GLOBAL METHODS OF PATTERN RECOGNITION

Helmut EICHINGER

CERN, Geneva, Switzerland

Global methods of track finding and of parameter estimation require sufficient information to be available at a certain point on a particle trajectory. Then vector functions can be looked for such that in the space spanned by these functions each track forms a cluster of points which is well separated from the clusters of the other tracks. A clever detector lay-out can simplify this task. However, current wire chamber detectors do not provide either enough, or precise enough, measurements at points on curved trajectories to allow the application of global pattern recognition methods – improved detectors should not only provide precise coordinate measurements but also sufficiently precise measurements of the track direction.

1. Pattern recognition in counter experiments

Pattern recognition in high energy physics counter experiments is usually split into the following tasks:

(i) to solve the problem of *grouping measurements*, provided by localization detectors to a particle's *trajectory candidate*,

(ii) to *confirm* this track hypothesis,

(iii) to *estimate* (physically interesting) track *parameters*. The final estimates to be obtained by the track fitting [1].

It has become common to regard pattern recognition as an *intrinsically difficult* problem, in the sense that no general and (somewhat) optimal strategy can be given to solve the problem [2].

As the general solution of the equations of motion for a particle moving in a magnetic field contains five integration constants, a particle's trajectory is generally characterized by *five parameters*. They might be:

(i) *two coordinates* of one point on it (usually near to the vertex region) and *two directions* (cosines or tangents or helix angles) at this point and the value of q/p ; where q is the electric charge of the particle and p is its momentum.

If the particle encounters dense media on its path through the detector:

(ii) the right *mass* has also to be assigned to a particular possible trajectory in order to be able to take account of the correct amount of energy loss and scattering,

(iii) and the use of *scattering angles* as additional track parameters might be an efficient way to describe facultative multiple or nuclear scattering.

But in the limit of high momentum particles, the

energy loss gets mass independent and multiple scattering decreases and, therefore, a track can be described by deterministic functions of the five start parameters only – $p_1(s_0), \dots, p_5(s_0)$.

A complete *track model* would then consist in knowing the particle's path and the track parameters as, for instance, explicit analytic functions of the curvilinear length s , i.e.

$$p_1 = x = f_x[p_1(s_0), \dots, p_5(s_0); s],$$

$$p_2 = y = f_y[p(s_0); s],$$

$$z = f_z[p(s_0); s],$$

$$p_3 = f_3[p(s_0); s],$$

$$p_4 = f_4[p(s_0); s],$$

$$p_5 = q/p = f_5[p(s_0); s],$$

when using cartesian coordinates and choosing $x(s_0)$ and $y(s_0)$ as the first two track parameters.

In principle, we have to distinguish between:

(i) the *ideal track*. It corresponds to a valid solution of the equations of motion and would be the particle's path described by the functions f .

(ii) The *physically true track*. It corresponds to the real path of the particle and the difference between this and the ideal track is due to stochastic processes (scattering).

(iii) The *measured track* constituted by the set of the n measurements $\{c_1, \dots, c_n\}$ – where n is the number of detectors (wire chambers) that localize a particle's path. The difference between this and the physically true track is due to the quantization errors induced by the measurement process (e.g. wire spacing, electronics) and due to positioning errors (e.g.

Fig. 1. The five-dimensional space in which all tracks have to lie.

alignment, chamber deformations).

In the n -dimensional space of the measurements, all particles lie near to a *five-dimensional* surface (fig. 1); the point on this *constraint surface* represents the ideal track — a solution of the equations of motion. Owing to the influence of the detector the point of the measurements lies somewhat distant from the constraint surface and the ultimate result of the track fit will be the foot of the measurements' point onto the constraint surface.

We can now summarize:

(i) that track finding consists in finding combinations of measurements which give rise to points reasonably near to the constraint surface,

(ii) and the positions of their foot points in the constraint surface determine the track parameters: the role of track fitting then is to determine this position as precisely as possible.

