CONTROLLING FORM ERRORS FROM 3D MEASURES

Jean-François Debongnie1 et Luc Masset2

1. *Professor, Manufacturing Laboratory, University of Liège*
2. *Research Assistant, Manufacturing Laboratory, University of Liège*

**ABSTRACT**

Owing to the fact that a correct evaluation of form errors is particularly difficult by the classical ways, it seems more efficient to develop numerical algorithms from 3D measures. Several methods are described, including direct and iterative ones. A new method is proposed, which is based on Lp‑norms and a specialized algorithm. It proved to be the most effective and robust one on every studied application.

# Introduction

Form error specifications are very frequently indicated on mechanical drawings, in order to ascertain the functionality of assemblies or mechanisms. Unfortunately, a direct measure of form errors, if not impossible, is very difficult.

As an example, when measuring a flatness error, the operator first seeks an appropriate orientation of the piece, where it is *more or less* parallel to the reference axes. He then measures the *parallelism* of the surface to the reference plane. It is clear that another operator, having made another "best" choice of the orientation, would obtain another flatness error.

In a circularity control, the classical base is a set of radii measured at different values of the azimuth. One has then to find two concentric circles which exactly enclose the measured profile, the center of which is chosen in order to obtain the minimum radial separation [**1**]. Such a determination is very delicate. By far more delicate is the determination of cylindricity errors, in which one has to find the most appropriate position and orientation of the axis. As reported by MEADOWS [**1**], researches were conducted on this problem in GDR and led to the TGL39097 norm in which *simplified* procedures are described, whose results are not always identical.

This is to say that classical methods fall down on the point of *objectivity*, so that the way is open for disputes between supplier and customer. It therefore seems necessary to treat the problem by a *numerical algorithm* avoiding any initial positioning effect.

The proposed method consists to start from 3D measures and to compute the form error with suited algorithms. The fundamental exigency concerning these algorithms is the ability to find *the exact value of the form errors*, that is, the minimum value of the separation of measured points. The objectivity exigency is then satisfied, as the solution is uniquely determined, at least concerning the error value.

The present paper is devoted to the description of several algorithms which lead to the correct value of the form error, including a new one, which is based on Lp-norms with very high values of p.

# General statement of the form error problem

## Separation Function

It is useful to start with a general and formal description of the form error problem. 3D measures lead to a compact set *K* of measured points. Any form tolerance problem may be expressed by a *separation* concept. In fact, straightness is defined by the separation of two extreme straight lines, circularity by the separation of two extreme concentric circles, flatness by the separation of two extreme planes and cylindricity by the separation of two extreme co-axial cylinders.

From a mathematical point of view, one may define a continuous separation function *f* (***x*** ,λ) depending on the coordinates ***x*** and on a set of parameters λ.

In straightness problems, the separation function is given by



***Figure 1.*** *Separation function and separation value for straightness*



For flatness errors, the measured surface being supposed approximately horizontal, the separation function will be



with n = (cos θ, sin θ cos φ, sin θ sin φ). The angles θ and φ are chosen so as to have the pole on the x axis, in order to avoid numerical difficulties with a quasi-vertical normal (fig. 2).

To treat circularity problems, let a and b be the coordinates of a center. The separation function is then (fig. 3)

***Figure 2.*** *Definition of angles* θ *et* φ





The case of cylindricity may be reduced to a circularity problem by defining an axis whose direction is



and then projecting the measured points (x ,y ,z) on the two orthogonal axes



***Figure 3.*** *Separation function and separation value for circularity*



The separation function is then



where (a , b) are the coordinates of a center and (xp ,yp) the coordinates of the point projected in the plane ( e 1, e 2 ).

## Separation value and form error of a compact *K*

The *separation value* of a compact *K*, for a given value of parameters λ, is defined by



As an example, for flatness problem, the separation value of the measured surface is its parallelism error to a plane perpendicular to n. This being posed, the *form error* of compact K is the least separation value of this compact when the set of parameters λ is varied,



The fact that this infimum is reached for some value λ, in other words, that it is a true minimum, may be proved from compacity properties which go beyond the scope of the present paper.

