Geometric tolerances verification: strategy optimization for CMM measurement

Relatore: Ch.mo Prof. Giovanni MORONI
Coordinatore: Ch.mo Prof. Bianca Maria COLOSIMO

Tesi di Dottorato di:
Stefano PETRÒ
Matr. D02250
Contents

1 Introduction ........................................... 7
   1.1 Uncertainty ........................................ 8
   1.2 Specifications in mechanics: Geometric Tolerances and Compliance to Tolerances .......... 10
       1.2.1 Evaluating Geometric Error ................. 14
   1.3 Thesis Aim ....................................... 19
       1.3.1 Thesis structure .............................. 20

2 State of the art ...................................... 22
   2.1 Blind strategies ................................. 22
       2.1.1 Sampling strategies according international standards .......................... 22
   2.2 More on blind sampling strategies .......... 29
   2.3 Adaptive sampling .............................. 34
   2.4 Manufacturing based sampling strategies .... 38
       2.4.1 Sample based sampling strategies ........ 39
       2.4.2 Signature based sampling strategies and models for signature ............... 44

3 Models describing the manufacturing signature .... 50
   3.1 Ordinary Least Squares model ............... 50
       3.1.1 Background and conventions .............. 51
       3.1.2 Single Confidence Intervals in Regression .......... 53
       3.1.3 Simultaneous Regression Intervals ....... 54
       3.1.4 Regression Splines ......................... 60
   3.2 Spatial Error Model .......................... 63
       3.2.1 Neighborhood structures for data on a regular lattice .................... 64
       3.2.2 Spatial Autoregressive Model ............ 65
   3.3 Examples of signature models ............... 66
       3.3.1 SARX(2) model for cylindrical surfaces ...... 66
       3.3.2 SARX(2) model for roundness profiles ... 69
3.3.3 Regression Spline model for cam profiles . . . . . . . . 71

4 On manufacturing based sampling strategies 74
  4.1 Extreme Points Selection and Principal Components Variables 74
    4.1.1 Case study for EPS and PCV . . . . . . . . . . . . . . 74
    4.1.2 Principal Components Variables . . . . . . . . . . . . . 75
    4.1.3 Extreme Points Selection . . . . . . . . . . . . . . . . 77
  4.2 Extended Zone . . . . . . . . . . . . . . . . . . . . . . . . . 80
  4.3 Fitted Deviation-based Extraction . . . . . . . . . . . . . . 81
  4.4 Consideration on these manufacturing signature based strategies 83

5 Optimal signature based sampling strategy 86
  5.1 Definition of the Tolerance Interval Based sampling strategy . 86
    5.1.1 Application of TIB strategy to roundness profiles . . . 88
  5.2 Criterion for choosing the right sample size . . . . . . . . . . 95
    5.2.1 Evaluation of $C_M$ . . . . . . . . . . . . . . . . . . 98
    5.2.2 Evaluation of $C_E$ . . . . . . . . . . . . . . . . . . 98
    5.2.3 Evaluation of $U$ . . . . . . . . . . . . . . . . . . 100

6 Conclusions and future aims 104

Bibliography 107

A Loop Control Algorithm 118
  A.1 State of the art . . . . . . . . . . . . . . . . . . . . . . . . 118
  A.2 A Brute Force Algorithm . . . . . . . . . . . . . . . . . 120
  A.3 Reducing the sample size: convergence problems . . . . 121
  A.4 Loop Control Algorithm . . . . . . . . . . . . . . . . . . . . 124
  A.5 Algorithm comparison . . . . . . . . . . . . . . . . . . . . 124
    A.5.1 Initialization of loop control algorithm . . . . . . . . 125
    A.5.2 Loop control algorithm against Carr-Ferreira algorithm 125
    A.5.3 Loop control algorithm against Voronoi algorithm . . 127
    A.5.4 Loop control algorithm against speed-up algorithm . . 128
    A.5.5 Loop control algorithm against steepest descent algo-
       rithm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 128
  A.6 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . 129
List of Figures

1.1 Zones according to [ISO 14253-1] (two-sided tolerance) . . . . . 12
1.2 Zones according to [ISO 14253-1] (one-sided tolerance, like geo-
metric tolerances). . . . . . . . . . . . . . . . . . . . . . . . . 13
1.3 CMM owned by Politecnico di Milano. . . . . . . . . . . . . . 15
1.4 Sampling points and tolerance zone for a nominally flat surface
(flatness tolerance). . . . . . . . . . . . . . . . . . . . . . . . . 16
1.5 Error components leading to uncertainties in CMM measure-
ment. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
1.6 Sampling a planar hollow surface: in the higher part, usual
practice; below, distribution suggested by the signature. . . . . . . 19
2.1 Classification of sampling strategies . . . . . . . . . . . . . . 23
2.2 Transmission characteristic for longwave-pass filter having cut-
off frequencies \( f_c = 15 \) UPR; 50 UPR; 150 UPR; 500 UPR;
1500 UPR . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
2.3 Flatness sampling strategies proposed in [ISO/TS 12781-2]. . . 27
2.4 Cylindricity sampling strategies proposed in [ISO/TS 12180-2]. 28
2.5 Blind sampling strategies for flatness (256 points in each stra-
tegy) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
2.6 Example of Hammersley sequence based sampling strategy for
flatness (256 points). . . . . . . . . . . . . . . . . . . . . . . . 31
2.7 Example of Halton-Zeremba sequence based sampling strategy
for flatness (256 points). . . . . . . . . . . . . . . . . . . . . 32
2.8 Possible shapes for a straightness profile: above, non convex
shape, below, completely convex shape. . . . . . . . . . . . . . 40
3.1 Intervals on the average vs prediction intervals . . . . . . . . . . 55
3.2 Non simultaneous vs simultaneous confidence intervals for the
regression equation . . . . . . . . . . . . . . . . . . . . . . . . . 57
3.3 Simultaneous confidence intervals for the regression equation
vs tolerance intervals. . . . . . . . . . . . . . . . . . . . . . . . 61
5.8 TIB strategy and tolerance intervals for roundness profile (40 points, \( \sigma = 0.01 \text{ mm} \)). ........................................ 94
5.9 Comparison of average error for uniform, EZ, and TIB strategies (full model). ........................................ 96
5.10 Comparison of error sample variance for uniform, EZ, and TIB strategies (full model). .......................... 97
5.11 Rejected fraction of produced parts. ......................... 101
5.12 Typical contour plot (isolines) for \( C_I \). ................. 101
5.13 Uncertainty behavior as function of the sample size for uniform, EZ, and TIB strategies. ......................... 102
5.14 Optimal sampling strategy for roundness profile in Eq. 5.2. ........................ 103

A.1 Essential subset ........................................ 122
A.2 Degenerate points distribution .......................... 123
A.3 Initialization of loop control algorithm ................ 126
A.4 Loop control algorithm against Carr-Ferreira algorithm .... 126
A.5 Loop control algorithm against Voronoi algorithm ....... 127
A.6 Loop control algorithm against speed-up algorithm ........ 128
A.7 Loop control algorithm against steepest descent algorithm . 129
List of Tables

3.1 The regressor functions $f_i (\theta, \zeta)$ as a function of the index $i$. . 68
3.2 $\mathbf{W}^{(1)}$ for a five points roundness profile. . . . . . . . . . . . . . 69
3.3 Covariance matrix for the parameters in Eq. 3.30. . . . . . . . . 71
Chapter 1

Introduction

Every company has essentially four “instruments” to beat its competitors: price, innovation, service, and quality. Quality in particular refers to the ability of satisfying the customer by selling him a “good” product. Quality is, of course, a quite subjective matter, being related to opinions of the customer on what is effectively “good”.

The search of good quality (without specifying quantitatively what “good quality” means) can be undertaken by trying to create a “quality culture” in the workshop, namely to encourage workers to pursue the objective of a better quality. However, in order to propose instruments able to guarantee quality, a clear definition of quality is required:

We define quality as fitness for use [Mon96].

Classically, “fitness for use” is defined by conformance to standards: a product “fits for use” if every requirement in the product project is satisfied. At present, “fitness for use” concept has moved to a less stringent definition, considering overall satisfaction of the customer, which is not ensured only by conformance to tangible requirements, but also to intangible ones, like a pleasant design, additional features, and so on.

Of course, there is a very easy way of improving the quality of products: checking every product and discarding every defective one ensures perfect final quality (if the check system is perfect, this issue will be addressed §1.1); however, this usually leads to excessive inspection costs for mass production. Therefore, the concept of “process quality” has been coupled to “product quality” concept: a manufacturing process may be considered good if the quality of manufactured parts is good, and it is good over time. The search of good quality over time leads to the introduction of quality systems based on “Statistical Process Control” techniques (e.g. control charts), which try to
ensure the manufacturing process does not deviate from its typical, standard (and possibly efficient) behavior.

Regardless of which kind of quality is being controlled, product quality or process quality, it is always based on measurement. A measurement system is required to state whether a product is conforms or is not conforms to a requirement. Even if this evaluation may be qualitative (i.e. a “specialist” observes the product, directly or by means of some instrument like a projector or a microscope, and states whether the product is conforming or not conforming), in most cases some device will perform a measurement, and on the basis of this measurement result the part is stated to be conforming or not conforming.

The use of measurement instrument leads to other problems. In a “perfect” world, measurement results should perfectly reflect the reality, so that if a sufficient number of measurement are performed any measurand could be known at the desired accuracy level. Unfortunately, measurement results will never be exact – some difference will be present anyway between measurement and real measured characteristic. This leads to quality issues: is a part judged conforming really conforming, or is it not?

1.1 Uncertainty

The presence of measurement errors has lead to the definition of a parameter describing the goodness of a measurement: uncertainty ($u$). Two international standards define uncertainty and its evaluation, the guide to “Expression of the Uncertainty of Measurement in Calibration” [EA-4/02] and the “International Vocabulary of basic and general terms in Metrology” or VIM [ISO/IEC GUIDE 99]. The definition of uncertainty that can be found in these standards is [ISO/IEC GUIDE 99]:

\[
\text{non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used}
\]

where the measurand is the characteristic of an object that is being measured.

There are two kind of errors in measurement, that will therefore influence uncertainty: systematic errors, and random errors. First of all let us define measurement error [ISO/IEC GUIDE 99]:

\[
\text{measured quantity value minus a reference quantity value}
\]
where a reference quantity value is a “quantity value used as a basis for comparison with values of quantities of the same kind”. Systematic error is then defined as [ISO/IEC GUIDE 99]

component of measurement error that in replicate measurements remains constant or varies in a predictable manner

that is, an error that is present in any measurement repetition, and that is (usually) constant, or at least predictable. If this error is linked to the measuring instrument adopted, it is usually called “bias” ($b$), even if this is not completely correct. Random error is slightly different, being defined as [ISO/IEC GUIDE 99]

component of measurement error that in replicate measurements varies in an unpredictable manner

so random error is equal to measurement error minus systematic error. According to the given definition, uncertainty is associated to the dispersion of the random error.

To evaluate uncertainty, two ways are possible: the first one is still based on repeated measurement, suppose $n$ measurement repetitions $t_i$ of a measurand are available, the $u$ is evaluated by the sample standard deviation of the $n$ values for $t_i$:

$$u = \sqrt{\frac{\sum_{i=1}^{n} (t_i - \bar{t})^2}{n-1}}$$ (1.1)

where $\bar{t}$ is the sample average of the $n$ measurement results.

Else, uncertainty may be evaluated by an “expert operator” directly or basing on nominal characteristics of the measuring equipment adopted.

Of course, when evaluating uncertainty both errors have to be considered (even if bias may be corrected, if convenient, see [EA-4/02]). Moreover, uncertainty usually does not have a single reason: several uncertainty sources may be identified, and combined; usually if there are $l$ uncertainty sources, each one giving rise to a $u_i$ uncertainty contribution, the composite uncertainty in a so called “black box” model [ISO/TS 14253-2] (i.e. the output value of a measurement is obtained in the same unit as the input, rather than by measurement of other quantities functionally related to the measurand) may be calculated as

$$u = \sqrt{\sum_{i=1}^{l} u_i^2}$$ (1.2)
The uncertainty defined in Eq. 1.1 and 1.2 is the so called “standard uncertainty”. It is usual practice to state an “extended uncertainty” $U = ku$ ($k$ is called the “coverage factor”). The reason for stating an extended uncertainty is probabilistic. Supposing $t$ is the result of a measurement, it may be stated that the “real value” of the measurand is contained in an interval $(t - U, t + U)$, with the probability of this statement to be true depending on $k$.

1.2 Specifications in mechanics: Geometric Tolerances and Compliance to Tolerances

As in any other production sector, in mechanics specification have to be satisfied to guarantee mechanical systems to work properly. Mechanical specification include, among others, material properties (hardness, tensile strength, and so on), structural integrity, surface finishing, and dimensional/geometric conformance. The latter is very common: nearly any mechanical design includes some dimension which has to be respected, so, as an instance, if a shaft has a nominal diameter of 20 mm, it may be admitted that its real diameter ranges from 20 to 20.01 mm – a tolerance stated on the size of a part. At present, dimensional specification are evolving from “size conformance” to “geometric conformance”, which leads to the definition of “geometric tolerances” [ISO 1101, ASME Y14.5M]:

**Tolerance, Geometric.** The general term applied to the category of tolerances used to control form, profile, orientation, location and runout [ASME Y14.5M].

**Form tolerances** limit the deviations of an individual feature from its ideal geometrical form [ISO 1101].

Practically, a geometric tolerance states how much a real part can differ from its ideal, designed geometry: even if mechanical surfaces are usually defined as planes, cylinders, spheres, etc, and this geometry influences functionality of the part, “perfect” surfaces are not achievable, so a limit has to be stated for the deviations from nominal shape. Differing from size tolerances, for which the tolerance limits are usually two-sided and defined by a “Lower Specification Limit” $LSL$ and an “Upper Specification Limit” $USL$, a geometric tolerance is usually one-sided [ISO 14253-1], so it defines only an $USL$ for the geometric error (a “perfect” part is always considered conforming). Geometric tolerances importance is increasing because of the performances the market requires from mechanical systems: if for simple parts
dimensional tolerancing may suffice to ensure functionality, more complex fits require more stringent definitions of the part.

Of course, the check of a geometric tolerance involves the estimate of the geometric error, and an uncertainty evaluation has to be proposed for this estimate. However, because of the measurement error presence, even if these data are available, it cannot be stated univocally whether part conforms or not, so a criterion to protect the exactness of the conformance statement from measurement error is required.

To define how to decide conformance or non conformance, some series of international standards have been proposed giving guidelines for this statement, like those from “The American Society of Mechanical Engineers”, ASME B89.7.3 [ASME B89.7.3.1, ASME B89.7.3.2, ASME B89.7.3.3], and “International Organization for Standardization”, ISO 14253 [ISO 14253-1, ISO/TS 14253-2, ISO/TS 14253-3]. In particular, [ASME B89.7.3.1] and [ISO 14253-1] propose “guidelines for decision rules” for “considering uncertainty in determining conformance to specifications”. These standards state that a decision rule must meet these conditions [Phi91]:

1. A decision rule must have a well-documented method of determining the location of the acceptance, rejection, and any transition zones (transition zones are optionally defined regions between acceptance or rejection).

2. Each zone of a decision rule must correspond to a documented decision that will be implemented should the result of a measurement lie in that zone. While this is implicit for the acceptance and rejection zones by definition, any transition zones must have their corresponding decision outcomes documented.

3. A decision rule must state the procedure for addressing repeated measurements of the same characteristic on the same workpiece or instrument.

4. A decision rule must state the procedure regarding data rejection, that is, rejection of “outliers”.

A decision rule should then essentially identify three “zones” for the measurement result:

1. **Conformance zone.** If the measurement result \( t \) falls in this area, then the related part is accepted.
2. **Non-Conformance zone.** If the measurement result $t$ falls in this area, then the related part is rejected.

3. **Uncertainty range.** If the measurement result $t$ falls in this area, it is not possible to state whether the part conforms or not conforms to the tolerance.

[ASME B89.7.3.1] and [ISO 14253-1] differ in that the ASME standard gives only guidelines on how to choose these regions, because the selection of a decision rule is considered a business decision, and the flexibility of having a continuum of rules ranging from stringent to relaxed acceptance or rejection is needed in order to satisfy a broad range of industries. ISO standard is more rigid; practically, [ISO 14253-1] gives the following definition for the Conformance Zone:

Specification zone reduced by the expanded uncertainty of measurement, $U$

where the Specification zone is, in the specific case a geometric tolerance, defined by “tolerance limits”:

Specified values of the characteristic giving upper and/or lower bounds of the permissible value

Similarly, the Non-Conformance zone is defined as:
Figure 1.2: Zones according to [ISO 14253-1] (one-sided tolerance, like geometric tolerances).

Zone(s) outside the specification zone extended by the expanded uncertainty of measurement, $U$

and the uncertainty range is

Range(s) close to the specification limit(s) where neither conformance nor non-conformance can be proved taking into account the uncertainty of measurement. [...] The uncertainty range(s) is(are) located around the specification limit (one-sided specification) or specification limits (two-sided specification) and has the width of $2 \times U$

Note that, as already stated, a geometric tolerance is usually one-sided, so only a single non-conformance zone and a single uncertainty range are
defined These zones are graphically identified in Fig. 1.1 and Fig. 1.2.

Mathematically, [ISO 14253-1] standard defines conformance and non-conformance zones, and uncertainty range as follows, for a geometric tolerance:

1. **Conformance Zone**: \((0, USL - U)\)
2. **Non-conformance Zone**: \((USL + U, +\infty)\)
3. **Uncertainty Range**: \((USL - U, USL + U)\)

Acceptance and rejection region have obvious interpretation: a part in the acceptance region is conforming, and a part in the rejection region is non conforming. A rule for managing parts in the transition region has to be stated. [ISO 14253-1] standard suggest that

the uncertainty of measurement always counts against the party who is providing the proof of conformance or non conformance and therefore making the measurement

and then considers two different situations: “supplier proving conformance”, and “customer proving non-conformance”. In the first situation, because conformance has to be proved, it is stated conservatively that parts in the transition region are non conforming, without any further investigation; on the contrary, in the second situation, if a part falls in the transition region it may be considered conforming. The standard specifies that the value of \(U\) to apply is the one of who is performing the test, i.e. the supplier in the first situation, and the customer in the second. This rule aims to pose stringent constraints statements – it is always harder to prove the “alternative hypothesis”, e.g., if conformance has to be proved, then non-conformance is supposed, and to prove conformance a result “far” (in a metric given by the uncertainty) from the tolerance limit is required.

### 1.2.1 Evaluating Geometric Error

Having described criteria on stating conformance to tolerance, let’s introduce how geometric error are measured.

[ISO 1101], when defining the concept of conformance in the specific case of geometric tolerance, states that

A geometrical tolerance applied to a feature defines the tolerance zone within which that feature shall be contained.
and defines the tolerance zone as:

space limited by one or several geometrically perfect lines or surfaces, and characterized by a linear dimension, called a tolerance.

From these definitions, the multidimensional nature of geometric tolerances is apparent: single size measurements are not sufficient to state whether a part conforms to geometric tolerance or not, because a whole surface (or a whole profile) has to be inspected (sampled) to assess it is contained in the tolerance zone.

Classical instrument for evaluating geometric errors include dial gauges, micrometers, and similar instruments; however, these instrument require a specific configuration of the test equipment for each tolerance to be checked. In order to achieve better flexibility, an instrument usually adopted is a “Coordinate Measuring Machine” (CMM, Fig. 1.3), a system which may sample some points on any surface, by mechanical contact (as common) or with non contacting systems (laser trigger or conoscopic holography etc.). Then, geometric error is evaluated: it is possible to fit a “substitute geometry” on the sampling points (e.g. by means of a Least Squares fitting), and then to evaluate the geometric error as maximum deviation of sampled points from
this geometry, or to directly search the tolerance zone, or to directly apply “Minimum Zone”, in which separation of two surfaces (whose shape is coherent with the nominal shape) enveloping the cloud of points is minimized. (see [Ant96] and appendix A for further details). Because only few points usually define the amplitude of the tolerance zone, only these points are really relevant to evaluate the geometric error. Fig. 1.4 illustrates an example of tolerance zone for flatness, and shows a cloud of sampled points, of which only those indicated with asterisks effectively define the geometric error. In this example, MZ has been adopted for fitting the substitute geometry.

In this measurement process, five categories of measurement uncertainty have been identified [Wil01] (Fig 1.5):

- **CMM hardware**, sources of uncertainty related to errors inherent to the design of the machine, its scales and geometry, its probing system, its dynamics, and the environment in which it is placed.
- **Workpiece**, these uncertainties relate to properties of the workpiece and measurement strategy interaction with the workpiece.

- **Sampling strategy**, errors due to inadequate sampling, the interaction of sampling strategy with form error, interactions of sampling with complex forms, uncertainty “magnification” due to uncorrect choice datums [ISO 1101].

