Geometric Tolerance Assessment

A B Forbes

September 1999
Geometric Tolerance Assessment

A B Forbes
Division of Information Technology and Computing
National Physical Laboratory
Teddington
Middlesex
United Kingdom
TW11 0LW

ABSTRACT

Tolerance assessment is a key step in maintaining and improving quality in manufactured products. This report describes how a class of design and tolerance specifications can be formulated in terms of parametrized curves and surfaces and parameter and form constraints. The tolerance assessment problem can then be addressed by optimization techniques. It is shown how many important tolerance assessment problems can be formulated as simple matching problems for which efficient solution algorithms can be developed.
## CONTENTS

1. **INTRODUCTION** .......................................................... 1
   1.1 *NOTES AND REFERENCES* ........................................... 2

2. **GENERAL FORMULATION OF TOLERANCE ASSESSMENT** .......... 2
   2.1 *PARAMETRIZED CURVES AND SURFACES* ................. 2
   2.2 *TOblahrances* .................................................. 2
   2.3 *MEASURED DATA POINTS* ...................................... 3
   2.4 *TOLERANCE ASSESSMENT* ..................................... 4
   2.5 *PROCESS CONTROL* ............................................ 4

3. **FEATURES OF TOLERANCE ASSESSMENT** ......................... 5
   3.1 *FEASIBILITY OF THE TOLERANCE CONSTRAINTS: OVER AND UNDER TOLERANCING* ............................................. 5
   3.2 *AMBIGUITY OF THE TOLERANCE SPECIFICATION* ........... 6
   3.3 *MEASUREMENT STRATEGY* ...................................... 7
   3.4 *MEASUREMENT ACCURACY* ..................................... 9
   3.5 *MATHEMATICAL PROGRAMMING AND TOLERANCE ASSESSMENT* . 10

4. **DECOMPOSITION OF MULTI-COMPONENT PROBLEMS** ............. 10

5. **GEOMETRIC ELEMENTS** ............................................. 13
   5.1 *NOTES AND REFERENCES* ...................................... 14

6. **TEMPLATE MATCHING** ............................................... 15

7. **PART MATING** ..................................................... 16
   7.1 *MATING BY MATCHING* ........................................ 17
   7.2 *NOTES AND REFERENCES* ...................................... 17

8. **APPROXIMATE METHODS** ........................................... 17
   8.1 *NOTES AND REFERENCES* ...................................... 19

9. **SUMMARY AND CONCLUSIONS** .................................... 19

10. **ACKNOWLEDGEMENTS** ............................................. 20

11. **REFERENCES** ..................................................... 20
1 INTRODUCTION

This report is concerned with mathematical techniques underlying computer-aided inspection methods used for checking whether a manufactured part is within tolerance and is aimed at those who want to develop or use tolerance assessment software.

Tolerancing questions feature strongly in the drive to improve quality in manufactured products at all stages. At the design stage, once the functional requirement of a part has been established, the designer will arrive at a part specification which is in some sense optimal. Due to variability inherent in the manufacturing process the designer will also have to specify how far the manufactured product can depart from its optimal specification and still meet its functional requirement. This is done by specifying various tolerances associated with the design.

At the manufacturing stage, parts will need to be inspected and compared with their design and tolerancing specification. The information gathered at this stage can be used in two ways. Firstly, parts which do not meet the specification can be rejected (possibly for reworking). Secondly, the behaviour of the manufacturing process can be monitored, providing a feedback control mechanism which can tune the manufacturing system so as to minimise the number of parts produced out of tolerance.

A major part of the inspection process is the comparison of the manufactured part with its specification. One approach is to manufacture hard gauges and physically compare the part with the gauge. For example, a part's maximum diameter can be checked by fitting it into a ring gauge. The advantage of this method is that it gives a fairly unambiguous answer. Among the disadvantages of the method is that it is inflexible. In particular, the tolerance specification is effectively defined by the number and use of the hard gauges. Also, the method gives no direct information about drifts in the behaviour of the manufacturing system.

In a computer-aided inspection setting, a part is measured by a coordinate measuring system (CMS) yielding a data set $X$ representing data points $x_i$, $i \in I$, related to the part's surface. The data set is then compared with the tolerancing information using tolerance assessment software. The advantages of this method are i) it can be flexible since the tolerancing information is represented in software, ii) it is amenable to automation and iii) the assessment software can also return information about the manufacturing process. A disadvantage of the method is that without additional information it can only make statements about the data set $X$ rather than the actual manufactured part.

The process outlined above gives rise to a number of questions:

1. Given a functional requirement and a design, what tolerancing information should be supplied in order that the manufactured part will achieve its functional requirement at minimal total cost (i.e. cost to the manufacturer and cost to the user over the part's lifetime)?

2. How can tolerancing information be represented in a form suitable for computer-aided assessment?

3. How can data points representing a part be compared with its tolerance specification and what information can be derived about the state of the manufacturing system?
4. What is the best data distribution to ensure maximum representativeness with respect to the tolerances?

This report concentrates on questions 2–4 with the main focus of attention on No. 3. We aim to formulate tolerance assessment problems in a way that allows for their solution through the application of optimisation techniques. Algorithms for solving tolerance assessment problems will be described in a companion report [20]. In section 2 we give a general formulation of tolerance assessment in terms of optimisation and mathematical programming. In section 3, we discuss aspects of tolerance assessment problems which are of practical importance. In sections 4 to 7 we describe various tolerance assessment problems in terms of the general framework. In section 8 we briefly consider how some assessment problems can be simplified. Section 9 contains our concluding remarks.

