Adaptive vertex fitting

2007 J. Phys. G: Nucl. Part. Phys. 34 N343

(http://iopscience.iop.org/0954-3899/34/12/N01)

J. Phys. G: Nucl. Part. Phys. **34** (2007) N343–N356 doi:10.1088/0954-3899/34/12/N01

RESEARCH NOTE FROM COLLABORATION

Adaptive vertex fitting

Wolfgang Waltenberger[1], **Rudolf Frühwirth**[1] **and Pascal Vanlaer**[2]

[1] Institute of High Energy Physics, Austrian Academy of Sciences, Vienna, Austria
[2] IIHE (ULB-VUB), Pleinlaan 2, B-1000 Brussels, Belgium

E-mail: walten@hephy.oeaw.ac.at, fru@hephy.oeaw.ac.at and pvanlaer@ulb.ac.be

Received 25 April 2007
Published 16 November 2007
Online at stacks.iop.org/JPhysG/34/N343

Abstract
Vertex fitting frequently has to deal with both mis-associated tracks and mis-measured track errors. A robust, adaptive method is presented that is able to cope with contaminated data. The method is formulated as an iterative re-weighted Kalman filter. Annealing is introduced to avoid local minima in the optimization. For the initialization of the adaptive filter a robust algorithm is presented that turns out to perform well in a wide range of applications. The tuning of the annealing schedule and of the cut-off parameter is described using simulated data from the CMS experiment. Finally, the adaptive property of the method is illustrated in two examples.

1. Introduction

The method of Least Squares is seen to be our best course when we have thrown overboard a certain portion of our data—a sort of sacrifice which has often to be made by those who sail the stormy seas of Probability.

F Y Edgeworth, 1887

Vertex fitting is the task of computing the location and the error of an interaction vertex from a given set of reconstructed tracks. A widely used method for this purpose is the Kalman filter [1, 2], which was implemented in the CMS reconstruction program ORCA [3] and is now available in the new framework CMSSW [4]. For implementation details the reader is referred to [5].

The Kalman filter is a least-squares estimator which minimizes the sum of the squared standardized distances of all tracks from the vertex position v:

$$\hat{v}_{\mathrm{LS}} = \operatorname*{argmin}_{v} L(v), \qquad \text{with} \qquad L(v) = \frac{1}{2}\sum_{i=1}^{n} \chi_i^2(v) = \frac{1}{2}\sum_{i=1}^{n} d_i^2(v)/\sigma_i^2. \tag{1}$$

Differentiation with respect to $\boldsymbol{v}$ gives the following equation for $\hat{\boldsymbol{v}}$:

$$\frac{\partial L(\boldsymbol{v})}{\partial \boldsymbol{v}} = \sum_{i=1}^{n} \chi_i(\boldsymbol{v}) \frac{\partial \chi_i}{\partial \boldsymbol{v}} = 0. \tag{2}$$

Usually the distance d_i is approximated by an affine function of $\boldsymbol{v}$ using a first-order Taylor expansion

$$d_i(\boldsymbol{v}) \approx c_i + \boldsymbol{a}_i^T \boldsymbol{v}. \tag{3}$$

Equation (2) then becomes a linear equation for $\hat{\boldsymbol{v}}$ and can be solved explicitly either globally or iteratively with the Kalman filter.

Least-squares estimators are known not to be robust, which means that they are sensitive to contaminated data, such as mis-associated tracks or mis-measured track errors. In one of the authors' PhD thesis [6] a few robustifications of the standard Kalman filter have been suggested, one of which has turned out to be a very powerful general-purpose technique: the adaptive vertex fitter (AVF). This paper deals almost exclusively with this most successful method. Techniques which have turned out to be less powerful are only hinted at; the more interested reader is referred to the aforementioned thesis. While this paper is intended to describe the method and motivate its default values, a CMS note [7] systematically compares the AVF against the classical methods.

2. The adaptive vertex fitter

The adaptive vertex fitter does not reject an outlying track; rather it down-weights the outlier with a weight w_i [8, 9]. The weight w_i depends on the compatibility of track i with the vertex, as measured by χ_i^2:

$$w_i\left(\chi_i^2\right) = \frac{\exp\left(-\chi_i^2/2T\right)}{\exp\left(-\chi_i^2/2T\right) + \exp\left(-\chi_c^2/2T\right)}. \tag{4}$$

The weight w_i can be interpreted as the probability that track i belongs to a vertex at $\boldsymbol{v}$. The constant χ_c^2 defines the threshold where the weight is equal to $1/2$; beyond this threshold a track is considered to be more likely an outlier than an inlier. The temperature T is a parameter that controls the shape of the functional dependence in equation (4). A zero temperature results in a step function and is equivalent to a hard cut at χ_c^2. Figure 1 shows the weight as a function of χ, with a cutoff at $\chi_c = 3$, for three different temperatures.

