Parameter Detection of Thin Films From Their X-Ray Reflectivity by Support Vector Machines

D. Strauß, G. Steidl, and U. Welzel

262/01

April 2001

Parameter Detection of Thin Films From Their X–Ray Reflectivity by Support Vector Machines

Daniel Strauß and Gabriele Steidl
Faculty of Mathematics and Computer Science
University of Mannheim
D–68131 Mannheim
Germany
strauss@keynumerics.com
steidl@math.uni-mannheim.de

Udo Welzel
Max Planck Institute for Metals Research
D–70174 Stuttgart
Germany
welzel@mf.mpi-stuttgart.mpg.de

Abstract

Reflectivity measurements are used in thin film investigations for determining the density and the thickness of layered structures and the roughness of external and internal surfaces. From the mathematical point of view the deduction of these parameters from a measured reflectivity curve represents an inverse problem. At present, curve fitting procedures, based to a large extent on expert knowledge are commonly used in practice. These techniques suffer from a low degree of automation.

In this paper we present a new approach to the evaluation of reflectivity measurements using support vector machines. For the estimation of the different thin film parameters we provide sparse approximations of vector–valued functions, where we work in parallel on the same data sets. Our support vector machines were trained by simulated reflectivity curves generated by the optical matrix method. The solution of the corresponding quadratic programming problem makes use of the *SVMTorch* algorithm.

We present numerical investigations to assess the performance of our method using models of practical relevance. It is concluded that the approximation by support vector machines represents a very promising tool in X–ray reflectivity investigations and seems also to be applicable for a much broader range of parameter detection problems in X–ray analysis.

1991 *Mathematics Subject Classification.* 49N10, 49N45, 41A63, 41A30.
Key words and phrases. Support vector machines, reproducing kernel Hilbert spaces, radial basis functions, X–ray reflectometry, optical matrix method

1 Introduction

Thin films appear in various fields of technology such as conductor line materials in integrated circuits, diffusion barriers or anticorrosion coatings, antireflection coatings in optics, and magnetooptic storages. Three important parameters for characterizing thin films are the density, the thickness, and the roughness of the surface. The *reflectometry*, i.e., the utilization of the X–ray reflectivity curve obtained at grazing incidences is an established non–destructive method for determining these parameters which is widely used in practical environments. This method involves two types of reflectivity curves. One curve is measured by hardware, see Figure 1, mainly build on the basis of conventional powder diffractometers and the other one is simulated by a physical model using a set of assumed model parameters.

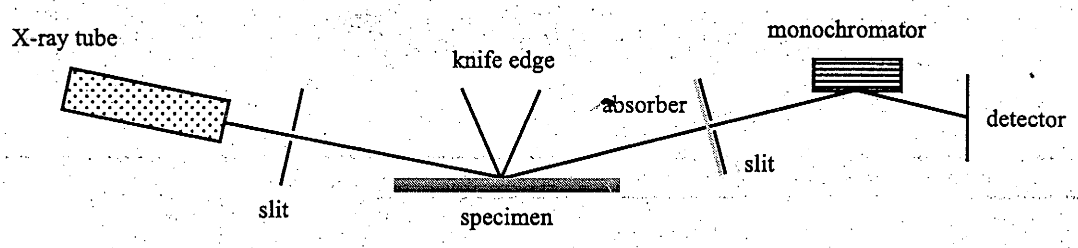

Figure 1: Setup for X–ray reflectivity measurements

Up to now, the measured and the simulated curves are fitted in an interactive trial and error procedure of changing the model parameters and comparing the concurrence of the curves, see [20]. This procedure is mainly based on expert knowledge and suffers, in general, from a low degree of automation.
In this paper we present a new approach to the evaluation of reflectivity measurements by means of Support Vector Machines (SVMs). SVMs were recently introduced by Vapnik [22] in statistical learning theory and have found wide applications in machine learning tasks such as regression, classification and novelty detection. In contrast to other multivariate approximation schemes such as feed forward backpropagation networks (FFBNs), the quadratic programming (QP) problem raised in the SVM approach guarantees a global solution. Moreover, it leads in general to a sparse approximation of the unknown function.
Based on SVMs the inverse problem of thin film parameter deduction is solved by the sparse approximation of a vector–valued function mapping the reflectivity curve directly onto the parameter set. Our SVMs work in parallel on the same data. The major advantage of our method is that it offers the possibility for an automation of the evaluation of reflectivity curves. Expert intervention is only involved for determining a few parameters for the raised QP problems. For routine applications, we have only a limited number of possible sample constitutions which have to be analyzed. Thus, the QP problems must be solved only once for a particular specimen constitution and the results can be stored for subsequent analysis. The training set for our SVMs is provided by simulated reflectivity curves using the *Optical Matrix Method* (OMM) including the effect of surface roughness. Thus, we are independent from measured data and can generate a large set of training associations. This results in large QP problems. For the solution of these problems we apply the recently developed *SVMTorch* algorithm [3, 4] which is based on several previous papers [15, 17, 6, 8].
In particular, we investigate a three–layer and four–layer model based on practical samples. We show that our method provides a good approximation of the underlying mapping.

This paper is organized as follows: In Section 2 we introduce the OMM which will provide our training set of associations. Section 3 deals with the SVM approach with respect to our setting. In Section 4 we present some numerical investigations showing the performance of our scheme. Conclusions of the paper are given in Section 5.

2 The Optical Matrix Method

The OMM is an established technique to model the reflectivity of thin films. The method goes back to Kiessig [9] who investigated the dispersion of X–rays of different wavelength in thin nickel films and showed that X–rays can be treated similar to the reflection of visible light. It was generalized by Parratt [16] who extended the results of Kiessig for multilayer

packages. In the following we introduce the OMM with a further extension by including the surface roughness according to Névot and Croce [14, 23].
Let us consider the reflection of X–rays at an interface of two media first. This can be described by the model of a planar electromagnetic wave hitting an ideal interface (mathematical plane). See Figure 2.