1.1 NOTES AND REFERENCES

An excellent survey of the types of problems associated with tolerancing is given in [32]. See also [1], and [29, 30, 31, 23] and the references therein. Tolerance assessment has become an important factor in a ‘total quality’ approach to design and engineering pioneered by Deming and Taguchi; see e.g. [24, 5] for discussions on this ‘quality philosophy’. The difficult problem of how to set tolerances to minimise total costs is considered in [27, 28]

2 GENERAL FORMULATION OF TOLERANCE ASSESSMENT

In this section we outline the main components of tolerance assessment: the representation of the ideal geometry and tolerance information, measured data points and the comparison of the measured data with the tolerance specification.

2.1 PARAMETRIZED CURVES AND SURFACES

We assume that the ideal design (on an engineering drawing or computer-aided design (CAD) system, for instance) can be represented as a union of components each of which is a parametrized curve or surface defined in a common frame of reference. For example, a circle in two dimensions is parametrized by its centre coordinates \((a, b)\) and radius \(r_0\) and a particular circle is specified by giving the values of these three parameters. Thus, the ideal design can be specified by a finite number \(n_K\) of ‘elements’ \(a_k \mapsto S_k(a_k), k = 1, \ldots, n_K\), denoted collectively by \(a \mapsto S(a)\) with \(a = (a_1, \ldots, a_{n_K})\), along with actual values for the parameters \(a_k\). For simple designs, the total number of parameters may be of the order of 10; for more complicated shapes including those designed using CAD tools, hundreds of parameters may be required. Note that the parameter values will refer to the fixed frame of reference of the design specification.

We sometimes refer to \(a \mapsto S(a)\) as the nominal form.

2.2 TOLERANCES

Tolerances set limits on how far the manufactured workpiece can depart from the ideal and still be regarded as acceptable.
Parameter tolerances. The simplest type of tolerances are those which relate only to the values of the parameters. For example, the ideal value for the radius of a circular arc may be 100mm but circles with radii between 98 and 102mm could be acceptable. Similarly the distance between two circle centres may ideally be 300mm but a separation between 295 and 305 could be acceptable. These two tolerances can be expressed as

\[ |r_0 - 100| \leq 2, \]

and

\[ |[(a_1 - a_2)^2 + (b_1 - b_2)^2]^{1/2} - 300 | \leq 5. \]

In general, each parameter tolerance can be expressed in terms of a function \( c \), say, of the parameters \( a \) as an inequality of the form

\[ c(a) \geq 0. \]

Each inequality usually involves only a subset of the parameters \( a \). We will denote the set of parameter tolerances by

\[ C(a) \geq 0. \]

Form tolerances. Parameter tolerances relate to ideal geometries such as a circle specified in the design. The manufactured artefact will not be an ideal curve or surface but will have form error, and tolerances will be required to define the limit of acceptability of this error. For example, a tolerance on circularity of a curve \( C \) might require that there exist concentric circles \( C_1 \) and \( C_2 \) of separation at most \( s \) such that \( C \) lies between \( C_1 \) and \( C_2 \). This can be expressed mathematically as: there exist \( a, b, r_1 \) and \( r_2 \) with

\[ |r_1 - r_2| \leq s \]

and such that for all \( \mathbf{x} = (x, y) \in C \),

\[ r_1 \leq [(x - a)^2 + (y - b)^2]^{1/2} \leq r_2. \]

In general, form tolerances relate points \( \mathbf{x} \) on the surface of the workpiece with the parameters \( a \) of the nominal shape through inequalities of the form

\[ f(x; a) \geq 0. \]

These constraints can be quite varied in nature involving not only departure from nominal form but also curvature and roughness, for example. We will denote the set of form tolerances defined on data points \( X \) by

\[ F(X; a) \geq 0. \]

The main difference between the parameter constraints and the form constraints is that the form constraints involve measured data while the parameter constraints do not.

2.3 MEASURED DATA POINTS

In computer aided inspection, the comparison of the manufactured artefact with its specification is made through a data set, gathered by a coordinate measuring system (CMS), which represents points lying on the surface\(^{(1)} \) of the artefact. If the design is given in terms of (ideal)

\(^{(1)} \)The problem of probe radius deconvolution is not considered here; see Cox and Forbes [12], for example.
parametrized surfaces $a_k \mapsto S(a_k)$, $k = 1, \ldots, n_K$, the artefact will be made up of corresponding surfaces $S_k$ and for each $S_k$, there will be a set of data points $X_k = \{x_i : i \in I_k\}$ lying in $S_k$, modulo measurement error. We will denote the combined set of measurements by $X$. Note that these data points will be given in the frame of reference internal to the CMS and in general this frame will be different from that used in the design specification.

2.4 TOLERANCE ASSESSMENT

The general tolerance assessment problem can now be stated as:

Given a design specification $$a \mapsto S(a),$$ and a workpiece $S$ with data points $X$ lying on $S$, find parameters $a$ (if they exist) which satisfy the parameter constraints $$C(a) \geq 0,$$ and form constraints $$F(X; a) \geq 0.$$

**Example:** Circle with tolerances on radius and circularity. Suppose a design specifies a circle of radius $50.000 \pm 0.050$mm with a circularity (maximum departure from a circle) of $0.020$mm and we collect data points $X = \{x_i = (x_i, y_i) : i = 1, \ldots, 50\}$ on the workpiece. The corresponding tolerance assessment problem\(^{(2)}\) is:

Find (if possible) $a$, $b$, and $r_0$ which satisfy the parameter constraints

$$0 \leq r_0 - 49.050,$$
$$0 \leq 50.050 - r_0,$$

and form constraints

$$0 \leq r_i - r_0 - 0.010,$$
$$0 \leq r_0 - r_i + 0.010,$$

where $r_i = [(x_i - a)^2 + (y_i - b)^2]^{1/2}$.