- **Algorithms**, algorithm suitability and selection, algorithm interaction with sampling density, and algorithm implementation dominate this error category.

- **Extrinsic factors**, these variables include apparently minor variations in operator–machine interaction (tightening of clamps, failure to wear gloves, etc.) coupled with conditions such as part cleanliness, presence of contaminates, etc.

Extrinsic factors effects may be reduced or even eliminated in obvious ways (e.g. a cleaner environment ensures a lower uncertainty), and it is usually difficult to modify the hardware influence, requiring consistent investments in instrumentation; the workpiece usually cannot be modified to improve its easiness of measurement. Finally, at present fitting algorithms are actually efficient enough to ensure their influence is unimportant. Therefore, the operator may mainly influence the sampling strategy to reach an optimal uncertainty.
Traditional approaches for planning the measurement strategy for geometric tolerances obtained on a manufactured component are independent from the specific process adopted for manufacturing that feature. In fact, a tolerance is most of the times computed by means of a generic measurement strategy, where the number and the position of the measured points are empirically determined (usually they are uniformly or randomly distributed on the ideal feature that has to be inspected). Of course, as the number of measured points increases, inspection costs increase as well, while uncertainty in estimated tolerance decreases because of the completer description of the surface: hence, sampling strategy must be designed in order to balance these two factors.

A closer analysis of the specific process involved in manufacturing the feature can add cleverness in designing measuring strategy, through a proper selection of the number and the position of the measuring points. This added cleverness can result in a reduction of sample size for a fixed level of uncertainty. Cleverness coming from the knowledge of the manufacturing process is related to the definition of the empirical model representing the “signature” left by the process on the manufactured feature. This manufacturing signature is the systematic pattern that characterizes all the features manufactured with that process. In fact, measured surfaces (or profiles) often present a systematic pattern and a superimposed random noise [Dow97]: the first one is mainly due to the process used in specific operation conditions, while the second is due to unpredictable factors which are usually known as “natural variability”. The systematic pattern constitutes what is called “the signature” of the process [Whi94].

The manufacturing signature, which may be considered part of the “workpiece” uncertainty sources, effectively interacts with the sampling strategy. As it has already been stated, only the most distant points from the fitted ideal surface will define the estimated geometric error, and because of systematic geometric error areas which deviate the most from the ideal geometry are quite repeatable throughout the production period, an higher concentration of sampling points in these areas is desirable.

To better understand why signature identification can improve the sampling strategy, consider as an example a planar surface that, due to the specific process used (e.g. face working on a lathe) is systematically hollow (Fig. 1.6 reports the systematic pattern once random noise is filtered out from data). A traditional, “blind” sampling strategy will distribute sampling points uniformly on the whole surface, while cleverness coming from the signature identification will suggest a different distribution of points because it could reasonably be supposed that points defining the tolerance error will be
found mainly on the edges and in the middle of the surface.

1.3 Thesis Aim

In order to control inspection related cost, this work mainly aims to propose a method which will be able to generate an optimal measuring strategy (points location and sample size) for the evaluation of geometric tolerances, and in particular form tolerances. The strategy will be optimal under an economic point of view. Because the principal component of form error is often the manufacturing signature, the manufacturing signature itself will trave the choice of the sampling points pattern.

Of course, these results will be supported by an introductory step defining actual state of the art for sampling strategies and signature models. Particular attention will be given to manufacturing based sampling strategy, which belong to the same sampling strategies family to which belongs the strategy proposed in this thesis.

Finally, the strategy definition methodology will be proposed. Because it is supposed that the signature has a statistical nature, the proposed stra-
strategy will be based on the statistical concept of “Regression Tolerance Intervals” [Lie63]. Tolerance Intervals allows only to define the location of sampling points, and not the number of sampling points to be taken; therefore, an economic criterion will be added which will able to establish the sample size. So the methodology is a two step methodology: in the first one, given the sample size, the strategy minimizing uncertainty is generated; then, trade off between uncertainty and measurement cost is found.

1.3.1 Thesis structure

This work is structured as follows:

- In chapter 2 a state of the art on sampling strategies and signature models will be proposed. This state of the art will start from present international standards on sampling strategies for geometric error estimate, and then introduce few works on sampling without informations on signature. Then, present works on signature and signature based inspection will be proposed.

- Chapter 3 will describe basics on manufacturing signature modeling; signature proposed model include:
  - Standard regression models, with the particular case of regression splines
  - Spatial Error Model

Few manufacturing signature examples will be proposed.

- Chapter 4 describes some methodology able to generate a manufacturing based strategy. Even if these model are valid and applicable, this work effectively aims to pass pass their defects and limitations. Proposed methodologies include:
  - Principal Components Variates
  - Extreme points selection
  - Extended Zone
  - Fitted Deviation-based Extraction

- Finally, chapter 5 will propose the Tolerance Intervals Based criterion supported by the economic criterion for the choice of the sample size. The proposed methodology will be applied to the case study of turned
roundness profiles, in order to compare it to uniform and “extended zone” strategies (see §2.1 and §4.2), showing its superior performances.

- Constituting a parallel argument, appendix A describes a generic fitting algorithm for estimating form error from a given cloud of points; an implementation of this “Loop Control Algorithm” for roundness has been adopted in this work.
Chapter 2

State of the art

Before introducing the methodology to generate the sampling strategy that is the aim of his work, methodologies proposed in literature will be introduced.

Sampling strategy may be categorized (Fig. 2.1). First of all, they can be split in “blind strategies”, “adaptive strategies”, and “manufacturing based strategies”.

2.1 Blind strategies

Lots of papers discussing efficiency of sampling strategies have been proposed. Most of the sampling strategies are “blind” in the sense that the sampling points pattern is decided once and forever, usually on the basis of some empirical criterion or expert judgment. Blind sampling strategies include those proposed in ISO standards.

2.1.1 Sampling strategies according international standards

The International Organization for Standardization (ISO), in the end of ’80, has undertaken a general project of renewal of standards on tolerancing. The new series of standards, defined “Geometrical Product Specifications” (GPS) has the aim of systematize the tolerancing chain from the design step in which tolerances are stated, till the verification step in which size and geometric errors are evaluated together with their measurement uncertainty. GPS series includes therefore standards on expressing tolerances, measuring equipment characteristics and verification, measurement procedures, and uncertainty evaluation for sizes, geometric tolerances and texture (roughness). However, a large part of these standards are still missing.

In particular, four series of standard deal with the definition and verification of form tolerances (cylindricity [ISO/TS 12180-1, ISO/TS 12180-2],

ordered set of specification operation(s)

where a specification operation is an

operation formulated using only mathematical or geometrical expressions or algorithms, or all these

Standards [ISO/TS 12180-2], [ISO/TS 12181-2], [ISO/TS 12780-2], and [ISO/TS 12781-2] give, in particular, “Complete specification operators” for, respectively, cylindricity, roundness, straightness, and flatness. A “Complete specification operator” is defined as

specification operator based on an ordered and full set of completely defined specification operation(s)

[ISO/TS 17450-2] states that “a complete specification operator is unambiguous and therefore has no specification uncertainty”, where the specification uncertainty is the “uncertainty inherent in an actual specification operator
when applied to a real feature” and the actual specification operator is the “specification operator derived from the actual specification given in the actual technical product documentation”.

A specification operation is essentially a set of rules which have to be respected to estimate, as an instance, a geometric error, and then to affirm this estimate has been performed according to a given international standard. Applying a Complete specification operator as defined in standards should ensure that no uncertainty due to measurement strategy should originate.

The various sampling strategies proposed in standards are now described.

**Roundness and Straightness**

Complete Specification Operators proposed for roundness, as described in [ISO/TS 12181-2], and for straightness (described in [ISO/TS 12780-2]) are very similar, and therefore only roundness strategy will be described; extending concepts from roundness to straightness is straightforward.

[ISO/TS 12780-2] states that

The complete specification operator defines the transmission band for the roundness profile, together with an appropriate stylus tip geometry.

These requirements imply that, if points are sampled in order to evaluate a roundness error, as in the case of CMM measurement, a filter has to be applied to sampling points before applying a fitting algorithm to them. This filter aims to reject noise components due to roughness and measurement error. At present, the standard require a “phase correct” filter to be adopted. ISO has proposed several standards describing several kinds of filters, including spline, wavelet and morphological filters, but [ISO/TS 12780-2] considers only the application of the gaussian filter described in [ISO 11562]. This filter has the transmission band illustrated in Fig. 2.2 and defined by

\[
\frac{a_1}{a_0} = e^{\pi(\alpha f_c)^2}
\]

\[
\alpha = \sqrt{\frac{\ln(2)}{\pi}}
\]

(2.1)

where \(a_0\) is amplitude of sine wave undulation before filtering, \(a_2\) is the amplitude of this sine wave after filtering, \(f_c\) is the cut-off frequency (in UPR) of the longwave-pass filter, and \(f\) is the frequency of the sine wave (in UPR). In this formulation, it is implicit that the roundness error may be described
as sum of sine wave, and that only the longwave part of these waves constitutes the form (roundness) error. In this optic, it is reasonable to express the cut-off frequency of the filter in “Undulation Per Revolution” (UPR), namely how many undulation the sine wave shows in a full revolution. When dealing with straightness, instead of a cut-off frequency $f_c$ is defined a cut-off length $\lambda_c$, usually expressed in millimeters.

Basing on the chosen cut-off frequency, the standard suggests to sample at least seven points for each cut-off, that is, the suggested sample size is $n \geq 7 f_c$; sampling points will be evenly spaced on the profile. This sample size ensures that only small amount of aliasing [Bra00, ISO/TS 12181-2, ISO/TS 12780-2] is present in the filtered profile due to the undersampling of high frequency (noise) components of the considered roundness profile.

Other requirements in the standard include the radius of the stylus tip (which is directly proportional to the diameter of the measured profile), the tip geometry (spherical), and the probing force (0 N).

Two observation have to be added: the standard does not propose any value for the cut-off frequency, so the sample size is effectively an operator’s choice, and the criterion “seven points per cut-off” leads to huge sample sizes - a 50 UPR filter, which is quite common, leads to an at least 350 sampling
points strategy, which is hard to adopt in a productive environment with a CMM with a touch trigger probe. This standard is effectively intended to be used with roundness-testing machines or straightness measurement instruments, capable of scanning large amount of points in short time; with a shop-floor device they are inapplicable.

Cylindricity and flatness

As for roundness and straightness, [ISO/TS 12180-2] and [ISO/TS 12781-2], describing respectively complete operators for Cylindricity and flatness, are very similar. Again, it is stated that “the complete specification operator defines the transmission band for the roundness profile, together with an appropriate stylus tip geometry”. However, no indication is given on the kind of filter which has to be adopted for flatness and cylindricity, nor the cut-off frequency is specified. Again, a least seven points per cut-off are suggested.

Like specification for roundness and straightness, the tip geometry for measuring flatness and cylindricity is required to be spherical, and the probing force should be 0 N. Because in estimation of flatness and cylindricity surfaces, not profiles like in straightness and roundness estimation, are involved, a singular sampled profile is not sufficient to completely describe the geometric feature. Standards therefore suggest how to place profiles on the surface, proposing rectangular grid, polar grid, triangular grid, Union Jack, parallel profile, and points extraction sampling strategies for flatness, and bird-cage, roundness profile, generatrix, and, again, points extraction sampling strategies for cylindricity. These sampling strategies are illustrated in Fig. 2.3 and 2.4. Each of these strategies is suitable for different situations: standards state that rectangular and polar extraction for flatness and bird-cage extraction for cylindricity are usually to be preferred; the other strategies are to be chosen if particular interest is posed in some behavior of the surface, e.g. if lobing is supposed to be relevant in a cylindrical surface, then roundness profile extraction should be adopted. In particular, for both flatness and cylindricity the only strategy involving single points sampling instead of profiles sampling is points extraction strategy: standards state that, for a points extraction strategy,

The density of points is typically lower than with the other extraction strategies listed above. This restricts the ability to assess the harmonic content of a feature of *flatness/cylindricity*. The lower number of points also presents problems when filtering. It is for this reason that the points extraction strategy is
Figure 2.3: Flatness sampling strategies proposed in [ISO/TS 12781-2].
Figure 2.4: Cylindricity sampling strategies proposed in [ISO/TS 12180-2].
not recommended unless only approximate estimates of the flatness/cylindricity parameters are required.

However, this is the only extraction strategy really applicable when using a touch trigger CMM, like in shop-floor environments.

Moreover, both [ISO/TS 12180-2] and [ISO/TS 12781-2] state, in a note, that

In practice it is unrealistic to achieve comprehensive coverage of the flatness/cylindricity feature given by the theoretical minimum density of points within an acceptable time span using current technology. Therefore more limited extraction strategies are employed that give specific rather than general information concerning the deviations from flat/cylindrical form.

This is one of the reason leading to this thesis work. Because, with actual technologies, it is impossible to efficiently measure some form tolerance, then strategies to reduce measurement uncertainty for geometric error is to be researched.

2.2 More on blind sampling strategies

Classical blind sampling strategies include random points distribution, stratified distribution, and uniform distribution [Tho92] (see Fig 2.5); the latter, involving sampling points distributed in a regular pattern on the surface, is probably the most adopted, because it is very easy to implement on most of the control softwares of CMMs. However, this sampling strategy may give some problem if the surface has some harmonic content (some undulation is present on the surface), and the frequency of the uniform sampling is identical or submultiple of the frequency of the harmonic content [Dow97]: this is a typical example of bad interaction between manufacturing signature a sampling strategy. The random strategy, in which points are, obviously, randomly distributed on the surface, usually does not present the problem of conflicting with the harmonic content of the surface; however, a fully random strategy may leave some zones of the surface not covered by any sampling points. Defects of uniform and random sampling strategies are solved by the stratified sampling strategy, in which the surface is split in some (usually equal) areas, and then a portion of the sampling points are randomly distributed in each area. This ensures a distribution more uniform than a random distribution, but still free of periodic behavior like a uniform distribution.
Among typical possible blind sampling strategies, Hammersley sequence [Ham64] based sampling strategy has proved to be efficient. As an instance, for square flat plan, the Hammersley sequence sampling strategy is defined as follows: if $n$ sampling points have to be taken, then define

$$P_i = \frac{1}{n}$$

$$Q_i = \sum_{j=0}^{k-1} b_{ij} 2^{-j-1}$$

(2.2)

where $P_i$ and $Q_i$ are normalized cartesian coordinates (namely they are in the interval $[0, 1]$, to obtain the actual coordinate on the surface to be inspected.
Hammersley sampling strategy

Figure 2.6: Example of Hammersley sequence based sampling strategy for flatness (256 points).

$P_i$ and $Q_i$ have to be multiplied by the sizes of the surface to be inspected) of the $i^{th}$ point, $b_i$ is the binary representation of $i$, and $b_{ij}$ is the $j^{th}$ bit of $b_i$ (so $j \in \{0, 1\}$), and $k = \lceil \log_2 n \rceil$, where $\lceil x \rceil$ is the smallest integer greater than or equal to $x$.

As apparent from Eq. 2.2, there is nothing random in an Hammersley strategy. However, as Fig. 2.6 shows, points are spread throughout the whole surface without any apparent pattern, and this should avoid any interaction with an eventual harmonic content in the surface, or any other systematic behavior. Lee et al. [Lee97] have shown that Hammersley sequence based sampling strategies outperforms both uniform and random sampling strategies for plan, cylinder, cone, and dome (sphere). These results have been obtained by means of a simulation study: one hundred surfaces per geometry are simulated, deviation from the nominal geometry being constituted by a polynomial part (which simulates the signature) and a random part (simulating roughness). Authors claim that

the Hammersley sequence has a nearly quadratic reduction
the number of points needed by uniform sampling for the same
level of accuracy as measured by discrepancy.

Moreover, it is suggested to integrate Hammersley sequence in a stratified
A strategy similar to Hammersley one is based on the Halton-Zeremba sequence [Woo95]. The Halton-Zeremba strategy is defined as follows:

\[
P_i = \frac{1}{n} = \sum_{j=0}^{k-1} b_{ij} 2^{-k-j}
\]
\[
Q_i = \sum_{j=0}^{k-1} b'_ij 2^{-j-1}
\]

where \(b'_{ij} = 1 - b_{ij}\) for \(j\) odd, else \(b'_{ij} = b_{ij}\). Halton-Zeremba sequence is defined only if \(n\) is a power of 2. Again, as Fig. 2.7 shows, the strategy is nor random, nor presents any systematic behavior. Kim and Raman [Kim00] claim the Halton-Zeremba strategy not only outperforms uniform and random strategies, but Hammersley strategy for flatness, too. This assertion is confirmed by an experimental study conducted on thirty plates.

The structure of the CMM influences the effectiveness of a measurement, too: different kind of machines show different performances. As an instance Chan et al. [Cha96] considers the verification of size and center location for a circular geometry. They consider only three points sampling strategies,
and these strategies differer because of the different spacing among points: taking as “standard” strategy the one in which sampling points are evenly distributed on the profile, the authors consider strategies with sampling points more an more concentrated in a particular zone of the profile. As expected, these strategies perform worse than the “standard” strategy. However, some few considerations on probe lobing have to be added: if adopting a touch trigger probe for the CMM, the authors show that the location of the three points influences the measurement accuracy, i.e. not only reciprocal position of sampling points is relevant, but position of points with respect to the CMM mechanics are relevant, too (this is an example of the so called “probe lobing”). Moreover, they show that, sampling a ring gauge characterized by a form error lower than 0.1 µm (having evaluated this form error by means of a roundness measurement machine), the estimated form error may reach 4 µm: this is due to the measurement error proper of the sampling of a single point. Anyway, in this work it will be supposed that the effect of these kind of error is not very relevant when dealing with standard mechanical part, i.e. manufacturing signature errors are at least one order of magnitude grater than the typical accuracy of the considered CMM. When dealing with ring gauges, reference plans, and similar, effect of probing error [ISO 10360-2] has to be considered, of course.

Finally, two works dealing with the choice of the right sample size are presented. In particular, at technique to estimate the expected evaluation error (essentially systematic measurement error) for the straightness error evaluation is proposed. Namboothiri and Shunmugam propose to state an upper bound for this error, and then their methodology is able to choose the right sample size. Some objections may be proposed: the criterion considers only systematic errors, but random error in measurements are at least as important as them, like usual uncertainty evaluations shows; analyzed sampling strategies are limited to random sampling; and an arbitrary choice of the error level may be suitable when calibrating a single artifact, but the choice of the sample size for quality check of mass production should be based on economic considerations. Anyway, an interesting consideration found in this article is

Furthermore, this study clearly underlines the importance of the sampling pattern to be followed on the surface during form error measurement. The guiding principle should be to catch that point that has got maximum error. If we get this point at the initial stages of measurement then further prolonging the measurement process is not necessary. Hence the measurement
time can be considerably reduced (please note this leads to reduced measurement costs, too).

This consideration guides toward adaptive sampling, which will be discussed in the following, and manufacturing signature based sampling, which is the main subject of this work.

Finally, two articles have been proposed by Lin and Lin [Lin01a, Lin01b] dealing with the use of “grey theory” [Den82] (which is typically applied in the control system sector) for evaluating the right sample size. This is a very different approach with respect to [Nam99], because it is someway “dynamic”. Instead of choosing a fixed sample size, the sample size is left free to change depending on the predicted value of the form error and measurement accuracy in the next inspected part. Errors prediction is performed by means of a grey theory model, which could be considered as an evolution of time series [Shu88] in which slope is considered too. The sampling strategy is supposed to be uniform, in these works, but the really interesting subject is the attention posed in the possibility the production process modifies its typical behavior, thus leading to the necessity of recalibrating not only the production process itself, but the measurement system, too. This is particularly true if a signature based sampling strategy is chosen because a modification in the process usually leads to a modification in the signature, thus making the signature based sampling strategy inefficient, and perhaps damaging, because it will maybe tend to sample points in areas of the features which are not the most deviating from the nominal geometry.

### 2.3 Adaptive sampling

Adaptive sampling [Tho92] is a quite different kind of sampling strategy with respect to those which have been described till now. While in a blind strategy, and even in a signature based sampling strategy, sample size and sampling points locations are chosen before starting sampling, in an adaptive sampling strategy they may change during the sampling, in order to achieve the required measurement accuracy. In fact, an adaptive sampling strategy is a multi-step sampling strategy in which information in the precedent measuring step is interpreted and basing on it next sampling points are chosen (both in terms of sample size and location). A criterion, usually based on required accuracy for some measurement, is adopted in order to stop sampling points collection. It could be considered an “extreme” signature based sampling strategy: here, the signature identification and the strategy definition is performed for each inspected part.
Being quite diffused as a sampling technique in several fields, a discussion on the use of adaptive sampling for coordinate metrology, for the special case of profile measurement, has been proposed by Edgeworth and Wilhelm [Edg99]. Their technique is aimed to identify and measure, in the successive steps of the algorithm, those areas of the profile presenting the maximum deviations from the ideal geometry. The algorithm starts with an initial sampling containing enough points “so that it is possible to fit the measurement data to the nominal profile”. Then, an associated integral feature [ISO 14660-1] (namely the ideal geometry which best fits sampling points according to some criterion, a “substitute geometry”) is fitted on points and both actual deviations from nominal geometry $\varepsilon$ and normal direction of the error profile $\varepsilon'$ from the surface are calculated.

