Applications of linear programming to engineering metrology

D G Chetwynd, BA, PhD
Centre for Micro-engineering and Metrology, Department of Engineering, University of Warwick, Coventry

Computer-based metrology now makes use of exchange algorithms for computing best-fit geometries in which the solution is obtained by a series of iterations each involving the exchange of one previously unused data point for one of the dominant points of the previous iteration according to formal rules. While conferring large advantages in specific circumstances, their major significance is that of being examples of a class of optimization of much wider applicability.

This paper examines the theoretical basis of these algorithms in linear programming and develops a general approach to the solution of this class from which new applications can be derived. To maintain an engineering context in the analysis, practical examples are used, mainly from the field of roundness measurement.

NOTATION

a, b, c, m	geometrical parameters of linear reference figures
h	half zone width between parallel lines
R	radius of circle
(r_i, θ_i) or (ρ, θ)	radial ordinates from instrument centre
S_i	takes value ± 1 to show type of contact of ith point
(x_i, y_i, z_i)	profile ordinate in Cartesian frame
Z	scalar function to be optimized
Δ_{ij}	determinant using ith and jth ordinates

Also, used in discussing the general nature of linear programs, Sections 3 and 4:

A, **K**	rectangular matrices of coefficients
b, **c**	column matrices of coefficients
x	column matrix of parameters
β	square matrix, inverse of the basis

1 INTRODUCTION

The ready availability of digital computation has influenced all branches of engineering metrology. Arguably, the greatest impact has occurred in the measurement of the shapes of workpieces in order to determine how well they comply with the ideal geometries specified in the design drawings. This involves the fitting of geometrically defined shapes, called reference figures, to sets of measured data. Traditionally, this was done by hand. Reference figures were fitted to magnified graphs of the surface profile by trial and error. National standards reflect this by incorporating intuitively sensible and useful criteria of fit such as enclosing the measurements within the smallest circumscribing circle or between two parallel figures of minimum separation. In producing automatic instruments, the intuitive skill of the human must be replaced by a mathematically rigorous algorithm which follows the requirements of the Standards while operating at a speed acceptable for industrial use. It is preferable, again for reasons of standardization, that the algorithms be readily expressible in geometric terms so as to be usable, and intelligible, by practising metrologists perhaps unfamiliar with powerful, modern methods of mathematics.

These requirements have been admirably met by a series of geometrical exchange algorithms (**1**). The reference figure to a set of data points is found by first fitting a trial figure to a subset of the data and then performing a series of iterations at each of which exactly one datum point which violates the criteria of fit is exchanged with one of the defining set to create a new trial solution. The power of the method lies in the rules governing permissible exchanges which ensure an orderly convergence on to the true solution. Unfortunately, the exchange rules are usually highly problem-specific. The concepts have a wide range of applicability but it cannot be assumed that the rules will work in other situations. Such generalizations must be built upon a mathematically rigorous foundation. The concern of this paper is to develop the theory, and an approach to derivation, which underlies all the metrological exchange algorithms. In order to give a physical context, specific examples are used for this purpose. Roundness analysis, and a little on flatness, is used since it is easily visualized, an important class of measurement in its own right and highlights all the points relevant to industrial metrology. The nature of the instrumentation and the profile distortions caused by it, which are of considerable metrological importance and which influence the choice of reference model (**1**, **2**), will not be discussed. Here it is assumed that, by some means, a set of data points has been obtained which represents either spot-heights of a real, nominally flat surface relative to an ideal plane placed approximately parallel to it or distances to points on the surface of a nominally circular cross-section measured radially from a point close to the true centre of that section.

It will become apparent that all the reference figures have certain features in common. A degree of imprecision in their definitions must be clarified in a manner sympathetic to Standards. Then, all may be expressed mathematically as constrained optimizations. Each

The MS was received on 26 July 1984 and was accepted for publication on 6 December 1984.

involves minimizing or maximizing an objective function while keeping other functional relationships within predefined bounds. If all these relationships can be expressed as linear functions of a set of parameters, the optimization becomes a linear program. The theory of linear programming is fundamental to a generalized understanding of the exchange algorithms.

2 THE NATURE OF REFERENCE FITTING

One common approach to fitting a reference figure to a set of data is to use least-squares estimation (usually with some degree of approximation), but here attention will focus on other methods much used by hand and less obviously computed.

