A Multi-Output Active Learning Method for the Uncertainty Quantification of PCB Lines

A Multi-Output Active Learning Method for the Uncertainty Quantification of PCB Lines

Michele Cusano, Riccardo Trinchero, Igor S. Stievano,
Stefano Grivet-Talocia, Paolo Manfredi
Politecnico di Torino, 10129 Torino, Italy
E-mail: michele.cusano@polito.it

Stefanie Schatt
Continental Automotive Technologies GmbH
93055 Regensburg, Germany

***Abstract*—This paper introduces a novel active learning method for iteratively building a suitable training set within the scope of Gaussian process models applied to the uncertainty quantification of multi-output system responses. In particular, the proposed strategy offers a significant advantage by extending the already existing generic techniques that typically account for a single scalar output only. This approach extends the training to all outputs, thus improving the accuracy of the overall response of the system in a more comprehensive manner.**

***Index Terms*—Active learning, Bayesian estimation, Gaussian processes, machine learning, surrogate modeling, training, uncertainty quantification.**

I. Introduction

Nowadays, the need for new, smarter strategies to perform both design and optimization of electronic systems, along with their uncertainty quantification (UQ), is rapidly increasing. To this end, recent studies have focused on the use of machine learning techniques [1], [2], [3], with particular attention to kernel-based methods, whose structure is simple and flexible. What makes them suitable in this kind of application is their good scalability to high-dimensional analyses, which also implies a reasonable cost for training the models.

Gaussian process (GP) models were proven to be particularly successful in the design exploration and optimization of electronic devices due to their capability of associating confidence bounds to model predictions [4]. One of the most critical steps in the achievement of a good model is the selection of the training data, which should be as small as possible, while still allowing to approximate well the true system behavior.

Several approaches were proposed in the past years concerning active learning (AL) methods tailored to the effective augmentation of the training dataset [2], [5], [6]. However, most strategies addressed the optimization of a single scalar output quantity, which was generally selected based on the level of variability exhibited.

In this paper, an effective multi-output AL strategy for GP surrogates applied to UQ tasks is proposed and investigated. The algorithm, starting from the already obtained insights on this topic, relies on the predictive variance of the output mean, focusing this time on the overall time/frequency domain. The methodology leverages the concepts of principal component analysis (PCA) compression and Cholesky factorization to make the computation efficient. The proposed AL method is illustrated based on the full-wave simulation of a microstrip line with a slot in the ground plane.

II. GP Regression for Multiple Responses

Let us assume a generic system whose frequency-domain behavior is described by

$$y = \mathcal{M}(f; \boldsymbol{x}), \tag{1}$$

where f denotes the frequency, $\boldsymbol{x} = (x_1, ..., x_d)^\top$ is a set of (uncertain) design parameters, and $\mathcal{M} : \mathbb{R}^+ \times \mathcal{X} \subseteq \mathbb{R}^d \to \mathbb{R}$. The system (1) is evaluated at discrete frequency points $\{f_p\}_{p=1}^P$, e.g., by means of a full-wave solver, producing a vectorial output response $\boldsymbol{y} = (y_1, \ldots, y_P)^\mathsf{T}$ for a given configuration of the design parameters $\boldsymbol{x}$.

A. GP Regression Settings for Single Outputs

The output of (1) for the p-th frequency point can be expressed as a particular realization of a GP (prior) with trend $\mu(\boldsymbol{x})$ and kernel function $k(\boldsymbol{x}, \boldsymbol{x'})$. Without loss of generality, we consider here $\mu(\boldsymbol{x}) = 0$ and an anisotropic Matérn 5/2 kernel function.

