CERN - DD/DP/66/7
J.M. Howie
October 1966

Some Tests for Checking the Smoothness of Measurements of Bubble Chamber Track Images

1. Introduction

In the analysis of bubble chamber events, a number of measurements are made along each track in each of the photographs (views). The tracks are measured in at least two views, (normally three) and a combination of the measurements in all of the views is used to make a three-dimensional reconstruction of the track within the bubble chamber. Before attempting any spatial reconstruction, one first makes a check on the smoothness of the measurements in each individual view, and if one finds a measurement which is inconsistent in comparison with the other measurements then one rejects this bad measurement. One implies here that any deviation from the smooth curve is caused by a bad measurement, but one must also admit the possibility of a track having a kink in it. To try and recognize a genuine kink in a track from a single two-dimensional picture is both difficult and dangerous, and this problem is best left to a three-dimensional analysis using all the available views. For the purpose of this report, one makes the implicit assumption that any deviation, which arises from a genuine kink in a track, will not have a large enough magnitude to cause a measurement to be rejected. Bad measurements can in principle arise in two ways : (1) the operator is either careless or inexperienced and really makes a bad measurement, or (2) the measuring machine produces a spurious digitizing, which when interpreted as a measurement, bears no relation to the other measurements along the track (except by pure coincidence).

The usual test that is made to check the smoothness of the measurements is to fit a circle through them and to reject any measurement which lies further away from the circle than a given tolerance. To safeguard against the possibility of rejecting good measurements by mistake, this tolerance generally has to be much larger than the known measurement error, because not all tracks appear circular on the film. The main reason for the non-circularity of some tracks is that particles with a comparatively low momentum tend to be slowed down within the chamber and deviate from their original path. There are other reasons why track images on the film are non-circular, but the energy loss effect is the most important one.

The simple circle fit test does however find most of the bad measurements, but there are others which it cannot possibly discover. In order to weed out all the obvious bad points, more elaborate checks are needed and a method with this aim in mind is described in the following sections.

2. Classification and Detection of Bad Measurements

It is convenient to classify bad measurements into three categories :

1. The measurement is so bad that it could not possibly correspond to a point in the chamber.

2. The bad measurement could refer to a point within the chamber, but it is still quite remote from the track.

3. The bad measurement lies somewhere in the region of the track.

Bad measurements of type (1) can easily be detected by knowing the range of values of the x and y coordinates of points on the film. If the coordinates of any measurement violate these known limits then the measurement can immediately be flagged as a bad measurement.

One may consider bad measurements of types 2 and 3 together and it will be supposed, for the argument, that there is just one bad measurement amongst a set of good ones. If one fits a circle by least squares through these measurements one should obtain one of two possibilities, which are shown in Figs 1(a) and 1(b). In Fig. 1(a) the bad measurement is in the region of the track and belongs to type (3). This kind of bad measurement may be detected in the usual way of comparing the deviations of individual measurements from the circle with a certain tolerance. This test can however be made more powerful by making the tolerance a function of the radius of curvature of the fitted circle. As was pointed out in the Introduction, only low energy tracks are liable to deviate appreciably from a circle, whereas high energy tracks (large radius of curvature) should fit a circle very closely.

In Fig. 1(b) one has a bad measurement of type (2). In this case the bad measurement fits the circle very well and would not be detected by looking at the individual deviations. Instead, though, one may consider the angular separations of neighbouring measurements along the circle. By the angular separation of two measurements one means the angle between the two radius vectors joining the centre of the circle to the measurements (see Fig. 1(c)). A bad measurement of the kind shown in Fig. 1(b) will clearly have a much larger angular separation from its neighbouring measurement, than the average angular separation between the other measurements, and hence this gives one a good means of detecting bad measurements of this type. Of course for this kind of test it is implicitly assumed that the measurements are correctly ordered along the circle, whereas when they are input to the geometry program, they could be in some other order, particularly the bad measurement, which in this case has almost surely been produced by a spurious digitising from the measuring device. Now one of the initial tasks of a geometry program is to order the measurements relative to the beginning of the track and it can be seen that a useful by-product of making a circle fit to the measurements is that it also provides an excellent method for ordering the points; i.e. the angular separation of each measurement from the first measurement can be computed and the measurements ordered accordingly.

3. Details of a method for rejecting bad measurements

The ideas presented in the last section described a set of tests which will detect bad measurements of various types. In this section these ideas are combined into a single method which may be used for detecting and rejecting bad measurements along a track. The method is presented as a series of stages which correspond to stages within a computer program. It is assumed that only one bad measurement is allowed per view of each track and if the method detects two bad measurements on one view, then that view is rejected for the track. (In practice, measurements rejected at stage 1 of the method could well be omitted from this count of bad measurements.)

Stage 1 : The x and y coordinates of each measurement are checked against preset maximum and minimum values. If either coordinate lies outside its permitted range, the measurement is rejected.

Stage 2 : A circle is fitted to the measurements and the radius and coordinates of the centre of the circle determined. An explanation of how one fits the 'best' circle to a set of points is given in Appendix I.

