Electronic Journal of Statistics
Vol. 18 (2024) 5091–5131
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2320

Consistency of some sequential experimental design strategies for excursion set estimation based on vector-valued Gaussian processes

Philip Stange and David Ginsbourger

Institute of Mathematical Statistics and Actuarial Science, University of Bern,
e-mail: philip.stange@unibe.ch; david.ginsbourger@unibe.ch

Abstract: We tackle the extension to the vector-valued case of consistency results for Stepwise Uncertainty Reduction sequential experimental design strategies established in [3]. This leads us in the first place to clarify, assuming a compact index set, how the connection between continuous Gaussian processes and Gaussian measures on the Banach space of continuous functions carries over to vector-valued settings. From there, a number of concepts and properties from [3] can be readily extended. However, vector-valued settings do complicate things for some results, mainly due to the lack of continuity for the pseudo-inverse mapping that affects the conditional mean and covariance function given finitely many pointwise observations. We apply obtained results to the Integrated Bernoulli Variance and the Expected Measure Variance uncertainty functionals employed in [9] for the estimation for excursion sets of vector-valued functions.

MSC2020 subject classifications: 60G15, 62L05.
Keywords and phrases: Convergence, Gaussian processes, sequential design of experiments, stepwise uncertainty reduction, excursion sets.

Received August 2023.

1. Introduction

Sequential design of experiments is an important statistical area dealing with the step by step assignment of resources (typically, experiments, measurements, simulations) towards reducing the uncertainty about some quantity of interest. Bect et al. have in [3] reinforced the theoretical foundations for the analysis of a large class of strategies that are built according to the stepwise uncertainty reduction (SUR) paradigm. This has enabled them to establish some broader consistency results for the considered strategies under the assumption that the function of interest is a sample path of the Gaussian process model used to construct the sequential design. [3] is based on the idea that each of the SUR sequential design strategies involves an uncertainty functional applied to a sequence of conditional probability distributions such that for any sequential design the

arXiv: 2310.07315

resulting sequence of random variables, that we will denote by $(H_n)_{n\in\mathbb{N}}$, is a supermartingale with respect to the filtration generated by the observations. This is called supermartingale property of the underlying uncertainty functional. In [3], a number of methodological developments and application areas of sequential design of experiments with scalar-valued Gaussian process models are recalled (notably in the introduction), and it is shown that two strategies for probability of excursion / excursion set estimation (based on the integrated Bernoulli variance and on the expected measure variance, respectively) and two global optimization strategies (based on the expected improvement and on the knowledge gradient criteria, respectively) enjoy established SUR consistency results. The present work is motivated by the investigation of theoretical guarantees of consistency to related approaches in the framework of vector-valued Gaussian process models, such as considered in [9]. In the latter reference, a bivariate Gaussian process was used to jointly model salinity and temperature fields to delineate the river plume in a considered domain at the interface between the Fjord of Trondheim and the ocean. Another related setting of learning many tasks simultaneously using kernel methods (multi-task learning) also arises in [2] and [8] and has turned out to significantly outperform standard single-task learning methods in some cases. The multi-output model for Gaussian processes was also recently studied and encouraged in [25].

In what follows we do establish extensions of SUR consistency results to vector-valued (multi-output) settings, with a focus on the situation where multiple quantities are all observed at the same time and may correlate with each other. While such extensions may seem quite natural, so far only very few works have used vector-valued Gaussian processes in theoretical settings, which has motivated us to investigate this aspect and establish links to Gaussian measures on corresponding function spaces. The latter is all the more crucial since the connection between Gaussian processes and Gaussian measures plays a central role in the theoretical constructions used in [3] to prove consistency of (scalar-valued) SUR sequential design strategies.

We assume throughout that the function of interest is an element in the space of continuous functions from a compact metric space $(\mathbb{X}, \mathfrak{d})$ to $\mathbb{R}^d$ (for $d \in \mathbb{N}$), denoted by $\mathcal{C}\left(\mathbb{X}; \mathbb{R}^d\right)$, and a sample path of the multivariate Gaussian process ξ that is used to construct the sequential design. This means $\xi = (\xi_1, ..., \xi_d)$ is a $\mathbb{R}^d$-valued Gaussian process (every ξ_i is a $\mathbb{R}$-valued Gaussian process) with continuous sample paths defined on a compact metric space $(\mathbb{X}, \mathfrak{d})$. Observations

$$Z_n = \xi\left(X_n\right) + \varepsilon_n$$

for $n \geqslant 1$ are to be made sequentially in order to estimate the quantity of interest. Furthermore, we assume the sequence of observation errors $(\varepsilon_n)_{n\in\mathbb{N}}$ to be independent of the Gaussian process ξ and distributed as independent centered Gaussian vectors.

We can then directly take over the definition of a SUR strategy from [3], which starts with the choice of a "measure of residual uncertainty" for the quantity of interest after n observations

$$H_n = \mathcal{H}\left(P_n^{\xi}\right),$$

which is a non-negative function of the conditional distribution P_n^{ξ} of ξ given $\mathcal{F}_n$, where $\mathcal{F}_n$ is the σ-algebra generated by $X_1, Z_1, ..., X_n, Z_n$. For $n \geqslant 0$, the SUR sampling criterion J_n associated with $\mathcal{H}$ is then a function from $\mathbb{X}$ to $[0, \infty)$ and defined for $x \in \mathbb{X}$ as the conditional expectation

$$J_n\left(x\right) = \mathbb{E}\left[H_{n+1} | \mathcal{F}_n, X_{n+1} = x\right],$$

assuming that H_{n+1} is integrable for any choice of $x \in \mathbb{X}$. The value of the sampling criterion $J_n\left(x\right)$ quantifies the expected residual uncertainty at time $n+1$ if the next observation is to be made at $x \in \mathbb{X}$. The sequential design is then constructed by minimizing the expected residual uncertainty over $\mathbb{X}$

$$X_{n+1} \in \underset{x \in \mathbb{X}}{\operatorname{argmin}}\ J_n\left(x\right).$$

Note that some of the statements from [3] carry over smoothly to the vector-valued case, with proofs needing moderate adjustments and further arguments to hold in the more general setting. However, other aspects of extending SUR consistency results to vector-valued settings pose novel challenges. One of the key observations in [3] is that the sampling criterion is continuous when the covariance of the underlying Gaussian process is bounded away from zero. Things become more complicated in the vector-valued case as the pseudo-inverse mapping presents discontinuities between matrices of different ranks. This affects in turn extending the existence result for SUR sequential design and deserves some more thorough consideration (see Proposition 3.12, Lemma 4.5 and Proposition 5.1).

Since only two of the example design strategies from [3] have a straightforward extension to vector-valued settings, we focus on them and establish consistency for the extensions of the algorithms of [3] dedicated to the excursion probability/ excursion set estimation. For $f = (f_1, ..., f_d) \in \mathcal{C}\left(\mathbb{X}; \mathbb{R}^d\right)$ define the set

$$\Gamma\left(f\right) := \left\{u \in \mathbb{X} : f\left(u\right) \in \mathbf{T}\right\},$$

where $\mathbf{T} \subset \mathbb{R}^d$ is some closed set. For the case of orthants $\mathbf{T} := [t_1, \infty) \times ... \times [t_d, \infty)$,

$$\Gamma\left(f\right) = \left\{u \in \mathbb{X} : f(u) \geqslant T\right\} = \left\{u \in \mathbb{X} : f_i\left(u\right) \geqslant t_i, i \in \{1, ..., d\}\right\}.$$

is called excursion set with excursion threshold $T = \left(t_1, ..., t_d\right)^{\top} \in \mathbb{R}^d$. Given a finite measure μ on $\mathbb{X}$, the first measure of residual uncertainty in the excursion case is called the integrated Bernoulli variance (IBV) and defined by

$$H_n^{IBV} = \int_{\mathbb{X}} p_n\left(1 - p_n\right) d\mu,$$

where $p_n(x) = P(\xi(x) \geqslant T | \mathcal{F}_n)$ denotes the excursion probability with respect to $\mathcal{F}_n$. Still with μ a finite measure on $\mathbb{X}$, the second measure of residual uncertainty is the variance of the excursion volume (EMV), defined by

$$H_n^{EMV} = \mathrm{Var}\left(\mu\left(\Gamma\left(\xi\right)\right)|\mathcal{F}_n\right).$$

Both measures of residual uncertainty also appear in [9]. On the first look these criteria may seem to be the same as in [3], since H_n^{IBV} and H_n^{EMV} are again functions to $\mathbb{R}$. In fact, the vector-valued aspect is hidden within the definition of the excursion set/probability and yet complicates theoretical investigations on the aforementioned residual uncertainties (see Section 5). Indeed, it is not possible to reduce the expressions for the residual uncertainties to a form where we can utilize the theory from [3] for real-valued Gaussian processes. Even in the case a multivariate Gaussian process $\xi = (\xi_1, ..., \xi_d)$ with pairwise independent components ξ_i, we would end up with residual uncertainties that require further investigation beyond what is supported in the existing literature. An easy way out would be to assume a multivariate Gaussian process with independent components and to select the sequential design points by alternating between different residual uncertainties $H_n^{(i)}$, each one corresponding to a different component ξ_i. However, this is clearly different from the approach in [9], that motivated our work, and not the multivariate extension of SUR sequential design that we would advertise.

In the next section we will prepare the ground for our theoretical investigations on SUR strategies for the vector-valued case. The connection between continuous $\mathbb{R}^d$-valued Gaussian processes and Gaussian random elements in $\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)$ will be tackled with Theorem 2.5. This connection will later be crucial for the proofs of the consistency results and the analysis of SUR sequential design, since we will work with the distribution P^ξ of the Gaussian process ξ (or more precisely with its conditional distribution given finitely many observations), which will turn out to be a Gaussian measure on $\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)$. In Section 3 we will define the statistical model and design problem more precisely as introduced in [3]. In Section 4 we will discuss uncertainty functionals and some properties that are important for the existence of SUR sequential design and state general sufficient conditions for the consistency of SUR sequential designs. In Section 5 we will finally apply these consistency results to the two SUR sequential designs introduced above. The proofs are postponed to the Appendix.

2. Gaussian processes and Gaussian random elements

Let $(\Omega, \mathcal{F}, P)$ be the underlying probability space. In this section we will focus on the connection between multivariate Gaussian processes and Gaussian random elements in the space of continuous functions from a compact metric space $(\mathbb{X}, \mathfrak{d})$ to $\mathbb{R}^d$, that we denote by $\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)$ for some fixed $d \in \mathbb{N}$. The most important statement of this section is Theorem 2.5, which shows that we can identify each continuous multivariate Gaussian process as Gaussian random element in $\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)$ with respect to its Borel σ-algebra and vice versa. The theory of

Gaussian random elements is based on [5] and [23]. See also Chapter 2.3 and 2.4 in [24] for similar results in a more general setting. For the proofs or more details on $\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)$ and its dual space see Appendix A.1.

Definition 2.1. Let $(B, \|\cdot\|)$ be a (real) Banach space. A B-valued random element $X : \Omega \to B$ is called Gaussian if for any bounded linear functional $L \in B^*$ the random variable $\langle X, L\rangle$ is Gaussian. The distribution ν of X is called a Gaussian measure on $(B, \|\cdot\|)$.

Definition 2.2. Let $(\mathbb{X}, \mathfrak{d})$ be a compact metric space and $\xi = (\xi(x))_{x\in\mathbb{X}}$ a stochastic process with state space $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$. ξ is called a multivariate (d-variate, vector-valued) Gaussian process if the finite-dimensional distributions of ξ are Gaussian, i.e. if

$$\left(\xi(x_1)^\top, ..., \xi(x_n)^\top\right)^\top$$

is a Gaussian vector in $\mathbb{R}^{dn}$ for every $n \geqslant 1$ and $x_1, ..., x_n \in \mathbb{X}$.

Remark 2.3. We sometimes write $\xi = (\xi_1, ..., \xi_d)$ for a multivariate Gaussian process, where $\xi_i(x) := \pi_i(\xi(x)) = \xi(x)_i$ is the i-th component of $\xi(x)$ for $i \in \{1, ..., d\}$ and $x \in \mathbb{X}$. By the definition of a multivariate Gaussian process this means ξ_i is a real-valued Gaussian process for all $i \in \{1, ..., d\}$.

Furthermore, recall that the finite-dimensional distributions of a multivariate Gaussian process ξ are determined by the mean function

$$m : \mathbb{X} \to \mathbb{R}^d,\ x \mapsto \mathbb{E}[\xi(x)]$$

and covariance function

$$k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}^{d\times d},\ (x, y) \mapsto \mathbb{E}\left[(\xi(x) - m(x))(\xi(y) - m(y))^\top\right].$$

It holds for the entries of the mean function

$$m(x)_i = \mathbb{E}[\xi_i(x)]$$

for $x \in \mathbb{X}$ and $i \in \{1, ..., d\}$ and for the entries of the matrix covariance function

$$k(x, y)_{ij} = \mathbb{E}[(\xi_i(x) - m(x)_i)(\xi_j(y) - m(y)_j)]$$

for $x, y \in \mathbb{X}$ and $i, j \in \{1, ..., d\}$. Hence the entries $k(x, y)_{ij}$ of the matrix $k(x, y)$ correspond to the covariance between the outputs $\xi_i(x)$ and $\xi_j(y)$ and describe the degree of correlation or similarity between them.

Consideration of sample path properties makes it possible to think of multivariate Gaussian processes as random elements (measurable maps) from the underlying probability space Ω to a function space $\mathbb{S} \subset \left(\mathbb{R}^d\right)^{\mathbb{X}}$. In the following we will see that the induced random element will also be Gaussian if we consider continuous sample paths. See also [15] for the case $\mathbb{X} \subset \mathbb{R}$ and $d = 1$.

It is a well known result that $\mathbb{S} := \mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)$ is a separable Banach space if we equip it with the supremum norm

$$\|f\|_\infty := \sup_{x\in\mathbb{X}} |f(x)|_{\max},$$

where $|x|_{\max} := \max_{i\in\{1,\dots,d\}} |x_i|$ is the maximum norm on $\mathbb{R}^d$ (see Theorem 4.19 in [12]). Similar to the space of continuous $\mathbb{R}$-valued functions it can be shown that for the Borel σ-algebra $\mathcal{S}$ on $(\mathbb{S}, \|\cdot\|_\infty)$ it holds

$$\mathcal{S} = \sigma\left(\{\delta_x : x \in \mathbb{X}\}\right),$$

where $\delta_x : \mathbb{S} \to \mathbb{R}^d$, $\delta_x(f) = f(x)$ are the evaluation maps for $x \in \mathbb{X}$ (see Chapter 1.2 in [23]).

Lemma 2.4. *Let ξ be a multivariate Gaussian process with continuous sample paths. Then the mean function m and covariance function k are continuous.*

Theorem 2.5. *A multivariate Gaussian process ξ with continuous sample paths is a Gaussian random element in $(\mathbb{S}, \|\cdot\|_\infty)$ with respect to the Borel σ-algebra $\mathcal{S}$ and its distribution is a Gaussian measure on this space. Vice versa, we can find for every Gaussian measure ν on $(\mathbb{S}, \mathcal{S})$ a multivariate Gaussian process with continuous sample paths that has distribution ν. The distribution of ξ is uniquely determined by the mean function m and covariance function k, so we use the notation $\xi \sim \mathcal{GP}_d(m, k)$.*

3. Conditioning on finitely many observations

Now that we have the grounding, we can revisit the construction from [3] around conditioning on finitely many observations. Many properties carry over, but we still require some careful thoughts in places. Especially Propositions 3.11 and 3.12 need some additional arguments if the underlying Gaussian process has values in $\mathbb{R}^d$. Proposition 3.12 comes as a surprise and will be the reason that the sample criterion has a more complicated discontinuity structure in the vector-valued case than in the scalar-valued case. All proofs have been moved to Appendix A.2.

We will assume that

1. $(\mathbb{X}, \mathfrak{d})$ is a compact metric space,
2. $\xi = (\xi(x))_{x\in\mathbb{X}}$ is a d-variate Gaussian process on the probability space $(\Omega, \mathcal{F}, P)$ with mean function m and covariance function k,
3. ξ has continuous sample paths,

and concentrate on the following model:

ξ can be observed at sequentially selected design points $X_1, X_2, \dots$ with additive independent heteroscedastic Gaussian noise. This means pointwise observations Z_k for $k = 1, 2, \dots$ are given by

$$Z_k = \xi(X_k) + \tau(X_k)\, U_k,$$

where $(U_k)_{k\in\mathbb{N}}$ denotes a sequence of independent and $\mathcal{N}_d(0, I_d)$-distributed random vectors, that are also independent of ξ, and $\tau : \mathbb{X} \to \mathbb{R}^{d\times d}$ denotes a known continuous function. Then $\mathcal{T} : \mathbb{X} \to \mathbb{R}^{d\times d}$, $\mathcal{T}(x) := \tau(x)\tau(x)^\top$ is also continuous and $\mathcal{T}(x)$ is symmetric and positive semi-definite for every $x \in \mathbb{X}$.

Furthermore, we define the filtration $(\mathcal{F}_n)_{n\geqslant 0}$ by

$$\mathcal{F}_n := \sigma\left(\left\{\bigcup_{i=1}^{n}(X_i, Z_i)\right\}\right)$$

for $n \geqslant 1$ and set $\mathcal{F}_0$ to be the trivial σ-algebra. $\mathcal{F}_n$ is the σ-algebra generated by the first n sequential design points and n according pointwise observations $X_1, Z_1, X_2, Z_2, \ldots, X_n, Z_n$ and we have $\mathcal{F}_n \subseteq \mathcal{F}_m$ for $n \leqslant m$. We finally define

$$\mathcal{F}_\infty := \sigma\left(\bigcup_{n\geqslant 1}\mathcal{F}_n\right) \subset \mathcal{F}.$$

Definition 3.1. A sequence $(X_n)_{n\geqslant 1}$ is called sequential design if X_n is $\mathcal{F}_{n-1}$-measurable for all $n \geqslant 1$.

