FILTERING, ROBUST AND ADAPTIVE METHODS FOR TRACK RECONSTRUCTION

A. Strandlie

Gjøvik University College
P. O. Box 191
N-2802 Gjøvik
Norway

ABSTRACT

The task of processing and analyzing data from experiments at the future Large Hadron Collider accelerator at CERN will be huge. Track reconstruction is one of the crucial parts of the data analysis chain in such experiments. After a brief introduction into the problem of track reconstruction, various methods for solving this task are presented. Emphasis is on state-of-the-art robust and adaptive algorithms, which mainly are applied in the most challenging experimental scenarios. Results from some simulation experiments are also presented.

1. INTRODUCTION

Track reconstruction is a vital part of the data analysis chain in a high-energy physics experiment. Starting out from a set of position measurements provided by a tracking detector situated close to the collision point of two particle beams, the goal is to find out how many particles have been created as well as their momenta and directions at the production point. Track reconstruction is traditionally divided into two separate subtasks: track finding and track fitting. Track finding is a pattern recognition task, which aims to divide the set of position measurements into several subsets. Each of these subsets contains measurements believed to come from one single particle. There is usually an additional subset containing noise or measurements from uninteresting tracks. Such a set is also often called a *track candidate*. The track fit tries – given the information of the position measurements in each of the subsets provided by the track finder – in an optimal way to estimate a set of parameters describing the state of the particle close to its production point (vertex). Such a set of parameters can for instance be momentum, direction angles and position coordinates at a reference surface.

Experiments at the Large Hadron Collider (LHC) accelerator at CERN – currently under construction for start-up in 2007 – will face unprecedented challenges for data analysis methods in general and track reconstruction algorithms in particular. Some of the algorithms have to be very fast in order to deal with a beam collision rate of 40 MHz. A very high collision rate is needed due to the fact that the really interesting particle reactions from a physics point of view have a relatively low production rate, so in most of these collisions there are no interesting reactions. In order to increase the probability of an interesting reaction, many particle collisions will occur simultaneously for each of the beam collisions. In those of the beam collisions which contain an interesting reaction there will therefore be additional non-interesting reactions. Fast algorithms will be used as a means of selecting the interesting collisions or signal events, and for these events the full information from the detectors will be directed onto mass storage. Assuming that one of about twenty simultaneous collisions correspond to an interesting reaction, most of the signals recorded by the detectors are created by particles coming from the other reactions and should therefore be considered as noise from the point of view of the track reconstruction algorithms. Therefore, track reconstruction also has to cope with a very noisy environment, in which the tracks of the signal event have to be reconstructed as precisely as possible. For this task more advanced and time-consuming algorithms will be applied. This paper will only focus on algorithms from the latter category.

2. TRACK RECONSTRUCTION ALGORITHMS

Statistical signal processing methods have for many decades been used in high-energy physics track reconstruction algorithms. Due to its attractive statistical properties, the least-squares method [1] was applied already from the time of the early bubble chamber experiments in the 1950's. A big revolution took part in the 1980's when it was realized that the Kalman filter [2] successfully could be applied to the same tasks. The Kalman filter turned out to possess some decisive advantages with respect to the standard formulation of the least-squares estimator. Due to the recursive nature of the Kalman filter, it can be used for track finding and track fitting concurrently. Also, the existence of the Kalman smoother – making it possible to obtain optimal estimates anywhere along the track – enables precise treatment of the effects of disturbance of particle trajectories due to material being present in the detectors. Due to such features, the Kalman filter is currently by far the most widely used approach for this task.

One of the challenges for the LHC experiments will be the search for particle tracks embedded in large amounts of noise due to measurements originating from particles from other, simultaneous collisions. A quite recently developed generalization of the Kalman filter – the so-called Deterministic Annealing Filter (DAF) [3] – has been shown to be superior to the Kalman filter in situations with large amounts of noise. The DAF combines the solid statistical properties of the Kalman filter with robust estimation features inspired by mean field theory from statistical physics.

Another challenge is the precise modelling of non-Gaussian effects encountered during reconstruction, such as tails of multiple Coulomb scattering, bremsstrahlung energy loss of electrons and non-Gaussian measurement uncertainties. Another generalization of the Kalman filter – the so-called Gaussian-sum filter (GSF) [4] – can treat non-Gaussian effects in a natural way and has also been shown to be superior to the Kalman filter in certain applications.

