

Form inspection using kernel methods

R C Gilbert, S Raman*, and T B Trafalis

School of Industrial Engineering, University of Oklahoma, Norman, Oklahoma, USA

The manuscript was received on 22 November 2008 and was accepted after revision for publication on 8 October 2009.

DOI: 10.1243/09544054JEM1433

Abstract: Form inspection of non-linear surfaces is a difficult task as suitable analytical models are often unavailable. This paper presents a mathematical model for surface inspection of face-milled plates and determination of the minimum zone based on a modification of the support vector machine (SVM) technique. The SVM approach is reformulated to regression problems using a different methodology than the 'largest margin' paradigm. In addition, this work derives extremely simple quadratic programming (QP) problems that allow for general symbolic solutions to non-linear regression problems. The results obtained from preliminary testing allow identification of processing tendencies so that a selective sampling procedure may be applied for inspecting future plates from that lot.

Keywords: support vector machines, regression, kernel methods, face milled plates, surface inspection

1 INTRODUCTION

The inspection of deformations on the surface of parts is facilitated by characteristic patterns left by manufacturing processes. Sampling parts [1] in a random fashion may miss process characteristics and errors and, in many cases, be inefficient. Nevertheless, if a processing model is available then manufacturing errors and their effects on the product can be evaluated [2–6]. However, models can be very difficult to derive, especially when multiple processes are applied on the part. This paper introduces a form of support vector regression (SVR) that evaluates process errors on parts and can therefore be used for form inspection and adaptive sampling. The subject of the paper is the adaptation of a machine learning approach based on support vector machines (SVMs) into simple algorithms in order to obtain more computationally efficient and accurate form inspection methods. The focus is on the flatness determination of face-milled plates. The objectives of this task are:

- (a) to check if probed plates have their actual surface profiles located within a certain distance from a nominal plane (tolerance zone) [7];
- (b) to optimize the probing procedure (sampling size, sampling pattern [8], measurement path, etc.) in order to accelerate the process and to diminish time-induced errors;
- (c) to improve overall accuracy.

The first part of this study relies on a regression approach that is introduced to estimate the size of the 'tolerance zone'. After collecting measurements on the plate surface with a coordinate measuring machine (CMM), the formulation is used to fit an improvised test model of that surface. The size of the tolerance zone can then be estimated by study of the amplitude variation of the regression model. These surface models are analysed in order to uncover a pattern in the deformation generated by the manufacturing process. This pattern is a precious clue in the optimization of the sampling schemes used by the CMM for probing the plates.

The paper is structured as follows. Section 2 introduces the methodology employed for the form inspection of face-milled plates. Section 3 discusses quadratic formulations for non-linear regression problems and some computational issues concerning these formulations. Experiments on several face-milled plates are provided in section 4. Finally, section 5 concludes the paper.

*Corresponding author: School of Industrial Engineering, University of Oklahoma, Norman, OK 73019, USA.
email: raman@ou.edu