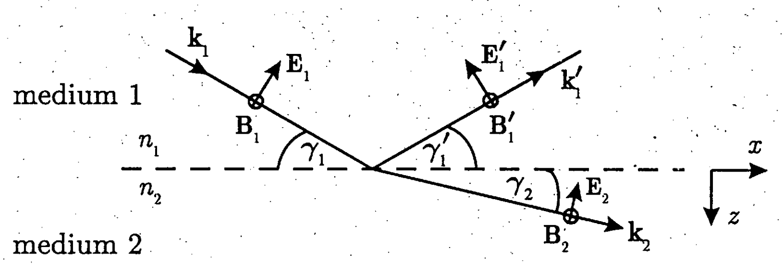

Figure 2: Refraction of a planar electromagnetic wave at an interface between two uniform homogeneous media (σ–polarization) where $\mathbf{k}_j$: wave vector, $\mathbf{E}_j$: electric field vector, $\mathbf{B}_j$: magnetic field vector, n_j: refractive index of medium j, γ_j: angle between interface and wave vector $\mathbf{k}_j$ $(j = 1, 2)$. The reflected vectors with their corresponding angles are prime marked.

Crossing the interface between the media, the X–rays are refracted according to *Snell's law*

$$\frac{\cos \gamma_1}{\cos \gamma_2} = \frac{n_2}{n_1}, \tag{1}$$

where n_j denotes the *refractive index* of medium j and γ_j the angle between the interface and the wave vector $\mathbf{k}_j$ $(j = 1, 2)$.
For electromagnetic radiation belonging to the X–ray range, the refractive index n in matter is smaller than 1 and can be expressed as

$$n = 1 - \delta - i\beta. \tag{2}$$

Here δ and β are the *dispersive correction* and the *absorptive correction*, respectively. Typical values are $\delta \approx 10^{-5}$ and $\beta \approx 10^{-7}$. These corrections are proportional to the mass density ρ of the medium.
If the angle γ_2 becomes zero, then the beam is totally reflected and medium 2 behaves like a perfect mirror. The corresponding angle γ_1 is called the *critical angle* γ_c and we have that $\cos \gamma_c = n_2/n_1$. See also the upper picture of Figure 4. If we consider the transition from vacuum $(n_1 = 1)$ to matter $(n_2 < 1)$ and neglect the absorptive correction β, then we obtain by (2) that $\cos \gamma_c \approx 1 - \gamma_c^2/2 = 1 - \delta_2$, i.e., $\gamma_c \approx \sqrt{2\delta_2}$. Thus, given γ_c, we can determine the refractive index of the medium and the mass density, respectively.
The intensities of reflected and refracted electromagnetic waves at an ideal interface are described by the *Fresnel equations* in classical electrodynamics, cf. [7]. At grazing incidence (small angles of γ_1) the polarization plays no role and we can turn to a scalar consideration. If $\hat{E}$ denotes the amplitude of the electric field, the Fresnel *reflection coefficient* r_F and the *transmission coefficient* t_F are given by

$$r_F = \frac{\hat{E}_1'}{\hat{E}_1} = \frac{\gamma_1 - \gamma_2}{\gamma_1 + \gamma_2}, \tag{3}$$

$$t_F = \frac{\hat{E}_2}{\hat{E}_1} = \frac{2\gamma_1}{\gamma_1 + \gamma_2}. \tag{4}$$

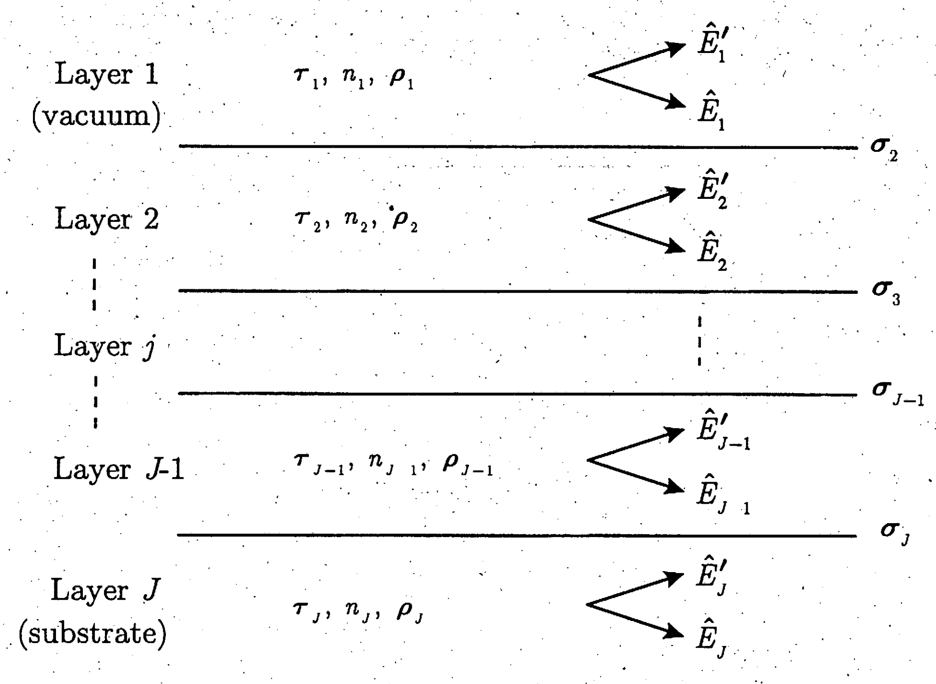

Figure 3: Multilayer consisting of J layers. $\hat{E}_j$ $(j = 1, 2, \ldots, J)$ represents the amplitude of electrical field in the middle of layer j.

The *reflectivity* ν is finally defined as squared ratio of the reflected and incident field amplitudes, i.e., $\nu = r_F^2$. Note, that with this definition it also holds that $\nu = I_1'/I_1$, where I_1' represents the reflected and I_1 the incident intensity.
A main application of the X–ray reflectometry is the characterization of multilayer packages on substrate. In the following, we consider a multilayer package consisting of J layers. Here the first layer represents the vacuum and the last layer is the substrate. These layers are characterized by their refractive index n_j, their *thickness* τ_j, their *mass density* ρ_j and by the *roughness* σ_j of the interface between consecutive layers j and $j+1$, see Figure 3. Note, that we included the surface / interface roughness which is, in short, the standard deviation from the mean height of a rough surface. As described, we have a transmission and reflection of the incident beam above some angle γ_c at an interface. Since the reflected beams are coherent, they interfere and modulate the reflectivity ν of the multilayer package as a function of the incidence angle $\gamma = \gamma_1$ in a characteristic manner. See [23] for detailed treatments.
Again, we have by Snell's relation (1) that