The straightness of profile is measured relative to a pair of parallel straight lines so placed that the profile is contained between them and their separation, measured in a direction perpendicular to the general trend of the profile, is minimized. The idea embodied in Standards, that the general trend, or orientation, of the surface can be identified prior to measurement is a source of practical difficulties. Here, it is taken to imply that there is a known Cartesian frame, conveniently called the instrument coordinate frame, relative to which mathematical formulations which are not rotationally invariant may be expressed without significant error. This is a reasonable expectation with data from surface metrology instruments but it should be used only with extreme care with, for example, data from a coordinate measuring machine. The flatness of a plane is similarly measured from a pair of minimum separation parallel planes.

Roundness can be measured from any of three reference circles: the minimum radius circumscribing circle; the maximum radius inscribing circle; and the minimum radial zone circles. Definitions of the first two are self-evident while the third consists of two concentric circles of minimum radial difference which contain the profile between them.

Other, more complex, geometrical forms are generally constructed from a combination of the above or from simple variants of them as in minimum enclosing sphere, minimum zone cylinders and so on.

Once the transition to a sampled data scheme is performed prior to digital computation, all these references may be written in immediately recognizable mathematical forms. For example, the straightness reference, expressed in instrument coordinates, is: given a sequence of Cartesian datum points (x_i, y_i)

$$\begin{aligned} &\text{minimize} \quad Z = h \\ &\text{subject to} \quad mx_i + c + h \geqslant y_i \\ &\qquad\qquad\quad mx_i + c - h \leqslant y_i \end{aligned} \tag{1}$$

for all (x_i, y_i) simultaneously. This illustrates a convenient parametrization, namely a single line (slope m, intercept c) together with a zone of acceptability of width $2h$ set symmetrically about it. Equation (1) is a linear program in (m, c, h). [It is also a simple form of the minimax polynomial fit for which the so called Stiefel exchange algorithm offers an efficient solution. This may be derived from, and owes its efficiency to the properties of, the associated linear program (**3**).]

Standards present a method, originally for calculating

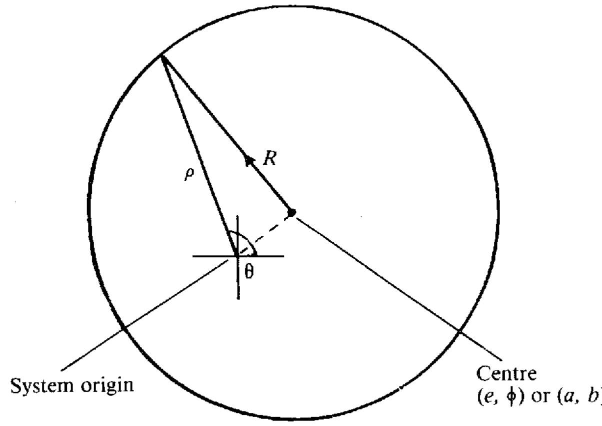

Fig. 1 Definitions for the eccentric circle

the least squares parameters, which has been extensively studied (**4**, **5**, **6**) and is known as the 'limacon approximation' for roundness measurement. Referring to the notation of Fig. 1, the eccentric circle is reduced, providing $e \ll R$:

$$\begin{aligned} \rho &= e \cos(\theta - \phi) + [R^2 - e^2 \sin^2(\theta - \phi)]^{1/2} \\ &\simeq e \cos(\theta - \phi) + R \end{aligned} \tag{2}$$

or

$$\rho \simeq a \cos \theta + b \sin \theta + R \tag{3}$$

This is a linearization of the parameters about the origin. Linearization about any other point involves considerable extra complexity of the coefficients. In practice, because roundness measuring instruments usually introduce a slight geometrical distortion into the data, fitting a limacon, equation (3), nearly always gives a more accurate measurement than would fitting a true circle (**6**). Whenever the limacon approximation is valid, the calculation of limiting reference circles becomes a linear program. For example, the minimum circumscribing figure to a set of data points (r_i, θ_i) is expressible as:

$$\begin{aligned} &\text{minimise} \quad Z = R \\ &\text{subject to} \quad a \cos \theta_i + b \sin \theta_i + R \geqslant r_i \end{aligned} \tag{4}$$

for all i. Others may be expressed in the form either of equation (1) or equation (4).

Before proceeding to develop algorithms from these formulations, it is useful to establish a practical context and a mathematical notation by first illustrating earlier work on reference circles and then, in Section 3, reviewing, extremely briefly, the main points of linear programming theory.

