

A supermartingale approach to Gaussian process based sequential design of experiments

Julien Bect, François Bachoc, David Ginsbourger

▶ **To cite this version:**

Julien Bect, François Bachoc, David Ginsbourger. A supermartingale approach to Gaussian process based sequential design of experiments. 2017. hal-01351088v2

HAL Id: hal-01351088

https://hal.science/hal-01351088v2

Preprint submitted on 26 Jul 2017 (v2), last revised 13 Sep 2018 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

arXiv: arXiv:1608.01118v2

A supermartingale approach to Gaussian process based sequential design of experiments

Julien Bect[*], **François Bachoc**[†] **and David Ginsbourger**[‡,§]

[*] *Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec, CNRS, Univ. Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France.*
e-mail: julien.bect@centralesupelec.fr

[†] *Toulouse Mathematics Institute, University Paul Sabatier, France.*
e-mail: francois.bachoc@math.univ-toulouse.fr

[‡] *Uncertainty Quantification and Optimal Design group, Idiap Research Institute, Martigny, Switzerland*
e-mail: ginsbourger@idiap.ch

[§] *Institute of Mathematical Statistics and Actuarial Science, Department of Mathematics and Statistics, University of Bern, Switzerland*
e-mail: ginsbourger@stat.unibe.ch

Abstract: Gaussian process (GP) models have become a well-established framework for the adaptive design of costly experiments, and notably of computer experiments. GP-based sequential designs have been found practically efficient for various objectives, such as global optimization (estimating the global maximum or maximizer(s) of a function), reliability analysis (estimating a probability of failure) or the estimation of level sets and excursion sets. In this paper, we deal with convergence properties of an important class of sequential design approaches, known as stepwise uncertainty reduction (SUR) strategies. Our approach relies on the key observation that the sequence of residual uncertainty measures, in SUR strategies, is generally a supermartingale with respect to the filtration generated by the observations. We study the existence of SUR strategies and establish generic convergence results for a broad class thereof. We also introduce a special class of uncertainty measures defined in terms of regular loss functions, which makes it easier to check that our convergence results apply in particular cases. Applications of the latter include proofs of convergence for the two main SUR strategies proposed by Bect, Ginsbourger, Li, Picheny and Vazquez (Stat. Comp., 2012). To the best of our knowledge, these are the first convergence proofs for GP-based sequential design algorithms dedicated to the estimation of excursions sets and their measure. Coming to global optimization algorithms, we also show that the knowledge gradient strategy can be cast in the SUR framework with an uncertainty functional stemming from a regular loss, resulting in further convergence results. We finally establish a new proof of convergence for the expected improvement algorithm, which is the first proof for this algorithm that applies to any GP with continuous sample paths.

MSC 2010 subject classifications: Primary 60G15, 62L05; secondary 68T05.
Keywords and phrases: Sequential Design of Experiments, Active Learning, Stepwise Uncertainty Reduction, Supermartingale, Uncertainty functional, Convergence.

1. Introduction

Sequential design of experiments is an important and lively research field at a crossroads between applied probability, statistics and optimization, where the goal is to allocate experimental resources step by step so as to reduce the uncertainty about some quantity, or function, of interest. While the experimental design vocabulary traditionally refers to observations of natural phenomena presenting aleatory uncertainties, the design of computer experiments—in which observations are replaced by

numerical simulations—has become a field of research *per se* [29, 41, 42] where Gaussian process models are massively used to define efficient sequential strategies in cases of costly evaluations. The predominance of Gaussian processes in this field is probably due to their unique combination of modeling flexibility and computational tractability, which makes it possible to work out sampling criteria accounting for the potential effect of adding new experiments. Defining, calculating and optimizing sampling criteria for various application goals have inspired a significant number of research contributions in the last decades [see, e.g., 3, 9–11, 15, 17–19, 21, 22, 37–39, 44, 45, 53]. Yet, available convergence results for the associated design strategies are quite heterogeneous in terms of their respective extent and underlying hypotheses [8, 23, 44, 46, 51]. Here we develop a probabilistic approach to the analysis of a large class of strategies. This enables us to establish generic convergence results whose broad applicability is subsequently illustrated on four popular sequential design strategies. The crux is that each of these strategies turns out to involve some *uncertainty functional* applied to a sequence of conditional probability distributions, and our main results rely on an associated supermartingale property.

Among the sampling criteria considered in our examples, probably the most famous one is the expected improvement (EI), that arose in sequential design for global optimization. Following the foundations laid by Mockus et al. [33] and the considerable impact of the work of Jones et al. [27], EI and other Bayesian optimization strategies have spread in a variety of application fields. They are now commonly used in engineering design [16] and, in the field of machine learning, for automatic configuration algorithms [see 45, and references therein]. Extensions to constrained, multi-objective and/or robust optimization constitute an active field of research [see, e.g., 5, 14, 15, 22, 37, 55]. In a different context, sequential design strategies based on Gaussian process models have been used to estimate contour lines, probabilities of failures, profile optima and excursion sets of expensive to evaluate simulators [see, notably, 3, 10, 21, 38, 39, 50, 54, 57].

The class of sequential design strategies that we consider here are built according to the *stepwise uncertainty reduction* (SUR) paradigm [see 3, 9, 53, and references therein]. Our main focus is on the convergence of these algorithms under the assumption that the function of interest is a sample path of the Gaussian process prior that is used to construct the sequential design. Almost sure consistency has been proven for the EI algorithm in [51], but only under the restrictive assumption that the covariance function satisfies a certain condition—the "No Empty Ball" (NEB) property—which excludes very regular Gaussian processes[1]. Moreover, to the authors' knowledge no proof of convergence has yet been established for algorithms dedicated to probability of excursion and/or excursion set estimation (referred to as *excursion case* henceforth) such as those of Bect et al. [3].

The proof scheme developed here, that relies notably on a supermartingale property defined for uncertainty functionals, allows us to address the excursion case and also to revisit the convergence of the knowledge gradient algorithm [17–19], as well as that of the EI algorithm —that was recently cast as a particular case of a SUR strategy; see [9]—without requiring the NEB assumption. Before outlining the paper in more detail, let us briefly introduce its general setting and, in particular, what we mean by SUR strategies. We will focus directly on the case of Gaussian processes for clarity, but the SUR principle in itself is much more general, and can be used with other types of

[1] On a related note, Bull [8] proves an upper-bound for the convergence rate of the expected improvement algorithm under the assumption that the covariance functon is Hölder, but his result only holds for functions that belong the the reproducing kernel Hilbert space (RKHS) of the covariance—a condition that, under appropriate assumptions, is almost surely not statisfied by sample paths of the Gaussian process according to Driscoll's theorem [31]. Another result in the same vein is provided by Yarotsky [56] for the squared exponential covariance in the univariate case, assuming the objective function is analytic in a sufficiently large complex domain around its interval of definition.

models [see, e.g., 11, 20, 32].

Let ξ be a real-valued Gaussian process defined on a measurable space $\mathbb{X}$—typically, ξ will be a continuous Gaussian process on a compact metric space, such as $\mathbb{X} = [0,1]^\ell$— and assume that evaluations $Z_n = \xi(X_n) + \epsilon_n$ are to be made, sequentially, in order to gather information about some quantity of interest. We will assume the sequence of observation errors $(\epsilon_n)_{n\in\mathbb{N}^*}$ to be independent of the Gaussian process ξ, and composed of independent centred Gaussian variables. The definition of a SUR strategy starts with the choice of a "measure of residual uncertainty" for the quantity of interest after n evaluations, which is a functional

$$H_n = \mathcal{H}\left(\mathrm{P}_n^\xi\right) \tag{1.1}$$

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 a given prior distribution P_0^ξ, assume that the H_n's are $\mathcal{F}_n$-measurable random variables. SUR sampling criteria are then defined as

$$J_n(x) \;=\; \mathrm{E}_{n,x}\left(H_{n+1}\right), \tag{1.2}$$

where $\mathrm{E}_{n,x}$ denotes the conditional expectation with respect to $\mathcal{F}_n$ with $X_{n+1} = x$ (assuming that H_{n+1} is integrable, for any choice of $x \in \mathbb{X}$). The value of the sampling criterion $J_n(x)$ at time n measures the expected residual uncertainty at time $n+1$ if the next evaluation is made at x. Finally, a (non-randomized) sequential design is constructed by choosing at each step the point that provides the smallest expected residual uncertainty—or, equivalently, the largest expected uncertainty reduction—, that is,

$$X_{n+1} \in \operatorname{argmin}_{x\in\mathbb{X}} J_n(x). \tag{1.3}$$

Given a finite measure μ over $\mathbb{X}$ and an excursion threshold $T \in \mathbb{R}$, a typical choice of measure of residual uncertainty in the excursion case [3] is the integrated indicator variance $H_n = \mathcal{H}\left(\mathrm{P}_n^\xi\right) = \int_{\mathbb{X}} p_n\,(1-p_n)\,d\mu$ (also called *integrated Bernoulli variance* in what follows) where $p_n(u) = \mathrm{P}_n\left(\xi(u) \geq T\right)$ and P_n denotes the conditional probability with respect to $\mathcal{F}_n$. Another related measure of uncertainty, for which a semi-analytical formula is provided in [10], is the *variance of the excursion volume*, $H_n = \operatorname{var}(\mu(\{u \in \mathbb{X} : \xi(u) \geq T\}))$. In the optimization case on the other hand, it turns out that the EI criterion is underlaid by the following measure of residual uncertainty [see, e.g., 9, Section 3.3]: $H_n = \mathrm{E}_n\left(\max\xi - M_n\right)$ where $M_n = \max_{i\leq n}\xi(X_i)$ and E_n refers to the conditional expectation with respect to $\mathcal{F}_n$. A similar construct can be obtained for the *knowledge gradient*, as developed later in the paper. It appears in all four cases that the associated measures of residual uncertainty possess the aforementioned supermartingale property, allowing us to establish in this paper convergence results for the associated strategies.

The paper is structured as follows. In Section 2 we present the class of GP models of interest and the sequential design context. In particular, we review some fundamental results on links between GP models with continuous sample paths and Gaussian measures on the space of continuous functions, and we address properties of conditioning and convergence of Gaussian measures that are instrumental for proving the main results of the paper. In Section 3, we introduce uncertainty functionals along with related concepts and results. Next, we define SUR and quasi-SUR strategies and give sufficient conditions of existence depending on their underlying uncertainty functional. At the heart of the section, we then establish general convergence results for such strategies. In particular, special care is devoted to a class of uncertainty functionals that are defined in terms

of the minimization of average loss functions. Ultimately, the convergence results are revisited in the introduced framework of *regular loss functions*, leading to a simplified treatment for a number of SUR strategies including three out of the four examples treated next. Section 4 details how the previous results apply in the excursion (*integrated Bernoulli variance* and *variance of excursion volume*) and Bayesian optimization (knowledge gradient and EI) cases, establishing both convergence to zero for the considered measure of residual uncertainty and convergence of the corresponding estimator to the quantity of interest, in the almost sure and L^1 sense.

2. Preliminaries: Gaussian process priors and sequential designs

2.1. Model

Let $(\xi(x))_{x\in\mathbb{X}}$ denote a Gaussian process with mean function m and covariance function k, defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and indexed by a metric space $\mathbb{X}$. Assume that ξ can be observed at sequentially selected (data-dependent) design points X_1, X_2, ..., with additive heteroscedatic Gaussian noise:

$$Z_n = \xi(X_n) + \tau(X_n)\, U_n, \qquad n = 1, 2, \ldots \tag{2.1}$$

where $\tau : \mathbb{X} \to \mathbb{R}_+$ gives the (known) standard deviation $\tau(x)$ of an observation made at the design point $x \in \mathbb{X}$, and $(U_i)_{i\geq 1}$ denotes a sequence of independent and identically distributed $\mathcal{N}(0,1)$ variables, independent from ξ. Let $\mathcal{F}_n$ denote the σ-algebra generated by $X_1, Z_1, \ldots, X_n, Z_n$.

Definition 2.1. *A sequence $(X_n)_{n\geq 1}$ will be said to form a (non-randomized) sequential design if, for all $n \geq 1$, X_n is $\mathcal{F}_{n-1}$-measurable.*

Standing assumptions 2.2. *We will assume in the rest of the paper that*

i) $\mathbb{X}$ *is a compact metric space,*
ii) ξ *has continuous sample paths,*
iii) $\tau : \mathbb{X} \to \mathbb{R}_+$ *is continous.*

Remark 2.3. Note that that the variance function τ^2 is not assumed to be strictly positive. Indeed, the special case where $\tau^2 \equiv 0$ is actually an important model to consider given its widespread use in Bayesian numerical analysis [see, e.g. 13, 24, 35, 40] and in the design and analysis of deterministic computer experiments [see, e.g., 2, 41, 42].

Remark 2.4. The setting described in this section arises, notably, when considering from a Bayesian point a view the following non-parametric interpolation/regression model with heteroscedastic Gaussian noise:

$$Z_n = f(X_n) + \tau(X_n)\, U_n, \qquad n = 1, 2, \ldots \tag{2.2}$$

with a continuous Gaussian process prior on the unknown regression function f. In this case, m and k are the *prior* mean and covariance functions of ξ.

2.2. Gaussian random elements and Gaussian measures on $\mathcal{C}(\mathbb{X})$

Let $\mathbb{S} = \mathcal{C}(\mathbb{X})$ denote the set of all continuous functions on $\mathbb{X}$. Since $\mathbb{X}$ is assumed compact, $\mathbb{S}$ becomes a separable Banach space when equipped with the supremum norm $\|\cdot\|_\infty$. We recall [see, e.g., 1, Theorem 2.9] that any Gaussian process $(\xi(x))_{x\in\mathbb{X}}$ with continuous sample paths on a compact metric space satisfies $\mathrm{E}(\|\xi\|_\infty) < \infty$.

Any Gaussian process $(\xi(x))_{x\in\mathbb{X}}$ with continuous sample paths can be seen as a Gaussian random element in $\mathbb{S}$. More precisely, the mapping $\xi : \Omega \to \mathbb{S}$, $\omega \mapsto \xi(\omega, \cdot)$ is $\mathcal{F}/\mathcal{S}$-measurable, where $\mathcal{S}$ denotes the Borel σ-algebra on $\mathbb{S}$, and the probability distribution P^ξ of ξ is a Gaussian measure on $\mathbb{S}$. A proof of these facts is provided in the Appendix A.1. The reader is referred to Vakhania et al. [48] and Ledoux and Talagrand [30] for background information concerning random elements and measures in Banach spaces, and to van der Vaart et al. [49] and Bogachev [6] for more information on the case of Gaussian random elements and measures.

We will denote by $\mathbb{M}$ the set of all Gaussian measures on $\mathbb{S}$. Any $\nu \in \mathbb{M}$ is the probability distribution of some Gaussian process with continuous sample paths, seen as a random element in $\mathbb{X}$. The mean function m_ν and covariance function k_ν of this Gaussian process are continous (see, e.g., Lemma 1 in [26]) and fully characterize the measure, which we will denote as $\mathcal{GP}(m_\nu, k_\nu)$. We endow $\mathbb{M}$ with the σ-algebra $\mathcal{M}$ generated by the evaluation maps $\pi_A : \nu \mapsto \nu(A)$, $A \in \mathcal{S}$. Using this σ-algebra, conditional distributions on $\mathbb{S}$—i.e., transition kernels—can be conveniently identified to random elements in $\mathbb{M}$ [see, e.g. 28, p. 105–106].

