# MULTIDIMENSIONAL SVM TO INCLUDE THE SAMPLES OF THE DERIVATIVES IN THE RECONSTRUCTION OF A FUNCTION

*Fernando Pérez-Cruz, Marcelino Lázaro, Antonio Artés-Rodríguez*

Departamento de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid
Avda. Universidad 30, 28911, Leganés (Madrid) SPAIN
{fernandop,marce,antonio}@ieee.org

## ABSTRACT

In this paper we propose a multidimensional regression estimation algorithm for estimating a function from its first derivatives. The proposed method is extended to introduce the information about the function itself and higher order derivatives. The proposed algorithm is able to exploit the dependency between the output variables to provide a better estimation of the function and it guarantees that the estimated derivatives belong to the same function. The method has been validated by synthetic test functions and it has been used to model a MESFET transistor including intermodulation distortion characterization, where the approximation of the derivatives of the characteristic function is mandatory.

## 1. INTRODUCTION

Regression approximation from a given dataset is a very common problem in a number of applications. In some of these applications, like econometrics [1], telemetry or device modeling [2], it is necessary to fit not only the underlying characteristic function but also its derivatives. For instance, to represent the intermodulation distortion (IMD) of a microwave device, it is necessary to approximate derivatives up to the same order of the intermodulation products to be considered [3]. Usually, up to the third order is considered [2].

Several methods have been employed to simultaneously approximate a function and its derivatives: splines, neural networks, filter banks, etc. For further details see [4] and references therein. Regression estimation from samples of the function and its derivatives becomes a multidimensional regression estimation problem, in which the output variables are dependent on each other. Therefore, an individual regression estimation of each output will be discarding the dependency between these outputs and will not be guaranteeing that the estimated derivatives belong to the same originating function.

Support Vector Machines (SVMs) are state-of-the-art tools for linear and nonlinear input-output knowledge discovery [5]. Support Vector Machines, given a labeled dataset and a nonlinear mapping to a higher dimensional space, have been proposed for solving pattern recognition [6] and regression estimation [7] problems. Recently, they have been extended for solving multidimensional regression estimation problems [8] using a quadratic $\varepsilon$-insensitive cost function.

In this paper, we propose to solve the estimation of a function from its first derivatives using as base algorithm the Multidimensional Support Vector Regressor (M-SVR) proposed in [8]. The extension for using the function values and higher order derivatives is straightforward, as we will show in Section 4. We have not included them in the initial algorithmic development, because we believe it does not give a further insight about the problem at hand and would probably obscure its presentation, making the paper hard to follow. The M-SVR will have to be modified in two ways to be able to solve the proposed problem. First, the algorithm will have to be adapted to the actual problem. Second, to avoid numerical instabilities, the cost function will be changed, leading to an algorithm easy to implement and fast to solve.

The rest of the paper is organized as follows. The problem of estimating the function from its first derivatives is detailed in Section 2, together with the needed modification over the M-SVR algorithm. The optimization is carried out using an Iterative Re-Weighted Least Square (IRWLS) procedure, which is detailed in Section 3. The method is extended to include the function and higher order derivatives in Section 4. In Section 5, we show by means of computer experiments the performance of the proposed approach. Section 6 shows the results obtained when the method is applied to model a MESFET transistor. Finally, we end the paper in Section 7 with some concluding remarks.

## 2. FUNCTION ESTIMATION FROM ITS DERIVATIVES

The problem we are going to solve is to estimate $f(\mathbf{x})$ from its first derivatives, given $n$ data points $(\mathbf{x}_1,\ldots,\mathbf{x}_n)$ in a $d$ dimensional space $\mathbf{x}_i \in \mathbb{R}^d$. Therefore for each input vector a $d$-dimensional label vector $\mathbf{y}_i \in \mathbb{R}^d$ will be available, where

$$\mathbf{y}_i = \left[ \left.\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_{i1}}\right|_{\mathbf{x}_i}, \left.\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_{i2}}\right|_{\mathbf{x}_i}, \ldots, \left.\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_{id}}\right|_{\mathbf{x}_i} \right] = \nabla_{\mathbf{x}} f(\mathbf{x}_i).$$

