# A general square exponential kernel to handle mixed-categorical variables for Gaussian process

Paul Saves, Youssef Diouane, Nathalie Bartoli, Thierry Lefebvre, Joseph Morlier

# A general square exponential kernel to handle mixed-categorical variables for Gaussian process

P. Saves*
*ONERA, DTIS, Université de Toulouse, Toulouse, France*
*ISAE-SUPAERO, Université de Toulouse, Toulouse, France*

Y. Diouane†
*Polytechnique Montréal, Montreal, QC, Canada*

N. Bartoli ‡, T. Lefebvre§
*ONERA, DTIS, Université de Toulouse, Toulouse, France*

J. Morlier¶
*ICA, Université de Toulouse, ISAE-SUPAERO, MINES ALBI, UPS, INSA, CNRS, Toulouse, France*

**Recently, there has been a growing interest for mixed categorical meta-models based on Gaussian process (GP) surrogates. In this setting, several existing approaches use different strategies. Among the recently developed methods, we could cite: GP models built using continuous relaxation of the variables, Gower distance based models or GP models derived from direct estimation of the correlation matrix.**

**In this paper, we present a kernel-based approach that extends continuous Gaussian kernels to handle mixed-categorical variables. The proposed kernel leads to a GP surrogate that generalizes continuous relaxation and Gower distance based GP models. The good potential of the proposed framework is shown on analytical mixed-categorical variables test cases. On different settings, our proposed GP models is as accurate as the state-of-the-art GP models.**

## I. Nomenclature

| | | |
|---|---|---|
| $n$ | = | number of continuous variables |
| $m$ | = | number of integer variables |
| $l$ | = | number of categorical variables |
| $\Omega \in \mathbb{R}^n$ | = | continuous space |
| $S \in \mathbb{Z}^m$ | = | integer space |
| $\mathbb{F}^l$ | = | categorical space |
| $L_i, i \in \{1, \ldots, l\}$ | = | number of levels for the $i^{th}$ categorical variable |
| $\theta^{cont}$ | = | vector of hyperparameters for the continuous part of the Gaussian process model |
| $k$ | = | correlation kernel |
| $\theta_j^{cont}, j \in \{1, \ldots, n+m\}$ | = | hyperparameter for the $j^{th}$ continuous or integer variable |
| $R^{cont}$ | = | correlation matrix for continuous and integer inputs |
| $\Theta^{cat}$ | = | hyperparameters for the categorical part of the Gaussian process model |
| $K_i$ | = | categorical kernel for the $i^{th}$ categorical variable |
| $\Theta_i$ | = | matrix of hyperparameters for the $i^{th}$ categorical variable |
| $R^{cat}$ | = | correlation matrix for categorical inputs |
| $\Theta = [\Theta^{cat}, \theta^{cont}]$ | = | hyperparameters for the Gaussian process model |
| $R = R^{cont} R^{cat}$ | = | correlation matrix for mixed integer inputs |

*PhD Student, Information Processing and Systems Department & Complexes Systems Engineering Department, paul.saves@isae-supaero.fr
†Professor, Mathematics and Industrial Engineering Department, youssef.diouane@polymtl.ca
‡Senior researcher, Information Processing and Systems Department, nathalie.bartoli@onera.fr, AIAA MDO TC Member.
§Research Engineer, Information Processing and Systems Department, thierry.lefebvre@onera.fr, AIAA Member.
¶Professor, Structural Mechanics, joseph.morlier@isae-supaero.fr, AIAA Member.

## II. Introduction

New aircraft configurations with a lower footprint on the environment (also known as Eco-aircraft design) have seen a resurgence of interest [1–3]. In this context, one targets to minimize the footprint on the environment of the aircraft using a *Multidisciplinary Design Analysis* (MDA) [4–6]. This is an example of an expensive-to-evaluate without derivative problem that could be encountered on industry. Therefore, it could be useful to use a surrogate model that simplifies by a lot an expensive model and gives a good approximation from a small data set of known configurations. An example of an industrial application of surrogate model in the context of aircraft design is given in Fig. 1.

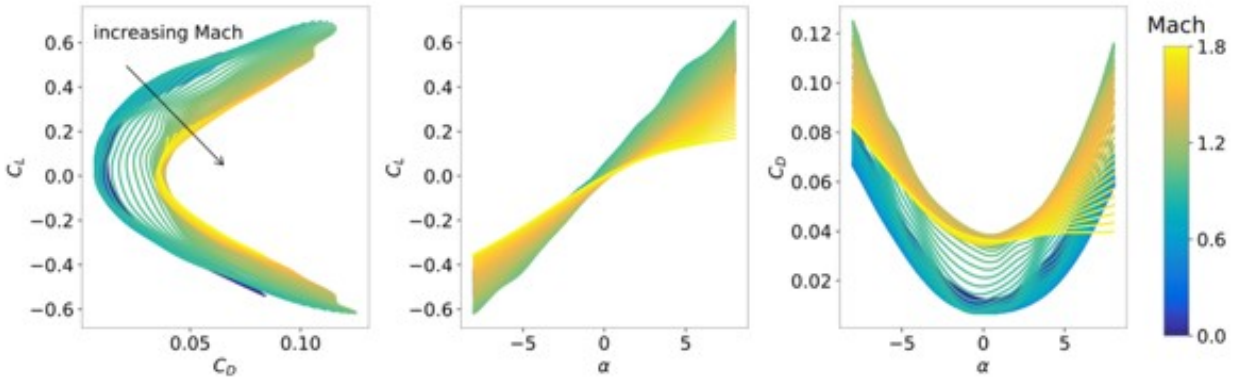



**Fig. 1 Drag polar and aerodynamic properties for an efficient supersonic air vehicle obtained from a surrogate model for different Mach speed and sweep angles [7, Figure 3].**

Nevertheless, in this context, the process generally involves mixed continuous-categorical design variables. For instance, the size of aircraft structural parts can be described using continuous variables; in case of thin-sheet stiffened sizing, they represent panel thicknesses and stiffening cross-sectional areas. The set of discrete variables can encompass design variables such as the number of panels, the list of cross sectional areas or the material choices.
In this work, we target to construct an inexpensive surrogate model $\hat{f}$ for a black-box simulation function of the form

$$f : \Omega \times S \times \mathbb{F}^l \to \mathbb{R}. \tag{1}$$

The function $f$ is typically expensive-to-evaluate simulations with no exploitable derivative information. $\Omega \subset \mathbb{R}^n$ represents the bounded continuous design set for the $n$ continuous variables. $S \subset \mathbb{Z}^m$ represents the bounded integer set where $L_1, \ldots, L_m$ are the numbers of levels of the $m$ quantitative integer variables on which we can define an order relation and $\mathbb{F}^l = \{1, \ldots, L_1\} \times \{1, \ldots, L_2\} \times \ldots \times \{1, \ldots, L_l\}$ is the design space for $l$ categorical qualitative variables with their respective $L_1, \ldots, L_l$ levels.

In this context, *Gaussian processes* (GP) [8–12], also called Kriging models, are known to be a good modelling strategy to define response surface models. Namely, we will consider that our unknown black-box function $f$ is a realization of an underlying GP of mean $\mu^f$ and of standard deviation $s^f$, i.e.,

$$f \sim \hat{f} = \mathrm{GP}\left(\mu^f, [s^f]^2\right). \tag{2}$$

For a general problem involving categorical or integer variables, several modeling strategies to build a mixed-categorical GP have been proposed [13–18]. Compared to a continuous GP, the major changes are in the estimation of the correlation matrix: the latter is essential for building estimates of $\mu^f$ and $s^f$. Similarly to the process of constructing a GP with continuous inputs, relaxation techniques [15] and Gower distance based models [16] use a kernel-based approach to estimate the correlation matrix. Other recent approaches try to estimate the correlation matrix directly independently of a kernel choice [13, 14, 18] which shows good results as these methods model completely the correlations. However, the direct estimation of the correlation matrix as proposed in [13, 14] is not adapted for high-dimensional problems as it is very expensive to compute all the required hyperparameters. In fact, dimension reduction methods such as principal components analysis (known as KPLS [19, 20], Kriging model with Partial Least Squares) require the construction of the correlation matrix via a kernel function. KPLS models are used to reduce to number of hyperparameters and to handle a large number of mixed inputs [21].

In this work, we target to extend the classical paradigm used for continuous inputs to cover the mixed categorical case. We will present a kernel-based approach that will lead to a unified approach for existing approximation methods [13–16]. A similar process for the estimation of the hyperparameters could be applied to both continuous and categorical inputs. The good potential of the proposed approach is shown over a set of analytical test cases. In this paper, a particular attention will be given to the Gaussian kernel, but our proposed approach can be straightforwardly extended to other existing kernels.

