1 Introduction

Manufacturing processes leave a very specific pattern or signature on the part. This provides a good basis for inspecting parts. Hence, sampling the part throughout uniformly or randomly, may altogether miss process characteristics and errors, and in many cases, be wasteful. In some cases, it is possible to quantify the manufacturing errors and their effect on the product, based on the model of the processing [2–4]. In some other cases, where multiple processes are applied on the same feature, the net effect of these processing errors will be too difficult to model accurately. It is hence suggested that the manufacturing errors in such cases can be captured using experimental analysis. The first few parts manufactured can give a good estimate of how the part must be sampled. Sampling can also provide information for more effective process control. We used a profound modification of the support vector regression (SVR) method to quantify process errors on products, and thereby provide a basis for adaptive sampling and form inspection. The subject of this paper is the use of converting the support vector machines into simpler formulations for easier solutions. Consequently, these simpler formulations are used to study simple and complex nonlinear form errors [5,6]. This paper suggests the mathematical foundations on this basis, for better form verification through coordinate metrology.

Support vector machines (SVMs) are algorithms used in supervised learning tasks such as statistical regression and binary classification. SVMs were proposed by Vapnik [7], about 25 years ago, but they only have been studied by the machine learning community since the mid-90s [8]. Kernel methods and SVMs were thoroughly investigated for the past ten years [9,10] as alternatives to Bayesian networks, decision trees, and artificial neural networks, for binary classification problems. SVM applications range from computer vision to speech recognition and cryptography. Nonetheless, despite several improvements for training SVMs on large sets of data [11,12], it is still computationally demanding to train SVMs quickly on very large batches of data and/or on noisy inputs, with or without missing features. The basic SVM problems are often large quadratic programming (QP) problems that are solved with computationally demanding interior point methods. The sequential minimal optimization (SMO) method proposed by Platt [11] is an attempt to circumvent the problem of training speed, though it relies on a heuristic that can sometimes be inadequate. Hence, in order to surpass these problems using SVMs, we introduce a simplified and straightforward approach to the construction of nonlinear nonparametric regression machines that gives exceptionally simple QP problems, very efficiently solved with classic numerical techniques. This is in extension of the previous work by Trafalis and Gilbert [13] for robust classification and regression for noisy data.

2 Quadratic Formulations for Nonlinear Regression

Consider a finite set of \( \ell \in \mathbb{N}^* \) observations \( (x_i, y_i) \) with \( x_i \in E \subseteq \mathbb{R}^n \) \( (n \in \mathbb{N}) \) and \( y_i \in \mathbb{R} \) for every \( i \in [1, \ell] \). The set \( E \) is a compact subset of \( \mathbb{R}^n \). These observations are the realizations of random variables that are considered to be independently drawn and identically distributed according to an unknown probability distribution. A vector \( x_i \) describes an entity characterized by \( n \) significant real-valued features (Note that categorical features can be described by real-valued features, using a multiple correspondence analysis). The associated target value \( y_i \) takes any value in \( \mathbb{R} \) that represents the output of an unknown function \( f: E \rightarrow \mathbb{R} \). Our regression problem is to find the real-valued nonlinear function \( f \) (without any given model), such that, for every \( i \in [1, \ell] \), we have that \( f(x_i) = y_i \), and such that, for any new observation \( (x_{\ell+1}), y_{\ell+1} \), the probability that \( f(x_{\ell+1}) = y_{\ell+1} \) is the highest possible.

Consider a set of \( \ell \) slack variables \( \xi_i \), \( i \in [1, \ell] \), such that \( f(x_i) - y_i = \xi_i \). Ideally, all slack variables would be null if such a function \( f \) was existing (and if it was perfectly known beforehand). However, the problems we consider are such that \( f \) is partially or totally unknown, and the only prior knowledge is that \( f \) is a smooth function. Therefore, all regression functions obtained by our approach will be continuous and smooth approximations of \( f \), for which the absolute value of the sum of the slack variables is minimized.