Stage 3 : Using the circle fitted in stage 2, one then determines the angular separation (θ) of each measurement from the first measurement, in a clockwise direction, say, the range of θ being from $0-2\pi$ radians. At this stage one does not know whether the track is going in a clockwise or anti-clockwise direction, but one can decide this by counting the number of measurements which have a θ value less than π, and the number for which θ lies in the range $\pi-2\pi$. If the majority of measurements have an angular separation in the range $\pi-2\pi$, then one assumes that the track is really travelling in an anti-clockwise direction and the angular separation θ is replaced by $2\pi-\theta$. For example, the track in Fig. 2 has 3 points which lie in the range $0-\pi$ and 8 points in the range $\pi-2\pi$ and is hence deduced to be going in an anti-clockwise direction.

One also has to be careful that one does not make the implicit assumption that the first measurement in the list of measurements is necessarily the first measurement along the track, though it is reasonable to assume that it does lie somewhere near the beginning. This situation can be allowed for by computing the angular separations of the measurements from a point somewhere before the first measurement instead of the first measurement itself. If one takes this point to be at a small angular separation, θ_o, from the first measurement then one simply changes the values of θ one has already computed to $(\theta + \theta_o)\ |2\pi|$ (i.e. modulo 2π). For example in Fig. 3, there are 5 measurements along a track and one assumes that their real order is 2, 1, 3, 4, 5, though they have been measured in the order 1, 2, 3, 4, 5. Measurement 2 would initially have a θ value of nearly 2π, whereas the other measurements would all have small θ values. However, after referring the measurements to the point P instead of to the first measurement, the θ value of measurement 2 has a value of the same order as the others.

These final θ values for the measurements are now used to order the measurements along the track in terms of increasing values of θ.

Stage 4 : After ordering the measurements, one can now compute the angular separation (ϕ_i) between consecutive measurements

$$\text{i.e.} \quad \phi_i = \theta_{i+1} - \theta_i$$

The largest value of ϕ_i is extracted and the mean value of the remainder found. If one supposes there are n measurements, then the mean ϕ is given by

$$\bar{\phi} = \frac{1}{(n-2)} \left[\sum_{i=1}^{n-1} \phi_i - \phi_{max.} \right]$$

One wishes to use these values of ϕ_i to try and extract bad measurements of the kind shown in Fig. 1(b). One assumes one has a bad measurement of this type if a gap between two measurements is too large, i.e. if $\phi_i > K\bar{\phi}$, where K is some pre-assigned constant. As one is only interested in removing one bad measurement and not more than one, it is safe to assume that a single bad measurement of this type will appear either as the first or last measurement in the reordered set. One now considers the various possibilities.

a) All values of $\phi_i < K\bar{\phi}$: - no measurement is rejected and one proceeds to stage 5.

b) More than one value of $\phi_i > K\bar{\phi}$: - reject the view for this track.

c) One value of $\phi_i > K\bar{\phi}$, but it is not ϕ_1 or ϕ_{n-1} : - in this case one has a large gap somewhere in the middle of the track, as shown in Fig. 4 and it is hard to say whether one should reject the view for this track or not. Our decision has been to keep the view, on the principle that the gap was probably made deliberatly (e.g. a confused region on the film).

d) $\phi_{n-1} > K\bar{\phi}$: - reject the last measurement. Before proceeding to stage 5 one should first return to stage 2 to refit the circle.

e) $\phi_1 > K\bar{\phi}$: - reject the first measurement and then as above in (d). However, in this case one first needs to make an additional check. Let us consider the situation in Fig. 5. The first measurement is a wild point and the other measurements, 2-6, are good. Because the measurements are ordered relative to the first measurement, the points 2-6 will have been put into reverse order, so that measurement 6 is now the second point, etc. In this case one cannot just reject the first measurement and carry on, as the remaining measurements would then be left in the wrong order. It is impossible to introduce checks for every conceivable arrangement of points, but some test is certainly necessary here, as it is a not improbable situation. A simple and probably sufficient test is to check whether the original second measurement is the same as the reordered second measurement. If they are not the same, then one supposes that the measurements have been reversed and they are put back into their original order.

<u>Stage 5</u> : At this stage one can presume that no widely inaccurate measurements are present amongst the data and also that one knows the parameters of a circle fitted to the measurements. One now proceeds to test for bad measurements of the type shown in Fig. 1(a). One makes the standard test of comparing the perpendicular deviations of the measurements from the circle with a certain tolerance. This tolerance could be taken as a constant for all tracks, but it is more profitable to make it a function of the radius of the fitted circle. A discussion on the form of this function is given in Appendix II. If the tolerance is exceeded, then the measurement, which gives the maximum deviation is removed and a second circle fitted to the remaining measurements. The deviations of the measurements from this new circle are now computed and in turn compared with the tolerance. If the tolerance is again exceeded, this implies that there are at least two bad measurements and hence the view should be rejected. Before rejecting the view completely though, there is one further test one should make. The assumption, which has has been implied so far is that a bad measurement will automatically have the largest deviation, whereas, although this is true in the vast majority of cases, it is known that there can be exceptions when the bad measurement appears at one end of the track. A typical example is shown in Fig. 6, where the bad measurement is the first point, but the second point is flagged as the one with the largest deviation. In order to take account of this possibility, one should therefore, in the cases where two measurements have been rejected, make a third fit, in which the two rejected measurements are replaced and instead, the end point nearest the first rejected measurement is removed. If this third fit is still not acceptable, then the view is rejected. This sequence of tests has been found to be perfectly satisfactory in practice.