The reminder of this paper is as follows. In Section III, a detailed review of the GP model for continuous and for categorical inputs is given. The extended kernel-based approach for constructing the correlation matrix is presented in Section IV. Section V presents academical tests as well as the obtained results. Conclusions and perspectives are finally drawn in Section VI.

## III. Gaussian process meta-models for mixed categorical inputs

In general, a GP model is used to fit a response surface model from an initial set of points, known as the Design of Experiments (DoE) [10, 11, 22]. The GP provides a mean response hypersurface as well as a pointwise estimation of its variance. In what comes next, let $n_t$ be the size of DoE data set $(x, y^f)$.

### A. GP meta-models for continuous inputs

In this subsection, we will only consider that all the design variables are continuous in problem (1): namely, the design space will be restricted to $\Omega \subset \mathbb{R}^n$. In this case, we assume that $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is defined only over the continuous design space: $n_t$ is the number of already evaluated points in $\mathbb{R}^n$ of the deterministic function $f$ and $\forall r \in \{1, .., n_t\}$. Let $x^r = (x_1^r, ..., x_n^r) \in \mathbb{R}^n$ be the $r^{th}$ point with its respective $n$ continuous variable values and ${y^f}_r \in \mathbb{R}$ be the associated values of $f(x^r)$ and denote the DoE as $(x, y^f)$. The stochastic model [23] writes as: $\hat{f}(x) = \mu(x) + \epsilon(x) \in \mathbb{R}$ with $\epsilon$ the error term between $f$ and the model approximation $\mu(x)$. The error terms are considered as independent and identically distributed random variables of variance $\sigma^2$. Let $R^{cont}$ be the correlation matrix between the input points whose elements are defined by

$$[R^{cont}]_{r,s} = Corr\left(\epsilon(x^r), \epsilon(x^s)\right)$$

The correlation function $Corr$ is computed using a kernel function $k$ that relies on $n$ hyperparameters $\theta^{cont}$ estimated typically using maximum likelihood estimator (MLE) [24]:

$$Corr(.,.) = k(.,., \theta^{cont})$$

Let $r^{cont}(x^*) = (Corr(\epsilon(x^*), \epsilon(x^1)), \ldots, Corr(\epsilon(x^*), \epsilon(x^{n_t})))$ for a given $x^*$ and $\mathbb{1}$ be the $n_t$ vector of ones, then, we have:

$$\mu^f(x^*) = \hat{\mu}^f + r^{cont}(x^*)^T [R^{cont}]^{-1} (\mathbf{y}^f - \mathbb{1}\hat{\mu}^f), \tag{3}$$

and

$$[s^f]^2(x^*) = [\hat{\sigma}^f]^2 \left[1 - r^{cont}(x^*)^T [R^{cont}]^{-1} r^{cont}(x^*) + \frac{(1 - \mathbb{1}^T [R^{cont}]^{-1} r^{cont}(x^*))^2}{\mathbb{1}^T [R^{cont}]^{-1} \mathbb{1}}\right], \tag{4}$$

where $\hat{\mu}^f$ and $\hat{\sigma}^f$, respectively, are the MLE of $\mu$ and $\sigma$ with respect to $\theta^{cont}$ given the DoE data set $(x, y^f)$. In these formulae, $\hat{f}$ and $[s^f]^2$ both depend on $R$ and $r$ which are characterized by the correlation kernel $k(., ., \theta^{cont})$. For two continuous inputs $x^r$ and $x^s$, the Gaussian kernel is defined as:

$$k(x^r, x^s, \theta^{cont}) = \prod_{j=1}^{n} \exp\left(-\theta_j^{cont} \left(x_j^r - x_j^s\right)^2\right) = \prod_{j=1}^{n} \exp\left(-|x_j^r - x_j^s| \theta_j^{cont} |x_j^r - x_j^s|\right) \tag{5}$$

Other kernels can be used like the Matérn 3/2 kernel [25]:

$$k(x^r, x^s, \theta^{cont}) = \prod_{j=1}^{n} \left(1 + \sqrt{3}\theta_j^{cont} \left|x_j^r - x_j^s\right|\right) \exp\left(-\sqrt{3}\theta_j^{cont} \left|x_j^r - x_j^s\right|\right) \tag{6}$$

As the hyperparameters are always multiplied by the distance between two points, they can be interpreted as being the inverse correlation length. A GP whose kernel is stationary and based on a distance between two points is well adapted to a continuous context and its extension to categorical or integer variables is not straightforward. In the next part, for the general mixed-categorical case, we will consider both continous and categorical hyperparameters $\Theta$.

### B. GP meta-models for mixed-categorical inputs

In this subsection, we are considering the case where the design variables could be categorical or integer. Namely, we assume that $f : \mathbb{R}^n \times \mathbb{Z}^m \times \mathbb{F}^l \mapsto \mathbb{R}$. The set $\mathbb{F}^l = \{1, \ldots, L_1\} \times \{1, \ldots, L_2\} \times \ldots \times \{1, \ldots, L_l\}$ is the design space for $l$ categorical qualitative variables with their respective $L_1, ..., L_l$ levels. Our goal is to build a GP surrogate model for $f$. In this case, the GP model will be constructed following the same methodology used for continuous design space (see Eq. (3) and Eq. (4)). The only changes are related to the construction of the correlation matrix $R$. In fact, for a given couple $(r, s) \in (\{1, \ldots, n_t\})^2$, let $w^r = (x^r, z^r, c^r) \in \mathbb{R}^n \times \mathbb{Z}^m \times \mathbb{F}^l$ and $w^s = (x^s, z^s, c^s) \in \mathbb{R}^n \times \mathbb{Z}^m \times \mathbb{F}^l$ two points from the design space. In this case, the correlation kernel [14] is given by the product of continuous and categorical kernels as:

$$k(w^r, w^s, \Theta) = k^{cat}\left(c^r, c^s, \Theta^{cat}\right) k^{cont}\left((x^r, z^r), (x^s, z^s), \theta^{cont}\right), \tag{7}$$

where $\Theta = [\Theta^{cat}, \theta^{cont}]$, the kernel $k^{cont}((.,.), (.,.), \theta^{cont})$ is constructed efficiently as before (with the continuous relaxation of the integer inputs $z$) and the term $k^{cat}(.,., \Theta^{cat})$ is a categorical kernel [26] that depends on a matrix of hyperparameters $\Theta^{cat}$. Using Eq. (7), we have

$$[R]_{r,s}(\Theta) = Corr(w^r, w^s) = [R]^{cat}_{r,s}(\Theta^{cat})[R]^{cont}_{r,s}(\theta^{cont}), \tag{8}$$

where $R^{cont}_{r,s}(\theta^{cont}) = k^{cont}((x^r, z^r), (x^s, z^s), \theta^{cont})$ and $R^{cat}_{r,s}(\Theta^{cat}) = k^{cat}(c^r, c^s, \Theta^{cat})$. In the general setting, the categorical kernel $k^{cat}$ needs to be chosen such that the correlation matrix $R^{cat}$ is symmetric positive definite (SPD) [13, 14]. For the categorical inputs, the kernel $k^{cat}(c^r, c^s, \Theta^{cat})$ is constructed on the following way. $\forall i \in \{1, \ldots, l\}$, $c^r_i$ is the level taken by the $i^{th}$ component of the input $c^r$. As in Pelamatti et al. [27], let $k^{cat}$ be formulated level-wise as:

$$k^{cat}(c^r, c^s, \Theta^{cat}) = \prod_{i=1}^{l} K_i\left(c^r_i, c^s_i, \Theta_i\right)$$

where every sub-kernel $K_i$ is associated with a correlation matrix $R_i$ that contains the correlations between the various levels of the categorical variable $i$. Namely, we have

$$R^{cat}_{r,s}(\Theta^{cat}) = \prod_{i=1}^{l} [R_i]_{r,s}(\Theta_i).$$

Thus, the hyperparameters $\Theta^{cat}$ can be seen as a concatenation of the set of matrices $\Theta_1, \ldots, \Theta_l$, i.e., $\Theta^{cat} = [\Theta_1, \ldots, \Theta_l]$. The full set of hyperparameters $\Theta$ will be estimated using the DoE data set $(x, y^f)$ via an MLE approach on the following way

$$\Theta^* = \arg\max_{\Theta}\left(-\frac{1}{2}{y^f}^T[R(\Theta)]^{-1}y^f - \frac{1}{2}\log|[R(\Theta)]| - \frac{n_t}{2}\log 2\pi\right), \tag{9}$$

where $R(\Theta)$ is computed using Eq. (8).