We define the estimated function $\hat{f}(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x})$, where $\mathbf{w}$ is a weight vector and $\phi(\cdot)$ is a nonlinear transformation of the input vector $\mathbf{x}$ to a higher dimensional space (the feature space, $\phi(\mathbf{x}) \in \mathbb{R}^{\mathscr{H}}$ and $\mathscr{H} \geq d$). We need to solve a multidimensional regression problem for finding $\mathbf{w}$, in which we need to reduce the error between the derivatives of the estimated function and the $\mathbf{y}_i$ vector:

$$\mathbf{e}_i = \mathbf{y}_i - \nabla_{\mathbf{x}} \hat{f}(\mathbf{x}_i) =$$
$$\left[ y_{i1} - \mathbf{w}^T \phi_1'(\mathbf{x}_i), y_{i2} - \mathbf{w}^T \phi_2'(\mathbf{x}_i), \ldots, y_{id} - \mathbf{w}^T \phi_d'(\mathbf{x}_i) \right],$$

where we have defined $\phi_j'(\mathbf{x}_i) = \left.\frac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}_{ij}}\right|_{\mathbf{x}_i}$.

A Multidimensional Support Vector Regressor (M-SVR) has been recently proposed [8]: given a labeled data set

$\{(\mathbf{x}_i,\mathbf{y}_i)\}$, $i=1,\cdots,n$, with $\mathbf{x}_i \in \mathbb{R}^d$ and $\mathbf{y}_i \in \mathbb{R}^Q$, we need to optimize

$$\min_{\mathbf{w}^j,b^j} \frac{1}{2}\sum_{j=1}^{Q}\|\mathbf{w}^j\|^2 + C\sum_{i=1}^{n} L(u_i),$$

where $u_i = \|\mathbf{e}_i\| = \sqrt{(\mathbf{e}_i^T\mathbf{e}_i)}$, $\mathbf{e}_i^T = \mathbf{y}_i^T - \phi^T(\mathbf{x}_i)\mathbf{W} - \mathbf{B}^T$, $\mathbf{W} = [\mathbf{w}^1,\ldots,\mathbf{w}^Q]$, $\mathbf{B} = [b^1,\ldots,b^Q]^T$ and

$$L(u) = \begin{cases} 0, & u < \varepsilon \\ u^2 - \varepsilon^2, & u \geq \varepsilon \end{cases},$$

is a quadratic epsilon-insensitive cost function.

This multidimensional problem needs to be modified to solve our particular function approximation problem. First, instead of having a vector $\phi(\mathbf{x})$ and a matrix $\mathbf{W}$ to construct the error vector, we have a unique weight vector $\mathbf{w}$ and a matrix that contains the derivatives of $\phi(\mathbf{x})$. The second modification is a slightly change over the cost function:

$$L(u) = \begin{cases} 0, & u < \varepsilon \\ u^2 - 2u\varepsilon + \varepsilon^2, & u \geq \varepsilon \end{cases}, \tag{1}$$

to make its derivative with respect to $u$ continuous and to avoid numerical instabilities. To sum up, we are left with the following unconstrained functional

$$L_P(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} L(u_i), \tag{2}$$

which needs to be minimized with respect to $\mathbf{w}$, where

$$u_i = \|\mathbf{e}_i\| = \sqrt{\mathbf{e}_i^T\mathbf{e}_i},$$
$$\mathbf{e}_i = \mathbf{y}_i - (\phi^{'}(\mathbf{x}_i))^T\mathbf{w},$$
$$\phi^{'}(\mathbf{x}_i) = [\phi_1^{'}(\mathbf{x}_i),\ldots,\phi_d^{'}(\mathbf{x}_i)],$$

and $L(u)$ is given by (1).