Essentially the same geometrical procedure for finding, for example, the minimum circumscribing circle has been proposed independently on at least two occasions (**7**, **8**) (Fig. 2):

1. Find the profile point furthest from the origin: a central circle passing through this is circumscribing but not minimum.
2. Move the circle centre towards the point, reducing its radius in order to maintain contact with the point, until a second contact is found.
3. Move the centre along the bisector of the angle between these contacts, holding onto both by

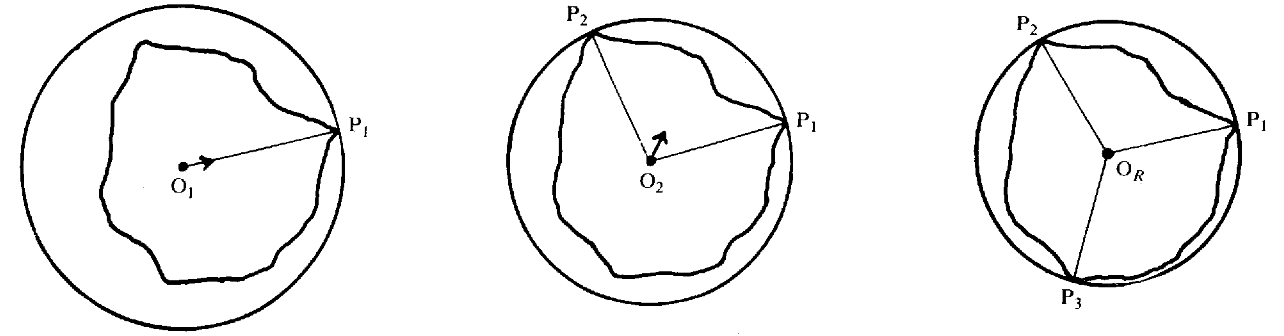

Fig. 2 A search procedure for the minimum circumscribing circle (7, 8)

reducing the radius, until a third point of contact is found.

As a circle has three independent parameters, three contacts are sufficient to define it and a solution is found. A computer program would use limacons rather than circles but follow the logic. Since a search along the whole locus of the centre is needed, there are many iterations, slow execution and, in practice, some difficulties over the determination of end-conditions. Nevertheless, the method has been used commercially with some success.

3 BASIC CONCEPTS IN LINEAR PROGRAMMING

A linear program is an optimization in which the objective function and all the constraints are linear in the parameters. Using vector notation, it can be expressed as:

$$\text{maximize} \quad Z = \mathbf{c}^T\mathbf{x}$$

$$\text{subject to} \quad \mathbf{A} \cdot \mathbf{x} \leqslant \mathbf{b}$$

where, for m positive parameters, $\mathbf{x}$, and n constraints, $\mathbf{c}$ is an m-vector, $\mathbf{b}$ an n-vector and $\mathbf{A}$ an $m \times n$ matrix.

It is known (there is extensive literature on this subject) that the optimum solution occurs when each of the constraints which is actively limiting that optimum is satisfied to its limit by one of the parameters. Hence only certain combinations of parameter values need be examined. An orderly search through these is obtained by using the simplex method in which iterations involve only elementary row operations on the matrix-vector representation. Simplex organizes these vectors as a partitioned matrix (a tableau):

$$\left[\begin{array}{c|c} \mathbf{K} & \mathbf{b} \\ \hline \mathbf{c}^T & Z \end{array}\right]$$

where $\mathbf{K}$ is $\mathbf{A}$ augmented by an $n \times n$ identity matrix and $\mathbf{c}$ is correspondingly extended by n zero-elements. This appends n 'slack variables' to the original parameters. If the ith parameter is limiting a particular constraint, the column $\mathbf{K}_i$, in $\mathbf{K}$, will have value $+1$ in the row corresponding to that constraint and zero in all other elements. The set of defining parameters so identified form the 'basis'. Initially the basis is the n slack variables. Iterations attempt to match parameters to constraints in such a way that Z is rapidly maximized. It is usual to maintain always the feasibility of the current iteration by ensuring that no constraint is ever violated, that is, that no element of $\mathbf{b}'$ becomes negative. The prime indicates the vector which currently occupies the position originally occupied by $\mathbf{b}$. At each iteration the largest positive element of $\mathbf{c}'^T$ is chosen and its column brought actively into the solution (this is the stategy of 'steepest descent'). When no positive elements remain in $\mathbf{c}'^T$, optimality has been achieved and the solution values are readily interpreted from the tableau. Equality constraints, which must, of course, be always exactly satisfied, do not take slack variables but equivalent 'artificial variables' are used as a device for starting the iterative procedure in an orderly manner. By definition, artificial variables cannot remain present in any feasible solution of the tableau.