Let \( k \) be a continuous function \( k: E \times E \rightarrow \mathbb{R} \) that is symmetric and positive semidefinite (such a function is called a positive semidefinite kernel or simply kernel), and suppose that \( f \) belongs to a reproducing kernel Hilbert space (RKHS) \( \mathcal{F} \), such that \( k \) is the reproducing kernel. From the reproducing property of RKHS, we have, for every \( i \in [1, \ell] \), \( f(x_i) = \langle k(\cdot, x_i), f \rangle \), where \( \langle \cdot, \cdot \rangle \) is the dot product.
product in $\mathcal{F}$. Now, let $S$ be the linear space spanned by the $k(\cdot, x_i)$ images of the vectors $x_i, i \in [1, \ell]$. Since $f$ is supposed to belong to $\mathcal{F}$, we have that the vectors $f = - \sum_{i=1}^{\ell} \alpha_i k(\cdot, x_i) + f^\perp$, where $f^\perp$ is orthogonal to $S$ and $\alpha_i \in \mathbb{R}, i \in [1, \ell]$. Thus,

$$f(x) = \langle k(\cdot, x), f \rangle = \sum_{i=1}^{\ell} \alpha_i k(\cdot, x_i) + f^\perp = \sum_{i=1}^{\ell} \alpha_i k(\cdot, x_i) = \sum_{i=1}^{\ell} \alpha_i k(\cdot, x_i)$$

where $K_{ij}$ is the $(i,j)$-th element of the $\ell \times \ell$ Gramian matrix $K$, made of all the dot products $(k(\cdot, x_i), k(\cdot, x_j)) = k(x_i, x_j)$. The matrix $K$ (also called kernel matrix) is a positive semidefinite matrix. For all $i \in [1, \ell]$, the set of constraints is as follows:

$$\sum_{j=1}^{\ell} \alpha_i K_{ij} - y_i = \xi_i \quad (1)$$

In other words, $f(x_i) - y_i$ is a slack variable that is minimized in the objective function (ridge regression). Though in real life problems, such a representation of $f$ might be too restrictive, especially when a kernel $k$ is chosen without clear ideas about the actual form of $f$ (furthermore, it is not possible to determine an expression for $f^\perp$ from the above). Thus, the function $f$ is replaced by $f + b$, where $b \in \mathbb{R}$ is a new variable added in Eq. (1). This leads to a new set of linear constraints for all $i \in [1, \ell]$, which is

$$\sum_{j=1}^{\ell} \alpha_i K_{ij} + b - y_i = \xi_i \quad (2)$$

where $b, \alpha_i$, and $\xi_i$ are variables, and where $y_i$ and $K_{ij}$ are given beforehand. If we consider the contributions of $f^\perp$ to be negligible, then we obtain an approximated rule $\tilde{f}$ of the form $\tilde{f} = \sum_{i=1}^{\ell} \alpha_i k(\cdot, x_i) + b$.

Concerning the objective function, we already know that the sum of the absolute values of the slack variables should be minimized, but we have to find a way to ensure that for any new observation $(x_{t+1}, y_{t+1})$, the probability that $f(x_{t+1}) = y_{t+1}$ is the highest possible. Hence, consider the set of functions $\mathcal{F}_B$ defined by

$$\mathcal{F}_B = \left\{ x \in E \rightarrow \sum_{i=1}^{\ell} \alpha_i k(x, x_i) \in \mathbb{R} : \alpha^T K \alpha \leq B^2 \right\} \subseteq \mathcal{F}$$

with $B \in \mathbb{R}^*$. It is proven that, given a probability $p \in [0, 1]$, the generalization error of $f \in \mathcal{F}_B$ is bounded with probability $p$ by an expression of the empirical Rademacher complexity of the class $\mathcal{F}_B$, denoted by $\hat{R}(\mathcal{F}_B)$, and the sum of the absolute values of the slack variables. The smaller $\hat{R}(\mathcal{F}_B)$ is, the smaller the generalization error is. From Ref. [9], we have $\hat{R}(\mathcal{F}_B) = 2B \sqrt{\log(K)/\ell} + 2|b|/\sqrt{\ell}$. Thus, it is natural to try to minimize $|b|$. $B$, and the sum of the absolute values of the slack variables, with respect to the constraints defined in Eq. (2). However, minimizing the quantity $B$ can also be done by trying to minimize the quantity $\alpha^T K \alpha$ for any vector $\alpha \in \mathbb{R}^\ell$. Furthermore, using the Cauchy–Schwarz inequality, and noticing that the Euclidean norm $\| \cdot \|_2$ is a consistent norm, we have $\alpha^T K \alpha = \langle \alpha, K \alpha \rangle = \| \alpha \|_2^2 \leq \| K \|_2 \| \alpha \|_2^2$. Therefore, since $K$ is given, we can approximate the problem of minimizing $B$, by minimizing the quantity $\alpha^T \alpha$. This approximation is quite loose but the conditioning of the Hessian matrix in the resulting QP problem is greatly improved. The final QP formulation of the nonlinear nonparametric regression problem then becomes