2.5 PROCESS CONTROL

A supplementary problem, important for the control of the manufacturing process, is to give a measure of the extent to which the tolerance constraints have been met. One measure can be given by solving an optimisation problem derived from the assessment problem such as:

Given a design specification $$a \mapsto S(a),$$

and a workpiece $S$ with data points $X$ lying on $S$, find parameters $a^*$ and $s^*$ to solve:

$$\max_{a,s} s,$$

\(^{(2)}\)See also section 3.2.
subject to the parameter constraints

\[ C(\mathbf{a}) \geq s, \]

and form constraints

\[ F(X; \mathbf{a}) \geq s. \]

If the maximum value is non-negative, then the solution \( \mathbf{a}^* \) will also solve the tolerance assessment problem and, the greater the maximum, the more within tolerance the part (as represented by the data \( X \)).

**Example: Circle with tolerances on radius and circularity II.** In the example of the circle in section 2.4, the related optimisation problem is

\[
\max \quad s \\
\text{s.t.} \quad r_0 - 49.050 \geq s, \\
50.050 - r_0 \geq s.
\]

subject to parameter constraints

\[
\begin{align*}
  r_0 - 49.050 & \geq s, \\
  50.050 - r_0 & \geq s,
\end{align*}
\]

and form constraints

\[
\begin{align*}
  r_i - r_0 + 0.010 & \geq s, \\
  r_0 - r_i + 0.010 & \geq s,
\end{align*}
\]

where \( r_i = [(x_i - a)^2 + (y_i - b)^2]^{1/2} \), as before. Note that if \( s = 0.010 \) at the solution then \( r_i = r_0 \) indicating that the points lie exactly on a circle. However, the radius \( r_0 \) need not be at its optimal value of 50.000. We can, though, formulate a *normalised* optimisation problem

\[
\max \quad s \\
\text{s.t.} \quad r_0 - 49.050 \geq 0.050s, \\
50.050 - r_0 \geq 0.050s.
\]

subject to parameter constraints

\[
\begin{align*}
  r_0 - 49.050 & \geq 0.050s, \\
  50.050 - r_0 & \geq 0.050s,
\end{align*}
\]

and form constraints

\[
\begin{align*}
  r_i - r_0 + 0.010 & \geq 0.010s, \\
  r_0 - r_i + 0.010 & \geq 0.010s.
\end{align*}
\]

If at the solution \( s \) is non-negative then the part will be within tolerance while if \( s = 1 \) the part will be perfect (with respect to the tolerances imposed).

### 3 FEATURES OF TOLERANCE ASSESSMENT

#### 3.1 FEASIBILITY OF THE TOLERANCE CONSTRAINTS: OVER AND UNDER TOLERANCING

Suppose we are given a design specification \( \mathbf{a} \rightarrow \mathcal{S}(\mathbf{a}) \) with ideal parameters \( \mathbf{a}_0 \), along with tolerance constraints \( C(\mathbf{a}) \geq 0 \) and \( F(X; \mathbf{a}) \geq 0 \). An immediate requirement is to confirm that the constraints are consistent. The simplest approach is to check that the ideal parameters \( \mathbf{a}_0 \)
satisfy the constraints by evaluating $C(a_0)$ and $F(x; a_0)$ for a sufficiently representative set of points $X^* = \{x_i^*: i \in I\}$ on $S(a_0)$.

Once it has been established that the constraints are consistent, it may be interesting to see how tight are the tolerances. For example, we may consider optimisation problems of the form

$$\max_p p^T G p,$$

where $G$ is diagonal scaling matrix, subject to the constraints

$$C(a_0 + p) \geq 0,$$

to examine how far the parameters can stray from their ideal value of $a_0$ and still satisfy the parameter tolerances.

We may also examine the implications on form by choosing a suitable distribution of points $X^* = \{x_i^*: i \in I\}$ lying on $S(a_0)$ and solving

$$\max_{i \in I} |e_i|$$

subject to

$$f(x_i + e_i; a_0) \geq 0.$$ 

By looking at optimisation problems such as these, it may be possible to show that combinations of tolerances are either too lax or too tight when compared with the designer's intention or the functional requirement. Importantly, these calculations can be performed at the design stage as part of the design process.

**Example: Tolerances on an arc of circle.** Consider the following specification of a circle: the circle is to pass through $(0, 5.0)$ and $(0, -5.0)$ and have a radius of $100.0 \pm 0.1$. We can show that if the radius is $99.9$ then the circle passes though $(0.1252, 0)$ while if the radius is $100.1$ the circle passes through $(0.1250, 0)$. This shows that a tolerance of $\pm 1\text{mm}$ on the radius implies a much tighter tolerance of $\pm 0.1\text{micrometres}$ on where the circle crosses the $x$-axis; see [13] for a discussion on tolerancing problems associated with partial arcs.

3.2 AMBIGUITY OF THE TOLERANCE SPECIFICATION

Tolerances defined using natural language terms are sometimes ambiguous.

**Example: Circle with tolerances on radius on circularity III.** In the example of section 2.4, the tolerance on the radius of the circle was interpreted as applying to the radius given by the average of the maximum and minimum distance from the points to the circle centre and this gave rise to constraints (1). Another interpretation of the tolerances gives rise to the following (different) assessment problem:

Find $a, b,$ and $r_0$ satisfying the parameter constraints

$$r_0 - 49.050 \geq 0,$$
$$50.050 - r_0 \geq 0,$$

(3) See section 3.3.
and form constraints
\[
\begin{align*}
\max_i r_i - r_0 & \geq 0, \\
r_0 - \min_i r_i & \geq 0, \\
0.020 - (\max_i r_i - \min_i r_i) & \geq 0,
\end{align*}
\]
i = 1, \ldots, 50.

The formal specification of tolerances in terms of parameter and form constraints removes the problem of ambiguity in tolerance specification and moreover facilitates greatly the automatic generation of tolerance assessment software from tolerancing specifications.