At any iteration, the columns which originally consisted of the identity matrix carry a complete and interpretable record of the row transformations carried out on the tableau. Likewise, the columns of the current basis carry the same information in the inverse of their original form. The computationally efficient method of revised simplex, does not update the full tableau but merely notes what would have been done at each iteration. Elements are only updated when specifically required for calculations, as, for example $\mathbf{b}' = \boldsymbol{\beta}\mathbf{b}$ where $\boldsymbol{\beta}$ is the inverse of the current basis.

While the total computation required rises with both m and n, it is particularly sensitive to n, the number of constraints, as the work required relates to that of inverting $n \times n$ matrices. It may, therefore, be advantageous to use a dual program. For any $m \times n$ linear program (termed the primal), an $n \times m$ dual may be defined as:

$$\left[\begin{array}{c|c} \mathbf{K} & \mathbf{c} \\ \hline \mathbf{b}^T & Z \end{array}\right]$$

where $\mathbf{K}$ is now the augmented form of $\mathbf{A}^T$ and the optimization has changed from minimization to maximization or vice versa. It contains exactly the same information as the primal, subject to the correct relative interpretation of specific elements.

4 DUAL LINEAR PROGRAMS IN METROLOGY

Straightness, flatness and all routine roundness measurements involve reference fitting which appears naturally as linear programs. For more complex geometries, the errors inherent in parameter linearization may be judged acceptable when weighed against the computational efficiency of simplex. All the resulting formulations, essentially as equations (1) or (4), have in common features indicative that the dual program will offer the most efficient solutions.

The sign-definiteness of parameters required for simplex cannot be guaranteed with metrological data and so each parameter is replaced by an ordered pair having equal magnitude but opposite sign. Even so, the number of constraints usually dominates the number of parameters. Thus a circumscribing limacon fit involves six parameters and the minimum zone seven, but typical measurements involve several hundred profile points each generating a constraint, two in the case of minimum zone. The sources of the difficulties encountered with early attempts at circle fittings are now apparent. They did not exploit the simplex method of searching only certain basic solutions and, further, they worked with a primal formulation involving, say, six parameters and 500 constraints, rather than a dual which, while having 500 parameters, has only six constraints.

In moving from the primal to the dual, the roles of vectors **b** and **c** are interchanged. If at any iteration the dual is maintained in a feasible condition (all elements of **c** positive), the corresponding primal would be interpreted as being in an optimal, but generally infeasible, condition. The implications of dual feasibility are critical to what is to follow. Consider a physical interpretation for the case of a circumscribing limacon (or circle). The primal feasibility condition amounts to starting with a figure which is too large but which certainly encloses the profile and then shrinking it to the smallest radius which still encloses the profile. Dual feasibility would entail choosing initially a figure which is the smallest to enclose some, but not necessarily all, of the data points (in the primal, optimal but infeasible) and then expanding it as little as possible so as to include all the data.

The sign-indeterminancy of the primal parameters implies that the dual program will consist of equality constraints. A general solution may therefore require the use of artificial variables in order to obtain an initial basic feasible solution. More significantly, for specific problems the optimal solution cannot contain any slack variables since each constraint must be exactly satisfied. If a primal has three parameters, the dual has three constraints. The corresponding geometric observation is that a circle is defined by exactly three contacts with the data.

Another common feature, of particular significance for the development of exchange algorithms, is that the primal objective function involves only one parameter, minimizing, for example, the radius, R, or zone, h. So, by correctly choosing the order of parameters, **c** may be written with $+1$ in the final element with zero elsewhere. Then at any iteration of the dual, **c**′ is identical to the final column of the inverse of the basis, $\boldsymbol{\beta}$, and so maintaining dual feasibility requires only that the final column of $\boldsymbol{\beta}$ has non-negative elements. Given this, any iteration will be 'legal', whether or not it be the most efficient option. From this point, the strategy is best demonstrated by example.

