A UNIFIED APPROACH TO THE MEASUREMENT ANALYSIS OF NOMINALLY CIRCULAR AND CYLINDRICAL SURFACES

D. G. Chetwynd

Ph.D. Thesis, 1980

University of Leicester

D. G. CHETWYND

A Unified Approach to the Measurement Analysis of Nominally Circular and Cylindrical Surfaces.

The customary procedures of roundness measurement have been developed in response to particular needs as they have arisen, incorporating approximations as appropriate. Consequently, the direct extension of these procedures to more complex measurements such as "cylindricity" is a questionable exercise. The present work develops a mathematically consistent description of the processes underlying the measurement and analysis of roundness. From this are derived analytical methods appropriate to measurements for which instrumentation is, in some cases, yet to become available. New, highly efficient algorithms for solving the minimum circumscribing, maximum inscribing and minimum zone reference figures are also produced.

The method adopted identifies important features of roundness measurement such as eccentricity and radius suppression as translations between co-ordinate frames associated with the workpiece and instrument. Reference figure fitting is expressed formally as a problem in optimisation and the standard methods of Operations Research applied to it. All four standard reference circles are re-examined in this way leading to generalisations of measurement conditions and improved solution methods. Earlier advocacy of the limacon as a reference figure is confirmed and extended. The relationship of circular and limacon references is studied and an eccentricity ratio shown to be a suitable control over the approximations used in practice.

The use of "limacon cylindroids" seems to provide a working approximation for the measurement of cylindricity. It is recommended that cylindrical reference figures be fitted by standard techniques of linear programming rather than by special algorithm.

CONTENTS

Preface

The work reported in this thesis was performed on a part-time basis at the Department of Engineering, University of Leicester mainly while the author was employed in the Metrology Research Group of Rank Taylor Hobson Ltd., Leicester and was supported by that company. It was completed at the Department of Engineering, University of Warwick.

The author is indebted to his academic and industrial supervisors, Dr. P. H. Phillipson and Prof. D. J. Whitehouse for their guidance and encouragement and to the management of Rank Taylor Hobson, particularly Mr. R. C. Spragg, for the facilities and support provided. Thanks are extended to colleagues at both establishments for contributing their time and knowledge to discussions; at Leicester, to Mr D. Kinsey with whom earlier collaboration led up to this work and Mr. P. T. Budd who patiently performed "blind" experiments; at Warwick, to Prof. J. A. Shercliff for allowing the necessary time to complete this work and Dr. F. A. McKee for his helpful comments on the manuscript.

Finally, my thanks go to Miss I. Spence for typing the manuscript and to my wife for her patience!

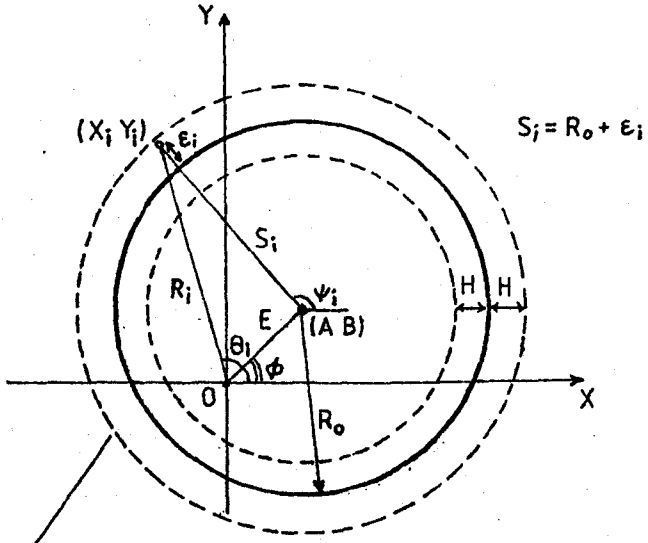
Nomenclature

Many of the terms used throughout this work have related but not identical existences within different frames of reference, here called instrument co-ordinates and chart co-ordinates. In these cases the property is allocated the same symbol from the English alphabet with the upper case character referring to instrument co-ordinates and the lower case to the chart co-ordinates. For example, from the list below:

$$a = MA$$

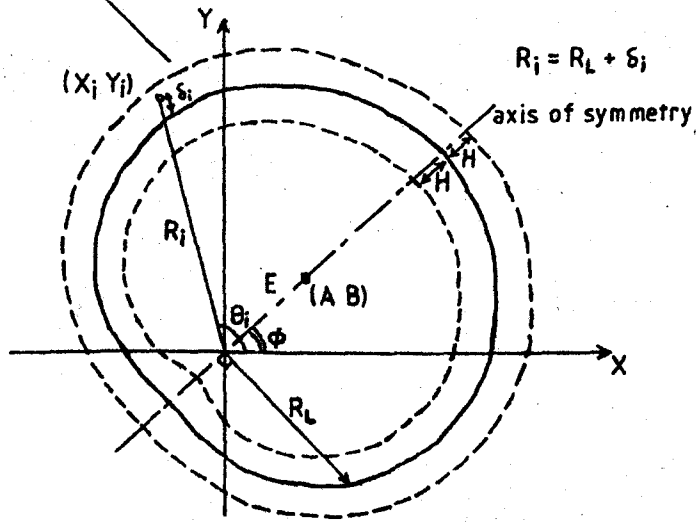
(A, B); (a, b)	- Cartesian representation of centre
(A_o, B_o)	- Intercepts of cylinder axis with XY plane
(A_1, B_1)	- Component of slope of cylinder axis from Z axis
$\underline{a} = (a \ b \ r_L)^T$	
$\underline{b}$	- Vectors used in the manipulation of linear
$\underline{c}$	- programmes
(E, Φ); (e, ϕ)	- Polar representation of centre
$f(\Psi, \eta)$	- Weighting function for profile measurement error
$2H$; $2h$	- Radial zone width of profile
$\underline{H}$	- Matrix of experimental points for least squares
$\underline{I}_n$	- n x n identity matrix
$\underline{K}$	- Matrix of experimental points for L. P.
L	- Radius suppression
L_o	- Axial length of cylinder
$\underline{l}$	- Direction cosines
M	- Magnification
m	- 'Slope' of line

N	- Number of data points
n	- Number of sides on a polygon
O	- Co-ordinate system origin
R_o; r_o	- Radius of circle
(R_i, θ_i); (r_i, θ_i)	- Profile radial data point from origin
R_L; r_L	- "Radius" (semi-diameter) of limacon
$\underline{R}$; $\underline{r}$ = $(r_1 \ \ldots \ r_i \ \ldots \ r_N)^T$	
(S_i, γ_i)	- Profile radial data point from centre
s	- Arbitrary radius of polar chart
W	- Active width of polar chart
(X_i, Y_i);	- Cartesian representation of profile data point
$\underline{X}_o$	- Point on cylinder axis
(X, Y, Z) (R, θ, Z)	- Principal axes of co-ordinate frames
$\underline{\beta}$	- Inverse of basis in linear programme
$\gamma = E/R_o$	- Eccentricity ratio
$\gamma_c = e/r_o$	- Eccentricity ratio on polar chart
Δ_{ijk}	- Determinant, co-factor of matrix
δ_i	- Out-of-roundness, residual from limacon
ε_i	- Out-of-roundness, residual from circle
ε_{max}	- Maximum radial divergence between limacon and circle
η	- Dimensionless ratio, E/S



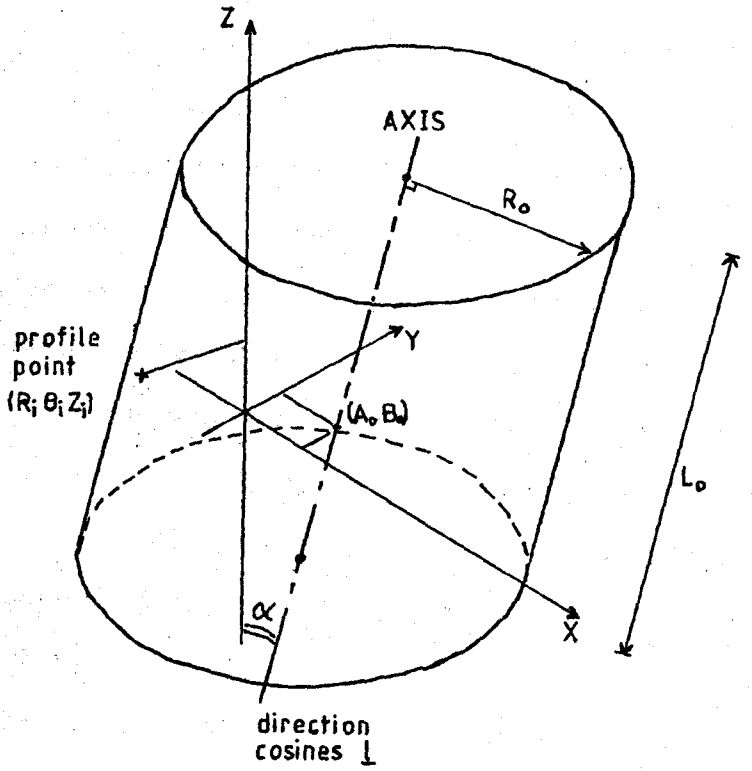
CIRCLE

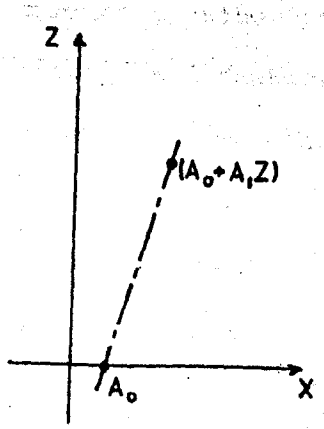
Profile points ∘ contained within pecked zone

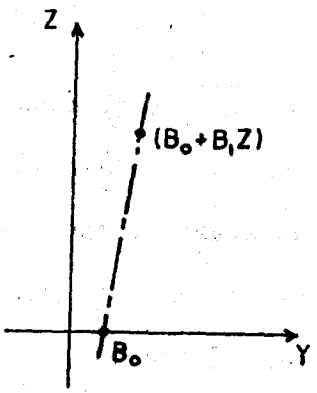


LIMACON

NOMENCLATURE: SINGLE PLANE







NOMENCLATURE: CYLINDERS

1. Introduction

As, over the past thirty years, surface metrology has become recognised as an important discipline in its own right so roundness measurement has grown to be one of its most significant branches. Traditionally "roundness measurement" has implied the comparison of single cross-sections of workpieces with perfect circles which are assumed to be the ideal. Using specialised measuring machines, an intuitive but perfectly workable basis for such measurements has developed and is embodied within National Standards. Recently, however, two developments have brought the potential limitations of current practice into importance. First, inexpensive digital computing has led to the expectation of more accurate or highly automated measurements. Secondly, there is a growing interest in so-called "integrated measurements" where the overall surface shape is to be described. This has caused the introduction of the currently rather poorly defined terms such as "cylindricity" and "sphericity".

Current state-of-the-art instrumentation, usually involving the generation of precise straight line and circular motions, seems capable of coping with most of these expectations provided that the data which it supplies can be adequately interpretted. However, it is not clear whether for this step existing methods can safely be extrapolated. Considering the basic measurement procedure illustrated by figure 1.1, the comparison of the imperfectly manufactured workpiece with its specification is seen to involve several processes.

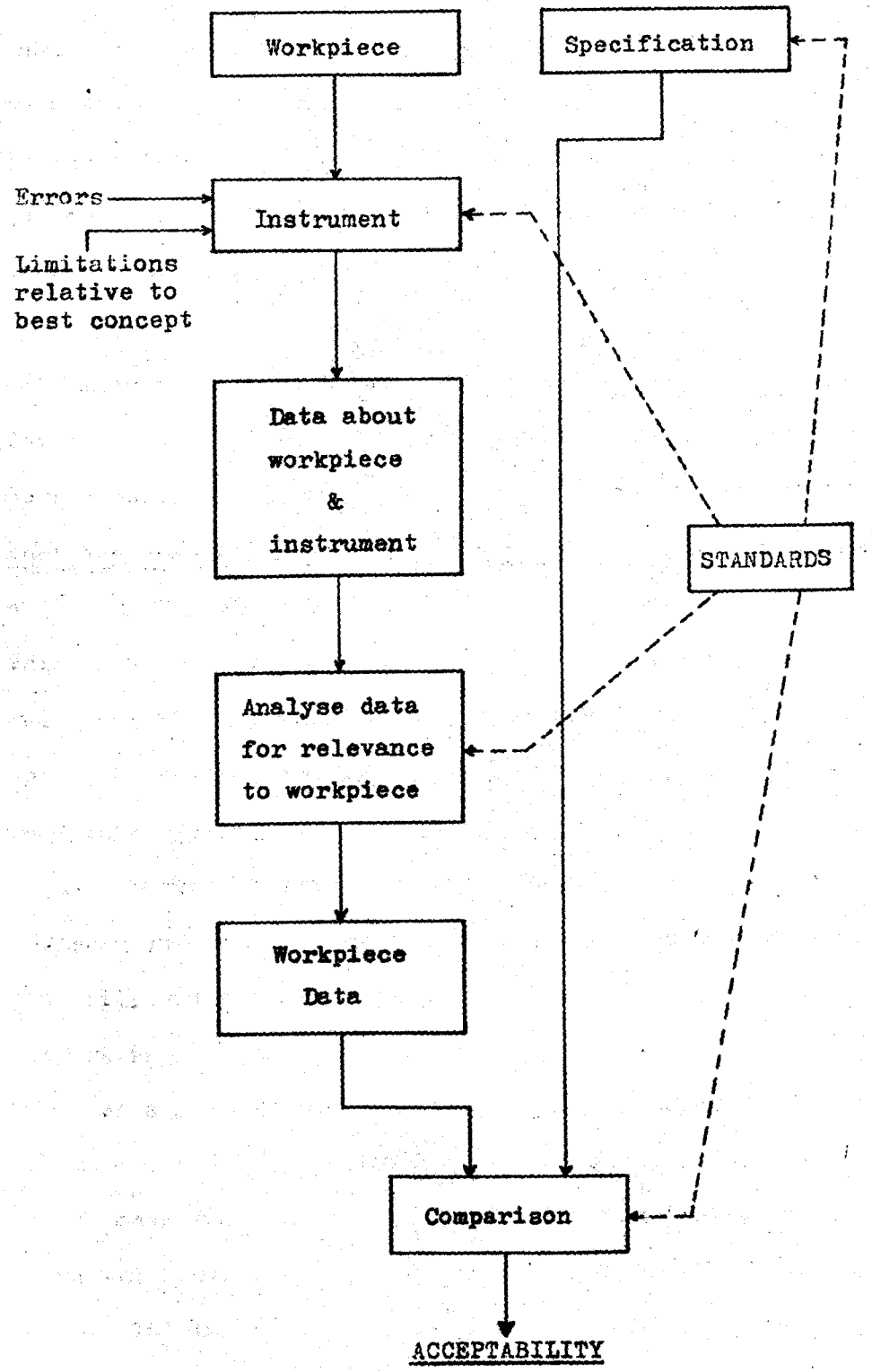


FIGURE 1.1: A Typical Measurement Procedure

Apart from the inevitable errors which occur with every measurement, there may be limitations in the instrument which cannot realistically be eliminated (see chapters 2 and 3 for more discussion). Thus it is necessary to extract from the data supplied that relevant to the workpiece so that only this is compared to the specification. Note also that at several stages, the procedures are influenced by Standards.

It is within the above context that this work is concerned with measurement analysis. Upon which techniques can be used for the analysis will depend the way in which instrumental developments should proceed. Ultimately it is to be hoped that the acceptance at inspection of manufactured workpieces will be governed by purely functional considerations, but this ideal is still well in the future. For the present, and therefore here, it will be necessary to work within the spirit of current standards, particularly with regard to reference figures.

This work studies and develops analytical methods suitable for both two and three dimensional measurements of nominally round parts. Its natural starting point lies in reappraising current roundness measurement practice.

On examining the history of roundness measurement the subject is found to be noticeably fragmented. The various topics have been treated virtually as if they were independent. There are natural reasons why this should be so, for example, many of the analytical methods have been developed as the need has arisen in response to particular, practical problems for

which previously existing methods were deemed unsuitable. On achieving a working method there has been little incentive to question deeply how it relates to those other methods. However, any attempt at extending the region of influence of these methods comes immediately against the difficulty that there is no unifying theory underlying the practice of roundness measurement. It is therefore regarded as the main task of this work that an attempt should be made to establish such a theory.

The present work is conceptually difficult to explain. Many of its concepts are of an abstract nature and some, at first sight, may seem only tenuously linked. This observation may be compared with that earlier concerning the fragmentation of the subject. In addition, many of the practically significant results and conclusions can stand alone and be successfully applied without deep knowledge of their inter-relationships. For these reasons the report will be allowed to retain a degree of fragmentation. Although its thesis is that there is an important, underlying structure and that the whole subject should be taken together, some chapters have been organised such that they could be extracted. Because of this the details of the history of the subject will be deferred until appropriate points in the text occur. In this introduction only a very brief overview will be given.

The whole subject of roundness measurement appears to have been born in almost its present form. National Standards (1, 2) indicate the importance of the independent spindle roundness instrument. There has been little conceptually important change in these since the pioneering work of Reason, whose survey (3) remains the major work on that subject. (Further discussions of instrumentation is given in

Chapter 2.) A major disadvantage of most roundness instruments is that the signal obtained as representing the deviations from circularity of the workpiece can be seriously distorted unless great care is taken to align the centre of the workpiece with the spindle axis. Various guidelines have been produced to restrict the degree of eccentricity present in the graphical output of the instrument (1, 3). Nevertheless some distortion will be present in practice and it adds greatly to the complexity of measurement analysis. This will affect the accuracy with which the reference circles (2) used for assessing roundness errors can be constructed. Reason (3) describes the nature of the distortion and gives a graphical method for reducing it but it seems that Whitehouse (4) first identified the distorted form with the geometric figure known as the limacon, claiming that its use as a reference figure is superior to using perfect circles under customary measurement conditions. (Section 3.1. gives more details.)

Standards allow four reference figures, the least squares circle and three which will here be grouped as "boundary references": the minimum circumscribing, maximum inscribing and minimum zone circles. An adequate solution technique for a restricted set of conditions has been known to the least squares circle for many years (3). More recently the method has been extended to include some incomplete, arcuate figures (4). All this work depends on what has come to be known as the "limacon approximation" and also makes further restrictive assumptions. Siddall (5) has queried the accuracy of the standard formulae for small arcs. Other work relating to this area, but not in a direct way, includes generating reference centres from centroid methods (6) and proposals for different standard formulae (see (7) and Appendix 3 for discussion of these.). Further discussion of least

squares methods is given in section 5.1.

The boundary references have been studied even less, most authorities apparently accepting the suggestion in Standards that "trial and error" methods be used for their fitting. Avdulov (8) and later Whitehouse (9), using a geometric method, propose basically the same method for bringing order into the search procedure. The method requires much work but is usable with high speed computers. For special purposes attempts at data reduction, identifying the major surface features, have been tried (10). Again the problem arises of intuitively simple ideas requiring much computation. Recent suggestions for finding minmum zone have included a non-uniformly weighted least squares calculation (11) and Monte Carlo methods based on the least squares centre (12). Neither will produce the exact solution in general. See also, section 6.1.

"Cylindricity" has yet to be formally defined by metrology Standards. Drawing Office Standards on geometric tolerancing (13) and speculations by Reason (14) have both been used to support a minimum zone approach although neither actually implies minimum zone. Other work relating to cylinder measurement (1, 15) has been restricted, for example, to axial straightness which is clearly less than should be implied by "cylindricity". Some recent work in Japan (16, 17) has calculated "least squares cylinders" by direct extension of the two-dimensional methods but seems not to have considered the implicit assumptions of doing so. Tsukada (18) questions what measurement schemes should be used with least squares cylinders and suggests that a helical trace around the workpiece is appropriate. There is almost no consideration of boundary cylinders, although the modified least squares method (11) has been extended to three dimensions. Section

11.1 contains a fuller survey of cylinder related work.

In accordance with the philosophy outlined earlier, this work first studies the underlying principles of roundness measurement. In particular it examines the sources of error relevant to reference figure calculation and identifies them with various specific points in the measurement process. This leads to the identification of the major properties derived from ideal and practical reference figures. Then, rather than follow semi-empirical methods as in earlier work, reference figure fitting is examined formally using the mathematical methods of optimisation theory. This leads to a full recognition of the role of the limacon in roundness measurement. Two of the techniques used are of particular interest, namely Parameter Space analysis and Linear Programming, both most commonly used in Operations Research. By specialisation of the Linear Programme solution for boundary reference limacons, highly efficient algorithms are developed. The nature of the approximations involved in using limacon references is examined in detail and the eccentricity ratio established as a control on the accuracy of the method. Some other reference figures are also briefly studied.

The practical usefulness of the methods introduced here will depend upon when the mathematically predictable errors become significantly large. One aim of this work is to reduce the necessity for precise centring, or, conversely, to increase the measurement capability of an instrument. This will be achieved only if the methods are tolerant of eccentricity levels considerably higher than those occurring in current practice. To discover whether this is so will require widespread experimental investigation. Initial tests reported here suggest that it is.

With a firm theoretical basis in two dimensions, the development of methods for finding cylindrical references is relatively straightforward. Solutions using linear least squares and linear programming are advocated. Special algorithms are not developed. The nature of the approximations is similar to those in two dimensions but the magnitudes are more severe. However, practically useful measurement systems can be produced.

All engineering is ultimately concerned with pragmatism and so with value judgements. This work sets out to discover what might be theoretically possible if suitable instrumentation were to be developed. Its value can therefore be judged against two criteria, usefulness under current conditions and potential usefulness to future research. Only the former can be quantified and so only comparisons of this work with current practice are made here. It is hoped that this concentration on the present will not mask possible implications for the future. Because of its standpoint, this work is distinctly not a textbook on roundness measurement, although it may be of interest to those writing such books. Any influence it might exert is expected to be upon the designers of instrumentation systems rather than their users.

Please note that throughout this work many diagrams are shown, for clarity, with highly exaggerated features. In particular the limacon is normally shown with a cusp developed so that it is easily distinguishable from a circle. It is not intended to imply that these represent sensible measurement conditions.

2. Current Practice in Roundness Measurement

2.1 Requisites for Roundness Measurement

There are three metrological questions which can be posed concerning the cross-section of a nearly circular component:

i. What are its deviations from circularity?

ii. What is its size?

iii. Where is its centre relative to some known point?

Depending upon the circumstances, the relative importance of these factors and the precision required of their measurement will vary. Examples of such combinations could include:-

a) A simple shaft. Here centre position may be of little importance, and, if used in a bearing, even quite significant radius variation might be tolerable to a compliant journal. However if the shaft is to run "true" the roundness must be reasonably good, which is also a condition required for the avoidance of cyclic loads on such a compliant journal.

b) High speed air-hydrodynamic bearing. A possible design might use a circular journal with a shaft having a deliberate, small amplitude lobing to give the necessary pressurisation. Shape measurement is now important for the control of the size of the lobes and not, as more commonly occurs, for checking that no significant deviation from circularity exists. The bearing would have small clearances and so radial tolerance on both components would have to be well controlled.

c) Crankshaft. The requirement on each bearing is probably similar to a) above, but in addition the individual bearings must be

aligned axially. Thus here out-of-roundness and centre position would be measured.

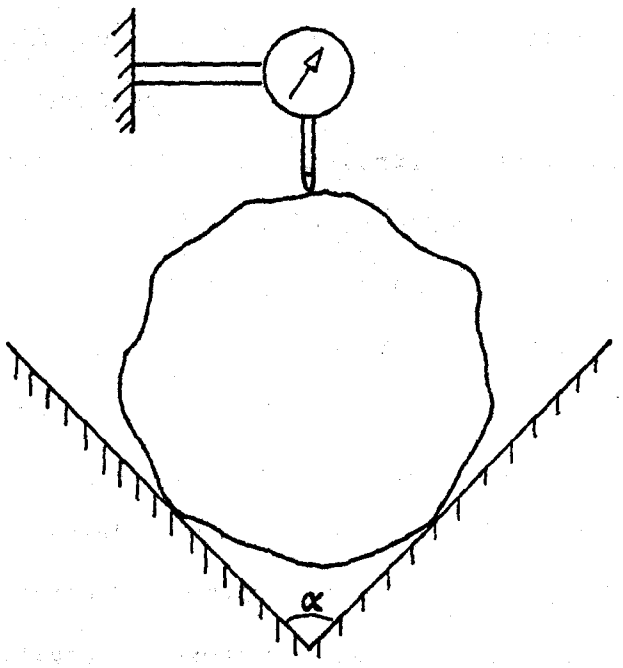
To determine the out-of-roundness of a section, it must be compared with a perfect circle. This could be either a specific circle (predetermined centre and radius) used in the manner of general profile measurement (BS 308) or, more commonly, any convenient circle suitably aligned to the section, since often the out-of-roundness is controlled simply by a peak-tc-valley amplitude measurement. When the section is not perfectly circular, the definition of radius and centre become to some extent matters of interpretation. Usually the relevant parameters of a circle associated with the section by a "best-fit" criterion are used, although centroids have also been suggested (6).

The range of shapes and sizes which might be absorbed within the field of roundness measurement is obviously as wide as the imagination can make it. However, the vast majority of practical examples, including virtually all those for which standard instrumentation might be supplied, fall in the same range as the typical components of, say, the automobile or aero-space industries. The most common range contains a peak-to-valley out-of-roundness in the order of 1μm to 10μm and occurs on parts with radii from a few millimetres to about 100mm. That roundness measurement should be mainly applied to small components is due in part to its association with high absolute precision. It is only when such precision is needed that the special techniques identifying roundness measurement as a specific branch of metrology become necessary. This work will comply with this grouping, being concerned mainly with techniques having the potential for resolutions and accuracies of better than 1μm.

2.2 Roundness Instrumentation

This work is concerned with the analysis of data representing roundness measurement rather than with the acquisition of such data, consequently the discussion of instrumental techniques is included only to establish a context for the remainder of the work. It is not intended to be either a survey or a detailed description of method (for which see references 1, 2, 5, 19). However, it should be stressed that throughout this work any practical difficulties of instrumental method which might occur will not generally be acknowledged in the text. This approach follows the argument that the limitations of data analysis must be established in their own right and only thereafter should limitations due to instruments be introduced. The areas which would benefit most from improvement should then stand out clearly. As a specific example this work is greatly concerned with the eccentricity between the component and the normal stylus-type roundness instrument but will not consider such important features as the effect of finite tip radius and tangential stiffness of the stylus assembly both of which are particularly significant on small components. In addition it will not question, other than by this statement, the relevance of the essentially two-dimensional cross-section measurement to the total, three-dimensional component.

Of the various practical methods used to measure roundness, nearly all involve the use of a mechanical rotation as a means of producing a circle with which to compare the workpiece. Perhaps the simplest in concept is the V-block method, figure 2.1, in which the part is rotated in the V-block while in contact with the probe of

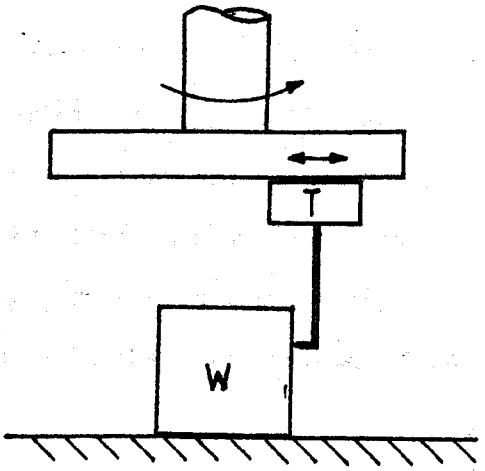


OUT-OF-ROUNDNESS = SWING OF INDICATOR (T.I.R.)

FIGURE 2.1: Roundness Measurement by V-block

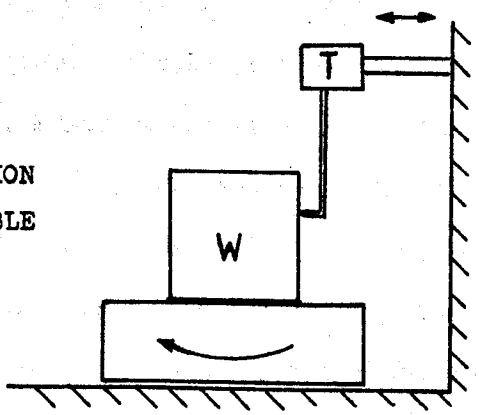
an indicator device. The total out-of-roundness is taken to be the maximum swing of the indicator. This method can give an indication of out-of-roundness and perhaps, radius but not centre since it is intrinsically a self aligning device. Its major drawback is that depending upon the angle of the V-block various orders of lobing can be suppressed. Various multiprobe techniques have evolved from V-block methods (20) but although useful for "in situ" work they show similar difficulties.

The problems with V-block methods arise because the out-of-roundness of the workpiece affects the accuracy of the rotation which is supposed to be generating the datum circle. Thus the logical move would be to a system with an independent precision spindle, the method generally used in proprietary instruments. (Indeed, Moore (19) goes so far as to claim that the independent spindle is the only method usable for precision work.) The spindle is used to produce a relative rotation between the workpiece and a transducer and so either may be rotated. Figure 2.2 shows the general scheme of such instruments. There are some significant and quite subtle differences between rotating workpiece and rotating transducer instruments concerning their structural and operational convenience together with their accuracy and their ability to perform other measurements but for the present purposes they need not be distinguished. Typical modern spindles on such instruments run with a maximum deviation from circularity in the region 0.25μm to 0.025μm. Normally this gives a limit to the achievable accuracy of the system although in some cases computational methods can achieve even smaller figures (20, 21). These instruments are ideal for measuring out-of-roundness and may also, sometimes by the use of optional attachments, be used for a



CROSS-SLIDE ON
PRECISION SPINDLE

T - TRANSDUCER & STYLUS
W - WORKPIECE



PRECISION
TURNTABLE

FIGURE 2.2: Roundness Measurement by Spindle Based Instrument

limited range of centre and radius measurement.

A totally different concept for measuring a circular workpiece would be not to compare it directly with a circle but simply to take a series of point measurements on the surface using a co-ordinate measuring machine and to analyse them mathematically. However, the resolutions and accuracies of present machines is sufficiently low that this would not be readily regarded as a valid method except, perhaps, on large workpieces, but, with this limitation, it would seem to be the ideal method of measuring the distance between centres.

Overall then, it appears that, at present, the conventional roundness instrument has little real competition for precision measurement and so most references, throughout this work, to instrumentation will refer to independent spindle methods. There is one fundamentally important limitation of such instruments. It is normal to measure only the radial variation between the surface of the workpiece and a nominal circle (represented by a known point in the transducer) having a radius almost equal to that of the workpiece. For display purposes this variation is magnified and superimposed onto a convenient arbitrary radius for plotting as a polar graph (or chart) of the errors. One reason for doing this seems straightforward: it would be impossible to magnify the whole radius by the amount (probably greater than 1000x) needed for the roundness errors to be easily detectable on a graph. Another facet to be considered is that if the total radius were measured, a range/resolution ratio for the transducer of as much as 10^7 might be needed, together with a very small absolute resolution. This performance cannot be obtained economically and so virtually all precision roundness measurement

involves this self-evidently named radius suppression. It is illustrated schematically in figure 3.1. Some modern instruments do have absolute radius measuring attachments but these work to a precision considerably less than that of the roundness transducer. So an additional problem for the later analysis of the profile is that the precision to which radius suppression is known is well below that of the measurement itself. It is well known that radius suppression is one of the factors which causes the distortion of the profile from its expected form when plotted on the chart. Reason shows some particularly good examples of this (3). Here further discussion will be defered until Section 3 to save repetition.

2.3 Roundness Analysis

If the data representing a roundness profile had been collected by a co-ordinate measuring machine, it would be quite clear that the first stage of analysis would be to calculate a circle from which to measure the points. With data from a normal instrument the same is true for although the instrument measures relative to a circle, there will inevitably be relative eccentricity between the spindle axis and the workpiece centre. Part of the profile signal will be the variation of distance between workpiece and transducer caused by this eccentricity and so the out-of-roundness should be measured from a reference figure which compensates for the eccentricity.

It is important to note that this way of approaching the subject of reference figures should not be taken to imply that they are needed solely because of eccentricity. As in other measurements, the roundness reference figure is that definable state relative to which the

data is expressed in the analysis. If a roundness instrument were perfectly set up, the datum circle of the instrument would be the reference figure. Hence the need to calculate the reference figure is a property of the measurement imperfections, notably, in this case, eccentricity.

In most roundness measurements virtually the whole effort of analysis is concerned with obtaining the parameters of the reference figure either because those values represent the end in themselves (as with eccentricity measurement) or since the out-of-roundness is normally measured only in terms of its maximum divergence from the reference, see figure 2.3.

Standardisation of the form of references to be used is important if records are to be compared and so four "best fit" circles according to different physical criteria are allowed. Various National Standards give virtually identical definitions of them but may differ over which are to be "preferred" (1, 2). In descending order of preference according to BS 3730 these are:

i. Least squares circle
ii. Minimum radial zone circles
iii. Minimum radius circumscribing circle
iv. Maximum radius inscribing circle

The definition of the least squares circle is discussed in detail in Chapter 5. The minimum zone circles are two concentric circles so placed that the profile is enclosed between them and their separation is minimised. The other definitions are reasonably self-evident. The terms Ring Gauge and Plug Gauge circles are sometimes applied to the latter two references. Some authorities have objected to this notation on the grounds that since the references act only on profiles,

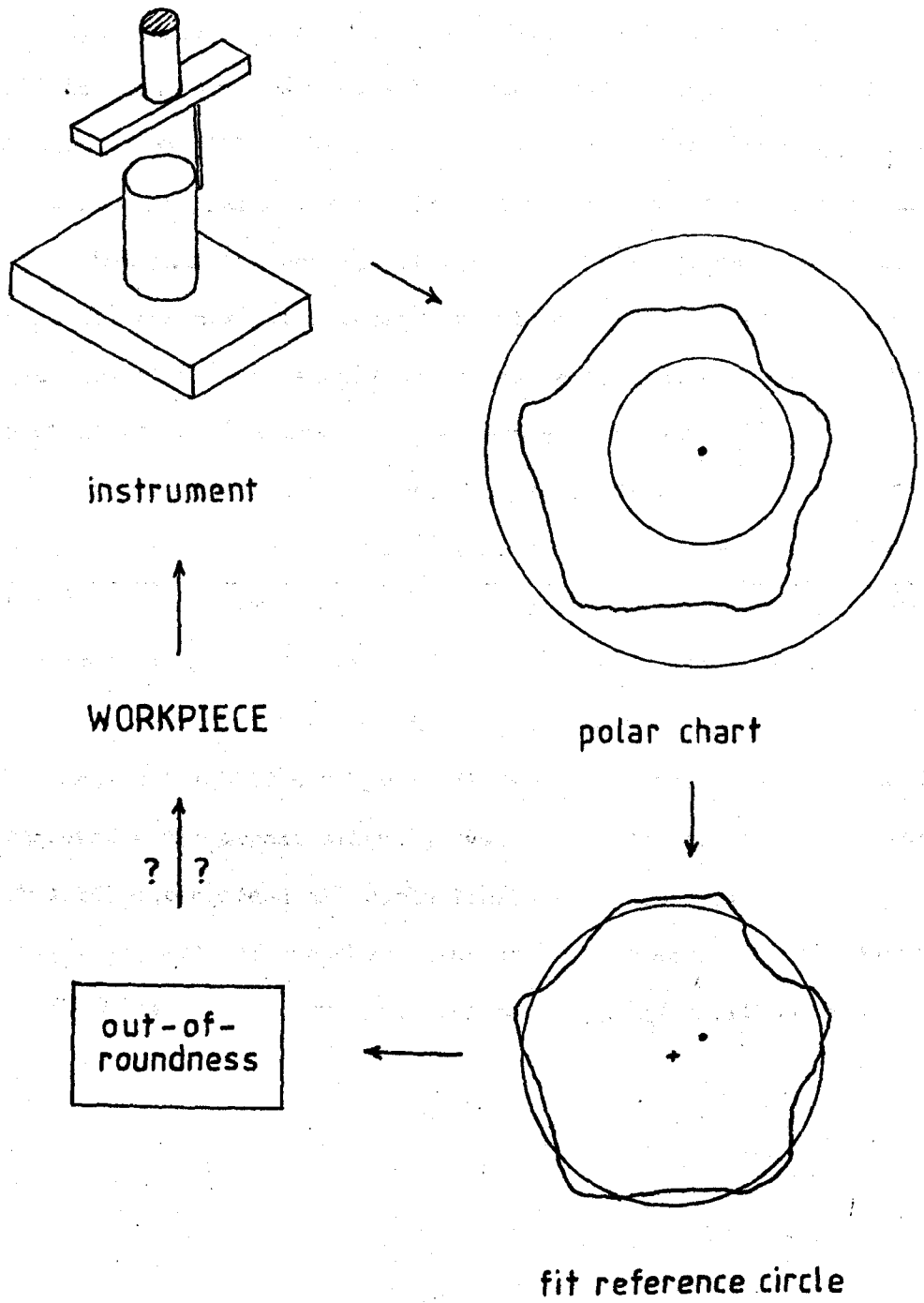


FIGURE 2.3: Stages of Standard Roundness Measurement Process

not three-dimensional surfaces, they are not equivalent to the use of limit gauges. However the terms are concise and graphic and will be used here, particularly since they are used in the current version of BS 3730. The Standard claims that the least squares circle is unique and that the zone width of the minimum zone is obviously unique but that the centres of minimum zone, ring gauge and plug gauge circles are not necessarily unique. It will become apparent that such statements should sometimes be treated with caution but in particular the ring gauge circle centre is always unique.

With the exception of least squares, for which an approximate formula is given, the standards accept that the reference circles will be found by trial and error using compasses or circular templates on polar charts. ANSI B89 gives some guidance for this process. It has been noted that the combination of eccentricity and radius suppression leads to some polar distortion so that a truly circular component would appear slightly oval on the chart. Therefore some authorities recommend allowable limits on eccentricity on the chart, usually in terms of the chart mean or inner radius (1, 3). Again to save repetition discussion will be postponed until the next chapter.

3. Roundness Analysis Re-assessed

3.1 The Measurement System

In keeping with the philosophy outlined in Chapter 1, an attempt will be made to bring all the material of Chapter 2 within a single descriptive framework. The first point in this process is to ask what is the fundamental reason for the whole process? Most commonly it is to determine to what extent a workpiece deviates from its expected, circular shape. Now the shape of the workpiece is a property of itself alone; the points making up its surface have a fixed relationship to each other (assuming that it is rigid) independent of its orientation in space. It is clear, then, that the shape, including the concept of a reference circle as the form from which the surface has deviated, exists in a frame of reference tied to the workpiece which will be called the component co-ordinate frame.

In order to be measured, the workpiece must be presented to an instrument and in being so its orientation is constrained by the limitation of that instrument. The measurement data is expressed relative to the instrument datum, that is, in an instrument co-ordinate frame of reference. Generally an attempt would be made to align the component co-ordinates with the instrument co-ordinates but this will never be perfect. The two frames can always be related by simple axis transformation. With roundness instruments the data does not appear in instrument co-ordinate form since it is immediately transformed by radius suppression and magnification to chart co-ordinates which is where the data is available for analysis. The

transformation between instrument and chart co-ordinates is not a change of axes but a point for point mapping which leaves the orientation of the axes unchanged.

The expression of the measurement in terms of the component, instrument and chart co-ordinate frames gives more than a convenient notational arrangement. By forcing each action of the measurement into a definite position. it is identified as belonging to a particular frame or to a transformation between frames. Its effect on the measurement is then easily identified.

Since the transformation between component and instrument co-ordinates is a change of axes only, then all shapes are preserved and a measurement in instrument co-ordinates relates directly to the profile in component co-ordinates. The conceptual reference circle of the component becomes an eccentric circle in instrument co-ordinates.

The nature of the transformation from instrument to chart co-ordinates is first radius suppression

$$r(\theta) \leftrightarrow r(\theta) - \text{constant}$$

and then magnification

$$r(\theta) \leftrightarrow r(\theta) \times \text{constant}$$

The transformation is a linear action on radial vectors and non-linear on all other relationships. Therefore a circle centred on the origin is transformed to another circle but an eccentric circle is transformed to a non-circular shape. The effect may be seen to be quantified by reference to Figure 3.1. In instrument co-ordinates an eccentric circle, centred at (E, ϕ) will be:

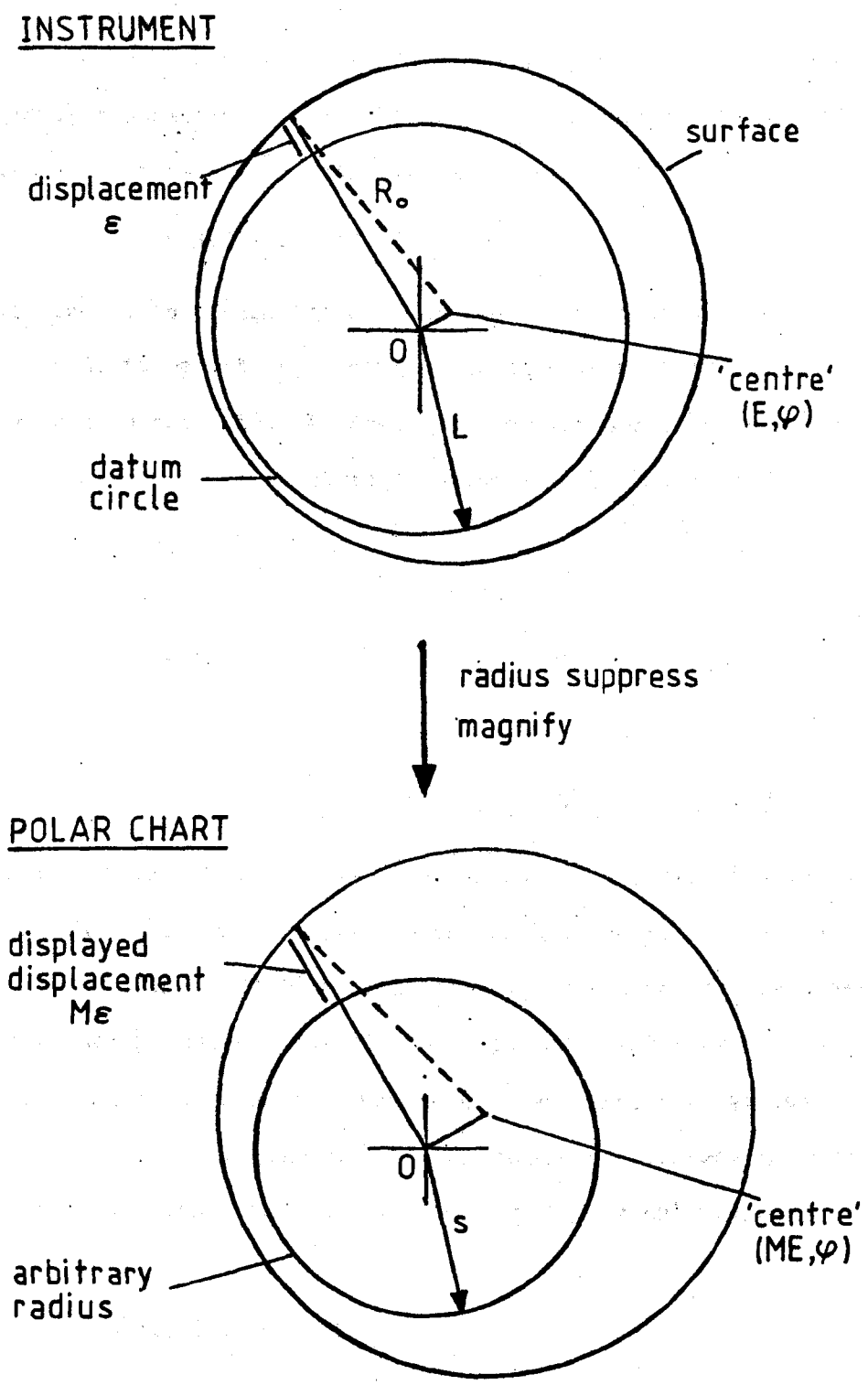


FIGURE 3.1: Radius Suppression in Roundness Measurement

$$R = E\cos(\theta-\phi) + [R_o^2 - E^2\sin^2(\theta-\phi)]^{\frac{1}{2}} \qquad 3.1$$

If this is then transformed into chart co-ordinates by applying radius suppression L, magnification M and then adding the arbitrary chart radius S, its polar chart will be:

$$r = M(R - L) + S$$

Although S is of significance in the actual plotted form of the profile, it can, being a simple additive term, normally be ignored in analysis. It does not, for instance, occur (or has zero value) in the electrical output signal available from most commercial instruments.

Applying the binomial expansion to equation 3.1 for $E < R_o$:

$$R = E\cos(\theta-\phi) + R_o\left(1 - \frac{\gamma^2}{2}\sin^2(\theta-\phi) - \frac{\gamma^4}{8}\sin^4(\theta-\phi) \ldots\right) \qquad 3.2$$

and if this is expressed in chart co-ordinates:

$$r = ME\cos(\theta-\phi) + M(R_o - L) + MR_o\left(-\frac{\gamma^2}{2}\sin^2(\theta-\phi) - \ldots\right. \qquad 3.3$$

where $\gamma = E/R_o$ and will here be termed the eccentricity ratio. The shape difference between equation 3.2 and 3.3 is caused because of the different relative weights applied to the constant and harmonic terms of their equivalent Fourier series. (Although this distortion has been known since the earliest days of roundness measurement, its explicit statement in terms of harmonic distortion appears to be quite recent. (4).)

3.2 The Assessment System

Although ultimately the assessment of roundness in chart co-ordinates is unavoidable, the above discussion attempts to

make clear that the analysis should take place in instrument co-ordinates. The reference circle calculation is an attempt to get back from instrument to component co-ordinates. Fitting a circle to the chart profile implicitly assumes that it represents, at least to an adequate approximation, the appearance on the chart of the reference circle.

The current practice and Standards (1, 2) of necessity allow the definition of circular references on the chart. Here the definition of such references will be kept only for instrument co-ordinates, but otherwise they will be regarded as those discussed in section 2.2. The four references divide naturally into two groups; the least squares solution which is by nature an averaging process and the other three references all of which are "peak sensitive". The general name "boundary references" will be used for these latter three. Mathematically, the solution to any of these references is a process of optimisation. In the case of least squares it is an unconstrained minimisation of the sum of squares of the residuals With the boundary references the optimisation is on a simple function but subject to constraints. Thus the ring gauge circle problem may be stated

$$\text{Minimise } Z = R_o$$

$$\text{Subject to } R_i \leqslant E\cos(\theta_i - \phi) + \left[R_o^2 - E^2\sin^2(\theta_i - \phi)\right]^{\frac{1}{2}} \qquad 3.4$$

$$\text{for all } i \text{ of the data set.}$$

that is, find the smallest value of R_o for which all data points (R_i, θ_i) lie inside the circle. The plug gauge problem is similar but maximises R_o subject to constraints of opposite sense. The formulation of the minimum zone circles can be approached in

various ways. One convenient method is to use a mean radius circle from which symmetrically placed half zone widths, H are established. This gives

$$\text{Minimise } z = H$$

$$\text{Subject to: } R_i \leq E\cos(\theta_i-\phi) + \left[R_o^2 - E^2\sin^2(\theta_i-\phi)\right]^{\frac{1}{2}} + H \qquad 3.5$$

$$R_i \geq E\cos(\theta_i-\phi) + \left[R_o^2 - E^2\sin^2(\theta_i-\phi)\right]^{\frac{1}{2}} - H$$

$$\text{for all } i \text{ of the data set.}$$

Each data point causes two constraints implying that more work will be needed to solve the minimum zone rather than the ring or plug circles.

Optimisation theory offers a complete general solution only to some fairly restricted sets of problems. The relevance of these restrictions to the fitting of reference circles will be explored in depth later (particularly in Chapter 4). Here it will suffice to observe that the major class of problems which are always fully solvable and efficiently calculable consists of those in which all functions have only a linear dependence upon the parameters with respect to which the optimisation is to be performed. It will be seen that this is not the case with either constraints 3.4 or 3.5. The least squares method will clearly never have such a linear relationship in the sum of residuals. It belongs to a different class which allows a general solution providing that the reference figure is linearly dependent on its parameters. (In common usage the term "least squares" is taken as implying "linear least squares".) These observations add weight to the doubts concerning the uniqueness of references quoted in, for instance, BS 3730. They also extend

these doubts to include the least squares circle.

Any difficulties concerned with fitting circular references in instrument co-ordinates will be compounded when working in chart co-ordinates by the fact that non-circles should be used on the chart. If circles are used on the chart, difficulties of fitting might be expected to be greater than in instrument co-ordinates since the effect of radius suppression and magnification will be to increase the amplitude of profile fluctuations relative to their mean value. In the past the approach has been to sidestep difficulties that do occur by convenient local approximations, see, for example, section 5.1. However, surely it is desirable that an overall solution should be sought. One approach would be to adopt an alternative reference figure which can be used consistently and which maintains an adequately close adherence to the philosophy of using circular references.

Among the properties of any proposed reference figure, the following can be readily identified as being of practical importance. It should closely model the form displayed in instrument or chart co-ordinates of an eccentric (or centred) circle in instrument co-ordinates. It should be readily translatable between instrument and chart co-ordinates, implying that its dependence on radius suppression must be very simple. On any graph a theoretically unique reference figure should exist for all four accepted fitting criteria. Definitive algorithms for all the fitting criteria should exist and be reasonably easy to implement in practice. It should be possible to use the reference 'by hand' directly on a graph. It would further be useful if the reference figure could be expressed as a continuous mathematical function. Additionally, in

order to maintain continuity of records, it would be preferable if, for at least a large set of restricted conditions, measurements with the reference figure could be easily related to those performed with circular references.

The circle does not fulfill all of these requirements.

Having regard for the advantages to be gained in the optimisation, a good candidate for a reference figure would be a linearisation of the circle in instrument co-ordinates. Radius suppression is a linear operation with respect to the radius vector of polar co-ordinates and so will have the desired simple effect on any figure which is linear in its radius term measured from the origin. Thus providing that it is a good approximation to a circle in instrument co-ordinates, a circle linearised in its parameters about the origin (position of zero eccentricity) will fulfill all the computational requirements listed above. The normal way of linearising a function is to use its truncated Taylor Series about the desired point of linearisation. For completeness, here, the general linearisation of equation 3.1 about a point (A_E, B_E) in instrument co-ordinates will be quoted. The first stage is to remove the parameter ϕ by re-expressing equation 3.1 in terms of the cartesian components of eccentricity, $A = E\cos\phi$, $B = E\sin\phi$:

$$R = A\cos\theta + B\sin\theta + [R_o^2 - (A\sin\theta - B\cos\theta)^2]^{\frac{1}{2}} \qquad 3.6$$

The Taylor expansion of this form gives:

$$R \simeq A\left(\cos\theta + \frac{(A_E\sin\theta - B_E\cos\theta)\sin\theta}{[R_E^2 - (A_E\sin\theta - B_E\cos\theta)^2]^{\frac{1}{2}}}\right) \qquad 3.7$$

$$+ B\left(\sin\theta + \frac{A_E(\sin\theta - B_E\cos\theta)\cos\theta}{[R_E^2 - (A_E\sin\theta - B_E\cos\theta)^2]^{\frac{1}{2}}}\right)$$

$$+ R_o \frac{R_E}{[R_E^2 - (A_E\sin\theta - B_E\cos\theta)^2]^{\frac{1}{2}}} \qquad 3.7$$

$$+ A_E\cos\theta + B_E\sin\theta + [R_E^2 - (A_E\sin\theta - B_E\cos\theta)^2]^{\frac{1}{2}}$$

This approximation is linear in A, B and R_o although it is highly non-linear in the initial parameter estimates A_E, B_E and R_E. When the linearisation is performed from the origin ($A_E = B_E = 0$) this becomes greatly simplified:

$$R \simeq A\cos\theta + B\sin\theta + R_o \qquad 3.8$$

Equation 3.8 is, of course, a form very well known in roundness measurement and is usually derived not by the method given here but by the simple truncation of the binomial series as shown in equation 3.2. The alternative approach is presented to stress two points. In this attempt to unify the mathematics of roundness measurement a specific choice is made to linearise the circle. Further there remains the possibility, not shown by the binomial expansion, of linearising not with respect to the origin but about any other point in the instrument co-ordinate frame. This second point is probably of little interest with current roundness instruments but could have application in other schemes such as co-ordinate measuring machine based techniques.