Another, more complete approach would be to say that if the first fit is not acceptable, then one tries many second fits, in which each measurement in turn is removed from the set of measurements. The fit, which produces the smallest sum of squares of the deviations, could then be considered to be the best fit and the measurement, which was omitted from this fit, becomes the rejected measurement. This approach, though, must become rather time-consuming when there are many points along the track, and for this reason, is not recommended.

4. Conclusions

The methods described in this report were incorporated into the CERN bubble chamber geometry program, THRESH [1], at the beginning of 1966. The previous smoothness test in THRESH was simply to compare the deviations of the measurements from the fitted circle with a constant tolerance. Before the new routines were introduced into the program, tests were carried out on a few hundred events. The results of these tests were very satisfactory. All known isolated measurement errors, which had remained undetected before, were now detected and accordingly deleted.

Acknowledgements

I would like to thank Mr. A. Meyer and Mrs V. Alles-Borelli for their help in the work described in Appendices I and II respectively.

Appendix I

On fitting a circle to a set of track measurements

One assumes from the outset that the 'best' circle fit to a set of points is defined as that fit which minimises the sum of the squares of the perpendicular deviations of the points to the circles; i.e. if the coordinates of the centre of the circle are (λ, μ), the radius of the circle is R and the distance of a point P (x_i, y_i) to the centre of the circle is r_i, then the deviation of the point P_i from the circle is

$$\delta_i = r_i - R$$

and the 'best' circle fit minimises $\Sigma\delta_i^2$. If one writes down the equation of the circle as

$$x^2 + y^2 - 2\lambda x - 2\mu y - c = 0$$

then

$$\delta_i = \sqrt{(x_i - \lambda)^2 + (y_i - \mu)^2} - \sqrt{c + \lambda^2 + \mu^2}$$

The problem of minimising $\Sigma\delta_i^2$ with respect to the three parameters λ, μ and c is a non-linear one and the solution would have to be found by an iterative method.

In practice it is not necessary to go to the lengths of an iterative procedure, which would, of course, be time-consuming for a computer program. Instead let us consider the quantity

$$\begin{aligned} \epsilon_i &= x_i^2 + y_i^2 - 2\,x_i - 2\mu y_i - c \\ &= (x_i - \lambda)^2 + (y_i - \mu)^2 - (c + \lambda^2 + \mu^2) \\ &= r_i^2 - R^2 \\ &= (r_i - R)\,(r_i + R) \\ &= \delta_i(2R + \delta_i) \approx 2R\delta_i \end{aligned}$$

One sees therefore that the quantity $\epsilon_i/2R$ is practically the same as δ_i and hence one can try to minimise the quantity $\Sigma(\epsilon_i/2R)^2$ instead.

One first performs a rotation and translation of the coordinate system which will make the track approximately parallel to the x-axis. Let (x_1, y_1) and (x_n, y_n) be the first and last measured points respectively. The point $(x_1 y_1)$ becomes the zero of the new coordinate system and the axes are rotated through an angle θ given by $\theta = \tan^{-1}\left[(y_n - y_1)/(x_n - x_1)\right]$ (see Fig. 7a). In this new coordinate system one sees that for all tracks which do not have too large a sagitta, the absolute value of the y coordinate of the centre of the circle is approximately the same as the radius (see Fig. 7b).

$$\text{i.e. } R \approx |\mu|$$

using this relation one may write

$$\epsilon_i' = \frac{\epsilon_i}{2\mu} \approx \frac{\epsilon_i}{2R} \approx \delta_i$$

66/1380/3/mg

The quantity ϵ_i' may be conveniently expressed as

$$\epsilon_i' = a_o(x_i^2 + y_i^2) - 2a_1 x_i - y_i - a_2$$

The problem of minimising $\Sigma(\epsilon_i')^2$ with respect to the parameters a_o a_1 and a_2 is a linear one and hence is directly solvable.

One has shown that the quantity ϵ_i' is practically equivalent to the real deviation δ_i for all tracks with small sagittae, which in fact comprises the majority of bubble chamber tracks. For tracks which have larger sagittae, one may still be reasonably confident that minimising $\Sigma(\epsilon_i')^2$ will produce the best circle fit. This is because tracks with large sagittae must as a consequence have small radii, and one can see that minimising $\Sigma(\epsilon_i')^2$ or even $\Sigma\epsilon_i^2$ for that matter, becomes less sensitive to the value of R, when R is small. The one region where the method will fail though, is in the region of the singularity, $\mu=0$; i.e. those tracks measured through an angle of about 180°. For these tracks, one must minimise some other quantity and since the value of R must be very small in this case, the quantity $\Sigma\epsilon_i^2$ should be quite sufficient.

It would be wrong though, to assume that minimising $\Sigma\epsilon_i^2$ will always produce the same answer as minimising $\Sigma\delta_i^2$. When the radius of the track is large and the measurements not particularly accurate, it is possible to arrive at quite different solutions. An example of this phenomenon is given in Fig. 8. In order to show the difference between the two fits more clearly, the y-scale has been greatly magnified, which results in the circle fits appearing as ellipses. The measured points are denoted by the large crosses; curve A is the circle fit obtained by minimising $\Sigma\epsilon_i^2$ and curve B is the circle produced by minimising $\Sigma(\epsilon_i')^2$.

