Linearized Track Fitting

Marco De Mattia

November 11, 2014

1 Introduction

The task of fitting tracks from a multitude of hits in the CMS silicon tracker is achieved offline by the CMS tracking software. The software uses a Kalman Filter approach to refine the information from hits in successive layers. Even in the offline environment, where time is in principle not a constraint, it is not feasible to explore all possible hit combinations. Seeds are instead built using layers that give 3D information (pixels, or double sided strip layers) to guide the search for hits in a narrower window. Even with this approach it is not feasible to perform track fitting within the 12.5 μs latency that will be available in the L1 trigger of CMS for the HL-LHC. New solutions must be found to reduce combinatorics and speed up the track fitting. One possible approach is to utilize associative memories (AM) to quickly group together subsets of the hits in roads that are potentially consistent with tracks. The coarseness of this step leads to multiple fake roads and an additional step is needed to refine the information, remove fakes, and estimate track parameters.

In this document we explore a method of fitting tracks out of a set of hits that is simple enough to be fast and implementable in FPGAs. This approach is based on a linear approximation of the dependence of the track parameters on the hit coordinates. The linear approximation will be acceptable only in a limited region of the phase space, i.e. an expansion of the function around a specific point. This method was utilized in the SVT at CDF and a review can be found in [1] and [3].

Other methods for track fitting at L1 trigger are being explored for CMS, such as Hough Transform and Retina Fit. They are not described in this document.

2 Linear Approximation

Each candidate track is represented as a list of n hit coordinates, and can be thought of as a point $\mathbf{x}$ within a set $\mathcal{C} \subset R^n$. This subset is limited by the physical boundaries on the hit coordinates. Of all the points in $\mathcal{C}$ only a subset $\mathcal{T} \subset \mathcal{C}$ can be reasonably consistent with tracks coordinates. The concrete subset $\mathcal{T}$ is defined by specific criteria, such as a χ^2 cut. Track finding is the process of deciding whether a set of coordinates corresponds to a track in $\mathcal{T}$.

Each track is fully defined by a set of $m < n$ parameters $\mathbf{p}$ (e.g. p_T, ϕ, η, d_0, z_0). In the case of perfect detector resolution, the hit coordinates are uniquely

determined by the parameters $\mathbf{p}$, and the set $\mathcal{T}$ reduces to an m-dimensional surface contained in $\mathcal{C}$, described by n parametric equations,

$$\mathbf{x} = \mathbf{x}(p_1, \ldots, p_m) \,. \tag{1}$$

By eliminating the parameters equation 1 can be rewritten as $n - m$ constraint equations,

$$f_i(\mathbf{x}) = 0; \ i = 1, \ldots, n - m \,. \tag{2}$$

A set of coordinates $\mathbf{x}$ corresponds to a track in $\mathcal{T}$ if and only if it satisfies the constraints equations.

In the case of finite detector resolution each hit coordinate has an associated uncertainty. The constraint functions f_i then become random variables that take values slightly different from zero. Geometrically, the surface $\mathcal{T}$ acquires some thickness. In this case, there is no general exact solution to the constraint equations.

2.1 Linearized Constraints

The starting system of equations is generally not diagonal. However, given that the f_i are random variables we can write their covariance matrix

$$F_{ij} = E[\ (f_i(\mathbf{x}) - E[f_i(\mathbf{x})])(f_j(\mathbf{x}) - E[f_j(\mathbf{x})]\ ] \,, \tag{3}$$

where $E[\ ]$ denotes the expected value. By expanding the constraint functions to first order around a point $\mathbf{x_0}$, and defining $\Delta x_r = x_r - x_{r0}$,

$$f_i(\mathbf{x}) \simeq \frac{\partial f_i}{\partial \mathbf{x}}(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) = \sum_{r=1}^{n} \frac{\partial f_i}{\partial x_r}(\mathbf{x}_0) \Delta x_r \,, \tag{4}$$

we can express, to first order, the covariance matrix of the constraint functions f as a function of the covariance matrix of the deltas of the coordinates:

$$
\begin{aligned}
F_{ij} &= E\left[\frac{\partial f_i}{\partial \mathbf{x}}(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) - E\left[\frac{\partial f_j}{\partial \mathbf{x}}(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)\right]\right] = \\
&= \sum_{rs} \frac{\partial f_i}{\partial x_r}\frac{\partial f_j}{\partial x_s}\ E\left[(\Delta x_r - E[\Delta x_r])(\Delta x_s - E[\Delta x_s])\right] = \\
&= \sum_{rs} \frac{\partial f_i}{\partial x_r}\frac{\partial f_j}{\partial x_s} \mathrm{Cov}(\Delta x_r, \Delta x_s)
\end{aligned}
\tag{5}
$$