The figure represented by equation 3.8 is known as the limacon. It has been advocated elsewhere (4) as being a better approximation to the radius suppressed form of an eccentric circle than is a circle drawn on the chart. This is seen from equation 3.3 which may be compared with the equivalent form of equation 3.8.

$$r = MA\cos\theta + MB\sin\theta + M(R_o - L) + S \qquad 3.9$$

$$= a\cos\theta + b\sin\theta + r_t \qquad 3.9$$

and the expanded form of a circle in chart co-ordinates having the same parameters:

$$r = e\cos(\theta-\phi) + [r_c^2 - e^2\sin^2(\theta-\phi)]^{\frac{1}{2}}$$

$$= ME\cos(\theta-\phi) + M(R_o-L) + r_c\left(-\frac{\gamma_c^2}{2}\sin^2(\theta-\phi) - \cdots\right)$$

where now $r_c = M(R_o-L) + S$

and $\gamma_c = e/r_c$

Assuming that the shape difference between circles and limacons is dominated by the second order term, the condition for the error between the limacon and the radius suppressed circle to be less than that between the radius suppressed and the true circle is approximately:

$$MR_o\gamma^2 < r_c\gamma_c^2 - MR_o\gamma^2$$

or $$2(M(R_o-L) + S) < MR_o$$

In a typical instrument, r_c might be about 50mm and so the limacon will represent a better model than a perfect circle on the chart on almost all measurements (magnification greater than 100 on 1mm radius components or only 10 on components of 10mm radius).

The limacon reference has known advantages over a circle on the chart and being linear in its parameters can be used conveniently for computation. It is amenable to least squares analysis and the boundary references using it become linear programming problems for which well established solution techniques are available. The computational methods will be considered in later chapters. Here the general properties of the limacon as a reference will be examined.

3.3 The Limacon Reference Figure

The limacon is a smoothly varying closed figure having constant diameter in that $R(\theta) + R(\theta+\pi)$ is a constant. It does have chords not passing through the origin which are larger than this diameter, There is some difficulty over terminology when using limacon references since the measurements made with these are always alluding to those with circles. Here the general approach will be to name a property of a limacon according to its equivalent in a circle. Consequently "diameter" will be defined as a chord passing through the origin and the name "radius" will be applied to the constant of the limacon, R_L, although strictly "semi-diameter" would be more accurate. The "eccentricity" may be defined as the amplitude of the sinusoidal term, $E = (A^2+B^2)^{\frac{1}{2}}$, and by inference the "centre" of the limacon will be the point (E, ϕ). Measurements on limacons are made from the origin not from the centre. The shape of limacon varies according to the degree of eccentricity present, being flattened from circularity in the opposite direction to that of the eccentricity. It develops a cusp, symmetrically placed about the direction $-\phi$, when the eccentricity exceeds half the radius and ceases to have a simple geometric form when the eccentricity exceeds the radius. In most diagrams in this work the limacon will be shown at high eccentricity and cusped for clarity in distinguishing it from circles. With current practical roundness instruments cusp behaviour is never likely to occur. Such properties as circumference and area also vary with eccentricity and so, unlike the circle, care must be taken to specify which property is being optimised when fitting a reference figure. Here it will be taken that it is the radius which is to be optimised since this is consistent

with the circle definitions quoted earlier and has the added advantage for the optimisation process of being a linear property of the limacon. The area of a limacon is $\pi(R_L^2 + E_{\frac{1}{2}}^2)^{\frac{1}{2}}$ and so will not necessarily be optimised at the same time as is radius. The minimum and maximum distances of the limacon periphery from the centre are R_L and $(R_L^2 + E^2)^{\frac{1}{2}}$ occurring respectively on the axis of symmetry (the line containing both the origin and centre) and on the line perpendicular to the axis of symmetry and passing through the origin.

By choosing to use a limacon reference and making all measurements with respect to it radial from the origin (note here that these are separate decisions even though usually taken automatically together - see also the discussion on which centre to use with circular references in (3)), the analytical difficulties concerning radius suppression may be overcome. Radial differences measured from the origin are always preserved under radius suppression and a limacon is transformed to another limacon. Although the translation of a limacon from instrument to chart co-ordinates, or vice versa, will change its shape, it will maintain the same mathematical properties and so the same analytical methods can be used in either co-ordinate frame. For example, if both the minimum radius circumscribing limacon to a set of data points in chart co-ordinates and those data points are transformed to instrument co-ordinates, the result, there, will be the minimum radius circumscribing limacon to the data. Additionally the radial distance of a particular point from the reference measured in either co-ordinate system will be the same (subject to scaling by the magnification). Neither of these observations will be generally true for circular references.

In instrument co-ordinates the true reference figure is an eccentric circle and the use of a limacon reference an approximation to it. The nature of the approximation is simply the truncation of the infinite series given by equation 3.2 and so a direct measure of its quality is given by the eccentricity ratio γ. If this is adequately small the limacon will be a good approximation to the circle and so a limacon will also be an equally good approximation to the distorted circle in chart co-ordinates. Using this fact a link between circular references can be produced; the circle in instrument co-ordinates is approximated by a limacon which is then radius suppressed to another limacon in chart co-ordinates which is itself approximated by a circle on the chart. The quality of the initial approximation depends upon the value of γ while that of the approximation in chart co-ordinates depends on the chart eccentricity ratio, $\gamma_c = e/r_L$. The recommendations given for limiting eccentricity in practice (1, 3) are, effectively, controls on γ_c, that is, they maintain the error due to using a circle on the chart to an adequately low level and do not refer at all to γ. Given current instrumental designs this is not unreasonable. R_o and r_L are likely to be of the same order of size and so γ_c will be larger than γ by a factor of the order of the magnification. The nature of the errors is further illustrated by Figure 3.2. In the American Standard (1) a maximum allowable error between the reference figure and the true shape of a radius suppressed circle is suggested as 0.25mm on the chart. The figure shows for one condition the value, E_{max}, of eccentricity in instrument co-ordinates which is allowable in order to just satisfy this criterion at different component radii with

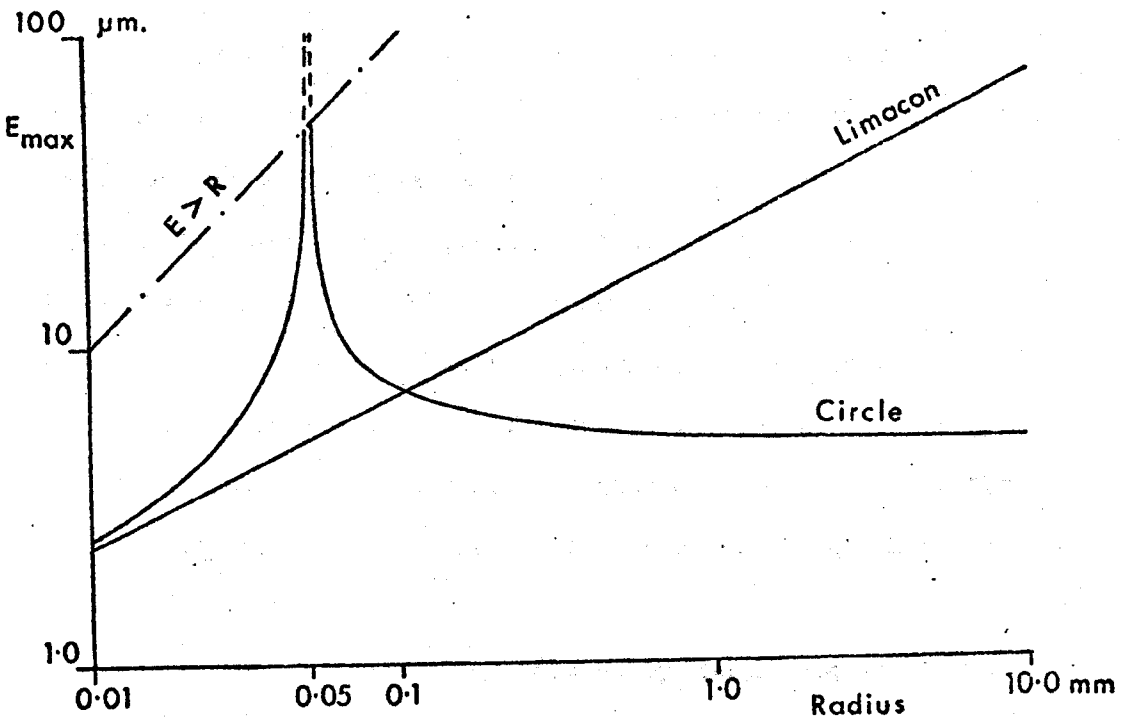


FIGURE 3.2: Allowable Eccentricity E_{max} to Cause Maximum Divergence between Circular Part and Reference of 0.25mm on the Chart Plotted Against Part Radius R_o (Radius Suppression, $L = R_o$)

circle and limacon references used on the chart. The assumed conditions are of complete radius suppression, $R_o = L$, a magnification of 1000 and a chart radius, S, of 50mm. The limacon allows larger errors except when the radius is reduced to the extent that MR_o is comparable or smaller than S. When $MR_o = S$ there is, in effect, no radius suppression and the circle appears undistorted on the chart however much eccentricity may be present. The advantage of the limacon is stressed by the additional line showing its value of E_{max} for conditions identical except that the magnification is 10000. The corresponding line for the circle would have the same shape as that shown but moved one decade downwards and one decade to the left so that it nowhere appears in the graph range of figure 3.2. The simple criteria for limiting eccentricity are based upon the 'flat' part of the graph for circular references.

For values of γ much less than unity the divergence between circle and limacon measured radially from the origin will be dominated by the lowest order term ignored by the truncation of equation 3.2, that is $(E^2/2R_o)\sin^2(\theta - \phi)$. The maximum divergence may then, for convenience of calculation and discussion, be taken as due only to that term:

$$\varepsilon_{max} = R_o - R_o(1 - \gamma^2)^{\frac{1}{2}}$$

$$\simeq \frac{R_o\gamma^2}{2} = \frac{E\gamma}{2} = \frac{E^2}{2R_o} \qquad 3.10$$

The acceptability of this magnitude of error will clearly depend upon the measurement being made. Controlling the approximation simply by the use of γ is suitable if it is only desired to hold the radial variation within some fraction of the circle radius. It is less clear whether it gives adequate control for measuring

out-of-roundness from the limacon rather than for the circle. Although there is an intuitive expectation that the out-of-roundness correlates with part radius (for instance, because the size of the mechanical loop needed will affect the stiffness of the manufacturing machine), over the normal range of workpiece sizes encountered in roundness measurement an absolute error magnitude control will also be needed. Figure 3.3 illustrates how γ relates to typical current applications by plotting the value of ϵ_{max} against R_o at constant γ. Also shown on the figure is the magnitude of spindle errors (minimum zone criterion) of typical instruments, 0.05μm to 0.2μm and an area designating the zonal out-of-roundness and radius of workpieces which might be most commonly handled by such instruments, 0.5μm to 10μm and 2mm to 50mm, respectively. This information is tentative, not being based upon a formal survey of instrument users but on casual marketing feedback from customers of Rank Taylor Hobson Ltd. (unpublished). From the figure it is apparent that holding γ below 0.001 is almost an absolute control whereas γ = 0.01 will most often generate rather larger errors (in relation to the out-of-roundness being measured) than would normally be accepted. When in this work a test is required on the validity of a particular operation, a standard condition will be to evaluate it at γ = 0.01 (assuming its validity increases as γ decreases) since this represents a sensible 'worst case' condition on the normal use of limacon methods. It also happens to represent something like the lowest eccentricity ratio which is likely to be encountered in current practice; for example 10μm eccentricity on a 1mm radius workpiece could be accomodated on most instruments for magnifications up to about 1000, that is to the low end of their most commonly used working range.

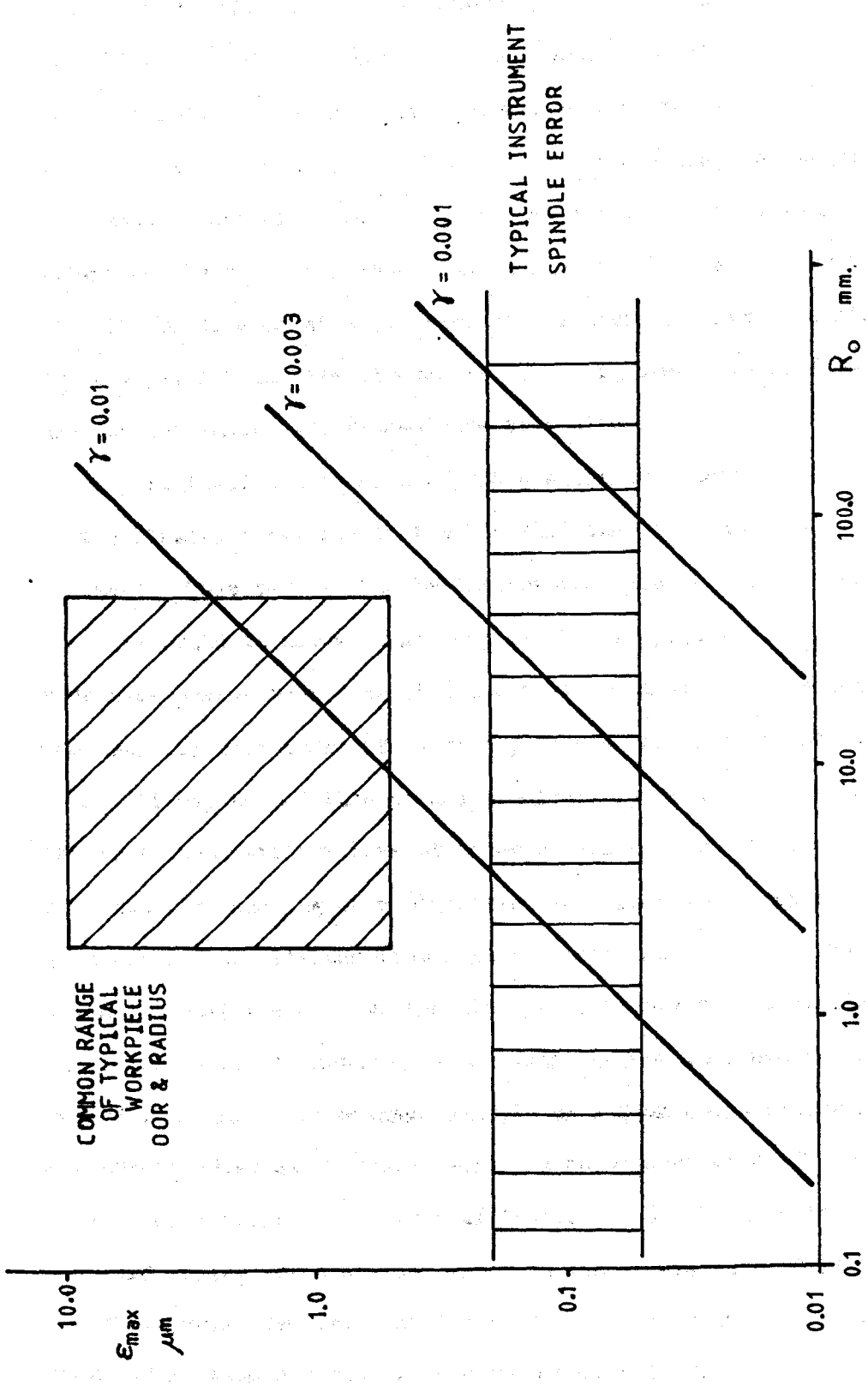


FIGURE 3.3: Maximum Divergence between Circle and Limacon References, ε_{max} Plotted against Part Radius R_o and Eccentricity Ratio γ.

Instrumentally, the achievement of eccentricity ratios of the order of 0.001 or better will not usually be difficult; even with 1mm radius components it requires only a precision of 1μm in the centring device. The limacon may, therefore, be regarded as a good, practical alternative to the use of a circle as a reference figure in instrument co-ordinates. In chart co-ordinates the limacon is much more tolerant of eccentricity than is a circle, for any given allowable divergence of the reference shape from that of the radius suppressed eccentric circle.

A practical consequence of the greater tolerance to eccentricity of the limacon reference is the "de-skilling" of roundness measurement. This term should be interpreted here in a broad sense. Not only could it allow relatively unskilled operators to perform roundness measurement accurately, but it could also, by reducing the need for refinement of centring, reduce the measurement cycle time. It may be possible to use considerably cheaper fixturing on the instrument without loss of accuracy and make the introduction of automatic handling more plausible. Alternatively greater accuracy can be obtained from a given instrument set-up. The potential 'improvement' of the limacon reference over a circle on the chart is often not realisable with current instruments and methods. Consider an extreme example of a 25mm component, which for normal measurement would require a magnification of 20000 (the highest available on typical instruments). The American Standard ruling for a maximum error in the reference of 0.25mm on the chart would require centring on a 50mm radius chart to better than 0.25μm with a circle reference but only to 22μm ($\gamma \cong 0.001$) with a limacon. However, on magnification the 22μm eccentricity

would appear as 440mm, much larger than the chart! Conventional instruments will never cope with, or need, such eccentricities but the better accuracy performance of the limacon is of direct significance when attempts to extend the instrument range computationally, with a consequent removal from the scheme of eccentric polar charts, are made (21). The extra tolerance of the limacon may also justify, for some applications, the higher cost of high range: resolution transducers (for example, (22)) by allowing savings elsewhere in the measurement system.

To summarise, the case for using limacon reference figures rests on the following properties. It well approximates a circle in instrument co-ordinates under sensible instrumental conditions. It can be used consistently in instrument and chart co-ordinate frames and is generally a more accurate representation of the desired reference on the chart than is a circle. It has considerable computational advantages and gives unique answers for all four reference criteria. The quality of the limacon approximation is judged, at least initially, by the eccentricity ratio in instrument co-ordinates. The use of a circular reference on the chart is very convenient for direct graphical working and according to this scheme is allowed by taking the circle as an approximation to the limacon in chart co-ordinates. The quality of this approximation is governed by the eccentricity ratio in chart co-ordinates.

4. Roundness References: Parameter Space Analysis

4.1 Concepts of Parameter Space

The fitting of reference figures is identified mathematically with problems of constrained optimisation and so it seems logical to apply the analytical methods of optimisation to the study of such references. In particular the form of their representation in so-called parameter space will be investigated. Parameter space is a well known concept in optimisation theory and operations research and similar ideas are used in control theory (for example state plane analysis). However, the method seems not to have been much used in other branches of engineering and, in particular, appears to be new to surface metrology. The basic ideas involved will therefore, be reiterated in the course of this discussion.

The basic proposition involved is that an optimisation problem with respect to a set of parameters x_1, x_2, ..., x_n can be represented in an n-dimension co-ordinate system which has each parameter allocated to an orthogonal axis. In such a space both the objective function (that is the function being optimised) and any constraints appear as hyper-surfaces. Clearly the purpose in doing this (except possibly in problems with only two or three parameters) is not directly to allow graphical solutions but because the geometry of the parameter space is of great importance. In some simple cases it could be that parameter space and the co-ordinate system in which the problem is formulated are the same (finding the minimum of $y = x^2$, would be a trivial example) but generally this is not so and it is important to make clear the distinction. To allow

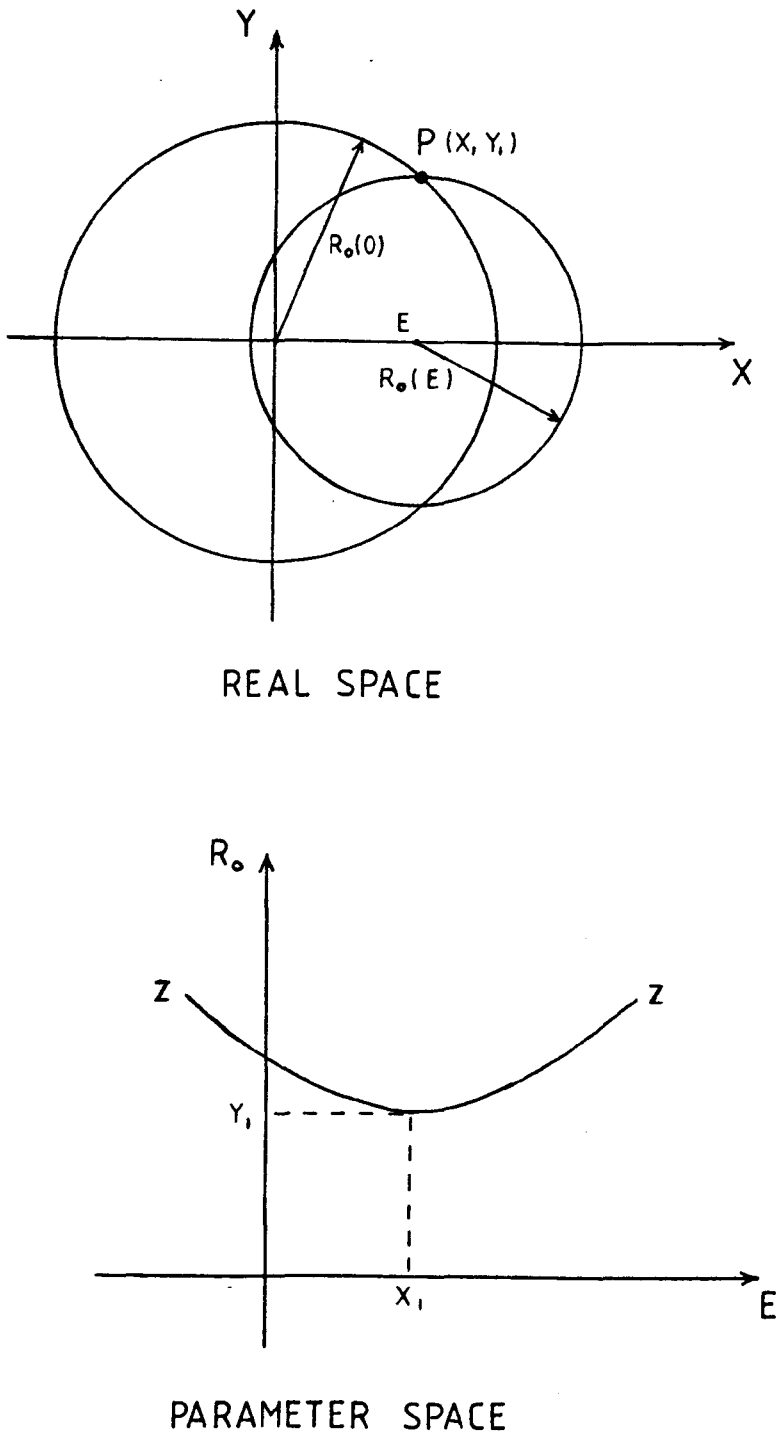


FIGURE 4.1: Circle on X - axis Passing Through Fixed Point: The Concept of Parameter Space Representation

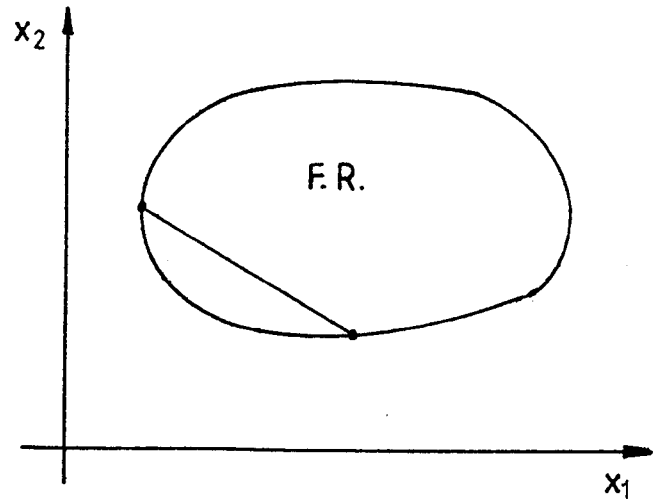
visualisation the discussion will deal mainly with examples having only two parameters but the results are totally general and apply to n-dimensional parameter space. Consider a simple example relevant to this current work, figure 4.1. A point P in the XY plane is given and it is desired to construct a circle with centre lying on the x-axis which passes through P. Defining the distance of the centre from the origin as E and the necessary radius at any value of E as R_o then the properties of the desired circle can be plotted in the two-dimensional parameter space with E and R_o as co-ordinates. Intuition dictates that the parameter space plot will be identical with the locus of the maximum Y value of the circle in the XY plane. The parameter space represenation may be seen not to be just a mathematical artefact but to have direct physical significance. If it is desired to discover what value of E gives the smallest circle, the line ZZ represents the objective function which has a single minimum at $E = X_1$ (again intuitively obvious). Line ZZ has another important property in dividing the ER_o plane into two areas. The area 'above' the line contains all those combinations of E and R_o which would produce a circle which would enclose the point P and the area below the line describes all the combinations of parameters which would fail to enclose the point. The line corresponds to a constraint of exactly the type required by equation 3.4. In the usual terminology the line represents the boundary between the feasible region (in which the combination of parameters obeys the condition) and the infeasible region for the circumscription of the point. Before pursuing this line of investigation of circle fitting, a few general properties of optimisation problems will be discussed in terms of parameter space.

The representation of a general optimisation problem would consist of a series of hyper-surfaces representing the constraints which together form an envelope which is the boundary of the feasible region and, in effect, a series of hyper-surfaces corresponding to constant values of the objective function. The desired result is to discover the largest or smallest value of this constant which generates an objective function hyper-surface intersecting the feasible region. This value will always occur with a hyper-surface which just touches the boundary of the feasible region. Except for a few special cases having a fully analytical solution, the discovery of the optimum will depend upon a search algorithm, usually iterative, which is expected to converge onto the solution. If the problem has several turning points (either or both of the objective functions and the constraints can cause this) then it is usually fairly easy to converge onto a local optimum but much more difficult to establish whether this represents the global optimum or not. Once an iterative procedure approaches a local optimum it will become entrapped and only by restarting the iterations from an alternative initial condition might an alternative, possibly more extreme, optimum be discovered. Unless it is known, in advance, how many local optima exist, it is almost impossible to guarantee that the global value has been found. By doing more and more computation the likelihood that the global value is found increases but certainty is never achieved. In most real problems the exact forms of constraints and objective functions are governed by measurements data and so their exact interaction is extremely difficult to predict. The only method then of discovering the global solution is to restrict operations to problems which have always

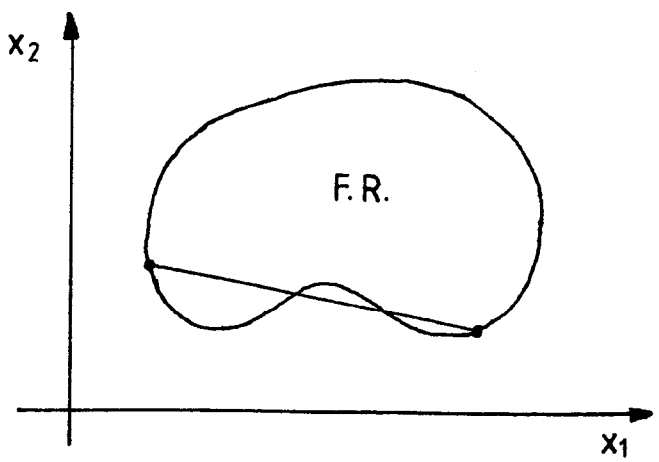
the same number of local optime. In practice this implies that the problem has just one optimum value. Also these problems represent the only class for which computational efficiencies are high enough to allow solution as part of a normal instrumentation system.

The existence of just a single optimum to a specific problem relates to its geometry in parameter space: both the objective function and the feasible region must be convex. Convexity here has its normal geometrical meaning that any two points within the region are joined by a straight line which lies totally within that region, see the examples of figure 4.2. Considering simply the attempt to minimise x_2 in figure 4.2 shows the significance of convexity of the feasible region. Since the overall feasible region is the intersection of the constraint surfaces, it will always be convex if all of the individual constraints are convex. If any constraint is non-convex, the intersection will be non-convex if the offending part of that constraint is active. Thus any single non-convex element in the problem removes the guarantee of a unique solution, although conversely, it does not indicate the definite presence of alternate optima: in figure 4.2 the maximisation of x_2 is unique even for the non-convex feasible region.

In an n-dimensional space a hyper-plane (in two dimensions, a straight line) has the unique property of dividing that space into two regions both of which are convex. Thus the linear property has a special place in optimisation theory: any optimisation problem which involves only linear functions of its parameters is always solvable by general methods and gives a unique optimal value. About no other functions can such a wide statement be made, although



CONVEX



NON-CONVEX

FIGURE 4.2: Convexity of a Feasible Region

other classes are solvable generally. There are three important classes which can be solved which have particular relevance here. Unconstrained optimisation with a convex (linear or quadratic) objective function (linear least squares is in this group); linear objective functions with linear constraints giving problems in linear programming; quadratic objective functions with linear constraints giving problems in quadratic programming. The latter two classes are relevant to the solution of the boundary references in roundness measurement.

4.2 The Analysis of Circular References

The previous section indicated the physical significance of the plotting of a constraint from a ring gauge circle problem in parameter space. In further studying the nature of the circle fits, the simplified ring gauge with two parameters will be used initially so that a pictorial representation can be given. Table 4.1 gives a set of data points chosen to illustrate these features. The data is symmetrical about the $\theta=0$ axis and the minimum radius circumscribing circle having its centre lying on that axis is required to be found. The simplified form of equation 3.4 will be

$$\begin{aligned} &\text{Minimise} \quad z = R_o \\ &\text{Subject to} \quad R_i \leqslant A\cos\theta_i + [R_o^2 - A^2\sin^2\theta_i]^{\frac{1}{2}} \end{aligned} \qquad 4.1$$

In the parameter space, AR_o, the boundary of the feasible region will be given by the equality condition of the constraint for any (R_i, θ_i):

Angle θ_i	Radial distance R_i
30°	1.0
45°	0.5
90°	0.5
135°	0.6
225°	0.6
270°	0.5
315°	0.5
330°	1.0

TABLE 4.1: Example Data, Symmetrical about $\theta=0^\circ$

$$R_o^2 = R_i^2 + A^2 - 2AR_i\cos\theta_i \qquad 4.2$$

These lines are shown in figure 4.3 for the data of Table 4.1. The feasible region for circumscribing circles lies above the boundary line which, since the boundary is hyperbolic, will be convex. The objective function is linear, the contours of constant value being lines parallel to the A-axis. Thus it is seen that there is a unique solution to the ring gauge, corresponding in the diagram to point J.

The equivalent plug gauge fit will be equation 4.1 with the direction of optimisation and inequality reversed, so the constraint lines will be identical to those of the ring gauge, equation 4.2 but the feasible region will now lie below the line. This region is non-convex, as is the intersection of the region from all the constraints. The contours of the objective function are, as before, lines parallel to the A-axis. In this example the non-convexity causes two independent local maxima, corresponding to points K and L on figure 4.3. The plug gauge is not necessarily unique. A peculiar feature of the plug gauge problem is demonstrated by figure 4.3 in that, since the constraint lines are hyperbolic, as A tends to very large positive or negative values, the feasible region allows very large R_0 to occur. This is a consequence of using sampled data and corresponds in physical terms to a circle initially inside the data shrinking until it is small enough to slip between the data points and once outside it can then expand for ever without enclosing the data. In this example, the centre must lie on the $\theta=0$ axis and so placing a data point on that axis should stop the circle 'escaping'. For $\theta_i = 0$, equation 4.2 becomes:

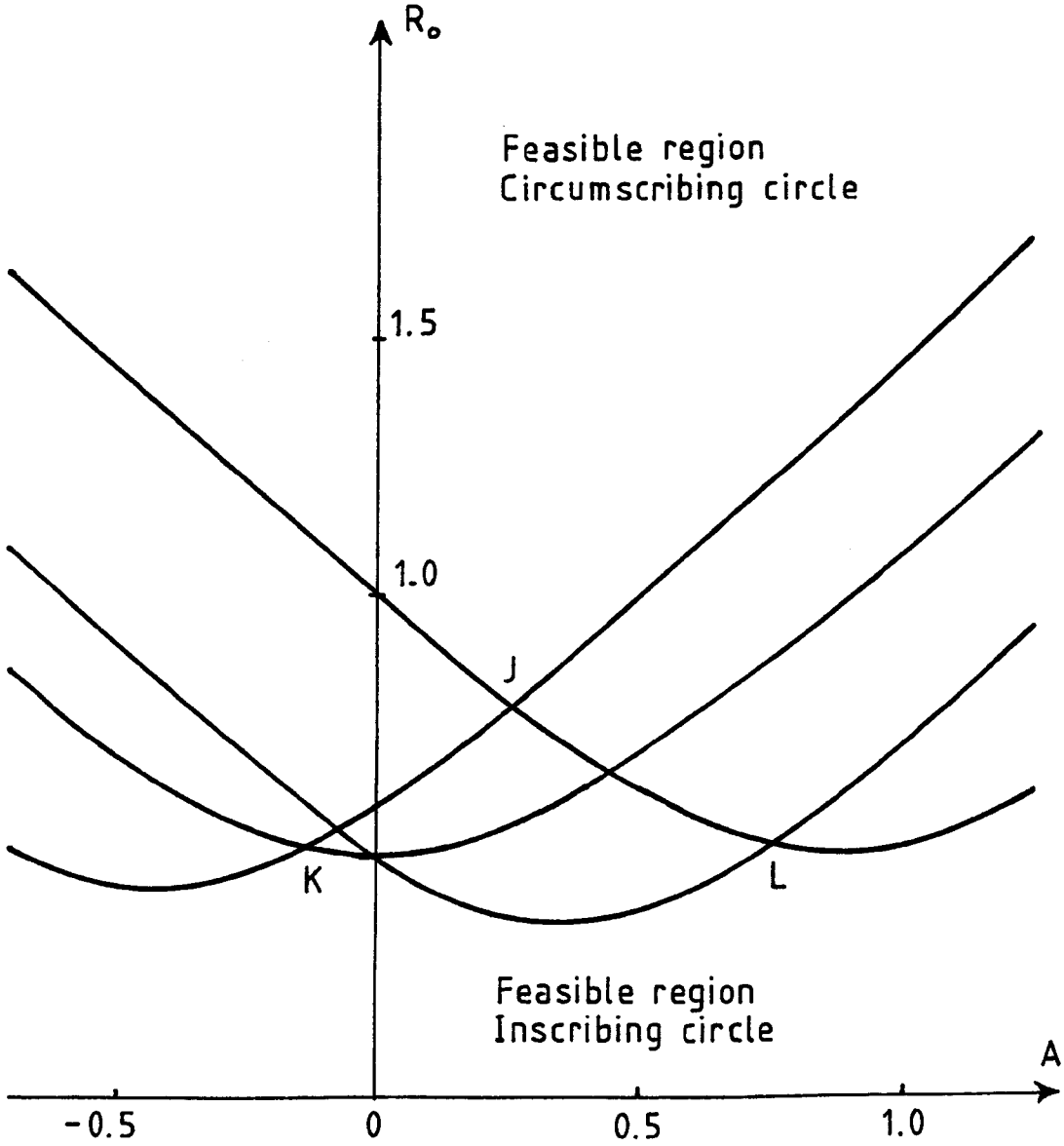


FIGURE 4.3: Parameter Space Representation of Circumscribing and Inscribing Circle Constraints (see Table 4.1)

$$R_o = R_i + A$$

which is a straight line intersecting the R_o and A axes and so limiting the extent of the feasible region as predicted. If the centre was allowed to leave the $\theta=0$ axis the circle would always be able to 'escape' through sampled data. Thus any attempt at a general solution to the plug gauge should include extra constraints to limit the feasible region to that part which has physical relevance.

Extending, now, this discussion to the general ring and plug gauge circles, a three dimensional parameter space ABR_o will be used. The form of the constraints will then be found directly from the equation 3.4 expressed in terms of A and B, see equation 3.6:

$$R_o^2 = R_i^2 + A^2 + B^2 - 2AR_i\cos\theta_i - 2BR_i\sin\theta_i \qquad 4.3$$

The simplfied form just discussed, equation 4.2, is the cross - section of this three dimensional figure in the plane B = 0. The full figure is hyperbolic in any cross-section parallel to the R_o axis and circular in any sectional plane parallel to the AB plane. It is convex in the circumscribing region and non-convex for the inscribing region and so the discussion already given applies equally well to the full form.

The consideration of the geometry of parameter space has thus derived a result which can also be found by considering the geometry of real space namely that there is always a unique minimum circumscribing circle to a set of data points but there is not necessarily a unique solution for the maximum inscribing circle. (Note, here, that BS 3730 (2) wrongly states that the ring gauge may be non-unique.)

The minimum zone circle, equation 3.5, has a four dimensional

parameter space ABR_oH and so no attempt will be made to sketch the feasible region of even a simplified problem. Again the objective function is linear, its constant contours being hyper-planes perpendicular to the H-axis. The outer circle constraint may be written as

$$H \geqslant R_i - A\cos\theta_i - B\sin\theta_i - [R_o^2 - (A\sin\theta_i - B\cos\theta_i)^2]^{\frac{1}{2}} \qquad 4.4$$

the constraint boundary being the equality condition and the feasible region lying above (in the sense at larger H) the boundary. The convexity of this feasible region is tested in the formal way that if $\underline{A}_1 = (A_1\ B_1\ R_1)^T$ and $\underline{A}_2 = (A_2\ B_2\ R_2)^T$ are vectors representing points on the boundary of the region any point on the straight line joining them must have greater value than H at the same values of A, B and R_o, so:

$$(1-\lambda)H(\underline{A}_1) + \lambda H(\underline{A}_2) \geqslant H((1-\lambda)\underline{A}_1 + \lambda\underline{A}_2) \qquad 4.5$$

where λ is a scalar constant taking any value between 0 and 1. The terms of equation 4.4 which are linear in the parameters will not alter the presence or absence of convexity caused by the other terms as can be seen from equation 4.5 from which they cancel. Convexity therefore depends upon the truth for all $\underline{A}$ and λ of the inequality:

$$-(1-\lambda)[R_1^2-(A_1\sin\theta_i-B_1\cos\theta_i)^2]^{\frac{1}{2}} - \lambda[R_2^2-(A_2\sin\theta_i-B\cos\theta_i)^2]^{\frac{1}{2}}$$
$$\geqslant -[((1-\lambda)R_1+\lambda R_2)^2 - \{((1-\lambda)A_1+\lambda A_2)\sin\theta_i - ((1-\lambda)B_1+\lambda B_2)\cos\theta_i\}^2]^{\frac{1}{2}}$$

From the physical reality of the problem originally posed, interest lies only in conditions for which the origin is enclosed by both circles for which condition the eccentricity is less than the radius and all three square roots will be positive. Multiplying

through by -1 and changing the sense of the inequality then makes both sides always positive and so they may be squared. After cancelling out terms common to both sides this operation yields:

$$(1-\lambda)\lambda[R_1R_2-(A_1\sin\theta_i-B_1\cos\theta_i)(A_2\sin\theta_i-B_2\cos\theta_i)]$$
$$\geq\lambda(1-\lambda)[\{R_1^2-(A_1\sin\theta_i-B_1\cos\theta_i)^2\}\{R_2^2-(A_2\sin\theta_i-B_2\cos\theta_i)^2\}]^{\frac{1}{2}}$$

Again the condition that the eccentricity is less than the radius ensures that both sides are positive and so the root can be removed by squaring. After squaring the inequality simplifies to give:

$$[R_1(A_2\sin\theta_i-B_2\cos\theta_i) - R_2(A_1\sin\theta_i-B_1\cos\theta_i)]^2 \geq 0$$

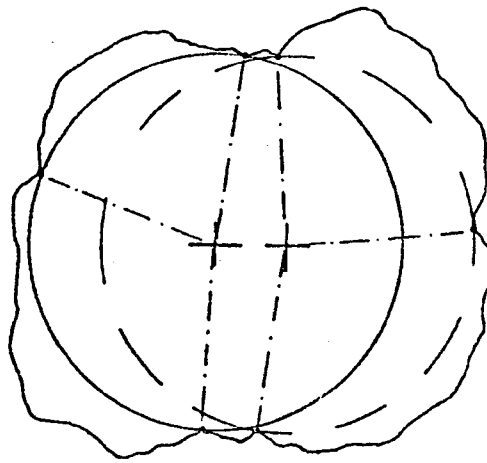
which is always true and so the equation 4.4 is convex.

The constraint of equation 3.5 pertaining to the inner circle may be analysed in the same way. It is:

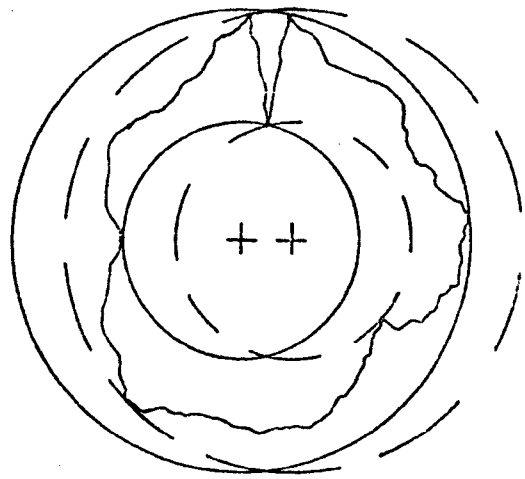
$$H \geq -R_i + A\cos\theta_i + B\sin\theta_i + [R_o^2-(A\sin\theta_i-B\cos\theta_i)^2]^{\frac{1}{2}}$$

Its right-hand-side is the negative of equation 4.4 and so describes a concave surface (convex for increasing H) if equation 4.4 is convex. Thus for the minimising problem the constraint is non-convex and so the whole feasible region may be non-convex. The minimum zone circle is formally shown to be potentially non-unique.

In fact the non-uniqueness of both plug gauge and minimum zone can be confirmed by graphical example, figure 4.4 but, particularly for the zone, the analysis is really needed to prove the circles shown are local optima. The profiles shown in figure 4.4 look much more like chart representations than real profiles in instrument co-ordinates. There are two reasons for this, one being that if it were not so the profile out-of-roundness would not be visible on the figure and the second being that the feature most likely to



MIC



MZC

FIGURE 4.4: Examples of Non-uniqueness for Plug Gauge Circle (MIC) and Minimum Zone Circles (MZC)

generate non-unique solutions (again, particularly for the zone) is a rapid, large change of radius. Such changes are much more common on the chart because the radius variation of the profile is magnified relative to its mean radius. It appears then that not only is the circle an inaccurate form of reference on the chart but it is also susceptible to significant non-uniqueness problems.

4.3 Solving Reference Circles

The previous section has demonstrated the behaviour of reference circles in parameter space, but so far no attempt has been made to develop a formal method of solving them. The formulations of the fitting problems given do not directly compare with any of the forms having known solutions quoted in section 4.1 and there are difficulties of practical application caused by non-convexity in some cases. The ring gauge is at least unique and so a general solution should exist. The constraints of the ring gauge, although convex, are non-linear while for standard solutions linear constraints are needed. However in this case (as first pointed out by P. H. Philipson) a change of variable can give linear versions which will be usable providing the objective function remains a suitable function of the new variables. From the cartesian expression for a circle, the constraints of the ring gauge may be written

$$R_o^2 \geqslant (X_i-A)^2 + (Y_i-B)^2 \quad ,$$

$$X_i = R_i\cos\theta_i \quad \text{and} \quad Y_i = R_i\sin\theta_i .$$

Multiplying out gives

$$R_i^2 \leqslant R_o^2 - E^2 + 2X_i.A + 2Y_i.B$$

$$= C + 2X_i.A + 2Y_i.B$$

so that the constraints are linear in the parameters A, B and $C = (R_o^2 - E^2)$. The original objective was to minimise R_o and, since $R_o > 0$, it is equally valid to minimise R_o^2. But

$$R_o^2 = C + E^2 = C + A^2 + B^2$$

which is a quadratic function of the parameters. The ring gauge can thus be formulated as a problem in quadratic programming:

$$\text{Minimise } z = C + A^2 + B^2$$

$$\text{Subject to: } 2R_i\cos\theta_i.A + 2R_i\sin\theta_i.B + C \geqslant R_i^2$$

and so standard methods are available for its solution. The methods are not, however, particularly easy and a large amount of computing would be needed for complete solution.

The plug gauge clearly has a formulation very similar to that just given. This indicates not that a general solution is available for it but that not all problems which may be formulated as quadratic programs are convex and so capable of being solved directly for the global optimum. The minimum zone circles do not yield at all to this approach since H cannot be expressed as a simple function of variables that linearise the constraints.

It seems clear that direct attempts at fitting circular references are rather unsatisfactory. The re-expression of the constraints in linear form is not particularly successful in terms of generating solutions but is nevertheless an operation of importance to the solution. For instance, since the ring and plug gauge circles have the same constraint boundaries but with feasible regions

lying on opposite sides of those boundaries, the only possibility for simultaneously obtaining convex regions for both is to have linear constraints. Since there is an incentive, in terms of computing effort, to simplify, the approach is to replace the circle constraints by boundaries which approximate them and which are linear. Each constraint can be linearised about a suitable point in parameter space by using a truncated Taylor's Series. Since, normally, instrumental constraints will ensure that the eccentricity is small a suitable point of linearisation for general consideration would be $A=B=0$, $R_o=R_i$. It will be noticed that this argument is the same as that pursued in section 3.2 and produces the same result: the linearisation of the constraint changes the problem from one of circle fitting to one of limacon fitting. This form of linearisation leaves the objective function unaltered and so limacon fitting, having linear constraints and linear objective functions, belongs to the class of linear-programmes.

4.4 The Analysis of Limacon References

The geometric properties of limacon references have been discussed in section 3.3 so all it is necessary to do here is to study the implications of the limacon reference in parameter space. Only the simple limacon form, representing linearisation of the circle constraints about $A=B=0$ will be considered here partly because of the complexity of the expression about any other point (equation 3.7) but mainly because that form is directly amenable to radius suppression and so would be used on virtually all real measurements.

The minimum radius circumscribing limacon can be stated:

$$\text{Minimise } z = R_o$$

$$\text{Subject to: } A\cos\theta_i + B\sin\theta_i + R_o \geq R_i \qquad 4.6$$

$$\text{for all } (R_i, \theta_i)$$

and again the maximum radius inscribing limacon will seek to maximise R_o subject to constraints identical apart from the sense of the inequality to equation 4.6. The feasible region for these figures using the simple example of section 4.2 is shown in figure 4.5. For both ring and plug limacons they are convex and a single minimum occurs at point J (equivalent to point J on figure 4.3) and a single maximum at point K (equivalent to the local maximum K of figure 4.3). The equivalent of point L in the earlier diagram is no longer in the feasible region here. The linear optimisation problem has one possible difficulty over the the definition of optimum parameter values, in that, since the feasible region boundary is made up of straight lines one section may be co-linear with the optimum contour of the objective function. This is illustrated by the section KQ of the inscribing region in figure 4.5. The single optimum radius exists but occurs over a finite continuous range of eccentricities: it can only occur over a single continuous region. (An easily visualised analogy of this behaviour is the inscribing circle to a rectangle!) Although shown in the figure as associated with the inscribing limacon, the same feature could occur with the circumscribing limacon. This is different to the case with a circle fit where the curvature of the constraint would ensure that the minimum occured at just one point along that constraint boundary.

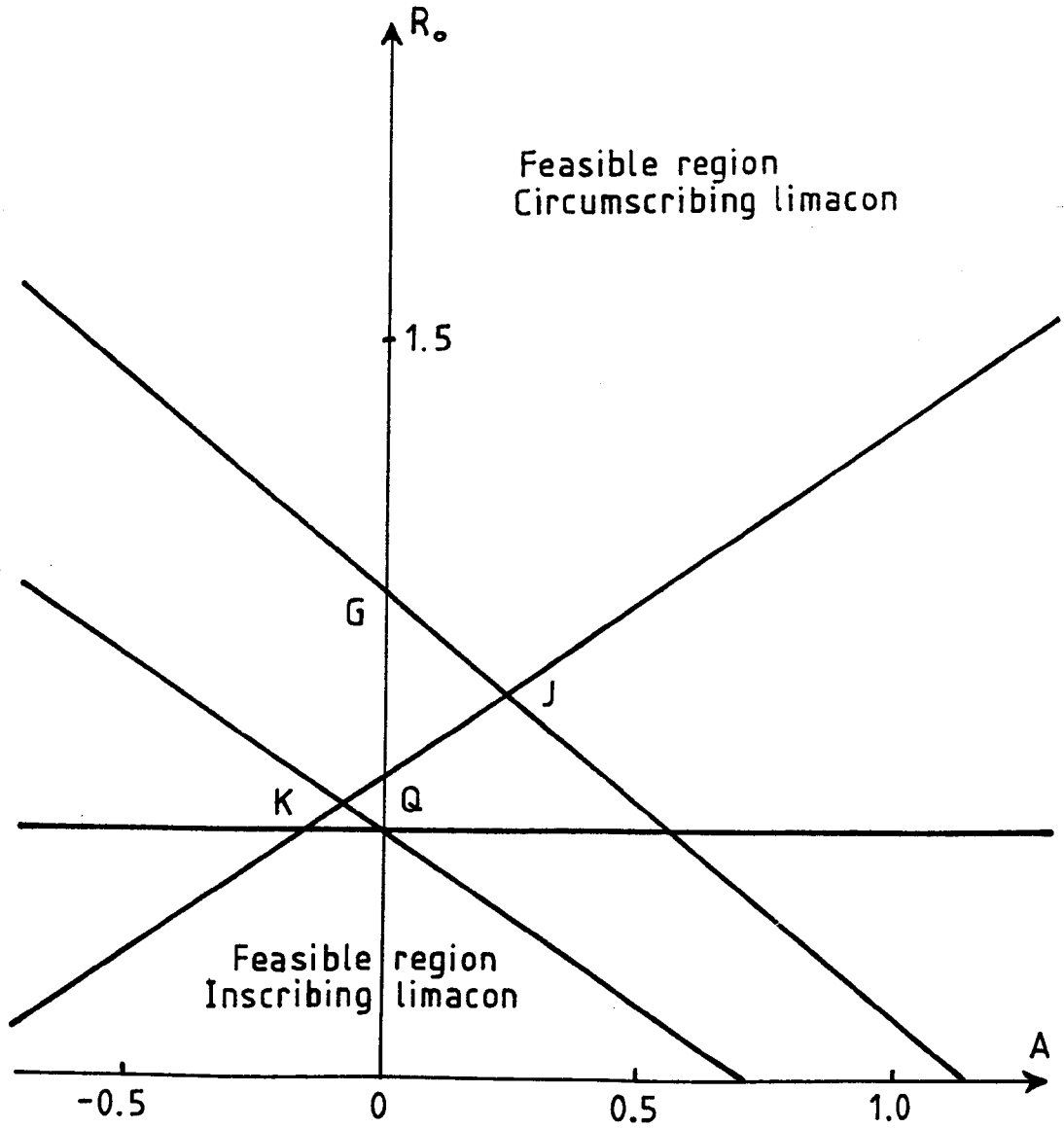


FIGURE 4.5: Parameter Space Representation of Circumscribing and Inscribing Limacons (see Table 4.1)

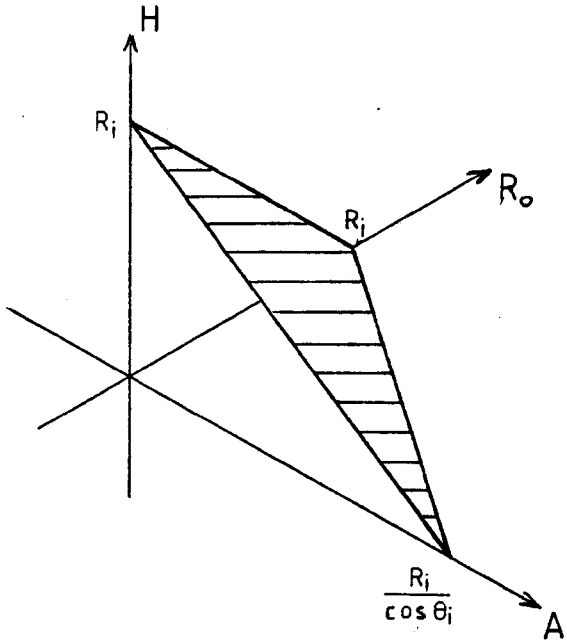
Figure 4.5 also illustrates the source of the computational advantages of linear programming: it can be seen that the optimum solution lies not only on the boundary of the feasible region but at the intersection of two constraints. Efficiency is gained because it is necessary to search not the entire boundary but only its vertices in the general n-dimensional parameter space.