<u>For curve A</u> : R = 213; $\Sigma\delta_i^2 = 0.0139$; $R^2\Sigma\delta_i^2 = 629$

<u>For curve B</u> : R = 796; $\Sigma\delta_i^2 = 0.0037$; $R^2\Sigma\delta_i^2 = 2353$

Both these circle fits, in fact, obviously have the wrong sign of curvature compared to the real curve, but the point of the example is simply to demonstrate that two different solutions are possible.

As an additional check to make sure that minimising $\Sigma(\epsilon_i')^2$ really did produce the same answer as minimising $\Sigma\delta_i^2$, a comparison of the results of the two methods was made for about 1000 tracks. To minimise $\Sigma\delta_i^2$ one has to start with a first approximation to the parameters λ, μ and c (say λ_o, μ_o and c_o), to express the parameters in terms of small deviations from the initial values, i.e. $\lambda = \lambda_o(1 + \delta\lambda)$ etc, and then to linearise the least squares equations in terms of the new parameters $\delta\lambda$, $\delta\mu$ and δc. After solving for $\delta\lambda$, $\delta\mu$ and δc, one replaces λ_o, μ_o and c_o by the new values of λ, μ and c respectively and then repeats the process. One judges the iteration to have converged when the increments $\delta\lambda$, $\delta\mu$ and δc become less than some pre-assigned value.

As expected, comparison of the results of the two fits for the 1000 tracks showed no significant difference whatsoever. The program for finding the solution which minimises $\Sigma\delta_i^2$ and the work involved in making the comparisons was carried out by Mr. A. Meyer* whilst working as a visitor in the DD Division of CERN.

For a further discussion on the subject of circle fits, the reader is referred to the geometry program report by Solnitz et al.[2].

* on leave from the Forschungsstelle für Physik Höher Energien, East Berlin

Appendix II

An expression for the tolerance as a function of the radius, in a circle fit test

As was explained in the Introduction, high energy particles travelling through the bubble chamber in a magnetic field leave tracks which generally appear very circular on the film, and the higher the energy of the particle, the smaller is the curvature. On the other hand, low energy particles are found to noticeably lose energy as they pass through the chamber, and their tracks correspondingly deviate from a circular path. When one applies a circle fit test to detect any bad measurements, one can use this knowledge to make a more accurate test, by expressing the tolerance, which one is using to throw out a measurement, as a function of the fitted radius of the circle. In order to determine a suitable form for this function, the following data was accumulated.

The measurements of about 600 track images were plotted on the Calcomp Plotter at CERN. These measurements were examined very carefully by eye to see if any bad measurement could be detected. The accuracy of the Plotter is such that good measurements lie perfectly on a smooth curve, and the resolution which one obtains by looking along the line of measurements (i.e. putting one's eye close to the plane of the plotted points) is such that errors of 10-20 μ in the film plane are clearly observed. If one detected a bad measurement (i.e. deviation from the smooth curve), this track image was discarded, but if all the measurements appeared to be good, then one recorded the radius of the circle fitted to these measurements and the maximum deviation of a measurement from this circle.

A plot was made of radius (R) against maximum deviation (d) and this is shown in Fig. 9, each track being represented by one dot. The radius is the radius of the track image projected onto the bottom surface of the front glass (given in cm), and the deviation is the deviation in the film plane (given in microns). In order to make a reasonable scale for the radius, the maximum radius shown is 1000 cms, but the behaviour of tracks with radii above this value is no different from other tracks with a large radius.

All deviations shown on this plot are genuine deviations of well measured points, therefore any tolerance, which one is going to use to reject bad measurements, must lie above this level. One sees that quite a low tolerance could be used for tracks with a large radius, but this would have to increase sharply as the radius gets smaller. Because of this behaviour, one has chosen to express the tolerance as a second order function of the radius for values of R between 0 and some value R_o, and to keep the tolerance fixed for $R > R_o$.

i.e. let d_2 be the tolerance for $R = 0$
and let d_1 be the tolerance for $R = R_o$

then the tolerance may be expr ssed as

$$\left.\begin{array}{ll} d = d_1 + (d_2 - d_1)\left(1 - \dfrac{R}{R_o}\right)^2 & \text{for } 0 < R < R_o \\ d = d_1 & \text{for } R \geqslant R_o \end{array}\right\} \quad (1)$$

For the data given in Fig. 9 one has used the values

$$d_1 = 17.5 \quad d_2 = 77.5 \quad R_o = 400$$

and in Fig. 9, this tolerance is drawn as the continuous curve above the plotted points.

It should be noted that the data contained in Fig. 9 was obtained from tracks belonging to a particular experiment (i.e. 5.7 Gev/c anti-proton beam in the Saclay 80 cm hydrogen bubble chamber), and will not necessarily be valid for other experiments. The general behaviour should be the same though, and one could use the function given above in equation (1) with appropriate values for the parameters d_1, d_2 and R_o.

The work involved in accumulating the data shown in Fig. 9 was carried out by Mrs V. Alles-Borelli*.

* now at the Instituto Nazionale di Fisica Nucleare, Bologna

References

1) Track chamber Program Library, section THRESH; CERN.