## IV. Towards a general correlation matrix representation for a Gaussian kernel

We propose a novel approach that tackles the problem of extending correlation kernels to categorical variables by replacing the distance between input points with the kernel that depends only on the hyperparameters. In this section, we will present first the mathematical framework for the Gaussian kernel which fits well with classical dimension reduction techniques such as KPLS [19]. The Gaussian kernel has a natural extension to the use of hyperparameters in a matrix form that is an usual form when handling categorical design variables. The proposed extended model will lead to a generalization of both continuous relaxation and Gower distance based methods.

For our purposes, the treatment of continuous inputs will not bring any additional difficulty. Thus, without loss of generality, we will consider only categorical inputs. Hence, in the following, we assume $\Theta = \Theta^{cat}$ and $R = R^{cat}$.

### A. An extended correlation matrix approach for Gaussian kernel

In the case of Gaussian kernel, a natural extension of hyperparameters to the matrix form can be as follows. Starting from (5) and replacing the vector $\theta^{cont}$ by a given symmetric matrix $\Theta$, we obtain the following kernel:

$$k(x^r, x^s, \Theta) = \prod_{j=1}^{n}\prod_{j'=1}^{n}\exp\left(-|x^r_j - x^s_j|[\Theta]_{j,j'}|x^r_{j'} - x^s_{j'}|\right) \tag{10}$$

If $\Theta$ is a diagonal matrix with only positive values, then the resulting kernel would be the exact Gaussian kernel for continuous inputs with the hyperparameters being the inverse correlation length. This kernel, given by Eq. (10) generalizes the continuous Gaussian kernel for a 2D correlation matrix.

Now, for a given $i \in \{1, \ldots, l\}$, let $c_i$ be a categorical variable characterized by $L_i$ levels. The mapping to a $L_i$-dimensional Hilbert space defined in such a way that the only non-zero coordinate of the image is 1 in the dimension associated to the mapped level is the so-called one-hot encoding [28]. Let $e_{c_i}$ be the one-hot encoding of $c_i$ that takes value 0 everywhere and value 1 on the dimension corresponding to the level taken by the category $c$ on the variable $i$, $e_{c_i} \in \{0, 1\}^{L_i}$. For example, if $c$ takes the $j^{th}$ level on the variable $i$, $\left(e_{c_i}\right)_j = 1$ and $e_{c_i} = [0, \ldots, 0, 1, 0, \ldots, 0] \in \mathbb{R}^{L_i}$. Inspired by Eq. (10), it is now possible to define a $L_i \times L_i$ symmetric matrix $\Theta_i$ using positive correlation values. This leads to the following general formulation which is the natural extension of Gaussian kernel when dealing with a matrix of hyperparameters:

$$K_i(c_i^r, c_i^s, \Theta_i) = \prod_{j=1}^{L_i} \prod_{j'=1}^{L_i} \exp\left(-\left|(e_{c_i^r})_j - (e_{c_i^s})_j\right| [\Theta_i]_{j,j'} \left|(e_{c_i^r})_{j'} - (e_{c_i^s})_{j'}\right|\right) \tag{11}$$

Hence by the definition of $e_{c_i^r}$ and $e_{c_i^s}$, if $c_i^r = c_i^s$, one deduces that $K_i(c_i^r, c_i^r, \Theta_i) = \exp(0) = 1$. Otherwise, if $c_i^r \neq c_i^s$, one gets

$$\begin{aligned} K_i(c_i^r, c_i^s, \Theta_i) &= \exp\left(-\sum_{j=1}^{L_i} \sum_{j'=1}^{L_i} \left|(e_{c_i^r})_j - (e_{c_i^s})_j\right| [\Theta_i]_{j,j'} \left|(e_{c_i^r})_{j'} - (e_{c_i^s})_{j'}\right|\right) \\ &= \exp\left(-\left([\Theta_i]_{c_i^r,c_i^r} + [\Theta_i]_{c_i^s,c_i^s} + [\Theta_i]_{c_i^r,c_i^s} + [\Theta_i]_{c_i^r,c_i^s}\right)\right) \\ &= \exp\left(-\left([\Theta_i]_{c_i^r,c_i^r} + [\Theta_i]_{c_i^s,c_i^s}\right)\right) \exp\left(-2[\Theta_i]_{c_i^r,c_i^s}\right). \end{aligned} \tag{12}$$

where $[\Theta_i]_{c_i^r,c_i^s}$ is the coefficient characterizing the correlation between the two discrete categorical levels taken by $c^r$ and $c^s$ in the $i^{th}$ categorical component.

In the following, as far as the matrices $\Theta_i$ respect a specific parameterization, we will show that our approach will guarantee to the correlation matrix $R$ to be SPD with unit diagonal and off-diagonal term values in $[0, 1]$ [29]. The latter properties are needed to avoid numerical issues during the computations, see Eq. (3) and Eq. (4).
In fact, for all $i \in \{1, \ldots, l\}$, we propose to use the following parameterization for the hyperparameter matrix $\Theta_i$:

$$\begin{aligned} [\Theta_i]_{j,j} &\geq 0 \\ [\Theta_i]_{j,j'} &= \frac{\log \epsilon}{2}([C_i C_i^T]_{j,j'} - 1) \quad \text{if } j \neq j' \end{aligned} \tag{13}$$

where for every categorical variable $i$, $C_i$ is a Cholesky lower triangular matrix that relies on $L_i(L_i - 1)/2$ elements in $[0, \frac{\pi}{2}]$ that represent the coordinates of a point on the surface of a sphere with a unit radius as in [13, 26] (see Appendix VII.A for a proof in this context). The parameter $\epsilon$ is a small tolerance ($1 > \epsilon > 0$). Equation (13) is chosen such that the elements of $R_i$ will be in $[0, 1]$, thanks to the hypersphere decomposition [30, 31]. In the following, we will show how this model generalizes all the previous works and proposes a new framework and new matrix configurations.

### B. Equivalence with other categorical kernels

The proposed parameterization of the matrix $R_i$ using Eq. (12) can be seen as the product of the continuous relaxation kernel [15] and the Gaussian homoscedastic hypersphere kernel, for a total of $\frac{L_i(L_i+1)}{2}$ hyperparameters per categorical variable. Our proposed approach guarantees that the correlation matrix is SPD and that all its elements are in $[0, 1]$ as in the continuous case. It follows that our proposed kernel is also equivalent to the Gaussian homoscedastic hypersphere model alone.

Note that the hypersphere decomposition, as proposed by Pelamatti et al. [13], allows negative correlation. Indeed, for categorical variables, values between -1 and 0 are considered because two levels can correspond to opposite effects whereas for continuous GP, the furthest the points, the smaller the correlation. In this case, it follows that our proposed Gaussian homoscedastic hypersphere model $K_i(c_i^r, c_i^s, \Theta_i) = \exp\left(-2[\Theta_i]_{c_i^r,c_i^s}\right)$ is not completely equivalent to the model proposed in [13] that does not use a positive kernel (like the Gaussian one). A negative correlation would be

close to 0 through Gaussian kernel. Indeed, it is equivalent if and only if the correlations are strictly positive, *i.e* on $[0, \frac{\pi}{2}]$(Appendix VII.A).

In what comes next, this GP model using our kernel-based approach will be called the homogeneous full model (defined in Eq. (12)). It allows to generalize existing approaches in the following way:

- If $\forall i \in \{1, \dots, l\}$ the matrices $\Theta_i$ are set to be diagonal with only positive values, then the obtained kernel will correspond to the continuous relaxation method one [15] with $L_i$ hyperparameters per categorical variable, *i.e.*,

$$K_i(c_i^r, c_i^s, \Theta_i) = \prod_{j=1}^{L_i} \exp\left(-[\Theta_i]_{j,j}\left((e_{c_i^r})_j - (e_{c_i^s})_j\right)^2\right), \forall c_i^r \neq c_i^s.$$

- If $\forall i \in \{1, \dots, l\}$ the matrices $\Theta_i$ have all the diagonal terms equal to zero, then, the obtained kernel will be reduced to the Gaussian homoscedastic hypersphere one [27] with $\frac{L_i(L_i-1)}{2}$ hyperparameters per categorical variable, *i.e.*,

$$K_i(c_i^r, c_i^s, \Theta_i) = \exp\left(-2[\Theta_i]_{c_i^r, c_i^s}\right) \tag{14}$$

- If $\forall i \in \{1, \dots, l\}$ the matrices $\Theta_i$ have all the off-diagonal terms equal to $[\Theta_i]_{cov} < 0$ and zero on the diagonal. In this case, the obtained kernel will be reduced to Gower distance based model [16] with only 1 hyperparameter, *i.e.*,

$$K_i(c_i^r, c_i^s, \Theta_i) = \exp\left(-2[\Theta_i]_{cov}\right), \forall c_i^r \neq c_i^s. \tag{15}$$

As one can see, our proposed framework generalizes different approaches by considering different matrix structures of the same model. In what follows, the full parameterization is called $mat\_FULL$, the diagonal representation (which is equivalent to continuous relaxation) will be called $mat\_CR$, the off-diagonal representation will be is called $mat\_GHH$, as it is the Gaussian homoscedastic hypersphere model. Last, when we will consider only one covariance, as in Gower distance, we will denote the model $mat\_GOWER$. Table 1 summarizes how to obtain the several existing approaches using our proposed framework.

