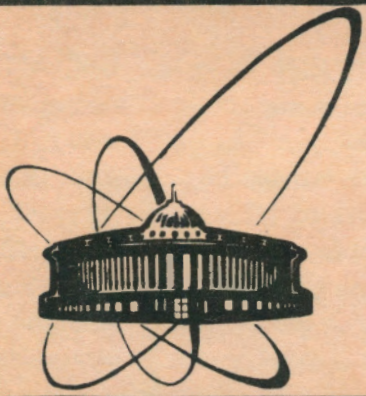

сообщения
объединенного
института
ядерных
исследований
дубна

E10-91-361

N.I.Chernov, I.V.Kisel, S.M.Korenchenko,
G.A.Ososkov

TRACK RECONSTRUCTION IN DISCRETE DETECTORS

1991

1 · Introduction

The least square fit (LSF) is the traditional method to fit a particle trajectory to points measured in such detectors as bubble and streamer chambers. The LSF application in these experiments is based on the independence and normal distribution of errors of measurements. But this is no longer valid for electronic experiments with discrete detectors like multiwire proportional chambers or strip detectors. They consist of a discrete set of detecting wires or strips in fixed positions which are hit (produce signals) whenever a particle track comes close to them. Here the errors of measurements are in fact correlated and their distribution is not normal but uniform with the width d specified by the wire or strip spacing.

This situation was discussed in [1, 2, 3]. It is known that the accuracy of track reconstruction in such detectors depends severely on the mutual disposition of wires (chambers). For different dispositions the mean accuracy may vary from d (in the worst case) to d/N (in the best one), where N is the number of chambers, i.e. the number of measurements. Compare to the canonical accuracy $d/\sqrt{N}$ of the LSF in detectors with the Gaussian distribution of errors! But one cannot expect $\sqrt{N}$ times improvement rejecting the LSF because our error distribution is not Gaussian. Besides, non-LSF procedures of track reconstruction exploiting the discrete nature of detectors are complicated and time-consuming. So, the LSF is still widely used in these experiments despite it being in no way justified.

We compared numerically the accuracy of track reconstruction by the LSF and by specific 'discrete' methods in a setup corresponding to the ARES experimental facility in JINR, Dubna [4]. Surprisingly, we found really $const \cdot \sqrt{N}$ times improvement comparing to the LSF! Besides, we describe here a simple procedure of track reconstruction in discrete detectors which is very easy to implement and rather fast. Our algorithm is applicable to straight line tracks as well as, in more important cases to circular ones. Some related problems are also discussed.

2 Basic notation and problem formulating

We consider detectors consisting of several parallel chambers ('coordinate planes'). Each chamber is performed as a row of identical rectangular cells. Each cell contains a wire which is triggered (hit) whenever a particle track crosses the cell. Denote (see fig. 1):

N — the number of chambers,

d — the cell width (along the chamber),

h — the cell height.

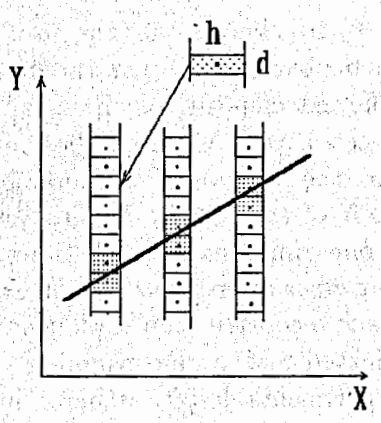

Figure 1: Basic notation.

A particle track triggers one or more adjacent wires in each chamber. Our problem is to find a track-candidate *crossing all* the cells with wire hits and *missing all* other ones.

This problem admits a reformulation, which is mathematically simpler. When several adjacent cells in a chamber contain signals we conclude that the track crosses the left side of the lower cell and the right side of the upper one or vice versa (fig. 2). Otherwise it would either not have crossed all the hit cells or have touched one of missing ones. This is obvious for straight tracks and 'almost' true for curved ones with sufficiently small curvature. We neglect possible errors like the one shown

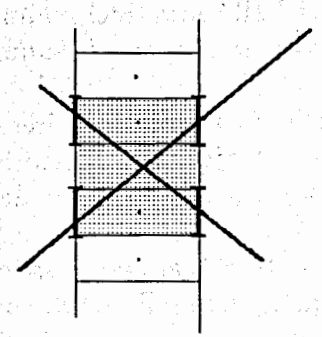

Figure 2: Replacement of a wire cluster by two vertical segments.