The matrix F_{ij} is a covariance matrix, therefore it is symmetric and it can be diagonalized through a rotation in the space of the f_is. The new coordinates $\tilde{f}_i$ can be scaled to unit variance so that the sum $\sum_i \tilde{f}_i^2$, with $\langle \tilde{f}_i^2 \rangle = 1$, follows a χ^2 distribution as long as the PDFs of the $\tilde{f}_i$s are gaussian. The linear expansion of the diagonalized base of $\tilde{f}_i$s can be written as

$$\tilde{f}_i \simeq \frac{\partial \tilde{f}_i}{\partial \mathbf{x}} \cdot (\mathbf{x} - \mathbf{x}_0) = \mathbf{v_i} \cdot \mathbf{x} + c_i \,. \tag{6}$$

Geometrically, this corresponds to approximating $\mathcal{T}$ with its tangent hyperplane in $\mathbf{x}_0$. The approximation will only work in a small region around $\mathbf{x}_0$, therefore,

to apply this method over a large region and retain good accuracy the region must be split in smaller parts where the linearization is performed.

In principle, one can derive the constraint functions analytically given the detector geometry, magnetic field, and material. In practice, however, these functions can be arbitrarily complicated and determining their derivatives from first principles might not be practical. A numerical method can be used to estimate the linearized constraints via the empirical covariance matrix. If we sample directly the phase space of $\mathcal{T}$ we can build this covariance matrix and study its principal components, some of which are the $\mathbf{v_i}$s of equation 6. We can do this by taking a sample of tracks (either from MC or from data), which by construction are sets of hits that belong to $\mathcal{T}$, and by computing the covariance matrix of their hit coordinates

$$\mathrm{M}_{ij} \simeq E[(\Delta x_i - E[\Delta x_i])(\Delta x_j - E[\Delta x_j])] \,. \tag{7}$$

Diagonalizing this matrix, i.e. finding the principal components, will allow to estimate the $\mathbf{v_i}$s. Note that the full correlation matrix of the coordinates contains more information than the constraints, since it has a higher dimensionality (n) than the constraints covariance matrix ($n-m$). In general, what one obtains are n eigenvectors and n corresponding eigenvalues. Of these eigenvalues m will be bigger than the others as their associated variances correspond to the dependence of the track parameters on the hit coordinates. Geometrically, their associated eigenvectors form a base on the hyperplane tangent to the surface $\mathcal{T}$ in $\mathbf{x}_0$. The remaining $n-m$ eigenvalues are smaller and correspond to the linearized constraints. Geometrically, their associated eigenvectors correspond to the coordinates perpendicular to the tangent hyperplane associated to the thickness of the surface $\mathcal{T}$ due to the finite detector resolution (uncertainties on the hit coordinates). We can take the $n-m$ eigenvectors with smaller eigenvalues, scale them to unit variance, and identify them with the $\tilde{f}_i$s. The advantages of this procedure are in its simplicity and in the possibility to perform it directly on data samples, thereby automatically accounting for possible detector misalignements and avoiding uncertainties that might affect the simulation.

With this method we can derive the coefficients $\mathbf{v_i}$ of the linearized constraints through which we can evaluate a χ^2 for each set of hits.

2.2 Linearized Track Parameters

The method described in the previous subsection naturally yields a subset of m eigenvectors that are associated to bigger eigenvalues and that can be related to the track parameters. However, these track parameters are not in a readily usable form (e.g. p_T, η, ϕ, d_0 and z_0). We want to find the transformation from the principal components of the hit coordinates to the track parameters in a usable form. We can invert, at least locally, equation 1 to express the parameters as a function of the hit coordinates

$$p_i = p_i(\mathbf{x}) \,. \tag{8}$$

We consider a linear approximation of this equation in the form

$$p_i \simeq \mathbf{w_i} \cdot (\mathbf{x} - \mathbf{x}_0) + p_i(\mathbf{x}_0) \tag{9}$$

or

$$\Delta p_i \simeq \mathbf{w_i} \cdot \mathbf{\Delta x} = \sum_{j=1}^{n} D_{ij} \Delta x_j \tag{10}$$

where we defined $\Delta p_i = p_i - p_i(\mathbf{x}_0)$ and where D_{ij} for $j \in [1, n]$ are the components of $\mathbf{w_i}$. The best coefficients D_{ij} are those that minimize the variance. We can estimate them on a sample of tracks using the least squares method and minimizing the sum of the squares of the deviations

$$\chi^2 = \sum_{ik} \left(\sum_j D_{ij} \Delta x_j^{(k)} - \Delta p_i^{(k)} \right)^2 , \tag{11}$$

where the index (k) denotes iteration on the tracks. Note that, unlike for the linearized constraints coefficients that could be derived also on data, this method can only be applied on MC because we need to know the true values of $\Delta p_i^{(k)}$

