Efficient Uncertainty Quantification using sensitivity information in Least Squares SVM

Karanvir S. Sidhu and Roni Khazaka
Department of Electrical and Computer Engineering, McGill University, Montréal, Québec, Canada, H3A 0E9
Email: karanvir.sidhu@mail.mcgill.ca, roni.khazaka@mcgill.ca

***Abstract*—This paper proposes a method to use the sensitivity information for efficient uncertainty quantification using the Least Squares-Support Vector Machine (LS-SVM) framework. The inclusion of the sensitivity information in the LS-SVM framework allows us to reduce the number of simulations. Finally, the proposed approach is verified with use of an circuit example with 25 random parameters.**

***Index Terms*—Machine Learning, SVM, LS-SVM, Parametric-Derivatives, Curse of dimensionality**

I. Introduction

In recent years, device scaling has led to process variations in the electronics. Several approaches, such as Polynomial Chaos and Rational Polynomial Chaos, have been suggested [1] to form a surrogate model for the output Quantity of Interest (QoI). In addition to the above, several machine learning techniques such as Least Squares-Support Vector Machines (LS-SVM) [2,3] and Gaussian Process Regression [4] have also been proposed for uncertainty quantification for the QoI. Although the above techniques can accurately compute the statistical properties of QoI, the number of simulations required in the above methods can be a bottleneck.

This paper proposes using sensitivity information with the LS-SVM method to compute the surrogate model for the QoI with fewer circuit simulations since computing sensitivity is less expensive than performing circuit evaluations. The use of sensitivity information for the Polynomial Chaos method has been proposed in [5], where it was shown that the sensitivity information can be used to compute the surrogate model for QoI accurately with fewer circuit simulations. Sensitivity information has been used in the context of Support Vector Regression in [6] for learning the derivatives of a given function. Moreover, the derivative information has also been used with Radial Basis Kernels in [7] to improve the training time of the Gaussian Processes. This paper adapts the technique in [6] and [7] for using it for uncertainty quantification in the field of circuit systems as most commercial circuit simulators can efficiently compute the sensitivity information.

II. Background

Consider a circuit for which we need to quantify the uncertainty of a given quatity of interest (QoI), given that some circuit parameters are randomly distributed. The approach used in this paper is to first compute a surrogate model $\mathcal{M}(\boldsymbol{\xi})$ that computes the QoI from the parameters $\boldsymbol{\xi} \in \mathbb{R}^d$, where d is the number of parameters. The surrogate model is then used in order to compute statistical properties of the QoI. Our approach builds upon the LS-SVM method in [8] where the model is defined as,

$$\mathcal{M}(\boldsymbol{\xi}) = \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{\xi}) + b \tag{1}$$

where $\boldsymbol{\phi}(\boldsymbol{\xi}) : \mathbb{R}^d \mapsto \mathbb{R}^D$ is a vector valued function, and $\boldsymbol{w} \in \mathbb{C}^D$ and $b \in \mathbb{C}$, are the unknown weights and the biases, respectively.

In order to compute the model parameters $\boldsymbol{w}$ and b, the LS-SVM approach in [8] solves the following constrained optimization problem,

$$\underset{\boldsymbol{w},b,\boldsymbol{\epsilon}}{argmin} \;\; \frac{1}{2}\boldsymbol{w}^*\boldsymbol{w} + \frac{1}{2}\gamma\boldsymbol{\epsilon}^*\boldsymbol{\epsilon} \tag{2}$$

such that,

$$\begin{aligned}\epsilon_k &= v(\boldsymbol{\xi}^{(k)}) - \mathcal{M}(\boldsymbol{\xi}^{(k)}) \\ &= v_k - \boldsymbol{w}^T \; \phi(\boldsymbol{\xi}^{(k)}) - b; k = 1 \dots K\end{aligned} \tag{3}$$