**Table 1 A summary of how our proposed kernel-based approach generalizes existing categorical models.**

| Models | $\Theta_i$ | $K_i(c_i^r, c_i^s, \Theta_i)$ | # of Hyperparam. |
|---|---|---|---|
| Our full model (i.e., $\Theta_i = mat\_FULL$) | $\begin{bmatrix} [\Theta_i]_{1,1} & & \textbf{\textit{Sym.}} & \\ [\Theta_i]_{1,2} & [\Theta_i]_{2,2} & & \\ \vdots & \ddots & \ddots & \\ [\Theta_i]_{1,L_i} & \dots & [\Theta_i]_{L_i-1,L_i} & [\Theta_i]_{L_i,L_i} \end{bmatrix}$ | $\exp\left(-\left([\Theta_i]_{c_i^r,c_i^r} + [\Theta_i]_{c_i^s,c_i^s}\right)\right)$ $\exp\left(-2[\Theta_i]_{c_i^r,c_i^s}\right)$ | $\frac{1}{2}L_i(L_i+1)$ |
| Continuous relaxation [15] (i.e., $\Theta_i = mat\_CR$) | $\begin{bmatrix} [\Theta_i]_{1,1} & & \textbf{\textit{Sym.}} & \\ 0 & [\Theta_i]_{2,2} & & \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 0 & [\Theta_i]_{L_i,L_i} \end{bmatrix}$ | $\exp\left(-\left([\Theta_i]_{c_i^r,c_i^r} + [\Theta_i]_{c_i^s,c_i^s}\right)\right)$ | $L_i$ |
| Gaussian homoscedastic hypersphere [13] (i.e., $\Theta_i = mat\_GHH$) | $\begin{bmatrix} 0 & & \textbf{\textit{Sym.}} & \\ [\Theta_i]_{1,2} & 0 & & \\ \vdots & \ddots & \ddots & \\ [\Theta_i]_{1,L_i} & \dots & [\Theta_i]_{L_i-1,L_i} & 0 \end{bmatrix}$ | $\exp\left(-2[\Theta_i]_{c_i^r,c_i^s}\right)$ | $\frac{1}{2}L_i(L_i-1)$ |
| Gower distance [16] (i.e., $\Theta_i = mat\_GOWER$) | $[\Theta_i]_{cov}\begin{bmatrix} 0 & & \textbf{\textit{Sym.}} & \\ 1 & 0 & & \\ \vdots & \ddots & \ddots & \\ 1 & \dots & 1 & 0 \end{bmatrix}$ | $\exp\left(-2[\Theta_i]_{cov}\right)$ | 1 |

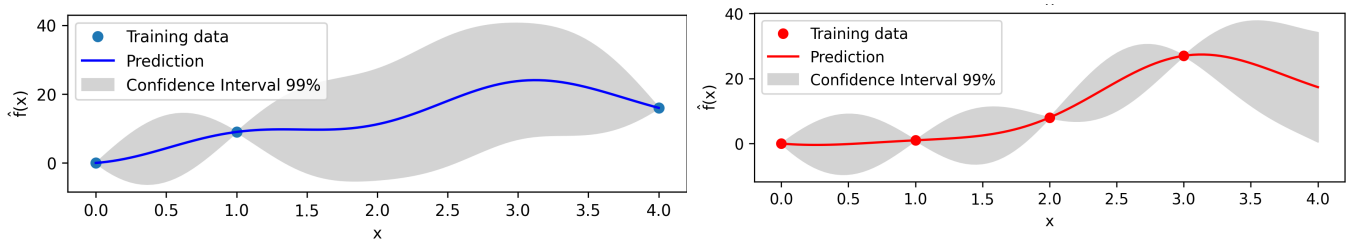
Finally, we note that, although our methodology is detailed only for Gaussian kernels, it is possible to extend our proposed approach to other kernels. In the next section, we will see how these models behave on different test cases.

## V. Results

In this section, we propose several illustrations and comparisons on three different test cases (with 1 or 2 continuous variables, 1 categorical variable up to 13 levels) to show the interest of our method and the equivalence with other models from the literature. The optimization of the likelihood as a function of the hyperparameters needs a performing algorithm, in this work, we are using COBYLA [32] to maximize this quantity. We note that the implementation of our proposed method has been released in the toolbox SMT v1.2* [33] and further developments are to appear in the next release.

### A. Hyperparameters equivalence

We start with a 2D test case with one continuous variable in $[0, 4]$ and one categorical variable with two levels (blue or red). We consider a DoE of 3 blue points and 4 red points (see Appendix VII.B for a detailed description of the test case). The mixed integer GP model is shown in Fig. 2 with the two associated levels (blue and red) and is obtained from continuous relaxation. As any other method leads to the same GP model of Fig. 2, we only report the optimal values of the hyperparameters in Tab. 2 to illustrate this equivalence. In this case, with only two levels, it is easy to prove that the Gower distance kernel and the homoscedastic hypersphere one are equivalent: there is only one categorical hyperparameter and we can check in Tab. 2 that $\exp(-0.23015) = 0.7944$. On this particular test case, all the correlations are positive and so, the Gaussian homoscedastic hypersphere and the original homoscedastic hypersphere models are equivalent as they can be restricted to angles in $[0, \frac{\pi}{2}]$ (see Appendix VII.A for a proof). A negative correlation would have been model by a close to 0 correlation through Gaussian kernel.




**(a) GP model for blue level** **(b) GP model for red level**

**Fig. 2 2D test case Gaussian process models for blue and red levels.**

**Table 2 Hyperparameter estimation of our proposed models versus existing approaches for the 2D test case.**

| Tested methods | $\theta_{red,blue}$ | $\theta_{red,red}$ | $\theta_{blue,blue}$ | $\theta^{cont}$ |
|---|---|---|---|---|
| Continuous relaxation [15] | - | 0.2300 | 1.2168e-06 | 16.576 |
| Our model (with $\Theta_i = mat\_CR$) | - | 0.2301 | 8.0294e-06 | 16.575 |
| Gower distance [16] | 0.2300 | - | - | 16.573 |
| Our model (with $\Theta_i = mat\_GOWER$) | 0.2301 | - | - | 16.573 |
| Homoscedastic hypersphere [13] | 0.7944 | - | - | 16.573 |
| Our model (with $\Theta_i = mat\_GHH$) | 0.7944 | - | - | 16.573 |

### B. Comparison of the different models to approximate some analytic functions

*1. Categorical cosine problem*

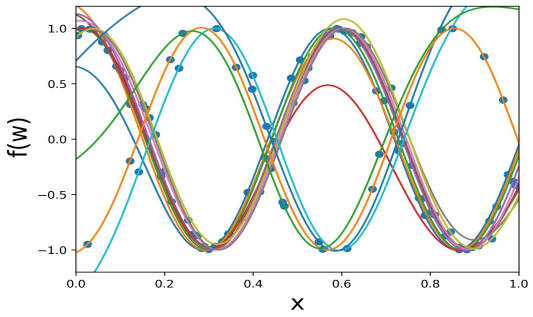
Let consider the cosine with two group example proposed by Roustant et al. in the paper that introduced the matrix parameterization of categorical correlation kernel [26]. The objective function $f$ depends on a continuous variable

*https://smt.readthedocs.io/en/latest/

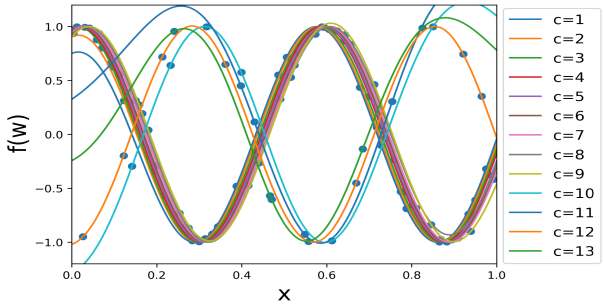
in $[0, 1]$ and on a categorical variable with 13 levels. Let $w = (x, c)$ be a given point with $x$ being the continuous variable and $c$ being the categorical variable, $c \in \{1, \ldots, 13\}$. There are two groups of curves corresponding to levels 1 to 9 and levels 10 to 13 with strong within-group correlations, and strong negative between-group correlations (see Appendix VII.D for a detailed description of the function).