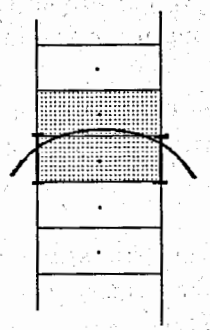

Figure 3: A possible error in replacement illustrated in fig. 2.

in fig. 3. Besides, the choice between two above possibilities can be easily made by rough approximation of the track direction, say, comparing the positions of wire hits in the leftmost and rightmost chambers (for straight tracks). We also neglect possible errors in such a choice. Then we replace each group ('cluster') of adjacent hits by only two segments (the corresponding sides of two utmost cells in this cluster). Our problem is now to find a track crossing all these segments. Note that we do not need to check whether our track touches other (missing) cells or not. Then we can digress from our chambers, cells and work only with a set of parallel segments, each of length d.

3 Probabilistic and geometrical considerations

Usually the estimation of unknown track parameters is based on probabilistic models and statistical methods, e.g. the maximum likelihood estimates. In our case the probabilistic model (at least in its idealized version) is very poor (see also [2]). Really, every track-candidate crossing all our segments has the likelihood 1 while any other track-candidate has the likelihood 0. The maximum likelihood method gives then a continuum family of tracks with the same likelihood 1. Nobody can distinguish

'most probable' tracks among them. We call all these tracks admissible and the corresponding region in the track parameter space the admissible region.

The situation can change if we admit malfunction of our wires, noisy hits and deflections of tracks from modelled trajectories. Such an approach is discussed below in sect. 8. Now we focus only on the idealized model.

Let us describe admissible regions in the track parameter space and then estimate the accuracy of track reconstruction. First we consider rectilinear tracks $y = ax + b$ (the coordinate system is shown in fig. 1). The admissible region in the parameter space (a, b) is bounded by admissible tracks containing one or more endpoints of our segments. So, if (x', y') is an endpoint of a given segment, then the equation

$$ax' + b - y' = 0 \tag{1}$$

determines a part of the boundary of the admissible region. It is then a straight line in the space (a, b), so our admissible region is always a convex polygon. Its projections to the coordinate axes give two segments which we denote by $[a_{min}, a_{max}]$ and $[b_{min}, b_{max}]$ respectively. Now it is reasonable to take

$$\hat{a} = \frac{1}{2}(a_{min} + a_{max}), \qquad \hat{b} = \frac{1}{2}(b_{min} + b_{max}) \tag{2}$$

as parameter estimates and the values

$$\Delta a = \frac{1}{2}(a_{max} - a_{min}), \qquad \Delta b = \frac{1}{2}(b_{max} - b_{min}) \tag{3}$$

as maximal parameter errors. The estimates (2) obviously provide the minimal values of maximal errors (3). The convexity of the admissible region implies that the point $(\hat{a}, \hat{b})$ always belongs to it.

4 Mean accuracy of the parameter estimates

Only a rough sketch for solving this problem is outlined further. Taking all the possible endpoints (x', y') of our segments (i.e. all the corner points of our cells in the detector) we obtain through the equation (1) a set of $2K_c$ straight lines in the parameter space (a, b). Here K_c stands for

the number of cells. These lines split the parameter space into many convex polygons. Now we are able to estimate roughly the typical errors $\Delta a, \Delta b$ in (3). Suppose we are interested in a region of size ~ 1 in the parameter space (for instance, $0 \le b \le 1, -1 \le a \le 1$). Let the length of each chamber be equal to 1, then the number of cells is $K_c = N/d$. So we have $2K_c = 2N/d$ lines which partition our parameter region into polygons. Suppose the typical size of a polygon is ε, so its typical area is ε^2. Then the number of (typical) polygons is $M \approx 1/\varepsilon^2$. So the total perimeter length of all the polygons is approximately $M \cdot \varepsilon \approx 1/\varepsilon$. On the other hand, counting the number of lines forming these polygons we estimate the perimeter length as $\approx 2K_c = 2N/d$. Hence, $\varepsilon \sim d/N$. Remember now that ε is the typical polygon size, i.e. it is exactly the same as Δa and Δb in (3). So we obtain the typical accuracies $\Delta a, \Delta b$ in track reconstruction as d/N, as was already mentioned in Introduction.

