Particle tracks fitted on the Riemann sphere

A. Strandlie [a,b,1], J. Wroldsen [a,2], R. Frühwirth [c,3], B. Lillekjendlie [a,d,4]

[a] *Faculty of Technology, Gjøvik College, P.O. box 191, N-2802 Gjøvik, Norway*

[b] *Department of Physics, University of Oslo, Oslo, Norway*

[c] *Institute for High Energy Physics of the Austrian Academy of Sciences, Vienna, Austria*

[d] *SINTEF, Oslo, Norway*

Abstract

We present a novel method of fitting trajectories of charged particles in high-energy physics particle detectors. The method fits a circular arc to two-dimensional measurements by mapping the measurements onto the Riemann sphere and fitting a plane to the transformed coordinates of the measurements. In this way, the non-linear task of circle fitting, which in general requires the application of some iterative procedure, is turned into a linear problem which can be solved in a fast, direct and non-iterative manner. We illustrate the usefulness of our approach by stating results from two simulation experiments of tracks from the ATLAS Inner Detector Transition Radiation Tracker (TRT). The first experiment shows that with a significantly lower execution time, the accuracy of the estimated track parameters is virtually as good as the accuracy obtained by applying an optimal, non-linear least-squares procedure. The second experiment focuses on track parameter estimation in the presence of ambiguous measurements. For this purpose, we have developed a new version of the Elastic Arms algorithm called the Elastic Planes algorithm. The algorithm produces results which are almost identical to the results from an optimal version of the Elastic Arms algorithm. The computational cost of our algorithm, however, is much lower.

---

[1] E-mail: `Are.Strandlie@hig.no`
[2] E-mail: `Joern.Wroldsen@hig.no`
[3] E-mail: `fruhwirth@hephy.oeaw.ac.at`
[4] E-mail: `Bjorn.Lillekjendlie@staff.sintef.no`

1 Introduction

The task of fitting circular arcs to a set of measurements is of vital importance in high-energy physics experiments. This is due to the fact that many of the tracking systems in collider experiments are put into relatively homogeneous magnetic fields, and the trajectories of charged particles will therefore be helix curves, or in the bending plane of the particles, circles. Since the particle momenta are directly connected to the curvature of these circles, it is important to be able to make reliable estimates of the circle parameters.

If the measurement errors can be adequately described by a Gaussian noise model, the method of least-squares is well-known to be the optimal one. For circular arcs, however, it is in general not possible to express the measurements as linear functions of the parameters describing the circle. This means that the simple and powerful tools of linear least-squares regression not can be used directly, and one has to adopt some strategy based on non-linear least-squares methods. Among the most popular of these are the Gauss-Newton algorithm and the closely related Levenberg-Marquardt algorithm. These basically work by expanding the function describing the relation between the circle parameters and the measurements in a Taylor series and collecting terms up to first order. This enables the application of linear regression methods, and the estimated parameters are used as a new expansion point of the function. The procedure is repeated until some convergence criterion is fulfilled. If there are no local minima in parameter space, this method will converge to the correct solution. However, due to the iterative nature of the procedure the computational cost can be quite large. In addition, one has to supply the algorithm with an initial guess of the circle parameters, which means that some preprocessing of the data has to be done.

A more simple and faster approach would be to go through only one step in the above iteration procedure and keep the outcome of the first step as the final estimate of the parameters. The quality of the estimate will of course depend strongly on the extent of which it is justified to keep only the first order terms of the series expansion. With a suitable choice of circle parameters, this approximation can in fact be very good. However, the initial guess of the parameters has to be reasonably close to the correct values. The need of preprocessing of the data is therefore still present.

We present in this paper a novel method of track fitting based on mapping two-dimensional measurements onto the Riemann sphere [1]. It is well-known from complex analysis that circles and lines in the plane uniquely map onto circles on the Riemann sphere [2]. Since a circle on the Riemann sphere uniquely defines a plane in space, there is a one-to-one correspondence between circles and lines in the plane and planes in space. The problem of fitting measurements to a circle in the plane is therefore changed into the problem of fitting the transformed measurements to a plane in space. This can be done in a fast and direct way. In contrast to the Gauss-Newton method, our algorithm does not require any iterations. There is also no need for an initial guess of the track parameters, something which enables our method to work without any preprocessing of the data. This is in our opinion is a decisive advantage over all the methods mentioned above.

We will by means of a simulation experiment of tracks in the ATLAS Inner Detector TRT show that our algorithm compares favourably with other methods - such as a non-linear least-squares fit and a linear, global least-squares fit - when it concerns computational speed. The accuracy of the estimates will also be shown to be negligibly worse than what is achieved with an optimal, non-linear procedure. Another experiment, which focuses on track fitting in the presence of ambiguous measurements, requires the use of sophisticated, adaptive methods as for instance the Elastic Arms algorithm [3]. It has recently been pointed out that the Elastic Arms algorithm can be formulated as an iteratively reweighted least-squares procedure [4]. We propose to do the fitting part of this procedure on the Riemann sphere, and the novel algorithm naturally arising from this idea will be called the Elastic Planes algorithm. The algorithm will be shown to be virtually as accurate as an optimal version of the original Elastic Arms algorithm, and it is significantly faster.

The paper will be organized as follows. In Section 2 we present the basic concepts of track fitting on the Riemann sphere. We introduce in Section 3 some modifications of the ideas from the original paper [1] in order to optimize the performance of the algorithm. The Elastic Planes algorithm is derived in Section 4. In Section 5 we state results from the different simulation experiments, and in Section 6 we draw conclusions and give some ideas of possible future work.

2 Basic concepts of track fitting on the Riemann sphere

A circle in the plane is in general described by a set of three parameters. This can, for instance, be the radius of curvature ρ and the Cartesian coordinates (u_0, v_0) of the centre point. We assume that the points $P = \{P_i\}$ to be fitted are given in polar coordinates: $P_i = (R_i, \phi_i)$, where $i = 1, \dots, N$. The transformation formulas for a point (R_i, ϕ_i) in the plane to a point (x_i, y_i, z_i) on the Riemann sphere [5] are given by

[5] The Riemann sphere has radius $1/2$ and centre coordinates $(0, 0, 1/2)$. The north pole of the sphere is therefore in the point $(0, 0, 1)$, and the south pole is in $(0, 0, 0)$.

$$x_i = R_i \cos\phi_i/(1 + R_i^2), \tag{1}$$
$$y_i = R_i \sin\phi_i/(1 + R_i^2), \tag{2}$$
$$z_i = R_i^2/(1 + R_i^2). \tag{3}$$

The transformed point can also be found in the following way. Consider the straight line going through $P_i = (R_i, \phi_i)$ and the north pole of the Riemann sphere. The intersection between this line and the Riemann sphere defines the transformed point.