## 3. RESOLUTION OF THE MULTIDIMENSIONAL SUPPORT VECTOR REGRESSOR

To optimize the proposed multidimensional regression estimation problem, we are going to follow an Iterative Re-Weighted Least Square (IRWLS) procedure which has been successfully applied for solving SVMs in [9, 10] and has been recently proven to converge to the SVM solution [11]. To construct an IRWLS procedure, we first obtain a first order Taylor expansion of $L(u)$, leading to the minimization of

$$L_P^{'}(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\sum_{i=1}^{n} L(u_i^k) + \left.\frac{dL(u)}{du}\right|_{u_i^k}[u_i - u_i^k]\right) \tag{3}$$

where $u_i^k = \|\mathbf{e}_i^k\|$ and $\mathbf{e}_i^k = \mathbf{y}_i - (\phi^{'}(\mathbf{x}_i))^T\mathbf{w}^k$. Then, we need to construct a quadratic approximation as follows:

$$L_P^{''}(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\sum_{i=1}^{n} L(u_i^k) + \left.\frac{dL(u)}{du}\right|_{u_i^k}\frac{(u_i)^2 - (u_i^k)^2}{2u_i^k}\right)$$
$$= \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{2}\sum_{i=1}^{n} a_i(\mathbf{e}_i^T\mathbf{e}_i) + CT, \tag{4}$$

where

$$a_i = \frac{C}{u_i^k}\left.\frac{dL(u)}{du}\right|_{u_i^k} = \begin{cases} 0, & u_i^k < \varepsilon \\ \frac{2C(u_i^k - \varepsilon)}{u_i^k}, & u_i^k \geq \varepsilon \end{cases}. \tag{5}$$

$CT$ are constant terms that do not depend on $\mathbf{w}$. This functional is a regularized weighted least square problem in which the weight $a_i$ depends on the previous solution, so one needs to iterate to find the fixed point solution.

The functional $L_P^{''}(\mathbf{w})$ is a quadratic approximation to $L_P(\mathbf{w})$ in (2) that presents the same value $L_P^{''}(\mathbf{w}^k) = L_P(\mathbf{w}^k)$ and gradient $\nabla_{\mathbf{w}} L_P^{''}(\mathbf{w}^k) = \nabla_{\mathbf{w}} L_P(\mathbf{w}^k)$ for $\mathbf{w} = \mathbf{w}^k$. Therefore, we can define $\mathbf{p}^k = \mathbf{w}^s - \mathbf{w}^k$ as a descending direction for $L_P(\mathbf{w})$, where $\mathbf{w}^s$ is the least square solution to (4), and we can use it to construct a line search method [12], i.e. $\mathbf{w}^{k+1} = \mathbf{w}^k + \eta^k\mathbf{p}^k$. The value of $\eta^k$ can be computed using a backtracking line search [12], in which $\eta^k$ is initially set to 1 and if $L_P(\mathbf{w}^{k+1}) \geq L_p(\mathbf{w}^k)$, it is iteratively reduced until a strict decrease in the functional in (2) is observed.

Now we are going to obtain $\mathbf{w}^s$, the solution to $L_P^{''}(\mathbf{w})$ in (4), equating its gradient to zero

$$\nabla_{\mathbf{w}} L_P^{''}(\mathbf{w}) = \mathbf{w} - \sum_{i=1}^{n}\phi^{'}(\mathbf{x}_i)\mathbf{e}_i a_i = \mathbf{0},$$

which can be expressed as follows:

$$\mathbf{w} + \sum_{i=1}^{n}\phi^{'}(\mathbf{x}_i)\mathbf{D}_{\mathbf{a}_i}(\phi^{'}(\mathbf{x}_i))^T\mathbf{w} = \sum_{i=1}^{n}\phi^{'}(\mathbf{x}_i)\mathbf{D}_{\mathbf{a}_i}\mathbf{y}_i.$$

We have defined $\mathbf{D}_{\mathbf{a}_i}$ as a $d \times d$ diagonal matrix with $a_i$ as its diagonal elements, $(\mathbf{D}_{\mathbf{a}_i})_{lk} = a_i\delta[l-k]$. Finally, we can express it in matrix notation as follows:

$$[\mathbf{\Phi}^T\mathbf{D}_{\mathbf{a}}\mathbf{\Phi} + \mathbf{I}]\mathbf{w} = \mathbf{\Phi}^T\mathbf{D}_{\mathbf{a}}\mathbf{Y}. \tag{6}$$

$\mathbf{\Phi}^T = [\phi^{'}(\mathbf{x}_1),\ldots,\phi^{'}(\mathbf{x}_n)]$, $\mathbf{D}_{\mathbf{a}}$ is a $nd \times nd$ diagonal matrix where each $d \times d$ submatrix is defined as $(\mathbf{D}_{\mathbf{a}})_{ij} = \mathbf{D}_{\mathbf{a}_i}\delta[i-j]$ and $\mathbf{Y} = [\mathbf{y}_1^T,\ldots,\mathbf{y}_n^T]^T$ is a $nd$-dimensional column vector.