To illustrate this, let us consider a simple detector facility consisting of N chambers of the length 1, situated along the section of X-axis of the length 1 (i.e. $0 \le x \le 1$). In the parameter space (a, b) the region $0 \le b \le 1,\ 0 \le a + b \le 1$ is considered, which is formed by the tracks crossing all our chambers. Its area equals 1 and it is partitioned by $2N/d$ lines, the total length of which is $2N/d$. Therefore the lower bound of parameter errors is $d/2N \le \Delta a, \Delta b$. Moreover, we can really attain this accuracy (up to a factor) by the following trick. Situating $N/3$ leftmost chambers in such a manner, that their relative y-displacements are, correspondingly, $3id/N$ $(i = 1, 2, \ldots, N/3)$, we guarantee that $\Delta b \le 3d/N$. In the same way, by displacing $N/3$ rightmost chambers we obtain $\Delta a \le 3d/N$, hence the accuracy of parameter estimates as $const \cdot d/N$ is really attained. We have only to estimate the factor $const$. Note that without any mutual chamber disposition along the y-axis the maximal errors Δa, Δb would be $\approx d$, i.e. worse than for LSF with the Gauss model.

The above considerations are only rough estimation of the *maximal* admissible accuracy of parameter estimates. In real experiments it depends severely on the region in the parameter space we are interested in and on the structure of the polygons in this parameter region. For instance, let these polygons be very different: there are a number of large and a lot of little ones with a small total area. Then the typical parameter errors would be specified by the sizes of large polygons which

could be much bigger than d/N (up to d, see also about the necessity to stagger chamber dispositions uniformly [1]).

Thus a requirement to the detector design comes out. The chamber construction and disposition should provide polygon sizes in the parameter space as small as possible. If their number is fixed, they should have as much homogeneous sizes as possible. The optimal disposition of chambers can be found by numerical analysis through minimizing the mean values of $\Delta a, \Delta b$ in (3).

The influence of the chamber height h should also be taken into account. Apparently, if $h \ll d$, the accuracy should get worse, since the left and right corner points of a cell would give almost the same line in the parameter space. If $h \approx d$, the accuracy should be good and then it would not improve with the h increasing — see also our sect. 7. We guess that if the typical angle α of the track slope is small (i.e. tracks go mostly horizontally), then h has to be chosen as $\geq d/\alpha$.

5 Numerical algorithm of the track reconstruction

Let all our segments be numbered from left to right: $1, 2, \ldots, 2N$. Take any two segments i, j (with $i < j$) and consider the straight line $y = a'x + b'$ joining the upper endpoint of the segment i and the lower one of the other segment j. It is easy to check that $a_{min} = \max_{\forall i<j}\{a'\}$ and $b_{max} = \min_{\forall i<j}\{b'\}$. Analogously, considering all the lines $y = a''x + b''$ joining the left endpoint of the segment i and the right one of the segment j $(i < j)$ we find $a_{max} = \min_{\forall i<j}\{a''\}$ and $b_{min} = \max_{\forall i<j}\{b''\}$. Further, we evaluate the values $\hat{a}, \hat{b}, \Delta a, \Delta b$ through (2), (3). Note that in this algorithm we do not need to check whether our lines cross all the segments or not. The corresponding FORTRAN routine contains only 20 – 30 lines and runs rather fast. Further optimization is also possible if one makes use of the same length d of all the segments and the fixed positions of all the chambers (for instance, all possible slopes a', a'' of the above lines can be computed in advance and tabulated).

6 Circular tracks

The extension of all our results to circular tracks presents no new actual difficulties. The circle equation $(x-a)^2 + (y-b)^2 = R^2$ can be written

in the form

$$x^2 + y^2 = 2ax + 2by + \gamma \qquad (4)$$

with a new parameter $\gamma = R^2 - a^2 - b^2$. For a fixed point (x, y) the equation (4) determines a plane in the 3-parameter space (a, b, γ). Therefore our admissible regions become convex *polyhedrons* in the space (a, b, γ).