The number of relaxed dimensions for continuous relaxation is 14. We draw a $14 \times 7$=98 points DoE by Latin Hypercube Sampling (LHS) [34] and plot the mean Gaussian process models on Fig. 3 for Gower distance, continuous relaxation and Gaussian homoscedastic hypersphere. The number of hyperparameters to optimize is therefore 2 for Gower, 14 for continuous relaxation and 79 for Gaussian homoscedastic hypersphere. Then, from these models, we compute the root mean square error (RMSE) as RMSE $= \sqrt{\sum_{i=1}^{n} \frac{1}{n} \left( \hat{f}(w_i) - f(w_i) \right)}$ where $n$ is the size of the validation set, $\hat{f}(w_i)$ is the prediction of our model at $w_i$ and $f(w_i)$ is the true value.

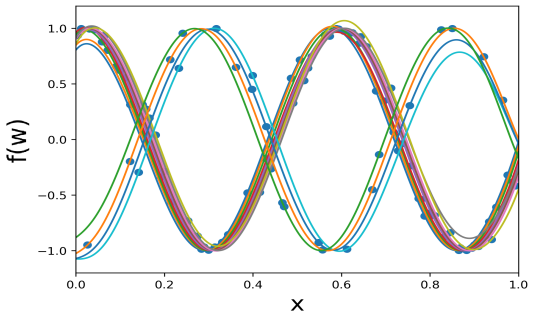
As expected, the more the hyperparameters, the better the model, as it can be seen on the decreasing RMSE. However, Gower distance takes 1.4 seconds to compute, continuous relaxation takes 24.5 seconds and Gaussian homoscedastic hypersphere takes 514.5 seconds to compute which motivates the use of a reduced order model. We can also consider the full homogeneous model with 92 hyperparameters; this model takes 642 seconds to compute and is worse than the Gaussian homoscedastic hypersphere. This model is not to consider for practical use cases.



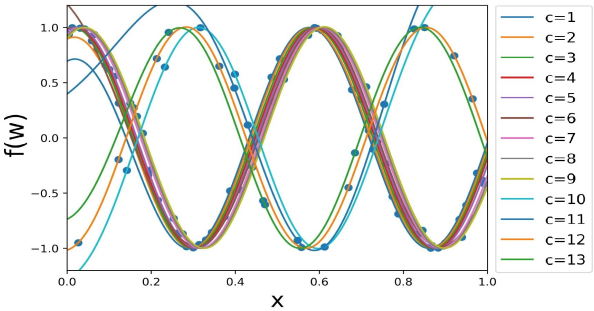

**(a) Our model (with $\Theta_1 = mat\_GOWER$): 2 hyperparameters, CPU time =1.4s, RMSE=30.079.**




**(b) Our model (with $\Theta_1 = mat\_CR$): 14 hyperparameters, CPU time =24.5s, RMSE=22.347.**




**(c) Our model (with $\Theta_1 = mat\_GHH$): 79 hyperparameters, CPU time =514.5s, RMSE=2.941.**



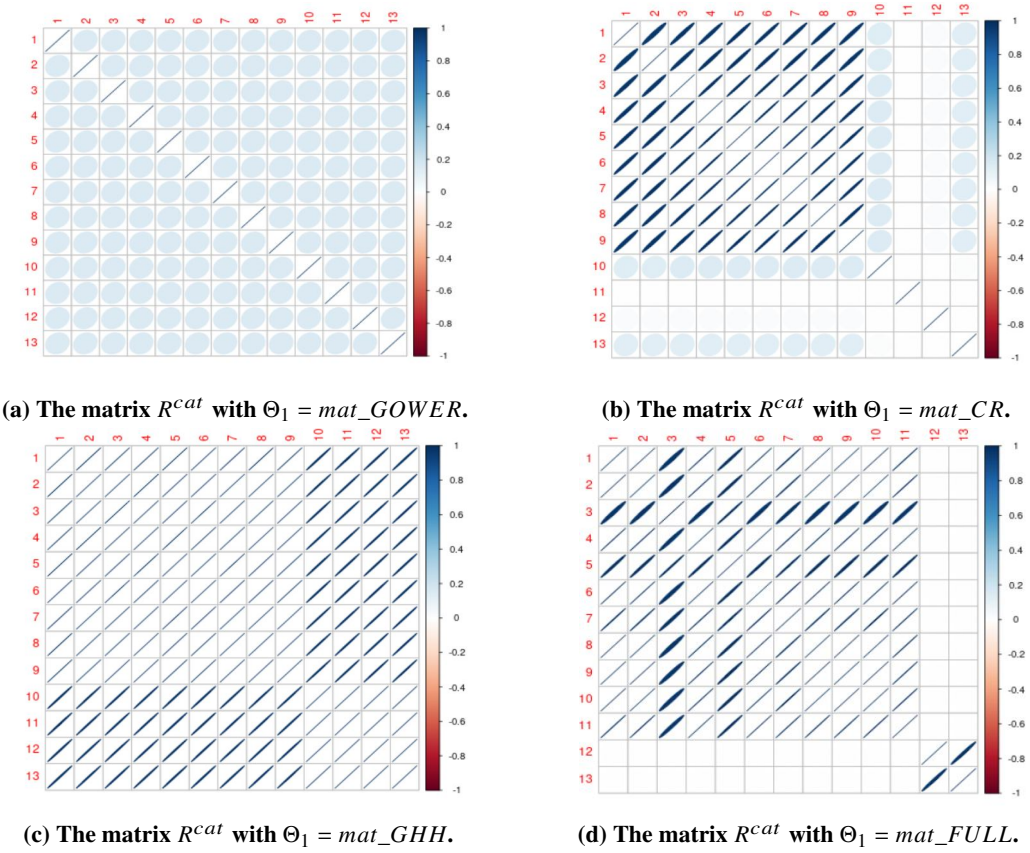

**(d) Our model (with $\Theta_1 = mat\_FULL$): 92 hyperparameters, CPU time =642s, RMSE=22.61.**

**Fig. 3 Mean predictions for the cosine problem with a 98 points DoE for our 4 proposed models.**

The estimated correlation matrix $R_i = R_1$ is shown in Fig. 4. For two given levels $\{r, s\}$, the correlation $[R_1]_{r,s}$ is in blue if the correlation is close to 1 and in red if the correlation is close to 0. We can see on the figure that the correlation between a level and itself is always 1. For Gower distance, there is only one estimated "mean correlation" as in Fig. 4a. For continuous relaxation (see Fig. 4b), we have $[R_1]_{r,s} = \exp(-\left([\Theta_1]_{r,r} + [\Theta_1]_{s,s}\right))$, therefore for the most important levels (1 to 9) are strongly correlated (in blue) with one another and the other levels (10 to 13) that should also have been also correlated are badly estimated because of the model limitation that neglected them. In contrast, the Gaussian parametrization (see Fig. 4c) of the Gaussian homoscedastic hypersphere decomposition model gives a good approximation of the real correlation and we see that there are two groups of highly correlated levels. The levels 1 to 9

are strongly similar, the levels 10 to 13 are strongly similar and the two groups are less similar. The latter correlations should have been different but we do not allow negative values through the Gaussian kernel.

As previously mentioned, the full homogeneous model (see Fig. 4d) that combines both continuous relaxation and Gaussian homoscedastic hypersphere adds irrelevant parameters making it more hard to optimize numerically while being equivalent to the Gaussian homoscedastic hypersphere model. For this reason, this model should not be considered for real applications.