2. Track finding methods

We not distinguish between the following track finding methods:

2.1. General track finding method

It requires the *a priori* knowledge of the constraint surface for all possible combinations of n different chambers for all possible n ($n > 5$) and would then consist in checking the distance of any combination of n measurements to the corresponding constraint surface.

It is obvious that this general combinatorial method is prohibitive [2].

The method becomes less involved if:

(i) all tracks have the same length of n measure-

ments; i.e. when all chambers are 100% efficient and when the detector lay-out is such that all tracks will be localized by all n chambers. Then we only have to check the distances of the measurement points to *one* constraint surface.

(ii) the detector set-up is such as to measure tracks in projections; the n -dimensional pattern space then decays into lower dimensional subspaces. So the number of combinations to be tried has been reduced.

(iii) the detector set-up and the phase space distribution of the particles to be detected are such that they allow subsampling of track candidates according to distinct detector telescopes.

2.2. Usual track finding methods

They consist in chaining measurements along somewhat *a priori* known track models, i.e. *a priori* known parameterized solutions of the equations of motion (helices, parabolas, etc.). The main conceptual inconvenience is that, *at the same time*,

(i) we have to test whether a particular measurement c_i lies on the track model and

(ii) we want to guess which are the best parameters representing the set of grouped measurements $\{c_i, \dots, c_k\}$ and which are needed for the approximate track model.

This approach to track finding is usually referred to as the approach of *track-following*. It is unsatisfactory from an academic point of view, as it is a completely open game. If in a particular application the percentage of correctly reconstructed trajectories is low, hardly any strategy other than trial and error are available to improve its efficiency.

However, given the present detectors, it is almost always the only practicable approach.

2.3. Global methods

The main requirement is the following: the detector has to be such that the pattern space gets structured in such a way that from the measurements we can guess some or all of track parameters immediately; in general, for trajectories in a magnetic field, this is only the case when each measurement device provides not just one single measurement but already a *group of measurements*.

A usual multiple proportional wire chamber provides the following localization information: a fixed coordinate given by the intersect of the wire plane with its normal and one variable coordinate in the wire plane.

Fig. 2. Parameters for (a) linear and (b) circular projections of trajectories.

For our purpose, we would require that apart from the fixed coordinate both additional space coordinates and also precise measurements of the direction of the trajectory at this space point were given. Then from each group of measurements the track parameters could be evaluated and a *track* would then be defined as a sufficiently large number of *groups providing similar track parameters*.

(i) Let us first consider the special case of trajectories showing up as straight lines in the xy -projection (fig. 2a): For measured x_i, y_i and $(dx/dy)_i$ the position of the intersect x_v with the x -axis is the only still unknown parameter for the track model

$$x = dx/dy(y + x_v)$$

and x_v can be evaluated from each group of measurements separately. Each single track having the same intersect x_v should then show up as a peak in the x_v -distribution. If it is known *a priori* that all tracks originate from only one vertex situated on the x -axis, then x_v is not only a track parameter but also an event parameter and we would have found the vertex of an event even before having found its tracks. If the vertex x_v is known *a priori*, then already the usual planar multiwire proportional chambers – placed for instance at fixed y values – provide enough information to allow global track finding; tracks with different inclination should show up as distinct peaks in the distribution of the quantity $(x_i - x_v)/y_i$.

If the vertex is known and if $x_i, y_i, (dx/dy)_i$ have been measured a unique situation is given: the track model can serve as constraint to check the track hypothesis on the basis of one group of measurements only since each group provides redundant information.

(ii) For trajectories showing up as circles in the xy -projection the situation is more complex (fig. 2.b). Only if one of the three parameters needed to

describe a circle is known *a priori* e.g. the vertex $(x_v, 0)$ of all tracks, can the two other parameters e.g. α, R or x_m, y_m or the momentum $p_x \propto By_m, p_y \propto Bx_m$ can be evaluated from the group of measurements $x_i, y_i, (dx/dy)_i$.

(iii) Given some *a priori* knowledge of the vertex, e.g. $(\dots, y_v = 0, z_v = 0)$, of helicoidal trajectories in space, measuring the space points, e.g. x_i, y_i, z_i and directions, e.g. $(dx/dy)_i, (dz/dy)_i$, would allow the three still unknown helix parameters to be evaluated.