The minimum zone limacon fit can be stated:

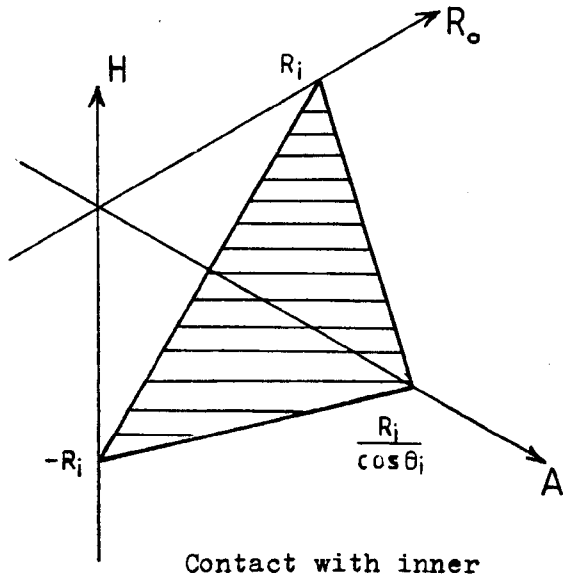
$$\begin{aligned} &\text{Minimise } z = H \\ &\text{Subject to: } H \geqslant R_i - A\cos\theta_i - B\sin\theta_i - R_o \\ &\qquad\qquad H \geqslant A\cos\theta_i + B\sin\theta_i + R_o - R_i \end{aligned} \qquad 4.7$$

The interaction of the two constraints can be illustrated for the condition B=0. They represent two planes in the three dimensional HAR_o space as shown in figure 4.6. These planes intersect in the H=0 plane and together form a V-shaped boundary which is everywhere in the region $H \geqslant 0$ as is needed for physical sense. Parallel vertices of the 'V' can occur for at most two points since the direction of the vertex in parameter space depends only on θ_i. Also, under sensible conditions of data, the sides of the 'V' can never become parallel to the plane H=constant. The intersection of the constraints causing the boundary of the feasible region will, therefore, involve either a point or a line not parallel to H=constant. Thus it appears that the minimum zone limacon formulation does not allow the continuous region of optimum parameters which occurs with the other boundary references, but always gives a completely unique answer.

As the boundaries of the limacon feasible regions are derived from those of the appropriate circle by truncating the Taylor Series of the latter, the value of the function and its first derivative will



Contact with outer



Contact with inner

FIGURE 4.6: Representation of Constraints from One Data Point in Reduced (A,R_o,H) Parameter Space for Minimum Zone Limacon

be the same for both at the point of linearisation. Thus the limacon constraint planes (in 3 or 4 dimensions) are tangents to the circle constraint surfaces on the R_o- (or H- for minimum zone) axis; compare fiures 4.3 and 4.5. Since the circle constraints are convex functions their tangents at any point never intersects them and so in the case of ring and plug gauges the limacon constraint lies always below (at smaller R_o) its equivalent circle constraint. The solution for either problem will therefore be lower for the limacon than the circle. The limacon fit consistently underestimates the radius of the ring and plug circles. If subscribing to the view that ring and plug references should be used on shafts and holes respectively, then for judging a fit between shaft and hole the 'error' of using a limacon for the plug is in the safer direction and that of the ring in the less safe direction (taking 'safe' to mean that they will fit if the measurement says so).

The minimum zone limacon constraints lie below the convex circle constraint and above the concave one and since it is not known which will be of most influence in a given situation it is not possible to state in general whether the limacon zone over- or under-estimates the circle zone.

In principle the least squares circle and limacon could also be studied in parameter space but other than the general statement about the need for convexity of the function being minimised little illumination is to be gained by doing so. The least squares reference belongs to a different group of problems to the boundary references and will be studied in the next chapter before returning to the practical application of these observations to the boundary references in Chapter 6.

4.5 Parameter Space and Chart Co-ordinates

The analysis so far carried out in this chapter, although completely relating to instrument co-ordinates, is in fact perfectly general. The fitting of circles and limacons in chart co-ordinates is governed exactly as has been discussed. The third form of reference boundary of interest is that of the true chart representation of the eccentric circle in instrument co-ordinates. To illustrate its general behaviour it is convenient to simplify the algebra by assuming that B=0 and that there is no imposed chart radius, S=0. The circle constraint, equation 4.2, can be expressed in terms of the chart parameters a, r_o and the chart data r_i using radius suppression L and magnification M:

$$\left(\frac{r_o}{M} + L\right)^2 = \left(\frac{r_i}{M} + L\right)^2 + \frac{a^2}{M^2} - \frac{2a}{M}\left(\frac{r_i}{M} + L\right)\cos\theta_i$$

Since M>0, this can be multiplied through by M^2 and re-arranged to give the boundary of the feasible region as:

$$r_o = \left[(r_i+ML)^2 - 2a(r_i+ML)\cos\theta_i+a^2\right]^{\frac{1}{2}} - ML \qquad 4.8$$

This is still a convex form in the same sense as equation 4.2 and if ML=0 reduces to that equation with merely different symbols used. Essentially there is no difference between equations 4.2 and 4.8. At large ML values the chart co-ordinate form may appear to be flattened with respect to the instrument co-ordinates form but this is a graphical effect caused since the a-axis is stretched M times with respect to the A-axis while there is not similar effect between the r_o- and R_o-axes. Because radial distance is an axis of the parameter space co-ordinate system, the representation there is, in

a sense, independent of radius suppression.

The boundaries for fitting circles on the chart do not relate to the stretched axis and so have much more curvature than the true boundaries. This again supports the view that plug circles on the chart are more likely to show multiple local maxima than are the plug circles properly fitted in instrument co-ordinates.

5. Least Squares References

5.1 Background

Generally speaking, the surface metrology community has accepted for many years the calculation of the "least squares circle" (LSC). With the exception of a few dissenters, mainly dimensional metrologists, (see Appendix 3 for a discussion of their viewpoint) the circle parameters are accepted as those quoted in BS 3730 in a derivation attributed by Reason (23) to R. C. Spragg (1960). The standard defines the centre of the least squares circle as that "point from which the sum of the squares of a sufficient number of equally spaced radial ordinates has a minimum value". This choice of wording is unfortunate for although the centre of the LSC does have this property, it is not the formal definition of least squares. Formally, the requirement is the minimisation of the sum of squares of residuals, that is of the radial distances of the ordinates from the reference figure. The derivation given in the Standards does correctly use the residuals but other workers have been led into erroneous assumptions by this wording, see, for example, (7).

The Standard claims to be attempting to fit a circle to data from the polar chart but to be doing so by the use of approximate formulae. As the solution of non-linear least squares is problematical, it is no surprise to find that the approximations made in the derivation are in effect a linearisation of the reference figure in its parameters. Thus the derivation is an exact solution for the least squares limacon! Two other points may be discovered in this derivation. Firstly it uses, as a simplifying device, a free switching between summations

of discrete data and definite integrals of an assumed continuous profile. The safety of this operation was perhaps overestimated since the only acknowledgement to it comes in the requirement that there be a "sufficient" number of points in the original definition. Secondly, by making these simplifying moves early in the derivation, it becomes immediately restricted to the special case: the well known results which are obtained are true only for the arrangement of data stated and no information about other configurations of data is available. Thus in saying that the least squares approach is not generally applicable to interrupted profiles, the Standard should only refer to these simplified formulae for determining the least squares limacon parameters.

Historically, this work seems to have been the first use of the limacon approximation in roundness measurement, although the separate identity of the limacon was not expressed. There was however some recognition of its useful properties on the chart. Reason (3) observes, without explanation, that the error in the least squares circle calculation cancels the error due to chart distortion (radius suppression). Of course, the cancellation is not exact since an infinite series has been truncated, but the remaining divergence would not usually be detectable by instrumentation as opposed to mathematical means.

The first acknowledgement of the separate existence of the limacon in the determination of least squares appears to be attributable to Whitehouse (4). This work basically followed earlier derivations of the limacon approximation and again only identified it for its good modelling of distorted circles on the chart. Its important contribution was to generalise the derivation of the least squares

limacon to profiles consisting of just an arc of the circle. However, in doing so it still used continuum mathematics on a sampled data problem and restricted conditions to equally spaced ordinates, rather than fully generalising. Using a direct implementation of Whitehouse's formulae, programmed by the present author, Siddall (22) demonstrated a significant reduction in parameter accuracy as smaller arcs were taken. In particular it was reported that some simple estimation methods, from which no great accuracy was expected, gave better results than the least squares method on arcs of less than about 45°. No specific explanation of this behaviour was offered at that time.

For all the sources of potential error contained in their derivations these formulae have proved extremely successful in practice, providing that reasonably good centring is maintained. The formulae for complete profiles have been used for many years with virtually no difficulties being suspected from the results and incomplete profiles have also been adequately processed (24). Their wide success is due, in part, to the nature of available instrumentation; the most convenient way of using roundness instruments corresponds generally to the use of the most favourable conditions for these formulae.

5.2 The Least Squares Limacon

The derivations discussed above all used the truncation of a binomial expansion to develop the limacon approximation. An equally good choice for simplifying the problem would have been to use Taylor's Series. Perhaps if this approach had been used the full implications of the limacon reference might have been realised earlier since the

association of Taylor's Series with linearisation is well known. Be that as it may, it is clear from the previous discussion that this study will rely heavily upon that linearity.

The reference figure which is really required will be a circle fitted to the data in instrument co-ordinates according to the principle of least squares. The first step, therefore, is to define the residuals of this problem. Referring to the notation given on Figure 5.1, the true residual is seen to be:

$$\varepsilon_i = S_i - R_o \qquad 5.1$$

At this stage, the only known point of the system is the origin and so the data must be expressed relative to the origin, as must the circle parameters. Thus equation 5.1 becomes:

$$\varepsilon_i = \left[(R_i - E\cos(\theta_i - \phi))^2 + E^2\sin^2(\theta_i - \phi)\right]^{\frac{1}{2}} - R_o \qquad 5.2$$

Even if this is expressed in terms of (A, B) rather than (E,ϕ) it is non-linear in its parameters. Since only the linear least squares is generally solvable, the expression for residuals must be linearised about some estimate (A_o, B_o) of the eccentricity. There is basically only one linearisation of any given figure, so this operation will result in a limacon having its origin at (A_o, B_o). The limacon thus has a fundamental role in obtaining a least squares solution in instrument co-ordinates.

In practice it will be necessary to work in chart co-ordinates. From section 3.2, it is seen to be highly desirable that, to get a reference figure which is easily interpretable across the radius suppression transformation, a limacon having its origin identical with that of the co-ordinate system be used. This, of course, is the same limacon approximation as used in previous work. The re-

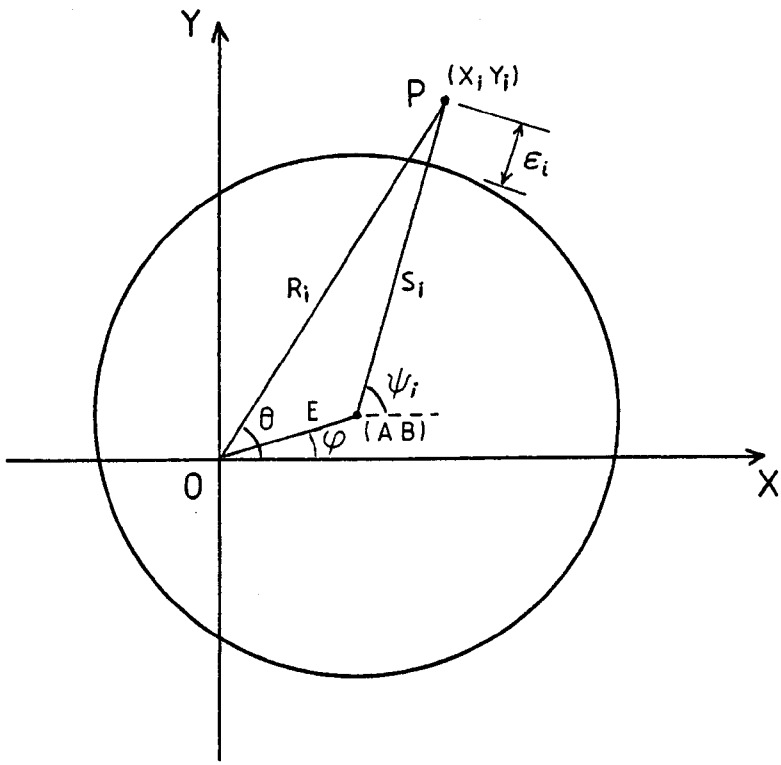


FIGURE 5.1: Measurement from Reference Circle: Residual Definition

development carried out here serves to stress the physical nature of that approximation. It is equivalent to linearisation of the circle in instrument co-ordinates about the point of zero eccentricity ($A_o=B_o=0$). It is, furthermore, a linearisation of the expression for the residuals about that point, this being the real key to the least squares solution. The actual residuals which are minimised will be:

$$\delta_i = R_i - (A\cos\theta_i+B\sin\theta_i+R_L) \qquad 5.3$$

Comparing this with equation 5.2 and Figure 5.1, the approximation for any individual residual is exactly that discussed in section 3.3. What the overall effect on the least squares estimate of the circle pareters will be is rather less predictable. The use of equation 5.3 is the best that can be achieved in chart co-ordinates since it is the only linearisation not requiring knowledge of the radius suppression.

The discussion here has stressed the lack of generality of earlier work on least squares limacons caused by the use of continuous mathematical assumptions. Thus here a fully general derivation of the discrete least squares limacon will be given. The techniques used in the derivation are standard theory but, as it appears that they are less well known than some other methods, a fuller outline of their solution than might otherwise be expected will be given.

The full set of residuals can be expressed as a matrix equation:

$$\begin{pmatrix} \delta_1 \\ \vdots \\ \delta_N \end{pmatrix} = \begin{pmatrix} R_1 \\ \vdots \\ R_N \end{pmatrix} - \begin{pmatrix} \cos\theta_1 & \sin\theta_1 & 1 \\ \vdots & \vdots & \vdots \\ \cos\theta_N & \sin\theta_N & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ R_L \end{pmatrix}$$

or, more compactly:

$$\underline{\delta} = \underline{R} - \underline{H} \cdot \underline{A} \qquad 5.4$$

$\underline{H}$ is seen to represent the experimental design in that it expresses the points at which measurements are taken and $\underline{R}$ represents the results of those measurements. The least squares condition can now be expressed:

$$\underset{\underline{A}}{\text{Min}} \sum_i \delta_i^2 = \underset{\underline{A}}{\text{Min}}\ \underline{\delta}^T.\underline{\delta}$$

where the superscript T indicates transposition. Now:

$$\begin{aligned} \hat{\underline{\delta}}^T\hat{\underline{\delta}} &= (\underline{R}^T - \hat{\underline{A}}^T\underline{H}^T)(\underline{R} - \underline{H}\hat{\underline{A}}) \\ &= \underline{R}^T\underline{R} - \underline{R}^T\underline{H}\hat{\underline{A}} - \hat{\underline{A}}^T\underline{H}^T\underline{R} + \hat{\underline{A}}^T\underline{H}^T\underline{H}\hat{\underline{A}} \end{aligned} \qquad 5.5$$

where the ^ symbol indicated that the values are those estimated from the experimental data (which is all that is physically possible!). It will be noticed that equation 5.5 is similar in form to a perfect square. As squares are easily minimised, the completion of the square is sought. Consider the quadratic form:

$$\begin{aligned} &(\hat{\underline{A}}\underline{H}^T\underline{H} - \underline{R}^T\underline{H})(\underline{H}^T\underline{H})^{-1}(\underline{H}^T\underline{H}\hat{\underline{A}} - \underline{H}^T\underline{R}) \\ &= \hat{\underline{A}}^T\underline{H}^T\underline{H}\hat{\underline{A}} - \underline{R}^T\underline{H}\hat{\underline{A}} - \hat{\underline{A}}^T\underline{H}^T\underline{R} + \underline{R}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{R} \end{aligned}$$

Substituting this into equation 5.5 gives:

$$\begin{aligned} \hat{\underline{\delta}}^T\hat{\underline{\delta}} &= (\hat{\underline{A}}\underline{H}^T\underline{H} - \underline{R}^T\underline{H})(\underline{H}^T\underline{H})^{-1}(\underline{H}^T\underline{H}\hat{\underline{A}} - \underline{H}^T\underline{R}) + \underline{R}^T\underline{R} - \underline{R}^T\underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{R} \\ &= (\hat{\underline{A}}\underline{H}^T\underline{H} - \underline{R}^T\underline{H})(\underline{H}^T\underline{H})^{-1}(\underline{H}^T\underline{H}\hat{\underline{A}} - \underline{H}^T\underline{R}) + \underline{R}^T(\underline{I} - \underline{H}(\underline{H}^T\underline{H})^{-1}\underline{H}^T)\underline{R} \end{aligned}$$

It is this sum of quadratics that is to be minimised. The second

expression is positive and independent of $\underline{\hat{A}}$, so the minimum over $\underline{\hat{A}}$ of $\underline{\hat{\delta}}^T\underline{\hat{\delta}}$ will occur when the first expression is minimum. Since this is a quadratic form, its smallest possible value will be zero which occurs when

$$\underline{H}^T\underline{H}\underline{\hat{A}} - \underline{H}^T\underline{R} = 0 \qquad 5.6$$

This condition upon $\underline{\hat{A}}$ thus represents the least sum of squares of residuals and is the required solution:

$$\underline{\hat{A}} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{R} \qquad 5.7$$

Substituting equation 5.4 into 5.7 allows the estimation errors to be expressed:

$$\underline{\hat{A}} - \underline{A} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{\delta} \qquad 5.8$$

In reality this is of little use since $\underline{\delta}$ cannot be derived from $\underline{\hat{\delta}}$ and so its use must depend upon un-substantiated assumptions. For good estimates $(\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{\delta}$ must be small requiring that the data fluctuates only slightly from the reference (a condition encountered elsewhere) and that $\underline{H}^T\underline{H}$ does not approach singularity. The condition for unbiased estimates (that is an expected error of zero) may be seen to be that the expected value of $\underline{\delta}$ is zero and that $\underline{H}$ and $\underline{\delta}$ are independent.

Returning now to the specific problem in hand, equation 5.7 can be fully expanded to give:

$$\begin{pmatrix} \hat{A} \\ \hat{B} \\ \hat{R}_L \end{pmatrix} = \begin{pmatrix} \Sigma\cos^2\theta & \Sigma\sin\theta\cos\theta & \Sigma\cos\theta \\ \Sigma\sin\theta\cos\theta & \Sigma\sin^2\theta & \Sigma\sin\theta \\ \Sigma\cos\theta & \Sigma\sin\theta & N \end{pmatrix}^{-1} \begin{pmatrix} \Sigma R\cos\theta \\ \Sigma R\sin\theta \\ \Sigma R \end{pmatrix} \qquad 5.9$$

where for clarity the indices have been ommitted: all sums are over i=1 to N with R_i, θ_i. So far the derivation is perfectly

general: no assumptions concerning the positioning of the values θ_i have been made. Equation 5.9 is valid for any profile whether arc, interrupted surface or a normal roundness graph.

Given the general solution for the least squares limacon, it is now time to investigate whether simplifications can be made by selection of specific patterns of θ_i. In particular it is desired to control $\underline{H}^T\underline{H}$ for two reasons; to make the solution of equation 5.9 easier and to ensure that it does not approach singularity. The most desirable action would thus be to diagonalise H^TH which requires that:

$$\sum_i \cos\theta_i = \sum_i \sin\theta_i = \sum_i \sin\theta_i \cos\theta_i = 0$$

To ensure that $\cos\theta$ and $\sin\theta$ simultaneously sum to zero requires that for each point θ there is another point at $\theta+\pi$, while for $\sin 2\theta$ to sum to zero there must be a point at $\theta+\pi/2$. Thus it is necessary to have four-fold symmetry of the data points, but not necessarily equally spaced points, although the most common arrangement would, one expects, be to have a multiple of four points spaced evenly around a full circle. Also, under these conditions:

$$\Sigma \cos^2\theta = \Sigma \frac{\cos 2\theta + 1}{2} = \frac{N}{2} = \Sigma \sin^2\theta$$

So that the well known "least squares circle formulae" are obtained:

$$\begin{pmatrix} \hat{A} \\ \hat{B} \\ \hat{R}_L \end{pmatrix} = \frac{1}{N} \begin{pmatrix} 2\Sigma R_i \cos\theta_i \\ 2\Sigma R_i \sin\theta_i \\ \Sigma R_i \end{pmatrix} \qquad 5.10$$

Two inaccuracies are seen to exist in the earlier full circle

derivations: the number of data points should be a multiple of four, not just an even number, and some arrangements of non-uniform samples are permissable so that some types of symmetrically interrupted profiles could be handled.

The relative ease with which equation 5.9 can be implemented tends to invalidate the development of specific schemes for a method optimised to a given situation. Even then the approach should make use of equation 5.9 rather than use continuous analysis since this automatically removes difficulties concerned, for instance, with the adequate representation of definite integrals from sampled data (see also section 5.4).

5.3 Least Squares Limacons as Reference Figures

The previous section has developed the least squares limacon from the desired circular reference in instrument co-ordinates. If it is to be practically useful two conditions must be satisfied: the estimation errors of the least squares limacon parameters must be adequately small and those parameters must then be relatable to the circle. The real data will exist in chart co-ordinates and so the first step must be to transfer a measurement there into instrument co-ordinates.

Since angles subtended at the origin are preserved under radius suppression, the $\underline{H}$ matrix is unaffected by that operation. Also the radial measurements are simply modified. The least squares estimate on the chart will be, by analogy with equation 5.7:

$$\begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c}_L \end{pmatrix} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T \begin{pmatrix} r_1 \\ \vdots \\ r_N \end{pmatrix}$$

Expressing this in terms of the instrument co-ordinates data gives:

$$\hat{\underline{a}} = M(\underline{H}^T\underline{H})^{-1}\underline{H}^T\begin{pmatrix} R_1-L \\ \vdots \\ R_N-L \end{pmatrix} \qquad 5.11$$

$$= M[\hat{\underline{A}} - (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{L}]$$

where $\underline{L}$ is an N-vector with each element having magnitude L. Now:

$$\underline{H}^T\underline{L} = L\begin{pmatrix} \Sigma\cos\theta_i \\ \Sigma\sin\theta_i \\ N \end{pmatrix}$$

which apart from the scalar multiplier is identical with the final row or column of $\underline{H}^T\underline{H}$. Thus the final reduction of equation 5.11 can be asserted without formal inversion from the observation that since $\underline{H}^T\underline{H}$ cannot be singular,

$$(\underline{H}^T\underline{H})^{-1}\,(\underline{H}^T\underline{H}) = \underline{I}_3$$

to give

$$\hat{\underline{a}} = M\left[\hat{\underline{A}} - \begin{pmatrix} 0 \\ 0 \\ L \end{pmatrix}\right]$$

This shows that for any possible measurement scheme applied consistently in instrument and chart co-ordinates the least squares limacon has, subject to magnification, the same centre and a radius term differing simply by the radius suppression.

The difficulties encountered when attempting to judge estimation errors have already been briefly introduced, so it is to be expected that only a few generalisations can be discussed here. The parameter estimates are only independent of each other when $\underline{H}^T\underline{H}$ is diagonal, that is under the four-fold symmetry conditions. Under

other conditions, which would be expected generally to become more difficult as the asymmetry increases, the interdependence may lead to large errors occuring in more than one parameter such that the whole set appears reasonable. This type of error can be induced by the use of imprecise computational techniques. In the specific case of the limacon, the only error of this type which can sensibly occur is for an inaccurate radius to be offset by an error in the centre position so that the reference figure passes reasonably through the data. Geometrical constraints limit the likelihood of such an occurrence to cases where all the data lies in one reasonably small arc. (See, also, section 5.4 for illustration of this.) As the matrix $\underline{H}^T\underline{H}$ will tend towards a singularity as the arc containing the data tends to zero, increased random estimation error is likely at small arcs. Overall it seems that the use of least squares limacons may break down on small arcs, but "small" remains an undefined term. Certainly increased care should be taken over all highly asymmetrical measurements.

The other major assumption of error analysis in least squares concerns the nature of the residuals, that they be random and independent of the measurement scheme. On a full roundness profile the residuals show by definition a periodicity with wavelength equal to the component circumference. Generally roundness errors are dominated by a small number of shapes each having a few undulations per revolution. Even the assumption of a pseudo-random sequence of residuals is therefore rather unlikely. Furthermore, since the limacon was derived from the eccentric circle by truncating a series, the residuals of a perfect circle will be periodic and could adversely affect the overall accuracy at high eccentricity, notably when the

eccentricity is significantly greater than the total out-of-roundness. Attempts to quantify the effects just discussed are so problem-specific that they will not be investigated here. To bring the discussion into context, it should be stressed that, for all these potential errors, no reports of serious difficulties being discovered with the least squares limacon have been found in the literature. In effect, it has been shown experimentally that the method is stable under practical conditions. There seems need, however, for a systematic investigation to be carried out before much weight can be given to such statements. It is not clear, for instance, how near to any limit of stability is the current practice.

The relationship between the parameters of least squares limacons and circles is only slightly more amenable to analysis. The question is not, in any case, totally clear cut: it asks whether the limacon gives a good estimation to a local minimum of the circle fitting problem. The establishment of error bounds for the correspondence depends on assumptions about the data which are not justifiable in a general discussion. Choosing "worst case" data would give bounds so widely separated as to be of no practical use. However, here, two observations will be made which may give some "feel" to the situation.

In deriving the limacon as an approximation to an eccentric circle by means of the binomial expansion (see section 3.1) the terms truncated represent a power series in $\sin^2(\theta-\phi)$, that is, one containing only even powers. Any even power of $\sin(\theta-\phi)$ may be reduced to a sum of harmonics which will contain only even terms and a constant. Thus the only odd harmonic in the expansion of the eccentric circle is that retained in the limacon. The estimate of the circle centre obtained by using the limacon should be exact. The estimate of

radius however is low since all the truncated terms contribute: the amplitude of the $\sin^2(\theta-\phi)$ term dominates the error so the circle. from equation 3.2, can be expressed:

$$R \simeq E\cos(\theta-\phi) + R_o\left[1-\frac{E^2}{2R_o^2}\left(\frac{1-\cos 2\theta}{2}\right)\right]$$

With a suitable measuring scheme, the application of equation 5.10 to this form will yield:

$$\hat{R} = R_o\left(1 - \frac{E^2}{4R_o^2}\right)$$

for full circle measurements. The presence of form errors on the circle will introduce odd harmonics into the expansion and so the exact correspondence of limacon and circle centres will no longer apply.

A second approach would be to consider the relationship between ϵ_i and δ_i in equations 5.2 and 5.3. Using polar notation for the centre:

$$\epsilon_i^2 = T_i^2 - 2R_oT_i + R_o^2$$

where:

$$T_i^2 = R_i^2 - 2ER_i\cos(\theta_i-\phi) + E^2$$

Substituting for R_i in terms of δ_i gives:

$$T_i^2 = \delta_i^2 + R_L^2 + 2\delta_iR_L + E^2\sin^2(\theta_i-\phi)$$

Now since the limacon parameters are determined according to least squares, $\sum_i\delta_i^2$ is, by definition, minimum and $\sum_i\delta_i$ is zero. In principle these facts can be used to express the sum of squares of residuals for the circle in terms of the residuals of the limacon, for instance:

$$\Sigma\varepsilon_i^2 = \Sigma(\delta_i + R_L)^2 + E^2\Sigma\sin^2(\theta_i-\phi) + nR_o^2$$
$$-2R_o\Sigma[(\delta_i + R_L)^2 + E^2\sin^2(\theta_i-\phi)]^{\frac{1}{2}}$$

In practice, the presence of the square root within a summation prevents real progress. For example, summations involving $\delta_i\sin^2(\theta_i-\phi)$ are formed which can only be solved for specific examples. Generally the limacon solution will not represent a minimum of $\Sigma\varepsilon_i^2$, but providing that E and $\underset{i}{\text{Max}}(\delta_i)$ are small compared to R_L and R_o then $\Sigma\varepsilon_i$ is likely to be close to a minimum. Again the terms "small" and "close" are not quantified.

Perhaps discussions such as those above are not of primary importance. Providing that conditions in instrument co-ordinates are such that a circle is adequately represented by a limacon, then if a stable, averaged reference is wanted, the convenience of linear least squares ensures the use of measurements relative to limacon references.

5.4 Measurement of Small Arcs

Although, generally, the description of experimental investigations is covered in chapters 9 and 10, one particular study carried out has more relevance here. This is to re-investigate the results quoted by Siddall (5), concerning the accuracy of parameter estimation on arcs. By using a glass reference hemisphere placed eccentrically on a roundness instrument and measured at relatively low magnification, a close approximation to the measurement of a perfect circle subject only to effects such as instrument noise could be made. After measurement, a series of calculations were

performed on the profile each terminating after a shorter arc and all starting at the same point. Using analysis based on the formulae quoted by Whitehouse (4) Siddall reported that the least squares radius estimate became rapidly unstable for arcs of less than about 60°.

Using the system described in chapter 9, these tests can be easily repeated. Since the profile is stored in a disc file there is added convenience in that tests, using identical data, of variations due to different algorithms can be investigated at leisure. The earlier work was restricted to comparing results taken from a single run of the program since the data was held in the main memory of the computer. Measurment conditions were similar to those quoted by Siddall. A hemisphere, with roundness better than 0.025μmMZC, was set up at near the maximum permissable eccentricity with the instrument magnification of 500. Profiles were data logged using 512 points to a full revolution with the direction of eccentricity being approximately along the +x, -x and -y directions, z being the spindle axis and -y the front of the worktable, in three measurements. Each of these was then processed over various arc lengths, always starting from $\theta = 0$ (+x axis) by two least squares limacon programs. The first (program DENT) used Whitehouse's formulae in an implementation believed to be identical with that used by Siddall. The second (program DLS2) used a direct implementation of the sampled least squares, setting up the matrix and solving equation 5.6. A Cholesky method was used for matrix operations (25).

The least squares limacon parameter estimates are shown in Table 5.1 for calculations according to the "traditional" formulae and in Table 5.2 for calculations from the direct implementation. In all

	Number of samples	Arc length o	$\hat{A}$ μm	$\hat{B}$ μm	$\hat{R}_L$ μm
a)	512	360	37.3	-4.1	1.8
	384	270	37.3	-4.1	1.8
	256	180	37.3	-4.2	1.8
	128	90	37.8	-3.8	1.3
	64	45	38.2	-3.6	.9
	32	22.5	19.3	-6.9	20.0
	16	11.25	-231	-29.0	271
	8	5.6	-944	-47.0	984
b)	512	360	-36.8	4.1	-1.3
	384	270	-36.8	4.1	-1.3
	256	180	-36.8	4.1	-1.3
	128	90	-36.6	4.3	-1.5
	64	45	-34.5	5.1	-3.7
	32	22.5	-15.3	8.4	-23.0
	16	11.25	198	26.5	-236
c)	512	360	-3.5	-37.4	2.4
	384	270	-3.5	-37.4	2.4
	256	180	-3.5	-37.4	2.4
	128	90	-3.4	-37.3	2.3
	64	45	-3.1	-37.2	1.9
	32	22.5	1.1	-36.6	-2.3
	16	11.25	47.6	-33.3	-48.9

TABLE 5.1: Least Squares Limacon Estimates from various arc lengths of eccentric hemisphere in three orientations using Whitehouse's formulae (program DENT).

	Number of samples	Arc length o	$\hat{A}$ μm	$\hat{B}$ μm	$\hat{R}_L$ μm
a)	512	360	37.3	-4.1	1.8
	384	270	37.3	-4.1	1.8
	256	180	37.3	-4.2	1.8
	128	90	37.8	-3.8	1.3
	64	45	38.8	-3.3	0.2
	32	22.5	34.3	-4.0	4.8
	16	11.25	34.1	-4.5	5.0
	8	5.6	45.9	-4.9	-6.8
b)	512	360	-36.8	4.1	-1.3
	384	270	-36.8	4.1	-1.3
	256	180	-36.8	4.1	-1.3
	128	90	-36.6	4.3	-1.5
	64	45	-35.2	4.8	-3.0
	32	22.5	-30.4	5.5	-7.8
	16	11.25	-62.2	2.6	24.1
c)	512	360	-3.5	-37.4	2.4
	384	270	-3.5	-37.4	2.4
	256	180	-3.5	-37.4	2.4
	128	90	-3.4	-37.3	2.3
	64	45	-3.4	-37.3	2.3
	32	22.5	-3.1	-37.3	1.9
	16	11.25	-19.1	-38.9	18.0

TABLE 5.2: Least Squares Limacon Estimates from various arc lengths of eccentric hemisphere in three orientations using the general solution (program DLS2).

cases the peak to valley and r.m.s. out-of-roundness for the full circle was 0.3 to 0.4μm and about 0.07μm respectively. These figures may be compared with digistation resolution of just under 0.1μm to judge the quality of the experimental approximation to the measurement of a perfect circle. From Table 5.1 it is seen that the general behaviour of the estimates is as given by Siddall but here the fall off of values is less rapid than he noted (for example, errors in the radial estimate of about 1μm at 90°, 10μm at 45° and 1000μm at 15°). The reasons for this difference are not known. It is possible that there are subtle differences in the algorithmic implementation of the two experiments but it is perhaps more likely that the earlier results had rather more "noise" on the profile. That there can be significant effects caused by algorithm is shown by comparison of Tables 5.1 and 5.2. In the latter the same pattern of behaviour is discovered but the size of the effect is much reduced by using the direct implementation. The difference is highlighted by Figure 5.2 which plots R_L against arc length for the case of the -x centring error. Also included is a sketch of Siddall's result (although it is not known what the centre orientation was used in his test). Note that a non-linear vertical scale is used on the graph.

The nature of the error in the estimates is that of a poor x-direction estimate of the centre being compensated by an opposite error in the radial estimate, as can be seen from the near constancy of $A+R_L$ in each set of measurements. The error in B is probably caused by "breakthrough" of the main error due to the non-independence of the estimates.

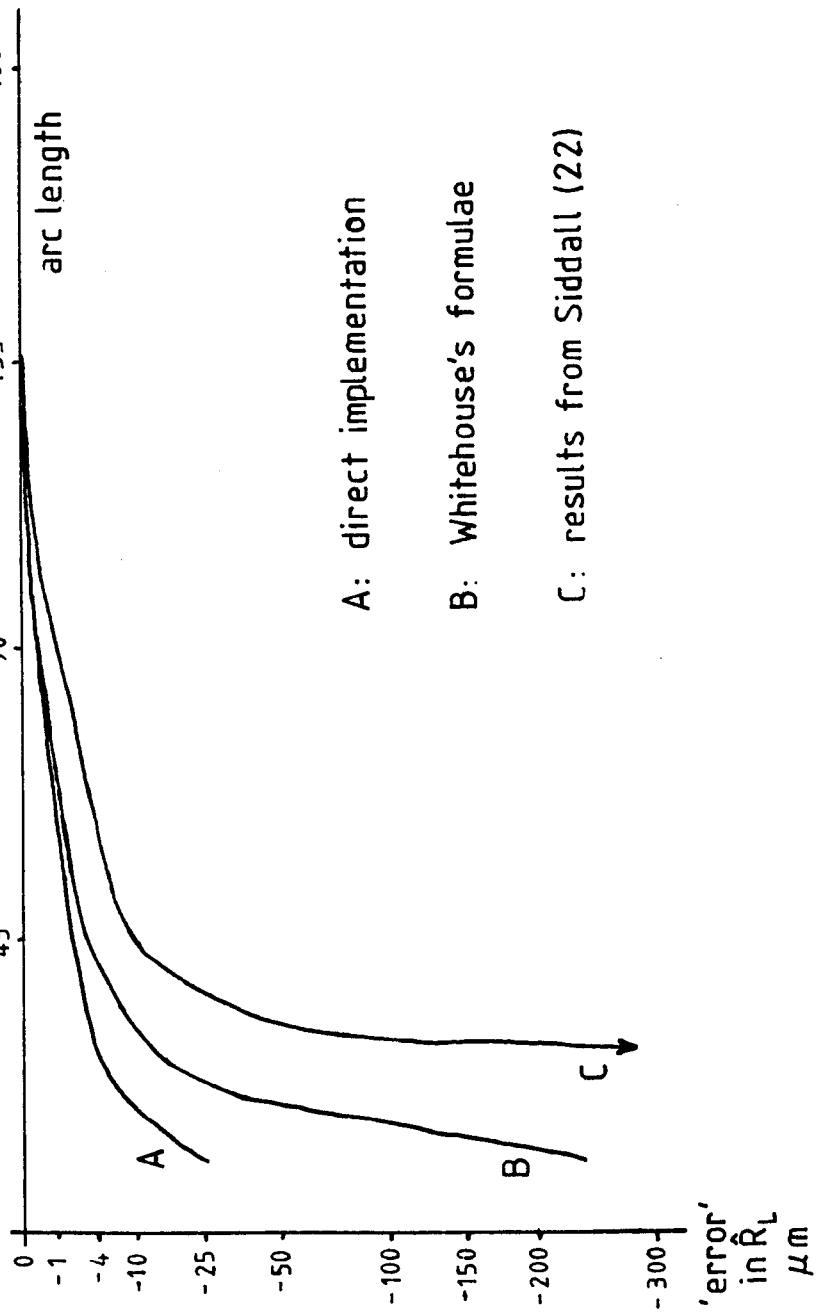


FIGURE 5.2: Comparison of "Error" in $\hat{R}_L$ from Various Arc Lengths

These results give renewed confidence to the use of least squares limacon methods on fairly small arcs. Although the errors depend upon the orientation, the direct implementation gives in all cases good results on arcs of less than 45° and useful results seem obtainable to below 30°. For the purposes of interactively recentring the workpieces an initial arc of just a few degrees would usually be adequate.

6 The Boundary References

6.1 Earlier Work

The grouping together of minimum zone, minimum circumscribing and maximum inscribing references which has been adopted here has been recognised since the earliest days of roundness measurement. In general, however, this grouping has been of a rather negative nature rather than the present positive assertion that their similarity is in their mathematical formulation. Thus national standards have tended to isolate the least squares approach by way of its uniqueness and computability, leaving the boundary references as possibly non-unique and solvable only by graphical 'trial and error' methods. Generally Standards subdivide the group, specifying ring and plug references as being non-preferred and advocating the minimum zone method. The American Standard (1) gives minimum zone as the preferred measurement with least squares as an alternative. In British Standards (2) no direct choice between minimum zone and least squares is made although it is possible to read into it an implied preference for least squares. As discussed in previous chapters, the Standards refer to the fitting of circles to chart co-ordinate data. BS 3730 gives no indication about the solution technique other than to state that repeated trials must be used and that first finding the ring and plug circle centres may help the discovery of the minimum zone circle centre. ANSI B89 gives a little more guidance in the form of statements about the geometry of the contacts between the profile and the reference circle(s) which will exist when the solution is

found. It gives no justification for those statements but they are basically the same as the condtions which will be derived in this discussion and were presumably obtained geometrically.

There have been some attempts to develop algorithms for the solution of the boundary reference circles, usually starting from a geometrical or topological description of the problem. The first attempt seems that of Avdulov (8) who posed the geometrical problem in terms of a mathematical description but derived methods which, certainly at that time, required far too much computational effort to be practical. Some years later Whitehouse (9), in effect, re-derived the scheme starting from the graphical concept of the problem. Since it is the more easily visualised, Whitehouse's approach to the ring gauge circle will be briefly outlined to illustrate the method. The definition of a circle in a plane has three degrees of freedom (the x- and y-positions of the centre and the radius.). Hence it needs three fixed points to be completely defined and so three contacts between the ring gauge and the profile will be required. The procedure for achieving these is shown schematically in figure 6.1. An initial guess is made for the position of the centre, O_1, and the point, P_1, on the profile most distant from O_1 found. A circle centred on O_1 passing through P_1 will circumscribe the profile but will not be the minimum (unless there happen to be several points equidistant from O_1). The circle may therefore be reduced by moving its centre along O_1P_1 while maintaining its contact with P_1 until at centre O_2 a second contact P_2 is found. Now the process continues by maintaining both P_1 and P_2 on the reducing circle and so the centre must lie on the bisector of the angle subtended by P_1 and P_2. When the third contact P_3 is found the ring

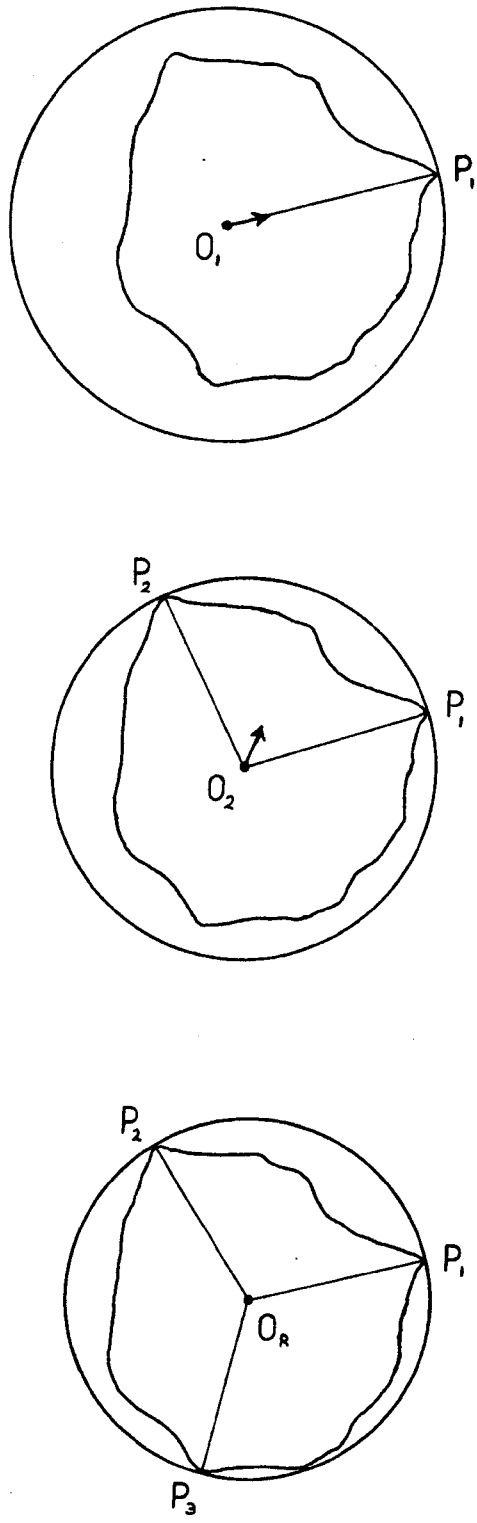


FIGURE 6.1: A Search Procedure for Finding the Ring Gauge Circle (28)

gauge centre O_R is established. The procedure for plug gauge circle is the opposite of this. First start with the point nearest the initial trial centre and then expand the circle by moving in directions opposite to those shown in figure 6.1. An attempt to achieve the minimum zone circles notes first that since there are now four degrees of freedom (the centre and two radii or one radius and a sone width) there will be four contacts and since the result will generally not involve either the plug or ring gauge as one of the zone circles these contacts should lie two on each of the zone circles rather than three and one. Armed with this information the minimum zone algorithm uses alternate moves from the ring and plug algorithms.

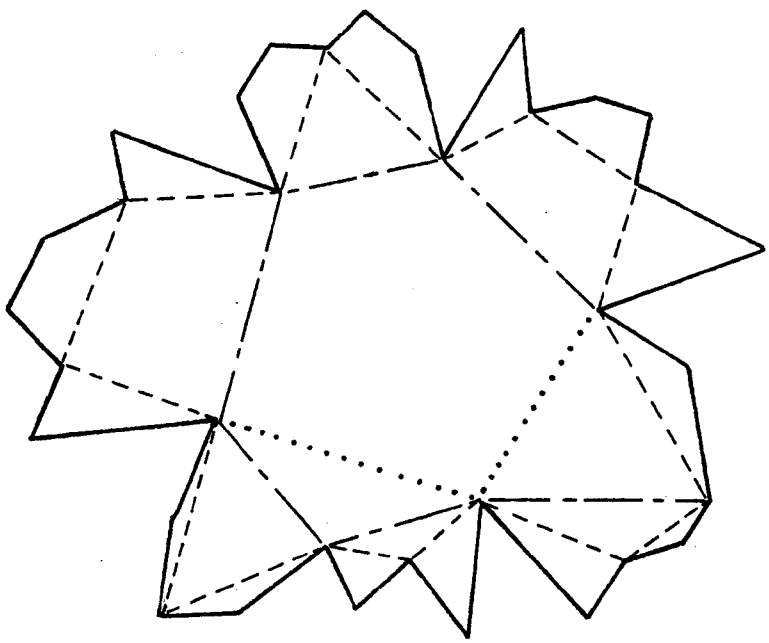
In the form described here these algorithms are not very useful. For instance they must be able to cope with such features as a two point diametral contact of the profile with its ring circle. Furthermore, they need their end points defined more completely as was given in ANSI B89. If the centre of the circumscribing circle to a triangle (which is of course unique) does not lie within that triangle, the latter contains an obtuse angle and so can be contained within a smaller circle by using its longest side as a diameter. So the existence of a three point contact geometry does not alone determine the ring gauge circle. The additional geometric condition that the points form a triangle enclosing the centre is needed. Similarly the four contacts of the minimum zone must lie alternately with angle from the centre on each zone circle.

These methods are very useful guides for finding reference circles graphically and can be quite efficient when coupled with the intuitive ability of the human operator. For automatic assessment

by computer they are not nearly so good since careful searches along the prescribed directions of centre movement are needed and this involves much calculation. Commercial systems using this approach have been produced (Rank Taylor Hobson Ltd. - 'Talydata') but they are based upon specialised fast processing units and tend to cut-off after a fixed number of iterations in order to keep the computation time to a reasonable limit. Their results are therefore likely to be close to the solution but may well not be precisely that solution. Also they do not calculate with circular references but with the algebraically simpler limacons. This is justified in terms of the modelling of distorted circular references on the chart, but the direct application to limacons of geometric algorithms developed for circles under such conditions has not been proven.

An alternative approach to reducing the amount of computation is to perform some form of data reduction exercise, a plausible approach since clearly large parts of a profile are not going to be invovled in the definition of a particular boundary reference (8). In the course of some other work, Scheiding (10) used a data reduction scheme which could also, in a sense, reduce the potential for non-uniqueness of the plug circle. This consisted of reducing the profile to an irregular, but convex, polygon by successively connecting the peaks or valleys of the profile by straight lines, see figure 6.2. Only the points remaining in contact with the polygon are used in the circle fit.

Other attempts at solving in the main minimum zone problems have been to use 'Monte-Carlo' techniques usually searching an area around the least squares centre (12). Such methods would require a



Profile

1st attempt

2nd attempt

Final figure

FIGURE 6.2: Scheiding's Method of Data Reduction Demonstrated on Plug Gauge (21)

lot of work if enough tests to provide resonable confidence in the result were to be performed. In any case, using random checks is really a last resort! The solution of a perturbed least squares problem has also been suggested (11). It is usual with all these types of approach to use limacons rather than circles for the computation - in particular when least suqares is used there is little choice about this.

6.2 Boundary Limacons as Problems in Linear Programming

Unlike the work just described, the approach here will be that of a deliberate choice to use limacons as reference figures, for the reasons and under the conditions discussed in earlier chapters. As was pointed out in that discussion, the linearity of limacons in their parameters means that the fitting of boundary limacons to a set of data is a problem in linear programming. Linear programming is a standard technique in the field of Operations Research and has a well-developed theory (see, for example, (26)). In this discussion a knowledge of the major principles of this theory and the solution techniques will be assumed and, for example, the standard terms will be used here without definition. However a very brief outline of the theory, defining these terms, is given in Appendix 2 for easy reference.

As all three references will be handled in the same manner, the solution will be discussed in terms of the solution of the minimum radius circumscribing limacon. In order to stress the nature of the relationship between the algebraic method of linear programming and the geometry of the physical fitting problem, reference will be

made to the simple example introduced in section 4.2 and table 4.1. The formal statement of the ring limacon, in terms of chart co-ordinate parameters, is represented in matrix form as:

$$\text{Subject to:} \quad \begin{pmatrix} \cos\theta_1 & \sin\theta_1 & 1 \\ \vdots & \vdots & \vdots \\ \cos\theta_N & \sin\theta_N & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ r_L \end{pmatrix} \geq \begin{pmatrix} r_1 \\ \vdots \\ r_N \end{pmatrix} \qquad 6.1$$

$$\text{Minimise:} \quad z = (0 \;\; 0 \;\; 1) \begin{pmatrix} a \\ b \\ r_L \end{pmatrix}$$

or, more compactly,

$$\text{Subject to:} \quad \underline{K}\,\underline{a} \geq \underline{r} \qquad 6.2$$

$$\text{Minimise:} \quad z = \underline{C}^T \underline{a}$$

The formulation of the plug limacon will be similar and that of the minimum zone expressible exactly in the form of equation 6.2 if the symbols are defined to included the extra parameter and constraints. In expanded form the minimum zone gives:

$$\text{Subject to:} \quad \begin{pmatrix} \cos\theta_1 & \sin\theta_1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \cos\theta_N & \sin\theta_N & 1 & 1 \\ -\cos\theta_1 & -\sin\theta_1 & -1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ -\cos\theta_N & -\sin\theta_N & -1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ r_L \\ h \end{pmatrix} \geq \begin{pmatrix} r_1 \\ \vdots \\ r_N \\ -r_1 \\ \vdots \\ -r_N \end{pmatrix}$$

$$\text{Minimise:} \quad z = (0 \;\; 0 \;\; 0 \;\; 1) \begin{pmatrix} a \\ b \\ r_L \\ h \end{pmatrix}$$

Returning now to the ring limacon, equation 6.2 represents the tableau which may be used for the simplex method of solution.

However before such a solution can be entertained, some modification is necessary because in these problems the non-negativity of the parameters cannot be guaranteed. That a and b may be negative in the final solution is self evident but, further, in chart co-ordinates (and so the reason for their use here) the radius r_L can also become negative. This is because the instrument is not constrained to ensure that the setting of the suppressed radius is less than the true radius of the workpiece: R_L in instrument co-ordinates would always be positive but $r_L = R_L - L$ in chart co-ordinates could be negative. Since non-negativity is a condition of the simplex method, the usual splitting of parameters into mutually exclusive positive and negative variables is needed:

$$a \rightarrow a^+ - a^-$$

where either $a^+ \geq 0$ and $a^- = 0$ or $a^+ = 0$ and $a^- \geq 0$. The simplex solution for the ring limacon is thus over six, not three, parameters and having N constraints will require also N slack variables. Bearing these comments in mind the initial tableau for the simplified example (having no b) is shown in figure 6.3. To comply with the usual way in which the rules for manipulating the tableau are given this is shown as the negative of equation 6.2, namely

$$\text{Maximise:} \quad z = -r_L$$

$$\text{Subject to:} \quad -\underline{K}.\underline{a} \leq -\underline{r}$$

The tableau is infeasible: since all the right hand sides are negative $r_L = a = 0$ is not a feasible solution (for this example this is clear from the parameter space representation, figure 4.5, from which a graphical solution could be obtained). In general, then, artificial variables would have to be added (and then iterated out

r_L^+	r_L^-	a^+	a^-	s_1	s_2	s_3	s_4	b'
-1.0	1.0	-0.866	0.866	1	0	0	0	-1.0
-1.0	1.0	-0.7071	0.7071	0	1	0	0	-0.5
-1.0	1.0	0	0	0	0	1	0	-0.5
-1.0	1.0	0.7071	-0.7071	0	0	0	1	-0.6
-1.0	1.0	0	0	0	0	0	0	0
1.0	-1.0	0.866	-0.866	-1.0	0	0	0	1.0
0	0	0.1589	-0.1589	-1.0	1	0	0	0.5
0	0	0.866	-0.866	-1.0	0	1	0	0.5
0	0	1.5731	-1.5731	-1.0	0	0	1	0.4
0	0	0.866	-0.866	-1.0	0	0	0	1.0
1.0	-1.0	0	0	-0.4494	0	0	-0.5505	0.7798
0	0	0	0	-0.899	1	0	-0.1010	0.4596
0	0	0	0	-0.4494	0	1	-0.5505	0.2798
0	0	1.0	-1.0	-0.6357	0	0	0.6357	0.2543
0	0	0	0	-0.4494	0	0	-0.5505	0.7798

FIGURE 6.3: Simplex Tableau Solution for Circumscribing Limacon

of the solution) to create an initial basic feasible solution. However, here it can be seen that setting r_L to +1 in the first constraint and then using it to eliminate from the others produces a basic feasible solution as shown in the second stage of the tableau. (This corresponds to point G in figure 4.5: the axis is acting as a constraint). The only non-basic variable with a positive co-efficient in the objective is now a^+ and so this is brought into the basis, in constraint 4 since this gives the smallest ratio of right hand side to column co-efficients. On completing this third step, the tableau is optimum and its coefficients can be interpreted in terms of the ring limacon. a^- and r_L^- are ignored. In the first constraint r_L is basic and has value 0.7798 and in the fourth constraint a is basic with value 0.2543. In the other constraints the slack variables are basic and lie inside the limacon by 0.4596 and 0.2798 and the out-of-roundness is the larger of these. The slack variables of the first and fourth constraints are non-basic and so have zero value, indicating that the corresponding data points, at 30^o and 135^o, lie on the limacon.

Even on this simple example, the work needed to iterate the tableau is not negligible. In practice there might commonly be 512 data points and so 512 slack variables are needed. The total size of the tableau could then be 519 by 513 elements and an enormous amount of calculation would be needed. The minimum zone having more parameters and more constraints would have a tableau 1032x1025; over one million elements would need updating at each iteration! Direct application of simplex to the boundary limacon fits is not really practicable therefore.