For the training step, a set of input-output observations $\{(\boldsymbol{x}_l^{\mathrm{tr}}, y_l^{\mathrm{tr}})\}_{l=1}^L$ is collected from the actual system (1) and used to condition the prior. By evaluating the (posterior) model at a discrete set $\{\boldsymbol{x}_*\}_{i=1}^N$ of input values, we obtain a set of predictions $\boldsymbol{y}^*$ with mean vector

$$\boldsymbol{m} = \boldsymbol{K}_* \boldsymbol{K}^{-1} \boldsymbol{y}^{\mathrm{tr}} \tag{2}$$

and covariance matrix

$$\boldsymbol{C} = \boldsymbol{K}_{**} - \boldsymbol{K}_*^\top \boldsymbol{K}^{-1} \boldsymbol{K}_*, \tag{3}$$

where $\boldsymbol{y}^{\mathrm{tr}} \in \mathbb{R}^L$ collects the training responses, whereas

- $\boldsymbol{K} \in \mathbb{R}^{L \times L}$ has entries $K_{i,j} = k\left(\boldsymbol{x}_i^{\mathrm{tr}}, \boldsymbol{x}_j^{\mathrm{tr}}\right)$;
- $\boldsymbol{K}_* \in \mathbb{R}^{L \times N}$ has entries $K_{*_{i,j}} = k\left(\boldsymbol{x}_i^{\mathrm{tr}}, \boldsymbol{x}_j^{*}\right)$;
- $\boldsymbol{K}_{**} \in \mathbb{R}^{N \times N}$ has entries $K_{**_{i,j}} = k\left(\boldsymbol{x}_i^{*}, \boldsymbol{x}_j^{*}\right)$.

Considering that each prediction y_i^* is a multivariate normal variable withe mean and covariance given by equations (2) and (3), one can compute the expectation and variance of the output mean as [4]

$$\mathrm{E}\left\{\hat{\mu}_y\right\} = \frac{1}{N} \sum_{i=1}^{N} m_i \tag{4}$$

$$\mathrm{Var}\left\{\hat{\mu}_y\right\} = \frac{1}{N^2}\sum_{i,j=1}^{N} C_{ij}. \tag{5}$$

The second equation will be particularly useful in the next section, related to the proposed AL scheme.

B. PCA Compression for Multiple Outputs

An elementary approach would be to train P separate models for each frequency component y_p. Anyway, this soon becomes intractable as the number of points P increases. Moreover, it would unavoidably yield a different sequence of training input samples, which is inconvenient as a single simulation is usually run for an entire frequency sweep.

A viable option is to resort to PCA compression, which allows obtaining, thanks to its linearity, a model for the original outputs by training separate GP models for the principal components only. Indeed, the training observations now consist of vectors $\boldsymbol{y}_l^{\mathrm{tr}} = \mathcal{M}(f, \boldsymbol{x}_l)$ and form a matrix dataset $\boldsymbol{Y}^{\mathrm{tr}} \in \mathbb{R}^{P\times L}$. This can be approximated by a truncated singular value decomposition as

$$\boldsymbol{Y}^{\mathrm{tr}} = \bar{\boldsymbol{y}} + \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\top} \approx \bar{\boldsymbol{y}} + \tilde{\boldsymbol{U}}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{V}}^{\top} = \bar{\boldsymbol{y}} + \tilde{\boldsymbol{U}}\tilde{\boldsymbol{Z}}, \tag{6}$$

where $\bar{\boldsymbol{y}} \in \mathbb{R}^P$ is the mean along the columns of $\boldsymbol{Y}^{\mathrm{tr}}$. The matrix $\tilde{\boldsymbol{Z}} = \tilde{\boldsymbol{U}}^{\top}(\boldsymbol{Y}^{\mathrm{tr}} - \bar{\boldsymbol{y}}) \in \mathbb{R}^{\tilde{n}\times L}$ can be seen as a collection of L training responses of the "principal components" z_n of the system, which relate to the original frequency components y_p as

$$y_p = \bar{y}_p + \sum_{n=1}^{\tilde{n}} U_{pn} z_n. \tag{7}$$

The number $\tilde{n}$ of principal components is obtained based on a predefined relative threshold on the singular values, which in this application is chosen to be 1%, and is typically $\ll P$.