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, i.e., the set of all random elements $\boldsymbol{\nu}$ in $(\mathbb{M}, \mathcal{M})$ such that $\boldsymbol{\nu} = \mathrm{P}\big(\xi \in \cdot \mid \mathcal{F}'\big)$ for some σ-algebra $\mathcal{F}' \subset \mathcal{F}$.

2.3. Conditioning on finitely many observations

It is well known that Gaussian processes remain Gaussian under conditioning with respect to pointwise evaluations, or more generally linear combination of pointwise evaluations, possibly corrupted by independent additive Gaussian noise. In the language of nonparametric Bayesian statistics (see Remark 2.4), Gaussian process priors are *conjugate* with respect to this sampling model. The following result formalizes this fact in the framework of Gaussian measures on $\mathbb{S}$, and states that the conjugation property still holds when the observations are made according to a sequential design.

Proposition 2.5. *For all $n \geq 1$, there exists a measurable mapping*

$$\begin{aligned} (\mathbb{X}\times\mathbb{R})^n \times \mathbb{M} &\to \mathbb{M}, \\ (x_1, z_1, \ldots, x_n, z_n, \nu) &\mapsto \mathrm{Cond}_{x_1,z_1,\ldots,x_n,z_n}(\nu), \end{aligned} \tag{2.3}$$

such that, for any sequential design $(X_n)_{n\geq 1}$, $\mathrm{Cond}_{X_1,Z_1,\ldots,X_n,Z_n}(\mathrm{P}^\xi)$ is a conditional distribution of ξ given $\mathcal{F}_n$.

A proof of this result is provided in Appendix A.3. In the rest of the paper, we will denote by $\mathrm{P}^\xi_n = \mathcal{GP}(m_n, k_n)$ the conditional distribution $\mathrm{Cond}_{X_1,Z_1,\ldots,X_n,Z_n}(\mathrm{P}^\xi)$ of ξ given $\mathcal{F}_n$, which can be seen as a random element in $(\mathbb{M}, \mathcal{M})$. Note that m_n (respectively k_n) is an $\mathcal{F}_n$-measurable process[2] on $\mathbb{X}$ (respectively $\mathbb{X}\times\mathbb{X}$), with continuous sample paths. Note also that $m_0 = m$ and $k_0 = k$. Conditionally to $\mathcal{F}_n$, the next observation follows a normal distribution:

$$Z_{n+1} \mid \mathcal{F}_n \;\sim\; \mathcal{N}\left(m_n(X_{n+1}),\, s_n^2(X_{n+1})\right), \tag{2.4}$$

where $s_n^2(x) = k_n(x,x) + \tau^2(x)$.

[2] i.e., a measurable process when considered as defined on $(\Omega, \mathcal{F}_n)$ instead of $(\Omega, \mathcal{F})$

2.4. Convergence in $\mathbb{M}$

We consider in this paper the following notion of convergence on $\mathbb{M}$:

Definition 2.6. *Let $\nu_n = \mathcal{GP}(m_n, k_n) \in \mathbb{M}$, $n \in \mathbb{N} \cup \{+\infty\}$. We will say that (ν_n) converges to ν_∞, and write $\nu_n \to \nu_\infty$, if $m_n \to m_\infty$ uniformly on $\mathbb{X}$ (i.e., $m_n \to m_\infty$ in $\mathbb{S}$) and $k_n \to k_\infty$ uniformly on $\mathbb{X} \times \mathbb{X}$.*

Remark 2.7. In other words, we consider the topology on $\mathbb{M}$ induced by the strong topology on the Banach space $\mathcal{C}(\mathbb{X}) \times \mathcal{C}(\mathbb{X} \times \mathbb{X})$, where $\mathbb{M}$ is identified to a subset Θ of this space through the injection $\nu \mapsto (m_\nu, k_\nu)$.

Let us now state two important convergence results in this topology, that will be needed in Section 3. In the first of them, and later in the paper, we denote by $\mathcal{F}_\infty = \bigvee_{n\geq 1} \mathcal{F}_n$ the σ-algebra generated by $\bigcup_{n\geq 1} \mathcal{F}_n$.

Proposition 2.8. *For any sequential design, 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 $\mathrm{P}^\xi_n \to \mathrm{P}^\xi_\infty$ almost surely.*

Proposition 2.9. *Let $\nu \in \mathbb{M}$ and let $(x_j, z_j) \to (x, z)$ in $\mathbb{X} \times \mathbb{R}$. Assume that $k_\nu(x, x) + \tau^2(x) > 0$. Then $\mathrm{Cond}_{x_j,\, z_j}(\nu) \to \mathrm{Cond}_{x,\, z}(\nu)$.*

Proof. See Appendix A.4 for proofs of both results. □

3. Stepwise Uncertainty Reduction

3.1. Uncertainty functionals and uncertainty reduction

As explained in the introduction, the definition of a SUR strategy starts with the choice of an uncertainty functional $\mathcal{H}$, which maps the posterior distribution of ξ to a measure of residual uncertainty for the quantity of interest.

More formally, let $\mathcal{H} : \mathbb{M} \to [0, +\infty)$ denote a measurable map[3]. Since P^ξ_n is an $\mathcal{F}_n$-measurable random element in $(\mathbb{M}, \mathcal{M})$, the residual uncertainty $H_n = \mathcal{H}(\mathrm{P}^\xi_n)$ is an $\mathcal{F}_n$-measurable random variable. The SUR sampling criterion introduced informally as $J_n(x) = \mathrm{E}_{n,x}(H_{n+1})$ in Equation (1.2) can be more precisely defined as

$$J_n(x) = \mathrm{E}_n\left(\mathcal{H}\left(\mathrm{Cond}_{x,\, Z_{n+1}(x)}\left(\mathrm{P}^\xi_n\right)\right)\right), \tag{3.1}$$

where $Z_{n+1}(x) = \xi(x) + U_{n+1}\tau(x)$. Then, observe that $J_n(x) = \mathcal{J}_x(\mathrm{P}^\xi_n)$, where the functional $\mathcal{J}_x : \mathbb{M} \to [0, +\infty]$ is defined for all $x \in \mathbb{X}$ and $\nu \in \mathbb{M}$ by

$$\mathcal{J}_x(\nu) = \iint_{\mathbb{S}\times\mathbb{R}} \mathcal{H}\left(\mathrm{Cond}_{x,\, f(x)+u\,\tau(x)}(\nu)\right)\, \nu(\mathrm{d}f)\, \phi(u)\, \mathrm{d}u \tag{3.2}$$

$$= \int_{\mathbb{R}} \mathcal{H}\left(\mathrm{Cond}_{x,\, m_\nu(x)+v\, s_\nu(x)}(\nu)\right)\, \phi(v)\, \mathrm{d}v, \tag{3.3}$$

with $s^2_\nu(x) = k_\nu(x, x) + \tau^2(x)$ and ϕ denotes the probability density function of the standard normal distribution. The mapping $(x, \nu) \mapsto \mathcal{J}_x(\nu)$ is $\mathcal{B}(\mathbb{X}) \otimes \mathcal{M}$-measurable (see Proposition A.8

[3] Only non-negative uncertainty functionals are considered in this paper, but our results could easily be adapted, with appropriate integrability conditions, to the case of signed uncertainty functionals.

in the appendix). As a consequence, J_n is an $\mathcal{F}_n$-measurable process for all n and thus, for any sequential design, $J_n(X_n)$ is a well-defined $\mathcal{F}_n$-measurable random variable.

A key observation for the convergence results of this paper is that many uncertainty functionals of interest—examples of which will be given in Section 4—enjoy the following property:

Definition 3.1. *A measurable functional $\mathcal{H}$ on $\mathbb{M}$ will be said to have the* supermartingale property *if, for any $\mathrm{P}_0^\xi \in \mathbb{M}$ and any sequential design $(X_n)_{n\geq 1}$, the sequence $(H_n)_{n\geq 1}$, with $H_n = \mathcal{H}(\mathrm{P}_n^\xi)$, is an $(\mathcal{F}_n)$-supermartingale.*

It is straightforward to prove the following characterization of this property:

Proposition 3.2. *$\mathcal{H}$ has the supermartingale property if, and only if, $\mathcal{J}_x(\nu) \leq \mathcal{H}(\nu)$ for all $x \in \mathbb{X}$ and $\nu \in \mathbb{M}$.*

The supermartingale property echoes with DeGroot's observation that "reasonable" measures of uncertainty should be decreasing on average for any possible experiment [12]. To discuss this connexion more precisely in our particular setting, let us consider the following definition.

Definition 3.3. *Let $\mathbb{M}_0$ denote a set of probability measures on $\mathbb{S}$. Let $\mathcal{M}_0$ denote the σ-algebra generated on $\mathbb{M}_0$ by the evaluation maps. For any random element $\boldsymbol{\nu}$ in $(\mathbb{M}_0, \mathcal{M}_0)$, let $\overline{\boldsymbol{\nu}}$ denote the probability measure defined by $\overline{\boldsymbol{\nu}}(A) = \mathrm{E}\,(\boldsymbol{\nu}(A))$, $A \in \mathcal{S}$. We will say that a measurable function $\mathcal{H}$ on $\mathbb{M}_0$ is* decreasing on average *(DoA) if, for any random element $\boldsymbol{\nu}$ in $(\mathbb{M}_0, \mathcal{M}_0)$ such that $\overline{\boldsymbol{\nu}} \in \mathbb{M}_0$, $\mathrm{E}\,(\mathcal{H}(\boldsymbol{\nu})) \leq \mathcal{H}(\overline{\boldsymbol{\nu}})$.*

Note that, if the set $\mathbb{M}_0$ is convex, DoA functionals on $\mathbb{M}_0$ are concave. The converse statement is expected to be false, however, since Jensen's inequality does not hold for all concave functionals in infinite dimensional settings [see 36, for extensions of Jensen's inequality under various assumptions]. The set $\mathbb{M}$ of all Gaussian measures on $\mathbb{S}$ is not convex, but all the uncertainty functionals presented in Section 4 can in fact be extended, if infinite values are allowed, to DoA functionals defined on its convex hull (i.e., on the set of all mixtures of Gaussian measures).

The supermartingale and DoA properties are easily seen to be connected as follows:

Proposition 3.4. *Let $\mathbb{M}_0$ denote a set of probability measures on $\mathbb{S}$. Assume that $\mathbb{M}_0$ is conjugate with respect to the sampling model (cf. Sections 2.1–2.3). Let $\mathcal{H}$ denote a (measurable) DoA functional on $\mathbb{M}_0$. Then $\mathcal{H}$ has the supermartingale property.*

Remark 3.5. The converse would hold for a sampling model that allows conditioning with respect to *any* σ-algebra [see 12, Theorem 2.1, for a proof with "concave" instead of "DoA"].

Let us conclude this section with an additional definition:

Definition 3.6. *A measurable functional $\mathcal{H}$ on $\mathbb{M}$ will be said to be* $\mathfrak{P}$-uniformly integrable *if, for any Gaussian random element ξ in $\mathbb{S}$, the family $(\mathcal{H}(\boldsymbol{\nu}))_{\boldsymbol{\nu}\in\mathfrak{P}(\xi)}$ is uniformly integrable.*

Proposition 3.7. *Let $\mathcal{H}$ denote a measurable functional on $\mathbb{M}$. If there exists $L^+ \in \cap_{\nu\in\mathbb{M}}\mathcal{L}^1\,(\mathbb{S}, \mathcal{S}, \nu)$ such that $|\mathcal{H}(\nu)| \leq \int_{\mathbb{S}} L^+ \mathrm{d}\nu$ for all $\nu \in \mathbb{M}$, then $\mathcal{H}$ is $\mathfrak{P}$-uniformly integrable.*

Proof. Let ξ denote a Gaussian random element in $\mathbb{S}$ and let $\boldsymbol{\nu} = \mathrm{P}\big(\xi \in \cdot \mid \mathcal{F}'\big) \in \mathfrak{P}(\xi)$. We have $|\mathcal{H}(\boldsymbol{\nu})| \leq \mathrm{E}\,(L^+(\xi) \mid \mathcal{F}')$, and the result follows from the uniform integrability of conditional expectations [see, e.g., 28, Lemma 5.5]. □

Remark 3.8. If $\mathcal{H}$ is $\mathfrak{P}$-uniformly integrable and has the supermartingale property then, for any sequential design, the sequence (H_n) is a uniformly integrable supermartingale (since $\{\mathrm{P}_n^\xi\} \subset \mathfrak{P}(\xi)$).

3.2. Existence of SUR and quasi-SUR strategies

A SUR sequential design is built by selecting at each step (possibly after some initial design) the next design point as a minimizer of the SUR sampling criterion J_n. More formally:

Definition 3.9. *We will say that* (X_n) *is a* SUR sequential design *(or* SUR strategy*) associated with the uncertainty functional* $\mathcal{H}$ *if it is a sequential design such that* $X_{n+1} \in \operatorname{argmin} J_n$ *for all* $n \geq n_0$*, for some integer* n_0*. Given a sequence* $\varepsilon = (\varepsilon_n)$ *of positive real numbers such that* $\varepsilon_n \to 0$*, we will say that* (X_n) *is an* ε-quasi-SUR *sequential design (or strategy) if it is a sequential design such that* $J_n(X_{n+1}) \leq \inf J_n + \varepsilon_n$ *for all* $n \geq n_0$*, for some integer* n_0*.*

In order to provide sufficient conditions for the existence of an ε-quasi-SUR strategy associated with a certain uncertainty functional $\mathcal{H}$, we will need to assume a form of continuity of the uncertainty functional on $\mathbb{M}$. Assuming $\mathcal{H}$ to be continuous, however, would be too strong a requirement, that some important examples would fail to satisfy (see Sections 4.1 and 4.2). The following weaker notion of continuity will turn out to be suitable for our needs:

Definition 3.10. *A measurable functional* $\mathcal{H}$ *on* $\mathbb{M}$ *will be said to be* $\mathfrak{P}$-continuous *if, for any Gaussian random element* ξ *in* $\mathbb{S}$ *and any sequence of random elements* $\boldsymbol{\nu}_n \in \mathfrak{P}(\xi)$ *such that* $\boldsymbol{\nu}_n \xrightarrow{a.s.} \boldsymbol{\nu}_\infty \in \mathfrak{P}(\xi)$*, the convergence* $\mathcal{H}(\boldsymbol{\nu}_n) \xrightarrow{a.s.} \mathcal{H}(\boldsymbol{\nu}_\infty)$ *holds.*

Remark 3.11. Note that the definition of "$\mathfrak{P}$-continuous" does not put any restriction on the distribution of the Gaussian random element ξ. In particular, we are not requiring that ξ should be such that any $\boldsymbol{\nu}_\infty \in \mathfrak{P}(\xi)$ belongs almost surely to the set of continuity points of $\mathcal{H}$.

We are now in a position to state a general existence result for ε-quasi-SUR strategies. Recall that $\mathbb{X}$ is assumed, thoughout the paper, to be a compact metric space (see Assumptions 2.2).

