrow \mathbf{y} - \mathbf{y}^m$.

4. For each iteration:

5. Compute $\mathbf{E} \leftarrow \mathbf{G}^{-1}\mathbf{B}_{\mathrm{x}}$, $\mathbf{H} \leftarrow \mathbf{E}^T\mathbf{B}_{\mathrm{x}}$, and $\mathbf{r} \leftarrow \mathbf{c}^T\mathbf{B}_{\mathrm{x}} + \mathbf{d}^T\mathbf{B}_{\mathrm{y}} - \mathbf{F}$.

6. Solve the set of equations $\begin{pmatrix} -\mathbf{H} & \mathbf{B}_{\mathrm{y}}^T \\ \mathbf{B}_{\mathrm{y}} & \mathbf{Y} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \mathbf{d}_1 \end{pmatrix} = \begin{pmatrix} \mathbf{r}^T \\ 0 \end{pmatrix}$ for $\boldsymbol{\alpha}$ and $\mathbf{d}_1$. Also remember the matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$, defined by the equations $\begin{pmatrix} \mathbf{W} & \mathbf{V}^T \\ \mathbf{V} & \mathbf{U} \end{pmatrix} = \begin{pmatrix} -\mathbf{H} & \mathbf{B}_{\mathrm{y}}^T \\ \mathbf{B}_{\mathrm{y}} & \mathbf{Y} \end{pmatrix}^{-1}$.

7. Compute the new value of $\mathbf{c}$ and χ^2 by $\mathbf{c}_1 \leftarrow -\mathbf{E}\boldsymbol{\alpha}$ and $\chi^2 \leftarrow -\mathbf{r}\boldsymbol{\alpha}$. In case the step needs to be cut, compute the value of ψ by $\psi \leftarrow -(\mathbf{r}+\mathbf{F})\boldsymbol{\alpha}$.

8. Compute the trial destination of this step, $\mathbf{x}$, $\mathbf{y}$, by $\mathbf{x} \leftarrow \mathbf{c}_1 + \mathbf{x}^m$ and $\mathbf{y} \leftarrow \mathbf{d}_1 + \mathbf{y}^m$.

9. Initialize the total cut size for this step, `this_step_cutsize` $\leftarrow 1$.

10. Call `constraint_fcn` at $\mathbf{x}$, $\mathbf{y}$. If `constraint_fcn` rejects the point, the step must be cut, as described below. Compute the sum of the absolute values of the constraint functions, `constraint_sum`. If it has decreased since the last step, or if it is below the convergence threshold `constraint_sum_eps`, the step is accepted. The first step is accepted no matter what the constraint functions are doing. Otherwise, the constraints have increased, and the step should be cut:

    (a) If this is the first attempt to cut this step, test to see if the χ^2 is stationary. If $|$`chisq` $-$ `chisq_last`$|$ $<$ `chisq_diff_eps`, try a directed step:

i. Solve the set of equations $\begin{pmatrix} -\mathbf{H} & \mathbf{B}_y^T \\ \mathbf{B}_y & \mathbf{Y} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{pmatrix} = \begin{pmatrix} -\mathbf{F}^T \\ 0 \end{pmatrix}$ for $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, using the values of $\mathbf{B}_y$ and $\mathbf{F}$ from the beginning of the step. Also remember $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$, as before.

ii. Compute $\boldsymbol{\gamma} = -\mathbf{E}\boldsymbol{\beta}$.

iii. Compute the destination of the directed step by $\mathbf{x} \leftarrow \mathbf{c} + \mathbf{x}^m + \boldsymbol{\gamma}$ and $\mathbf{y} \leftarrow \mathbf{d} + \mathbf{y}^m + \boldsymbol{\delta}$.

iv. Call `constraint_fcn`. If $\mathbf{x}$, $\mathbf{y}$ is an acceptable point and the value of `constraint_sum` is now less than the value on the previous step, then accept the directed step and make the assignments

- `chisq` $\leftarrow$ `chisq_last` $- (\mathbf{F} + 2\mathbf{r})\boldsymbol{\beta}$
- $\mathbf{c}_1 \leftarrow \mathbf{x} - \mathbf{x}^m$
- $\mathbf{d}_1 \leftarrow \mathbf{y} - \mathbf{y}^m$

Otherwise, continue with the attempt to cut the step.

(b) If too many attempts have already been made to cut this step, give up and return an error.