where $v_k \in \mathbb{C}$ is the QoI evaluated at sample $\boldsymbol{\xi}^{(k)} \in \mathbb{R}^d$, and K is the number of samples. Note that $\epsilon_k \in \mathbb{C}$ is the error between the QoI v_k and the model, $\mathcal{M}(\boldsymbol{\xi}^{(k)})$, evaluated at $\boldsymbol{\xi}^{(k)}$. For brevity, we denote $\boldsymbol{\phi}(\boldsymbol{\xi}^{(k)})$ as $\boldsymbol{\phi}_k$ throughout the remainder of this paper. The above constrained optimization problem is solved in [8] using the Lagrangian approach. In this paper we propose a new method that takes into account sensitivity information, and therefore requires fewer samples to compute the model.

III. Proposed Approach

In order to reduce the required number of simulations, we propose adding the sensitivity of QoI as an extra constraint to the optimization problem in (2). Using the sensitivity constraint, the optimization problem in (2) becomes,

$$\underset{\boldsymbol{w},b,\boldsymbol{\epsilon}}{argmin} \;\; \frac{1}{2}\boldsymbol{w}^*\boldsymbol{w} + \frac{1}{2}\gamma\boldsymbol{\epsilon}^*\boldsymbol{\epsilon} + \frac{1}{2}\gamma' \sum_{i=1}^{d}\sum_{k=1}^{K} \bar{\eta}_{i,k}\eta_{i,k} \tag{4}$$

such that,

$$\begin{aligned}\epsilon_k &= v_k - \boldsymbol{w}^T \; \phi_k - b \\ \eta_{1,k} &= v_{1,k} - \boldsymbol{w}^T \; \phi_{k,1} \\ &\vdots \\ \eta_{d,k} &= v_{d,k} - \boldsymbol{w}^T \; \phi_{k,d}\end{aligned}$$

where, $v_{i,k}$ denotes the sensitivity of v_k with respect to the parameter ξ_i, i.e. $v_{i,k} \equiv \frac{\partial v\left(\boldsymbol{\xi}^{(k)}\right)}{\partial \xi_i}$. Similarly, $\phi_{i,k}$ denotes the

979-8-3503-1798-5/23/$31.00 ©2023 IEEE

derivative of ϕ_k with respect to the parameter ξ_i at sample $\boldsymbol{\xi}^{(k)}$, i.e., $\frac{\partial \phi\left(\boldsymbol{\xi}^{(k)}\right)}{\partial \xi_i}$, and $\eta_{i,k}$ denotes the discrepancy between $v_{i,k}$ and the derivative of the model with respect to the parameter ξ_i at sample $\boldsymbol{\xi}^{(k)}$. To solve the optimisation problem in (4), we can write the Lagrangian as,

$$
\begin{aligned}
L(\boldsymbol{w}, b, \boldsymbol{\epsilon}, \alpha_k, \beta_{1,k}, \ldots, \beta_{d,k}) &= \frac{1}{2}\boldsymbol{w}^* \boldsymbol{w} + \frac{1}{2}\gamma \sum_{k=1}^{K} \bar{\epsilon}_k \epsilon_k \\
&+ \frac{1}{2}\gamma' \sum_{i=1}^{d}\sum_{k=1}^{K} \bar{\eta}_{i,k}\eta_{i,k} - \sum_{k=1}^{K} \alpha_k \left(\epsilon_k - v_k + \boldsymbol{w}^T \phi_k + b\right) \\
&- \sum_{i=1}^{d} \left(\sum_{k=1}^{K} \beta_{i,k} \left(-v_{i,k} + \boldsymbol{w}^T \phi_{i,k}\right) \right)
\end{aligned}
$$