Theorem 3.12. *Let* $\mathcal{H}$ *denote a measurable uncertainty functional on* $\mathbb{M}$*. Assume that* $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$*, where* $\mathcal{H}_0(\nu) = \int_{\mathbb{S}} L_0 \,\mathrm{d}\nu$ *for some* $L_0 \in \cap_{\nu \in \mathbb{M}} \mathcal{L}^1(\mathbb{S}, \mathcal{S}, \nu)$*, and* $\mathcal{H}_1$ *is* $\mathfrak{P}$*-uniformly integrable,* $\mathfrak{P}$*-continuous and has the supermartingale property. Then,*

a) for any sequential design, the sample paths of J_n *are continuous on* $\{x \in \mathbb{X} : s_n^2(x) > 0\}$*;*
b) for any sequence $\varepsilon = (\varepsilon_n)$ *of positive real numbers, there exists an* ε*-quasi-SUR sequential design* $(X_n)_{n \geq 1}$ *associated with* $\mathcal{H}$*.*

Proof. We will assume without loss of generality that $\mathcal{H}_0 = 0$, since $\mathcal{H}_0$ only adds a constant term (i.e., a term that does not depend on x) to the value of the sampling criterion.

Let us first prove Assertion (a). Since $J_n(x) = \mathcal{J}_x(\mathrm{P}_n^\xi)$, it is equivalent to prove that the result holds at $n = 0$ for any $\mathrm{P}_0^\xi \in \mathbb{M}$. Assume then that $n = 0$, fix $x \in \mathbb{X}$ such that that $s_0^2(x) = k(x,x) + \tau^2(x) > 0$, and let (x_j) denote a sequence in $\mathbb{X}$ such $x_j \to x$. Recall from Equation (3.1) that $J_0(x) = \mathcal{J}_x(\mathrm{P}_0^\xi) = \mathrm{E}\big(\mathcal{H}\big(\mathrm{Cond}_{x,\, Z_1(x)}\big(\mathrm{P}_0^\xi\big)\big)\big)$. Set $\boldsymbol{\nu}_k = \mathrm{Cond}_{x_k,\, Z_1(x_k)}\big(\mathrm{P}_0^\xi\big)$ and $\boldsymbol{\nu}_\infty = \mathrm{Cond}_{x,\, Z_1(x)}\big(\mathrm{P}_0^\xi\big)$. We have $\boldsymbol{\nu}_k \in \mathfrak{P}(\xi)$ for all $n \in \mathbb{N} \cup \{+\infty\}$, and $\boldsymbol{\nu}_k \to \boldsymbol{\nu}_\infty$ by Proposition 2.9. It follows that $\mathcal{H}(\boldsymbol{\nu}_k) \xrightarrow{\text{a.s.}} \mathcal{H}(\boldsymbol{\nu}_\infty)$ since $\mathcal{H}$ is $\mathfrak{P}$-continuous, and thus $\mathcal{J}_{x_k}(\mathrm{P}_0^\xi) = \mathrm{E}\big(\mathcal{H}(\boldsymbol{\nu}_k)\big) \to \mathrm{E}\big(\mathcal{H}(\boldsymbol{\nu}_\infty)\big) = \mathcal{J}_x(\mathrm{P}_0^\xi)$ since $(\mathcal{H}(\boldsymbol{\nu}_k))$ is uniformly integrable. Assertion (a) is proved.

Consider now the following compact subsets of $\mathbb{X}$:

$$B_{n,\gamma}(\omega) = \left\{ s_n(\omega, \cdot) \geq \gamma^{-1} > 0 \right\}, \tag{3.4}$$

$$A_{n,\gamma}(\omega) = B_{n,\gamma}(\omega) \cap \left\{ J_n(\omega, \cdot) \leq \inf J_n(\omega, \cdot) + \varepsilon_n \right\}. \tag{3.5}$$

Observe that, on the event $\{s_n \not\equiv 0\} \in \mathcal{F}_n$, $B_{n,\gamma}(\omega)$ is non-empty when γ is large enough. Since $J_n(\omega,\cdot) \leq H_n(\omega)$ by Proposition 3.2, and $J_n(\omega,x) = H_n(\omega)$ for any x such that $s_n^2(x) = 0$, $A_{n,\gamma}(\omega)$ is also non-empty for large values of γ on $\{s_n \not\equiv 0\}$. Moreover, since $\mathbb{X}$ is a compact metric space, it is easily proved that $\omega \mapsto A_{n,\gamma}(\omega)$ is an $\mathcal{F}_n$-measurable random closed set, and thus admits [see, e.g., 34, Theorem 2.13] an $\mathcal{F}_n$-measurable selection $X_{n+1}^{(\gamma)}$, i.e., an $\mathbb{X}$-valued random variable such that $X_{n+1}^{(\gamma)} \in A_{n,\gamma}$ on the event $\{A_{n,\gamma} \neq \varnothing\}$. Finally, let $\tilde{x}$ denote an arbitrary fixed point in $\mathbb{X}$. Setting

$$X_{n+1} = \begin{cases} \tilde{x} & \text{if } s_n \equiv 0, \\ X_{n+1}^{(k)} & \text{if } A_{n,k} \neq \varnothing \text{ and } A_{n,l} = \varnothing,\ \forall l < k. \end{cases} \tag{3.6}$$

provides the desired ε-quasi-SUR strategy and thus finishes the proof. □

In some situations, it is possible to prove directly the continuity of the sampling criteria J_n (see Section 4.4 for an example), in which case a stronger existence result can be formulated as in [51], that does not even require the supermartingale property:

Theorem 3.13. *Let $\mathcal{H}$ denote a measurable uncertainty functional on $\mathbb{M}$, such that, for all $\nu \in \mathbb{M}$, $x \mapsto \mathcal{J}_x(\nu)$ is finite and continuous on $\mathbb{X}$. Then,*

a) for any sequential design, the sample paths of J_n are continuous on $\mathbb{X}$;
b) there exists a SUR sequential design $(X_n)_{n\geq 1}$ associated with $\mathcal{H}$.

Proof. Assertion a) follows trivially from the fact that $J_n(x) = \mathcal{J}_x(\mathrm{P}_n^\xi)$, and a SUR sequential design is again obtained using the mesurable selection theorem for random closed sets. □

3.3. General convergence results

Given a measurable uncertainty functional $\mathcal{H} : \mathbb{M} \to [0,+\infty)$ with the supermartingale property, let $\mathcal{G}_x : \mathbb{M} \to [0,+\infty)$ denote the corresponding *expected gain functional*

$$\mathcal{G}_x(\nu) = \mathcal{H}(\nu) - \mathcal{J}_x(\nu), \tag{3.7}$$

and let $\mathcal{G} : \mathbb{M} \to [0,+\infty)$ denote the *maximal expected gain functional*:

$$\mathcal{G}(\nu) = \sup_{x\in\mathbb{X}} \mathcal{G}_x(\nu). \tag{3.8}$$

Remark 3.14. Following [12], $\mathcal{G}_x$ could be called the "information" brought by an evaluation at x about the quantity of interest. This is consistent with the usual definition of mutual information, when $\mathcal{H}$ is taken to be the posterior Shannon entropy of some discrete quantity of interest.

Denote by $\mathbb{Z}_\mathcal{H}$ and $\mathbb{Z}_\mathcal{G}$ the subsets of $\mathbb{M}$ where the functionals $\mathcal{H}$ and $\mathcal{G}$, respectively, vanish. The inclusion $\mathbb{Z}_\mathcal{H} \subset \mathbb{Z}_\mathcal{G}$ always hold: indeed, $0 \leq \mathcal{J}_x \leq \mathcal{H}$ for all x by Proposition 3.2, thus $0 \leq \mathcal{G}_x \leq \mathcal{H}$, and therefore $0 \leq \mathcal{G} \leq \mathcal{H}$. The reverse inclusion plays a capital role in the following result, which provides sufficient conditions for the almost sure convergence of SUR strategies, and more generally ε-quasi-SUR strategies, associated with uncertainty functionals that enjoy the supermartingale property.

Theorem 3.15. *Let $\mathcal{H}$ denote a non-negative, measurable functional on $\mathbb{M}$ with the supermartingale property. Let (X_n) denote a quasi-SUR sequential design for $\mathcal{H}$. Then $\mathcal{G}(\mathrm{P}_n^\xi) \to 0$ almost surely. If, moreover,*

i) $H_n = \mathcal{H}\big(\mathrm{P}_n^\xi\big) \to \mathcal{H}\big(\mathrm{P}_\infty^\xi\big)$ *almost surely,*
ii) $\mathcal{G}\big(\mathrm{P}_n^\xi\big) \to \mathcal{G}\big(\mathrm{P}_\infty^\xi\big)$ *almost surely (or, equivalently,* $\mathcal{G}\big(\mathrm{P}_\infty^\xi\big) = 0$ *almost surely);*
iii) $\mathbb{Z}_\mathcal{H} = \mathbb{Z}_\mathcal{G}$*;*

then $H_n \to 0$ *almost surely.*

Proof. Since X_{n+1} is $\mathcal{F}_n$-measurable, we have:

$$\begin{aligned} J_n\left(X_{n+1}\right) &= \mathrm{E}_n\left(\mathcal{H}\left(\mathrm{Cond}_{x,Z_{n+1}(x)}\big(\mathrm{P}_n^\xi\big)\right)\right)_{|x=X_{n+1}} \\ &= \mathrm{E}_n\left(\mathcal{H}\left(\mathrm{Cond}_{X_{n+1},Z_{n+1}}\big(\mathrm{P}_n^\xi\big)\right)\right) = \mathrm{E}_n\left(H_{n+1}\right). \end{aligned} \tag{3.9}$$

Set $\Delta_{n+1} = H_n - H_{n+1}$ and $\overline{\Delta}_{n+1} = \mathrm{E}_n\left(\Delta_{n+1}\right) = H_n - \mathrm{E}_n\left(H_{n+1}\right)$. The random variables $\overline{\Delta}_n$ are non-negative since (H_n) is a supermartingale and, using that (X_n) is an ε-quasi-SUR design, we have for all $n \geq n_0$:

$$\overline{\Delta}_{n+1} = H_n - \mathrm{E}_n\left(H_{n+1}\right) = H_n - J_n\left(X_{n+1}\right) \geq H_n - \inf_{x\in\mathbb{X}} J_n(x) - \varepsilon_n, \tag{3.10}$$

i.e., since $J_n(x) = \mathcal{J}_x\big(\mathrm{P}_n^\xi\big)$ and $\mathcal{G}_x = \mathcal{H} - \mathcal{J}_x$,

$$\overline{\Delta}_{n+1} \geq \sup_{x\in\mathbb{X}} \mathcal{G}_x\big(\mathrm{P}_n^\xi\big) - \varepsilon_n = \mathcal{G}\big(\mathrm{P}_n^\xi\big) - \varepsilon_n. \tag{3.11}$$

Moreover, for any n, we have $\sum_{k=0}^{n-1} \Delta_k = H_0 - H_n$, and therefore

$$\mathrm{E}\left(\sum_{k=0}^{n-1} \overline{\Delta}_k\right) = \mathrm{E}\left(\sum_{k=0}^{n-1} \Delta_k\right) = \mathrm{E}\left(H_0 - H_n\right) \leq \mathrm{E}\left(H_0\right) < +\infty.$$

It follows that $\mathrm{E}\left(\sum_{k=0}^{\infty} \overline{\Delta}_k\right) < +\infty$, and thus $\overline{\Delta}_n \to 0$ almost surely. As a consequence, $\mathcal{G}\big(\mathrm{P}_n^\xi\big) \to 0$ almost surely, since $0 \leq \mathcal{G}\big(\mathrm{P}_n^\xi\big) \leq \overline{\Delta}_{n+1} + \varepsilon_n$.

Let now Assumptions i–iii hold. It follows from the first part of the proof that $\mathcal{G}\big(\mathrm{P}_n^\xi\big) \to 0$ almost surely. Thus, $\mathcal{G}\big(\mathrm{P}_\infty^\xi\big) = 0$ almost surely according to Assumption ii. Then $\mathcal{H}\big(\mathrm{P}_\infty^\xi\big) = 0$ since $\mathbb{Z}_\mathcal{G} \subset \mathbb{Z}_\mathcal{H}$, and the conclusion follows from Assumption i. □

Remark 3.16. Note that the conclusions of Theorem 3.15 still hold partially if it is only assumed that the condition $J_n\left(X_{n+1}\right) \leq \inf J_n + \varepsilon_n$ holds infinitely often, almost surely: in this case the conclusion of the first part of the theorem is weakened to $\liminf \mathcal{G}\big(\mathrm{P}_n^\xi\big) = 0$, but the final conclusion ($H_n \to 0$ a.s.) remains the same.

Assumptions i and ii) of Theorem 3.15 hold if $\mathcal{H}$ and $\mathcal{G}$, respectively, are $\mathfrak{P}$-continuous. Checking that $\mathcal{G}$ is $\mathfrak{P}$-continuous, however, is not easy in practice. The following results provides sufficient conditions that are easier to check.

Theorem 3.17. *Let the assumptions of Theorem 3.12 hold. Then, for any quasi-SUR sequential design,* $\mathcal{G}\big(\mathrm{P}_\infty^\xi\big) = 0$ *almost surely.*

Proof. We will assume without loss of generality that $\mathcal{H}_0 = 0$. Indeed, it is easy to check that $\mathcal{H}_0$ contributes the same additive term to both $\mathcal{H}$ and $\mathcal{J}_x$, and thus has no influence on the value of $\mathcal{G}_x = \mathcal{H} - \mathcal{J}_x$.

