

**Analysis and Fitting
of Certain Problem Data ***

Keith W. Hartman[†] and Richard Blankenbecler

*Stanford Linear Accelerator Center
Stanford University, Stanford, California 94309
hartman@r2tech.com and rzbth@slac.stanford.edu*

Abstract

The fitting of data that contains problem measurements and the possible improvement in the fits in situations in which the measured quantities satisfy auxiliary differential constraints is discussed. Problem data is defined to include (1) cases in which the data may contain poorly measured quantities but exact constraints must be satisfied and (2) cases in which random external perturbations, such as multiple scattering, may locally violate the constraint equations. Algorithms with a direct physical interpretation are developed that treat these situations. Applications to the measurement of magnetic fields and to detector surveying by using track data will be presented.

*Work supported by the Department of Energy, contract DE-AC03-76SF00515.

[†]Present address: R2 Technology, Los Altos, CA.

1 Introduction and Motivation

In the real world, the measurement of global quantities over a large volume are beset by a number of difficulties. In the paper by Ohlsson, Peterson, and Yuille[1] an elegant and useful method for track fitting in the presence of noise was given. This work was extended by Ohlsson[2] and by Blankenbecler[3]. In the present paper, a treatment of certain classes of measurement/fitting problems will be treated by using an approach inspired by these works. It will be shown that seemingly disparate problems can be treated here using a general and unified method; the two problems chosen for explicit treatment here are measuring a magnetic field over a large volume and surveying a detector using particle trajectory data. Excellent discussions of the problems of fitting magnetic fields over large volumes can be found in G. F. Pearce et.al.[4] and in the lectures of D. Newton[5].

An example problem that introduces the concepts and methods to be utilized in more complex situations will be given in the next section. Following that discussion, a more complete and general treatment will be given. Application to the problems of magnetic field measurement and to detector surveying by the use of tracks will then be given. The fundamental idea used here is to introduce weight parameters suggested by the physics of the measurement process. These parameters are then determined during the fitting procedure. They allow suspect data points to be subordinated in order to improve the fit to the data and the consistency with the constraint equations.

2 Example – Position Errors

2.1 Measurements, Constraints and Zips

Consider a set of measurements of the magnetic field B performed at the the space points r . The results are written as the vector

$$\vec{B} = \vec{B}(r) . \tag{1}$$

where an example of \vec{B} would be the measurement of the three components of the magnetic field at the point r .

In the sample problem to be treated here is that the measurements are not accurately located in the longitudinal direction. In this case, the data is taken in separate blocks, which we will term ‘zips’. Within each zip the data points are consistent and correctly located and spaced, but the zips themselves may have an uncertain longitudinal location relative to each other. Of course, much more general uncertainties can be treated by this method.

In this case, it is convenient to group the data into individual zips. The coordinates system is written as (x, y, z) and the measurement lattice as $x = i\Delta x$, $y = j\Delta y$, and $z = k\Delta z$. A zip consists of all the data taken at a given value of x and y . This is denoted as

$$\vec{B}(i, j) = \vec{B}(i, j, z + d(i, j)) , \tag{2}$$

where $d(i, j)$ is the unknown displacement parameter in the z-direction for this zip. These parameters will be determined by requiring that the measurements obey Maxwell’s equations throughout the measured volume.

If the parameters $d(i, j)$ were known, then Maxwell’s equations could be checked by evaluating the numerical derivatives of the fields. On the other hand, if Maxwell’s equations are imposed, the $d(i, j)$ can be determined to within measurement errors.

To this end, and following Ref[3] very closely, we introduce a total effective energy, or cost, function for each space point,

$$T(i, j, k) = (\vec{\nabla} \cdot \vec{B})^2 + (\vec{\nabla} \times \vec{B})^2 , \quad (3)$$

and for each zip,

$$T_{\text{zip}}(i, j) = \sum_k T(i, j, k) . \quad (4)$$

If Maxwell's equations are satisfied, then $T(i, j, k)$ will vanish for all values of (i, j, k) .

The value of $T_{\text{zip}}(i, j)$ gives a measure of the total violation along the zip labeled by (i, j) .

The energy is now written as

$$E(\mathbf{w}) = \sum_{i,j} [w_{ij} T_{\text{zip}}(i, j) + \sigma (1 - w_{ij})] . \quad (5)$$

where the $\mathbf{w} = \{w_{ij}\}$ are appropriate weights for the associated zip measurements.

The weight w_{ij} is one if the zip energy $T_{\text{zip}}(i, j)$ is to be fully included in the fitting process. It is zero if this zip is bogus and should be entirely discarded but this choice will yield an energy cost of σ , a preset parameter. The weights w_{ij} will be determined in the fitting process.

It is also possible to associate a weight with the measurement of the fields at each space point (i, j, k) in order to retain the valid points in a zip and to discard the bogus points, and this case will be discussed later. In any case, it will prove convenient and expedient to first treat the zips and to get a first estimate of the shifts $d(i, j)$ and the validity of each zip. Then a more refined treatment of each individual point can be done more efficiently.