Definition 3.2. For $A \in \mathbb{R}^{n\times m}$ the (Moore-Penrose) pseudo-inverse of A is defined as the matrix $A^\dagger \in \mathbb{R}^{m\times n}$ satisfying the properties

1. $AA^\dagger A = A$ and $A^\dagger AA^\dagger = A^\dagger$,
2. $\left(AA^\dagger\right)^\top = AA^\dagger$ and $\left(A^\dagger A\right)^\top = A^\dagger A$.

Remark 3.3.
1. By the Theorem of Moore and Penrose the pseudo-inverse always exists and is unique (see [14]).
2. If A is a square matrix with full rank, then $A^\dagger = A^{-1}$ (see [14]).
3. The mapping $A \mapsto A^\dagger$ is measurable (see [18]).
4. In contrast to the usual matrix inversion mapping $A \mapsto A^{-1}$ for invertible matrices, the pseudo-inverse mapping $A \mapsto A^\dagger$ is in general not continuous. However, continuity of this mapping is provided on sets with constant matrix rank. This means $A_n^\dagger \to A^\dagger$, if $A_n \to A$ and there exists $n_0 \in \mathbb{N}$ such that $\operatorname{rank}(A_n) = \operatorname{rank}(A)$ for all $n \geqslant n_0$ (see [16, 20]).

Theorem 3.4. *For any $\xi \sim \mathcal{GP}_d(m, k)$, $\boldsymbol{X}_n = (X_1, \ldots, X_n) \in \mathbb{X}^n$, $\boldsymbol{Z}_n = (Z_1, \ldots, Z_n) \in \mathbb{R}^{d\times n}$, as defined above, the conditional mean and covariance function of ξ given $\boldsymbol{Z}_n = \boldsymbol{z}_n$ and assuming a deterministic design $\boldsymbol{X}_n = \boldsymbol{x}_n$ are given by*

$$m_n(x; \boldsymbol{x}_n, \boldsymbol{z}_n) = m(x) + K(x, \boldsymbol{x}_n)\,\Sigma(\boldsymbol{x}_n)^\dagger\left(vec(\boldsymbol{z}_n) - vec(m(\boldsymbol{x}_n))\right),$$
$$k_n(x, y; \boldsymbol{x}_n) = k(x, y) - K(x, \boldsymbol{x}_n)\,\Sigma(\boldsymbol{x}_n)^\dagger K(y, \boldsymbol{x}_n)^\top,$$

where we define $\Sigma(x) := k(x, x) + \mathcal{T}(x)$ for $x \in \mathbb{X}$ and use the matrix convention

$$m(\boldsymbol{x}_n) := \begin{pmatrix} m(x_1) & \cdots & m(x_n) \end{pmatrix} \in \mathbb{R}^{d\times n},$$
$$K(x, \boldsymbol{x}_n) := \begin{pmatrix} k(x, x_1) & \cdots & k(x, x_n) \end{pmatrix} \in \mathbb{R}^{d\times nd},$$

$$\Sigma(\boldsymbol{x}_n) := K(\boldsymbol{x}_n) + \mathcal{T}(\boldsymbol{x}_n) \in \mathbb{R}^{nd\times nd},$$

with

$$K(\boldsymbol{x}_n) := \begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n) \\ k(x_2,x_1) & \ddots & & \vdots \\ \vdots & & \ddots & k(x_{n-1},x_n) \\ k(x_n,x_1) & \cdots & k(x_n,x_{n-1}) & k(x_n,x_n) \end{pmatrix},$$

$$\mathcal{T}(\boldsymbol{x}_n) := \begin{pmatrix} \mathcal{T}(x_1) & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathcal{T}(x_n) \end{pmatrix}.$$

Remark 3.5. 1. For the case $d = 1$ the formula reduces to the form

$$m_n(x;\boldsymbol{x}_n,\boldsymbol{z}_n) = m(x) + K(x,\boldsymbol{x}_n)\,\Sigma(\boldsymbol{x}_n)^{\dagger}(\boldsymbol{z}_n - m(\boldsymbol{x}_n))^{\top},$$
$$k_n(x,y;\boldsymbol{x}_n) = k(x,y) - K(x,\boldsymbol{x}_n)\,\Sigma(\boldsymbol{x}_n)^{\dagger} K(y,\boldsymbol{x}_n)^{\top}$$

with $\Sigma(\boldsymbol{x}_n) = (k(x_i,x_j) + \mathcal{T}(x_i)\delta_{i,j})_{1\leqslant i,j\leqslant n} \in \mathbb{R}^{n\times n}$. Hence the expressions for m_n and k_n are consistent with the ones for $\mathbb{R}$-valued Gaussian processes as in [3].

2. Conditionally on $\mathcal{F}_n$, the next observation Z_{n+1} follows a multivariate normal distribution: $Z_{n+1}|\mathcal{F}_n \sim \mathcal{N}_d(m_n(X_{n+1}), \Sigma_n(X_{n+1}))$, where

$$\Sigma_n(x) := k_n(x,x) + \mathcal{T}(x).$$

3. For the case $n = 0$ we have $\Sigma(x) = \Sigma_0(x) = k(x,x) + \mathcal{T}(x) \in \mathbb{R}^{d\times d}$ as expected for our model.
4. The conditional mean m_n (or posterior mean in Bayesian statistics) is also called Kriging predictor or rather co-Kriging in our case, since we have the joint Kriging of multiple data inputs. For the univariate response setting, there exists abundant literature to it (see [6, 17] and the references therein). For the multivariate response see also [2, 25].

As shown in Section 2 we can think of the multivariate Gaussian process ξ as a Gaussian random element in the separable Banach Space $(\mathbb{S}, \|\cdot\|_\infty)$ equipped with the Borel σ-algebra $\mathcal{S}$. Let furthermore $\mathbb{M}$ be the space of Gaussian measures on $\mathbb{S}$. We equip $\mathbb{M}$ with the σ-algebra $\mathcal{M}$ generated by the evaluation maps $\pi_A : \nu \to \nu(A)$ for $A \in \mathcal{S}$. As shown in Theorem 2.5 we have the following connections:

Any measure $\nu \in \mathbb{M}$ corresponds to the distribution P^ξ of some continuous multivariate Gaussian Process ξ with mean m and covariance function k. Hence we can write

$$\nu = P^\xi = \mathcal{GP}_d(m,k),$$

since the distribution is uniquely determined by m and k.

On the other hand the probability distribution $P^\xi = \mathcal{GP}_d(m, k)$ of some continuous multivariate Gaussian process is a Gaussian measure on $\mathbb{S}$ and hence an element in $\mathbb{M}$.

Definition 3.6. Given a Gaussian random element ξ in $\mathbb{S}$, we will denote by $\mathfrak{P}(\xi)$ the set of all Gaussian conditional distributions of ξ. That is the set of Gaussian random measures $\boldsymbol{\nu}$ such that $\boldsymbol{\nu} = P(\xi \in \cdot | \mathcal{F}')$ for some σ-algebra $\mathcal{F}' \subset \mathcal{F}$.

Remark 3.7.

1. Note that we use a bold letter $\boldsymbol{\nu}$ to denote a random element in $\mathbb{M}$ (Gaussian random measure) and a normal letter ν to denote a point in the space $\mathbb{M}$ (Gaussian measure).
2. $\boldsymbol{\nu} = P(\xi \in \cdot | \mathcal{F}')$ is not necessarily Gaussian for an arbitrary σ-algebra $\mathcal{F}' \subset \mathcal{F}$ and hence not always a random element in $\mathbb{M}$. However, the next Proposition shows that it holds for the σ-algebra $\mathcal{F}_n$ generated by a sequential design with corresponding pointwise observations.

Proposition 3.8. *Let $n \geqslant 1$. There exists a measurable mapping*

$$\mathbb{X}^n \times \mathbb{R}^{d \times n} \times \mathbb{M} \to \mathbb{M},$$
$$((x_1, \dots, x_n), (z_1, \dots, z_n), \nu) \mapsto Cond_{x_1, z_1, \dots, x_n, z_n}(\nu)$$

such that $Cond_{X_1, Z_1, \dots, X_n, Z_n}\left(P^\xi\right)$ is the conditional distribution of ξ given the σ-algebra $\mathcal{F}_n$ for any $P^\xi \in \mathbb{M}$ and sequential design $(X_n)_{n\geqslant 1}$ with pointwise observations $(Z_n)_{n\geqslant 1}$.

Remark 3.9. In a Bayesian context, P^ξ can be seen as the prior distribution and for $n \in \mathbb{N}$

$$P_n^\xi := \mathrm{Cond}_{X_1, Z_1, \dots, X_n, Z_n}\left(P^\xi\right)$$

as the posterior distribution after observing $(Z_k)_{1\leqslant k\leqslant n}$ at the sequentially selected design points $(X_k)_{1\leqslant k\leqslant n}$. By Proposition 3.8 we have that P_n^ξ is a $\mathcal{F}_n$-measurable random element in $\mathbb{M}$ with $P_n^\xi = P(\xi \in \cdot | \mathcal{F}_n)$ and hence $P_n^\xi \in \mathfrak{P}(\xi)$ for all $n \in \mathbb{N}$. If $P^\xi = \mathcal{GP}_d(m, k)$, then it holds $P_n^\xi = \mathcal{GP}_d(m_n, k_n)$, where the $\mathcal{F}_n$-measurable (and random) conditional mean function m_n and conditional covariance function k_n are given as in Theorem 3.4.

Definition 3.10. Let $(\nu_n = \mathcal{GP}_d(m_n, k_n))_{n\geqslant 1}$ be a sequence of Gaussian measures in $\mathbb{M}$. We will say that $(\nu_n)_{n\geqslant 1}$ converges to $\nu_\infty \in \mathbb{M}$ if

$$m_n \to m_\infty \quad \text{in} \quad \mathcal{C}\left(\mathbb{X}; \mathbb{R}^d\right),$$
$$k_n \to k_\infty \quad \text{in} \quad \mathcal{C}\left(\mathbb{X} \times \mathbb{X}; \mathbb{R}^{d\times d}\right)$$

with respect to the corresponding supremum norms $\|\cdot\|_\infty$ on the function spaces. Notation: $\nu_n \to \nu_\infty$.

Proposition 3.11. *Let $\mathcal{F}_\infty$ be the σ-algebra generated by $\bigcup_{n\geqslant 1} \mathcal{F}_n$. For any Gaussian random element ξ in $\mathbb{S}$, defined on any probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and for any sequential design $(X_n)_{n\geqslant 1}$, the conditional distribution of ξ given*

$\mathcal{F}_\infty$ admits a version P_∞^ξ which is an $\mathcal{F}_\infty$-measurable random element in $\mathbb{M}$, and it holds

$$P_n^\xi \xrightarrow[n\to\infty]{a.s.} P_\infty^\xi.$$

Proposition 3.12. *Let $\nu = \mathcal{GP}_d(m_\nu, k_\nu) \in \mathbb{M}$, $\Sigma_\nu(x) := k_\nu(x,x) + \mathcal{T}(x) \in \mathbb{R}^{d\times d}$ and assume $(x_i, z_i) \to (x, z)$ as $i \to \infty$ in $\mathbb{X} \times \mathbb{R}^d$ with x_i, x in the set*

$$C_k := \{x \in \mathbb{X} : rank(\Sigma_\nu(x)) = k\}$$

and $z_i, z \in \mathbb{R}^d$ for $i \in \mathbb{N}$ and some $k \in \{1, ..., d\}$. Then we have

$$Cond_{x_i,z_i}(\nu) \xrightarrow[i\to\infty]{} Cond_{x,z}(\nu).$$

Remark 3.13. The above Proposition illustrates an important difficulty that arises when one turns from a scalar-valued Gaussian process $\mathcal{GP}(m,k)$ to a multivariate Gaussian process $\mathcal{GP}_d(m,k)$. For a Gaussian process $\mathcal{GP}(m,k)$ and just one observation $(\mathbf{x}_1, \mathbf{z}_1) = (x, z) \in \mathbb{X} \times \mathbb{R}$, the convergence mentioned above only depends on the inverse of a scalar, whereas for a multivariate Gaussian processes $\mathcal{GP}_d(m,k)$ and the observation $(\mathbf{x}_1, \mathbf{z}_1) = (x, z) \in \mathbb{X} \times \mathbb{R}^d$ we already have to deal with matrix inversion. This yields some difficulties for the limit as $x_i \to x$.

4. SUR sequential design and its existence in the multivariate setting

We will start by recalling some definitions from [3] that can instantly be extended to our multivariate setting, since they only depend on the space $\mathbb{M}$ of Gaussian measures on $\mathbb{S}$. However, the existence of SUR sequential design in the multivariate case comes with some pitfalls caused by Proposition 3.12 in the previous section. We will nevertheless prove that under some special assumptions the SUR sequential design indeed exists. All proofs have been moved to Appendix A.3.

Definition 4.1. An uncertainty functional on $\mathbb{M}$ is a measurable function

$$\mathcal{H} : \mathbb{M} \to [0, \infty)$$

with $\min_{\nu\in\mathbb{M}} \mathcal{H}(\nu) = 0$. The residual uncertainty after n observations, for a Gaussian random element ξ in $\mathbb{S}$ and a sequential design $(X_n)_{n\geqslant 1}$, is defined as the $\mathcal{F}_n$-measurable random variable

$$H_n := \mathcal{H}\left(P_n^\xi\right)$$

for $n \geqslant 0$.

Definition 4.2. Let $\mathcal{H}$ be an uncertainty functional on $\mathbb{M}$.

1. $\mathcal{H}$ has the supermartingale property if for any Gaussian random element ξ in $\mathbb{S}$, defined on any probability space $(\Omega, \mathcal{F}, P)$, and any sequential design $(X_n)_{n\geqslant 1}$ the sequence of residual uncertainties $(H_n)_{n\geqslant 0}$ is a $(\mathcal{F}_n)_{n\geqslant 0}$-supermartingale.
2. $\mathcal{H}$ is $\mathfrak{P}$-uniformly integrable if for any Gaussian random element ξ in $\mathbb{S}$, defined on any probability space $(\Omega, \mathcal{F}, P)$, the family $(\mathcal{H}(\boldsymbol{\nu}))_{\boldsymbol{\nu}\in\mathfrak{P}(\xi)}$ is uniformly integrable.
3. $\mathcal{H}$ is $\mathfrak{P}$-continuous if for any Gaussian random element ξ in $\mathbb{S}$, defined on any probability space $(\Omega, \mathcal{F}, P)$, and any sequence of random measures $(\boldsymbol{\nu}_n)_{n\geqslant 1} \subset \mathfrak{P}(\xi)$ such that $\boldsymbol{\nu}_n \xrightarrow[n\to\infty]{a.s.} \boldsymbol{\nu}_\infty \in \mathfrak{P}(\xi)$ it holds
$$\mathcal{H}(\boldsymbol{\nu}_n) \xrightarrow[n\to\infty]{a.s.} \mathcal{H}(\boldsymbol{\nu}_\infty).$$

For the definition of stepwise uncertainty reduction (SUR) sequential design strategies we need to define some important functionals on $\mathbb{M}$. For any $x \in \mathbb{X}$ observe the mapping $\mathcal{J}_x : \mathbb{M} \to [0, \infty]$ defined by

$$\begin{aligned}\mathcal{J}_x(\nu) &= \int_{\mathbb{R}^d}\int_{\mathbb{S}} \mathcal{H}\left(\mathrm{Cond}_{x,f(x)+\tau(x)u}(\nu)\right)\nu(df)\,\phi_d(u)\,du\\ &= \int_{\mathbb{R}^d} \mathcal{H}\left(\mathrm{Cond}_{x,m_\nu(x)+\Sigma_\nu(x)^{\frac{1}{2}}u}(\nu)\right)\phi_d(u)\,du,\end{aligned}$$

where $\Sigma_\nu(x)^{\frac{1}{2}} \in \mathbb{R}^{d\times d}$ is the unique symmetric and positive semi-definite square root matrix of $\Sigma_\nu(x) = k_\nu(x,x) + \mathcal{T}(x)$, ϕ_d is the density of $\mathcal{N}_d(0, I_d)$ and $\nu = \mathcal{GP}_d(m_\nu, k_\nu)$.

Proposition 4.3. *The mapping*

$$\mathcal{J} : \mathbb{X} \times \mathbb{M} \to [0, \infty],\ (x, \nu) \mapsto \mathcal{J}_x(\nu)$$

is $\mathcal{B}(\mathbb{X}) \otimes \mathcal{M}$-measurable.

Definition 4.4. Let ξ be a Gaussian random element in $\mathbb{S}$, $(X_n)_{n\geqslant 1}$ be a sequential design and $\mathcal{F}_n$ the σ-algebra generated by $X_1, Z_1, ..., X_n, Z_n$. The SUR sampling criterion J_n associated to an uncertainty functional $\mathcal{H}$ on $\mathbb{M}$ is defined as the function $J_n : \mathbb{X} \to [0, \infty]$, where

$$J_n(x) := \mathcal{J}_x\left(P_n^\xi\right) = \mathbb{E}_n\left[\mathcal{H}\left(\mathrm{Cond}_{x,Z(x)}\left(P_n^\xi\right)\right)\right] := \mathbb{E}\left[\mathcal{H}\left(\mathrm{Cond}_{x,Z(x)}\left(P_n^\xi\right)\right)|\mathcal{F}_n\right]$$

with $Z(x) = \xi(x) + \tau(x)U$ and $U \sim \mathcal{N}_d(0, I_d)$ independent of $\xi, U_1, ..., U_n$ as defined in the introduction of Section 3.

1. $(X_n)_{n\geqslant 1}$ is called a SUR sequential design associated to the uncertainty functional $\mathcal{H}$, if
$$X_{n+1} \in \underset{x\in\mathbb{X}}{\operatorname{argmin}}\ J_n(x)$$
for all $n \geqslant n_0$ with $n_0 \in \mathbb{N}$.

2. Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence of non-negative real numbers with $\varepsilon_n \to 0$ as $n \to \infty$. $(X_n)_{n\geqslant 1}$ is called an ε-quasi SUR sequential design associated to the uncertainty functional $\mathcal{H}$, if it holds

$$J_n(X_{n+1}) \leqslant \inf_{x\in\mathbb{X}} J_n(x) + \varepsilon_n$$

for all $n \geqslant n_0$ with $n_0 \in \mathbb{N}$.