The principal components z_n can be modeled with $\tilde{n}$ individual GP surrogates with corresponding mean and covariance $\tilde{\boldsymbol{m}}_n$ and $\tilde{\boldsymbol{C}}_n$ as in (2) and (3), respectively. From these reduced outputs, the original outputs are recovered via (7). Thanks to the linearity of PCA, the mean and covariance of the predictions for the p-th output are obtained as [4]

$$\boldsymbol{m}_p = \bar{y}_p + \sum_{n=1}^{\tilde{n}} U_{pn}\tilde{\boldsymbol{m}}_n \tag{8}$$

and

$$\boldsymbol{C}_p = \sum_{n=1}^{\tilde{n}} U_{pn}^2 \tilde{\boldsymbol{C}}_n, \tag{9}$$

respectively, for $p = 1, ..., P$.

III. Proposed Multi-Output AL

The proposed strategy to iteratively increment the training dataset builds upon the stochastic AL method presented in [7], which adds new points with the aim of reducing the predictive variance of the output mean given by (5). Here, we extend the method to the multi-output case, considering the overall response instead of a single frequency.

When using PCA compression, we are dealing with $\tilde{n}$ separate models. At each iteration we focus on the principal component with the largest predictive variance of the corresponding mean as the optimization criterion, i.e.,

$$n^{\star} = \arg\max_{n} \mathrm{Var}\left\{\hat{\mu}_{z_n}\right\} = \arg\max_{n} \frac{1}{N^2}\sum_{i,j=1}^{N} \tilde{C}_{n,ij}. \tag{10}$$

In this way, we track the overall model variation, focusing at each iteration on the least accurate component to improve.

Since the term $\boldsymbol{K}_{**}$ in (3) is not affected by the expansion of the training dataset with a new point $\boldsymbol{x}_{\mathrm{new}}$, we focus on the update of the second term, which we express as

$$\Delta\tilde{\boldsymbol{C}}_{n^{\star}}^{(\nu+1)} = \begin{bmatrix}\boldsymbol{K}_{*}^{(\nu)\top} & \boldsymbol{\kappa}_{*}^{\top}\end{bmatrix}\begin{bmatrix}\boldsymbol{K}^{(\nu)} & \boldsymbol{\kappa}^{\top}\\ \boldsymbol{\kappa} & \kappa\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{K}_{*}^{(\nu)}\\ \boldsymbol{\kappa}_{*}\end{bmatrix} \tag{11}$$

where ν denotes the iteration index, matrices $\boldsymbol{K}^{(\nu)}$ and $\boldsymbol{K}_{*}^{(\nu)}$ are available from the previous iteration, whereas $\kappa_{*,i} = k\left(\boldsymbol{x}_{\mathrm{new}}, \boldsymbol{x}_i^{*}\right)$, $\kappa_i = k\left(\boldsymbol{x}_i^{\mathrm{tr}}, \boldsymbol{x}_{\mathrm{new}}\right)$, and $\kappa = k\left(\boldsymbol{x}_{\mathrm{new}}, \boldsymbol{x}_{\mathrm{new}}\right)$ depend on the new candidate point.

In essence, (11) describes the decrease of the predictive variance of the principal component mean when a new point $\boldsymbol{x}_{\mathrm{new}}$ is added to training dataset. Therefore, the best new point is selected as the one solving the optimization problem

$$\boldsymbol{x}_{\mathrm{new}}^{\star} = \arg\max_{\boldsymbol{x}_{\mathrm{new}}\in\mathcal{X}} \sum_{i,j=1}^{N} \Delta\tilde{\boldsymbol{C}}_{n^{\star},ij}^{(\nu+1)}. \tag{12}$$

Solving the optimization problem involves the evaluation of the cost function (11) for a possibly large number of candidate points $\boldsymbol{x}_{\mathrm{new}}$. This in turn requires the inversion of a matrix in which only one row/column is changed. To speed-up this calculation, the Cholesky factorization of the inner matrix is computed directly as an update of the Cholesky factorization of matrix $\boldsymbol{K}^{(\nu)} = \boldsymbol{L}\boldsymbol{L}^{\top}$.