The system in (6) can be solved using kernels. In order to do so, we are going to apply the Representer theorem [5] which states that the optimal solution can be constructed as a linear combination of the training samples in the feature space, i.e. $\mathbf{w} = \mathbf{\Phi}^T\beta$, which can be replaced into (6) leading to:

$$[\mathbf{\Phi}^T\mathbf{D}_{\mathbf{a}}\mathbf{\Phi} + \mathbf{I}]\mathbf{\Phi}^T\beta = \mathbf{\Phi}^T\mathbf{D}_{\mathbf{a}}\mathbf{Y}. \tag{7}$$

Pre-multiplying (7) by the pseudo-inverse of $\mathbf{\Phi}^T\mathbf{D}_{\mathbf{a}}$:

$$(\mathbf{D}_{\mathbf{a}}\mathbf{\Phi}\mathbf{\Phi}^T\mathbf{D}_{\mathbf{a}})^{-1}\mathbf{D}_{\mathbf{a}}\mathbf{\Phi}[\mathbf{\Phi}^T\mathbf{D}_{\mathbf{a}}\mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{\Phi}^T]\beta = \mathbf{Y},$$

and extracting common factor $\mathbf{\Phi}^T\mathbf{D}_{\mathbf{a}}$ from the brackets, leads to

$$(\mathbf{D}_{\mathbf{a}}\mathbf{\Phi}\mathbf{\Phi}^T\mathbf{D}_{\mathbf{a}})^{-1}\mathbf{D}_{\mathbf{a}}\mathbf{\Phi}\mathbf{\Phi}^T\mathbf{D}_{\mathbf{a}}[\mathbf{\Phi}\mathbf{\Phi}^T + \mathbf{D}_{\mathbf{a}}{}^{-1}]\beta = \mathbf{Y}.$$

This equation can be simplified to

$$[\mathbf{H} + \mathbf{D}_{\mathbf{a}}{}^{-1}]\beta = \mathbf{Y}, \tag{8}$$

where we have defined $\mathbf{H} = \mathbf{\Phi}\mathbf{\Phi}^T$, which is a kernel matrix only formed by inner products of $\phi_j^{'}(\mathbf{x}_i)$.

The IRWLS procedure for solving the multidimensional regression problem to find a function from its derivatives can be summarized in the following steps:

1. Initialization: set $\beta^0 = \mathbf{0}$, $u_i = \|\mathbf{y}_i\|$ and compute $a_i$ from (5).
2. Compute $\beta^s = [\mathbf{H} + \mathbf{D_a}^{-1}]^{-1}\mathbf{Y}$ and set $\eta^k = 1$.
3. Set $\beta^{k+1} = \beta^k + \eta^k[\beta^s - \beta^k]$ if $L(\beta^{k+1}) < L(\beta^k)$ go to Step 5.
4. Set $\eta^k = \rho\eta^k$ with $0 < \rho < 1$ and go to Step 3.
5. Recompute $u_i$ and $a_i$, set $k = k+1$ and go to Step 2 until convergence.

## 4. EXTENSIONS

The extension of the proposed method to include samples from the function and from higher order derivatives is straightforward. In this case, the vectors $\mathbf{y}_i$ and $\mathbf{e}_i$ will be constructed with all the available information and the procedure in Section 2 and 3 can be easily replicated. To illustrate this point, we propose the following example, where $\mathbf{y}_i$ and $\mathbf{e}_i$ are, respectively

$$\mathbf{y}_i = \left[ f(\mathbf{x}_i), \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_{i1}}\bigg|_{\mathbf{x}_i}, \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_{i2}}\bigg|_{\mathbf{x}_i}, \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_{i1}\partial \mathbf{x}_{i2}}\bigg|_{\mathbf{x}_i} \right],$$

$$\mathbf{e}_i = \Big[ y_{i1} - \mathbf{w}^T\phi(\mathbf{x}_i), y_{i1} - \mathbf{w}^T\phi_1^{'}(\mathbf{x}_i), \\ y_{i2} - \mathbf{w}^T\phi_2^{'}(\mathbf{x}_i), y_{i3} - \mathbf{w}^T\phi_{1,2}^{''}(\mathbf{x}_i) \Big].$$