2 METHODOLOGY OF FORM INSPECTION

2.1 Theoretical background

Measurements of the surface of a face-milled plate are given as ℓ pairs $(\mathbf{x}_i, y_i)_{i=1}^{\ell}$, where $\mathbf{x}_i \in E$, with E (the probed surface) being a compact subset of \mathbb{R}^2 , and $y_i \in \mathbb{R}$ is the elevation of the surface at a given point \mathbf{x}_i . The target value y_i is assumed to be the summations of a nominal term, a deviation term, and an error term such that $y_i = f(\mathbf{x}_i) + \delta_i + \varepsilon_i$, for all $i \in [1, \ell]$ where f is the representative function of the nominal plane of the plate, δ_i is the deviation from this nominal profile, and ε_i is a random error. It is assumed that the ε_i 's are normally, independently, and identically distributed with a mean equal to zero, and that they are all uncorrelated. The δ_i 's are considered to be generated by an unknown but continuous deformation pattern during the manufacturing process. The continuity assumption will reasonably hold if the surfaces of the plates are not obviously damaged before their inspection by a CMM. The function f representing a plane is interpolated by linear regression. If the regression approach of section 3 is used, then it is equivalent to choosing a linear kernel $k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$ (a kernel is a continuous, symmetric, and positive semi-definite function) and plugging it in our quadratic programming (QP) formulation. This approach defines a family of coefficients $(\alpha_i^0)_{i \in [1, \ell]}$ and a scalar b_0 so that the function $f_0: \mathbf{x} \mapsto \sum_{j=1}^{\ell} \alpha_j^0 \langle \mathbf{x}_j, \mathbf{x} \rangle + b_0$ interpolates the function f . Once the function f_0 is determined, the same non-linear regression interpolates the quantities $y_i - f_0(\mathbf{x}_i)$. The resulting interpolating function f_1 , defined by $f_1: \mathbf{x} \mapsto \sum_{j=1}^{\ell} \alpha_j^1 k(\mathbf{x}_j, \mathbf{x}) + b_1$, estimates the deviations δ_i . As discussed in the next section, adequate choice of a non-linear kernel k for f_1 is critical. In this study, a multi-quadric kernel defined by $k(\mathbf{x}, \mathbf{y}) = (\|\mathbf{x} - \mathbf{y}\|^2 + \sigma^2)^{1/2}$, is chosen where $\sigma > 0$ is a parameter large enough in order to reduce the influence of the small local coordinate perturbations between points close to each other. This kernel leads to a smooth and continuous representation of the deviation surfaces of the studied plates. Furthermore, it generalizes the shape of the deviation surface such that the accuracy of the representation is not seriously affected by missing portions of the measurements. Other adequate choices for the kernel are possible, notably the so-called translation-invariant kernels such as the Gaussian radial basis function, which is defined by $k(\mathbf{x}, \mathbf{y}) = \exp[\|\mathbf{x} - \mathbf{y}\|^2 / (2\sigma^2)]$.

2.2 Flatness determination

If the residuals $r_i = y_i - [f_0(\mathbf{x}_i) + f_1(\mathbf{x}_i)]$ are independent and normally distributed with a $100(1 - a)\%$

confidence interval for the mean that is centered on zero with a half-width (denoted by c_a) of the order of the accuracy of the measurements, then the function f_1 is deemed to estimate the deviation surface with good accuracy. This paper will try to confirm this trend by running a cross-validation and confirm that the normalized root mean square deviation is as low as possible and the linear correlation coefficient is as close to one as possible. A face-milled plate will be considered flat if the form of the representation of f_1 on E is within the pre-established minimum zone. Thus, once f_1 is determined by our approach, a global non-linear optimization of f_1 on its domain E gives the quantity $d = |\max_{\mathbf{x} \in E} f_1(\mathbf{x}) - \min_{\mathbf{x} \in E} f_1(\mathbf{x})|$. If the confidence level for the interval of the mean of the residuals is high ($a = 0.01$, for example) and if c_a is small relative to the accuracy of the probe, then the plates will be considered flat if the quantity $d + 2c_a$ is less than or equal to the desired variation spread. In order to confirm this result, the location of the estimated optimal points is probed again to see if the experimental results match the estimations.

2.3 Further comments on the methodology

In the model, the function f represents the nominal (and perfect) plane where the experimental points should lie if the plate was perfectly manufactured. The desired shape of f is known beforehand (it is a plane), thus this regression technique is not the only one that can determine the interpolated plane f_0 since a least-squares regression could do this task too. Nevertheless, the traditional non-linear least-squares technique will fail to interpolate to the deviation surface (represented by f_1 here) since no analytical model exists for this surface. In this study on face-milled plates, a smooth translation-invariant kernel function is chosen that is not influenced by brutal local variations, and that defines locally how distances and angles are measured on the interpolated deviation surface. In other words, the model-driven approach of the non-linear least-squares technique has been replaced, which 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 infinitesimal cracks and deformations (that cannot be measured by a probe-type CMM anyway). In cases where the function f does not represent a plane, linear regression is no longer appropriate in order to determine f_0 and this step then has to rely on a non-linear least-squares technique since the desired shape of f is known beforehand. Furthermore, there is a supplementary difficulty to this task: the measurements \mathbf{x}_i

will be slightly translated and rotated in space when compared with the ideal profile of the part. This error is due to imperfections in the alignment and the set-up of the measuring tools. To counter this phenomenon, it is possible to implement a least-squares approach that corrects the misalignment and interpolates the quantities $y_i - f_0(\mathbf{x}_i)$ correctly (see Dowling *et al.* [9] for a discussion of surface alignment techniques). Then, once these quantities are obtained, the treatment is resumed as usual by using the QP formulation to determine the function f_1 .