For each couple of points, each one characterized by its actual value of $\varepsilon$ and $\varepsilon'$, the surface error is fitted by a third degree polynomial:

$$\varepsilon(h) = a_0 + a_1 h + a_2 h^2 + a_3 h^3 \quad h \in [0, 1]$$  \hspace{1cm} (2.4)

where $h$ is a parameter indicating the relative position of the fitting point and terms $a_i$ are fitting parameters for each couple of points. The resulting model is a cubic spline, thus characterized by $C^1$ continuity. Two metrologically relevant properties of this model are outlined by Edgeworth and Wilhelm: “the interpolation method described will not produce an error surface with more waves than the true surface”, and “the interpolation model will also converge with the true error surface as sample size increases”.

On the basis of this fitting, some more points have to be sampled. Edgeworth and Wilhelm propose three general rules:

1. Sample at the extreme location(s) on the interpolating curve (namely where the fitted curve deviates the most from the nominal geometry) if the error curve exceeds the bounding box formed by the sample points plus an offset value.

2. If the interpolation curve does not violate the sampling boundaries described above, but the distance between sample points is greater than the maximum spacing between sample points, probe at the midpoint between the pair of samples.

3. If the spacing between sample points is less than or equal to the minimum spacing between sample points parameter, do not sample further, regardless of the previous two rules.
On the subject of the error curve (how and how much the fitted spline lacks in describing the real error curve), Edgeworth and Wilhelm propose the following maximization of the fitting uncertainty (result is expressed in [mm]):

\[
 u = \sqrt{0.5P_U^2 + 0.03125d^2N_U^2} 
\]  

(2.5)

where \( P_U \) is the standard probing uncertainty, expressed in [mm], \( N_U \) is the standard uncertainty on the measurement of surface normal (expressed in [arc min]), and \( d \) is the distance between sampling points (sampling rate). Combining this evaluation of measurement error with criteria above, it is possible to construct an adaptive sampling algorithm.

Some observations are to be proposed on this algorithm. The authors state that

The required sample size is proportional to part quality. The most accurate parts will require the least sample points; whereas, lower quality parts will require a greater number of total samples.

Moreover, the methodology implicitly supposes the surface is \( C^1 \) continuous, because cubic splines adopted to fit the model are \( C^1 \) continuous, but this is not true in real surfaces, which are characterized by the presence of roughness, which is effectively constituted by micro scale discontinuities. Another difficulty is the lack of automatic measurement of surface normals in several CMM systems; it is suggested to numerically estimate normals in this situation. Finally, “process planning errors that result in discontinuities, such as burrs or gouges, would be difficult to detect consistently using the search algorithms”. However, the major difficulty for this algorithm, and in general for any adaptive sampling algorithm, is that

for general application, the CMM must have the ability to find a collision-free path from its current position to the new targets

and this is quite seldom found in actual CMMs software environment.

More recently, Badar et al. [Bad03], proposed a different approach to adaptive sampling in metrology, in particular for the evaluation of form error. Because the form error estimation, if performed by a MZ criterion, requires only points in the areas which deviate the most from the nominal geometry to be sampled, Badar et al. treat the problem of choosing the next sampling point as a numerical optimization problem, and in particular they adopt a pattern search algorithm. The advantage of pattern search with respect to classical optimization methods is that
pattern search methods are direct search methods that neither compute nor explicitly approximate derivatives of objective function $f$, instead they require only values of $f$ to proceed.

This is may be very useful in metrology where, because of inherent limitations of CMM hardware and practical considerations on the presence of sharp discontinuities in the inspected surfaces due to roughness, it is often difficult to obtain reliable evaluations of “derivatives of the objective function”, which are not required by a pattern search algorithm. The authors examine several pattern search algorithm, including Coordinate Search, Hooke-Jeeves Pattern Search, Tabu Search, and Hybrid Search, and conclude that Tabu Search and Hybrid Search are most suitable for metrology. As an instance, Tabu Search is introduced.

Tabu search [Glo97] is a meta-heuristic intelligent problem solving technique, which guides a local heuristic search procedure to explore the solution space beyond local optimality. In doing so, the method incorporates adaptive memory and responsive exploration. The memory is structured as short-term, intermediate-term, and long-term. The function is evaluated at all neighborhood points with respect to starting point $(x_0, y_0)$. For a step length of $\Delta s$, four neighborhood points will be $(x_0 + \Delta s, y_0)$, $(x_0, y_0 + \Delta s)$, $(x_0 - \Delta s, y_0)$, and $(x_0, y_0 - \Delta s)$, if diagonal neighbors are not considered. A move is made to the point with the “best” $f$ value. If all the neighbors result in worse $f$ value, then the least bad is selected. For the next iteration, point $(x_0, y_0)$ becomes tabu (forbidden). A move is defined to be tabu if any of its attributes is tabu-active for certain tenure. Tabu tenure/size determines the length of the tabu list. Tabu list is a short-term memory to store the coordinates of the most recent moves. This prevents the reversal of the search to any point in the tabu list. If needed, aspiration criteria can be applied to override tabu status of a move. The search continues based on the tabu size, bad moves allowed, and maximum iterations allowed. In many applications, short-term memory may itself result in high quality solutions; but in general to make Tabu search significantly stronger, intermediate and long-term memory are incorporated to intensify and diversify the search.

Applying this algorithm to the evaluation of flatness, the authors show that sampling just 18 points they can reach an accuracy of 87.5% with respect to a sample size of 121 points.

Some limitations in this kind of algorithm may be found in the need of arbitrary parameters like the value of $\Delta s$ (lower values of $\Delta s$ lead to a greater sample size, but to a better accuracy, too), maximum number of “bad moves”, and so on. Moreover, $\Delta s$ influences the maximum resolution
of the sampling, namely the maximum sample size, and, again, like [Edg99],
the CMM control system has to be able to reach the next sampling points
without colliding with the measured surface.

Later, Badar et al. [Bad05] proposed another paper in which they prosed
an approach that could be considered partially adaptive and partially sig-
nature based: the choice of the initial sample set was signature based, then
the inspection proceeded by means of Tabu search. However, signature was
nor formally described, nor a formal algorithm was proposed for choosing
signature based sampling points. In Badar et al. [Bad05] approach an expert
operator simply has to choose the initial set looking at the typical manufact-
uring signature.

Another approach which can considered to be mixed - adaptive plus sig-
nature - has been proposed by Rossi [Ros01] for roundness. In the first of two
steps, Rossi’s approach tries to identify the presence of lobing on a roundness
profile. Then, in the second and last step, if lobing is present, next sampling
points will be placed in peaks and valleys of this lobed geometry; else, a
standard uniform sampling is adopted. Statistical criteria are proposed for
choosing whether the lobing is significant or not. The full methodology is
quite complicated, so the reader is invited to refer to the article for full ex-
planation.

Rossi’s approach is someway opposite with respect to Badar et al. [Bad05]
approach: in the latter, the signature is supposed to exist in the measured
surface, then an initial sampling strategy is proposed basing on it, and finally
sampling proceeds in an adaptive way; in Rossi’s approach initially signature
presence is uncertain, then, basing on an identification procedure, sampling
strategy is chosen according to the ascertained signature (which can be con-
stituted by lobes or pure errors): only after recognizing the signature the
sampling strategy is fully determined.

Rossi’s approach results have been proved to be quite good, in particular
when lobing is present, leading to a reduction in sample size to about one
third of the sample size of the uniform strategy characterized by the same
accuracy.

2.4 Manufacturing based sampling strategies

Manufacturing based sampling strategies are sampling strategies in which the
presence of the signature is exploited in order to choose the best sampling
points pattern.

Again, manufacturing based sampling strategies can be split into two
main subcategories: sampling strategies based on the signature, and sample based strategies.

2.4.1 Sample based sampling strategies

Let’s start from the latter. Sample based sampling strategies assume a signature is present in the measured surface, but never require a model for it. In order to choose the sampling points pattern, an algorithm is directly applied to sampled points in some dense samples coming from a specific manufacturing process, and therefore characterized by a specific signature.

Sample based sampling strategies may, again, be split in two subclasses: reconstructive and non reconstructive strategies.

Sample based sampling strategies without reconstruction

The following papers discuss “sample based sampling strategies without reconstruction”.

A “sample based” methodology is the so called “Harmonic Fitting Method”. In this approach based on a DFT representation of the signature (in a DFT model components of the profile/surface are sine waves or undulated surfaces – the chosen model is a DFT model regardless of the real behavior of the signature, so this is not a “signature based” method), Capello and Semeraro [Cap01a] propose the use of signature in an economic model [Cap01b] aimed at choosing the number of sampling points. The objective function is selected in order to balance inspection costs and costs related to possible errors in quality inspection, due to uncertainty in tolerance estimate. First of all, they develop analytical relations linking DFT coefficients and least squares substitute geometry for cone, cylinder, straight line, plan, and circle. For these features, DFT representation allows to calculate average radius, inclination of the cylinder axis, inclination of the normal to the plane, and so on. Then authors derive the empirical statistical distribution of the estimate errors as a function of the number of sampled points, while considering the effect of aliasing on DFT components which define the substitute geometry; aliasing impact is evaluated basing on the signature. The optimal number of points that has to be measured is thus computed considering the trade off between error computed in estimating tolerance of interest vs. inspection costs. The main limitation of this approach is that, being based on DFT, a uniformly spaced distribution of points is required (Harmonic Fitting Method aims mainly to the definition of the right sample size), and that the method is suitable only for location and size tolerances.
To evaluate both the sampling points locations and the sample size for flatness, a method has been proposed by Raghunandan and Rao [Rag07]. The methodology to choose the right sampling points pattern is based on the assumption that, in a minimum zone flatness evaluation, points defining the tolerance zone are always located on the convex hull of the points cloud [Sam99]. Therefore, having densely sampled a single part, “exact” (based on the dense sample) MZ tolerance $t$ and convex hull are calculated for the part. Then, a subset of the sampling points belonging to the convex hull is randomly chosen, and MZ tolerance $\hat{t}$ for this sample is evaluated. If $\hat{t} > ct$ ($c < 1$ expresses a required level of accuracy for the flatness error evaluation), then the sampling strategy is defined, else more points from the convex hull are randomly added until the required level of accuracy is achieved. The proposed strategy is applied to two more parts in order to confirm the effectiveness of the strategy itself. The methodology is signature based because it is the presence of the signature that ensures the convex hulls of inspected surfaces are similar part to part. An objection that can

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.8.png}
\caption{Possible shapes for a straightness profile: above, non convex shape, below, completely convex shape.}
\end{figure}
be proposed to this methodology is that the sampling strategy is defined on
the observation of a single part, so it may tend to be very part specific, i.e.
it is “perfect” for the observed part, but not for any other part. A more
serious defect may be suggested for the choice of sampling points on the con-
 vex hull, if the signature shape is convex itself. An example for two straight
lines is proposed in Fig 2.8. Both shapes presents critical areas for sampling
in the middle and at the edges. The shape above is not convex: there
fore, the Raghunandan and Rao criterion should place points in the right zones.
Instead, the shape below is convex: its convex hull includes every point,
and therefore Raghunandan and Rao criterion will spread sampling points
randomly on the whole surface. Therefore, this sampling strategy planning
methodology at least lacks generality. However, its main advantage is that,
as the authors state,

In the absence of a manufacturing process model that can
give information regarding the nature of the error on the surface
and consequently the location of the points, arriving at a smaller
sample size that would yield similar results is difficult

and this approach does not, as already stated, require any signature model.

Colosimo et al. [Col07b] adapt a multivariate statistic technique, “Prin-
cipal Component Variables” (PCV), in order to define the sampling strategy
for geometrical tolerances. In the same paper another approach is proposed,
based on using the signature to identify the points that mainly influence the
geometric tolerance ("Extreme Points Selection", EPS). Authors shows that,
with reference to the case of face-worked plans, PCV ensures more repeatable
estimates, while EPS results are less biased. These criteria will be further
discussed in §4.1.2 and §4.1.3

Sample based sampling strategies with reconstruction

Of course, when a sample is taken from a surface, the sample itself will
not completely describe the surface, but the surface will be someway ap-
proximated by it. In particular, it is usually supposed that nothing may
be assessed on the behavior of the surface between sampling points, namely
the surface is known (with a certain level of uncertainty) only in sampled
locations.

Reconstructive geometric error estimation tries to reduce uncertainty as-
sessing something on the behavior of the surface in points which have not
been sampled by “reconstructing” them, i.e. fitting (if it is allowed that
sampled points do not belong to the reconstructed surface) or interpolating
(if sampling points always belong to the reconstructed surface) the points cloud. Of course, the level of uncertainty of reconstructed points will be higher than the uncertainty of sampled points, but this approach may allow to reconstruct those peaks that define the tolerance zone for observed part (even if they have not been sampled), thus reducing the overall geometric error estimation uncertainty.

In a sample based reconstruction the technique for reconstructing the surface is chosen \textit{a priori}, without considering the real behavior of the surface. Therefore, general models will have to be adopted, like Kriging, Shannon reconstruction, Spline fitting, Kernel regression, Loess, and so on. However, the idea of reconstructing implicitly supposes the presence of a signature, i.e. a behavior of the surface – the form error is considered not completely random.

In literature, two works may be found proposing technique and experiences on “\textit{a priori}” reconstruction, both proposed by Yang and Jackman. In the first one, Yang and Jackman [Yan00] propose the reconstruction to be performed by means of Kriging. Kriging is a quite diffused technique in the field of geology, social and environmental sciences, meteorology, and in general of spatial statistics [Cre93], namely that part of statistic studying phenomena which are characterized by a spatial evolution. Of course, a manufactured surface may be considered as an instance of a spatial phenomena, and then treated by Kriging. In a Kriging reconstruction there is a first step in which the surface is characterized by modeling the point to point statistical correlation as a function of the point to point (signed) distance; then, expected behavior of the surface is predicted, supposing the surface follows a “Gaussian Process” model, by means of a conditioned expected value [Joh02]. Mathematical basics of Kriging are really complex, so the reader is addressed to manuals like [Cre93] for further explanation on it; here it will be outlined only that Yang and Jackman choose “universal Kriging” as preferable technique for manufactured surface reconstruction, because universal Kriging may easily manage problem of misalignment of the measured part in the coordinate system, and may be adopted regardless of the sampling points location, so it may be easily coupled to a random or Hammersley sequence based sampling.

Fitting by Kriging poses several problems: it is not so easy to model the surface correlation structure, in fact this structure is effectively linked to the kind of surface inspected, namely its signature. Therefore, to correctly fit the Kriging model a signature study should be performed, thus making this technique no more \textit{a priori}, as intended by the authors. Moreover, the estimate of surface correlation parameters require several sampling points:
a technique which has been originally adopted to reduce the sample size
tends effectively to increase the sample size. Despite these limits, Kriging
reconstruction has a very interesting property:

Kriging provides a standard deviation map for the artificial
surface. How to utilize this error information to characterize the
mean and standard deviation of form error estimation for a given
number of uniform sampling points on general surfaces requires
further study.

The possibility of deriving informations regarding the measurement uncer-
tainty from the Kriging model is very interesting.

In order to overcome Kriging limits, Yang and Jackman, in a more re-
cent paper [Yan02], suggested to adopt “Shannon sampling”. Here “Shan-
non reconstruction” for a monodimensional features (profiles) is introduced;
extension to multidimensional features (surfaces) may be found in [Zay93]
or [Bra00].

Suppose the form error profile may be described as a \textit{band-limited} signal
[Bra00], with maximum frequency \( W \), and that this signal is sampled
with an interval \( 1/(2W) \). In this situation, the signal theory teaches that
the form error profile may be exactly reconstructed by

\[
y(h) = \sum_{i=-\infty}^{\infty} y\left(\frac{i}{2W}\right) \frac{\sin\{2\pi W[h - i/(2W)]\}}{2\pi W[h - i/(2W)]} \tag{2.6}
\]

where the \( y(h) \) is the so called “local form deviation” (see [ISO/TS 12180-1,
deviation relative to Cylinder, Circle, Straight Line and Plan). The ex-
istence of a sampled point in the location \( i/(2W) \) is guaranteed by the
assumed sampling frequency. Shannon reconstruction is based on a sim-
ple assumption: suppose \( n \) points have been uniformly sampled in the in-
terval \([a,b]\), then it may be assumed that \( y(i/(2W)) = 0 \) \( \forall i \notin [1,n] \),
and that \( W = \lfloor n - 1/2(b-a)\rfloor \), where \([x]\) is the greatest integer lower
than \( x \). Basing on these two assumption any sampled signal may be re-
constructed by means of Eq. 2.6; of course, if \( W \geq \lfloor n - 1/2(b-a)\rfloor \) and
\( y(i/(2W)) \neq 0 \) \( \forall i \notin [1,n] \), as usual, the reconstruction is only approxi-
mated.

Yang and Jackman show that if the violation of the hypothesis \( W \leq \lfloor n - 1/2(b-a)\rfloor \) is not very relevant, then the Shannon reconstructed sig-
nal is a very good reconstruction; an instance is proposed involving flatness,
showing that the systematic error in flatness error evaluation is reduced if
the fitting algorithm for error evaluation is applied on the reconstructed surface instead of the original sampled points. However, if Shannon sampling is applied, then the sampling strategy has to be necessarily uniform.

Implicitly signature based reconstructive geometric error estimation could be regarded as a different kind of fitting algorithm (namely the algorithm allowing to extrapolate a geometric error evaluation from a cloud of points), not a sampling strategy. An a priori model works directly on sampling points taken by means of a blind strategy, and tries to improve measurement uncertainty by means of a deeper analysis of available information.

2.4.2 Signature based sampling strategies and models for signature

These sampling strategies require a model to be proposed for the signature; therefore, some few models are proposed.

Models describing signature may be split into two main categories: numerical models, and analytical models. Numerical models are usually easier to generate, requiring a reduced modeling effort for the operation - it could be said that a numerical model automatically is automatically generated from the available data; however, a numerical model is usually difficult to understand - the information it contains is not clearly visible. On the contrary, an analytical model is easier to understand but requires a bit more effort.

The methodology this thesis proposes is based on an analytical model of the signature.

Numerical models

Numerical models are typically based on the mathematical methodology called the “Karhunen-Loève (KL) Transform” [Tum00, Fuk90], which is also known as “Eigenshapes Analysis” [Hen99], and which is strictly linked to the statistical technique known as “Principal Components Analysis” [Joh02]. KL transform is a technique capable of analyzing a series of datasets in order to find similarities between the datasets themselves. In the case of manufactured mechanical part, datasets are samples of point coming from different inspected parts.

Suppose a series of \( M \) mechanical parts manufactured by means of the same manufacturing system (so that it may be supposed they share a common manufacturing signature) are sampled according the same \( n \) points sampling strategy (which should be as dense as possible: to identify the manufacturing
signature the best description of surfaces is required). Therefore, sampling points correspond from one feature to the other. Then, each sample may be rearranged in a vector $y_i$, $i \in \{1, 2, \ldots, M\}$. Tumer et al. [Tum00] suggest to rearrange each sample in a vector $y_i$, supposing just a single value is sufficient to completely characterize the local form error (which may be, e.g., the radius for a roundness profile – angular positions of sampled points are chosen a priori, so only the radius has effectively to be sampled). Then, in [Tum00] a mean profile is derived as $\bar{y} = \left(\sum_{i=1}^{M} y_i\right)/M$, and deviations from the average are calculated: $d_i = y_i - \bar{y}$. Henke et al. [Hen99] model is similar, but does not involve centering, i.e. if Henke et al. eigenshapes model is applied, in the following formulas $d_i$ should be replaced by $y_i$, and in Eq. 2.9 term $\bar{y}$ should not appear. Then a covariance matrix is calculated as follows:

$$V = \frac{1}{M} \sum_{i=1}^{M} d_i d_i^T = \frac{1}{M} D D^T$$  \hspace{1cm} (2.7)

where $D$ is the matrix obtained rearranging vectors $d_i$ in a single matrix. It is interesting to note that $V$ is symmetric, $\text{rank}(V) \leq M$ and $\det(V) \geq 0$.