### 3.3 MEASUREMENT STRATEGY

One of the most important factors governing the effectiveness of tolerance assessment is the measurement strategy used to gather the data points \(X\): i) the number of data points to collect, and ii) their location on the workpiece surface. The strategy adopted will depend on a) functional requirement, b) stability of the manufacturing process and c) the nature of the tolerances. Clearly, the more data points that are collected, the more reliable the tolerance assessment is likely to be. However, economic considerations will limit the number of data points it is possible to gather and it will be necessary to balance the need for reliable information with these constraints.

**Example: circularity of a lobed circle.** Consider the problem of determining the circularity of a lobed circle. This problem has practical importance since many manufacturing processes introduce lobing to surfaces of revolution. A circle with \(q\) lobes is described in polar coordinates by
\[
r(\theta) = r_0 + A \cos(q\theta + \theta_0);
\]
the circularity of such a circle is \(2A\). If we sample the circle at angles \(\theta_i\), \(i \in I\), then the estimate of circularity obtained from the points is bounded above by
\[
2a = \max_i A \cos(q\theta_i + \theta_0) - \min_i A \cos(q\theta_i + \theta_0).
\]

We wish to choose \(\theta_i\) so that this bound is as close to \(2A\) as possible. If there are \(m\) points uniformly spaced around the circle, it is not difficult to show that if \(m\) and \(q\) have no common factor, then
\[
A \geq a \geq A \cos \frac{\pi}{m}, \quad m \text{ even,}
\]
\[
A \cos^2 \frac{\pi}{2m} \geq a \geq A \cos \frac{\pi}{2m}, \quad m \text{ odd.}
\]
We note that for \(m\) odd, \(2a\) underestimates the circularity by at least a factor of \(\cos \frac{\pi}{2m}\) and we can therefore take as our estimate of \(A\)
\[
\hat{a} = a / \cos \frac{\pi}{2m} \geq A \cos \frac{\pi}{2m}.
\]
Table 1 shows the value of \(\cos \frac{\pi}{2m}\) for \(m\) small and prime. Five points will detect 95% of the lobing if the number of lobes \(q\) is not a multiple of 5, while eleven points will detect 99% of the lobing if \(q\) is not multiple of 11. The use of an even number of points for detecting lobing is not recommended. Figure 1 shows two distributions of six points on a three lobed circle; one set marked * detects 100% of the lobing while the other set marked o fails to detect any lobing.
Table 1  Values of $\cos \frac{\pi}{2m}$ for small primes.

<table>
<thead>
<tr>
<th>$m$</th>
<th>$\cos \frac{\pi}{2m}$</th>
<th>$m$</th>
<th>$\cos \frac{\pi}{2m}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.9511</td>
<td>19</td>
<td>0.9966</td>
</tr>
<tr>
<td>7</td>
<td>0.9749</td>
<td>23</td>
<td>0.9977</td>
</tr>
<tr>
<td>11</td>
<td>0.9898</td>
<td>29</td>
<td>0.9985</td>
</tr>
<tr>
<td>13</td>
<td>0.9927</td>
<td>31</td>
<td>0.9987</td>
</tr>
<tr>
<td>17</td>
<td>0.9957</td>
<td>37</td>
<td>0.9991</td>
</tr>
</tbody>
</table>

By contrast the seven points shown in Figure 2 detects 97.5% of the lobing, slightly above the theoretical minimum given in Table 1.

![Figure 1](image1.png)

Figure 1. Two sets of six points uniformly spaced on a three-lobed circle. The points marked * detect 100% of the lobing while the points marked o fail to detect any form deviation.

![Figure 2](image2.png)

Figure 2. Seven points uniformly spaced on a three-lobed circle detect at least 97% of the lobing.

The analysis in the example above exploits a known model for the expected form error, namely lobing. In the general case, only partial models for the form error will be known and it may be
necessary to use empirical models such as spline curves and surfaces to take into account the errors in form introduced by the manufacturing process.

In more detail, suppose the nominal form is given by \( a \mapsto S(a) \). We assume that the manufactured artefact \( S \) can be described as

\[
S = S(a) + \mathcal{E}(\mathbf{v}),
\]

where the \( \mathcal{E}(\mathbf{v}) \) models the form error in terms of parameters \( \mathbf{v} \) which represent known form errors (such as lobing) and/or empirical errors. The measurement strategy problem is:

Given tolerances on form represented by \( F(S,a) \geq 0 \), choose a finite subset of points \( X \subseteq S \) such that if

\[
F(X;a) \geq 0,
\]

then the tolerances on \( S \) also hold.

A probabilistic approach to this problem can be adopted where it is assumed that the parameters \( a \) and \( \mathbf{v} \) belong to known statistical distributions. For any given \( X \) we determine (through simulation or otherwise) the probability that \( X \) will indicate correctly that the part is within tolerance. This information can be fed into a cost function \( V \) depending on, for example, the potential costs of making a wrong decision and the cost of performing the measurements \( X \). The optimal distribution can then be sought:

\[
\min_X V(X).
\]

It is clear that finding a good definition for such a function is a non-trivial matter but in many ways it is key to developing an effective tolerance assessment and quality improvement strategy. Again, this type of calculation can be performed at the design stage.

3.4 MEASUREMENT ACCURACY

In the discussion on measurement strategy above it is implicitly assumed that the measurements \( X \) represent points lying on the surface of the artefact \( S \). In fact each data point \( x_i \) will be subject to a measurement error \( e_i \); so that

\[
x_i = x_i^* + e_i,
\]

with \( x_i^* \in S \) denoting the point on \( S \) nearest \( x_i \). The tolerance assessment is performed using \( X \) and if the errors \( e_i \) are significant compared to the form error (or the tolerances on the form error), the effectiveness of assessment may be seriously undermined.