2) F.T. Solmitz, A.D. Johnson and T.B. Day : Three View Geometry Program. UCRL Alvarez Group "Programmers' Note" P 117 (page 38). May, 1965.

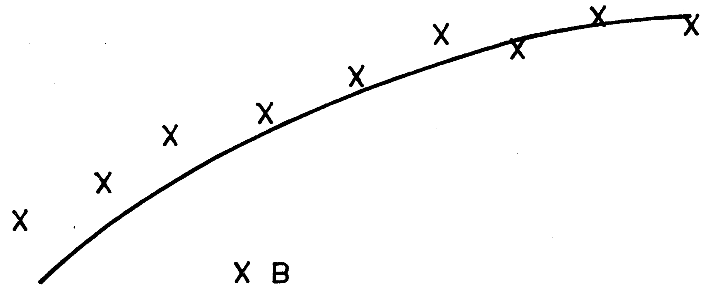


FIG. 1(a)

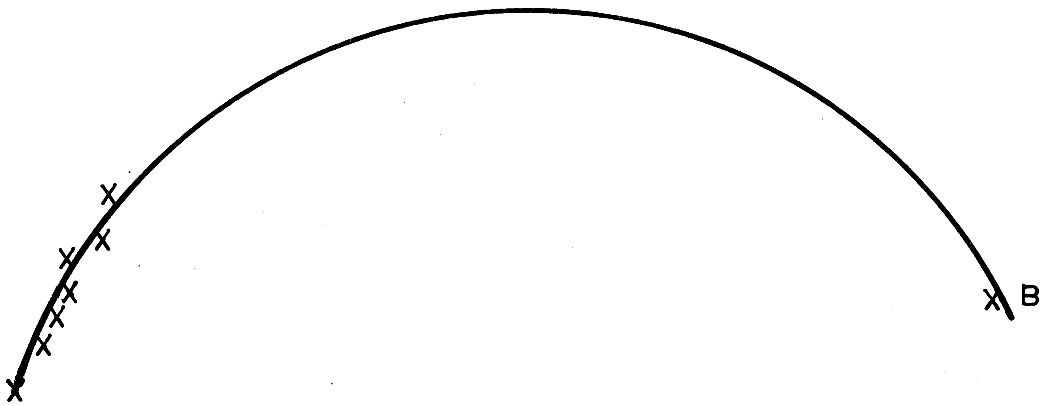


FIG. 1(b)

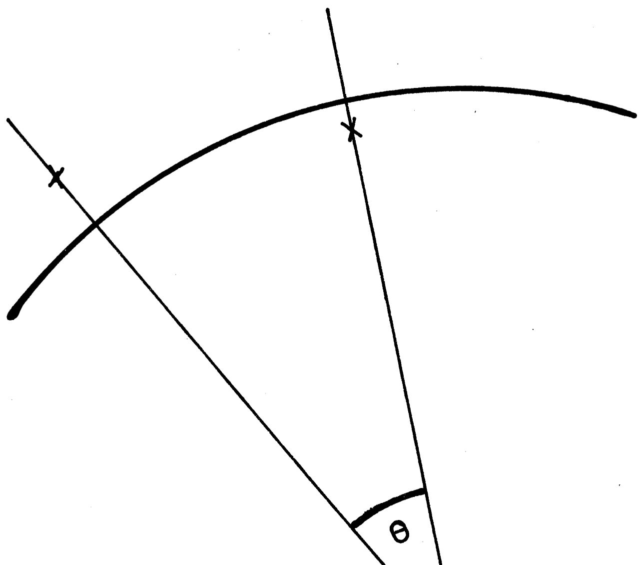


FIG.1(c)