$$\frac{\cos\gamma_j}{\cos\gamma} = \frac{n_1}{n_j}$$

so that the angles γ_j are determined by the incidence angle and by the refractive indices of the media.
Given the above parameters of the layers, the reflectivity $\nu(\gamma)$ of the whole multilayer package can be calculated by the OMM:
Let k_0 denote the absolute value of the vacuum wave vector. Then the relation between the amplitudes $\hat{E}_j$, $\hat{E}'_j$ and $\hat{E}_{j+1}$, $\hat{E}'_{j+1}$ in the middle of the j–th and $(j+1)$–th layer, respectively, reads

$$\begin{pmatrix} \hat{E}_j \\ \hat{E}'_j \end{pmatrix} = \mathbf{R}^{(j,j+1)} \begin{pmatrix} \hat{E}_{j+1} \\ \hat{E}'_{j+1} \end{pmatrix}, \tag{5}$$

where the entries of the transition matrix $\mathbf{R}^{(j,j+1)}$ are given by [23]

$$\begin{aligned}
R_{11}^{(j,j+1)} &= \frac{\gamma_j+\gamma_{j+1}}{2\gamma_j}\; e^{-\frac{1}{2}k_0^2(\gamma_j-\gamma_{j+1})^2\sigma_{j+1}^2}\; e^{-\mathrm{i}\frac{k_0}{2}(\gamma_j\tau_j+\gamma_{j+1}\tau_{j+1})}\,,\\
R_{12}^{(j,j+1)} &= \frac{\gamma_j-\gamma_{j+1}}{2\gamma_j}\; e^{-\frac{1}{2}k_0^2(\gamma_j+\gamma_{j+1})^2\sigma_{j+1}^2}\; e^{-\mathrm{i}\frac{k_0}{2}(\gamma_j\tau_j-\gamma_{j+1}\tau_{j+1})}\,,\\
R_{21}^{(j,j+1)} &= \frac{\gamma_j-\gamma_{j+1}}{2\gamma_j}\; e^{-\frac{1}{2}k_0^2(\gamma_j+\gamma_{j+1})^2\sigma_{j+1}^2}\; e^{\mathrm{i}\frac{k_0}{2}(\gamma_j\tau_j-\gamma_{j+1}\tau_{j+1})}\,,\\
R_{22}^{(j,j+1)} &= \frac{\gamma_j+\gamma_{j+1}}{2\gamma_j}\; e^{-\frac{1}{2}k_0^2(\gamma_j-\gamma_{j+1})^2\sigma_{j+1}^2}\; e^{\mathrm{i}\frac{k_0}{2}(\gamma_j\tau_j+\gamma_{j+1}\tau_{j+1})}\,.
\end{aligned}$$

The first factors on the right–hand side of the above equations stem from the Fresnel equations (3), (4). The exponential terms in the middle represent the damping due to the interface roughness. The last terms carry the shifts in phase, depending on the thickness of the layer. They mainly describe the interference of the rays reflected at the various interfaces. The substrate is considered as infinitely thick, i.e., $\hat{E}'_J$ equals zero. Now successive application of (5) yields for the amplitudes in the vacuum

$$\begin{pmatrix} \hat{E}_1 \\ \hat{E}'_1 \end{pmatrix} = \mathbf{R}^{(1,2)}\mathbf{R}^{(2,3)}\cdots\mathbf{R}^{(J-1,J)}\begin{pmatrix} \hat{E}_J \\ 0 \end{pmatrix}. \tag{6}$$

Finally, the reflectivity of the whole multilayer package can be obtained by

$$\nu = \left(\frac{\hat{E}'_1}{\hat{E}_1}\right)^2. \tag{7}$$

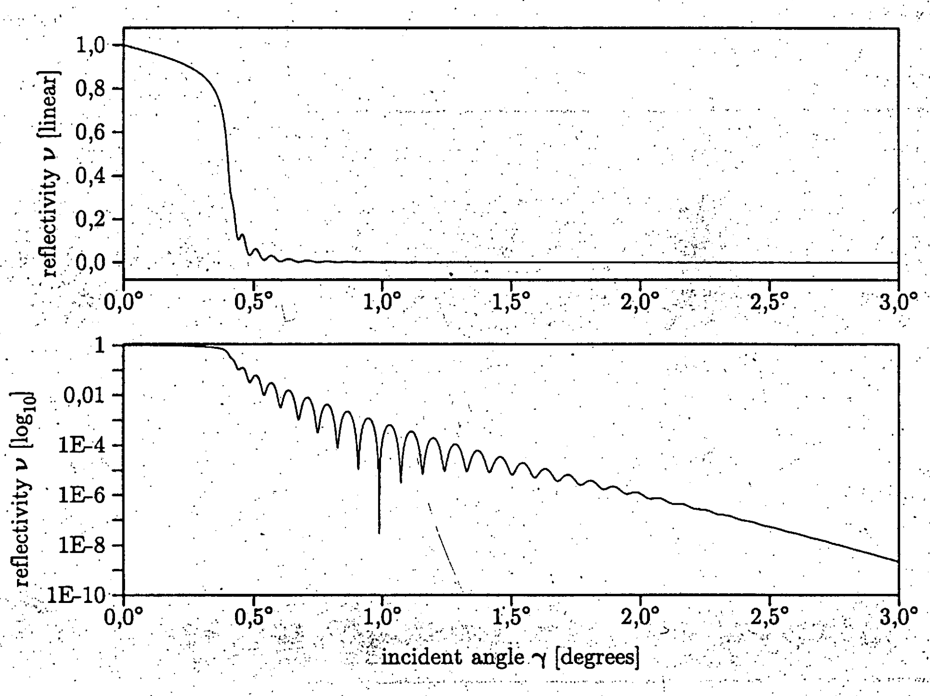

Figure 4: An exemplary reflectivity curve for $J = 4$ simulated by the OMM.

Figure 4 shows the reflectivity $\nu = \nu(\gamma)$ ($\gamma \in [0°, 3°]$) simulated by the above OMM for a fixed multilayer package consisting of vacuum, molybdenum, silicon oxide, silicon substrate, i.e., $J = 4$. Note that there is no abrupt cross–over from total reflection to transition. This is due to the absorption which smears an abrupt change. Thus, an angle γ_c can hardly be defined in presence of strong absorption. Without absorption, the reflectivity would be 1 below a critical angle γ_c. More information about the morphological analysis of reflectivity curves can be found in [11].

