CERN-Data Handling Division
DD/82/14
F. James
September 1982

FITTING TRACKS IN WIRE CHAMBERS
USING THE CHEBYSHEV NORM INSTEAD OF LEAST SQUARES

(submitted to Nucl. Inst. & Methods)

DD-FJ-jt

**FITTING TRACKS IN WIRE CHAMBERS
USING THE CHEBYSHEV NORM INSTEAD OF LEAST SQUARES**

F. James

**Data Handling Division
CERN, Geneva**

September, 1982

Abstract

It is proposed that for fitting tracks in certain detectors such as multiwire proportional chambers, the Chebyshev norm may be better than the usual least-squares method. Simulation results for tracks in ideal MWPCs show that the new method is more efficient, offers a better goodness-of-fit criterion, and produces more reliable track-parameter uncertainties than least squares.

TRACK-FITTING USING THE CHEBYSHEV NORM INSTEAD OF LEAST SQUARES

	<u>page</u>
1. INTRODUCTION	2
Least Squares	3
The Gauss-Markov Theorem	3
Maximum likelihood for uniform error distribution	4
2. THE NUMERICAL METHODS	5
The track fit	5
The parameter uncertainties	6
3. STATISTICAL CONSIDERATIONS	7
The efficiency of the Chebyshev estimator	7
Error symmetrization	8
4. PHYSICAL CONSIDERATIONS	8
Multiple scattering	9
Clusters	9
Determining the "narrow" window size.	11
Delta rays and "wide" windows.	13
"Constant" windows.	13
Pattern recognition errors and crossing tracks	14
5. RESULTS ON SIMULATED TRACKS	14
Experiment NA14	15
The distribution of parameter estimates	15
The parameter errors and goodness-of-fit	17
6. CONCLUSIONS	17
Comparison of the new method with that of Least Squares	17
REFERENCES	18

1. INTRODUCTION

The traditional method of fitting particle trajectories to position measurements is that of least squares. This method is certainly justified for detectors such as bubble chambers for which measurements of points on tracks are approximately Gaussian-distributed and independent. The same method is however also used for quite different detectors such as multiwire proportional chambers (MWPC) in which position measurements are neither Gaussian nor independent. The resulting difficulties have long been recognised, namely:

1. The track-parameter uncertainties as calculated by standard least-squares techniques are often incorrect (generally too small).
2. The overall distribution of the chisquare statistic does not follow the expected law.
3. The value of the sum of squares of residuals after fit is not a reliable measure of the goodness of the fit. (This is related to the preceding point.)

Typically various "corrections" are applied to compensate for these shortcomings. For example, parameter errors may have to be scaled up, or detector planes may be omitted in the fit, and chisquare cutoffs for acceptance of fits may have to be determined empirically since the chisquare distribution obtained may be very far from that expected ideally.

A more fundamental approach has been taken by Drijard, Ekelof, and Grote 1980 who show that the effects of correlations between wire planes can largely explain the observed behaviour of least-squares fits, particularly the slower-than-expected convergence as more wire planes are added. More recently, Duerdoth 1982 has extended these ideas and applied them successfully to a particular MWPC detector.

We present here a more general approach to the fitting of tracks in any detectors which are "quantized", that is, where a signal in one element (or in several adjacent elements) assures that the track trajectory passed within a certain "window", but where the probability of having passed through any given point within this window is a priori uniform. The best example of such a detector is the MWPC (for which we give actual results below) but these techniques can probably be applied as well to silicon strip detectors and other quantized devices.

Although this study was originally motivated by a desire to understand and correct the anomalous behaviour of least-squares track fitting in wire chambers, it has resulted in a totally different method of track fitting, complete with its own theory of parameter estimation, parameter uncertainty calculation, and goodness-of-fit test, which could also be useful at the pattern recognition level.