After including the weights the fit equation (equation (2)) reads

$$\sum_{i=1}^{n} w_i\left(\chi_i^2(\boldsymbol{v})\right) \chi_i(\boldsymbol{v}) \frac{\partial \chi_i}{\partial \boldsymbol{v}} = 0. \tag{5}$$

The influence of track i is now reduced by a factor $w_i\left(\chi_i^2\right)$. As the weights depend on the vertex position $\boldsymbol{v}$, an iterative procedure is required to solve equation (5). The weights are computed for an initial vertex position, and the vertex is estimated using these weights. These two steps are repeated until convergence. The resulting estimator can be regarded as an M-estimator [10], i.e. the result of minimizing an objective function of the form

$$M(\boldsymbol{v}) = \sum_{i=1}^{n} \rho(\chi_i(\boldsymbol{v})). \tag{6}$$

In the special case $\rho(\chi_i) = \chi_i^2$, the least-squares estimator is recovered. Obviously the M-estimator is a solution of the equation

$$\frac{\partial M(\boldsymbol{v})}{\partial \boldsymbol{v}} = \sum_{i=1}^{n} \psi(\chi_i(\boldsymbol{v})) \frac{\partial \chi_i}{\partial \boldsymbol{v}} = 0, \qquad \text{with} \quad \psi(\cdot) = \rho'(\cdot). \tag{7}$$

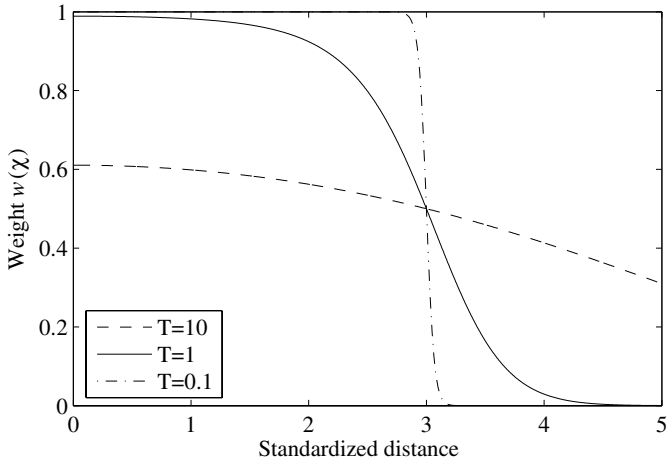

Figure 1. The weight function of equation (4) at three different temperatures.

A comparison with equation (5) shows that with our choice of weights

$$\psi(\chi) = \chi \frac{\exp(-\chi^2/2T)}{\exp(-\chi^2/2T) + \exp\left(-\chi_c^2/2T\right)}. \tag{8}$$

As ψ vanishes for the limit of large χ, the M-estimator is of the redescending type [10]. Integrating ψ over χ yields the ρ-function of the adaptive estimator:

$$\rho(\chi) = \tfrac{1}{2}\chi^2 - T \ln\left(\exp(\chi^2/2T) + \exp\left(\chi_c^2/2T\right)\right) + T \ln\left(1 + \exp\left(\chi_c^2/2T\right)\right). \tag{9}$$

The constant of integration has been chosen such that $\min \rho = 0$. Figure 2 shows the shape of the function $\rho(\chi)$ for three different values of the temperature T, with the same cut ($\chi_c = 3$) as in figure 1. If the temperature is at $T = 1$, the M-estimator is very close to a least-squares estimator for tracks within the cut, whereas for tracks beyond the cut, the contribution to the objective function is nearly constant. As a consequence, the vertex position is influenced very little by the outliers.

As mentioned before, the definition of the weights in equation (4) introduces the notion of a temperature T. This temperature can be used to employ a deterministic annealing approach that helps to avoid falling into local minima. The estimation starts at a user-defined initial temperature $T_{\text{ini}} > 1$. The temperature is then lowered in each step in a well-defined sequence that converges to 1. The iteration is stopped as soon as:

- the temperature is equal to 1 and
- the vertex candidate position has not changed by more than one micron.

The implementation of the adaptive vertex fitter method is straightforward, given a Kalman filter implementation that accounts for the notion of track weights. Details of the implementation are given in [5]. An example of an adaptive vertex fit is visualized in figure 3.

3. Test samples

All case studies described in this paper have been performed with ORCA version 8.2.0. All events are without any simulated pile-up. If not stated otherwise, the default parameters

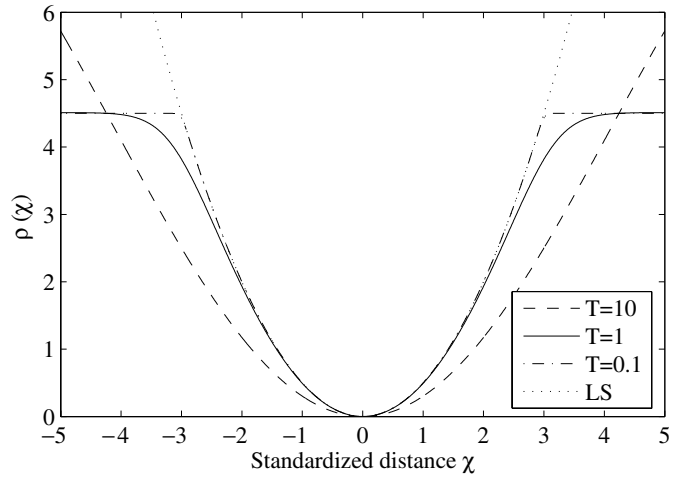

Figure 2. The function ρ of equation (9) at three different temperatures.