Let $x \in \mathbb{X}$. We have $\mathcal{H}\big(\mathrm{P}_n^\xi\big) \xrightarrow{\text{a.s.}} \mathcal{H}\big(\mathrm{P}_\infty^\xi\big)$ since $\mathcal{H}$ is $\mathfrak{P}$-continuous, $\mathcal{G}_x\big(\mathrm{P}_n^\xi\big) \xrightarrow{\text{a.s.}} 0$ by Theorem 3.15, and thus

$$\mathcal{J}_x\big(\mathrm{P}_n^\xi\big) \xrightarrow{\text{a.s.}} \mathcal{H}\big(\mathrm{P}_\infty^\xi\big). \tag{3.12}$$

Let $\mathrm{P}_{n,x}^\xi = \mathrm{Cond}_{x,Z(x)}$, with $Z(x) = \xi(x) + \tau(x)\,U$ and $U \sim \mathcal{N}(0,1)$ independent from ξ and the U_n's, and observe that $\mathcal{J}_x\big(\mathrm{P}_n^\xi\big) = \mathrm{E}_n\left(\mathcal{H}\big(\mathrm{P}_{n,x}^\xi\big)\right)$. Consider then the decomposition:

$$\mathcal{J}_x\big(\mathrm{P}_n^\xi\big) = \mathrm{E}_n\left(\mathcal{H}\big(\mathrm{P}_{n,x}^\xi\big) - \mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\right) + \mathrm{E}_n\left(\mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\right). \tag{3.13}$$

It follows from Theorem 6.23 in Kallenberg [28] that

$$\mathrm{E}_n\left(\mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\right) \xrightarrow[n\to\infty]{\text{a.s., } L^1} \mathrm{E}_\infty\left(\mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\right). \tag{3.14}$$

Moreover, note that

$$\begin{aligned}\mathrm{P}_{n,x}^\xi &= \mathrm{Cond}_{X_1,\,Z_1,\,\ldots,\,X_n,\,Z_n,\,x,\,Z(x)}\left(\mathrm{P}_0^\xi\right)\\ &= \mathrm{Cond}_{x,\,Z(x),\,X_1,\,Z_1,\,\ldots,\,X_n,\,Z_n}\left(\mathrm{P}_0^\xi\right)\end{aligned}$$

is the conditional distribution of ξ at the $(n+1)^{\text{th}}$ step of the modified sequential design $\big(\widetilde{X}_n\big)$, where $\widetilde{X}_1 = x$ and $\widetilde{X}_{n+1} = X_n$ for all $n \geq 1$, with a modified sequence of "noise variables" $\big(\widetilde{U}_n\big)$ defined by $\widetilde{U}_1 = U$ and $\widetilde{U}_{n+1} = U_n$ for all $n \geq 1$. Note also that $\mathrm{P}_{\infty,x}^\xi$ corresponds to the conditional distribution with respect to the σ-algebra generated by $\widetilde{X}_1, \widetilde{Z}_1, \widetilde{X}_2, \widetilde{Z}_2 \ldots$, where the $\widetilde{Z}_n$'s have been defined accordingly. As a result,

$$\mathrm{E}_n\left(\mathcal{H}\big(\mathrm{P}_{n,x}^\xi\big) - \mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\right) \xrightarrow[n\to\infty]{L^1} 0 \tag{3.15}$$

since $\mathcal{H}$ is $\mathfrak{P}$-continuous and $\mathfrak{P}$-uniformly integrable. Combine Equations (3.13), (3.14) and (3.15), to prove that $\mathcal{J}_x\big(\mathrm{P}_n^\xi\big) \to \mathrm{E}_\infty\big(\mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\big)$ in L^1. Then, it follows a comparison with Equation (3.12) that $\mathcal{H}\big(\mathrm{P}_\infty^\xi\big) = \mathrm{E}_\infty\big(\mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\big)$ almost surely, and therefore

$$\mathcal{G}_x\big(\mathrm{P}_\infty^\xi\big) = \mathcal{H}\big(\mathrm{P}_\infty^\xi\big) - \mathrm{E}_\infty\left(\mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\right) = 0 \quad \text{almost surely.} \tag{3.16}$$

To conclude, note that by Assertion (a) of Theorem 3.12 the sample paths of $J_\infty : x \mapsto \mathrm{E}_\infty\big(\mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\big)$ are continuous on $\big\{x \in \mathbb{X} : s_\infty^2(x) > 0\big\}$. Let $\{x_j\}$ denote a countable dense subset of $\mathbb{X}$. We have proved that, almost surely, $\mathcal{G}_{x_j}\big(\mathrm{P}_\infty^\xi\big) = 0$ for all j. Using the continuity of J_∞ on $\big\{s_\infty^2 > 0\big\}$, and the fact that $\mathcal{G}_x = 0$ on $\big\{s_\infty^2 = 0\big\}$, we conclude that, almost surely, $\mathcal{G}_x\big(\mathrm{P}_\infty^\xi\big) = 0$ for all x, and therefore $\mathcal{G}\big(\mathrm{P}_\infty^\xi\big) = 0$, which concludes the proof. □

3.4. Uncertainty functionals based on a loss function

Let us now consider specific uncertainty functionals $\mathcal{H}$ of the form

$$\mathcal{H}(\nu) = \inf_{d\in\mathbb{D}} \int_{\mathbb{S}} L(f,d)\,\nu(\mathrm{d}f) = \inf_{d\in\mathbb{D}} \overline{L}_\nu(d), \tag{3.17}$$

where $\mathbb{D}$ is a set of "decisions", $L : \mathbb{S} \times \mathbb{D} \to [0, +\infty]$ a "loss function" such that $L(\cdot, d)$ is $\mathcal{S}$-measurable for all $d \in \mathbb{D}$, and $\overline{L}_\nu(d) = \int_{\mathbb{S}} L(f, d)\, \nu(\mathrm{d}f)$. All the examples that will be discussed in Section 4 can be written in this form.

The following result formalizes an important observation of DeGroot [12, p.408] about such uncertainty functionals—namely, that they always enjoy the DoA property introduced in Section 3.1 (and thus can be studied using Theorem 3.15).

Proposition 3.18. *Let $\mathcal{H}$ denote a measurable functional on $\mathbb{M}$. If $\mathcal{H}$ is of the form (3.17), then it is DoA on $\mathbb{M}$, and consequently has the supermartingale property.*

Proof. The result follows directly from the fact that $\mathcal{H}$ is the infimum of a family of linear functionals ($\nu \mapsto \overline{L}_\nu(d)$, for all $d \in \mathbb{D}$) that commute with expectations in the following sense: for any random element $\boldsymbol{\nu}$ in $\mathbb{M}$ and any $d \in \mathbb{D}$,

$$\mathrm{E}\left(\overline{L}_{\boldsymbol{\nu}}(d)\right) = \mathrm{E}\left(\int_{\mathbb{S}} L(f, d)\, \boldsymbol{\nu}(\mathrm{d}f)\right) = \overline{L}_{\overline{\boldsymbol{\nu}}}(d),$$

where $\overline{\boldsymbol{\nu}}$ is defined as in Definition 3.3. □

An uncertainty functional of the form (3.17) is clearly $\mathcal{M}$-measurable if the infimum over d can be restricted to a countable subset of $\mathbb{D}$ (since the linear functionals $\nu \mapsto \overline{L}_\nu(d)$ are $\mathcal{M}$-measurable by Lemma A.3). This is true, for instance, if $\mathbb{D}$ is separable and $d \mapsto \overline{L}_\nu(d)$ is continuous for all ν. The following result provides more general sufficient conditions for the measurability of $\mathcal{H}$, that can cope with the case where $\overline{L}_\nu$ takes infinite values—an important example of which is discussed in Section 4.4. The reader is referred to Molchanov [34, Section 2.1] for the definition of Effros measurability.

Proposition 3.19. *Assume that $\mathbb{D}$ is a Polish space. For all $\nu \in \mathbb{M}$, set*

$$\mathbb{D}_{L,\nu} = \left\{ d \in \mathbb{D} : \overline{L}_\nu(d) < +\infty \right\}. \tag{3.18}$$

If the following conditions are satisfied:

i) L is $\mathcal{S} \otimes \mathcal{B}(\mathbb{D})$-measurable;
ii) $\overline{L}_\nu$ is continuous on $\mathbb{D}_{L,\nu}$ for all ν;
iii) $\mathbb{D}_{L,\nu}$ is closed for all ν, and $\nu \mapsto \mathbb{D}_{L,\nu}$ is Effros-measurable with respect to $\mathcal{M}$;

then $\mathcal{H}$ is $\mathcal{M}$-measurable.

Remark 3.20. Assumption iii) is always satisfied if $\overline{L}_\nu$ does not take infinite values (i.e., when $\mathbb{D}_{L,\nu} = \mathbb{D}$).

Proof. To establish the measurability of $\mathcal{H}$ we follow closely the proof of the upper semi-continous part of Lemma 2.1 of Hiai and Umegaki [25]. Since $\nu \mapsto \mathbb{D}_{L,\nu}$ is $\mathcal{M}$/Effros-measurable, the set $\mathbb{M}_L = \{\nu \in \mathbb{M} : \mathbb{D}_{L,\nu} \neq \varnothing\}$ is in $\mathcal{M}$ and there exists a Castaing representation of $\mathbb{D}_{L,\nu}$, i.e., a countable family of mesurable selections $U_k : \mathbb{M} \to \mathbb{D}$ such that $\mathbb{D}_{L,\nu} = \mathrm{cl}\{U_k(\nu)\}$ on $\mathbb{M}_L$ [see, e.g., 34, Theorem 2.3, 3) $\Rightarrow$ 5)]. Therefore, by continuity of $\overline{L}_\nu$ on $\mathbb{D}_{L,\nu}$,

$$\mathcal{H}(\nu) = \begin{cases} +\infty & \text{if } \nu \in \mathbb{M}_L, \\ \inf_k \overline{L}_\nu(U_k(\nu)) & \text{otherwise.} \end{cases} \tag{3.19}$$

Since $(\nu, d) \mapsto \overline{L}_\nu(d)$ is $\mathcal{M} \otimes \mathcal{B}(\mathbb{D})$-measurable by Lemma A.3, $\nu \mapsto \overline{L}_\nu(U_k(\nu))$ is measurable for all k, and therefore $\mathcal{H}$ is measurable as the infinimum of a countable family of measurable functions. □

Remark 3.21. The assumption of continuity of $\overline{L}_\nu$ over $\mathbb{D}_{L,\nu}$ can be relaxed to lower semi-continuity, in which case it can be proved (following the lower semi-continuous part of Lemma 2.1 in Hiai and Umegaki [25]) that $\nu \mapsto \mathcal{H}(\nu) = \inf_d \overline{L}_\nu(d)$ is $\mathcal{M}_0^\mu$-measurable for any probability measure μ on $(\mathbb{M}, \mathcal{M})$, where $\mathcal{M}^\mu$ denotes the completion of $\mathcal{M}$ with respect to μ.

Three of the examples of SUR sequential designs from the literature that will be analyzed in Section 4 are based on regular non-negative loss functions in the following sense.

Definition 3.22. *We will say that a non-negative loss function $L : \mathbb{S} \times \mathbb{D} \to [0, +\infty)$ is* regular *if*

i) $\mathbb{D}$ *is a separable space,*
ii) *for all $d \in \mathbb{D}$, $L(\cdot, d)$ is $\mathcal{S}$-measurable,*
iii) *for all $\nu \in \mathbb{M}$, $\overline{L}_\nu$ takes finite values and is continuous on $\mathbb{D}$,*

and if the corresponding functionals $\mathcal{H}$ and $\mathcal{G}$ satisfy:

iv) $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$, *where $\mathcal{H}_0(\nu) = \int_{\mathbb{S}} L_0 \, \mathrm{d}\nu$ for some $L_0 \in \cap_{\nu \in \mathbb{M}} \mathcal{L}^1(\mathbb{S}, \mathcal{S}, \nu)$, and $\mathcal{H}_1$ is $\mathfrak{P}$-uniformly integrable and $\mathfrak{P}$-continuous,*
v) $\mathbb{Z}_{\mathcal{H}} = \mathbb{Z}_{\mathcal{G}}$.

The following result is provided as a convenient summary of the results that hold for uncertainty functionals based on regular non-negative loss functions.

Corollary 3.23. *Let $\mathcal{H}$ denote a functional of the form (3.17) for some non-negative loss function L. If L is regular, then*

a) *$\mathcal{H}$ is a measurable functional that satisfies the assumptions of Theorems 3.12, 3.15 and 3.17.*

In particular,

b) *for any sequence $\varepsilon = (\varepsilon_n)$ of positive real numbers, there exists an ε-quasi-SUR sequential design $(X_n)_{n \geq 1}$ associated with $\mathcal{H}$,*
c) *for any quasi-SUR design, $H_n = \mathcal{H}(\mathrm{P}_n^\xi) \to 0$ almost surely.*

4. Applications to popular sequential design strategies

We now present applications of the previously established results to four popular sequential design strategies, two of them adressing the excursion case (Sections 4.1 and 4.2), and the other two adressing the optimization case (Sections 4.3 and 4.4). For each example, the convergence results are preceded by details on the associated loss functions, uncertainty functionals and sampling criteria.

4.1. The integrated Bernoulli variance functional

Here we focus on the case where $\mathbb{X}$ is endowed with a finite measure μ and we let $T \in \mathbb{R}$ be a given excursion threshold. For any measurable function $f : \mathbb{X} \mapsto \mathbb{R}$, we let $\Gamma(f) = \{u \in \mathbb{X} : f(u) \geq T\}$ and $\alpha(f) = \mu(\Gamma(f))$. The quantities of interest are then $\Gamma(\xi)$ and $\alpha(\xi)$. Let $p_n(u) = \mathrm{E}_n\left(\mathbb{1}_{\Gamma(\xi)}(u)\right) = \mathrm{P}_n\left(\xi(u) \geq T\right)$. A typical choice of measure of residual uncertainty in this case is the integrated indicator—or "Bernoulli"—variance [3]:

$$H_n = \int_{\mathbb{X}} p_n \left(1 - p_n\right) d\mu, \tag{4.1}$$

which corresponds to the uncertainty functional

$$\mathcal{H}(\nu) = \int_{\mathbb{X}} p_\nu \,(1 - p_\nu)\, d\mu, \qquad \nu \in \mathbb{M}, \tag{4.2}$$

where $p_\nu(u) = \mathrm{P}_\nu\,(\xi(u) \geq T)$. See [10] for more information on the computation of the corresponding SUR sampling criterion J_n.

The functional (4.2) can be seen as the uncertainty functional induced by the loss function

$$\begin{aligned} L : \mathbb{S} \times \mathbb{D} &\to \mathbb{R}_+, \\ (f, d) &\mapsto \|\mathbb{1}_{\Gamma(f)} - d\|^2_{L^2(\mathbb{X})}, \end{aligned} \tag{4.3}$$

where $\mathbb{D} \subset L^2(\mathbb{X})$ is the set of "soft classification" functions on $\mathbb{X}$ (i.e., measurable functions defined on $\mathbb{X}$ and taking values in $[0, 1]$). Indeed, for all $\nu \in \mathbb{M}$,

$$\overline{L}_\nu(d) = \mathrm{E}_\nu\,(L(\xi, d)) = \|p_\nu - d\|^2_{L^2(\mathbb{X})} + \int p_\nu(1 - p_\nu)\,\mathrm{d}\mu$$

is minimal for $d = p_\nu$, and therefore $\mathcal{H}(\nu) = \inf_{d\in\mathbb{D}} \overline{L}_\nu(d)$.

The following theorem establishes the convergence of SUR (or quasi-SUR) designs associated to this uncertainty functional using the theory developed in Section 3.4 for regular loss functions.