3 The Support Vector Machine Approach

In this section, we introduce the SVM approach with respect to our problem. For a more detailed treatment of SVMs we refer to standard literature on this topic, e.g., [22].
As in the previous section we consider a multilayer package consisting of J layers. We are interested in determining the thickness τ_j, the mass density ρ_j and the roughness σ_j ($j = 2, \ldots, J-1$) from the reflectivity $\nu^m = \nu^m(\gamma)$ measured for different incidence angles $\gamma \in [0, \kappa]$. Note that we have indeed only $J-2$ layers of interest since the parameters of vacuum and substrate are known. Let $L = 3(J-2)$. Set $\boldsymbol{\tau} = (\tau_2, \ldots, \tau_{J-1})^T$, $\boldsymbol{\rho} = (\rho_2, \ldots, \rho_{J-1})^T$ and $\boldsymbol{\sigma} = (\sigma_2, \ldots, \sigma_{J-1})^T$. For $\gamma_k = \frac{\kappa k}{N-1}$ ($k = 0, \ldots, N-1$), let $\boldsymbol{\nu}^m = (\nu^m(\gamma_0), \ldots, \nu^m(\gamma_{N-1}))^T$.
Up to now the following time consuming interactive trial and error technique was mainly used to solve the above problem: Choose $\boldsymbol{\tau}$, $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ and compute $\boldsymbol{\nu} : \mathbb{R}^L \to \mathbb{R}^N$

$$\boldsymbol{\nu} = \boldsymbol{\nu}(\boldsymbol{\tau}, \boldsymbol{\rho}, \boldsymbol{\sigma}) = (\nu(\gamma_k; \boldsymbol{\tau}, \boldsymbol{\rho}, \boldsymbol{\sigma}))_{k=0}^{N-1\,T} \tag{8}$$

by the OMM. Compare $\boldsymbol{\nu}$ and $\boldsymbol{\nu}^m$. If $\boldsymbol{\nu}$ is a "good" approximation of the measured vector $\boldsymbol{\nu}^m$, than associate the parameters $\boldsymbol{\tau}$, $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ with the multilayer package, otherwise select other parameters and repeat the procedure. Unfortunately, this technique is to a large extent based on expert knowledge since fitting algorithms can only be used for a refinement of the curve fitting [20]. Thus, this technique suffers from a low degree of automation and can be time consuming.
In the following, we propose an approach by SVMs which seems to be superior to other possible automation methods, e.g., FFBNs [19], for our purposes. FFBNs were already used to solve inverse problems in X–ray analysis, e.g., Long et al. [13] applied FFBNs for the identification of fluorescence spectra and Wern and Ringeisen [25] used them for the evaluation of residual strain/stress gradients from X–ray diffraction data. However, these networks suffer from two major drawbacks: they can be trapped into local minima during learning and their architecture must be determined empirically.
In contrast to FFBNs the SVM complexity depends on the data. There are only a few parameters to adjust. Training a SVM requires the solution of a QP problem which yields a global solution. Furthermore, training a SVM does not depend directly on the dimensionality of the input space. In general, SVMs provide a sparse approximation of the unknown function so that we can efficiently evaluate the approximate function. Due to the flexible kernel substitution, a variety of approximation schemes can be implemented by SVMs.
Assume that we are given a set of M associations

$$\{(\boldsymbol{\nu}_i, \mathbf{p}_i) \in \mathbb{R}^N \times \mathbb{R}^L : i = 1, \ldots, M\}$$

where $\mathbf{p}_i = (\boldsymbol{\tau}_i, \boldsymbol{\rho}_i, \boldsymbol{\sigma}_i)$ and $\boldsymbol{\nu}_i = (\nu(\gamma_1; \mathbf{p}_i), \ldots, \nu(\gamma_{N-1}; \mathbf{p}_i))^T$. Note that we can provide a large number of associations by using the OMM. We are interested in a function $\mathbf{F} : \mathbb{R}^N \to \mathbb{R}^L$ so that $\mathbf{F}(\boldsymbol{\nu}_i)$ approximates $\mathbf{p}_i$ ($i = 1, \ldots, M$), i.e, we want to approximate the inverse of $\boldsymbol{\nu}$ in (8). We intend to determine the functions F_l ($l = 1, \ldots, L$) of the vector–valued function $\mathbf{F}$ simultaneously.
To avoid multiindices, we fix $l \in \{1, \ldots, L\}$ in the following and set

$$f(\boldsymbol{\nu}) = F_l(\boldsymbol{\nu}), \; y_i = p_{i,l}.$$

Our SVM introduction follows mainly the lines of Wahba [24].
Let $K(\cdot,\cdot)$ be a positive definite function on $\mathbb{R}^N \times \mathbb{R}^N$ and let $\mathcal{H}_K$ denote the reproducing kernel Hilbert space (RKHS) with reproducing kernel K. For more information on RKHS see [1]. Suppose that we are given a set of training data $(\boldsymbol{\nu}_i, y_i)$ $(i = 1, \ldots, M)$. Set $\mathbf{f} = (f_1, \ldots, f_M)^T$, where $f_i = f(\boldsymbol{\nu}_i)$.
We are interested in finding a function $f = f_\lambda$ of the form $h + d$ $(h \in \mathcal{H}_K, d \in \mathbb{R})$ which minimizes

$$\lambda \sum_{i=1}^{M} V_\epsilon(y_i - f_i) + \frac{1}{2} ||h||^2_{\mathcal{H}_K}, \tag{9}$$

where

$$V_\epsilon(x) = \max\{0, |x| - \epsilon\}$$

denotes Vapnik's *ϵ–insensitive loss function* [22]. By the Representer Theorem [10, 24] the minimizer of (9) can be written in the form

$$f(\boldsymbol{\nu}) = \sum_{j=1}^{M} c_j K(\boldsymbol{\nu}, \boldsymbol{\nu}_j) + d \tag{10}$$

so that

$$\mathbf{f} = \mathbf{Kc} + d\mathbf{e}. \tag{11}$$

Here $\mathbf{K} = (K(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j))_{i,j=1}^M$, $\mathbf{c} = (c_1, \ldots, c_M)^T$ and $\mathbf{e}$ denotes the vector with M entries 1. Using this notation we are looking for $\mathbf{c} \in \mathbb{R}^M$ and $d \in \mathbb{R}$ minimizing

$$\lambda \sum_{i=1}^{M} V_\epsilon(y_i - f_i) + \frac{1}{2} \mathbf{c}^T \mathbf{Kc}.$$