Now, compute eigenvalues $\lambda_i$ and eigenvectors $e_i$ of $V$, so that

$$V e_i = \lambda_i e_i$$  \hspace{1cm} (2.8)

This is always possible because of properties matrix $V_i$ and $\lambda_i \geq 0 \ \forall i$. The eigenvalues $\lambda_i$ are then ordered, and the relevant features are selected by choosing the first $m \leq M$ dominant eigenvalues. Each original sample vector $y_i$ is then fitted with lower dimensionality by adding the KL linear combination to the sample mean:

$$\hat{y}_i = \bar{y} + \sum_{j=1}^{m} c_{i,j} e_j$$  \hspace{1cm} (2.9)

$$c_{i,j} = d_i^T e_j$$

The chosen $m$ eigenvectors effectively constitute a model, because the original $y_i$ can be regarded as a linear combination of them, plus the average profile $\bar{y}$ and an error $\varepsilon_i$, which is constituted by the weighted sum of discarded eigenvectors:

$$y_i = \hat{y} + \sum_{j=m+1}^{M} c_{i,j} e_j = \hat{y} + \varepsilon_i$$  \hspace{1cm} (2.10)
Moreover, the percentage of variability between profiles explained by the chosen $m$ eigenvectors is \[ \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{M} \lambda_i} \], so if, as suggested, eigenvector linked to the higher eigenvalues are chosen, few eigenvectors should suffice to explain the most part of the variability: if a signature is present, the first eigenvalues tend to be larger than the last ones.

A methodology for choosing the correct value of $m$ still lacks: in order to choose how many principal components suffice to model the signature, Colosimo and Pacella [Col07c] propose to apply Wold’s cross-validation approach [Wol78], which comes from the Principal components analysis.

The cross-validation procedure consists in subdividing data into a number of different subgroups. At each step of the procedure, a subgroup is deleted from the data and re-estimated by performing the KL transform on the remaining data set. Within this step, the estimation of the deleted observations is performed by retaining a different number $m_j \in \{1, 2, \ldots, M\}$ of the first eigenvectors. For each possible choice of $m_j$, the $Q$ statistics is computed as the sum of the squared errors obtained by reconstructing the original $y_i$ via Equation 2.9, i.e. by using the first $m_j$ eigenvectors:

\[
Q_{i(m_j)} = (y_i - \hat{y}_{i(m_j)})^T (y_i - \hat{y}_{i(m_j)}) \quad i \in \{1, 2, \ldots, M\}
\] (2.11)

The procedure is then repeated for all the deleted subgroups and the prediction error sum of squares ($PRESS$) statistic is obtained as the grand average of the $Q_{i(m_j)}$ divided by $n$. $PRESS(m_j)$ is hence used to denote the $PRESS$ statistics obtained by retaining the first $m_j$ eigenvectors. In order to choose whether to add the $(m_j + 1)^{th}$ eigenvector, Wold [Wol78] proposed to compute the ratio

\[
R = \frac{PRESS(m_j + 1)}{\sum_{i=1}^{M} Q_{i(m_j)}}
\] (2.12)

which compares the $PRESS$ obtained by using $m_j$ PCs with the sum of squared differences between data observed and estimated by $m_j$ eigenvectors. If the ratio $R$ is smaller than 1, then a better prediction is obtained by using $m_j + 1$ instead of $m_j$ eigenvectors.

As apparent from this brief discussion, KL transform may directly generate a model from the available data. However, care should be posed to the interpretation of eigenvectors: even if Henke et al. [Hen99] and Colosimo and Pacella [Col07c] have shown that they may be linked to typical defects (i.e. signature) effectively present on the surface, this link has to be identified “by eye” by an expert operator, so it is not guaranteed to be correct. A KL model is surely a powerful instrument for a qualitative analysis of data, but quantitative conclusions on basing on it should be evaluated carefully.
Analytical models

As already mentioned, the methodology proposed in this thesis is based on an analytical model. Therefore, some analytical models will be illustrated in chapter 3. In this section, only a brief description of available literature will be proposed.

The basic structure of an analytical model consists in a series of analytical functions describing possible defects of the modeled feature (e.g. undulation for a roundness profile, convexity or concavity for flatness surface, and so on); then these function are fitted (e.g. by means of regression) on available data and it is chosen, by means of some suitable criterion, which features are really present on the surface, and which have to be discarded. The obtained model has an obvious interpretation, each component of the signature reflects a defect on the surface; however, possible present defects have to be proposed \textit{a priori}, and criteria for choosing effectively relevant features are often quite difficult to adopt - the expert choice may be required to state which components are relevant.

An approach based on “Discrete Fourier Transform” (DFT) \cite{Bra00} analysis is proposed by Cho and Tu \cite{Cho01}. Authors’ intention is to create a “database” which can allow to simulate sampling of roundness profiles. Since the adoption of DFT can be geometrically interpreted as a sum of sinusoids, the manufacturing signature is described in Cho and Tu’s paper by the statistical distribution of amplitudes (i.e. the absolute value of the coefficients of the DFT) of harmonics contained in a profile. The authors analyze the case of cylindrical parts made by turning, and, using experimental data, they identify a beta distribution for each harmonic amplitude (where parameters of the beta distribution differ harmonic from harmonic). However, the model proposed by Cho and Tu may lead to anomalous results. Because a model in which amplitudes of harmonicas are completely random does not take into account the correlation between these harmonics, resulting simulated profiles often seems to be constituted by noise only, without any signature.

Two non DFT models are presented by Henke \textit{et al.} \cite{Hen99}. According to the first model, a profile (or a surface) subject to a geometric tolerance can be considered as a sum of analytical functions, chosen by the analyst depending on the specific case faced. As an instance, the authors develop the case of cylindrical surfaces (holes), and they propose a model which defines radius variations using sinusoidal functions in circumferential direction, and Chebyshev polynomials in axial direction. A criterion to sort functions in order of importance has been proposed, even if the methodology lack a criterion stating how many functions have to be retained (which is left to
the personal evaluation of the operator). An experimental stage is presented to check the adequacy of the model. The second method defines the signature using “eigenshapes” and has already been described in the precedent paragraph.

Henke et al.’s methodology lacks a proper description of autocorrelation in residuals, which is often present in real situations. Three papers address this problem: in the first one [Col04] the problem is solved by applying the Cochrane and Orcutt [Coc49] procedure, which implicitly models the autocorrelation by means of a time series model; the second [Col07b] explicitly models autocorrelation by means of a spatial autocorrelation model [Cre93]. Both these papers consider roundness profile as case study; because the case study in this thesis is based on the model developed in [Col07b], this model will be further illustrated in §3.3.2. Finally, a similar model (i.e. with spatially correlated errors) has been proposed for cylindrical shafts [Col07a], and will be described in §3.3.1.

Back to explicitly signature based sampling strategies

Three works [Mor07a, Sum02] have been proposed dealing with the problem of signature based sampling strategies.

A technique called “Fitted Deviation-based Extraction” (FDE) has been proposed in [Mor07a]; this methodology has been applied to the profile of contoured cams, modeled by a “regression spline” model [Mor07c]. It consists in simply taking as sampling points those points the most deviating from the nominal shape in the fitted model. A weighting function is adopted which prevents points from concentrating excessively. However, this methodology is only empirical: there is no idea on how to tune the weighting function to avoid excessive sampling points density. Further considerations on FDE will be proposed in §4.3.

Summerhayes et al. [Sum02] propose a signature based sampling strategy which, according to the authors, may be adopted with or without reconstruction. This “Extended Zone” (EZ) strategy is based on the so called “V-method” [Mon04]: the criterion aims to minimize the variability in reconstructed points, supposing the signature model is (or may be dealt with as) an “Ordinary Least Squares” Regression model (see §3.1). This EZ method is strongly linked to the criterion to choose sampling points locations proposed in this work, and therefore Extended Zone method will be more deeply described in §4.2.

Moroni and Pacella [Mor08a] propose a reconstruction approach that is very similar to Yang and Jackman [Yan02]’s, as both of them adopt Shan-
non reconstruction in order to fit the model. The two approach differ in the
greater attention Moroni and Pacella pose to the harmonic content in the
feature: Yang and Jackman adopt Shannon reconstruction “blindly”, with-
out considering the real behavior of the feature; instead Moroni and Pacella
analyze the influence of real behavior of the feature on the effectiveness of
reconstruction. In particular, they consider Cho and Tu [Cho01] model for
roundness, which is completely harmonic, plus a specific ARMAX [Shu88]
model for roundness, and show that effectiveness of reconstruction depends
not only on sample size, but on real behavior of the signature, too.

Finally, the problem of choosing the right sampling locations and sam-
ple size, when the inspected phenomenon has a spatial evolution (as a ma-
chined surface/profile has), has been addressed in the field of meteorology
and hydrology [Dix96]. Because in this field a “sampling point” consists in
a meteorological station or a drilled well, which are expensive and cannot be
moved after being built, a great effort has been undertaken to propose meth-
ods to choose the right “sampling strategy”. The only limitation to direct
application of these methods to industrial metrology is the lack of systematic
behavior in meteorological and hydrological phenomena (i.e. no signature
is considered, nor probably exists). Anyway, the economical model leading
the choice of the right sample size proposed in this thesis derives from
hydrology and meteorology approach, so these approaches s have been cited.
Moreover, the idea of applying a “Simulated Annealing”, as described in §5.1,
has been found in a paper from Nunes et al. [Nun04], which considers the
design of Groundwater Monitoring Networks.
Chapter 3

Models describing the manufacturing signature

This chapter is aimed to describe the signature model on which the methodology this work proposes is based on. This model is characterized by the absence of “correlation” in residuals; because a complete analysis of form error behavior and residuals has shown that real surfaces often present correlation, a “Spatial Error Model”, which spatially models correlation, will be depicted in §3.2, and adopted to prove robustness of the methodology facing correlation in §5.1.1.

In this thesis, the problem of choosing the model for describing the signature will not be addressed. It will be supposed that the set of functions $f_1(h), f_2(h), \ldots, f_p(h)$ (see Eq. 3.4), and eventually the correlation structure of residuals, are already known. If the reader is interested in signature model identification, please refer to [Hen99, Tum00, Col04, Col07a, Col07c] and the other references cited in §2.4.2.

3.1 Ordinary Least Squares model

Linear regression [Dra98] is one of the most applied statistical techniques to model data. It allows to build and validate models on experimental data, and then to predict how the inspected phenomenon will behave where it has not been observed. However, because of the statistical nature of the phenomenon itself, which is affected by errors, prediction will not be exact, so a tool to evaluate this error is needed. A possible tool to do so is a confidence interval, which states, with a given probability, limits for the “true” value of the predicted variable.

Paying attention to the true meaning of regression fitting, it can be adopted to describe manufacturing signature: in the following sections, basics
on regression will be given, with a particular emphasis on defining confidence intervals for regression.

3.1.1 Background and conventions

Before starting describing regression properties, let’s point out some convention and basic definition on regression and Multivariate Normal distribution, which is strictly linked to regression.

Let’s consider a generic random vector $x_i$ constituted by $p$ random variables, a single random variable $y_i$ and a (usually unknown) coefficients vector $\beta$; a regression model for these variables is defined as

$$y_i = x_i^T \beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2)$$ (3.1)

thus, $y_i$ is a linear combination of variables in $x_i$, plus a random error $\varepsilon$. Now, suppose that $n$ samples have been taken, thus yielding $n$ values of $y$ corresponding to $n$ (not necessarily different, i.e. replicates are allowed) values of $x$. The above expression shall be rewritten as

$$y = X\beta + \varepsilon \quad \varepsilon \sim N_n(0, \sigma^2 I)$$ (3.2)

namely $\varepsilon$ is distributed according a multivariate normal distribution [Joh02], and variables in the random vector $\varepsilon$ are independent and homoscedastic. Supposing both $\beta$ and $\sigma^2$ are unknown, by means of Ordinary Least Squares (OLS) these data can be used to obtain estimate $\hat{\beta}$, $\hat{\sigma}^2$ and the average value of $y_i$:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$
$$\hat{y}_i = x_i^T \hat{\beta} = x_i^T (X^T X)^{-1} X^T y$$
$$\hat{y} = X\hat{\beta}$$ (3.3)

$$\hat{\sigma}^2 = s^2 = \frac{\varepsilon^T \varepsilon}{n - p} = \frac{(y - \hat{y})^T (y - \hat{y})}{n - p}$$

Estimates obtained by means of these equations are linear combination of gaussian random variables, so they are normal, too [Joh02].

However, when dealing with a geometric feature (surface or profile), $x_i$ and $y_i$ are not completely generic: they are function of the generic location $h$ (e.g. $h$ can be an angular coordinate for a roundness profile, or a vector of cartesian coordinate if a flat plan has to be described, and so on): $x_i = x(h)$
and \( y_i = y(h) \). Eq. 3.1 may be rewritten as
\[
y(h) = \beta_1 f_1(h) + \beta_2 f_2(h) + \cdots + \beta_p f_p(h) = x(h)^T \beta + \epsilon
\]
(3.4)

\[
x(h) = \begin{bmatrix} f_1(h) \\ f_2(h) \\ \vdots \\ f_p(h) \end{bmatrix}
\]

that is, \( y(h) \) is a linear combination of \( p \) known functions \( f_i(h) \) (depending on location \( h \)), plus an independent and normally distributed random noise, with null expected value (i.e. a “white noise”). Eq. 3.4 can be considered as the general form of an analytical model for regression.

This is the general formulation of the regression model that will be considered for the development of the proposed methodology. Now, its properties are introduced.

**Multivariate Normal Distribution: some remarks**

Most of the properties of the OLS regression model are based on the assumption \( \epsilon \sim N_n(0, \sigma^2 I) \), therefore, few concepts on multivariate normal distribution and confidence intervals for it are introduced.

Consider a generic random vector \( y \) of \( n \) components. This vector is said to follow a Multivariate Normal Distribution [Joh02] if and only if every linear combination \( a_1 y_1 + a_2 y_2 + \cdots + a_n y_n \) follows an univariate normal distribution. The behavior of \( y \) is completely defined by its means vector \( \mu \) and its covariance matrix \( \Sigma \).

Suppose \( r (r \geq n) \) samples have been taken from this distribution. Then, let \( \mu \) to be estimated by the sample mean \( \bar{y} \) and \( \Sigma \) by the sample covariance matrix \( S \); a confidence region characterized by the confidence level \( 1 - \alpha \) can be defined for the real value of \( \mu \) by applying [Joh02]
\[
P \left( r(\bar{y} - \mu)^T S^{-1}(\bar{u} - \mu) > \frac{(r - 1)n}{(r - n)} F_{\alpha}(n, r - n) \right) = \alpha
\]
(3.5)

where \( F_{\alpha}(n, r - n) \) is the \( \alpha \) (right tailed) quantile of a Fisher distribution with \( n \) and \( r - n \) degrees of freedom. The confidence region is then an iperellipsoid centered on \( \bar{y} \) and having axis length and direction depending on \( S \).

Basing on Eq. 3.5, Scheffé [Mil66] proposed the following expression which allows to define simultaneous confidence intervals for *any* linear combination \( a^T y \) of variables in \( y \):
\[
P \left( \frac{(a^T \bar{\mu} - a^T \mu)^2}{a^T S a} > \frac{(r - 1)n}{r(r - n)} F_{\alpha}(n, r - n), \forall a \in \mathbb{R}^n \right) = \alpha
\]
(3.6)
This statement can be geometrically interpreted as a projection of the iperellipsoid defined in Eq. 3.5 in the direction defined by \( a \). Intervals calculated by means of Eq. 3.6 are simultaneous, so if any number of linear combination \( a^T \hat{y} \) is calculated and intervals are calculated for them, the probability that the expected value of every linear combination belongs to its confidence interval is greater than or equal to \( 1 - \alpha \), the equality is verified if infinite linear combination are considered.

### 3.1.2 Single Confidence Intervals in Regression

As it has already been stated, \( \hat{\beta} \) and \( \hat{y}(h) \) are normally distributed if random variables in \( \epsilon \) are normal; in particular, if variables in \( \epsilon \) are also independent and homoscedastic it has been demonstrated [Dra98] that

\[
\hat{\beta} \sim N_p \left( \beta, \sigma^2 (X^T X)^{-1} \right)
\]  

\[
\hat{y}(h) \sim N \left( x(h)^T \beta, x(h)^T \sigma^2 (X^T X)^{-1} x(h) \right)
\]

The knowledge of these distributions allows one to easily state single confidence intervals on estimated regression coefficients, regression equation and future observations [Joh02].

**Confidence Region on Regression Coefficients**

If we add confidence region for the sample mean of a multivariate normal sample in (3.5) to distribution in (3.7), we obtain the following confidence region:

\[
P \left( \frac{(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)}{s^2} > p F_{\alpha}(p, n - p) \right) = \alpha
\]  

\[
s^2 = \frac{\hat{\epsilon}^T \hat{\epsilon}}{n - p} = \frac{(y - \hat{y})^T (y - \hat{y})}{n - p}
\]

**Single Confidence Interval on the Regression Equation**

As apparent from Eq. 3.3, \( \hat{y}(h) \) is a linear combination of normal variables, and therefore is normal itself; a confidence interval for the average behavior
of the phenomenon at point \( x(h) \) follows from Eq. 3.8:

\[
P \left( \frac{\left( \hat{y}(h) - x(h)^T \beta \right)^2}{s^2 \left( x(h)^T (X^T X)^{-1} x(h) \right)} > \left[ t_{\alpha/2} (n-p) \right]^2 \right) = \alpha
\]  

(3.10)

where \( t_{\alpha/2}(n-p) \) is the (right tailed) \( \alpha/2 \) quantile of the Student t distribution with \( n-p \) degrees of freedom.

**Single Confidence Interval on a Future Observation**

If the *average* behavior of the phenomenon is not of interest, but the behavior of a *single* future observation at a particular location \( h \) is, a “prediction interval” has to be applied. According to Eq. 3.4, \( y(h) \) is constituted by an average behavior (which has already been analyzed in 3.1.2) plus a random error \( \varepsilon \). Therefore the confidence interval shall be larger than the interval on the Regression Equation:

\[
P \left( \frac{(\hat{y}(h) - y(h))^2}{s^2 \left( 1 + x(h)^T (X^T X)^{-1} x(h) \right)} > (t_{\alpha/2}(n-p))^2 \right) = \alpha
\]  

(3.11)

Fig. 3.1 compares intervals on the average behavior of the regression and prediction intervals (the model is \( y = \beta_0 + \beta_1 h + \varepsilon \), with \( \hat{\beta}_0 = 0, \quad \hat{\beta}_1 = 1, \quad s = 0.5, \) and ten equally spaced sampling points). As expected, prediction intervals are wider than intervals on the average.

### 3.1.3 Simultaneous Regression Intervals

If Eq. 3.10 and Eq. 3.11 are repeatedly applied so that we have an interval on several points \( h \), the overall confidence level will not be \( 1-\alpha \), but lower, as stated by Bonferroni inequality [Mil66]. Therefore, if we wish to simultaneously put limits \( \hat{y}(h) \) for several values of \( h \), an alternative technique is required (for any further issues on simultaneous inference, please refer to [Mil66]).

In particular, Hahn [Hah72] distinguishes four kinds of simultaneous intervals in regression:

1. Simultaneous intervals to contain the true expected value of \( y(h) \), \( h \in \{h_1, h_2, \ldots, h_k\} \).
2. Simultaneous intervals to contain the entire regression equation.

3. Simultaneous intervals on the true average of $y(h), h \in \{h_1, h_2, \ldots, h_k\}$ in $m_i, i \in \{1, 2, \ldots, k\}$ future observations.

4. Simultaneous tolerance intervals.

The first and the second intervals are effectively specific cases of the third one, and will be considered together: essentially, these intervals are generalization of the “Single Confidence Interval on the Regression Equation” defined in 3.1.2. We can add one kind more of simultaneous confidence interval [Lie61], which generalizes “single prediction intervals” described in 3.1.2. Tolerance intervals are linked to prediction intervals, too, but with some difference.

Each of these intervals has different meaning, and serves to solve different problems. Let’s illustrate them.
Simultaneous intervals on the true average of \( y(h_i), i \in [1,k] \) in \( m_i \) future observations

Suppose that a sample of dimension \( n \) has been taken, and a regression model has been fitted on it. We wish to simultaneously state \( k \) intervals on future values of the true average of \( y(h), h \in \{h_1, h_2, \ldots, h_k\} \) (namely \( x(h) \beta \)), knowing that each experimental condition (predictors combination) will appear \( m_i \) times. Hahn [Hah72] proposes the following solution:

\[
P \left( \frac{\left( \hat{y}(h_i) - x(h_i)^T \beta \right)^2}{s^2 \left( \frac{1}{m_i} + x(h_i)^T (X^T X)^{-1} x(h_i) \right)} \right) > \left( u_{\alpha/2}(k, n - p, 0) \right)^2, i \in [1,k] \leq \alpha
\]

(3.12)

Where \( u_{\alpha}(k, r, \rho) \) is the (right tailed) \( \alpha \) quantile of the distribution of the absolute maximum of \( k \) correlated variables distributed according a student \( t \) with \( r \) degrees of freedom, with reciprocal correlation coefficient \( \rho \). These confidence interval are “conservative”, i.e. the real confidence level is “at least” \( \alpha \); equality holds when \( m_i = m_j, \forall i, j \).