where α_k and $\{\beta_{i,k}\}_{i=1}^{d}$ are the Lagrange multipliers for the QoI constraints and the sensitivity constraints for all d-parameters, respectfully. To find the model, $\mathcal{M}\left(\boldsymbol{\xi}\right)$, we compute the partial derivative of the Lagrangian, $L(\boldsymbol{w}, b, \epsilon_k, \eta_{1,k}, \ldots, \eta_{d,k}, \alpha_k, \beta_{1,k}, \ldots, \beta_{d,k})$, with respect to all the variables and set those equal to zero, and we obtain the following equations,

$$\frac{\partial L}{\partial b} = 0 \Leftrightarrow \sum_{k=1}^{K} \alpha_k = 0 \tag{5}$$

$$\frac{\partial L}{\partial \epsilon_k} = 0 \Leftrightarrow \epsilon_k = \frac{1}{\gamma}\alpha_k \tag{6}$$

$$\frac{\partial L}{\partial \eta_{i,k}} = 0 \Leftrightarrow \eta_{i,k} = \frac{1}{\gamma'}\beta_{i,k} \;; i = 1, \ldots, d \tag{7}$$

$$\frac{\partial L}{\partial \boldsymbol{w}} = 0 \Leftrightarrow \boldsymbol{w} = \sum_{k=1}^{K} \alpha_k \phi_k + \sum_{i=1}^{d} \left(\sum_{k=1}^{K} \beta_{i,k}\phi_{i,k} \right) \tag{8}$$

$$\frac{\partial L}{\partial \alpha_k} = 0 \Leftrightarrow \boldsymbol{w}^T \phi_k + b + \epsilon_k = v_k \tag{9}$$

$$\frac{\partial L}{\partial \beta_{i,k}} = 0 \Leftrightarrow \boldsymbol{w}^T \phi_{i,k} + \eta_{i,k} = v_{i,k} \;;\; i = 1, \ldots, d \tag{10}$$

Upon substituting (6), (7), and (8) in (9) and (10) we obtain the following set of equations at a given sample $\boldsymbol{\xi}^{(k)}$,

$$\sum_{k=1}^{K} \alpha_k = 0 \tag{11}$$

$$\sum_{j=1}^{K} \alpha_j \phi_j^T \phi_k + \sum_{i=1}^{d}\sum_{j=1}^{K} \beta_{i,j}\phi_{i,j}^T \phi_k + b + \frac{1}{\gamma}\alpha_k = v_k \tag{12}$$

$$\sum_{j=1}^{K} \alpha_j \phi_j^T \phi_{1,k} + \sum_{i=1}^{d}\sum_{j=1}^{K} \beta_{i,j}\phi_{i,j}^T \phi_{1,k} + \frac{1}{\gamma'}\beta_{1,k} = v_{1,k} \tag{13}$$

$$\vdots$$

$$\sum_{j=1}^{K} \alpha_j \phi_j^T \phi_{d,k} + \sum_{i=1}^{d}\sum_{j=1}^{K} \beta_{i,j}\phi_{i,j}^T \phi_{d,k} + \frac{1}{\gamma'}\beta_{d,k} = v_{d,k} \tag{14}$$

Writing the above set of equations at all samples $\left\{\boldsymbol{\xi}^{(k)}\right\}_{k=1}^{K}$, we obtain the following system,

$$
\begin{bmatrix}
0 & \mathbf{1}^T & \mathbf{0}^T & \mathbf{0}^T & \ldots & \mathbf{0}^T \\
\mathbf{1} & \boldsymbol{\Omega} + \frac{1}{\gamma}\mathbb{I} & \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_2 & \ldots & \boldsymbol{\Omega}_d \\
\mathbf{0} & \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_{1,1} + \frac{1}{\gamma'}\mathbb{I} & \boldsymbol{\Omega}_{2,1} & \ldots & \boldsymbol{\Omega}_{d,1} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \boldsymbol{\Omega}_d & \boldsymbol{\Omega}_{d,1} & \boldsymbol{\Omega}_{2,d} & \ldots & \boldsymbol{\Omega}_{d,d} + \frac{1}{\gamma'}\mathbb{I}
\end{bmatrix}
\begin{bmatrix} b \\ \boldsymbol{\alpha} \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_d \end{bmatrix}
=
\begin{bmatrix} 0 \\ \boldsymbol{v} \\ \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_d \end{bmatrix}
\tag{15}
$$