Lemma 4.5. *Let $\mathcal{H}$ be a measurable uncertainty functional on $\mathbb{M}$ that is $\mathfrak{P}$-continuous, $\mathfrak{P}$-uniformly integrable and has the supermartingale property.*

1. *For any sequential design $(X_n)_{n\geqslant 1}$ the sample paths of*

$$J_n : \mathbb{X} \to [0,\infty),\ J_n(x) = \mathbb{E}_n\left[\mathcal{H}\left(Cond_{x,Z(x)}\left(P_n^{\xi}\right)\right)\right]$$

are continuous on the random sets

$$C_{n,k} := \{x \in \mathbb{X} : rank(\Sigma_n(x)) = k\} \subseteq \mathbb{X}$$

for $n \in \mathbb{N}$ and $k = 0, ..., d$.

2. *Assume that the covariance function k of the underlying Gaussian process ξ is positive definite and $\mathcal{T}(x) = \tau(x)\tau(x)^{\top}$ is positive definite for all $x \in \mathbb{X}$. Then there exists a SUR sequential design $(X_n)_{n\geqslant 1}$ associated with $\mathcal{H}$.*
3. *There exists an ε-quasi SUR sequential design $(X_n)_{n\geqslant 1}$ associated with $\mathcal{H}$.*

Definition 4.6. Let $\mathcal{H}$ be an uncertainty functional on $\mathbb{M}$ that has the supermartingale property.

1. The expected gain functional at $x \in \mathbb{X}$ is defined by

$$\mathcal{G}_x : \mathbb{M} \to [0,\infty),\ \mathcal{G}_x(\nu) := \mathcal{H}(\nu) - \mathcal{J}_x(\nu).$$

2. The maximal expected gain functional is defined by

$$\mathcal{G} : \mathbb{M} \to [0,\infty),\ \mathcal{G}(\nu) := \sup_{x\in\mathbb{X}} \mathcal{G}_x(\nu).$$

5. Consistency of multivariate excursion set estimation under SUR sequential design

In the previous section we have recalled some desirable properties of a SUR sequential design strategy that guarantee existence and continuity of the sample criterion on a partition of the domain $\mathbb{X}$. The following Proposition is the key to proving consistency of SUR sequential design in the case of multivariate excursion set estimation. The Proposition (see proof in Appendix A.4) follows from two Theorems in [3] that are also stated for completeness in Appendix A.4. The proofs of the Theorems have been adjusted to the multivariate setting.

Proposition 5.1. *Let $\mathcal{H}$ be an uncertainty functional on $\mathbb{M}$, $(X_n)_{n\geqslant 1}$ be an ε-quasi SUR sequential design for $\mathcal{H}$ and $\mathcal{G}$ the associated maximal expected gain functional. Assume that*

1. *$\mathcal{H}$ is $\mathfrak{P}$-continuous, $\mathfrak{P}$-uniformly integrable and has the supermartingale property,*
2. *$\{\nu \in \mathbb{M} : \mathcal{H}(\nu) = 0\} = \{\nu \in \mathbb{M} : \mathcal{G}(\nu) = 0\}$.*

Then it holds

$$H_n = \mathcal{H}\left(P_n^{\xi}\right) \xrightarrow[n\to\infty]{a.s.} 0.$$

5.1. Integrated Bernoulli variance (IBV)

In this subsection we are turning to the integrated Bernoulli variance, that is for example used for uncertainty reduction in [9] in the case of river plume mapping as already mentioned in the Introduction. The proofs are inspired by [3] and can be found in Appendix A.4.

Let ξ be a Gaussian random element in $\mathbb{S}$. The residual uncertainty of the integrated Bernoulli variance (IBV) is defined as the random variable

$$H_n^{IBV} := \int_{\mathbb{X}} p_n(u)\left(1 - p_n(u)\right)\mu(du) = \int_{\mathbb{X}} \mathrm{Var}\left(\mathbf{1}_{\Gamma(\xi)}(u)\,|\mathcal{F}_n\right)\mu(du),$$

where $\mathcal{F}_n$ is the σ-algebra generated by n observations and

$$p_n(u) := \mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u)\,|\mathcal{F}_n\right] = P\left(\xi(u) \geqslant T|\mathcal{F}_n\right).$$

More generally, we can define the corresponding uncertainty functional $\mathcal{H}^{IBV}$ by the mapping

$$\mathcal{H}^{IBV} : \mathbb{M} \to [0,\infty),$$
$$\nu \mapsto \int_{\mathbb{X}} p_\nu(u)\left(1 - p_\nu(u)\right)\mu(du),$$

where $p_\nu(u) := \int_{\mathbb{S}} \mathbf{1}_{\Gamma(f)}(u)\,\nu(df)$. Note that $\mathcal{H}^{IBV}$ is clearly an uncertainty functional on $\mathbb{M}$. Furthermore, let $\mathcal{G}^{IBV}$ be the associated maximal expected gain functional.

We want to use Proposition 5.1 to show

$$H_n^{IBV} = \mathcal{H}^{IBV}\left(P_n^{\xi}\right) \xrightarrow[n\to\infty]{a.s.} 0$$

for any ε-quasi SUR sequential design $(X_n)_{n\geqslant 1}$ for $\mathcal{H}^{IBV}$. In the following we will check the assumptions of the Proposition.

Lemma 5.2. *$\mathcal{H}^{IBV}$ is $\mathfrak{P}$-uniformly integrable and has the supermartingale property.*

Recall that for every sequence $(\boldsymbol{\nu}_n)_{n\geqslant 1} \subset \mathfrak{P}(\xi)$ such that $\boldsymbol{\nu}_n \to \boldsymbol{\nu}_\infty \in \mathfrak{P}(\xi)$ almost surely, it holds

$$\mathcal{H}^{IBV}(\boldsymbol{\nu}_n) \xrightarrow[n\to\infty]{a.s.} \mathcal{H}^{IBV}(\boldsymbol{\nu}_\infty)$$

and that $\boldsymbol{\nu}_n \in \mathfrak{P}(\xi)$ with $\xi \sim \mathcal{GP}_d(m,k)$ if there exist σ-algebras $\mathcal{G}_n \subset \mathcal{F}$ such that $\boldsymbol{\nu}_n = P(\xi \in \cdot|\mathcal{G}_n)$ and $\boldsymbol{\nu}_n(\omega)$ is a Gaussian measure on $\left(\mathcal{C}\left(\mathbb{X},\mathbb{R}^d\right), \|\cdot\|_\infty\right)$. In this case we can write $\boldsymbol{\nu}_n = \mathcal{GP}_d(m_n,k_n)$ for some random mean function m_n and random covariance function k_n. We have

$$\boldsymbol{\nu}_n \xrightarrow[n\to\infty]{a.s.} \boldsymbol{\nu}_\infty \in \mathfrak{P}(\xi),$$

if

$$m_n \xrightarrow[n\to\infty]{a.s.} m_\infty$$

uniformly on $\mathbb{X}$ and

$$k_n \xrightarrow[n\to\infty]{a.s.} k_\infty$$

uniformly on $\mathbb{X}\times\mathbb{X}$, where

$$\boldsymbol{\nu}_\infty = \mathcal{GP}_d(m_\infty,k_\infty) = P(\xi \in \cdot|\mathcal{G}_\infty)$$

for some $\mathcal{G}_\infty \subset \mathcal{F}$. Furthermore, we can write

$$\mathcal{H}^{IBV}(\boldsymbol{\nu}_n) = \int_{\mathbb{X}} g\left(P\left(\xi(u) \geqslant T|\mathcal{G}_n\right)\right)\mu(du)$$

for the bounded continuous function $g:[0,1]\to\left[0,\frac{1}{2}\right]$, $x\mapsto x(1-x)$, so the claim follows by the Dominated Convergence Theorem if we can show

$$P(\xi(u)\geqslant T|\mathcal{G}_n) = P(\xi_1(u)\geqslant t_1,...,\xi_d(u)\geqslant t_d|\mathcal{G}_n) \xrightarrow[n\to\infty]{a.s.} P(\xi(u)\geqslant T|\mathcal{G}_\infty).$$

We have for almost all $\omega\in\Omega$ and all $u\in\mathbb{X}$ by definition of the multivariate Gaussian process that $\xi(u)\sim\mathcal{N}_d(m(u),k(u,u))$ and

$$\nu_n(u,\omega) := \mathcal{L}\left(\left(\xi(u)|\mathcal{G}_n\right)(\omega)\right) = \mathcal{N}_d\left(m_n(u)(\omega), k_n(u,u)(\omega)\right)$$

for all $n\in\mathbb{N}\cup\{\infty\}$ with

$$m_n(u)(\omega) \xrightarrow[n\to\infty]{} m_\infty(u)(\omega),$$
$$k_n(u,u)(\omega) \xrightarrow[n\to\infty]{} k_\infty(u,u)(\omega)$$

by the almost sure uniform convergence of m_n and k_n. This already implies $\nu_n(u,\omega) \xrightarrow{w} \nu_\infty(u,\omega)$ as $n\to\infty$ for almost all $\omega\in\Omega$ and all $u\in\mathbb{X}$, so by the Portmanteau Theorem

$$\begin{aligned}\nu_n(u,\omega)(\mathbf{T}) =& P(\xi_1(u)\geqslant t_1,...,\xi_d(u)\geqslant t_d|\mathcal{G}_n)(\omega)\\ \xrightarrow[n\to\infty]{}& P(\xi_1(u)\geqslant t_1,...,\xi_d(u)\geqslant t_d|\mathcal{G}_\infty)(\omega)\end{aligned}$$

$$=\nu_\infty(u,\omega)(\mathbf{T}),$$

if

$$\nu_\infty(u,\omega)(\partial\mathbf{T}) = P(\exists i \in \{1,...,d\} : \xi_i(u) = t_i|\mathcal{G}_\infty)(\omega) = 0.$$

This is clearly the case if for all $j \in \{1,...,d\}$ it holds $k_\infty(u,u)(\omega)_{jj} > 0$ or $m_\infty(u)(\omega)_j \neq t_j$, but turns out to be more difficult in other cases. For the prove that $\mathcal{H}^{IBV}$ is $\mathfrak{P}$-continuous, we need to construct a suitable finite decomposition of $\mathbb{X}$ to check the convergence in each case.

Lemma 5.3. *Define the functions* $F_1 : \mathbb{X} \times \mathcal{P}(\{1,...,d\}) \to [0,\infty)$ *and* $F_2 : \mathbb{X} \times \mathcal{P}(\{1,...,d\}) \times \Omega \to [0,\infty)$ *by*

$$F_1(u,J) = \sum_{j\in J} k(u,u)^2_{jj}$$

and

$$F_2(u,J,\omega) = \sum_{j\in J}\left(m_\infty(u)(\omega)_j - t_j\right)^2 + k_\infty(u,u)(\omega)^2_{jj}.$$

For $J_1, J_2 \subseteq \{1,...,d\}$ *and* $\omega \in \Omega$ *fixed, let* $B_{J_1,J_2}(\omega) \subseteq \mathbb{X}$ *be the set of all* $u \in \mathbb{X}$ *such that*

1. $F_1(u,J_1) = 0$
2. $F_2(u,J_2,\omega) = 0$
3. *For every* $J_1' \supset J_1$ *and* $J_2' \supset J_2$ *it holds* $F_1(u,J_1') > 0$ *and* $F_2(u,J_2',\omega) > 0$.

Then $\mathbb{X}$ *can be written as the disjoint union*

$$\bigcup_{J_1,J_2\subseteq\{1,...,d\}} B_{J_1,J_2}(\omega)$$

and it holds:

1. *If* $J_2 \nsubseteq J_1$, *then* B_{J_1,J_2} *is almost surely a* μ*- null set.*
2. *If* $J_2 \subseteq J_1$, *then*

$$P(\xi(u) \geqslant T|\mathcal{G}_n) \xrightarrow[n\to\infty]{a.s.} P(\xi(u) \geqslant T|\mathcal{G}_\infty)$$

for all $u \in B_{J_1,J_2}$.

Lemma 5.4. $\mathcal{H}^{IBV}$ *is* $\mathfrak{P}$*-continuous.*

Proof. Using the decomposition for $\mathbb{X}$ from the above Lemma 5.3 and recalling that the finite union of P-null sets is again a P-null set, we get by the first property in Lemma 5.3

$$\mu(\mathbb{X}) \overset{a.s.}{=} \mu(A),$$

where the random subset A is defined by

$$A(\omega) := \bigcup_{\substack{J_1,J_2\subseteq\{1,...,d\}\\ J_2\subseteq J_1}} B_{J_1,J_2}(\omega).$$

Hence we can conclude with $g : [0,1] \to \left[0, \frac{1}{2}\right], x \mapsto x(1-x)$ that

$$\begin{aligned} &\int_{\mathbb{X}} g\left(P\left(\xi(u) \geqslant T \mid \mathcal{G}_n\right)\right) \mu(du) \\ \overset{a.s.}{=} &\int_A g\left(P\left(\xi(u) \geqslant T \mid \mathcal{G}_n\right)\right) \mu(du) \\ \xrightarrow[n\to\infty]{a.s.} &\int_A g\left(P\left(\xi(u) \geqslant T \mid \mathcal{G}_\infty\right)\right) \mu(du) \\ \overset{a.s.}{=} &\int_{\mathbb{X}} g\left(P\left(\xi(u) \geqslant T \mid \mathcal{G}_\infty\right)\right) \mu(du). \end{aligned}$$

by the second property in Lemma 5.3 and dominated convergence. □

Lemma 5.5.

$$\left\{\nu \in \mathbb{M} : \mathcal{H}^{IBV}(\nu) = 0\right\} = \left\{\nu \in \mathbb{M} : \mathcal{G}^{IBV}(\nu) = 0\right\}.$$

Note that $\{\nu \in \mathbb{M} : \mathcal{H}(\nu) = 0\} \subseteq \{\nu \in \mathbb{M} : \mathcal{G}(\nu) = 0\}$ always holds as shown in [3]. In the proof of the above Lemma (see Appendix) we will only focus on the reverse inclusion.

Theorem 5.6. *If $(X_n)_{n\geqslant 1}$ is an ε-quasi SUR sequential design for $\mathcal{H}^{IBV}$, then it holds*

$$H_n^{IBV} \xrightarrow[n\to\infty]{a.s.} 0.$$

Furthermore, it holds almost surely and in L^1 that,

$$\int_{\mathbb{X}} \left(\mathbf{1}_{\xi(u)\geqslant T} - p_n(u)\right)^2 \mu(du) \xrightarrow[n\to\infty]{} 0.$$

5.2. Excursion Measure Variance (EMV)

Let ξ be a Gaussian random element in $\mathbb{S}$. Another popular measure of the residual uncertainty with respect to the excursion set that is used in [3], is the variance of the excursion volume or excursion measure variance (EMV) defined by

$$H_n^{EMV} := \mathbb{E}\left[\left(\alpha(\xi) - \mathbb{E}\left[\alpha(\xi) \mid \mathcal{F}_n\right]\right)^2 \mid \mathcal{F}_n\right] = \operatorname{Var}\left(\alpha(\xi) \mid \mathcal{F}_n\right),$$

where $\mathcal{F}_n$ is the σ-algebra generated by n observations and $\alpha(\xi) := \mu(\Gamma(\xi))$. More generally, we can in this case define the corresponding uncertainty functional $\mathcal{H}^{EMV}$ by the mapping

$$\begin{aligned} \mathcal{H}^{EMV} : \mathbb{M} &\to [0,\infty) \\ \nu &\mapsto \int_{\mathbb{S}} \left(\alpha(f) - \bar{\alpha}_\nu\right)^2 \nu(df), \end{aligned}$$

where $\bar{\alpha}_\nu := \int_{\mathbb{S}} \alpha(f)\, \nu(df)$. Note that $\mathcal{H}^{EMV}$ is clearly an uncertainty functional on $\mathbb{M}$. Furthermore, let $\mathcal{G}^{EMV}$ be the associated maximal expected gain functional.

We want again to use Proposition 5.1 to show

$$H_n^{EMV} = \mathcal{H}^{EMV}\left(P_n^{\xi}\right) \xrightarrow[n\to\infty]{a.s.} 0$$

for any ε-quasi SUR sequential design $(X_n)_{n\geqslant 1}$ for $\mathcal{H}^{EMV}$. The ideas for the proofs are again based on results shown in [3] and can be found in Appendix A.4.

Lemma 5.7. *$\mathcal{H}^{EMV}$ is $\mathfrak{P}$-uniformly integrable and has the supermartingale property.*

Using Fubini's Theorem one can see that for the $\mathfrak{P}$-continuity of $\mathcal{H}^{EMV}$ it is necessary to deal with the covariance of $\mathbf{1}_{\Gamma(\xi)}(u_1)$ and $\mathbf{1}_{\Gamma(\xi)}(u_2)$ at two points $u_1, u_2 \in \mathbb{X}$.