1.1 LEAST SQUARES

The justification for using least squares is that it is equivalent to the maximum likelihood method when the residuals (measurement errors) are Gaussian-distributed. This can be seen very easily, since the (log-) likelihood function is just the sum of logarithms of Gaussian terms:

$$\begin{aligned}
 L(\underline{g}) &= \prod_i e^{-\frac{(t_i(\underline{g}) - y_i)^2}{2\sigma_i^2}} \\
 \ln L(\underline{g}) &= \sum_i \ln e^{-\frac{(t_i(\underline{g}) - y_i)^2}{2\sigma_i^2}} \\
 &= \sum_i -\frac{(t_i(\underline{g}) - y_i)^2}{2\sigma_i^2}
 \end{aligned}$$

which is, apart from a factor of $-1/2$, the well-known quadratic form of chisquare. The best fit is then obtained by maximizing (or dividing by -2 and minimizing) this expression with respect to the vector of track parameters $\underline{g}$. Notice first-of-all that if the measurements y_i are correlated, then the quadratic form must include off-diagonal terms:

$$\chi^2 = \sum_{ij} (t_i(\underline{g}) - y_i) V_{ij} (t_j(\underline{g}) - y_j)$$

But notice also that even if one takes account of such correlations, the method still corresponds to maximum likelihood if and only if the residuals $t_i - y_i$ are Gaussian-distributed. That is, the quadratic form of chisquare is a direct result of the assumption of Gaussian measurement errors, and we should expect that it would no longer be optimal when residuals are distributed in some other way.

1.1.1 The Gauss-Markov Theorem

Justification for the use of least squares is often based on the well-known Gauss-Markov theorem which states that under the usual regularity conditions and for linear models, the linear least-squares estimator has the minimum variance among all estimators which are unbiased and linear functions of the data. Restricting oneself to linear combinations of the data may have been an important consideration fifty years ago, but in the age of the computer it begins to appear ridiculous. Indeed we will give a counterexample later in this paper, in which the best estimator is considerably better than least-squares, and in addition it is just as easy to compute, although it is not strictly speaking a linear function of the observations.

I prefer to state the Gauss-Markov Theorem as follows: Any estimator which is more efficient than least-squares cannot be reduced to a linear function of the observables. Consider for example the class of estima-

tors which minimize deviations between models and observations using the L_p norm, that is minimizing the sum of the p^{th} powers of the absolute values of the residuals. Then for all values of p from one to infinity, only $p=2$ (least squares) yields an estimator that can be reduced to a linear function of the observables. In recent years, especially due to the interest in robust estimation, there has been considerable study of at least two other members of this infinite class of estimators, namely for $p=1$ and $p=\infty$, which both possess very interesting statistical properties and for both of which reasonably fast algorithms exist. Since they are not linear functions of the data, the Gauss-Markov theorem has nothing to say about these estimators.

1.2 MAXIMUM LIKELIHOOD FOR UNIFORM ERROR DISTRIBUTION

It is natural to seek a solution to our problem by considering the likelihood function for the case of a uniform (rectangular) error distribution. (It is now convenient to consider the likelihood rather than the log-likelihood.) The individual factors in the likelihood function are functions which take on either the value one (if the trajectory passes through the corresponding window) or zero (if it does not). The overall value of the likelihood is therefore either one (if the trajectory passes through all the windows) or zero (if it misses at least one).

The maximum value of this likelihood function will in general be taken on not at a point, but over some region in the space of the parameters describing the track trajectory. Inside this region, where $L=1$, all trajectories are equally "good" since they would all give rise to the same wire hits as were observed. The entire space outside this region corresponds to trajectories which are impossible since they would have resulted in different wire hits. If the track is described by only two parameters, we can visualize this likelihood function as resembling a mesa rising out of the Arizona desert, where $L=0$ in the desert and $L=1$ on the top of the mesa. Real tracks are usually described by five parameters, but five-dimensional mesas are more difficult to visualize.

Although all trajectories corresponding to points in the allowed region are equally good, it is natural to choose the point in the middle of this region as the solution of the fit. The parameter uncertainties are completely defined by specifying the allowed region (the walls of the mesa) since all combinations of parameters lying on the top of the mesa are equally good fits. The walls of the mesa correspond then to the error ellipse of least-squares theory, except that they must be understood in the sense of an absolute bound rather than a statistical uncertainty.