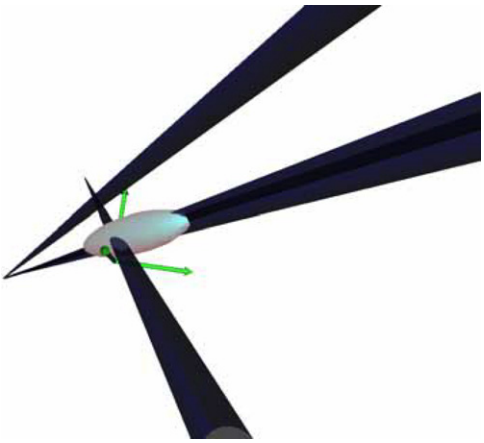

Figure 3. Result of an adaptive fit. The fitter was supplied with four tracks ($K^+K^-\mu^+\mu^-$), one of which is incompatible with the other three. Two tracks are highly collimated and appear as one in the plot. The fitter completely ignores the outlying track. The size of the ellipsoid has been multiplied by a factor of ten. The three arrows behind the vertex have a length of 100 μm in the 'ellipsoid scale' and a length of 1 mm in the 'track scale'.

of the adaptive fitter are used. Track reconstruction is performed by ORCA's default track reconstruction method. Four different kinds of event topologies are considered:

- $c\bar{c}$ jets: high multiplicity events. This topology implements a benchmark for fitting primary vertices with secondary vertices as a 'background'. The total transverse jet energy—summing over all jets—is 100 GeV or above, and the jets are in the tracker barrel region ($|\eta| < 1.4$). The primary vertex is fitted using all tracks within the jet cone found by the ORCA class `PersistentJetFinder` (with default values).
- $q\bar{q}$ jets: similar to the $c\bar{c}$ case, but there are fewer secondary vertices with fewer tracks and a higher average distance to the primary vertex. Also, the primary vertices tend to

contain more tracks. Again, the total jet transverse energy is >100 GeV, in the barrel region only. The primary vertex is, again, fitted using all tracks within the jet cone.

- $\tau \to \pi^{\pm}\pi^{\pm}\pi^{\mp}$: a 3-prong vertex that will be a good benchmark for fitting highly collimated low-multiplicity secondary vertices. Contamination comes from mis-measured tracks. Tracks matching the simulated pions from the τ-decay have been selected. The events have been obtained by producing a light MSSM Higgs $\{h^0 \to \tau^+\tau^- \to 6\pi\}$ and selecting three-prong τ decays.
- $B_s \to J/\psi\varphi \to KK\mu\mu$: a 4-prong vertex (if reconstructed correctly) that will serve as another secondary vertex fitting test case. The simulated events were pre-selected such that both muons have a $p_T > 2$ GeV/c. Again, the data contain mis-measured tracks. It does not contain mis-associated tracks.

The rest of this section will give a few more details of the event characteristics in the four test samples.

3.1. Association criteria

Association between the simulated and the reconstructed vertices has been performed on a 'by tracks' basis. This means that a reconstructed vertex is associated with the simulated vertex with which most tracks are in common. The simulated and reconstructed tracks are, in turn, associated 'by hits' using the ORCA class TrackAssociatorByHits. So, in order for a reconstructed track to be assigned properly, it has to share more hits with the correct simulated track than with any other track in the sample, independent of the absolute number of shared hits.

3.2. Event topology of $c\bar{c}$ jets

8903 events have been analysed. Figure 4 shows one such event. Associations between reconstructed and simulated objects have been performed according to subsection 3.1. In this event sample, a primary vertex has an average track multiplicity of about 14. Typically between one and three secondary simulated vertices are to be found in an event, with a total of about 6 ± 2 (reconstructed) tracks attached to them.

Note that at the generator level, the track multiplicity of a charmed meson decay is between two and three.

3.3. Event topology of $q\bar{q}$ jets

8936 events have been analysed. Compared with the $c\bar{c}$ case, the $q\bar{q}$ has a few more tracks attached to the primary vertex (on average $\approx$17 instead of 14), there are fewer or even no secondary vertices. This discrepancy comes from the fact that most $q\bar{q}$ jets are prompt decays—their secondary vertex is indistinguishable from the primary vertex.

3.4. Kinematics of $\tau \to \pi\pi\pi$

6404 events have been analysed, 5110 of which have all three π's reconstructed. The remaining 1294 events have only two reconstructed π's which were assignable to the corresponding simulated track. As with the previous cases, associators which follow the definitions given in subsection 3.1 have been employed. In order to be able to reconstruct a vertex without prior information, at least two tracks are needed. Therefore, this channel tests the low-multiplicity limits of vertex reconstruction. The events are energetic: the average p_T is around 70 GeV/c.

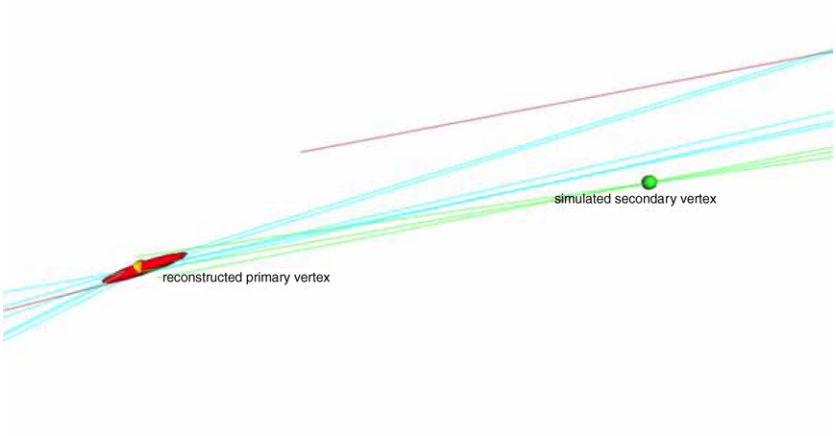

Figure 4. Snapshot of a $c\bar{c}$ event. The two 'contamination' tracks and the three secondary vertex tracks are clearly visible. The ellipsoid represents the reconstructed vertex error. For better visibility it is magnified by a factor of ten.