It can be pointed out that, when \( k = 1 \) and \( m_i = 1 \), Eq. 3.12 reduces to Eq. 3.11, and, when \( k = 1 \) and \( m_i \to \infty \), Eq. 3.12 reduces to Eq. 3.10.

More generally, if \( m_i \to \infty, \forall i \in [1,k] \), Eq. 3.12 reduces to [Hah72]

\[
P \left( \frac{\left( \hat{y}(h_i) - x(h_i)^T \beta \right)^2}{s^2 \left( x(h_i)^T (X^T X)^{-1} x(h_i) \right)} \right) > pF_{\alpha}(p, n - p), \forall h \leq \alpha
\]

(3.13)

thus solving the problem of stating “simultaneous intervals to contain the true average of \( y(h_i), i \in [1,k] \)”.

However, this solution can manage only simultaneous interval on a finite number \( k \) of experimental conditions, i.e. sampling points locations. If “simultaneous intervals to contain the entire regression equation” are required, then Scheffé technique [Mil66] can be applied, combining Eq. 3.6 and Eq. 3.9:

\[
P \left( \frac{\left( \hat{y}(h) - x(h)^T \beta \right)^2}{s^2 \left( x(h)^T (X^T X)^{-1} x(h) \right)} \right) > pF_{\alpha}(p, n - p), \forall h \leq \alpha
\]

(3.14)

It should be pointed out that in regression usually we have constraints on \( x(h) \), e.g. when two or more predictors are function of an independent
Figure 3.2: Non simultaneous vs simultaneous confidence intervals for the regression equation

parameter, or a predictor is always positive, and especially when $h$ cannot assume any value, namely when the inspected profile/surface is finite; in this case, intervals in Eq. 3.14 are conservative (i.e. the real confidence level is greater than $1 - \alpha$).

Fig. 3.2 proposes a comparison for non simultaneous (Eq. 3.10) and simultaneous (Eq. 3.14) confidence intervals for the regression equation (same model as in Fig. 3.1). Simultaneous intervals are wider; however, it can effectively be stated that, with confidence level $1 - \alpha$, every point of the average regression equation will fall in its interval, and therefore simultaneous confidence interval constitutes a $1 - \alpha$ confidence band.
Simultaneous Confidence Intervals on future observations in $k$ experimental conditions

As Eq. 3.14 generalizes Eq. 3.10 by applying Scheffé technique, Lieberman [Lie61] has demonstrated that it is possible to generalize equation Eq. 3.11:

$$P \left( \frac{(\hat{y}(h_i) - y(h_i))^2}{s^2 \left( 1 + x(h_i)^T (X^TX)^{-1} x(h_i) \right)} > kF(p, n - p), i \in [1, k] \right) = \alpha$$  (3.15)

However, if Eq. 3.14 and Eq. 3.15 are compared, it can be pointed out that Eq. 3.14 does not depend on the number of simultaneous intervals stated, while Eq. 3.15 does; therefore, because amplitude of simultaneous prediction intervals is directly proportional to $k$, if several intervals are to be defined, limits defined by Eq. 3.15 can grow very large, thus making impossible to create a “confidence band” for predictions. Miller [Mil66] suggests to switch to tolerance intervals, if a band is required.

Simultaneous Tolerance Intervals

Tolerance intervals in regression are a concept very similar to “process natural tolerance interval” [Mon96], which defines limits between which a fraction $\gamma$ of the future observations will fall, with probability $1 - \alpha$.

A Tolerance Interval in Regression is defined in terms of predictions [Lie63]: a future value of $y(h)$, given the independent variables $x(h)$, will lie in $[L_{x(h)}(\gamma), U_{x(h)}(\gamma)]$ with probability at least $\gamma$ with confidence level $1 - \alpha$ for any $x(h)$ and any or fixed $\gamma$. The proper interpretation of this probability statement is that the “confidence $1 - \alpha$” refers to the sample from which the regression line is estimated and the “probability at least $\gamma$” to the sampling distribution of future observations. If for a single regression line one asserts that the proportion of future observations falling within the given tolerance limits (for any $x(h)$) is at least $\gamma$, and similar statements are made repeatedly for different regression lines, then for 100 $(1 - \alpha)$% of the different regression lines the statements will be correct.

A conservative solution to the problem of simultaneously state several (infinite) tolerance intervals has been proposed by Lieberman and Miller [Lie63].

If $\beta$ and $\sigma$ are known, then the tolerance interval could be easily expressed as

$$P(K(\gamma)\sigma \geq |y(h) - x(h)\beta|) = \gamma$$

$$\gamma = \Phi(K(\gamma)) - \Phi(-K(\gamma))$$  (3.16)
Where $\Phi(a)$ is the standard normal cumulative distribution calculated in $a$. Intervals can then be stated both for $\beta$ and $\sigma$. Because simultaneous tolerance intervals are desired, for $\beta$ apply equation Eq. 3.14, which may be rewritten as

$$P\left( |\mathbf{x}(h)^T \hat{\beta} - \mathbf{x}(h)^T \beta| \leq s \sqrt{p} \mathbf{x}(h)^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}(h) F_{\alpha/2}(p, n - p), \forall h \right) = 1 - \alpha/2 \tag{3.17}$$

Then, let us introduce an upper bound for $\sigma$:

$$P \left( \sigma \leq s \sqrt{\frac{n - p}{\chi^2_{1-\alpha/2}(n - p)}} \right) = 1 - \alpha/2 \tag{3.18}$$

Where $\chi^2_{1-\alpha/2}(n-p)$ is the (right tailed) $1 - \alpha/2$ quantile of the $\chi^2$ distribution with $(n-p)$ degrees of freedom. Finally, intervals in Eq. 3.17 and Eq. 3.18 are put together by means of Bonferroni inequality. This results in the following statement:

$$P \left( |\mathbf{x}(h)^T \beta \pm K(\gamma) \sigma - \mathbf{x}(h)^T \hat{\beta}| \leq s \left[ \sqrt{p} F_{\alpha/2}(p, n - p) \mathbf{x}(h)^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}(h) + K(\gamma) \sqrt{\frac{n - p}{\chi^2_{1-\alpha/2}(n - p)}} \right], \forall h \right) \geq 1 - \alpha \tag{3.19}$$
which leads to the following tolerance interval:

\[
L_{x(h)}(\gamma) = x(h)^T \hat{\beta} - s \left[ \sqrt{pF_{\alpha/2}(p, n-p)x(h)^T(X^TX)^{-1}x(h)} + \Phi_{(1-\gamma)/2} \frac{n-p}{\chi_{1-\alpha/2}^2(n-p)} \right]
\]

\[
U_{x(h)}(\gamma) = x(h)^T \hat{\beta} + s \left[ \sqrt{pF_{\alpha/2}(p, n-p)x(h)^T(X^TX)^{-1}x(h)} + \Phi_{(1-\gamma)/2} \frac{n-p}{\chi_{1-\alpha/2}^2(n-p)} \right]
\]

The confidence band in Eq. 3.20 should be taught as the area where, \(\forall h\), if several samples are taken, for at least a fraction \(1 - \alpha\) of these samples a fraction \(\gamma\) of the observations will fall.

Several different solutions have been proposed for the problem of stating simultaneous tolerance intervals (see for instance [Lie63] or [Lim88]), but they are always more complex than Eq. 3.19, requiring percentiles which are hard to calculate or the solution of transcendental equation, nor they ensure the stated intervals to be tighter than intervals stated by Eq. 3.19.

The meaning of tolerance intervals is quite similar to the prediction intervals one, but, by means of Eq. 3.19, a confidence band, which cannot be calculated with Eq. 3.15, can be easily defined.

Fig. 3.3 compares simultaneous intervals for the regression equation and tolerance intervals (same model as in Fig. 3.1, \(\gamma = 0.9\)). As expected, tolerance intervals a wider. In metrology, this area has to be considered as the area to which, with a confidence level \(1 - \alpha\), a fraction \(\gamma\) of the manufactured profiles/surfaces will belong.

### 3.1.4 Regression Splines

A particular regression model is the “Regression Splines” model: here, the regression spline model for profiles is proposed.

If functions \(f_j(h)\) in Eq. 3.4 are estimated through a spline expressed in “Truncated Power Series” (TPS) form, we obtain the so-called regression
spline model. A spline in TPS form is a piecewise polynomial function expressed in analytic explicit form by “+” functions: $(a)_+ = \max(0, a)$. These “+” functions represent a basis for the spline space [DB01]. The general analytic TPS form of a spline is:

$$y(t) = \beta_{00} + \beta_{01}t + \ldots + \beta_{0k}t^k + \sum_{i=1}^{n} \beta_{ik}(t - \tau_i)_+^k + \varepsilon$$  (3.21)

where $k$ is the degree of the spline, $n$ the number of knots, $\tau = \{\tau_i\}_{i=1}^{n}$ the set of knots, and $\beta = \{\beta_{ij}\}_{i=0\ldots n, j=0\ldots k}$ the set of coefficients. Here, $h$ has been substituted by $t$ to emphasize $t$ is a simple parameter. In this form, the choice of the knots corresponds to the choice of the model describing the signature.

Regression spline in Eq. 3.21 can describe only monodimensional splines (e.g. the radius of a circular profile as a function of a the angular position
of the considered point [Mor07c]). If more multidimensional (typically three-dimensional, in the space) profiles have to be modeled, several regression splines may be grouped together so that each one models a coordinate of the generic point of the generic profile:

\[
P(t) = \begin{cases} 
  x(t) = \beta^x_{00} + \beta^x_{01} t + \ldots + \beta^x_{0k} t^k + \sum_{i=1}^{n} \beta^x_{ik} (t - \tau_i)^+_k + \epsilon_x \\
  y(t) = \beta^y_{00} + \beta^y_{01} t + \ldots + \beta^y_{0k} t^k + \sum_{i=1}^{n} \beta^y_{ik} (t - \tau_i)^+_k + \epsilon_y \\
  z(t) = \beta^z_{00} + \beta^z_{01} t + \ldots + \beta^z_{0k} t^k + \sum_{i=1}^{n} \beta^z_{ik} (t - \tau_i)^+_k + \epsilon_z 
\end{cases}
\]  
(3.22)

Again, \( t \) is a simple parameter. Note that the structure of the residual \( \epsilon \) is slightly different (it is constituted by three components \( \epsilon_x, \epsilon_y, \) and \( \epsilon_z \)). Each coordinate profile has to be treated separately.

Regression spline is very adapt to model profiles which cannot be described by more “conventional” models (polynomials, sine waves etc.). It is difficult to link a particular piecewise polynomial to a specific process defect, but the regression spline allows to model even surfaces characterized by sharp variations. Fig. 3.4 shows an example of a regression spline modeling the profile of shoe sole.

**Figure 3.4:** Shoe sole modeled by means of a regression spline.
3.2 Spatial Error Model

If the hypothesis of residuals being a white noise is violated, the distributions in Eq. 3.7 and 3.8 do not hold anymore. If it may demonstrated that the real distributions of residuals is $\varepsilon \sim N_p (0, \Sigma)$ (a multivariate normal distribution with a null mean and a generic correlation structure), the so called “Generalized Least Squares” (GLS) estimators (in contrast with OLS estimators) may be defined [Dra98]:

$$\hat{\beta}_{GLS} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y$$ \hspace{1cm} (3.23)

and definitions of $\hat{y} (h)_{GLS}$ etc. follow from this. These estimators own the following properties:

$$\hat{\beta}_{GLS} \sim N_p (\beta, \sigma^2 (X^T \Sigma^{-1} X)^{-1})$$ \hspace{1cm} (3.24)

$$\hat{y} (h)_{GLS} \sim N \left( x(h)^T \beta, x(h)^T \sigma^2 (X^T \Sigma^{-1} X)^{-1} x(h) \right)$$ \hspace{1cm} (3.25)

Even if this formulation would allow to state confidence interval for the regression coefficients, regression equation, and so on, the real problem is the knowledge of $\Sigma$, which cannot be estimated from data, unless some hypothesis is stated for the structure of this matrix. Because the geometric surfaces/profiles have a spatial nature, models adapt to describe the correlation structure of them relate the correlation to the “closeness” of sampling points: nearer sampling points will show a stronger (positive or, seldom, negative) correlation. This kind of model is studied in “Spatial Statistics”; for any issue on Spatial Statistics, please refer to [Cre93].

Three kind of error models are usually considered: point process, in which the distribution of events which may or may not happen in a particular location is modeled; geostatistical models, in which correlation is modeled as a function of the distance point to point; and lattice models, in which measurements are not point to point but refer to areas of the surface, and correlation depends on “borders” of inspected areas (the inspected surface/profile is split in sub-areas and correlation structure is defined basing on this division). Machined surface should be treated as geostatistical data; however, inspection strategies are usually the same for the whole production, so lattice model are easier to apply to sampled surface; moreover, lattice models may be easily be extended to closed surface (like cylinders of roundness profiles) which are often found in mechanical application. Therefore, it will be supposed that
3.2.1 Neighborhood structures for data on a regular lattice

For lattice data, it is of capital importance the concept of “neighborhood”, because two areas being neighbor will have a stronger correlation than two point not being neighbor. For data on an irregular lattice, neighborhood structure is usually obvious: think to borders delimiting countries, as an instance.

For data on a regular lattice, like those coming from the uniform sampling strategy shown in figure 2.5, at least two “classical” neighborhood structure may be defined (Fig. 3.5):
• *Tower structure*, in which two areas are considered neighborhood if they share a *side* of their perimeter.

• *Queen structure*, in which two areas are considered neighborhood if they share a *side* and/or a *vertex* of their perimeter.

This are said *first lag* neighborhood structures. In general, two areas will be said to be neighbors of the $q^{th}$ lag if it is required to get across $q$ borders to go from area one to area two. An instance of lag two neighborhood for queen and tower structure is given in Fig. 3.6.

The neighborhood structure can be defined by a matrix $n \times n W$, where $n$ is the number of areas (sampling points). The first order neighborhood matrix $W^{(1)}$ will be characterized by elements $w_{i,j}^{(1)} = 1$ if area $i$ and area $j$ are first lag neighbors, $w_{i,j}^{(1)} = 0$ else. Definition of higher order neighborhood matrices follow.

Neighborhood matrices are often expressed in “row standardized” form, in which terms $w_{i,j}^{(q)} \in \{0, 1\}$ are substituted by $w_{i,j}^{(q)*} = w_{i,j}^{(q)} / \sum_{s=1}^{n} w_{i,s}^{(q)}$.

### 3.2.2 Spatial Autoregressive Model

Suppose a spatial phenomenon (like a manufactured surface or profile) is characterized by a null expected value. It may be modeled as a “first order autoregressive model” [Cre93] if it can be described by

$$y = \rho W^{(1)} y + \varepsilon$$

where $\varepsilon \sim N_n (0, \sigma^2 I)$

(3.26)

where higher values for $\rho$ indicate a stronger spatial correlation, and $I$ is the unit matrix. Coefficient $\rho$ is often unknown but may be evaluated from data, usually by means of maximum likelihood.

This model can be easily generalized to higher order of spatial correlation, giving rise to “Spatial Autoregressive model of order $q$” SAR($q$):

$$y = \sum_{i=1}^{q} \rho_i W^{(i)} y + \varepsilon$$

$\varepsilon \sim N_n (0, \sigma^2 I)$

(3.27)

In a SAR($q$) model, $W^{(s)}$ are usually expressed in row standardized form. Therefore, each $y (h)$ term may be interpreted as weighted sum of the averages of its neighbors of the various lags, plus a random error.
It may be demonstrated that, if model in Eq. 3.27 holds, then
\[ y \sim N_n \left( 0, \sigma^2 (I - B)^{-1} (I - B^T)^{-1} \right) \]
\[ B = \sum_{i=1}^{q} \rho_i W^{(i)} \]

That is, \( y(h) \) is distributed according a multivariate normal random vector, with a correlation structure defined by coefficients \( \rho_i \) error variance \( \sigma^2 \), and neighborhood structure.

Regression with Spatially Autocorrelated Errors

Regression and SAR\((q)\) models may be mixed, giving rise to the “Regression with Spatially Autocorrelated Errors” SAR\((q)\) [Cre93] model:
\[ y = X\beta + u \]
\[ u = \sum_{i=1}^{q} \rho_i W^{(i)} u + \varepsilon \]
\[ \varepsilon \sim N_n \left( 0, \sigma^2 I \right) \]

A model of this form may solve the problem of simultaneously estimating the \( \Sigma \) covariance matrix and the \( \beta \) coefficients in a GLS model by iteratively calculating, by means of GLS, \( \hat{\beta} \), then obtaining an estimate of \( u \) on which to estimate, by maximum likelihood, \( \rho_i \) and \( \sigma^2 \), going back to the calculation of \( \hat{\beta} \), and so on, till convergence. Therefore, SAR\((q)\) model is suitable for modeling signature when residuals are spatially correlated.

### 3.3 Examples of signature models

In this chapter some models for the signature are proposed. In particular, models for cylindrical surfaces, roundness profiles, and cam profiles are proposed. These models are based on real case studies.

#### 3.3.1 SARX(2) model for cylindrical surfaces

The following model has been proposed in [Col07a].

This case study refers to 100 C20 carbon steel cylinders, which were supplied in 30 mm diameter rolled bars and machined to a final diameter of 26 mm (cutting speed equal to 163 m/min, feed equal to 0.2 mm/rev). Two
cutting steps of 1 mm depth each were required to arrive to the final diameter. Each machined surface was eventually measured by using a Zeiss Prismo Vast HTG CMM, whose Maximum Permissible Errors are $MPE_p = 1.8 \, \mu m$ and $MPE_e = 2 + L/300 \, \mu m$ [ISO 10360-2]. A set of 68 generatrices was sampled on each turned specimen by continuous scanning. In particular a set of 61 points were measured on each of the 68 generatrices. Therefore, $61 \times 68 = 4148$ points were measured on each of the 100 machined surfaces. Fig. 3.7 shows the actual shape of the form error observed in one of the machined items.

In the case of cylindrical features, Henke et al. [Hen99] presented an analytical model for aiding the interpretation or the relationship between manufacturing processes and the typical deviation from the ideal cylinder. The approach combines Chebyshev polynomials and Fourier (periodic) functions for representing axial errors and circular errors, respectively. Fig. 3.8 shows some typical errors observed on cylindrical components modeled by combining Chebyshev polynomials and Fourier functions.
Figure 3.8: Different types of form error for cylindrical surfaces, as classified by Henke et al. [Hen99]: trilobed, taper, hourglass, barrel, banana.

<table>
<thead>
<tr>
<th>Order of the Chebyshev polynomial</th>
<th>Order of the periodic component (Fourier)</th>
<th>( f_i(\theta, \zeta) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>( \cos(2\theta) )</td>
</tr>
<tr>
<td>0</td>
<td>-2</td>
<td>( \sin(2\theta) )</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>( \cos(3\theta) )</td>
</tr>
<tr>
<td>0</td>
<td>-3</td>
<td>( \sin(3\theta) )</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>( T_1(\zeta) )</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>( T_1(\zeta) \cos(2\theta) )</td>
</tr>
<tr>
<td>1</td>
<td>-2</td>
<td>( T_1(\zeta) \sin(2\theta) )</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>( T_1(\zeta) \cos(3\theta) )</td>
</tr>
<tr>
<td>1</td>
<td>-3</td>
<td>( T_1(\zeta) \sin(3\theta) )</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>( T_2(\zeta) )</td>
</tr>
</tbody>
</table>

Table 3.1: The regressor functions \( f_i(\theta, \zeta) \) as a function of the index \( i \).

The proposed model is a SARX(2) model (see Eq. 3.27). Tab. 3.1 summarizes regressors in the model (i.e. constituting vector \( \mathbf{x}(h) \)); in this model location \( h \) can be indicated in a cylindrical reference system, so there are coordinates: the height \( z \) along the axis of the cylinder, and the angle \( \theta \). Note that \( T_0(\zeta) \) represents a second-type Chebyshev polynomial of order \( i \), e.g., \( T_0(\zeta) = 1, T_1(\zeta) = 2\zeta, \) and \( T_1(\zeta) = 4\zeta^2 - 1 \), where \( \zeta = \frac{2(z - z_{\text{min}})}{z_{\text{max}} - z_{\text{min}}} - 1 \). From Tab. 3.1 it is clear that some of the possible combinations of the Chebyshev polynomials and the periodic functions are not included as regressor functions. As examples, \( T_0(\zeta) \) (representing the least-square cylinder radius or \( T_0(\zeta) \cos(\theta) \) and \( T_0(\zeta) \sin(\theta) \) associated to the translation of the least-square cylinder axis are not included in this table. In fact, the model is describing the deviation of the observed data from the ideal cylinder.