The above exercise demonstrates the basic principles of a complete solution to the problem of fitting with uniform error distribution, in the sense that it allows us to understand the meaning of a best fit, a good fit, and the uncertainties of the fitted track parameters. Unfortunately it does not allow us to find these solutions numerically since one cannot maximize the likelihood function described here with standard maximization techniques, all of its derivatives being almost everywhere zero.

2. THE NUMERICAL METHODS

2.1 THE TRACK FIT

The numerical technique used to obtain the solutions described above is to minimize the maximum deviation between hypothetical and measured track positions instead of minimizing the sum of the squares of the residuals. In mathematical terms we use the Chebyshev norm¹ rather than the L_2 norm. The quantity to be minimized is then:

$$\begin{aligned} C(\underline{a}) &= \max_i |(t_i(\underline{a}) - y_i)/\sigma_i| \\ &= \max_i |r_i(\underline{a})| \end{aligned}$$

where $t_i(\underline{a})$ is the position where the trajectory described by parameters $\underline{a}$ would cross the i^{th} plane, y_i is the middle of the window established by the detector at the i^{th} plane (i.e. the position of the wire which was hit), and σ_i is the distance from the middle of the window to the edge (i.e. half the wire spacing). Notice that σ is no longer a standard deviation, it is here an absolute bound.

It is clear that for any set of parameters $\underline{a}$:

$$\text{if } C(\underline{a}) < 1, \quad L(\underline{a}) = 1$$

$$\text{and if } C(\underline{a}) > 1, \quad L(\underline{a}) = 0.$$

Therefore the Chebyshev statistic $C_{\min}$ is the goodness-of-fit statistic with critical value 1, and if the Chebyshev solution has a value of $C_{\min} > 1$, then there is no allowed region (the track fit is not good; no trajectory will pass through all the windows).

Furthermore, the Chebyshev solution, if $C_{\min} < 1$, will in some sense be in the middle of the allowed region, so that we can consider that both the problem of finding the best fit and that of determining the goodness of fit will be solved if we can perform a Chebyshev fit. Fortunately such techniques have long been used in numerical analysis, particularly in approximation of functions, where it is desired to minimize the maximum deviation between a function and its approximation. As with least squares, only the linear Chebyshev problem has a straightforward algorithm, so we proceed by linearization and iterations. The algorithm used here for the linear Chebyshev solution is that of Barrodale and Phillips 1975, based on the simplex method of linear programming.

¹ The Chebyshev norm is the L_∞ norm since minimizing the maximum residual is equivalent to minimizing the sum of the p^{th} powers of the residuals in the limit as $p \rightarrow \infty$.

The Barrodale-Phillips algorithm exists in the form of a FORTRAN subroutine which accepts as input the vector of residuals $\underline{r}$ and the matrix of first derivatives of each r_i with respect to each element of $\underline{g}$. It returns as output the solution of the linear Chebyshev problem, the point $\underline{g}$ for which $C(\underline{g})$ takes on its smallest value, assuming each r_i is just a linear function of $\underline{g}$. It also returns the vector of residuals $\underline{r}$ at the solution and the value of $C_{\min}$.

In practice the Chebyshev fit is computationally more expensive than a least-squares fit. For the fits described below, the linear Chebyshev solving was considerably slower than the matrix inversion required for the usual least-squares analysis, but for reasons not yet understood by the author, the linearity hypothesis seems to be obeyed better so that fewer iterations are required for the complete Chebyshev fit. Depending on the track parameterization used, this may mean fewer trackings through the magnetic field which may compensate the longer time per iteration.

2.2 THE PARAMETER UNCERTAINTIES

The parameter errors are simply the distances from the solution point to the extremities of the allowed region in the different directions, and the boundary of the allowed region is the set of all $\underline{g}$ for which $C(\underline{g})=1$. Thus the positive error bound on the first track parameter α_1 is the largest value of α_1 for which $C(\underline{g})=1$. Similarly the negative error bound is the smallest value of α_1 for which $C(\underline{g})=1$, and so on for the other components of $\underline{g}$. Geometrically this is very similar to the MINOS² error analysis used for highly non-linear least-squares fits.