5 MINIMUM RADIUS CIRCUMSCRIBING LIMACON

Transferring the primal, geometrical statement of the minimum radius circumscribing limacon, equation (4), to the dual, and omitting the artificial variables which cannot contribute to the final solution, the initial tableau can be written as a minimization:

$$\begin{array}{ccccc|c} \cos\theta_1 & \cdots & \cos\theta_i & \cdots & \cos\theta_n & 0 \\ \sin\theta_1 & \cdots & \sin\theta_i & \cdots & \sin\theta_n & 0 \\ 1 & \cdots & 1 & \cdots & 1 & 1 \\ \hline -r_1 & \cdots & -r_i & \cdots & -r_n & \end{array}$$

At any iteration giving a feasible solution, the basis will be formed from three of these columns. So, taking three general contact points at θ_i, θ_j and θ_k,

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

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. The determinant of $\boldsymbol{\beta}^{-1}$ 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 $\boldsymbol{\beta}$. The non-negativity of the elements of the third column of $\boldsymbol{\beta}$ thus requires that these co-factors, Δ_{ij}, Δ_{jk}, Δ_{ki}, must have the same sign where:

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

and similarly for the others. Using Cartesian coordinates, 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

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

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) and dividing the xy plane into the two areas where $\Delta_{ir} > 0$ and where $\Delta_{ir} < 0$. The line is also the locus of all points having θ_i as their argument. Noting the order of indices, dual feasibility 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 $\Delta_{jr} = 0$ or $\Delta_{kr} = 0$. If point k is to lie on the opposite side of $\Delta_{ir} = 0$ from point j and on the opposite side of $\Delta_{jr} = 0$ from point i, it can only occupy the sector shown in Fig. 3. As it is only in this geometry that Δ_{ik} and Δ_{jk} will have opposite signs, as required for dual feasibility, the following theorem, termed here 'the 180° rule', is proved.

A circumscribing limacon on 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.

A complete simplex iteration for the minimum radius circumscribing limacon in the dual consists of 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

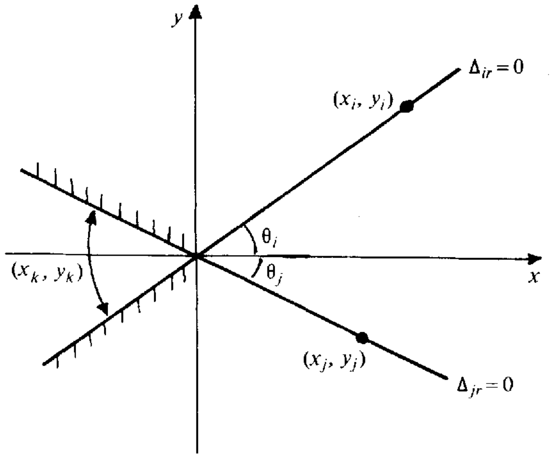

Fig. 3 Geometry of the dual-feasibility condition for the circumscribing limacon

such a way that dual feasibility is maintained. The 180° rule allows the general iteration to be simplified to the following exchange algorithm:

1. Choose any three data points such that no two adjacent ones subtend an angle at the origin of more than 180°.
2. Construct a reference limacon through these three points.
3. If no data points lie outside this limacon the solution is found. Otherwise choose the point which violates the reference by the largest amount.
4. Replace one of the reference points by this new point such that the 180° rule is still obeyed and go back to 2).

The exchange between any new point and the contacts is always unique, as illustrated in Fig. 4.

An exchange algorithm depends upon the iterations moving monotonically towards an optimum solution in order to guarantee that cyclical exchanges do not occur.

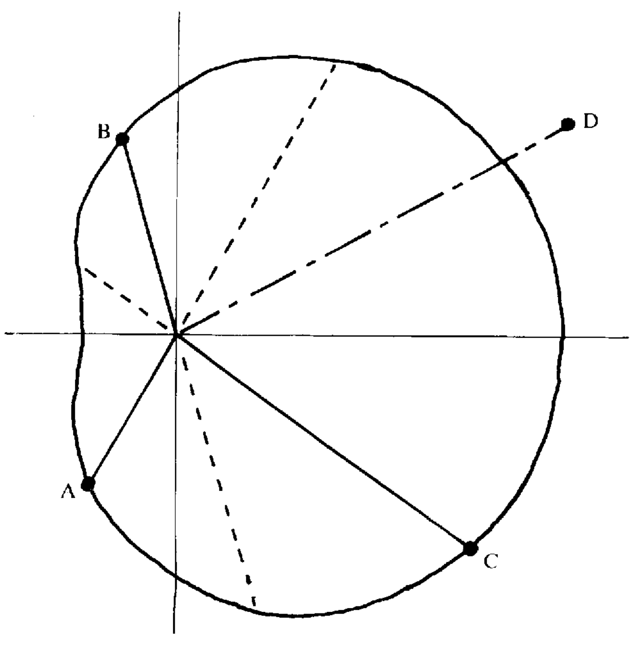

Fig. 4 Feasible point exchange for the circumscribing limacon: only D replacing C continues to satisfy the 180° rule. (Note, diagram is exaggerated)

Here this is the case for, as the exchange is unique at each iteration, it must be identical to the variable change at the simplex iteration of a linear program, and that is known to converge monotonically.