A plane in space is in general described by four parameters $\{c, n_1, n_2, n_3\}$, where $\boldsymbol{n}^T = (n_1, n_2, n_3)$ is a unit length normal vector of the plane and c is a signed distance from the plane to the origin. All points (x, y, z) in space satisfying

$$c + n_1 x + n_2 y + n_3 z = 0 \tag{4}$$

are lying in this plane. In order to map a circle in the plane onto the Riemann sphere, one first picks three points on the circle. These points are mapped onto the sphere by Equations (1), (2) and (3), and the transformed points are used to form two non-parallel vectors in space. The normalized cross product of these two vectors is the normal vector of the plane, and c is then easily found by insertion into Equation (4).

Fitting a plane in space to the measurements on the Riemann sphere will be defined as the minimum of

$$S = \sum_{i=1}^{N} (c + n_1 x_i + n_2 y_i + n_3 z_i)^2 = \sum_{i=1}^{N} d_i^2 \tag{5}$$

with respect to c, n_1, n_2 and n_3, subject to the constraint $n_1^2 + n_2^2 + n_3^2 = 1$. In other words, we want to minimize the sum of squared distances d_i^2 from the points to the plane. We start by finding the minimum of S with respect to c. This is straightforwardly done by solving $\partial S/\partial c = 0$, and this gives

$$c = -\boldsymbol{n}^T \overline{\boldsymbol{r}}, \tag{6}$$

with $\overline{\boldsymbol{r}}^T = (\overline{x}, \overline{y}, \overline{z})$. Here $\overline{\boldsymbol{r}}$ can be interpreted as the mean vector of the data, with $\overline{x} = \sum_i x_i/N$, $\overline{y} = \sum_i y_i/N$ and $\overline{z} = \sum_i z_i/N$. Omitting an unimportant constant factor N, the cost function S can now be written

$$S = \boldsymbol{n}^T \boldsymbol{A} \boldsymbol{n}, \tag{7}$$

where

$$\boldsymbol{A} = \frac{1}{N} \cdot \sum_{i=1}^{N} (\boldsymbol{r}_i - \overline{\boldsymbol{r}}) (\boldsymbol{r}_i - \overline{\boldsymbol{r}})^T, \tag{8}$$

and $\boldsymbol{r}_i^T = (x_i, y_i, z_i)$. The matrix $\boldsymbol{A}$ can be recognized as the sample covariance matrix of the measurements. We now want to find the minimum of S with respect to $\boldsymbol{n}$. In order to solve this task, we begin by expressing $\boldsymbol{n}$ in the basis consisting of the eigenvectors $\{\boldsymbol{u}_j\}$ of $\boldsymbol{A}$: $\boldsymbol{n} = \sum_{j=1}^{3} a_j \boldsymbol{u}_j$. Since $\boldsymbol{A}$ is a symmetric matrix, it can be expressed as $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$ [5], where the columns of $\boldsymbol{U}$ consist of the eigenvectors $\{\boldsymbol{u}_j\}$ and $\boldsymbol{D}$ is a matrix with the eigenvalues $\{\lambda_j\}$ of $\boldsymbol{A}$ on the main diagonal and zeros otherwise. This gives

$$S = \sum_{j=1}^{3} a_j^2 \cdot \lambda_j, \tag{9}$$

with $\sum_{j=1}^{3} a_j^2 = 1$. From this expression it is easy to see that the minimum of S is obtained by choosing $\boldsymbol{n}$ to be the eigenvector corresponding to the smallest eigenvalue of $\boldsymbol{A}$. Thus, the minimization problem basically reduces to finding the eigenvalues and eigenvectors of a 3×3 matrix, something which in principle can be done analytically. The solution can therefore be found in a simple and direct manner. After having determined the normal vector $\boldsymbol{n}$, c is directly given by insertion into Equation (6).

The approach of finding the eigenvectors and eigenvalues of the covariance matrix of the data is also called Principal Component Analysis (PCA) [6], a technique often used in the fields of Pattern Recognition and Image Analysis. In this language, the above solution is the principal vector corresponding to the least principal value. It can be shown [6] that there is no direction in space in which the projected measurements have a smaller variance than the direction of this principal vector. It should come as no surprise that the criterion of smallest variance is exactly the same as the minimization of the cost function given in Equation (5).

The mapping from the parameters of the plane to the parameters of the circle is easily derived by expressing Equations (1), (2) and (3) in Cartesian coordinates (u, v) and inserting these into Equation (4). This gives an expression which is quadratic in u and v. Knowing that the general formula of a circle is $(u - u_0)^2 + (v - v_0)^2 = \rho^2$, the centre coordinates and radius of curvature can be identified as

$$u_0 = -\frac{n_1}{2\,(c + n_3)}, \tag{10}$$

$$v_0 = -\frac{n_2}{2\,(c + n_3)}, \tag{11}$$

$$\rho^2 = \frac{n_1^2 + n_2^2 - 4c\,(c + n_3)}{4\,(c + n_3)^2}. \tag{12}$$

Note that there is a singularity for $c = -n_3$. This only happens in the case where the plane is going through the north pole of the Riemann sphere. Planes going through the north pole of the sphere corresponds to straight lines in the plane, and for straight lines the above parameterization has no meaning. However, in the case where c is very close or equal to $-n_3$ it is possible to map the parameters of the plane to another set of circle parameters, say, the curvature, the angle between the tangent to the track and the x-axis, and the distance from the origin at the point of closest approach to the origin. This mapping is well-defined for both circles and straight lines. We have tested out both methods on the track sample used in the experiments described later on, and numerically the results have in this case turned out to be totally equivalent.

Even though we know the centre coordinates and the radius of curvature, the information about the orientation of the circle is lacking. This might be relevant for instance in a collider experiment when tracking reasonably high-energetic particles through a tracking detector. For such particles, the measurements are positioned along only a minor fraction of the circle, and it is therefore meaningful to speak about a track direction. The orientation of the fitted circle can be defined to depend on the direction of the normal vector. Since the centre coordinates and the radius of curvature are invariant with respect to a transformation $\boldsymbol{n} \to -\boldsymbol{n}$ (c is automatically changing sign when $\boldsymbol{n} \to -\boldsymbol{n}$), we have two possible normal vectors for each circle, corresponding to the two different orientations of the circle. One possible convention is to say that a positive sign of n_3 corresponds to a counterclockwise orientation of the circle.

In other situations, as for instance in determining the parameters of a circular pattern in a RICH detector, the measurements are in principle spread around the whole circle. The method of fitting presented herein can be used also in such a case, but for this application it does not make sense to define an orientation of the circle.

The algorithm for track fitting on the Riemann sphere will then be as follows.