Lemma 5.8. *For $J_2^i \subseteq J_1^i \subseteq \{1, ..., d\}$ with $i = 1, 2$ it holds*

$$Cov\left(\mathbf{1}_{\Gamma(\xi)}(u_1), \mathbf{1}_{\Gamma(\xi)}(u_2) | \mathcal{G}_n\right) \xrightarrow[n\to\infty]{a.s.} Cov\left(\mathbf{1}_{\Gamma(\xi)}(u_1), \mathbf{1}_{\Gamma(\xi)}(u_2) | \mathcal{G}_\infty\right)$$

for all $u_1 \in B_{J_1^1, J_2^1}$ and $u_2 \in B_{J_1^2, J_2^2}$

Lemma 5.9. *$\mathcal{H}^{EMV}$ is $\mathfrak{P}$-continuous.*

Proof. Using Fubini's Theorem we have

$$\begin{aligned}
\mathcal{H}^{EMV}(\boldsymbol{\nu}_n) =& \mathbb{E}\left[\left(\alpha(\xi) - \mathbb{E}\left[\alpha(\xi) | \mathcal{G}_n\right]\right)^2 | \mathcal{G}_n\right] \\
=& \mathbb{E}\left[\left(\int_{\mathbb{X}} \mathbf{1}_{\Gamma(\xi)}(u)\, \mu(du)\right)^2 | \mathcal{G}_n\right] - \mathbb{E}\left[\int_{\mathbb{X}} \mathbf{1}_{\Gamma(\xi)}(u)\, \mu(du) | \mathcal{G}_n\right]^2 \\
=& \int_{\mathbb{X}}\int_{\mathbb{X}} \mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u_1)\, \mathbf{1}_{\Gamma(\xi)}(u_2) | \mathcal{G}_n\right] \mu(du_1)\, \mu(du_2) \\
&- \int_{\mathbb{X}} \mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u_1) | \mathcal{G}_n\right] \mu(du_1) \int_{\mathbb{X}} \mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u_2) | \mathcal{G}_n\right] \mu(du_2) \\
=& \int_{\mathbb{X}}\int_{\mathbb{X}} \operatorname{Cov}\left(\mathbf{1}_{\Gamma(\xi)}(u_1), \mathbf{1}_{\Gamma(\xi)}(u_2) | \mathcal{G}_n\right) \mu(du_1)\, \mu(du_2).
\end{aligned}$$

We can now use the same decomposition $\mathbb{X} = \bigcup_{J_1, J_2 \subseteq \{1,\ldots,d\}} B_{J_1,J_2}$ as in Lemma 5.3 and already know that B_{J_1,J_2} is almost surely a μ- null set if $J_2 \nsubseteq J_1$. Hence the claim follows by Lemma 5.8 and the Dominated Convergence Theorem as

$$\begin{aligned}
&\int_{B_{J_1^2,J_2^2}(\omega)} \int_{B_{J_1^1,J_2^1}(\omega)} \operatorname{Cov}\left(\mathbf{1}_{\Gamma(\xi)}(u_1), \mathbf{1}_{\Gamma(\xi)}(u_2) | \mathcal{G}_n\right)(\omega)\, \mu(du_1)\, \mu(du_2) \\
\xrightarrow[n\to\infty]{}& \int_{B_{J_1^2,J_2^2}(\omega)} \int_{B_{J_1^1,J_2^1}(\omega)} \operatorname{Cov}\left(\mathbf{1}_{\Gamma(\xi)}(u_1), \mathbf{1}_{\Gamma(\xi)}(u_2) | \mathcal{G}_\infty\right)(\omega)\, \mu(du_1)\, \mu(du_2)
\end{aligned}$$

for almost all $\omega \in \Omega$. ∎

To apply Proposition 5.1 it remains to show that $\mathcal{H}^{EMV}$ and $\mathcal{G}^{EMV}$ vanish on the same subset of $\mathbb{M}$. It can be deduced by the same steps as in part (f) in the proof of Theorem 4.3 in [3] that $\alpha(\xi) - \mathbb{E}[\alpha(\xi)]$ is orthogonal to $L^2(\Omega, \sigma(Z(x)), P)$ for all $x \in \mathbb{X}$, where $Z(x) = \xi(x) + \tau(x) U$, $U \sim \mathcal{N}_d(0, I_d)$ independent of ξ, since even for a multivariate Gaussian process ξ we have that $\alpha(\xi)$ is only a random variable. The bottleneck is to conclude that $\alpha(\xi) - \mathbb{E}[\alpha(\xi)]$ is also orthogonal to $L^2(\Omega, \sigma(\xi(x)), P)$, which can be handled by the following Lemma.

Lemma 5.10. *Let $V = (V_1, ..., V_d)$ and $W = (W_1, ..., W_d)$ be independent Gaussian random vectors in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and U be a random variable in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, all defined on the probability space $(\Omega, \mathcal{F}, P)$. Assume that W is independent of (U, V) and that U is orthogonal to $L^2(\Omega, \sigma(V + W), P)$, that means for every $\sigma(V + W)$-measurable and square-integrable random variable $X : \Omega \to \mathbb{R}$ we have $\mathbb{E}[UX] = 0$. Then U is also orthogonal to $L^2(\Omega, \sigma(V), P)$.*

Lemma 5.11.

$$\left\{\nu \in \mathbb{M} : \mathcal{H}^{EMV}(\nu) = 0\right\} = \left\{\nu \in \mathbb{M} : \mathcal{G}^{EMV}(\nu) = 0\right\}.$$

Combining Lemmas 5.7, 5.9 and 5.11 we get by means of Proposition 5.1 and the same martingale convergence arguments for $(\mathbb{E}[\alpha(\xi) | \mathcal{F}_n])_{n \in \mathbb{N}}$ as in the proof of Proposition 4.5 in [3] the following Theorem.

Theorem 5.12. *If $(X_n)_{n \geqslant 1}$ is an ε-quasi SUR sequential design for $\mathcal{H}^{EMV}$, then it holds*

$$H_n^{EMV} \xrightarrow[n \to \infty]{a.s.} 0.$$

Furthermore, we have almost surely and in L^1 that,

$$\mathbb{E}[\alpha(\xi) | \mathcal{F}_n] \xrightarrow[n \to \infty]{} \alpha(\xi).$$

6. Conclusion

We have successfully extended the consistency results for the SUR sequential design strategies based on the integrated Bernoulli variance functional (IBV) and the variance of the excursion volume functional (EMV), that address the estimation of the excursion set problem, as introduced and proven for the univariate setting in [3], to the multivariate setting based on multivariate Gaussian processes $\xi = (\xi_1, ..., \xi_d)$ with sample paths in the function space $\mathcal{C}(\mathbb{X}; \mathbb{R}^d)$.

The authors of [3] have furthermore proven consistency for the knowledge gradient functional and the expected improvement functional. However, multi-objective optimization with multivariate Gaussian processes is beyond the scope of this paper and not considered. Nevertheless, our results, i.e. from Section 2, can be of interest for further research in this area.

The multivariate setting for excursion set estimation also arises in [9] and our work provides a (slightly relaxed) theoretical foundation for the techniques that are used in the paper for river plume mapping. Note that the excursion sets that we are addressing have the special form

$$\Gamma(\xi) = \{u \in \mathbb{X} : \xi(u) \geqslant T\},$$

due to the orthants $\mathbf{T} := [t_1, \infty) \times \dots \times [t_d, \infty)$ that we are considering. It remains to be checked if the general case of arbitrary closed sets $\mathbf{T} \subseteq \mathbb{R}^d$ also holds.

Further studies should also include the convergence rate of the SUR sequential design to provide a full theoretical support for their effectiveness. An important question in this context is also whether the correlation (similarity) of the Gaussian processes $(\xi_1, \dots, \xi_d)$ has an enhancing effect on the convergence rate.

Appendix A: Proofs and auxiliary results

A.1. Proofs of Section 2

Proof. **(Lemma 2.4)**
Let $\xi = (\xi_1, \dots, \xi_d)$ be a multivariate Gaussian process with continuous sample paths. Then we have for $x \in \mathbb{X}$ and any sequence $(x_n)_{n \in \mathbb{N}}$ in $\mathbb{X}$ with $x_n \to x$ (with respect to the metric $\mathfrak{d}$ on $\mathbb{X}$) that $\sum_{i=1}^d (\xi_i(x_n) - \xi_i(x))^2 \xrightarrow{a.s.} 0$. Hence $\xi_i(x_n)$ is a Gaussian random variable for every $i \in \{1, \dots, d\}$ and $n \in \mathbb{N}$ with $\xi_i(x_n) \xrightarrow{a.s.} \xi_i(x)$, so Lemma 1 in [10] implies $\sum_{i=1}^d \mathbb{E}\left[(\xi_i(x_n) - \xi_i(x))^2\right] \to 0$. This already implies continuity of the mean function m since

$$\|m(x_n) - m(x)\|_2^2 \leqslant \sum_{i=1}^d \mathbb{E}\left[(\xi_i(x_n) - \xi_i(x))^2\right] \to 0.$$

For continuity of the covariance function let also $y \in \mathbb{X}$ and $(y_n)_{n \in \mathbb{N}}$ be any sequence in $\mathbb{X}$ with $y_n \to y$. Since

$$k(x_n, y_n) = \mathbb{E}\left[\xi(x_n)\xi(y_n)^\top\right] - m(x_n)m(y_n)^\top,$$

it only remains to show convergence of the first term on the right hand side. This can be checked component-wise using Cauchy-Schwarz and

$$\mathbb{E}\left[(\xi_i(x_n) - \xi_i(x))^2\right] \to 0$$

for all $i \in \{1, \dots, d\}$. □

Recall that the (continuous) dual space of a real normed vector space $(X, \|\cdot\|)$ is defined by

$$X^* := \{L : X \to \mathbb{R} : L \text{ is linear and continuous}\}$$

and that the operator norm $\|\cdot\|_{op}$ on the dual space X^* is given by

$$\|L\|_{op} := \inf\{c \geqslant 0 : |Lx| \leqslant c\|x\| \text{ for all } x \in X\}.$$

Two well known results are that L is bounded if and only if it is continuous and that $\|L\|_{op} = \sup_{\|x\| \leqslant 1} |Lx|$. Furthermore, we have the following basic result for the dual space of (finite) product spaces:

Let $\left(X_1, \|\cdot\|_{X_1}\right), ..., \left(X_d, \|\cdot\|_{X_d}\right)$ be real Banach spaces with dual spaces $X_1^*, ..., X_d^*$. Define the space

$$X_1 \times ... \times X_d := \{(x_1, ..., x_d) : x_i \in X_i, i \in \{1, ..., d\}\}$$

with norm $\|(x_1, ..., x_d)\| = \max_{i \in \{1,...,d\}} \|x_i\|_{X_i}$ and

$$X_1^* \times ... \times X_d^* := \{(L_1, ..., L_d) : L_i \in X_i^*, i \in \{1, ..., d\}\}$$

with norm $\|(L_1, ..., L_d)\|_* = \sum_{i=1}^d \|L_i\|_{op, X_i}$. Then the following statements hold.

1. $(X_1 \times ... \times X_d, \|\cdot\|)$ and $\left(X_1^* \times ... \times X_d^*, \|\cdot\|_*\right)$ are Banach spaces.
2. $J : X_1^* \times ... \times X_d^* \to (X_1 \times ... \times X_d)^*$ defined by

$$J(L_1, ..., L_d)(x_1, ..., x_d) = \sum_{i=1}^d L_i x_i$$

is an isometric isomorphism.

Let $(\mathbb{X}, \mathfrak{d})$ be a compact metric space. The product space $\mathcal{C}(\mathbb{X}) \times ... \times \mathcal{C}(\mathbb{X})$ and $\mathcal{C}\left(\mathbb{X}; \mathbb{R}^d\right)$ with the supremum norm

$$\|f\|_\infty = \sup_{x \in \mathbb{X}} \|f(x)\|_{\max}$$

are isomorphic and have the same topological structure. They even form an isometric isomorphism if we consider the product space with the norm

$$\|(f_1, ..., f_d)\|_{d,\infty} = \max_{i \in \{1,...,d\}} \|f_i\|_\infty .$$

Proposition A.1. *Let $(\mathbb{X}, \mathfrak{d})$ be a compact metric space. The dual space of $\mathcal{C}\left(\mathbb{X}; \mathbb{R}^d\right)$ is isometrically isomorphic to $\mathfrak{M}(\mathbb{X}) \times ... \times \mathfrak{M}(\mathbb{X})$, where $\mathfrak{M}(\mathbb{X})$ is the space of finite signed measures on $\mathbb{X}$ equipped with the Borel σ-algebra. This means for every $L \in \mathcal{C}\left(\mathbb{X}; \mathbb{R}^d\right)^*$ there exist finite signed measures μ_i for $i \in \{1, ..., d\}$ such that for all $f = (f_1, ..., f_d) \in \mathcal{C}\left(\mathbb{X}; \mathbb{R}^d\right)$ it holds*

$$L(f) = \sum_{i=1}^d \int_{\mathbb{X}} f_i(x) \mu_i(dx) .$$

Proof. Since $\mathbb{X}$ is a compact metric space, every measure in $\mathfrak{M}(\mathbb{X})$ is also a Radon measure. The statement that the dual space of $\mathcal{C}(\mathbb{X})$ is the space $\mathfrak{M}(\mathbb{X})$ of finite signed measures is well-known as Riesz–Markov Representation Theorem, see Chapter 14 in [1]. Since $\mathcal{C}(\mathbb{X}) \times ... \times \mathcal{C}(\mathbb{X})$ and $\mathcal{C}\left(\mathbb{X}; \mathbb{R}^d\right)$ are isometrically isomorphic, this shows that $(\mathcal{C}(\mathbb{X}) \times ... \times \mathcal{C}(\mathbb{X}))^*$ is isometrically isomorphic to $\mathfrak{M}(\mathbb{X}) \times ... \times \mathfrak{M}(\mathbb{X})$. ☐

Proof. **(Theorem 2.5)**
We first show that ξ is a stochastic process with continuous sample paths if and only if it is a random element in $\left(\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right), \|\cdot\|_\infty\right)$ with respect to its Borel σ-algebra. Assume that ξ is a stochastic process with continuous sample paths. ξ is measurable with respect to the σ-algebra $\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right) \cap \mathcal{B}\left(\mathbb{R}^d\right)^{\mathbb{X}}$ by Lemma 4.1 in [11]. Since $\mathcal{B}\left(\mathbb{R}^d\right)^{\mathbb{X}}$ is generated by the projection (or evaluation) maps $\pi_x : \left(\mathbb{R}^d\right)^{\mathbb{X}} \to \mathbb{R}^d$ for $x \in \mathbb{X}$, we have that ξ is also measurable with respect to the Borel σ-algebra on $\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)$. This makes ξ a random element in $\left(\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right), \|\cdot\|_\infty\right)$. The other direction is clear since the evaluation maps are linear and continuous.

Assume that ξ is a multivariate Gaussian processes. By the above part ξ is a random element in $\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)$ and it is Gaussian if $\langle \xi, L\rangle$ is a Gaussian variable for all $L \in \mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)^*$. By Proposition A.1 there exist finite signed Borel measures μ_i on $(\mathbb{X}, \mathfrak{d})$ such that

$$\langle \xi, L\rangle = \sum_{i=1}^d \int_{\mathbb{X}} \xi_i\left(x\right) \mu_i\left(dx\right).$$

Let D be a countable dense subset of $\mathbb{X}$. Since Borel measures and Baire measures are equivalent on compact metric spaces (see Chapter 7 and 8 in [4]) and every finite measure μ_i is also a Radon measure since it is regular, there exists a sequence of linear combinations of Dirac measures $\delta_{x_k^{(i)}}$ with $x_k^{(i)} \in D$ such that

$$\sum_{k=1}^n a_k^{(i)} \delta_{x_k^{(i)}} \xrightarrow[n\to\infty]{w} \mu_i,$$

where $a_k^{(i)} \in \mathbb{R}$, by Example 8.1.6 in [4] (see also Example 8.16 in [19] or Chapter 15 in [1]) for every $i \in \{1, ..., d\}$. By the definition of multivariate Gaussian processes we know that

$$\sum_{i=1}^d \sum_{k=1}^n a_k^{(i)} \xi_i\left(x_k^{(i)}\right)$$

is a Gaussian variable for every $n \in \mathbb{N}$ and

$$\begin{aligned}\sum_{i=1}^d \sum_{k=1}^n a_k^{(i)} \xi_i\left(x_k^{(i)}\right) &= \sum_{i=1}^d \int_{\mathbb{X}} \xi_i\left(x\right) \left(\sum_{k=1}^n a_k^{(i)} \delta_{x_k^{(i)}}\right)(dx) \\ &\xrightarrow[n\to\infty]{a.s.} \sum_{i=1}^d \int_{\mathbb{X}} \xi_i\left(x\right) \mu_i\left(dx\right).\end{aligned}$$

Since the almost sure limit of a sequence of Gaussian random variables is again Gaussian, we conclude that $\langle \xi, L\rangle$ is Gaussian.

Assume now that ξ is a Gaussian random element. We know that ξ induces a Gaussian vector

$$\left(L_1\left(\xi\right), ..., L_d\left(\xi\right)\right)$$

for $L_i \in U_i$, where U_i are subsets of the dual space of $\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)$ for $i \in \{1,...,d\}$. Taking the evaluation maps $\delta_x : \mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right) \to \mathbb{R}^d, f \mapsto f(x)$ and the projection maps $\pi_i : \mathbb{R}^d \to \mathbb{R}, x \mapsto x_i$ we have that $\pi_i \circ \delta_x : \mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right) \to \mathbb{R}$ is linear and continuous for every $i \in \{1,...,d\}$, $x \in \mathbb{X}$. Hence we can define a $\mathbb{R}^d$-valued process by $(\pi_1 \circ \delta_x(\xi),...,\pi_d \circ \delta_x(\xi))_{x\in\mathbb{X}}$, whose finite-dimensional distributions are Gaussian.

The finite-dimensional distributions of ξ are uniquely determined by m and k and hence by Proposition 4.2 in [11] the last claim follows, since the σ-algebras coincide as already mentioned in the first part of the proof. □

Remark A.2. ξ is a continuous multivariate Gaussian process with zero mean function and covariance function $k : \mathbb{X}\times\mathbb{X} \to \mathbb{R}^{d\times d}$ if and only if ξ is a centered Gaussian random element in the separable Banach space B with covariance operator $K_\xi : B^* \to B$. The covariance operator can be derived from the covariance function k by

$$L \mapsto K_\xi L = \left[y \in \mathbb{X} \mapsto \left(\sum_{k=1}^{d} \int_{\mathbb{X}} k(x,y)_{kl}\, \mu_k(dx) \right)_{l=1,...,d} \right],$$

where μ_k are the finite signed measures from Proposition A.1. Given the covariance operator $K_\xi : B^* \to B$ we can derive the covariance function by

$$(x,y) \mapsto \left(\left\langle K_\xi \delta_x^k, \delta_y^l \right\rangle\right)_{k,l=1}^{d},$$

where $\delta_x^k := \pi_k \circ \delta_x$ and $\delta_y^l := \pi_l \circ \delta_y$ for $x, y \in \mathbb{X}$ and $k, l \in \{1,...,d\}$.