Theorem 4.1. *The loss function* (4.3) *is regular in the sense of Definition 3.22, so that all the conclusions of Corollary 3.23 apply. In particular* $\mathcal{H}(\mathrm{P}_n^\xi) \xrightarrow{\text{a.s.}} 0$ *for any quasi-SUR design associated with* $\mathcal{H}$*.*

Proof. The proof consists in six points, as follows:

a) $\mathbb{D}$ is separable;

The space $L^2(\mathbb{X})$ is a separable metric space since $\mathbb{X}$ is a separable measure space (see, e.g., Theorem 4.13 in [7]). Hence $\mathbb{D}$ is also separable.

b) for all $d \in \mathbb{D}$, $L(\cdot, d)$ is $\mathcal{S}$-measurable;

Indeed, $f \mapsto \int_{\mathbb{X}} \left(\mathbb{1}_{f(x)\geq T} - d(x)\right)^2 \mathrm{d}\mu(x)$ is $\mathcal{S}$-measurable by Fubini's theorem since the integrand is $\mathcal{S} \otimes \mathcal{B}(\mathbb{X})$-jointly measurable in (f, x).

c) for all $\nu \in \mathbb{M}$, $\overline{L}_\nu$ takes finite values and is continuous on $\mathbb{D}$;

Here $\overline{L}_\nu$ is clearly finite since the loss is upper-bounded by $\mu(\mathbb{X})$, and its continuity directly follows from the continuity of the norm.

d) $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$, where $\mathcal{H}_0(\nu) = \int_{\mathbb{S}} L_0 \,\mathrm{d}\nu$ for some $L_0 \in \cap_{\nu\in\mathbb{M}}\mathcal{L}^1\,(\mathbb{S}, \mathcal{S}, \nu)$, and $\mathcal{H}_1$ is $\mathfrak{P}$-uniformly integrable;

Here this holds with $L_0 = 0$ and $\mathcal{H}_1 = \mathcal{H}$. Indeed, $\mathcal{H}$ is trivially $\mathfrak{P}$-uniformly integrable since the loss is upper-bounded.

e) $\mathcal{H}_1$ is $\mathfrak{P}$-continuous;

Let $\xi : (\Omega, \mathcal{F}, \mathrm{P}) \times \mathbb{X} \to \mathbb{R}$ be a Gaussian process on $\mathbb{S}$ with mean and covariance functions m and k. Let $(\boldsymbol{\nu}_n)$ be a sequence of random elements in $\mathfrak{P}(\xi)$ so that a.s. $\nu_n \to \nu_\infty$. For $n \in \mathbb{N}\cup\{\infty\}$, let m_n

and k_n be the mean and covariance functions of ν_n. Let also $\sigma^2(u) = k(u,u)$ and $\sigma_n^2(u) = k_n(u,u)$. For $u \in \mathbb{X}$ and $n \in \mathbb{N} \cup \{\infty\}$, let

$$g_n(u) = g\left(\bar{\Phi}\left(\frac{T - m_n(u)}{\sigma_n(u)}\right)\right),$$

where $g(p) = p(1-p)$ and $\bar{\Phi}(t) = P(Z \geq t)$ where Z is a standard Gaussian variable, with the convention that $\bar{\Phi}(0/0) = 1$. Using this notation, for $n \in \mathbb{N} \cup \{+\infty\}$ and for almost all $\omega \in \Omega$,

$$\mathcal{H}(\nu_n) = \int_{\mathbb{X}} g_n(u)\,\mathrm{d}\mu(u) = \int_{A(\omega)} g_n(u)\,\mathrm{d}\mu(u) \tag{4.4}$$

where $A(\omega) = \{u \in \mathbb{X} : \sigma(u) > 0, \sigma_\infty(\omega,u) = 0, m_\infty(\omega,u) \neq T\} \cup \{\sigma(u) > 0, \sigma_\infty(\omega,u) > 0\}$, as proven below. The property then follows using that, since $\nu_n \to \nu_\infty$ almost surely, for almost all $\omega \in \Omega$ and for all $u \in A(\omega)$, $m_n(u) \to_{n\to\infty} m_\infty(u)$ and $\sigma_n(u) \to_{n\to\infty} \sigma_\infty(u)$. Furthermore, either $\sigma_\infty(u) > 0$ or $\sigma_\infty(u) = 0, m_\infty(u) \neq T$. Hence, we have that $g(\bar{\Phi}([m_n(u) - T]/\sigma_n(u))) \to_{n\to\infty} g(\bar{\Phi}([m_\infty(u) - T]/\sigma_\infty(u)))$. So, for almost all $\omega \in \Omega$ we can apply the dominated convergence theorem and obtain that

$$\mathcal{H}(\nu_n) = \int_{A(\omega)} g_n(u)\,\mathrm{d}\mu(u) \to_{n\to\infty} \mathcal{H}(\nu_\infty) = \int_{A(\omega)} g_\infty(u)\,\mathrm{d}\mu(u).$$

Let us now prove Equation (4.4). Observe first that, for u so that $\sigma(u) = 0$, we have $\sigma_n(u) = 0$ for all $n \in \mathbb{N} \cup \{\infty\}$ since $\nu_n \in \mathfrak{P}(\xi)$. Hence, $g_n(u) = 0$ when $\sigma(u) = 0$. Thus, setting $B(\omega) = \{u \in \mathbb{X};\, \sigma(u) > 0,\, \sigma_\infty(u) = 0,\, m_\infty(u) = T\}$, we have

$$\mathcal{H}(\nu_n) = \int_{A(\omega)} g_n(u)\,\mathrm{d}\mu(u) + \int_{B(\omega)} g_n(u)\,\mathrm{d}\mu(u).$$

Then, since $(\omega,u) \mapsto m_\infty(\omega,u)$ is jointly measurable by continuity of m_∞ for all $\omega \in \Omega$, we obtain from Fubini's theorem:

$$\mathrm{E}(\mu(B(\omega))) = \int_{\mathbb{X}} \mathbb{1}_{\sigma(u)>0}\,\mathrm{E}(\mathbb{1}_{\sigma_\infty(u)=0}\mathbb{1}_{m_\infty(u)=T})\,\mathrm{d}\mu(u) = 0. \tag{4.5}$$

This follows from $0 = \mathbb{1}_{\sigma_\infty(u)=0}\sigma_\infty^2(u) = \mathrm{E}(\mathbb{1}_{\sigma_\infty(u)=0}(\xi(u) - m_\infty(u))^2|\mathcal{F}'_\infty)$ so that, almost surely, $\mathbb{1}_{\sigma_\infty(u)=0}(\xi(u) - m_\infty(u))^2 = 0$ and thus $\mathbb{1}_{\sigma_\infty(u)=0}\mathbb{1}_{m_\infty(u)=T} = \mathbb{1}_{\sigma_\infty(u)=0}\mathbb{1}_{\xi(u)=T}$. Finally, as $\xi(u)$ is a Gaussian variable, we have $\mathrm{E}(\mathbb{1}_{\sigma_\infty(u)=0}\mathbb{1}_{m_\infty(u)=T}) = 0$ when $\sigma(u) > 0$.

f) $\mathbb{Z}_\mathcal{H} = \mathbb{Z}_\mathcal{G}$.

Let $\nu \in \mathbb{Z}_\mathcal{G}$ and let $\xi \sim \nu$. Let m, k, σ^2 be defined as above. Let $U \sim \mathcal{N}(0,1)$ be independent of ξ. Since $\mathcal{G}(\nu) = 0$, we have from Fubini theorem and the law of total variance

$$\int_{\mathbb{X}} \mathrm{var}\left[\mathrm{E}\left(\mathbb{1}_{\xi(u)\geq T}|Z_x\right)\right] d\mu(u) = 0$$

for all $x \in \mathbb{X}$, where $Z_x = \xi(x) + \tau(x)U$. Hence, for all $x \in \mathbb{X}$, for almost all $u \in \mathbb{X}$, we have

$$\mathrm{var}\left(\bar{\Phi}\left(\frac{T - m(u) - \frac{k(x,u)\,(Z_x - m(x))}{\sigma^2(x)+\tau^2(x)}}{\sqrt{\sigma^2(u) - \frac{k(x,u)^2}{\sigma^2(x)+\tau^2(x)}}}\right)\right) = 0,$$

which implies that $k(x,u)=0$ (as can be proven without difficulty by separating the cases of nullity and non-nullity of the denominator). Hence, if there exists x^* for which $\sigma^2(x^*)=k(x^*,x^*)>0$, we obtain a contradiction, since then $k(x,u)>0$ in a neighboorhood of x^* by continuity. We conclude that $\sigma^2(x)=0$ for all $x\in\mathbb{X}$, and therefore $\mathcal{H}(\nu)=0$. □

In the next proposition, we refine Theorem 4.1 by showing that it entails a consistent estimation of the excursion set $\Gamma(\xi)$.

Proposition 4.2. *For any quasi-SUR design associated with $\mathcal{H}$, as $n\to\infty$, almost surely and in L^1,*

$$\int_{\mathbb{X}} \left(\mathbb{1}_{\xi(u)\geq T} - p_n(u)\right)^2 d\mu(u) \to 0$$

and

$$\int_{\mathbb{X}} \left(\mathbb{1}_{\xi(u)\geq T} - \mathbb{1}_{p_n(u)\geq 1/2}\right)^2 d\mu(u) \to 0.$$

Proof. From the proofs of **e)** and **f)** in the proof of Theorem 4.1, it follows that a.s.

$$\int_{\mathbb{X}} \left(\mathbb{1}_{\xi(u)\geq T} - p_n(u)\right)^2 d\mu(u) = \int_{A(\omega)} \left(\mathbb{1}_{\xi(u)\geq T} - p_n(u)\right)^2 d\mu(u).$$

Also, for all $u\in A(\omega)$, $p_n(u)$ goes almost surely to $\mathbb{1}_{\xi(u)\geq T}$ as $n\to\infty$ since $\sigma_\infty\equiv 0$ a.s. from the proof of **f)** in Theorem 4.1 and the conclusion of this theorem. Hence the first part of the proposition follows by applying the dominated convergence theorem twice. The proof of the second part of the proposition is identical. □

4.2. The variance of excursion volume functional

Following up on the previous example of Section 4.1, we now consider the alternative measure of residual uncertainty from [3, 10]:

$$H_n = \mathcal{H}\left(\mathrm{P}_n^\xi\right) = \mathrm{var}_n(\alpha(\xi)) = \mathrm{var}_n(\mu(\Gamma(\xi))). \tag{4.6}$$

The corresponding sampling criterion is

$$J_n(x) = \mathrm{E}_n\,\mathrm{var}_n(\alpha(\xi)|Z_{n+1}(x)),$$

with $Z_{n+1}(x)$ as in Eq. 2.1 with X_n replace by x and U_n replaced by U_{n+1}. This uncertainty again derives from a loss function, and even a simpler one as here $L(f,d)=(\alpha(f)-d)^2$ with $\mathbb{D}=\mathbb{R}$ leads to

$$\overline{L}_{\mathrm{P}_n^\xi}(d) = \mathrm{E}_n[(\alpha(\xi)-d)^2] = \mathrm{var}_n(\alpha(\xi)) + (\mathrm{E}_n(\alpha(\xi))-d)^2$$

whereof $\overline{L}_{\mathrm{P}_n^\xi}$ reaches its infimum for $d=\mathrm{E}_n(\alpha(\xi)$ and $H_n=\inf_{d\in\mathbb{D}}\overline{L}_{\mathrm{P}_n^\xi}(d)$. As in the last section, the theorem below ensures that L is regular and that the further pre-requisites are met towards establishing almost sure convergence of H_n to 0 for corresponding (quasi-)SUR sequential designs.

Theorem 4.3. *The loss function $L(f,d)=(\alpha(f)-d)^2$, where $d\in\mathbb{D}=\mathbb{R}$, is regular in the sense of Definition 3.22, so that all the conclusions of Corollary 3.23 apply. In particular $\mathcal{H}(\mathrm{P}_n^\xi)\xrightarrow{\text{a.s.}}0$ for any quasi-SUR design associated with $\mathcal{H}$.*

Proof. The proof consists in the same six points as in the proof of Theorem 4.1. Here **a)** and **c)** are obvious. Let us now prove the four remaining points:

b) We have $L(f,d) = (\int_{\mathbb{X}} \mathbb{1}_{f(u)\geq T} d\mu(u) - d)^2$, so that for fixed d, $L(f,d)$ is a $\mathbb{S}$-measurable function of f, as can be shown similarly as in the proof of Theorem 4.1.

d) We use again $L_0 = 0$ for this criterion. The functional $\mathcal{H}_1 = \mathcal{H}$ is trivially $\mathfrak{P}$-uniformly integrable since $0 \leq \mathcal{H}_1 \leq \frac{1}{4}\mu(\mathbb{X})^2$ by Popoviciu's inequality.

e) Let us now show that $\mathcal{H}_1 = \mathcal{H}$ is $\mathfrak{P}$-continuous. We use the same notation ξ, σ, σ_n, m and m_n, for $n \in \mathbb{N} \cup \{+\infty\}$, as in the proof of Theorem 4.1. We have

$$\mathcal{H}(\nu_n) = \int_{\mathbb{X}} \int_{\mathbb{X}} c_n(u_1,u_2) d\mu(u_1) d\mu(u_2),$$

where $c_n(u_1,u_2) = \mathrm{cov}(\mathbb{1}_{\xi(u_1)\geq T}, \mathbb{1}_{\xi(u_2)\geq T} | \mathcal{F}'_n)$ for $n \in \mathbb{N} \cup \{+\infty\}$. Using the same notation $A(\omega)$ and a similar reasoning as in the proof of Theorem 4.1, we have that for $n \in \mathbb{N} \cup \{+\infty\}$ and ω-almost surely,

$$\mathcal{H}(\nu_n) = \int_{A(\omega)} \int_{A(\omega)} c_n(u_1,u_2) d\mu(u_1) d\mu(u_2).$$

For $j = 1,2$ and $u_j \in A(\omega)$, we have either $\sigma_\infty(u_j) > 0$ or $\sigma_\infty(u_j) = 0, m_\infty(u_j) \neq T$. Hence, for almost all $\omega \in \Omega$, for $u_1 \in A(\omega)$ and $u_2 \in A(\omega)$, we obtain $c_n(u_1,u_2) \to_{n\to\infty} c_\infty(u_1,u_2)$ by Lemma 4.4 (proven later):

Lemma 4.4. *Let $m_n = (m_{n1}, m_{n2})^t \to (m_1,m_2)^t = m$ as $n \to \infty$. Consider a sequence of covariance matrices Σ_n so that*

$$\Sigma_n = \begin{pmatrix} \sigma_{n1} & \sigma_{n12} \\ \sigma_{n12} & \sigma_{n2} \end{pmatrix} \to_{n\to\infty} \begin{pmatrix} \sigma_1 & \sigma_{12} \\ \sigma_{12} & \sigma_2 \end{pmatrix} = \Sigma.$$

Assume that for $i = 1,2$ we have $m_i \neq T$ or $\sigma_i > 0$. Let $Z_n \sim \mathcal{N}(m_n, \Sigma_n)$ and $Z \sim \mathcal{N}(m,\Sigma)$. Then as $n \to \infty$, $\mathrm{cov}(\mathbb{1}_{\{Z_{n1}\geq T\}}, \mathbb{1}_{\{Z_{n2}\geq T\}}) \to \mathrm{cov}(\mathbb{1}_{\{Z_1\geq T\}}, \mathbb{1}_{\{Z_2\geq T\}})$.

Hence, for almost all $\omega \in \Omega$, we can use the dominated convergence theorem to obtain that $\mathcal{H}(\nu_n) \to_{n\to\infty} \mathcal{H}(\nu_\infty)$. Hence, $\mathcal{H}$ is $\mathfrak{P}$-continuous.

f) Observe that, for any $\nu \in \mathbb{M}$,

$$\begin{aligned} \mathcal{G}(\nu) = 0 &\iff \forall x \in \mathbb{X},\ \mathcal{G}_x(\nu) = 0 \\ &\iff \forall x \in \mathbb{X},\ \mathrm{E}\left(\mathrm{var}(\alpha(\xi) \mid Z_x)\right) = \mathrm{var}(\alpha(\xi)) \\ &\iff \forall x \in \mathbb{X},\ \mathrm{var}\left(\mathrm{E}(\alpha(\xi) \mid Z_x)\right) = 0 \\ &\iff \forall x \in \mathbb{X},\ \alpha(\xi) - \mathrm{E}(\alpha(\xi)) \perp L^2(Z_x), \end{aligned} \tag{4.7}$$

where $\xi \sim \nu$ and $Z_x = \xi(x) + \tau(x)\,U$, with $U \sim \mathcal{N}(0,1)$ independent from ξ.