2.2 Zip Annealing and Minimization

Our mathematical problem is to find the global minimum of the energy. An efficient method for treating this problem is simulated annealing (see refs[1] and [3]). The

following manipulations are to develop a robust yet simple fitting procedure. There are many unexpected features to the method; the procedure yields physical interpretations that have real meanings and are of considerable utility. The first step is to introduce a Boltzmann distribution for the relevance variables and any fit parameters

$$P[\mathbf{w}] = \frac{1}{Z} e^{-\beta E[\mathbf{w}]} , \quad (6)$$

where \mathbf{w} is the set of relevance variables, The inverse temperature β is introduced to control the search for the minimum energy by providing an overall scaling of the error measures. Finally, the partition (normalization) function is

$$Z = \sum_w e^{-\beta E[\mathbf{w}]} \quad (7)$$

in which the sum goes over *all* allowed values of the relevance parameters w_{ij} .

Now let us compute the *marginal* probability that describes the distribution of fitting parameters for a *uniform* distribution of assignments \mathbf{w} (each $w_{ij} = 0$ or 1)

$$P = \sum_{w_{ij}} P[w_{ij}] \equiv \frac{1}{Z} e^{-\beta E_{\text{eff}}(\mathbf{d})} , \quad (8)$$

where a total *effective* energy E_{eff} has been defined as the sum over the effective energies of each zip:

$$E_{\text{eff}}[\mathbf{d}] = \sum_{i,j} E_{\text{eff}}[i, j; \mathbf{d}] . \quad (9)$$

The evaluation of the sum over the allowed values of \mathbf{w} is straightforward. The result is

$$P = \frac{1}{Z} \prod_{i,j} D(i, j) \quad (10)$$

$$\text{where} \quad D(i, j) = \left\{ e^{-\beta T_{\text{zip}}(i,j)} + e^{-\beta\sigma} \right\} , \quad (11)$$

and the effective energy of $\text{zip}(i, j)$ is

$$E_{\text{eff}}[i, j; \mathbf{d}] = -\frac{1}{\beta} \log D(i, j; \mathbf{d}) = -\frac{1}{\beta} \log \left\{ e^{-\beta T_{\text{zip}}(i,j)} + e^{-\beta\sigma} \right\} . \quad (12)$$

Recall that $T_{\text{zip}}(i, j)$ could be written more explicitly as $T_{\text{zip}}(i, j; \mathbf{d})$.

We are looking for the most probable fit values of the displacement parameters \mathbf{d} , and the relevance parameters \mathbf{w} that minimize the energy in the limit of low temperatures, high β . In this limit, the wrong assignment configurations, i.e. those with a large value of $T_{\text{zip}}(i, j)$, i.e. $T_{\text{zip}}(i, j) \gg \sigma$, are exponentially suppressed in the marginal probability. Thus σ is a physical parameter, and represents the validity limit of the permitted zip errors.

The standard procedure now is to minimize the effective energy E_{eff} for a small value of β , and then to follow the fit parameters for a range of increasing values of β in order to avoid being trapped in a local minima.

Using the gradient descent method, at each stage in the iteration the parameters are changed by

$$\delta \mathbf{d} = -\eta \vec{\nabla}_{\mathbf{d}} E_{\text{eff}}[\mathbf{d}] = -\eta \sum_{i,j} \vec{\nabla}_{\mathbf{d}} E_{\text{eff}}[i, j, \mathbf{d}]. \quad (13)$$

This simply states that if the derivative of the energy with respect to a particular zip displacement is positive, reduce that displacement. The scalar parameter η is used to control the rate of approach to the minimum.

Interpretation: An explicit evaluation of the derivatives leads to a simple but useful interpretation of this process and of the quantities involved in the calculation. Note that the derivative of the energy $E(\mathbf{w})$ with respect to $T_{\text{zip}}(i, j)$ is the assignment w_{ij}

$$\frac{\partial E(\mathbf{w})}{\partial T_{\text{zip}}(i, j)} = w_{ij}, \quad (14)$$

while the derivative of the effective energy $E_{\text{eff}}[\mathbf{d}]$ is

$$\frac{\partial E_{\text{eff}}[\mathbf{d}]}{\partial T_{\text{zip}}(i, j)} = \frac{e^{-\beta T_{\text{zip}}(i, j)}}{\{e^{-\beta \sigma} + e^{-\beta T_{\text{zip}}(i, j)}\}}. \quad (15)$$

This suggests that one introduce the (thermalized) probability that the $(i, j)^{\text{th}}$ zip

measurement is relevant as

$$\langle w_{ij} \rangle = \frac{e^{-\beta T_{\text{zip}}(i,j)}}{\{e^{-\beta\sigma} + e^{-\beta V[j,p]}\}}, \quad (16)$$

and the probability of being irrelevant as $(1 - \langle w_{ij} \rangle)$.

The gradient descent equations can now be written as

$$\delta \vec{p}_t = -\eta \sum_{i,j} \langle w_{ij} \rangle \vec{\nabla}_d T_{\text{zip}}(i,j; \mathbf{d}). \quad (17)$$

Thus it is seen that points with a small value of the relevance probability $\langle w_{ij} \rangle$ are ignored in the determination of the parameters.