However, the natural circle parameters a, b, R (as well as a, b, γ) are inconvenient for tracks with small curvature since a, b, R grow to infinity. All we really need to know of a track is where it enters our device (i.e. the first chamber), where it exits from the last chamber and finally its curvature. So we take three new parameters y_L, y_R and κ $(= \pm 1/R)$ shown in fig. 4 to describe circular tracks. The curvature κ is supplied with the sign '+' for convex arcs and '−' for concave ones. Now the parameter region (y_L, y_R, κ) must be strongly bounded.

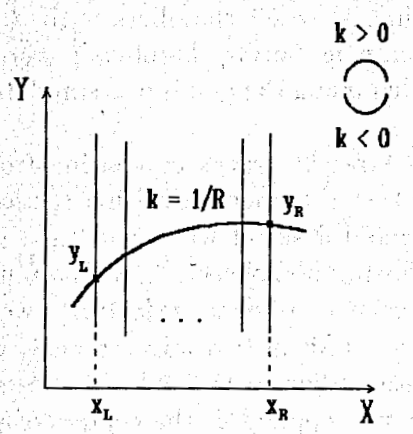

Figure 4: Basic notation for circular tracks.

The admissible regions in our parameter space (y_L, y_R, κ) are far from polyhedrons. They are not even convex. However, a simple numerical procedure for parameter estimation is possible. Let $(y_{L,min}, y_{L,max})$, $(y_{R,min}, y_{R,max})$ and $(\kappa_{min}, \kappa_{max})$ be the projections of the admissible region to the coordinate axes. These are the range of our three variables y_L, y_R, κ inside the admissible region. As in sect. 4, it is reasonable to take $\widehat{y_L} = \frac{1}{2}(y_{L,min} + y_{L,max})$ etc. with the maximal errors $\Delta y_L = \frac{1}{2}(y_{L,max} - y_{L,min})$ etc.

Let, as in sect. 5, our segments be numbered from left to right: $1, 2, \ldots, 2N$. Take any three segments i, j, k $(i < j < k)$ and draw a circle through the lower endpoints of the segments i, k and the upper one of the middle segment j. Compute the parameters y'_L, y'_R, κ' of the circle. Now it is easy to check that $y_{L,min} = \max_{\forall i<j<k}\{y'_L\}$, $y_{R,min} = \max_{\forall i<j<k}\{y'_R\}$ and $\kappa_{max} = \min_{\forall i<j<k}\{\kappa'\}$. Analogously, we find $y_{L,max}, y_{R,max}$ and κ_{min} drawing circles through the upper endpoints of the segments i, k and the lower one of the segment j. Note again that we do not need to check whether our circles cross all other segments or not. To speed up the procedure some improvements as in sect. 5 are also available.

7 Numerical experiment

We choose for numerical study an analogue of the experimental facility ARES in JINR, Dubna [4]. It contains $N = 10$ chambers with $d = 2$ mm, $h \approx 4$ mm. The distance between neighboring chambers is about 25 mm. The displacement of each chamber along the y-axis was simulated randomly from 0 to d.

We simulated several thousands of straight tracks emanating from random points of the middle cell of the first chamber and with a random slope from 0^o to 60^0. Each track triggered a set of wires and then we obtained a set of vertical segments via the method of sect. 2. No malfunction, no noisy hits and no deviations from rectilinear trajectories were admitted. We reconstructed the tracks by LSF in two ways. Firstly, we applied LSF to the centers of the groups of adjacent wire hits ('clusters') in all the chambers. Secondly, the LSF was applied to the centers of all the above vertical segments. The results of these two fits were almost identical (both are marked by the same line in fig. 5).

We also performed a 'discrete' fit described in sect. 5. Fig. 5 shows that the root mean square (RMS) deviation of the reconstructed tracks from the modelled ones is about 1.5 times smaller for our 'discrete' tracks than for the least-square ones.

Fig. 5 also shows the dependence of RMS errors on the cell height h. As we have predicted in sect. 4, the RMS deviation decreases with the h increasing but it reaches a plateau (for $h \approx d$) and does not improve any further.

We also made the same experiment but with a constant angle of the

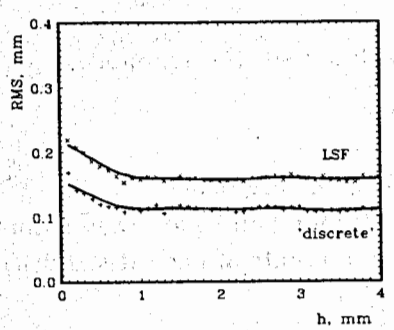

Figure 5: The RMS deviations of the reconstructed tracks from the modelled ones for the 'discrete' algorithm (+) and for the LSF (×).