This is equivalent to the following constraint optimization problem

$$\min_{\mathbf{c},d,\mathbf{u},\mathbf{u}^*} \lambda\,(\mathbf{e}^T\mathbf{u} + \mathbf{e}^T\mathbf{u}^*) + \frac{1}{2}\mathbf{c}^T\mathbf{Kc} \tag{12}$$

subject to

$$\begin{aligned} \mathbf{u} &\geq \mathbf{0},\ \mathbf{u}^* \geq \mathbf{0}, \\ \mathbf{y} - \mathbf{Kc} - d\mathbf{e} &\leq \epsilon\mathbf{e} + \mathbf{u}, \\ -\mathbf{y} + \mathbf{Kc} + d\mathbf{e} &\leq \epsilon\mathbf{e} + \mathbf{u}^*. \end{aligned}$$

The dual problem with Lagrange multipliers $\boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}, \boldsymbol{\beta}^*$ reads

$$\max_{\mathbf{c},d,\mathbf{u},\mathbf{u}^*,\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\beta},\boldsymbol{\beta}^*} L(\mathbf{c}, d, \mathbf{u}, \mathbf{u}^*, \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}, \boldsymbol{\beta}^*)$$

$$\begin{aligned} L(\mathbf{c}, d, \mathbf{u}, \mathbf{u}^*, \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}, \boldsymbol{\beta}^*) &= \lambda(\mathbf{e}^T\mathbf{u} + \mathbf{e}^T\mathbf{u}^*) + \frac{1}{2}\mathbf{c}^T\mathbf{Kc} - \boldsymbol{\beta}^T\mathbf{u} - \boldsymbol{\beta}^{*T}\mathbf{u}^* \\ &\quad -\boldsymbol{\alpha}^T(\epsilon\mathbf{e} + \mathbf{u} - \mathbf{y} + \mathbf{Kc} + d\mathbf{e}) - \boldsymbol{\alpha}^{*T}(\epsilon\mathbf{e} + \mathbf{u}^* + \mathbf{y}^* - \mathbf{Kc} - d\mathbf{e}) \end{aligned}$$

subject to

$$\frac{\partial L}{\partial \mathbf{c}} = 0, \; \frac{\partial L}{\partial \mathbf{u}} = 0, \; \frac{\partial L}{\partial \mathbf{u}^*} = 0, \; \frac{\partial L}{\partial d} = 0, \tag{13}$$
$$\boldsymbol{\alpha} \geq 0, \; \boldsymbol{\alpha}^* \geq 0, \; \boldsymbol{\beta} \geq 0, \; \boldsymbol{\beta}^* \geq 0.$$

Now $\mathbf{0} = \frac{\partial L}{\partial \mathbf{c}} = \mathbf{K}\mathbf{c} - \mathbf{K}\boldsymbol{\alpha} + \mathbf{K}\boldsymbol{\alpha}^*$ implies that

$$\mathbf{c} = \boldsymbol{\alpha} - \boldsymbol{\alpha}^*.$$

Further, by $\frac{\partial L}{\partial \mathbf{u}} = \mathbf{0}$ and $\frac{\partial L}{\partial \mathbf{u}^*} = \mathbf{0}$ it follows $\boldsymbol{\beta} = \lambda \mathbf{e} - \boldsymbol{\alpha}$ and $\boldsymbol{\beta}^* = \lambda \mathbf{e} - \boldsymbol{\alpha}^*$, respectively. Finally, $\frac{\partial L}{\partial d} = 0$ can be rewritten as $\mathbf{e}^T(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = 0$. Then the above optimization problem becomes

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} -\frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T \mathbf{K} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) - \epsilon \mathbf{e}^T(\boldsymbol{\alpha} + \boldsymbol{\alpha}^*) + \mathbf{y}^T(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) \tag{14}$$

subject to

$$\mathbf{e}^T(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = 0,$$
$$\mathbf{0} \leq \boldsymbol{\alpha}, \boldsymbol{\alpha}^* \leq \lambda \mathbf{e}.$$

This QP problem is usually solved in SVM literature. It requires resources of order M^2. Thus, it can be very challenging for standard QP–routines if M becomes large. On the other hand, the set of training associations should be large to provide a dense sampling of the unknown function. Recently, the so–called *SVMTorch* algorithm has been introduced by Collobert and Bengio [3, 4] for solving large scale problems. Based on an idea in [15], in every iteration step of *SVMTorch* a small subset of variables is selected as working set and the QP problem is solved with respect to this working set. If the working set consists only of two variables, the partial QP problems can be solved analytically. Working sets of two variables were also used for classification tasks in the so–called Sequential Minimal Optimization [17] and for regression in [6]. These working sets often imply a faster convergence of the QP algorithm than larger sets [4]. The decision rule for the choice of the working set goes back to [26] and was used in [8] for classification problems. Furthermore, a shrinking phase is used to exclude variables that are stuck to 0 or λ for a longer phase of iterations so that these variables will probably not change anymore. These variables can be removed from the optimization problem such that a more efficient overall optimization is obtained. If no shrinking is used, the convergence of the *SVMTorch* algorithm was proved in [2] for a working set of size two and for an arbitrary working set in [12] under some restrictions.
Once we have computed $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$, we obtain the function

$$f(\boldsymbol{\nu}) = \sum_{j=1}^{M} K(\boldsymbol{\nu}, \boldsymbol{\nu}_j)(\alpha_j - \alpha_j^*) + d. \tag{15}$$

The support vectors are those $K(\cdot, \boldsymbol{\nu}_j)$ for which $\alpha_j - \alpha_j^* \neq 0$, i.e., since $\alpha_j \alpha_j^* = 0$ $(i = 1, \ldots, M)$, those for which $\alpha_j > 0$ or $\alpha_j^* > 0$. Only the summands in (15) including support vectors do not vanish.
With respect to the computation of the constant d we notice the following: The Kuhn–Tucker conditions in (12) are satisfied by

$$\begin{aligned} \alpha_i(\epsilon + u_k - y_k + f_k) &= 0, \\ \alpha_i^*(\epsilon + u_k^* + y_k - f_k) &= 0, \\ (\lambda - \alpha_i)u_i &= 0, \\ (\lambda - \alpha_i^*)u_i^* &= 0. \end{aligned}$$

Thus, we have for $0 < \alpha_i < \lambda$ that $u_i = 0$ and consequently that $f_i = y_i - \epsilon$. By (15) we obtain

$$f_i = \sum_{j=1}^{M} K(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j)(\alpha_j - \alpha_j^*) + d = y_i - \epsilon\,,$$

which implies $d = y_i - \epsilon - \sum_{j=1}^{M} K(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j)(\alpha_j - \alpha_j^*)$.