Example: flatness of surface tables. In the assessment of the flatness of a surface table, the gradients between between pairs of points lying on the surface of the table are measured. These gradients are used to estimate the minimal separation of two parallel plane containing the surface of the table (i.e., its flatness). However the errors in the gradient measurements mean that even if the table was perfectly flat, the flatness estimate would be non-zero; see [7, 9]

For simple tolerance assessment, the minimal requirement on measurement accuracy is that the measurements can be used to distinguish with a high probability of correctness between parts which are within and out of tolerance. For process control, it is also necessary to detect drifts in the manufacturing process and this will demand greater accuracy.
3.5 MATHEMATICAL PROGRAMMING AND TOLERANCE ASSESSMENT

The tolerance assessment problem is, as stated in section 2.4:

Find parameters \( \mathbf{a} \) (if they exist) to satisfy the constraints

\[
\begin{align*}
C(\mathbf{a}) & \geq 0, \\
F(X; \mathbf{a}) & \geq 0.
\end{align*}
\]

(2)

A set of parameters \( \mathbf{a} \) satisfying the constraints is known as a feasible point; the constraints divide parameter space into feasible regions and infeasible regions. Thus the tolerance assessment problem is to determine a point in a feasible region. Most problems of this nature are solved by formulating an associated optimisation problem whose solution will provide a feasible point if one exists. For tolerance assessment we consider the optimisation problem

\[
\max_{\mathbf{a}, s} s
\]

subject to the constraints

\[
\begin{align*}
C(\mathbf{a}) & \geq s, \\
F(X; \mathbf{a}) & \geq s.
\end{align*}
\]

Note that this is the same formulation as that given for the process control problem in section 2.5. In general, algorithms to solve (3) will start with an initial estimate of \( \mathbf{a} \), determine the maximal \( s \) satisfying the constraints and then iteratively adjust \( \mathbf{a} \) so as to increase \( s \). If it is only required to find a feasible \( \mathbf{a} \), the algorithm can be stopped as soon as \( s \) becomes positive. Otherwise the algorithm can continue until a (local) maximum has been found. In this way, two problems can be addressed by the same algorithm: a part can be passed as acceptable as soon as a positive \( s \) has been found, and information about the manufacturing process can be found by determining the maximal \( s \).

In geometrical terms, the algorithm starts at a point in the infeasible region and makes progress towards the feasible region. \( s \) becomes positive when the boundary has been crossed and the maximal \( s \) is an estimate of the radius of the feasible region. If the part is out of tolerance then the feasible region is empty.

If the constraints in (3) are linear in the parameters \( \mathbf{a} \) then the optimisation problem is a linear programming problem which can be solved using the simplex algorithm. More likely, the constraints will be non-linear and non-linear mathematical programming techniques will be required. For complicated artefacts with many components defined by many parameters, the tolerance assessment problem in its full generality presents considerable difficulties. General purpose optimisation software, for example sequential quadratic programming routines available in the NAG software library [21], can be applied to these problems but these algorithms may not be suitable for implementation in real time applications. In the following sections we consider ways of simplifying the assessment problems.

4 DECOMPOSITION OF MULTI-COMPONENT PROBLEMS

For many multi-component problems, the form constraints decompose into sets of form constraints, one for each component \( S_k \), so that the optimisation problem can be written as
subject to the constraints

\[
C(a) \geq s, \quad D(X_k; a_k) \geq s, \quad k = 1, \ldots, n_K.
\]

**Example: two hole problem.** Suppose two holes are each required to have radius at least 10mm and be set at 100±0.2mm apart. These tolerance constraints can be encoded as

\[
[(a_1 - a_2)^2 + (b_1 - b_2)^2]^{1/2} - 99.8 \geq 0, \\
100.2 - [(a_1 - a_2)^2 + (b_1 - b_2)^2]^{1/2} \geq 0,
\]

and

\[
[(x_i - a_1)^2 + (y_i - b_1)^2]^{1/2} - 10.0 \geq 0, \quad i \in I_1, \\
[(x_i - a_2)^2 + (y_i - b_2)^2]^{1/2} - 10.0 \geq 0, \quad i \in I_2.
\]

The two sets of form constraints are independent of each other in that they have no parameter in common: the link between the parameters is provided by the parameter constraints.

One method of finding an approximate solution of (4) is to solve a sequence of problems

\[
\max_{a_k, s_k} s_k
\]

subject to

\[
D(X_k; a_k) \geq s_k,
\]

to determine parameters \( \hat{a}_k \) and then evaluate the parameter constraints using these parameters.

**Example: two hole problem II.** For the two hole problem we can solve the two maximum inscribed problems

\[
\max_{a_k, b_k, s_k} s_k
\]

subject to

\[
[(x_i - a_k)^2 + (y_i - b_k)^2]^{1/2} - 10.0 \geq 0, \quad i \in I_k,
\]

to determine \( \hat{a}_k, \hat{b}_k \), and \( s_k, k = 1, 2 \), and then evaluate

\[
[(\hat{a}_1 - \hat{a}_2)^2 + (\hat{b}_1 - \hat{b}_2)^2]^{1/2}.
\]

There are a number of advantages in proceeding this way:

**Reduction in complexity.** The optimisation subproblems are of reduced complexity since they only involve a subset of the optimisation parameters \( a \). In very broad terms, if the \( k \)th subproblem involves \( n_k \) parameters then the computational effort for the \( k \)th subproblem will be proportional to \( n_k^2 \) compared with \((\sum n_k)^2\) for the original problem. The subproblems will likely be easier to solve so that the iterative algorithms will take fewer iterations leading to further gains in efficiency.
Tolerance information on each component. The solution of the $k$th subproblem will provide information about the form of the $k$th subcomponent. For the general problem, this information is not immediately available; its solution may indicate that the workpiece is out of tolerance without showing which individual subcomponents (if any) are out of form.