It is evident that we should not think that we have already established a track just by having been able to evaluate a possible (and complete) set of track parameters. As each measured quantum carries errors and as random signals may be present in the detectors (background signals), we need redundant information to be able to test a track hypothesis.

Only if several groups of measurements give rise to – within the errors – the same track parameters will we consider that this set of groups of measurements constitutes a track.

As each group of measurements corresponds to a point in the space spanned by the unknown parameters, the problem of *track finding in the pattern space* of the measurements has thus become the problem of *cluster finding in the parameter space*. *Separated clusters* are identified as *separated tracks* (fig. 3a).

There are global methods where the cluster search

Fig. 3. Parameter space: (a) separated tracks show up as separated clusters. (b) The influence of scaling. (c) The right parameter choice.

can only be started when all points in the parameter space have been evaluated.

For instance, a graph theoretical approach would be to build up a *minimum spanning tree* by linking points of nearest distance *in the parameter space*. By then cutting particular edges, the tree will decay into the desired clusters.

The minimum spanning tree in the parameter space can also be considered as the *minimum spanning tree in the three-dimensional world space* if non-Euclidian metric is used.

For usual detectors, where it is not possible to guess all the track parameters already from every single group of measurements, we may approximate the non-Euclidian distance definition; and when only space coordinate measurements are given, we would end up with looking for a minimum spanning tree in world space using Euclidian metric. Such an approach can be called global only to that extent that by using it we still try to find all tracks in one single go; this primitive Euclidian minimum spanning tree will link the right data points only if the Euclidian inter-track distances are larger than the distances between points on a single track. One major problem with the minimum spanning tree in world space is the track separation since other criteria but the distance may have to be used to cut the tree into pieces.

Anyhow, if the cluster search can be started only when all data points have been considered, the track finding would get quite clumsy and time consuming; for instance, the execution time of the minimum spanning tree cluster search method increases with $N_{\text{tot}}^2 - N_{\text{tot}}$ being the number of points that have to be linked [3].

By studying the possible cluster distributions in the parameter space for a given experiment, it might be possible to derive a *classification scheme* for the analysis of real data and this scheme would allow a decision to be made about which cluster a point of measurements belongs to *even before* that and *irrespectively* of whether the other measurements have already been taken into account.

Only global methods of track finding based on such classification schemes might be competitive with the usual heuristic methods (like track following). Such global methods of track finding then have the great advantage:

(i) that *no combinatorial trying* of grouping measurements to track candidates has to be carried out and

(ii) that *no sequential decisions* as in track following have to be taken.

3. Global track finding by classification

As the advantages of such global track finding sounded so promising I investigated whether this approach could not be used for the analysis of tracks in the homogeneous magnetic field of the *UA1-central-detector*, since its image chambers provide space points and some direction information along almost exactly helicoidal trajectories.

The central detector of the UA1 collaboration is described in two other contributions to these proceedings [4] * and they should be read in order to get an idea of how trajectories of charged particles of the expected complex events will be measured. When traversing the sensitive volume of a single anode wire the ionizing particle will cause pulses at the ends of the wire (fig. 5):

(i) from the beginning of these pulses we obtain a coordinate value along the drift direction — $x_{d,1}$; its accuracy is expected to be $\sigma \approx 250 \mu\text{m}$. This is the accuracy that we can reach when all measurement errors due to alignment, geometrical imprecision and deformation of the chamber structure have been subtracted. These wire positioning errors are quite individual for different wires and thus it is practically impossible to correct for them at the earliest level of pattern recognition. Any track finding should therefore be accepted with a precision of about $\sigma \approx 2 \text{ mm}$.

(ii) Fast analog-to-digital converters (FADC) sample the pulses every 32 ns; the first FADC (see fig. 1 in Calveti et al. [4]) give the track's position along the wire by the charge division method; i.e. the z-coordinate measurement is the mean of the ratio of the bin contents of the left and right end-pulses: $z = \langle z_i \rangle$ with $z_i = d_i(L)/I_i(R)$ where I_i is the content of the i th bin of length 32 ns. The z-coordinate will be measured up to an accuracy of $\sigma \leq 1\%$ of the wire length — i.e. $\sigma \leq 2 \text{ mm}$.