At this point, notice that if the nearest neighbor approximation is used to evaluate the divergence and curl in Maxwell's constraints, each $T_{\text{zip}}(i,j; \mathbf{d})$ depends only upon the nearest neighbor d's. That means that

$$\begin{aligned} -\beta \frac{\partial E_{\text{eff}}[\mathbf{d}]}{\partial d_{i,j}} &= \frac{\partial}{\partial d_{i,j}} \log D(i,j; \mathbf{d}) \\ &+ \frac{\partial}{\partial d_{i,j}} [\log D(i+1,j; \mathbf{d}) + \log D(i-1,j; \mathbf{d})] \\ &+ \frac{\partial}{\partial d_{i,j}} [\log D(i,j+1; \mathbf{d}) + \log D(i,j-1; \mathbf{d})]. \end{aligned} \quad (18)$$

Thus

$$\delta d_{i,j} = -\eta \frac{\partial E_{\text{eff}}[\mathbf{d}]}{\partial d_{i,j}}, \quad (19)$$

where

$$\begin{aligned} \frac{\partial E_{\text{eff}}[\mathbf{d}]}{\partial d_{i,j}} &= \langle w_{ij} \rangle \frac{\partial}{\partial d_{i,j}} T_{\text{zip}}(i,j; \mathbf{d}) \\ &+ \langle w_{i+1j} \rangle \frac{\partial}{\partial d_{i,j}} T_{\text{zip}}(i+1,j; \mathbf{d}) + \langle w_{i-1j} \rangle \frac{\partial}{\partial d_{i,j}} T_{\text{zip}}(i-1,j; \mathbf{d}) \\ &+ \langle w_{ij+1} \rangle \frac{\partial}{\partial d_{i,j}} T_{\text{zip}}(i,j+1; \mathbf{d}) + \langle w_{ij-1} \rangle \frac{\partial}{\partial d_{i,j}} T_{\text{zip}}(i,j-1; \mathbf{d}). \end{aligned} \quad (20)$$

The search for the minimum energy is carried out by starting the iterations at a small value of β , i.e. large temperature, and iterating several times through the zips using

the above equation to modify the values of the displacements \mathbf{d} . The value of β is then increased and the process repeated. This continues until a satisfactory minimum energy is achieved.

Absolute Zip Location: The above procedure will allow the entire zip array to move during the minimization process. This can be cured by either defining one of the zips to be the origin of the coordinate system by not varying its value of d . Another procedure is to keep the average displacement zero. After each iteration through the array, the average value of the displacements is computed and then reset to zero by translating each zip.

3 Measurements and Constraints

Consider a set of measurements performed at the the space points r_j . The results are written as the vector

$$\vec{M}_j = \vec{M}(r_j) . \tag{21}$$

where \vec{M} can have an arbitrary number of components, each component corresponding to one of the measurements made at the point r_j . An example would be the measurement of the three components of the magnetic field at the point r_j .

The problem is to fit the measurements with the general function

$$\vec{F}(r, p) , \tag{22}$$

where p is a set of arbitrary parameters to be determined by the data. To this end, and following Ref[3] very closely, we introduce a total effective energy, or cost, function that is linear in the energy of each measurement:

$$E_{\text{total}} = \sum_j w_j V[j, p] , \tag{23}$$

where w_j is an appropriate weight for the j^{th} measurement. The energy is

$$V[j, p] = \frac{|\vec{F}(p, r_j) - \vec{M}(r_j)|^2}{e_j^2}, \quad (24)$$

where e_j^2 is the square of the appropriate error measure for the j^{th} measurement, and the weight w_j is one if the j^{th} point is to be fully included in the fitting process, or zero if it is bogus and should be discarded.

Constraints: If the measured quantity is a physical ‘field’ or a particle trajectory, both of which must satisfy differential equations, then there are constraints on its values at different spatial points. Write these equations formally as

$$\Delta \vec{F}(r, p) = 0 \quad \text{and} \quad T[j, p] = |\Delta \vec{F}(r_j, p)|^2, \quad (25)$$

where, for example, if the constraint is the vanishing of the second derivative of the position $F(r, p)$, this is of the form

$$\Delta \vec{F}(r, p) = \vec{F}(r_{j+1}, p) - 2\vec{F}(r_j, p) + \vec{F}(r_{j-1}, p). \quad (26)$$

If the constraint is Maxwell’s equations, then $\Delta \vec{F}$ obviously takes a different form. The function $T[j, p]$ depends upon the measurements in the neighborhood of the point j . These differential requirements must be imposed at all space points. They then relate the fitting function at neighboring space points and hence act as a non-local ‘smoothing’ constraint. Thus there is an additional energy or cost that must be minimized, namely the violation of these constraints. Imposing the equations of motion at the points r_j , the total energy becomes

$$E_{\text{total}} = \sum_j \{ W T[j, p] + w_j V[j, p] \}, \quad (27)$$

The parameters p are chosen to minimize the total energy function. Thus the fit is a compromise between the measured values and the constraint equations as measured by the ratio of the weights W and the w_j . The basic scheme, to be used many

times in this paper, is to note that the errors will be due to the violation of either $T[j, p]$ or $V[j, p]$. In this case, it is the measurements in $V[j, p]$ that are suspect, so the parameters w_j is introduced to allow the weight of a particular measurement sequence to be varied to optimize the fit by conserving the constraint equations.

4 Noise and Bogus Data

One obvious omission in the above discussion is how to decide during the fitting process which if any of the data should not be included in the fit. The treatment of such bogus data is straightforward. It is implemented by allowing the weight variable w_j to be determined as a parameter of the fitting process; it will become a relevance variable for the associated measurement point. The energy is written as

$$E_{\text{total}} = \sum_j E[j, p] = \sum_j \{ WT[j, p] + w_j V[j, p] + \nu(1 - w_j)^2 \}. \quad (28)$$

Thus if w_j is assigned the value one, the j^{th} data point contributes $V[j, p]$ to the energy; if w_j is assigned the value zero, the energy cost is ν . For either value of the weight w_j , the constraint is to be enforced with weight W .