Equation 11 presents a global $\chi^2 = \sum_i \chi_i^2$ computed using the deviations for all track parameters. Each parameter is estimated through a separate linear equation and these equations are independent. Therefore, minimizing the χ^2 in equation 11 is equivalent to minimizing the χ_i^2 independently for each track parameter, and we only write it as a single value for convenience. This is important because the uncertainties between different tracks (k) can be assumed to be the same ($\sigma_{(k)} = \sigma$ for all k) and in this case the least squares expression above is equivalent to a χ^2 (if the errors are also gaussian) apart from a multiplicative factor $1/\sigma^2$ which does not affect the result of the minimization. Furthermore, the errors have expectation zero, are uncorrelated and have equal variances as they all have the same σ, therefore from the Gauss-Markov theorem [2], we know that the ordinary least squares estimator is the best linear unbiased estimator of the coefficients. In this case best means giving the lowest variance of the estimate.

To find the coefficients that minimize the χ^2 let us consider the derivative

$$\frac{1}{2} \frac{\partial \chi^2}{\partial D_{rs}} = \sum_{ik} \left(\sum_j D_{ij} \Delta x_j^{(k)} - \Delta p_i^{(k)} \right) \sum_q \delta_{ir} \delta_{qs} \Delta x_q^{(k)} = $$
$$= \sum_k \left(\sum_j D_{rj} \Delta x_j^{(k)} - \Delta p_r^{(k)} \right) \Delta x_s^{(k)} . \tag{12}$$

By solving the equation obtained by setting the derivative equal to zero we can find the best coefficients. We have

$$\sum_j D_{rj} \sum_k \Delta x_j^{(k)} \Delta x_s^{(k)} = \sum_k \Delta p_r^{(k)} \Delta x_s^{(k)} . \tag{13}$$

Given N tracks, the second half of equation 13 is $N-1$-times the element $C_{rs}^{(p)}$ in the empirical correlation matrix between track parameters and hit coordinates (where x_0 and p_0 are taken as the expected values, or the means) since

$$C_{rs}^{(p)} = \sum_{k=1}^{N} \frac{\Delta p_r^{(k)} \Delta x_s^{(k)}}{N-1} , \tag{14}$$

4

and part of the left term is $N-1$-times the empirical covariance matrix of the hit coordinates $C_{js}^{(v)}$ since

$$C_{js}^{(v)} = \sum_{k=1}^{N} \frac{\Delta x_j^{(k)} \Delta x_s^{(k)}}{N-1} \ . \tag{15}$$

We can finally write

$$\sum_j D_{rj} C_{js}^{(v)} = C_{rs}^{(p)} \ , \tag{16}$$

or in matrix form

$$D \cdot C^{(v)} = C^{(p)} \tag{17}$$

We conclude that the matrix D that performs the linear transformation from the hit coordinates to the track parameters and with coefficients that minimize the variances can be obtained from the covariance matrix of the hit coordinates and from the correlation matrix between track parameters and hit coordinates as

$$D = C^{(p)} \cdot C^{(v)-1} \ . \tag{18}$$

This result is valid independently of the form of the correlation matrix $C^{(v)}$. However, if it is diagonal, i.e. if we are in the base of the principal components, then the inverse matrix $C^{(v)-1}$ is still diagonal with elements $1/\lambda_{ii}$ on the main diagonal, where λ_{ii} are the eigenvalues of $C^{(v)}$. In this case equation 18 simplifies to

$$D_{ij} = \sum_r C_{ij}^{(v)}/\lambda_{jj} \ , \ \text{for } i,j \in [1,n] \ . \tag{19}$$

References

[1] Ashmanskas B et al. The cdf silicon vertex trigger. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 518(1–2):532 – 536, 2004. Frontier Detectors for Frontier Physics: Proceedings.

[2] Gauss and Markov. Gauss-Markov Theorem. `http://en.wikipedia.org/wiki/GaussMarkov_theorem`. [Online; accessed Nov-10-2014].

[3] Luciano Ristori and Giovanni Punzi. Triggering on heavy flavors at hadron colliders. *Annual Review of Nuclear and Particle Science*, 60(1):595–614, 2010.