A.2. Proofs of Section 3

Lemma A.3. *Let $\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)$ and $\mathcal{C}\left(\mathbb{X}\times\mathbb{X};\mathbb{R}^{d\times d}\right)$ be endowed with their Borel σ-algebra. Define the mappings*

$$m_\bullet : \mathbb{M} \to \mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right), \nu \mapsto m_\nu$$
$$k_\bullet : \mathbb{M} \to \mathcal{C}\left(\mathbb{X}\times\mathbb{X};\mathbb{R}^{d\times d}\right),\ \nu \mapsto k_\nu$$
$$\Psi := (m_\bullet, k_\bullet)$$

and let $\Theta \subset \mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right) \times \mathcal{C}\left(\mathbb{X}\times\mathbb{X};\mathbb{R}^{d\times d}\right)$ be the range of Ψ with trace σ-algebra Σ_Θ. Then $\Psi : \mathbb{M} \to \Theta$ is $\mathcal{M}/\Sigma_\Theta$-measurable and its inverse $\Psi^{-1} : \Theta \to \mathbb{M}$ is $\Sigma_\Theta/\mathcal{M}$-measurable.

Proof. The mappings $m_\bullet$ and $k_\bullet$ are $\mathcal{M}/\mathcal{S}$-measurable and $\mathcal{M}/\mathcal{B}\left(\mathcal{C}\left(\mathbb{X};\mathbb{R}^{d\times d}\right)\right)$-measurable, respectively. The statement for $m_\bullet$ follows by Proposition A.1 as in Lemma A.1 and A.2 in [3]. The statement for $k_\bullet$ follows by the same arguments if we consider the isometric isomorphism

$$A = (a_{ij})_{1\leqslant i,j\leqslant d} \mapsto (a_{11},...,a_{1d},a_{21},...,a_{dd}),$$

between $\mathbb{R}^{d\times d}$ and $\mathbb{R}^{d^2}$. That Ψ is $\mathcal{M}/\Sigma_\Theta$-measurable follows now easily by the measurability of $m_\bullet$ and $k_\bullet$. Ψ^{-1} is $\Sigma_\Theta/\mathcal{M}$-measurable if and only if $(m,k)\mapsto \mathcal{GP}_d(m,k)(A)$ is measurable for all $A\in\mathcal{B}\left(\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)\right)$. The latter holds for all cylinder sets of the form $A=\bigcap_{k=1}^n\left\{f\in\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right):f(x_k)\in\Gamma_k\right\}$ with $x_k\in\mathbb{X}$ and $\Gamma_k\in\mathcal{B}\left(\mathbb{R}^d\right)$ for $k=1,...,n$ and hence for all $A\in\mathcal{B}\left(\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)\right)$ by Dynkin's π-λ Theorem. ☐

Lemma A.4. *For all $n\geqslant 1$, the mapping*

$$\begin{aligned}\tilde{\kappa}_n:\mathbb{X}^n\times\mathbb{R}^{d\times n}\times\Theta\to\Theta,\\ (\mathbf{x}_n,\mathbf{z}_n,(m,k))\mapsto\left(m_n(\cdot;\mathbf{x}_n,\mathbf{z}_n),k_n(\cdot;\mathbf{x}_n)\right)\end{aligned}$$

is $\left(\mathcal{B}(\mathbb{X}^n)\otimes\mathcal{B}\left(\mathbb{R}^{d\times n}\right)\otimes\Sigma_\Theta\right)/\Sigma_\Theta$-measurable, where $\mathbf{x}_n=(x_1,...,x_n)\in\mathbb{X}^n$ and $\mathbf{z}_n=(z_1,...,z_n)\in\mathbb{R}^{d\times n}$.

Proof. $\xi-m_n(\cdot;\mathbf{x}_n,\mathbf{z}_n)$ is again a multivariate Gaussian process with continuous sample paths and has the covariance function $k_n(\cdot;\mathbf{x}_n)$ for any deterministic design $\mathbf{x}_n=(x_1,...,x_n)\in\mathbb{X}^n$ which implies indeed $\left(m_n(\cdot;\mathbf{x}_n,\mathbf{z}_n),k_n(\cdot;\mathbf{x}_n)\right)\in\Theta$. The claim for the measurability of the function $\tilde{\kappa}_n$ follows from the continuity of the mappings $(m,x)\mapsto m(x)$, $(k,x)\mapsto k(x,\cdot)$, $(k,x,y)\mapsto k(x,y)$ and the measurability of the mapping $X\mapsto X^\dagger$ in combination with the explicit expressions for $m_n(\cdot;\mathbf{x}_n,\mathbf{z}_n)$ and $k_n(\cdot;\mathbf{x}_n)$ from Theorem 3.4. ☐

Proof. **(Proposition 3.8)**
Any $\nu\in\mathbb{M}$ is the distribution of a multivariate Gaussian process ξ and uniquely determined by its mean m and covariance k, so we can write $\nu=P^\xi=\mathcal{GP}_d(m,k)$ and define $\kappa_n:\mathbb{X}^n\times\mathbb{R}^{d\times n}\times\mathbb{M}\to\mathbb{M}$ by

$$\kappa_n(\mathbf{x}_n,\mathbf{z}_n,\nu)=\mathcal{GP}_d\left(m_n(\cdot;\mathbf{x}_n,\mathbf{z}_n),k_n(\cdot;\mathbf{x}_n)\right)=:P_n^\xi$$

with $\mathbf{x}_n=(x_1,...,x_n)\in\mathbb{X}^n$ and $\boldsymbol{z}_n=(z_{1,...,}z_n)\in\mathbb{R}^{d\times n}$. The measurability of κ_n follows by the previous two Lemmas and the equality

$$\kappa_n(\mathbf{x}_n,\mathbf{z}_n,\nu)=\Psi^{-1}\left(\tilde{\kappa}_n(\mathbf{x}_n,\mathbf{z}_n,\Psi(\nu))\right).$$

We need to show that P_n^ξ is a conditional distribution of ξ given the σ-algebra $\mathcal{F}_n$ generated by the sequential design $(X_n)_{n\geqslant 1}$ and pointwise observations $(Z_n)_{n\geqslant 1}$. By the defining property of the conditional expectation this holds if we can prove

$$\mathbb{E}\left[UP_n^\xi(\Gamma)\right]=\mathbb{E}\left[U\mathbf{1}_{\xi\in\Gamma}\right]$$

for any $U=\prod_{i=1}^n\varphi_i(Z_i)$ with measurable $\varphi_i:\mathbb{R}^d\to\mathbb{R}$ and $\Gamma\in\mathcal{B}\left(\mathcal{C}\left(\mathbb{X};\mathbb{R}^d\right)\right)$ of the form $\Gamma=\bigcap_{j=1}^J\{\xi(\bar{x_j})\in\Gamma_j\}$ with $\bar{x_j}\in\mathbb{X}$ and $\Gamma_j\in\mathcal{B}\left(\mathbb{R}^d\right)$ for $j=1,...,J$, since it extends to any $\mathcal{F}_n$-measurable U (recall that X_n is $\mathcal{F}_{n-1}$-measurable and by iteration it can be written as a measurable function of $Z_{n-1},...,Z_1$) and any set in $\mathcal{S}$ by Dynkin's π-λ Theorem, since $\mathcal{C}(\mathbb{X})\times...\times\mathcal{C}(\mathbb{X})$ and $\mathbb{S}$ with the supremum norm $\|f\|_\infty=\sup_{x\in\mathbb{X}}\|f(x)\|_{\max}$ are isomorphic and have the same topological structure. Indeed, the above statement follows by applying the equality

$$\kappa_{n+m}\left(\mathbf{x}_{n+m}, \mathbf{z}_{n+m}, \nu\right) = \kappa_m\left(\mathbf{x}_{n+1:n+m}, \mathbf{z}_{n+1:n+m}, \kappa_n\left(\mathbf{x}_n, \mathbf{z}_n, \nu\right)\right)$$

recursively to $P_n^\xi = \kappa_n\left(\mathbf{x}_n, \mathbf{z}_n, P^\xi\right)$. □

Proof. **(Proposition 3.11)**
By Proposition 3.8 we have that the conditional distribution of ξ given $\mathcal{F}_n$ is of the form $P_n^\xi = \mathcal{GP}_d\left(m_n, k_n\right)$ and that ξ is a Bochner-integrable random element with values in $\mathbb{S}$. We can define a Lévy-martingale by $\left(\mathbb{E}\left[\xi|\mathcal{F}_n\right]\right)_{n\in\mathbb{N}}$ with respect to the filtration $\left(\mathcal{F}_n\right)_{n\in\mathbb{N}}$, which is again a random element with values in $\mathbb{S}$ for every $n \in \mathbb{N}$, and by the Convergence Theorem for Lévy-martingales (see Theorem 6.1.12 in [21]) we have

$$\mathbb{E}\left[\xi|\mathcal{F}_n\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[\xi|\mathcal{F}_\infty\right]$$

in $\mathbb{S}$ with respect to the supremum norm, P-almost surely and in $L^1\left(\Omega, \mathcal{F}, P\right)$. The limit $m_\infty := \mathbb{E}\left[\xi|\mathcal{F}_\infty\right]$ is again a random element with values in $\mathbb{S}$ and $\mathcal{F}_\infty$-measurable by definition. We clearly have $m_n = \mathbb{E}\left[\xi|\mathcal{F}_n\right]$ since both are elements in $\mathbb{S}$ with $m_n\left(x\right) = \mathbb{E}\left[\xi\left(x\right)|\mathcal{F}_n\right] = \delta_x\left(\mathbb{E}\left[\xi|\mathcal{F}_n\right]\right)$ for all $x \in \mathbb{X}$, where δ_x denotes the (linear and continuous) evaluation function $\delta_x : \mathbb{S} \to \mathbb{R}^d$ with $\delta_x\left(f\right) := f\left(x\right)$. We conclude

$$m_n \xrightarrow[n\to\infty]{a.s.} m_\infty$$

uniformly on $\mathbb{X}$.

Assume now for simplicity that ξ is a centered Gaussian random element so the covariance function reduces to $k^c\left(x,y\right) = \mathbb{E}\left[\xi\left(x\right)\xi\left(y\right)^\top\right]$. We can define a random element ξ^2 in the separable Banach space $\mathcal{C}\left(\mathbb{X}\times\mathbb{X}; \mathbb{R}^{d\times d}\right)$ with supremum norm

$$\|f\|_\infty := \sup_{x\in\mathbb{X}}\left(\max_{i,j\in\{1,\ldots,d\}}\left|f\left(x\right)_{ij}\right|\right)$$

by $\xi^2\left(x,y\right) := \xi\left(x\right)\xi\left(y\right)^\top$ for $x, y \in \mathbb{X}$. Note that it holds

$$\begin{aligned}
\sup_{x,y\in\mathbb{X}}\left\|\xi\left(x\right)\xi\left(y\right)^\top\right\|_F &= \sup_{x,y\in\mathbb{X}} \operatorname{tr}\left(\left(\xi\left(x\right)\xi\left(y\right)^\top\right)^\top\left(\xi\left(x\right)\xi\left(y\right)^\top\right)\right)^{\frac{1}{2}} \\
&= \sup_{x,y\in\mathbb{X}}\left(\xi\left(x\right)^\top\xi\left(x\right)\right)^{\frac{1}{2}} \operatorname{tr}\left(\xi\left(y\right)\xi\left(y\right)^\top\right)^{\frac{1}{2}} \\
&= \sup_{x,y\in\mathbb{X}}\left\|\xi\left(x\right)\right\|_2\left\|\xi\left(y\right)\right\|_2 \\
&= \left(\sup_{x\in\mathbb{X}}\left\|\xi\left(x\right)\right\|_2\right)^2.
\end{aligned}$$

Using the equivalence of norms on finite-dimensional vector spaces and Fernique's Theorem, this yields $\mathbb{E}\left[\left\|\xi^2\right\|_\infty\right] < \infty$ and hence ξ^2 is Bochner-integrable.

By the same reasoning as above we have $\mathbb{E}\left[\xi^2|\mathcal{F}_n\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[\xi^2|\mathcal{F}_\infty\right] =: k_\infty^c$ in $\mathcal{C}\left(\mathbb{X}\times\mathbb{X};\mathbb{R}^{d\times d}\right)$ with respect to the supremum norm, P-almost surely and in $L^1\left(\Omega,\mathcal{F},P\right)$. Hence $k_n^c = \mathbb{E}\left[\xi^2|\mathcal{F}_n\right]$, since both are elements in $\mathcal{C}\left(\mathbb{X}\times\mathbb{X};\mathbb{R}^{d\times d}\right)$ with $k_n^c\left(x,y\right) = \mathbb{E}\left[\xi\left(x\right)\xi\left(y\right)^\top|\mathcal{F}_n\right] = \delta_{x,y}\left(\mathbb{E}\left[\xi^2|\mathcal{F}_n\right]\right)$ for all $(x,y)\in\mathbb{X}\times\mathbb{X}$ and hence

$$k_n^c \xrightarrow[n\to\infty]{a.s.} k_\infty^c$$

uniformly on $\mathbb{X}\times\mathbb{X}$. The general case (non-centered) follows in combination with the first part since

$$\begin{aligned} k_n\left(x,y\right) &= \mathbb{E}\left[\left(\xi\left(x\right)-m\left(x\right)\right)\left(\xi\left(y\right)-m\left(y\right)\right)^\top|\mathcal{F}_n\right] \\ &= k_n^c\left(x,y\right) - m_n\left(x\right)m_n\left(y\right)^\top. \end{aligned}$$

Hence we conclude

$$k_n \xrightarrow[n\to\infty]{a.s.} k_\infty$$

uniformly on $\mathbb{X}\times\mathbb{X}$.

Let now Q denote any conditional distribution of ξ given $\mathcal{F}_\infty$. The $\mathcal{F}_\infty$-measurable random measure Q is then almost surely Gaussian (follows as in the proof of Proposition 2.9 in [3] using the characteristic function for random vectors) and thus taking

$$P_\infty^\xi\left(\omega,\cdot\right) = \begin{cases} Q\left(\omega,\cdot\right) & ,\omega\in\Omega_0 \\ \mathcal{GP}_d\left(0,0\right) & ,\omega\in\Omega\backslash\Omega_0 \end{cases}$$

we have constructed a $\mathcal{F}_\infty$-measurable random element in $\mathbb{M}$ such that

$$P_n^\xi \xrightarrow[n\to\infty]{a.s.} P_\infty^\xi.$$

□

Proof. **(Proposition 3.12)**
Let $\nu = \mathcal{GP}_d\left(m_\nu,k_\nu\right)\in\mathbb{M}$ and let $\left(x_i,z_i\right)\to\left(x,z\right)$ in $\mathbb{X}\times\mathbb{R}^d$. For any $i\in\mathbb{N}\cup\{\infty\}$ we have $\mathrm{Cond}_{x_i,z_i}\left(\nu\right) = \mathcal{GP}_d\left(m_1\left(\cdot;x_i,z_i\right),k_1\left(\cdot,\cdot;x_i\right)\right)$ where m_1 and k_1 are given as in Theorem 3.4. It follows easily by uniform continuity of m_ν (see Lemma 2.4) that $m_1\left(\cdot;x_i,z_i\right)\to m_1\left(\cdot;x,z\right)$ uniformly on $\mathbb{X}$ and by uniform continuity of k_ν (Lemma 2.4, $\mathbb{X}\times\mathbb{X}$ compact) together with the continuity of $M\mapsto M^\dagger$ for matrices with the same rank that $k_1\left(\cdot,\cdot;x_i\right)\to k_1\left(\cdot,\cdot;x\right)$ uniformly on $\mathbb{X}\times\mathbb{X}$. □

A.3. Proofs of Section 4

Proof. **(Lemma 4.5)**
1.) Without loss of generality assume $\mathcal{H}_0 = 0$, since it only adds a constant

term. Furthermore $J_n(x) = \mathcal{J}_x\left(P_n^\xi\right)$ and hence it is equivalent to prove that the result holds for all $P^\xi \in \mathbb{M}$ at $n = 0$.

Assume now that $n = 0$, $x \in \mathbb{X}$ such that

$$\Sigma(x) = \Sigma_0(x) = k(x, x) + \mathcal{T}(x) \in \mathbb{R}^{d \times d}$$

has rank k for some $k \in \{0, ..., d\}$ and let $(x_i)_{i \in \mathbb{N}}$ be a sequence in $C_{0,k}$ with $x_i \xrightarrow{i \to \infty} x$ ($C_{0,k}$ is separable as subset of a separable metric space for all $k \in \{0, ..., d\}$). Recall that it holds

$$J_0(x) = \mathcal{J}_x\left(P^\xi\right) = \mathbb{E}\left[\mathcal{H}\left(\mathrm{Cond}_{x, Z(x)}\left(P^\xi\right)\right)\right],$$

so if we take $\boldsymbol{\nu}_i := \mathrm{Cond}_{x_i, Z(x_i)}\left(P^\xi\right)$ for $i \in \mathbb{N}$ and $\boldsymbol{\nu}_\infty := \mathrm{Cond}_{x, Z(x)}\left(P^\xi\right)$ we have $\boldsymbol{\nu}_i \in \mathfrak{P}(\xi)$ and by Proposition 3.12 $\boldsymbol{\nu}_i \xrightarrow[i \to \infty]{a.s.} \boldsymbol{\nu}_\infty$. It follows

$$\mathcal{H}(\boldsymbol{\nu}_i) \xrightarrow[i \to \infty]{a.s.} \mathcal{H}(\boldsymbol{\nu}_\infty)$$

by $\mathfrak{P}-$ continuity of $\mathcal{H}$ and by the above equality finally

$$\mathcal{J}_{x_i}\left(P^\xi\right) \xrightarrow[i \to \infty]{} \mathcal{J}_x\left(P^\xi\right)$$

since $(\mathcal{H}(\boldsymbol{\nu}_i))_{i \in \mathbb{N}}$ is uniformly integrable.