The opening angle between the π's is small, and consequently the vertex resolution is worse than in the $B_s \to J/\psi\varphi \to K^+K^-\mu^+\mu^-$ sample.

3.5. Kinematics of $B_s \to J/\psi\varphi \to K^+K^-\mu^+\mu^-$

9803 events have been analysed. 7451 events have all four tracks reconstructed 'correctly': all four of them are assigned to the corresponding simulated track using the standard associators. In 2088 cases one track was not reconstructed correctly in the above sense; the analysis was performed with only three reconstructed secondary tracks. In 264 cases, two tracks are missing. If one or two tracks are missing, the vertex can still be reconstructed, but a full kinematic analysis is no longer possible. The events are soft; the average total p_T of a jet is around 13 ± 6 GeV/c. The vertex resolution is therefore considerably better than in the $\tau \to \pi\pi\pi$ sample.

4. Technical aspects of the adaptive method

This section describes the various parts of the adaptive method. The technical choices that had to be taken are presented and justified.

4.1. Track (re-)linearization

No matter what track parametrization is used, a charged track in a magnetic field cannot be described exactly by a linear model. In order to deal with this nonlinearity, the exact track model is approximated by a linear model (see equation (3)). The linear expansion is recomputed if the estimated vertex has moved too far from the expansion point. For CPU performance reasons, track relinearization should only be performed when needed. The current implementation recomputes the linear approximation when the current vertex estimate moves by more than a certain threshold in the transverse plane. The default for this threshold is currently at 100 μm. Another possibility is hinted at in [11]: the definiteness of the matrix of second derivatives of the model can be used to determine whether the current estimate is

still in the domain of attraction of the global maximum. Further studies in this direction are desirable.

4.2. Initial estimate of vertex position

A robust initial estimate of the vertex location is important in the adaptive estimation. It not only defines around which points the tracks are linearized, but also the initial assignment probabilities (weights) of the tracks. If the adaptive method is interpreted as an optimization procedure, then the initial estimate can be seen as its global aspect. It is imperative that it resides close to where the global optimum of the adaptive estimate is. The method by which it is produced must therefore be robust with a high break-down point [12].

4.2.1. The default algorithm. The input for all linearization point finders is a container of reconstructed tracks. The output is a point in a three-dimensional Euclidean space. Details of the implementation are given in [5].

Many different algorithms have been tried. For the sake of brevity we shall in this note restrict ourselves to the presentation of the default method: the fraction of sample mode with weights (FSMW, [13]), and show a comparison with a few other methods that have been tried [6]. This default method is based on the *crossing points* of the tracks. A crossing point is defined as the algebraic mean of the two points of closest approach of two tracks. To a crossing point we attach a weight which is a function of the distance of the two tracks, such that a smaller distance between the two tracks gives a larger weight to their crossing point. The weight function is defined by

$$w = (d + d_{\min})^n, \tag{10}$$

where d denotes the distance between the points of closest approach. The default values are $n = -0.5$, $d_{\min} = 10\ \mu$m.

The FSMW finds the mode (point of highest density) of the crossing points, separately in each of the three spatial coordinates. Each one-dimensional mode finding starts by finding the smallest 'weighted' interval that covers at least f per cent of all data points, where f is a parameter of the algorithm. A weighted interval is defined as the length of the interval divided by the sum of all weights of the contained points. The procedure is iterative: it is recursively applied to the previously found interval, until an interval with two points remains. Finally, the mode of this particular spatial coordinate is the average of the remaining two points. Applying this iterative procedure separately to each spatial coordinate results in the final three-dimensional mode of the crossing points. The default value of f is 0.4 in the current implementation.

For performance reasons, the weights of the crossing points are ignored until the number of data points drops below a certain threshold. The default value for this threshold is 5. CPU performance was also the reason behind the decision to implement an upper limit for how many crossing points are considered. Since there is one crossing point for each track pair, their total number grows quadratically with the number of tracks. The default value of the upper limit is 400. If more track pairs are available, the most 'interesting' ones are chosen, 'interesting' being defined by

(a) using as many different tracks as possible,
(b) as high-energetic tracks as possible (the length of the full 3D track momentum vector is currently used) and
(c) mixing as much as possible high-energy tracks with low-energy tracks, as far as this is compatible with (a) and (b).

Table 1. Resolutions and failure rates of different linearization point finders (LPF) for the four test samples. The results in this table refer to the initial vertex estimate. See the text for a detailed description.

Sample	LPF	RMS (μm)	σ_{Fit} (μm)	>2 mm (‰)	t
$c\bar{c}$	ISMS	86	43	5	7.7
(primary vertex)	HSM	90	47	2	3.7
	LMS	373	63	66	3.6
	FSMW	92	49	3	4.4
$q\bar{q}$	ISMS	80	39	5	9.3
(primary vertex)	HSM	85	41	3	5.7
	LMS	358	51	88	5.6
	FSMW	88	44	3	5.8
$\tau \to \pi\pi\pi$	ISMS	630	458	146	0.1
(secondary vertex)	HSM	632	455	146	0.2
	LMS	632	455	146	0.2
	FSMW	617	436	140	0.2
$J/\psi\varphi \to K^+K^-\mu\mu$	ISMS	290	80	41	0.5
(secondary vertex)	HSM	293	84	41	0.5
	LMS	554	272	154	0.3
	FSMW	289	85	40	0.5

4.2.2. Performance analysis of the FSMW. The performance of the FSMW method has been analysed and compared against a few other algorithms. The HSM is a special case of the FSMW, with $f = 0.5$ and all weights equal to 1. The LMS, in the one-dimensional case relevant here, is the midpoint of the shortest interval containing at least half of the data points. Finally, the ISMS is a three-dimensional iterative method. A detailed description can be found in [6].