where, $\boldsymbol{v} \in \mathbb{C}^K$ contains the QoI evaluations at K samples and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d \in \mathbb{C}^K$ contain the sensitivity of QoI with respect to all d-parameters. Also, $\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d \in \mathbb{C}^K$ are the vectors of unknowns Lagrange multipliers; these can be calculated by solving the system in (15). $\boldsymbol{\Omega} \in \mathbb{R}^{K\times K}$ is the Kernel matrix defined below,

$$
\boldsymbol{\Omega} =
\begin{bmatrix}
\phi_1^T\phi_1 & \phi_2^T\phi_1 & \ldots & \phi_K^T\phi_1 \\
\phi_1^T\phi_2 & \phi_2^T\phi_2 & \ldots & \phi_K^T\phi_2 \\
\vdots & \vdots & \ddots & \vdots \\
\phi_1^T\phi_K & \phi_2^T\phi_K & \ldots & \phi_K^T\phi_K
\end{bmatrix}
\tag{16}
$$

In (15), $\boldsymbol{\Omega}_p \equiv \frac{\partial \boldsymbol{\Omega}}{\partial \xi_p} \in \mathbb{R}^{K\times K}$ is the derivative of the Kernel matrix, $\boldsymbol{\Omega}$, with respect to the parameter ξ_p, it is defined as

$$
\frac{\partial \boldsymbol{\Omega}}{\partial \xi_p} =
\begin{bmatrix}
\phi_1^T\phi_{p,1} & \phi_2^T\phi_{p,1} & \ldots & \phi_K^T\phi_{p,1} \\
\phi_1^T\phi_{p,2} & \phi_2^T\phi_{p,2} & \ldots & \phi_K^T\phi_{p,2} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_1^T\phi_{p,K} & \phi_2^T\phi_{p,K} & \ldots & \phi_K^T\phi_{p,K}
\end{bmatrix}
\tag{17}
$$

The term $\phi_1^T\phi_{k,p}$ in (17) is the derivative of the Kernel, $\kappa(\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(k)})$, with respect to ξ_p at sample $\boldsymbol{\xi}^{(k)}$; it is defined as [9],

$$\phi_1^T\phi_k^{(p)} \equiv \frac{\partial}{\partial \xi_p}\kappa(\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(k)}) \tag{18}$$

Similarly, $\boldsymbol{\Omega}_{p,q} \equiv \frac{\partial^2 \boldsymbol{\Omega}}{\partial \xi_p \partial \xi_q} \in \mathbb{R}^{K\times K}$ is the derivative of the Kernel matrix with respect to parameters ξ_p and ξ_q.

$$
\frac{\partial^2 \boldsymbol{\Omega}}{\partial \xi_p \partial \xi_q} =
\begin{bmatrix}
\phi_{q,1}^T\phi_{p,1} & \phi_{q,2}^T\phi_{p,1} & \ldots & \phi_{q,K}^T\phi_{p,1} \\
\phi_{q,1}^T\phi_{p,2} & \phi_{q,2}^T\phi_{p,2} & \ldots & \phi_{q,K}^T\phi_{p,21} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{q,1}^T\phi_{p,K} & \phi_{q,2}^T\phi_{p,K} & \ldots & \phi_{q,K}^T\phi_{p,K}
\end{bmatrix}
\tag{19}
$$

Similarly, in (19) $\phi_{q,j}\phi_{p,k}$ denotes the derivative of the Kernel with respect to parameters ξ_q and ξ_p at samples $\boldsymbol{\xi}^{(j)}$ and $\boldsymbol{\xi}^{(k)}$, respectively (17). It can be written as,

$$\phi_{q,j}\phi_{p,k}^T \equiv \frac{\partial^2}{\partial \xi_q \partial \xi_p}\kappa(\boldsymbol{\xi}^{(j)}, \boldsymbol{\xi}^{(k)}) \tag{20}$$