Since each constraint generates a slack variable while the

variables generate no extra terms in the tableau, the solution of linear programming problems having relatively few constraints even with many variables is more efficient than those with few variables and many constraints. The boundary limacons fall into the latter group and so an increase in efficiency can be obtained by solving the dual linear program rather than the primal so far considered. An additional simplification occurs with the current problem since a pair of variables associated with a sign-unrestricted variables in the primal transform to a single equality constraint in the dual and vice versa. Thus a ring limacon having, say, an array of variables of 6x512 and needing 512 slack variables in the primal has in the dual an array of 512x3 and needs only 3 surplus variables. The total dual tableau would have 516x4 elements and so require less than one hundredth of the work needed on the primal for an iteration.

The dual solution for the simple example, namely:

$$\begin{aligned} &\text{Minimise:} \quad z = -\underline{r}^T\underline{y} \\ &\text{Subject to:} \quad -\underline{K}^T\underline{y} \geq -\underline{c} \end{aligned} \qquad 6.5$$

where $\underline{y}$ are the variables of the dual to which no direct physical interpretation is attached. This statement may be compared to that of equation 6.4. The tableau solution is shown in figure 6.4. Noting that all dual constraints are here equalities, the negative signs demanded in the constraints by equation 6.5 have been cancelled. Also because of the equality constraints the origin is not a feasible solution and artificial variables (t_1 and t_2) have been added. These are driven from the solution by the Wagner method, a variation of the "two phase method" (27). This uses a secondary objective function, w, in which each co-efficient is the negative sum of the corresponding

	y_1	y_2	y_3	y_4	t_1	t_2	b'
	1.0	1.0	1.0	1.0	1	0	1.0
	0.866	0.7071	0	-0.7071	0	1	0
	-1.0	-0.5	-0.5	-0.6	0	0	0
w	-1.866	-1.7071	-1.0	-0.2929			
	0	0.1835	1.0	1.8165	1	-1.1547	1
	1	0.8165	0	-0.8165	0	1.1547	0
	0	0.3165	-0.5	-1.4165	0	1.1547	0
w	0	-0.1835	-1.0	-1.8164	0	2.1547	
	0	0.1010	0.5505	1	0.5505	-0.6357	0.5505
	1	0.899	0.4495	0	0.4495	0.6357	0.4495
	0	0.4596	0.2798	0	0.7798	0.2542	0.7798
w	0	0	0	0	1.0	1.0	

Note: $w = -\sum t_i$ is iterated to drive out artificial veriables

FIGURE 6.4: Simplex Tableau of Dual Programme for Circumscribing Limacon

co-efficients in the constraints for which an artificial variable is introduced. In this example the iterating out of the artificial variables happens to coincide with the optimum result for the main objective function. In terms of physical interpretation, the correspondence of results between the optimal tableaux in figures 6.3 and 6.4 can be compared. In particular the co-efficients of the dual objective function give the most important information: those corresponding to artificial variables give the parameter values and those corresponding to $\underline{y}$ give the negative of the primal slack variable values and so show the contact points and 'out-of-roundness'.

Even with such a small problem the computationl advantage of the dual is demonstrated by these two tableaux. Although the artificial variables introduced a secondary objective function, the total number of element operations on the tableau in the dual is less than two thirds of those in the primal.

When there are many variables and relatively few constraints, considerable computational savings can be made by using the techniques of the revised simplex method. It is then only necessary to update a square matrix of size governed by the number of constraints plus the active rows and columns of the tableau at each iteration. Thus revised simplex on the dual is a good technique for solving boundary limacons. The minimum zone limacon fit would have about 1050 element calculations per iteration, a figure to be compared with that for the primal quoted earlier. The majority of practical measurements can be arranged to have simply (usually uniformly) spaced data points. The main body of the tableau depends only upon the position of those points and, under these conditions, elements of it could be constructed as required rather than the whole tableau stored. Using the revised

simplex on the dual the amount of necessary reconstruction at each iteration is small enough to be practical and so considerable savings in the necessary computer memory can be made. In fact the method allows the solution to come easily within the computational power of the types of mini-computer which can reasonably be used in ordinary instrumentation systems.

There are techniques which are formally more efficient than revised simplex but as they tend to require more complicated software, their overall advantage to the present problems may not be clear. This question will not be pursued here. It is sufficient to have demonstrated that linear programming on limacons offers a practically useful method of solving boundary reference fits. All methods will require considerable computations and are certainly unsuitable for use 'by hand'. The approach here will be to examine the implications of these methods to the original geometry with two potential benefits: specialised high efficiency algorithms and a method at least intelligible as a graphical procedure.

6.3 Implications of Dual Feasibility

All the simplex methods discussed above seek to optimise an objective function while maintaining a feasible solution by ensuring the non-negativity of the 'right-hand-side' of the constraints in the tableau. Because of the reversal of roles between primal and dual, any state which is feasible in the dual corresponds to an optimal condition in the primal but since the dual is not optimal, the primal is infeasible. A first step is to identify the geometry of these various conditions. In figure 6.5 are shown the four condi-

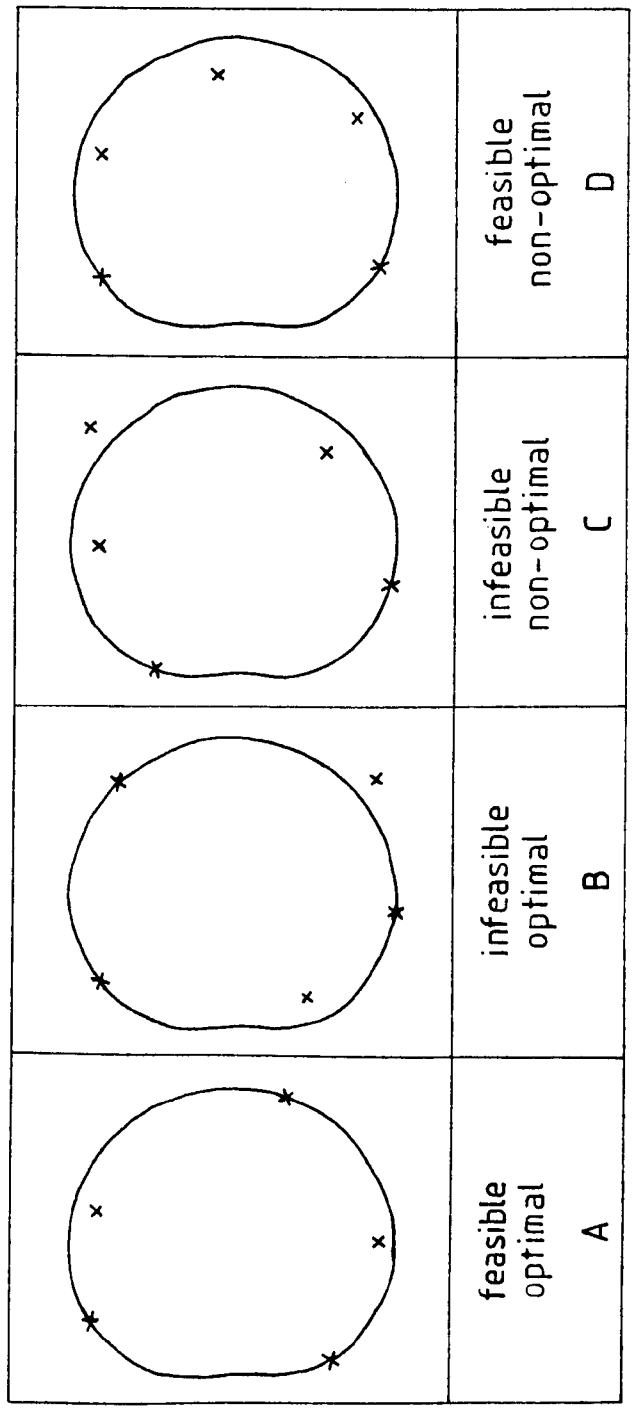


FIGURE 6.5: Possible Conditions Occurring during the Search for a Minimum Circumscribing Limacon to a Set of Five Data Points

tions that can occur (the labels there refer to the primal) when seeking the smallest limacon to enclose a set of data points. The solution of course is the feasible, optimal condition and the same in primal and dual representations. Simplex solution to the primal maintains feasibility and so proceeds through stages all of type D. Simplex solution to the dual maintains, in primal terms, optimality and so proceeds always through type B stages. The geometric interpretation of an optimal, infeasible condition is a limacon which does not enclose all the points but which is the smallest one which can enclose those within it. The difference between the solution of the primal or the dual is thus very simply stated geometrically, the primal starts with a figure too large but all-encompassing and shrinks it as far as possible while the dual starts with a figure too small and expands it as little as possible. The earlier workers all, therefore, effectively attempted primal solutions: the previous section indicates the amount of work this approach could entail.

The dual solution for ring limacon involves three equality constraints which must be exactly satisfied at each iteration. Thus the solution will always involve three contact points between the data and the reference figure. No two point contact such as could occur with the ring circle is possible.

The maintenance of dual feasibility requires that the 'right-hand-side' of the tableau remains non-negative. However, at any iteration the theory underlying the method of revised simplex shows that the current 'right-hand-side' can be found from its initial form and the inverse of the current basis:

$$\underline{b}' = \underline{\beta}\,\underline{b}$$

where $\underline{\beta}$ is the inverse of the basis, that is the inverse of the matrix consisting of the columns from the tableau of the basic variables. In the dual each column of $\underline{\beta}^{-1}$ will be a column of $\underline{K}^T$, $\underline{b} = \underline{c}$ and each basic variable corresponds to a zero slack variable in the primal (section 6.2), that is, to a contact between the indicated data point and the current trial limacon. Any condition on $\underline{b}'$ can be interpreted as a geometrical condition related to the contact points.

With all the boundary limacons, $\underline{c}$ is a simple form having all elements zero except the last which is unity. This property is quite common in practical situations being, of course, the request to optimise a particular parameter: another relevant example from surface metrology, the assessment of flatness, is discussed in Appendix 1. In such problems the non-negativity of $\underline{b}$ requires simply the non-negativity of the final column of $\underline{\beta}$ and so a quite simple dependence on the contact point geometry is to be expected. The nature of this geometry for the limacons will now be examined for each reference.

6.4 The Minimum Circumscribing Limacon

The dual of the ring limacon linear programme involves only equality constraints. Its solution will involve artificial variables but these cannot remain in the basis of a feasible solution. There are no slack/surplus variables and so all columns of the basis must be from the constraint matrix $\underline{K}$ at any general iteration. (The possibility of an insistence on a positive radius causing an inequality constraint in the dual is not worrying: the surplus variable

so produced could be in the basis only if the radius were zero!). Taking three general contact points at θ_i, θ_j and θ_k the basis would be:

$$\underline{\beta}^{-1} = \begin{pmatrix} \cos\theta_i & \cos\theta_j & \cos\theta_k \\ \sin\theta_i & \sin\theta_j & \sin\theta_k \\ 1 & 1 & 1 \end{pmatrix}$$

No significance (such as $\theta_i < \theta_j$, for example) can be read into this matrix; the relative positioning of columns depends upon the workings of revised simplex in previous iterations. Now the determinant of $\underline{\beta}^{-1}$ is given by the sum of the co-factors of its third row, that is by the same co-factors which identify the elements of the third column of $\underline{\beta}$. The non-negativity of the elements of the third column of $\underline{\beta}$ thus requires that these co-factors all have the same sign. So Δ_{ij}, Δ_{jk}, Δ_{ki} must agree in sign where:

$$\Delta_{ij} = \begin{vmatrix} \cos\theta_i & \cos\theta_j \\ \sin\theta_i & \sin\theta_j \end{vmatrix}$$

$$= \cos\theta_i \sin\theta_j - \sin\theta_i \cos\theta_j$$

$$= \sin(\theta_j - \theta_i)$$

and similarly for the others. The condition depends then only on differences of angles, relationships which are independent of the rotation of the axis system. By using such a rotation one angle, θ_i say, can be set to zero and the required condition then becomes the similarity of sign of $\sin(\theta_j)$, $\sin(\theta_k - \theta_j)$ and $\sin(-\theta_k)$. Both θ_j and θ_k must lie in the range 0 to 360° with respect to θ_i initially and so, now, according to whether the co-factors are positive or negative the conditions must be

either $0 < \theta_j < 180^\circ$ or $180^\circ < \theta_j < 360^\circ$

$180^\circ < \theta_k < 360^\circ$ $0^\circ < \theta_k < 180^\circ$

$0 < \theta_k - \theta_j < 180^\circ$ $0^\circ < |\theta_k - \theta_j| < 180^\circ$

So the requirement for dual feasibility is that the three angle differences $|\theta_j - \theta_i|$, $|\theta_k - \theta_j|$ and $|\theta_i - \theta_k|$ are all less than 180°.

Before proceeding, an alternative approach to deriving this condition will be introduced for completeness since it is useful in studying the nature of the feasibility condition. This relies on a direct geometric interpretation of the co-factors in the measurement plane and conveniently uses Cartesian co-ordinates. The co-factor can be expressed:

$$\Delta_{ij} = \frac{1}{r_i r_j} \begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix}$$

and related to this is a function

$$\Delta_{i.} = \frac{1}{r_i r} \begin{vmatrix} x_i & x \\ y_i & y \end{vmatrix} = 0$$

which (apart from an indeterminacy at the origin, of little importance here) is a straight line passing through (x_i, y_i) and $(0,0)$ which divides the xy plane into two areas, one where $\Delta_{i.} > 0$ and the other where $\Delta_{i.} < 0$. The line also represents the locus of all points having θ_i as their argument. Noting the order of indices, the feasibility condition requires that Δ_{ij} and Δ_{ik} have opposite sign and so lie on opposite sides of the line. An exactly similar argument applies to the other points and $\Delta_{j.} = 0$ or $\Delta_{k.} = 0$. Given the points and their associated lines, the third must lie as an opposite with respect to both. Thus the only positions it can

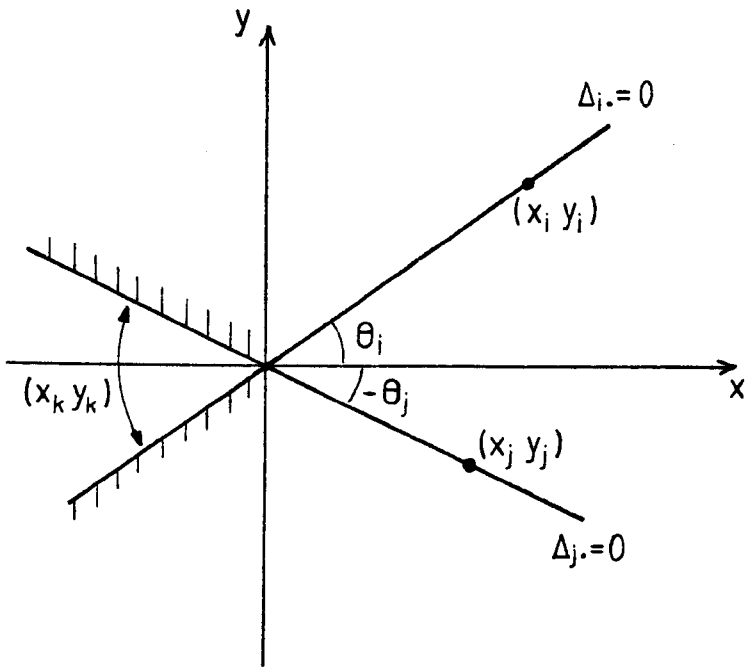


FIGURE 6.6: Geometric Form of Dual Feasibility Condition for Ring Limacon

occupy are in the sector defined by the two lines on the opposite side from the origin to their respective points, see figure 6.6.

However the result is derived, the following theorem is a direct consequence of dual feasibility.

The 180° Rule for Minimum Radius Circumscribing Limacon

A circumscribing limacon with a given origin to a set of points is the minimum radius circumscribing limacon to those points if it is in contact with three of them such that no two adjacent contact points subtend an angle at the origin of more than 180°, where the term "adjacent" implies that the angle to be measured is that of the sector not including the third contact point.

(It must be noted in passing that the statement that the minimum radius circumscribing limacon in chart co-ordinates transforms to the minimum radius circumscribing limacon in instrument co-ordinates relies on the facts that (a) circumscribing figures remain circumscribing figures, (b) three contact points in chart co-ordinates obey the 180° rule and hence, since angles are unchanged, their transformations into instrument co-ordinates also obey the rule.)

If the radial values of the three contact points are identical the fitted limacon degenerates to a circle centred at the origin. Thus the above rule has the following well-known corollary.

The 180° Rule for Minimum Radius Circumscribing Circle

The circumscribing circle to a set of data points is the minimum radius circumscribing circle if no two adjacent of the three contact points of the circle with the data subtend an angle of more than 180° at the centre of the circle. The implications in the word "adjacent" are as for the limacon case.

The complete simplex iteration for finding the minimum radius circumscribing limacon in the dual can be summarised as selecting any point which violates the reference (conventionally, the point giving the largest violation is chosen) and substituting it for one of the points defining the reference in such a way that dual feasibility is maintained. The 180° rule gives to the problem a geometric concept for maintaining dual feasibility which is simpler than the general simplex iteration. This complete algorithm is as follows.

Minimum Radius Circumscribing Limacon Algorithm

a) Choose any three data points such that no two adjacent ones subtend an angle at the origin of more than 180°.

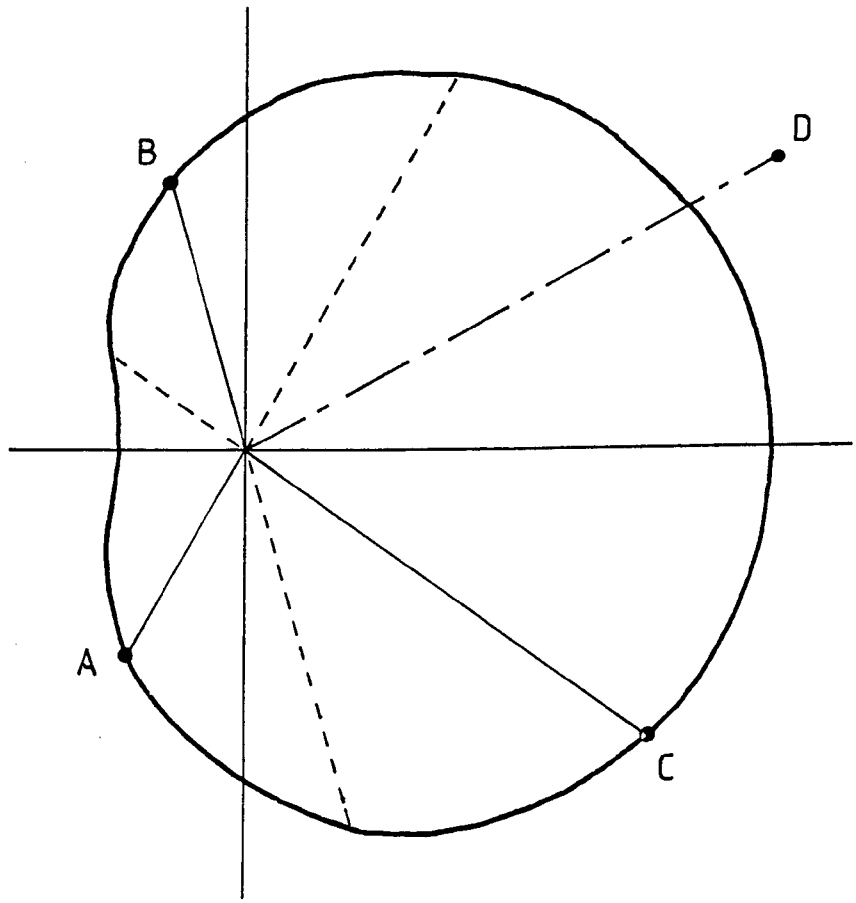
b) Construct a reference limacon through these three points.

c) If no data points lie outside this limacon the solution is found. Otherwise choose the point which violates the reference by the largest amount.

d) Replace one of the reference points by this new point such that the 180° rule is still obeyed and go back to b).

This procedure is an exchange algorithm and is illustrated by figure 6.7. The exchange between any new point and the contacts is always unique.

In general, the efficiency of an exchange algorithm depends upon the iterations moving monotonically towards an optimum solution in order to guarantee that cyclical exchanges do not occur. Here this is the case for as the exchange is unique at each iteration it is identical with that chosen by the simplex method on the dual linear programme, and that is known to converge monotonically.



A, B, C obey the 180° rule

A, B, D obey the 180° rule

Therefore D replaces C

FIGURE 6.7: Feasible Point Exchange for Circumscribing Limacon

6.5 The Maximum Inscribing Limacon

Formal study of the plug limacon could follow the same path as that carried out in the previous section. More directly it can be observed that the general form of the primal problem is:

$$\text{Maximise:} \quad z = \underline{c}^T \underline{a}$$

$$\text{Subject to:} \quad \underline{K}\,\underline{a} \leqslant \underline{r}$$

where the symbols are exactly as used in section 6.2. This form may be compared with equation 6.4, a statement of the ring limacon problem. They are identical apart from sign which could be absorbed into the vectors. Thus a complete method for the solution of the maximum radius inscribing limacon is:

a) Change the sign of all radial data points.

b) Apply the ring limacon algorithm to this modified data.

c) The plug limacon parameters are those found in b) with signs reversed.

This ability to use the same algorithm on two problems, which is computationally very attractive, is a further consequence of the parameter linearity of the limacon.

The inversion of the data reflects each point through the origin and does not alter relative angles subtended there. Thus there is 180° rule for inscribing limacons. If a limacon is inscribing to a set of points and contacts three of them such that no two adjacent contacts subtend an angle of more than 180° at the origin, it is the maximum radius inscribing limacon for that set of points.

6.6 The Minimum Zone Limacons

The linear program for minimum zone has been stated in equation 6.3, as a primal problem. The zone width must be a positive quantity and so unlike the other boundary limacons a sign restricted variable, h, is present. The dual thus has three equality constraints and one inequality. Since artificial variables cannot be present in a feasible solution, the basis of the dual must consist of either four columns from the original constraints or three such columns and the single slack variable. This latter form is always feasible and corresponds to a zero width zone being fitted to three points. It has no relevance to the problem at hand and will be ignored henceforth.

The set of primal constraints from which the basis is to be chosen has two equal subsets comprising those relating to 'inner' or 'outer' contacts of the zone. The basis may select four of these subject only to the provision that the same point from inner and outer sets cannot be used simultaneously (to do so would represent a physical impossibility for a non-zero zone width). The basis at a general iteration of the dual can be represented:

$$\underline{\beta}^{-1} = \begin{pmatrix} S_i\cos\theta_i & S_j\cos\theta_j & S_k\cos\theta_k & S_l\cos\theta_l \\ S_i\sin\theta_i & S_j\sin\theta_j & S_k\sin\theta_k & S_l\sin\theta_l \\ S_i & S_j & S_k & S_l \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

The variables S_i to S_l take only values +1 or -1 and indicate whether the contact comes from the outer or inner set respectively. As with the ring limacon, dual feasibility insists that the co-factors of the final row of $\underline{\beta}^{-1}$ must have the same sign. These co-factors

will be:

$$-S_jS_kS_l\,\Delta_{jkl};\quad S_iS_kS_l\,\Delta_{ikl};\quad -S_iS_jS_l\,\Delta_{ijl};\quad S_iS_jS_k\,\Delta_{ijk}$$

where:

$$\Delta_{ijk} = \begin{vmatrix} \cos\theta_i & \cos\theta_j & \cos\theta_k \\ \sin\theta_i & \sin\theta_j & \sin\theta_k \\ 1 & 1 & 1 \end{vmatrix}$$

etc.

Consider, now the function:

$$\Delta_{ij.} = \begin{vmatrix} \cos\theta_i & \cos\theta_j & \cos\theta \\ \sin\theta_i & \sin\theta_j & \sin\theta \\ 1 & 1 & 1 \end{vmatrix} = 0$$

which can be used to test the sign of Δ_{ijk} and Δ_{ijl}. In the measurement plane it is represented by two lines from the origin in directions θ_i and θ_j, that is passing through the i^{th} and j^{th} points respectively see figure 6.8. Unlike the equivalent picture for the ring limacon, the lines do not reflect through the origin: the immutability of the third row of $\Delta_{ij.}$ means that columns are not linearly related if, say, $\theta = -\theta_i$. By differentiating $\Delta_{ij.}$ with respect to θ it is shown that it has just two turning points (at $\theta = \pm(\theta_i + \theta_j)/2$) and that the slopes at $\theta=\theta_i$ and $\theta=\theta_k$ are in opposite senses. Thus there are no zeros of $\Delta_{ij.}$ except those shown in figure 6.8 and those lines represent the boundary between positive and negative values of $\Delta_{ij.}$ (The marked areas are correct for the figure but it is not necessary that the reflex angle sector represent the positive region).

Now, if the k^{th} and l^{th} points are the same type of contact,

$$S_k = S_l$$

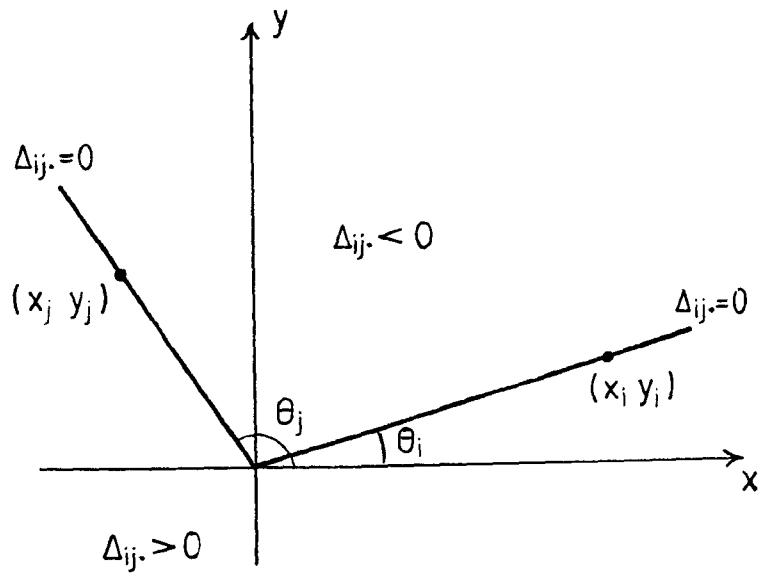


FIGURE 6.8: Geometric Form of Dual Feasibility Condition for Minimum Zone Limacons

and for the third and fourth column co-factors to have the same sign, Δ_{ijk} and Δ_{ijl} must differ in sign, whereas if they are of different types these two determinants will have the same sign. Thus if the k^{th} and l^{th} points are of different type they will be adjacent to each other in the measurement place, while if they are the same type they will lie alternately with the i^{th} and j^{th} points. The change of sign associated with interchanging columns of a determinant ensure that an exact parallel of this analysis can be carried out for any pair of the four points. Dual feasibility therefore demands that any pair of contacts of the same type lie in alternation with the other pair as angle increases and that any pair of contacts of the opposite type must lie adjacent to each other. This can be satisfied only if the four points are alternately inner and outer contacts with respect to the origin. This condition is purely a control on angular position with respect to the origin and so is unaffected by radius suppression. Hence minimum zone limacons in chart co-ordinates transform to minimum zone limacons in instrument co-ordinates.

Given four contact points between the data and the reference limacons, the substitution of a new point for one of them is always unique if the above alternate contact rule is obeyed both before and after the substitution. Thus a complete minimum zone algorithm may be based on this rule.

Minimum Zone Limacon Algorithm

Following the mathematically guaranteed convergence of the simplex solution of the dual linear programme, the minimum radial zone limacon can be found by:

a) Choose, arbitrarily, four data points.

b) Fit to these points a reference limacon such that each of the

four points lies equidistant from it and such that the points lie alternately to either side of the reference with increasing angles.

c) If no other points are further from the reference, the solution is found.

d) Otherwise substitute the point which lies furthest from the reference for one of the four defining points such that the new set of points lie alternately either side of the reference and return to b).

It may be noted that this alternating contact requirement is not unique to this problem. It also occurs, for instance, with the Stiefel Exchange Algorithm for finding the minimax polynomial to a set of data points. In fact the Stiefel Exchange Algorithm has been derived from a dual linear programming problem in a method parallel to that used here (28).

7. Boundary Limacons and Boundary Circles

7.1 Introduction

In the previous chapter, definitive and efficient algorithms were developed for the solution of boundary references using limacons. From the arguments of chapter 3, if only chart co-ordinate data is available (the radius suppression being unknown) their use is established since they represent the best available approximation to the desired reference shape. In effect this amounts to using boundary limacons in instrument co-ordinates and so investigation of the relationship between boundary limacons and boundary circles in instrument co-ordinates is indicated. From this two important issues must be judged: how accurately can the true reference parameters be estimated if only chart co-ordinate data is available and when with instrument co-ordinate data (if it were available) would it be acceptable to use the convenience of the limacon rather than the circle? Of course, under conditions when the second of these is true, then the first will also be adequately accurate. In analysis only the situation in instrument co-ordinates need be considered.

Rules have been established for determining all the optimum boundary limacons, the optimum ring circle and the presence of local optima of the plug and minimum zone circles. In chapter 3 the question of the quality of the approximation of a limacon to a circle was examined and found to be acceptable under many practical conditions. What is yet to be established is whether the parameters found for a boundary limacon relate to those of the boundary circle to

the same data. If it happens that the appropriate circle parameters can neither be derived from nor closely approximated by those of the limacon then the measurement of out-of-roundness is in doubt. The limacon reference may be behaving as a good approximation to the wrong circle!

In the following sections the correspondence of parameters between circle and limacon will be examined for each of the three boundary references. It is because of the difficulties over multiple optima that only the ring gauge can be expected to give complete answers and so this will be given pride of place and analysed in more detail than the others. Some thought will be given in section 7.5 to the possibility of improving upon the approximation of the limacon in a manner which would have practical relevance to measurement systems.

7.2 Ring Circles and Limacons

The parameters of both circle and limacon fits are dependent upon the geometry of the data which is not known in advance. It is not appropriate, therefore, to seek a general (algebraic) function which links the two sets of parameters. Instead an attempt will be made to establish one set from the other in terms of a value subjected to a definite error bound. Since good calculation methods exist for the limacon fit, the circle parameters will be expressed in terms of those of the limacon.

A limacon always has an axis of symmetry, the line joining origin and centre, and so when discussing purely the geometry of the figure no loss of generality is incurred by setting this axis co-

incident with the x-axis. This procedure will be adopted here since it simplifies the notation. A limacon

$$R(\varphi) = R_L + E\cos\varphi$$

has minimum zone circles (as can be seen from the alternate contact rule) shown in figure 7.1. Their centre is coincident with that of the limacon and their radii are $(R_L^2 + E^2)^{\frac{1}{2}}$ and R_L. Any circle which encloses the ring limacon certainly encloses the data also and will be larger than the ring circle. The outer of the minimum zone circles is not the smallest circle which will enclose the limacon and so its radius, $(R_L^2 + E^2)^{\frac{1}{2}}$, is a definite over-estimate of the ring circle radius.

In the parameter space representation of the ring gauge problem (section 4.4) the limacon constraint lies always below its equivalent circle constraint. So every point on a limacon constraint lies in the infeasible region for the ring circle, except, possibly at E=0 where the two surfaces touch. Thus the ring limacon solution lies outside the ring circle feasible region and since that region is convex,the circle solution must have a larger value of radius than has the limacon solution. Thus the limacon radius, R_L, and so the inner of the minimum zone circles represents a lower bound on the ring circle radius. (Note, incidentally, that this is a good illustration of the power of parameter space analysis: the rigorous proof of this lower bound by geometry is surprisingly complicated (29)).

Other bounds on radius can be established. The minimum radius circumscribing circle to the ring limacon is a smaller upper bound than that quoted above. However the complexity of the expression for its radius is such as to discourage its use unless precision is

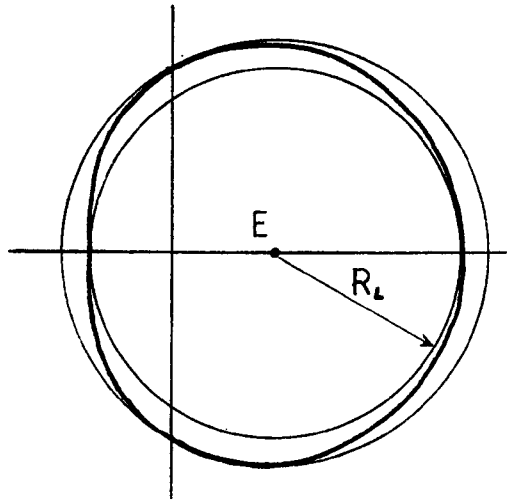


FIGURE 7.1: Minimum Zone Circles to a Limacon

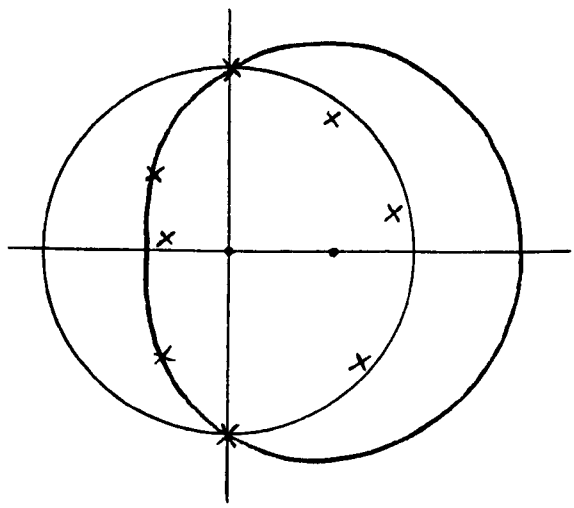


FIGURE 7.2: Minimum Circumscribing Circle and Limacon with Diametral Contacts

of very great importance. The circle constructed through the three contact points of the ring limacon has important properties if those points obey the 180^{o} rule with respect to its centre. If all the data lies within it, it is the ring gauge, otherwise it is smaller than the ring gauge and so represents a high lower bound on ring gauge circle radius. Thus five circles have been defined, being in order of increasing radius:

i) Inscribing circle, radius R_L, to the ring limacon

ii) Circle through contact points of ring limacon and data, 180^{o} rule obeyed for the circle.

iii) Ring Gauge Circle

iv) Minimum radius circumscribing circle to the ring limacon

v) Circle, radius $(R_L^2 + E^2)^{\frac{1}{2}}$, circumscribing the ring limacon

The circles denoted by i) and v) certainly over-estimate the uncertainty in the ring circle radius. An estimate of the ring circle radius will be their mean value:

$$\hat{R}_o = (R_L + (R_L^2 + E^2)^{\frac{1}{2}})/2$$

and, using the identity

$$P^2 - Q^2 = (P + Q)(P - Q) \qquad 7.1$$

their radial separation is $E^2/2\hat{R}_o$.

The ring circle radius is located with an absolute tolerance band

$$R_o = \hat{R}_o(1 \pm \gamma^2/4)$$

where, here, $\gamma = E/\hat{R}_o$, which for small γ will not differ significantly from the true eccentricity ratio as given in section 3.1.

Even taking as a test case the "typically poor" centring condition $\gamma = 0.01$ the ring circle radius can be estimated from the

limacon parameters to better than 25 parts per million. Relating this to the polar chart of a roundness instrument, since the maximum possible eccentricity which could be tolerated would be half the chart annulus width, W, that is ME<W/2, the error ther would be no greater than $\pm W\gamma/8$. For this example the maximum error on the chart could only be 1/8% of the chart width. The accuracy of this ring circle radius estimation is perfectly adequate for almost all practical applications. The closer estimates are available if necessary. On the other hand, the error in taking R_L as an estimate of the circle radius is at most only 50 parts per million at $\gamma = 0.01$ and this value has the advantage of being usable independently of an accurate knowledge of the radius suppression in a comparator system.

Although the circle radius is well estimated under all conditions, subject to γ, the same is not true for the estimation of centre position. Physically sensible examples can readily be constructed in which the difference between the centre of ring limacons and circles is a significant fraction of the true eccentricity. Mostly this effect is caused by the necessity of the ring limacon having three contact points with the data. If a diametral contact existed, the linear programme would introduce eccentricity without affecting that diameter until a third contact was found. Circle fitting does not involve such behaviour. Figure 7.2 illustrates this extreme of the divergence between circle and limacon: it can be seen that the radii of the two figures are the same while the difference in eccentricity is large and that the 180° rule is obeyed marginally for both figures.

In one sense, the best estimate of the ring circle is the circle through the contact points of the ring limacon, providing that the 180° rule is obeyed for the circle. If these contacts are (R_i, θ_i), (R_j, θ_j) and (R_k, θ_k) they are related by

$$R_i = R_L + A_L\cos\theta_i + B_L\sin\theta_i \qquad 7.2$$

with similar expression for j and k where R_L, A_L, B_L are the parameters of the ring limacon. From its cartesian equation the circle through these points is found from the simultaneous equations

$$R_o^2 = (R_i - A_o\cos\theta_i - B_o\sin\theta_i)^2 + (A_o\sin\theta_i - B_o\cos\theta_i)^2 \qquad 7.3$$

with again similar expressions for j and k and R_o, A_o, B_o being the circle parameters. Substituting equation 7.2 into 7.3 and re-arranging gives:

$$\frac{R_o^2 - R_L^2}{2R_L} = E_{LOi}\left(1 + \frac{E_{LOi}}{2R_L}\right) + \frac{(A_o\sin\theta_i - B_o\cos\theta_i)^2}{2R_L} \qquad 7.4$$

where $E_{LOi} = (A_L - A_o)\cos\theta_i + (B_L - B_o)\sin\theta_i$

Now E_{LOi} will be expected to be no larger than the order of size of the limacon eccentricity which is in practice small compared to R_L. To a good approximation, therefore:

$$\frac{R_o^2 - R_L^2}{2R_L} \simeq E_{LOi} + Q_i$$

where $Q_i = (A_o\sin\theta_i - B_o\cos\theta_i)^2/2R_L$. Exactly parallel expressions can be derived using the j and k subscripts. From these three are obtained:

$$E_{LOi} - E_{LOj} = Q_i - Q_j$$

$$E_{LOi} - E_{LOk} = Q_i - Q_k$$

which may be solved to give:

$$A_L - A_o = \frac{E_o^2}{2R_L}\left\{\frac{(\sin\theta_i-\sin\theta_k)\left[\sin^2(\theta_i-\phi_o)-\sin^2(\theta_i-\phi_o)\right]}{\sin(\theta_j-\theta_i)+\sin(\theta_i-\theta_k)} \ldots\right.$$

$$\left. \ldots \frac{-(\sin\theta_i-\sin\theta_j)\left[\sin^2(\theta_k-\phi_o)-\sin^2(\theta_i-\phi_o)\right]}{+\sin(\theta_k-\theta_j)}\right\}$$

$$B_L - B_o = \frac{E_o^2}{2R_L}\left\{\frac{(\cos\theta_i-\cos\theta_k)\left[\sin^2(\theta_j-\phi_o)-\sin^2(\theta_i-\phi_o)\right]}{\sin(\theta_j-\theta_i)+\sin(\theta_i-\theta_k)} \ldots\right.$$

$$\left. \ldots \frac{-(\cos\theta_i-\sin\theta_j)\left[\sin^2(\theta_k-\phi_o)-\sin^2(\theta_i-\phi_o)\right]}{+\sin(\theta_k-\theta_j)}\right\}$$

Where (E_o, ϕ_o) is the polar form of (A_o, B_o). Although these expressions can only be used for specific cases, some general behaviour can be seen. Noting that

$$\sin^2(\theta_j-\phi_o) - \sin^2(\theta_i-\phi_o) = \sin(\theta_j-\theta_i)\sin(\theta_i+\theta_j-2\phi_o)$$

and that since the 180^orule with respect to the origin must be obeyed for the points so that all terms in the denominator have the same sign, it is seen that the expressions are made of combinations of terms having one of the following forms:

$$\left|\frac{\sin(\theta_j-\theta_i)}{\Delta}\right| < 1$$

$$\left|\sin\theta_i - \sin\theta_k\right| < 2$$

or $$\left|\sin(\theta_i+\theta_j-2\phi_o)\right| < 1$$

Even if all the terms are allowed to take their least favourable values simultaneously it is not possible for their combination to exceed ± 4 and so

$$\left| A_o - A_L \right| \leq \frac{2E_o^2}{R_L} \simeq 2\gamma E_o$$

is an over-estimation of the uncertainty band. The same uncertainty is found in $B_o - B_L$. For small γ, which was in any case assumed in deriving this expression, the limacon centre is a good estimate of the centre of the circle through the contact points. If this circle is a good approximation to the ring circle then the ring limacon parameters represent a good approximation to the ring circle parameters.

Given the above observations it is of some importance whether or not the limacon contact points obey the 180° rule with respect to the centre as well as the origin. Whenever they do a reasonable estimate of the ring circle parameters can be guaranteed. Consider the situation shown in figure 7.3. The 180° rule with respect to O is obeyed if the figure is a ring limacon and so there will be contacts on both sides of line PB. If there are contacts with both sides of line PC the 180° rule with respect O' is obeyed. So difficulties only occur if the only contact to the right of PB lies in the sector BOC. Applying the sine rule to triangle POC gives

$$(R_L + E\sin\psi)(\sin\psi\cos K - \cos\psi\sin K) = R\sin K$$

Since ψ and K cannot, physically, exceed $\pi/2$ this may safely be expressed in terms of $\tan\psi = t$ and $\tan K = E/R_L = \gamma$ to give:

$$(1 + t^2)^{\frac{1}{2}}(t - \gamma) = \gamma(1 + \gamma t) \qquad 7.5$$

The diagram makes it clear that $t \geq \gamma$, so all terms are positive and squaring has no deleterious effect. After regrouping equation 7.5 yields:

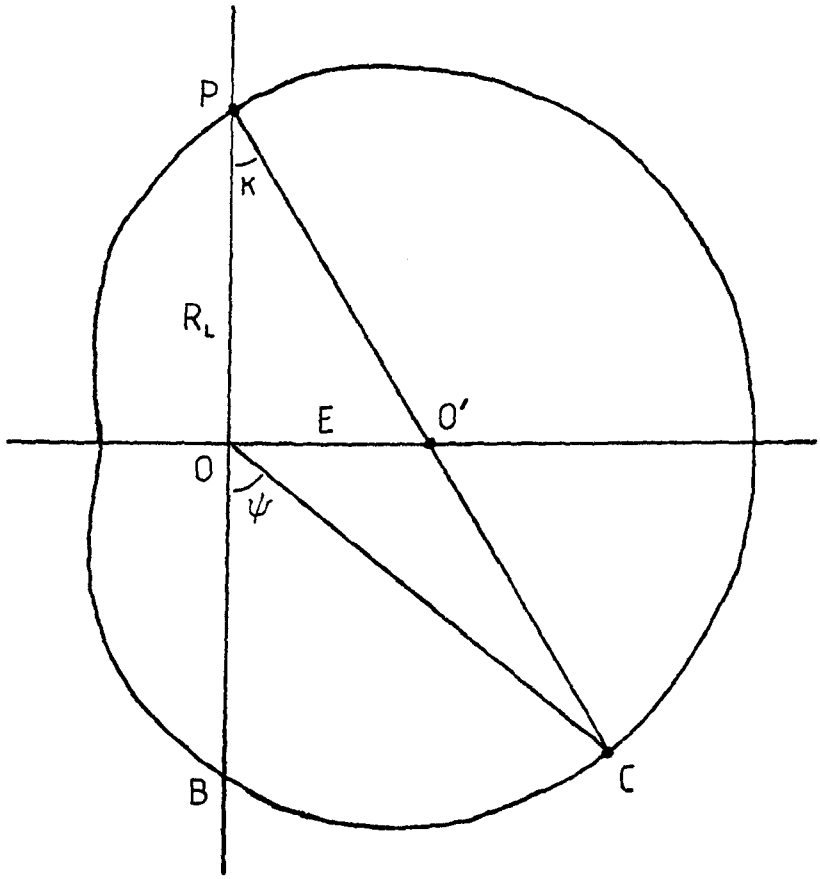


FIGURE 7.3: Conditions on the Satisfaction of the 180° Rule with respect to the Centre O'

$$t\left(t^3 - 2\gamma t^2 + (1+\gamma^2-\gamma^4)t - 2\gamma(1+\gamma^2)\right) = 0 \qquad 7.6$$

The solution t=0 has no significance here. The remaining cubic can be seen from the sign alternation of its co-efficients to have no real negative roots and one, or three, positive real roots. This corresponds to the geometrical sense of figure 7.3. For $\gamma \ll 1$ a general solution of equation 7.6 is not attempted. Instead by neglecting all terms in γ^2 and γ^3 a factorisable expression is revealed:

$$(t - 2\gamma)(t^2 + 1) = 0$$

from which the real solution is interpreted as

$$\tan\psi = 2\gamma \qquad 7.7$$

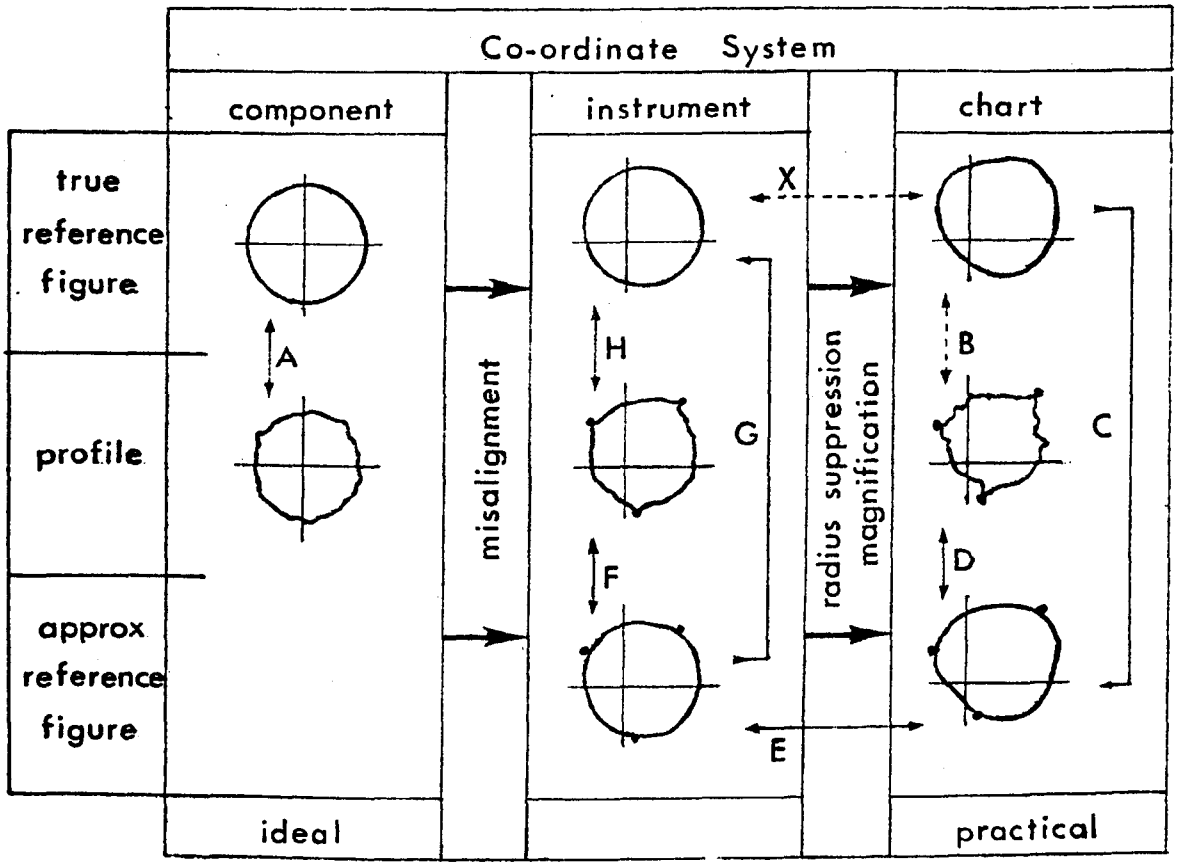
(This approximate formula is in fact very close to the true solution: a numeric solution of equation 7.6 using Newton's method showed that even with $\gamma = 0.1$, using equation 7.7 gives an error of only about one part in 10^4.)

Figure 7.3 shows, in effect, the worst situation which can occur in this context, that is when a contact point lies at ⁻. Thus in equation 7.7 is embodied a complete criterion for the 180° rule to be obeyed relative to the centre. Providing each pair of adjacent contact points subtend an angle at the origin of greater than $\tan^{-1}(2\gamma)$ then the ring limacon parameters may be used directly to estimate the ring circle parameters. In a sampled data system, this criterion can become absolute: if the angular sampling interval exceeds $\tan^{-1}(2\gamma)$ then, apart possibly from diametral contact which is easily checked, the 180° rule must always be satisfied relative to the centre. As an example, with 512 equispaced points the condition is always satisfied providing $\gamma < 0.0061$.

As a consequence of these analyses, it seems that, although

not a certainty, there is a high probability that under practical conditions the circle through the ring limacon contact points is a good approximation to the ring circle. Further the test on the angle between contact points is a simple way of detecting possible bad cases. The work described in section 10.2 gives evidence for this conclusion.

A complete cycle of operations relating the initially desired measurement to that actually attainable with a normal roundness measuring instrument, that is one working in chart co-ordinates, can be established for ring gauge, subject to the conditions above, namely that γ must be kept small. This cycle is illustrated in figure 7.4. It is deliberately not explained in detail since it is intended to provoke further thought about its implications. The basic flow around the figure is as follows. It is desired to ascertain the deviation of a workpiece cross-section from true circularity for an application where the dominant outer-features are of particular significance. Placing the workpiece on the instrument requires that an eccentric circle, optimised as the ring gauge circle, should be used as a reference. Given difficulties over the accurate knowledge of radius suppression and also those of calculation, this equivalent distorted circle in chart co-ordinates cannot be used. As a linearised approximation to this distorted circle a limacon is fitted onto the chart profile. Because of the nature of radius suppression, the contact points of the ring limacon in instrument co-ordinates are identified with those of the ring limacon on the chart. Thus the ring limacon parameters for instrument co-ordinates can be established and from them an adequate representation of the ring



A is ideally required measurement.

B practical fit not viable since X not accurately known.

C distorted circle modelled by limacon.

D fit can be found, contact points defined.

E linear parameters can be transferred.

F found from D since E preserves contact points.

G upper and lower bounds to circle parameters.

H defined as parameter value and tolerance from F and G.

FIGURE 7.4: Summary of Ring Gauge Circle Measurement Philosophy

circle parameters is found in terms of a value and an associated tolerance band. Thus, in principle, the cycle is closed, and so a relation to the desired measurement achieved, by using two approximations. Firstly the modelling of an eccentric circle by a limacon and then, at the end, the estimation of the ring circle parameters from those of the ring limacon. Both of these steps take place in instrument co-ordinates.

The final stages of closing the cycle require the use of the true radius and so the radius suppression must be known. However estimates having only linear dependence on radius suppression can be used for the ring circle parameters so that comparisons at fixed, unknown, radius suppression may be performed with ring circles.

7.3 Plug Circles and Limacons

Given a problem in optimisation, such as the plug circle, where there may be several local maxima only one limit may be ascertained with certainty in a useful form. The lower bound to any local maximum must also be a lower bound to the global value but the upper bound to the local maximum may still be smaller than the global value. Thus only a weak upper bound can usually be obtained. In the case of the plug circle, it is clear that any circle which encloses all the data cannot be positioned so that it inscribes that data (as opposed to a position at which it does not enclose any data points, see section 4.2). So the ring circle may be specified as an upper bound to the plug circle! Usually, however, such a bound is not of much use for measurement purposes.

Using the same style of argument as adopted with the ring circle, information concerning a local maximum of the plug circle can be obtained. A lower bound for the plug circle radius is obtained from the inscribing circle, radius R_L, to the plug limacon. The circle constructed through the three contact points of the plug limacon with the profile will be, if those points obey the 180° rule with respect to its centre, either a local maximum inscribing circle (when none of the profile lies inside it) or an over-estimate of the radius of a local maximum. If this condition is obeyed, the radius of the circle through the points which lie on the limacon cannot be greater than that of a circumscribing circle to that limacon. Thus a pair of bounds, similar to those for the ring circle, is established. A local optimum of the plug circle exists with radius bounded by values R_L and $(R_L^2 + E^2)^{\frac{1}{2}}$, where R_L and E are parameters of the plug limacon, providing that the defining contact points of that limacon obey the 180° rule with respect to the centre of the circle constructed through them. This circle may be treated as a best estimate of the local maximum and the same analysis regarding its centre carried out as was performed for the ring circle in the previous section.