IV. Application Example and Numerical Results

The proposed multi-output AL technique is applied to the UQ of the insertion loss of a microstrip line with a discontinuity in the ground plane. The stochastic input parameters are the location and the length of the slot in the ground plane, while the magnitude of S_{21}, analyzed with CST Studio Suite® for a frequency sweep spanning from dc to 10 GHz, is considered as the output. The Reader is referred to [7] and references therein for additional details on this test case, for which the insertion loss was investigated at a single frequency.

The training and validation datasets are built based on an overall set of 1000 input-output observations. Specifically, $N = 300$ observations are assumed as a reference and reserved for an independent validation, whereas the remaining 700 samples are used as candidates from which to pick the training samples. The training starts with $L = 10$ observations only. Then, 15 iterations are performed to expand the training dataset by selecting the best candidate from the remaining points according to the proposed AL scheme. The GP models are trained in MATLAB® using the Statistics and Machine Learning Toolbox™ toolbox.

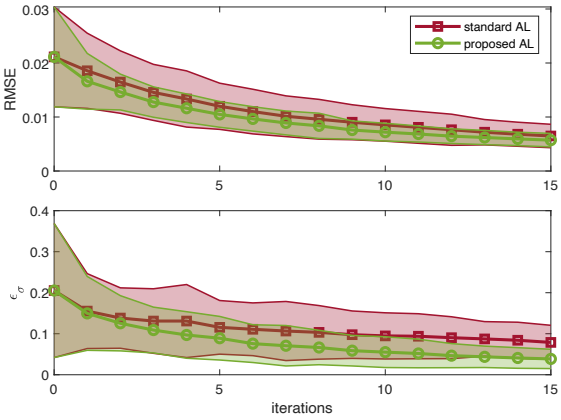

Fig. 1. RMSE (top) and ϵ_σ (bottom) over 50 runs as the iterations increase. Red curves: standard AL [5]; green curves: proposed AL.

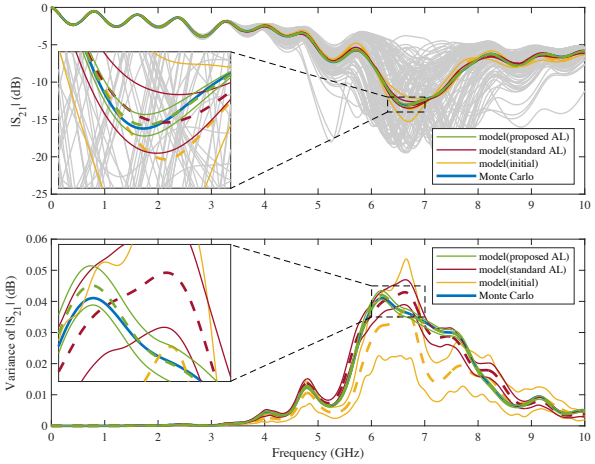

Fig. 2. Mean (top) and variance (bottom) of S_{21} over frequency. Blue: reference from Monte Carlo analysis; yellow: initial GP model (10 samples); red and green: GP models after 15 iterations with standard and proposed AL, respectively; gray lines: samples from the Monte Carlo analysis.