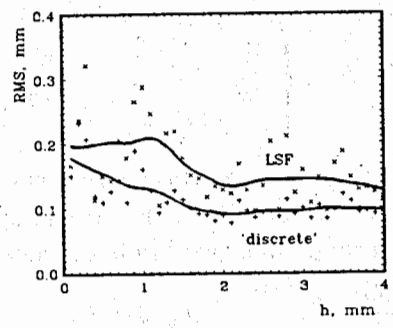

Figure 6: The same as in fig. 5 but with a fixed track slope $(= 10^o)$.

track slope $(= 10^o)$. For each value of h we simulated its own displacement of the chambers along the y-axis. Note the RMS values for the LSF-tracks are much more sensitive to the chamber disposition – see fig. 6. Interpretation of this phenomenon can be found in [1].

8 Deviations from the idealized model

Coming back to sect. 3, we admit now malfunction of the wires, noisy hits and deflection of tracks from modelled trajectories (for instance, as a result of multiple scattering). In these cases the existence of an admissible track crossing all our segments is no longer guaranteed. So we have to choose 'the most probable' track-candidate which, however, misses some of our segments. To this end we use (see also [2]) the Chebyshev (minimax) metric:

$$\mathcal{L}(a, b) = \max_i\{d_i\} \to \inf, \qquad (5)$$

where d_i denotes the distance of the track $y = ax + b$ to the i-th segment ($i = 1, 2, \ldots, 2N$):

$$d_i = \begin{cases} 0, & \text{if the track crosses the segment,} \\ \dfrac{|y_i - ax_i - b|}{\sqrt{a^2+1}} & \text{otherwise.} \end{cases} \qquad (6)$$

Here (x_i, y_i) means the nearest endpoint of the segment.

The function $\mathcal{L}(a, b)$ is obviously continuous and piecewise smooth. In general, it has two kinds of fractures (i.e. points of derivatives discontinuity):

Fractures of the first kind: a jump from one segment at which the maximum in (5) is attained to another one;

Fractures of the second kind: a jump from 0 to $|y_i - ax_i - b|/\sqrt{a^2+1}$ in the expression (6).

Due to our assumption $\mathcal{L}(a, b) \neq 0$ (otherwise one can find an admissible track and solve the problem as in sect. 5). Therefore the fractures of the second kind must be eliminated.

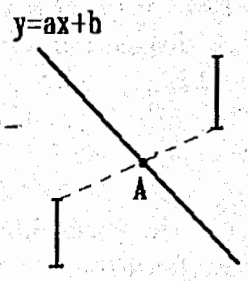

Figure 7: A track equidistant from two segments.

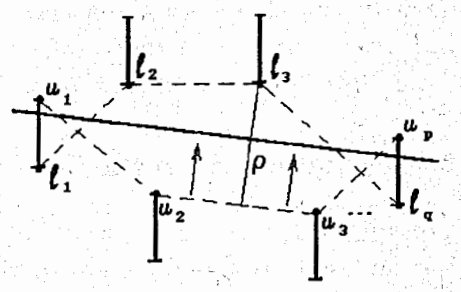

Figure 8: How the algorithm works.

The fractures of the first kind occurred when a track $y = ax + b$ is equidistant from some two of our segments, i.e. it crosses the middle point

between them (fig. 7). If (x', y') are the coordinates of the middle point, then the equation $ax' + b - y' = 0$ gives a straight line in the parameter space separating different regions of the $\mathcal{L}(a, b)$ smoothness. Hence we obtain, as in sect. 4, a partition of the plane (a, b) into convex polygons where our function $\mathcal{L}$ is smooth.