**(a) The matrix $R^{cat}$ with $\Theta_1 = mat\_GOWER$.**

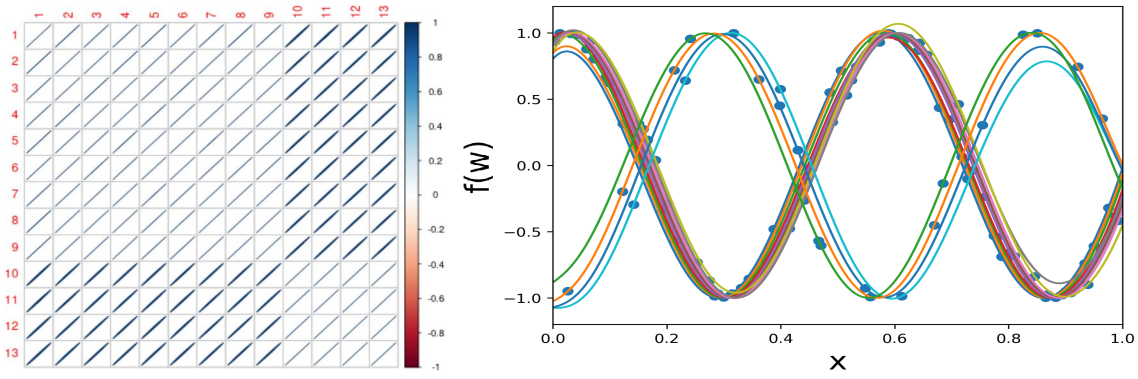
**(b) The matrix $R^{cat}$ with $\Theta_1 = mat\_CR$.**

**(c) The matrix $R^{cat}$ with $\Theta_1 = mat\_GHH$.**

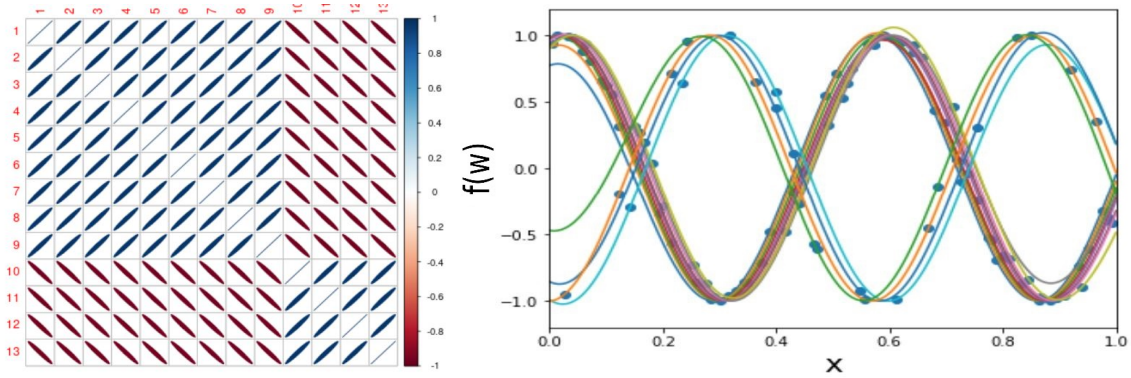
**(d) The matrix $R^{cat}$ with $\Theta_1 = mat\_FULL$.**

**Fig. 4 Obtained correlation matrices for the cosine problem using a DoE of 98 points.**

We compare in Fig. 5 our Gaussian hypersphere decompositon model (see Fig. 5a) with the homoscedastic hypersphere model [27] that allows negative values (see Fig. 5b). By imposing the correlation to be the matrix correlation of Fig. 5b in the case of Homoscedastic hypersphere, we obtain a likelihood of around 210 against 138 for the Gaussian homoscedastic hypersphere model with only positive values. Nevertheless, even if the Homoscedastic hyperphere model is more general than our model, the two RMSE are of the same order of magnitude, indicating similar performances on that particular test case.




**(a) Our model (with $\Theta_1 = mat\_GHH$): 79 hyperparameters, RMSE=2.941**

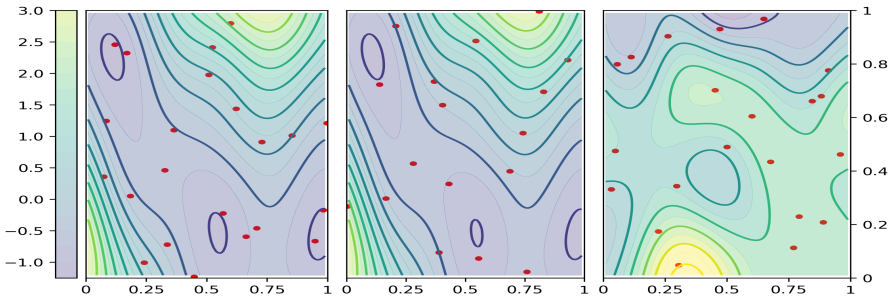



**(b) Homoscedastic hypersphere model [27]: 79 hyperparameters, RMSE=5.280**

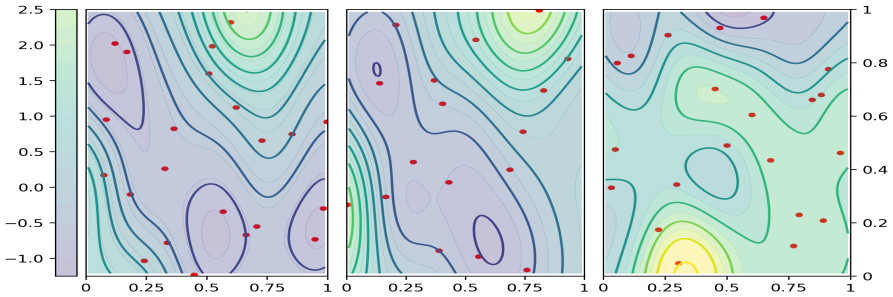
**Fig. 5 Comparison results between our model and Homoscedastic hypersphere [27] on the cosine problem using a DoE of 98 points.**

*2. Categorical Branin function*

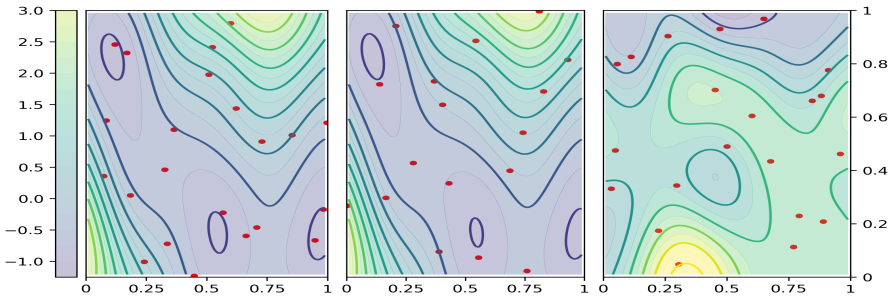
Let the function $f$ to model be the modified categorical Branin function [16]. This problem has 3 variables: two continuous variables in $[0, 1]$ and one categorical variable with three levels and the two first levels are totally correlated. We draw a DoE of 60 points by LHS to compare the given GP models and compute the error terms. To begin with, we start by plotting the models built with the different methods in Fig. 6 and compute their respective RMSE to compare them with the original formulation of the methods. The obtained values are given in Fig. 6 from a validation base of size 30603 that corresponds to 101 points from 0 to 1 in every continuous direction for every level (see Appendix VII.C for a detailed description of the function).



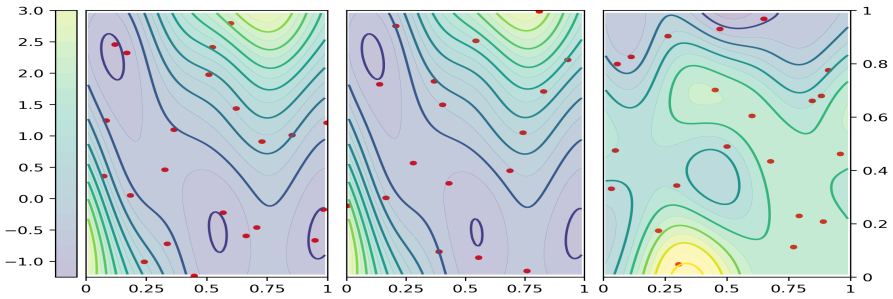
**(a) Our model (with $\Theta_1 = mat\_CR$): RMSE = 60.707**



**(b) Our model (with $\Theta_1 = mat\_GOWER$): RMSE = 77.793**



**(c) Our model (with $\Theta_1 = mat\_GHH$): RMSE = 60.721**



**(d) Our model (with $\Theta_1 = mat\_FULL$): RMSE = 60.707**

**Fig. 6 Mean predictions (over the three levels) for the categorical Branin problem using a DoE of 60 points (in red in the curve plots).**

In this test case, we clearly show that the full model, the continuous relaxation model and the Gaussian homoscedastic one are the same. However, we still have numerical instabilities but using the Gaussian homoscedastic kernel as proposed in this paper instead of using the raw matrix leads to a better estimation of the hyperparameters. For homoscedastic hypersphere, the RMSE that we found is 63.021. The Gower distance model is the only one that differs visually and that differs consequently in error from the others (77.8 instead of 60.7). As mentioned in Section V.A, with 3

levels, Gaussian homoscedastic hypersphere and continuous relaxation are equivalent methods, so these results were theoretically expected. However, when comparing with the original homoscedastic hypersphere, the Gaussian model differs as the third level is negatively correlated from the two others and because the Gaussian kernel returns only positive value. In this case, the estimated correlation is 0.03 with the Gaussian kernel against -0.15 with the original one.