On the other hand, the physical situation may be such that there is a random ‘noise’ in the system which does not affect V , or the measurement value, but does affect the nonlocal constraint by modifying the differential equations. One such example is multiple scattering which can insert a random ‘kink’ into the path. This is handled by allowing the formalism to ignore the constraint at a point, but to still utilize the data before and after the scattering. After modifying the relevance variable in an obvious way, the energy becomes

$$E_{\text{total}} = \sum_j E[j, p] = \sum_j \{ w_j T[j, p] + V[j, p] + \nu(1 - w_j)^2 \}. \quad (29)$$

5 Simulated Annealing

Our mathematical problem is to find the global minimum of the energy. An efficient method for treating this problem is simulated annealing (see refs[1] and [3]). The following manipulations are to develop a robust yet simple fitting procedure. There are many unexpected features to the method; the procedure yields physical interpretations that have real meanings and are of considerable utility. The first step is to introduce a Boltzmann distribution for the relevance variables and the fit parameters

$$P[\mathbf{w}; \mathbf{p}] = \frac{1}{Z} e^{-\beta E[\mathbf{w}; \mathbf{p}]} , \quad (30)$$

where \mathbf{w} ($= \{w_j\}$) is the set of relevance variables, \mathbf{p} ($= \{p_t\}$) is the set of track parameters, and β is an inverse temperature. It is introduced to control the search for the minimum energy by providing an overall scaling of the error measures. Finally, the partition (normalization) function is

$$Z = \sum_w \sum_p e^{-\beta E[\mathbf{w}; \mathbf{p}]} \quad (31)$$

in which the sum goes over *all* allowed values of \mathbf{w} and \mathbf{p} .

Now let us compute the *marginal* probability that describes the distribution of fitting parameters for a *uniform* distribution of assignments \mathbf{w} ($= 0$ or 1)

$$P[\mathbf{p}] = \sum_{w_j} P[w_j; \mathbf{p}] \equiv \frac{1}{Z} e^{-\beta E_{\text{eff}}[\mathbf{p}]} , \quad (32)$$

where an *effective* energy E_{eff} has been defined

$$E_{\text{eff}}[\mathbf{p}] = \sum_j E_{\text{eff}}[j, \mathbf{p}] . \quad (33)$$

The evaluation of the sum over the allowed values of \mathbf{w} is straightforward. The result is.

$$P[\mathbf{p}] = \frac{1}{Z} \prod_j D(j) \quad (34)$$

$$\text{where } D(j) = \left\{ e^{-\beta \nu} + e^{-\beta V[j, p]} \right\} e^{-\beta W T[j, p]} , \quad (35)$$

and the effective energy of the j^{th} measurement is

$$E_{\text{eff}} [j, \mathbf{p}] = -\frac{1}{\beta} \log D(h) = -\frac{1}{\beta} \log \left\{ e^{-\beta\nu} + e^{-\beta V[j,p]} \right\} + WT[j, p] . \quad (36)$$

6 Treatment of Fluctuations

In this section a more general approach to random errors on both measured quantities and tracks will be considered. Consider the problem defined by

$$E[j, p] = w_j T[j, p] + V[j, p] + \mu(1 - w_j)^2 , \quad (37)$$

$$\text{with } T[j, p] = (t_j - \delta_t^j)^2 \quad \text{and} \quad V[j, p] = (v_j - \delta_v^j)^2 , \quad (38)$$

where δ_t^j is a random perturbation of the nonlocal constraints at the point j , such as would be caused by multiple scattering for example, and δ_v^j is a random error in the measurement of the track location. These are physically different and independent effects. Different j values are uncorrelated. The assignment variable w_j is used to incorporate or to ignore the constraint.

Proceeding as in the previous section, the *marginal* probability is now evaluated by summing the Boltzmann distribution for E over the assignments \mathbf{w} ($= 0$ or 1) and by integrating over the random variables δ_t and δ_v . It will be assumed that the δ 's for each j value are distributed according to

$$dP[\delta_t^j, \delta_v^j] = \frac{d\delta_t^j d\delta_v^j}{Z_t^j Z_v^j} e^{-\delta_t^2/(2\langle T[j,p] \rangle) - \delta_v^2/(2\langle V[j,p] \rangle)} , \quad (39)$$

where $Z_t^j Z_v^j = 2\pi\sqrt{\langle T[j,p] \rangle \langle V[j,p] \rangle}$. The marginal probability can now be computed as

$$P = \sum_w \int dP[\delta_t, \delta_v] \frac{1}{Z} e^{-\beta E} = \sum_w \prod_j \frac{1}{Z(j, w)} e^{-\beta E_{\text{eff}}(j, w)} , \quad (40)$$

$$E_{\text{eff}}(j, w) = \frac{w_j T[j, p]}{1 + 2\beta w_j \langle T[j, p] \rangle} + \frac{V[j, p]}{1 + 2\beta \langle V[j, p] \rangle} + \mu(1 - w_j)^2 , \quad (41)$$

where $Z(j, w) = Z(0)\sqrt{1 + 2\beta w < T[j, p] >}$. Proceeding as in the previous section, eqn(36) becomes in this situation

$$E_{\text{eff}}(j) = V_{\text{eff}}(j) - \frac{1}{\beta} \log \left\{ e^{-\beta\mu} + e^{-\beta T_{\text{eff}}(j)} \right\}, \quad (42)$$

$$V_{\text{eff}}(j) = \frac{V[j, p]}{1 + 2\beta < V[j, p] >} \quad (43)$$

$$T_{\text{eff}}(j) = \frac{T[j, p]}{1 + 2\beta < T[j, p] >} + \frac{1}{2\beta} \log(1 + 2\beta < T[j, p] >). \quad (44)$$

The correct statistical measure of the fluctuations is correctly introduced as β increases.