(iii) From the length of the pulses we could deduce some information about the angle α — i.e. the complement of the angle between the drift direction and the trajectory projected into the xy -plane. The precision of this direction measurement depends on the accuracy with which the end of the pulses can be given. The lower accuracy limit is given by defining then end of the pulses to have occurred when the content of the next 32 ns bin is less than a given threshold. By studying the shape of the pulses it

* Ref. [4] gives the most recent description of the UA1 detector.

Fig. 4. The distribution of the parameter k for the data points of one event estimated by a histogram.

should however be possible to improve this accuracy considerably.

(iv) From a systematic variation of the ratios $I_i(L)/I_i(R)$ we could also deduce some information about the inclination of the track with respect to the wire direction; up to what accuracy such measurements could be obtained is still unknown.

Given the geometric lay-out of the UA1 detector and the above summarized qualities of the *position and direction measurements* we can see that we can try to use a global method only for the recognition of the *circular track projections* in the plane normal to the magnetic field (i.e. the xy -plane). And this only, if at least *one circle parameter* is already *known a priori*. It would be most advantageous to know the vertex position – e.g. $(x_v, y_v = 0)$; then, measuring a line segment $x_i, y_i, (dy/dx)_i$, the circle parameters α_v and R for instance would be obtained by

$$\operatorname{tg} \alpha_v = \frac{(dy/dx)_i [(x_v - x_i)^2 + y_i^2]}{(x_v - x_i)^2 - y_i [y_i + 2(dy/dx)_i (x_v - x_i)]},$$

$$R^2 = [1 + (dy/dx)_i^2] \times \left| \frac{(x_v - x_i)^2 + y_i^2}{2[(x_v - x_i)(dy/dx)_i + y_i]} \right|^2.$$

Looking for circles in the xy -plane would then mean a search for clusters in the two-dimensional parameter space spanned for instance by the functions $\operatorname{tg} \alpha_v$ and R^2 .

Before we can continue to try to recognize tracks and to evaluate their parameters using the above space point and direction measurements we have first to return to a discussion of the *features of classification schemes* of global pattern recognition methods.

The most common classification scheme for detecting clusters is to split the parameter space into cells and to then investigate the different cell contents (e.g. *histogram estimator*). In the case of a one-dimensional parameter space [e.g. when the inclination k of straight lines can be evaluated from a point measurement and an *a priori* known vertex – $k = k(x_i, y_i, x_v, y_v = 0)$] the distribution of the evaluated parameter k for a given event might look like that drawn in fig. 4 where it has been approximated by a histogram.

When (almost) all groups of measurements have been used to build up the distribution, a (almost) “completely” filled bin will indicate a track. For instance, in fig. 4 the N entries in the interval around

Fig. 5. Wire signal: (a) a trajectory traversing the sensitive volume of a single anode wire; the drift direction $(-x_d)$ is inclined to the wire plane since the magnetic field is parallel to the wire (z -direction). (b) A sketch of the projection of the trajectory onto the plane normal to the wire.

k_1 indicate a track with the approximate inclination k_1 ; the bin around k_2 contains $2N$ entries, thus a “jet” of two tracks.

Of course, owing to the binning it may happen that one trajectory fills up adjacent bins. This *influence of the binning* may get reduced

(i) by using two histograms with overlapping binning,

(ii) or we could use other estimators: for instance, for the *Rosenblatt estimator* we would enter the result k_i obtained for each group of measurements as a rectangle around k_i with height $\frac{1}{2}$ and content 1.

When using the histograms we have also be aware of another important influence of the bin width. Any change in the bin width corresponds to some kind of *scaling* in the parameter space. And scaling might distort the picture in – for instance – a two-dimensional parameter space in such a way that we could end up with a different partitioning of the data points into clusters (fig. 3b).