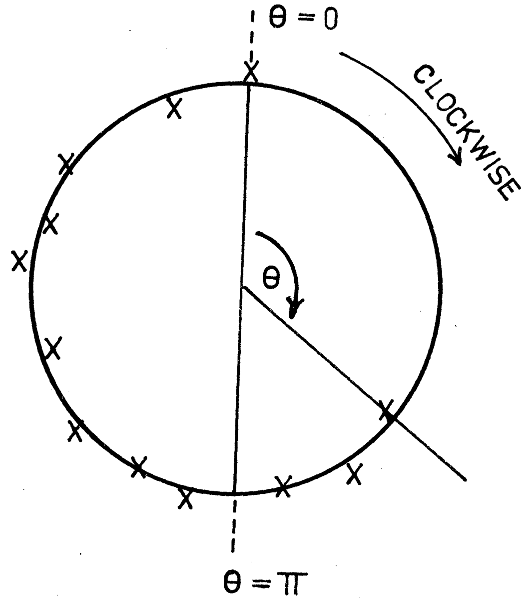


FIG. 2

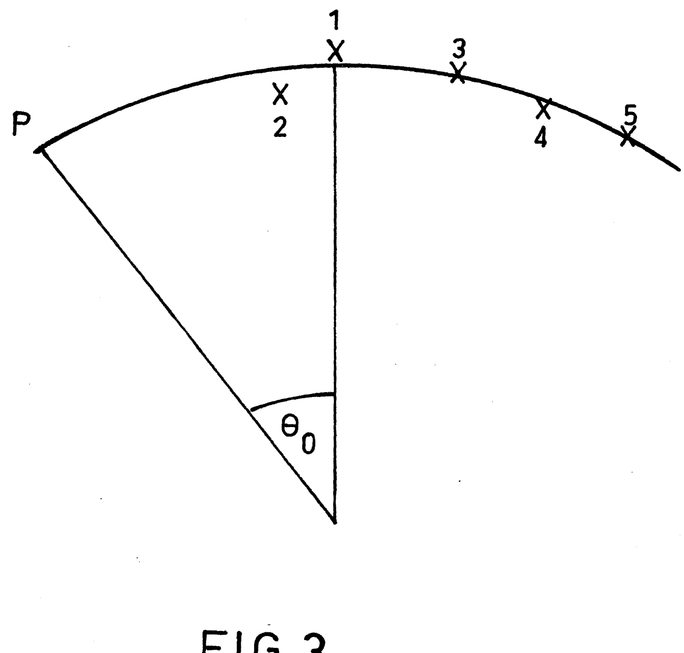


FIG. 3

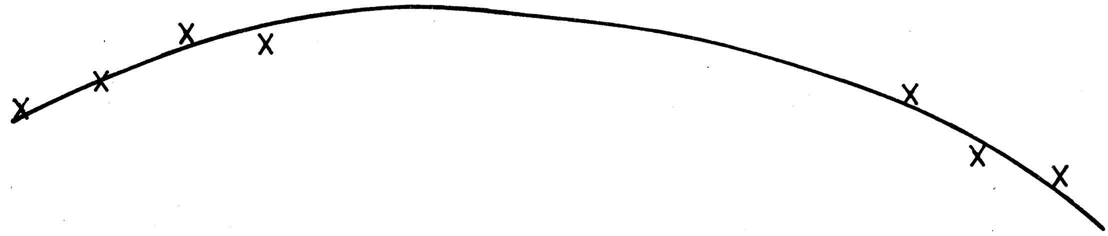


FIG. 4

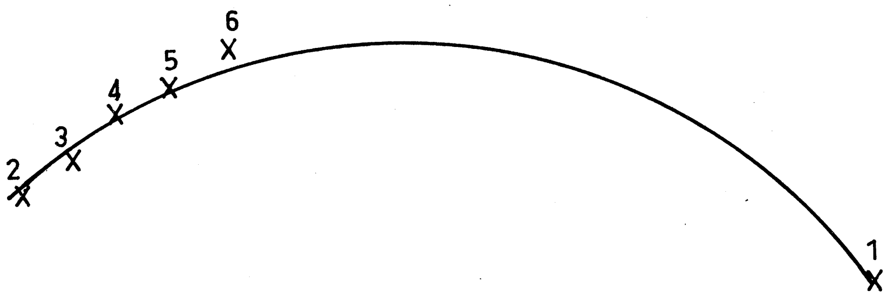


FIG. 5

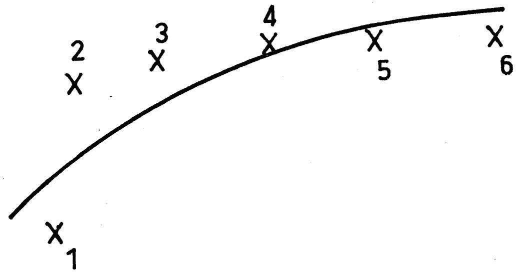


FIG. 6

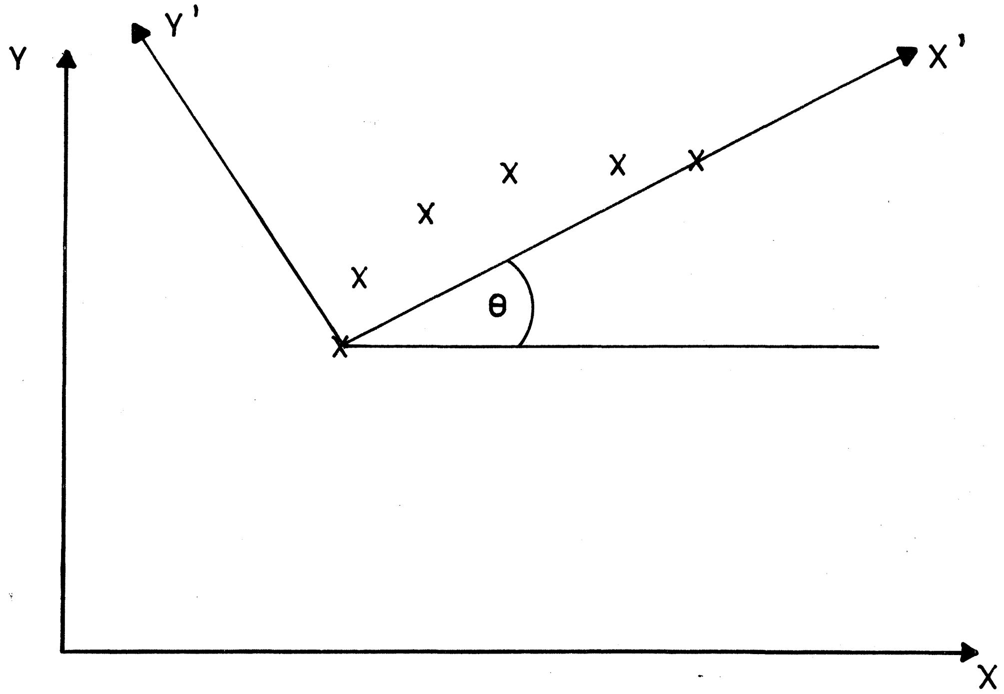


FIG. 7(a)