3. THE KALMAN FILTER

The state of a particle anywhere along its trajectory is described by a set of parameters, for instance two position coordinates, two direction angles and the momentum at some reference surface. This set constitutes the *state vector* in the Kalman filter terminology. The Kalman filter aims to update the estimate of the state vector recursively, as information from the position measurements along the particle track is included.

The filter proceeds by alternating prediction and update steps. In the prediction step, the state vector is propagated from one detector layer to another. This propagation can be linear or non-linear, depending on the specific solution of the equations of motion of a particle, the so-called *track model*. The characteristics of the track model depends on the configuration of the magnetic field and on the choice of track parameters. For a vanishing magnetic field the particle tracks are straight lines, leading to a truly linear track model. In a homogeneous magnetic field the track model is a helix, which for all relevant sets of track parameters yields a non-linear track model.

After a propagation step, effects of the material in the detector layer under consideration are included in the Kalman filter formalism. Such effects are typically ionization energy loss and deflection of the particle trajectory due to multiple Coulomb scattering. Ionization energy loss is usually included as a deterministic correction to the state vector, whereas multiple scattering leads to modification of a few terms in the covariance matrix of the state vector. The latter effect corresponds to *process noise* in the usual terminology of the Kalman filter.

After the addition of material effects, information from the position measurement in the sensor is included in the Kalman updator, leading to an updated estimate of the state vector. This updated state vector is propagated to the next detector layer with a measurement, and the alternating procedure is repeated until there are no more measurements left in the set given by the track finder.

There are several advantages of such a recursive approach with respect to a standard linear regression or a global least-squares fit, which takes the information from all measurements into account at once. One advantage is that the Kalman filter avoids the inversion of a possibly large covariance matrix of the measurements, which is inherent in a global least-squares fit. This covariance matrix usually is non-diagonal due to correlations between different measurements introduced by multiple scattering. Another advantage is that material effects can be taken into account locally by optimal use of the information of the measurements already included in the estimation procedure, as the updated estimate of the state vector is available at any stage of the filter. This is for instance relevant for the effects of multiple scattering, which needs an estimate of the momentum of the particle for the determination of the distribution of the deflection angle. A global least-squares fit has to use a preliminary guess of the momentum given by the track finder, which potentially can be quite crude.

The recursive nature of the Kalman filter also enables it to be applied as a track finder [5]. At a given stage of the filter, there might be several measurements in the next detector layer compatible with the prediction. In other words, given a track segment at the current stage of the filter, several measurements compete for being included in the track candidate. The simplest solution is to select the measurement closest to the predicted state and discard the others. If the density of measurements is high enough, such a procedure leads to too many premature exclusions of correct measurements and subsequently a too high fraction of lost tracks. In the general case a combinatorial approach is therefore needed, where the track candidate is split into several branches which are followed in parallel [6]. After the end of the detector is reached, the best – according to some selection criterion – of the branches are kept as the final track candidate.

The application of the Kalman filter in track reconstruction was pioneered in the DELPHI experiment at the LEP collider at CERN. Due to its many attractive features it has gained enormous popularity and is today by far the most widely used method for this application. It will also be used at the LHC and will most likely constitute the baseline algorithm for track reconstruction in the LHC experiments.

4. THE DETERMINISTIC ANNEALING FILTER

If the density of measurements in the vicinity of a track is high enough, a track finding procedure based on a combinatorial Kalman filter strategy might lead to a large number of branches to be followed in parallel. The Kalman filter is particularly sensitive to a large number of ambiguous measurements in the early phases of the track finding, where the information from the measurements already included in the track candidate can be quite poor. In such a situation it is plausible that an approach utilizing information from the entire set of measurements in the track candidate for the resolution of ambiguities can lead to a better discrimination between good and wrong measurements.

One such approach is the Elastic Arms algorithm [7]. Given a set of measurements from a track embedded in measurements from other, nearby tracks and electronic noise, it attempts to solve the combinatoric optimization problem of assigning only the correct hits to the track and discarding the rest by minimizing a global *energy function* with respect to the track parameters and a set of *binary assignment variables*. There is one such variable for each of the measurements included in the estimation procedure, i.e. for each of the measurements being reasonably close to the track. The value is one if the hit is assigned to the track and zero otherwise. The assignment variables within a detector layer are constrained to either sum up to one or to zero. If the sum is one, then one of the variables has the unit value while the rest have to be zero. If the sum is zero, all of the assignment variables have value zero. In other words, this constraint implies that either none or only one of the measurements in the detector layer under consideration are assigned to the track.