Finally, when some data is more reliable or less noisy, or the range of the derivatives is clearly different, a weighted norm is more convenient for $u_i$, i.e. $u_i = \sqrt{\sum_{j=1}^{Q} c_j e_{ij}^2}$, where $Q$ is the dimension of $\mathbf{y}_i$ and $c_j$ are the corresponding weights with each dimension of $\mathbf{y}_i$. It is straightforward to find out that, in the algorithm, this just supposes to include the weights in the diagonal matrix $\mathbf{D_a}$ as $(\mathbf{D}_{\mathbf{a}_i})_{lk} = a_i\delta[l-k]c_k$.

## 5. SYNTHETIC EXPERIMENTS

First of all, the performance of the proposed method has been validated by synthetic experiments. 8 functions sampled from a two-dimensional input space, proposed in [13], have been used. Their analytical expressions can be found in [13] or [4].

The following methods will be compared. The conventional SVR will be used when only the samples of the function are used. When the samples of the function and the first order derivatives are jointly used, the proposed method (labeled "M-SVR" in the following) is employed. Finally, when only the samples of the two first order derivatives are used, again the proposed method (labeled "M-SVRd" in this case) is applied. In all cases, Gaussian kernels are employed. The Signal to Error Ratio (SER), expressed in dB, between the true function/derivatives and its corresponding reconstruction has been used as figure of merit.

For the sake of a fair comparison, in the results to be presented a similar number of total samples is used for all methods. A Signal to Noise Ratio (SNR) of 10 dB has been considered for the samples of both function and derivatives. SVR was trained with a uniform grid of $19 \times 19$ sampling points (361 samples for the function), M-SVRd with $13 \times 14$ sampling points (364 samples, 182 for each derivative) and M-SVR with $11 \times 11$ sampling points (363 samples, 121 for the function and for each derivative). The optimal kernel size (variance), as well as the parameters for all methods ($C$ and $\varepsilon$), have been determined by cross-validation.

Table 1 compares the performance of the three methods under test in terms of the SER (dB) obtained in the reconstruction of the 8 test functions and their first order derivatives. To help the comparison task, the mean value over the 8 function has also been included. For the derivatives, the mean value obtained for both of them, with respect to $x_1$ and with respect to $x_2$, is presented.

The methods including the samples of the derivatives, M-SVR and M-SVRd, clearly out-perform the SVR in the reconstruction of the derivatives. Moreover, they also improve the reconstruction of the function. Consequently, the information of the derivatives is clearly helpful in the reconstruction of a function. In this case, the M-SVRd provides better results for the derivatives while the M-SVR is better for the reconstruction of the function.

## 6. NONLINEAR SMALL-SIGNAL MODELING OF A MESFET FOR INTERMODULATION DISTORTION

In this section, the proposed method is used to reproduce the intermodulation distortion (IMD) behavior of a microwave *Metal Semiconductor Field Effect* (MESFET) transistor. To address this problem, the model must accurately represent not only the nonlinear current-voltage ($I/V$) characteristic, but also its derivatives up to the same order of the intermodulation products to be considered [3]. In applications with amplifiers and mixers, the usual is to consider up to the third order IMD [2]. Therefore, the model must approximate up to the third order derivatives.

The characteristic function of a MESFET transistors relates the drain to source current $I_{ds}$, with both the drain to source, $V_{ds}$, and gate to source, $V_{gs}$, voltages. Here we are going to model the static drain to source current appearing in the MESFET equivalent circuit (see [14] for details). Typically, the transistor is polarized in a bias point, ($V_{ds0}$,$V_{gs0}$), and the incremental drain-to-source and gate-to-source voltages,$v_{ds}$ and $v_{gs}$, are applied over this DC polarization. With these premises, it is necessary to accurately reproduce the $I_{ds} = f(V_{ds0}, V_{gs0}, v_{ds}, v_{gs})$ dependence and its derivatives with respect to the incremental voltages. Some traditional methods [15, 14] propose to implement a truncated (up to the third order) Taylor series expansion representing $I_{ds}$ in a small interval around the bias point,i.e.