Results can be obtained for estimating plug circle parameters which exactly parallel those for the ring circle. They are weaker, however, on two counts. Firstly they refer only to a local maximum of the plug circle having its centre in the vicinity of that of the plug limacon. Secondly the tolerance bound for the radius of this local maximum is not true under all conditions but depends upon the geometry of the limacon contact points. Under practical conditions the necessary geometry has already been shown to be likely to occur,

so the restriction may not be too serious. The plug limacon is found from constraints for fitting circles which have been linearised about the point of zero eccentricity and will therefore tend to identify a local maximum having small eccentricity. The selection of which local maximum is discovered is, in principle, governable by the initial instrumental set-up. Over-all, although there are profound mathematical and philosophical difficulties with this method, it appears that as an engineering technique it is useful providing that some care and forethought is given to the measurement and its interpretation.

7.4 Minimum Zone Circles and Limacons

The potential non-uniqueness of solution to the minimum zone circles causes difficulties similar to those encountered with the plug circles compounded by a more complex geometry. Again the only absolute bound on the zone value which can be established is that falling short of optimum, that is an upper bound to a minimisation.

The minimum zone circles to the minimum zone limacons clearly have a separation greater than that of the minimum zone circles to the profile. The zone over-estimate can thus be formed by circles having radii (R_L-H) and $((R_L+H)^2 + E^2)^{\frac{1}{2}}$. Using the identity, equation 7.1, the zone estimate would then be:

$$2\hat{H}.2\hat{R}_o = (R_L + H)^2 + E^2 - (R_L - H)^2$$

$$= \left(2H + \frac{\gamma E}{2}\right) . 2R_L$$

This suggests use of the limacon radius as an estimate of the minimum zone circles mean radius and gives the over-estimate of the zone

as $(2H + \gamma E/2)$. Generally it would be adequate to take simply 2H as the measurement of the radial difference of the zone circles under practical measurement conditions.

A pair of concentric circles constructed so that each passes through either the inner or outer pair of contacts between the minimum zone limacons and the profile may give a bound on the minimum zone circles. If the four contact points satisfy the alternation condition with respect to the centre of those circles and also enclose the whole profile between them then they are local minimum zone circles. If they do not totally enclose the profile but obey the alternation conditions then it is only possible to enclose the whole profile between a larger zone that is centred near to the centre of those circles. Thus they offer a form of lower bound to the minimum zone and perhaps a "best estimate" in the sense adopted for the ring circle in section 4.2. The conditions that the points have alternate contacts relative to both the origin and the centre of the fitted circles can be seen from figure 7.5. The points P and Q are encountered in the opposite order from O' than from O. It is clear that this reversal will not happen if O' is anywhere to the same side of the straight line through P and Q as is O. The allowable eccentricity to maintain the order depends upon the relative size of R and H, the angle, α, subtended by the points at the origin and the direction of eccentricity relative to the points. Given this number of variables, the general analysis will be pursued no further.

The minimum zone approach is particularly relevant to the assessment of the magnitude of out-of-roundness. However the physical relevance of the reference parameters, for instance the

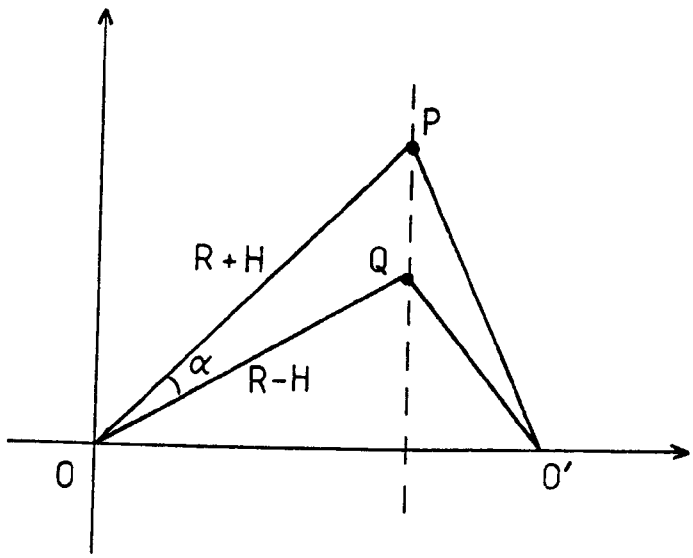


FIGURE 7.5: Preservation of Conditions of Alternating Contacts from a Centre O' for Minimum Zone Figures

centre position, seems less than that of the other references. They relate, if only loosely, to physical situations that a workpiece might encounter while the minimum zone does not obviously do so. The estimation accuracy of centres of local minimum zones will not therefore be investigated here. Methods broadly similar to those used in section 7.2 should give some indication if needed. The most common measurement is likely to involve testing whether the out-of-roundness exceeds some specified tolerance and for this just an over-estimate of the zone (providing it is not excessively large) will suffice. The direct use of the limacon zone value seems indicated for normal purposes.

7.5 Out-of-Roundness Measurement

The main concern of this chapter has been to establish the level of agreement which can be expected between reference limacons and the true reference circles. This is typified by the approach illustrated in figure 7.4 in which it is seen how the ring gauge circle relates to the ring gauge limacon. If the out-of-roundness is to be measured, the best estimate of the ring circle (based upon the fitted limacon) is constructed and imperfections of form measured from it. It would clearly be more convenient (and is indeed customary practice) for measurements to be taken directly from the fitted limacon. Doing this introduces two sources of error. First the circle and limacon do have different shapes, so that the relative distance of individual data points from each reference will be different even if all other features were in agreement. Secondly the limacon is defined, and the data points measured, with respect

to the co-ordinate system origin and not from the reference centre. Both of these error sources are of a purely geometrical nature and can be analysed. However a third source of error, not fully analysable, is that discussed earlier in this chapter; the circle which the reference limacon best represents may well not be the desired reference circle. In particular the doubt over the position of the reference circle centre is important. Most out-of-roundness values are expressed as peak-valley so that constant radial errors cancel out.

One case where a reasonable analysis is possible is that depicted in figure 7.6 where the ring gauge circle and ring gauge limacon have the same defining contact point. It has been shown that for this condition the distance of the circle centre from (E, 0) is small compared to E and so any error in assuming its centre to be at (E, 0) will be of second order compared to the effect being examined. A data point P, (R, θ) from the origin and (S, ψ) from the centre, is measured as being ε from the circle and δ from the limacon. From the cosine rule:

$$R^2 = S^2 + E^2 + 2ES\cos\psi$$

$$= S^2(1 + \eta^2 + 2\eta\cos\psi) \qquad 7.8$$

where $\eta = E/S$ is a ratio akin to and of the order of size of γ. In the present context R and S will always be positive and so square roots may be taken. Also

$$\varepsilon = R_o - S$$

and

$$\delta = (R_L + E\cos\theta) - R$$

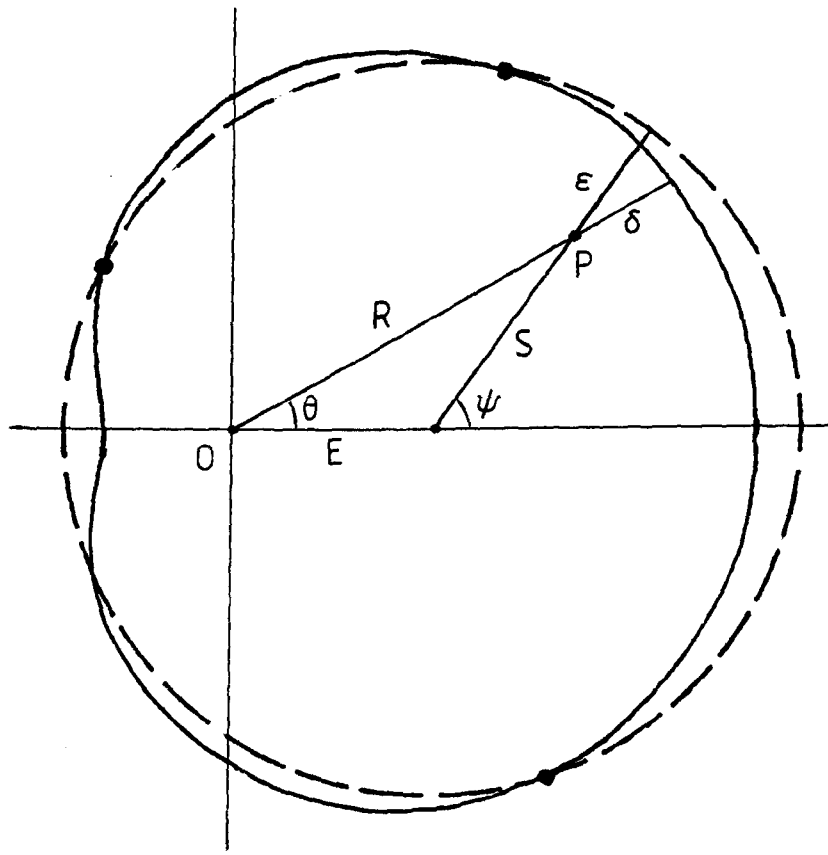


FIGURE 7.6: Minimum Circumscribing Circle and Limacon Having Same Contact Points with the Data

$\cos\theta$ may be eliminated from this by equating projections along the x-axis:

$$\delta = R_L + E\left(\frac{S\cos\psi + E}{R}\right) - R$$

which on substituting for R from equation 7.8 and simplifying gives:

$$\delta = R_L - S.\frac{1 + \eta\cos\psi}{(1 + \eta^2 + 2\eta\cos\psi)^{\frac{1}{2}}}$$

Hence:

$$\delta = R_L - (R_o - \varepsilon).f(\psi, \eta)$$

As ψ describes a circle, the function f will take a maximum value of unity and a minimum value of $(1 - \eta^2)^{\frac{1}{2}}$. Even at eccentricity ratios of about 0.01, f varies from unity by no more than a few parts in 10^5. The modification to the small displacement ε caused by multiplying it by f will be practically undetectable. However R_o is a much larger value and multiplying it by f could cause a variation significant compared to δ and ε. So to a very good approximation:

$$\delta - \varepsilon = R_L - R_o f(\psi, \eta)$$

It has been shown that $R_L \leq R_o \leq R_L(1 + \gamma^2)^{\frac{1}{2}}$ and so the limits of $\delta - \varepsilon$ can be stated in terms of both R_0 and f simultaneously taking their maximum and minimum values. For small η and γ this gives

$$-R_L\frac{\gamma^2}{2} \leq \delta - \varepsilon \leq R_L\frac{\eta^2}{2}$$

Given the similarity, under practical conditions, of η and γ, a practical bound is:

$$|\delta - \varepsilon| \leq R_L\frac{\gamma^2}{2} = \frac{\gamma E}{2}$$

By applying the sine rule to the diagram, figure 7.6, it is

found that:

$$\psi - \theta = \sin^{-1}(\eta \sin\theta) = \sin^{-1}\left[\frac{\eta \sin\psi}{(1+\eta^2+2\eta\cos\psi)^{\frac{1}{2}}}\right]$$

The angular error oscillates about zero with an amplitude of $\sin^{-1}\eta$, which is of the order of $\sin^{-1}\gamma$.

Estimating limits on the error of measuring peak to valley and not just locating one point relative to the limacon rather than the circle is complicated by the possibility that different data points may define the maximum with respect to the two references. Clearly, though, the error in peak to valley can never exceed twice the errors of a single point. It would seem that measuring from the limacon rather than the circle will for the present case be quite satisfactory under current practice. If high eccentricity measurements are made, γ still presents a simple control on the level of error.

The above analysis is specific to the case of ring gauge limacon and circle references which have (to a good approximation) the same centre. It does not apply to any other condition. It can be used intuitively to give some illumination to the more general but imponderable problem. If it is accepted as a practical decision (and often there will be little choice) that the centre of the reference circle will be taken as identical to that of the reference limacon, then the errors inherent in so doing will not be seriously compounded by measuring the out-of-roundness from the limacon.

7.6 "Improving" the Reference Shape

The previous sections have examined how a limacon reference

compares to the true circular reference. The comparison must take place in instrument co-ordinates and so an implicit assumption has been made that the radius suppression is known. In some cases, although the analysis needs knowledge of radius suppression, the results may be taken directly from the limacon without greatly increasing the uncertainty band of the estimate when related to the circular reference. The overall conclusion might be summarised that while the limacon is not ideal its use its use is the best that can realistically be done. This is certainly true of work in chart co-ordinates when the radius suppression is ill-defined.

Notwithstanding the above conclusion, there remains the question of whether, if adequate knowledge of the radius suppression exists, a better reference than the limacon is available. The most obvious reason for this is that the quality of the limacon approximation depends upon the eccentricity ratio and while this is usually adequately controlled in practice at present there may be a move towards working at larger values. There is currently increasing interest in working at higher eccentricities either to save time by reducing centring operations on the instrument or to allow automatic measurements without the expense of very accurate placement systems. Also there is increasing use of small radius components in precision work. Together these two trends may create conditions where the limacon is inadequate.

If the limacon is too approximate but the circle too difficult to calculate than some compromise between them must be sought. This idea is not altogether new. Instruments have been proposed, by the Perthen organisation in Germany and probably elsewhere, in which,

after calculating the least squares circle (using a limacon), the reference line drawn onto the graph was constructed from the power series expansion of the circle, equation 2.3, truncated after the second order term. This idea has not been actively pursued: it has practical difficulties concerning the necessary precision of calculation once absolute radius has to be used which can only be overcome by the recent advances in inexpensive digital computation. It is now timely to look again at this type of approach. Here, since definitive methods are available for both limacons and circle fitting, it will be studied with reference to the ring gauge.

The direct solution of the ring circle as a quadratic programming problem (section 4.3) is not considered to be practicable using the computational power that could reasonably be included with a roundness measuring system. Much more work would be required than for a linear program since the solution would have to use the primal and a direct jump between vertices in parameter space is not allowed, it being necessary to search the whole boundary. The approach mentioned above is also not directly useful. Using a reference figure modified according to the limacon reference parameter may make matters worse because if the limacon does not adequately represent a circle then it will not fit the data as would a circle. Another method which could be adopted is to fit a limacon reference and then simulate the recentring of the instrument to that reference by performing a co-ordinate translation on the data after which a new reference limacon can be fitted with, presumably, a more favourable eccentricity ratio. A disadvantage of this scheme, apart from the high precision of the arithmetic needed for the translation, is that the translated data would not be uniformly spaced even if the original data was and

uniform spacing gives quite large gains in computational efficiency.

The method which is suggested here combines features from both the other approaches. Its basic logic is that rather than translate the data to a region of space where the limacon appears less distorted with respect to a circle, the data should be distorted in a manner similar to the difference between the figures. The ring limacon to this distorted data should behave in a similar fashion to the ring circle to the original data. If an initial ring limacon to the data indicates parameters $\hat{R}_o$ $\hat{E}$ and $\hat{\phi}$ as the first estimate of the ring circle, the distortion to be introduced into that data is the second term of the power series expansion of the circle using these estimates. Thus the constraints of the refitting are simply:

$$A\cos\theta_i + B\sin\theta_i + R_o \geq R_i - \frac{\hat{E}^2}{2\hat{R}_o}\sin^2(\theta_i - \hat{\phi})$$

which may be compared with equation 6.1. This method retains the advantage that the data remains uniformly spaced with respect to the origin so that exactly the same ring limacon fitting routine is used both before and after the data modification. It is therefore quite efficient. Another point is that since only shape relative to the origin is modified the change is not dependent upon radius suppression other than for the initial calcualtion of $\hat{E}^2/2\hat{R}_o$. The second fitting calculation can therefore be carried out effectively in chart co-ordinates with a consequent saving in necessary arithmetic precision.

Under the conditions which might be expected to occur with practical systems, the "correction term" is likely to remain on the level of exactly that. At most the distortion it accounts for might represent a few per cent of the total measurment value. This implies

that even a quite large percentage error within the estimate of the correction value will be acceptable (a 10% error in a correction of 5% of value has a probably negligible effect on the value itself). Thus it seems that even quite crude estimates of the radius suppression would be adequate for the purpose of the refinement considered here. Certainly a precision equal to that of the measurement data of the profile would not be required.

It remains an open question whether there is much demand for refinement of the limacon models of circular references. This will become clear with experience of new instruments and new measurement problems. The decisions on their use are likely to remain empirical and cost related for a significant time into the future. However useable refinements are shown here to be available when needed, some practical evidence being given in section 10.4.

8. Polygons as Reference Figures

8.1 Introduction

Even if corrector terms are applied to reference systems using limacons (section 7.6) there remains the restriction with such methods that the origin must lie inside the figure, that is the eccentricity ratio cannot exceed unity. In some measurements, notably when using a co-ordinate measuring machine, the data may well be expressed relative to an origin which is nowhere near the profile. (Note that this could only occur in instrument co-ordinates: radius suppression is meaningless under these conditions.) To use limacon methods it is necessary to find an initial estimate of the profile centre which may be used as an origin for re-expressing the data. An alternative would be to use some other approximation to a circle which retains advantage of computation. One such alternative could be to approximate the circle piecewise by a series of straight lines, that is the circle represented as a polygon. This concept is not the same as that used by Scheiding (section 6.4 and (10)) where straight lines are used as a data reduction device rather than as a reference figure.

Each side of a polygon is a section from a straight line in the measurement plane and so can be described by parameters in which it is linear. So providing that all the sides can be handled together in a simple way the polygon reference is capable of a linear parameterisation and consequently should be relatively easily optimised. In fact it turns out to be possible to express all the four reference circles in terms of equivalent polygons in which the

sides have fixed directions relative to the frame of reference. In the case of the least squares fit, however it is necessary to make assumptions about what is meant by residuals and the generalisation is rather inelegant. It seems likely that limacon and centroid methods will prove a more satisfactory approach to the average reference figure (30). Here as an example of the approach the minimum circumscribing polygon will be examined. Also, in section 8.3, reference squares will be used to illustrate the nature of boundary reference fits. This discussion will be parallel to that given for circles but gives a different viewpoint of the cause of the ensuing difficulties.

8.2 The "Ring Polygon"

In defining a polygon reference it will be assumed that the regular figure is constructed from the envelope of a series of straight lines, of pre-defined slope, each having the same perpendicular distance from a single point. This point is the "centre" of the polygon and the perpendicular distance is the "radius". These parameters are, in fact, those of the inscribing circle to the polygon.

For the circumscribing figure, each line must lie further from the centre than the distance of each data point projected onto the normal of that line, providing the point lies to the same side of the centre as does the line. The enclosure of the data is conveniently ensured by using pairs of parallel lines, see figure 8.1. A line of slope m passing a point (A, B) at a perpendicular distance R can be expressed in the cartesian plane XY as:

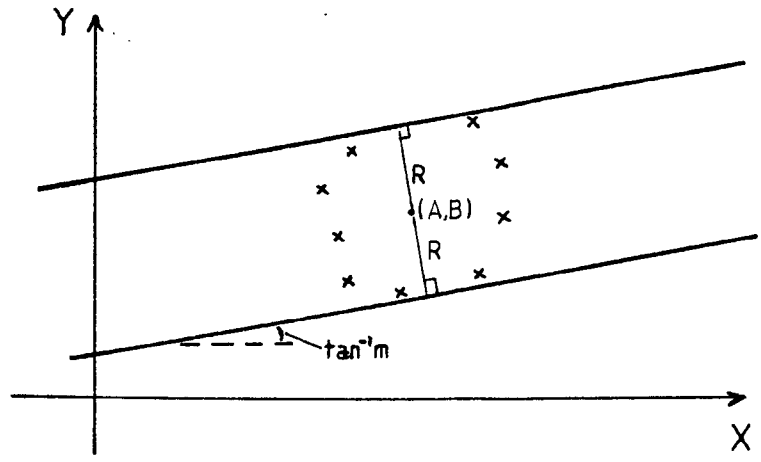


FIGURE 8.1: Enclosing Data between a Pair of Parallel Lines of Slope m. Polygon, Centre (A,B) and "Radius" R is made up from such Pairs.

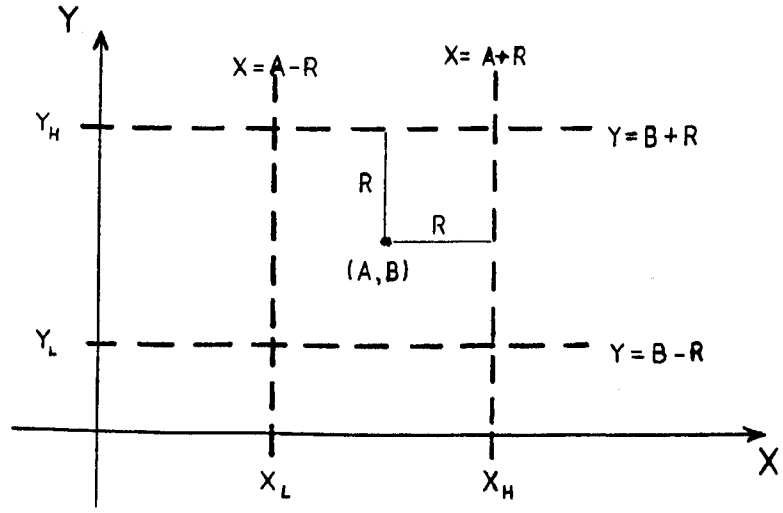


FIGURE 8.2: A Circumscribing Square Reference

$$(Y - B) = m(X - A) + R(1 + m^2)^{\frac{1}{2}} \qquad 8.1$$

Thus for the pair of lines to enclose the data, the following constraints must be satisfied for all data points (X_i, Y_i) simultaneously:

$$\begin{aligned} (Y_i - B) &\leqslant m(X_i - A) + R(1 + m^2)^{\frac{1}{2}} \\ (Y_i - B) &\geqslant m(X_i - A) - R(1 + m^2)^{\frac{1}{2}} \end{aligned} \qquad 8.2$$

This form becomes difficult when $m=\infty$ (lines parallel to the Y aixs) and so additional constraints, also required to be true for all data points are introduced:

$$\begin{aligned} (X_i - A) &\leqslant R \\ (X_i - A) &\geqslant -R \end{aligned} \qquad 8.3$$

The constraints are linear in the parameters (A, B, R) and as R is to be minimised, linear programming may be used to find the complete polygon reference which will involve such sets of constraints over a set of values of m corresponding to the orientation of the sides of the polygon. It is obvious from figure 8.1 that only one (or possibly a series lying on a straight line) data point can influence the closeness of approach of each of the parallel lines to the data. In a parameter space (A, B, R) representation this shows as the relevant constraint boundaries for a given slope of line being parallel planes. The points limiting each of the line pairs will be those having the largest and smallest values for $(Y_i - mX_i)$. Only these two points can possibly be active in defining the sides of slope m to the polygon. Thus the linear programming exercise can be significantly simplified by breaking it down into two linear programmes, the first of which is to identify

the potential defining points and is trivial in the sense that no iteration is required.

Denoting the slopes of the various lines making up the polygon as m_j and letting Q_{+j} and Q_{-j} be the maximum and minimum values of $(Y_i - m_j X_i)$ over all i, then the required optimisation is, from equation 8.2:

$$\text{Minimise:} \quad R \tag{8.4}$$

$$\text{Subject to:} \quad R \geq \frac{m_j A - B + Q_{+j}}{(1 + m_j^2)^{\frac{1}{2}}}$$

$$R \geq \frac{-m_j A + B - Q_{-j}}{(1 + m_j^2)^{\frac{1}{2}}}$$

for all j.

Again forms equivalent to equation 8.3 would be used to account for $m_j = \infty$. The linear program of equation 8.4 is solved directly, the dual being the more efficient form to use. If there are N data points and n sides to the polygon the total work involved is n searches through N points to find the values for Q followed by a linear program in three variables and n constraints (vice versa in the dual). At least with fairly small numbers of sides, the technique is a plausible one in terms of the effort needed to use it.

8.3. Reference Squares

The simplest reference polygon consists of a square in which the sides lie parallel to the co-ordinate axes. The four straight lines defining this figure are:

$$X = A+R = X_H \; ; \quad X = A-R = X_L$$

$$Y = B+R = Y_H \; ; \quad Y = B-R = Y_L$$

Figure 8.2 shows this square, stressing that only a section of each line is active in defining the figure.

The circumscribing figure has been examined in the previous section and so here it suffices to note that the constraints required to ensure that the square encloses the data are:

$$\underset{i}{\text{Max}}(X_i) \leqslant X_H \; ; \quad \underset{i}{\text{Max}}(Y_i) \leqslant Y_H$$

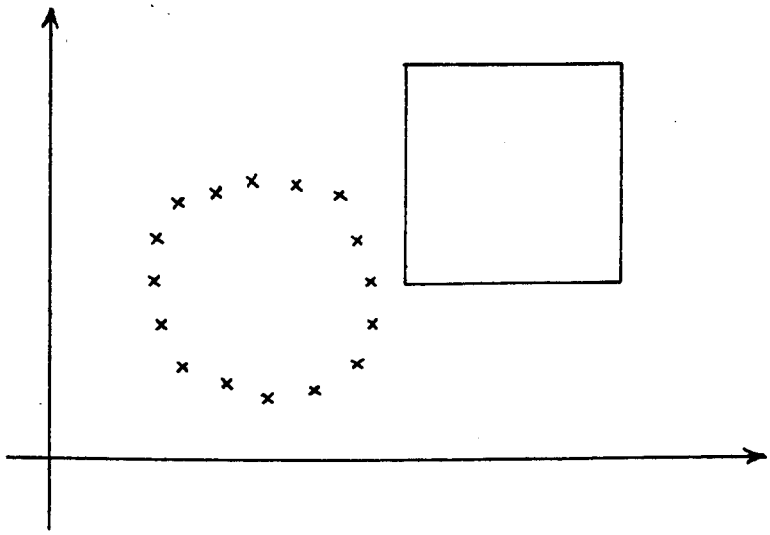
$$\underset{i}{\text{Min}}(X_i) \geqslant X_L \; ; \quad \underset{i}{\text{Min}}(Y_i) \geqslant Y_L$$

All constraints must be satisfied simultaneously. It is seen from figure 8.2 that the sections of each line not relevant to the formation of the square are automatically excluded by the constraints.

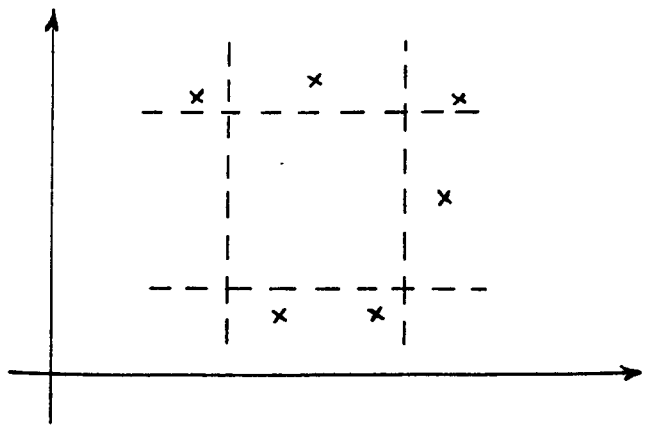
Consider now, the definition of an inscribing square. It is not sufficient to demand simply that the data is not contained within the square since this will allow solutions in which data and reference are totally unconnected, see figure 8.3. The other example in figure 8.3 concerns the problem of how to avoid the initially inscribing figure from 'escaping' when there is a significant gap in the data: it is perhaps not clear what interpretation may be placed on the phrase "maximum inscribing" in this situation. Some constraints must be applied to tighten the definition. If the data sensibly describes a closed figure then insistence that points lie on all sides of the square is a suitable condition, giving:

$$\underset{i}{\text{Max}}(X_i) \geqslant X_H \; ; \quad \underset{i}{\text{Max}}(Y_i) \geqslant Y_H$$

$$\underset{i}{\text{Min}}(X_i) \leqslant X_L \; ; \quad \underset{i}{\text{Min}}(Y_i) \leqslant Y_L$$



A - Excluding Square



B - Inscribing - but what happens if attempting to Maximise

FIGURE 8.3: Difficult Cases for an Inscribing Square Reference

Another approach could be to insist that the inscribing figure lies totally within the minimum circumscribing figure, which does not have such problems of definition. This again would generate four constraints on X_H X_L Y_H and Y_L but not related directly to the data points. These four constraints must be satisfied simultaneously.

These constraints force the reference to lie within the field of the data but do not cause it to be inscribing: it is now necessary to introduce constraints to cause the exclusion of all data points. It is obvious that, unlike the circumscribing square, there must be data points lying to both sides of each of the defining lines. Further it does not matter where a point is with respect to say, $Y = Y_H$ providing it does not satisfy $X_L < X_i < X_H$. A point lies outside the square if any of the following four conditions are true:

$$X_i \geq X_H \; ; \quad Y_i \geq Y_H$$

$$X_i \leq X_L \; ; \quad Y_i \leq Y_L$$

The difference in the behaviour of the constraints for circumscribing and inscribing polygons is directly describable by logic operators. In the circumscribing figure the constraints are combined as an AND function whereas with the inscribing figure thay are combined as an INCLUSIVE OR. Because of the inclusive property of constraints on the inscribing square some parts of the constraint boundaries in parameter space disappear and immediately the constraint planes do not completely divide parameter space there appears the possibility of non-convexity even with linear functions. There is in principle a different solution for the optimum for each valid choice amongst

the ORed constraints, although not all may make physical sense. The four constraints limiting its position (which are ANDed) bound the region of parameter space containing these solutions. (As a trivial example of what is happening consider attempting to find the gap in the sequence 1,2,3,7: the answer may be obvious but mathematically there are three gaps between 1 and 2, 2and 3, 3 and 7 which must be examined.)

Under conditions where it makes sense to use an inscribing reference, it should be possible to make the choice among the possible ORed constraints on physical grounds rather than having to search them. One possibility might be to select the constraint according to the geometric position of the particular data point relative to the centroid of all the data points. This would identify whether to use X_H or X_L and also Y_H and Y_L but it is less clear how to best make it choose between the remaining X and the remaining Y. In any case it is clear that the ORed constraints cause extra work and greater uncertainty about the result.

The minimum zone reference is, in terms of the constraints required, a combination of the circumscribing and inscribing cases. There will again be ORed constraints from the inscribing part. Similar problems to those encountered with the inscribing square will occur, but in this case it is more difficult to establish simple methods of choosing between the possible combinations of constraints.

9. Practical Assessment Systems

9.1 General

Whatever theoretical niceties may be developed, the measurement of roundness, as, indeed, all other measurements, will remain essentially a pragmatic exercise. A measurement is made because of uncertainty concerning the nature of the profile. Any method which clarifies that uncertainty sufficiently for the purpose in hand is acceptable for performing the measurement and generally the most economical approach would be adopted.

Narrowing the discussion to the use of reference figures with roundness measuring instruments does not alter the motivation enshrined in the above statement. If the solution is to be found by hand then the only really practical method is to fit circles onto the polar chart since circles are easily handled by templates and compasses. The penalty to be paid for this simplicity is the need for rather precise centring. The plotting of limacons is possible (see Appendix 4) but would be expected to entail more work than is involved in the refinement of the centring. The geometry of the exchange algorithms developed here may be useful in confirming the solution obtained but is unlikely to be used formally because of the powerful intuitive ability of eye and brain.

If automatic solution of the reference figures is required the situation is greatly changed. With analogue systems only least squares limacons can be accurately found (by using simple Fourier analysis) and so the application of some form of digital computer to the instrument is implied. The reason for using automatic assess-

ment is presumably either to increase accuracy or, more commonly, to reduce the total measurement time. In this case there are clear constraints on the type of computing that is feasible. Cost is an important consideration for it may be more economical to use several simple instruments than one more efficient, but much more expensive, system, particularly when the possibility of breakdown is taken into account. Computational time is also significant, although not always for what, at first sight, may seem the obvious reasons: if the computation can proceed in parallel with workpiece unloading and the setting of the next piece, it may not affect the measurement cycle time at all! The requirements for algorithms which are to be practically useful are that they can operate on mini- or micro-computers which might be found on-line to the instrument while retaining an adequately fast (this value can only be related to specific situations) performance and a sensibly better accuracy than that of the instrument so that the overall system performance is not down-graded by their use.

There are occasions, including some within this work, where it can be asserted from a purely mathematical analysis that one method is superior to another. Such an assertion does not require testing experimentally. However in an engineering context rather than a mathematical one this may no longer be so since the meaning of 'superior' becomes blurred. Here it is considered to be necessary to develop working versions of all the plausible methods which have been considered, not to test whether they work but in order to judge their relative cost-effectiveness under practical conditions. This process should not, of course, be taken to extremes: for example, from Chapter 6 it is clear that the direct solution of the primal

linear programme for ring limacon is rendered an implausible approach by the existence of the dual.

Having established that there would be the need to develop and test working measurement systems whatever the rigour of their theory, it should be stressed that there is need of direct experimental study in this work. It is necessary to gather practical experience of real measurement problems both to further the study of reference systems and 'for posterity'. The practical engineer or metrologist is aided in his judgement of his own problems by being able to call upon the experimental data of others. Those working in standardisation should always be seeking to simplify procedures and will, therefore, require information on which to judge whether, for instance, there is significant advantage to be gained by retaining alternative forms of analysis.

The discussions in chapter 7 make it clear that the quality of the limacon as an estimator of the true circular reference depends upon the particular geometry of a measurement. There is clearly a need to discover experimentally whether the occurence of adverse geometries is likely to be frequent. This will require that large amounts of work are performed, preferably by many authorities. In the next chapter the various studies that would be needed are illustrated by some pilot experiments performed under various instrumental conditions with minimum circumscribing references. This chapter will be concerned first to describe the experimental system which was developed and then to examine the general performance of the techniques proposed here.

9.2 The Experimental System

The basic components of a roundness analysis system had been developed prior to the start of the project described here in the Research Department of Rank Taylor Hobson, Ltd. mainly by the author and Kinsey (31). This consists of a modified Talyrond 73 (rotating transducer) instrument connected on-line to a Hewlett Packard 2116C computer with a moving-head disc operating system. The hardware and software of the interface existed along with some least squares limacon programs and programs for the other reference limacons using the same algorithms as in the Talydata accessory supplied by the company. These programs appeared to define the state-of-the-art at that time as summarised in sections 5.1 and 6.1.

The only modification to the Talyrond which has significance to this work is the attachment to the transducer carriage of an annular disc containing 512 equi-angularly spaced holes which were read photo-electrically to give the sampling positions for data-logging the profile. This system was designed to give a very good repeatability of position of samples relative to the spindle-axis. The absolute positional accuracy of the sampler was not of over-riding importance in its original application and was not tested for the current work in which, again, repreatability of sampling is the main requirement. The profile was taken from the instrument amplifier output as a signal of $\pm$1V, representing $\pm$25mm on the chart whereas the actual polar chart width of the instrument is 40mm. This was passed to a standard Hewlett Packard 10 bit successive approximation analogue to digital converter and thence to the computer. Other digital interfaces automatically monitored the instrument magnifica-

tion and other status lines. Apart from standard peripherals the computer had available an X-Y driven oscilloscope under program control upon which profiles could be plotted and the ability to draw profile and reference figures onto the polar chart of the Talyrond. The computer was equipped with 16K words of 16 bit memory of which about 11K was available to the user, the rest being concerned with the operating system. Any program which could be run on this system without needing direct use of the backing store for either data or program overlay would thus be of a size sufficiently small to be feasible for use in a 'stand-alone' measurement system. The programming language normally adopted was FORTRAN IV and generally single precision floating point arithmetic was used. This consisted of a 24 bit mantissa and an 8 bit exponent to give a precision of 6 to 7 decimal digits. Occasionally when absolute radius was involved directly in the computation double precision was used (giving 12 to 13 decimal digits), but with software floating point procedures these operations are expensive in both time and memory requirements and so were avoided whenever possible.

For the experiments needed in this study, the major requirement is the ability to compare the behaviour of different methods under identical conditions. A profile storage system is therefore required since repeated on-line measurements are bound to show small fluctuations. The data base constructed made direct use of the operating system file structure. This has several disadvantages caused by the relative crudeness of the operating system, for instance files cannot be created or extended dynamically and no read/write security can be operated. In the present circumstances these are not critically important for all roundness data files contain one revolution

of information (512 points on this system) plus similar status data and so all may be the same size. Additionally on a system to be accessed by only one or two people, the creation of the files on a removable disc cartridge used only for that purpose affords, with care, reasonable security for short term experimentation. The compensating advantage of this approach is that virtually the whole task of the data transfer is handled by the 'executive' and the software overhead to the user program is comparable in size to that of the on-line data-logging subroutine: both on-line and disc file based versions of programs can be developed simultaneously. The data files all consist of a continuous sequence of 512 16-bit words, each containing one ten bit profile point and up to 128 words of other relevant data including instrument magnification, the number of data points (so enabling incomplete profiles to be handled) and an alpha-numeric identifier of up to 72 characters length. Access to the files is by write and read subroutines ROWR and RORE which have calling sequences paralleling those of the data-logging routine.

Detailed description of all the programs developed for this system will not be given but extracts from the listings of some of the important programmes are given in Appendix 5, for illustration. Also particular points in the implementation of the algorithms are included in the discussion of the next section. As all were written in the same language by the same person they should be stylistically similar so that their relative performance will give some guidance to the relative efficiency of the algorithms themselves. A list, with brief descriptions, of the disc file programs used here follows:

RODD: Data logging from Talyrond 73 to disc file

DENT: Least squares limacon, including arcs, found by the 'traditional' formulae

DLS2: Least squares limacon, any data configuration, by direct solution of equations 5.9

DING: Ring limacon by exchange algorithm

DIPL: Ring or plug limacon by exchange algorithm and data inversion.

DEVS: Ring or plug limacon by direct use of revised simplex on the dual linear program

DINM: Version of DING allowing modification of data from console for "sensitivity tests".

DIN2: Ring limacon to 2nd order corrected data (see section 7.5)

DRIC: Finds and tests the circle constructed through the ring limacon contact points

DRC2: Interactive program, based upon DRIC, for finding the ring circle

DIMA: Minimum zone limacons by exchange algorithm

On-line revisions exist for some of these with names identical apart from the first letter: VENT, RING, RIPL, REVS, MIMA. Program VENT is a modified version of a program developed, by the author, in 1972/3.

9.3 Algorithm Implementation and General Operating Experience

All the programs were designed to work with equi-angular samples so that the profile data could be represented by simply the radial

value (after radius suppression) with the angle of each point being infered from its relative position in the data array. The only use made of angular information in any of the algorithms under consideration is in terms of either sine or cosine. This information was picked up from a simple sine/cosine look-up table using the same index variable used to recover the radial values from their array. If gaps are to be allowed in the profile, the same sampling scheme is followed but the radial points within gaps are set to an "impossible" value which is to be checked for and ignored in the ensuing calculation. This approach minimises the space needed for data storage and speeds the access time for the frequently needed sine and cosine values.

Implementation of both forms of least squares limacon programs is perfectly straightforward, in neither case were any special precautions found necessary. They are similar in size, the heart of DLS2 being a little shorter and less complicated than DENT but calling upon a subroutine for the solution of simultaneous linear equations (a Choleski method was used). Calculation times for a full circle of data were about three seconds for DENT and five for DLS2. The direct implementation is thus slightly larger and slightly slower than the "traditional" method. It is, however, much more general allowing any distribution of data rather than a single arc. It also has superior performance under some conditions, see section 5.4..

The ring limacon exchange algorithm was given a special starting procedure. The data is initially searched for its largest diameter $(r(\theta_i) + r(\theta_i+\pi))$ and the initial trial limacon taken to have centre at the mid point of this diameter and radius equal to half of it. The largest violation of the limacon by the profile then gives a third

contact which is bound to obey the 180^{o} rule and so the procedure given in section 6.4 is entered. If there were no violating point from this initial test, the procedure stops with a two point contact whereas the general limacon solutions would always search for a third contact (see the discussion of section 6.2). This process gives what might be regarded as the most logical solution, that of diametral contacts, in the situations for which a range of valid centre positions occurs. No special techniques were needed in the implementation of the algorithm but some care is needed in deciding whether a point violates the reference when using floating point calculations. Arithmetic rounding errors tend to cause the iteration procedure to oscillate about the final solution because of infinitesimal violations. This is easily overcome by comparing the difference between data and reference not to zero but to a small finite value larger than the rounding errors. A level of 0.1 of the data quantisation interval was used in the present programs.

The direct implementation of revised simplex on the dual linear program for ring limacon followed closely the standard methods but gained some increased efficiency by not using general purpose linear program subroutines. The solution was developed as code "in-line" with the data and sine/cosine table addressing. Again the only precaution needed was to protect against rounding errors causing oscillations in the iterations: here a discrimination level of just 0.001 of a quantisation interval was adequate for this purpose. No special start for finding potential diametral contacts was used. The iterations were started by using artificial variables which were then to be driven from the solution using the so-called Wagner method (see section 6.2). This is not a particularly efficient way of starting

this solution since the first fully feasible solution introduced on driving out the artificials relates not to the data but to the geometry of the measurement scheme. Instead of taking three iterations over removing artificials an initial feasible solution could be given to the program, in terms of a fixed basis, since in every solution of this implementation the measurement scheme is the same.

As would be expected of a more highly specialised routine the exchange algorithm version was rather smaller than the revised simplex version (about 60 lines of FORTRAN as opposed to 80) and also operated somewhat faster. Typically the exchange algorithm required only about 4 iterations to discover the solution taking perhaps 7 or 8 seconds. Including removal of artificial variables, revised simplex required about six iterations taking a little more than ten seconds. By using a special starting method instead of artificials variables, similar operating times would be expected from both methods. The nature of the operations required for the exchange algorithm is such that they are not performed efficiently in FORTRAN implementations. In ASSEMBLER language implementations a total exchange algorithm can readily be produced which occupies only 200 to 300 words of program. Since software floating point operations dominate the operational time, the change to ASSEMBLER does not significantly alter the operating speed, although with integer arithmetic a noticeable speeding up would be expected.

The comments about the ring gauge exchange algorithm apply also to the minimum zone exchange algorithm. An arbitrary start was taken by using the points at 0°, 90°, 180° and 270°. The solution takes about 45 lines of FORTRAN and operates typically in 10 to 12 seconds. For comparison the previously existing program took about 80 seconds and even then could only converge onto the solution from

much smaller initial eccentricities than could the exchange algorithm. The exchange algorithm has similar storage requirements to the earlier method. Again the use of ASSEMBLER improves the efficiency of implementing the exchange algorithm. It seems quite feasible that routines for all three boundary limacon references could be included in a "package" of about 500 words plus storage for data and sine/cosine tables.

Of the programs developed to study "improvements" to the ring limacon little need be said since they follow exactly the theoretical suggestions made earlier in this work. All use as their starting point the exchange algorithm solution for ring limacon, that is the end point of program DING. All also require an estimate of the true workpiece radius to be supplied by the operator. Program DIN2 calculates the second order term of the circle power series expansion based upon the eccentricity found for the limacon and subtracts this from the data, which is then reprocessed by the normal limacon exchange algorithm. This correction is calculated in single precision floating point since only division by the large radial value is involved. Program DRIC used the radius information to calculate the circle passing through the defining contacts of the data with the ring limacon. Here the radius is used intimately in the calculation and double precision arithmetic is needed. From the centre so calculated a reference line which is the circle expansion truncated after the second order term is compared to the data. This reference is used rather than the true circle since it can be calculated in single precision, with consequent savings, and has a precision adequate for use with the data and calculation performed elsewhere in the program. Program DRC2 is basically DRIC in which the operator

can specify the points through which the circle should be fitted. The program reports information on data points violating the reference and also on the relationship of the contacts to the centre in terms of the 180° rule. It is possible, therefore, for the operator to iterate, making intuitive jumps if necessary, towards the true ring circle fit.

10. Roundness References in Practice

10.1 Limacon References

Based upon the computer system described in chapter 9, a series of experiments have been performed to study the behaviour of roundness references under practical conditions. Generally these are pilot experiments performed on a small scale and are intended firstly to give an indication of reference behaviour and secondly to show the type of investigation which it is believed should be carried out on a large scale in order to build up a documented experience upon which the relative merits of approximations may be judged. For most of these experiments the ring gauge has been used because it gives a unique solution for all the reference figures under consideration. Some of the tests require data collected for specific conditions and will be described independently but for others a more general data base of "typical profiles" was appropriate.

A set of 100 profiles were taken and stored in disc-files from nominally circular objects which were readily to hand within the Research and Development Department of Rank Taylor Hobson. It is believed that these represent a range typical of that for which roundness instruments are most commonly used. The parts include elements of roller and ball bearings, finished shafts and holes and some plastic mouldings. Most commonly used materials are included. The profiles were data-logged with instrument set-up according to the following guideline. The workpiece was to be set such as to allow the use of a magnification sufficiently large for the out-of-roundness to be clearly visible on the chart, but beyond this requirement, no special care over centring was taken. In an attempt to

FILE	MAG'n	SYSTEM RESOLUTION	RADIUS (mm)	H or S	RING P-V	LIMACON E	γ
RD001	500	.1	6	S	11.32	2.37	.0003
RD002	5000	.01	9	S	6.18	2.93	.0003
RD003	5000	.01	36	S	5.15	1.82	.00005
RD007	5000	.01	44	S	6.09	0.19	.000004
RD010	10000	.005	22.5	S	1.52	0.18	.000008
RD013	500	.1	21	S	50.04	19.44	.00093
RD016	2000	.025	60	S	6.69	1.13	.00002
RD019	5000	.01	60	S	2.40	0.05	.000001
RD022	10000	.005	60	S	1.00	0.60	.00001
RD025	1000	.005	37.5	S	1.23	0.17	.000005
RD028	10000	.005	12.5	S	0.74	0.15	.00001
RD032	500	.1	12.5	S	23.60	3.88	.0003
RD035	10000	.005	16.5	S	1.98	0.52	.00003
RD039	2000	.025	2	S	9.85	1.62	.0016
RD040	1000	.05	2	S	9.73	18.63	.018
RD041	500	.1	2	S	9.28	44.65	.044
RD042	5000	.01	6.75	S	0.16	0.86	.0001
RD043	5000	.01	4.5	H	1.05	0.83	.0002
RD044	5000	.01	4.5	H	1.19	0.79	.0002
RD045	5000	.01	6.75	S	2.44	1.26	.0002
RD046	5000	.01	6.75	S	1.70	2.10	.0003
RD047	5000	.01	4.5	H	1.81	2.47	.0005
RD048	5000	.01	6.75	S	1.47	1.52	.0002
RD049	5000	.01	4.5	H	2.46	1.79	.0004
RD050	5000	.01	4.5	H	1.52	1.39	.0003
RD051	5000	.01	6.75	S	1.72	0.57	.00009
RD052	5000	.01	6.75	S	1.13	2.80	.0004
RD053	5000	.01	4.5	H	1.36	3.05	.0007
RD054	5000	.01	4.5	H	1.15	1.70	.0004
RD055	5000	.01	6.75	S	1.33	1.83	.0003
RD056	5000	.01	6.75	S	2.25	1.42	.0002
RD057	5000	.01	4.5	H	1.86	1.09	.0002
RD058	5000	.01	8.6	H	1.22	1.65	.0002
RD059	5000	.01	12.0	S	1.80	1.88	.0002

TABLE 10.1: Profiles in data base. Dimensions in μm except where stated.

RD060	1000	.05	4.76	H	33.83	11.65	.002
RD061	1000	.05	4.76	H	13.64	7.01	.001
RD062	500	.1	4.76	H	44.86	6.46	.001
RD063	1000	.05	4.76	H	42.98	2.06	.0004
RD065	1000	.05	2.5	H	15.39	5.61	.002
RD066	1000	.05	4.37	H	20.64	7.47	.0017
RD067	1000	.05	4.37	H	31.82	6.28	.0015
RD068	5000	.01	6.35	H	3.29	2.02	.0003
RD069	5000	.01	2.38	H	2.07	0.73	.0003
RD070	5000	.01	2.38	H	2.27	0.46	.0002
RD071	500	.1	4.1	H	40.24	4.46	.001
RD072	2000	.025	6.35	H	4.64	4.89	.0008
RD073	5000	.01	1	S	4.05	0.23	.0002
RD076	5000	.01	2.5	S	2.45	0.73	.0003
RD079	5000	.01	22	S	0.71	1.75	.00008
RD080	5000	.01	22	S	0.58	2.09	.0001
RD081	1000	.05	17.5	S	20.01	3.97	.0002
RD082	5000	.01	17.5	S	4.00	1.31	.00008
RD083	500	.1	14.5	H	40.85	11.46	.0008
RD084	2000	.025	3.9	H	5.58	2.50	.0006
RD085	2000	.025	3.9	H	4.84	4.21	.001
RD087	2000	.025	12.15	S	10.21	6.44	.0005
RD088	5000	.01	11	S	6.28	1.75	.0002
RD089	2000	.025	12.15	S	5.48	3.61	.003
RD090	2000	.025	12	H	10.99	4.21	.0004
RD091	2000	.025	15	S	6.11	3.23	.0002
RD092	20000	.0025	4	S	1.48	0.27	.00007
RD093	20000	.0025	4	S	1.29	0.25	.00006
RD094	20000	.0025	4	S	0.22	0.46	.0001
RD095	5000	.01	12	S	3.5	2.75	.0002
RD096	5000	.01	12	S	3.30	3.13	.0003
RD097	2000	.025	12	S	7.29	3.40	.0003
RD098	500	.1	6.25	S	85.43	15.58	.0024
RD099	2000	.025	1.5	S	11.64	1.38	.0009

TABLE 10.1: (continued)

RD100	10000	.005	2.25	S	3.86	0.56	.0002
RD101	10000	.005	2.25	S	2.46	0.20	.00009
RD102	5000	.01	12	S	4.42	2.90	.00002
RD103	20000	.0025	12.7	S	0.19	0.98	.00008
RD104	5000	.01	2.38	S	5.32	2.37	.001
RD105	5000	.01	3.18	S	4.38	1.26	.0004
RD106	5000	.01	3.98	S	7.05	3.73	.001
RD107	5000	.01	3.98	S	6.73	1.75	.0004
RD108	5000	.01	4.75	S	2.22	3.73	.0008
RD109	5000	.01	3	S	3.71	1.81	.0006
RD110	5000	.01	3	S	5.58	1.11	.0004
RD111	5000	.01	3	S	1.98	3.15	.001
RD112	5000	.01	3	S	3.57	3.30	.001
RD113	5000	.01	3	S	3.54	3.10	.001
RD114	5000	.01	3	S	3.62	0.91	.0003
RD115	2000	.025	5.6	S	10.67	8.60	.0015
RD116	5000	.01	5.6	S	2.58	0.73	.0001
RD117	5000	.01	5.6	S	4.44	0.99	.0002
RD119	5000	.01	3	S	3.18	1.84	.0006
RD120	5000	.01	4.75	S	7.53	1.40	.0003
RD121	5000	.01	2.75	S	4.96	0.57	.0002
RD122	5000	.01	2.75	S	5.58	0.90	.0003
RD123	5000	.01	4.75	S	6.08	0.64	.0001
RD124	5000	.01	4.75	S	9.18	4.51	.001
RD125	2000	.025	3.85	H	9.77	3.13	.0008
RD126	2000	.025	3.85	H	9.75	0.85	.0002
RD127	2000	.025	3.85	H	9.79	2.31	.0006
RD128	5000	.01	3.15	H	5.45	0.22	.00007
RD129	10000	.005	5.5	H	3.40	0.69	.0001
RD130	10000	.005	5.5	H	1.94	1.81	.0003
RD131	10000	.005	6.25	H	4.08	0.29	.00005
RD132	5000	.01	2.25	H	4.16	1.75	.0008

TABLE 10.1: (continued)

avoid bias, these measurments were not performed by the author. Table 10.1 gives a brief summary of the most important measurement conditions for this set, using minimum circumscribing limacons to assess eccentricity and out-of-roundness. Radii between 5mm and 100mm were measured and magnifications in the range 500 to 20000 used. The majority of parts had out-of-roundness values between 1 μm and 10 μm, which represents, perhaps, the most common range of use of roundness instruments. About one third of the profiles were from holes.

Other workers have reported comparisons of results from applying different reference circles to the same polar charts (see, for example, references 3 and 8). A natural experiment to perform is thus a similar comparison of the current data base using limacon references. This was performed using programs DENT, DIPL and DIMA The results were summarised in the same method as that used by the other workers: out-of-roundness is expressed in terms of the ratio of that for the reference being studied to the minimum zone value and variations between eccentricities of different references normalised as the difference of the centre from the minimum zone centre divided by the zone width. Figures 10.1, 10.2 and 10.3 show histograms of the distribution of the values so obtained with least squares, minimum circumscribing and maximum inscribing limacons relative to that minimum zone limacon. The general form of the histograms agrees quite well with the results quoted by others.