(1) Map the points of the track candidate onto the Riemann sphere by Equations (1), (2) and (3).
(2) Obtain the parameters $\{c, n_1, n_2, n_3\}$ as the solution of the constrained minimization problem described above.
(3) Map the parameters of the plane back to the circle parameters (u_0, v_0, ρ) by aid of Equations (10), (11) and (12).
(4) If the application requires, determine the orientation of the circle from the direction of the normal vector.

3 Modifications of the original algorithm

We will in Section 5 state results from simulation experiments of high-energetic tracks from a drift tube detector: the ATLAS Inner Detector TRT. In the barrel part of the ATLAS TRT, the detector elements are arranged in cylindrical layers, and in our simulations the magnitude of the measurement error in $R\phi$ is constant throughout the detector. This is of course only approximately correct when comparing with a real drift tube. The drift distance given by such a detector element will be the shortest distance from the track to the straw centre. For high-energetic tracks this distance is almost the same as the distance in $R\phi$, but there is a small difference for particles with lower energy.

The distance in $R\phi$ between the two points (R_i, ϕ_i) and $(R_i, \phi_i + \Delta\phi_i)$ is of course $R_i \Delta\phi_i$. The corresponding distance between the transformed points on the Riemann sphere is

$$r_i \Delta\phi_i = \frac{R_i \Delta\phi_i}{1 + R_i^2}, \tag{13}$$

where $r_i = \sqrt{x_i^2 + y_i^2}$. This can be seen by using the expressions from Equations (1) and (2) for x_i and y_i and by realizing that the angle ϕ_i in the plane maps onto the same angle in the parallel plane containing the two transformed points, i.e. the plane defined by the equation $z = R_i^2 / \left(1 + R_i^2\right)$. In other terms, a distance in $R\phi$ in the plane maps onto a distance in $r\phi$ on the Riemann sphere with a radius dependent scaling factor $\left(1 + R_i^2\right)^{-1}$. Even though the Gauss-Markov conditions [7] are fulfilled for the measurements in the plane, this is not the case for the transformed measurements, and the distances on the Riemann sphere have to be scaled with the above mentioned scaling factor in order to compensate for this. Moreover, since high-energy tracks map onto planes being relatively vertical, the distances from the points to the plane can be assumed to not deviate much from the distances in $r\phi$. Thus, the following modified cost function

$$S_{\text{mod}} = \sum_{i=1}^{N} \left(1 + R_i^2\right)^2 \cdot d_i^2 = \sum_{i=1}^{N} p_i \cdot d_i^2, \tag{14}$$

can be hoped to reasonably well fulfill the Gauss-Markov conditions. The simulation results from Section 5 will show that this is indeed the case.

The minimization of the above cost function with respect to the plane parameters c, n_1, n_2 and n_3 is a weighted least-squares problem. Even though the direct PCA analogy does not exist for the weighted case, it can be treated along exactly the same lines as the non-weighted least-squares problem of

Section 2. Firstly, we try to solve $\partial S_{\text{mod}}/\partial c = 0$. This gives

$$c = -\boldsymbol{n}^T \boldsymbol{r}_{\text{cg}}, \tag{15}$$

with the "centre of gravity"-vector $\boldsymbol{r}_{\text{cg}}^T = (x_{\text{cg}}, y_{\text{cg}}, z_{\text{cg}})$, where

$$x_{\text{cg}} = \frac{\sum_{i=1}^N p_i x_i}{\sum_{j=1}^N p_j}, \tag{16}$$

$$y_{\text{cg}} = \frac{\sum_{i=1}^N p_i y_i}{\sum_{j=1}^N p_j}, \tag{17}$$

$$z_{\text{cg}} = \frac{\sum_{i=1}^N p_i z_i}{\sum_{j=1}^N p_j}. \tag{18}$$

The cost function can then be written (again neglecting a constant factor N)

$$S_{\text{mod}} = \boldsymbol{n}^T \boldsymbol{A}_{\text{w}} \boldsymbol{n}, \tag{19}$$

where the weighted covariance matrix $\boldsymbol{A}_{\text{w}}$ is given as

$$\boldsymbol{A}_{\text{w}} = \frac{1}{N} \cdot \sum_{i=1}^N p_i \left(\boldsymbol{r}_i - \boldsymbol{r}_{\text{cg}}\right) \left(\boldsymbol{r}_i - \boldsymbol{r}_{\text{cg}}\right)^T. \tag{20}$$

As before, the vector $\boldsymbol{n}$ minimizing the above cost function is the eigenvector corresponding to the smallest eigenvalue of $\boldsymbol{A}_{\text{w}}$. With the normal vector given, c is determined from Equation (15).

It is possible to support the reasoning leading to the structure of the above modified cost function with some analytical results. Given a circle in the plane and a point (R_i, ϕ_i), we would like to derive a relation between the distance from the point to the circle and the distance d_i from the mapped point (x_i, y_i, z_i) to the corresponding plane in space (parameterized by the set $\{c, n_1, n_2, n_3\}$). By applying the transformation formulas given in Equations (1), (2) and (3) and the expression of d_i given in Equation (5), we end up with the following result:

$$d_{i,R\phi} \cdot \cos\left(\arcsin\left(\frac{1}{2}\kappa R_i\right)\right) = \frac{d_i \cdot (1 + R_i^2)}{\sqrt{1 - n_3^2}}. \tag{21}$$

Here $d_{i,R\phi}$ is the distance in $R\phi$ from the point to the circle, while κ denotes the curvature of the circle. The above expression is correct to first order series expansions of $\sin(d_{i,R\phi})$ and $\cos(d_{i,R\phi})$, which is a very good approximation for reasonably small $d_{i,R\phi}$. It is also assumed that the impact parameter of the track is not too large. The expression on the left-hand side of Equation (21) can by trigonometric considerations be seen to be equal to the orthogonal or the shortest distance $d_{i,\perp}$ from the point (R_i, ϕ_i) to the circle. See Figure 1 for a schematic drawing of the circle and the different types of distances. We can then write

$$d_{i,\perp} = \frac{d_i \cdot (1 + R_i^2)}{\sqrt{1 - n_3^2}}, \tag{22}$$

so minimizing S_{mod} is in fact equivalent to minimizing the sum of squared orthogonal distances from the points to the circle. (The square root in the denominator of Equation (22) would be the same for all terms in a cost function consisting of a sum of squared orthogonal distances, and the square root can therefore be omitted.) When dealing with real, experimental data minimizing these orthogonal distances is exactly what we would like to do.