Regressors actually included in Tab. 3.1 have been chosen because statis-
tically significant. Then 2nd has been chosen for the spatially autoregressive part of the model because lower order models would not guarantee uncorrelated residuals.

At present, this model has been applied to the on-line monitoring of cylindrical surfaces [Col07a], outperforming traditional monitoring approaches.

### 3.3.2 SARX(2) model for roundness profiles

Even if a SARX model is usually adopted for modeling surfaces, it may be easily adapted in order to model profiles, too. Therefore a SARX model has been applied to roundness: differing from a time series model [Shu88], with a spatial model, closeness of the profile may be easily treated by conveniently defining matrices $W^{(i)}$; Tab. 3.2 proposes the (non standardized) first order neighborhood matrix for a profile sampled on five points.

This model was developed on the same case study described in §3.3.1. A roundness profile has been taken from as far as possible from the part of the specimen which was blocked in the self-centering chuck during machining (because this is the edge of the specimen showing the most apparent signature). The resulting one hundred profiles, each sampled on 748 points, are shown in Fig. 3.9. As apparent from this figure, profiles show essentially undulations. Moreover, it may be demonstrated that the overall model is a SARX(2) model, therefore, the overall model is as follows:

$$ y(h) = r(\theta) = \frac{2}{748} [\beta_1 \cos(2\theta) + \beta_2 \sin(2\theta) + \beta_3 \cos(3\theta) + \beta_4 \sin(3\theta)] + u $$

$$ u = \sum_{i=1}^{2} \rho_i W^{(i)} u + \varepsilon $$

$$ \varepsilon \sim N_n(0, \sigma^2 I) $$

(3.30)
Colosimo et al. [Col08] show that, if estimating both regression coefficients $\beta$ and spatial autoregression model terms $\rho$, they show a multivariate normal distribution characterized by the mean vector in 3.31 and covariance matrix in Tab. 3.3.

$$
E[\beta] = \begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\rho_1 \\
\rho_2
\end{bmatrix} = \begin{bmatrix}
-0.0341 \\
0.0313 \\
0.0080 \\
-0.0322 \\
0.3021 \\
0.2819
\end{bmatrix}
$$ (3.31)

Finally, $\sigma = 0.014$ mm.

In this model the regression and spatial autoregression parameter are not defined by a single value, they are defined by a statistical distribution. This shows what is called the “between profiles” variability: the signature is not perfectly constant, it shows some variability part to part. Even if every part
share the same model, regression and spatial correlation coefficients may vary within specified limits.

This model will be taken as a case study for the application of the sampling strategy planning methodology this thesis proposes. Because the proposed methodology cannot manage explicitly between profile variability, nor presence of correlation in noise, a modification has to be adopted. In particular, an OLS regression model with $E[\beta]$ as regression coefficients will be considered. Because spatially autoregressive errors show a correlation structure as in Eq. 3.29, variance of the residuals will be the average of diagonal elements in matrix $\sigma^2 (I - B)^{-1} (I - B^T)^{-1}$. Robustness of the so calculated derived will be discussed in §5.1.1.

### 3.3.3 Regression Spline model for cam profiles

A set of cams has been produced and measured (Fig. 3.10). Cutting parameters used during contouring are: cutting speed: 550 m/min; feed: 0.02
mm/rev; cutting depth: 0.1 mm. Cams have been measured by means of the same CMM device used for cylindrical surfaces described in §3.3.1; 2014 sampling points were taken on each cam. Because, as apparent from the right side of Fig 3.10, form error profiles present sharp variations, a regression spline model is taught to be the right one to model this signature.

A first model for these cams, characterized by the presence of nine knots, has been proposed in [Mor07a]. This model, which is shown if Fig. 3.11 together with node positions and clouds of sampled points, is said to come from an “homogeneous” approach [Mor07b], which is more adapt for profiles not characterized by sharp variations. Therefore, in [Mor07b] a “inhomogeneous” approach, which leded to a 70 knots model, has been proposed, able to model accurately sharp variations in profile 3.12.

However, it should be noted that none of these two models respect the condition of normality and independence of residuals.
Figure 3.12: Cloud of sampling points, position of selected knots (vertical lines), and fitted spline for the cam profiles according to an inhomogeneous approach.
Chapter 4

On manufacturing based sampling strategies

Before proposing Tolerance Interval Based sampling strategy, few methodologies to plan a sampling strategy basing on signature are introduced. In particular, PCV is based on techniques, linked to “principal Components Analysis”, which have been found in literature, but never applied to metrology (PCV is based on an implicit definition of signature); EPS and FDE have been proposed during researches which lead to Tolerance Interval Based sampling strategy, and are, respectively, a methodology which does not require an explicit definition of the signature, and a methodology strictly linked to the signature model; finally, EZ is a technique specific for metrology, found in literature, that is strictly linked to the Tolerance Intervals Based methodology, so it explained briefly.

4.1 Extreme Points Selection and Principal Components Variables

Both “Extreme Points Selection” (EPS) and “Principal Components Variables” (PCV) extraction strategies have been proposed by Colosimo et al. in [Col07b]. Because these methodology may be better understood with reference to the case study they have originally been applied to, this case study is first introduced.

4.1.1 Case study for EPS and PCV

The case study consists of planar surfaces obtained by face turning. The test item are a set of 130 austempering iron disks of 90 mm diameter. 120 points were measured on each disk, as shown in Fig. 4.1: this sampling strategy can be considered as a very sparse “rectangular grid” [ISO/TS 12781-2]. To
sample the surface a “Zeiss Prismo VAST HTG”, already described in §3.3.1, was used.

For each machined feature, the whole set of 120 points was used to estimate the flatness tolerance by using a Minimum Zone (MZ) algorithm; the standard uncertainty for this reference value is about 0.7 µm (according to previous experimental studies on similar features).

Fig. 4.2 shows the surface obtained by averaging the 130 measured surfaces. The presence of a signature is quite evident: the surface, that nominally should be a flat plane, is concave. Some reasons can be suggested for this behavior: the axis of the lathe used to machine the surface were perhaps not perpendicular, or some strange inflection of the tool may have happened.

4.1.2 Principal Components Variables

Principal Components Analysis (PCA) is one of the best known methods to reduce the dimensionality of a multivariate data set, while preserving as much as possible of the variability present in the data set. An exhaustive description of PCA can be found in standard text as [Jac91]. When dealing with geometric tolerances, the sampled points can be modeled as a multivariate random vector of size $n$ (see §2.4.2). With reference to flatness, the
distances (with sign) of the sampled points from the nominal plan, say \( z \), can be modeled as a \( n \)-variate random vector.

Based on PCA, several variable selection methods have been proposed in the literature [Bea67, Jol72, McC84]. The main idea behind these selection procedures is to retain just few variables (i.e., few points) among the initial set of \( n \) variables (i.e., an initial set of points sampled with a dense inspection strategy) while retaining most of the variability observed in this initial set. Among the different approaches of Principal Component variable selection, it has been shown that the RM method is able to achieve effective results at low computational costs [Cad01]. The method [Cad01, Cad04] is based on an optimality index which selects a subset of \( k \) variables (i.e., a subset of points) which maximizes the percentage of the total variance accounted for by the \( k \) variable subset.

An exact identification of the best \( k \) variable subset, would require an exhaustive enumeration of all the possible subset of size \( k \) that can be found starting from the original set of \( n \) variables. Unfortunately, this exhaustive
Figure 4.3: Histogram of extreme points occurrences.

evaluation requires a prohibitive computational time. Therefore, different heuristic algorithms have been proposed [Cad04]. A possible algorithm is a simulated annealing algorithm for subset selection (a freely available software implementing this algorithm can be found among packages running with “R” [Cer07]).

This sampling strategy will be referred to as “Principal Component Variables” (PCV) in the following.

4.1.3 Extreme Points Selection

When estimating geometric tolerances via the Minimum Zone algorithm (MZ), only few points held all the information needed, namely the “essential subset” (see [Ant96] and appendix A). As an instance, when estimating roundness, just four points define the inscribed and the circumscribed circle characterizing the MZ estimate. These four points constitute the essential subset (ES). If a manufacturing signature is present, points in the ES will tend to appear approximately in the same position. In fact, the manufacturing signature is defined as the systematic pattern characterizing all the items produced with a specific process. When all the machined items have the same systematic pattern, it is likely that the ES will be constituted by a specific subset of points which are always in the same positions.

Therefore, the proposed procedure consists in identifying these subset of relevant points by using a dense sampling strategy on a first set of machined
items. For each machined item, a “1” is associated to the “extreme” sampling points which influence the MZ estimate for that item. Therefore, a simple histogram which represents the number of times that a given point is “extreme” in the sample of machined items, allows to select the position of points that has to be included in the sampling strategy.

Fig. 4.3 shows the EPS histogram, where the abscissa shows the label of each point according to Fig. 4.1, while the ordinate represents the number of times each point resulted “extreme” (i.e., the point influenced the MZ estimate) in the whole set of 130 disks in the reference sample.

**Effectiveness of PCV and EPS**

To prove PCV and EPS strategies effectiveness, the case study introduced in §4.1.1 is considered, and strategies are compared to a standard polar-like [ISO/TS 12781-2] sampling strategy an “Extended Zone” (EZ) [Sum02] strategy, which will be introduced in §4.2.

Let \( i \) denote the \( i \)th flatness surface obtained by face turning an austempering iron disk, \( i \in \{1, 2, \ldots, 130\} \). For each surface, a MZ flatness tolerance estimate can be firstly computed by using the whole set of 120 points represented in Fig. 4.1. For each item \( i \) this value of the tolerance is taken as reference. When just a subset of the whole set of 120 points is considered for the \( i \)th item, the corresponding estimate of the MZ tolerance will be different from the reference value. In particular, the estimated tolerance will tend to be lower than the reference value, because when few points of the whole set are selected, it is unlikely that these points are just the extreme ones. Therefore, the error on the MZ tolerance estimate (i.e., the bias between the reference value and the MZ estimate computed by using just a subsample of the whole set of points) is considered as performance criterion.

Let \( n_j \) denote the sample size of the subset of points. For any given sample size \( n_j \), each sampling strategy allows one to select a specific subset of the measurement points for the \( i \)th item. Since this subset affects the MZ tolerance estimate, each sampling strategy can in principle induce different errors in the MZ estimate.

Fig. 4.4 shows the comparison of the polar strategy with the signature-based measuring strategies (EZ, PCV, EPS). The figure reports the confidence interval on the median error obtained by using the different approaches (on the whole set of 130 items) when the subset size varies, \( n_j \in \{9, 13, 19, 25, 33\} \). The median was used instead of the mean because error data were not normally distributed. It shows that, with the exception of samples of size 9 and 13, the signature-based strategies perform always better
than the polar approach. In particular, the EPS approach proposed in this paper performs always better than all the other ones, despite of the actual value of the sample size.

Eventually, Fig. 4.5 shows the position of the points chosen by the different signature-based sampling strategies (for a sample size $n=33$). As clear from this pictures, all these sampling strategies tend to concentrate sampling points at the center and on the borders of the machined surface, according with the signature pattern shown in 4.2. This well-behaved pattern is particularly clear for the EPS algorithm, which resulted as the best performing approach.
4.2 Extended Zone

Extended zone [Sum02] is quite similar to the Tolerance Interval Based strategy, being characterized by a statistical approach. It has been considered in [Col07b] for comparison, too.

First of all, a model for the signature has to be identified; [Sum02] suggests a methodology described in [Hen99]. This model can be an analytical (OLS regression) model, or a numerical model based on eigenshapes as briefly described in §2.4.2; regardless of the chosen model, the model itself is treated as an OLS regression model (so all conventions and properties in chapter 3 apply), and the criterion for choosing sampling points locations does not change, being based on the distribution of $\hat{y}(h)$ in Eq. 3.8.

Consider an $N$ points dense sampling strategy for the surface of interest: if the signature model is known, a regression matrix $N \times p \mathbf{X}$ may be associated to this strategy. Suppose now only a subset of $n$ of points have been sampled; by means of Eq. 3.8 the variance of $\hat{y}(h)$ can be calculated anyway, for every
point in the original dense sampling. In particular, a matrix $N \times N \mathbf{H}$ may be defined, which is constituted only by zeros, with the exception of diagonal terms $h_{i,i}$, which are equal to one if the point $i$ from the $N$ points sampling is included in the $n$ points sampling. It may be demonstrated that, if only $n$ points are sampled,

$$
\hat{y} \sim N_N \left( \mathbf{X} \mathbf{\beta}, \sigma^2 \mathbf{X} (\mathbf{X}^T \mathbf{H} \mathbf{X})^{-1} \mathbf{X}^T \right)
$$

(4.1)

Summerhayes et al. [Sum02] propose to apply the so called “V-method” [Mon04], i.e. to minimize (fixed $n$):

$$
\min_{\mathbf{H}} \text{var} \hat{y}(h) = \min_{\mathbf{H}} \frac{\sigma^2}{N} \text{tr} \left( \mathbf{X} (\mathbf{X}^T \mathbf{H} \mathbf{X})^{-1} \mathbf{X}^T \right)
$$

(4.2)

as a criterion to choose the location for sampling points, thus defining the “Extended Zone” (EZ) method; an instance of an EZ strategy for the roundness model described in §3.3.2 is given in Fig. 4.6. This criterion, as apparent depends only on $\mathbf{X}$ (which depends only on the choice of the set of sampling points locations $\{h_1, h_2, \ldots, h_n\}$, fixed $n$), namely on the model structure describing the manufacturing signature model (sinusoidal, polynomial, etc.).

The EZ method was originally designed as reconstructive signature-based method: the criterion in Eq. 4.2 serves to choose the right sampling points for reconstruction, which is performed by means of OLS regression as in Eq. 3.3. Summerhayes et al. [Sum02] affirm that form error estimates with an EZ strategy plus reconstruction are better (at least for the cylindricity of holes); however, they admit that with actual CMM software, reconstruction cannot be performed. Anyway, they state (and verify, for holes) that the EZ strategy is effective even without reconstruction.

### 4.3 Fitted Deviation-based Extraction

“Fitted Deviation-based Extraction” (FDE) was first proposed in [Mor07a] for the specific case of contoured cams already proposed in §3.3.3. The criterion is really simple: the areas the farthest from the nominal have to be more densely sampled. With this method those point characterized by the maximum deviations between signature model and nominal profile are selected. Then, in order to avoid an unnecessary sampling points concentration, the deviations in the area around the chosen point are filtered out from the selection procedure.
If, at step $i$ of the algorithm, for $h = h_i$ signature shows the maximum expected deviation $\hat{y}(h_i)$, the point corresponding to $h_i$ is added to the sampling pattern. To filter out points surrounding $h_i$, corresponding deviations are reduced according to:

$$
\hat{y}(h_j)’ = \hat{y}(h_j) \left(1 - \exp\left(-\frac{d_{i,j}^2}{k}\right)\right) \quad (4.3)
$$

where $d_{i,j}^2$ is the distance of the $i$th point from the $j$th, and $k$ is a coefficient chosen by the user; higher values of $k$ generate more uniform distributions. The procedure is iterated until the desired number of points is reached. An example of FDE strategy is proposed in Fig. 4.7: as apparent sampling points are concentrated in the zones of the profile which most deviate from the nominal profile.

In order to check FDE effectiveness, three sampling strategy have been identified: uniform, Hammersley’s sequence, and the FDE strategy here proposed. As performance indicator uncertainty has been evaluated according to [ISO/TS 15530-3].

A comparison of the uncertainty of these six tolerance evaluation method is shown in Fig. 4.8. If only few points are sampled FDE shows bad results.
These results could be explained as follow: using direct FDE the form error evaluation algorithm fails in aligning them to the nominal profile, so tolerance estimates are very far from the calibrated tolerance value. If a large amount of points (more than 150) is sampled FDE shows a slightly improved measurement uncertainty.

### 4.4 Consideration on these manufacturing signature based strategies

Defects of these methodologies, even if valid and always applicable, have lead to the definition of Tolerance Interval Based strategy. Here are some observation to justify the search for a better methodology.

Both PCV and EPS techniques are sample based strategies, thus not requiring an explicit model for signature. Even if this is doubtless an advantage, because a criterion not requiring the effort of identifying a model
for the signature is probably more easy to adapt to a production environment, the lack of knowledge the missing signature identification leads to means that a methodology is adopted without understanding “why” it works. Moreover, without a model identification, efficient techniques proposed in [Col07a, Col07c, Col08] for on-line process control are not applicable. Therefore, if a model has to be identified for process control issues, then it reasonable to adopt a signature model based strategy. Finally, sample based strategy are exposed to the risk that the available data do not completely define the real behavior of the geometric feature (insufficient data have been acquired); a superior knowledge of the feature behavior anyway, model based strategies are less exposed to this risk.

The EZ method has its own limits. As already stated in §4.2, it considers only the regressors structure \( X \) to define the strategy. Therefore, this criterion is not based on the whole signature: as apparent from, as an instance, Eq. (3.30), a signature is defined by the \( \beta \) and \( \sigma^2 \) terms, too. Not considering them could lead to totally inadequate strategies. In fact, when estimating...
form error, as it has already been stated, zones which on average deviate the most from the ideal shape are of major interest, and these zones depend not only on $x(h)$, but on $\beta$, too (Fig. 4.9 shows an average roundness profile characterized by the same $X$ as Fig. 4.6, but different $\beta$). Moreover, as $\sigma^2$ increases, sampling points should be less concentrated, because random spikes in the shape are more probable in zones which seldom show the maximum deviation, so to minimize the risk of not sampling these zones, points should be more dispersed throughout the profile/surface. Anyway, the strong statistical base of the EZ method is surely interesting.

Finally, FDE strategy strongly depends on the definition of coefficient $k$, which influences the strategy in a way that is not easy to quantify. Probably an higher value for variance of residuals $\sigma^2$, namely geometric errors which are characterize more by noise than by signature, should indicate the need of an higher value of $k$, but the link between $\sigma^2$ and the optimal $k$ is not clearly identifiable. Moreover, FDE strategy has shown some problems when dealing with small size samples.

Therefore, in chapter 5 a methodology which tries to summarize both th advantages of EZ (string link with the statistical model) and FDE (identification of critical areas for sampling) will be researched. Finally, none of these models is able to state the sample size together with the sampling points locations. A methodology to choose the right sample size is required.
Chapter 5

Optimal signature based sampling strategy

After proposing, in precedent chapters, a brief discussion on present techniques and norms for sampling surfaces in order to evaluate geometric error, and therefore to state conformance to geometric tolerances, a methodology is introduced so as to improve sampling strategies by reducing the measurement uncertainty (without increasing the sample size) exploiting the interaction between the sampling strategy itself and the manufacturing signature (see [Wil01] and Fig. 1.5).

Moreover, because a criterion for choosing the sample size, which is the main influence factor for measurement costs, should consider the trade off between uncertainty and costs as an objective, and economic objective function will be proposed to find the optimal sample size.

5.1 Definition of the Tolerance Interval Based sampling strategy

As it has been assessed several times in precedent chapters, a signature based sampling strategy, in order to be effective, should identify those area in which critical points (points which effectively define the form error evaluated by means of Minimum Zone), i.e. it should identify areas which most often deviate from nominal shape. It has been stated that FDE [Mor07a] is able to do so, but it lacks a second requirement: concentration of points should be inversely proportional to the “random” part of the form error. The methodology is going to be proposed tries to pass this limit, basing on the statistical nature of the profile like EZ method [Sum02].

Look at the Regression Tolerance interval defined in Eq. 3.20: it should be noted that the local amplitude of the tolerance interval, i.e. $U_{X(h)}(\gamma) - L_{X(h)}(\gamma)$, will usually depends on the density of sampling points: if there
Figure 5.1: TIB strategy (asterisks) and confidence bands for it compared to uniform strategy ones (linear model).

are several sampling points around a location, then the tolerance interval in that location will be tight. Therefore, the statement “sampling points should concentrate where the real average feature deviates the most from the ideal feature” corresponds to “the tolerance interval should be tight where the real average feature deviates the most from the ideal feature”.

A distribution of sampling points ensuring this is satisfied is the one generated optimizing the following function (“Tolerance Interval Based” criterion, TIB):

$$\arg\min_{\{h_1, h_2, ..., h_n\}} \left\{ \max_h \left[U_{x(h)}(\gamma)\right] - \min_h \left[L_{x(h)}(\gamma)\right] \right\}$$  \hspace{1cm} (5.1)

TIB criterion corresponds to the search of that sampling strategy which minimizes the total amplitude of the confidence band; because this total amplitude mainly depends on zones of the profile which most often deviate from the nominal shape, sampling points tend to concentrate in these zones, but
some points are anyway located far from them: this is due to the presence of random errors, and as random error magnitude increase, points tend to be sparser. The solution of Eq. (5.1) requires the sample size \( n \) to be defined \textit{a priori}; if \( n \) is left free to change, it will tend to increase indefinitely, so that a very (at most infinitely) high sampling points density is reached. Fig. 5.1 compares tolerance bands for a standard uniform distribution of sampling points, and for a TIB strategy (same model as in Fig. 3.1, \( \gamma = 0.9 \)). As apparent, tolerance intervals tends to be larger where the profile is not sampled, but are tighter near the edges, which show values most differing from zero.