The results are summarized in table 1. The column labelled with 'RMS' denotes the RMS of the resolution plot of the z-coordinate of the initial vertex estimate. The z-coordinate is used because the differences between the various methods are particularly pronounced in this variable. 'σ_{Fit}' refers to the standard deviation of a Gaussian distribution fitted into the resolution distribution. A least-squares fit has been used, assuming Gaussian errors on the uncertainty in each bin. '>2 mm' denotes the failure to find a linearization point whose z-coordinate is within 2 mm from the true vertex position. Note that the given 'RMS' as well as 'σ_{Fit}' refer to the distributions that have been truncated according to the 'failure' criterion, i.e. at 2 mm. Finally, the column labelled 't' denotes the average time spent per event, in milliseconds, on a 2.8 GHz Intel Celeron processor.

The FSMW performs well in all scenarios and was therefore chosen as the default. The only non-iterative method, the LMS, is worse in high-multiplicity events, as it does not make full use of the information. ISMS and HSM perform very well in the high-multiplicity samples and could be used as an alternative method in this case. On the other hand, they show somewhat worse results in the τ channel. This channel is peculiar; in the case of only three data points the mode finder cannot give a precise location estimate. Also, many algorithmic differences between the methods disappear. As a matter of fact, the HSM and the LMS methods coincide—in this special case they are one and the same algorithm. Only the concept of weights (FSMW) is able to add to the quality of the estimate in this case.

The CPU consumption of all methods is acceptable, but it is evident that the ISMS algorithm scales more poorly in terms of computing time.

Table 2. Influence of the linearization point on the resolution of the final adaptive vertex fit. The results in this table refer to the final vertex estimate. See the text for further explanations.

Sample	LPF	RMS (μm)	σ_{Fit} (μm)	>2 mm (‰)	t
$c\bar{c}$	Zero	191	28	110	18.9
(primary vertex)	MC	72	30	1	12.1
	LS	78	30	11	25.4
	FSMW	72	30	2	16.7
$q\bar{q}$	Zero	194	26	115	25.1
(primary vertex)	MC	53	28	2	15.7
	LS	88	27	14	34.9
	FSMW	55	27	2	21.4
$\tau \to \pi\pi\pi$	Zero	726	663	183	2.2
(secondary vertex)	MC	591	348	106	1.9
	LS	590	354	114	4.7
	FSMW	589	352	114	1.9
$J/\psi\varphi \to K^+K^-\mu\mu$	Zero	453	158	281	6.7
(secondary vertex)	MC	260	72	29	3.8
	LS	262	72	31	8.0
	FSMW	261	72	36	4.0

4.2.3. Influence of the linearization point on the final estimate. It is interesting to study the importance of the initialization of the adaptive fitter by looking at the final resolution of the vertex fit. To this end, the default linearization point finder was compared against the three 'artificial' finders:

- a Monte Carlo-based finder that uses the simulated vertex as the fitter's initialization ('MC'),
- a finder that always returns the point (0, 0, 0) ('Zero') and
- a finder that returns the result of the linear least-squares fitting method as the linearization point ('LS').

No beam spot constraints were applied. The results, i.e. the resolution of the final fitted vertex, are given in table 2. The columns match those given in table 1. For the final fit the adaptive vertex fitter with default cut and annealing schedule was employed ($\chi_c = 3.0, T = 256, 64, 16, 4, 1, 1, \ldots$).

The table shows that the initialization indeed *does* matter. The 'Zero' linearization point finder scores poorly. The comparison between FSMW and the linear method is interesting insofar as the linear method (which itself was initialized with the FSMW method) is mathematically equivalent to starting the adaptive fitter by assigning equal weights to all tracks. It can be seen that this leads to an increase of the residual tails. Note also that in the $c\bar{c}$ sample the adaptive fit with the simple 'Zero' linearization point finder takes longer than the fit with the sophisticated FSMW. The reason is that a better linearization point speeds up the fitting procedure because fewer iterations are necessary for convergence. The comparison of FSMW with Monte Carlo indicates that there might be some space for improvement, albeit not very much.

4.3. Annealing schedule

A few annealing schedules have been tried out. Table 3 compares some of them in the $q\bar{q}$ event sample, in which the primary vertex is fitted. The results are similar in the other three

Table 3. The choice of the annealing schedule influences the result. The dots ('...') refer to geometric annealing schedules with $r = 2$. The table depicts the results for the $q\bar{q}$ sample.