A full combinatorial exploration of all the valid configurations of the assignment variables is intractable if the density of measurements in the vicinity of the track is high. The algorithm therefore applies an approximation frequently used in condensed matter physics - so-called *mean field theory*. The various configurations are assumed to be distributed according to the Boltzmann distribution of statistical physics, meaning that configurations with low values of the energy function occur more frequently than configurations with high energy. A marginal probability density is calculated by summing over all allowed configurations of the assignment variables, defining an *effective energy* which only depends on the measurements and on the track parameters. Starting out from an initial temperature, the effective energy is minimized at successively lower temperatures, taking the zero temperature limit in the end. This procedure is called *annealing*. During the annealing, the track template is moved from its initial position towards the subset of measurements originating from the true track, while the other measurements are down-weighted. At very low temperatures, only measurements being very close to the track have an effect on the estimate of the track parameters.

One drawback of this algorithm is that the energy function to be minimized is non-quadratic, leading to a slow and cumbersome, non-linear minimization procedure. Another drawback is that the track model has to be analytic, i.e. linear, circular or helical, implying that detector material effects such as energy loss and multiple Coulomb scattering have to be neglected.

An alternative way of minimizing the energy function at a given temperature is to apply the EM algorithm [8]. It turns out that this leads to an iterative procedure of minimizing a quadratic function. The quadratic function can be minimized by any kind of least-squares estimator, for instance the Kalman filter. The resulting algorithm is called the Deterministic Annealing Filter (DAF). It has the decisive advantage over the original formulation of the algorithm that energy loss and multiple scattering correctly can be taken into account [3]. It is also significantly faster, since a non-linear minimization procedure is avoided.

The DAF was first implemented for the purpose of resolving so-called left-right ambiguities in the Transition Radiation Tracker (TRT) of the ATLAS experiment at CERN. The ATLAS TRT is a detector consisting of *drift tubes* [9]. When a charged particle traverses a drift tube, the recorded signal is the distance from the centre wire to the point of closest approach of the particle. A priori it is not known at which side of the wire the particle has passed, and two potential measurement positions have to be considered by the reconstruction algorithm. An example of a track from the ATLAS TRT and a least-squares fit giving equal weight to all measurements is shown in Fig. 1. The fit is clearly biased by the ambiguous measurements, as well as by outlying measurements not belonging to the track. The result of the DAF applied to the same track is shown in Fig. 2. The algorithm is obviously able to down-weight the outliers and resolve the ambiguities in an efficient way. Running on a large sample of simulated tracks, the standard deviations of the residuals – i.e difference between estimated and true value – of the estimated parameters of the DAF was shown to be not more than about 7 % worse than the standard deviations of the corresponding residuals obtained by a least-squares fit to only the correct measurements [3].

The DAF has also been implemented in the reconstruction software of the CMS experiment at CERN [10]. The algorithm has been systematically compared to the Kalman filter and shown to be superior in scenarios which are difficult from a track reconstruction point of view. Fig. 3 shows residuals of one of the track position coordinates close to the interaction point for narrow bundles of tracks originating from decays of b-quarks. The resolution of the DAF is seen

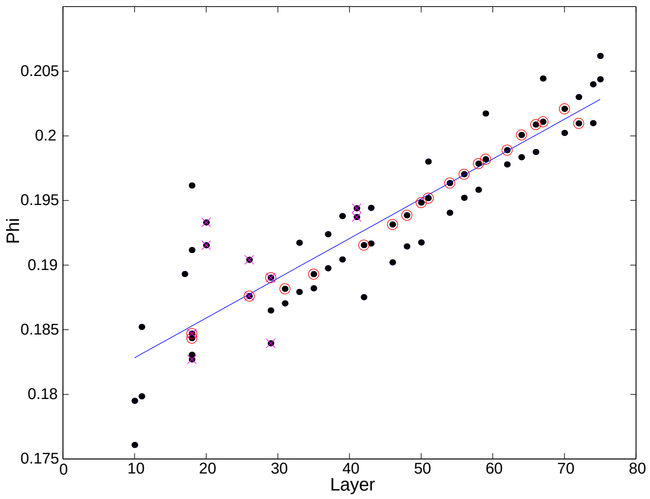

Fig. 1. Track from the ATLAS TRT and the result of a least-squares fit giving equal weight to all measurements. The track is circular, but in this projection – plotting the azimuthal angle ϕ versus layer number – it is approximately straight. The circles indicate measurements being closer than 1 mm to the estimated track.