$$\min_{(\alpha, b) \in \mathbb{R}^{\ell+1}} \alpha^T \alpha + b^2 + C \xi^T \xi$$

subject to $K \alpha + b - y = \xi$ \quad (3)

where $y \in \mathbb{R}^\ell$ is a vector made of all $y_i$ for $i \in [1, \ell]$, and $1 \in \mathbb{R}^\ell$ is a vector made of ones. This problem tries to find an approximate regression function $\tilde{f} = \sum_{i=1}^{\ell} \alpha_i k(\cdot, x_i) + b$, such that, for the majority of the observations $(x_i, y_i)$, we have $\tilde{f}(x_i) = y_i$, and such that for any new observation $(x_{t+1}, y_{t+1})$, the probability that $\tilde{f}(x_{t+1}) = y_{t+1}$ is the highest possible. The constant $C \in \mathbb{R}^*$ in the objective function is a trade-off parameter between giving more importance in minimizing the sum of the absolute values of the slack variables, or more importance in having a small empirical Rademacher complexity.

Problem 3 is written canonically as

$$\min_{u \in \mathbb{R}^\ell} \| u^T H u / 2 + A u \|_2 \quad (4)$$

where $u^T = (\alpha^T, \xi^T, b)$, $A = [K, -I, I]$, with $I$ being the $\ell \times \ell$ identity matrix, and $H$ is a diagonal matrix with its diagonal $\Delta$ equal to $\Delta^2 = 2(I_{\ell}^T C I_{\ell}, 1)$. Since $H$ is a symmetric and invertible matrix, we can derive the Lagrangian dual problem of Eq. (4), which is

$$\min_{v \in \mathbb{R}^\ell} v^T Q v / 2 + y^T v \quad (5)$$

where $Q = (K^2 + I_{\ell} / C + I_{\ell} 1_{\ell}^T / 2)$ [14]. Since the Hessian matrix of this new QP problem is positive definite, the unique optimal solution of this problem is $\hat{u} = Q^{-1} y$, and the optimal $\alpha$ and $b$ of the primal problem (5) are given by

$$K^2 + I_{\ell} + 1_{\ell} 1_{\ell}^T d = y, \quad \alpha = K d, \quad b = I_{\ell}^T d \quad (6)$$

Note that if the kernel matrix $K$ is not invertible, then the value of $C$ must not be too large. Otherwise, the resulting linear system will be ill conditioned.

## 3 Computational Issues

### 3.1 Invertible Kernel Matrix

If the kernel matrix $K$ is invertible, we can derive approximated solutions in cases when the trade-off parameter $C$ is large. From equations in Eq. (6), we have

$$(K^2 + I_{\ell}) d_0 = y, \quad \alpha = K d_0, \quad b = I_{\ell}^T d_0 \quad (7)$$

The linear system in Eq. (7) can be solved by two approaches: by using direct methods or by using iterative methods.

If $K^2$ is well conditioned, then it is possible to solve the linear system $(K^2 + I_{\ell}) d_0 = y$ using a generalized minimal residual method (GMRES) [15]. This method has a computational cost of $O(\ell^2)$ floating-point operations since $K^2$ is not sparse. If the matrix $K^2$ is not so well conditioned, then a diagonal matrix $d_0$ is added to $K^2$ in order to improve the condition number. However, this approach is the same as reintroducing the trade-off parameter $C$ with the obvious relation $\epsilon = 1 / C$.

The second approach for solving the linear system in Eq. (7) is to use direct methods. Using the binomial inverse theorem, we have

$$K^2 + I_{\ell} 1_{\ell} 1_{\ell}^T = K^{-2} - K^{-2} 1_{\ell} 1_{\ell}^T (1 + I_{\ell} K^{-2} 1_{\ell})$$

and then equations in Eq. (7) can be written as

$$d_0 = I_{\ell} - \frac{K^{-2} 1_{\ell} 1_{\ell}^T}{1 + I_{\ell} K^{-2} 1_{\ell}} K^{-2} y, \quad \alpha = K d_0, \quad b = I_{\ell}^T d_0$$

which leads to the resolution of the following two linear systems:

041001-2 / Vol. 131, August 2009 Transactions of the ASME
The first direct method to consider, if the conditioning of the kernel matrix is enough, is a QR decomposition $K=QR$, where $Q$ is an $\ell \times \ell$ orthogonal matrix, and $R$ is an $\ell \times \ell$ upper triangular matrix. In this case, the matrix $Q$ is not involved in the computation of $K$, since $K=QR$. Hence, we obtain the solution of the first system by solving successively the triangular systems $Rz_1^{(1)}=y$ and $Rz_2^{(2)}=z_0^{(1)}$. The solution of the second system is given by solving successively the triangular system $Rz_1^{(2)}=1_\ell$ and $Rz_2^{(2)}=z_0^{(2)}$. The computational complexity of the QR decomposition is $O(\ell^2)$, and solving each one of the four triangular systems is only $O(\ell^2)$. Thus, the computational cost is higher than for the GMRES iterative method, however, the conditioning of the matrix $K$ is better than the one of $K^2$.