Annealing schedule	RMS (μm)	σ_{Fit} (μm)	>2 mm (‰)	t
1	68	28	2	17.3
4 3 2 1	59	27	2	18.2
8 4 ...	58	27	2	21.8
32 16 ...	59	27	2	21.4
256 64 16 4 1	55	27	2	22.2
512 256 ...	61	27	3	27.5
2048 1024 ...	65	27	2	29.8
8192 4096 ...	68	27	2	33.2

topologies (see [5]). As before, no beam spot constraints were applied. Again, the RMS and the (Gaussian) fitted σ of the z-coordinate are given. The label '>2 mm' denotes the failure to reconstruct a vertex whose z-coordinate is within 2 mm from the true vertex position, including truly failed fits (in which cases exceptions were thrown). The 't' column lists the average time spent per event, given in milliseconds. The '...' refers to geometric annealing schedules with an annealing ratio $r = 2$. The time was measured on a 2.8 GHz Intel Celeron processor and an annealing schedule of $T = (256, 64, \ldots)$ has been chosen as the default.

4.4. Choosing a χ^2_c

Also a good default χ^2_c criterion needed to be found. To this end the same procedure as before has been applied: RMS, σ_{Fit}, fraction of outliers and CPU time have been evaluated as a function of χ^2_c. The result for the $c\bar{c}$ channel, again without any beam spot constraints, is shown in figure 5. Again, the results for the other channels look similar [5].

The effect is more pronounced in the $c\bar{c}$ and $q\bar{q}$ sample, because of the larger number of outliers (secondary tracks). In the samples with small track multiplicity the results hardly depend on the choice of χ^2_c. In the end a default value of $\chi_c = 3$ has been chosen.

4.5. Prior information on the vertex position

A vertex fit can also make use of a prior knowledge of the vertex. This prior information is used as a linearization point and as an additional observation with finite errors. The number of degrees of freedom of the reconstructed vertex is thus raised by three. The adaptive fitter can deal with such a prior information. One use case for this feature is to feed a fitter with the knowledge of the beam profile. This makes sense if it is known that the vertex that is to be fitted is a primary vertex. The analyses shown in this paper do not exploit any such prior information.

4.6. Failures

The adaptive vertex fitter returns a failure condition ('supplied fewer than two tracks'), if the user supplies one or no tracks. The class also returns a failure condition ('fewer than two significant tracks'), if, after the iterative fit, fewer than two significant tracks were found. Significant in this context means that the weight is above a certain threshold, which defaults to 0.01.

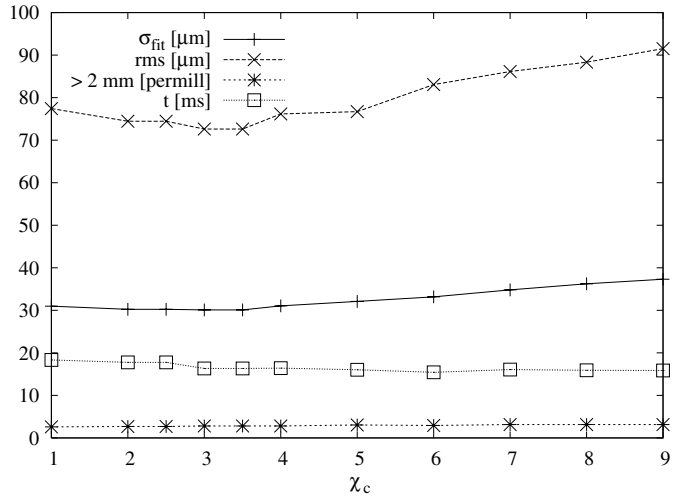

Figure 5. Fit results in the $c\bar{c}$ channel as a function of χ_c.

5. Case studies

This section is dedicated to two use cases that are intended to further illustrate the algorithmic procedure. Subsection 5.1 shows how the track weights change in each iteration step. Subsection 5.2 studies the algorithmic behaviour at its low-multiplicity limits.

5.1. Evolution of track weights in a $c\bar{c}$ event

Figure 6 shows how the track weights change in each iteration step in the adaptive method for one particular $c\bar{c}$ event, shown in figure 4. The 11 tracks from the primary vertex are contaminated with three tracks from secondary vertices, plus two more tracks that could not be associated with any vertex. It is interesting to note that in this particular topology the fitter down-weights one of the primary tracks in the beginning. Only when the bias on the fit from the mis-associated tracks decreases is the fitter capable of 'deciding for' keeping this track.

5.2. Track weights in a τ event

The adaptive method has originally been designed for high-multiplicity vertices with mis-associated tracks. It is thus interesting to study the behaviour of the method in low-multiplicity vertices. To this end we investigate how the adaptive method behaves in cases of failure of the ORCA class TrimmedKalmanVertexFitter (TKVF, see [7]), a least-squares fitter with iterative removal of incompatible tracks.

The τ-subsample for which the TrimmedKalmanVertexFitter fails contains 221 events with only two reconstructed tracks and 169 with all three tracks reconstructed correctly. Figure 7 shows the highest versus second highest track weights obtained by the default AVF, run on this τ-subsample (left-hand plot). It can be seen that in the majority of the cases, the vertex is pulled towards a single track. Already the second highest track weight is zero or close to zero in most cases. The right-hand plot of figure 7 shows how the second highest track weight $w_{(2)}$ affects the distribution of the standardized residuals of the fitted vertices' z-coordinate. For the