(c) Set up the size by which to cut this step, ϵ. Normally, this is set to a constant `cutsize`. However, it is fairly common to have a number of steps in a row, all of which get cut to about the same size. To speed up this case, remember the total cut size from the last step. If this is the first cut for this step, and the last step was also cut, set ϵ to twice the final cut size from the last step, provided that this value is less than `cutsize`.

(d) Compute the the new total cut size for this step by using the assignment `this_step_cutsize` $\leftarrow \epsilon \cdot$ `this_step_cutsize`. If the total cut size falls below a threshold, `this_step_cutsize` $<$ `min_tot_cutsize`, give up and return an error.

(e) Cut the step, by $\mathbf{c}_1 \leftarrow \epsilon \cdot \mathbf{c}_1 + (1-\epsilon)\mathbf{c}$ and $\mathbf{d}_1 \leftarrow \epsilon \cdot \mathbf{d}_1 + (1-\epsilon)\mathbf{d}$.

(f) Calculate the new χ^2 and ψ from (A.56) and (A.58), and the new $\mathbf{x}$ and $\mathbf{y}$ from $\mathbf{x} \leftarrow \mathbf{c}_1 + \mathbf{x}^m$ and $\mathbf{y} \leftarrow \mathbf{d}_1 + \mathbf{y}^m$.

(g) Return to step 10 to test the cut step.

11. At this point, there's an acceptable step. Shuffle the variables around to prepare for the next step:

- `z2` $\leftarrow$ |`chisq` $-$ `chisq_last`|
- `chisq_last` $\leftarrow$ `chisq`
- `constraint_sum_last` $\leftarrow$ `constraint_sum`
- `last_step_cutsize` $\leftarrow$ `this_step_cutsize`
- $\mathbf{c} \leftarrow \mathbf{c}_1$
- $\mathbf{d} \leftarrow \mathbf{d}_1$

12. Test for convergence by requiring that the system satisfy `z2` $<$ `chisq_diff_eps` and `constraint_sum` $<$ `constraint_sum_eps`

for two iterations in a row. If these conditions are satisfied, exit the fit loop and go to step 15.

13. Count the total number of steps taken. If it exceeds `maxit`, give up and return an error.

14. Return to step 4 for the next iteration.

15. This fit has succeeded. Calculate the error matrices $\mathbf{Q}$ and $\mathbf{R}$ from (A.38) and (A.40), and the vectors of pull functions $\boldsymbol{\xi}$ and $\boldsymbol{\phi}$ from (A.48) and (A.50).

A.2 Evaluation of Constraint Functions

Now turn to the problem of evaluating the kinematic constraints and their gradients. In order to do this, the variables used to describe an event must be specified, as well as the forms of the constraints themselves.

An event consists of a collection of N final state particles with four-momenta $\mathbf{p}_i$. Each particle is described by three variables: the absolute value of its three-momentum P_i and its direction in $\eta - \phi$ space η_i and ϕ_i. The masses of the particles are taken to be constants, and denoted m_i.

There can be at most one neutrino in the final state, with four-momentum $\mathbf{p}_\nu$. The neutrino is also described by three variables: the z-component of its momentum p_ν^z, and ${k_T}^x$ and ${k_T}^y$, which are defined as

$$
\begin{aligned}
{k_T}^x &= p_\nu^x + \sum_i p_i^x \\
{k_T}^y &= p_\nu^y + \sum_i p_i^y.
\end{aligned}
\tag{A.69}
$$

The neutrino, of course, is considered massless.

Now for the constraints. Two general forms of mass constraint are implemented. The first asserts that some collection of particles has a fixed invariant mass (such as in the W mass constraints). The second asserts that two collections of particles have the same invariant mass (as in the top mass equality constraint). The first of these can be written

$$F = \left(\sum_{i \in C} \mathbf{p}_i \right)^2 - M^2, \tag{A.70}$$

and the second

$$F = \left(\sum_{i \in C_1} \mathbf{p}_i \right)^2 - \left(\sum_{i \in C_2} \mathbf{p}_i \right)^2 . \tag{A.71}$$

The sums in parentheses can be expanded out as

$$\left(\sum_{i \in C} \mathbf{p}_i \right)^2 = \sum_{i \in C} m_i^2 + 2 \sum_{\substack{i,j \in C \\ i<j}} \mathbf{p}_i \cdot \mathbf{p}_j . \tag{A.72}$$

Note that since the masses m_i are treated as constant, their gradients are zero. Thus, the problem of calculating the gradients of the constraints reduces to calculating the gradients of the dot product of two four-vectors.