The accuracy is assessed based on the mean of the relative errors on the variance at each frequency point, i.e.,

$$\epsilon_\sigma = \frac{1}{P}\sum_{p=1}^{P}\frac{\left|\hat{\sigma}^2_{y_p} - \sigma^2_{y_p}\right|}{\sigma^2_{y_p}}, \tag{13}$$

where $\sigma^2_{y_p}$ is the variance of the reference observations in the validation set and $\hat{\sigma}^2_{y_p}$ the corresponding prediction of the GP model, and the mean of the RMSE between observations and predictions, i.e.,

$$\text{RMSE} = \frac{1}{P}\sum_{p=1}^{P}\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\hat{y}_{p_j} - y_{p_j}\right)^2}. \tag{14}$$

The performance of the proposed AL strategy is compared against a more traditional strategy that uses the maximum predictive variance of the principal components (instead of their mean) as optimization criterion [5], which corresponds to the maximum diagonal entry in the covariance matrix $\tilde{\boldsymbol{C}}_n$ computed according to (3).

Figure 1 shows the behavior of the above error metrics for the two AL strategies and for 50 independent runs. For each run, the same initial set of 10 training samples, randomly picked from the candidate set, is considered for the two approaches. Then, 15 additional training samples are iteratively selected based on the corresponding algorithm. The line with markers represent the median error over the 50 runs, whereas the shaded area represent the dispersion of the results (2σ interval). The proposed AL method is shown to outperform the standard deterministic strategy, as it provides faster convergence and tighter confidence bounds as the number of iterations (i.e., added samples) increases.

The moments of S_{21} predicted over frequency with the GP models are shown in Fig. 2 for one execution of the two AL schemes. The dashed red and green curves refer to the GP predictions obtained by applying the conventional and the proposed AL strategies, respectively, whereas the solid lines are their respective 95% confidence intervals. The prediction of the starting model obtained with only 10 samples (yellow curves) are included for comparison. Once again, the proposed method is found to substantially outperform the standard AL method by yielding a far more accurate model.

V. Conclusions

This paper presented a new methodology to iteratively train a multi-output GP model in the context of UQ. The proposed AL scheme accounts for the overall frequency-domain response of the system and was applied to UQ of the insertion loss of a microstrip line with ground plane discontinuity, for which it achieved better convergence and lower error compared to a standard AL strategy for the same number of samples.

References

[1] M. B. Yelten, T. Zhu, S. Koziel, P. D. Franzon, and M. B. Steer, "Demystifying surrogate modeling for circuits and systems," *IEEE Circuits Syst. Mag.*, vol. 12, no. 1, pp. 45–63, 2012.

[2] M. Swaminathan, H. M. Torun, H. Yu, J. A. Hejase, and W. D. Becker, "Demystifying machine learning for signal and power integrity problems in packaging," *IEEE Trans. Compon., Packag., Manuf. Technol*, vol. 10, no. 8, pp. 1276–1295, Aug. 2020.

[3] T. Nguyen, B. Shi, H. Ma, E.-P. Li, X. Chen, A. C. Cangellaris, and J. Schutt-Aine, "Comparative study of surrogate modeling methods for signal integrity and microwave circuit applications," *IEEE Trans. Compon., Packag., Manuf. Technol*, vol. 11, no. 9, pp. 1369–1379, Sep. 2021.

[4] P. Manfredi and R. Trinchero, "A probabilistic machine learning approach for the uncertainty quantification of electronic circuits based on Gaussian process regression," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 41, no. 8, pp. 2638–2651, 2022.

[5] R. Trinchero and F. Canavero, "Use of an active learning strategy based on Gaussian process regression for the uncertainty quantification of electronic devices," *Eng. Proc.*, vol. 3, no. 1, p. 3, 2020.

[6] H. Mohammadi, P. Challenor, D. Williamson, and M. Goodfellow, "Cross-validation-based adaptive sampling for Gaussian process models," *SIAM/ASA J. Uncertain. Quantif.*, vol. 10, no. 1, pp. 294–316, Feb. 2022.

[7] P. Manfredi, "A stochastic active learning strategy for Gaussian process models with application to the uncertainty quantification of signal integrity," in *Proc. IEEE Elect. Des. Adv. Packag. Syst. Symp. (EDAPS)*, 2023, pp. 1–3.