7 Minimization

We are looking for the most probable fit function, i.e. with the parameter values and relevance variables that minimize the energy in the limit of low temperatures, high β . In this limit, the wrong assignment configurations, i.e. those with a finite $V[j, p]$, are exponentially suppressed in the marginal probability. The standard procedure is to minimize the effective energy $E_{\text{eff}}[\mathbf{p}]$ for a small value of β , and then to follow the fit parameters for a range of increasing values of β in order to avoid being trapped in a local minima.

Using the gradient descent method, at each stage in the iteration the parameters of each track t are changed by

$$\delta \vec{p}_t = -\eta \vec{\nabla}_{\mathbf{p}} E_{\text{eff}}[\mathbf{p}] = -\eta \sum_j \vec{\nabla}_{\mathbf{p}} E_{\text{eff}}[j, \mathbf{p}]. \quad (45)$$

The scalar parameter η is used to control the rate of approach to the minimum.

Interpretation: An explicit evaluation of the derivatives leads to a simple but useful interpretation of this process and of the quantities involved in the calculation. Note that the derivative of the energy $E[j, p]$ with respect to $V[j, p]$ is the assignment w_j .

The derivative of the effective energy with respect to the same variable when there are no fluctuations present is

$$\frac{\partial E_{\text{eff}} [j, \mathbf{p}]}{\partial V[j, p]} = \frac{e^{-\beta V[j, p]}}{\{e^{-\beta \nu} + e^{-\beta V[j, p]}\}} . \quad (46)$$

This suggests that one introduce the (thermalized) probability that the j^{th} measurement is relevant as

$$\langle w_j \rangle = \frac{e^{-\beta V[j, p]}}{\{e^{-\beta \nu} + e^{-\beta V[j, p]}\}} , \quad (47)$$

and the probability of being irrelevant as $(1 - \langle w_j \rangle)$.

The gradient descent equations can now be written as

$$\delta \vec{p}_t = -\eta \sum_j \left\{ \langle w_j \rangle \vec{\nabla}_p V[j, p] + W \vec{\nabla}_p T[j, p] \right\} . \quad (48)$$

Thus it is seen that points with a large value of the relevance probability, $\langle w_j \rangle$ dominate the determination of the parameters while the constraints are imposed at every point no matter what the quality of the measurement is at that point.

8 Zips

In many cases, the data is taken in separate blocks, which have been termed ‘zips’. Within each zip the data points are consistent and correctly correlated, but the zips themselves may have problems relative to each other. In this case, it is convenient to group the data into individual zips and write the partition function of eq(35) in the form

$$P[\mathbf{p}] = \frac{1}{Z} \prod_z D[z] \quad (49)$$

$$\text{where} \quad D[z] = \prod_{j \text{ in } z} \left\{ e^{-\beta \nu} + e^{-\beta V[j, p]} \right\} e^{-\beta W T[j, p]} , \quad (50)$$

fluctuations are neglected, and the effective energy of zip z is

$$E_{\text{eff}} [z, \mathbf{p}] = -\frac{1}{\beta} \log D[z] = -\frac{1}{\beta} \sum_{j \text{ in } z} \log D(j) \quad (51)$$

$$= -\frac{1}{\beta} \sum_{j \text{ in } z} \log \left\{ e^{-\beta\nu} + e^{-\beta V[j,p]} \right\} + W \sum_{j \text{ in } z} T[j,p] . \quad (52)$$

This is to be contrasted with starting over at eqn(23) and eqn(28) and assuming that an entire zip is either relevant or irrelevant by introducing the zip weights w_z and writing

$$E_{\text{total}} = \sum_z \left\{ W T[z,p] + w_z V[z,p] + \sigma(1 - w_z)^2 \right\} , \quad (53)$$

$$\text{where } T[z,p] = \sum_{j \text{ in } z} T[j,p] \quad \text{and} \quad V[z,p] = \sum_{j \text{ in } z} V[j,p] . \quad (54)$$

Following the same line of argument as before, the result for the conditional probability after averaging over the zip weights w_z is

$$P[\mathbf{p}] = \frac{1}{Z} \prod_z D_{zip}[z] , \quad (55)$$

$$\text{where } D_{zip}[z] = \left\{ e^{-\beta\sigma} + e^{-\beta V[z,p]} \right\} e^{-\beta W T[z,p]} \quad (56)$$

$$\text{or } D_{zip}[z] = \left\{ e^{-\beta\sigma} + \prod_{j \text{ in } z} e^{-\beta V[j,p]} \right\} \prod_{j \text{ in } z} e^{-\beta W T[j,p]} \quad (57)$$

$$. \quad (58)$$

This is a different form than that given in eqn(52). All the measurements in a given zip are treated equally in eqn(55). In contrast here one has

$$E_{\text{zip}} [z, \mathbf{p}] = -\frac{1}{\beta} \log D_{zip}[z] \quad (59)$$

$$= -\frac{1}{\beta} \log \left\{ e^{-\beta\sigma} + e^{-\beta V[z,p]} \right\} + W T[z,p] . \quad (60)$$

The noise cost parameter σ determines if an entire zip is used in the parameter determination or whether it is totally ignored. In eqn(52) on the other hand, the relevance of each individual measurement is judged by comparing to the noise parameter ν . One choice for these parameters is to set $\sigma \approx N\nu$, where N is the number of bad points in a zip that disqualifies it from the fitting procedure. The information in any good measurement points can be recovered by using the result of the zip analysis as the starting point for the point-by-point formulation of eqn(28).