Let $\nu \in \mathbb{Z}_{\mathcal{G}}$. Using Lemma A.9, it follows from Equation (4.7) that $\alpha(\xi) - \mathrm{E}(\alpha(\xi)) \perp L^2(\xi(x))$, for all $x \in \mathbb{X}$. In particular, $\alpha(\xi) - \mathrm{E}(\alpha(\xi)) \perp \mathbb{1}_{\xi(x)\geq T}$, for all $x \in \mathbb{X}$, and thus

$$\mathrm{var}\left(\alpha(\xi)\right) = \int \mathrm{cov}\left(\alpha(\xi), \mathbb{1}_{\xi(x)\geq T}\right) \mu(\mathrm{d}x) = 0,$$

which concludes the proof. □

Proof of Lemma 4.4. By the convergence of moments and Gaussianity, $(Z_{n1}, Z_{n2}) \to^d_{n\to\infty} (Z_1, Z_2)$. Furthermore, from the assumptions the cumulative distribution functions of Z_1 and Z_2 are continuous at T so that by the Portemanteau theorem, $P(Z_{ni} \geq T) \to_{n\to\infty} P(Z_i \geq T)$. In addition, $Y := \min(Z_1, Z_2)$ then also has a continuous cumulative distribution function at T and, as $\mathrm{E}(\mathbb{1}_{\{Z_1\geq T\}}\mathbb{1}_{\{Z_2\geq T\}}) = P(Y \geq T)$, we similarly get that $n \to \infty$, $\mathrm{E}(\mathbb{1}_{\{Z_{n1}\geq T\}}\mathbb{1}_{\{Z_{n2}\geq T\}}) \to \mathrm{E}(\mathbb{1}_{\{Z_1\geq T\}}\mathbb{1}_{\{Z_2\geq T\}})$ which suffices to conclude. □

Similarly as before, in the next proposition, we show that Theorem 4.3 yields a consistent estimation of the excursion volume.

Proposition 4.5. *For any (quasi-)SUR design associated with $\mathcal{H}$, as $n \to \infty$, almost surely and in L^1, $\mathrm{E}_n(\alpha(\xi)) \to \alpha(\xi)$.*

Proof. Let $\alpha = \alpha(\xi)$. From Theorem 4.3, $\mathrm{var}_n(\alpha) \to^{a.s.}_{n\to\infty} 0$ so that, by dominated convergence, $\mathrm{E}(\mathrm{var}_n(\alpha)) \to_{n\to\infty} 0$. Hence, $\mathrm{E}(\mathrm{E}_n[(\mathrm{E}_n(\alpha) - \alpha)^2]) \to_{n\to\infty} 0$ so that $\mathrm{E}_n(\alpha) \to^{L^1}_{n\to\infty} \alpha$. Now, $\mathrm{E}_n(\alpha)$ is a martingale bounded in L^1 so, by Theorem 6.23 in [28], it converges a.s. to a random variable. Thus $\mathrm{E}_n(\alpha) \to^{a.s.}_{n\to\infty} \alpha$. □

4.3. The Knowledge Gradient functional

Coming to the topic of sequential design for global optimization, we now focus on the knowledge gradient criterion that has been used within Gaussian Process modelling following [18, 19]. The knowledge gradient sampling criterion (with maxima taken over the whole domain $\mathbb{X}$) is defined by

$$\widetilde{J}_n(x) = \mathrm{E}_n\left(\max_{u\in\mathbb{X}} \mathrm{E}_n[\xi(u)|Z_{n+1}(x)]\right) - \max_{u\in\mathbb{X}} \mathrm{E}_n[\xi(u)].$$

Here we consider a slightly modified criterion that defines the same strategy and fits more naturally in our framework (note that $\widetilde{J}_n$ is to be maximized while J_n is to be minimized as the criteria of the previous sections):

$$J_n(x) = \mathrm{E}_n[\max_{u\in\mathbb{X}} \xi(u)] - \mathrm{E}_n\left(\max_{u\in\mathbb{X}} \mathrm{E}_n[\xi(u)|Z_{n+1}(x)]\right).$$

Indeed the latter corresponds to the uncertainty

$$H_n = \mathcal{H}\left(\mathrm{P}^\xi_n\right) = \mathrm{E}_n[\max_{u\in\mathbb{X}} \xi(u)] - \max_{u\in\mathbb{X}} \mathrm{E}_n[\xi(u)]. \tag{4.8}$$

This time again, $\mathcal{H}$ derives from a loss function, with $\mathbb{D} = \mathbb{X}$ and $L(f, d) = \max_{u\in\mathbb{X}} f(u) - f(d)$, leading to

$$\overline{L}_{\mathrm{P}^\xi_n}(d) = \mathrm{E}_n[\max_{u\in\mathbb{X}} \xi(u)] - \mathrm{E}_n[\xi(d)]$$

whereof $\overline{L}_{\mathrm{P}^\xi_n}$ reaches its infimum for $d \in \arg\max_{u\in\mathbb{X}} \mathrm{E}_n[\xi(u)]$ and $H_n = \inf_{d\in\mathbb{D}} \overline{L}_{\mathrm{P}^\xi_n}(d)$. Still following the same route as in the last two sections, we have:

Theorem 4.6. *The loss function $L(f, d) = \max_{u\in\mathbb{X}} \xi(u) - \xi(d)$, where $d \in \mathbb{D} = \mathbb{X}$, is regular in the sense of Definition 3.22, so that all the conclusions of Corollary 3.23 apply. In particular $\mathcal{H}(\mathrm{P}^\xi_n) \xrightarrow{\text{a.s.}} 0$ for any quasi-SUR design associated with $\mathcal{H}$.*

Proof. The proof consists in the same six points as in the proof of Theorem 4.1.

a) $\mathbb{X}$ is a compact metric space, hence separable.

b) The mapping $L(\cdot, d) : f \mapsto \max f - f(d)$ is continuous on $\mathbb{S}$, hence $\mathcal{S}$-measurable.

c) $\overline{L}_\nu : d \mapsto \int \max f \, \mathrm{d}\nu - m_\nu(d)$ is continuous since $m_\nu \in \mathbb{S}$ for all $\nu \in \mathbb{M}$.

d) Let $L_0(f) = \max f$. Then $L_0 \in \cap_{\nu\in\mathbb{M}} \mathcal{L}^1(\mathbb{S}, \mathcal{S}, \nu)$ since $|L_0(f)| \leq L^+(f) \coloneqq \max|f|$ and $\mathrm{E}(\max_{\mathbb{X}}|\xi|) < \infty$ for any continuous Gaussian process ξ on a compact metric space $\mathbb{X}$. Moreover, it follows from Proposition 3.7 that $\mathcal{H}_1 : \nu \mapsto -\max m_\nu$ is $\mathfrak{P}$-uniformly integrable, since $|\mathcal{H}_1(\nu)| \leq \int L^+ \mathrm{d}\nu$.

e) We have that $\mathcal{H}_1(\nu) = \inf_{d\in\mathbb{X}} \int_{\mathbb{S}} (-f(d)) d\nu(f) = -\max_{u\in\mathbb{X}} m(u)$, where m is the mean function of ν. Consider a sequence of measures $\nu_n \in \mathbb{M}$, with mean functions m_n converging, as $n \to \infty$, to $\nu_\infty \in \mathbb{M}$ with mean function m_∞, in the sense of Definition 2.6. Then, m_n converges uniformly to m_∞ as $n \to \infty$, so that $\max_{u\in\mathbb{X}} m_n(u)$ converges as $n \to \infty$ to $\max_{u\in\mathbb{X}} m_\infty(u)$. Hence $\mathcal{H}_1(\nu_n)$ converges as $n \to \infty$ to $\mathcal{H}_1(\nu_\infty)$.

f) Let $\nu \in \mathbb{Z}_{\mathcal{G}}$ and let $\xi \sim \nu$. Let m, k, σ^2 be defined, w.r.t. ξ, as in the proof of Theorem 4.1. Let $v_x = \xi(x) + \tau(x)U$ with $U \sim \mathcal{N}(0,1)$ independently of ξ. Let x^* satisfy $m(x^*) = \max_{u\in\mathbb{X}} m(u)$. We have, after some standard computations, for all $x \in \mathbb{X}$,

$$0 = \mathcal{G}_x(\nu) = \mathrm{E}(\max_{u\in\mathbb{X}} \mathrm{E}(\xi(u)|v_x)) - m(x^*).$$

This implies that, for all $x, y \in \mathbb{X}$,

$$\mathrm{E}\left[\max(\mathrm{E}(\xi(y)|v_x), \mathrm{E}(\xi(x^*)|v_x)) - \mathrm{E}(\xi(x^*)|v_x)\right] = 0.$$

Hence, for all $x, y \in \mathbb{X}$,

$$\mathrm{E}\left[\left(\mathrm{E}(\xi(y)|v_x) - \mathrm{E}(\xi(x^*)|v_x)\right)^+\right] = 0.$$

This implies, by Gaussianity, that for all $x, y \in \mathbb{X}$,

$$\mathrm{var}\left(\mathrm{E}(\xi(y)|v_x) - \mathrm{E}(\xi(x^*)|v_x)\right) = 0.$$

Hence, for all $x, y \in \mathbb{X}$,

$$\begin{aligned} 0 &= \mathrm{var}\left(m(y) - m(x^*) + \mathbb{1}_{\sigma^2(x)+\tau^2(x)>0} \frac{k(x,y) - k(x,x^*)}{\sigma^2(x) + \tau^2(x)} (v_x - m(x))\right) \\ &= \mathbb{1}_{\sigma^2(x)+\tau^2(x)>0} \frac{(k(x,y) - k(x,x^*))^2}{\sigma^2(x) + \tau^2(x)}. \end{aligned}$$

Hence, for all $x, y \in \mathbb{X}$, whether or not $\sigma^2(x) + \tau^2(x) = 0$, we have

$$k(x,y) = k(x,x^*). \tag{4.9}$$

Now, for all $x, y \in \mathbb{X}$,

$$\begin{aligned}\operatorname{var}(\xi(x) - \xi(y)) &\leq 4 \max_{x \in \mathbb{X}} \operatorname{var}(\xi(x) - \xi(x^*)) \\ &\leq 4 \max_{x \in \mathbb{X}} |k(x,x) + k(x^*,x^*) - 2k(x,x^*)| \\ &\leq 4 \max_{x \in \mathbb{X}} |k(x,x) - k(x,x^*)| + 4 \max_{x \in \mathbb{X}} |k(x^*,x^*) - k(x,x^*)| \\ &= 0\end{aligned}$$

from (4.9). Hence, for all $x, y \in \mathbb{X}$, $\xi(x) - m(x) = \xi(y) - m(y)$ a.s. Hence $\mathrm{E}(\max_{u \in \mathbb{X}} \xi(u)) = \mathrm{E}[m(x^*) + \xi(x^*) - m(x^*)] = m(x^*) = \max_{u \in \mathbb{X}} m(u)$. Hence $\mathcal{H}(\nu) = 0$. □

In the next proposition, we refine Theorem 4.6, by showing that the loss goes to zero for an optimal decision in $\arg\max_{u \in \mathbb{X}} \mathrm{E}_n[\xi(u)]$.

Proposition 4.7. *Let x_n^* be any ($\mathcal{F}_n$-measurable) sequence in $\arg\max_{u \in \mathbb{X}} \mathrm{E}_n[\xi(u)]$. Then, for any (quasi-)SUR design associated with $\mathcal{H}$, as $n \to \infty$, almost surely and in L^1, $\xi(x_n^*) \to \max \xi$.*

Proof. From the proof of **f)** in the proof of Theorem 4.6, and the fact that $\mathcal{H}(P_\infty^\xi) = 0$ a.s., it follows that a.s. $\operatorname{var}_\infty(\xi(x) - \xi(y)) = 0$ for all $x, y \in \mathbb{X}$. Hence, a.s. for all $x, y \in \mathbb{X}$ $\xi(x) - \xi(y) = m_\infty(x) - m_\infty(y)$. Let $x^* \in \arg\max \xi$. We then have a.s.

$$\limsup_{n \to \infty} (\xi(x^*) - \xi(x_n^*)) = \limsup_{n \to \infty} (m_\infty(x^*) - m_\infty(x_n^*)) = \limsup_{n \to \infty} (m_n(x^*) - m_n(x_n^*)) \leq 0.$$

Hence $\xi(x_n^*) \to_{n \to \infty} \max \xi$ a.s. The convergence in the L^1 sense is obtained by the dominated convergence theorem. □

4.4. The Expected Improvement functional

In this subsection, we address the celebrated expected improvement sequential strategy Jones et al. [27], Mockus et al. [33]. We address the case of exact observations for which $\tau(x) = 0$ for all $x \in \mathbb{X}$. Then, the expected improvement strategy is defined by

$$X_{n+1} \in \operatorname{argmax}_{x \in \mathbb{X}} \mathrm{E}_n \left(M_{n+1,x} - M_n \right), \tag{4.10}$$

with $M_n = \max_{\sigma_n(u)=0} \xi(u)$ and $M_{n+1,x} = \max_{\sigma_n(u|x)=0} \xi(u)$, where $\sigma_n(u|x)$ is the conditional variance of $\xi(u)$ given $\xi(x)$ under P_n. We remark that, in the case when $\sigma_n(u) = 0$ if and only if $u \in \{X_1, ..., X_n\}$, the strategy can be written more familiarly as

$$X_{n+1} \in \operatorname{argmax}_{x \in \mathbb{X}} \mathrm{E}_n \left([\xi(x) - \max(\xi(X_1), ..., \xi(X_n))]^+ \right). \tag{4.11}$$

The criterion in Equation (4.10) brings more generality than that of Equation (4.11), since it allows, for instance, for Gaussian processes whose sample paths satisfy symmetry properties.

It is plain that the expected improvement strategy can be written as $X_{n+1} \in \operatorname{argmax}_{x \in \mathbb{X}} \mathrm{E}_n(\mathcal{H}(\operatorname{Cond}_{x,\xi(x)}(P_n^\xi)))$, with

$$\mathcal{H}(\nu) = \mathrm{E}(\max \xi) - \max_{\sigma(u)=0} m(u),$$

where $\xi \sim \nu$ and ν has mean and standard deviation functions m and σ. We now associate a loss function to $\mathcal{H}$. We let $\mathbb{D} = \mathbb{X} \times \mathbb{R}$, endowed with the sigma-algebra $\mathcal{B}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{X} \times \mathbb{R})$ (as these two spaces are separable). We let $d = (x^*, z^*)$,

$$L\,(f, d) = \begin{cases} \max f - z^* & \text{if } f(x^*) = z^* \\ +\infty & \text{otherwise} \end{cases} \tag{4.12}$$

$$\overline{L}_\nu\,(d) = \begin{cases} \mathrm{E}\,(\max \xi) - m(x^*) & \text{if } \sigma(x^*) = 0 \text{ and } m(x^*) = z^* \\ +\infty & \text{otherwise.} \end{cases} \tag{4.13}$$

Then, $\bar{L}_\nu$ is minimized at $d^* = (x^*, m(x^*))$, with $x^* \in \operatorname{argmax}_{\sigma=0} m$ so that $\inf_{d\in\mathbb{D}} \bar{L}_\nu(d) = \mathcal{H}(\nu)$.