Upon solving (15), we obtain the values for the Lagrange multipliers, substituting (8) in (1), we obtain the expression for the surrogate model as a function of Lagrange multipliers,

$$\mathcal{M}(\boldsymbol{\xi}) = b + \sum_{k=1}^{K} \alpha_k \kappa(\boldsymbol{\xi}^{(k)}, \boldsymbol{\xi}) + \sum_{i=1}^{d}\sum_{k=1}^{K} \beta_{i,k} \frac{\partial \kappa(\boldsymbol{\xi}^{(k)}, \boldsymbol{\xi})}{\partial \xi_i} \tag{21}$$

As we can see in the above expression, the surrogate model computed using the proposed approach uses kernel derivatives as opposed to the LS-SVM based model [3]. Due to this, the model in (21) is accurate despite using fewer samples than the LS-SVM technique, as demonstrated in the next section.

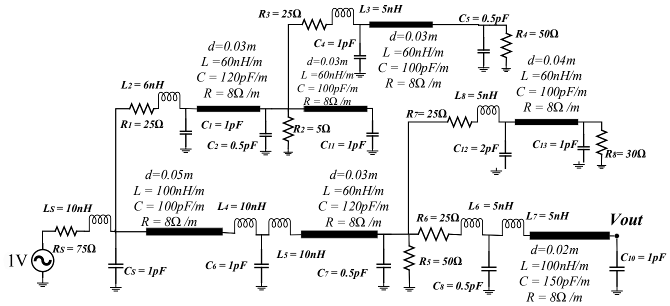

Fig. 1. Interconnect circuit.

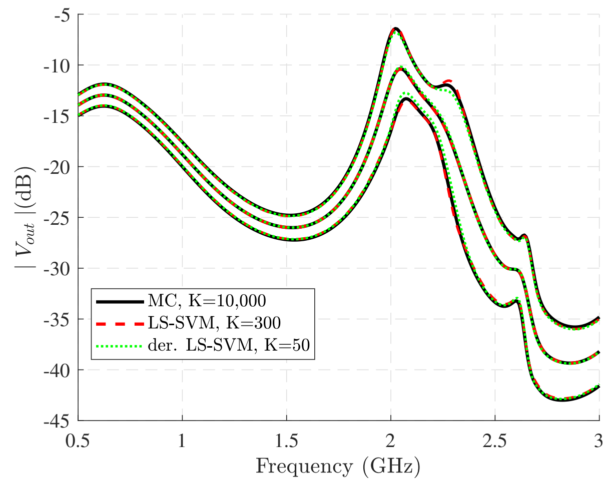

Fig. 2. Mean $\pm 3\sigma$ plot for $|V_{out}|$ of the interconnect circuit from 500MHz to 3GHz.

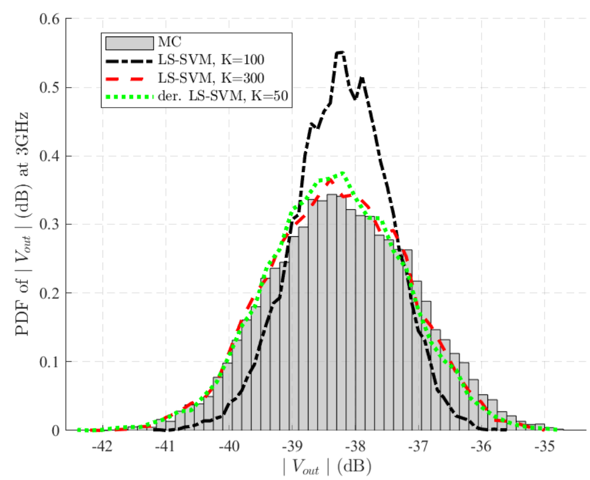

Fig. 3. Probability Distribution Function of $|V_{out}|$ for the interconnect circuit at 3GHz.