It appears that there is a distinct tendency for the least squares and minimum zone limacons to have separate identities, although large differences are to be expected only rarely. On average the least squares out-of-roundness is about 10% more than the minimum zone. The apparent tendency for there to be a 0.1 zone

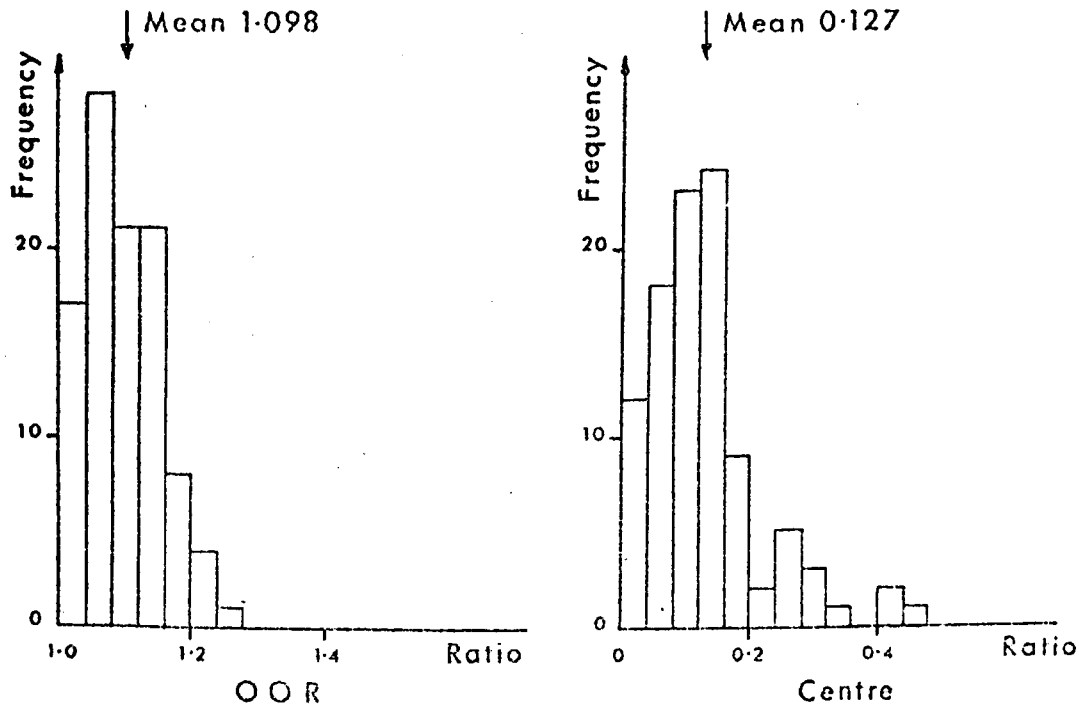


FIGURE 10.1: Comparison of Least Squares Limacons to Minimum Zone Limacons on Typical Parts

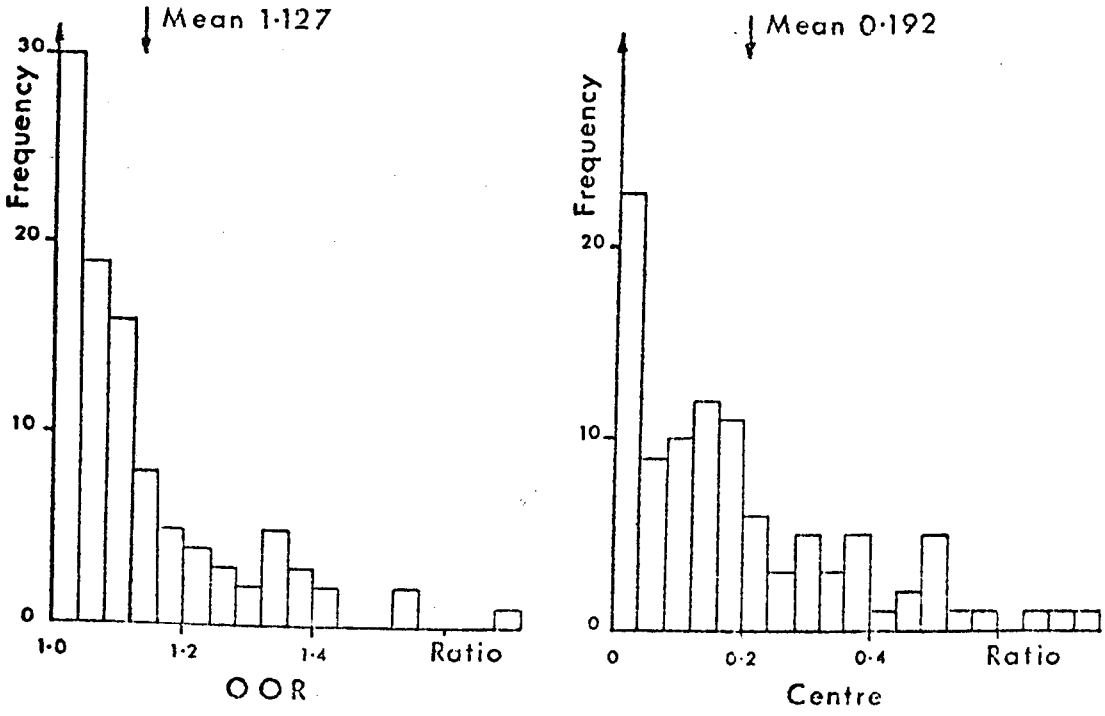


FIGURE 10.2: Comparison of Minimum Circumscribing Limacons to Minimum Zone Limacons on Typical Parts

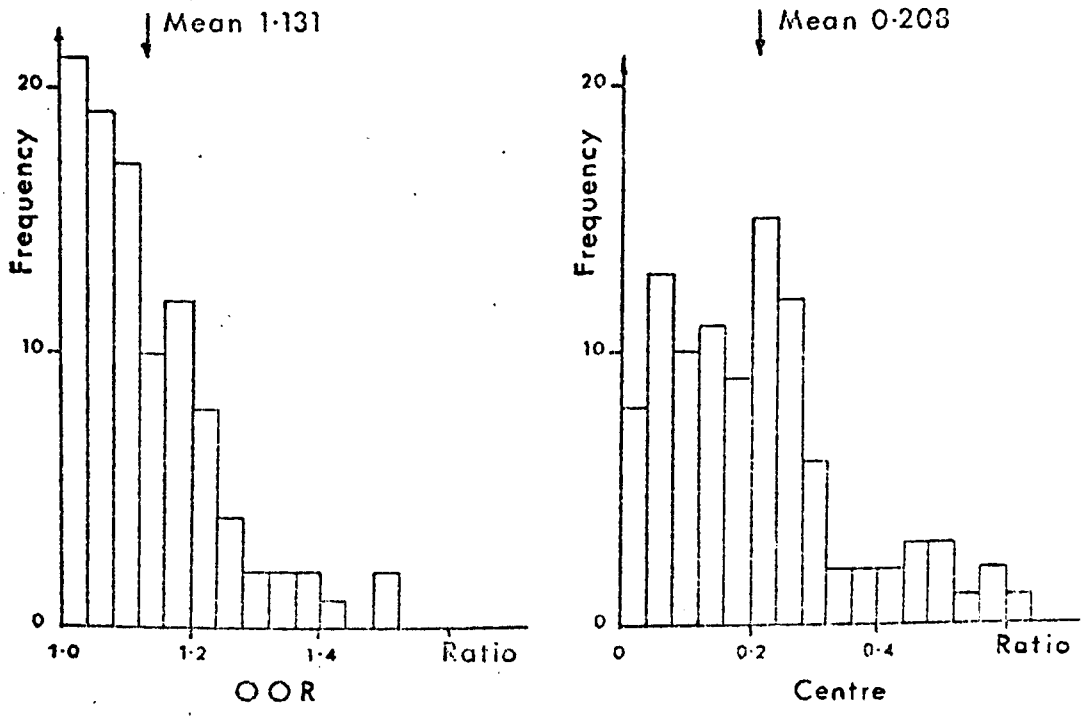


FIGURE 10.3: Comparison of Maximum Inscribing Limacons to Minimum Zone Limacons on Typical Parts

width separation of the centres does not have an obvious explanation. Both the ring and plug limacons are more likely to give values close to the minimum zone although the greater spread of results causes mean differences higher than those found with least squares. The parameters of the plug limacon seem to correlate with minimum zone rather less than do the other references. Again there is no clear evidence why this should be so. There was, for instance, no particular tendency for inscribing and minimum zone to agree on holes and circumscribing to agree with minimum zone on shafts. However the preponderance of shafts in the data base could be a contributor to the observed behaviour. The sensitivity of plug gauge to a single scratch on a shaft could be cited, for example.

10.2 Ring Limacons and Circles

By using programs DING, DIN2, DRIC and DRC2 it is possible to investigate in some detail the behaviour of circumscribing references over the set of typical profiles contained in the data base. The standard comparison of references should be to the ring circle fitted in instrument co-ordinates. This may be found using the interactive iterative procedure in DRC2 based upon an estimate (or nominal value) of the true radius. The first concern therefore is the sensitivity of this solution to the radius estimate supplied.

A check on sensitivity was performed by testing a component with nominally repeated measurement tracks for well centred and quite eccentric conditions for parameter variation when different radius values were given to DRC2. The part chosen had a nominal radius of 12mm and an out-of-roundness of about 3μm. It was measured at x500 magnification. With the centred profile (residual eccentricity about

7 μm) no parameter change of even 0.1 of a quantisation level was detected when using radius estimates of 13mm, 12mm, 11mm and 5mm! With an eccentricity of about 45 μm the same behaviour was found; a little more variation occured with the estimate of 5mm, the worst change being 0.5 of the quantisation interval. On the eccentric data an additional check using a radius estimate of 25mm was performed and again no variation greater than 0.2 quantisation levels was found. Since radial estimates to better than $\pm$1mm are easily obtained no difficulties from sensitivity are anticipated.

The use of program DRIC identifies whether the circle through the ring limacons contact points is indeed the ring circle. From the 100 profiles there were 15 cases in which violations of the circle by the data were reported. However in 8 cases this was merely arithmetic rounding error, the violating point being one of the contacts. The most serious violation was by an amount of one sixteenth of a quantisation interval. For the seven cases where violating points really existed just a single iteration in program DRC2 yielded the true ring circle. On one profile this resulted in a change of parameter value of 0.02 μm, on two others changes of 0.01 μm were found and on the others no change (at the level of 0.01 μm) was found.

Using a slightly reduced set of profiles (94 from the data base) the variation of centre position for ring limacons and circles was examined. Comparisons were performed on eccentricity (A and B) values rounded to 0.01 μm. In 70 cases there were no differences then detected between the centre of the limacon and the centre of the circle through the limacon contacts. In all other cases the differences were of 0.01 μm (only once did the error occur in both axes for the same profile). Thus in no case was a difference which could be

regarded as significant obtained. For the seven profiles for which the ring circle was distinct from the circle through the contacts the true ring circle again had a centre indistinguishably close to the limacon centre except in one case. For this one profile (RD108) a difference of slightly more than 0.02 μm was found which is possibly just significant. It is still a difference of less than 1% of the eccentricity value.

It is fair to conclude that, since these figures are small compared to typical roundness instrument performance specification, in instrument co-ordinates, the ring limacon adequately estimates the ring circle under normal measurement conditions. The use of refinements such as the second order data correction of program DIN2 are clearly superfluous here. Note that the eccentricity ratio (see Table 10.1) rarely exceeds 0.001 in this set of profiles.

The suite of programs may also be used to study typical cases of fitting reference figures on the polar chart. This is done by specifying not the true radius but a value of 50mm divided by the magnification. The program still operates in "instrument co-ordinates" but the false conditions it has been given make this equivalent to a chart co-ordinate frame in which a nominal chart radius of 50mm is used. Operating program DRIC in this manner reveals that on 25 profiles violations of the circle through the limacon contact points were found. Of these only one was attributable to arithmetic error and 23 have violations larger than the quantisation interval. On other profiles, program DRC2 indicated that although the circle through the contacts circumscribed the profile, those contacts did not obey the 180° rule with respect to its centre. The whole set of profiles were therefore checked and attempts made to iterate

them using DRC2 with the following results. In 41 cases out of the 100 the limacon contact points did not define the minimum circumscribing circle on the chart. Of these, in 29 cases the points disobeyed the 180° rule and in 24 the circle did not circumscribe the data. More ominous was the fact that on 15 occasions the attempt to iterate to the ring circle was abandonned in frustration. On at least a few occasions this appeared to be caused by the presence of diamteral contacts to the ring circle, a case difficult to include within the iterative procdure being adopted. It seems fair to observe, however, that since the author has a considerable background in the theory underlying the process, if he failed to make progress, then most operators might be similarly frustrated.

Since in instrument co-ordinates there was never a significant difference between limacon and circle, the limacon fitted onto the chart profile is the figure representing the truly desired reference. Thus the results just quoted can be rephrased as: in 41% of the tests the minimum circumscribing circle to the chart profile did not represent the ring gauge circle to the component. This divergence is a reflection of the much higher eccentricity ratios which occur on the chart. Of the 100 profiles, 99 had a chart eccentricity ratio greater than 0.01, 61 were greater than 0.1 and 24 exceeded 0.2. Reason's recommendations for the control of centring error based on chart eccentricity ratios of 15% or 7% for high accuracy work (3) were exceeded on 42 and 78 occasions respectively. It is clear that these criteria are much more severe than is necessary to control errors when using limacon references. They are once more confirmed as being about right for controlling the fitting of circles on the chart. Further it is seen that quite commonly the operator

will accept as "not too bad" a degree of eccentricity which is not adequate for using circles on the chart.

This experiment involving measurement of contact positions led to two other pieces of information which may be of interest. In figure 10.4 is shown the relative occurrence of the angles subtended at the origin by the ring limacon contact points grouped according to the largest, smallest and intermediate angles occurring for each profile. These results are taken over a subset of the data base using the last 84 profiles from table 10.1. Some of the features are caused by the form of grouping the smallest angle must be less than 120° and the largest greater than that but less than 180° (from the 180° rule!). The intermediate angle is similarly constrained to always exceed 90°. Two features stand out. Firstly there is a very good probablity that a measurement will yield a contact angle not much less than 180° while there is a significant but lesser chance of getting a very small angle. It is however more likely that a small angle will be close to zero than of the order of 20°. This would seem to indicate a tendency towards there being nearly diametral contacts on many occasions with the limacons third contact being found with reasonably even likelihood around the periphery. The set of profiles contains many having an important oval component although not often does ovality alone dominate the measurement. The second remarkable feature is the near absence of largest and smallest angles near to 120°. The occurrence of three more or less equally spaced contacts seems to be quite rare in practice. Just one profile showed almost equal angles. no other having a variation of less than 10° between the contact angles.

Since all the collected profiles come from independently set up

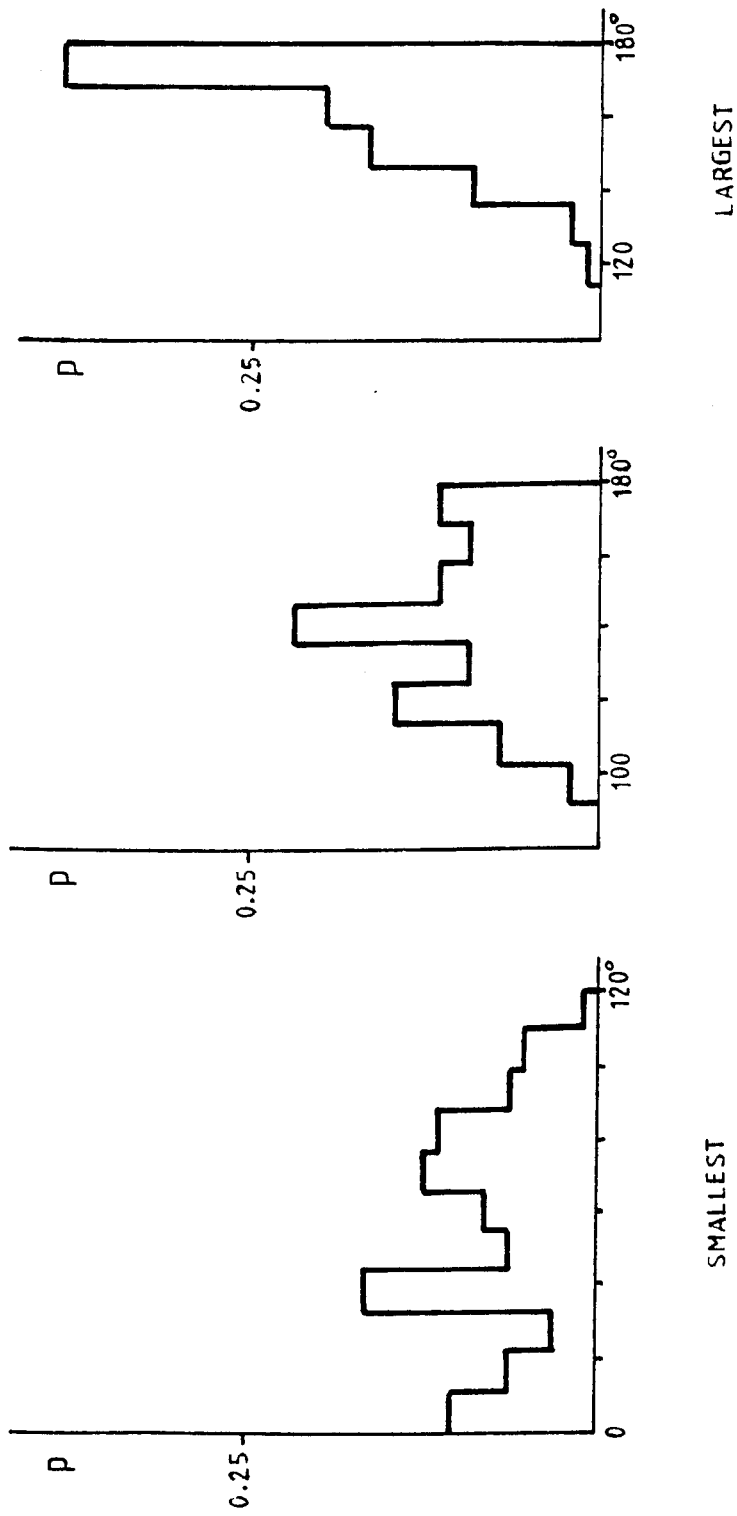


FIGURE 10.4: Distributions of Angles Subtended at Origin by the Contact Points between Profiles and their Limacons

measurements the direction of eccentricity would be expected to be random. In practice this is only partly bourne out. In the profiles of this data base there appeared to be a rather low occurrence of eccentricities in the direction towards the rear of the instrument, particularly so in the second quadrant. More evidence would be needed to show whether this variation is statistically important. If so it appears that the explanation lies in the behaviour of the operator rather than the instrument as such, perhaps by tending to use the controls in a particular sequence when centring. This was not investigated here as it does not seem to be relevant, but there may be experiments in which such an unexpected source of systematic error could be important.

10.3 Profiles Under Differing Eccentricities

In the previous section results concerned the general behaviour of profiles from many workpieces each measured under conditions judged to be reasonable. An alternative study is to consider what happens with profiles taken from the same workpiece at different eccentricities. If only the eccentricity is adjusted between a series of measurements the profiles will be nominally from the same track around the workpiece. Comparison of these profiles is a test of the complete measurement system, not just of that part concerned with establishing reference figures.

Measurements were made on eight workpieces of various roundness quality: six lying in the most common range of use, about 1 to 7 μm peak to valley, the other two being much more extreme with out-of-roundness of about 25 μm and 50 μm. Each was measured "well centred"

on the highest appropriate instrument magnification and then at two, or sometimes three, stages of decentring using lower magnifications as necessary. "Well centred" is a term having considerable vagueness: in practice it ranges from a fraction of a μm on the very round parts to a few μm on the "worst" profiles. Certainly a figure of the order of 1 μm is being considered. The largest eccentricity in each set was also not specified closely. Using programs DING, DIN2 and DRIC each set was analysed for variability of ring gauge reference figure. The largest eccentricity ratio in the set is 0.0033 so all readings fit within the compass of those for which limacon models are considered good. Only the well centred profiles satisfy the circle on chart criteria based on chart eccentricity. As with much work of this type the full results obtained consist of tables of nearly identical numbers which are difficult to assimilate and virtually impossible to summarise graphically. The figures will not be given here. Using ring gauge references based upon the limacon, the limacon fit to 2nd order corrected data and the circle through the limacon contacts, the parameters and out-of-roundness measured for each individual profile never varied by as much as the quantisation interval. On just one occasion, that with the largest eccentricity ratio, the second order correction caused a single change of contact point. The circle through the contact points was never violated by more than a fraction of a quantisation interval. These results do little more than add weight to the conclusions of the previous section concerning the adequacy of the limacon approximation.

Of the eight sets, there were two for which changes of peak to valley greater than 20% were found over the range of eccentricities. Other changes were not more than 5% and sometimes totally absent.

In five cases the orientation of the contact points varied markedly as the eccentricity was altered. There were also minor alterations such as a contact moving from one sample position to an adjacent one. Some of these changes are undoubtedly due to slight variations in the nominally repeated profiles: the orientation of a surface feature to the origin will vary with the eccentricity. The grosser effects observed are not likely to be caused just by these variations. Given the consistency of the reference fits the measurements can be taken as genuine reflection of changes in the true ring gauge out-of-roundness. The instrument has 'seen' different profiles on different runs. It is probably significant that the largest effects are observed on profiles having a "spiky" nature.

Although the tests just described give some indications of the effects of eccentricity as they might occur in practice, the number of inaccurately known variables in the instrumental system make it very difficult to do more than a qualitative comparison. As a particular example, there is no accurate check on whether the eccentricity measured corresponds to that of the workpieces. A better test is to arrange an independent method of placing the component at different points within the working range of the instrument while leaving the instrument itself undisturbed throughout a series of measurements. In case there might be a favourable direction of eccentricity for the system, a scheme in which the component is moved to all positions over a regular matrix covering the working range is appropriate. The only results available for reporting here from such an experiment relate to a pilot scheme intended originally just as a means of testing the concepts and mechanics of the experimental procedure. A very crude movement

device based on the standard instrument accessories was used to generate the matrix of positions: a more kinematic system should be used in a full experiment of this type. The Talyrond 73 centring and levelling table gives a small flat worktable which can be set very precisely perpendicular to the spindle axis by using the instrument itself with a right-angled attachment to its transducer and which can therefore be used as a reference plane for moving the testpiece. This table also has a shallow plate with a 90° V and a spring plunger assembly running in a T-slot for locating a workpiece. With the V-plate locked in one position the workpiece could be moved by packing the edges of the V with slip gauges. It is necessary to preserve the angular orientation of the test piece with respect to the spindle and so it was not mounted directly onto the worktable. Instead a glass target from an alignment telescope system was used as a carrier. This component is a glass disc about 18mn thick having two flat, parallel faces so that it may be slid across the worktable and its top face remain perpendicular to the spindle. Its other feature consists of two location flats on its edges which are 90° apart and which give a natural lock to the orientation of the disc when used to bear against the slip gauges. The arrangement is shown schematically in figure 10.5. For these first tests no physical constraint was placed on the positioning of the slip gauges along the sides of the V-plate. Pencil marks were used as a guide to avoid gross errors. The test piece was mounted on the glass flat and secured with plasticine. With the combinations of slip gauges corresponding to the centre of the matrix of positions being checked, the instrument was set to give a well centred profile. A magnification suitable for handling the largest eccentricity of the set was

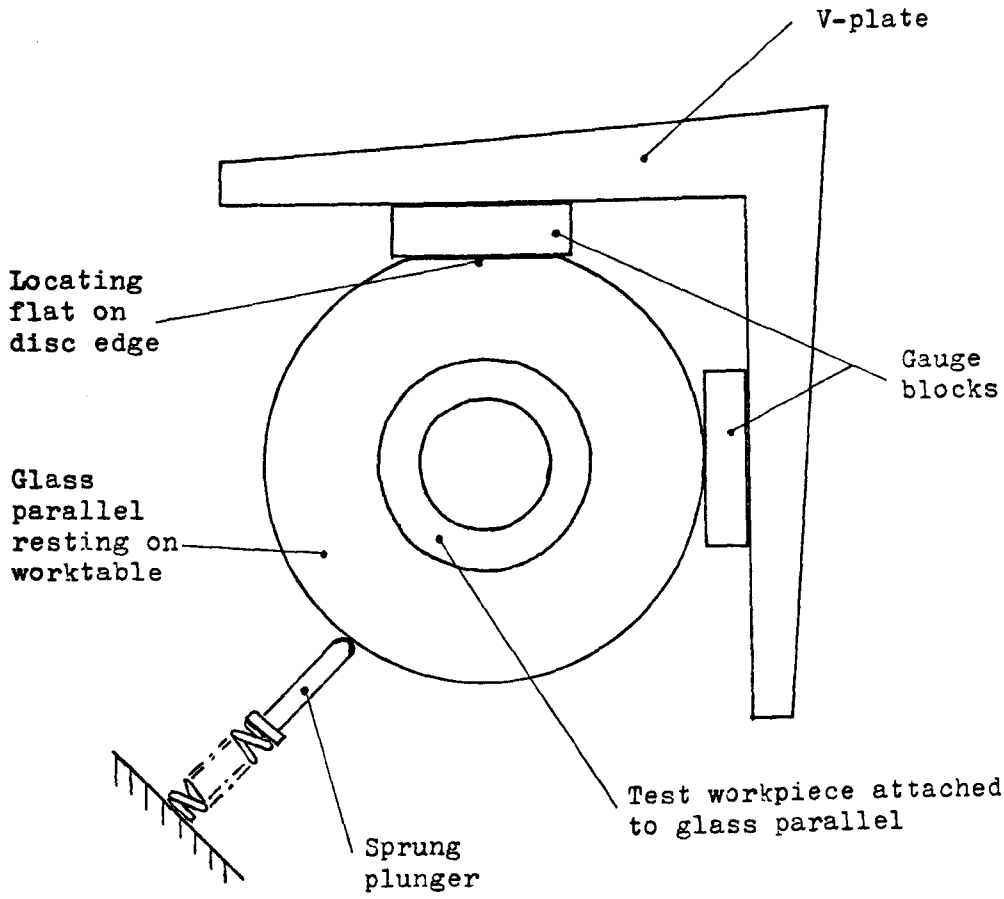


FIGURE 10.5: Concept of Mechanical Layout for the "Matrix Experiment"

then selected after which no further adjustment of the instrument controls was made throughout the sequence of tests. A matrix giving 25 positions being all combinations of 0, $\pm$12.7 and $\pm$25.4 μm in both x and y directions was tested (Imperial gauges were being used). Two distinct groups of slip gauges were used to generate the x and y steps and the combinations were selected to give minimum errors according to their calibration certificate. (In fact their accuracy is so much superior to that of the experimental relocation that this finesse was hardly necessary). The experimental order started with the measurement at both x and y eccentricities of +25 μm and then proceeded, varying x more rapidly, in a zig-zag pattern to finish at -25 μm in both cases.

An independent test was made of the relocation repeatability which could be expected. On repeated runs of the same profile without disturbing the system in between a variation of about $\pm$0.02 μm was obtained on the centre position. Removing and replacing the same slip gauges between runs degraded this repeatability. Without using even pencil marks to guide the relocation a set of seven repeats caused a total spread in centre position of 2.5 μm in the y-direction but about 6 μm in the x-direction. With care and the use of guide marks repeatability was generally within $\pm$1 μm.

The test piece was the ground edge (not the bearing track) of a ball race inner having a radius of 12mm. The magnification used was x500, giving a resolution of about 0.1 μm in the profile data. The 25 profiles selected were measured against all four reference limacons. The behaviour of all four was remarkably similar and so only the ring gauge, for which other information was also collected, will be discussed here. Table 10.2 shows the centre positions, rela-

Centre positions relative to middle test

for MINIMUM CIRCUMSCRIBING LIMACON

Matrix points should be nominal 5.10^{-4}ins = 12.7μm

Talyrond magnification = 500, resolution 0.1μm

Values in μm

						$\bar{B}$
	05	04	03	02	01	s_B
	-17.91	-7.91	2.53	15.51	30.64	21.96
	22.26	22.56	22.04	21.40	21.56	0.48
	06	07	08	09	10	
	-17.62	-7.06	5.91	17.67	29.77	10.56
	8.28	10.43	11.34	11.36	11.38	1.34
	15	14	13	12	11	
	-20.75	-12.21	A = 0	14.50	29.39	-0.05
	0.83	0.60	B = 0	-0.45	-1.23	0.83
	16	17	18	19	20	
	-20.55	-9.51	0.24	12.81	25.98	-13.39
	-11.33	-11.27	-16.35	-14.90	-13.10	2.23
	25	24	23	22	21	
	-22.48	-10.38	2.38	14.51	27.32	-22.14
	-21.57	-21.59	-21.87	-22.55	-23.11	0.67
$\bar{A}$	-19.86	-9.41	2.21	15.00	28.68	
s_A	2.06	2.03	2.38	1.78	1.91	

Table 10.2 Matrix Experiment

5	4	3	2	1
3.24 3.25	3.13 3.15	3.08 3.14	3.18 3.19	3.22 3.22
6	7	8	9	10
3.01 3.00	3.11 3.11	3.17 3.18	3.15 3.15	2.93 2.92
15	14	13	12	11
3.13 3.12	3.09 3.08	3.19 3.19	3.18 3.17	3.24 3.21
16	17	18	19	20
3.09 3.07	3.23 3.23	3.00 3.00	3.03 3.03	3.14 3.09
25	24	23	22	21
3.11 3.11	2.99 3.01	2.92 2.92	3.16 3.14	3.19 3.18

LIMACON: Mean = 3.12, s.d. = 0.09, spread = 0.32

CIRCLE: Mean = 3.11, s.d. = 0.09, spread = 0.33

Matrix Experiment

Peak to Valley out-of-roundness relative to ring limacon (upper figure) and ring circle (lower figure). All values in μm.

Table 10.3

tive to the middle obtained, in a layout representing the position of the result within the matrix of measurements. Also shown are the mean and standard deviations of the 'constant' term along each rank and file. The standard deviations show again a poorer repeatability in the x direction than in y. In the y direction the repeatability is generally of the order of $\pm 1\ \mu m$ as expected. Table 10.3 gives in similar layout the peak to valley out-of-roundness relative to the ring limacon and the circle through its contact points (here taken as the ring circle). There is very little variation either between positions or between references. Note that the standard deviation of the whole set with either reference is very close to the quantisation interval. Confirmation that the spread could be caused by the 1-bit uncertainty of data logging was made by means of a sensitivity analysis using program DINM. Using profiles from this set variations greater than the total spread of peak to valley occuring in the experiment, could be obtained by 1-bit changes to the amplitude of the contact points. The contacts on the testpiece were well spread: the two smaller enclosed angles being about 85° a and 120°. Other than switching between the contact actually being registered on one of adjacent samples, they did not vary in position throughout the whole set.

The rather large variations from the expected centre positions were further analysed in an attempt to discover their cause. Using the x-ordinates a least squares plane fit was performed and the hypothesis that this plane had linear relationship with x and no y or constant terms tested by means of the Student-t. This showed that the variations from expected values were not statistically significant and so are probably attributable to experimental

inaccuracies rather than specific trends. However taking the ordinates in the measurement order there is some evidence of serial correlation. The numbers are rather smaller than would be desired, but if a zero crossing test is accepted serial correlation is significant at the 99% confidence limit. This casts a little doubt over the analysis.

Overall this experiment serves only to confirm that the algorithms are better than the physical experimental method. It would certainly be of interest to do further work along the lines of this test using a more accurate procedure but the impression is that the limacon references will not be tested "to destruction" by normal instrumental techniques.

10.4 Measurements at Large Eccentricity Ratios

One set of conditions where current, conventional instrumentation can supply high eccentricity ratios in instrument co-ordinates is when near maximum amounts of eccentricity occur on very small parts. This area of measurement may have some importance in the future since the trend towards miniaturisation could call for the automatic or semi-automatic measurement of small components which are difficult to handle and place precisely. By equipping the Talyrond with a special small stylus (made from a sapphire gramophone pick-up) it was possible to track repeatably around pins of radius about 0.5mm at eccentricities of more than 150 μm. Instrument co-ordinates eccentricity ratios of greater than 0.3 were obtained. These readings were taken with a magnification of x100 which shows practically the behaviour pattern discussed in section 3.2 and figure

True Radius	Reference	A μm	B μm	P-V μm *
0.432mm	L	148.5	-40.6	28.9
	L2	150.6	-45.4	16.1
γ = 0.34	CL	149.2	-34.0	+8.2, -15.7
	CI	145.2	-34.8	22.5
	CC	147.0	-36.4	19.9
0.508mm	L	162.4	49.1	28.5
	L2	162.0	47.3	9.1
γ = 0.33	CL	162.7	48.2	+4.5, -3.6
	CI	160.3	46.1	10.8
	CC	160.0	46.3	11.1
0.517mm	L	169.7	11.4	34.1
	L2	168.2	6.3	6.8
γ = 0.33	CL	169.7	8.1	+3.8, -5.2
	CI	168.7	5.8	5.9
	CC	168.5	5.6	6.6
1.753mm	L	155.0	-79.2	9.8
	L2	154.8	-79.2	2.1
γ = 0.10	CL	155.0	-79.2	+0.5, -1.5
	CI	154.8	-79.2	2.1
	CC	147.3	-75.9	28.2

* where a reference circle is violated, Peak and Valley are given separately

TABLE 10.4: Ring Gauge at High Eccentricity, low Radius. See text for abbreviations.

3.2. With a 50mm chart radius the profiles are not radius suppressed on the chart. Table 10.4 shows the data obtained from three pins of about 0.5mm radius and one rather larger part for comparison using the various ring reference criteria discussed throughout this work. In the table the following abbreviations are used: L limacon; L2 limacon on 2nd order 'corrected' data, program DIN2; CL circle through limacon contacts, program DRIC; CI and CC ring circles in instrument and chart co-ordinates respectively, program DRC2. The quantisation interval for these results is about 0.5 μm.

Taking always the circle in instrument co-ordinates as the true measurement it is immediately obvious that the limacon model has broken down. The peak to valley is always much too large and the circle through its contacts is consistently violated by the profile. Generally the second order correction works well. However on parts of about 0.5mm radius the circle on the chart gives the nearest value for out-of-roundness. On the larger radius part, however, the chart circle has become a very poor reference while with the slight reduction of eccentricity ratio the second order correction is virtually exact. An interesting feature of these readings is the remarkable similarity of the eccentricity of all the references: even when giving wildly inaccurate peak to valley readings the divergence of centres is hardly significant.

This demonstration is very encouraging for future development in roundness measurement. It has been demonstrated that the data "correction" approach is a sound practical weapon for use when the limacon reference alone ceases to be sufficiently accurate.

11. The Assessment of Circular Cylinders

11.1 Cylindricity Measurement

Although the term "cylindricity" is quite widely used and embodies a generally recognised concept, namely how does the surface of a workpiece vary from a perfect cylinder, there is no generally accepted specific definition. There is no metrology standardisation on cylinder measurement although codes of drawing office practice define methods of tolerancing deviations from cylindricity by insisting that the real surface lies between two co-axial perfect cylinders. This agrees with the definition given by Reason(14) who states that cylindricity errors are conceived as lying between two co-axial cylinders, figure 11.1. The exact nature of these cylinders with respect to the profile is not stated but the context suggests that they might obey a minimum zone or possibly least squares relationship to it. This is consistent with the generally prevailing view that cylindricity is measured as an extension of roundness and so the same analytical methods are to be expected. This method has more relevance than normal tolerancing approaches based upon totally predefined ideal forms if, as seems likely in practice, cylindricity is to be measured on radius suppressing instruments.

There has been, in virtually all work on cylindricity measurement, an inherent assumption that roundness instruments having a straight datum parallel to their spindle axes would be used and this has greatly influenced the approach to the analysis. It builds directly upon the methods and assumptions of roundness analysis without expressing any concern regarding the validity of doing so.

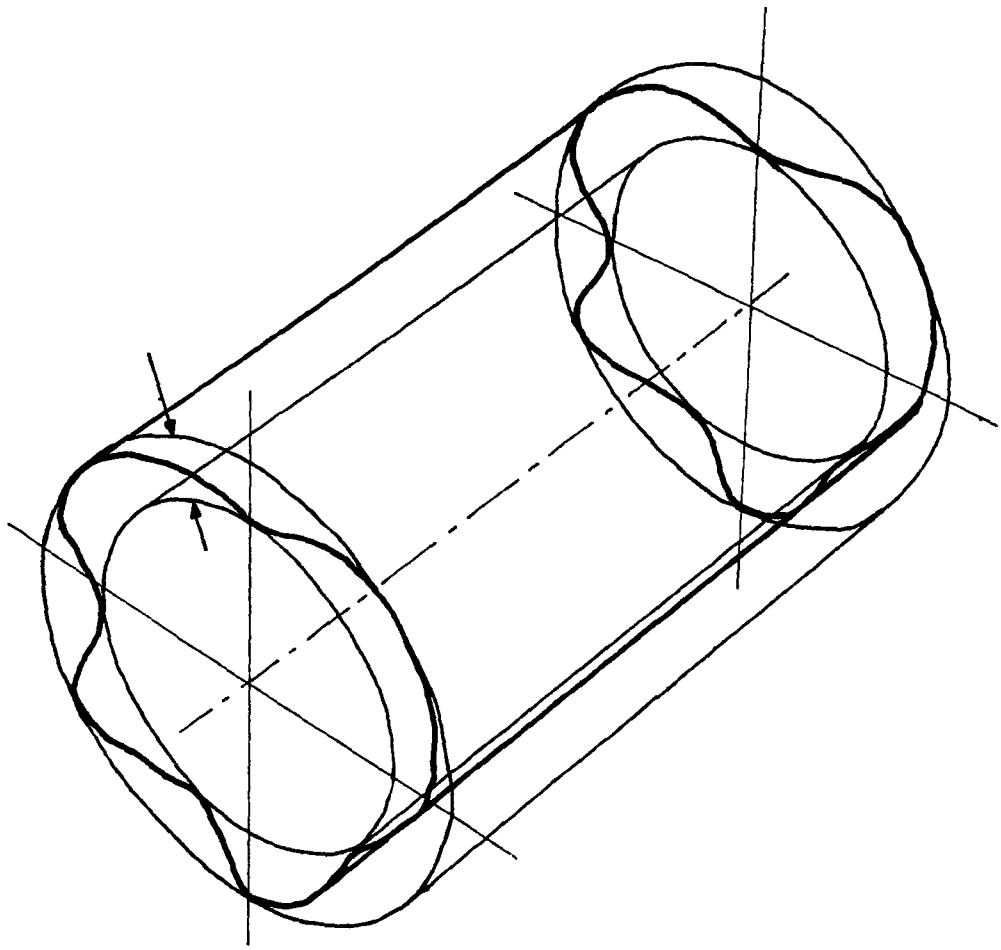


FIGURE 11.1: Commonly Accepted Definition of Cylindricity

Although the ANSI roundness standard(1) does not discuss cylindricity it does observe that the axis of a workpiece is defined by the centres of reference circles at a series of cross-sections and that, therefore, different axes are to be expected according to which reference criterion is used. It does not recommend a method for assessing the lack of straightness of this axis. This type of measurement for axial straightness bears only an indirect relationship to cylindricity and it is important that their separate identities be recognised. An axial straightness measurement can give out-of-roundness information at each cross-section but cannot assess an overall "out-of-cylindrical" value by direct means. A cylindricity measurement would create an overall reference cylinder from which "out-of-cylindrical" information may be directly taken and will specify a direction for the best fit straight axis but does not give an immediate summary of, for instance, the axial straightness. The complete description of the error of form of a cylinder will therefore require both of these analyses to be applied, that is both overall and cross-sectional references seem to be necessary. Further evidence for there being a fair degree of independence between these measurements comes, for instance, from Lotmar's work (15) measuring the radial variation along a ring gauge, at the expense of losing all information about the axis, on a roundness instrument not equipped for cylindrical measurement.

One proposed method for defining cylindricity which does not rely, at least at first sight, on the usual reference figures is due to Iizuka and Goto (16, 17) when a deformed cylinder is described in terms of an axis consisting of orthogonal polynomials in z and cross-sections of constant z are described as Fourier series using r and θ ; r, θ , z being a cylindrical co-ordinate system. It seems that this method was first adopted more to demonstrate the use of

least-squares by which the parameters can be found than for metrological reasons but it is a method of describing the surface of which more use could probably be made. Harmonic analysis in roundness measurement has been variously proposed, see, for example (32) and (33) and the extension to a polynomial axis seems natural. It is, however, still restricted by the need to interpret part of the "profile error" as caused by residual misalignment of the workpiece to the instrument co-ordinate system. To do this they allow that the first harmonic of each cross-section and the linear polynomial along the axis are wholly, and the only, terms caused by misalignment. Thus the limacon approximation is being applied at each cross-section to account for eccentricity there and the least squares straight line through the centres of these limacons is taken as the tilt error between workpiece and instrument.

Other workers have implicitly assumed the use of a "reference cylinder" which is in fact a limacon on each cross-section perpendicular to the z-axis with the centres of these limacons lying on a straight line. This is true even of methods which do not actually measure such cross-sections, such as schemes using a helical trace around the workpiece (18). Virtually all reported work is concerned with least squares methods. One partial exception is an attempt to discover the minimum zone cylinders (11) from the least squares solution. A search method is given but considered to be too inefficient and an alternative is proposed which uses a weighted least squares approach in which the weights relate to the residuals of an unweighted least squares solution so that the major peaks and valleys are emphasised. This method is of course an estimation of the minimum zone cylinders rather than a solution. It still relies upon the

validity of the limacon approximation at every cross-section.

Here the measurement of cylindricity will be examined from the viewpoint that it will be required to produce extensions of the roundness standards and the methods are required for the solution of least squares, minimum zone, minimum circumscribing and maximum inscribing cylinders in instrument co-ordinates. The methods and philosophy used are similar to those adopted here for roundness analysis and given detailed treatment in earlier chapters; the cylinder analysis given will be brief.

11.2 Reference Figures for Cylinder Measurement

None of the literature describing work on the measurement of cylinders makes use of cylindrical reference figures. Nearly always the same implicit assumption is made, namely that the cross-sectional shape in a given plane is unaltered as the workpiece alignment is altered. The reason for this constancy of approach probably arises from the nature of the instruments used in the measurement. In effect they produce profiles representing sections of a cylinder on planes perpendicular to the z-axis of the instrument co-ordinate frame. The cylinder is represented by a series of circles of the same radius placed perpendicular to the z-axis and having their centres lying on a straight line. In practice these circles are almost inevitably approximated by limacons. As the distinction of these different forms can be of some importance, a distinct terminology will be adopted here. "Cylinder" will be reserved strictly for describing a figure in which all cross-sections perpendicular to its axis are identical with respect to that axis. Unless specifically

stated otherwise a right circular cylinder is implied. Other cylinder-like figures which do, however, have a different geometry will be called "cylindroids". Distinction is also made between "tilt" in which all points of a figure are rotated by the same amount relative to the co-ordinate system and "skew" in which the axis of the figure is so rotated but the cross-sections remain parallel to their original positions. Figure 11.2 illustrates the difference between these by showing a tilted cylinder and a skew circular cylindroid. A skew circular cylindroid may also be described as a scalene cylinder, the former notation being preferred here because of it is more descriptive of the physical situation being studied. In the case of a cylindroid, shape is described in terms of the cross-section parallel to the unskewed axis. The reference figure used commonly in cylinder measurement is then a skew limacon cylindroid. It should be noted that since the axis is skewed, the eccentricity at different heights will vary and so the skew limacon cylindroid does not have a constant cross-sectional shape. It does have constant geometrical properties on its cross-sections.

An investigation of reference figures suitable for measuring cylindricity must start from a statement of the form of a true cylinder oriented arbitrarily in the space described by a set of instrument co-ordinates. The circular cylindrical surface is defined by the property that all its points have the same perpendicular distance (radius) from a straight line (the axis). This is conveniently described using direction cosines and a vector notation. The axis is fully defined by a set of direction cosines $\underline{l}_1$ and a point $\underline{X}_o$ through which it passes. The perpendicular distance of a

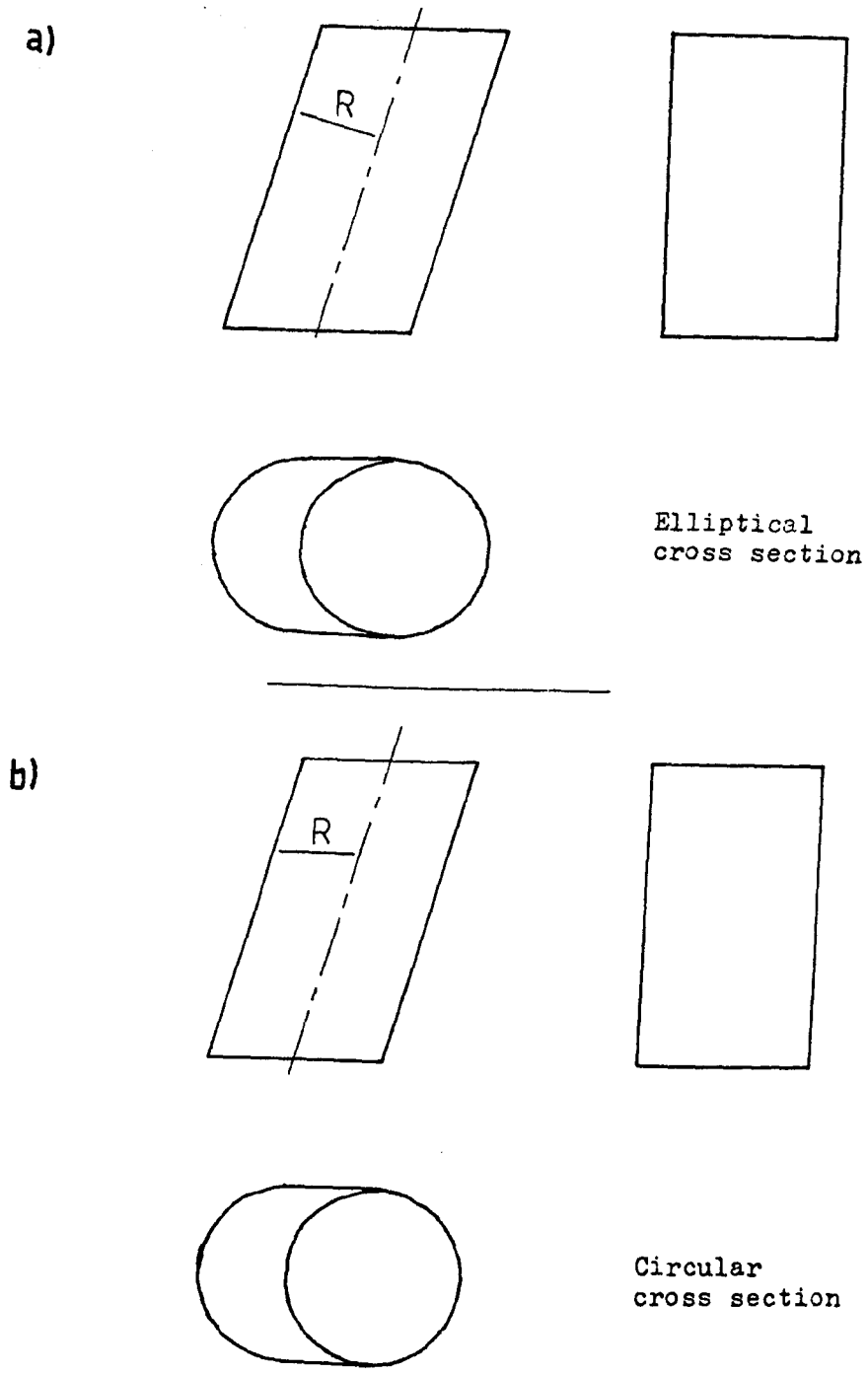


FIGURE 11.2: Comparison of a) Tilted Circular Cylinder
and b) Skewed Circular Cylindroid

general point $\underline{X}$ from this line is given by:

$$p = |\underline{X} - \underline{X}_o| \sin\alpha \qquad 11.1$$

where α is the angle between the axis and the line joining $\underline{X}$ to $\underline{X}_o$. The direction cosines $\underline{l}$ of this joining line will be:

$$\underline{l} = \frac{\underline{X} - \underline{X}_o}{[(\underline{X} - \underline{X}_o)^T(\underline{X} - \underline{X}_o)]^{\frac{1}{2}}} \qquad 11.2$$

and the angle α is then found:

$$\cos\alpha = \underline{l}_1^T\underline{l} \qquad 11.3$$

Substituting equations 11.2 and 11.3 into 11.1 and working, for convenience, with p^2 since p is a scalar length gives:

$$p^2 = (\underline{X} - \underline{X}_o)^T(\underline{X} - \underline{X}_o) - ((\underline{X} - \underline{X}_o)^T\underline{l}_1)^2$$

To define the cylinder, all points having $p = R_o$ are required, so a complete description is:

$$R_o^2 = (\underline{X} - \underline{X}_o)^T(\underline{I}_3 - \underline{l}_1\underline{l}_1^T)(\underline{X} - \underline{X}_o) \qquad 11.4$$

where $\underline{I}_3$ is the 3-square identity matrix.

Within the context of normal cylindricity measurement, a less generally applicable description of the cylinder can be used to give a better 'feel' to the parameters describing it. Also experience of two-dimensional roundness measurement shows the type of operations likely to be needed on the reference (for example, linearisations) and the forms of parametrisation which are convenient to handle. It may be assumed that the axis of a cylinder being measured will not be far misaligned from the instrument Z-axis (that is the axis of the instrument spindle) and so its description in terms of deviation from that axis has advantages. In a direct parallel with the

description of eccentricity, cartesian components of these deviations are used. The intersection of the axis with the Z = 0 plane will be at (A_o, B_o) and the slopes from the Z-axis of the projections of the cylinder axis into the XZ- and YZ-planes will be A_1 and B_1. Any point on the axis is then defined by the co-ordinates $(A_o + A_1Z, B_o + B_1Z, Z)$. The slopes A_1 and B_1 relate simply to the direction cosines so that:

$$\underline{I}_3 - \underline{l}_1\underline{l}_1^T = \frac{1}{(1+A_1^2+B_1^2)} \begin{pmatrix} 1+B_1^2 & -A_1B_1 & -A_1 \\ -A_1B_1 & 1+A_1^2 & -B_1 \\ -A_1 & -B_1 & A_1^2+B_1^2 \end{pmatrix}$$

and on multiplying out equation 11.4 gives:

$$R_o = \frac{1}{(1+A_1^2+B_1^2)^{\frac{1}{2}}} \Big[(X-A_o)^2(1+B_1^2) + (Y-B_o)^2(1+A_1^2) + Z^2(A_1^2+B_1^2) \qquad 11.5$$
$$-2(X-A_o)(Y-B_o)A_1B_1 - 2(X-A_o)A_1Z - 2(Y-B_o)B_1Z \Big]^{\frac{1}{2}}$$

The conversion of equation 11.5 from cartesian to cylindrical polar co-ordinates gives the equation of a tilted cylinder as:

$$R(\theta, Z) = \Bigg[\frac{(A_o + A_1Z + A_oB_1^2 - A_1B_oB_1)\cos\theta}{1} \cdots$$
$$\frac{+ (B_o + B_1^2 + B_oA_1^2 - A_oA_1B_1)\sin\theta}{+ (B_1\cos\theta - A_1\sin\theta)^2} \Bigg] \qquad 11.6$$
$$+ \frac{R_o(1 + A_1^2 + B_1^2)^{\frac{1}{2}}}{(1 + (B_1\cos\theta - A_1\sin\theta)^2)^{\frac{1}{2}}} \cdot$$
$$\Bigg[1 - \frac{((A_o + A_1Z)\sin\theta - (B_o + B_1Z)\cos\theta)^2}{R_o^2(1 + (B_1\cos\theta - A_1\sin\theta)^2)} \Bigg]^{\frac{1}{2}}$$

In this form both the similarity to and the differences from the simple eccentric circle, equation 3.6, can be seen. The cross-

section in a plane of constant Z is an ellipse with minor semi-diameter R_o and major semi-diameter $R_o(1 + A_1^2 + B_1^2)^{\frac{1}{2}}$, its major axis having the same direction as the cylinder axis projected into the XY plane. The direction of the ellipse does not correspond to the direction of eccentricity in the plane since this latter value includes the contribution of A_o and B_o.

The cylinder is clearly non-linear in its parameters and furthermore can be shown to exhibit a non-convex feasibility region in parameter space for all the possibly required reference conditions: minimum radial zone, maximum radius inscribing and minimum radius circumscribing. The formal proof of non-convexity will not be included here but simple demonstrations illustrate its truth. The non-convexity of minimum zone and inscribing circles has already been demonstrated in chapter 4 and as the circle is a special case of equation 11.6 the equivalent cylinders are also non-convex. The non-convexity of the minimum circumscribing cylinder can be shown by the existence of examples, such as figure 11.3, in which distinctly separate local minima are shown. In the figure, the fit shown is a minimum radius circumscribing figure since the radius of the cylinder corresponds to the minor semi-diamter of the elliptical section. By symmetry it is clear that four such positions exist and it is readily seen that it is impossible to move between them without involving a figure of larger radius. This example shows that non-uniqueness does not occur only in the relatively unimportant (in the present context) sense that a cube may be circumscribed by a cylinder having its axis perpendicular to any face .