6 MINIMUM ZONE LIMACONS

The primal expression of the minimum zone limacon fit can be written with all constraints in the same sense as:

minimize $Z = h$

subject to

$$\begin{bmatrix} \cos\theta_i & \sin\theta_i & 1 & 1 \\ -\cos\theta_i & -\sin\theta_i & -1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ R \\ h \end{bmatrix} \geqslant \begin{bmatrix} r_i \\ -r_i \end{bmatrix}$$

for all i simultaneously.

The constraints consist of two equal subsets comprising those relating to 'inner' (on $-r_i$) or 'outer' (on $+r_i$) contacts of the zone. The zone width must be a positive quantity and so the primal has a sign restricted variable, h. The dual thus has three equality constraints and one inequality and its feasible basis must consist of either four columns from the original constraints or three such columns and the single slack variable. This latter 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 basis can be any four columns chosen freely from inner or outer contact sets provided only that the same point cannot be used simultaneously from both sets (this is physically impossible for a non-zero zone width). So:

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

where the variables S_i to S_l take only values $+1$ or -1 and indicate whether the contact is with the outer or inner line respectively.

The co-factors of the final row of $\boldsymbol{\beta}^{-1}$, which must have the same sign for dual feasibility, will be:

$$-S_j S_k S_l\,\Delta_{jkl};\; S_i S_k S_l\,\Delta_{ikl};\; -S_i S_j S_l\,\Delta_{ijl};\; S_i S_j S_k\,\Delta_{ijk}$$

where the notation is defined in the function:

$$\Delta_{ijr} = \begin{bmatrix} \cos\theta_i & \cos\theta_j & \cos\theta \\ \sin\theta_i & \sin\theta_j & \sin\theta \\ 1 & 1 & 1 \end{bmatrix} = 0$$

which is used to test the sign of Δ_{ijk} and Δ_{ijl}. It is readily shown (Fig. 5), that this function divides the measurement plane into two regions within which lie all positive or negative values of Δ_{ijr}. (Note that it is not necessarily the reflex sector which holds positive values.)

Now, if the kth and lth points are contacts of the same type, $S_k = S_l$ and Δ_{ijk} and Δ_{ijl} must differ in sign if all co-factors are to have the same sign. Conversely if contacts of different type, $S_k \neq S_l$ and Δ_{ijk} and Δ_{ijl} must have the same sign. Figure 5 shows that to maintain the co-factors with the same sign, points k and l must lie alternately with points i and j as angle increases if they are contacts of the same type. If k and l are contacts of

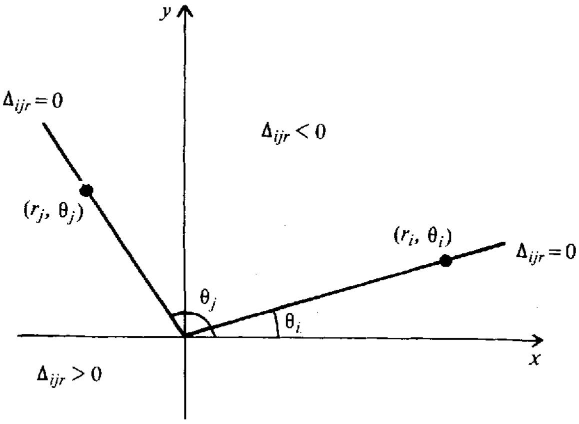

Fig. 5 Geometry of the dual-feasibility condition for the minimum zone limacons

different types they must lie adjacent to each other. This situation exists simultaneously for all point pairs. The geometrical interpretation is that dual feasibility is maintained only if the four points contact alternately the inner and outer limit of the zone as angle, from the measurement origin, is swept.

As before, any single point exchange is unique and its relationship to simplex guarantees convergence of the following exchange algorithm:

1. Choose arbitrarily four data points.
2. Fit to these a reference limacon such that they are radially equidistant from it and lie alternately to either side of it with increasing angle.
3. If no other points are further from the reference the solution is found.
4. 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 to either side of the reference and return to 2.