If there is no final-state neutrino in the event, then there are the two additional constraints

$$0 = {k_T}^x = \sum_i p_i^x$$
$$0 = {k_T}^y = \sum_i p_i^y. \tag{A.73}$$

These constraints will be dealt with separately below.

Unfortunately, the variables used to describe an event are not very convenient for calculating the gradients. So the strategy used requires several steps: first, the gradients are calculated with respect to a more convenient set of variables. Then, the gradients are converted to the variables actually desired by a series of Jacobian transformations. For this initial set of variables, the polar angle θ will be used instead of the pseudorapidity η. In addition, the neutrino will be treated just like any other final-state particle, and parameterized using P, θ, and η.

So, the problem is to evaluate the gradients of $\mathbf{p}_1 \cdot \mathbf{p}_2$ with respect to the variables P_1, P_2, θ_1, θ_2, ϕ_1, and ϕ_2. Since, by definition,

$$E = \sqrt{P^2 + m^2}$$
$$p^x = P \sin\theta \cos\phi$$
$$p^y = P \sin\theta \sin\phi \tag{A.74}$$
$$p^z = P \cos\theta,$$

the dot product

$$\mathbf{p}_1 \cdot \mathbf{p}_2 = E_1 E_2 - p_1^x p_2^x - p_1^y p_2^y - p_1^z p_2^z \tag{A.75}$$

can be written as

$$\begin{aligned}\mathbf{p}_1 \cdot \mathbf{p}_2 = &\sqrt{P_1^2 + m_1^2}\sqrt{P_2^2 + m_2^2} - \\ &P_1 P_2 \left(\sin\theta_1 \sin\theta_2 \cos(\phi_1 - \phi_2) + \cos\theta_1 \cos\theta_2\right).\end{aligned} \tag{A.76}$$

The partial derivative with respect to P_1 is thus

$$\begin{aligned}\frac{\partial\, \mathbf{p}_1 \cdot \mathbf{p}_2}{\partial P_1} &= \frac{P_1}{\sqrt{P_1^2 + m_1^2}}\sqrt{P_2^2 + m_2^2} - P_2(\ldots) \\ &= \frac{\mathbf{p}_1 \cdot \mathbf{p}_2}{P_1} - \frac{E_1 E_2}{P_1} + \frac{P_1}{E_1} E_2 \\ &= \frac{\mathbf{p}_1 \cdot \mathbf{p}_2}{P_1} + \frac{P_1^2 - E_1^2}{P_1 E_1} E_2 \\ &= \frac{1}{P_1}\left(\mathbf{p}_1 \cdot \mathbf{p}_2 - m_1^2 \frac{E_2}{E_1}\right).\end{aligned} \tag{A.77}$$

Similarly for P_2,

$$\frac{\partial\, \mathbf{p}_1 \cdot \mathbf{p}_2}{\partial P_2} = \frac{1}{P_2}\left(\mathbf{p}_1 \cdot \mathbf{p}_2 - m_2^2 \frac{E_1}{E_2}\right). \tag{A.78}$$

For the angular gradients, the result is

$$\frac{\partial\, \mathbf{p}_1 \cdot \mathbf{p}_2}{\partial \theta_1} = -P_1 P_2 \left(\cos\theta_1 \sin\theta_2 \cos(\phi_1 - \phi_2) - \sin\theta_1 \cos\theta_2\right). \tag{A.79}$$

Noting that

$$\begin{aligned}\sin\phi &= p^y / p_T \\ \cos\phi &= p^x / p_T \\ \sin\theta &= p_T / P \\ \cos\theta &= p^z / P\end{aligned} \tag{A.80}$$

(where $p_T = \sqrt{p^x + p^y}$), (A.79) can be rewritten as

$$\begin{aligned}\frac{\partial\, \mathbf{p}_1 \cdot \mathbf{p}_2}{\partial \theta_1} &= -P_1 P_2 \left[\frac{p_1^z}{P_1}\frac{p_{T2}}{P_2}\frac{1}{p_{T1}}\frac{1}{p_{T2}}(p_1^x p_2^x + p_1^y p_2^y) - \frac{p_{T1}}{P_1}\frac{p_2^z}{P_2}\right] \\ &= p_{T1} p_2^z - \frac{p_1^z}{p_{T1}}(p_1^x p_2^x + p_1^y p_2^y).\end{aligned} \tag{A.81}$$