$$I_{ds} = I_{ds0} + G_m v_{gs} + G_d v_{ds} + G_{m2} v_{gs}^2 + G_{md} v_{gs} v_{ds} + \\ G_{d2} v_{ds}^2 + G_{m3} v_{gs}^3 + G_{m2d} v_{gs}^2 v_{ds} + G_{md2} v_{gs} v_{ds}^2 + G_{d3} v_{ds}^3,$$

where $I_{ds0}$ is the DC drain current and, $(G_m, \cdots, G_{d3})$ are coefficients related to the $n$th-order derivatives of the $I/V$ characteristic evaluated at the bias point.

In [14], a *Generalized Radial Basis Function Network* (GRBF) is used to approximate the 10 coefficients involved in the Taylor expansion individually. Here we propose to employ the M-SVR method to approximate the characteristic function and its derivatives. This method has the advantage of allowing to extend the range of validity of the model if the set of derivatives is measured at different values of $v_{ds}$ and $v_{gs}$, which is not possible by using a Taylor based model. Moreover, the interdependence of the derivatives is exploited by the method, which can be clearly helpful, as the results with synthetic data has probed. In the following, the capability of this method for this application is tested by comparing its performance with the method proposed in [14].

For a real NE72084 MESFET, $I_{ds0}$ and the 9 derivatives ($G$ coefficients) have been measured (by the method in [15]) in 533 different bias points, which gives a total of 5330 samples. 147 bias points have been assigned to the training set and the remaining 386 bias points to the test set. A GRBF with 75 activation units has been used for the method in [14]. Not a noticeable improvement was observed by increasing the network size. Table 2 compares the SER of each method for all the coefficients/derivatives.

Both methods provide very close results, which shows the ability of the proposed method for this application, even when only samples in a single point of the $v_{ds}$ and $v_{gs}$ variables are available. Moreover, it can be seen that the method provides slightly better results in the reconstruction of the higher order derivatives, which are the noisiest ones.

## 7. CONCLUSION

A new multidimensional regression method to estimate a function from the samples of the function and its derivatives has been presented. The proposed method exploits the interdependence of the derivatives to improve the reconstruction task. Experimental results have shown that this approach improves the performance obtained when the function is directly estimated from its samples. This demonstrates that function estimation can benefit from the use of samples from the derivatives.

The proposed method has been applied to the nonlinear modeling of a MESFET transistor to reproduce the intermodulation behavior. In this case, the derivative measures are only available at the bias point, which limits the capability of the proposed method since only derivatives at $v_{ds} = v_{gs} = 0$ are given. However, the results provided by the method are similar to the classical approaches based on Taylor series expansions. Moreover, the method has the potentiality to extend the small-signal range of the method if measures of the derivatives are available for a wider range, unlike Taylor based methods. Actually, serious efforts are dedicated to develop accurate methods to measure the derivatives of $I_{ds}$ in large signal. From these measures, taking into account the results obtained with synthetic data, the proposed method is expected to provide outstanding results.

Approximation of the function

| | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | **Mean** |
|---|---|---|---|---|---|---|---|---|---|
| SVR | 19.1 | 27.0 | 24.0 | 21.9 | 29.3 | 25.6 | 26.8 | 16.3 | **23.8** |
| M-SVRd | 21.0 | 29.0 | 27.1 | 20.1 | 27.7 | 26.2 | 27.2 | 16.4 | **24.3** |
| M-SVR | 21.5 | 29.9 | 27.7 | 23.1 | 29.6 | 27.3 | 28.5 | 16.7 | **25.5** |

Approximation of the first order derivatives (mean value)

| | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | **Mean** |
|---|---|---|---|---|---|---|---|---|---|
| SVR | 12.3 | 11.3 | 11.2 | 14.9 | 12.1 | 12.2 | 11.4 | 11.6 | **12.1** |
| M-SVRd | 17.8 | 18.0 | 18.7 | 17.9 | 16.1 | 17.3 | 16.3 | 14.6 | **17.1** |
| M-SVR | 17.0 | 17.0 | 17.9 | 18.2 | 15.6 | 16.5 | 15.5 | 13.3 | **16.4** |