Contrary to the case of the previous three criteria, the loss function above is not regular, as we show in the next proposition.

Proposition 4.8. *The loss function given in* (4.12) *is not a regular loss function.*

Proof. Assume that L is a regular loss function. Then, for any sequential design of experiments (X_n), $\mathrm{E}_n(L_0(\xi))$ is a martingale bounded in L^1 so that it converges to $\mathrm{E}_\infty(L_0(\xi))$ a.s. Hence, from Proposition 2.8 and by assumption on $\mathcal{H}_1$, $\mathcal{H}(P_n^\xi)$ converges to $\mathcal{H}(P_\infty^\xi)$ a.s. Also, by the same martingale argument, $\mathrm{E}_n(\max \xi)$ converges to $\mathrm{E}_\infty(\max \xi)$ a.s. Hence, $\max_{\sigma_n(u)=0} m_n(u)$ converges to $\max_{\sigma_\infty(u)=0} m_\infty(u)$.

We now show that this last convergence does not hold for a certain choice of the Gaussian process ξ on $\mathbb{X} = [0, 1]$, which yields a contradiction. We consider ξ with mean $m : \mathbb{X} \to \mathbb{R}$, $u \mapsto m(u) = u$ and covariance function $k : \mathbb{X}^2 \to \mathbb{R}$, $(u, v) \mapsto k(u, v) = \exp(-(u - v)^2)$. We let (X_n) be a deterministic dense sequence on $[0, 1/3]$. Then, as follows from the proof of Proposition 1 in [52], we have $\sigma_\infty(u) = 0$ for all $u \in [0, 1]$. Hence, $\max_{\sigma_\infty(u)=0} m_\infty(u) = \max_{u\in[0,1]} \xi(u)$. Also, since $X_k \in [0, 1/3]$ for all $k \in \mathbb{N}$, we have $\max_{\sigma_n(u)=0} m_n(u) \leq \max_{u\in[0,1/3]} m_n(u)$. This upper bound converges to $\max_{u\in[0,1/3]} \xi(u)$ a.s. Hence we have a.s. $\max_{u\in[0,1/3]} \xi(u) = \max_{u\in[0,1]} \xi(u)$. This last equality can not hold because by symmetry $\max_{u\in[0,1/3]}(\xi(u) - u)$ and $\max_{u\in[2/3,1]}(\xi(u) - u)$ have the same distribution. □

Because of the previous proposition, we apply the more general results of Section 3 in order to show the consistency of the expected improvement. We first apply Proposition 3.19 to show that $\mathcal{H}$ is $\mathcal{M}$-measurable so that H_n is a well-defined random variable for any quasi-SUR strategy.

Proposition 4.9. *The conditions of Proposition 3.19 hold. As a consequence, $\mathcal{H}$ is $\mathcal{M}$-measurable.*

Proof. Proposition 3.19 i) holds since all the operations involved in L are $\mathcal{S} \times \mathcal{B}(\mathbb{D})$ measurable. In particular $f, x^* \mapsto f(x^*)$ is continuous (as is easy to see) and hence measurable. Proposition 3.19 ii) holds since m is a continuous function for all $\nu \in \mathbb{M}$. The set $\mathbb{D}_{L,\nu} = \{(x^*, z^*); |m(x^*) - z^*| + \sigma(x^*) = 0\}$ is clearly closed for all ν since σ and m are continuous. Also, from Lemma A.3, since $f, d \mapsto f(d)$ and $f, d \mapsto f(d)^2$ are $\mathbb{S} \times \mathcal{B}(\mathbb{D})$ measurable, the functions $\nu, d \mapsto m(d)$ and $\nu, d \mapsto \sigma(d)$ are $\mathcal{M} \otimes \mathcal{B}(\mathbb{D})$-measurable. Hence, the function $(\nu, x^*, z^*) \mapsto |m(x^*) - z^*| + \sigma(x^*)$ is $\mathcal{M} \otimes \mathcal{B}(\mathbb{D})$-measurable. Also, for all ν, $(x^*, z^*) \mapsto |m(x^*) - z^*| + \sigma(x^*)$ is continuous. Hence, from Lemma A.4, the set $\mathbb{D}_{L,\nu}$ is Effros-measurable with respect to $\mathcal{M}$. Hence Proposition 3.19 iii) holds. □

We now prove the consistency of the expected improvement strategy, based on applying Theorem 3.15, which shows that the maximum expected gain $\mathcal{G}(P_n^\xi)$ vanishes asymptotically.

Proposition 4.10. *For any quasi-SUR sequential design associated with $\mathcal{H}$, as $n \to \infty$, almost surely and in L^1, $\max m_n - M_n \to 0$, $H_n \to 0$ and $M_n \to \max \xi$.*

Proof. We have that (H_n) is an $\mathcal{F}_n$ supermartingale since $\sigma_n(.)$ decreases with n so that $\mathcal{H}$ has the supermartingale property. Hence Theorem 3.15 applies and $\mathcal{G}(P_n^\xi) \to^{a.s.}_{n\to\infty} 0$. Observe also that $\mathcal{G}(P_n^\xi) = \sup_{x\in\mathbb{X}} \mathrm{E}_n[M_{n+1,x} - M_n] \geq \sup_{x\in\mathbb{X}} \mathrm{E}_n[(\xi(x) - M_n)^+]$ so that a.s.

$$\max_{u\in\mathbb{X}} \gamma(m_n(u) - M_n, \sigma_n^2(u)) \to_{n\to\infty} 0, \tag{4.14}$$

with $\gamma(a,b) = \mathrm{E}((Z_{a,b}^+))$, where $Z_{a,b} \sim \mathcal{N}(a,b)$.

Recall from Section 3 in Vazquez and Bect [51] that γ is continuous and satisfies

- $\gamma(z, s^2) > 0$ if $s^2 > 0$,
- $\gamma(z, s^2) \geq z > 0$ if $z > 0$.

Note that, from Proposition 2.8 a.s. $\limsup_{n\to\infty} \max_{u\in\mathbb{X}} |m_n(u)| < +\infty$, and that $0 \leq \sigma_n(u) \leq \max_{u\in\mathbb{X}} \sigma(u) < +\infty$. Hence, from (4.14) and from the properties of γ, we have a.s. $\sigma_\infty(u) = 0$ for all $u \in \mathbb{X}$ and $\limsup_{n\to\infty} \max_{u\in\mathbb{X}}(m_n(u) - M_n) \leq 0$. Also, cleary $M_n \leq \max_{u\in\mathbb{X}}(m_n(u))$ so that a.s. $\max_{u\in\mathbb{X}} m_n(u) - M_n \to_{n\to\infty} 0$. Hence a.s. $H_n - (\mathrm{E}_n(\max\xi) - \max_{u\in\mathbb{X}} m_n(u)) \to_{n\to\infty} 0$. We also have a.s. $\mathrm{E}_n(\max\xi) - \max_{u\in\mathbb{X}} m_n(u)) \to_{n\to\infty} \mathrm{E}_\infty(\max\xi) - \max_{u\in\mathbb{X}} m_\infty(u)$, from Proposition 2.8 and the martingale convergence theorem (see, e.g., Theorem 7.23 in [28]). Also, a.s. $\mathrm{E}_\infty(\max\xi) - \max_{u\in\mathbb{X}} m_\infty(u) = 0$, since a.s. $\sigma_\infty(u) = 0$ for all $u \in \mathbb{X}$. Hence, we have shown that $H_n \to^{a.s.}_{n\to\infty} 0$. Also, since almost surely m_n converges uniformly to ξ (as $\sigma_\infty = 0$) we have $\max_{u\in\mathbb{X}} m_n(u) \to_{n\to\infty} \max\xi$ a.s. and so $M_n \to_{n\to\infty} \max\xi$ a.s.

We conclude the proof by observing that all three convergence results also hold in the L^1-sense by the dominated convergence theorem. □

Finally, we remark that Proposition 4.10 improves the consistency result of [51], since it does not impose the no-empty-ball property on the covariance function k. Hence, Proposition 4.10 also holds with very smooth Gaussian processes, or with Gaussian processes which sample paths have symmetry properties.

Appendix A: Proofs

A.1. Random element in $\mathcal{C}(\mathbb{X})$

Let $\mathbb{X}$ be a compact metric space and let $\mathbb{S} = \mathcal{C}(\mathbb{X})$ be the separable Banach space of all continuous functions on $\mathbb{X}$. Denote by $\mathcal{S}$ the Borel σ-algebra on $\mathbb{S}$.

Proposition A.1. *Let $(\xi_x)_{x\in\mathbb{X}}$ denote a stochastic process defined on $(\Omega, \mathcal{F}, \mathrm{P})$, indexed by $\mathbb{X}$. Assume that $(\xi_x)_{x\in\mathbb{X}}$ has continuous sample paths (i.e., that $(x \mapsto \xi_x(\omega)) \in \mathbb{S}$, for all $\omega \in \Omega$) and define $\xi : \Omega \to \mathbb{S}$, $\omega \mapsto \xi.(\omega)$. Then,*

i) ξ is a random element in $\mathbb{S}$ (i.e., it is $\mathcal{F}/\mathcal{S}$-measurable);
ii) if $(\xi_x)_{x\in\mathbb{X}}$ is a Gaussian process, P^ξ is a Gaussian measure on $(\mathbb{S}, \mathcal{S})$.

Proof. Let $\varphi \in \mathbb{S}'$ and let μ_φ denote the unique signed measure on $\mathbb{S}$ such that $\varphi(f) = \int_{\mathbb{X}} f \,\mathrm{d}\mu_\varphi$. Then

$$\varphi(\xi) : \ \omega \mapsto \int_{\mathbb{X}} \xi_x(\omega)\, \mu_\varphi(\mathrm{d}x)$$

is measurable by Fubini's theorem (because $(\omega, x) \mapsto \xi_x(\omega)$ is measurable). Since $\mathbb{S}'$ separates the points of $\mathbb{S}$ (Hahn-Banach theorem), and since $\mathbb{S}$ is a separable Banach space, we conclude from Proposition 1.10 of Vakhania et al. [48] that ξ is measurable, which proves i).

Assume now that ξ is Gaussian. Assume further, without loss of generality, that ξ is centered. Let H denote the closed linear span of $(\xi_x)_{x\in\mathbb{X}}$ in $L^2(\Omega, \mathcal{F}, \mathrm{P})$. For any $\varphi \in \mathbb{S}'$, we can write $\varphi(\xi) = \eta_1 + \eta_2$, with $\eta_1 \in H$ and $\eta_2 \perp H$. Then

$$\mathrm{var}(\eta_2) = \mathrm{E}(\eta_2\, \varphi(\xi)) = \mathrm{E}\left(\eta_2 \int_{\mathbb{X}} \xi_x\, \mu_\varphi(\mathrm{d}x)\right) = \int_{\mathbb{X}} \mathrm{E}(\eta_2\, \xi_x)\; \mu_\varphi(\mathrm{d}x) = 0.$$

Therefore $\varphi(\xi) \in H$, which proves ii) since H is a Gaussian space. □

Remark A.2. Since $\mathcal{C}(\mathbb{X})$ is a Polish space, all Borel measures are actually Radon measures, and therefore the Gaussian probability measures that we are dealing with are actually Radon Gaussian measures, as studied in depth by Bogachev [6, Chapter 3].

A.2. Measurability results

Lemma A.3. *Let $(\mathbb{E}, \mathcal{E})$ denote a measurable space. Let $\varphi : \mathbb{S} \times \mathbb{E} \to [0, +\infty]$ denote an $\mathcal{S} \otimes \mathcal{E}$-measurable function. Then the function $\mathbb{M} \times \mathbb{E} \to [0, +\infty]$, $(\mathrm{P}, v) \mapsto \int \varphi(f, v)\, \mathrm{P}(\mathrm{d}f)$, is $\mathcal{M} \otimes \mathcal{E}$-measurable.*

Proof of Lemma A.3. The result is clear for any $\varphi = \mathbb{1}_{A\times B}$, with $A \in \mathcal{S}$ and $B \in \mathcal{E}$. Indeed, $\int \varphi(f, v)\, \mathrm{P}(\mathrm{d}f) = \pi_A(\mathrm{P})\, \mathbb{1}_B(v)$, where π_A denotes the evaluation map $\mathrm{P} \mapsto \mathrm{P}(A)$, and the restriction of π_A to $\mathbb{M}$ is $\mathcal{M}$-measurable. It can be extended to any $\varphi = \mathbb{1}_\Gamma$, with $\Gamma \in \mathcal{S} \otimes \mathcal{E}$, using a standard monotone class argument, and then to any $\mathcal{S} \otimes \mathcal{E}$-measurable function by linearity and increasing approximation by simple functions. □

Lemma A.4. *Assume that $\mathbb{D}$ is a locally compact metric space and that there exists a measurable function $\psi : \mathbb{M} \times \mathbb{D} \to \mathbb{R}$ such that, for all ν, $\psi(\nu, \cdot)$ is continous and let $\mathbb{D}_\nu = \{\psi(\nu, \cdot) = 0\}$. Then $\nu \mapsto \mathbb{D}_\nu$ is Effros-measurable with respect to $\mathcal{M}$.*

Proof. Level sets of measurable processes with continuous paths are random closed sets in locally compact spaces [34, Example 1.2.iv, page 4]. The reason is that, in locally compact spaces, Effros-measurability can be tested using compact sets instead of open sets [34, Section 2.1, page 26], and the infinimum and supremum of a continuous process over a compact set are measurable (because compact subsets of a metric space are separable). □

In the following lemma, the Banach space $\mathcal{C}(\mathbb{X} \times \mathbb{X})$ is endowed with its Borel σ-algebra.

Lemma A.5. *The mappings $m_\bullet : \mathbb{M} \to \mathbb{S}$, $\nu \mapsto m_\nu$ and $k_\bullet : \mathbb{M} \to \mathcal{C}(\mathbb{X}\times\mathbb{X})$, $\nu \mapsto k_\nu$ are measurable.*

Proof. The mapping $m_\bullet$ is measurable if, and only if, $\nu \mapsto \varphi(m_\nu)$ is measurable for all $\varphi \in \mathbb{S}'$ [see, e.g., 48, Theorem 2.2]. Let $\varphi \in \mathbb{S}'$: there exists a unique signed measure μ_φ on $\mathbb{X}$ such that $\varphi(f) = \int_{\mathbb{X}} f\, \mathrm{d}\mu_\varphi$. It is then easy to check with Fubini's theorem that $\varphi(m_\nu) = \int \varphi(f)\, \nu(\mathrm{d}f)$, and the conclusion follows from Lemma A.3. The measurability of $k_\bullet$ is established in a similar way, working on $\mathbb{X} \times \mathbb{X}$ instead of $\mathbb{X}$. □

Let $\Theta \subset \mathbb{S} \times \mathcal{C}(\mathbb{X} \times \mathbb{X})$ denote the range of $\Psi = (m_\bullet, k_\bullet)$, and let $\mathcal{T}$ denote the trace on Θ of the Borel σ-algebra on $\mathbb{S} \times \mathcal{C}(\mathbb{X} \times \mathbb{X})$.