## VI. Conclusion and perspectives

In this work, we have proposed a kernel-based class of models that extends continuous Gaussian kernels to handle mixed-categorical variables. We have observed on test cases that our proposed models generalizes Gower distance and continuous relaxation based models. Numerical illustrations on analytical problems showed the good potential of the proposed models. Further works will consider to include dimension reduction techniques (such as KPLS) to improve the computational efficiency of our model and be able to tackle higher dimensional problems. A rigorous classification between the proposed models as well as a proof of the SPD nature of the correlation matrices within our proposed model shall be investigated.


## Acknowledgements

This work is part of the activities of ONERA - ISAE - ENAC joint research group. The research presented in this paper has been performed in the framework of the AGILE 4.0 project (Towards Cyber-physical Collaborative Aircraft Development) and has received funding from the European Union Horizon 2020 Programme under grant agreement n° 815122. The authors are grateful to the partners of the AGILE 4.0 consortium for their contribution and feedback.


## VII. Appendix

### A. Positive homoscedastic hypersphere

**Theorem 1.** *We have equivalence between homoscedastic hypersphere [35] and Gaussian homoscedastic hypersphere if and only if the correlations are positive ($> \epsilon$) or by restricting the angles to be on $[0, \frac{\pi}{2}]$ that is the first quadrant.*

*Proof.* Let $F$ be the polyspherical change of coordinates [36], such that $F(\theta_1, \ldots, \theta_{n-1}) = \xi_1, \ldots, \xi_n$ with $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{S}^{n-1}$ and $\theta = (\theta_1, \ldots, \theta_{n-1}) \in [0, 2\pi]^{n-1}$, for a given integer $n$.

$$\begin{aligned} &F : [0, 2\pi]^{n-1} \rightarrow \mathbb{S}^{n-1} \\ &\xi = F(\theta) \end{aligned} \tag{16}$$

The formula for the so-called "polyspherical change of coordinates" given by Vilenkin et al. [36], restricted to real values, is:

$$\begin{aligned} \xi_1 &= \sin\theta_{n-1} \ldots \sin\theta_2 \sin\theta_1 \\ \xi_2 &= \sin\theta_{n-1} \ldots \sin\theta_2 \cos\theta_1 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \xi_{n-1} &= \sin\theta_{n-1} \cos\theta_{n-2} \\ \xi_n &= \cos\theta_{n-1} \end{aligned} \tag{17}$$

The function $F$ is bijective as a change of coordinates [36]. Correlations are values in $[-1, 1]$ so we consider that we have "unit radius" [31] ($r = 1$). $\xi$ are the so-called "cartesian coordinates" and $\theta$ are the so-called "spherical coordinates".

This leads to the formulation of Rebonato [31] for a $m \times m$ matrix. Let $b_{ij}$ be the elements of a triangular superior matrix $B$ of size $m \times m$. The $m \times (m-1)$ corresponding $b_{ij}$ elements are determined from the $m \times (m-1)$ $\theta_{ij}$ as:

$$\begin{aligned} b_{ij} &= \cos\theta_{ij} \prod_{k=1}^{j-1} \sin\theta_{ik}, \text{ if } j \in \{1, \ldots, m-1\} \\ b_{ij} &= \prod_{k=1}^{j-1} \sin\theta_{ik}, \text{ if } j = m \end{aligned} \tag{18}$$

Then, we take $C = BB^T$ as a Cholesky decomposition and this is the so-called "hypersphere decomposition" [30]. By restricting our coordinates $\theta$ to be in $[0, \frac{\pi}{2}]$, we still have a bijection, by construction. This bijection takes values between $\theta = [0, \frac{\pi}{2}]$ and $\xi = F([0, \frac{\pi}{2}]) = [0, 1]$. This is the so-called first quadrant as in Fig. 7.

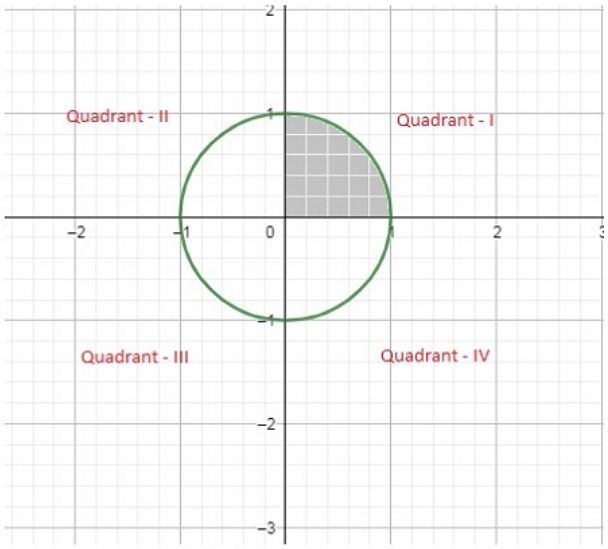



**Fig. 7 First quadrant on the sphere $\mathbb{S}^1$.**

Therefore, we know that

$$\forall \xi \in [0,1]^n, \exists!\ \theta \in \left[0, \frac{\pi}{2}\right]^n : F(\theta) = \xi.$$

The function g defined as $g(\xi) = \xi' = \left(\exp(-\log(\epsilon)(\xi_1 - 1)), \ldots, \exp(-\log(\epsilon)(\xi_{(m(m-1))} - 1))\right)$ is bijective if restricted to $[\epsilon, 1]$.

Therefore,

$$\forall \xi' \in [\epsilon,1]^n, \exists!\ \theta \in \left[0, \frac{\pi}{2}\right]^n : F(\theta) = \xi'.$$

For a given $\xi^* > \epsilon$ that corresponds to the optimal correlation according to the maximum of likelihood estimator, with the homoscedastic hypersphere method, $\exists!\ \theta^* \in [0, \frac{\pi}{2}] : F(\theta^*) = \xi^*$.
For the same $\xi^* > \epsilon$ that corresponds to the optimal correlation, with the Gaussian homoscedastic hypersphere method, $\exists!\ \theta'^* \in [0, \frac{\pi}{2}] : F(g(\theta'^*)) = \xi^*$.

Therefore, the two methods are equivalent over $[\epsilon, 1]$. By choosing $\epsilon$ sufficiently small, the two methods are equivalent to model positive correlations. Note that we obtain the same optimal correlation and the same likelihood but the values of $\theta^*$ and $\theta'^*$ may differ even if being uniquely defined on $[0, \frac{\pi}{2}]$.

□

**B. 2D blue/red test case**

This test case has one categorical variable with two levels: blue or red and one continuous variable in $[0, 4]$.

The blue DoE is the following: x= $\{0, 1, 4\}$, y=$\{0, 9, 16\}$

The red DoE is the following: x= $\{0, 1, 2, 3\}$, y=$\{0, 1, 8, 27\}$

Therefore, we have a DoE consisting of 7 points either blue or red, with continuous value ranging between 0 and 4 and taking value between 0 and 27.

**C. Categorical Branin function case**

This test case has one categorical variable with three levels and two continuous variables in $[0, 1]$. Let $w = (x_1, x_2, c)$ be a given point with $x_1$ and $x_2$ being the continuous variables and $c$ being the categorical variable, $c \in \{0, 1, 2\}$.

$$f(w) = \frac{1}{51.9496}\left(\left(\left(15x_2 - \frac{5}{4\pi^2}(15x_1 - 5)^2 + \frac{5}{\pi}(15x_1 - 5) - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(15x_1 - 5) + 10\right) - 54.8104\right), \quad \text{if c} = 0$$

$$f(w) = \frac{95}{5194.96}\left(\left(\left(15x_2 - \frac{5}{4\pi^2}(15x_1 - 5)^2 + \frac{5}{\pi}(15x_1 - 5) - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(15x_1 - 5) + 10\right) - 54.8104\right), \quad \text{if c} = 1$$

$$f(w) = -\log\left\{\left|\frac{1}{51.9496}\left(\left(\left(15x_2 - \frac{5}{4\pi^2}(15x_1 - 5)^2 + \frac{5}{\pi}(15x_1 - 5) - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(15x_1 - 5) + 10\right) - 54.8104\right)\right|\right\}^{\frac{1}{2}} + x_1^2 - 2x_2^2 + 1.03, \quad \text{if c} = 2$$

The DoE is given by a LHS of 60 points.

Our validation set is a evenly spaced grid of 101 points in $x_1$ ranging from 0.01 to 0.99, 101 points in $x_2$ ranging from 0.01 to 0.99 for every of the three categorical levels (0, 1, 2) for a total of 30603 points.