Similarly,

$$\frac{\partial\,\mathbf{p}_1\cdot\mathbf{p}_2}{\partial\theta_2} = p_{T2}p_1^z - \frac{p_2^z}{p_{T2}}(p_1^x p_2^x + p_1^y p_2^y). \tag{A.82}$$

Finally,

$$\begin{aligned}\frac{\partial\,\mathbf{p}_1\cdot\mathbf{p}_2}{\partial\phi_1} &= P_1P_2\sin\theta_1\sin\theta_2\sin(\phi_1-\phi_2)\\ &= P_1P_2\frac{p_{T1}}{P_1}\frac{p_{T2}}{P_2}\frac{1}{p_{T1}}\frac{1}{p_{T2}}(p_1^y p_2^x - p_1^x p_2^y)\\ &= p_1^y p_2^x - p_1^x p_2^y\end{aligned} \tag{A.83}$$

and

$$\frac{\partial\,\mathbf{p}_1\cdot\mathbf{p}_2}{\partial\phi_2} = p_1^x p_2^y - p_1^y p_2^x. \tag{A.84}$$

If there is no neutrino in the final state, then the gradients of the k_T constraints (A.73) also need to be calculated. This reduces to evaluating the derivatives of the x and y components of a four-momentum p^x, p^y with respect to P, θ, and ϕ. From (A.74), these gradients are

$$\begin{aligned}\frac{\partial p^x}{\partial P} &= p^x/P\\ \frac{\partial p^y}{\partial P} &= p^y/P.\end{aligned} \tag{A.85}$$

Also,

$$\begin{aligned}\frac{\partial p^x}{\partial\phi} &= -p^y\\ \frac{\partial p^y}{\partial\phi} &= p^x.\end{aligned} \tag{A.86}$$

And finally,

$$\begin{aligned}\frac{\partial p^x}{\partial\theta} &= P\cos\theta\cos\phi\\ &= p^x\cot\theta\\ \frac{\partial p^y}{\partial\theta} &= p^y\cot\theta.\end{aligned} \tag{A.87}$$

Now the variables need to be transformed from the set in which the gradients were evaluated to the final set used for describing the event. The first step will be to change the variables for the neutrino from spherical (P_ν, θ_ν, ϕ_ν) to rectangular ($p_\nu^x = x$, $p_\nu^y = y$, $p_\nu^z = z$) coordinates. This transformation is independent of all other event variables. The equations of transformation can be written

$$\begin{aligned} P &= \sqrt{x^2+y^2+z^2} \\ \theta &= \tan^{-1}\frac{p_T}{z} = \tan^{-1}\frac{\sqrt{x^2+y^2}}{z} \\ \phi &= \tan^{-1}\frac{y}{x}. \end{aligned} \tag{A.88}$$

Thus, by the chain rule,

$$\begin{aligned} \frac{\partial}{\partial x} &= \frac{\partial P}{\partial x}\frac{\partial}{\partial P} + \frac{\partial \phi}{\partial x}\frac{\partial}{\partial \phi} + \frac{\partial \theta}{\partial x}\frac{\partial}{\partial \theta} \\ &= \left(\frac{x}{P}\right)\frac{\partial}{\partial P} - \left(\frac{y}{{p_T}^2}\right)\frac{\partial}{\partial \phi} + \left(\frac{xz}{P^2 p_T}\right)\frac{\partial}{\partial \theta} \\ \frac{\partial}{\partial y} &= \frac{\partial P}{\partial y}\frac{\partial}{\partial P} + \frac{\partial \phi}{\partial y}\frac{\partial}{\partial \phi} + \frac{\partial \theta}{\partial y}\frac{\partial}{\partial \theta} \\ &= \left(\frac{y}{P}\right)\frac{\partial}{\partial P} + \left(\frac{x}{{p_T}^2}\right)\frac{\partial}{\partial \phi} + \left(\frac{yz}{P^2 p_T}\right)\frac{\partial}{\partial \theta} \\ \frac{\partial}{\partial z} &= \frac{\partial P}{\partial z}\frac{\partial}{\partial P} + \frac{\partial \phi}{\partial z}\frac{\partial}{\partial \phi} + \frac{\partial \theta}{\partial z}\frac{\partial}{\partial \theta} \\ &= \left(\frac{z}{P}\right)\frac{\partial}{\partial P} - \left(\frac{p_T}{P^2}\right)\frac{\partial}{\partial \theta}. \end{aligned} \tag{A.89}$$