3 QUADRATIC PROGRAMMING FORMULATIONS FOR NON-LINEAR REGRESSION

3.1 Introduction to support vector machines

Support vector machines are learning algorithms (introduced by Vapnik [10]) that are used in tasks such as statistical classification and regression. After their introduction in 1982, the machine learning community only started to actively investigate SVMs after 1995 as alternatives to artificial neural networks and Bayesian networks [11–13]. SVMs have since been applied to meteorology, computer vision, speech recognition, and cryptography. Despite the introduction of several improvements for building SVMs on large datasets [14, 15], developments stalled and it still remains a computational challenge to efficiently construct SVMs from large batches of data. The main reason for such difficulties is that the mathematical programming problems that need to be solved when SVMs are learning are large QP problems that use interior points methods (IPMs). IPMs are typically computationally demanding for large-scale problems, but some heuristics can drastically reduce computational times and memory requirements. Platt [14] introduced sequential minimal optimization (SMO), which breaks down the SVM problem into smaller auxiliary problems. Each auxiliary problem can be solved analytically, thus reducing the overall computational complexity in most cases. In another approach to circumvent the scalability problems of SVMs, it was chosen to simplify the mathematical programming problems involved in SVR. This leads to simple mathematical programming problems that can be solved using standard and computationally efficient numerical techniques. This is an extension of previous work by Trafalis and Gilbert [16] for robust classification and regression for noisy data.

3.2 Mathematical developments

Consider a set of $\ell \in \mathbb{N}^*$ pairs of observations (\mathbf{x}_i, y_i) with $\mathbf{x}_i \in E \subseteq \mathbb{R}^n$ ($n \in \mathbb{N}^*$) and $y_i \in \mathbb{R}$ for every $i \in [1, \ell]$ where E is a compact subset of \mathbb{R}^n . The independent

observations are distributed according to an *unknown* probability distribution. Each target value y_i is paired with a feature vector \mathbf{x}_i and represents the output of an unknown function $f: E \rightarrow \mathbb{R}$. The regression problem here is to find the function f such that for every $i \in [1, \ell]$, $f(\mathbf{x}_i) = y_i$, and such that $f(\mathbf{x}) \approx y$ for any new observation (\mathbf{x}, y) . In order to match a realistic setting, a set of ℓ slack variables ξ_i , $i \in [1, \ell]$ is introduced such that $f(\mathbf{x}_i) - y_i = \xi_i$. The slack variables are quantities to be minimized in order to solve the regression problem. The sum of the absolute values of the slack variables represents the empirical error for the function f .

In the mathematical developments detailed in the appendix, the regression function shown has the form $f(\mathbf{x}) = \sum_{i=1}^{\ell} \alpha_i k(\mathbf{x}_i, \mathbf{x}) + b$, where $\alpha_1, \dots, \alpha_{\ell}$ and b are solutions of a linear system of equations. The function k , called the *kernel*, has special properties and is constructed based on the characteristics of the regression problem. The kernel used for the applications of this paper was introduced in section 2. The linear system to be solved is $(\mathbf{K}^2 + \mathbf{I}_{\ell}/C + \mathbf{1}_{\ell}\mathbf{1}_{\ell}^T)\mathbf{x} = \mathbf{y}$ where $C > 0$, $\mathbf{y} \in \mathbb{R}^{\ell}$ is a vector made of all y_i for $i \in [1, \ell]$, and $\mathbf{1} \in \mathbb{R}^{\ell}$ is a vector made of ones. The (i, j) th element of the $\ell \times \ell$ Gram matrix \mathbf{K} is made of all the dot products $\langle k(\cdot, \mathbf{x}_i), k(\cdot, \mathbf{x}_j) \rangle = k(\mathbf{x}_i, \mathbf{x}_j)$ for every i and j in $[1, \ell]$. The optimal α and b are given by $\alpha = \mathbf{K}\mathbf{x}$ and $b = \mathbf{1}_{\ell}^T \mathbf{x}$.

4 EXPERIMENTS ON FACE-MILLED PLATES

4.1 Experimental set-up

The QP formulation was tested on a Dell Precision Workstation 530 equipped with two 2.4 GHz Intel Xeon processors and 2 GiB of RAM; the codes were developed under MATLAB 7.4. Two batches of four and five face-milled plates were produced for the experiment. The cutting parameters for the plates were as follows.