### D. Categorical cosine case

This test case has one categorical variable with 13 levels and one continuous variable in $[0, 1]$. Let $w = (x, c)$ be a given point with $x$ being the continuous variable and $c$ being the categorical variable, $c \in \{1, \ldots, 13\}$.

$$f(w) = \cos\left(\frac{7\pi}{2}x + \left(0.4\pi + \frac{\pi}{15}c\right) - \frac{c}{20}\right), \quad \text{if c} \in \{10, \ldots, 9\}$$

$$f(w) = \cos\left(\frac{7\pi}{2}x - \frac{c}{20}\right), \quad \text{if c} \in \{10, \ldots, 13\}$$

The reference landscapes of the objective function (with respect to the categorical choices) are drawn on Fig. 8.

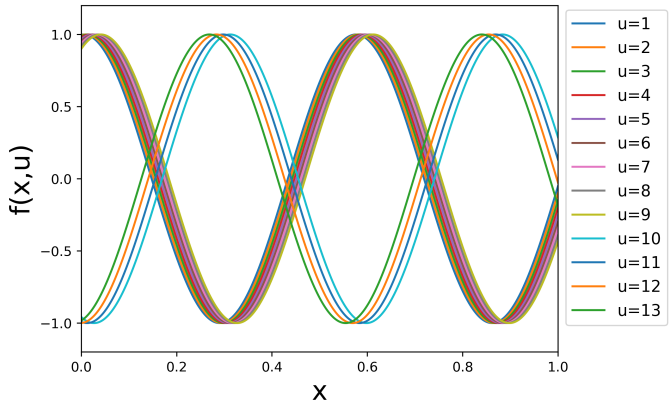



**Fig. 8 Landscape of the cosine test case from [26].**

The DoE is given by a LHS of 98 points. Our validation set is a evenly spaced grid of 101 points in $x_1$ ranging from 0.01 to 0.99, 101 points in $x_2$ ranging from 0.01 to 0.99 for every of the three categorical levels (0, 1, 2) for a total of 30603 points.

## References


[1] Duriez, E., and Morlier, J., "HALE multidisciplinary design optimization with a focus on eco-material selection," *Aerospace Europe Conference*, 2020.

[2] Priem, R., Gagnon, H., Chittick, I., Dufresne, S., Diouane, Y., and Bartoli, N., "An efficient application of Bayesian optimization to an industrial MDO framework for aircraft design," *AIAA AVIATION 2020 FORUM*, 2020, p. 3152.

[3] Ciampa, P. D., and Nagel, B., "The AGILE Paradigm: the next generation of collaborative MDO," *18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2017.

[4] Lambe, A., and Martins, J. R. R. A., "A unified description of MDO architectures," *9th World Congress on Structural and Multidisciplinary Optimization*, 2011.

[5] Lambe, A., and Martins, J. R. R. A., "Extensions to the design structure matrix for the description of multidisciplinary design, analysis, and optimization processes," *Structural and Multidisciplinary Optimization*, Vol. 46, 2012, pp. 273–284.

[6] Martins, J. R. R. A., and Lambe, A., "Multidisciplinary Design Optimization: A Survey of Architectures," *AIAA Journal*, Vol. 51, 2013, pp. 2049–2075.

[7] Jasa, J. P., Brelje, B. J., Gray, J. S., Mader, C. A., and Martins, J. R. R. A., "Large-Scale Path-Dependent Optimization of Supersonic Aircraft," *Aerospace*, Vol. 7, No. 10, 2020.

[8] Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. 13, 1998, p. 455–492.

[9] Močkus, J., "On bayesian methods for seeking the extremum," *Optimization Techniques IFIP Technical Conference Novosibirsk*, 1974.

[10] Rasmussen, C. E., and Quiñonero-Candela, J., "A Unifying View of Sparse Approximate Gaussian Process Regression," *Journal of Machine Learning Research*, Vol. 6, 2005, p. 1939–1959.

[11] Forrester, A., Sobester, A., and Keane, A., *Engineering Design via Surrogate Modelling: A Practical Guide*, Wiley, 2008.

[12] Sasena, M. J., Papalambros, P., and Goovaerts, P., "Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization," *Engineering Optimization*, Vol. 34, 2002, pp. 263–278.

[13] Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., "Efficient global optimization of constrained mixed variable problems," *Journal of Global Optimization*, Vol. 73, 2019, p. 583–613.

[14] Roustant, O., Padonou, E., Deville, Y., Clément, A., Perrin, G., Giorla, J., and Wynn, H., "Group kernels for Gaussian process metamodels with categorical inputs," *arXiv e-prints*, 2018.

[15] Garrido-Merchán, E. C., and Hernández-Lobato, D., "Dealing with categorical and integer-valued variables in Bayesian Optimization with Gaussian processes," *Neurocomputing*, Vol. 380, 2020, pp. 20–35.

[16] Halstrup, M., "Black-Box Optimization of Mixed Discrete-Continuous Optimization Problems," Ph.D. thesis, TU Dortmund, 2016.

[17] Rufato, R. C., Diouane, Y., Henry, J., Ahlfeld, R., and Morlier, J., "Creating Recommender Systems for Industrial Engineering Problems Using a Mixed Categorical-Continuous Data-Driven Method," *ECCOMAS CSAI 2021*, 2021.

[18] Cuesta-Ramirez, J., Le Riche, R., Roustant, O., Perrin, G., Durantin, C., and Gliere, A., "A comparison of mixed-variables Bayesian optimization approaches," 2021.

[19] Bouhlel, M., Bartoli, N., Regis, R., Otsmane, A., and Morlier, J., "Efficient Global Optimization for high-dimensional constrained problems by using the Kriging models combined with the Partial Least Squares method," *Engineering Optimization*, Vol. 50, 2018, pp. 2038–2053.

[20] Bouhlel, A. M., Bartoli, N., Otsmane, A., and Morlier, J., "Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction," *Structural and Multidisciplinary Optimization*, Vol. 53, 2016, pp. 935–952.

[21] Saves, P., Nguyen Van, E., Bartoli, N., Lefebvre, T., David, C., Defoort, S., Diouane, Y., and Morlier, J., "Bayesian optimization for mixed variables using an adaptive dimension reduction process: applications to aircraft design," *AIAA SciTech 2022*, San Diego, United States, 2022.

[22] Kim, S. H., and Boukouvala, F., "Surrogate-Based Optimization for Mixed-Integer Nonlinear Problems," *Computers & Chemical Engineering*, Vol. 140, 2020.

[23] Duvenaud, D., "Automatic model construction with Gaussian processes," Ph.D. thesis, University of Cambridge, 2014.

[24] Rossi, R. J., *Mathematical statistics: an introduction to likelihood based inference*, John Wiley & Sons, 2018.

[25] Sacks, J., Schiller, S. B., and Welch, W. J., "Designs for Computer Experiments," *Technometrics*, Vol. 31, No. 1, 1989, pp. 41–47.

[26] Roustant, O., Padonou, E., Deville, Y., Clément, A., Perrin, G., Giorla, J., and Wynn, H. P., "Group kernels for Gaussian process metamodels with categorical inputs," *HAL*, 2019.

[27] Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., and Guerin, Y., *Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems*, Springer International Publishing, 2020, pp. 189–224.

[28] Golovin, D., Solnik, B., Moitra, S., Kochanski, G., Karro, J., and Sculley, D., "Google Vizier: A Service for Black-Box Optimization," *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, New York, NY, USA, 2017, p. 1487–1495.

[29] Qian, P. Z. G., Wu, H., and Wu, C. F. J., "Gaussian Process Models for Computer Experiments With Qualitative and Quantitative Factors," *Technometrics*, Vol. 50, No. 3, 2008, pp. 383–396.

[30] Zhou, Q., Qian, P. Z. G., and Zhou, S., "A Simple Approach to Emulation for Computer Models With Qualitative and Quantitative Factors," *Technometrics*, Vol. 53, No. 3, 2011, pp. 266–273.

[31] Rebonato, R., and Jaeckel, P., "The Most General Methodology to Create a Valid Correlation Matrix for Risk Management and Option Pricing Purposes," *Journal of Risk*, Vol. 2, 2001.

[32] Powell, M. J. D., *A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation*, Springer Netherlands, 1994, pp. 51–67.

[33] Bouhlel, M. A., Hwang, J. T., Bartoli, N., Lafage, R., Morlier, J., and Martins, J. R. R. A., "A Python surrogate modeling framework with derivatives," *Advances in Engineering Software*, 2019, p. 102662.

[34] Jin, R., Chen, W., and Sudjianto, A., "An efficient algorithm for constructing optimal design of computer experiments," *Journal of Statistical Planning and Inference*, Vol. 134, No. 1, 2005, pp. 268–287.

[35] Pelamatti, J., "Mixed-variable Bayesian optimization : application to aerospace system design," Theses, Université de Lille, Mar. 2020.

[36] Klimyk, A., and Vilenkin, N. Y., *Representations of Lie groups and special functions*, Springer, 1995.