The second direct method to consider, if the conditioning of the kernel matrix is not enough, is a Cholesky decomposition $K=LL^T$, where $L$ is an $\ell \times \ell$ lower triangular matrix. Here we obtain the solution of the first system by solving successively the triangular systems $Lz_1^{(1)}=y$, $Lz_2^{(1)}=z_0^{(1)}$, $Lz_1^{(2)}=z_0^{(1)}$, and $Lz_2^{(2)}=z_0^{(2)}$. The solution of the second system is given by solving successively the triangular systems $Lz_1^{(2)}=1_\ell$, $Lz_2^{(2)}=z_0^{(2)}$, $Lz_1^{(2)}=z_0^{(2)}$, and $Lz_2^{(2)}=z_3$. The computational complexity of the Cholesky decomposition is $O(\ell^3)$, and solving each one of the eight triangular systems costs only $O(\ell^2)$. Of course, it is more computationally expensive than a simple QR decomposition, however, the conditioning of the problem is vastly improved.

Once $z_1^{(1)}$ and $z_2^{(1)}$ are obtained, we derive the optimal solutions $\alpha$ and $\beta$ by computing,

$$d_0=z_1^{(1)}-\frac{1}{1+\lambda}z_2^{(2)}, \quad \alpha=Kd_0, \quad \beta=1^T_d_0$$

Note that it is required that no duplicated observation should be left, otherwise, the kernel matrix $K$ would be singular. Also, even if no duplicates exist, the matrix $K$ will be ill conditioned if some observations have too similar feature values. In other words, the cloud of observations should be evenly spread so that dot products of face-milled plates, cylinders, cones, and spheres with a coordinate measuring machine (CMM), and we fit models of these surfaces. Then we use the developed QP formulation for studying the variations of amplitudes of the models. From this nonlinear nonparametric regression analysis, it is possible to estimate how large the error zone of the probed surface is. The deformation surfaces that are generated by our regression technique are then used for the second part of this study. These regression models are analyzed in order to detect a pattern in the shape of the surfaces that can indicate which part should be further inspected to improve the determination of the error zone. Furthermore, this pattern is a precious clue in the optimization of the sampling schemes used by our CMM for probing the plates, since it gives ideas about the general shape of the deformations.

### 4.2 Problem Formulation

Measurements on a surface are given as $\ell$ samples $(x_i)^T$, where $x_i \in \mathbb{R}^3$. The ideal surface can be defined by an equality of the type $\Phi(x) = 0$, where $\Phi: \mathbb{R}^3 \rightarrow \mathbb{R}$ is an implicit function for which one of the Cartesian coordinates of $x \in \mathbb{R}^3$ is dependent on the other two. We also know that there always exists, locally, a function $\phi$, such that this dependency can be written explicitly, i.e., (supposing that in our case that the third Cartesian component is depending explicitly on the other two) we can write an equality of the form $y_1 = \phi(x_1, x_2)$ for $(x_1, x_2)$, belonging to a restricted domain of $\mathbb{R}^2$.

These measurements must be corrected, prior to the regression analysis, because of possible misalignment in the setup procedure of the CMM. To do so, we introduce a translation vector $b \in \mathbb{R}^3$ and a $3 \times 3$ rotation matrix $A$, defined by the three Euler angles $\psi \in [0, \pi]$ (precession), $\theta \in [0, \pi]$ (nutation), and $\varphi \in [0, 2\pi]$ (proper rotation). Then we have to determine the optimal angles and the optimal translation, such that a certain quantity is minimized. This quantity is defined by

$$\sum_{i=1}^{\ell} (y_i - \Phi(x_i) - \phi(x_1, x_2))^2$$

where $y_i = A(\psi, \theta, \varphi)x_i + b$, $e_i^T = (0, 0, 1)$, and $n_i$ is a unit vector normal to the surface at the point of coordinate $(y_1, y_2, \phi(x_1, x_2))$.