1. *Workpiece* aluminium 6061-T6, 10.16 cm \times 10.16 cm \times 1.27 cm.
2. *Tool* $\varnothing 7.62$ cm cutter, seven inserts with carbide coating.
3. *Machining conditions* coolant, cutter speed = 750 m/min, step depth = 0.254 mm, cutting feed = 1.524 m/min (first batch), and 1.778 m/min (second batch).

The plates were inspected with a Brown & Sharpe Microval PFX probe-type CMM equipped with a M3 straight stylus reference A-5000-3552 ($\varnothing 2$ mm ruby ball, $\varnothing 1.4$ mm stainless steel stem, 21 mm overall length, 1 g mass). The plates were labelled from 1–1 to 1–4 for the first batch, and from 2–1 to 2–5 for the

second batch. The plates were visually inspected before being measured and plates 1–1 and 1–2 were discarded owing to faulty machining. The probe-type CMM used in this work does not produce significant quantities of measurements in a short period of time; thus the datasets rarely exceeded 300 points and computational times were extremely small. However, some optical CMMs are capable of producing significant quantities of measurements in a single pass, which leads to longer computational times during the interpolation. The points of contact were chosen such that they formed a uniform mesh on the surface of the plates. The CMM thus traced a zig-zag pattern at a fast pace when taking measurements. Different densities of mesh were used for the same plates in order to compare the accuracy of the results on different meshes. In this study, uniform meshes were chosen in order to maximize the surface information on the whole plate. However, other good choices of contact-point sampling exist. Some of these involve the use of low-discrepancy sequences used in quasi-Monte Carlo methods. In brief, the idea behind using these low-discrepancy sequences is to obtain a discrete sampling such that evaluation of the volume delimited by the surface of interest can be controlled and obtained with a limited number of measurements. The main interest is to be able to extract the maximum amount of information about the shape of the surface with the minimum coordinate evaluations possible. Thus, if the surface of interest is sufficiently smooth, low-discrepancy sequences provide excellent sampling schemes.

4.2 Results and discussion

A test run was made on plate 2–2 with a total of 293 measurements. The parameter σ of the multi-quadric kernel and the trade-off parameter of the optimiza-

tion problems were tuned until the residuals ε_i were small enough and deemed to be random and uncorrelated. Results of the non-linear regressions can be seen in Fig. 1 where the deviation surface has the shape of a saddle.

Note that all the plates that were inspected have similar deviation surfaces. These surfaces are somehow saddle-shaped with the ‘valley’ part oriented along the direction of the cutter pass. More tests on differently manufactured plates will confirm if this is a general behaviour or if this is an artifact produced by the particular machining conditions. If the shape is found to be general, then it will be a precious hint for inspecting plates with optimized meshes since, on a saddle, the extrema are located on the sides. On seven plates, a set of four plates was inspected with different meshes. Using the formulations of section 2, the estimates $d + 2c_a$, with $a = 0.01$, of the variations of the tested plates are shown in Table 1.

The variation spreads for plates 1–4 and 2–3 are consistent with the visual aspect of the plates. Their surfaces appeared to be better finished and hence their flatness values reflected the same. Their variation spreads are about 2.4 times bigger than the accuracy of the probe. At this scale the errors are not negligible with regard to the variations of the deviation surface, and they can be easily witnessed on the representative graphs (see Fig. 2).

Nevertheless, the errors at such scales represent more random perturbations than real deformations on the surface of the plates and the choice of kernel shows some robustness to these perturbations. Thus, realistic and reliable deviation surfaces can be obtained even at scales a few times more than the probe accuracy. Different meshes were also tested on the same plates in order to assess the accuracy of the regression model. The tests performed on plate 1–3