Besides scaling, also the *choice of the functions* that shall span *the parameters space* may allow peaks to appear or clusters that are quite different – especially in a multi-dimensional parameter space. Whether, for instance, the radius R or R^2 or $1/R$ should be chosen as one circle parameter depends on the topology of the expected events. And for the point distribution as shown in fig. 3c, the set of parameters $\{p'_1, p'_2\}$ has obviously to be preferred to the set $\{p_1, p_2\}$.

When interpreting the estimated distribution we have to be aware of the possible *background signals* which have given rise to additional entries.

Additional entries may also be caused by *ambiguities* of measurements. Only the value of (dx/dy) is measured by the UA1 detector and not its sign; we thus obtain two values for α_v and R . As consequence the *threshold* in the cell structure has to be set approximately at $N_{th} = N_{tot}/N_{cell}$ where N_{tot} is the total number of groups of measurements and N_{cell} the number of populated cells – provided a flat distribution is found for all additional entries; otherwise the threshold has to be a more elaborated function of the track parameters.

The underlying assumptions in the above statement, saying that a bin content xN allows us to conclude the presence of x tracks, is that all tracks have the same “length” N of measurements irrespectively of the corresponding parameter value.

In most detectors, however, straight tracks for instance of different inclination would be seen by a

different number of wire chambers (detector limits). The value N for which a particular cell can then be considered to be filled up will thus depend on the parameter value: $N = N(k)$.

For the UA1 detector this *number* N is a strongly *varying* function of x_v , R and α_v because of its *anisotropic geometry*.

It is obvious that the *precision* of the predictions made from the groups of measurements needs to be good. The better the track parameters can be evaluated the better the tracks will show up separated in the parameter distribution. The more imprecise the measurements then the more the estimated distribution will be smeared out.

(i) For most detectors, the precision of the predictions differs with the value of the track parameter; thus a *variable bin width* should be chosen such as to match the *variable precision*.

(ii) But the precision does not only depend on the track parameters. For instance, when evaluating the inclination k from the pair of coordinates x_i , y_i (x = fixed coordinate, y = variable coordinate) and a not exactly known vertex position x_v the precision of k decreases with increasing k :

$$k = y_i/(x_i - x_v)$$

$$dk = [1/(x_i - x_v)] dy_i - [k/(x_i - x_v)] dx_v.$$

For this example, the parameter is the more precise the larger the lever arm $x - x_v$. This could be taken as an argument for building a *large detector* enabling long *distance space measurements*.

However, a *tiny detector* has to be recommended when the vertex of straight lines should be evaluated using *direction measurements*

$$x_v = x_v[x_i, y_i, (dx/dy)_i].$$

A short lever arm $x_v - x$ will improve the precision of x_v and thus allow better track separation.

When both distance and direction measurements have to be used the detector should be simultaneously large and tiny. The only way out is to measure the space coordinates and the directions as precisely as possible; for the UA1 detector it would be of great help if $(dx/dy)_i$ – i.e. the length of the pulses – could be measured about one order of magnitude more precisely.

But anyhow, the main hindrance in using a global method of pattern recognition for the circle finding in the xy -projection of the UA1 central detector is that the x -coordinate of the interaction point – the event vertex – is known only up to ± 30 cm. The y -

and z-coordinate of the interaction point however are known to be approximately zero. Thus, in the xy -projection a badly known vertex position is combined with precise coordinate measurements, in the xz - or yz -projection a well known vertex is combined with the badly measured z-coordinate. In order to be able to make use of global methods of pattern recognition, should we not then propose to turn the whole UA1 detector by 90° around the vertical coordinate axis....

Up to now we have been dealing with *pure global pattern recognition*.