7 MINIMUM ZONE STRAIGHT LINES AND PLANES

The minimum separation parallel, straight lines 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 to occur is that relative to an nth order polynomial, the data must have $(n + 2)$ maxima and minima all of equal magnitude. The solution can be found by the Stiefel exchange algorithm which proceeds by fitting the polynomial according to this condition to $(n + 2)$ points and then exchanging points further away from it than those points into the defining set while maintaining the 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 (Fig. 6).

The minimum zone planes can be expressed, in instrument coordinates:

$$\text{minimize} \quad Z = h$$

$$\text{subject to} \quad ax_i + by_i + c + h \geqslant z_i$$

$$ax_i + by_i + c - h \leqslant z_i$$

for all data points (x_i, y_i, z_i)

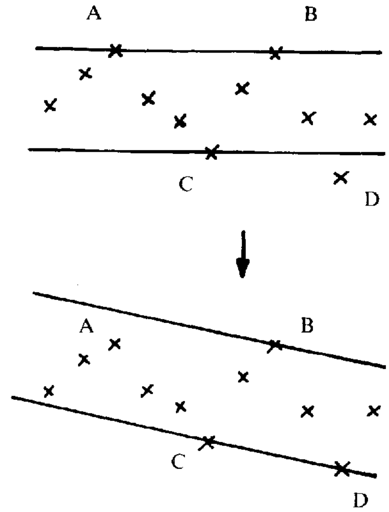

Fig. 6 Minimum zone straight lines (the Stiefel exchange): only D replacing A preserves alternating property.

a, b and c are sign unrestricted and $h \geqslant 0$. 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 of the dual which can be expressed (see Section 6 for the reasoning):

$$\boldsymbol{\beta}^{-1} = \begin{bmatrix} S_i x_i & S_j x_j & S_k x_k & S_l x_l \\ S_i y_i & S_j y_j & S_k y_k & S_l y_l \\ S_i & S_j & S_k & S_l \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

where S_i, etc. take values $+1$ or -1 according to whether (x_i, y_i, z_i) contacts the upper or lower of the minimum zone planes. As before, dual feasibility is guaranteed if all terms in the final column of $\boldsymbol{\beta}$ are positive, which will be true providing that:

$$-S_j S_k S_l \Delta_{jkl}; \quad -S_i S_j S_l \Delta_{ijl}$$

$$S_i S_k S_l \Delta_{ikl}; \quad S_i S_j S_k \Delta_{ijk}$$

all have the same sign. Consider the determinant equation representing the boundary between positive and negative regions of Δ_{jkl}:

$$\Delta_{jkr} = \begin{vmatrix} x_j & x_k & x \\ y_j & y_k & y \\ 1 & 1 & 1 \end{vmatrix} = 0$$

It is a plane parallel to the z-axis (since it is independent of z), passing through points (x_j, y_j) and (x_k, y_k). Dual feasibility requires that if $S_i = S_l$ (contacts with the same plane) Δ_{jki} and Δ_{jkl} must have different signs and vice versa. So if the ith and lth contacts are with the same plane they lie on opposite sides of $\Delta_{jkr} = 0$ but if they contact different planes they lie both to the same side of $\Delta_{jkr} = 0$. A parallel argument shows that the same is true for all pairs of points.

These relationships show the relative positions of contacts which give dual feasibility. The two ways of satisfying them are shown in the plan views of Fig. 7. There can 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

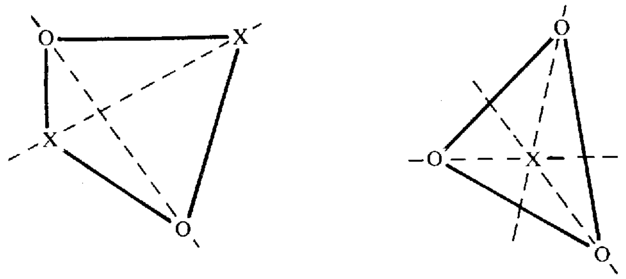

Fig. 7 Plan view of contact geometries for minimum zone planes. O and X represent contacts with different planes.

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 pattersn. 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 in an automatic system is becoming unclear.

8 ALGORITHM PRACTICALITY

It may, at first glance, seem surprising than limacon fitting rather than the apparently simpler case of flat surfaces has been used as the primary example. The final observation of Section 7 gives some explanation and leads us to question the relative efficiencies of different algorithmic approaches. This paper, being primarily concerned with mathematical foundations, will not report detailed performance tests but it would be inappropriate to ignore such practicalities. Truly comparative cross-system benchmarks are notoriously difficult to obtain so discussion will be limited to some general comments about experience obtained on sixteen-bit mini- and microcomputers.