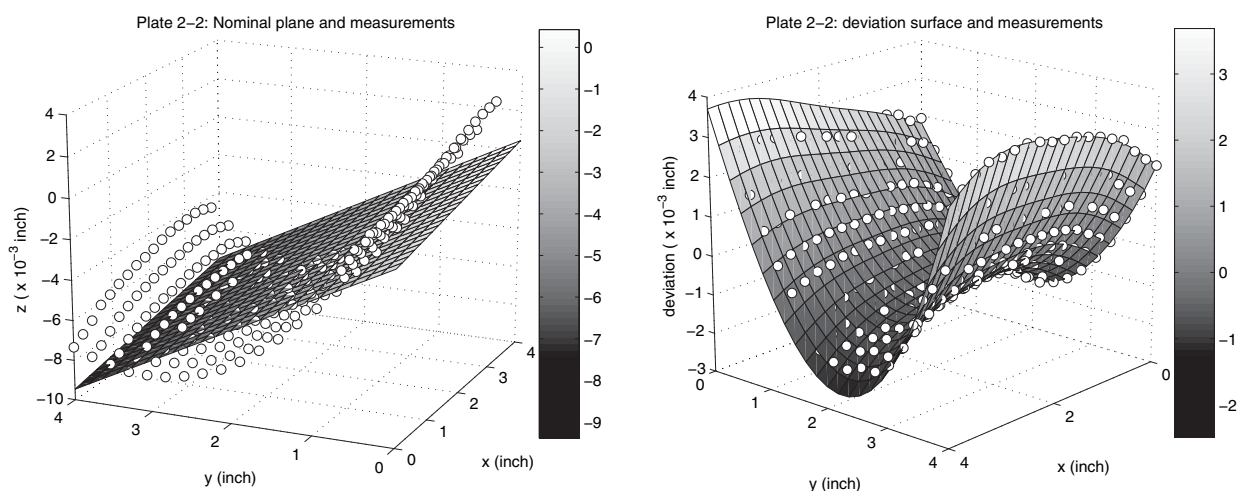


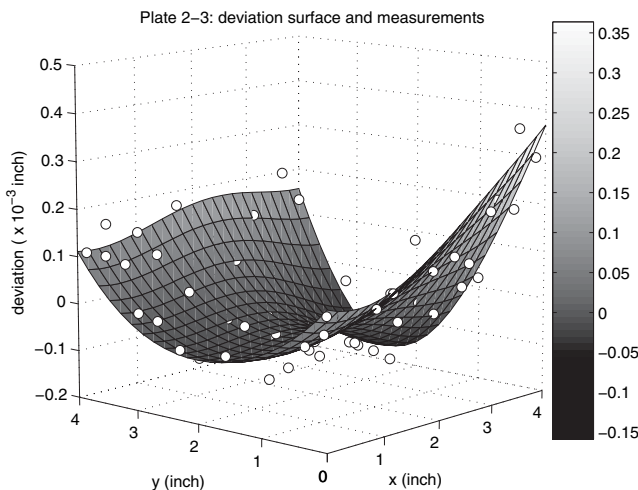
Fig. 1 Interpolated nominal plane of plate 2–2 (left) and interpolated deviation surface of plate 2–2 (right).

Table 1 Variation spreads of the plates

Plate ref.	Number of points	Variation spread (μm)
1-3	348	57.9 ± 3.8
1-4	81	18.3 ± 3.8
2-2	293	166.9 ± 3.8
2-3	64	19.3 ± 3.8

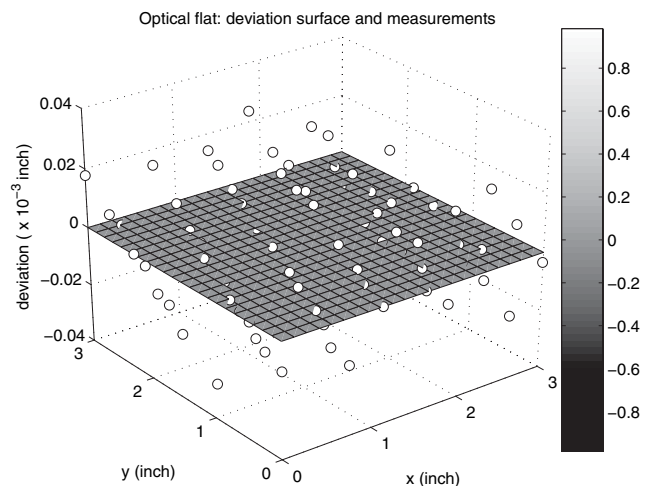
Table 2 Variation spread versus number of points in the mesh

Number of points	Variation spread (μm)
36	58.7 ± 3.8
64	61.2 ± 3.8
81	56.4 ± 3.8
138	61.0 ± 3.8
174	60.2 ± 3.8
202	61.7 ± 3.8
219	58.4 ± 3.8
283	58.4 ± 3.8
348	57.9 ± 3.8

**Fig. 2** Interpolated deviation surface of plate 2-3

are summarized in Table 2, which indicates the estimated variation spread according to the number of points in the mesh.