We are now in the position of being able to define a χ^2 of the track. For a real

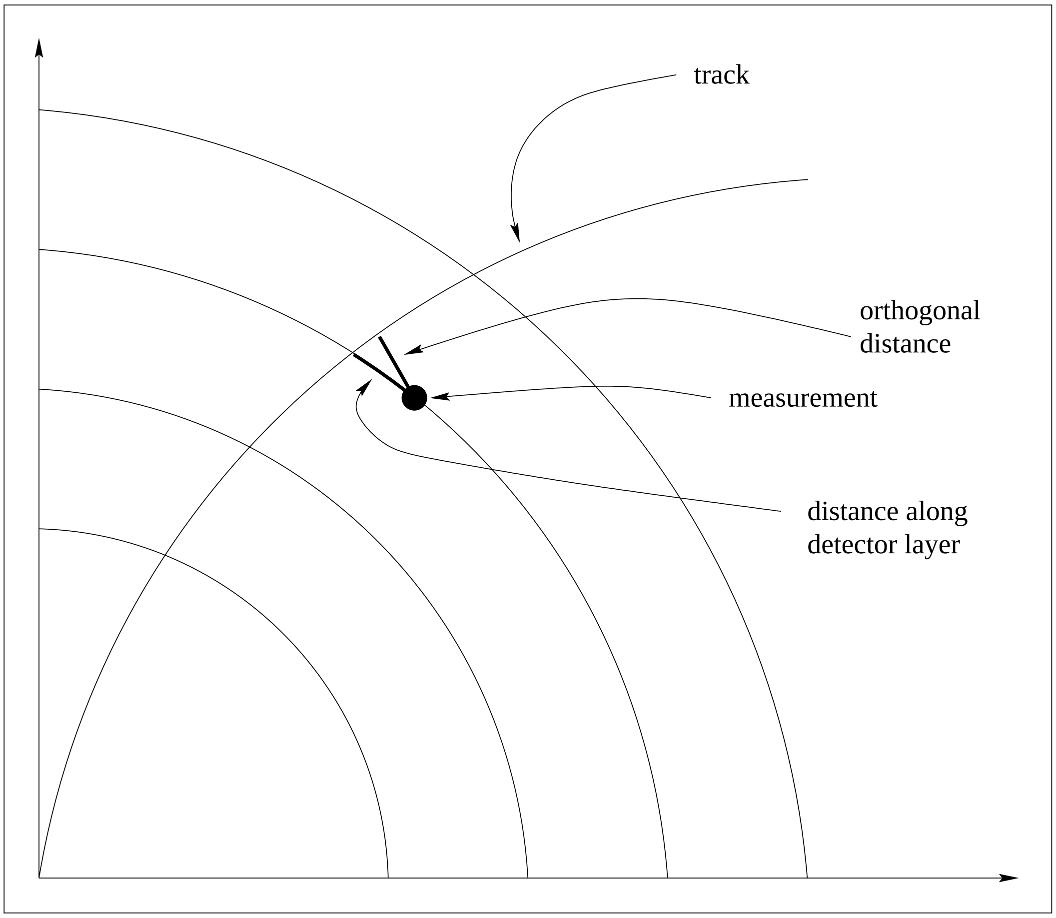


Fig. 1. This is a simplified drawing of a track going through a detector consisting of cylindrical detector layers. A measurement i and the distances $d_{i,R\phi}$ and $d_{i,\perp}$ from the measurement to the track are indicated.

drift tube, the quantity

$$\chi_{\perp}^2 = \sum_{i=1}^{N} \left(\frac{d_{i,\perp}}{\sigma} \right)^2 \tag{23}$$

should be approximately χ^2 distributed with $N-3$ degrees of freedom. The quantity σ denotes the measurement error of the drift tube. In the simulation experiments of this paper, however, we encounter the simplified situation of a measurement error being constant in magnitude in $R\phi$. Thus, for our specific case the following, modified χ^2,

$$\chi_{R\phi}^2 = \sum_{i=1}^{N} \left(\frac{d_{i,R\phi}}{\sigma} \right)^2 , \tag{24}$$

should follow the χ^2 distribution even more closely. The validity of these statements will be confirmed by the results of Section 5.

4 The Elastic Planes algorithm

In a study [4] of track fitting in the presence of ambiguous measurements and noise, the Elastic Arms algorithm (EA) has been shown to be a very powerful tool. However, in order to achieve the optimal performance of the algorithm, it is necessary to apply advanced methods of unconstrained optimization for the minimization of the effective energy. This makes the EA quite time consuming. In the single-track case, the effective energy is given as

$$E_{\text{eff}} = -\frac{1}{\beta} \sum_k \log \left(n_k \cdot e^{-\beta\lambda} + \sum_{j=1}^{n_k} e^{-\beta M_{jk}} \right) , \tag{25}$$

where β is the inverse temperature, the sum over k is a sum over detector layers, n_k is the number of points in detector layer k, λ is a squared cutoff distance and M_{jk} is the squared distance from point j in layer k to the track. In the same paper it was pointed out that by using the EM algorithm [8], the EA can be formulated as an iteratively reweighted least-squares procedure. One step of this procedure consists of minimizing the function

$$Q\left(\boldsymbol{p}|\boldsymbol{p'}\right) = \sum_k \sum_{j=1}^{n_k} M_{jk} p'_{jk} \tag{26}$$

with respect to $\boldsymbol{p}$. Here $\boldsymbol{p}$ and $\boldsymbol{p}'$ denote two different values of the parameter vector defining the track. The prime on the weights $\left\{p'_{jk}\right\}$ means that they are a function of $\boldsymbol{p'}$ and therefore considered as constants during the minimization.

The weights or assignment probabilities in the EA are given as

$$p_{jk} = \frac{e^{-\beta M_{jk}}}{n_k e^{-\beta\lambda} + \sum_{l=1}^{n_k} e^{-\beta M_{lk}}}. \tag{27}$$

The solution of the minimization part of the EM algorithm was obtained by a Kalman filter, and the resulting algorithm was called the Deterministic Annealing Filter (DAF).

In this work, we propose an alternative formulation of the EA based on the idea of doing the minimization part of the EM algorithm as a weighted least-squares fit on the Riemann sphere. The method of obtaining such a fit is described in Section 3. Since the distances now involved are the d_i's from Equation (14), the weights will for this algorithm be a product of the assignment probabilities given above and the geometrical factor p_i from Equation (14). The "arms" or circular arc templates are here mapped onto planes in space, and we therefore call this new algorithm the *Elastic Planes (EP)* algorithm. The major advantages of the EP over the original formulation of the EA are the following:

- Being virtually as precise as the EA, the EP is significantly faster.
- In contrary to the EA, there is no need for applying advanced, time-consuming optimization methods.
- As a consequence of the previous point, the number of parameters that are to be tuned during the optimization is significantly less for the EP than for the EA.