2.) Let $n \in \mathbb{N}$. Assume that we have a deterministic design such that $X_i = x_i$ and $Z_i = Z_i(x_i)$ for all $i \leqslant n$. We will first prove that $k_n(x, x)$ is positive definite for all $n \in \mathbb{N}$ and $x \in \mathbb{X} \setminus \{x_1, ..., x_n\}$. As already mentioned in the proof of Theorem 3.4 we know that $\left(\xi(x)^\top, \boldsymbol{Z}_n^\top\right)^\top$, where $\boldsymbol{Z}_n := \left(Z_1^\top, ..., Z_n^\top\right)^\top$, is a Gaussian vector by definition of multivariate Gaussian processes with covariance matrix

$$\begin{pmatrix} \mathrm{Var}(\xi(x)) & \mathrm{Cov}(\xi(x), \boldsymbol{Z}_n) \\ \mathrm{Cov}(\boldsymbol{Z}_n, \xi(x)) & \mathrm{Var}(\boldsymbol{Z}_n) \end{pmatrix} = \begin{pmatrix} k(x, x) & K(x, \boldsymbol{x}_n) \\ K(\boldsymbol{x}_n, x) & \Sigma(\boldsymbol{x}_n) \end{pmatrix},$$

using the same notation as in Theorem 3.4. The covariance matrix above is positive definite whenever $x \in \mathbb{X} \setminus \{x_1, ..., x_n\}$, since for every $v_0, v_1, ..., v_n \in \mathbb{R}^d$, at least one non-zero, it holds

$$\sum_{i,j=0}^{n} v_i^\top k(x_i, x_j) v_j + \sum_{i=1}^{n} v_i^\top \mathcal{T}(x_i) v_i > 0$$

by positive definiteness of k, where we define $x_0 := x$. As shown in Theorem 3.4 the covariance of $\xi(x)$ given $\boldsymbol{Z}_n$ is

$$k_n(x, x) = k(x, x) - K(x, \boldsymbol{x}_n)\Sigma(\boldsymbol{x}_n)^{-1} K(x, \boldsymbol{x}_n)^\top,$$

which is exactly the Schur complement of the covariance matrix stated above. By [26] the Schur complement of a positive definite matrix is again positive definite and hence we conclude that $k_n(x, x)$ is positive definite for all $x \in \mathbb{X} \setminus \{x_1, ..., x_n\}$.

For $x \in \{x_1, ..., x_n\}$ we obtain again a Gaussian vector $(\xi(x_i), \boldsymbol{Z}_n)$ with covariance matrix

$$\begin{pmatrix} \mathrm{Var}(\xi(x_i)) & \mathrm{Cov}(\xi(x_i), \boldsymbol{Z}_n) \\ \mathrm{Cov}(\boldsymbol{Z}_n, \xi(x_i)) & \mathrm{Var}(\boldsymbol{Z}_n) \end{pmatrix}$$
$$= \begin{pmatrix} k(x_i, x_i) & K(x_i, \boldsymbol{x}_n) \\ K(\boldsymbol{x}_n, x_i) & K(\boldsymbol{x}_n) \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \mathcal{T}(\boldsymbol{x}_n) \end{pmatrix}.$$

The first matrix is positive semi-definite as covariance matrix of a Gaussian vector (one entry is double) and the second matrix contains the positive definite matrix $\mathcal{T}(\boldsymbol{x}_n)$. Since $k(x_i, x_i)$ is also positive definite, we conclude

$$(v_1, v_2)^\top \begin{pmatrix} \mathrm{Var}(\xi(x_i)) & \mathrm{Cov}(\xi(x_i), \boldsymbol{Z}_n) \\ \mathrm{Cov}(\boldsymbol{Z}_n, \xi(x_i)) & \mathrm{Var}(\boldsymbol{Z}_n) \end{pmatrix} (v_1, v_2) > 0$$

for all $(v_1, v_2) \in \mathbb{R}^d \times \mathbb{R}^{nd}$, $(v_1, v_2) \neq 0$ and hence the positive definiteness of the covariance matrix of $(\xi(x_i), \boldsymbol{Z}_n)$. This means also the Schur complement $k_n(x_i, x_i)$ is positive definite. We finally conclude that $k_n(x, x)$ is positive definite for all $x \in \mathbb{X}$.

If $k_n(x, x)$ is positive definite, then $\Sigma_n(x)$ has full rank since $\mathrm{rank}(\Sigma_n(x)) \geqslant \max\{\mathrm{rank}(k_n(x, x)), \mathrm{rank}(\mathcal{T}(x))\} = d$ and hence the sample paths of J_n are continuous on $\mathbb{X}$ by 1.).

Since J_n has continuous sample paths and $\mathbb{X}$ is compact, we know that $A_n(\omega) := \{x \in \mathbb{X} : J_n(x)(\omega) = \inf_{x \in \mathbb{X}} J_n(x)(\omega)\}$ is a non-empty closed set for every $\omega \in \Omega$. Hence the mapping $\omega \mapsto A_n(\omega)$ is a $\mathcal{F}_n$-measurable random closed set that admits a $\mathcal{F}_n$-measurable selection X_{n+1}, i.e. a $\mathbb{X}$-valued random element such that $X_{n+1}(\omega) \in A_n(\omega)$ for all $\omega \in \Omega$ (see Theorem 2.13 in [13]).

3.) Define for $\omega \in \Omega$ and $0 \leqslant k \leqslant d$ the sets

$$C^{\leqslant}_{n,k}(\omega) := \{x \in \mathbb{X} : \mathrm{rank}(\Sigma_n(x))(\omega) \leqslant k\},$$

$$M_n(\omega) := \left\{x \in \mathbb{X} : J_n(x)(\omega) \leqslant \inf_{y \in \mathbb{X}} J_n(y)(\omega) + \epsilon_n\right\}.$$

For each $0 \leqslant k \leqslant d$ the set $C^{\leqslant}_{n,k}(\omega)$ is compact as closed subsets of the compact metric space $\mathbb{X}$ by continuity of the mapping $x \mapsto (\Sigma_n(x))(\omega)$ and closedness of the set $\{M \in \mathbb{R}^{d \times d} : \mathrm{rank}(M) \leqslant k\}$. Since $J_n(\cdot)(\omega)$ is by definition constant if $\Sigma_n(\cdot)(\omega) \equiv 0$ and hence $M_n(\omega) = \mathbb{X}$, we assume that $\omega \in \Omega$ is such that $\Sigma_n(\cdot)(\omega) \not\equiv 0$. Without loss of generality, we assume $\inf_{y \in \mathbb{X}} J_n(y)(\omega) < H_n(\omega)$ and consider the following two cases:

Assume $C^{\leqslant}_{n,0}(\omega) \cap M_n(\omega) \neq \varnothing$. Then it holds for $x_* \in C^{\leqslant}_{n,0}(\omega) \cap M_n(\omega)$ that $\Sigma_n(x_*)(\omega) = 0$ and hence $J_n(x_*)(\omega) = H_n(\omega)$, which yields

$$\inf_{y \in \mathbb{X}} J_n(y)(\omega) \leqslant J_n(x)(\omega) \leqslant \inf_{y \in \mathbb{X}} J_n(y)(\omega) + \epsilon_n$$

for all $x \in \mathbb{X}$ and hence $M_n(\omega) = \mathbb{X}$.

Assume $C^{\leqslant}_{n,0}(\omega) \cap M_n(\omega) = \varnothing$. Then there must exist an integer $1 \leqslant k_* \leqslant d$ such that $C^{\leqslant}_{n,k}(\omega) \cap M_n(\omega) = \varnothing$ for all $0 \leqslant k < k_*$ and $C^{\leqslant}_{n,k_*}(\omega) \cap M_n(\omega) \neq \varnothing$.

Indeed, if $(x_j)_{j\in\mathbb{N}} \subset \mathbb{X}$ is a sequence such that $J_n(x_j)(\omega) \to \inf_{y\in\mathbb{X}} J_n(y)(\omega)$ for $j\to\infty$, then we have for some j large enough that

$$J_n(x_j)(\omega) \leqslant \inf_{y\in\mathbb{X}} J_n(y)(\omega) + \epsilon_n$$

and $J_n(x_j)(\omega) < H_n(\omega)$, which implies $\operatorname{tr}(\Sigma_n(x_j))(\omega) > 0$ and hence $x_j \in C^{\leqslant}_{n,k}(\omega) \cap M_n(\omega)$ for some $1 \leqslant k \leqslant d$. Hence the sets $C^{\leqslant}_{n,k}(\omega) \cap M_n(\omega)$ are not empty for some k large enough. Taking the smallest integer $k_* \geqslant 1$ such that $C^{\leqslant}_{n,k_*}(\omega) \cap M_n(\omega) \neq \varnothing$ gives the desired result.

Choosing k_* this way means

$$C^{\leqslant}_{n,k_*}(\omega) \cap M_n(\omega) \subset C_{n,k_*}$$

and since J_n is continuous on C_{n,k_*} by 1.), we conclude that $C^{\leqslant}_{n,k_*}(\omega) \cap M_n(\omega)$ is closed and hence compact. This means the mapping $\omega \mapsto C^{\leqslant}_{n,k_*}(\omega) \cap M_n(\omega)$ is an $\mathcal{F}_n$-measurable random closed set and we can chose an $\mathcal{F}_n$-measurable selection $X^{(k_*)}_{n+1}$ that takes values in this random closed set (see Theorem 2.13 in [13]). The desired ε-quasi SUR sequential design $(X_n)_{n\geqslant 1}$ can hence be defined by

$$X_{n+1} = \begin{cases} x & , M_n = \mathbb{X} \\ X^{(k_*)}_{n+1} & , \text{else} \end{cases}$$

for some arbitrary $x \in \mathbb{X}$. □

Proof. **(Proposition 4.3)**
Using the notation from the proof of Proposition 3.8, we see that

$$\mathcal{J}_x(\nu) = \int_{\mathbb{R}^d} \mathcal{H}\left(\kappa_1\left(x, m_\nu(x) + \Sigma_\nu(x)^{\frac{1}{2}} u, \nu\right)\right) \phi_d(u)\, du.$$

Using Lemma A.4 in the Appendix and the measurability of κ_1 (see proof of Proposition 3.8), we see that the integrand is a $\mathcal{B}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{M}$ measurable function of (x, u, ν). The claim follows by Fubini's Theorem. □

A.4. Proofs of Section 5

Theorem A.5. *Let $\mathcal{H}$ be an uncertainty functional on $\mathbb{M}$ that has the supermartingale property, $\mathcal{G}$ the associated maximal expected gain functional and $(X_n)_{n\geqslant 1}$ be an ε-quasi SUR sequential design for $\mathcal{H}$.*

1. *Then it holds*
$$\mathcal{G}\left(P^\xi_n\right) \xrightarrow[n\to\infty]{a.s.} 0.$$
2. *If moreover*
   - *(a)* $H_n = \mathcal{H}\left(P^\xi_n\right) \xrightarrow[n\to\infty]{a.s.} \mathcal{H}\left(P^\xi_\infty\right)$,
   - *(b)* $\mathcal{G}\left(P^\xi_n\right) \xrightarrow[n\to\infty]{a.s.} \mathcal{G}\left(P^\xi_\infty\right)$,

(c) $\mathbb{Z}_{\mathcal{H}} := \{\nu \in \mathbb{M} : \mathcal{H}(\nu) = 0\} = \{\nu \in \mathbb{M} : \mathcal{G}(\nu) = 0\} =: \mathbb{Z}_{\mathcal{G}}$,

then we have

$$H_n \xrightarrow[n\to\infty]{a.s.} 0.$$

Proof. See Theorem 3.12 in [3]. The proof works the same way for multivariate Gaussian processes. □

Theorem A.6. *Let $\mathcal{H}$ be an uncertainty functional on $\mathbb{M}$ and $\mathcal{G}$ the associated maximal expected gain functional. Assume that we can decompose $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$, where*

1. *$\mathcal{H}_0(\nu) = \int_{\mathbb{S}} L_0 d\nu$ for some $L_0 \in \bigcap_{\nu\in\mathbb{M}} \mathcal{L}^1(\mathbb{S}, \mathcal{S}, \nu)$ and*
2. *$\mathcal{H}_1$ is $\mathfrak{P}$-continuous, $\mathfrak{P}$-uniformly integrable and has the supermartingale property.*

Then, for any ε-quasi SUR sequential design associated with $\mathcal{H}$, it holds

$$\mathcal{G}\left(P_\infty^\xi\right) \overset{a.s.}{=} 0.$$

Proof. Using Theorem A.5 and Proposition 3.11, it is straightforward to obtain

$$\mathcal{G}_x\left(P_\infty^\xi\right) = \mathcal{H}\left(P_\infty^\xi\right) - \mathbb{E}_\infty\left[\mathcal{H}\left(P_{\infty,x}^\xi\right)\right] \overset{a.s.}{=} 0$$

for each $x \in \mathbb{X}$ by following the same steps as in the proof of Theorem 3.16 in [3]. To conclude, we need to show that

$$\mathcal{G}\left(P_\infty^\xi\right) = \sup_{x\in\mathbb{X}} \mathcal{G}_x\left(P_\infty^\xi\right) \overset{a.s.}{=} 0.$$

Define with $\Sigma_\infty(x) = k_\infty(x,x) + \mathcal{T}(x)$ the random sets

$$C_{k,\infty} := \{x \in \mathbb{X} : \operatorname{rank}(\Sigma_\infty(x)) = k\}$$

and

$$C_{k,\infty}^{\leqslant} := \{x \in \mathbb{X} : \operatorname{rank}(\Sigma_\infty(x)) \leqslant k\}$$

for $k = 0, ..., d$. For $k \in \{0, ..., d\}$ the sets $C_{k,\infty}^{\leqslant}$ are closed subsets of $\mathbb{X}$ ($\Sigma_\infty(\omega)$ is continuous for all $\omega \in \Omega$ and $\left\{M \in \mathbb{R}^{d\times d} : \operatorname{rank}(M) \leqslant k\right\}$ is closed) and hence compact and separable with $C_{k,\infty}^{\leqslant} \subseteq C_{k+1,\infty}^{\leqslant}$, $k \in \{0, ..., d-1\}$, and $\mathbb{X} = C_{d,\infty}^{\leqslant}$. By the previous Lemma we know that the sample paths of J_∞ are continuous on each set $C_{k,\infty}$ and hence $x \mapsto \mathcal{G}_x\left(P_\infty^\xi\right)$ has continuous sample paths on $C_{k,\infty}$.

Let $\{x_i\}_{i\in\mathbb{N}}$ be a countable dense subset of $C_{1,\infty}^{\leqslant}$. In the previous part we have seen that $\mathcal{G}_{x_i}\left(P_\infty^\xi\right) = 0$ for all $i \in \mathbb{N}$, almost surely. Using the continuity of $x \mapsto \mathcal{G}_x\left(P_\infty^\xi\right)$ on $C_{1,\infty}$ and the fact that $\mathcal{G}_x\left(P_\infty^\xi\right) = 0$ on $C_{0,\infty}$, we conclude

$$\sup_{x\in C_{1,\infty}^{\leqslant}} \mathcal{G}_x\left(P_\infty^\xi\right) = 0$$

almost surely. Assume now that we have shown $\mathcal{G}_x\left(P_\infty^\xi\right) = 0$ on $C_{k,\infty}^{\leqslant}$ almost surely. Let $\{x_i\}_{i\in\mathbb{N}}$ be a countable dense subset of $C_{k+1,\infty}^{\leqslant}$. Then $\mathcal{G}_{x_i}\left(P_\infty^\xi\right) = 0$ for all $i \in \mathbb{N}$ almost surely and using the continuity of $x \mapsto \mathcal{G}_x\left(P_\infty^\xi\right)$ on $C_{k+1,\infty}$ and $\mathcal{G}_x\left(P_\infty^\xi\right) = 0$ on $C_{k,\infty}^{\leqslant}$ we conclude

$$\sup_{x\in C_{k+1,\infty}^{\leqslant}} \mathcal{G}_x\left(P_\infty^\xi\right) = 0$$

almost surely. This leads in the end for $k = d$ to

$$\mathcal{G}\left(P_\infty^\xi\right) = \sup_{x\in\mathbb{X}} \mathcal{G}_x\left(P_\infty^\xi\right) = 0$$

almost surely, since $\mathbb{X} = C_{d,\infty}^{\leqslant}$. ☐

Proof. **(Proposition 5.1)**
Take $\mathcal{H}_0 = 0$ and $\mathcal{H}_1 = \mathcal{H}$. Then we have for any ε-quasi SUR sequential design associated with $\mathcal{H}$ that $\mathcal{G}\left(P_\infty^\xi\right) \overset{a.s.}{=} 0$ by Theorem A.6. By the first part of Theorem A.5 it holds $\mathcal{G}\left(P_n^\xi\right) \xrightarrow[n\to\infty]{a.s.} 0$ and hence $\mathcal{G}\left(P_n^\xi\right) \xrightarrow[n\to\infty]{a.s.} \mathcal{G}\left(P_\infty^\xi\right)$. By $\mathfrak{P}$-continuity we also have $\mathcal{H}\left(P_n^\xi\right) \xrightarrow[n\to\infty]{a.s.} \mathcal{H}\left(P_\infty^\xi\right)$ and hence all conditions for the second part of Theorem A.5 are satisfied. We conclude $H_n \xrightarrow[n\to\infty]{a.s.} 0$. ☐

A.4.1. Proofs of Section 5.1 (IBV)

Proof. **(Lemma 5.2)**
$\mathcal{H}^{IBV}$ is $\mathfrak{P}$-uniformly integrable, since the uncertainty functional is upper-bounded. Indeed, we have for every Gaussian random element ξ in $\mathbb{S}$ and $u \in \mathbb{X}$

$$\mathrm{Var}\left(\mathbf{1}_{\Gamma(\xi)}(u)\right) \leqslant \frac{1}{4},$$

since $\mathbf{1}_{\Gamma(\xi)}(u) \in [0,1]$. Hence we have for any measure $\nu \in \mathbb{M}$ taking $\xi \sim \nu$ that

$$\left|\mathcal{H}^{IBV}(\nu)\right| = \int_{\mathbb{X}} \mathrm{Var}\left(\mathbf{1}_{\Gamma(\xi)}(u)\right)\mu(du) \leqslant \frac{\mu(\mathbb{X})}{4},$$

since μ is a finite measure over $\mathbb{X}$. Furthermore, H_n^{IBV} is an $\mathcal{F}_n$-measurable random variable and integrable by definition of the conditional expectation. Furthermore, we have for $p_n(u) = \mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u)\,|\mathcal{F}_n\right]$ by the tower property and Jensen's inequality

$$\begin{aligned}\mathbb{E}\left[H_n^{IBV}|\mathcal{F}_{n-1}\right] &= \int_{\mathbb{X}}\left(\mathbb{E}\left[p_n(u)\,|\mathcal{F}_{n-1}\right] - \mathbb{E}\left[p_n(u)^2\,|\mathcal{F}_{n-1}\right]\right)\mu(du)\\ &\overset{a.s.}{\leqslant} \int_{\mathbb{X}}\left(p_{n-1}(u) - p_{n-1}(u)^2\right)\mu(du)\end{aligned}$$

$$= H_{n-1}^{IBV}.$$

□

Proof. **(Lemma 5.3)**
That $\mathbb{X}$ can be written as the disjoint union

$$\bigcup_{J_1,J_2\subseteq\{1,...,d\}} B_{J_1,J_2}(\omega)$$

follows already by assumption 3. of the Lemma.