No significant effects were found between the density of the mesh and the estimated variation spreads for the four tested plates. All the meshes had similar estimated variation spreads whatever the number of points (tests were carried out with 36 points or more). Thus, if this behaviour is found to be general, relatively small uniform meshes could be used for flatness inspection since they reduce time-induced errors during the measurements. Finally, in order to test the robustness to perturbations of this choice of kernel, some measurements were taken on a $7.62\text{ cm} \times 7.62\text{ cm}$ portion of an optical flat ($\phi 12.7\text{ cm}$ Lapmaster) on which the accuracy is certified to one tenth light band. This accuracy is much higher than the probe accuracy of the CMM used in this study ($9\text{ }\mu\text{m}$ on the position of a particular point along each axis) and therefore deviations from the nominal plane numerically obtained by this method are only artifacts and not real deviations. If this choice is robust, then the deviation surface should be a plane passing by the coordinate origin and parallel to the canonical xOy plane. This deviation surface (reproduced in Fig. 3) is at first sight completely flat and inside a cloud of noisy measurements. This test confirms the relative immunity of this choice to errors caused by chance variations.

**Fig. 3** Interpolated deviation surface of an optical flat

5 CONCLUSIONS

This work successfully derived simplified QP formulations using kernel methods for non-linear regression problems. These extremely simple QP problems give general symbolic solutions that can be solved efficiently with very classic numerical techniques. Furthermore, it has been shown that this regression technique can be used to estimate efficiently the flatness of face-milled plates. Preliminary results show that, despite the fact that the plates had different manufacturing specifications, the deviation surfaces are usually saddle-shaped. This observation advocates the use of samplings that judiciously choose the points of measurements as close as possible to the sides of the plates for maxima determination instead of the interior; the minima are more likely to be found in the central region. Uniform sampling with a limited number of initial points of measurements are reasonable. More points may be chosen with a refined search within each of these regions. The plates tested also showed the same amplitude of errors ($3.81\text{ }\mu\text{m}$) for both small uniform samplings (36 points) and large uniform samplings

(up to 348 points). Ultimately, adaptive sampling is the overall objective of this research project and it can be concluded that preliminary results on these sampling schemes give much encouragement. The successful test on an optical flat has shown that the chosen approach can be deemed relatively noise-tolerant.

ACKNOWLEDGEMENTS

The authors wish to acknowledge NSF under grant DMI0427966 for providing financial support for this work. The opinions expressed herein are those of the authors and not necessarily those of the NSF.

© Authors 2010

REFERENCES

- 1 Kim, W.-S. and Raman, S. On the selection of flatness measurement points in coordinate measuring machine inspection. *Int. J. Machine Tools Mf.*, 2000, **40**(3), 427–443.
- 2 Badar, M. A., Raman, S., and Pulat, P. S. Intelligent search-based selection of sample points for straightness and flatness estimation. *J. Mfg Sci. Engng*, 2003, **125**(2), 263–271.
- 3 Badar, M. A., Raman, S., and Pulat, P. S. Experimental verification of manufacturing error pattern and its utilization in form tolerance sampling. *Int. J. of Machine Tools Mf.*, 2005, **45**(1), 63–73.
- 4 Badar, M. A., Raman, S., Pulat, P. S., and Shehab, R. L. Experimental analysis of search-based selection of sample points for straightness and flatness estimation. *J. Mfg Sci. Engng*, 2005, **127**(1), 96–103.
- 5 Raman, S., Basu, A., and Kilic, D. S. Analysis of fractal frictional contact in machining. *Proc. IMechE, Part B: J. Engineering Manufacture*. 2007, **221**(B7), 1119–1128. DOI: 10.1243.09544054JEM740.
- 6 Raman, S. and Chellaperumal, J. Re-examining the machining frictional boundary conditions using fractals. *Proc. IMechE, Part B: J. Engineering Manufacture*, 2002, **216**(B7), 993–997. DOI: 10.1243.09544050260174193.
- 7 Murthy, T. S. R. and Abidin, S. Z. Minimum zone evaluation of surfaces. *Int. J. Machine Tool Des. Res.*, 1980, **20**(2), 123–136.
- 8 Collins, C. E., Fay, E. B., Aguirre-Cruz, J. A., and Raman, S. Alternate methods for sampling in coordinate metrology. *Proc. IMechE, Part B: J. Engineering Manufacture*, 2007, **221**(B6), 1041–1052. DOI: 10.1243.09544054JEM734.
- 9 Dowling, M. M., Griffin, P. M., Tsui, K.-L., and Zhou, C. Statistical issues in geometric feature inspection using coordinate measuring machines. *Technometrics*, 1997, **39**(1), 3–17.
- 10 Vapnik, V. N. *Estimation of dependences based on empirical data*, 1982 (Springer, New York).
- 11 Vapnik, V. N. *The nature of statistical learning theory*, 1995 (Springer, New York).
- 12 Shawe-Taylor, J. and Cristianini, N. *Kernel methods for pattern analysis*, 2004 (Cambridge University Press, Cambridge).
- 13 Schölkopf, B. and Smola, A. J. *Learning with kernels: support vector machines, regularization, optimization, and beyond*, 2002 (MIT Press, Cambridge).
- 14 Platt, J. C. Using sparseness and analytic QP to speed training of support vector machines. *Advances in Neural Information Processing Systems*, Vol. 11 (Eds M. S. Kearns *et al.*), 1999, pp. 557–563 (MIT Press Cambridge, MA).
- 15 Musicant, D. R. and Mangasarian, O. L. Large scale kernel regression via linear programming. *Machine Learning*, 2002, **46**(1–3), 255–269.
- 16 Trafalis, T. B. and Gilbert, R. C. Robust classification and regression using support vector machines. *Eur. J. Opl Res.*, 2006, **173**(3), 893–890.
- 17 Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. *Nonlinear programming: theory and algorithms*, third edition, 2006 (Wiley-Interscience, Hoboken, NJ).