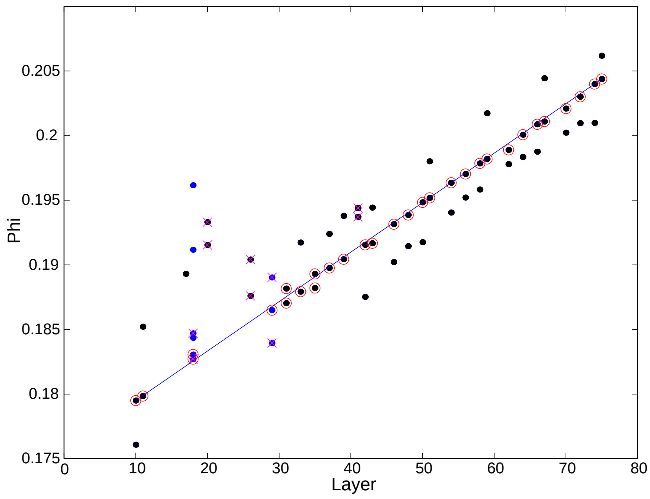

Fig. 2. Track from the ATLAS TRT and the result of a fit by the DAF. The circles indicate measurements being closer than 1 mm to the estimated track.

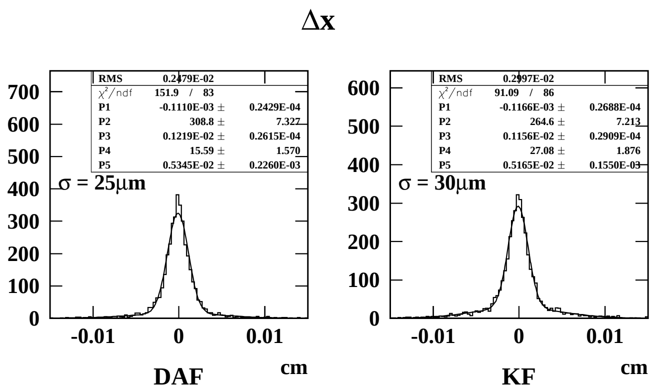

Fig. 3. Residuals of estimated position coordinate at a reference surface close to the interaction vertex.

to be clearly better than that of the Kalman filter. In Fig. 4, a histogram of the so-called χ^2 probability is shown. This quantity is the probability transform of the sum of squared, normalized distances between the measurement positions and the fitted track positions. If the covariance matrix of the track parameters is estimated correctly, and if the hits are correctly assigned to the tracks, the histogram should be reasonably flat. Even though none of the histograms exhibit perfect behaviour, the one of the DAF is clearly more compatible with a flat distribution than the one of the Kalman filter.

5. THE GAUSSIAN-SUM FILTER

The Kalman filter is known to be optimal when all probability densities involved in the estimation procedure are Gaussian. This is because the Kalman filter is a least-squares estimator. When not all of the involved densities are Gaussian, it is plausible that a non-linear estimator which better takes the actual shape of the distributions into account can do better. One such estimator is the Gaussian-sum filter (GSF) [4], which is adequate when the probability densities can be described or approximated by Gaussian mixtures. We recall that a Gaussian mixture is a weighted sum of single Gaussians. The state vector of the GSF also becomes a Gaussian mixture, described by a set of mean values, covariance matrices and weights. An attractive feature of the GSF is that it resembles several Kalman filters running in parallel, implying that the backbone of this non-linear estimator is a linear procedure.

Like the Kalman filter, the GSF works by alternating prediction and update steps. After a prediction step, material effects such as multiple scattering and energy loss are included. Inclusion of material effects amounts to a convolution of the probability densities, which leads to a multiplication of the number of components in the state vector mixture, if material effects are modelled by Gaussian mixtures. If measurement noise also is modelled by a Gaussian mixture, there is a similar multiplication of the number of components during the update procedure. The component weights are updated in a non-linear manner, depending on the distance between the predicted components and the measurement.