The next step is to change from the x and y components of the neutrino momentum p_ν^x and p_ν^y to ${k_T}^x$ and ${k_T}^y$, as defined by (A.69). That is, the transformation is

$$x_\nu, y_\nu, z'_\nu, P'_i, \phi'_i, \theta'_i \rightarrow {k_T}^x, {k_T}^y, z_\nu, P_i, \phi_i, \theta_i \tag{A.90}$$

and the equations of transformation are

$$\begin{aligned}
x_\nu &= k_T{}^x - \sum_i P_i \sin\theta_i \cos\phi_i \\
y_\nu &= k_T{}^y - \sum_i P_i \sin\theta_i \sin\phi_i \\
z'_\nu &= z_\nu \\
P'_i &= P_i \\
\phi'_i &= \phi_i \\
\theta'_i &= \theta_i.
\end{aligned} \tag{A.91}$$

Thus, a Jacobian transform gives

$$\begin{aligned}
\frac{\partial}{\partial k_T{}^x} &= \frac{\partial}{\partial x_\nu} \\
\frac{\partial}{\partial k_T{}^y} &= \frac{\partial}{\partial y_\nu} \\
\frac{\partial}{\partial z_\nu} &= \frac{\partial}{\partial z'_\nu} \\
\frac{\partial}{\partial P_i} &= \frac{\partial P'_i}{\partial P_i}\frac{\partial}{\partial P'_i} + \frac{\partial x_\nu}{\partial P_i}\frac{\partial}{\partial x_\nu} + \frac{\partial y_\nu}{\partial P_i}\frac{\partial}{\partial y_\nu} \\
&= \frac{\partial}{\partial P'_i} - \frac{1}{P_i}\left(x_i\frac{\partial}{\partial x_\nu} + y_i\frac{\partial}{\partial y_\nu}\right) \\
\frac{\partial}{\partial \phi_i} &= \frac{\partial \phi'_i}{\partial \phi_i}\frac{\partial}{\partial \phi'_i} + \frac{\partial x_\nu}{\partial \phi_i}\frac{\partial}{\partial x_\nu} + \frac{\partial y_\nu}{\partial \phi_i}\frac{\partial}{\partial y_\nu} \\
&= \frac{\partial}{\partial \phi'_i} + y_i\frac{\partial}{\partial x_\nu} - x_i\frac{\partial}{\partial y_\nu} \\
\frac{\partial}{\partial \theta_i} &= \frac{\partial \theta'_i}{\partial \phi_i}\frac{\partial}{\partial \theta'_i} + \frac{\partial x_\nu}{\partial \phi_i}\frac{\partial}{\partial x_\nu} + \frac{\partial y_\nu}{\partial \phi_i}\frac{\partial}{\partial y_\nu} \\
&= \frac{\partial}{\partial \theta'_i} - \left(x_i\frac{\partial}{\partial x_\nu} + y_i\frac{\partial}{\partial y_\nu}\right)\cot\theta_i.
\end{aligned} \tag{A.92}$$

Finally, the polar angles need to be converted from θ to η. This is independent for each object. The transformation is

$$\theta = 2\tan^{-1} e^{-\eta}. \tag{A.93}$$

Thus,

$$\frac{\partial\theta}{\partial\eta} = \frac{2}{1+e^{-2\eta}}\left(-e^{-\eta}\right) = -\frac{2}{e^{\eta}+e^{-\eta}} = -\frac{1}{\cosh\eta} \tag{A.94}$$

and the associated Jacobian transform is

$$\frac{\partial}{\partial\eta} = \frac{\partial\theta}{\partial\eta}\frac{\partial}{\partial\theta} = -\frac{1}{\cosh\eta}\frac{\partial}{\partial\theta}. \tag{A.95}$$

There is one more transformation which might need to be made. If the lepton in the event is a muon, then the variable used to describe its momentum is not actually P but instead its inverse

$$K = 1/P. \tag{A.96}$$

The Jacobian transformation for this case is

$$\frac{\partial}{\partial K} = \frac{\partial P}{\partial K}\frac{\partial}{\partial P} = -\frac{1}{K^2}\frac{\partial}{\partial P} = -P^2\frac{\partial}{\partial P}. \tag{A.97}$$