Degenerate solutions. If in solving the amalgamated problem the maximal $s$ is determined by the form constraints of one subcomponent alone, then the parameters determining the other subcomponents are likely to be ill-defined in that changing them has no effect on the maximum of $s$. This phenomenon is known as degeneracy. Although general purpose optimisation software should be able to cope with degenerate solutions, their presence usually degrades the effectiveness of the algorithms. The individual subproblems are less likely to be degenerate.

Example: two hole problem III. In the two hole problem, if the first hole is significantly smaller than the second, then the coordinates of the second centre are ill defined; see Figure 3.

![Figure 3](image-url)

There are also a number of drawbacks for the approach:

Solutions are approximate. The solution parameters $\hat{a}_k$ to the subproblems may not satisfy the parameter constraints even although the part satisfies the tolerances. This reflects the fact that solving the sequence of subproblems is not equivalent to solving the original problem. However, the solution parameters $\hat{a}_k$ can be used as starting estimates in solving the complete problem.

Ill-posed subproblems. Although the complete tolerance assessment problem may be well-posed it is possible that the subproblems are not.

Example: two hole problem IV. Figure 4 shows a set of three points on each of the two circles. The complete tolerancing problem defined by the parameter and form constraints is well posed for this data in that there is a well-defined local maximum for $s$: the constraints on
the separation of the circle centres limit the size of the radii of the inscribed circles. However, each of the subproblems is ill-posed since the maximum inscribed circle to each set of the three points is not defined.

![Figure 4](image)

Figure 4. In the two hole problem, the constraints on the separation of the circle centres limit the size of the radii of the inscribed circles.

The conditioning of the individual subproblems depends on the measurements $X_k$. The locations of each set of measurements should be determined so as to lead to well-posed subproblems and in this way avoid situations like the one illustrated above. These questions should be addressed at the design stage when determining measurement strategy.

5 GEOMETRIC ELEMENTS

We have argued that the decomposition of tolerance assessment into components is a sensible approach. In this section, we look at geometric tolerance assessment problems associated with the simple geometric elements: lines, planes, circles, spheres, cylinders, and cones.

Methods of parametrizing these elements are considered in [10, 17, 2]. The parametrization of partial arcs and caps of spheres is considered in [18, 13]. Given a parametrized element $a \mapsto S(a)$ we can calculate the distance $d(x;a)$ from a point $x$ to $S(a)$ in terms of the parameters $a$.

**Zone tolerances.** Zone tolerances specify the maximum deviation $\tau$ from nominal form:

Given a surface $S$ of nominal form $S(a)$, there must exist $a$ such that for all $x \in S$

$$|d(x;a)| \leq \tau.$$  

Given a set of data points $X = \{x_i : i \in I\}$ lying on the surface of $S$ the related optimization problem can be expressed as

$$\min_{a,s} s$$

subject to

$$|d(x_i;a)| \leq s, \quad i \in I.$$
If the minimal $s$ is less than $\tau$, then the tolerance has been met. The optimisation problem above is a type of Chebyshev ($\text{minimax}$, $L^\infty$) approximation problem. The commonly used terms straightness, flatness, circularity, sphericity, cylindricity and conicity refer to this type of tolerance.

**Inscribing and circumscribing tolerances.** Inscribing and circumscribing tolerances require that an artefact can either fit inside an element of fixed size or can contain an element of fixed size. The element for which these problems are defined are the circle in a fixed plane, sphere and cylinder. The associated optimisation problems are i) the maximum inscribing problem:

Given data points $X = \{x_i : i \in I\}$,

$$\max_{a,s} s$$

subject to

$$d(x_i; a) \leq s, \quad i \in I,$$

and ii) the minimum circumscribing problem:

Given data points $X = \{x_i : i \in I\}$,

$$\min_{a,s} s$$

subject to

$$d(x_i; a) \geq s, \quad i \in I;$$

For both problems $d$ is the distance from a point to the circle or sphere centre or cylinder axis.

Most functional requirements will lead to a mixture of form and parameter tolerances, typically a zone tolerance in addition to tolerances on radius (or angle for a cone).

Despite the apparent simplicity of the element tolerance assessment problems, the nonlinear nature of the constraint equations means that these problems are far from 'elementary'. For lines, planes, circles in the plane, and spheres it is in fact possible to reparametrize the elements so that the constraint equations are linear and this does indeed lead to simpler optimisation algorithms. For the cylinder, cone and circle in three dimensions (or torus) no such reparametrisation is possible and effective algorithms have to embrace a minimal amount of optimisation technology.

### 5.1 Notes and References

For discussions on optimisation subject to constraints see for example, [14, 22, 16]. Nonlinear Chebyshev approximation problems have received much attention, including [26, 34, 25, 11]. In particular, successful algorithms for solving linear Chebyshev approximation problems are available e.g. [6]. Tolerancing for circles is set out in BS3730 [8] and algorithms for the related optimisation problems are described in [3]. Algorithms for solving the minimum zone, minimum circumscribed and maximum inscribed problems are described in [4]. These algorithms can be readily adapted to incorporate constraints on the element parameters.
6 TEMPLATE MATCHING

In this section we consider one of the simplest types of tolerance assessment problems where the only parameters are those defining a change in the frame of reference. It is also one of the most common in practice. Suppose as before that \( a \mapsto S(a) \) represents the nominal form with ideal parameters \( a_0 \). The surface \( S(a_0) \) specifies a fixed surface or template in a fixed frame of reference. Suppose also that data points \( X \) are gathered on the surface of a workpiece by a CMS. As noted in section 2.3, the coordinates of the data points \( x_i \) will be in a different frame of reference from that of \( S \). The template matching problem is to find a frame transformation \( T \) which maps the data onto the template in such a way that the transformed data points will satisfy the form constraints.

To formulate this problem more explicitly, we write the transformation \( T \) as a combination of translations and rotations depending on parameters \( t \).