It has to be pointed out that the DAF, also being an EM algorithm, has many of the same advantages over the EA. In addition, since the backbone of the DAF is a Kalman filter, process noise like multiple Coulomb scattering can easily be incorporated in the formalism. The covariance matrix of the estimated track parameters is also supplied by the DAF, while it is not obvious how to calculate this for the EP algorithm. With respect to these aspects, the DAF is a more powerful algorithm than both the EA and the EP. However, the DAF is comparable to the EA when it concerns CPU time consumption. In cases where multiple scattering can be neglected, the EP is therefore a significantly faster alternative than the other two methods.

5 Simulation experiments of tracks from the ATLAS TRT

5.1 The ATLAS Inner Detector TRT

ATLAS is one of the two general-purpose experiments in the LHC project at CERN. According to the current schedule, the LHC will start to take data during the year 2005. In the ATLAS experiment, the Inner Detector is the main tracking device (see Figure 2). The outermost part of the Inner Detector is the Transition Radiation Tracker (TRT), and the simulation experiments of this paper consist of tracks from the barrel part of the TRT. The TRT is a drift tube ("straw") detector. The straws in the barrel are parallel to the beam pipe and arranged in cylindrical layers. The radii of the layers range from about 56 cm to about 106 cm, and there are 75 layers in total. There is a separation of about 6.8 mm between the layers, and the individual straws in each layer are also separated with about 6.8 mm. The diameter of each straw is 4 mm. There are in total about 50000 straws in the barrel part of the TRT. More details concerning this detector can be found in the ATLAS Inner Detector Technical Design Report [9].

The simulated measurements are given in polar coordinates. For each measurement, we therefore have information about the layer number, the radius

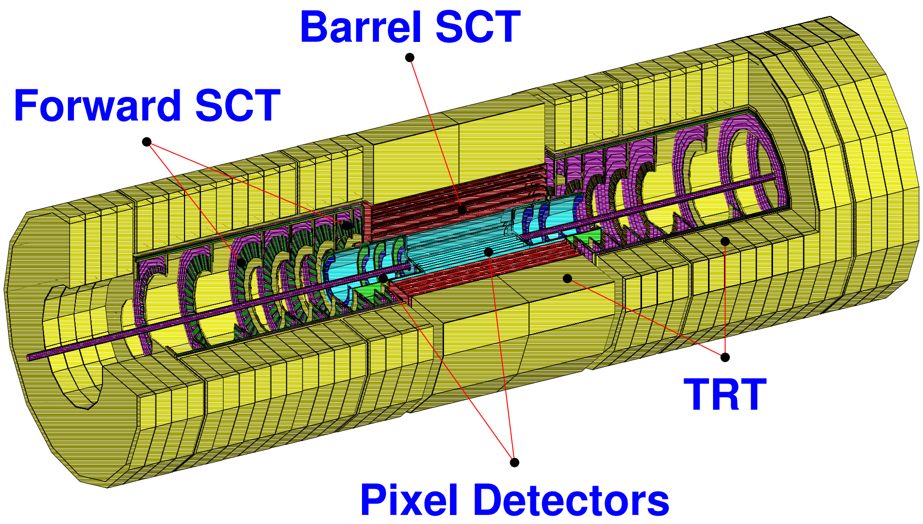


Fig. 2. The ATLAS Inner Detector.

of this layer, the polar angle of the centre of the straw, the absolute value of the drift distance and the number of the particle creating the hit. During the reconstruction, it is not known at which side of the sense wire the particle has passed. The observations are therefore in principle ambiguous.

We have simulated a sample of 9800 tracks coming from the origin and going entirely through the barrel. The information about the location of the vertex is of course not used during the reconstruction; we therefore invoke a full three parameter fit to the data points. All tracks have a transverse momentum p_T larger than 1 GeV/c. We have neglected all process noise in the form of material effects, but a measurement error of 250 μm is simulated. The track sample used in this work is the same as the one used in a couple of recent publications [4,10].

For some of the methods tested below, we need an initial estimate of the track parameters. This is provided by a least-squares fit in the $R\phi$-projection of all points in a track candidate to a straight line [4]. Since tracks with a p_T above 1 GeV/c coming from a place reasonably close to the origin are almost straight lines, this initial estimate should be sufficiently close to the true track.

5.2 Results without mirror hits

In the first simulation experiment, we will focus on the basic statistical properties of the different methods. For this reason, we have turned the mirror hits off during the simulation. This corresponds to the case of a perfect pattern recognition, i.e. all points in a track candidate belong to the track to be fitted. An example of such a track is shown in Figure 3.

We evaluate the precision of the methods by the generalized variance V of the residuals of the estimated track parameters $\hat{\boldsymbol{x}}$ with respect to the true parameters $\boldsymbol{x}_{\mathrm{true}}$. This is defined as the determinant of the sample covariance matrix $\boldsymbol{C}$ of $\hat{\boldsymbol{x}}$,

$$\boldsymbol{C} = \frac{1}{N_{\mathrm{t}}} \sum_{i=1}^{N_{\mathrm{t}}} \left(\hat{\boldsymbol{x}}_i - \boldsymbol{x}_{\mathrm{true}}\right) \left(\hat{\boldsymbol{x}}_i - \boldsymbol{x}_{\mathrm{true}}\right)^{\mathrm{T}}, \tag{28}$$

where N_{t} denotes the number of tracks in the sample.

The results of the experiment are summarized in Table 1. We here state the generalized variance V_{rel} and the CPU time consumption t_{rel} relative to the method with the best performance. We have focused on five methods: an iterative, non-linear least-squares procedure (NLS) based on the Levenberg-Marquardt algorithm [7], a global least-squares fit (GLS) [11], a Kalman filter (KF) [12], a least-squares fit to a parabola by the conformal mapping method

(CM) [13] and the least-squares fit on the Riemann sphere (RF). For the first three algorithms, the CPU time consumption is given both with and without the track parameter initialization procedure. For the Riemann fit, the generalized variance is given for the two cases with and without the modifications mentioned in Section 3.