## Unicity

More awkward is the question o *unicity* of this minimum. Can there exist more than one set of parameters λ leading to the same minimum value of the separation ? It can be proved that for circularity, the solution is unique. In contrary, for straightness, flatness and cylindricity, the uniqueness is not necessarily guaranteed, and counter-examples are easy to exhibit. However, each known counter-example supposes some particular symmetries which are seldom encountered in practical problems, so that the question of unicity, however serious from a theoretical point of view, seems not to have practical incidences.

## Formulation based on the concept of deviation

There exists a second equivalent formulation of the problem of form errors. Let us introduce a supplementary parameter ρ and define the *deviation* as



Defining the *maximum deviation* on *K* by



it is easy to see that



the minimum being reached when



This property leads to a new definition of the form error, namely



In this context, the determination of the form error consists in finding the surface of equation



from which the maximum deviation is a minimum. It is thus a problem of *best uniform approximation*, very close to the problem of function approximation in the sense of Chebyshev [**2**,**3**,**4**,**5**].

# Approximate computation of form errors by least square approximation

The uniform approximation being a difficult nonlinear problem, most authors proposed to replace it by a minimization of the root of the sum of the squared deviations,



where *dev* ( *i* ) = *dev* (*xi*, λ , ρ ). This leads to a value (λ2 , ρ2) of the set of parameters, from which an approximate value of the form error is obtained, namely,



This value is of course an *upper bound* of the error form, since λ2 and ρ2 are not the optimal values of the parameters. Experiences show that this method may lead to an appreciable excess on the form error and that the orientation parameters (λ ,ρ) are very sensitive to the distribution of measured points, tending to conform to zones where they are numerous. As a conclusion, this method, which does not agree with the standards, cannot be considered as reliable.

# Direct methods

*Direct methods* are defined as processes in which the exact value of the form error is reached by a *finite* (although not necessarily low) number of operations. Such methods may be based on a geometrical analysis of the problem or on a transposition of reasoning ways of the classical theory of Chebyshev approximations.

## Straightness error

As the straightness of some line *K* is defined by the separation of two straight lines containing the given line, it is clear that these two straight lines also contain the convex closure of *K* . The first step is therefore the determination of this closure. This may be done by several classical methods. It may now be proved that any separation which remains valid when the orientation of the two lines is perturbed cannot be optimal. As a consequence, the optimal separation is such that one at least of the two right lines contains one side of the convex closure. So, it suffices to compute the separation values for each case where one of the two lines is a side of the convex closure, the form error being the least of them. This algorithm is interesting from two points of view. Firstly, it is easy to code. Secondly, it permits to isolate cases where the optimal separation is not unique. As an example, figure 4 shows a profile which is defined by 6 points and for which two different orientations D1 and D2 lead to the same minimum separation. The least squares criterium leads to the line LS. The variation of the separation value with the orientation φ of the straight lines is given in figure 5. Note that the minima are sharp, a point that will be referred later.

***Figure 4.*** *A case of non-unicity of the straightness problem*



***Figure 5.*** *Separation value versus angle φ for the straightness problem of figure 4*



It is clear that in this case, the non-uniqueness is due to a very particular symmetry of the set of measured points, a situation that is seldom encountered in practical applications. However, such a non-unique definition of the optimal orientation would lead to philosophical problems if, as an example, the measured line has to be used as a simulated reference for parallelism.

## Direct evaluation of flatness error

The same way of reasoning may be transposed to the flatness problem. However, numerical difficulties may arise when constructing the convex closure of the measured points, from the fact that in practical cases, the measured points are at a low distance of a true plane. A very careful, conditioning-oriented procedure has to be used, so that the computation cost may be somewhat greater than expected. After a lot of tests, we finally adopted an algorithm that for n measured points, oscillates between an *O* ( *n* 2) and an *O* ( *n* 3 ) complexity.

In this case, the optimal separation is constituted by two planes having one of the following properties.