If there are more than just a few instances of component multiplication, the number of components in the state vector mixture quickly becomes high and the filter too slow

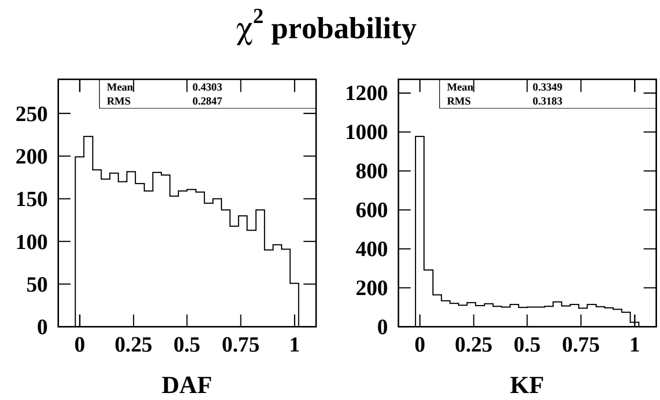

Fig. 4. The probability transform for the track χ^2 of the DAF and of the Kalman filter.

for most practical purposes. A procedure for limiting the maximum number of components in the filter is therefore needed. Such a limiting procedure should be performed by minimal loss of the information inherent in the state vector mixture. Successful approaches for some applications attempt to merge components being close to one another according to some distance measure. A simpler procedure is to drop components with weights smaller than a given cutoff. The latter approach has to be used carefully, however, since it for some applications leads to an undesirable deterioration of the statistical properties of the filter.

The GSF has been implemented for the treatment of bremsstrahlung energy loss effects in the reconstruction of electron tracks in the CMS experiment at CERN [11]. The bremsstrahlung distribution of relative energy loss in a material layer – i.e energy after an electron has traversed the layer divided by the incident energy – is highly non-Gaussian. A plot of an approximation of this distribution by Bethe and Heitler is shown in Fig. 5. A six-component Gaussian-mixture approximation of the Bethe-Heitler distribution has been used in the GSF. The GSF yields a better precision of the estimate of the momentum of the electron than the Kalman filter, as shown in Fig. 6. The estimated state vector of the GSF also gives a better representation of the actual distribution of the track parameters than the Kalman filter. In Fig. 7, the probability transform of the estimated q/p (charge divided by momentum) of the GSF and of the Kalman filter are shown. Each entry in this histogram amounts to the cumulative distribution function of the estimated state vector at the true value of the track parameter. If the distribution of the estimated state vector is close to the correct distribution, this histogram should be reasonably flat. In contrast to the GSF, the distribution of the state vector for the Kalman filter is a single Gaussian. The results presented in this section come from a simulation experiment, so the true values of the track parameters are known.

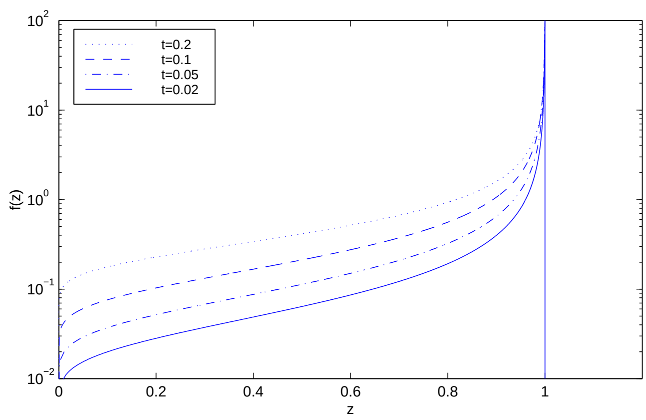

Fig. 5. Bethe-Heitler model of relative energy loss in a material layer, for a set of thicknesses t of the layer. The value of t is given in units of a quantity called the *radiation length*.

6. SUMMARY AND OUTLOOK

In this paper, a set of algorithms used for the reconstruction of tracks of charged particles in modern high-energy physics experiments have been presented. For the most difficult experimental conditions, robust and adaptive algorithms have been shown to be superior to the Kalman filter.

Even though the LHC is not operational before 2007, one already plans an upgrade of the accelerator. Such an upgrade will lead to conditions with even more simultaneous collisions in the detectors of the experiments, and it is still not clear whether current track reconstruction algorithms will be able to cope with the increased noise level at a future, upgraded LHC. More research is clearly needed in order gain a quantitative understanding of the dimension of these problems, and new algorithms possibly have to be derived for their solutions.