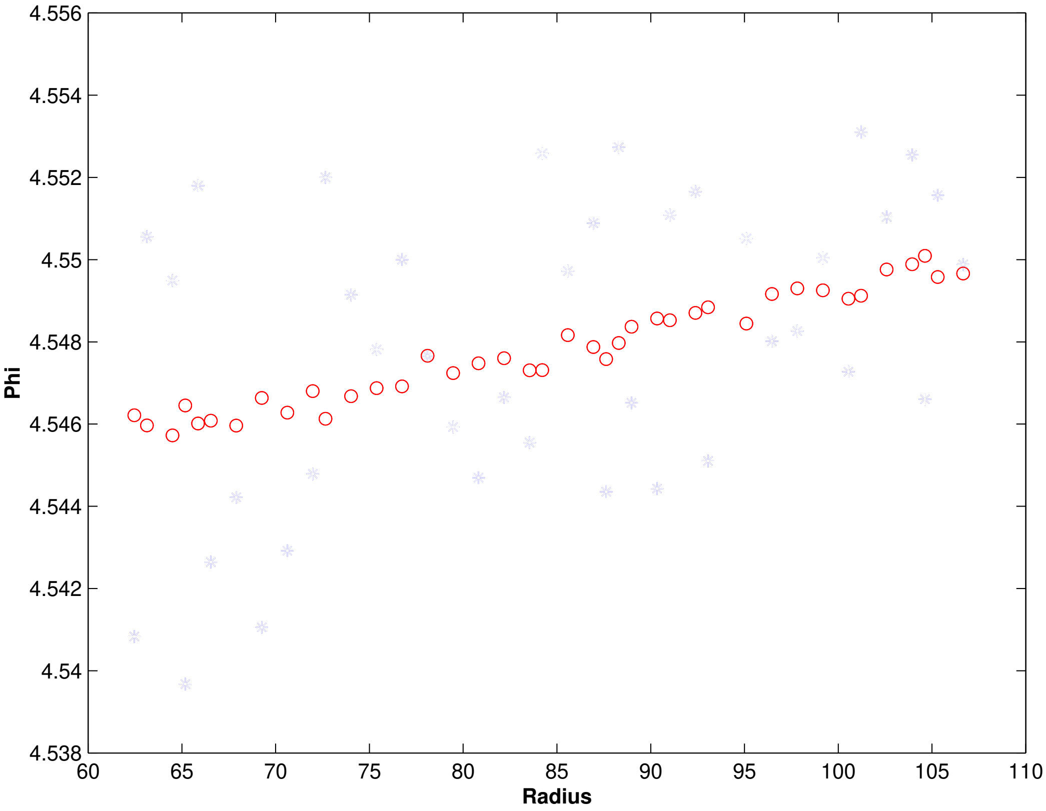


Fig. 3. A track without mirror hits, shown in the $R\phi$-projection. The radii are given in centimetres.

Method	V_{rel}	t_{rel}
NLS without initialization	1.000	36.3
NLS with initialization	1.000	41.4
GLS without initialization	1.001	15.9
GLS with initialization	1.001	21.1
KF without initialization	1.001	28.2
KF with initialization	1.001	33.3
RF with modifications	1.003	1.00
RF without modifications	1.055	1.00
CM	1.582	1.03

Table 1
The relative generalized variance and the relative time consumption for five different algorithms.

The CM method is similar to the RF in the way that it maps a non-linear problem onto a linear one, and the fit is obtained non-iteratively. Hence, the computational load of the CM should be comparable to the RF. If one requires that the circles go through the origin, the circles map onto straight lines in the transformed space. In order to be able to make a reasonable comparison, we have implemented the three-parameter version of this method which allows for a non-zero impact parameter. The CM works only in the case of a small impact parameter. The RF, however, maps *any* circle onto a plane in space and is therefore in general not restricted to this requirement.

As expected, the NLS method is the most accurate. However, the GLS, the Kalman filter and the RF with modifications are negligibly less accurate. The GLS estimates are obtained in an approximation of linear track parameter extrapolation. In the case of no process noise, the GLS procedure is equivalent to the KF, and these two methods therefore give identical results. The original RF is also performing well, but the difference to the first three methods is significant. With the modifications included, it is virtually as good as the NLS, the GLS and the KF. The CM performs much worse than all the other methods.

Concerning CPU time consumption, the RF and the CM are by far the fastest algorithms. This is true also if we neglect the track parameter initialization. It has to be noted that this initialization is a vital part of the three other methods. In order to make a realistic comparison, the initialization should therefore be included. Due to some overhead in the code, this procedure is slower than what is expected. Since the initialization basically is a linear least-squares fit, it should be possible to make it as fast as the RF or the CM. Being an iterative method, the NLS is the slowest of all. The GLS is number two after the RF. More than 80% of the execution time is used to extrapolate track parameters and build up matrices of derivatives, so the fit itself is less than three times slower than the RF fit. Again, the building-up of derivative matrices is an indispensable part of the GLS and has to be recognized as such. The KF is slower than the GLS. If process noise cannot be neglected, however, the KF is able to include this in a natural and efficient way and is therefore the natural choice.

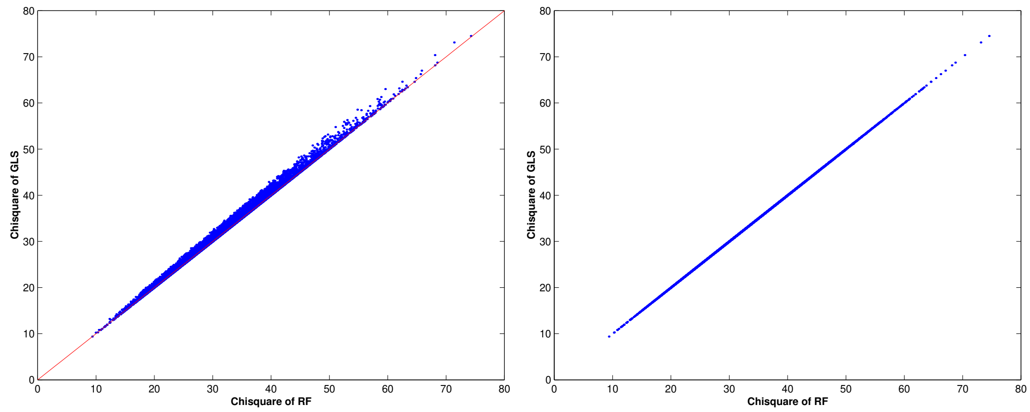
In Section 3 we have looked at two different quantities which could define a χ^2 of a fitted track for the RF. Figure 4 shows scatter plots of these two candidates against the corresponding χ^2 of the GLS. The correlations are seen to be large in both cases, but it is obvious from (a) that χ^2_{GLS} is consistently larger than or equal to $\chi^2_{\perp}$. This is very reasonable, since the distances in $R\phi$ always are larger than or equal to the orthogonal distances. From (b) it is clear that the correspondence between χ^2_{GLS} and $\chi^2_{R\phi}$ is almost exact. This is confirmed by the results shown in Figure 5. In (a) we show a histogram of the tail probability of the RF using $\chi^2_{\perp}$ to calculate the χ^2, while in (b) the same quantity is shown with $\chi^2_{R\phi}$ as a basis for the calculation. (The tail probability is defined as $1 - F(\chi^2)$, where F is the cumulative distribution function of the χ^2-distribution. If a χ^2-statistic really is χ^2-distributed, a histogram of the corresponding tail probability should be reasonably flat.) In (c) a histogram of the tail probability of the GLS is plotted. The $\chi^2_\perp$ obviously has a tendency to yield too high tail probabilities. This is also reflected in the mean value, which is 0.518. The histograms of (b) and (c) are virtually indistinguishable, and the mean value is in both cases 0.502.