1. One plane contains a face of the convex closure, the other one containing an apex.
2. Both planes contain non parallel edges of the convex closure.

Explorating all edge couples is theoretically of complexity *O* (*n* 4), but this research may be limited by *a priori* excluding a lot of couples which are not interesting.

This algorithm, although conceptually simple, necessitates a very careful implementation to be reliable. Here also, cases of non-uniqueness may be produced, but we never encountered this problem in practical applications.

## Direct evaluation of circularity

A transposition of classical results concerning Chebyshev approximation [**2**,**3**], leads to the following results

1. The optimal separation is unique;
2. There exist at least four extremal points, such that if ordered by increasing azimuth values, they alternatively lie on the great and on the little circle. This property constitutes a *characterization* of the optimal separation;
3. If one finds a circle centered at some point *C* = (*a* , *b*) and having a radius ρ, such that there exists four points ordered in azimuth P1, P2, P3 and P4 that verify



with α = 0 or 1, i.e., they are alternatively situated inside and outside the considered circle, then the circularity error verifies



This property may be used to compute the circularity error. The procedure consists to arbitrarily select four measured points of increasing azimuth. Let be Pi = (*x* i,yi) these points. One then seeks a point C = ( c , d) which is simultaneously at the same distance from P1 and P3 and at the same distance from P2 and P4. This leads to the linear system



Point *C* is then automatically the center of a circle verifying condition (17). Noting



is follows from the above properties b and c that



This constitutes a direct method of evaluating the circularity error. Unfortunately, it is of complexity *O* ( *n* 4) and is therefore hardly practicable when the number of measured points exceeds 100, say.

## Cylindricity

To our knowledge, no direct method exists for the cylindricity error.

# Nelder-mead simplex method

In opposition with direct methods, iterative algorithms may be used, which tend to minimize the maximum deviation. Unfortunately, this function is not very regular. In particular, in the neighbourhood of the minimum, the graph of the maximum deviation is funnel-shaped, it is to say that if (λ0, ρ 0) is the minimal point, one has



Moreover, this graph presents a lot of *thalwegs* where the gradient is no more uniquely determined (fig. 6). From this, it results that classical methods based on a gradient evaluation cannot be applied directly. It seems thus more easy to use methods which only use the values of the function itself. In this direction, BALLU *et al* [**5**] and PAULY [**7**] proposed to apply the Nelder-Mead simplex method. From an extensive experimentation, it appeared that with this method, the final result is very sensitive to the initial choice of a simplex, without any criterium to determine if the true optimum is reached or not. We therefore consider that the simplex method is not sufficiently reliable.

***Figure 6.*** *Separation value versus coordinates of the center (circularity error)*



# P-norm method

## General principle

As the maximum deviation, which has to be minimized, is not a regular function, it is temptating to try to replace it by an approximate function which is more regular. In fact, the least squares method, as exposed in section 3, may be interpreted as an application of this idea, where *maxdev* is replaced by ∥*dev*∥2. But here, the substitute differs too strongly from the original function so that the results are only rough approximations. A better choice is given by the p-norms which are defined by



and verify the following property



*n* being the number of measured points. From this follows that when *p* tends to infinity, the p-norm of the deviations uniformly converges to the maximum deviation on any compact of the space of parameters. It is interesting to note that the successive values of the norms form *decreasing* sequence and that, conversely, the successive values of the p-means



form an *increasing* sequence, both converging to *maxdev*. This property is the key of a convergence control (fig. 7). Moreover, it may be proved that

* the minimum value of the p-norm converges to the minimum value of the maximum deviation, that is to say, half of the form error;
* the value of the parameter set (λ, ρ) minimizing the p-norm converges to the value of this set which minimizes the maximum deviation.

Using this property is not a new idea, as it was already proposed by GOCH [**4**], but this author never surpassed values of *p* of 50, probably due to the implied numerical difficulties. In fact, we will see that *by far higher values of p* are necessary in order to obtain a precise determination of the form error.