Both to allow analytical solutions to reference fitting and because there will be practical need to work with radius suppressed

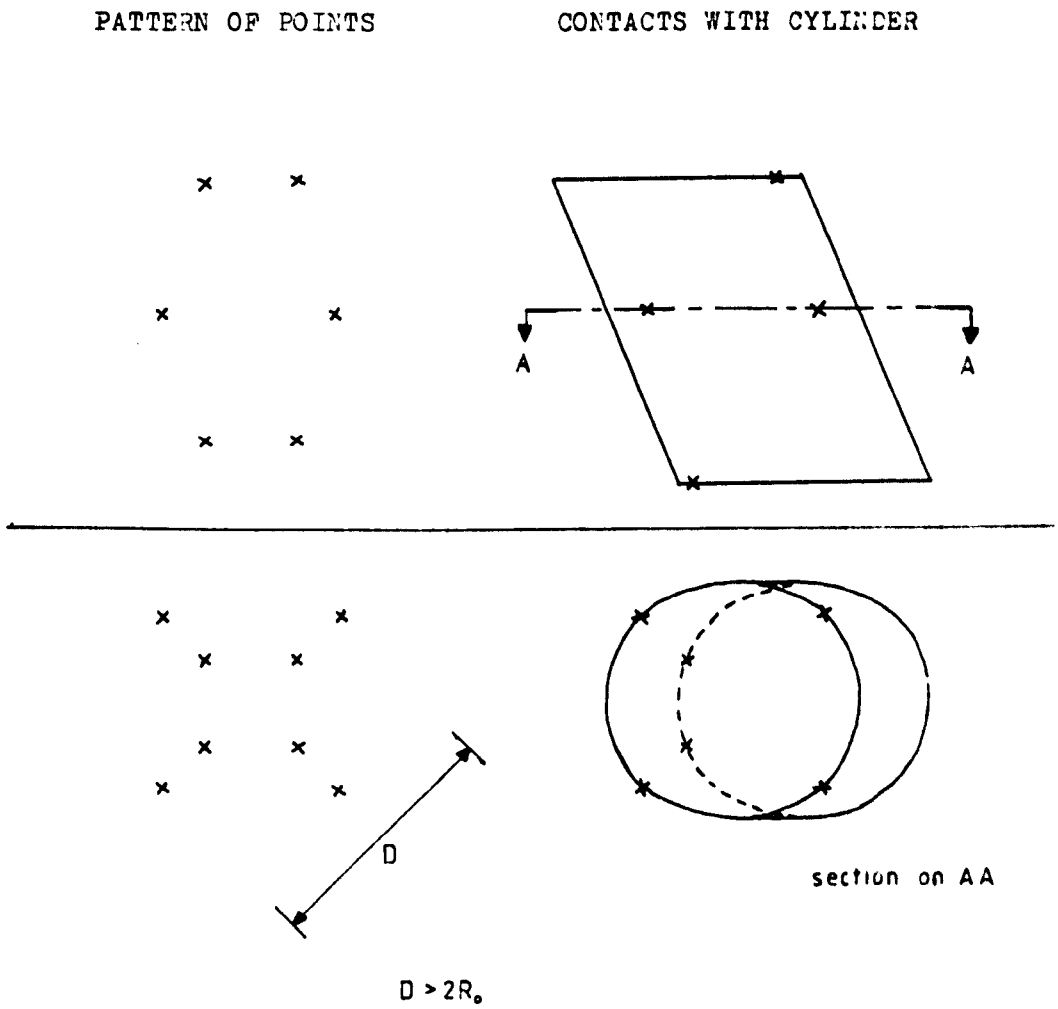


FIGURE 11.3: Example of Non-uniqueness in Minimum Circumscribing Cylinder Fitting

data, a reference figure linear in its parameters is desired. This may be found either by the direct application of Taylor expansions (it is easier to work with equation 11.5 and then convert the result to polar co-ordinates) or by removal of relatively small terms from equation 11.6 in a manner akin to the truncation of the binomial series in deriving the limacon from the circle. The linearisation of the cylinder about the point of perfect alignment ($A_o=B_o=A_1=B_1=0$) is shown to be the skew limacon cylindroid

$$R(\Theta, Z) = (A_o + A_1Z)\cos\Theta + (B_o + B_1Z)\sin\Theta + R_o \qquad 11.7$$

A comparison of equations 11.6 and 11.7 shows how much information is totally disregarded by this linearisation. In particular, there is no remaining term concerned with the ellipticity of the cross-section. For small parameter values, the differences between equations 11.6 and 11.7 will be dominated by the second order term of the power series expansion, namely:

$$\frac{R_c}{2}(A_1\cos\theta + B_1\sin\theta)^2 - \frac{1}{2R_c}((A_o+A_1Z)\sin\Theta - (B_o+B_1Z)\cos\Theta)^2$$

The nature of these error terms is emphasised if they are re-expressed as:

$$\frac{\tan^2\alpha\, R_c}{4}(1 + \cos 2(\Theta - \phi_\alpha)) - \frac{E^2(Z)}{4R_c}(1 - \cos 2(\Theta - \phi_E(Z)))$$

where α is the angle of the axis to the Z-axis and ϕ_α and ϕ_E are the directions of tilt and total eccentricity in the XY plane. The eccentricity terms E and ϕ_E depend upon Z whereas the terms due to pure tilt do not. The acceptability of the model depends upon the maximum value of eccentricity ratio which occurs at any plane (which will be at one end of the axis length over which measurements

are taken) and also upon the magnitude of the tilt compared to the absolute radius. As written above, the first term in the error can be identified with the representation of the tilted cylinder in terms of a skew circular cylindroid while the second term relates to the approximation of the circular cross-sections of that cylindroid by limacons.

The above discussion is naturally also of concern to the measurement of roundness profiles on cylindrical objects. It is quite common for tilt to be the major cause of eccentricity in a reading, particularly when using fixtures which cannot support the workpiece in the plane of measurement and under such conditions the phases ϕ_α and ϕ_E will be broadly similar so that the possible sources of second harmonic errors reinforce each other. On the other hand the error in the radial term could be rather smaller than would be expected simply from the limacon approximation.

11.3 Practical Considerations of Cylindroid References

The development of the skew limacon cylindroid from the cylinder is a parameter linearisation. Thus the immediate consequence to measurement practice is that exactly the same assessment techniques may be used as have been used here for roundness assessment. The cylindroid may also be used as a medium for transferring between instrument and chart co-ordinate frames. Its behaviour under radius suppression is exactly the same as that of the limacon since the suppression operates in directions perpendicular to the Z-axis. The magnification usually associated with the translation to chart co-ordinates has one extra effect on the cylindroid since generally

it would be expected that different values of magnification would be applied in the radial and axial directions. The slope of the cylindroid axis from the measurement axis will be multiplied by the ratio of the magnifications in these directions.

The shape difference between limacon cylindroid and cylinder is subject to more sources of variation than is that between limacon and circle but again similar methods can be used to control them. The amplitude of the second harmonic of the horizontal section through the cylinder will be, under practical measurement conditions, the effective error in that particular cross-section of the cylindroid. A worst condition for its size is that the harmonics generated by tilt and eccentricity are in phase, when the combined amplitude will be $R_{c/4}(\tan^2\alpha + \gamma^2(Z))$, $\gamma(Z)$ being the eccentricity ratio at the cross-section. Thus a quite conservative check method is to use $(\tan^2\alpha + \gamma^2_{max})^{\frac{1}{2}}$ as a control parameter in exactly the manner that γ is used for roundness measurement. It should be stressed that the values of α likely to be encountered within current practices are very small. The total tilt adjustment on some commercially available instruments is only a few minutes of arc, so values of $\tan\alpha=0.001$ would not be regarded as particularly small. In the majority of situations the limit on tilt will come from its effect on the allowable eccentricity: if the axial length of cylinder over which the measurement is performed is L_o, there must be at least one plane where the eccentricity is at least $L_{o/2}\tan\alpha$, so γ_{max} will exceed $\tan\alpha$ whenever the length of cylinder exceeds its diameter (as it may, also, if this condition is not satisfied.).

The ellipticity introduced by tilting a cylinder is difficult to account for in reference figure modelling since, apart from the

problems of working with a non-linear parameterisation, there are other causes of elliptical cross-sections with which interactions can take place. Using, for example, best fit 'ellipses', probably modelled by just the second harmonic of the Fourier series, on cross-sections will not usually yield directly information about tilt. This relates to the observation that while every tilted cylinder can be described alternatively, and equivalently, as a skew elliptical cylindroid, the vast majority of elliptical cylindroids do not describe tilted circular cylinders. Given a good estimate of the cylinder axis and knowledge of the true part radius, the amplitude and phase of the elliptical component can be calculated and could be used in a second stage of determining the reference, possibly in the form of a correction factor such as that proposed in section 7.5

MISSING

12 Limacon Cylindrical References

12.1 Least Squares Cylindroids

The skew limacon cylindroid is linear in its parameters and so the least squares solution for residuals

$$\delta_i = R_i - ((A_o+A_1Z_i)\cos\theta_i + (B_o+B_1Z_i)\sin\theta_i + R_L) \qquad 12.1$$

can be stated directly (see section 5.2 for details of the method). In matrix form, the parameter estimates are given by the solution of:

$$\begin{vmatrix} \Sigma\cos^2\theta & \Sigma\sin\theta\cos\theta & \Sigma Z\cos^2\theta & \Sigma Z\sin\theta\cos\theta & \Sigma\cos\theta \\ \Sigma\sin\theta\cos\theta & \Sigma\sin^2\theta & \Sigma Z\sin\theta\cos\theta & \Sigma Z\sin^2\theta & \Sigma\sin\theta \\ \Sigma z\cos^2\theta & \Sigma z\sin\theta\cos\theta & \Sigma Z^2\cos^2\theta & \Sigma Z^2\sin\theta\cos\theta & \Sigma Z\cos\theta \\ \Sigma z\sin\theta\cos\theta & \Sigma z\sin^2\theta & \Sigma Z^2\sin\theta\cos\theta & \Sigma Z^2\sin^2\theta & \Sigma Z\sin\theta \\ \Sigma\cos\theta & \Sigma\sin\theta & \Sigma Z\cos\theta & \Sigma Z\sin\theta & N \end{vmatrix} \begin{vmatrix} A_o \\ B_o \\ A_1 \\ B_1 \\ R_L \end{vmatrix}$$

$$= \begin{vmatrix} \Sigma R\cos\theta \\ \Sigma R\sin\theta \\ \Sigma RZ\cos\theta \\ \Sigma RZ\sin\theta \\ \Sigma R \end{vmatrix}. \qquad 12.2$$

where to save space indices have been ommitted: R, θ, and Z all have subscript i and all summations are over i = 1 to N.

The added complexity of the three dimensional problem means that there is even higher motivation than with the simple limacon for choosing measurement schemes which allow simplification of the coefficient matrix. This is unlikely to be possible on incomplete surfaces and so only full cylindrical surfaces will be considered. For these it is probable that a sampling scheme having a high degree of uniformity would be used for instrumental

as well as arithmetic convenience. Since, also, on a roundness measuring instrument it is normally advisable for best accuracy to keep the spindle rotating constantly throughout the measurement, two patterns of measurement are suggested: a series of cross-sections at pre-determined heights Z_i or a helical traverse.

If a series of cross-sections is used and each sampled identically, the summations over all the data in equation 11.2 can be replaced by a double summation over the points in each plane and the number of planes, for example:

$$\sum_{i=1}^{N} Z_i \cos\theta_i = \sum_{k=1}^{m} Z_k \sum_{j=1}^{n} \cos\theta_{jk}$$

where there are m sections each of n points, mn = N. Now if the sum over j satisfies the fourfold symmetry identified in section 5.2 for the simplification of the least squares limacon solution, at each plane the summations over $\cos\theta$, $\sin\theta$ and $\sin\theta\cos\theta$ will be zero and so also will be the sums of these terms over all the planes. The matrix of coefficients then becomes quite sparse:

$$\begin{pmatrix} \Sigma\Sigma\cos^2\theta & 0 & \Sigma Z\Sigma\cos^2\theta & 0 & 0 \\ 0 & \Sigma\Sigma\sin^2\theta & 0 & \Sigma Z\Sigma\sin^2\theta & 0 \\ \Sigma Z\Sigma\cos^2\theta & 0 & \Sigma Z^2\Sigma\cos^2\theta & 0 & 0 \\ 0 & \Sigma Z\Sigma\sin^2\theta & 0 & \Sigma Z^2\Sigma\sin^2\theta & 0 \\ 0 & 0 & 0 & 0 & mn \end{pmatrix}$$

Noting that those terms involving $\cos^2\theta$ correspond with A_0 and A_1 and similarly with $\sin^2\theta$ and B_0 and B_1, further interpretation of this matrix is possible. The radius of the least squares limacon cylindridroid is the mean value of all the radial data points and its axis is the least squares straight line through the centres of the least squares limacons on the cross-sectional planes.

This measurement scheme has, apart from computational simplicity, two advantageous features: the information on both axial straightness and cylindricity measurements is produced simultaneously and, depending upon exactly what is to be measured, there is considerable scope for data reduction during the course of the measurement.

There are other ways of selecting measuring schemes which lead to simplifications similar to, but not as complete as, the above when using measurement in cross-section. No details of them will be given here.

The helical traverse method is attractive from an instrumentation point of view. However, computationally, it loses the advantage of having Z and θ independent and so evaluation must be over the whole data set in one operation. It would be expected that samples would be taken at equal increments of θ and since Z depends linearly on θ this allows various schemes for simplifying equation 12.2 quite considerably. Again, only one scheme will be discussed here. If it can be arranged that the total traverse encompasses an exact even number of revolutions of the part and that there is a multiple of four samples in every revolution then defining the origin such that Z = 0 at the mid-point of the traverse will cause all summations of odd functions of Z and θ to be zero, as will all those in simply $\sin\theta$, $\cos\theta$ or $\sin\theta\cos\theta$. The coefficient matrix then becomes:

/continued over page

$$\begin{pmatrix} \Sigma\cos^2\theta & 0 & 0 & \Sigma z\sin\theta\cos\theta & 0 \\ 0 & \Sigma\sin^2\theta & \Sigma z\sin\theta\cos\theta & 0 & 0 \\ 0 & \Sigma z\sin\theta\cos\theta & \Sigma z^2\cos^2\theta & 0 & 0 \\ \Sigma z\sin\theta\cos\theta & 0 & 0 & \Sigma z^2\sin^2\theta & \Sigma z\sin\theta \\ 0 & 0 & 0 & \Sigma z\sin\theta & N \end{pmatrix}$$

The original set of five simultaneous equations reduced to a set of two and a set of three with considerable computational saving.

One failing of the helical traverse relative to the measurement of cross-sections is that no information relating directly to axial straightness is produced. Overall, it would seem that there needs to be fairly strong instrumental reasons for a helical traverse to be used, particularly as there would appear to be more types of surface discontinuity which can be excluded from the measurement by the judicious choice of cross-sectional heights than from choice of helix pitch.

The demonstration carried out in chapter 5 that the least squares limacon survives unchanged relative to the profile under the radius suppression transformation can be applied in identical fashion to the limacon cylindroid. Given only the provision that the z axis scaling is unchanged the cylindroid parameters can be used in chart or instrument co-ordinates by applying magnification and suppressed radius in the normal way.

One property of the limacon fit which does not apply to the cylindroid is the observation that the estimate for centre is exact (section 5.3). Reference to equation 11.6 reveals that there are additional terms which contribute slightly to the odd harmonics in the case of the cylindroid. Taking the second order term of the binomial expansion of the first part of equation 11.6 suggests that the fundamental is changed only by about $1 + \frac{\tan^2\alpha}{4}$:1 so that the

estimate of axis from the cylindroid should still be good in practice. This is, however, a further warning that there is a greater degree of approximation between cylinder and cylindroid than between circle and limacon. Although still a good approximation the cylindroid can stand rather less abuse than the simpler situations!

12.2 Boundary Value Cylindroids

As with the least squares fit, the solution to the boundary limacon cylindroids is a direct extension of the methods used in two dimensions. The general solutions examined in section 6.2 can be merely extended to include two extra parameters. All three boundary cylindroids are found by solving their linear programs, revised simplex on the dual being the most appropriate of the standard techniques. The inscribing cylindroid can be solved as a circumscribing cylindroid by inversion of the radial data, in the manner described in section 6.5, if so desired.

Generally the motivation for seeking exchange algorithms to replace the general method reduces as the geometrical complexity of the problem increases for the manipulations required to calculate the exchange become more involved. The properties of cylinder and cylindroid examined in the previous chapter would suggest that exchange conditions might be complicated and so offer no real advantages over the use of simplex. That this is indeed the case will be illustrated by just one example. The minimum circumscribing limacon cylindroid has the formulation:

Minimise R_L

Subject to $(A_o+A_1Z_1)\cos\theta_i + (B_o+B_1Z_1)\sin\theta_i + R_L \geq R_i$

for all i.

Following the method used in section 6.4, the basis of the dual linear program will be:

$$\begin{pmatrix} \cos\theta_1 & \cos\theta_2 & \cos\theta_3 & \cos\theta_4 & \cos\theta_5 \\ \sin\theta_1 & \sin\theta_2 & \sin\theta_3 & \sin\theta_4 & \sin\theta_5 \\ Z_1\cos\theta_1 & Z_2\cos\theta_2 & Z_3\cos\theta_3 & Z_4\cos\theta_4 & Z_5\cos\theta_5 \\ Z_1\sin\theta_1 & Z_2\sin\theta_2 & Z_3\sin\theta_3 & Z_4\sin\theta_4 & Z_5\sin\theta_5 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The co-factors of the final row of this matrix must have the same sign if dual feasibility is to be maintained. Using the geometric interpretation of this condition the boundary between positive and negative sign regions of these co-factors will be given, in cartesian form, by equations such as:

$$\Delta_{234.} = \frac{1}{R_2R_3R_4R}\begin{vmatrix} X_2 & X_3 & X_4 & X \\ Y_2 & Y_3 & Y_4 & Y \\ Z_2X_2 & Z_3X_3 & Z_4X_4 & ZX \\ Z_2Y_2 & Z_3Y_3 & Z_4Y_4 & ZY \end{vmatrix} = 0$$

It is not easy to describe the surface defined by such an equation but it is clear that it is not a simple geometric shape. The permissible orientations of points for defining a dual feasible cylindroid depend upon the interaction of four curves of this type, making their full description all but impossible. The search for an exchange algorithm can be conscientiously abandoned at this stage. The recommended method of solving boundary limacon cylindroids is by direct solution of the dual linear programme.

Although the full geometric description of dual feasibility

is not available, it is easy to discover particular sets of conditions which satisfy that requirement. This has the merit that an initial feasible solution for the dual programme can be specified, so saving the use of artificial variables. One such condition consists of three points in one cross-sectional plane which obey the 180° rule with respect to the Z axis and any two other points which do not lie in that same plane and which are not diametrically opposed to one another. If this occurs, two of the operational co-factors of the basis become unconditionally zero and it is then possible to invent ambiguous exchange conditions. This corresponds to degeneracy in the linear program and casts a slight doubt over whether it is always possible to obtain a solution by this method.

In passing it is worth observing that many of the difficulties encountered with cylinder fitting are caused because the figure is defined with respect to a line rather than a point. The natural extension of the circle to three dimensions is to maintain uniformity about a point by moving to a sphere. By direct extension of section 6.4, the "limacon spheroid" can be shown to have a dual feasibility condition in which any two defining points lie to opposite sides of a plane defined by the other two points and the origin. This will be seen to be exactly analogous to the situation shown in figure 6.6. Thus a very simple geometry is preserved, although calculating its implications for point exchange may still not be worthwhile compared to the direct use of the dual programme.

Given the greater degree of non-convexity in cylinder formulations than in circle ones, it is to be expected that there

will be less that can be said with certainty concerning bounds to cylinder fits from cylindroid fits. The only definite assertions are those giving overestimates for the minimum zone and minimum circumscribing radius and an underestimate of the maximum inscribing radius. The conceptual mechanism used is the same in all three cases. For the circumscribing cylinder the data is first enclosed by the minimum circumscribing limacon cylindroid. This figure is then enclosed within a circular cylindroid and around this is placed an elliptical cylindroid representing a cylinder having the same axis tilt as that of the limacon cylindroid. Figure 12.1 attempts to illustrate this procedure showing only the plane crucial to the exercise. The circular cylindroid which can enclose the limacon cylindroid will be controlled by the cross-section in which the limacon has the largest eccentricity and so needs the largest circumscribing circle. A convenient circle is that centred at the limacon centre having radius $(R_L^2 + E_{max}^2)^{\frac{1}{2}}$ and a circular cylindroid having this radius and the same axis as the limacon cylindroid will be a circumscribing figure. In fitting any ellipse about the circle, the minor semi-diameter remains equal to the circle radius and as the minor diameter corresponds to the dimension of the untilted cylinder a reasonable overestimate of the cylinder radius is $(R_L^2 + E_{max}^2)^{\frac{1}{2}}$. This is the same as the overestimate for the circumscribing circle except in one respect. It is quite likely that the limacon having maximum eccentricity, being at one end of the measured cylinder, will not contact the profile at all and so the error between the overestimate and the profile may be rather larger than in the two dimensional case. The original limacon cylindroid is sufficiently

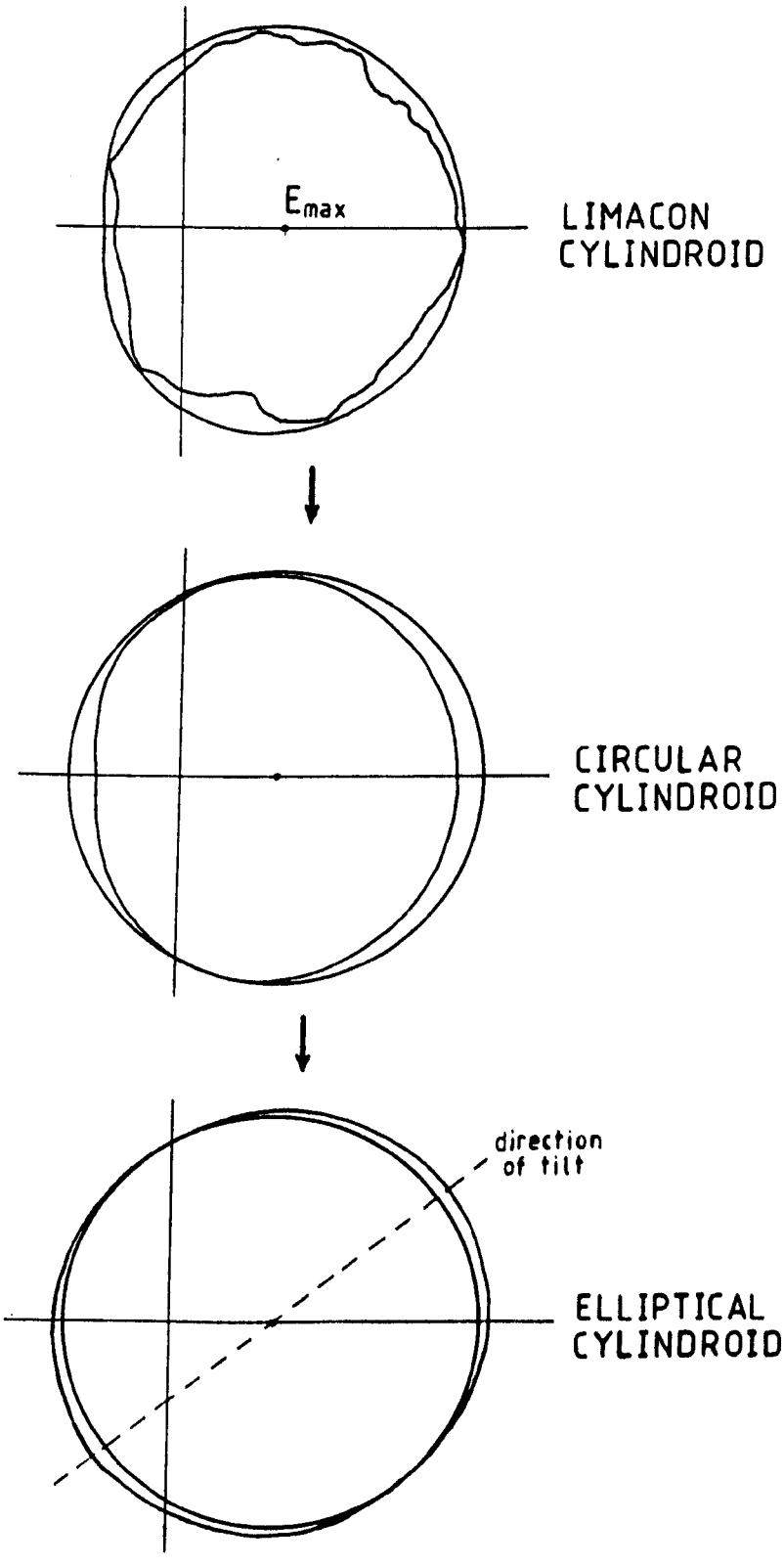


FIGURE 12.1: A Circumscribing Cylinder, Overestimating the Minimum Radius Figure

large to enclose the effects of ellipticity due to tilt and so such effects are included in the process. The degree of effect that is entailed cannot be specified: it depends upon the nature of the peaks of the data.

The estimate of inscribing cylinder entails first placing a circular cylindroid within the limacon cylindroid. This is straightforward: by inspection the circular cylindroid having radius equal to the limacon and the same axis will be its maximum inscribing figure. The elliptical cylindroid must lie inside this circular cross-section, however, and so its major semi-diameter will be R_L. Thus the radius of the underestimate of the inscribing cylinder will be $R_L\cos\alpha$.

The two cylinders defined for inscribing and circumscribing cases have the same axis and so may be used to overestimate the minimum zone. Their difference gives a zone of width well approximated (for $E_{max} \ll R_L$) by $H(1+\cos\alpha) + R_L(1-\cos\alpha) + \frac{E_{max}^2}{2(R_L+H)}$ where $2H$ is the cylindroid zone width

In specific cases it will be possible to establish tighter limits on these fits but it is not possible to generalise them fully. For example since the inscribing elliptical cylindroid does not touch the appropriate limacon cylindroid it could be expanded until it does do so. The degree of expansion will depend upon the direction of tilt relative to the orientation of the limacon, both unknowns in the general solution.

Unlike the case with the circle it cannot be guaranteed that a cylinder can be fitted through the five defining points of the limacon cylindroid (although with practical degrees of alignment

error it is likely that it will be so). Also the failure to identify fully the geometry of feasible contacts prevents the use of arguments similar to those adopted in Chapter 7. Theoretical work on cylindrical reference bounds seems dogged by areas of vagueness such as this which prevent the making of rigorous statements even when common sense suggests that things are so. In fact there is very little practical evidence about the fitting of cylinders and almost all statements are conjecture based on experience of two dimensional measurements. Perhaps what is most needed at this stage is a detailed and extensive practical study of cylinder and cylindroid fitting which may lead at least to an empirical approach to the whole subject.

12.3 Cylindricity Assessment in Practice

The facilities which could be made available during the course of the work reported here were not suitable for performing the exercise suggested in the previous section. The difficulty lay in the fact that a roundness instrument with a straightness datum was not available on line to the computer system. Thus although it was possible to develop and test programs for cylindricity measurement by measuring cross-sections positioned manually on the Talyrond 73, the data so obtained could not be considered as a sufficiently accurate representation of the cylinder for the results to be included in a survey of real conditions. What could be judged was the likely performance to be expected of working systems.

A least squares limacon cylindroid system was developed very easily from the simplified measurement scheme described in section 11.2. Equiangular samples in each of a series of equally spaced cross-sections were used. The required calculations are then quite simple: only about

30 lines of FORTRAN are needed to program the whole of the reference calculation. The two main decisions needed of a system design concern the amount of data to be stored and the method of presenting the results. The volume of data which is readily collected from a cylindrical surface can rapidly overwhelm a small computer system. If only axial straightness information is required the difficulty is removed since only the centre of each cross-section's least squares limacon need be stored plus, of course, the data from one profile which is being processed at any one time. However cylindricity measurement requires that all the data is retained so that it can be compared to a common reference not known until the end of the measurement sequence. Fortunately it seems that for the majority of workpieces likely to be measured the dominating form errors will be slow moving, usually only a few undulations per revolution in the cross-section and maybe only an "S-bend" in the axis. In the field of precision engineering, the nature of conventional machine tools helps to ensure occurence of such forms. Providing that it is accepted that cylindricity is a measurement of major deviations of form (having, for example, a relationship to roundness similar in some aspects to that between straightness and surface texture) then a relatively coarse sampling, with, consequently, a lesser number of data points being used, can be contemplated. In the system described here a sampling rate of 64 per revolution was used, with a maximum of 8 planes being allowable. The total data storage is then the same as for one roundness profile using 512 points. If information about short wavelengths of the surface is also required it is often sufficient to measure each plane at a high sampling rate and supply roundness measurements for them as individual, independent readings while retaining only a fraction of the

total points so collected for the overall cylinder analysis. This would require, say, just an extra 512 data points to be stored.

If the information required from a cylindricity measurement is much more complicated than an overall zone width, a purely numerical presentation of that information, while complete, is very difficult to assimilate. To achieve a rapid impact a graphical method is wanted. It seems that a direct extension of the polar chart into three dimensions using, say, an isometric projection on a visual display unit is not very useful: the distortions caused by radius suppression and relative magnification make all but the simplest shapes virtually unintelligible (34). This technique is also very expensive in terms of computational effort. Perhaps the best method of showing all the surface detail in one view is to use a contour map of the opened out cylinder surface (18), for example. Interpretation of a map requires some degree of care and experience and it remains computationally expensive to produce easily readable forms.

In the system developed here an attempt was made to present graphical information in forms with which the operator could be expected to be familiar: the two-dimensional roundness graph and orthographic projections following normal mechanical design drawing conventions. Axial straightness was shown in terms of front and side elevations of the axis. Cylindricity was shown in terms of plan views along the least-squares axis. Either a complete view showing the envelope of the superimposed profiles correctly aligned to the cylindroid axis or the views of aligned individual cross-sections were shown. On the visual display unit the individual sections can be displayed in sequence each holding for a short

interval. Providing that the switch from one section to the next is quite rapid this scan gives a good subjective indication of the real nature of surface features. The differences of features such as small holes and axial or helical scratches are readily recognised, for example. Figure 12.2 illustrates the concepts of this display. A schematic display of distinct radial variation, such as taper, was also included but proved less successful. There was a tendency for the display to be confused with either axial information or with vertical cross-sections through the cylinder depending on the display technique used.

An operating system based upon this program was built for demonstration but was dispatched to exhibitions and so could not be fully evaluated (35). The calculation time was about 5 seconds using floating point single precision arithmetic and a sine/cosine look-up table. This compares well to the time taken to acquire the data: because of the vertical shift between measurements two spindle revolutions are needed for each plane, so 8 sections would take about 160 seconds. Although this system has never been used practically several special measurement systems based upon various parts of it have been supplied commercially by Rank Taylor Hobson Ltd.

A minimum circumscribing limacon cylindroid program based on revised simplex has also been developed. The programming technique is basically similar to that described in chapter 9 for the two dimensional analysis. Again up to 8 planes each of 64 equiangular points were used, but unequal spacing of the planes was allowed. Since the mechanism of revised simplex does not change the program performance is very similar to that for the simple limacon. The

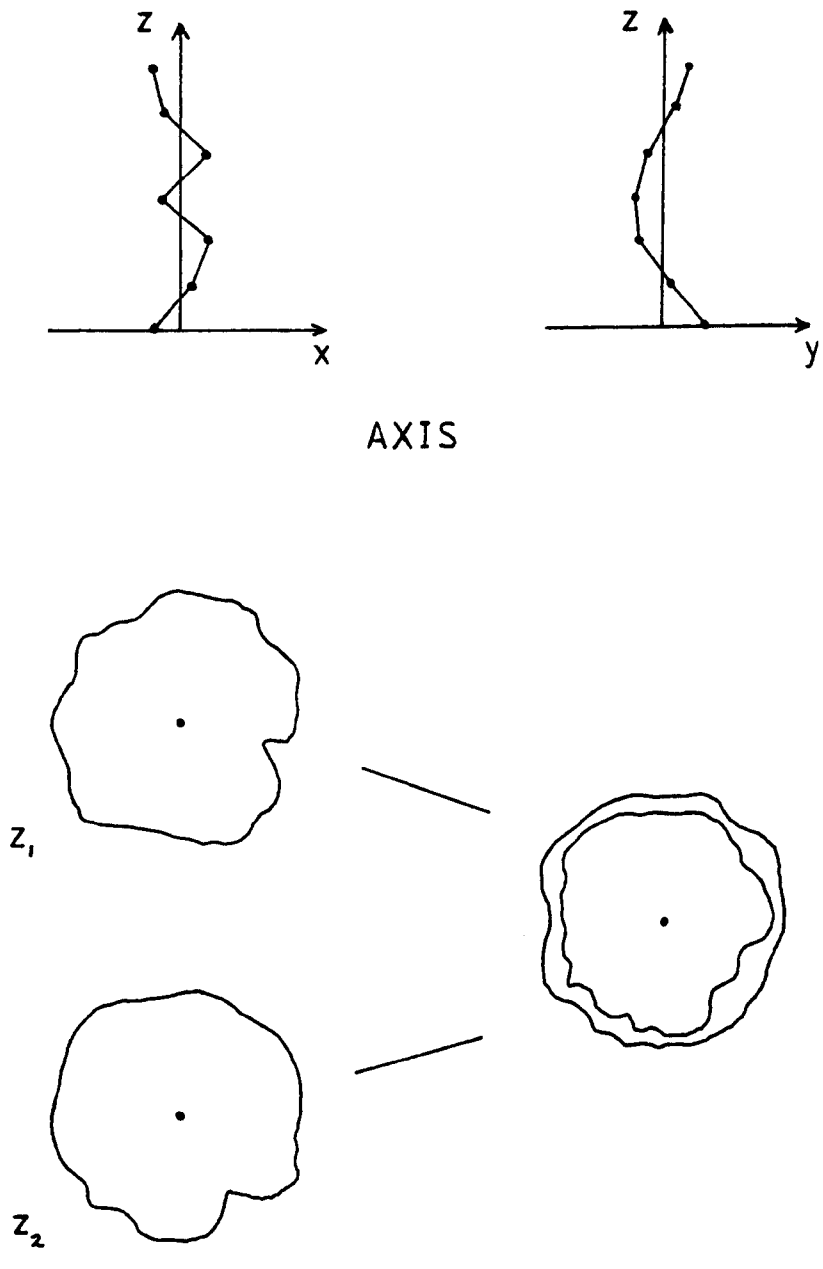


FIGURE 12.2: Display of Deviations from a Perfect Cylinder

program is slightly longer since extra effort is needed for the fetching and carrying of the extra variables. For the same reason it runs a little more slowly. However the operational time is dominated by the calculation of the reference at each iteration which depends mainly on the number of sine and cosine operations and so on the number of data points. Thus 8 planes of 64 points iterate almost as fast as a single plane of 512 points. The overall calculation time will be longer since with five parameters more iterations will be required particularly when, as in the present implementation, artificial variables are used. All the same, calculation times of not much more than 15 seconds were recorded which is considered to be quite acceptable.

13 Conclusions

At its inception it was intended that the primary aim of the work described here would be the development of methods for measuring cylindricity using instrumentation which is commercially available. Subject to the (expected) necessity of using approximations in the reference figure this aim has been fulfilled. However a consideration of the amount of this report which is dedicated directly to cylindricity indicates that the development of algorithms for cylinder fitting by the use of limacon cylindroids is a theoretically trivial extension of the ground covered in the earlier chapters. It is believed that the final, practically useful results obtained fron the analysis are not, in one sense, the most important feature of this work. The underlying groundwork concerning the mechanics and philosophy of both measurement and analysis makes these, and other results in related fields, relatively easy to produce. As with many other successful techniques it seems quite straightforward once it has been done!

The author anticipates that few would challenge the opinion that the greatest single contributor to the field of roundness measurement is R. E. Reason. So great is his influence that the whole subject reflects his approach. It relies upon mechanical instrumentation of very high quality with the measurements and necessary analysis being performed in a pragmatic and highly intuitive manner: the "kinematic" derivation of the polar distortion on the chart is a classic example of this approach (ref. 19). That it has been a successful method is bourne out by well over

twenty years of practical experience. The approach does, however, have some drawbacks. Being essentially mechanical/intuitive, approximations in the analysis become embedded within the derivation so that it is not always clear whether an approximation has been made at all. Being also a pragmatic method, the explicit assumptions, and any analysis built upon them are judged in terms of the performance of instrumentation available when they were made. Consequently the direct extension of the existing analysis to new forms of measurement which might be needed cannot be regarded as a "safe" procedure. While limits upon misalignment which will give adequately accurate measurements are known and tested, there is no information embodied in the early theoretical work to indicate whether improvements in instrumentation, or even a completely different approach to instrumentation, will be fully exploited by those limits. Currently these doubts are reflected in many areas: the analysis of cylindricity and the implications of modern developments in high range/resolution transducers and high accuracy co-ordinate measuring machines are just a few examples.

This work has adopted a deliberate policy of re-analysing the roundness measurement process in a mathematical manner in an attempt to produce a consistent, general theory upon which to build. This has led to the re-derivation of well-known relationships. No excuse is made for this since it is a necessary part of the total exercise. While it is hoped that the approach used here will supercede the earlier methods as a theoretical tool it does not invalidate using earlier results in context. (To draw a famous parallel: the work of Einstein has not affected the results or the continued use of Newtonian mechanics in "ordinary" situations but it offers a more complete description of rather less mundane problems.)

The general approach to roundness measurement analysis (and parallels may be drawn in other disciplines) proposed here is based upon the concept of the three frames of reference which occur during the measurement process: component, instrument and chart co-ordinates. The 'errors' and 'distortions' which afflict the data and particularly the analysis of reference figures can be associated with the transformations between these planes. Thus misalignment associates with the transformation from component to instrument co-ordinates and radius suppression with that from instrument to chart co-ordinates. (These effects may be contrasted with, for example, transducer inaccuracy which is associated specifically with the instrument co-ordinate frame.) The key to working consistently with these frames lies in producing a system model which has a geometry essentially independent of the transformations. The transformation between instrument and chart is a linear operation on radial values and so a linear model is a natural choice. Thus the linearised circle, the limacon, is identified as a figure of particular significance. It is not just a convenient approximation for avoiding unpleasant mathematics.

All reference figure fitting involves optimisation and for well controlled optimisation linear parameters are a desirable feature. Given complete linearisation the well established methods of linear least squares and linear programming can be invoked. Under many measurement conditions the use of limacons or limacon cylindroids which have the required linearity can be justified and standard solution techniques can be applied to them. The development of specialised highly efficient methods, such as

exchange algorithms, then follows as a logical step.

There is no point in restating here any detail of these processes or of the important questions concerning the relationship of circles and limacons. Having a general framework within which these have been discussed, the decision on whether they are useful is a pragmatic one. The evidence of this work shows that limacon approximations are highly usable in practice. It shows, also, under what conditions they become less useful and suggests alternatives which may then be used.

The revolution in instrumentation over the last few years has been due mainly to the availability of cheap digital computation. This work is well matched to this phenomenon since many of the consequences of the analysis can only be exploited digitally. To minimise the cost of computing with modern micro-processors efficiency is required both in storage requirements and in operation time so that the slowest suitable system can be used to provide the answers within a reasonable period. The algorithms developed here offer big improvements over previous methods both in this respect and in terms of guarantee-able accuracy. Systems, based upon the work described here, have already been supplied commercially in which throughput time is critical and high accuracy is demanded, conditions which the previously used algorithms could not possibly have satisfied without the use of hardware floating point processors. The cost savings in such cases are self-evident.

Appendix 1

Minimum Zone Lines and Planes

The methods developed in the main text for deriving boundary references are not applicable only to the measurement of roundness but to any process which can be expressed in a similar way. As a further example of the method the derivations of references for another important area of surface metrology, straightness and flatness, will be outlined here. This will give also another opportunity to compare solutions in two and three dimensions.

Probably the most commonly applied measurements for straightness and flatness work with respect to a reference defined in terms of two parallel lines or planes which enclose the profile while having the minimum vertical separation: that is minimum zone lines and planes are wanted. As the reference figures are actually linear, the use of linear programming methods is virtually automatic.

For the minimum enclosing lines there is no need to perform a formal derivation for they belong to the well documented class of minimax polynomials, that is curves having the smallest possible maximum divergence from the data. The condition for this is that relative to an n^{th} order polynomial, the data must have $n+2$ maxima and minima all of equal magnitude. The solution is found by means of the Stiefel Exchange Algorithm which proceeds by fitting the polynomial according to this condition to $n+2$ points and then bringing points further away than those points into the defining set while maintaining that condition. In terms of the minimum zone straight lines there will be three points, two contacting one line and one

the other in an alternate sequence which are iterated by exchanges, see figure A1.1.

The minimum zone planes problem can be expressed:

$$\begin{aligned} &\text{Minimise} && h \\ &\text{Subject to} && ax_i + by_i + c + h \geqslant z_i \\ & && ax_i + by_i + c - h \leqslant z_i \\ &\text{for all data points } (x_i, y_i, z_i) \end{aligned}$$

a, b and c are sign unrestricted and $h \geqslant 0$. Based on the arguments given in the main text the dual of this linear programme will be used. Noting that $h = 0$ is valid only for the trivial condition that all points are co-planar then it may be asserted that four points will be represented in the basis which for the dual can be expressed (see section 6.6 for the reasoning):

$$\underline{B} = \begin{pmatrix} S_i x_i & S_j x_j & S_k x_k & S_l x_l \\ S_i y_i & S_j y_j & S_k y_k & S_l y_l \\ S_i & S_j & S_k & S_l \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

where S_i, etc. are variables taking values +1 or -1 according to whether (x_i, y_i, z_i) contacts the upper or lower of the minimum zone planes. Any optimal solution to the primal problem must demonstrate dual feasibility. Here the primal objective function depends only on the fourth parameter and so dual feasibility is guaranteed if all terms in the final column of $\underline{B}^{-1}$ are positive. This in turn will be true providing that:

$$\begin{aligned} -S_j S_k S_l \Delta_{jkl} \qquad & -S_i S_j S_l \Delta_{ijl} \\ S_i S_k S_l \Delta_{ikl} \qquad & S_i S_j S_k \Delta_{ijk} \end{aligned}$$

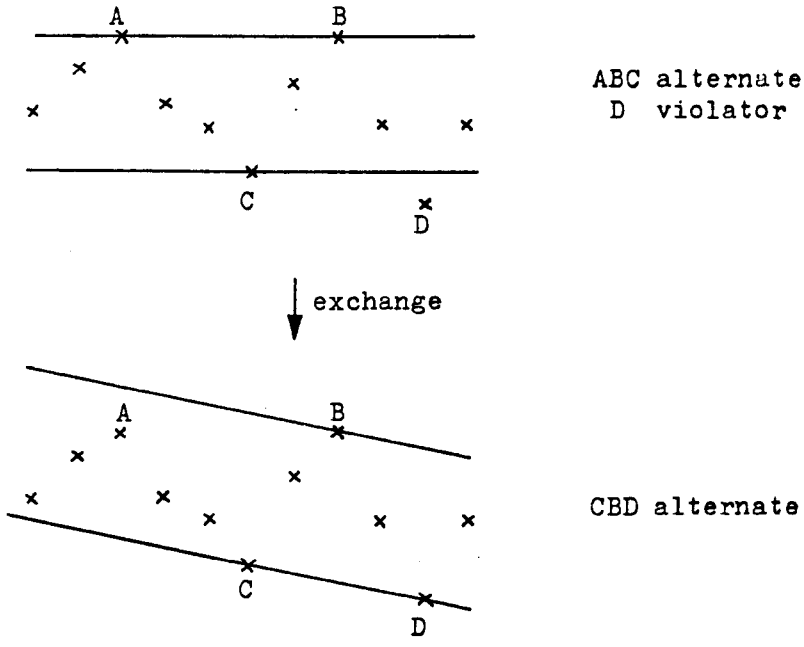


FIGURE A1.1: Minimum Zone Straight Lines: the Stiefel Exchange

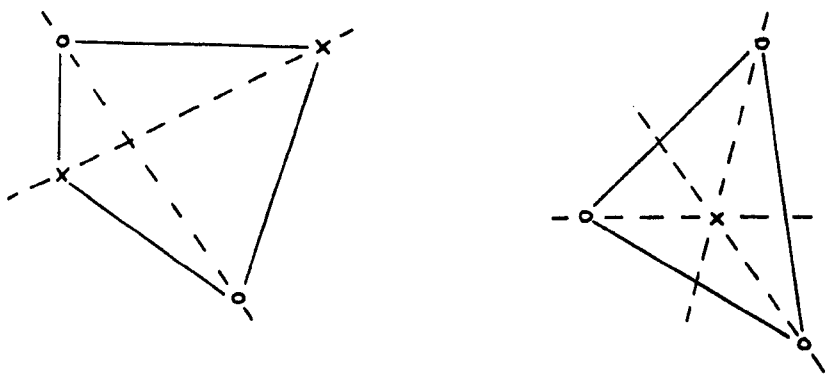


FIGURE A1.2: Plan of Contact Geometries for Minimum Zone Planes

O,X Represent Contacts with each of the two Planes

where:

$$\Delta_{jkl} = \begin{vmatrix} x_j & x_k & x_l \\ y_j & y_k & y_l \\ 1 & 1 & 1 \end{vmatrix} \qquad \text{etc.}$$

all have the same sign.

Consider the determinant:

$$\Delta_{jk.} = \begin{vmatrix} x_j & x_k & x \\ y_i & y_k & y \\ 1 & 1 & 1 \end{vmatrix} = 0$$

This is the boundary between regions in which Δ_{jkl} is positive and those where it is negative. Geometrically the boundary is a plane parallel to the z-axis (since it is independent of z) and passing through points (x_j, y_j) and (x_k, y_k). The dual feasibility condition requires that if $S_i = S_l$ (contacts with the same plane) Δ_{jkl} and Δ_{jki} must have different signs and vice versa. So if the i^{th} and l^{th} contact are with the same plane they lie on opposite sides of $\Delta_{jk.} = 0$ but if they contact different planes they lie both to the same side of $\Delta_{jk.} = 0$. A parallel argument shows that the same is true for all pairs of points.

The relative positions of contacts which give dual feasibility are given by these relationships. There are two ways of satisfying them as shown in the plan views of figure A1.2. There may be two contacts with each of the minimum zone planes in which case the plan of lines joining the alternate types must form a convex quadrilateral or a three:one split in which case the single contact must lie in the plan of the triangle formed by the other three contacts.

It is readily demonstrated that there is a unique exchange for

any new point in order that these relationships be preserved and so a workable exchange algorithm may be based upon these patterns. While its use by hand is very easy, the number of decisions which are involved in making the exchange is quite high and the development of a computer program to perform them is quite complex. For example the decision to switch from 3:1 to 2:2 arrangements of contacts is intuitively obvious but needs quite an effort to explain! Even with this most simple of three dimensional zone fits the advantage of using specific exchange algorithms rather than a general revised simplex solution for automatic system is becoming unclear.

Appendix 2

A Brief Summary of Linear Programming Concepts

The purpose of this note is to provide in very compact form an indication of how the theory of linear programming used in this work fits together. It may be used either as a reminder for those knowing, but not familiar with, the theory or as a guide to the most relevant sections of specialist textbooks on the subject. It covers only major features: important subtleties are not included. The notation used here has been kept in a form similar to that used in many textbooks, it does not therefore correspond with the notation used in the main body of this text.

Linear programming seeks solutions to optimisation problems in which the objective function and all constraints depend linearly on the problem parameters. In vector/matrix notation such problems are:

$$\text{Maximise} \quad z = \underline{C}^T \underline{x}$$

$$\text{Subject to} \quad \underline{A}\,\underline{x} \leqslant \underline{b}$$

where $\underline{C}$ and $\underline{x}$ are n-vectors, b an m-vector and $\underline{A}$ an nxm matrix of coefficients.

The first stage of solution is to switch to more easily handled equality constraints by introducing m extra parameters representing the amounts by which $\underline{A}\,\underline{x}$ is less than $\underline{b}$ and called slack variables (or sometimes surplus variables when having $\geqslant$ type constraints). This extension allows the problem to be stated as:

$$\text{Maximise} \quad z = \underline{C}^T \underline{x}$$

$$\text{Subject to} \quad \underline{A}_E \underline{x} = \underline{b} \quad ; \quad \underline{A}_E = \left[\underline{A} \,|\, \underline{I}_m\right]$$

It is required that the parameters, $\underline{x}$, be non-negative for the formal

solution of this programme. (There are special devices for allowing sign indeterminant parameters.)

In a parameter space where each x_i is expressed along an othogonal axis, the constraints form a set of hyperplanes which enclose a "feasible region" in which lie all combinations of the parameters satisfying all the constraints. As a consequence of the problem linearity this region will be convex and bounded by planes forming vertices when sets of n hyperplanes intersect (in three dimensions the region would be a polyhedron). Theorems of linear programming indicate that there will be one optimum value of the objective function and that it will coincide with a vertex of the feasible region. At this vertex a subset of exactly n of the m constraints will be satisfied exactly which in terms of the extended $\underline{x}$ vector indicates that the elements corresponding to the other constraints will be zero. These elements are termed non-basic. The defining variables constitute the basis. It is always possible to express both the constraints and the objective function in canonical form, that is as a constant plus the weighted sum of the non-basic variables by re-ordering the columns of the matrix and the elements of the vectors. As the non-basic variables are zero this is clearly a manipulative device which gives the current values of eg. the objective function and also indicates from the weighting co-efficients those non-basic variables which could, by becoming positive, increase the objective function.

The simplex solution to linear programming problems consists of manipulating the constraints in canonical form such that a move from vertex to vertex through the feasible region is performed in such a way that the objective function increases at every iteration. The

process consists of choosing from the canonical form of the objective function the non-basic variable which has the largest positive co-efficient, that is the variable which by changing value will most affect the value of the objective function. This variable will be brought into the basis and consequently one of the basic variables must be removed. As there must be a move to a new vertex, the new variable must be made as large as possible. Thus it is possible to identify the basic variable it replaces as being the one active in the constraint for which the new point can take the smallest value before making it infeasible. The process continues in this manner until the canonical form of the objective function has no positive co-efficients at which point the optimum is found. To start this process automatically it is necessary to have the parameter space origin as a vertex. This then becomes the initial basic feasible solution, all parameters are zero and all slack variables basic. If the origin is not a feasible point and no other initial basic feasible solution can be specifiec in advance a device involving extra variables called artificial variables is used to modify the problem so that the new origin is included. The artificials are iterated out preferentially and when all are removed, the problem sits at a basic feasible solution of the original problem from which the normal methods can proceed.

The usual manner of performing the calculations does not specifically manipulate the canonical forms but uses the simplex tableau. The tableau is an array of co-efficients of the form:

$$\begin{array}{c|c} \underline{A}_E & \underline{b} \\ \hline \underline{c}^T & z \end{array}$$

which is used because all the operations necessary to move from

vertex to vertex through the simplex iterations are elementary row operations on this tableau. The largest positive co-efficients of $\underline{C}$ gives the variable which will enter the basis. The test for replacement is that the $\underline{b'}$ column of the updated tableau must remain positive. The updating of the tableau is them easily performed since the columns of the tableau corresponding to the basic variables constitute an identity matrix embedded into the tableau. Thus once it is known which row will contain the 1, the full update is done by normalising this row and subtracting suitably weighted amounts of it from the other rows to reduce the rest of the column to zeros.

It may be noted that the operation of the simplex tableau is similarly mechanically to the standard method of inverting a matrix by a series of elementary row operations. At any iteration of the tableau the columns which originally constituted the identity matrix, the original basic (usually being the slack variables), contains a complete record of the combined effect of all the iterations to date and so they form the inverse of the matrix made of the initial values of the columns constituting the current basis. The inverse of the current basis thus contains all the information necessary to reconstruct the complete tableau. The method of Revised Simplex exploits this particular feature. The information stored consists of the original tableau and an updated version of the original basis. At each iteration matrix multiplications using this inverse of the basis and the original $\underline{C}$ and $\underline{b}$ vectors allow the current $\underline{C'}$ and $\underline{b'}$ vectors to be calculated. The appropriate row of the $\underline{A}_E$ matrix $\underline{P}_j$ corresponding to the largest element of $\underline{C'}$ is updated to $\underline{P'}_j$ by another matrix multiplication and then using $\underline{P'}_j$ and $\underline{b'}$ the necessary elementary row operation for the current iteration found in the normal way.