The gradient descent method for determining the parameters \mathbf{p} are now given in terms of the zips:

$$\delta \vec{p}_t = -\eta_z \vec{\nabla}_p E_{\text{eff}}[\mathbf{p}] = -\eta_z \sum_z \vec{\nabla}_p E_{\text{zip}}[z, \mathbf{p}]. \quad (61)$$

The scalar parameter η_z is used to control the rate of approach to the minimum.

The derivative of the effective energy with respect to $V[z, p]$ suggests that one introduce the (thermalized) probability that the z^{th} zip is relevant as

$$\langle w_z \rangle = \frac{e^{-\beta V[z, p]}}{\{e^{-\beta \sigma} + e^{-\beta V[z, p]}\}}. \quad (62)$$

The gradient descent equations can now be written as

$$\delta \vec{p}_t = -\eta_z \sum_z \left\{ \langle w_z \rangle \vec{\nabla}_p V[z, p] + W \vec{\nabla}_p T[z, p] \right\}. \quad (63)$$

Thus it is seen that zips with a large value of the relevance probability $\langle w_z \rangle$ dominate the determination of the parameters while the constraints are imposed at every point no matter what the quality of the zip measurement is at that point.

Zips were defined to be an internally consistent data series. To allow the fitting routine to move the zips relative to each other, it is convenient to add a displacement parameter d_z for each zip to the set of parameters p and write

$$V[z, p] = \sum_{j \text{ in } z} \frac{1}{e_j^2} |\vec{M}(r_j - d_z) - \vec{F}(p, r_j)|^2, \quad (64)$$

where the displacement parameters d_z are also to be varied to minimize the energy.

9 Surveying a Detector

In this section an important variation of the fitting problem will be discussed. Consider a detector comprised of a set of planes separated by a fixed amount and a set of particle trajectories that pass through the detector and are registered by each plane.

The problem here is to use the known character of the particle paths to locate the detector planes to high accuracy. To this end, define

$$\begin{aligned}
M[t; n] &= \text{approximate location of track } t \text{ at plane } n \\
d[n] &= \text{displacement of plane } n \text{ (to be determined)} \\
M[t; n] + d[n] &= \text{true location of track } t \text{ at plane } n \\
F[t; n] &= \text{true trajectory location at plane } n
\end{aligned} \tag{65}$$

The cost or energy for track t is composed of local and nonlocal terms

$$E[t] = V[t; d] + T[t; d] = \sum_n \{V[t; n, d] + T[t; n, d]\} . \tag{66}$$

The local or potential term is

$$V[t; n, d] = \frac{1}{e_n^2} |M[t; n] + d[n] - F[t; n]|^2 . \tag{67}$$

The differential equation for the track relates the trajectory location at neighboring planes. Thus the nonlocal term is of the general form

$$T[t; n, d] = (\Delta F[t; n, d])^2 \tag{68}$$

$$\Delta F[t; n, d] = A_+[n, d]F[t; n + 1] - A_0[n, d]F[t; n] + A_-[n, d]F[t; n - 1] , \tag{69}$$

where the $A_{\pm,0}[n, d]$ are given by particle dynamics and the plane spacings. For example, if there is no magnetic field, the tracks are straight lines, and the planes are uniformly spaced, then $A_+[n, d] = A_-[n, d] = 1$, $A_0[n, d] = 2$. Note that $\Delta F[t; n, d]$ is then simply the numerical second derivative (which should vanish for a straight path). More accurate and more nonlocal forms for this constraint can be utilized as well. The ultimate would be fitting the path with a straight line through the detector, but this would unnecessarily discard useful data.

There is a problem with the above formulation. If the particle decays or undergoes a large multiple scattering, then the track will not be well represented by the minimum

of Eq.(68), and the track should not be forced to obey the equation of motion at that point. However, the track data before and after the ‘kink’ can be used.

9.1 Tracks

First a formulation that either retains or discards the data from an entire track will be presented. Introduce a weight variable w_t which is either zero or one and write

$$E_{\text{total}} = E[\mathbf{w}; \mathbf{d}] = \sum_t \left\{ V[t; d] + w_t T[t; d] + \sigma(1 - w_t)^2 \right\} , \quad (70)$$

where σ is the energy cost of discarding the differential constraints on track t .

Proceeding as before, the Boltzmann distribution for the weights \mathbf{w} and the plane displacement parameters \mathbf{d} is introduced as

$$P[\mathbf{w}; \mathbf{d}] = \frac{1}{Z} e^{-\beta E[\mathbf{w}; \mathbf{d}]} , \quad (71)$$

where \mathbf{w} ($= \{w_t\}$) is the set of relevance variables, \mathbf{d} ($= \{d[n]\}$) is the set of plane displacements, β is an inverse temperature and Z is the normalization factor. The *marginal* probability is now introduced by summing over the assignments \mathbf{w} ($= 0$ or 1)

$$P[\mathbf{d}] = \sum_{w_t} P[w_t; \mathbf{d}] \equiv \frac{1}{Z} e^{-\beta E_{\text{eff}}[\mathbf{d}]} , \quad (72)$$

where the *effective* energy E_{eff} is

$$E_{\text{eff}}[\mathbf{d}] = \sum_t E[t; \mathbf{d}] , \quad (73)$$

$$\text{with} \quad E[t; \mathbf{d}] = V[t; d] - \frac{1}{\beta} \log \left\{ e^{-\beta\sigma} + e^{-\beta T[t; d]} \right\} . \quad (74)$$

Notice that this is the same form as Eq.(60) but that T and V are interchanged. This could be anticipated by comparing Eq.(70) with Eq.(53).