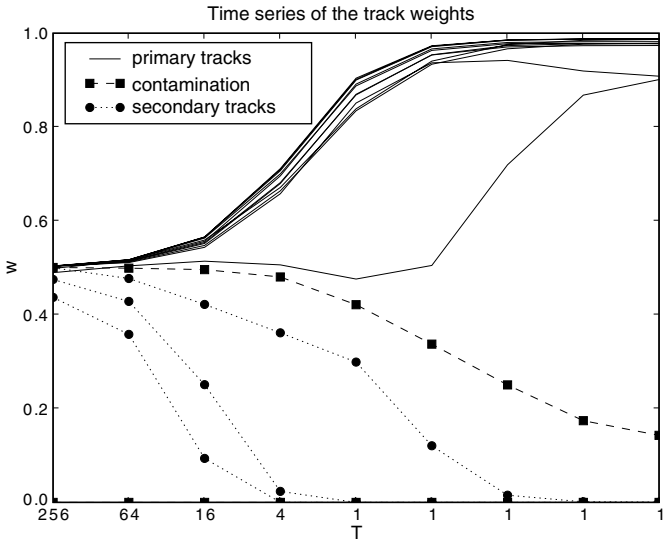

Figure 6. Evolution of the track weights of a $c\bar{c}$ primary vertex fit. One of the contamination tracks has zero weight at all temperatures.

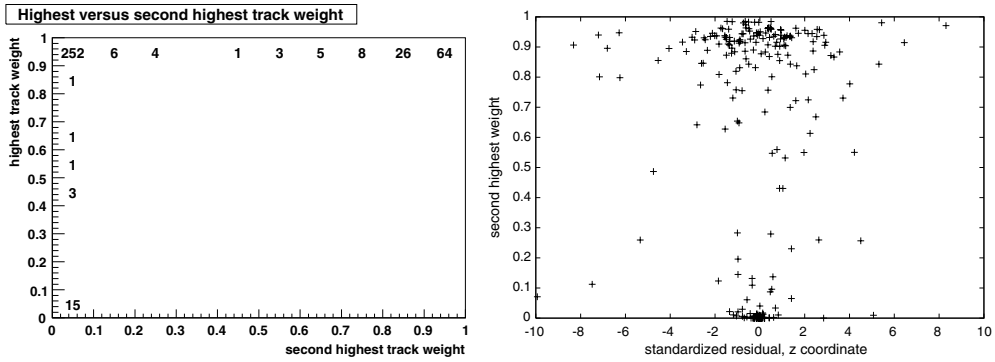

Figure 7. Track weights and standardized residuals of the AVF, in the τ-subsample for which the ORCA class TrimmedKalmanVertexFitter fails. The left plot shows the highest track weights plotted against the second highest track weights. On the right, the second highest track weight is plotted against the residuals of the z-coordinates of the vertices.

events with $w_{(2)} \approx 0$ the vertex errors tend to be overestimated and the standardized residuals cluster near 0.0. This fact can also be seen in figure 8. The pronounced peak in the left plot comes from these 'one-track' events. Introducing a cut on the second highest track weight of $w_{(2)} \geqslant 0.01$ removes the peak (right plot), at the price of throwing some events away. Finally, figure 9 repeats the plots of figure 8 (fitted with the superposition of two Gaussians), only this time the complete event sample is used. This study justifies the choice of the minimum weight for a track to contribute significantly to the vertex (see subsection 4.6).

One main advantage of the AVF over 'hard-assigning' algorithms like the TKVF is particularly visible in this example. When given three mutually incompatible tracks, the TKVF cannot but fail. Not so the AVF. Since tracks are never fully discarded in this 'soft-assigning' algorithm, a vertex can still be found. An *a posteriori* decision of what to do with the vertex can be (and is) made, based on the track weights with respect to the final vertex.

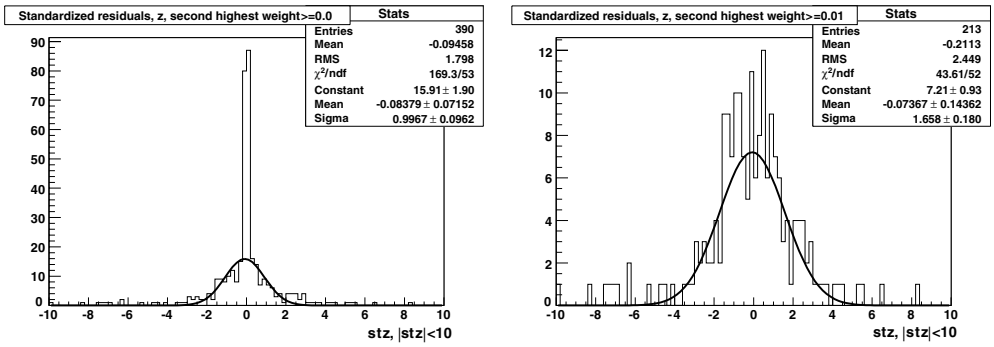

Figure 8. Standardized residuals, with all events of the τ-subsample for which the ORCA class TrimmedKalmanVertexFitter fails (left), introducing a threshold on the second highest track weight (right).