This is equivalent to projecting the discrepancies between the measurements and the ideal shape on a unit vector that is normal to the ideal surface after rotating and translating the measurements, and then summing them. Also, this is a way to minimize from a very limited set of measurements any systematic error caused by a faulty axis determination on the CMM. Once the adjustments are made, we compute the quantities

$$z_i = |(n_i, e_i^T)(y_i - \phi(x_1, x_2))|$$

and we suppose that the $z_i$'s are the summations of two terms. One term represents the actual deformation, and the other one denotes a random error, i.e., $z_i = \delta_i + \varepsilon_i$, where $\delta_i$ is the slack from the ideal surface, and $\varepsilon_i$ is a random error of measurement. It is assumed that the $\varepsilon_i$'s are normally, independently, and identically distributed with a zero mean, and that they are all uncorrelated. The $\delta_i$'s are considered to be generated by an unknown but continuous deformation pattern during the manufacturing process. This is a shape of interest because it indicates how the manufacturing process is making the actual surface diverge from its ideal profile. Furthermore, it also indicates where and how similar surfaces should be probed, such that the deformation profile can be determined quickly. The continuity assumption of the deformation surface will reasonably hold if there are no obvious damages on the parts before their inspection by a CMM. Using our QP formulations, the nonlinear nonparametric regression on the quantities $z_i$ gives an interpolating function $f$ for the $\delta_i$'s of the following form:
\[ f: y \in \mathbb{R}^2 \rightarrow \sum_{i=1}^{\ell} \alpha_i k\left(y, \left[ y_i, y_j \right]\right) + b \]

where \( k \) is a kernel function. Naturally, the choice of the kernel is critical, and we chose in this study a Gaussian RBF kernel defined by

\[ k(y_1, y_2) \in \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \exp\left(-\frac{|x-y|^2}{2\sigma^2}\right) \in \mathbb{R} \quad (11) \]

where \( \sigma > 0 \) is a parameter that typically represents an average distance between each data point \((y_i_1, y_i_2)\). The role of the Gaussian kernel is to make local “bumps” of radius \( \sigma > 0 \) that add up in order to interpolate a continuous surface from a mesh of discrete points (it is similar to the Parzen window method that is used for probability density function estimation). This kernel leads to a smooth and continuous representation of the deviation surfaces of the studied plates. Furthermore, it generalizes, locally, the shape of the deviation surface, such that the accuracy of the representation is not seriously affected by missing portions of the measurements. However, the deformation surface will no longer be interpolated by \( f \) outside a certain range that depends on \( \sigma \).

If the residuals \( z_i - f((y_i)_1, (y_i)_2) \) are uncorrelated, independent, and normally distributed with a 100(1-\( \alpha \))% confidence interval for the mean centered approximately at \( \mu_a=0 \) with a half-width \( c_a \) of the order of the accuracy of the measurements, then the function \( f \) is deemed to estimate the deformation surface with a relatively good accuracy. We also try to confirm this trend by running a cross validation and verify that the median relative absolute deviation is as low as possible. The surface of a part will be considered to conform to the standards if its interpolated deformation surface is within the pre-established minimum zone. Hence, we need to determine the variability of the deformation surface. Therefore, once the function \( f \) is determined by our approach, we perform a global optimization of \( f \) on its domain \( E \) in order to obtain the quantity as follows:

\[ d = \max_{x \in E} f(x) - \min_{x \in E} f(x) \]

Then, if the confidence level for the interval of the mean of the residuals is high (\( \alpha = 0.01 \), for example), and if \( \mu_a = 0 \) and \( c_a \) are small with regards to the accuracy of the probe, then the surface will conform to standards if the quantity \( d + 2(\mu_a + c_a) \) is less or equal to the desired tolerance. In order to confirm this result, we probe again the location of the estimated optimal points, and see if the experimental results match the estimations.

We should notice here that the traditional nonlinear least-square technique will fail to interpolate the deformation surface (represented by the function \( f \) in our notations) since there exists no analytical model for this surface. In this study, we chose a smooth translation-invariant kernel that is not influenced by large local variations (for a large parameter \( \sigma \), see Eq. (11)). This kernel just defines locally how distances and angles are measured on the interpolated deviation surface. Consequently, we have replaced the model-driven approach of the nonlinear least-square technique that requires a profound study about how the plate was manufactured by a local deformation approach that only takes into account the gradient at which the deformation surface can vary. This technique effectively removes the need for a deformation model, but provides only an approximated interpolation surface that ignores nanoscopic cracks and deformations (which cannot be measured by a probe-type CMM).