***Figure 7.*** *P-norm and P-mean versus p*



## Implementation

Replacing the maximum deviation by the p-norm of the deviations leads to a relative error that, from (**24**), may be evaluated as



This is to say that in order to obtain a given precision εp, the necessary value of *p* is



Current values of the number of measured points may be of the order of 1000. In this case, a 10-5 precision necessitates



As it can be seen, these are *very high* values of the exponent. Such exponentiations cannot be practically performed without a proper scaling of the deviations, that may be described as follows

1. Firstly, one writes


in order to avoid any overflow problem. The sum is now constituted of terms that never surpass unity;

1. On any computers, high exponentiations of low numbers is also impossible. It is then necessary to discard any value of the deviation verifying


η being a negligibly small number that however could be taken in account by the computer (10-100, say). This to avoid any numerical difficulty without a significative change of the result.

Using this double artifice, we currently work with p values ranging from 106 to 109.

## Minimization algorithm

The original idea was a direct minimization of the p-norm with a sufficiently high value of *p*, by a Newton-Raphson process. But the greater is *p*, the less regular is the p-norm, and the more long and uncertain is the iterative process. Actually, high values of *p* require a very good initial solution to converge. One may therefore imagine the following procedure.

An increasing sequence *p*1 = 2 < *p*2 < *p*3 · · · is chosen, and one successively minimizes the pi-norm with the optimal point of the pi-1-norm as a starting point. But such a procedure would be very long. Experiences showed that it may be largely shortened as follows

1. Find the minimum for *p* = 2 (least squares);
2. Adopt for the (*pi*) sequence a geometrical progression with √2 as a ratio. For each value of *p*, only one Newton-Raphson iteration is performed, provided no divergence is detected;
3. Convergence is controlled by making use of Jensen's inequality [**8**] from which for a given system of deviations, if *p* < *q*, one has



This implies that for successive iterations, when the process works good, the p-norms will decrease, firstly because *p* increases and secondly because the solution is supposed to be improved. So, the least norm that was obtained during the process is stored at each time. If, after a given iteration, the new norm is found greater than the stored one, the value of *p* is locked until a norm is already obtained which is lower than the stored value.

This algorithm was tested on a large amount of problems, both real and artificial, and compared to all the precedingly described methods. It was found to be the fastest (typically, 40 iterations for *p* = 106) and also the most reliable one.

# Applications

## Presentation

Measures have been performed on three actual mechanical parts, namely

* a cylindrical bar,
* a suspension arm,
* a camshaft bearing support.

These measures were used as data in order to test and compare the different possible algorithms (least squares, simplex, p-norms and direct methods, when available). Table 1 summarizes the different codes which were used in each case. All these codes were beforehand largely tested with arbitrary data, one part of which being randomly generated, the second part being analytical surfaces such as conoïds, ellipses, ... It is perhaps interesting to point out that such "simple" surfaces proved to be the most severe tests.

***Table 1.*** *Implemented codes concerning the 4 form errors (LS=Ieast squares, CC=convex closure, 4P=4 points method. SI=simplex. PN=P-norm)*



## Cylindrical bar

The measured cylindrical bars were cut on a lathe, in the clamped-supported configuration. The seven obtained bars were

measured on 22 circles corresponding to different values of z, each circle being defined by 72 measures. From these results, the following computations may be done

* straightness of the generating lines,
* circularity of the sections,
* cylindricity of the bar.

### Straightness

As illustrated on figure 8, the straightness error, as calculated from the least square procedure is higher than the exact value to an amount of 10 to 15%. The greatest observed difference amounts to 20%.

***Figure 8.*** *Straightness errors on 72 generators of the bar*



### Circularity

As illustrated on figure 9, the optimal separation circles coincide with two interior and two exterior points of the profile which are alternated, as predicted by the theory. In this case also, the least square method led to largely exaggerated values. Figure 10 illustrates the obtained differences on the 22 measured circles.