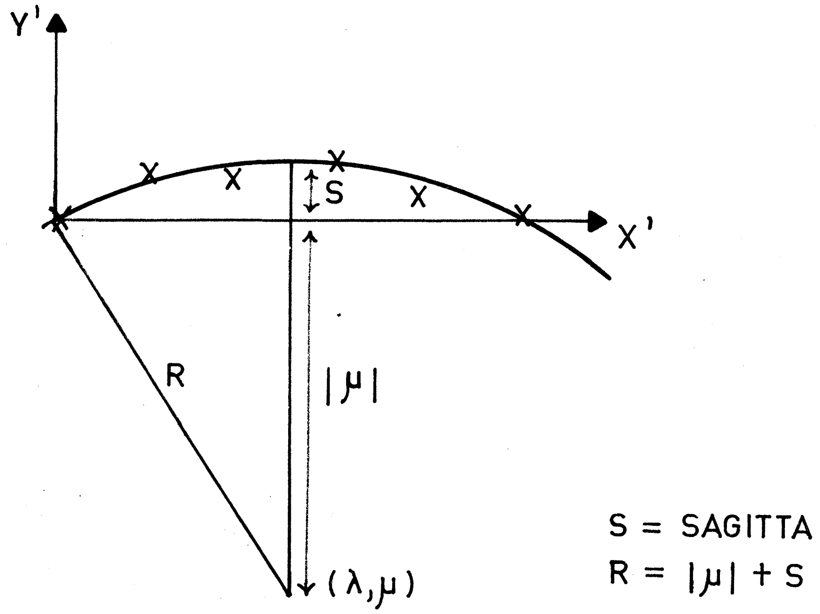


FIG. 7(b)