### 4.3 Experimental Setup

#### 4.3.1 Software. We have tested our QP formulation with a Dell Precision Workstation 530, equipped with two 2.4 GHz Intel Xeon processors and 2 Gbyte of RAM. Our codes were developed by our own means under MATLAB 7.4.

#### 4.3.2 CMM. We inspected the surfaces with a Brown & Sharpe MicroVal PFx™ 454 CMM, equipped with a touch trigger probe head. This CMM has a linear displacement accuracy of 5.1 \( \mu \)m along each axis, and a measurement repeatability of 3.8 \( \mu \)m. The touch trigger probe head is a Renishaw PH9/PH10 manual probe head, capable of holding M2 and M3 stylus. Two styls were used in the following experiments:

- a Renishaw M2 stylus ref. A-5003-0577: \( 0.7 \) mm ruby ball, \( 0.5 \times 20 \) mm\(^2 \) tungsten carbide stem, mass: 0.32g
- a Renishaw M3 straight stylus ref. A-5000-3552: \( 2 \) mm ruby ball, \( 1.4 \times 21 \) mm\(^2 \) stainless steel stem, mass: 1g

Our probe-type CMM does not produce significant quantities of measurements in a short period of time, thus, our data sets rarely exceed 300 points and the computational times are extremely small.

#### 4.4 Results. We reported all our results in Table 1. This table contains the widths of the error zone of every surface that we inspected. All figures are given in \( \mu \)m and the accuracy is \( \pm 18 \mu \)m.

#### 4.4.1 Face-Milled Plates. A batch of five face-milled plates was produced for the experimentation. The cutting parameters used were as follows:

- workpiece: aluminum 6061-T6, \( 4 \times 4 \times 0.5 \) in.
- tool: \( 3 \) in. cutter, seven inserts with carbide coating
- machining conditions: coolant, cutter speed=7500 rpm, step depth=0.01 in., cutting feed=60 ipm

The plates were labeled from 1 to 5 and were visually inspected before being measured. Plate 5 was discarded due to inconsistent machining surface. All plates were inspected with a M3 stylus ref. A-5000-3552, and the sampling meshes used were all uniform. Special sampling sequences are not used to ensure proper comparison.

The ideal shape of the surface of these plates shall be a square portion of a plane. Our ideal shape is defined by the equation \( x_3 = 0 \), and therefore, the positional correction consist in finding the Euler angles \( \psi, \theta, \) and the scalar \( b_3 \) such that the quantity

\[ \sum_{i=1}^{\ell} \left[ (\sin \psi \sin \theta \cos \psi + \cos \theta) x + b_3 \right]^2 \]

is minimized. Once we have the optimal \( \psi, \theta, \) and \( b_3 \), we compute the quantities \( z_i \) as defined in Eq. (10), and we determine a regression surface from these \( z_i \)‘s using our QP formulations and a Gaussian kernel. The typical interpolated deformation surfaces are saddle shaped, with the “hill” sides being the sides that were clamped during machining (see Fig. 1).

The width of the error zones of our plate are quite small (50–70 \( \mu \)m) with respect to the accuracy of the CMM (\( \pm 9 \mu \)m on the position of a particular point along each axis). Therefore, if some face-milled plates are flatter than these, the \( z_i \)’s are likely to be random fluctuations of the CMM, and the shapes of the regression surface will not be realistic anymore. The particular shape of the deviation surfaces allows to determine the width of the error zone with few points. The results in Table 1 shows that nine contact points of a uniform mesh are already enough to estimate this width, precisely. Naturally, since these deformation surfaces are saddle shaped, the only points of interest are on the borders of the plates. That is why three points at equal distance along each side are enough to obtain a good regression surface, and therefore, a good estimate of the width of the error zone. Table 1 shows that 3 \( \times \) 3 uniform meshes give estimates as good as 12 \( \times \) 12 meshes with respect to the accuracy (18 \( \mu \)m). Furthermore, our results have no large variations when increasing the sample size for all the plates. This advocates for the consistency of our approach.

Our results indicate that good samples for the form inspection of
face-milled plates are contact points of mesh grids grossly aligned with the plate sides. This is, also, a consequence of the saddle shape of their deformation surfaces.

### 4.4.2 Cylinder

One-half cylinder was produced using an end milling operation for the experimentation. This was done so that the surface generated will have a specific machining pattern that is ideal for measurement and analysis with a CMM. The cutting parameters used were as follows:

- workpiece: aluminum 7075-T6, \( \varnothing 3 \) in, 1.35 in. long
- tool: \( \varnothing 0.5 \) in ball nose end mill with a high speed steel (HSS) cutter
- machining conditions: step over=0.025 in.