The numerical technique used here to find the error bounds defined above is as follows: In order to find the positive error bound on the k^{th} parameter, find the linear Chebyshev solution to the following situation.

1. Let the parameters of the best Chebyshev fit be $\underline{g}=\underline{g}_0$. The residuals $\underline{r}$ and their first derivatives are known at this point.
2. Add a fictitious residual r_{n+1} with value slightly greater than one and all derivatives zero except for the k^{th} which is slightly negative.
3. Find the linear Chebyshev solution for this modified problem. The value of α_k at this solution is the upper error bound of the k^{th} parameter.

² This technique is named after a subroutine of the program MINUIT which implements it. MINUIT is described in James and Roos 1975 and the MINOS analysis in greater detail in James 1980.

The above procedure is followed for each parameter in turn. For negative error bounds, the derivative mentioned in step 2 is set slightly positive instead of slightly negative.

Let us see how this method works. Since the solution to the original fit gave $C < 1$, the additional residual which is everywhere almost equal to one will assure that the solution to the modified problem will occur where $C \approx 1$. If it turns out to be too far from one, additional iterations may be performed with the original value of r_{n+1} and its derivative adjusted to force the solution to be as close as desired to $C=1$. The fact that the k^{th} derivative is slightly negative (positive) will assure that the solution gives the largest (smallest) value of α_k for which $C=1$.

3. STATISTICAL CONSIDERATIONS

3.1 THE EFFICIENCY OF THE CHEBYSHEV ESTIMATOR

Whereas it is clear that the hypothesis-testing properties of the Chebyshev method are always superior to those of the least-squares method for the situations treated here, we have not yet any measure of the relative efficiency of the two methods for the estimation of parameters. Since the Chebyshev method always finds a maximum likelihood solution (if one exists), whereas the least-squares does not necessarily, we expect the former to be asymptotically superior and hopefully even optimal, but we would like to have a more precise idea of what to expect on theoretical grounds.

A somewhat simpler but closely related situation in which the behaviour of different estimators has been studied in some detail is that of the estimation of the center of a distribution [Rice and White 1964, summarized in Eadie et al 1971, p. 187]. The exact problem is: Given a sample of N values $x_i, i=1, N$, drawn randomly and independently from a uniform distribution of unknown central value, what are the variances of the least-squares and Chebyshev estimators of the center of the distribution? The answer³ is:

$$\begin{aligned} \text{Least-squares: } \mu &= \text{sample mean} = (\sum x_i)/N \\ V(\mu) &= 1/12N \end{aligned}$$

$$\begin{aligned} \text{Chebyshev: } \mu &= \text{sample midrange} = (x_{\max} + x_{\min})/2 \\ V(\mu) &= 1/(2N^2 + 6N + 4) \end{aligned}$$

For very large N , the variance of the Chebyshev estimate is smaller by a factor of $N/6$, so that the Chebyshev error is smaller by a factor $\sqrt{N/6}$. For small N , the difference is less striking, but the Chebyshev

³ The variance for the Chebyshev estimate is given in its asymptotic limit $1/2N^2$ in Rice and White 1964. The expression given here and valid for all N was calculated by the author.

estimate remains better for all N right down to $N=3$. For $N=1$ and $N=2$, the two estimators are of course the same.

For situations as complicated as the track-fitting problem, the statistical properties of both Chebyshev and least-squares estimators will depend on the exact configuration of wires and tracks, and we can only speculate that on the average their behaviour should be similar to that of the location estimators given above. Then we would expect the least-squares estimator to be as good as Chebyshev only for six wire planes (since there are now five free parameters instead of one), with the Chebyshev estimator becoming progressively superior as the number of wire planes increases. In the absence of correlations, the Chebyshev estimator should become twice as good as least-squares (i.e. one-fourth the variance) when the number of wire planes $N+5 \approx 25$. As both Drijard et. al. 1980 and Duerdoth 1982 have pointed out, one effect of the correlations is to change the N -dependence of the estimator efficiencies. However, there is no reason to expect the relative advantage of the Chebyshev method to be lost, since it takes account of these correlations correctly.