The minimization of Eq. 5.1 is not an easy task. First of all, it depends on \( \hat{\beta} \) and \( \hat{\sigma}^2 \), which depend on a present sampling, so a part should be sampled according to some \( \{h_1, h_2, \ldots, h_n\} \), then the actual value of \( \{\max_h [U_{x(h)} (\gamma)] - \min_h [L_{x(h)} (\gamma)]\} \) should be evaluated, on the basis of this value a new choice of \( \{h_1, h_2, \ldots, h_n\} \) should be taken, part resampled, and so on, until convergence: the proposed strategy will be very specific for this part, and the iteration of measurement/evaluation is effectively impossible to perform. However, if a model for the manufacturing signature is supposed to be known (namely something like the model in Eq. 3.30 as been identified from several densely inspected workpieces) then values for \( \beta \) and \( \sigma^2 \) are, with some uncertainty, known; because \( \hat{\beta} \) and \( \hat{\sigma}^2 \) are simply estimates for these value, to solve Eq. 5.1 \( \beta \) and \( \sigma^2 \) shall substitute \( \hat{\beta} \) and \( \hat{\sigma}^2 \).

Moreover, the function is strongly non linear and presents several local minima; therefore classical optimization algorithms fail in solving it. Therefore, a “Simulated Annealing” [Kir83] algorithm has been chosen.

Simulated Annealing (SA) algorithm, given a set of \( N \) points, seeks a subset which is optimal, as a surrogate for the whole set, with respect to Eq. (5.1). An initial points subset is selected and passed to a SA algorithm. The algorithm then selects a random subset in the “neighborhood” of the current subset (namely the new subset will be similar to the original subset, differing only for few points), and decides whether to replace the current subset according to the Simulated Annealing rule, i.e. either always, if the alternative subset’s value of \( \{\max_h [U_{x(h)} (\gamma)] - \min_h [L_{x(h)} (\gamma)]\} \) is lower; else with probability \( \exp \left( \frac{D_i - D_{i-1}}{T} \right) \), where \( T \) decreases throughout the iterations of the algorithm, and \( D_i \) is the value of the objective function at iteration \( i \).

5.1.1 Application of TIB strategy to roundness profiles

To prove performances of TIB criterion pass performances of other sampling strategies, a simulation will be carried out based on model found in [Col08]
and described in §3.3.2. As apparent from Eq. 3.20 and 5.1, TIB criterion cannot be applied if a variability of regression coefficients is present (part-to-part variability) or residuals are spatially correlated. Therefore, a simplification of model described in equation 3.30 will be adopted for simulation, that is, an OLS regression model:

\[
y(h) = r(\theta) = \beta_1 \cos(2\theta) + \beta_2 \sin(2\theta) + \beta_3 \cos(3\theta) + \beta_4 \sin(3\theta) + \varepsilon
\]

\[
\beta = \begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\rho_1 \\
\rho_2
\end{bmatrix} = \begin{bmatrix}
-0.0018 \text{ mm} \\
0.0016 \text{ mm} \\
0.0004 \text{ mm} \\
-0.0017 \text{ mm} \\
0.3021 \text{ mm} \\
0.2819 \text{ mm}
\end{bmatrix}
\]

\[
E[\varepsilon] = 0 \quad \sqrt{\text{var}(\varepsilon)} = \sigma_0 = 0.0017 \text{ mm}
\]

The value of \(\sigma_0 = 0.0017 \text{ mm}\) has been chosen considering the multinormal nature of the spatial residuals: their structure is described in Eq. 3.28, so diagonal terms in matrix \(\sigma^2(I - B)^{-1}(I - BT)^{-1}\) represent the overall variance of each spatially correlated residual. Therefore, in order to adopt a variance of the noise similar to the variance of the residuals in the real model, \(\sqrt{\text{tr}\left[\sigma^2(I - B)^{-1}(I - BT)^{-1}\right]} / n = \sigma_0 = 0.0017 \text{ mm}\) (where \(n\) is the number of sampling points) has been considered for the variance of independent residuals in simulations. The TIB criterion optimized sampling strategy for this model is shown in Fig. 5.2, together with tolerance intervals for both uniform and optimized strategy; confidence level \(1 - \alpha\) for the tolerance interval and expected fraction between limits \(\gamma\) are both 95%.

A set of one thousand circular profiles has been generated; for each profile, 748 (original sample size during signature model identification) evenly-spaced points have been simulated; the MZ form error \(T_i\) value calculated from these points has been considered as reference value [ISO/IEC GUIDE 99] (form error has been evaluated by means of the “Loop Control Algorithm” described in [Mor08b] and in appendix A). Then, average error \(\bar{E}\) and error sample variance \(\sigma^2_E\) were considered as effectiveness parameters:

\[
\bar{E} = \frac{1000}{\sum_{i=1}^{1000} T_i - t_{i,n,s}} \\
\sigma^2_E = \frac{1}{999} \sum_{i=1}^{1000} (T_i - t_{i,n,s})^2
\]
Figure 5.2: TIB strategy and tolerance intervals for roundness profile (40 points).

where $t_{i,n,s}$ is the $i^{th}$ form error evaluated by a strategy with $n$ sampling points and the extraction strategy $s$ - uniform, EZ, and TIB strategies have been considered for comparison. $\bar{E}$ is always non negative. As apparent, $\bar{E}$ is related to measurement bias, while $\sigma_E$ relates to measurement uncertainty.

Fig. 5.3 and 5.4 show the behavior of the three proposed strategies as the sample size varies ($n \in \{6, 10, 15, 20, 30, 40, 50, 60, 80, 100\}$) together with 95% confidence intervals. As apparent, TIB strategies show “estimate errors” which are both smaller on average and less dispersed. It has to be pointed out that uniform and EZ strategies do not show significantly different results, too.

To further explain the behavior of TIB strategy, the comparison is replicated varying $\sigma_0 \in \{0.0001, 0.0002, 0.0005, 0.0010, 0.0020, 0.0050\}$ (see Fig. 5.5 and 5.6). As apparent, TIB strategies outperform the other strategies regardless of the value of $\sigma_0$. It should be noted, however, that as $\sigma_0$ increases the usefulness of TIB strategies decreases. This may be explained considering that as random errors standard deviation increases, the signature loses
Figure 5.3: Comparison of average error for uniform, EZ, and TIB strategies.

Figure 5.4: Comparison of error sample variance for uniform, EZ, and TIB strategies.
Figure 5.5: Average errors with 95% confidence interval
Figure 5.6: Error variance with 95% confidence interval
Figure 5.7: TIB strategy and tolerance intervals for roundness profile (40 points, $\sigma = 0.0001$ mm).

Figure 5.8: TIB strategy and tolerance intervals for roundness profile (40 points, $\sigma = 0.01$ mm).
his relevance in determining form error, and therefore a signature based sampling strategy is less effective - if there is not a clearly visible signature, why should a signature based sampling strategy perform better than a standard strategy?

Moreover, Fig. 5.7 and 5.8 show how TIB strategy behaves as the residual standard deviation varies. In fact, the strategy in Fig. 5.7, which is calculated for a signature model identical to the one in Eq. 5.2, with the exception of $\sigma$, which is lower, shows more concentrated sampling points; on the contrary, Fig. 5.8, in which the sampling strategy is calculated for an higher value of $\sigma$, shows a lower concentration of sampling points. This is consistent with the fact that a more evident signature (i.e. characterized by a lower value of $\sigma$) requires points more concentrated, to reduce uncertainty, while if the signature is not clearly visible, sampling points should be less concentrated.

Robustness of TIB strategy

The real behavior of roundness profiles is not correctly modeled by the OLS regression model Eq. 5.2, but by the SARX model in 3.30. Therefore, one could ask if, in a real application, TIB strategy is still effective. To prove the effectiveness of TIB strategy, another simulation, this time adopting the real model in Eq. 3.30, is proposed. Again, one thousand profiles are measured and performances of uniform, EZ, and TIB strategies are compared in term of average error and errors sample variance (Eq. 5.3). Results are shown in Fig. 5.9 and 5.10: TIB strategies still outperform other strategies in term of average error; no difference can be seen in errors sample variance. Therefore, it may be concluded that TIB strategy outperforms standard strategies even if regression coefficients variability and residuals spatial correlation is present.

5.2 Criterion for choosing the right sample size

As stated in §5.1, when applying TIB criterion the sample size has to be chosen before the Simulated Annealing algorithm (or any other optimization algorithm) is started, in order to avoid the sample size to increase indefinitely. Therefore, in order to completely define the optimal sampling strategy, a criterion to choose the right sample size has to be coupled to TIB criterion. The proposed criterion is based on the minimization of the overall inspection cost.

Two sources of cost may be identified in a quality check process: cost of the measurement required to state whether a part is conforming or not
conforming to tolerance, and error related costs.

Measurement costs are quite evident: a measurements has to be performed, so it has to be paid for. However, as it has been described in chapter 1, because of measurement uncertainty some error is alway present in measurement results. Because a measurement error always exists, it is possible that a good part (namely a part complying with the tolerance) is rejected, or a defective part is accepted. The first kind of error has an obvious cost: a part that could be sold is not sold. Cost related to the second kind of error are less directly assessable. Perhaps the part has to be assembled, so the finished product in which it will have to fit will have to be discarded; or the part will be sold, and this will usually lapse the corporate image.

Usually, to reduce the influence of measurement error measurement uncertainty has to be reduced; unfortunately, low uncertainty measurements will be more expensive than high uncertainty measurements; reducing errors frequency when stating conformance or not conformance (which implies the choice of low uncertainty measurements) will therefore increase measurement

Figure 5.9: Comparison of average error for uniform, EZ, and TIB strategies (full model).
costs: a trade-off between inspection error costs and measurement costs has to be researched, and this economic criterion will drive the choice of the right sample size: in a CMM measurement usually uncertainty is inversely proportional to sample size.

The basic structure of the cost function, which will evaluate inspection costs, will be

$$C_I = C_M + C_E$$

where $C_I$ is the cost of inspecting a single part, $C_M$ is the cost of measuring a single part, and $C_E$ is the expected cost of taking a wrong decision about the conformance of the part.

Evaluation of $C_M$ and $C_E$ is required.
5.2.1 Evaluation of $C_M$

For a CMM, cost related to the sampling strategy depends essentially on measuring time: as measuring time increases, cost increases. The measuring time itself then depends on the sample size, so we can express $C_M$ as

$$C_M = c_M \tau = c_M \tau_p n = c_p n$$  \hspace{1cm} (5.5)$$

where $c_M$ is the hourly cost of the CMM, $\tau$ is the measurement time, $n$ is the sample size, and $c_p$ is the cost of sampling a single point.

In Eq. (5.5) the cost related to each sampling point is supposed to be constant, namely each point requires the same time to be sampled. This is not always true, because point to point distances may differ, and then travel time varies depending on points locations; however, because contact time for a CMM is usually predominant, we consider this difference omissible. Moreover, time required to set-up the machine, align the part etc. is not considered, because it does not depend on sampling strategy, and then is not differential.

5.2.2 Evaluation of $C_E$

As it has already been mentioned, two possible errors are possible when inspecting a part: rejecting a conforming part, and discarding a non conforming part. Of course, a criterion considering the measurement uncertainty when deciding conformance or non conformance of parts has to be stated: in the following, the seller points of view criterion proposed in [ISO 14253-1] standard will be adopted (see §1.2).

In the following the case of geometric tolerances, in which usually only an upper specification limit $USL$ is present, will be considered; anyway, it is not difficult to extend equations to bilateral specifications. $C_E$ may then be expressed as follows:

$$C_E = c_{gr} P \left( t > USL - U \cap T \leq USL \right) + c_{da} P \left( t \leq USL - U \cap T > USL \right)$$  \hspace{1cm} (5.6)$$

where $c_{gr}$ is the cost of rejecting a conforming part, $c_{da}$ is the cost of accepting a non conforming part, $U$ is the extended measurement uncertainty, $T$ is the real geometric error, namely the value of the measurand (so a part is non conforming if $t > USL$ and conforming if $T \leq USL$), and $t$ is the measured geometric error (so a part is rejected if $t > USL - U$ and accepted if $t \leq USL - U$).
Therefore, statement $P(t > USL - U \cap T \leq USL)$ means “probability of rejecting a conforming part” and $P(t \leq USL - U \cap T > USL)$ is the “probability of rejecting a non conforming part”.

The evaluation of these terms is very difficult. Evaluation of $c_{gr}$ is quite simple, being equal to value of the considered part if a rejected part is once and for all discarded; differently, evaluation of $c_{da}$ may be very difficult, being related to issues like corporate image, probability of a lot to be refused by the customer, possibility that a finished product with a single non conforming part in it has to be discarded or may be repaired, and so on. Evaluation of $P(t > USL - U \cap T \leq USL)$ and $P(t \leq USL - U \cap T > USL)$ is even more difficult, requiring the knowledge of the joint distribution of $T$ and $t$, which, in the case of geometric error estimation, are usually both non normal and correlated random variables. Therefore, a simpler function is proposed.

First of all, consider that $P(T \leq USL - U \cap T > USL)$ is usually very small: the [ISO 14253-1] criterion of reducing $USL$ by an amount $U$ and to use this limit when stating the conformance of a part has been proposed in order to limit this probability. Therefore, having accepted the [ISO 14253-1] criterion for stating conformance to specifications, it will be assumed that $c_{da}P(T \leq USL - U \cap T > USL)$ is negligible.

Then, to avoid the need of the joint distribution of $t$ and $T$, this final form is proposed for $C_E$:

$$C_E = c_{gr}P(USL - U < T < USL)$$

(5.7)

Probability $P(USL - U < T < USL)$ may be interpreted as the fraction of conforming parts that will be averagely discarded because of the presence of the measurement uncertainty (Fig. 5.11). This is not exactly the “probability of rejecting a conforming part”, but is surely strictly linked to it. The identification of the distribution of $T$ is not very easy, but is surely easier than the identification of the joint distribution of $T$ and $t$; probably the easiest way is to densely sample (in order to make measurement uncertainty as small as possible) a sufficiently large number parts, and from these measurement results to identify the probability distribution – of course this is not exactly the distribution of $T$, but if the uncertainty is small enough, it is a good approximation, and, if the manufacturing signature has to be identified, dense samplings should be already available. It should be remarked that the “true” value of measurand is impossible to know (according the latest definitions of [ISO/IEC GUIDE 99], it is not even defined), too, so the evaluation coming from the dense sampling of parts is its best approximation.

Fig. 5.12 shows a contour plot illustrating the typical behavior of $C_I$ as a function of extended uncertainty $U$ and sample size $n$. The graph has been
plotted considering $c_{gr} = 10 \, \varepsilon, \, c_p = 0.027 \, \varepsilon, \, USL = 0.025 \, \text{mm}$, and the distribution of $T$ is the real distribution of the original one hundred profile from which model in Eq. 3.30 has been obtained.

### 5.2.3 Evaluation of $U$

As apparent from Eq. (5.7), in order to calculate $C_I$ an evaluation of $U$ is required. One could suppose of choosing a sample size, generate a sampling strategy (blind, signature based, . . .) characterized by this sample size, evaluate someway the uncertainty, evaluate the cost function, and iterate until the optimal sample size is found; however, this is unpracticable.

[ISO/TS 15530-3] standard proposes a method to evaluate CMM measurement uncertainty: repeated measurements on one or more calibrated workpieces are performed, then from these results terms $u_p$ (sample standard deviation of repeated measurements) and $b$ (measurement bias, i.e. average of repeated measurements minus the calibrated value) are estimated; two more uncertainty contribution are added: $u_{cal}$ (uncertainty on the calibrated value of the reference workpiece) and $u_W$ (contribution to the uncertainty due to production variability); and finally $U$ is evaluated as

$$U = k \sqrt{u_{cal}^2 + u_p^2 + u_W^2 + |b|}$$  \hspace{1cm} (5.8)

Now, if the signature is know, it may be supposed that $u_W = 0$, because production variability is known. Moreover, if signature is known, in simulations performed in order to evaluate measurement strategy effectiveness (see §5.1.1) it may be supposed that $u_{cal} = 0$. Finally, terms of repeatability $u_p$ and bias $b$ are linked respectively to $\sigma_E$ and $\bar{E}$ in §5.1.1, therefore, uncertainty shall be evaluated as

$$U = k \sqrt{\sigma_E^2 + |\bar{E}|}$$ \hspace{1cm} (5.9)

This expression for uncertainty is still not sufficient in order to optimize the sample size: a model for the uncertainty as a function of the sample size is required. However, if a simulation like the one proposed in §5.1.1 has been performed, some evaluation of uncertainty as a function of the sample size is possible and, from these results, a model for the relationship between uncertainty and sample size can be obtained, e.g. by means of regression.

*Choosing the optimal sample size for roundness*

From $\sigma_E^2$ and $\bar{E}$ in fig 5.3 and 5.4, combined with a simplified evaluation of extended uncertainty ($k = 2$), extended uncertainty as a function of sample
Figure 5.11: Rejected fraction of produced parts.

Figure 5.12: Typical contour plot (isolines) for $C_I$. 
size $n$ in Eq. 5.9 can be evaluated, by means of OLS regression, in the form

\[
\begin{align*}
\text{Uniform:} & \quad U &= 0.0029 + \frac{0.0260}{\sqrt{n}} \text{ mm} \\
\text{Extended Zone:} & \quad U &= 0.0026 + \frac{0.0276}{\sqrt{n}} \text{ mm} \\
\text{Tolerance Interval Based:} & \quad U &= 0.0014 + \frac{0.0280}{\sqrt{n}} \text{ mm}
\end{align*}
\]

which is illustrated in Fig. 5.13. As expected TIB strategy shows an uncertainty which is lower than both uniform and EZ strategy.

Finding the optimal sample size is now quite simple: it is sufficient to superimpose Fig. 5.12 and 5.10, and the $C_I$ isoline which is tangent to the uncertainty function will indicate the optimal sample size (together with the optimal cost) for the considered sampling strategy (uniform, EZ, or TIB). This solution is proposed in Fig. 5.14: the most cheap sampling strategy is, as expected, TIB strategy. In particular, the optimal uniform strategy is
constituted by 43 points, the optimal EZ strategy is constituted by 41 points, and the optimal TIB strategy is constituted by 37 points, so the adoption of TIB has effectively lead to a reduction of sample size.

**Figure 5.14:** Optimal sampling strategy for roundness profile in Eq. 5.2.
Chapter 6

Conclusions and future aims

Manufacturing signature is a powerful instrument for solving several production related issues. A better understanding of what each manufacturing process effectively produces on manufactured parts is very important in the search for a better and better quality of production. Manufacturing signature applications include aid for an improved and standardized of roughness, size and form tolerances, manufacturing parameters adjusting methodologies, powerful statistical process control techniques, and, as in this thesis, optimization of sampling strategy.

In this thesis a model has been proposed for the exploitation of manufacturing signature in order to plan an optimal inspection strategy namely sampling points locations and sample size choice. Essentially, basing on a model of the manufacturing signature, the proposed method searches for those areas of the inspected feature which usually deviates the most from the nominal behavior of the feature itself. The criterion is of statistical nature: the minimization of the overall amplitude of the “Regression Tolerance Interval” for the manufacturing signature has been chosen. Because the amplitude of tolerance interval is locally related to the concentration of sampling points, while the overall amplitude (maximum of the upper limit of the interval minus minimum of the lower limit) depends both on maxima and minima of the average behavior of the signature and on noise superimposed on it, the search of the minimization of the overall amplitude leads to an optimal sampling strategy, with an higher concentration of sampling points in areas the most deviating from the nominal behavior, concentration which is inversely proportional to noise, thus avoiding excessive concentration of sampling points when the manufacturing signature is not very evident.

Having obtained a sampling strategy that, given the sample size, reduces measurement uncertainty, a link between the sample size and the optimal uncertainty is given. Then a criterion has been chosen for choosing the right
sample size. Because as sample size increases uncertainty, and then measurement error related costs, decreases, but contemporaneously measurement time increases, and then measurement costs increase too, an economical trade off has to be found. An objective function has been proposed, which evaluates inspection costs as a function of sample size and measurement uncertainty. The cost function considers the expected fraction of conforming parts which, because of measurement error, are declared to be not conforming to evaluate error cost, and the measurement cost of a single point (considered constant) to evaluate the measurement cost.