The gradient descent method for determining the displacement parameters of each plane n are now given as:

$$\delta \vec{d} = -\eta_d \vec{\nabla}_d E_{\text{eff}}[\mathbf{d}] = -\eta_d \sum_t \vec{\nabla}_d E[t; \mathbf{d}] . \quad (75)$$

The scalar parameter η_d is used to control the rate of approach to the minimum.

The derivative of the effective energy with respect to $T[t; d]$ suggests that one again introduce a (thermalized) probability that the data from track t is relevant as

$$\langle w_t \rangle = \frac{e^{-\beta T[t; d]}}{\{e^{-\beta \sigma} + e^{-\beta T[t; d]}\}} . \quad (76)$$

The gradient descent equations can now be written as

$$\delta \vec{d} = -\eta_d \sum_t \left\{ \vec{\nabla}_d V[t; d] + \langle w_t \rangle \vec{\nabla}_d T[t; d] \right\} . \quad (77)$$

Thus it is seen that tracks with a large value of the relevance probability $\langle w_t \rangle$ dominate the determination of the plane displacements. Tracks which can not be made good solutions of the path differential equation by varying the plane displacements d are de-weighted automatically in the fitting procedure.

9.2 SubTracks

The above procedure may throw out useful data. For example, if a track undergoes a multiple scattering in the detector, the entire track will be thrown out in spite of the fact that the track segments before and after the scatter can provide useful data if the ‘offending’ kink in the track is discarded in the analysis. This problem can be dealt with, at the expense of introducing a rather mild complication, by introducing a relevance variable w_t^n which is either zero or one and writing

$$E_{\text{total}} = E[\mathbf{w}; \mathbf{d}] = \sum_t \sum_n \left\{ V[t; n, d] + w_t^n T[t; n, d] + \mu(1 - w_t^n)^2 \right\} , \quad (78)$$

where μ is the energy cost of discarding the differential constraints on track t at the location of the kink. That is, the differential equations of free motion are not imposed because there is an impulse delivered to the particle at this location. Proceeding as before, the Boltzmann distribution for the weights \mathbf{w} and the plane displacement

parameters \mathbf{d} is introduced as

$$P[\mathbf{w}; \mathbf{d}] = \frac{1}{Z} e^{-\beta E[\mathbf{w}; \mathbf{d}]} , \quad (79)$$

where \mathbf{w} ($= \{w_t^n\}$) is the set of relevance variables, \mathbf{d} ($= \{d[n]\}$) is the set of plane displacements, β is an inverse temperature and Z is the normalization factor. The *marginal* probability is now introduced by summing over the assignments \mathbf{w} ($= 0$ or 1)

$$P[\mathbf{d}] = \frac{1}{Z} \prod_{t,n} e^{-\beta V[t;n,d]} \left(e^{-\beta T[t;n,d]} + e^{-\beta \mu} \right) \equiv \frac{1}{Z} e^{-\beta E_{\text{eff}}[\mathbf{d}]} , \quad (80)$$

where the *effective* energy E_{eff} has become

$$E_{\text{eff}}[\mathbf{d}] = \sum_t E[t; \mathbf{d}] = \sum_t \sum_n E[t; n, \mathbf{d}] , \quad (81)$$

$$\text{with} \quad E[t; n, \mathbf{d}] = V[t; n, d] - \frac{1}{\beta} \ln \left(e^{-\beta T[t;n,d]} + e^{-\beta \mu} \right) . \quad (82)$$

The gradient descent method for determining the displacement parameters of each plane n are now given as:

$$\delta \vec{d} = -\eta_d \vec{\nabla}_d E_{\text{eff}}[\mathbf{d}] = -\eta_d \sum_t \sum_n \vec{\nabla}_d E[t; n, \mathbf{d}] . \quad (83)$$

The scalar parameter η_d is used to control the rate of approach to the minimum.

Once again, the derivative of the effective energy with respect to $T[t; n, d]$ suggests that one introduce the (thermalized) relevance probability for the data from the n^{th} plane of track t as

$$\langle w_t^n \rangle = \frac{e^{-\beta T[t;n,d]}}{\{e^{-\beta \mu} + e^{-\beta T[t;n,d]}\}} . \quad (84)$$

If the path equation error measure $T[t; n, d]$ is smaller than μ , then $\langle w_t^n \rangle \sim 1$. However, if the inequality is reversed, $\langle w_t^n \rangle \sim 0$, and this data point is discarded automatically. The gradient descent equations can now be written as

$$\delta \vec{d} = -\eta_d \sum_t \sum_n \left\{ \vec{\nabla}_d V[t; n, d] + \langle w_t^n \rangle \vec{\nabla}_d T[t; n, d] \right\} . \quad (85)$$

Thus it is seen that track segments with a large value of the relevance probability $\langle w_t^n \rangle$ dominate the determination of the plane displacements.