3.2 ERROR SYMMETRIZATION

The parameter errors can be made symmetric if the fitted parameter values are chosen to be the middle of the allowed region rather than the Chebyshev solution. Such a procedure is somewhat arbitrary since it is not invariant under non-linear transformations of the parameters α , but this particular choice facilitates later stages in the event analysis.

On the other hand, if we adopt the "unsymmetrized" Chebyshev solution, we will have to carry around asymmetric errors for the rest of the analysis, but then the solution is invariant under even non-linear transformations of the parameters, that is, it is independent of the way the tracks are parameterized.

Fortunately, the numerical experiments described below indicate that the symmetrized estimates are slightly more efficient than the unsymmetrized, so we have used symmetrized estimates here. Of course asymptotically the errors should become symmetric so it should not matter which solution is taken.

4. PHYSICAL CONSIDERATIONS

In the above chapters we have outlined a method of track fitting and error analysis which is in some sense statistically optimal for the situation assumed. Unfortunately, optimal statistical efficiency is generally coupled with extreme sensitivity to the correctness of the assumptions made, for the simple reason that statistical optimality is achieved by taking advantage of all the information about the problem. If some of this "information" is not quite true, the results may be catastrophic. Let us consider the assumptions made and their consequences.

4.1 MULTIPLE SCATTERING

In real detectors, there is always some matter to cause multiple Coulomb scattering (and also of course large-angle nuclear scattering, but we do not consider that here). Multiple scattering causes a small, variable, and random deviation of the true particle trajectory from the ideal trajectory which one would get from simply integrating the Lorentz force on the particle in the magnetic field. Since one does not know how to parametrize this deviation, it is simpler to consider that there is an additional measurement error on each point, with the known statistical properties of multiple scattering, namely Gaussian-distributed with a width increasing like the square root of the amount of matter traversed. It is of course not a true measurement error, but since we do not know at present any other way to treat it, we are stuck with an apparent distribution of residuals which will be a convolution of the uniform MWPC distribution and the Gaussian multiple-scattering distribution. The importance of this effect depends on the relative widths of the two components. The multiple scattering in MWPC is often small compared to the detector resolution.

In order to isolate the effects of measurement errors and estimation efficiency, we study here tracks for which the multiple scattering is negligible. For the simulation results given below, the multiple scattering has been 'turned off'.

4.2 CLUSTERS

The term "cluster" we define here to mean the triggering of two or more adjacent wires by a single track because the trajectory actually passed within the region of sensitivity of more than one wire in the same plane. This happens because the trajectory is oblique to the detector plane so that it passes through the sensitive zone of more than one wire. (Other apparent clusters, caused by delta rays and crossing tracks, are treated below.)

We use the following model for cluster formation. Consider the diagram of a wire plane given in figure one. The symbol o designates wires not triggered, and x is for the wires producing a signal. The distance g (the gap) is between the two cathode plates which define the sensitive region of the wire plane. A wire will give a trigger whenever a charged particle crosses its sensitive region given by the rectangle surrounding it. As an example, the slanted line represents a track oblique to the wire plane which triggers four wires.

In the comparison between Chebyshev and least-squares fitting for which we give results below, we consider only ideal chambers in which clusters are formed exactly according to the above model, with no wire inefficiencies and no delta rays. There was however some noise due to pattern recognition errors.

Figure 1. Zones of wire sensitivity in one plane.

4.2.1 Determining the "narrow" window size.

If we assume the model of cluster formation described above, and if we know before the fit the approximate projected angle of the track with respect to the normal to the wire plane (which we must know in order to have starting values for the fit), then we can determine a window within which the track must have crossed the median wire plane. This window is often much smaller than the cluster size and is in no case wider than one wire spacing.

Figure 2. Determining narrow window (first case).