These results are supported by an exhaustive analysis of the available literature on sampling strategies, and, to prove the effectiveness of the proposed methodology, the real case of turned circular profiles has been considered.

One of the main defects of signature based sampling strategy is that, if the signature changes, then any signature based sampling strategy may loose its effectiveness, or even generate a very high measurement uncertainty, sampling the “wrong” areas of the surface/profile. Therefore, signature based sampling strategy has always to be coupled to a statistical process control technique capable of guaranteeing the signature to be constant. Techniques of this kind have been proposed in [Col07a, Col07c, Col08].

Future aims in signature analysis include:

- The link between signature an measurement uncertainty evaluation is probably stronger than what was suggested in §5.2.3. A better way to evaluate uncertainty from the signature (a sort of “virtual CMM” [Wil01]) should be investigated. This kind of activity is effectively argument of international standards to come, which are very promising for an integration in the cost function.

- Even if with some limitation, sample based sampling strategies are easier to apply than signature based strategies. Therefore, improving this kind of strategies may be very interesting, for a shop-floor application.

- The proposed cost function is just a simplification of the complete cost function. A way to adopt the “full” cost structure, namely to better link between signature and cost related to its defects, should be investigated. In particular, the structure of the multivariate statistical distribution of measurement result and the “reference” (true) geometric error, should be investigated, being probably relevant in the definition of the uncertainty.

- Finally, for each specific manufacturing (turning, grinding, milling,...) the signature should depend on process parameter. A link between
signature and parameters could help in the choice of the right parameter in order to efficiently respect the tolerance for the product. Moreover, if the signature model is known as a function of process parameter, the signature identification step could be avoided.
Bibliography


[ASME B89.7.3.3] The American Society of Mechanical Engineers, New York, *ASME B89.7.3.3: Guidelines for Assessing the Reliability of Dimensional Measurement Uncertainty Statements*, 2002


[Car95b] K. Carr, P. Ferreira, *Verification of Form Tolerances Part II: Cylindricity and Straightness of a Median Line*, Precision Engineering 17, 144–156, 1995


(GPS) – Coordinate measuring machines (CMM): Technique for determining the uncertainty of measurement – Part 3: Use of calibrated workpieces or standards, 1st ed., 2004


[Sam00] G. L. Samuel, M. S. Shunmugam, Evaluation of circularity from coordinate and form data using computational geometric techniques, Precision Engineering 24, 2000


[Sum02] K. D. Summerhays, R. P. Henke, J. M. Baldwin, R. M. Cassou, C. W. Brown, Optimizing discrete point sample patterns and measurement data analysis on internal cylindrical surfaces with systematic form deviations, Precision engineering 26, 105–121, 2002


Appendix A

Loop Control Algorithm

According to [ISO 1101], “A geometrical tolerance applied to a feature defines the tolerance zone within which that feature shall be contained”. To evaluate whether a real feature complies to the tolerance zone or not, the tolerance has to be estimated. This estimation involves the sampling of a cloud of points on the surface, and then the fitting of the cloud: a “substitute geometry” is identified for it, according to some criterion. In this appendix, the problem of quickly fitting the substitute geometry will be addressed.

It is quite unusual to sample high density samples. Standard algorithms can efficiently manage up to some hundreds of points. But fitting a 300000 points cloud can take up to fifteen minutes, which is unacceptable in productive environment. Even if probably a directly sampled 300000 points sample will never be found, reconstruction algorithms, described in §2.4.1 will propose this situation.

In this chapter roundness profile fitting and roundness tolerance estimation will be the main subject, using, as required by [ISO/TS 12181-2] standard, Minimum Zone (MZ) criterion (the application of which requires the solution of a strongly nonlinear problem); because examples concerning circularity are hard to be clearly illustrated, some instance from straightness when explaining the essential subset will be taken. However, the proposed algorithm can be easily extended to any sort of geometry to reduce the computational effort, if any exact but slow algorithm is known for that geometry.

A.1 State of the art

[ISO/TS 12181-2] suggests that the Minimum Zone criterion should be applied to evaluate roundness; it consists in finding two concentric circles containing the sampled profile having the minimum separation: these circles will define the “tolerance zone”.
MZ criterion involves the solution of a strongly non-linear problem; solving it with standard optimization techniques usually leads to local minima or non convergence problems. Several alternative algorithms have been proposed, which can be split in four categories: Computational Geometry based approaches, Brute Force algorithms, Optimization Techniques, and Approximated Approaches.

Several algorithms based on Computational Geometry have been proposed, like [Roy92, Nov97, Sam99, Sam00, Hua01]. These algorithms usually require the Nearest and the Farthest Voronoi diagram to be calculated for the points cloud; then the MZ center is found in the intersection of the two diagrams. The drawback of methods based on Voronoi Diagrams, that ensure the exact solution, are Voronoi diagrams themselves, that require long time to be calculated.

Another possible approach is “brute force”: if an exact solution for the MZ problem is available for a subset of the sampled points, the tolerance zone is calculated for any possible subset; then, among those “feasible” zones containing the whole cloud of points, the one having the minimum separation is chosen. An algorithm of this kind will be developed in §A.2. However, if the sample size is greater than a hundred points, the required computational effort becomes excessive. Therefore, a reduction in the number of subsets to be considered has yo be searched.

Some algorithms have then been proposed in [Che80, Che85], where it is highlighted the relation of this approach with linear programming; the proposed solution is based on the employ of “limacons”, so the solution is approximated.

A reference work in the field of fitting algorithms is [Ant96], MZ problem, local and global optimality criteria are exactly pointed out, and a classification of fitting algorithms is proposed. This algorithm is strictly linked to [Dri93], which develops optimality theory, and proposes the algorithm developed and improved in this chapter. This algorithm will be briefly explained later.

Some more “combinatorial” approaches have been proposed in literature: they find instances in [Raj99, Jyw99, Dha02], but these algorithm do not seem to guarantee convergence, and are outclassed by the “Steepest Descent” algorithm proposed in [Zhu03]. This algorithm is the quickest found in literature, but it has two defects: it cannot manage some degenerate situation (this deficiency is admitted by the authors themselves), and the solution is guaranteed to be optimal if and only if the initial center position is not far from the real MZ center.

Because MZ fitting is essentially an optimization problem, linear program-
ming and other optimization techniques have been applied to solve it. An instance of these methods is [Tsu84], which applies a modification of Nelder and Mead simplex method to estimate roundness tolerance. However, MZ problem is not linear, so the solution is approximated.

Optimization method can then be modified, thus generating approximated methods, in which a sequence of linear problems is solved, so that solutions converge to the exact one. The most famous algorithm of this kind is Carr-Ferreira [Car95a, Car95b] algorithm, which has been originally developed for flatness, straightness, and cylindricity; in [Gas98, Pac04] it has been shown that this algorithm can be generalized to circularity. Even if the found solution is a very good approximation of the exact solution, this algorithm requires a long time if there are over than one thousand points.

### A.2 A Brute Force Algorithm

The fitting algorithm that will be proposed requires a Brute Force Algorithm: this is a possible implementation for roundness.

Solutions for clouds of points constituted by up to 3 points are straightforward.

The four points cloud is the basic situation that must be solved. Two situations are possible: two points on the inner circumference and two on the outer, and three points on a circumference and the remaining one on the other. Let’s concentrate on the first situation: points can be grouped as follows

\[
\begin{align*}
1, 2 & \quad 3, 4 \quad [j = 1] \\
1, 3 & \quad 2, 4 \quad [j = 2] \\
1, 4 & \quad 2, 3 \quad [j = 3]
\end{align*}
\] (A.1)

For each couple, a possible solution can be calculated by solving the following linear system:

\[
\begin{align*}
2 (x_2 - x_1) a_j + 2 (y_2 - y_1) b_j & = x_2^2 + y_2^2 - (x_1^2 + y_1^2) \\
2 (x_4 - x_3) a_j + 2 (y_4 - y_3) b_j & = x_4^2 + y_4^2 - (x_3^2 + y_3^2)
\end{align*}
\] (A.2)

The first equation refers, here, to the first couple of points, and the second to the second couple; \(a_j\) and \(b_j\) are the center coordinates for the proposed set. In the second situation, the following clusterings have to be considered:

\[
\begin{align*}
1, 2, 3 & \quad 4 \quad [j = 4] \\
1, 2, 4 & \quad 3 \quad [j = 5] \\
1, 3, 4 & \quad 2 \quad [j = 6] \\
2, 3, 4 & \quad 1 \quad [j = 7]
\end{align*}
\] (A.3)
An then solve:

\[
2 (x_2 - x_1) a_j + 2 (y_2 - y_1) b_j = x_2^2 + y_2^2 - (x_1^2 + y_1^2)
\]
\[
2 (x_1 - x_3) a_j + 2 (y_1 - y_3) b_j = x_1^2 + y_1^2 - (x_3^2 + y_3^2)
\]  
(A.4)

As it can be seen, only the points in the triple are used. After calculating the seven possible solutions, defining \( r_{ij} \) as the distance from the \( i \)th points to the \( j \)th solution center, the MZ solution can be identified by selecting

\[
t = \min_j \left( \max_i r_{ij} - \min_i r_{ij} \right)
\]  
(A.5)

When considering clouds of points constituted by \( n \) points, a subset of four points can be extracted in

\[
k = \frac{n!}{4! (n - 4)!}
\]  
(A.6)

different ways, and then there will be \( 7k \) possible centers. Then, applying A.5 to \( 7k \) centers, the MZ solution is easily found.

### A.3 Reducing the sample size: convergence problems

As already stated, the application of Brute Force Algorithms (BFA) leads to long calculation time: if \( m \) points are required to define the tolerance, and the sample size is \( n \), then

\[
N = \frac{n!}{m! (n - m)!}
\]  
(A.7)

subsets have to be inspected. Then, criteria to choose which subsets have to be considered, while ensuring the found solution is optimal, need to be found.

A theorem has been demonstrated [Dri93] stating that, if a particular tolerance zone is minimal for a subset of the points cloud, and the whole set belongs to the tolerance zone itself, then this tolerance zone is the MZ problem solution.

Moreover, let’s define the essential subset: let us consider a sample set \( S \), and a subset \( S_r \) of this sample. Let us suppose the optimal solution for \( S_r \) is optimal for \( S \), too. Now, discard a point from \( S_r \): if the optimal solution for this \( S_{r-1} \) set changes, and it changes regardless of the discarded point, then \( S_r \) is an “Essential Subset” (ES) of \( S \).
This definition tells us that, in a cloud of points, only few points own the complete information about the MZ tolerance, i.e. only few point are needed to calculate the exact MZ solution even for large sample size.

It can be taught that only contacting points (those points lying on the border of the tolerance zone) constitute the essential subset, so any essential subset for a particular geometry should be constituted by the same number of points. Even if this is usually true, this is not always the case. Anomalous situation are rare for roundness, but appear quite frequently in straightness evaluation, so a straightness example will be proposed. Let’s consider the four points set shown in Fig. A.1; as it is evident, the tolerance zone is the one indicated by continuous lines, and point A does not seem to affect the zone itself: it should be discarded from the essential subset. But if point A is eliminated, the new tolerance zone is the one indicated by dashed lines. Then, no point can be eliminated: the essential subset is constituted by these four points.

A “speed-up” algorithm, improving BFA performances, can then be introduced: first, let’s take a subset of the sampling points (usually, in the beginning a single point is chosen), and, by means of BFA, let’s estimate an
MZ tolerance for it. If the related tolerance zone covers the whole points set, then the MZ solution is found; else, the subset is checked for “essentiality”, and redundant points, if present, are removed. Finally, a point external to the tolerance zone is added and the algorithm is iterated.

A degenerate case is well known for roundness: an essential subset constituted by six points is possible if there are three points on the inner circumference and three on the outer, each triple forming an equilateral triangle; the triangles vertexes correspond (Fig. A.2). Because the reciprocal distance of corresponding vertexes is equal to the amplitude of the tolerance zone, this points distribution may be used to test fitting algorithm effectiveness; however, it is obvious that this point distribution is not possible in any real case.

It is intuitive that speed-up algorithm cannot solve this case: a modification of this algorithm is then to be researched.

Figure A.2: Degenerate points distribution
A.4 Loop Control Algorithm

The proposed algorithm is a modification of the speed-up algorithm that is quicker than the original algorithm, and able to solve degenerate cases.

The essentiality check is often more computationally intensive than the zone fitting; but usually situations wherein a non-contacting point of the set belongs to the essential subset seldom appear. Therefore, the essentiality check is eliminated, but instead to directly discard the non-contacting point(s) of the subset. Of course, removing this step would lead to a decrease of computation time, but convergence of the algorithm is not ensured anymore: the algorithm could begin an endless loop.

But a loop situation can be easily detected: it suffices to compare the actual subset to those previously considered; if a loop situation happens, applying the BFA to the union of all of the subsets that have been considered since the first appearance of the “loop subset” ensures the loop breaks off.

A further improvement of the algorithm consists in choosing as next point to include in the subset the point the most external from the tolerance zone, as suggested in [Che80, Che85].

The resulting algorithm is called “Loop Control Algorithm” (LCA).

No crash situation has been found for this algorithm, unlike steepest descent [Zhu03] algorithm and speed-up [Ant96] algorithm. It is fast (perhaps the fastest, at least if it is compared to our software implementation of other algorithms). further improvements have to be searched in a better implementation of the BFA: as it will be shown in Appendix A.2, the used version of it is very “rough”. However, when considering large points sets, the algorithm bottle neck is not the BFA application, but the constraints verification (algorithm termination condition), which is hard to improve.

A.5 Algorithm comparison

Loop control algorithm has been implemented for roundness, and compared with several algorithms for MZ roundness found in literature and implemented using the programming language “Matlab 7.1”. The same computer (a 1.6 GHz Intel Pentium Mobile IV based laptop) has been used for every algorithm. Profiles used for testing where randomly generated, i.e. they were generated, in polar coordinates, according to
\[ \theta_i = 2\pi \frac{i - 1}{N} \]
\[ r(\theta_i) = r_0 + w_1 \]

where \( r_0 \) is the “nominal” radius, and \( w_i \) are uniformly and independently distributed random variables. One hundred profiles were used for every combination of algorithm and sample size \( N \). In the following graphs, averages and 95% confidence interval on computational time have been plotted on the y-axis, while the x-axis indicates the sample size.

Before doing so, implemented algorithms were tested, were possible, applying them on the degenerate case described in paragraph A.3; they all have proven to converge to the real value. Algorithms not able to deal with degenerate situations were then compared to modified Carr-Ferreira algorithm [Pac04], yielding the same results.

A.5.1 Initialization of loop control algorithm

First of all, how the Loop Control Algorithm (LCA) behaves if the starting conditions vary has to be investigated. MZ algorithms are usually initialized using a Least Squares (LS) solution, but it is unknown whether this reduces computational time for LCA or not.

Then, Fig. A.3 shows how LCA behaves when initialized in nominal center, LS center, or a completely randomly chosen center. As it can be seen, if the initialization is not random, the algorithm converges quite quickly; as apparent, LS initialization increases required time, with respect to nominal initialization: this shows that fitting LS center to have a better starting condition does not justify the time required to calculate it. Because it seems to be the best initialization, nominal center will be used as initialization in the following comparisons.

A.5.2 Loop control algorithm against Carr-Ferreira algorithm

Carr-Ferreira algorithm (CFA) is a reference algorithm in the field of approximated algorithms. As every solution of this kind, it requires a series of linear problems converging to the optimal solution, and the solution of these problems requires a long time if sample size is large. Let’s compare CFA with LCA. As Fig. A.4 shows, LCA is always quicker than CFA; then,
Figure A.3: Initialization of loop control algorithm

Figure A.4: Loop control algorithm against Carr-Ferreira algorithm
remembering that they share the same accuracy and advantages in terms of convergence, there is no reason to use CFA instead of LCA.

A.5.3 Loop control algorithm against Voronoi algorithm

Voronoi based Algorithms (VA) are quite hard to implement, but it is possible that simple computation of Voronoi diagrams would require longer than the complete application of LCA. Therefore, a VA was not implemented, but LCA was compared to Voronoi diagram definition, that is already implemented in standard versions of Matlab, required time.

In can be pointed out (Fig. A.5) that, for small size samples (up to one thousand points), computation of VA is quicker than LCA, but then VA required time grows rapidly passing LCA. However, it should be remembered that this is the computation time for a single Voronoi diagram: the complete application of a VA requires two diagrams, and then the best center search, so time can become longer even for small clouds of points. Therefore, it seems that LCA is better.
A.5.4 Loop control algorithm against speed-up algorithm

Speed-up algorithm (SUA) is not able to manage degenerate cases described in paragraph A.3. However, because it is the algorithm from which LCA derives, they have to be compared. A “modified SUA” (MSUA) was compared, too, where the choice of the point added to the subset as the farthest from the MM circumference was introduced. Fig. A.6 proposes this comparison. LCA is much faster than MSUA, as expected. It can be noticed that MSUA is a lot better than SUA, too, therefore, most of the improvement in LCA is due to the point selection criterion. However, SUA and MSUA are not able to solve degenerate sets, while LCA is, thus ensuring a superior completeness.

A.5.5 Loop control algorithm against steepest descent algorithm

Steepest Descent Algorithm (SDA) is probably the most complex and promising method found in literature. It does not require much calculations, so it is probably the fastest algorithm. Effectively (Fig. A.7), SDA is faster than LCA for samples up to two thousand points. However, in judging these algorithms, robustness has to be considered: as stated in paragraph A.1, if it is not correctly initialized, SDA can incur in local minima, while LCA always
converges to the global optimum. Moreover, SDA creators themselves admit their method is not able to manage degenerate situations, so perhaps LCA is slightly better than SDA.

**A.6 Conclusions**

It was shown that, with just few modifications, a BFA for fitting the MZ circumference, that is usually considered to be slow, may become faster than classical approximated algorithms. It must be noticed that the found solution is a global optimum, too.

This algorithm can give, as an additional output, the contacting points delimiting the tolerance zone and, with a little modification, the essential subset; this information allows to construct empirical distributions for contacting points, that is useful if best sampling points pattern has to be found [Col07b].

Differently from other combinatorial quick approaches, like for instance SDA, LCA can manage degenerate situations; even if degenerate situations should never appear in real inspections, this gives LCA a greater completeness.

Finally, this is a very fast algorithm, requiring a little computational time: this is relevant if sample size is greater than ten thousand points.
Ringraziamenti

Innanzitutto, ringrazio tutti quelli che hanno avuto importanza nel mio percorso di dottorando e che non cito espressamente nel seguito.

Ringrazio Dio, Che ha creato il mio, e non solo il mio, intelletto, senza il quale non stareste leggendo queste parole, e mi ha dato la forza per superare i momenti di crisi incontrati in questa che considero la mia prima esperienza lavorativa.


Ringrazio la Prof.ssa Bianca Maria Colosimo, coordinatrice del nostro dottorato e, di fatto, anche se il ruolo non mi pare esista per i dottorati, correlatrice della mia tesi, per la sua impagabile capacità di beccare non solo gli errori concettuali in questa tesi, ma anche quelli legati alla mia incapacità di trasmettere i concetti - vedi il mio orribile inglese... 

Ringrazio tutti gli altri docenti della Sezione di Tecnologie e Sistemi di Produzione per la loro disponibilità a sopportare alle mie (ampie) lacune di tecnologo.

Ringrazio l’ing. Francesco Cacciatore e l’ing. Pasquale Aquilino, anima pratica di questa tesi, perché se ogni tanto mi ricordo che l’ingegneria non è solo ricerca, è in buona parte merito loro.

Ringrazio poi gli altri colleghi ai vari livelli della gerarchia politecnica, e in particolare gli ingegneri Moreno Castelnuovo, Laura Catellani, Marco Cavallaro, e Maurizio Tomasella, miei compagni d’ufficio, per aver sopportato la mia petulanza e l’abitudine a parlottare tra me e me mentre lavoro.

Ringrazio i miei genitori, per tutte le volte che mi hanno detto “lascia perdere” vedendomi troppo preoccupato per il dottorato.

Ringrazio i miei amici, in particolare il dott. Mario Galbiati e la dott.ssa Anna Teruzzi, entrambi dottori in ricerca (c’è sempre stata della rivalità tra noi, almeno da parte mia – se prenderò il dottorato è anche colpa loro), e l’ing. Elisa Galbiati (che ha avuto il coraggio di chiedermi aiuto sia per la tesi di primo che di secondo livello, su argomenti di cui non sono esattamente un esperto, per semplice fiducia nelle mie capacità di dottorando).

Un ringraziamento particolare a tutte le mie ragazze della pallavolo, come allenatore, e ai miei compagni di squadra, come giocatore – non elenco tutti i nomi perché sarebbero troppi e per non dover creare un ordine di preferenza. Mi avete sempre ricordato che nella vita di un uomo ci sono cose molto più importanti del lavoro: fiducia, impegno, convinzioni, e valori.

Grazie!