The half cylinder was inspected with a M2 stylus ref. A-5003-0577, using uniform grid meshes of 64, 96, 189, 315, and 630 contact points.

The ideal shape of our half cylinder of radius \( \rho > 0 \) is defined by

\[
x_3 = \phi(x_1) = \begin{cases} 
\sqrt{\rho^2 - x_1^2} & \text{if } -\rho \leq x_1 \leq \rho \\
0 & \text{otherwise}
\end{cases}
\]

Thus, for the positional correction, it is just necessary to determine the three Euler angles \((\psi, \theta, \varphi)\), a translation \(b_1\) along the \(x\)-axis, and a translation \(b_3\) along the \(z\)-axis. The translation along the \(y\)-axis can be dropped since \(x_3\) is not a function of \(x_1\). In our case, we find that \(\psi=0.125\) deg, \(\theta=0.032\) deg, \(\varphi=0.102\) deg, \(b_1=-10\) \(\mu\)m, and \(b_3=17\) \(\mu\)m when using the \(63 \times 10\) mesh.

The interpolated deformation surface for the largest sample can be seen in Fig. 2. All interpolated surfaces for every sample size appear to have the shape of longitudinal waves along the \(x\)-axis, with 11 peaks total. No substantial deformations along the \(y\)-axis were detected in any sample. The longitudinal wave can be interpreted by the movement of the cutter during the machining process. During machining, the cutter moves linearly between reference points that belong to the ideal shape. Therefore, the real trajectory of the cutter is a chain of small linear trajectories that closely matches the ideal cylindrical profile (piecewise approximation). Although, it also involves periodic deviations from the ideal shape, resulting in the wave pattern observed on our cylinder.

The consequence of this longitudinal wave pattern is that, on cylinders that are manufactured similarly, meshes must have a lot of contact points along the \(x\)-axis, and can have less points along the \(y\)-axis. In our case, 32 contact points along the \(x\)-axis and 2 contact points along the \(y\)-axis (on each side) were enough to determine the width of the error zone with the same accuracy than with the full \(63 \times 10\) mesh grid (see Table 1).

### 4.4.3 Cone

A 2.5 in. high aluminum cone with a \(\varnothing 3\) in. base was manufactured using a taper turning process, and it was inspected with samples of 8, 64, and 256 data points. The sampling used a Hammersley distribution or contact points were chosen randomly.
randomly at the surface. A M2 stylus ref. A-5003-0577 was used for recording the measurements. The ideal shape of our cone of base radius $r > 0$ and height $h > 0$ is defined by

$$x_3 = \phi(x_1, x_2) = \begin{cases} \\ h(1 - \sqrt{x_1^2 + x_2^2}/p) & \text{if } \sqrt{x_1^2 + x_2^2} \leq p \\ 0 & \text{otherwise} \end{cases}$$

In this case, the positional correction resides in determining the optimal Euler angles $\psi$ and $\theta$, and a translation vector $b \in \mathbb{R}^3$, since this cone is invariant by the rotation along the $z$-axis. Furthermore, the dot product between the unit normal vector to the surface $n$, and the canonical vector $e_i$, is constant and defined everywhere on the surface, except at the tip. Since the probability of shifting one of the measurements precisely at the tip of the cone is negligible, the positional correction of this cone consists in minimizing the following quantity:

$$\sum_{i=1}^{\ell} ((y_i)_3 - \phi((y_i)_1, (y_i)_2))^2$$

with respect to $\psi$, $\theta$, and $b \ (y_i = A(\psi, \theta)x + b)$, where $A(\psi, \theta)$ is a rotation matrix defined by the Euler angles $\psi$ and $\theta$.

For the 256 points “Hammersley 1” sampling, we find $\psi = 63.328$ deg, $\theta = 17.2$ deg, $b_1 = 25 \mu$m, $b_2 = -170 \mu$m, and $b_3 = 1413 \mu$m. The results in Table 1 show that the width of the error zone of the cone is very important (about 700 $\mu$m) with the largest deviations being at the tip of the cone and at its base (see Fig. 3). This is possibly due to the fact that the angle at the top of the real cone is not exactly what it should be. Normally, the angle at the top should be 61.927 deg, but this angle, in reality, is 1.153 deg smaller, which implies that the real cone does not fit the ideal cone at the tip and at the base. Also, if the real angle was 60.774 deg, then the deformation surface will look such as the one in Fig. 4. If the angle is modified, then the deformations are extreme only at the tip of the cone (center of Fig. 4), and the average width of the error zone given by the two 256 points Hammersley samples is 160 $\mu$m. This result advocates for contact point samplings that radiate from the tip, with a density of points that decreases regularly until it reaches the base of the cone. A high density of points at the tip is crucial for the accuracy of the results. For example, in Table 1, the widths of the error zone for the eight point Hammer-
samples are significantly different than the ones obtained by samples with more points at the tip. Furthermore, the results given by the eight point random samples were completely inaccurate and were not reported in Table 1.