Consider the track shown in figure two, which triggers wires no. 3,4,5 and 6. Any trajectory at the same angle which triggers the same wires cannot lie higher than that marked A (otherwise it would trigger wire 2) nor lower than B (which would trigger 7). The window giving the measurement uncertainty is therefore that indicated by the two arrows and is rather narrow even though the cluster size is four.

Figure 3. Determining narrow window (second case).

A slightly different situation is shown in figure three, where the trajectory cannot be higher than A because otherwise it would not have triggered the lower of the two wires hit, and similarly not lower than B in order to have triggered the upper of the two wires.

We will define windows determined in this way (using both the constraints described above) as "narrow" windows. Clearly in certain favorable situations (such as a track nearly normal to the wire plane which nevertheless triggers two wires) the window may become arbitrarily narrow. In practice we must put a lower limit on this width due to the tolerances in construction of the wire chambers and the limited validity of our model of clustering.

4.2.2 Delta rays and "wide" windows.

We define a delta ray as being the situation where more than one adjacent wires are triggered, one (or more) directly by the particle in question and the other(s) by an electron produced by the particle in question scattering on the chamber gas or on some other matter in the detector. In typical wire chambers operated at atmospheric pressure, about 5% of the wire hits are caused by delta rays.

Delta rays should be treated differently from clusters since some of the wire hits in the "cluster" were not triggered by the particle whose track is being fitted. The question is whether we can distinguish experimentally between delta rays and real clusters.

In many cases it should indeed be possible to distinguish between delta rays and clusters, namely those cases where the delta ray triggers more wires than could ideally be triggered by a track of the known incident angle. In those cases, the analysis to determine the "narrow" window described above would yield a negative window, which would be an indication that something in addition to the track had triggered some wire(s), either a delta ray or another crossing track. Then we would use a "wide" window, determined by using only the constraint that the track must not pass through the sensitive zones of wires that were not triggered, but not requiring that it must pass through the sensitive zones of all the wires that were triggered. The Chebyshev fit remains optimal in this situation, taking perfectly into account the loss of localization both in goodness-of-fit determination and parameter estimation. The effect of the spurious wire hits on the Chebyshev fit is essentially negligible. The least-squares estimator, on the other hand, would be affected since it would try to pull the trajectory over toward the electron.

4.2.3 "Constant" windows.

A common method of treating clusters is to assign to all measurements the same window width (error bar) equal to one wire-spacing, independent of cluster size. (This would be the largest possible "narrow" window width, so we would expect a systematic overestimate of the errors.) Constant windows must be used if one has no knowledge of the track angle, or no faith at all in the accuracy of construction of the wire chambers or the model of cluster formation given here. As such, we would expect constant windows to be robust but not very efficient. In the simulation tests described below we compare narrow windows and constant windows.

4.3 PATTERN RECOGNITION ERRORS AND CROSSING TRACKS

A pattern recognition error occurs when one or more wire hits is associated with the wrong track. This happens for a variety of reasons and with different frequencies for different experiments, but is essentially impossible to avoid completely and is especially serious in events with high charged multiplicities. When the error is due to crossing tracks and produces a cluster part of which is actually due to the track in question, the situation is identical to delta rays and may be treated as described above.

When none of the hits is due to the correct track, it may sometimes be possible to detect the error simply because the cluster size is incompatible with the approximately known angle of the track. In particular, if the cluster size is too small for a wide-angle track, then (assuming 100% wire efficiency) we know it cannot be associated with this track. If the cluster size is too big, it would appear to be a delta ray, in which case the window would be widened so that the effect on the fit would be small.

We would of course hope that, with its very precise measure of goodness-of-fit, the Chebyshev method could actually find and eliminate some of the wrongly-associated measurements. In the simulation described below, this was in fact the case. However, there will clearly be some cases where it is impossible to eliminate a bad point because it is still possible to find a trajectory which satisfies the bad window as well as all the good ones. The effect of this bad point on the fit will of course be negligible since the best Chebyshev fit with the bad point would still have fallen in the allowed Chebyshev region without that point (otherwise it would be detectable). However, the parameter errors now attributed may be smaller due to the additional constraint of the bad point. This would of course always be the case with least-squares fits, but for the Chebyshev method the bad point would affect the errors only if it was an active constraint during the error analysis. It is difficult to compare the seriousness of the effect for the two methods.