A more practical approach might be to *combine global methods* and *track following*. We would use global classification mainly to subdivide all the measured data into subsamples — i.e. by defining *track roads* — in such a way that each subsample contains more or less all measurements which could be caused by the same particle and more or less no noise and no measurements that have been caused by different particles. When searching tracks by help of track following, less combinations had to be tried out. For the UA1 detector we have thought of classifying the data according to the value x_i/y_i — i.e. to subdivide them according to sectors in the xy -plane since the trajectories we are looking for have only small curvature. Originally we thought to perform this sector search in the space of the transformed coordinates

$$x'_i = x_i / (x_i^2 + y_i^2),$$

$$y'_i = y_i / (x_i^2 + y_i^2).$$

For, if the vertex were at $x_v = 0, y_v = 0$ we would thus transform all circles coming from the vertex in the xy -plane into straight lines in the $x'y'$ -plane [5]. The distance between these straight lines and the origin would then be a measure of the inverse of the projected particle momentum p_{xy} . In the limit of infinite p_{xy} the straight lines would pass through the origin ($x' = 0, y' = 0$) — but in this limit the trajectories are also straight lines in the xy -plane and could thus be easily classified by cutting the xy -plane into sectors — if only the vertex were known.

Another way would be to reduce the amount of redundant information along trajectories with about one hundred space points on them by first reconstructing line segments out of the coordinate measurements of (8 or 16) adjacent wires [6]; these line segments might then provide better direction information and thus allow a global pattern recognition method. But then as the line segments are already present in a somewhat ordered form, track following

has to be preferred over any global method since it is at least as fast and far less problematic and anyhow the only method enabling nearby tracks to be separated (jets) by making use of already reconstructed parts of the trajectories and some combinatorial means.

4. Conclusions

My experience with global methods of pattern recognition has led me to conclude that these methods could be applied in the analysis of high energy physics experiments under the following conditions:

(i) whenever a particle's trajectory is localized by a detector device, the measurement of several quantities has to be provided simultaneously; the best would be if all space coordinates and direction angles of the trajectory at this point were measured.

(ii) From each such group of measurements it must be possible to evaluate some track parameters as the values of some vector functions, which have to be chosen such that the groups of measurements caused by the same track show up as cluster in the parameter space.

Whether these functions are complicated or easy mainly depends on the detector lay-out and the magnetic field.

(iii) A particle's path through the detector has to be localized sufficiently frequently (at least 10 times) and

(iv) the geometrical lay-out of the detector has to be such that the number of N of localizations along each track does not vary too much with the track parameters. This variation has to be at least slower than the variation of the precision of the evaluated parameters.

If the parameter distributions have been studied well enough, a simple classification scheme in the parameter space might be found and allow a decision to be made about to which possible track candidate a point of measurements in the parameters space belongs even before, and irrespectively of whether, all the other measurements have already been taken into account.

The advantages of such global pattern recognition can be classified as:

(i) that it would be fast, since no combinatorial trying of grouping measurements has to be undertaken and

(ii) that it could allow for "on-line" track finding

and/or vertex finding. For instance, we could fill up histograms or boolean matrices on-line and stop this filling up at any time to investigate the structure of the parameter distribution – e.g. in order to look for a trigger particle [7].

With this summary of requirements for the application of global pattern recognition methods – and also the problems of other methods of pattern recognition – I would propose the following:

First, physicists should plan future wire chamber detectors in closer collaboration with the people responsible for pattern recognition, and second, the main point I want to make at this wire chamber conference is that future detectors should not only measure coordinates but also direction angles and ultimately space points and line segments.

References

- [1] H. Eichinger and M. Regler, Review of Track Fitting Methods in Counter Experiments, CERN-Yellow-Report (1980) to be published.
- [2] H. Wind, CERN-EP-Seminar (December, 1979).
- [3] C.T. Zahn, IEEE Trans. Computer C-20 (1971) 68; R.L. Page, Algorithm 479, Comm. ACM 17, 6 (1974) 321.
- [4] M. Barranco Luque et al., these proceedings, p. 175. M. Calvetti et al., these proceedings, p. 255.
- [5] M. Bazin, Princeton University PPAD 534E (1964).
- [6] A. Norton, UA1 collaboration (1980) private communication.
- [7] C. Verkerk, CERN-DD/74/27 (1974); J. Boucrot et al. Institut National de Physique Nucleaire et de Physique des Particules, Universite Paris-Sud, LAL-79/33 (1979).