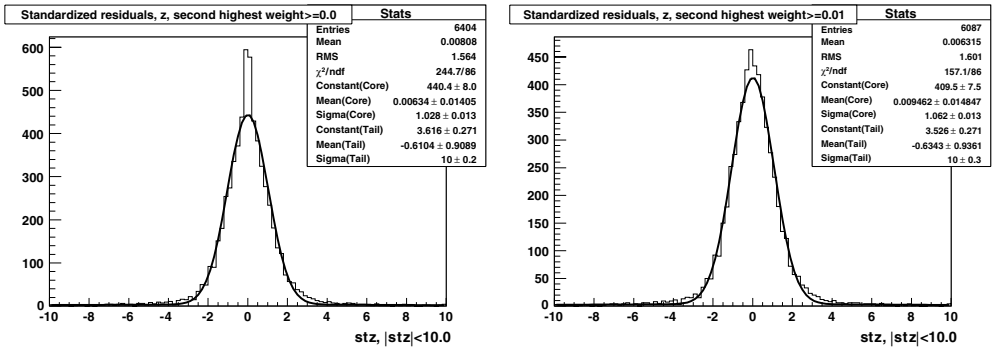

Figure 9. Standardized residuals of the AVF, for the full event sample, with (right) and without (left) a threshold on the second highest track weight.

6. Interpretation of the track weights

After fitting a vertex with a linear fitter, physicists usually discard vertices which fail a certain χ^2-probability cut. Only then one usually continues with the analysis. When using an adaptive fitter, the issue is more subtle. The χ^2-probability is not trivial to interpret; the information is now in the track weights, albeit at a more fine-grained level: a track with $w < 0.5$ is by construction an outlier; one with $w > 0.5$ is an inlier. 'Cutting' at anything other than 0.5 is discouraged; it is statistically meaningless.

So the user now implicitly defines a cut on the tracks when choosing χ^2_c. It is equivalent to cutting at a certain track's χ^2-probability, knowing that an individual track contributes two degrees of freedom to the vertex fit.

So what should one really do with the final vertex, knowing that the 'goodness of fit' information is hidden in the track weights? The authors believe that this a question of the specific use case. Consider the case of fitting $J/\psi\varphi \to KK\mu\mu$. Assume that the result of a fit is that three track weights are close to one, while the fourth weight is close to zero. The question of discarding the vertex is a question of what is relevant. If it is important that the vertex with its four daughter particles be reconstructed fully and correctly, then discarding this event is a possibility. If only the lifetime information of the mother particle is the relevant information, then the reconstructed vertex seems a perfectly legitimate candidate.

7. Summary and outlook

Let us not throw away data all too hastily. Instead, let us weigh and re-weigh the data, consider and reconsider alternative models. Only if we must, at the latest possible stage, shall we distinguish between 'in' and 'out', discriminate between signal and noise.

The authors (a formal answer to Mr Edgeworth)

The adaptive vertex fitter is a general-purpose algorithm that can be used in a very wide range of applications. Its most particular asset is the fact that no specific information on the type or level of contamination is required (see also [14]). This feature must be valued highly, considering the challenging LHC environment that has to be faced. Vertex fitting is used in a few high-level tasks such as *b*-tagging or kinematic fitting. It is not yet clear which consequences the introduction of a soft track-to-vertex association will have on such higher level code. It can be expected, though, that the extra information that is contained in the track weights can be exploited also in these parts of the analysis.

Acknowledgments

The authors would like to thank the CMS referees Danek Kotlinski and Ian Tomalin and an anonymous referee for their valuable comments and suggestions which have led to substantial improvements of the presentation.

References

[1] Frühwirth R 1987 Application of Kalman filtering to track and vertex fitting *Nucl. Instrum. Methods Phys. Res.* A **262** 444
[2] Frühwirth R, Kubinec P, Mitaroff W and Regler M 1996 Vertex reconstruction and track bundling at the LEP collider using robust algorithms *Comput. Phys. Commun.* **96** 189
[3] CMS Collaboration *ORCA, CMS OO Reconstruction* http://cmsdoc.cern.ch/orca
[4] *CMS Collaboration CMSSW, CMS SoftWare* http://cmsdoc.cern.ch/cms/cpt/Software/html/General/
[5] Frühwirth R, Waltenberger W and Vanlaer P 2007 Adaptive Vertex Fitting (*CMS NOTE* 2007-008) (Geneva: CERN)
[6] Waltenberger W 2004 Development of vertex finding and vertex fitting algorithms for CMS *PhD Thesis* TU Wien, CMS TS-2006/12. See also publications.teilchen.at/ww_diss.pdf
[7] Speer T, Prokofiev K, Frühwirth R, Waltenberger W and Vanlaer P 2006 Vertex Fitting in the CMS Tracker (*CMS NOTE* 2006-032) (Geneva: CERN)
[8] Ohlsson M, Peterson C and Yuille A 1992 Track finding with deformable templates—the elastic arms approach *Comput. Phys. Commun.* **71** 77
[9] Frühwirth R and Strandlie A 1999 Track fitting with ambiguities and noise: a study of elastic tracking and nonlinear filters *Comput. Phys. Commun.* **120** 197
[10] Hampel F R, Ronchetti E M, Rousseeuw P J and Stahel W A 1986 *Robust Statistics: The Approach Based on Influence Functions* (New York: Wiley)
[11] Eichinger H and Regler M 1981 Review of Track-Fitting Methods in Counter Experiments (*Technical Report CERN* 81-06) (Geneva: CERN)
[12] Rousseeuw P J and Leroy A M 1987 *Robust Regression and Outlier Detection* (New York: Wiley)
[13] Bickel D R and Frühwirth R 2006 On a fast, robust estimator of the mode: comparisons to other robust estimators with applications *Comput. Stat. Data Anal.* **50** 3500
[14] D'Hondt J, Frühwirth R, Vanlaer P and Waltenberger W 2004 Sensitivity of robust vertex fitting algorithms *IEEE Trans. Nucl. Sci.* **51** 2037