5. RESULTS ON SIMULATED TRACKS

The methods described here, in addition to the usual least-squares methods, have been implemented in the programs to fit wide-angle charged tracks for CERN experiment NA14, where they have been used on both real data and simulated data. The simulation program uses the package GEANT [Brun et al, 1978] and the software used for data reduction, pattern recognition, and track and vertex fitting is a special version of TRIDENT [Lassalle et al 1980] adapted for NA14. The results presented below were obtained using this chain of programs.

5.1 EXPERIMENT NA14

CERN experiment NA14 is a high-energy photoproduction experiment using a tagged photon beam which interacts in a dense target placed in the centre of an AEG magnet. Behind this magnet is a set of 7 MWPC known collectively as "stack one". The rest of the spectrometer, consisting of the main analyzing magnet GOLIATH, two more stacks of wire chambers, Cherenkov counters, gamma-ray calorimeters, a muon filter, and various scintillator hodoscopes, was not used in the fitting of the tracks described here.

In the configuration used here, stack one consists of 28 planes of wires grouped in seven chambers of four planes each. The wire spacing was 2 mm. and the chamber thickness 1.4 cm. Less than half the simulated tracks considered here actually traverse all 28 planes before leaving the detector.

5.2 THE DISTRIBUTION OF PARAMETER ESTIMATES

For the purposes of determining the parameter errors introduced solely by the measurement and fitting procedures, a special Monte Carlo run was produced, in which there was no multiple scattering. We present here the distributions of differences between true and fitted values of track parameters under different fitting conditions. These distributions include only the effects of measurement errors in the wire chambers and the efficiency of the pattern recognition and fitting algorithms. Actual experimental uncertainties will be considerably larger due to the effects of multiple scattering, delta rays, chamber inefficiencies and misalignments, and uncertainties on the vertex measurement in the target.

Table 1 gives the winsorized⁴ root-mean-square deviations between true and fitted track parameters for a sample of 233 simulated tracks. Winsorization was necessary because of a few tracks with problems which gave rise to extremely large errors with all fitting methods and would have drowned out the interesting effect if treated exactly, the interesting effect being the width of the central peak of well-behaved tracks ($\approx 95\%$ of the total).

⁴ A sample of values of x is said to be winsorized with respect to $x_{\min}$ and $x_{\max}$ if all values $x < x_{\min}$ are moved to $x_{\min}$ and all $x > x_{\max}$ are moved to $x_{\max}$. This makes the sample moments like the mean and variance less sensitive to the extreme values.

Table 1
R.M.S. Deviations between true and fitted track parameters

1a. for all 233 tracks

	z_{vtx} (cm)	$1/p$ (%)	λ (rad)	ϕ (rad)
	----	-----	-----	-----
Chebyshev fit "narrow"	.074	.0067	.00042	.00047
Least Squares "narrow"	.081	.0075	.00043	.00051
Chebyshev fit "constant"	.086	.0091	.00048	.00066
Least Squares "constant"	.084	.0080	.00048	.00058

1b. for the longest 104 tracks

	z_{vtx}	$1/p(\%)$	λ	ϕ
	----	-----	-----	-----
Chebyshev fit "narrow"	.034	.0046	.00014	.00028
Least Squares "narrow"	.058	.0059	.00022	.00036
Chebyshev fit "constant"	.047	.0075	.00018	.00053
Least Squares "constant"	.044	.0065	.00020	.00044

The above results show clearly that under the idealized conditions studied here and using the correct "narrow" errors, the Chebyshev method is more efficient than least squares. As expected, the superiority is greater for tracks which traverse more wire planes. The average number of planes hit by the 104 longest tracks is about 25, so these results are even consistent with our expectation that the r.m.s. Chebyshev errors should be smaller by a factor of somewhat less than two for 25 wire planes.