4 Numerical Investigation

In this section we present some numerical investigations for assessing the performance of our SVM approach. First of all, we emphasize that the constitution of the specimen to be analyzed is known a priori. Thus, we know the bulk values of the mass densities. The thickness and roughness depend on the production process and lower and upper limits are also known such that the physical domain of admissible parameters can be bounded prior the investigation. In other words, for a given specimen the ranges of F_l ($l = 1, \ldots, L$) are bounded intervals $\mathcal{I}_l = [a_l, b_l]$, where $a_l, b_l \in \mathbb{R}$. Of course, tight bounds lead to a problem that is much easier to treat. A specimen independent approximation seems to be infeasible since the range of physically admissible values becomes too large.

The accuracy of approximation can be slacked by the insensitivity ϵ_l for the individual parameter since a perfect match between the physical specimen parameters and the ones deduced from the OMM simulation can not be achieved in practice due to measurement inaccuracies and discrepancies from theoretical model assumptions. Unfortunately, such effects are not given quantitatively so far and recent results on the choice of ϵ_l, e.g., based on noise models [18], cannot be applied here. Therefore, the insensitivity can only be estimated by expert experience.

With respect to our (ideal) synthetic data we choose a very large constant λ which approximates infinity. In this way, we obtain a vector–valued function $\mathbf{F}$ with elements $F_l = h_l + d_l$ ($h_l \in \mathcal{H}_K$; $d_l \in \mathbb{R}$; $l = 1, \ldots, L$) having at most a deviation of ϵ_l from the target film parameters of the simulated curve. Note that ϵ_l heavily determines the degree of sparsity of the representation of F_l.

Another issue is the choice of the reproducing kernel $K(\cdot,\cdot)$. Here we follow the proposal of Smola and Schölkopf [21] to use Gaussian kernels, i.e., $K(\mathbf{x},\mathbf{y}) = e^{-\frac{1}{s^2}\|\mathbf{x}-\mathbf{y}\|_2^2}$ if there only exists a general smoothness assumption about the mapping. However, Gaussian kernels involve the Euclidean distance between the morphological features of two distinct curves. Due to the characteristic cross–over from total reflection to penetration in reflectivity curves, this distance measure is highly sensitive to morphological dissimilarities near the critical angle. On the other hand, dissimilarities for larger incident angles do nearly not influence the evaluation although they are not necessarily of minor importance. For weighting the morphological features more balanced, we work with $\sqrt{\nu}$, i.e., with the Fresnel reflection coefficient r_F (3) instead of the reflectivity. Hence the kernel evaluation becomes

$$K(\boldsymbol{\nu}, \boldsymbol{\nu}_j) = e^{-\frac{1}{s^2}\sum_{k=0}^{N-1}\left(\sqrt{\nu_k}-\sqrt{\nu_{k,j}}\right)^2}. \tag{16}$$

The constant s is a free parameter and must be determined empirically. Here we make use of the fact that small values of s lead to a fast convergence of the algorithm but result in an overfitting. Cristianini et al. [5] used this fact for dynamically adapting s during SVM learning for classification tasks. We begin with small values and then successively increase s until a satisfactory result is obtained on a test set separated from the learning set of association.

For our investigation, let us first consider a model with $J = 3$ layers consisting of a molybdenum film between vacuum and silicon substrate with $\rho_3 = 2.2\text{g/cm}^3$ and $\sigma_3 = 7\text{Å}$. We use a training set of $M = 5000$ associations $\{(\boldsymbol{\nu}_i, \mathbf{p}_i) \in R^N \times R^3 : i = 1, \dots, M\}$ provided by OMM simulations $\boldsymbol{\nu}_i$ with $\kappa = 2$, $N = 1000$, and uniformly distributed random numbers as model parameters $p_{i,l} \in \mathcal{I}_l$ $(l = 1, \dots, L)$. The resulting QP problems are solved by employing the *SVMTorch* method sketched in the previous section with a working set of size two. Note that shrinking can significantly speed up the calculation. The price we have to pay is the uncertainty whether the algorithm converges to the desired solution or not. Therefore, if shrinking is used the results should be controlled on the training set. In our numerical experiments it is controlled that shrinking does not affect the results, i.e., the error on the training set is within the predefined ϵ_l bound.
For assessing the generalization performance of our scheme and the quality of our approximation we use an independent test set $\{(\tilde{\boldsymbol{\nu}}_i, \tilde{\mathbf{p}}_i) \in R^N \times R^3 : i = 1, \dots, T\}$ of $T = 10000$ associations generated with uniformly distributed random numbers $\tilde{p}_{i,l} \in \mathcal{I}_l$ as model parameters and the corresponding OMM simulations $\tilde{\boldsymbol{\nu}}_i$, where again $\kappa = 2$ and $N = 1000$. Let us introduce the following error notation with respect to ϵ_l

$$\eta_{i,l} = \max\{0, |F_l(\tilde{\boldsymbol{\nu}}_i) - \tilde{p}_{i,l}| - \epsilon_l\} \quad (i = 1, \dots, T)$$

with mean

$$\overline{\eta}_l = \frac{1}{T}\sum_{i=1}^{T} \eta_{i,l}$$

and maximum

$$\hat{\eta}_l = \max_{i=1,\dots,T}\{\eta_{i,l}\}.$$

The results as well as the a priori given interval $\mathcal{I}_l$, the insensitivity ϵ_l, and the number of support vectors (NSV) are given in Table 1. For the density, the interval is given by $a_2 =$ 0.7bulk and $b_2 =$ bulk. As noticeable, $\overline{\eta}_l$ is small and also $\hat{\eta}_l$ is within tolerable bounds with respect to the range $b_l - a_l$. Thus, we have indeed found a function $\mathbf{F}$ which reflects well the dependency of the thin film parameters on the corresponding reflectivity curve simulated by the OMM. Note, that there is great variance in the NSVs which indicates how the complexity of the SVMs is individually adapted to the particular mappings F_l $(l = 1, \dots, L)$. Especially, the mass density of the first film can be represented by a simple model due to its direct relation to the cross–over from total reflection to penetration, i.e., the most significant morphological feature of the curve.