10 Multiple Scattering

The present formalism allows a more direct treatment of multiple scattering. This extension was briefly discussed in section 5. As before, consider a detector comprised of a set of planes separated by a fixed amount and a set of particle trajectories that pass through the detector and are registered by each plane. At each plane, multiple scattering will introduce a random momentum transfer, or ‘kink’, in the path. To treat this effect, define the kinetic energy as

$$T[t; n, d, f_n^t] = |\Delta F[t; n, d] - \Delta \mathbf{f}[t; n, d]|^2, \quad (86)$$

where $\Delta \mathbf{f}[t; n, d]$ is a random variable arising from multiple scattering; it will lead to the RMS variation in the differential path relation $\Delta F[t; n, d]$ that rises from multiple scattering occurring from the n^{th} plane up to the $(n+1)^{\text{st}}$ plane. The local potential energy is still given by eqn(67). The total energy and hence the Boltzmann distribution then become a function of \mathbf{w} , \mathbf{d} and $\Delta \mathbf{f}$.

Following closely the treatment given earlier in section 5, the *marginal* probability is now evaluated by summing the Boltzmann distribution over the assignments \mathbf{w} (= 0 or 1) and by integrating over the random variables $\Delta \mathbf{f}$. It will be assumed that the $\Delta \mathbf{f}$ are distributed according to

$$dP[\Delta \mathbf{f}] = \prod_{n,t} \frac{d\Delta f[t; n, d]}{Z_n} e^{-\Delta f[t; n, d]^2 / (2\langle T[n, d] \rangle)}, \quad (87)$$

where $\langle \Delta f[t; n, d] \rangle = 0$ and $\langle \Delta f[t; n, d] \Delta f[t'; n, d] \rangle = \delta_{t,t'} \langle T[n, d] \rangle$, i.e., different tracks are not correlated. These averages do not depend upon the particular track under consideration but may vary from plane to plane since detector element each may have its own multiple scattering characteristics.

The marginal probability can now be computed as

$$P[\mathbf{d}] = \sum_{w_t} \int dP[\Delta \mathbf{f}] P[w_t; \mathbf{d}; \Delta \mathbf{f}] \equiv \frac{1}{Z} e^{-\beta E_{\text{eff}}[\mathbf{d}]}, \quad (88)$$

$$E_{\text{eff}}[\mathbf{d}] = \sum_t \sum_n E_{eff}[t; n, \mathbf{d}] . \quad (89)$$

In this case one finds

$$E_{eff}[t; n, \mathbf{d}] = V[t; n, d] - \frac{1}{\beta} \log \left\{ e^{-\beta\mu} + e^{-\beta T_{eff}[t; n, d]} \right\} , \quad (90)$$

$$T_{eff}[t; n, d] = \frac{T[t; n, d]}{1 + 2\beta \langle T[n, d] \rangle} + \frac{1}{2\beta} \log(1 + 2\beta \langle T[n, d] \rangle) , \quad (91)$$

where $T[t; n, d]$ is given by eqns(68)-(69). The second term in eqn(91) is a normalization correction; it does not depend upon t and could be absorbed into an effective cost parameter μ if desired. In the limit of no multiple scattering, $\langle T[n, d] \rangle \sim 0$ and these equations reduce to the results of the previous section. The relevance probability in this case is

$$\langle w_t^n \rangle = \frac{e^{-\beta T_{eff}[t; n, d]}}{\left\{ e^{-\beta\mu} + e^{-\beta T_{eff}[t; n, d]} \right\}} . \quad (92)$$

Limiting Behavior: The above result simplifies in the limit that *every* contribution from *all* tracks are accepted. This is achieved formally by letting μ become large so that the cost of ignoring a contribution is unacceptable:

$$E_{eff}[t; n, \mathbf{d}] \simeq V[t; n, d] + T_{eff}[t; n, d] , \quad (93)$$

where $T_{eff}[t; n, d]$ is given by eqn(91). The last term in eqn(91) can now be fully absorbed into the normalization factor Z of $P[\mathbf{d}]$. Note that the denominator in the first term in eqn(91) is unity in the limit of small beta, but as β increases, the RMS path error relevant to the marginal probability cannot be driven smaller than the minimum expected from multiple scattering, namely $\langle T[n, d] \rangle$.

11 Acknowledgments

We wish to thank William Dunwoodie, George Irwin and Carsten Peterson for many useful discussions.

References

- [1] M. Ohlsson, C. Peterson, A. Yuille, “Track Finding with Deformable Templates—The Elastic Arms Approach,” *Computer Physics Communications* **71**, 77 (1992). An extensive list of earlier references are given here.
- [2] M. Ohlsson, “Extensions and Explorations of the Elastic Arms Algorithm,” *Computer Physics Communications* **77**, 19 (1993).
- [3] Richard Blankenbecler, “Deformable templates — revisited and extended — with an OOP implementation,” *Computer Physics Communications* **81**, 318-334 (1994). See also “A unified treatment of track reconstruction and particle identification,” *Computer Physics Communications* **81**, 335-342 (1994).
- [4] A Method of Field Parameterization for Spectrometer Magnets. G.F. Pearce et. al., *Nuc. Inst. Meth.* **115**, 371 (1978).
- [5] The Magnetic Field Mapping of Detector Magnets. D. Newton, Ed. S. Turner, Proceedings of CERN Accelerator School Magnetic Measurement and Alignment, March 1992, CERN 92-05, 283 (1992).