Let ξ be a random element in $\mathbb{S}$ with $\xi \sim \mathcal{GP}_d(m,k)$ and let $\boldsymbol{\nu}_n \in \mathfrak{P}(\xi)$ for all $n \in \mathbb{N} \cup \{\infty\}$ such that $\boldsymbol{\nu}_n \to \boldsymbol{\nu}_\infty$ almost surely. By definition of $\mathfrak{P}(\xi)$ there exist σ-algebras $\mathcal{G}_n$ such that $\boldsymbol{\nu}_n = P(\xi \in \cdot|\mathcal{G}_n)$ for all $n \in \mathbb{N} \cup \{\infty\}$. For the first property note that for $u \in \mathbb{X}$ and $j \in \{1,...,d\}$ we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{k_\infty(u,u)_{jj}=0}\left(\xi(u)_j - m_\infty(u)_j\right)^2\right]\\
=&\mathbb{E}\left[\mathbf{1}_{k_\infty(u,u)_{jj}=0}\mathbb{E}\left[\left(\xi(u)_j - m_\infty(u)_j\right)^2|\mathcal{G}_\infty\right]\right]\\
=&\mathbb{E}\left[\mathbf{1}_{k_\infty(u,u)_{jj}=0}k_\infty(u,u)_{jj}\right]\\
=&0
\end{aligned}$$

by the tower property and since $k_\infty(u,u)$ is $\mathcal{G}_\infty$-measurable. This means

$$\mathbf{1}_{k_\infty(u,u)_{jj}=0}\left(\xi(u)_j - m_\infty(u)_j\right)^2 \overset{a.s.}{=} 0,$$

since it is a non-negative random variable, and hence for almost all $\omega \in \Omega$ we have that $\sum_{j\in J_2} k_\infty(u,u)(\omega)_{jj}^2 = 0$ implies

$$\xi(u)(\omega)_j = m_\infty(u)(\omega)_j$$

for all $j \in J_2$. Note that $(u,\omega) \mapsto m_\infty(u)(\omega)$ and $(u,\omega) \mapsto k_\infty(u,u)(\omega)$ are jointly measurable by continuity of the sample paths $u \mapsto m_\infty(u)(\omega)$ and $u \mapsto k_\infty(u,u)(\omega)$ for all $\omega \in \Omega$. By Fubini-Tonelli we conclude

$$\begin{aligned}
&\mathbb{E}\left[\mu\left(B_{J_1,J_2}\right)\right]\\
=&\int_{\mathbb{X}}\left(\mathbf{1}_{\sum_{j\in J_1} k(u,u)_{jj}^2=0}\prod_{j\in J_1^c}\mathbf{1}_{k(u,u)_{jj}>0}\right)\cdot\\
&\mathbb{E}\left[\mathbf{1}_{\sum_{j\in J_2} k_\infty(u,u)_{jj}^2=0}\prod_{j\in J_2}\mathbf{1}_{m_\infty(u)_j=t_j}\cdot\right.\\
&\left.\prod_{j\in J_2^c}\max\left(\mathbf{1}_{k_\infty(u,u)_{jj}>0},\mathbf{1}_{m_\infty(u)_j\neq t_j}\right)\right]\mu(du)
\end{aligned}$$

$$\begin{aligned}
&\leqslant \int_{\mathbb{X}} \left(\mathbf{1}_{\sum_{j\in J_1} k(u,u)_{jj}^2=0} \prod_{j\in J_1^c} \mathbf{1}_{k(u,u)_{jj}>0} \right) \cdot \\
&\quad E\left[\mathbf{1}_{\sum_{j\in J_2} k_\infty(u,u)_{jj}^2=0} \prod_{j\in J_2} \mathbf{1}_{\xi(u)_j=t_j} \right] \mu(du) \\
&\leqslant \int_{\mathbb{X}} \left(\mathbf{1}_{\sum_{j\in J_1} k(u,u)_{jj}^2=0} \prod_{j\in J_1^c} \mathbf{1}_{k(u,u)_{jj}>0} \right) \mathbb{E}\left[\prod_{j\in J_2} \mathbf{1}_{\xi(u)_j=t_j} \right] \mu(du) \\
&=0,
\end{aligned}$$

where the last equality follows by the assumption that there exists $j^* \in \{1,...,d\}$ with $j^* \in J_2$ and $j^* \in J_1^c$. Indeed, we have for the multivariate Gaussian process ξ that $\xi(u)_{j*} \sim \mathcal{N}\left(m(u)_{j*}, k(u,u)_{j*j*}\right)$ which implies $\xi(u)_{j*} \overset{a.s.}{\neq} t_{j*}$ since $k(u,u)_{j*j*} > 0$ and hence

$$\begin{aligned}
\mathbb{E}\left[\prod_{j\in J_2} \mathbf{1}_{\xi(u)_j=t_j} \right] &= P\left(\bigcap_{j\in\ J_2} \left\{ \xi(u)_j = t_j \right\} \right) \\
&\leqslant \min_{j\in J_2} P\left(\xi(u)_j = t_j \right) \\
&\leqslant P\left(\xi(u)_{j*} = t_{j*} \right) \\
&= 0.
\end{aligned}$$

Since $\mu(B_{J_1,J_2})$ is almost surely non-negative, the first property follows.

We will now turn to the second property. For $j \in \{1,...,d\}$ we have that $k(u,u)_{jj} = 0$ implies $k_n(u,u)_{jj} \overset{a.s.}{=} 0$ for all $n \in \mathbb{N} \cup \{\infty\}$. Indeed, for a random variable X and σ-algebra $\mathcal{F}$

$$\mathbb{E}\left[(X - \mathbb{E}[X])^2 \right] = 0$$

implies $X \overset{a.s.}{=} \mathbb{E}[X]$ and hence also $\mathbb{E}[X|\mathcal{F}] \overset{a.s.}{=} X$. We conclude

$$\mathbb{E}\left[(X - \mathbb{E}[X|\mathcal{F}])^2 |\mathcal{F} \right] \overset{a.s.}{=} 0.$$

Hence for almost all $\omega \in \Omega$ and all $n \in \mathbb{N} \cup \{\infty\}$ we have for $u \in B_{J_1,J_2}(\omega)$ that

$$\xi(u)(\omega)_j = m_n(u)(\omega)_j = m(u)_j = \begin{cases} a_j & , j \in J_1 \backslash J_2 \\ t_j & , j \in J_2 \end{cases}$$

for some $a_j \in \mathbb{R}$, $a_j \neq t_j$. We will use the notation

$$\xi(u)_{J_1^c} := \left(\xi(u)_j \right)_{j\in J_1^c}$$

and

$$t_{J_1^c} := (t_j)_{j \in J_1^c}$$

for the sub-vectors containing only the elements with index in J_1^c. Then combining the previous result with the Portmanteau Theorem yields for almost all $\omega \in \Omega$ for all $u \in B_{J_1,J_2}(\omega)$

$$\begin{aligned} P(\xi(u) \geqslant T|\mathcal{G}_n)(\omega) =& P\left(\xi(u)_{J_1^c} \geqslant t_{J_1^c}|\mathcal{G}_n\right)(\omega) \prod_{j \in J_1 \setminus J_2} \mathbf{1}_{m(u)_j \geqslant t_j} \\ \xrightarrow[n\to\infty]{} & P\left(\xi(u)_{J_1^c} \geqslant t_{J_1^c}|\mathcal{G}_\infty\right)(\omega) \prod_{j \in J_1 \setminus J_2} \mathbf{1}_{m(u)_j \geqslant t_j} \\ =& P(\xi(u) \geqslant T|\mathcal{G}_\infty)(\omega), \end{aligned}$$

since $\xi(u)_{J_1^c}$ is again a Gaussian vector with almost surely convergent conditional mean and covariance. Indeed, for all elements in $j \in J_1^c$ (which also means $j \notin J_2$ by assumption) we have $k_\infty(u,u)(\omega)_{jj} > 0$ or $m_\infty(u)(\omega)_j \neq t_j$ which implies for the boundary of $\mathbf{T}_{J_1^c} = \bigtimes_{j \in J_1^c} [t_j \infty)$, that

$$\begin{aligned} P\left(\xi(u)_{J_1^c} \in \partial \mathbf{T}_{J_1^c}|\mathcal{G}_\infty\right)(\omega) &= P\left(\exists j \in J_1^c : \xi(u)_j = t_j|\mathcal{G}_\infty\right)(\omega) \\ &= P\left(\bigcup_{j \in J_1^c} \left\{\xi(u)_j = t_j\right\}|\mathcal{G}_\infty\right)(\omega) \\ &\leqslant \sum_{j \in J_1^c} P\left(\xi(u)_j = t_j|\mathcal{G}_\infty\right)(\omega) \\ &= 0 \end{aligned}$$

and hence we can apply the Portmanteau Theorem, which concludes the second property. □

Proof. **(Lemma 5.5)**
It remains to show $\left\{\nu \in \mathbb{M} : \mathcal{H}^{IBV}(\nu) = 0\right\} \supseteq \left\{\nu \in \mathbb{M} : \mathcal{G}^{IBV}(\nu) = 0\right\}$. Let $\nu = \mathcal{GP}_d(m,k) \in \left\{\nu \in \mathbb{M} : \mathcal{G}^{IBV}(\nu) = 0\right\}$ and $\xi \sim \nu$, then it holds by the law of total variance (see Theorem 8.2 in [11])

$$\begin{aligned} 0 &= \sup_{x \in \mathbb{X}} \mathcal{G}_x^{IBV}(\nu) \\ &= \sup_{x \in \mathbb{X}} \mathcal{H}^{IBV}(\nu) - \mathcal{J}_x^{IBV}(\nu) \\ &= \sup_{x \in \mathbb{X}} \int_{\mathbb{X}} \mathrm{Var}\left(\mathbf{1}_{\Gamma(\xi)}(u)\right)\mu(du) - \mathbb{E}\left[\int_{\mathbb{X}} \mathrm{Var}\left(\mathbf{1}_{\Gamma(\xi)}(u)|Z(x)\right)\mu(du)\right] \\ &= \sup_{x \in \mathbb{X}} \int_{\mathbb{X}} \mathrm{Var}\left(\mathbf{1}_{\Gamma(\xi)}(u)\right) - \mathbb{E}\left[\mathrm{Var}\left(\mathbf{1}_{\Gamma(\xi)}(u)|Z(x)\right)\right]\mu(du) \\ &= \sup_{x \in \mathbb{X}} \int_{\mathbb{X}} \mathrm{Var}\left(\mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u)|Z(x)\right]\right)\mu(du). \end{aligned}$$

This means for all $x \in \mathbb{X}$, with $Z(x) = \xi(x) + \tau(x) U$, $U \sim \mathcal{N}_d(0, I_d)$ independent of ξ, that we have

$$\begin{aligned}\operatorname{Var}\left(\mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u) | Z(x)\right]\right) &= \operatorname{Var}\left(P\left(\xi(u) \geqslant T | Z(x)\right)\right) \\ &= \operatorname{Var}\left(\mathcal{N}_d\left(m_1(u), k_1(u,u)\right)(\mathbf{T})\right) \\ &= 0\end{aligned}$$

for μ-almost all $u \in \mathbb{X}$. Note that $\mathcal{N}_d(m_1(u), k_1(u,u))$ is a random measure with

$$\mathcal{N}_d\left(m_1(u), k_1(u,u)\right)(\omega) = \mathcal{N}_d\left(m_1(u)(\omega), k_1(u,u)\right),$$

where $m_1(u)$ and $k_1(u)$ are given by

$$\begin{aligned}m_1(u) &= m(u) + k(u,x)\Sigma(x)^{\dagger}\left(Z(x) - m(x)\right), \\ k_1(u,u) &= k(u,u) - k(u,x)\Sigma(x)^{\dagger} k(u,x)^{\top}.\end{aligned}$$

Since $k_1(u,u)$ does not depend on $\omega \in \Omega$ and $\operatorname{Var}\left(\mathcal{N}_d\left(m_1(u), k_1(u,u)\right)(\mathbf{T})\right) = 0$, we conclude that $m_1(u)$ has to be P-almost surely constant and hence

$$\begin{aligned}\operatorname{Var}\left(m_1(u)\right) &= \operatorname{Var}\left(k(u,x)\Sigma(x)^{\dagger}\left(Z(x) - m(x)\right)\right) \\ &= k(u,x)\Sigma(x)^{\dagger}\operatorname{Var}\left(Z(x)\right)\left(k(u,x)\Sigma(x)^{\dagger}\right)^{\top} \\ &= k(u,x)\Sigma(x)^{\dagger}\Sigma(x)\Sigma(x)^{\dagger}k(u,x)^{\top} \\ &= k(u,x)\Sigma(x)^{\dagger}k(u,x)^{\top} \\ &= 0 \in \mathbb{R}^{d\times d}.\end{aligned}$$

Since $\Sigma(x)$ is symmetric positive semi-definite, we know that $\Sigma(x)^{\dagger}$ is symmetric positive semi-definite and hence $k(u,x)\Sigma(x)^{\dagger}k(u,x)^{\top} = 0$ if and only if $k(u,x)\Sigma(x)^{\dagger} = 0$. Indeed, if $A \in \mathbb{R}^{d\times d}$, $B \in S_d^+$ and $ABA^{\top} = 0$, then

$$ABA^{\top} = AQQ^{\top}A^{\top} = AQ\left(AQ\right)^{\top} = 0$$

for $Q \in \mathbb{R}^{d\times d}$ with $B = QQ^{\top}$ and hence $\|AQ\|_F^2 = \operatorname{tr}\left(\left(AQ\right)^{\top}AQ\right) = 0$ implying $AQ = 0$ and finally $AQQ^{\top} = AB = 0$ (the other direction is trivial). Using the properties $\Sigma(x)\Sigma(x)^{\dagger}\Sigma(x) = \Sigma(x)$ and $\left(\Sigma(x)^{\dagger}\Sigma(x)\right)^{\top} = \Sigma(x)^{\dagger}\Sigma(x)$ of the pseudo-inverse $\Sigma(x)^{\dagger}$, this implies for every $x \in \mathbb{X}$

$$k(u,x)\Sigma(x) = 0 \in \mathbb{R}^{d\times d}$$

for μ-almost all $u \in \mathbb{X}$. This also yields $k(u,u)\Sigma(u) = 0$ and hence

$$\operatorname{tr}\left(k(u,u)\Sigma(u)\right) = \operatorname{tr}\left(k(u,u)^2\right) + \operatorname{tr}\left(k(u,u)\mathcal{T}(u)\right) = 0$$

for μ-almost all $u \in \mathbb{X}$.

We have

$$\operatorname{tr}\left(k(u,u)^2\right) = \operatorname{tr}\left(k(u,u)^{\top}k(u,u)\right) = \|k(u,u)\|_F^2 \geqslant 0$$

and

$$\operatorname{tr}\left(k\left(u,u\right)\mathcal{T}\left(u\right)\right)\geqslant 0,$$

since $\mathcal{T}(u)$ and $k(u,u)$ are both symmetric and positive semi-definite. Indeed, we can write for $A, B \in S_d^+$ that

$$\operatorname{tr}\left(AB\right)=\operatorname{tr}\left(AQQ^\top\right)=\operatorname{tr}\left(Q^\top AQ\right)=\sum_{i=1}^{d}q_i^\top Aq_i\geqslant 0$$

for $Q\in\mathbb{R}^{d\times d}$ with $B=QQ^\top$ by the Spectral Theorem (see Theorem 1.3.1 in [22]).

We conclude $\|k(u,u)\|_F^2=0$ and hence $k(u,u)=0$ for μ-almost all $u\in\mathbb{X}$, which yields $P(\xi(u)\geqslant T)=\mathbf{1}_{m(u)\geqslant T}$ and finally

$$\begin{aligned}\mathcal{H}^{IBV}(\nu)&=\int_{\mathbb{X}}\operatorname{Var}\left(\mathbf{1}_{\Gamma(\xi)}\left(u\right)\right)\mu\left(du\right)\\&=\int_{\mathbb{X}}P\left(\xi\left(u\right)\geqslant T\right)\left(1-P\left(\xi\left(u\right)\geqslant T\right)\right)\mu\left(du\right)\\&=0.\end{aligned}$$

□

Proof. **(Theorem 5.6)**
The first statement follows by combining the three Lemmas in Section 5.1 with Proposition 5.1. For the second part note that, as seen in the proof of Lemma 5.3, we have $k_\infty(u,u)(\omega)=0\in\mathbb{R}^{d\times d}$ for μ-almost all $u\in\mathbb{X}$, since $\mathcal{H}^{IBV}\left(P_\infty^\xi(\omega)\right)=0$ for almost all $\omega\in\Omega$. Furthermore, we have for the random set

$$A=\bigcup_{\substack{J_1,J_2\subseteq\{1,\ldots,d\}\\J_2\subseteq J_1}}B_{J_1,J_2}$$

that

$$\mu\left(\mathbb{X}\right)=\mu\left(A\left(\omega\right)\right)$$

for almost all $\omega\in\Omega$ and

$$p_n\left(u\right)\left(\omega\right)=P\left(\xi\left(u\right)\geqslant T|\mathcal{F}_n\right)\left(\omega\right)\xrightarrow[n\to\infty]{}P\left(\xi\left(u\right)\geqslant T|\mathcal{F}_\infty\right)\left(\omega\right)=p_\infty\left(u\right)\left(\omega\right)$$

for all $u\in A(\omega)$ by Lemma 5.3. Combining both statements we conclude

$$p_n\left(u\right)\left(\omega\right)\xrightarrow[n\to\infty]{}\mathbf{1}_{\xi(u)(\omega)\geqslant T}$$

for P-almost all $\omega\in\Omega$ and μ-almost all $u\in\mathbb{X}$, since $k_\infty(u,u)\overset{a.s.}{=}0$ yields $p_\infty(u)=\mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u)\,|\mathcal{F}_\infty\right]\overset{a.s.}{=}\mathbf{1}_{\Gamma(\xi)}(u)$. The claim follows by using the Dominated Convergence Theorem. □

A.4.2. Proofs of Section 5.2 (EMV)

Proof. **(Lemma 5.7)**
$\mathcal{H}^{EMV}$ is $\mathfrak{P}$-uniformly integrable since we have for any measure $\nu \in \mathbb{M}$ with $\xi \sim \nu$ that

$$\left|\mathcal{H}^{EMV}(\nu)\right| \leqslant \mathbb{E}\left[\alpha(\xi)^2\right] \leqslant \mu(\mathbb{X})^2 < \infty,$$

since $\alpha(\xi) = \mu(\Gamma(\xi))$ and $\Gamma(\xi) \subseteq \mathbb{X}$ almost surely.