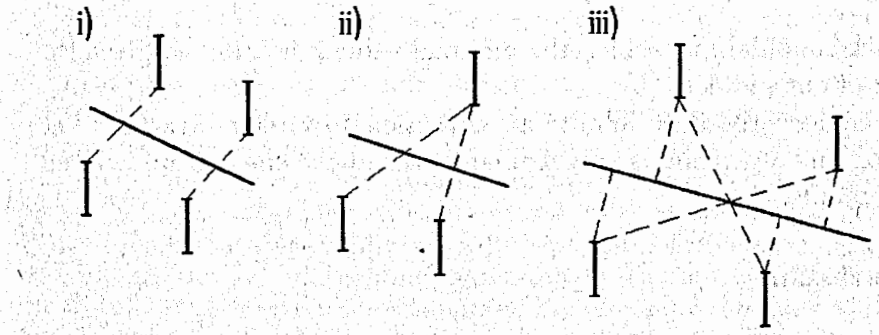

Figure 9: Three possibilities for tracks equidistant from three or more segments.

Although the function $\mathcal{L}(a, b)$ is *not* piecewise linear, its minimum is always attained at a vertex of the above polygons. This can be checked by elementary geometric considerations and we omit the details. Vertices of our polygons correspond to tracks equidistant from three or more of our segments. All possible situations are shown in fig. 9. Notice that in all the cases the track runs parallel to a line joining some two ends of our segments.

Consider the set of the upper endpoints of our segments and construct its convex hull. Denote $u_1, \ldots, u_p$ the upper endpoints in the lower part of the hull (from left to right) as in fig. 8. Analogously let $l_1, \ldots, u_q$ denote the lower endpoints of our segments constituting the upper part of the convex hull of all the lower endpoints.

The following algorithm for the function $\mathcal{L}$ minimization comes out:

1. Draw the line through any two neighboring points u_i, u_{i+1} and look for a point l_j which is the most distant from this line. Denote the corresponding distance by ρ.

2. Transfer our line to half a distance closer to the lower endpoint found at step 1.

3. Repeat steps 1-2 for all pairs u_i, u_{i+1} and then repeat steps 1-2 once again but for the opposite endpoints: take lower endpoints l_i, l_{i+1}, look for an upper one u_j and so on.

The track-candidate for which the minimal value ρ is attained gives the solution of our problem (5).

The above procedure admits an extension to circular tracks. The corresponding algorithm is very similar to the above one and we will not go into details.

9 Conclusion

Our study showed the striking results of taking into account the discrete nature of electronic detectors and the art of their staggering in experimental setup.

However it is necessary to keep in mind the idealized character of the single track models considered. Therefore at the next stage of our study we are going to consider most important practical circumstances, like imperfect detection efficiency, noisy points and random cluster size due to track crossing, multiple scattering etc. All these factors may decrease our algorithm contribution. The special careful exploration is needed to optimize algorithms and obtain quantitative estimations.

References

[1] I. Duerdoth, Nucl. Instr. and Meth. 203 (1982) 291.

[2] F. James, Nucl. Instr. and Meth. 211 (1983) 145.

[3] I. M. Ivanchenko e.a., Commun. JINR, P10-89-436, Dubna, 1989.

[4] V. A. Baranov e.a., Nucl. Instr. and Meth. B17 (1986) 438.

Received by Publishing Department
on July 30, 1991.

Чернов Н.И. и др. E10-91-361

Восстановление треков в дискретных детекторах

В работе проведен вероятностный и геометрический анализ данных, получаемых на экспериментальных установках таких, как многопроволочные пропорциональные камеры или стриповые детекторы. Теоретические исследования показали, что учет эффекта дискретности дает улучшение точности оценки параметров треков в $\sqrt{N}$ раз по сравнению с традиционным методом наименьших квадратов (N - число измерений). Для идеализированной модели прямолинейных треков предложен простой, удобный и быстрый алгоритм их восстановления, обобщенный также на случай дуговых треков. Численный анализ на основе параметров установки АРЕС подтвердил теоретические оценки. Предложено расширение алгоритма для учета таких факторов реальных измерений, как многократное рассеяние.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1991

Chernov N.I. et al. E10-91-361

Track Reconstruction in Discrete Detectors

The probabilistic and geometrical analysis is given for data from experimental facilities like multiwire proportional chambers or strip detectors. The theoretical study showed that taking into account the discrete nature of data gives us the accuracy of track reconstruction $\sqrt{N}$ better as compared with conventional least square fitting (here N is the number of measurements). For an idealized straight track model a simple, handy and fast algorithm is proposed for track reconstruction. It is modified to the case of circular tracks. Numerical analysis on the basis of the parameters of the ARES facility confirms the above theoretical estimates. The algorithm extension is proposed to take into account such factors of real observation as multiple scattering.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1991