APPENDIX

Quadratic programming formulation

Let k be a kernel and suppose that f belongs to a reproducing kernel hilbert space (RKHS) \mathcal{F} with k as the reproducing kernel. From the reproducing property of a RKHS, for every $i \in [1, \ell]$, $f(\mathbf{x}_i) = \langle k(\cdot, \mathbf{x}_i), f \rangle$. Let \mathbf{K} be the $\ell \times \ell$ Gram matrix \mathbf{K} made of all the dot products $\langle k(\cdot, \mathbf{x}_i), k(\cdot, \mathbf{x}_j) \rangle = k(\mathbf{x}_i, \mathbf{x}_j)$ for every i and j in $[1, \ell]$. The set of constraints for the regression function f is $f(\mathbf{x}_i) + b - y_i = \xi_i$ for all $i \in [1, \ell]$, with $f(\mathbf{x}_i) = \langle k(\cdot, \mathbf{x}_i), f \rangle = \sum_{j=1}^{\ell} \alpha_j \mathbf{K}_{ij}$ and $b \in \mathbb{R}$. In the resulting mathematical programming problem, b , α_j s, and ξ_i s are variables and y_i and \mathbf{K}_{ij} are given. The actual interpolating function then becomes $\tilde{f} = f + b = \sum_{i=1}^{\ell} \alpha_i k(\cdot, \mathbf{x}_i) + b$.

For a probability $p \in [0, 1]$, it is shown [12] that the generalization error of \tilde{f} is bounded with probability p by $2\alpha^T \mathbf{K} \alpha [\text{tr}(\mathbf{K})]^{1/2} / \ell + 2|b| / \sqrt{\ell}$. Therefore, minimizing the empirical error and the generalization error while keeping the regression constraints in check is accomplished by the QP problem

$$\min_{(\alpha, \xi, b) \in \mathbb{R}^{2\ell+1}} \{ \alpha^T \alpha + b^2 + C \xi^T \xi : \mathbf{K} \alpha + b \mathbf{1} - \mathbf{y} = \xi \}, \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^{\ell}$ is a vector made of all y_i for $i \in [1, \ell]$ and $\mathbf{1} \in \mathbb{R}^{\ell}$ is a vector made of ones. This is a form of Tikhonov regularization whose solution is given by solving the linear system $(\mathbf{K}^2 + \mathbf{I}_{\ell}/C + \mathbf{1}_{\ell} \mathbf{1}_{\ell}^T) \mathbf{d} = \mathbf{y}$ [17]. Optimal α and b are derived from the solution \mathbf{d} by $\alpha = \mathbf{K} \mathbf{d}$ and $b = \mathbf{1}_{\ell}^T \mathbf{d}$.