Table 1: SER (dB) for the reconstruction of the 8 test functions and their first order derivatives using a similar number of total samples

| | $I_{ds0}$ | $G_{ds}$ | $G_m$ | $G_{md}$ | $G_{d2}$ | $G_{m2}$ |
|---|---|---|---|---|---|---|
| GRBF | 43.5 | 44.1 | 45.2 | 41.5 | 33.5 | 44.8 |
| M-SVR | 42.5 | 43.4 | 45.6 | 41.5 | 33.7 | 44.9 |

| | $G_{m2d}$ | $G_{md2}$ | $G_{d3}$ | $G_{m3}$ | **Mean** |
|---|---|---|---|---|---|
| GRBF | 22.7 | 22.0 | 21.4 | 21.2 | **34.0** |
| M-SVR | 22.8 | 22.6 | 21.8 | 21.3 | **34.0** |

Table 2: SER (dB) in the reconstruction of the coefficients/derivatives of the MESFET

## REFERENCES


[1] John C. Hull, *Options, Futures, and Other Derivatives*, Prentice-Hall, New York, NY, 3 edition, 1997.

[2] S. A. Maas and D. Neilson, "Modeling MESFETs for intermodulation analysis of mixers and amplifiers," *IEEE Transactions Microwave Theory and Techniques*, vol. 38, no. 12, pp. 1964–1971, 1990.

[3] S. A. Maas, "How to model intermodulation distortion," *IEEE MTT-S International Microwave Symposium Digest*, pp. 149–151, 1991.

[4] M. Lázaro, I. Santamaría, C. Pantaleón, J. Ibáñez, and L. Vielva, "A regularized technique for the simultaneous reconstruction of a function and its derivatives with application to nonlinear transistor modeling," *Signal Processing*, no. 83, pp. 1859–1870, 2003.

[5] Bernhard Schölkopf and Alex Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.

[6] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *5th Annual ACM Workshop on COLT*, D. Haussler, Ed., Pittsburgh, PA, 1992, pp. 144–152, ACM Press.

[7] V. N. Vapnik, S. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," in *Neural Information Processing Systems*, M. Mozer, M. Jordan, and T. Petsche, Eds., Cambridge, MA, 1997, pp. 169–184, M.I.T. Press.

[8] F. Pérez-Cruz, G. Camps, E. Soria, J. Pérez, A. R. Figueiras-Vidal, and A. Artés-Rodríguez, "Multi-dimensional function approximation and regression estimation," in *ICANN02*, Madrid, Spain, 2002, Springer.

[9] F. Pérez-Cruz, P. L. Alarcón-Diana, A. Navia-Vázquez, and A. Artés-Rodríguez, "Fast training of support vector classifiers," in *Advances in Neural Information Processing Systems 13*, Cambridge, MA, Nov. 2000, M.I.T. Press.

[10] F. Pérez-Cruz, A. Navia-Vázquez, P. L. Alarcón-Diana, and A. Artés-Rodríguez, "An IRWLS procedure for SVR," in *Proceedings of the EUSIPCO'00*, Tampere, Finland, Sept. 2000.

[11] F. Pérez-Cruz, C. Bousoño-Calzón, C.-H. Lin, and A. Artés-Rodríguez, "Convergence of the irwls procedure to the support vector machine solution," *IEEE Transactions on Neural Networks*, Submitted, 2003.

[12] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 1999.

[13] V. Cherkassky, D. Gehring, and F. Mulier, "Comparison of adaptive methods for function estimation from samples," *IEEE Transactions on Neural Networks*, vol. 7, no. 4, pp. 969–984, July 1996.

[14] I. Santamaría, M. Lázaro, C. J. Pantaleón, J. A. García, A. Tazón, and A. Mediavilla, "A nonlinear MESFET model for intermodulation analysis using a generalized radial basis function network," *Neurocomputing*, vol. 25, pp. 1–18, 1999.

[15] J. C. Pedro and J. Pérez, "Accurate simulation of GaAs MESFETs intermodulation using a new drain-source current model," *IEEE Transactions Microwave Theory and Techniques*, vol. 42, no. 1, pp. 25–33, 1994.