H_n^{EMV} is $\mathcal{F}_n$-measurable and integrable by definition of the conditional expectation. Furthermore, it holds by Jensen's inequality and the tower property that

$$\begin{aligned}\mathbb{E}\left[\operatorname{Var}(\alpha(\xi)|\mathcal{F}_n)|\mathcal{F}_{n-1}\right] &= \mathbb{E}\left[\mathbb{E}\left[\alpha(\xi)^2|\mathcal{F}_n\right] - \mathbb{E}\left[\alpha(\xi)|\mathcal{F}_n\right]^2|\mathcal{F}_{n-1}\right] \\ &= \mathbb{E}\left[\alpha(\xi)^2|\mathcal{F}_{n-1}\right] - \mathbb{E}\left[\mathbb{E}\left[\alpha(\xi)|\mathcal{F}_n\right]^2|\mathcal{F}_{n-1}\right] \\ &\overset{a.s.}{\leqslant} \mathbb{E}\left[\alpha(\xi)^2|\mathcal{F}_{n-1}\right] - \mathbb{E}\left[\alpha(\xi)|\mathcal{F}_{n-1}\right]^2 \\ &= \operatorname{Var}(\alpha(\xi)|\mathcal{F}_{n-1}).\end{aligned}$$

□

Proof. **(Lemma 5.8)**
Let ξ be a random element in $\mathbb{S}$ with $\xi \sim \mathcal{GP}_d(m,k)$ and let $\boldsymbol{\nu}_n \in \mathfrak{P}(\xi)$ for all $n \in \mathbb{N} \cup \{\infty\}$ such that $\boldsymbol{\nu}_n \to \boldsymbol{\nu}_\infty$ almost surely. By definition of $\mathfrak{P}(\xi)$ there exist σ-algebras $\mathcal{G}_n$ such that $\boldsymbol{\nu}_n = P(\xi \in \cdot|\mathcal{G}_n)$ for all $n \in \mathbb{N} \cup \{\infty\}$.

Note first that for almost all $\omega \in \Omega$ and all $n \in \mathbb{N} \cup \{\infty\}$ we have for $u_i \in B_{J_1^i,J_2^i}(\omega)$, $j \in J_1^i$ and $i \in \{1,2\}$

$$\xi(u_i)(\omega)_j = m_n(u_i)(\omega)_j = m(u_i)_j = \begin{cases} a_j^i & , j \in J_1^i \backslash J_2^i \\ t_j & , j \in J_2^i \end{cases}$$

for some $a_j^i \in \mathbb{R}$ with $a_j^i \neq t_j$.

Combining this with the Portmanteau Theorem leads to

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u_i)|\mathcal{G}_n\right](\omega) =& P(\xi(u_i) \geqslant T|\mathcal{G}_n)(\omega) \\ =& P\left(\xi(u_i)_{(J_1^i)^c} \geqslant t_{(J_1^i)^c}|\mathcal{G}_n\right)(\omega) \prod_{j \in J_1^i \backslash J_2^i} \mathbf{1}_{m(u_i)_j \geqslant t_j} \\ \xrightarrow[n\to\infty]{}& P\left(\xi(u)_{(J_1^i)^c} \geqslant t_{(J_1^i)^c}|\mathcal{G}_\infty\right)(\omega) \prod_{j \in J_1^i \backslash J_2^i} \mathbf{1}_{m(u_i)_j \geqslant t_j} \\ =& P(\xi(u_i) \geqslant T|\mathcal{G}_\infty)(\omega) \\ =& \mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u_i)|\mathcal{G}_\infty\right](\omega)\end{aligned}$$

for almost all $\omega \in \Omega$ for all $u_i \in B_{J_1^i,J_2^i}(\omega)$ and $i \in \{1,2\}$, as we have already seen in the second claim in the proof of Lemma 5.4. We conclude

$$\mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u_1)|\mathcal{G}_n\right](\omega)\,\mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}(u_2)|\mathcal{G}_n\right](\omega)$$

$$\xrightarrow[n\to\infty]{}\mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}\left(u_1\right)|\mathcal{G}_\infty\right](\omega)\,\mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}\left(u_2\right)|\mathcal{G}_\infty\right](\omega).$$

Similarly, we have again by the Portmanteau Theorem

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}\left(u_1\right)\mathbf{1}_{\Gamma(\xi)}\left(u_2\right)|\mathcal{G}_n\right](\omega)\\
=&P\left(\xi\left(u_1\right)\geqslant T,\xi\left(u_2\right)\geqslant T|\mathcal{G}_n\right)(\omega)\\
=&P\left(\xi\left(u_1\right)_{(J_0^1)^c}\geqslant t_{(J_1^1)^c},\xi\left(u_2\right)_{(J_1^2)^c}\geqslant t_{(J_1^2)^c}|\mathcal{G}_n\right)(\omega)\prod_{\substack{j\in J_1^i\setminus J_2^i\\ i=1,2}}\mathbf{1}_{m(u_i)_j\geqslant t_j}\\
\xrightarrow[n\to\infty]{}&P\left(\xi\left(u_1\right)_{(J_1^1)^c}\geqslant t_{(J_1^1)^c},\xi\left(u_2\right)_{(J_1^2)^c}\geqslant t_{(J_1^2)^c}|\mathcal{G}_\infty\right)(\omega)\prod_{\substack{j\in J_1^i\setminus J_2^i\\ i=1,2}}\mathbf{1}_{m(u_i)_j\geqslant t_j}\\
=&P\left(\xi\left(u_1\right)\geqslant T,\xi\left(u_2\right)\geqslant T|\mathcal{G}_\infty\right)(\omega)\\
=&\mathbb{E}\left[\mathbf{1}_{\Gamma(\xi)}\left(u_1\right)\mathbf{1}_{\Gamma(\xi)}\left(u_2\right)|\mathcal{G}_\infty\right](\omega),
\end{aligned}$$

since $\left(\xi\left(u_1\right)^\top,\xi\left(u_2\right)^\top\right)^\top$ is a Gaussian vector and

$$\begin{aligned}
&P\left(\bigcup_{i=1,2}\left\{\xi\left(u_i\right)_{(J_1^i)^c}\in\partial\mathbf{T}_{(J_1^i)^c}\right\}|\mathcal{G}_\infty\right)(\omega)\\
\leqslant&\sum_{i=1,2}P\left(\xi\left(u_i\right)_{(J_1^i)^c}\in\partial\mathbf{T}_{(J_1^i)^c}|\mathcal{G}_\infty\right)(\omega)\\
=&0.
\end{aligned}$$

Combining both statements yields for almost all $\omega\in\Omega$ and all $u_i\in B_{J_1^i,J_2^i}(\omega)$, $i=1,2$, that

$$\operatorname{Cov}\left(\mathbf{1}_{\Gamma(\xi)}\left(u_1\right),\mathbf{1}_{\Gamma(\xi)}\left(u_2\right)|\mathcal{G}_n\right)(\omega)\xrightarrow[n\to\infty]{}\operatorname{Cov}\left(\mathbf{1}_{\Gamma(\xi)}\left(u_1\right),\mathbf{1}_{\Gamma(\xi)}\left(u_2\right)|\mathcal{G}_\infty\right)(\omega).$$

□

Proof. **(Lemma 5.10)**
Assume without loss of generality that all random elements U,V and W are centered. By Fernique's Theorem

$$\mathbb{E}\left[\exp\left(c\left\|V\right\|\right)\right]=\int_{\mathbb{R}^d}\exp\left(c\left\|x\right\|\right)P^V\left(dx\right)<\infty$$

for some $c>0$ and hence by Theorem 3.2.18 in [7] the space of polynomials Π^d is dense in the space $L^2\left(\mathbb{R}^d,\mathcal{B}\left(\mathbb{R}^d\right),dP^V\right)$. The polynomials in Π^d are here given by

$$p\left(x\right)=\sum_{\alpha\in A}c_\alpha x^\alpha,$$

where A is some (finite) subset of $\mathbb{N}_0^d$ and the product $x^\alpha=x_1^{\alpha_1}...x_d^{\alpha_d}$ is called monomial in the variables $x_1,...,x_d$ for $\alpha\in\mathbb{N}_0^d$ and $c_\alpha\in\mathbb{R}$. The integer

$|\alpha| = \alpha_1 + ... + \alpha_d$ is called total degree of the monomial x^α. This means every $f \in L^2\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right), dP^V\right)$ can be approximated by a sequence of polynomials $\left(x \mapsto \sum_{\alpha \in A_n} c_\alpha x^\alpha\right)_{n \in \mathbb{N}}$ in Π^d, i.e.

$$\int_{\mathbb{R}^d} \left(f(x) - \sum_{\alpha \in A_n} c_\alpha x^\alpha\right)^2 P^V(dx) \xrightarrow[n\to\infty]{} 0.$$

By the factorization Lemma every element in $Z \in L^2(\Omega, \sigma(V), P)$ can be written as $f(V)$ for some measurable function $f : \mathbb{R}^d \to \mathbb{R}$ with

$$\mathbb{E}\left[f(V)^2\right] = \int_{\mathbb{R}^d} f(x)^2 P^V(dx) < \infty.$$

Hence the above statement yields the existence of a polynomial sequence such that

$$\mathbb{E}\left[\left(Z - \sum_{\alpha \in A_n} c_\alpha V^\alpha\right)^2\right] = \mathbb{E}\left[\left(f(V) - \sum_{\alpha \in A_n} c_\alpha V^\alpha\right)^2\right] \xrightarrow[n\to\infty]{} 0.$$

Assume now that U is not orthogonal to $L^2(\Omega, \sigma(V), P)$. Then there exists an element $Z \in L^2(\Omega, \sigma(V), P)$ such that $\mathbb{E}[UZ] \neq 0$ and since

$$\left|\mathbb{E}[UZ] - \mathbb{E}\left[U \sum_{\alpha \in A_n} c_\alpha V^\alpha\right]\right| \leqslant \mathbb{E}\left[U^2\right]^{\frac{1}{2}} \underbrace{\mathbb{E}\left[\left(Z - \sum_{\alpha \in A_n} c_\alpha V^\alpha\right)^2\right]^{\frac{1}{2}}}_{\xrightarrow[n\to\infty]{} 0}$$

by Cauchy-Schwarz there must exist $N \in \mathbb{N}$ such that for all $n \geqslant N$ we have $\mathbb{E}\left[U \sum_{\alpha \in A_n} c_\alpha V^\alpha\right] \neq 0$ and by the linearity of the expectation there must hence be a multi-index α with total degree $|\alpha| = k_0$ for some $k_0 \in \mathbb{N}_0$ such that $\mathbb{E}[UV^\alpha] \neq 0$.

This means the set

$$I_{k_0} := \left\{\alpha \in \mathbb{N}_0^d : \mathbb{E}[UV^\alpha] \neq 0, |\alpha| = k_0\right\}$$

is non-empty. By reducing k_0 by one in each step, we can calculate the sets I_k for every $0 \leqslant k \leqslant k_0$ and define by k_* the smallest k such that I_k is non-empty. Then we have for every multi-index $\lambda \in I_{k_*}$

$$\mathbb{E}\left[UV^\lambda\right] = \mathbb{E}\left[U \prod_{i=1}^d V_i^{\lambda_i}\right] \neq 0$$

and

$$\mathbb{E}\left[U \prod_{i=1}^d V_i^{k_i}\right] = 0,$$

if $k_i \leqslant \lambda_i$ for all $i \in \{1, ..., n\}$ with at least one strict inequality. Indeed, if this would not be the case we would have a contradiction with the minimality of k_*.

Using the independence, we conclude

$$\begin{aligned}
\mathbb{E}\left[U \prod_{i=1}^{d} (V+W)_i^{\lambda_i}\right] &= \mathbb{E}\left[U \prod_{i=1}^{d} \left(\sum_{k=0}^{\lambda_i} \binom{\lambda_i}{k} V_i^k W_i^{\lambda_i - k}\right)\right] \\
&= \mathbb{E}\left[U \sum_{k_1,\ldots,k_d=0}^{\lambda_1,\ldots,\lambda_d} \prod_{i=1}^{d} \binom{\lambda_i}{k_i} V_i^{k_i} W_i^{\lambda_i - k_i}\right] \\
&= \sum_{k_1,\ldots,k_d=0}^{\lambda_1,\ldots,\lambda_d} \prod_{i=1}^{d} \binom{\lambda_i}{k_i} \mathbb{E}\left[U \prod_{i=1}^{d} V_i^{k_i}\right] \mathbb{E}\left[\prod_{i=1}^{d} W_i^{\lambda_i - k_i}\right] \\
&= \mathbb{E}\left[U \prod_{i=1}^{d} V_i^{\lambda_i}\right] \\
&\neq 0
\end{aligned}$$

and hence U can not be orthogonal to $L^2\left(\Omega, \sigma\left(V+W\right), P\right)$. □

Acknowledgments

The authors would like to thank two anonymous referees and an Associate Editor for their constructive comments that improved the quality of this paper.

Funding

The work was supported by the Institute of Mathematical Statistics and Actuarial Science of the University of Bern.

References

[1] Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3 ed. Springer, Berlin. MR1321140

[2] Álvarez, M. A., Rosasco, L. and Lawrence, N. D. (2012). Kernels for Vector-Valued Functions: A Review. *Foundations and Trends® in Machine Learning* **4** 195 – 266.

[3] Bect, J., Bachoc, F. and Ginsbourger, D. (2019). A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli* **25** 2883 – 2919. MR4003568

[4] Bogachev, V. I. (2007). *Measure Theory* **1**. Springer, Berlin.

[5] Bogachev, V. I. (1998). *Gaussian Measures. Mathematical Surveys and Monographs*, **62** American Mathematical Society, Providence, RI. MR1642391

[6] Cressie, N. A. C. (1993). *Statistics for Spatial Data. Wiley Series in Probability and Statistics*. John Wiley & Sons.

[7] Dunkl, C. F. and Xu, Y. (2014). *Orthogonal Polynomials of Several Variables*, 2 ed. *Encyclopedia of Mathematics and its Applications*, **155**. Cambridge University Press, Cambridge. MR3289583

[8] Evgeniou, T., Micchelli, C. A. and Pontil, M. (2005). Learning Multiple Tasks with Kernel Methods. *Journal of Machine Learning Research* **6** 615 – 637. MR2249833

[9] Fossum, T. O., Travelletti, C., Eidsvik, J., Ginsbourger, D. and Rajan, K. (2021). Learning excursion sets of vector-valued Gaussian random fields for autonomous ocean sampling. *The Annals of Applied Statistics* **15** 597 – 618. MR4298973

[10] Ibragimov, I. A. and Rozanov, I. A. (1978). *Gaussian random processes. Applications of Mathematics*, **9**. Springer, New York, NY. MR0543837

[11] Kallenberg, O. (2021). *Foundations of Modern Probability*, 3 ed. *Probability Theory and Stochastic Modelling*, **99**. Springer, Cham. MR4226142

[12] Kechris, A. S. (1995). *Classical descriptive set theory. Graduate Texts in Mathematics*, **156**. Springer, New York, NY. MR1321597

[13] Molchanov, I. S. (2017). *Theory of Random Sets*, 2 ed. *Probability Theory and Stochastic Modelling*, **87**. Springer, London. MR3751326

[14] Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* **51** 406 – 413. MR0069793

[15] Rajput, B. S. and Cambanis, S. (1972). Gaussian Processes and Gaussian Measures. *Annals of Mathematical Statistics* **43** 1944 – 1952. MR0370728

[16] Rakočević, V. (1997). On continuity of the Moore-Penrose and Drazin inverses. *Matematichki Vesnik* **49** 163 – 172. MR1611717

[17] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. The MIT Press, Cambridge, MA. MR2514435

[18] Schönfeld, P. (1973). A note on the measurability of the pseudo-inverse. *Journal of Econometrics* **1** 313 – 314.

[19] Simon, B. (2011). *Convexity: An Analytic Viewpoint. Cambridge Tracts in Mathematics*, **187**. Cambridge University Press, Cambridge. MR2814377

[20] Stewart, G. W. (1969). On the Continuity of the Generalized Inverse. *Siam Journal on Applied Mathematics* **17** 33 – 45. MR0245583

[21] Stroock, D. W. (2011). *Probability Theory: An Analytic View*, 2 ed. Cambridge University Press, Cambridge. MR2760872

[22] Tao, T. (2012). *Topics in Random Matrix Theory. Graduate Studies in Mathematics*, **132**. American Mathematical Society, Providence, RI. MR2906465

[23] Vakhania, N. N., Tarieladze, V. I. and Chobanyan, S. A. (1987). *Probability Distributions on Banach Spaces. Mathematics and its Applications (Soviet Series)*, **14**. Springer Netherlands, Dordrecht. MR1435288

[24] van der Vaart, A. W. and van Zanten, J. H. (2008). *Reproducing kernel Hilbert spaces of Gaussian priors* In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh* 200 – 222. Institute of Mathematical Statistics. MR2459226

[25] van der Wilk, M., Dutordoir, V., John, S. T., Artemev, A., Adam, V. and Hensman, J. (2020). A Framework for Interdomain and Multioutput Gaussian Processes. *ArXiv arXiv:2003.01115*.

[26] Zhang, F. (2005). *The Schur complement and its applications. Numerical Methods and Algorithms*, **4**. Springer US, Boston, MA. MR2160825