A typical roundness 'profile' would have 512 equally spaced radial ordinates each resolved over a ten- or twelve-bit working range. Exchange algorithm systems have now been working with data of this type in both industrial and research environments for several years and their robustness has been extablished. Even with an arbitrary choice of points for the initial basis, the exchange algorithm virtually always solves for the minimum circumscribing limcacon in five or less iterations, while the minimum zone only occasionally needs more than five, on real engineering profiles. The earlier (primal based) algorithms were run with poorly defined end-conditions, typically making thirty-two relatively coarse-stepped iterations and then thirty-two finer steps after which the process was terminated with a result assumed close to the desired optimum. The new techniques yield at least a tenfold saving in the number of iterations as well as giving a fully determined convergence and, so, better accuracy. With both algorithms the iteration is dominated by the almost identical computation and checking of the updated figure, so the program size and the cycle times are closely similar on similar machines programmed in the same language. A tenfold speed increase is also obtained.

The direct use of revised simplex on dual programs representing limacon fitting has been studied using a specially developed package containing only the subroutines essential for solving this class of problem. Memory requirements are only slightly larger than those of exchange algorithms and execution is typically about 20 per cent slower. This is due to the simple way artificial variables are treated. This difference can be removed at the cost of extra program length.

These comparisons were made between programs written in FORTRAN IV using software floating point. Iteration cycle times were typically between one and two seconds so the speed increase of the dual-based methods is over a range of some importance in a production environment, say a reduction from over a minute to a few seconds for the total calculation.

The limacon fits have simple exchange rules which can be expressed in a few numerical comparisons and logic operations. Thus in a specialized system both a size reduction and a speed increase would be obtained by replacing the direct use of revised simplex on the dual by an exchange algorithm. However, the exchange logic is specific, so if several different references are to be implemented there will be less shared code. With more complex geometries it is of even greater importance that the efficiency of dual-based methods is obtained. Yet, with even the simplest three-dimensional case the exchange rules are becoming quite complicated. The indications to date suggest that computer-based instrumentation may standardize towards a modified version of revised simplex rather than the pure exchange algorithms.

9 CONCLUSIONS

This paper has shown how the formal application of mathematical theory, in this case mathematical programming, can have a dramatic effect on a field such as surface metrology which has historically used an intuitive approach. A new generation of algorithms for limiting value roundness references has been produced which are more precise and run at least ten times faster than their predecessors with similar memory requirements. Duality theory has shown that the 'obvious' geometrical method is not the best approach: the new method seems obvious once it has been expressed!

The study of exchange algorithms gives a very clear insight into the geometrical implications of reference fitting. This is of great metrological significance, for measurement controls should always be based on engineering relevance rather than a mathematically convenient abstraction. The exchange algorithm also provides a good method for hand solution should that prove necessary, as it may in the context of standardization. Relatively flexible measurement systems are likely to use a more general implementation of a revised simplex algorithm. This is no cause for concern: both are firmly based on the same theoretical foundation.

ACKNOWLEDGEMENT

The ideas leading to this work were conceived during the author's employment with Rank Taylor Hobson Limited. He gratefully acknowledges the helpfulness of discussions with colleagues there and, particularly with Dr P. H. Phillipson at the University of Leicester who was much involved in the formulation of the approach used here.

REFERENCES

1 **Chetwynd, D. G.** and **Phillipson, P. H.** An investigation of reference criteria used in roundness measurement. *J. Phys. E. Sci. Instrum.*, 1980, **13**, 530.

2 **Reason, R. E.** *Report on the Measurement of Roundness*, 1966 (Rank Taylor Hobson).

3 **Osborne, M. R. and Watson, G. A.** On the best linear Chebyshev approximation. *Comput. J.*, 1968, **10**, 172–177.

4 **BS 3730: 1964** *Methods for the assessment of departures from roundness*, 1964 (British Standards Institution).

5 **Whitehouse, D. J.** A best fit reference line for use in partial arcs. *J. Phys. E. Sci. Instrum.*, 1973, **6**, 921–4.

6 **Chetwynd, D. G.** Roundness measurement using limacons. *Precision Engineering*, 1979, **1**, 137–141.

7 **Avdulov, A. N.**, *et al.* Some aspects of measurement of roundness. *Proc. Inst Mech. Engrs*, 1968, **182**, Pt. 3K, 425.

8 **Whitehouse, D. J.** Modern trends in the measurement of surfaces. *Proc. Inst. Belgium Soc. Mech. Engrs*, 1975, **21**.