**Example: Affine transformation in two dimensions.** A general affine transformation \( T \) in two dimensions is expressed in terms of three parameters \( t = (x_0, y_0, \theta) \) by

\[
\begin{bmatrix}
  \hat{x} \\
  \hat{y}
\end{bmatrix} = \begin{bmatrix}
  \cos \theta & \sin \theta \\
  -\sin \theta & \cos \theta
\end{bmatrix} \begin{bmatrix}
  x - x_0 \\
  y - y_0
\end{bmatrix}.
\]

(7)

The template tolerance assessment problem is:

Given a template \( S(a_0) \) and data points \( X = \{x_i : i \in I\} \) representing points on an artefact \( S \) of nominal form \( S(a) \), find \( t \) to satisfy the form constraints

\[
F(\hat{X}; a_0) \geq 0,
\]

where \( \hat{X} = \{\hat{x}_i : \hat{x}_i = T(t;x_i), \ i \in I\} \).

The associated optimisation problem is:

\[
\max_{t,s} s
\]

subject to

\[
F(\hat{X}; a_0) \geq s.
\]

**Example: Zone tolerances on a template.** We can consider zone tolerances on the template of the form

\[
|d(x; S(a_0))| \leq \tau,
\]

where \( d(x; S(a_0)) \) is the distance from a point \( x \) to the surface \( S(a_0) \). Given a data set \( X = \{x_i : i \in I\} \) of points on the surface of an artefact \( S \), the tolerance assessment problem is:

Find \( t \) such that

\[
|\hat{d}(x_i; t)| \leq \tau, \ i \in I,
\]

where

\[
\hat{d}(x; t) = d(\hat{x}, S(a_0)) = d(T(t;x); S(a_0)).
\]
The associated *matching* problem is
\[
\min_{t,s} \quad s
\]
subject to
\[
|\hat{d}(x;t)| \leq s, \quad i \in I.
\]
Note that (8) is a Chebyshev approximation problem in the parameters \(t\).

In some circumstances the tolerance \(\tau\) will be different for different parts (or components) of the template and this can be catered for by modifying the definition of \(d(x;S(a_0))\).

The major advantage in formulating tolerance assessment problems as template matching problems is that no matter how complicated the shape is, the only optimisation parameters are the transformation parameters \(t\) since \(a_0\) is fixed. This means that practical algorithms can be implemented for this type of tolerance assessment.

## 7 PART MATING

One of the most important tolerance assessment problems is to decide if two parts \(S_X\) and \(S_Y\) will fit together and if so how closely. We assume that the surfaces of parts \(S_X\) and \(S_Y\) each have nominal form \(a \leftarrow S(a)\) and we try to find a *separating surface* \(S(a)\) such that \(S_X\) lies on one side of \(S(a)\) and \(S_Y\) lies on the other. This information can be encoded by *separation constraints* of the form
\[
d(x;S(a)) \geq 0 \geq d(y;S(a)),
\]
for \(x \in S_X\) and \(y \in S_Y\). We will denote the set of separation constraints generated by measurements \(X\) and \(Y\) on \(S_X\) and \(S_Y\), respectively, by
\[
D(X;a) \geq 0 \geq D(Y;a).
\]

The general part mating problem can now be formulated as:

Find \(a\) satisfying the parameter constraints
\[
C(a) \geq 0,
\]
the separation constraints (9) and the form constraints (if any)
\[
F_X(X,a) \geq 0,
F_Y(Y,a) \geq 0.
\]

The problem implicitly requires one or two frame transformations to relate \(X\) and \(Y\) to the same frame of reference.

**Example: separating circles.** Consider a hole plate with \(n_K\) holes and corresponding plug plate with \(n_K\) circular plugs. The obvious question is: will the plugs fit in the hole plate? If \(X_k = \{x_i: i \in I_k\}\) \((Y_k = \{y_j: j \in J_k\})\) are data sets gathered from circles in the hole (plug) plate, then this mating problem can be expressed as:

Find an affine transformation \(T\) of the form given by equation (7) and circle centres \(a_k\) and radii \(r_k\) such that
\[
\|x_i - a_k\| - r_k \geq 0 \geq \|y_j - a_k\| - r_k, \quad i \in I_k, \quad j \in J_k, \quad k = 1, \ldots, n_k,
\]
where \(y_j = T(t;y_j)\).
7.1 MATING BY MATCHING

The part mating problem outlined above asks if a particular part $S_X$ represented by data points $X$ will mate with a second part $S_Y$ represented by data points $Y$. In practice, it is more useful to set separate tolerances for the $S_X$ and $S_Y$ parts so that any pair satisfying these tolerances will mate. This can be done by defining a template $S(a_0)$ for the separating surface. The associated assessment problem decomposes into two template matching problems:

Given data sets $X$ and $Y$, find i) transformation parameters $t_X$ satisfying the one-sided constraints

$$D(\hat{X};a_0) \geq 0,$$

and form constraints

$$F(\hat{X};a_0) \geq 0;$$

and ii) transformation parameters $t_Y$ satisfying the one-sided constraints

$$0 \geq D(\hat{Y};a_0),$$

and form constraints

$$F(\hat{Y};a_0) \geq 0,$$

where

$$\hat{X} = \{ \bar{x}_i = T(t_X; \bar{x}_i), \ i \in I \},$$

$$\hat{Y} = \{ \bar{y}_j = T(t_Y; \bar{y}_j), \ j \in J \}.$$  

We emphasise that template matching problems are much easier to solve than separating surface problems.

7.2 NOTES AND REFERENCES

[33] also considers the hole and plug plate problem.

8 APPROXIMATE METHODS

One method of simplifying the tolerance assessment problem is to replace it by easier approximate problem! Before discussing how this can be done in practice, we make a few general observations.