7. REFERENCES

[1] R. Frühwirth et al., “Data Analysis Techniques in High-Energy Physics”. Cambridge University Press, 2000.

[2] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting”. Nuclear Instruments and Methods in Physics Research **A262** (1987) 444.

[3] R. Frühwirth and A. Strandlie, “Track fitting with ambiguities and noise: a study of elastic tracking and non-linear filters”. Computer Physics Communications **120** (1999) 197.

[4] R. Frühwirth, “Track fitting with non-Gaussian noise”. Computer Physics Communications **100** (1997) 1.

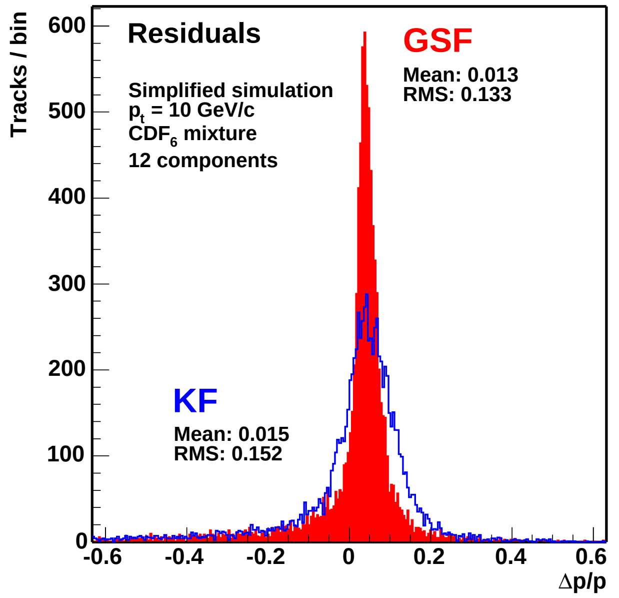

Fig. 6. Residuals of estimated momentum – relative to the absolute value of the momentum – for the GSF and for the Kalman filter.

[5] P. Billoir, "Progressive track recognition with a Kalman-like fitting procedure". Computer Physics Communications **57** (1989) 390.

[6] R. Mankel, "A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system". Nuclear Instruments and Methods in Physics Research **A395** (1997) 169.

[7] M. Ohlsson, C. Peterson and A. Yuille, "Track finding with deformable templates – the elastic arms approach". Computer Physics Communications **71** (1992) 77.

[8] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm". Journal of the Royal Statistical Society **B 39** (1977) 1.

[9] ATLAS Inner Detector Technical Design Report. CERN/LHCC 97-16, CERN, Geneva, 1997.

[10] M. Winkler, "A comparative study of track reconstruction methods in the context of CMS physics". PhD Thesis, CMS Thesis 2002-015/Tracking, Vienna University of Technology and CERN, 2002.

[11] W. Adam, R. Frühwirth, A. Strandlie and T. Todorov, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC". CMS Note 2005/001, CERN, Geneva, 2005.

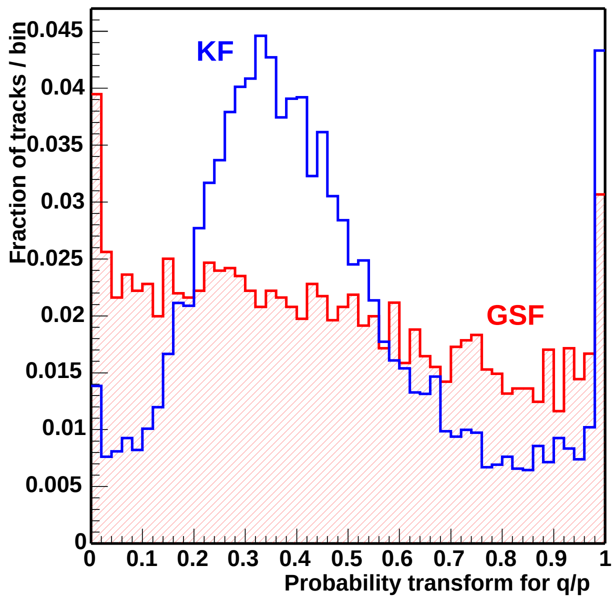

Fig. 7. Probability transform for q/p for the GSF and for the Kalman filter.