parameter	a_l	b_l	ϵ_l	$\overline{\eta}_l$	STD	$\hat{\eta}_l$	NSV
τ_2 [Å]	502	754	5.0	0.07	0.28	6.41	782
ρ_2 [g/cm^3]	7.14	10.2	0.1	$8 \cdot 10^{-5}$	$5 \cdot 10^{-4}$	0.007	24
σ_2 [Å]	0	10.0	0.2	0.067	0.12	1.22	801

Table 1: Results for an independent random test set of $T = 10000$ reflectivity curves for a model with $J = 3$ layers. Here the mean $\overline{\eta}_l$ is given with the standard deviation (STD). Note, that thickness and roughness is given in Ångström where 1Å=10^{-10}m.

A specimen consisting of the layers described above was also investigated by using the Siemens D500 X–ray diffractometer equipped with a knife edge for reflectivity measurements. The setup is shown schematically in Figure 1.
The resulting reflectivity curve $\boldsymbol{\nu}^{(m)}$ is shown in Figure 5 by the scattered points. Here we plotted $r_F^{(m)} = \sqrt{\nu^{(m)}}$ since this information is evaluated by the SVMs with Gaussian kernel due to (16). The evaluation of our computed function $\mathbf{F}$ for this curve yields

$$\mathbf{F}(\boldsymbol{\nu}^{(m)}) = \left(631\text{Å}, 8.60\text{g/cm}^3, 7.89\text{Å}\right)^T.$$

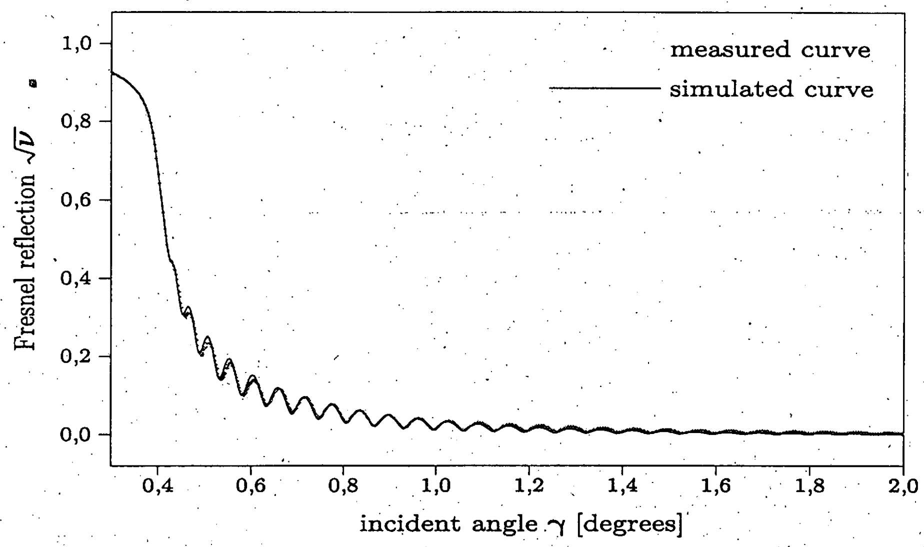

Figure 5: Comparison of a measured reflectivity curve and corresponding OMM simulation using the map $\mathbf{F}$.

Substituting this results in the OMM, the solid curve in Figure 5 is obtained. As noticeable, the measured and simulated curves offer a high degree of concurrence.

Let us now consider a model with $J = 4$ layers consisting of a metastable solution of oxygen in molybdenum (second layer) and a silicon oxide film (third layer) between vacuum and silicon substrate with $\rho_4 = 2.32\text{g/cm}^3$ and $\sigma_4 = 10\text{Å}$. For instance, such layers are used for realizing diffusion barriers. Here we stick to the very same settings described above for generating the training and test set, respectively, which allow us to compare the results. To be more precise, we have a training set of $M = 5000$ associations $\{(\boldsymbol{\nu}_i, \mathbf{p}_i) \in R^N \times R^6 : i = 1, \ldots, M\}$ provided by OMM simulations $\boldsymbol{\nu}_i$ with $\kappa = 2$, $N = 1000$, and uniformly distributed random numbers as model parameters $p_{i,l} \in \mathcal{I}_l$ $(l = 1, \ldots, L)$ and a corresponding independent test set $\{(\tilde{\boldsymbol{\nu}}_i, \tilde{\mathbf{p}}_i) \in R^N \times R^6 : i = 1, \ldots, T\}$ of $T = 10000$ associations. For the density we have again that $a_l = 0.7\text{bulk}$ and $b_l = \text{bulk}$ $(l = 2, 3)$. The other intervals corresponding to this specimen are given in Table 2 with the results of the analysis. Here our method offers nearly the same performance as for the simpler system analyzed before. One exception is τ_3 which yields a relatively large maximal error. However, the mean error is even here within tolerable bounds. As before, we have found a function $\mathbf{F}$ which reflects the dependency.
Note, that we have a low contrast of the silicon oxide layer with respect to the silicon substrate, i.e., the difference of the electron densities is low. For this reason, the reflectivity curve is relatively insensitive to the parameters of the silicon oxide layer, leading to an increased

parameter	a_l	b_l	ϵ_l	$\overline{\eta}_l$	STD	$\hat{\eta}_l$	NSV
τ_2 [Å]	80	120	1	$6 \cdot 10^{-4}$	0.01	0.38	41
τ_3 [Å]	400	600	5	1.71	4.08	39.00	2267
ρ_2 [g/cm^3]	6	8.58	0.086	$1.1 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$	0.01	22
ρ_3 [g/cm^3]	1.54	2.20	0.022	$1.1 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	0.04	581
σ_2 [Å]	0	10.0	0.2	0.028	0.06	0.79	1182
σ_3 [Å]	0	10.0	0.2	0.058	0.11	1.14	1950

Table 2: Results for an independent random test set of $T = 10000$ reflectivity curves for a model with $J = 4$ layers.

complexity, i.e., a larger NSVs, of the underlying mappings for the third layer as compared to the second layer.