***Figure 9.*** *Circularity - separation as calculated by the LS approximation and exact separation*



### Cylindricity

The cylindricity errors, as calculated by the different available methods, are compared in table 2. It may be observed that the simplex method never reaches so low values as those obtained with the p-norms.

***Figure 10.*** *Circularity errors of the bar for different values of z*



***Table 2.*** *Cylindricity of the 7 bars*



As no direct method exists in cylindricity, it could be argued that it is not guaranteed that the values obtained with the p-norms are the optimal ones. However, from an analogy with the other form errors, one may conjecture that *the computed separation cylinders have to contain at least 6 measured point*. This from the observed fact that the number of contact points between the optimal separation and the measured points exceeds the number of variables from one unit (see table 3). In the present case, the number of variables being equal to 5, the expected number of contact points is 6, a fact that was verified in all studied cases.

***Table 3.*** *The number of contact points between the optimal separation surfaces and the measured surface, as reported in function of the number of variables*



## Suspension arm

The analysed suspension arm presents two milled planes A and B on which a set of points were measured, as represented on figure 11. Th tolerances concerned the flatness of each plane and their parallelism. Flatness errors, as obtained by the different available methods, are reported in table 4.

***Figure 11.*** *Measured points on the suspension arm*



***Table 4.*** *Flatness and parallelism of planes A and B*



## Camshaft bearing support

The upper face of the part represented in fig. 12 is milled with 5 different cutting conditions. Results concerning each case are reported in table 5.

***Figure 12.*** *Measured surface of the camshaft bearings support*



***Table 5.*** *Flatness of the 5 camshaft bearings supports*



# Conclusions

The first conclusion is that the correct value of the form error, as defined by the standards, cannot be reached by a direct measure, since it corresponds to a sharp minimum. Numerical methods are thus necessary to obtain an objective result.

Turning now to the developed algorithms, one first concludes that the least square method is not satisfactory, leading to an exaggerated value of the form error and a too rough approximation of the separation surfaces.

Direct methods, when available, are attractive since they give the exact solution through a finite process. However, in the case of circularity, the number of operations is very large, so that the method is of less practical interest. Moreover, no direct method is known for the cylindricity error.

Concerning the simplex method, its robustness appeared insufficient, especially in the cylindricity case. This sensitivity to the initial choice of a simplex renders this method not very attractive for practical applications.

Finally, in each of the numerous performed tests, the p-norm method led to the correct result, i.e. the same as with direct methods. It demonstrated to be fast, robust and able to treat the four studied form errors. It seems therefore to be the most suited method for this kind of problems.

# References

[1] J.D. MEADOWS, *Geometric dimensioning and tolerancing*. Marcel Dekker Inc., New York, Basel, Hong Kong, 1995

[2] G. HACQUES*, Algorithmique numérique*, Colin, Paris, 1971

[3] I.S. BEREZIN, N.P. ZHIDKOV, *Computing methods*, vol. 1, Pergamon Press, Oxford, 1965

[4] G. GOCH, *Efficient multi-purpose algorithm for approximation and alignment problems in coordinate measurement techniques*, CIRP Annals, vol. 39/1, pp. 553-556, 1990

[5] A. BALLU, P. BOURDET, L. MATHIEU, *The processing of measured points in coordinate metrology in agreement with the definition of standardized specifications*, CIRP Annals, vol. 40/1, pp. 491-494, 1991

[6] J.A. NELDER, R. MEAD, *A simplex method for function minimization*, Computer Journal, vol. 7, pp. 308-313, 1965

[7] P. PAULY, *Etablissement d'algorithmes de calcul d'erreurs de forme et de position*, Final thesis, University of Liège, 1996

[8] L.P. LEBEDEV, I.I. VOROVICH, G.M.L. GLADWELL, *Functional analysis*, Kluwer Academic Publishers, Dordrecht, Boston, London, 1996

[9] J.F. DEBONGNIE, L. MASSET, *Sur l'évaluation des défauts de forme à partir de mesures tridimensionnelles*, European Journal Mech. Eng., vol. 43, no. 1, pp. 13-29