5.3 Results with mirror hits

In the second simulation experiment, the mirror hits are turned on. Figure 6 shows the same track as the one shown in Figure 3, but now the mirror hits are included as well.

We have studied the performance of five different versions of the Elastic Arms algorithm. Three of them are based on the iteratively reweighted least-squares formulation, while the rest follow the "standard" approach by minimizing the effective energy. For this minimization we have used ordinary gradient descent during the annealing phase, but at the final temperature we have applied the DFP algorithm [14].

The first algorithm is the DAF. The second one, the global least-squares method (GLS) with annealing, is an alternative which is faster than the DAF. It basically works by doing the minimization part of the EM algorithm with a GLS method instead of with a Kalman filter. In the case of negligible process



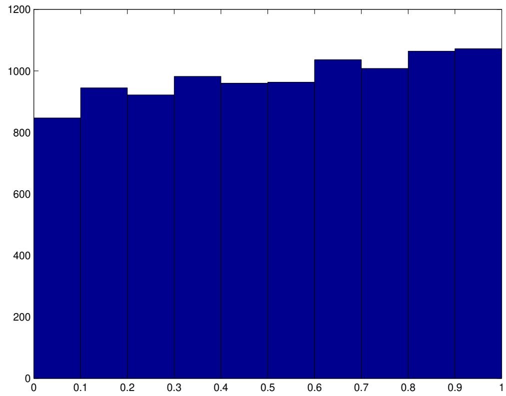
(a) $\chi^2_\perp$ against χ^2_{GLS}

(b) $\chi^2_{R\phi}$ against χ^2_{GLS}

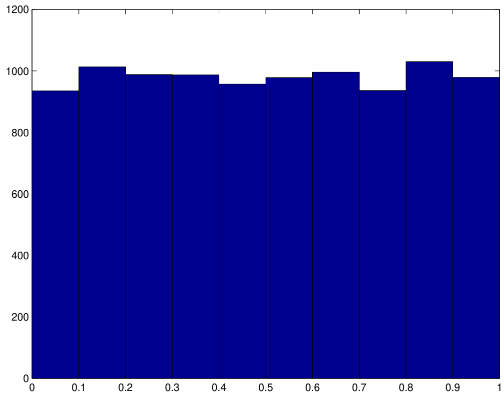
Fig. 4. We here plot $\chi^2_\perp$ and $\chi^2_{R\phi}$ against χ^2_{GLS}. In (a) the straight line (with slope 1) is shown as a guide to the eye.

noise, this method is equivalent to the DAF. The third method, the Elastic Planes (EP) algorithm, is doing the minimization part of the EM algorithm on the Riemann sphere. Otherwise, it is similar to the first two methods. The fourth algorithm is the standard EA algorithm as presented by Ohlsson, Peterson and Yuille. The fifth algorithm is minimizing an effective energy with the same structure and using the same tools as the standard EA, but the difference is that the distances are given on the Riemann sphere instead of in the plane. It is therefore related to the EP algorithm (EM formulation) in the same way as the standard EA algorithm is related to the GLS with annealing.

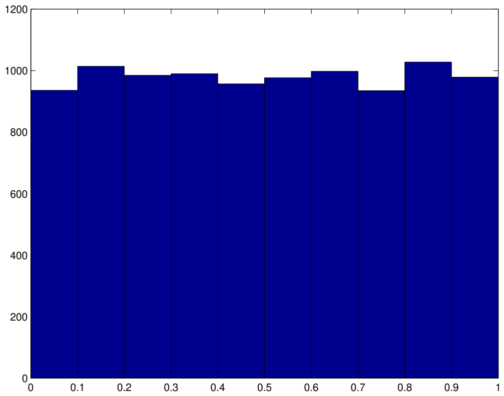
The results are again given relative to the method with the best performance. The initialization procedure is now included in the calculation of the time consumption for all methods. This is due to the fact that now all the algorithms have to be provided with an initial guess of the track parameters.



(a)



(b)



(c)

Fig. 5. Here histograms of tail probabilities of the RF ((a) and (b)) and the GLS are shown.

The results from this experiment are summarized in Table 2. As expected, it is seen that the DAF and the GLS with annealing produce equivalent results when it concerns the accuracy of the estimates. The EP algorithm (EM formulation) with the modifications presented in Section 3 is worse by a factor of 1.01 in generalized variance. Please note that this on average corresponds to only about 0.15% larger standard deviations of the track parameter estimates, something which in an experimental setting can be assumed to be negligible.

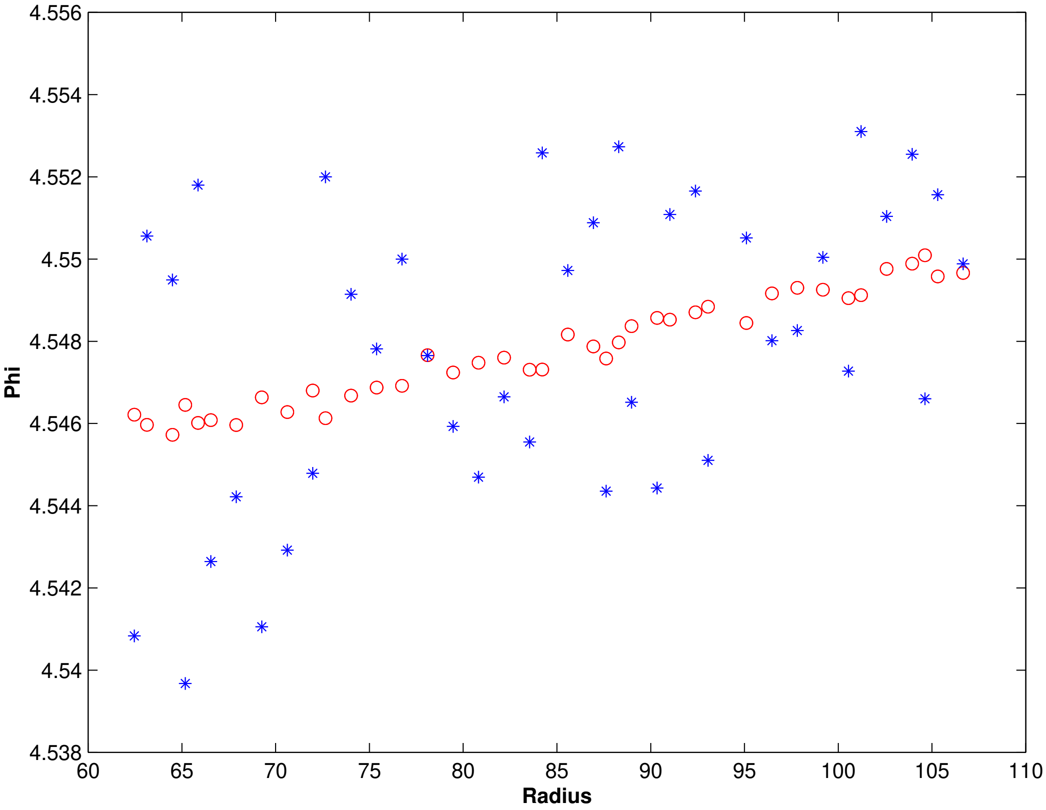


Fig. 6. This is the same track as in Figure 3, but here the mirror hits are included. The track hits are shown as open rings, the mirror hits as asterisks.