Since this operation is only performed on the inverse of the basis, not the complete tableau less total work has to be done. The savings are particularly evident in tableau having large numbers of columns and relatively few rows since the basis is a square matrix of dimension equal to the number of rows (that is to the number of constraints).

An important theorem of linear programming is that concerning duality. The forms considered so far relate directly to the physical nature of the problem being considered, and constitute the primal linear program. The theorem states that for any primal:

$$\begin{aligned} &\text{Maximise} && z_p = \underline{C}^T \underline{x} \\ &\text{Subject to} && \underline{A}\,\underline{x} \leqslant \underline{b} \\ & && \underline{x} \geqslant 0 \end{aligned}$$

there exists a dual linear programme:

$$\begin{aligned} &\text{Minimise} && z_d = \underline{b}^T \underline{y} \\ &\text{Subject to} && \underline{A}^T \underline{y} \geqslant \underline{C} \\ & && \underline{y} \geqslant 0 \end{aligned}$$

The dual has n constraints and m variables (y_i to which no direct physical meaning is attached). The optimum solution to both primal and dual give tableaux which are re-arranged version of each other. All information concerning the optimum may be obtained by solving either primal or dual. The major advantage of this manipulation is that if the original problem has many constraints and few variables, its dual will have few constraints and many variables. The dual can then be solved very efficiently by using revised simplex. The reversal of roles of the $\underline{b}$ and $\underline{C}$ vectors between primal and dual leads also to a series of correspondence conditions at iterations prior to the optimum being achieved. Examples of their use are given in the main text.

Appendix 3

Alternative Centre Definitions for Mean Circles

The following pages contain the text of a paper, written by the author and presented at the IMEKO VIII Conference in Moscow by Mr. V. S. Lukianov, "Dimensioning nominal circles: the resolution of conflicting ideas.". This is reference 30 of the main text.

Its relevance here is in the description of difficulties encountered by making different, sometimes inappropriate, assumptions regarding the mathematical formulation of derivations for the centre of mean circles to a set of data points.

DIMENSIONING NOMINAL CIRCLES: THE RESOLUTION OF CONFLICTING IDEAS

by

D. G. CHETWYND

Metrology Research Group, Rank Taylor Hobson,

Leicester, England.

The centre of the "least squares circle is a common starting point for measuring circular objects, which different schools of thought calculate by different formulae: an unsatisfactory situation. This paper investigates the methods of both schools and shows them to be approximations to the general solution. The applicabilities of the simple forms are compared. The relationship between least squares centres and profile centroids is also briefly examined.

Keywords: Metrology, Least Squares Circle

1. Introduction

Traditionally the methods and equipment of the dimensional and surface metrologists ahve been almost totally distinct. In the case of nominally circular components shape has been measured relative to the rotation of the precision spindle of a roundness measuring instrument. Generally only the variations between surface and datum are recorded, the absolute size being lost (the instrument employs "radius suppression"). The dimensions have largely been measured by diametral techniques such as calipers or with variations on ring or

plug gauges. In determining the distance between features, the centres have been defined mechanically, for example by locating tapered pins into holes.

With modern methods, the distinction between the fields of metrology is becoming less clear. Co-ordinate measuring machines with position sensitive electrical probes now have resolutions which, while still much coarser than that of surface metrology instruments, are capable of detecting the major shape errors of surfaces. Further by their nature such machines must define circular features in terms of the position of several points on the periphery of those features. Thus both the dimensional and surface metrologist are faced with the problem of defining from a set of data points lying on the circumference of a nominal circle, but subject to errors both of measurement and of the surface, some form of "best-fit" circle to be used as a basis for the measurement. A commonly used criterion in both schools is the least squares (minimum quadrature) circle. Usually simple formulae for estimating the centre of this circle are quoted. The radius is then defined as the mean distance of the data points from this centre.

It is here that conflict arises for generally, given a set of data points (x_i, y_i), the dimensional and surface metrologists respectively regard the least squares centre to be at co-ordinates:

$$\left(\frac{1}{N}\sum_{1}^{N} x_i, \frac{1}{N}\sum_{1}^{N} y_i\right) \quad \text{or} \quad \left(\frac{2}{N}\sum_{1}^{N} x_i, \frac{2}{N}\sum_{1}^{N} y_i\right)$$

This difference could clearly be of some importance. It is examined by briefly reviewing the derivation of both these forms, and then discussing the general least squares circle problem.

2. Least Squares Circle: Approximations

National Standards for surface metrology define the centre of the least squares circle as that point from which the sum of the squares of the radial distances of a sufficient even number of equally spaced radial ordinates is minimised (1). While this statement is true, it is not strictly a definition of the least squares method and it is easy to read into it more than is actually implied.

Consider the situation depicted in Figure 1. It is desired to find the circle, centred at (a, b) and radius R, which is the best fit to a set of data points (x_i, y_i), or in polar co-ordinates (r_i, θ_i), of which point P is an example. From the above definition it would appear that the centre is that point for which

$$\sum_1^N s_i^2 = \sum_1^N ((x_i - a)^2 + (y_i - b)^2) \qquad (1)$$

is minimised. The minimum is obtained by equating the partial differentials to zero. For example:

$$\frac{\partial}{\partial a} \Sigma s_i^2 = \sum_1^N -2(x_i - a) = 0 \qquad (2)$$

giving

$$a = \frac{1}{N} \sum_1^N x_i \quad \text{and} \quad b = \frac{1}{N} \sum_1^N y_i \qquad (3)$$

This is one of the derivations given by Farkas (2). It has not made specific use of the statement that the ordinates should be equally spaced.

The derivation given in the above Standard takes the following form. The polar expression for an eccentric circle,

$$\rho = a\cos\theta + b\sin\theta + (R^2 - (a\sin\theta - b\cos\theta)^2)^{\frac{1}{2}} \qquad (4)$$

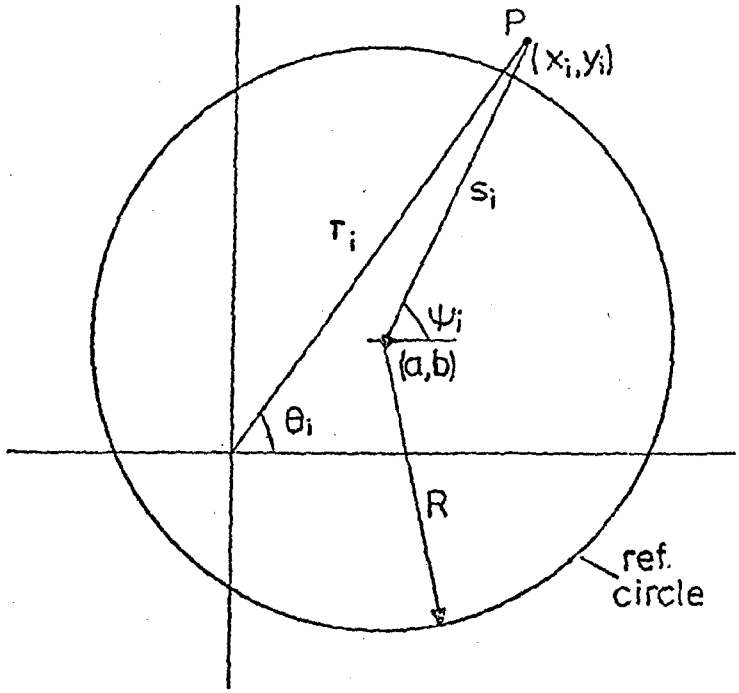


FIGURE A3.1: Reference Circle Definition

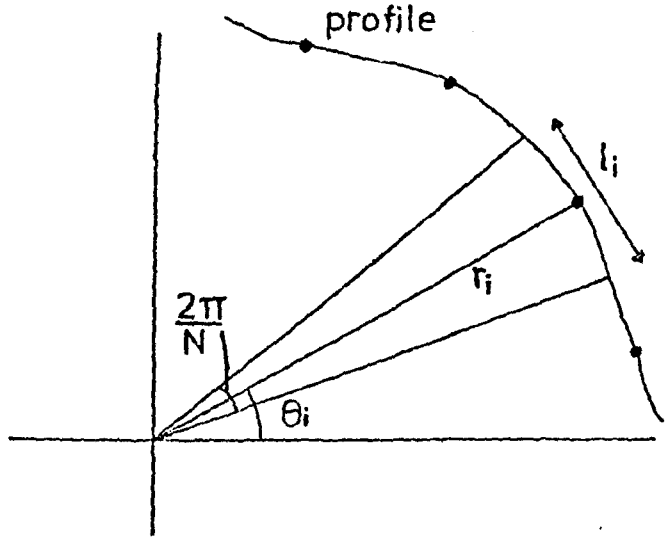


FIGURE A3.2: Profile Centroid Calculation

is approximated by

$$\rho = a\cos\theta + b\sin\theta + R \qquad (5)$$

for $a^2 + b^2 \ll R^2$, a condition almost always imposed by the nature of roundness measuring instruments. This is the limacon approximation which has particular advantages in surface metrology, notably its ability to compensate for radius-suppression (3). The least squares solution now requires that

$$I = \int_0^{2\pi} (r-\rho)^2 d\theta = \int_0^{2\pi} (r-a\cos\theta - b\sin\theta - R)^2 d\theta \qquad (6)$$

is minimised. Again equating partial differentials to zero gives, for example:

$$\frac{\partial I}{\partial a} = \int_0^{2\pi} -(r-a\cos\theta - b\sin\theta - R)\cos\theta \; d\theta = 0 \qquad (7)$$

$$a = \frac{1}{\pi}\int_0^{2\pi} r\cos\theta \, d\theta \qquad (8)$$

Now given sufficient evenly spaced points, the integral is replaced by a summation to give:

$$a = \frac{2}{N}\sum_1^N r_i\cos\theta_i \quad \text{and} \quad b = \frac{2}{N}\sum_1^N r_i\sin\theta_i \qquad (9)$$

Here the equispaced data condition is used to justify the replacement of integrals by summations but is still not used directly in the derivation.

3. The Least Squares Circle

The data which is available for the calculation of the reference circle is a series of discrete points. Thus a formal derivation should use a sampled data scheme not a continuous one. By analyzing

a problem in continuous form and then switching to a sampled form, the true nature of any assumptions made can be hidden. Formally least squares seeks to minimise the sum of squares of the residuals, that is the deviations of the data points from the solution figure, of any set of data points. In the case of circle fitting it is first, then, necessary to define what is meant by a residual.

Probably most metrologists and mathematicians would agree that the most logical form for the residuals is in terms of their radial distance from the circle centre:

$$\varepsilon_i = s_i - R \tag{10}$$

using again the notation of Figure 1. This form is implied by the definition given in Standards but is not quite that used in the associated derivation, where residuals are measured radially from the origin.

Using equation (10), the required minimisation will be of

$$I = \sum_1^N \left[((x_i-a)^2 + (y_i-b)^2)^{\frac{1}{2}} - R \right]^2 \tag{11}$$

where the positive square root will always be taken since it is a scalar length. This problem is non-linear in its parameters a, b and R and so is unlikely to be directly solvable. It may well exhibit multiple local minima. The solution will require that all the partial differentials are zero simultaneously giving:

$$\frac{\partial I}{\partial R}: \quad \sum_1^N \; (x_i-a)^2 + (y_i-b)^2)^{\frac{1}{2}} - R \; = \; 0$$

$$\frac{\partial I}{\partial a}: \quad \sum_1^N \; (x_i-a) - \frac{R(x_i-a)}{((x_i-a)^2 + (y_i-b)^2)^{\frac{1}{2}}} \; = \; 0 \tag{12}$$

$$\frac{\partial I}{\partial b}: \quad \sum_1^N \; (y_i-b) - \frac{R(y_i-b)}{((x_i-a)^2 + (y_i-b)^2)^{\frac{1}{2}}} \; = \; 0$$

The first of these indicates that the best fit radius will be the mean distance of the data points from the defined centre. The other two equations show the necessary condition for the 1/N solution:

$$R \sum_{1}^{N} \cos\theta_i = R \sum_{1}^{N} \sin\theta_i = 0 \qquad (13)$$

For this to be true, the data points must occur in diametral pairs with respect to the circle centre. The method thus suffers the theoretical objection that the data must be aligned to an initially unknown point and so an adequate experimental scheme cannot be produced.

Since only the linear least squares problem gives a general solution, the expression for the residuals must be linearised. This may be done by taking the first order terms of the Taylor series expansion of equation (10) about a point (x_o, y_o) close to the solution (a, b). (Since equation (10) is linear initially in R this does not enter the problem.) The expansion gives:

$$\varepsilon_i \simeq ((x-x_o)^2+(y-y_o)^2)^{\frac{1}{2}} - R - \frac{a(x-x_o) + b(y-y_o)}{((x-x_o)^2+(y-y_o)^2)^{\frac{1}{2}}} \qquad (14)$$

If (x_o, y_o) may be taken as the co-ordinate system origin, this reduces directly to the limacon approximation form. This will be used in the discussion here. The residuals for the set of data poits points may be expressed by the matrix equation.

$$\begin{pmatrix} \cos\theta_1 & \sin\theta_1 & 1 \\ \vdots & \vdots & \vdots \\ \cos\theta_N & \sin\theta_N & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ R \end{pmatrix} = \begin{pmatrix} r_1 \\ \vdots \\ r_N \end{pmatrix} - \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

$$H \cdot \underline{a} = \underline{r} - \underline{\varepsilon} \qquad (15)$$

The theory of least squares then gives as the solution:

$$\underline{\hat{a}} = (H^T H)^{-1} H^T \underline{r} \qquad (16)$$

H^T being the transpose of matrix H. So that the general solution to the least squares limacon can be stated as:

$$\begin{pmatrix} \Sigma\cos^2\theta & \Sigma\sin\theta\cos\theta & \Sigma\cos\theta \\ \Sigma\sin\theta\cos\theta & \Sigma\sin^2\theta & \Sigma\sin\theta \\ \Sigma\cos\theta & \Sigma\sin\theta & N \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{R} \end{pmatrix} = \begin{pmatrix} \Sigma r\cos\theta \\ \Sigma r\sin\theta \\ \Sigma r \end{pmatrix} \qquad (17)$$

where the limits and indices have been omitted for clarity, all being over i=1,N. There is strong motivation to attempt to select the measurement scheme, the values of θ_i, such that $H^T H$ is made diagonal. Not only does this make the solution very simple but also it makes the parameter estimates independent and helps to restrict estimation errors. This can be arranged by selecting a four-fold symmetry of points, such that for each θ_i there are also points taken at $\theta_i+\pi$ and $\theta_i\pm\pi/2$. Then the solution of equation (17) reduces to that given in Standards, equation (9). Such a scheme is quite practical since the angular position of the points are defined with respect to the origin, a point known before measurement. Note, however, that it is not sufficient to have any reasonably large even number of equispaced points as implied in Standards: the number must be a multiple of four. The general solution, equation (17), will naturally cope with any distribution of data points. It can thus deal with sections having holes in them and with the partial arc problems which have previously used special methods, again based on a continuous (integral) analysis (4). On such problems the off-diagonal terms will have an effect, in some cases exceeding in magnitude the diagonal terms. The strong interdependence of the parameter estimates can then cause large errors. The small difference in calculated values of the matrix co-efficients found by assuming integral forms may compound such errors and the continuum approach will break down.

The assumptions required to generate 1/N and 2/N types of solution can only both be satisfied when the centre of the circle lies at the origin. Then since both x_i and y_i must sum to zero for either approximation to be valid at that point, both give the correct result and there is no conflict. However, the fact alone that these sums are zero does not guarantee that the centre lies at the origin: there may be off-diagonal terms interacting with the radius estimate.

4. Centroid Methods

If a closed, non-circular profile is to be measured, a proposal (5) for defining a central reference point is to use the centroid of that profile, assuming it to be a uniform lamina. Other assumptions about the profile are also possible in this context. In effect, one quoted method of deriving the 1/N type centre solutions discussed here assumes the data points to be the positions of point masses. A third possibility here is to consider the profile to be a uniform wire. Assume that N radial measurements are equally spaced at angles θ_i about an origin. Providing N is reasonably large, the length of a sector associated with the ith point (see Figure 2) will be

$$l_i = r_i \cdot \frac{2\pi}{N} \tag{18}$$

and its associated x-moment:

$$M_x = l_i r_i \cos\theta_i \tag{19}$$

so that the x-co-ordinate of the centroid is

$$a = \left(\sum_1^N r_i^2 \cos\theta_i\right) / \left(\sum_1^N r_i\right) \tag{20}$$

the y-co-ordinate being similar but involving $\sin\theta_i$. Although this expression does not relate directly with the least squares solutions,

it is comparable. If the data points lie nearly on a circle and can be expressed as a mean value with small variations, δ_i, from it and if, also, $\cos\theta_i$ and $\sin\theta_i$ are arranged to sum to zero then equation (20) can be reduced to:

$$a \simeq \frac{2}{N}\sum_{1}^{N}\delta_i\cos\theta_i \qquad (21)$$

Thus there is some link between the centroid of the profile treated as a line and the least squares limacon. The association of the least squares approximations discussed here with different centroid methods is evidence of their different physical implications.

5. Concluding Remarks

The apparent conflict between different schools of metrology over the definition of the least squares circle centre is illusory. Both versions of standard formulae are approximations, their relative validity depending upon the actual positioning of the data points. The difficulties have arisen through lack of appreciation of these approximations. Some of the causes of this are informative: the general acceptance of a rather loose definition for what should be a precise concept and the masking of implicit assumptions by using continuous, integral analysis for essentially sampled systems.

The limacon approximation gives good centre estimates providing a reasonably accurate estimate of the centre is available (3). Its particular advantages when using a roundness instrument are such that its position is firmly established. Given also the normal practice and limitations of such instruments, the use of the 2/N-type formulae will often be safe. With, for instance, data from a co-

ordinate measuring machine it is much less clear what approximations may be used. Providing that the points are spread fairly regularly around the circumference of the section the best accuracy will probably be obtained by using the general form of the least squares limacon, an initial estimate of the centre being obtained by, say, the 1/N type approximation.

The profile centroid could also be used as an operational centre. From the discussion of the least squares problem it appears that treating the data as point sources is an oversimplification, but whether the profile is best regarded as a line or lamina is an open question, perhaps depending on the application. Indeed the question of "correctness" may be irrelevant: the important principles are that a method has good mathematical stability and is applied in a consistent manner by all users.

6. REFERENCES

1. BS 3730: British Standards Institute, 1972
2. FARKAS, J: Beitrage zur Mathematisch-statistischen Ermittlung von Formabweichungen. IV Oberflachenkolloquium, Karl Marx Stadt, 1976.
3. CHETWYND, D.G: Roundness Measurement using limacons. Metrology and Properties of Engineering Surfaces, Leicester 1979.
4. WHITEHOUSE, D.J: A best fit reference line for use in partial arcs. J. Phys. E: Sci. Instrum. 6, 1973.
5. THOMAS, T R and FORTUNE, N: An on-line computer system for measuring sectional geometry. J. Phys. E: Sci. Instrum. 8, 1975.

Appendix 4

Construction Aids for Limacon Roundness References

The following pages contain the main text of a note having the above title written initially as Rank Taylor Hobson Technical Note T60, September 1975 by the present author.

It describes the concept of the design of "limacon compasses", that is a mechanical device for drawing limacons and discusses the practicality of fitting limacons by template methods.

A third possiblity, not discussed, would be the use of an optical system by which a polar chart and the picture of a limacon on a cathode ray tube are superimposed. The construction of such a device is certainly possible without particular difficulty but would inevitably be quite expensive.

Construction Aids for Limacon Roundness References

Introduction

It is now widely established in theory and appreciated in practice that the use of circular reference lines for measuring out-of-roundness is inaccurate if the component is set eccentrically with respect to the measuring instrument axis. The shape of an eccentric component becomes distorted in the instrument output due to the radius suppression which is introduced. This distorted shape has been shown to be well approximated for practical situations by the limacon, which has the form

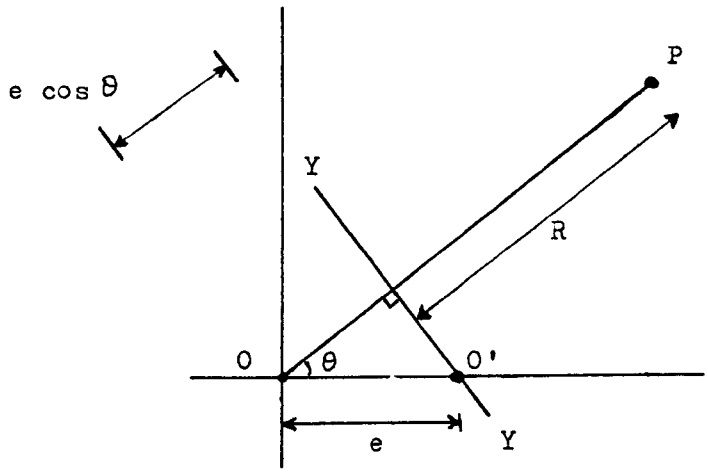
$$r(\theta) = R + e\cos\theta$$

where R is the component radius and e the relative eccentricity of component and instrument. As well as being a better approximation to the distorted shape than an offset circle, the limacon has the advantage of being easier to compute either by analogue or digital hardware. A disadvantage of the limacon and the reason that it has not been more widely adopted is that it is not easily estimated by hand from a profile graph. It is easily possible to produce reasonable looking fits for circular references (best fit or zonal) on graphs by using a template of concentric circles onto the graph with a pair of compasses. It is felt that the use of the limacon will not become a standard method (with circles used, for convenience, as an approximation for rapid hand checking of graphs) unless a potentially usable, although not necessarily convenient, 'hand' method is available. The most important aid to 'hand' assessment would be a pair of compasses which draw limacons instead

of circles since this could be used both to draw in the lines and to perform the search for best fit. A further aid to finding the fit would be a set of limacon templates for differing eccentricities, or, even better, an adjustable template.

Limacon Compasses

It is possible to construct an instrument, similar to a pair of compasses, which will describe a limacon figure. For a particular, given shape the easiest device to use would be a form of cam-follower attached to a radius arm which will superimpose a radial variation from the set value as the device is rotated. The device required here, however, should be easily and continuously variable in terms of eccentricity and the resultant deviations from circularity. Thus replaceable cams are not an attractive proposition. Instead a geometrical construction may be used. The diagram of figure 1 shows the construction of point P on the periphery of the limacon centred at O with eccentricity e such that the apparent circular centre is at O'. The total length from O to P will then be R + ecos providing that the line YY is perpendicular to OP. The limacon compasses can be built directly in the form indicated by figure 1, the main features being indicated schematically by figure 2. The device consists of a fixed length radius arm on the end of which is the pencil P. The other end of this arm is attached to a cross-member and also to a telescopic arm which holds the pivot point, O.A slot in the cross-member is at right angles to the radius arm and engages a fixed pin O'. The distance OO' defines the eccentricity of the limacon. As the radius arm is rotated about O, the



Line YY is always perpendicular to line OP

$$r(\theta) = R + e \cos\theta$$

FIGURE A4.1: Geometry of Limacon Compasses

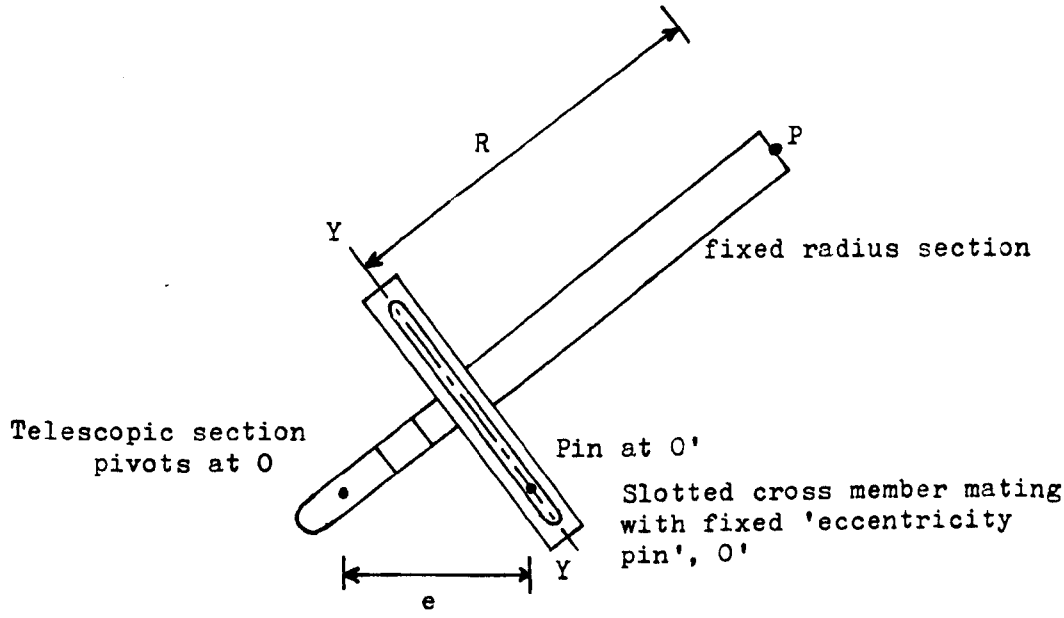
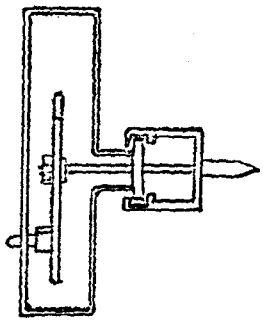


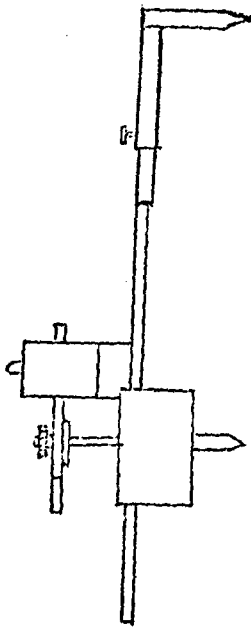
FIGURE A4.2: Schematic form of Limacon Compasses

fixed positions of O and O' together with the rigid right angle between YY and OP will cause the length of the telescopic section to vary as $e\cos\theta$ and so the pencil will draw a limacon figure. The sketches of figure 3 indicate practical way of obtaining the necessary movements. The feasibility of such a system has been demonstrated by the construction of a working cardboard model. For use on Talyrond graphs it is useful to have a base with a central spigot which can be used to locate the pivot of the compasses onto the chart centre.

Templates

It seems unlikely that an adjustable template, that is one which shows a set of concentric limacons which have a settable eccentricity, could be produced in a practical way. The most obvious approach would be to use the constant diameter property of the figure. A set of equal diametral wires, which all are constrained to pass through the mid point of a fixed equal wire could be constructed such that as the ray at right angles to the fixed wire is moved, the other rays move lesser amounts according to the limacon shape. If sufficient rays were used an elastic web attached to them would conform closely to the desired limacons. The mechanism needed to move the rays would be the right angled slides used in the compasses (section 2) and the whole would quickly become bulky. In any case if very many rays are used, the transparency of the template will be curtailed. Another possible mechanism would be to link the rays by a circumferential web of stiff elastic material (a spring!). If now the movable ray is adjusted, the constraints of minimum bending energy





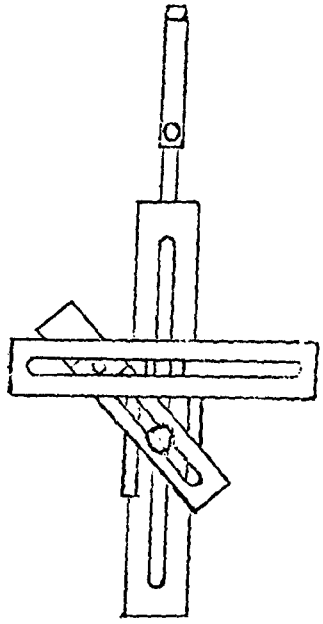


FIGURE A4.3: Sketches of Limacon Compasses

will tend to pull the other rays into a suitable shape. (Whether this shape is exactly a limacon has not been proved.) This system would probably need less rays and so be more transparent. Its difficulty lies in the variation of the circumference of the limacon with it's eccentricity - some method of tensioning the web will need to be devised.

A set of templates each consisting of limacons at one eccentricity could very easily be produced. The whole range of eccentricities which can occur on a Talyrond graph could probably be covered adequately by a relatively small number of templates, say 5-10. It would be necessary to ensure that all templates were located to the proper centre so that a base plate with a spigot for mounting both the graph and the templates would be needed. This system is likely to be not too tedious to operate and should give a better approximation to the true reference than a circular template.

Appendix 5.

Selected Extracts from Programs

Included here are extracts from the listings of some of the programs developed for this work (see chapter 9). An attempt has been made to preserve just the essential working heart of these programs so that the extracts indicate just the work involved in fitting the reference and not that concerned with acquiring data and displaying results. The listings are in FORTRAN IV but are only slightly annotated and it is not envisaged that they be studied unless specific need arises. They are included mainly for completeness. No flow-charts are given.

Certain conventions occur in all the extracts. The profile data (r_i) are in an array Y(I). Sine and cosine look-up tables having suitably indexed intervals are used, values being held in arrays S(I) and C(I) or SI(I) and CO(I). Variable SF is a scaling factor converting the original units of the data to μm and TP is the constant 2π. Generally variables A, B, R are the limacon or circle parameters as used in this text. However exceptions occur in those programs using matrix manipulations, DLS2 and DEVS, where arrays A(3) and PI(3) respectively contain the parameters. In those programs B( ) is an array having other manipulative use. The number of data points in a full revolution (usually 512) is represented by N and in cases where there may be less than a full revolution of data to process, the actual number of points is in NOG.

P R O G R A M D E N T

Least squares limacon by traditional formulae

```
0030          TH2=TP*(NOG-1)
0031          SR=.5*(Y(1)+Y(NOG))
0032          SRS=.5*Y(NOG)*SI(NOG)
0033          SRC=.5*(Y(1)+Y(NOG)*CO(NOG))
0034          DO 10 I=2,NOG-1
0035          SR=SR+Y(I)
0036          SRC=SRC+Y(I)*CO(I)
0037          SRS=SRS+Y(I)*SI(I)
0038       10 CONTINUE
0039          C4=SI(NOG)
0040          C5=1.-CO(NOG)
0041          N2=2*NOG-1
0042          IF(N2.GT.N)N2=N2-N
0043          C6=TH2/2-SI(N2)/4
0044          C7=(1.-CO(N2))/4
0045          C8=TH2/2+SI(N2)/4
0046          A=C6-C5*C5/TH2
0047          B=C4/TH2
0048          C=C4*C5/TH2-C7
0049          D=C5/TH2
0050          F=C8-C4*C4/TH2
0051          E=A*F-C*C
0052          XB=(A*(SRC-B*SR)+C*(SRS-D*SR))/E*TP
0053          YB=(F*(SRS-D*SR)+C*(SRC-B*SR))/E*TP
0054          R=(SR*TP-XB*C4-YB*C5)/TH2
```

31-38 Integration by trapezium rule

PROGRAM DLS2

Least squares limacon by formal evaluation

```
0033   C  SET UP REJECTED BITS
0034       YMIN=1E-20
0035   355 WRITE(1,350)
0036   350 FORMAT("REJECT: _")
0037       READ(1,*)IS,IF
0038       IF(IS.LE.0.OR.IF.GT.512)GOTO 360
0039       DO 365 I=IS,IF
0040   365 Y(I)=YMIN
0041       GO TO 355
0042   C  GENERAL L.S. SOLUTION
0043   360 DO 9 I=1,3
0044       B(I)=0
0045       DO 9 J=1,3
0046     9 HH(I,J)=0
0047       DO 10 I=1,NOG
0048       IF(Y(I).EQ.YMIN)GOTO 10
0049       HH(1,1)=HH(1,1)+1
0050       HH(1,2)=HH(1,2)+CO(I)
0051       HH(1,3)=HH(1,3)+SI(I)
0052       HH(2,2)=HH(2,2)+CO(I)*CO(I)
0053       HH(2,3)=HH(2,3)+CO(I)*SI(I)
0054       HH(3,3)=HH(3,3)+SI(I)*SI(I)
0055       B(1)=B(1)+Y(I)
0056       B(2)=B(2)+Y(I)*CO(I)
0057       B(3)=B(3)+Y(I)*SI(I)
0058    10 CONTINUE
0059       HH(2,1)=HH(1,2)
0060       HH(3,1)=HH(1,3)
0061       HH(3,2)=HH(2,3)
0062       CALL ULTR(HH,U,3,3,IER)
0063       IF(IER.NE.0)GOTO 1
0064       CALL SLEQ(U,B,A,3,3)
```

33-41 Interactive loop to remove profile sections from evaluation

62-64 ULTR and SLEQ subroutines give Choleski solution to $\underline{HH}.\underline{A} = \underline{B}$ for $\underline{A}$

PROGRAM DIMA

Minimum zone limacons by exchange algorithm

```
C  FIRST ESTIMATE USE QUADRANT POINTS
      K=0
      IAD(1)=1
      IAD(2)=129
      IAD(3)=257
      IAD(4)=385
      A=(Y(1)-Y(257))/2.0
      B=(Y(129)-Y(385))/2.0
      R=(Y(1)+Y(129)+Y(257)+Y(385))/4.0
      H=(Y(1)+Y(257)-Y(129)-Y(385))/4.0
    9 ZAD(3)=ZAD(1)
      ZAD(2)=-ZAD(1)
      ZAD(4)=ZAD(2)
      K=K+1
C  FIND LARGEST NON-ENCLOSED POINT
      J=0
      DMAX=ABS(H)-.1
      DO 10 I=1,N
      YR=ABS(Y(I)-(R+A*C(I)+B*S(I)))
      IF(YR.LE.DMAX)GO TO 10
      DMAX=YR
      J=I
   10 CONTINUE
      IF(J.EQ.0)GO TO 6
      EJ=Y(J)-(R+A*C(J)+B*S(J))
C  POSITION J IN IAD
      DO 11 I=1,4
      IF(J.LT.IAD(I))GO TO 12
   11 CONTINUE
      I=5
   12 IF(ISIGN(1,H)*(-1)**(I-1).NE.ISIGN(1,EJ))I=I-1
      IF(I.LE.0)GO TO 31
      IF(I.LE.4)GO TO 13
      IAD(1)=IAD(2)
      IAD(2)=IAD(3)
      IAD(3)=IAD(4)
      I=4
      GO TO 13
   31 IAD(4)=IAD(3)
      IAD(3)=IAD(2)
      IAD(2)=IAD(1)
      I=1
   13 IAD(I)=J
C  FIND NEW REFERENCE

      A=((Y(I1)-Y(I3))*(S(I2)-S(I4))-(Y(I2)-Y(I4))*(S(I1)-S(I3)))/
     #  ((C(I1)-C(I3))*(S(I2)-S(I4))-(C(I2)-C(I4))*(S(I1)-S(I3)))
      IF(ABS(S(I1)-S(I3)).LT.ABS(S(I2)-S(I4)))GO TO 14
      B=((Y(I1)-Y(I3))-A*(C(I1)-C(I3)))/(S(I1)-S(I3))
      GO TO 15
   14 B=((Y(I2)-Y(I4))-A*(C(I2)-C(I4)))/(S(I2)-S(I4))
   15 R=(Y(I1)+Y(I2)-A*(C(I1)+C(I2))-B*(S(I1)+S(I2)))/2.0
      H=Y(I1)-(R+A*C(I1)+B*S(I1))
      GO TO 9
```

PROGRAM DIPL

Ring or plug limacon by exchange algorithm

```
0026  C  FIRST ESTIMATE: MAX DIAMETER
0027        J=1
0028        V=Y(1)+Y(N2+1)
0029        DO 10 I=2,N2
0030        T=Y(I)+Y(I+N2)
0031        IF(T.LE.V)GO TO 10
0032         V=T
0033        J=I
0034     10 CONTINUE
0035        R=V/2.
0036        V=(Y(J)-Y(J+N2))/2.
0037        A=V*C(J)
0038        B=V*S(J)
0039        K=0
0040        CALL VIOL(A,B,R,MAX)
0041        IF(MAX.EQ.0)GO TO 6
0042  C  THREE POINT CONTACT: ORDER J,J+N2,MAX
0043        I3=J+N2
0044        IF(MAX.LT.I3)GO TO 11
0045        I2=I3
0046        I3=MAX
0047     13 I1=J
0048        GO TO 15
0049     11 IF(MAX.LT.J)GO TO 12
0050        I2=MAX
0051        GO TO 13
0052     12 I2=J
0053        I1=MAX
0054  C  CALCULATE REFERENCE I1,I2,I3
0055     15 S12=S(I1)-S(I2)
0056        S32=S(I3)-S(I2)
0057        C12=C(I1)-C(I2)                    ►(C12*S32-C32*S12)
0058        C32=C(I3)-C(I2)
0059        A=((Y(I1)-Y(I2))*S32-(Y(I3)-Y(I2))*S12)/
0060        IF(ABS(S12).LT.ABS(S32))GO TO 35
0061        B=(Y(I1)-Y(I2)-A*C12)/S12
0062        GOTO 36
0063     35 B=(Y(I3)-Y(I2)-A*C32)/S32
0064     36 R=Y(I1)-A*C(I1)-B*S(I1)
0065        K=K+1
```

28 N2 is 180°

41 No violation of reference by data: end of iterations

43-53 Not diametral contact, sort out angles I1, I2, I3 of three contact points

```
0066          CALL VIOL(A,B,R,MAX)
0067          IF(MAX.EQ.0)GO TO6
0068    C  PLACE MAX IN REFERENCE USING PI CRITERION
0069          IF(MAX.GT.I1)GO TO 20
0070          IF(I2-MAX.LT.N2)GO TO 21
0071          I3=I2
0072          I2=I1
0073       21 I1=MAX
0074          GO TO 26
0075       20 IF(MAX.GT.I2)GO TO 22
0076          IF(I3-MAX.GT.N2)GO TO 21
0077       23 I2=MAX
0078          GO TO 26
0079       22 IF(MAX.GT.I3)GO TO 24
0080          IF(MAX-I1.LT.N2)GO TO 23
0081          GO TO25
0082       24 IF(MAX-I2.LT.N2)GO TO 25
0083          I1=I2
0084          I2=I3
0085       25 I3=MAX
0086       26 GO TO 15
0087    C  END OF ITERATIONS
```

69-86 Exchange contacts, maintaining order, according to 180^{0} rule

```
0140          SUBROUTINE VIOL(A,B,R,MAX)
0141          COMMON Y(512),S(640)
0142          DIMENSION C(512)
0143          EQUIVALENCE (C,S(129))
0144          N=512
0145          MAX=0
0146          V=0.
0147          V=0.1
0148          DO 30 I=1,N
0149          T=Y(I)-(A*C(I)+B*S(I)+R)
0150          IF(T.LE.V)GO TO30
0151          V=T
0152          MAX=I
0153       30 CONTINUE
0154          RETURN
```

140 VIOL finds in MAX address of largest violation of limacon (A,B,R) by data. 0 means no violation.

Ring limacon by direct use of revised simplex on dual linear programme

```
0031  C   REVISED SIMPLEX: SET UP INITIAL CONDITIONS
0032        ART=.TRUE.
0033        Y(513)=0.
0034        DO 20 I=1,3
0035        BP(I)=0
0036        CB(I)=0.
0037        IC(I)=513
0038        DO 20 J=1,3
0039        B(I,J)=0
0040        IF(I.EQ.J)B(I,J)=1.
0041   20   CONTINUE
0042        BP(3)=1.
0043        KOUNT=0
0044  C   GET LATEST SIMPLEX MULTIPLIERS: PI=A,B,R
0045       15 DO 16 I=1,3
0046       16 PI(I)=-B(1,I)*CB(1)-B(2,I)*CB(2)-B(3,I)*CB(3)
0047  C   FIND PIVOT COL  MAX C'(J) FOR C'(J)>0
0048  C   IF ARTIFICIALS PRESENT USE WAGNER W FOR C
0049        J=0
0050        TM=.001
0051        DO 10 I=1,N
0052        IF(ART)GO TO 22
0053        T=Y(I)+(C(I)*A+S(I)*D+R)
0054        GO TO 21
0055     22 T=C(I)*(1.+A)+S(I)*(1.+D)+1.+R
0056     21 IF(T.LE.TM)GO TO 10
0057        TM=T
0058        J=I
0059     10 CONTINUE
0060        IF(J.EQ.0)GO TO 30
0061        KOUNT=KOUNT+1
0062  C   CALCULATE CURRENT VALUES PJ OF CHOSEN COLUMN
0063  C   AND FIND ROW K FOR MIN. THETA VALUE
0064        TM=1E30
0065        K=0

0066        DO 11 I=1,3
0067        PJ(I)=B(I,1)*C(J)+B(I,2)*S(J)+B(I,3)
0068        IF(PJ(I).LE.0.)GO TO 11
0069        T=BP(I)/PJ(I)
0070        IF(T.GT.TM)GO TO 11
0071        IF(T.EQ.TM)GO TO 31
0072     23 TM=T
0073        K=I
0074     11 CONTINUE
0075        IF(K.EQ.0)GO TO 32
```

PROGRAM DEVS continued

```
0076  C     PIVOT IS (K,J): SET BASIC VAR IN CONSTRAINT ORDER
0077  C     UPDATE BASIS AND R.H.S.
0078            IC(K)=J
0079            CB(K)=Y(J)
0080            IF(ART)CB(K)=C(J)+S(J)+1
0081            DO 12 I=1,3
0082       12   B(K,I)=B(K,I)/PJ(K)
0083            BP(K)=BP(K)/PJ(K)
0084            DO 13 I=1,3
0085            IF(I.EQ.K)GO TO 13
0086            BP(I)=BP(I)-BP(K)*PJ(I)
0087            DO 14 J=1,3
0088       14   B(I,J)=B(I,J)-B(K,J)*PJ(I)
0089       13   CONTINUE
0090            GO TO 15
0091  C     OPTIMALITY FOUND
0092       30   IF(.NOT.ART)GO TO 6
0093            ART=.FALSE.
0094            DO 36 I=1,3
0095       36   CB(I)=Y(IC(I))
0096            GO TO 15
0097  C     INFEASIBILITY !
0098       32   WRITE(1,110)
0099      110   FORMAT("ERROR CONDITION")
0100            GO TO 1
0101  C     SIMPLEX TIE BREAK FOR THETA
0102       31   IF(K.EQ.0)GO TO 32
0103            DO 33 KK=1,N
0104            TI=(B(I,1)*C(KK)+B(I,2)*S(KK)+B(I,3))/PJ(I)
0105            TK=(B(K,1)*C(KK)+B(K,2)*S(KK)+B(K,3))/PJ(K)
0106            IF(TI.EQ.TK)GO TO 33
0107            IF(TI.GT.TK)GO TO 11
0108            GO TO 23
0109       33   CONTINUE
0110            GO TO 11
0111  C     END OF ITERATIONS
```

Program is basic revised simplex. Note identification of variables and arrays.

ART - (logical) true when infeasible so using Wagner w

B - inverse of basis

BP - R.H.S. vector of tableau

CB - basic elements of objective function

D - y-component of eccentricity

PI - simplex multipliers

Also I(1),(2),(3) equivalent to A,D,R

P R O G R A M D R C 2

Interactive ring circle fitting

Only interactive section is given, it is proceeded by finding the ring limacon by exchange algorithm

```
0096  C   "CIRCLE" CALCULATIONS
0097        WRITE(1,110)
0098    110 FORMAT(" RADIUS (MM) _")
0099        READ(1,*)R0
0100        R0=R0*1D3/SF
0101     60 CONTINUE
0102        T1=TP*(I1-1)
0103        T2=TP*(I2-1)
0104        T3=TP*(I3-1)
0105        CALL C3P(Y(I1),Y(I2),Y(I3),T1,T2,T3,R0,A,B,R)
0106        T1=R0+R
0107        G=SQRT(A*A+B*B)/T1
0108        P=0.
0109        V=0.
0110        MAX=0
0111        DO 40 I=1,N
0112        T=Y(I)-A*C(I)-B*S(I)-R+(A*S(I)-B*C(I))**2/T1/2
0113        IF(P.GE.T)GO TO 42
0114        P=T
0115        MAX=I
0116     42 CONTINUE
0117        IF(V.GT.T)V=T
0118        IF(ISSW(8))41,40
0119     41 IF(T.GT.0.)WRITE(6,140)I,T
0120    140 FORMAT(" POS:",I4," Y=",F8.2)
0121     40 CONTINUE
0122        WRITE(6,111)G
0123    111 FORMAT(" CIRCLE DATA -2ND. ORDER-  GAMMA=",F8.6)
0124        P=P*SF
0125        V=V*SF
0126        T1=A*SF
0127        T2=B*SF
0128        T3=R*SF
0129        WRITE(6,112)T1,T2,T3,P,V
0130    112 FORMAT(" A=",F8.3," UM  B=",F8.3," UM  R=",F8.3," UM"/
           1   " MAX PEAK =",F8.5," UM  MIN-VALLEY =",F8.4," UM"/)
0131  C   INTERACTIVE CIRCLE FIT
0132        T1=R0+R
0133        CALL COPS(I1,A,B,T1)
0134        CALL COPS(I2,A,B,T1)
0135        CALL COPS(I3,A,B,T1)
0136        WRITE(6,145)MAX
0137    145 FORMAT(" MAX. POINT AT",I4)
0138        IF(ISSW(9).GE.0)GOTO 1
0139        WRITE(1,160)
0140    160 FORMAT(" NEW CONTACTS?")
0141        READ(1,*)I1
0142        IF(I1.LE.0)GO TO 1
```

PROGRAM DRC2 continued

```
0144          CALL COPS(I1,A,B,T1)
0145          READ(1,*)I2
0146          CALL COPS(I2,A,B,T1)
0147          READ(1,*)I3
0148          CALL COPS(I3,A,B,T1)
0149          GOTO 66
0150        4 CONTINUE
0151          END
0152          SUBROUTINE COPS(I,A,B,R)
0153          COMMON Y(512),S(648)
0154          DIMENSION C(512)
0155          EQUIVALENCE(C,S(129))
0156          XO=A*C(I)+B*S(I)+SQRT(R*R-(A*S(I)-B*C(I))**2)
0157          XP=XO*C(I)
0158          YP=XO*S(I)
0159          XO=2.*A-XP
0160          YO=2.*B-YP
0161          IOP=ATAN2(YO,XO)/.01227
0162          IF(IOP.LT.0)IOP=IOP+512
0163          IOP=IOP+1
0164          WRITE(6,100)I,IOP
0165      100 FORMAT(" CONTACT ",I4," OPPOSITE ",I4)
0166          RETURN
```

99 RO absolute radius - double precision

108-121 find peak & valley relative to specified circle. Peak is zero if no violation

152 COPS gives index of other end of circle diameter

```
0184          SUBROUTINE C3P(Y1,Y2,Y3,T1,T2,T3,R0,A,B,R)
0185          DOUBLE PRECISION R3,D,R1,R2,R3,TA,TB
0186          R3=R0+Y3
0187          R2=R0+Y2
0188          R1=R0+Y1
0189          D=(DSIN(DBLE(T2-T1))/R3+DSIN(DBLE(T1-T3))/R2
0190         1  +DSIN(DBLE(T3-T2))/R1)*200
0191          TA=((R1*DSIN(DBLE(T2))-R2*DSIN(DBLE(T1)))/R3
0192         1    +(R3*DSIN(DBLE(T1))-R1*DSIN(DBLE(T3)))/R2
0193         2    +(R2*DSIN(DBLE(T3))-R3*DSIN(DBLE(T2)))/R1)/D
0194          TB=-((R1*DCOS(DBLE(T2))-R2*DCOS(DBLE(T1)))/R3
0195         1    +(R3*DCOS(DBLE(T1))-R1*DCOS(DBLE(T3)))/R2
0196         2    +(R2*DCOS(DBLE(T3))-R3*DCOS(DBLE(T2)))/R1)/D
0197          R3=R1-TA*DCOS(DBLE(T1))-TB*DSIN(DBLE(T1))
0198          R2=TA*DSIN(DBLE(T1))-TB*DCOS(DBLE(T1))
0199          R1=R3*DSQRT(1D0+R2/R3*R2/R3)
0200          A=TA
0201          B=TB
0202          R=R1-R0
0203          RETURN
```

184 C3P calculates circle through three points, specified in terms of R, δR_i, θ_i to maintain precision

References

1. ANSI B89.3.1 - American National Standards: Measurement of Out-Of-Roundness. ASME 1972.
2. BS 3730 - Assessment of Departures from Roundness. BSI 1974.
3. REASON, R. E. - Report on the Measurement of Roundness, Rank Taylor Hobson, 1966.
4. WHITEHOUSE, D. J. - A best fit reference line for use in partial arcs - J. Phys. E. Sci. Instrum. 6 1973
5. SIDDALL, G. J. - Ph.D. Thesis, University of Aberdeen, 1975
6. THOMAS, T. R., FORTUNE, N. - An On-Line Computer System for Measuring Sectional Geometry, J. Phys. E: Sci. Instrum. 8 1975
7. FARKAS, J. - Beitrage zur Mathematisch-statistichen, Ermittlung von Formabsweichungen, IV Oberflachenkolloquium, Karl Marx Stadt 1977.
8. AVDULOV, A. N. et al - Some Aspects of Measurement of Roundness - Proc. I. Mech. E. Vol. 182 pt. 3K 1968
9. WHITEHOUSE, D. J. - Modern Trends in the Measurement of Surfaces - Proc. Inst. Belgium Soc. Mech. Eng. Vol 21 1975
10. SCHEIDING, U., von WEINGRABER H. - Bezugssystem Stutzzylinder - z. ind. Fertig. 66 73-76 1976
11. GOTO, M., IIZUKA, K. - An analysis of the relation between the minimum zone deviation and the least squares deviation in circularity and cylindricity - Paper 10.8 Proc. Int. Conf. Prod. Eng. New Dehli 1977.

12. MURTHY, T. S. R. et al - Evaluation of Spherical Surfaces - Metrology and Properties of Engineering Surfaces, Int. Conf., Leicester 1979.

13. BS 308: Part 3 - Geometric Tolerancing - BSI 1972

14. REASON, R. E. in M. J. NEALE (ed) Tribology Handbook: F2 Surface Topography - Newnes Butterworth 1973

15. LOTMAR, W - The use of the Talyrond 73 apparatus for the measurement of diameter variation of cylinders with height J. Phys. E: Sci. Instrum. 7, 1974

16. IIZUKA, K., GOTO, M. - Application of multiple regression analysis in measurement and calibration - ACTA IMEKO 1973

17. GOTO, M., IIZUKA, K. - A method for evaluating form errors of cylindrical parts - J. of Japan Soc. Prec. Eng. 41,5 1975

18. TSUKADA, T et al - An evaluation of form errors of cylindrical machine parts by spiral tracing method Proc 18th MTDR 1977

19. MOORE, W. R. - Foundations of Mechanical Accuracy. Section 4 Roundness - The Moore Special Tool Co. USA 1970

20. WHITEHOUSE, D. J. - Some theoretical aspects of error separation techniques in surface metrology - J Phys. E: Sci Instrum. 9 1976

21. CHETWYND, D. G., SIDDALL, G. J. - Improving the accuracy of roundness measurement - J. Phys. E: Sci. Instrum. 9 1976

22. GARRATT, J. D. - MSc. Thesis, University of Aberdeen 1977

23. REASON, R. E. - Report on the Measurement of Roundness - Rank Taylor Hobson, 1966

24. SIDDALL, G. J., CHETWYND, D. G. - Measurement of Surface Geometry of Hip Prostheses - I. Mech. E. "Engineering in Medicine" 4,3 1975

25. BULL, G - Computational Methods and Algol - Harrap, 1966

26. WAGNER, H. M. - Principles of Operations Research 2nd Edition Prentice Hall 1975

27. HADLEY. G. - Linear Programming Addison Wesley 1969

28. OSBORNE, M. R., WATSON, G. A. - On the best Chebyshev Approximation - Computer Jnl. 10 1968

29. CHETWYND, D. G., PHILLIPSON, P. H. - An Investigation of Reference Criterion used in Roundness Measurement - J. Phys. E., in press, 1980

30. CHETWYND, D. G. - Dimensioning Nominal Circles: The Resolution of Conflicting Ideas - ACTA IMEKO 1979

31. KINSEY, D., CHETWYND, D. G. - Some aspects of the use of digital computers to the on-line measurement of surfaces ACTA IMEKO 1973

32. WIRTZ, A., MADUDA, M. - Fertigungsgerechte Rundheitsmessung

33. DAMIR, M. N. H. - Approximate Harmonic models for roundness profiles Int. Conf. Metrology and Properties of Eng. Surfaces Leicester 1979

34. WHITEHOUSE, D. J., HUNT, R. T. - Private Communication

35. ANON - High Speed Cylindricity Measurement Talynews 2, Rank Taylor Hobson 1975