4.4.4 Sphere. One 2.5 in. half sphere was manufactured using a contour milling process, and it was inspected with samples of 8, 16, 64, 128, 256, and 512 data points. The samples used were a Hammersley distribution or points were chosen randomly at the surface. A M2 stylus ref. A-5003-0577 was used for recording the measurements. The ideal shape of our half sphere of radius \( \rho \) is defined by

\[
x_3 = \phi(x_1, x_2) = \begin{cases} 
\sqrt{\rho^2 - x_1^2 - x_2^2} & \text{if } \sqrt{x_1^2 + x_2^2} \leq \rho \\
0 & \text{otherwise}
\end{cases}
\]

In this case, the positional correction resides in determining a translation vector \( \mathbf{b} \in \mathbb{R}^3 \), since a sphere is invariant by the rotation along any axis. Consequently, the positional correction of this half sphere consists in minimizing the following quantity:

\[
\sum_{\ell=1}^{\ell} ((x_1, x_2)_3 + b_3 - \phi(x_1, x_2, b_1, b_2))^2
\]

with respect to \( \mathbf{b} \in \mathbb{R}^3 \).

For the 256 point Hammersley 1 sampling, we find \( b_1 = 1 \) µm, \( b_2 = -3 \) µm, and \( b_3 = 2227 \) µm. Results in Table 1 show that no particular sampling within the ones tested is privileged. All samplings give more or less the same width for the error zone, which is about 225 µm. Figure 5 indicates that the highest deformations are on the top of the half sphere and on its base. Therefore, appropriate sampling should focus on these two areas in an even manner. Finally, Fig. 5 (right) shows that the manufactured half sphere has a base, a little bit larger than specified, and that its top is slightly flattened. Actually, our analysis show that the radius is in fact 0.008 in. larger than specified, which is why the half sphere is larger at the base and flat at its top. Thus, if the radius was changed to 1.258 in., the average width of the error zone would drop from 225 µm to 70 µm.

5 Conclusion

We have successfully derived simplified QP formulations using kernel methods for nonlinear regression problems. Contrarily to Kriging and other interpolation methods, the proposed approach is finding an interpolation rule (or function), which is different than computing a set of interpolated values like in the other “traditional” methods. The advantage of this approach is the ability to generalize the behavior of the deformations outside the domain where most samples were collected. Also, whatever the chosen kernel is, the approach requires absolutely no statistical assumptions on the collected data, which is another key difference with the traditional methods. The extremely simple QP problems that have been derived give general symbolic solutions that can be solved efficiently with very classic numerical techniques. The actual speed mainly depends on four factors: the form of the shape to be analyzed, the choice of the numerical algorithms, their

![Fig. 4: Interpolated deformation surface of the cone if the desired angle at the top was 60.774 deg](image)

![Fig. 5: Interpolated deformation profile of the cylinder (left). Cross view of the sphere at y=0 in. with its interpolated deformation surface amplified ten times (left)](image)
implementation into the machine, and the type of computer used to run the program. In this paper, the numerical algorithms that are chosen are intrinsically fast due to their relatively low theoretical computational complexities. The rest is a matter of good coding in a suitable language and a fast machine. Furthermore, we showed that our regression technique can be used to estimate efficiently the flatness of face-milled plates, the cylindricity of cylinders, the conicity of cones, and the sphericity of spheres. On face-milled plates, preliminary results show that the deformation surfaces have a saddle shape for which the hill sides are the sides that were clamped during machining. These results indicate smarter sampling techniques for quickly assessing the flatness of other face-milled plates. Our observations on other types of surface pilling techniques for quickly assessing the flatness of other face-clamped during machining. These results indicate smarter sampling techniques for quickly assessing the flatness of other face-milled plates. Our observations on other types of surface piling techniques for quickly assessing the flatness of other face-clamped during machining. These results indicate smarter sampling techniques for quickly assessing the flatness of other face-milled plates.

Acknowledgment

The authors wish to acknowledge the National Science Foundation (NSF) under Grant No. DMI0427966 for providing financial support for this work.

References