Method	V_{rel}	t_{rel}
DAF	1.00	13.6
GLS with annealing	1.00	3.3
EP (EM formulation) with modifications	1.01	1.0
EP (EM formulation) without modifications	1.19	1.0
EA (standard formulation)	1.00	28.1
EP (standard formulation) with modifications	1.01	17.4
EP (standard formulation) without modifications	1.23	17.4

Table 2
The relative generalized variance and the relative time consumption for five different algorithms.

As expected, the EP algorithm (standard formulation) behaves similarly to the EP algorithm (EM formulation) when it concerns the accuracy of the estimated parameters. The standard EA algorithm is, also as expected, as accurate as the DAF/GLS with annealing.

As for the CPU time consumption, the EP algorithm (EM formulation) is the winner. However, the gain by applying the Riemann fit is less here than in the case of no mirror hits. This is mainly due to three reasons. Firstly, the initialization procedure is needed also for the EP algorithm. Secondly, the calculation of the weights takes approximately the same amount of time for the EP as for the other algorithms. Thirdly, since one needs to build up the derivative matrices only once for each track, the relative speed of the GLS increases as the number of iterations increases. All these effects contribute to make the relative difference between the EP and the GLS with annealing less than the relative difference between the RF and the GLS. Anyway, the EP is still significantly faster than the GLS with annealing, and a faster initialization procedure would have made the relative difference between the EP and the GLS with annealing even larger than the number presented in Table 2. The DAF is more than 13 times slower than the EP (EM formulation). As expected, the two algorithms minimizing the effective energy are the slowest. Note, however, that the EP (standard formulation) is significantly faster than the EA and almost as fast as the DAF. This can be explained by noting that the expressions of the distances are much simpler on the Riemann sphere (given in Equation (5)) than in the plane, so the effective energy and its derivatives are therefore computationally less expensive. The standard EA is almost 30 times slower than the fastest version of the EP.

6 Conclusions and outlook

We have in this paper presented a novel algorithm for track fitting in high-energy particle physics detectors. It is based on the idea of transforming the task of fitting circular arcs in the plane to fitting planes in space. This is done by mapping the measurement coordinates onto the Riemann sphere. The resulting problem can be solved in a fast and non-iterative manner.

In order to further optimize the performance of the RF algorithm, we have introduced some modifications of the ideas from the original paper [1]. This results in a method which is virtually as precise as an optimal, non-linear procedure, and it is significantly faster than other, popular estimators as the Kalman filter and the global least-squares method. Moreover, it is comparable in computational complexity and much more precise than the conformal mapping method. However, it is not straightforward to obtain an expression of the covariance matrix of the estimated parameters. Future efforts will be put into an attempt of solving this problem.

By employing the EM formulation of the Elastic Arms algorithm, we have presented a novel version of this called the Elastic Planes algorithm. It turns out to be negligibly less precise than an optimal version of the Elastic Arms, and it is much faster. With respect to computational speed, the EP algorithm compares favourably also with the DAF and a global least-squares method with annealing. It has to be noted, however, that the DAF incorporates process noise in a natural way. In cases where for instance multiple Coulomb scattering cannot be neglected, the DAF is the natural algorithm to choose. It can also be noted that the CPU time consumption for the different algorithms refers to an implementation in `MATLAB`. For compiler languages as `C` or `C++`, the timing behaviour might be slightly different.

In this work, we have applied the RF and the EP algorithm to relatively high-energetic tracks, i.e. tracks which are reasonably straight. For these tracks we have also seen that the Kalman filter and the GLS method work very well. If we consider the task of fitting a circle to measurements from a Ring Imaging Cherenkov Detector (RICH), the Kalman filter is not a good solution. The Riemann fit is well defined also in this case, and it is therefore an alternative method for solving such a problem. Sometimes one also needs robust circle fitting algorithms when analyzing data from a RICH detector [15]. The EP algorithm should be perfectly suited for this, and it would be very interesting to investigate its behaviour on experimental data.

We have in this paper focused on the task of fitting circular arcs. Data coming from tracking detectors in solenoidal magnetic fields are in general three-dimensional, and one will in such cases rather fit a helix to the measurements. It should be possible to extend the Riemann fit to this more general situation without losing the advantages of a fast and direct solution method. This will be the topic of a future study.

References

[1] B. Lillekjendlie, Circular arcs fitted on a Riemann sphere. *Computer Vision and Image Understanding* 67 (1997) 311.

[2] L. V. Ahlfors, *Complex analysis: an introduction to the theory of analytic functions of one complex variable*. Int. series in pure and applied mathematics, McGraw-Hill, 1979.

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

[4] 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.

[5] H. Anton, *Elementary Linear Algebra.* John Wiley & Sons, 6th. ed., 1991.

[6] K. V. Mardia, J. T. Kent and J. M. Bibby, *Multivariate Analysis.* Academic Press Ltd., 6th printing, 1997.

[7] A. Sen and M. Srivastava, *Regression Analysis.* Springer-Verlag, New York, 1990.

[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] A. Strandlie and J. Zerubia, Particle tracking with iterated Kalman filters and smoothers: the PMHT algorithm. *Computer Physics Communications* 123 (1999) 77.

[11] R. K. Bock, H. Grote, D. Notz and M. Regler (editor), *Data analysis techniques for high-energy physics experiments.* Cambridge University Press, Cambridge, 1990.

[12] R. Frühwirth, Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods* A262 (1987) 444.

[13] M. Hansroul, H. Jeremie and D. Savard, Fast circle fit with the conformal mapping method. *Nuclear Instruments and Methods* A270 (1988) 498.

[14] W. Press et al., *Numerical Recipes in C.* Cambridge University Press, Cambridge, 1988.

[15] G. Agakichiev et al., Cherenkov ring fitting techniques for the CERES RICH detectors. *Nuclear Instruments and Methods in Physics Research A* 371 (1996) 243.