Comparing the efficiencies of both estimators with "narrow" and "constant" error bars, we see that the Chebyshev efficiency is very sensitive to having the correct errors, whereas the least-squares method is so robust that it becomes more efficient than Chebyshev when constant errors are used. (The apparent superiority of least-squares with constant errors compared with narrow errors in two cases is not a real effect, but only a statistical fluctuation which remains because of insufficient winsorization.)

The final result is that the Chebyshev method is considerably more efficient than least squares, especially when there are a large number of detector planes, but only if the measurement uncertainties are well understood.

5.3 THE PARAMETER ERRORS AND GOODNESS-OF-FIT

We can study the appropriateness of the track-parameter errors assigned by the various fitting procedures by considering the distribution of "pulls" which we define as fitted parameter value minus true value, all divided by the calculated error. For least-squares fits to independent Gaussian measurements, we would expect all pulls to be distributed as standard normal, with mean zero and variance one. For Chebyshev fits, we do not know what to expect, (that is, the distribution will depend on the exact configuration) except that all pulls should be inside the interval $(-1,1)$.

The results of the simulation show that for "constant" windows, least-squares gives approximately correct errors, underestimated by only about 10%, which could be accountable to pattern recognition problems. However, the overall distribution of fit χ^2 is then very much shifted to low values, about half the tracks having a χ^2 probability of more than 95%. If "narrow" windows are used, the distribution of χ^2 probability becomes nearly uniform, but then the parameter errors are systematically underestimated by as much as a factor of two. This is consistent with the known inappropriateness of least squares for this problem.

The Chebyshev pulls all lie between -1 and 1 except for a few tracks with pattern recognition problems. The shape of the distribution is hard to determine with the limited sample size of this study, but it is peaked in the middle, and might be triangular or a truncated Gaussian. It is unlikely that the distribution of the Chebyshev statistic C_{min} has any particular significance, since it cannot be distribution-free.

6. CONCLUSIONS

6.1 COMPARISON OF THE NEW METHOD WITH THAT OF LEAST SQUARES

Based on both the behaviour expected theoretically and that found in practice under idealized conditions, the Chebyshev method appears clearly superior to least squares for fitting tracks in MWPC. Its estimation efficiency is greater for all tracks, and is strikingly better for tracks with many wire-plane measurements. It provides a reliable test for goodness-of-fit and a reliable criterion for rejection of points or widening of windows, which the least squares method -- even under ideal conditions -- does not.

The practical problems of background noise, multiple scattering, imperfect detection efficiency and random cluster size, which occur in real chambers may be expected to degrade the performance of the Chebyshev method more than that of least squares, if not treated carefully. But in view of the improved tools which the Chebyshev method offers for detecting and taking account of these effects, it is certainly a better base upon which to build future improved practical methods.

REFERENCES

- Barrodale, I., and C. Phillips, ALGORITHM 495, Solution of an overdetermined System of Linear Equations in the Chebyshev Norm, ACM TOMS, 1 (1975), 264-270.
- Brun, R., R. Hagelberg, M. Hansroul, and J. C. Lassalle, GEANT, simulation program for particle physics experiments, CERN report DD/78/2.
- Duerdoth, I., Track fitting and resolution with digital detectors, Nucl. Instr. and Meth., to be published (1982)
- Drijard, D., T. Ekelof, and H. Grote, On the reduction in space resolution of track detectors caused by correlations in the coordinate quantization, Nucl. Instr. and Meth., (1980) 389
- Eadie, W. T., D. Drijard, F. E. James, M. Roos, and B. Sadoulet, Statistical Methods in Experimental Physics. North-Holland, Amsterdam (1971)
- James, F., and M. Roos, MINUIT - a system for function minimisation and analysis of the parameter errors and correlations, Comput. Phys. Commun. 10 (1975) 343.
- James, F., Interpretation of the shape of the likelihood function around its minimum, Comput. Phys. Commun., 20 (1980) 29-35
- Lassalle, J. C., F. Carena, and S. Pensotti, TRIDENT, CERN Report DD/EE/79-2
- Rice, J. R. and J. S. White, Norms for smoothing and estimation, SIAM Review 6 (1964), 243