5 Conclusions

We presented a new SVM based approach for detecting the parameters of thin films from their reflectivity curves. To be independent from measured data, we employed the the optical matrix method for the generation of training associations. We investigated a three–layer and a four–layer model. Our method with 5000 training associations exhibited a good approximation of the underlying mapping for a large test set of 10000 simulated curves in both cases.
We conclude that parameter detection of thin films by SVMs represents a new and very promising scheme which approaches the problem by multivariate sparse approximation. Our method offers the possibility for an automation of the evaluation of reflectivity curves. An application of this method for a broader range of parameter detection problems in X-ray analysis seems to be possible. However, our approach is novel to the field of reflectometry from its statement and cannot be founded on any results obtained before. Therefore, some constants given here by heuristics are first attempts and can, of course, not be seen as optimal in general. Although we also have successfully analyzed measured data, more investigations are needed to evaluate whether our method offers the same performance in measurement practice. We also hope that further interdisciplinary research will illuminate some relations of the physical behaviours and the multivariate mappings such that we can incorporate more a priori knowledge in our task.

Acknowledgement: The authors like to thank Dr. P. Lamparter (Max Plank Institute for Metals Research, Germany) for critically reading the manuscript.

References

[1] N. ARONSZAJN, *Theory of reproducing kernels*, Trans. Amer. Math. Soc., 68 (1950), pp. 337–404.

[2] R. COLLOBERT AND S. BENGIO, *On the convergence of SVMTorch, an algorithm for large-scale regression problems*, Technical Report IDIAP–RR–00–24, Institut Dalle Molle d'Intelligence Artificelle Perceptive, Martigny, Switzerland, 2000.

[3] ——, *Support vector machines for large-scale regression problems*, Technical Report IDIAP–RR–00–17, Institut Dalle Molle d'Intelligence Artificelle Perceptive, Martigny, Switzerland, 2000.

[4] ——, *SVMTorch: Support vector machines for large-scale regression problems*, Journal of Machine Learning Research, 1 (2001), pp. 143–160.

[5] N. CRISTIANINI, C. CAMPBELL, AND J. SHAWE-TAYLOR, *Dynamically adapting kernels in support vector machines*, NeuroCOLT Technical Report NC-TR-98-017, Royal Holloway College, University of London, UK, 1998.

[6] G. W. FLAKE AND S. LAWRENCE, *Efficient SVM regression training with SMO*, Technical Report, NEC Research Institute, 1999.

[7] J. D. JACKSON, *Classical Electrodynamics*, John Wiley & Sons, New York, 1998.

[8] T. JOACHIMS, *Making large-scale support vector machine learning practical*, in Advances in Kernel Methods – Support Vector Learning, B. Schölkopf, C. Burges, and A. J. Smola, eds., Cambridge, MA, 1999, pp. 169–184.

[9] H. KIESSIG, *Interferenz von Röntgenstrahlen an dünnen Schichten*, Ann. Physik, 10 (1931), pp. 769–778.

[10] G. S. KIMELDORF AND G. WAHBA, *Some results on Tchebycheffian spline functions*, J. Anal. Applic., 33 (1971), pp. 82–95.

[11] B. LENGELER, *X-ray reflection, a new tool for investigating layered structures and interfaces*, in Advances in X-Ray analysis, C. S. Barrett, ed., vol. 35, New York, 1992, pp. 127–135.

[12] C.-J. LIN, *On the convergence of the decomposition method for support vector machines*, Technical Report, National Taiwan University, Taipei, Taiwan, 2000. to appear in: IEEE Transactions on Neural Networks 2001.

[13] X. LONG, N. HUANG, F. HE, AND X. PENG, *An artificial neural network analysis of low-resolution x-ray fluorescence spectra*, Advances in X-Ray Analysis, 40, CD-ROM (1997).

[14] L. NÉVOT AND P. CROCE, *Caractérisation des surfaces par réflexion rasante de rayons x. application à l'étude du polissage de quelques verres silicates*, Rev. Phys. App., 15 (1980), pp. 761–779.

[15] E. OSUNA, R. FREUND, AND F. GIROSI, *An improved training algorithm for support vector machines*, in Neural Networks for Signal Processing VII – Proc. of the 1997 IEEE Workshop, Amelia Island, FL, 1997, pp. 276–285.

[16] L. G. PARRATT, *Surface studies of solids by total reflection of X-rays*, Phys. Rev., 95 (1954), pp. 359–370.

[17] J. C. PLATT, *Fast training of support vector machines using sequential minimal optimization*, in Advances in Kernel Methods – Support Vector Learning, B. Schölkopf, C. Burges, and A. J. Smola, eds., Cambridge, MA, 1999, pp. 185–208.

[18] M. PONTIL, S. MUKHERJEE, AND F. GIROSI, *On the noise model of support vector machine regression*, A.I. Memo No. 1651, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Cambridge, MA, 1998.

[19] B. D. RIPLEY, *Pattern Recognition and Neural Networks*, Cambridge University Press, New York, 1996.

[20] SIEMENS, *The New Siemens X-Ray Reflectometer – A Tool with Outstanding Capabilities*, Siemens Report, 1994.

[21] A. J. SMOLA AND B. SCHÖLKOPF, *From regularization operators to support vector kernels*, in Advances in Neural information processing systems 10, San Mateo, CA, 1998, pp. 343–349.

[22] V. VAPNIK, *The Nature of Statistical Learning Theory*, Springer, New York, 1995.

[23] B. VIDAL AND P. VINCENT, *Metallic multilayers for X-rays using classical thin-film theory*, Applied Optics, 23 (1984), pp. 1794–1801.

[24] G. WAHBA, *Support vector machines, reproducing kernel hilbert spaces and the randomized GACV*, in Advanced in Kernel Methods – Support Vector Learning, B. Schölkopf, C. Burges, and A. J. Smola, eds., Cambridge, MA, 1999, pp. 293–306.

[25] H. WERN AND M. RINGEISEN, *Evaluation of residual stress gradients by diffraction methods with wavelets, a neural network approach*, in Proceedings of SPIE: Nondestructive Evaluation Techniques for Aging Infrastructure & Manufacturing, 1999, pp. 318–328.

[26] G. ZOUTENDIJK, *Methods of Feasible Directions. A study in linear and non-linear programming*, Elsevier, Amsterdam, 1960.