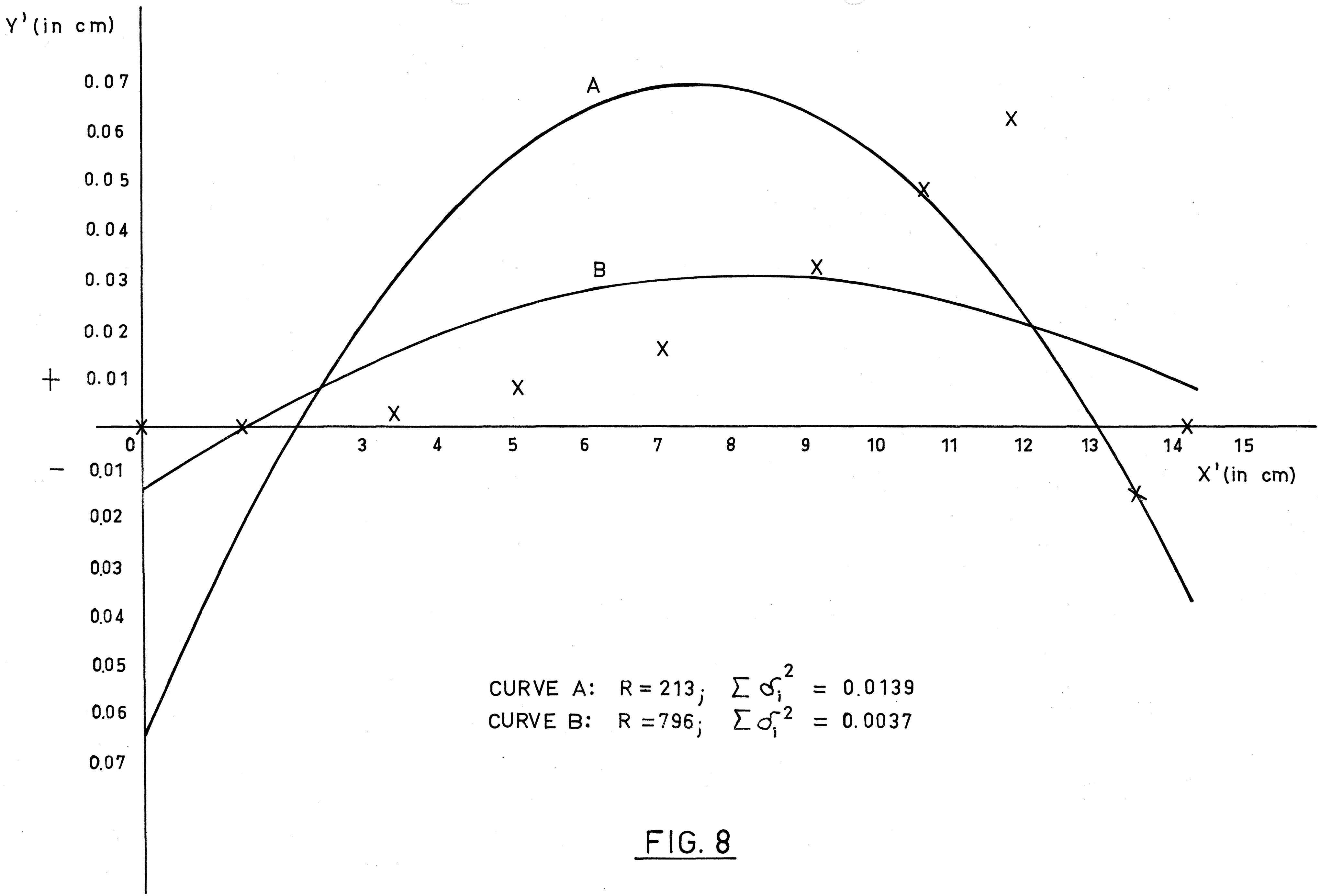
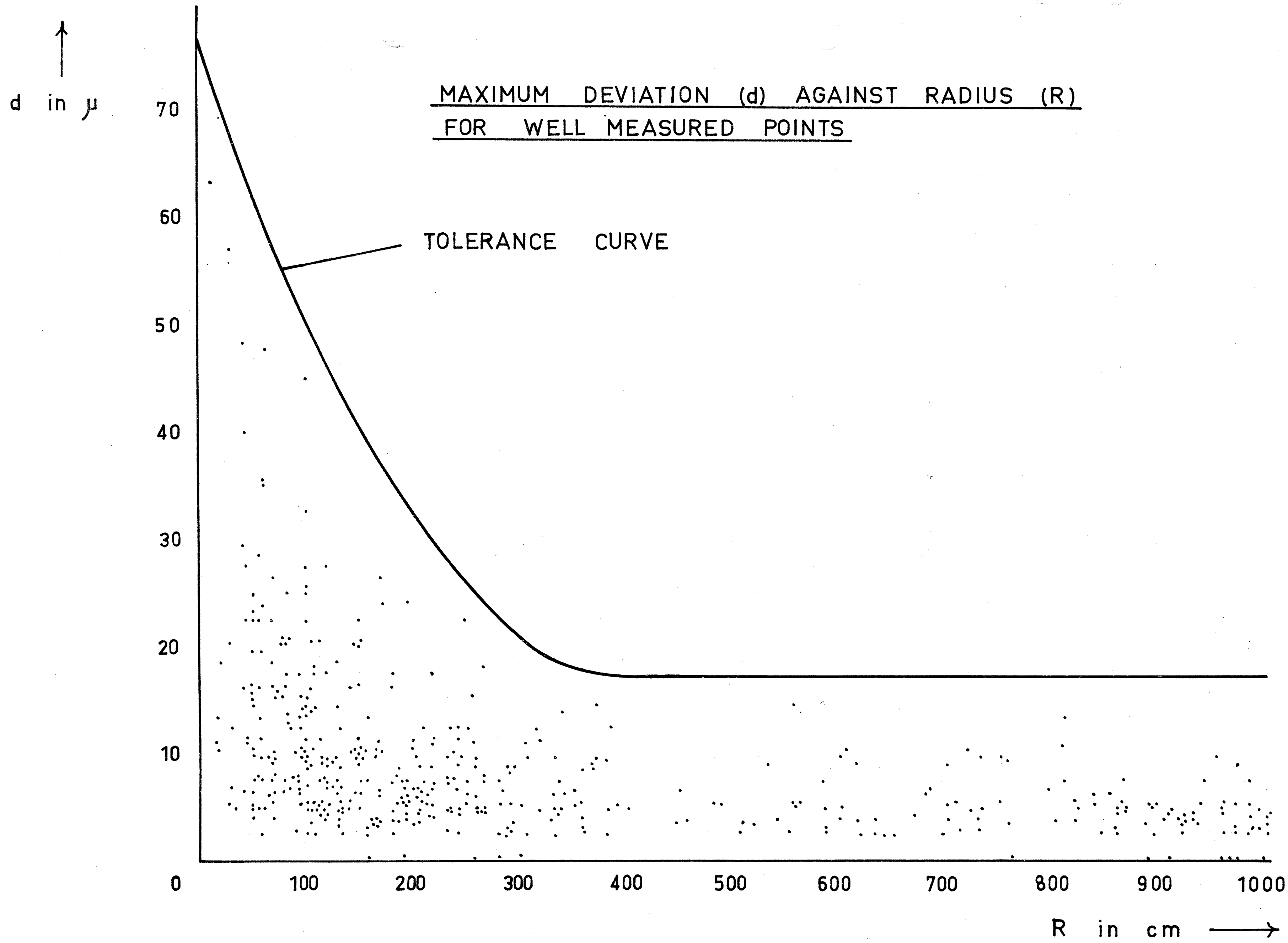

Y'(in cm)
0.07
0.06
0.05
0.04
0.03
0.02
+ 0.01
0
3
4
5
6
7
8
9
10
11
12
13
14
15
X'(in cm)
− 0.01
0.02
0.03
0.04
0.05
0.06
0.07
A
B
CURVE A: R = 213; Σ δ_i² = 0.0139
CURVE B: R = 796; Σ δ_i² = 0.0037

FIG. 8



d in μ
MAXIMUM DEVIATION (d) AGAINST RADIUS (R)
FOR WELL MEASURED POINTS
TOLERANCE CURVE
70
60
50
40
30
20
10
0
100
200
300
400
500
600
700
800
900
1000
R in cm

FIG. 9