IV. Numerical Example

We performed the frequency analysis for 300 frequency points between 500MHz to 3GHz for the circuit shown in Figure 1 to test the performance of the proposed approach. We used $d = 25$ random parameters following a uniform distribution for the proposed approach. The circuit components used as random parameters are named, R_s, L_s, C_s, R_1, L_2, C_1, C_2, R_2, C_{11}, R_3, L_3, C_4, C_5 R_4, L_4, C_6, L_5, C_7, R_7, L_8, C_{12}, C_{13}, R_8, R_5, and R_6. We used a 10% relative uniform variation around their respective mean values shown in Figure 1. We compared the performance of the proposed approach with the Monte-Carlo method, which was run with $10,0000$ samples. For both the LS-SVM method and the proposed derivative LS-SVM (der. LS-SVM) method, we used a polynomial kernel of degree 3. The LS-SVM required $K = 300$ simulations to accurately compute the surrogate model for QoI, whereas the proposed approach only needed $K = 50$ samples. As it can be seen in Figure 2, the proposed approach accurately computes the mean and standard deviation of the V_{out}. Figure 3 shows the PDF obtained with $K = 100$ and $K = 300$ simulations for the LS-SVM approach. As can be seen in Figure 3, the PDF obtained with $K = 100$ simulations using LS-SVM techniques does not match the PDF obtained using the Monte-Carlo technique, while the PDF obtained with $K = 50$ simulations using the proposed der. LS-SVM technique accurately models the PDF of $|V_{out}|$.

V. Conclusion

This paper presented the sensitivity information to reduce the computational complexity of the LS-SVM approach. For the example shown, it was demonstrated that the proposed approach could compute the statistical properties of the QoI with similar accuracy with six times fewer simulations compared to the LS-SVM method.

References

[1] P. Manfredi and S. Grivet-Talocia, "Rational polynomial chaos expansions for the stochastic macromodeling of network responses," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 1, pp. 225–234, 2019.

[2] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, pp. 293–300, 1999.

[3] M. Larbi, R. Trinchero, F. G. Canavero, P. Besnier, and M. Swaminathan, "Analysis of parameter variability in an integrated wireless power transfer system via partial least-squares regression," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 10, no. 11, pp. 1795–1802, 2020.

[4] P. Manfredi and R. Trinchero, "A probabilistic machine learning approach for the uncertainty quantification of electronic circuits based on gaussian process regression," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 8, pp. 2638–2651, 2021.

[5] K. S. Sidhu, M. T. Kassis, and R. Khazaka, "Efficient regression-based polynomial chaos using adjoint sensitivity," in *2019 IEEE 28th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*. IEEE, 2019, pp. 1–3.

[6] M. Lázaro, I. Santamaría, F. Pérez-Cruz, and A. Artés-Rodríguez, "Support vector regression for the simultaneous learning of a multivariate function and its derivatives," *Neurocomputing*, vol. 69, no. 1-3, pp. 42–61, 2005.

[7] D. Eriksson, K. Dong, E. Lee, D. Bindel, and A. G. Wilson, "Scaling gaussian process regression with derivatives," *Advances in neural information processing systems*, vol. 31, 2018.

[8] N. Soleimani and R. Trinchero, "Compressed complex-valued least squares support vector machine regression for modeling of the frequency-domain responses of electromagnetic structures," *Electronics*, vol. 11, no. 4, p. 551, 2022.

[9] J. E. Johnson, V. Laparra, A. Pérez-Suay, M. D. Mahecha, and G. Camps-Valls, "Kernel methods and their derivatives: Concept and perspectives for the earth system sciences," *Plos one*, vol. 15, no. 10, p. e0235885, 2020.