The tolerance assessment problem as stated in section 3.5 is one of finding a point in the feasible region $R$ for the constraints. Any method which generates points which lie in the feasible region will be satisfactory. Thus we would like to find a simpler tolerancing problem whose feasible region $\tilde{R}$ has a high degree of overlap with that of the original problem. Furthermore, we can hope that by finding a feasible point well inside $\tilde{R}$ (for example by finding the maximum of the related optimisation problem) there will be a high probability that the point so generated will be feasible for $R$. We may be able to choose a measurement strategy which improves the probability, perhaps to a certainty.

If we can indeed find a simplified tolerance assessment problem with good properties then we may ask why the original assessment problem was posed in the first place. With adequate design, the tolerance specification should be such that the related assessment problem is as
simple as possible subject to meeting the functional requirements. More likely, we can find an approximate problem which generates points to solve the original problem in the majority of cases and can be used as starting estimates for feasible points for the remainder. Note that to show a part is out of tolerance, the original problem must be solved.

We turn now to methods of simplifying the assessment problem.

Reducing the number of parameters. We have already described a number ways in which this can be done: see sections 4, 6, 7.1.

Replacing constrained optimisation by unconstrained optimisation.

Example: Least-squares approximation. Zone tolerance assessment problems generate Chebyshev approximation problems where we are required to minimise the maximum of a set of functions $|d(x; a)|$: a constrained optimisation problem. If instead we minimise the root-mean-square of these functions we get a least squares unconstrained optimisation problem

$$\min \sum_{i \in I} d^2(x_i; a),$$

which can be solved using the Gauss-Newton algorithm, for example, to determine the (locally) optimal $\hat{a}$. We can then (over-)estimate the zone separation by

$$\frac{\max d(x_i; \hat{a}) - \min d(x_i; \hat{a})}{2}.$$

Replacing nonlinear problems by linear problems.

Example: minimum zone circle. The minimum zone circle problem is:

Given a data set $X = \{x_i : i \in I\}$,

$$\min_{a, b, r_0} \max_i |r_i - r_0|$$

where

$$r_i = \left[(x_i - a)^2 + (y_i - b)^2\right]^{1/2}.$$

We consider instead the problem

$$\min_{a, b, r_0} \max_i |r_i^2 - r_0^2|.$$  \hspace{1cm} (11)

If we set $\rho = a^2 + b^2 - r_0^2$, we have

$$r_i^2 - r_0^2 = x_i^2 + y_i^2 - 2x_i a - 2y_i b + \rho,$$

which is linear in the parameters $a$, $b$ and $\rho$. Thus (11) is a linear Chebyshev approximation problem which can be solved by linear programming methods.
We can compare the solutions to the two problems. Suppose the minimum $\tau$ of (10) is attained with radii $r_i^*$, $r_0^*$ and the minimum $\sigma$ of (11) is attained with radii $\tilde{r}_i$, $\tilde{r}_0$. Then,

$$\tau \leq \max_i |\tilde{r}_i - \tilde{r}_0| = \max_i \{|\tilde{r}_i^2 - \tilde{r}_0^2|/(\tilde{r}_i + \tilde{r}_0)\},$$

$$\leq \max_i |\tilde{r}_i^2 - \tilde{r}_0^2|/\min_i (\tilde{r}_i + \tilde{r}_0),$$

$$\leq \sigma/(\tilde{r}_0 + \min_i r_i),$$

$$\leq \sigma/(\tilde{r}_0 + (\tilde{r}_0 - \sigma)^{1/2}).$$

Thus solving the linearised minimum zone circle problem gives an upper bound for $\tau$.

Conversely,

$$\sigma \leq \max_i |r_i^* - r_0^*| = \max_i \{|(r_i^* + r_0^*)|/\max_i |r_i - r_0|\},$$

$$\leq \max_i (r_i^* + r_0^*)\tau,$$

$$\leq (2r_0^* + \tau)\tau.$$

We can rearrange the last equation to read

$$\tau \geq \sigma/(r_0^* + (r_0^* + \sigma)^{1/2}),$$

which will give a lower bound for $\tau$ if we have an upper bound for $r_0^*$. Unfortunately, this will not always be the case.

While approximate methods can be very valuable for finding feasible points and supplying initial estimates of parameters, their indiscriminate use could lead to confusion. For example, a supplier may insist a part is within tolerance and a customer maintain it is not simply because they use different approximations for the tolerance assessment problem. Ideally, the customer and supplier should agree the design and tolerance specification, the measurement strategy and method of assessment early in the design process.

8.1 NOTES AND REFERENCES

See, for example [22, 15] for methods for least squares approximation. Algorithms for least squares element fitting are described in [17, 19].

9 SUMMARY AND CONCLUSIONS

The tolerance assessment problem is vital to effective quality improvement and it is important that all issues, from functional requirement, tolerance specification, measurement strategy to assessment algorithms are thought through and analysed at the design stage.

Design and tolerance information can be stated in terms of nominal form, parameter constraints, and form constraints involving measured data points. The associated tolerance assessment problem requires finding a point (a set of parameter values) which is feasible with respect to the constraints. This problem can be solved using mathematical programming techniques which in addition can supply important information about the behaviour of the manufacturing process.
Tolerance assessment problems in their full generality require significant algorithmic sophistication and computational effort. However, many important tolerance assessment problems can be posed as simplified matching problems for which practical algorithms can be developed and implemented as software.

10 ACKNOWLEDGEMENTS

The work reported on here was carried out as part of the ‘Mathematical Algorithms for Metrology’ project in the National Physical Laboratory’s Length programme. This programme is in support of the National Measurement System and is funded by the Department of Trade and Industry. Project members G T Anthony, B P Butler, Dr M G Cox, S A Hannaby and Dr P M Harris have contributed to this work. In particular, Gerald Anthony and Maurice Cox gave many helpful comments on drafts of this report.

11 REFERENCES