Lemma A.6. *Ψ is a bi-measurable mapping from $(\mathbb{M}, \mathcal{M})$ to $(\Theta, \mathcal{T})$.*

Proof. The measurability of Ψ follows from Lemma A.5. Since $\mathcal{M}$ is generated by the evaluation maps (see Section 2.2), Ψ^{-1} is measurable if, and only if, $(m,k) \mapsto [\mathcal{GP}(m,k)](A)$ is measurable for all $A \in \mathcal{S}$. This is easily checked for any finite intersection of the form $A = \cap_k \{f \in \mathbb{S} \mid f(x_k) \in \Gamma_k\}$, where $(x_k) \in \mathbb{X}^n$ and $(\Gamma_k) \in \mathcal{B}(\mathbb{R})^n$. The result extends to the ball σ-algebra $\mathcal{S}_0$ using a standard monotone class argument, which concludes the proof since $\mathcal{S}_0 = \mathcal{S}$ [see, e.g., 4]. □

A.3. The conditioning operator

Let $\underline{Z}_n = (Z_1, \ldots, Z_n)$ and $\underline{X}_n = (X_1, \ldots, X_n)$. For any $(m,k) \in \Theta$, $\underline{x}_n \in \mathbb{X}^n$ and $\underline{z}_n \in \mathbb{R}^n$, it is well-known that conditional mean and covariance functions of $(\xi(x))_{x\in\mathbb{X}}$ given $\underline{Z}_n = \underline{z}_n$, assuming a deterministic design $\underline{X}_n = \underline{x}_n$ (see Section 2.1), are given by

$$m_n(\,\cdot\,; \underline{x}_n, \underline{z}_n) = m + k(\cdot, \underline{x}_n)\, K(\underline{x}_n)^\dagger\, (\underline{z}_n - m(\underline{x}_n)) \tag{A.1}$$

$$k_n(\,\cdot\,; \underline{x}_n) = k - k(\,\cdot, \underline{x}_n)\, K(\underline{x}_n)^\dagger\, k(\underline{x}_n, \cdot\,), \tag{A.2}$$

where $K(\underline{x}_n)^\dagger$ denotes the pseudo-inverse of $K(\underline{x}_n) = (k(x_i,x_j) + \tau(x_i)\delta_{i,j})_{1\le i,j\le n}$, $k(\underline{x}_n, \cdot\,)$ and the other notations should be self-explanatory.

Lemma A.7. *$\tilde{\kappa}_n : (\underline{x}_n, \underline{z}_n, (m,k)) \mapsto (m_n(\,\cdot\,; \underline{x}_n, \underline{z}_n),\, k_n(\,\cdot\,; \underline{x}_n))$ is a measurable mapping from $\mathbb{X}^n \times \mathbb{R}^n \times \Theta$ to Θ, where Θ is endowed with the σ-algebra $\mathcal{T}$ defined in the preceding section.*

Proof. First observe that for any $\underline{x}_n$, $k_n(\,\cdot\,; \underline{x}_n)$ is the covariance function of $\xi - m_n(\,\cdot\,; \underline{x}_n, \underline{Z}_n)$, which is a Gaussian process with continuous sample paths. Thus, $(m_n(\,\cdot\,; \underline{x}_n, \underline{z}_n),\, k_n(\,\cdot\,; \underline{x}_n))$ is indeed an element of Θ. The result then follows from the continuity of $(m,x) \mapsto m(x)$, $(k,x) \mapsto k(x,\,\cdot\,)$, and $(k,x,y) \mapsto k(x,y)$, and the measurability of $K \mapsto K^\dagger$ [43]. □

Proof of Proposition 2.5. Let $\kappa_n : \mathbb{X}^n \times \mathbb{R}^n \times \mathbb{M} \to \mathbb{M}$ denote the mapping defined by

$$\kappa_n(\underline{x}_n, \underline{z}_n, \nu) = \mathcal{GP}\left(m_n(\,\cdot\,; \underline{x}_n, \underline{z}_n),\, k_n(\,\cdot\,; \underline{x}_n)\right), \tag{A.3}$$

where $\nu = \mathcal{GP}(m,k) \in \mathbb{M}$. Observe that, using the notations introduced in the previous section, $\kappa_n(\underline{x}_n, \underline{z}_n, \nu) = \Psi^{-1}\left(\tilde{\kappa}_n(\underline{x}_n, \underline{z}_n, \Psi(\nu))\right)$: thus, it follows from Lemmas A.6 and A.7 that κ_n is measurable. Standard algebraic manipulations then show that

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

whence it is easy to prove recursively that $\mathrm{P}_n^\xi := \kappa_n\left(\underline{X}_n, \underline{Z}_n, \mathrm{P}^\xi\right)$ satisfies the property $\mathrm{E}\left(U\, \mathrm{P}_n^\xi(\gamma)\right) = \mathrm{E}\left(U\, \mathbb{1}_{\xi\in\Gamma}\right)$ for any sequential design (X_i), any $\mathcal{F}_n$-measurable U of the form $U = \Pi_{i=1}^n \varphi_i(Z_i)$ and any $\Gamma \in \mathcal{S}$ of the form $\Gamma = \cap_{j=1}^J \{\xi(\tilde{x}_j) \in \Gamma_j\}$, with $\tilde{x}_j \in \mathbb{X}$, $\Gamma_j \in \mathcal{B}(\mathbb{R})$, $1 \le j \le J$. The result extends to any $\mathcal{F}_n$-measurable U and any $\Gamma \in \mathcal{S}$ thanks to a monotone class argument, which proves that P_n^ξ is a conditional distribution of ξ given $\mathcal{F}_n$. Proposition 2.5 is thus established with $\mathrm{Cond}_{x_1,z_1,\ldots,x_n,z_n} = \kappa_n\left(\underline{x}_n, \underline{z}_n, \,\cdot\,\right)$.

□

Proposition A.8. *The mapping $(x,\nu) \mapsto \mathcal{J}_x(\nu)$ is $\mathcal{B}(\mathbb{X}) \otimes \mathcal{M}$-measurable.*

Proof. Observe that $\mathcal{J}_x(\nu)$ can be rewritten as

$$\mathcal{J}_x(\nu) = \int_{\mathbb{R}} \mathcal{H}\left(\kappa_1\left(x,\, m_\nu(x) + v\, s_\nu(x),\, \nu\right)\right) \phi(v)\, \mathrm{d}v, \tag{A.4}$$

where $s_\nu^2 = k_\nu(x,x) + \tau^2(x)$ and κ_1 is defined as in the proof of Proposition 2.5. Using Lemma A.6 and the measurability of κ_1, the integrand in the right-hand side of Equation (A.4) is easily seen to be a $\mathcal{B}(\mathbb{X})\otimes\mathcal{M}\otimes\mathcal{B}(\mathbb{R})$-measurable function of (x,ν,v). The result follows from Fubini's theorem. □

A.4. Convergence in $\mathbb{M}$

Proof of Proposition 2.8. Recall from Proposition 2.5 that the conditional distribution of ξ given $\mathcal{F}_n$ is of the form $\mathrm{P}_n^\xi = \mathcal{GP}(m_n, k_n)$. Moreover, ξ is a Bochner-integrable $\mathbb{S}$-valued random element: indeed, it is measurable by Proposition A.1, and $\|\xi\|_\infty$ is integrable (see, e.g., Theorem 2.9 in [1]). The conditional expectation $\mathrm{E}\left(\xi \mid \mathcal{F}_n\right)$ of ξ given $\mathcal{F}_n$ is thus well-defined as a $\mathbb{S}$-valued random element (since $\mathbb{S} = \mathcal{C}(\mathbb{X})$ is a separable Banach space; see, e.g., Theorem 5.1.12 in [47]) and is easily seen to coincide with m_n. As a consequence, it follows from Theorem 6.1.12 in [47] that m_n converges uniformly, almost surely and in $L^1\left(\Omega,\mathcal{F},\mathrm{P}\right)$, to $m_\infty \coloneqq \mathrm{E}\left(\xi \mid \mathcal{F}_\infty\right)$. The limit m_∞ is, by definition of the conditional expectation, an $\mathcal{F}_\infty$-measurable random element in $\mathbb{S}$.

Let us now prove that the sequence k_n converges uniformly to a continuous function k_∞. Since $\mathrm{P}_n^\xi = \mathrm{Cond}_{X_1,Z_1,\ldots,X_n,Z_n}(\mathrm{P}^\xi)$ by Proposition 2.5, and since the sequence of conditional covariance functions depends only on the design points X_i (not on the observed values Z_i), we can reduce without loss of generality to the case of a deterministic design ($X_i = x_i \in \mathbb{R}$, for all $i \in \mathbb{N}$) and consider the associated *deterministic* sequence (k_n). Let $\mu = \sum_{i=1}^p \mu_i \delta_{\tilde{x}_i}$ denote any finitely supported measure on $\mathbb{X}$, and let $\sigma_n^2(\mu) = \sum_{i,j=1}^p \mu_i\mu_j k_n(\tilde{x}_i,\tilde{x}_j)$ denote the conditional variance of $Z = \sum_{i=1}^p \mu_i \xi(\tilde{x}_i)$ given $\mathcal{F}_n$. Because Z and the observations are jointly Gaussian, the sequence $\left(\sigma_n^2(\mu)\right)_{n\geq 1}$ is decreasing and therefore converges to a limit $\sigma_\infty^2(\mu)$, for all μ. Thus,

$$k_n(x,y) = \frac{1}{4}\left(\sigma_n^2\left(\delta_x+\delta_y\right) - \sigma_n^2\left(\delta_x-\delta_y\right)\right) \xrightarrow[n\to\infty]{} \frac{1}{4}\left(\sigma_\infty^2\left(\delta_x+\delta_y\right) - \sigma_\infty^2\left(\delta_x-\delta_y\right)\right),$$

which proves convergence to a limit $k_\infty(x,y)$. Moreover, we have for any $x,y,x',y' \in \mathbb{X}$:

$$\begin{aligned} \left|k_n\left(x,y\right) - k_n\left(x',y'\right)\right| &\leq \sigma_n(\delta_x)\,\sigma_n\left(\delta_y-\delta_{y'}\right) + \sigma_n(\delta_{y'})\,\sigma_n\left(\delta_x-\delta_{x'}\right) &\text{(A.5)}\\ &\leq \sigma_0(\delta_x)\,\sigma_0\left(\delta_y-\delta_{y'}\right) + \sigma_0(\delta_{y'})\,\sigma_0\left(\delta_x-\delta_{x'}\right). &\text{(A.6)}\end{aligned}$$

Letting n go to $+\infty$ in the left-hand side, we conclude that k_∞ is continuous. To see that the convergence $k_n \to k_\infty$ is uniform, consider the sequence of functions $\mathbb{X}^2 \to \mathbb{R}$, $(x,y) \mapsto \sigma_n^2\left(\delta_x+\delta_y\right)$. This is a decreasing sequence of continuous functions, which converges pointwise to the continuous function $(x,y) \mapsto \sigma_\infty^2\left(\delta_x+\delta_y\right)$. Since $\mathbb{X}^2$ is compact, the convergence is uniform by Dini's first theorem. The same argument applies to $(x,y) \mapsto \sigma_n^2\left(\delta_x-\delta_y\right)$ and therefore to k_n by polarization.

Finally, let Q denote any conditional distribution of ξ given $\mathcal{F}_\infty$. We will prove that the $\mathcal{F}_\infty$-measurable random measure Q is almost surely a Gaussian measure. Let $x \in \mathbb{X}$ and let ϕ_x denote the (random) characteristic function of $\mathrm{Q}\circ\delta_x^{-1}$. It follows from Theorem 6.23 in Kallenberg [28] that, for all $u \in \mathbb{R}$, $\phi_x(u) = \mathrm{E}_\infty\left(e^{iu\xi(x)}\right) \overset{\text{a.s.}}{=} \lim_{n\to\infty} \mathrm{E}_n\left(e^{iu\xi(x)}\right)$. Since

$$\mathrm{E}_n\left(e^{iu\xi(x)}\right) = e^{ium_n(x)}\, e^{-\frac{1}{2}k_n(x,x)u^2} \xrightarrow[n\to\infty]{\text{a.s.}} e^{ium_\infty(x)}\, e^{-\frac{1}{2}k_\infty(x,x)u^2},$$

we conclude from the continuity of ϕ_x and Levy's theorem that $\mathrm{Q} \circ \delta_x^{-1} = \mathcal{N}(m_\infty(x), k_\infty(x,x))$ almost surely. The argument extends to any image measure of the form $\mathrm{Q} \circ h^{-1}$, with $h = (\delta_{y_1}, \ldots, \delta_{y_m})$. Considering first the case where the y_j's are taken in a countable dense subset of $\mathbb{X}$ and then using the continuity of the elements of $\mathbb{S}$, we conclude that there is an almost sure event $\Omega_0 \in \mathcal{F}_\infty$ such that, for $\omega \in \Omega_0$, $(\delta_x)_{x\in\mathbb{X}}$ is a Gaussian process defined on the probability space $(\mathbb{S}, \mathcal{S}, \mathrm{Q}(\omega,\cdot))$. Then it follows from Proposition A.1 (with $\Omega \equiv \mathbb{S}$, $\xi \equiv \mathrm{Id}$, etc.) that $\mathrm{Q}(\omega,\cdot)$ is a Gaussian measure for all $\omega \in \Omega_0$. Finally, letting

$$\mathrm{P}_\infty^\xi(\omega,\cdot) = \begin{cases} Q(w,\cdot) & \text{if } w \in \Omega_0, \\ \mathcal{GP}(0,0) & \text{otherwise,} \end{cases}$$

we have constructed an $\mathcal{F}_\infty$-measurable random element in $\mathbb{M}$ such that $\mathrm{P}_n^\xi \to \mathrm{P}_\infty^\xi$ a.s. for the topology introduced in Definition 2.6, thereby concluding the proof. □

Proof of Proposition 2.9. Let $\nu = \mathcal{GP}(m,k) \in \mathbb{M}$ and let $(x_j, z_j) \to (x_\infty, z_\infty)$ in $\mathbb{X} \times \mathbb{R}$. For any $j \in \mathbb{N} \cup \{+\infty\}$, we have $\mathrm{Cond}_{x_j,z_j}(\nu) = \mathcal{GP}(m_1(\,\cdot\,; x_j, z_j), k_1(\,\cdot\,; x_j)$, where m_1 and k_1 are given by Equations (A.1)–(A.2). It is then easy to check that $m_1(\,\cdot\,; x_j, z_j)$ and $k_1(\,\cdot\,; x_j)$ converge uniformly to $m_1(\,\cdot\,; x_\infty, z_\infty)$ and $k_1(\,\cdot\,; x_\infty)$, respectively, using the uniform continuity of k over $\mathbb{X} \times \mathbb{X}$ and the fact that $K \mapsto K^\dagger$ is continuous at any invertible matrix (a scalar in this case). □

A.5. Miscellaneous

Lemma A.9. *Let U, V and W be real-valued random variables such that*

1. *W is independent from (U,V),*
2. *V and W are Gaussian.*

If U is orthogonal to $L^2(V+W)$, then U is orthogonal to $L^2(V)$.

Remark A.10. The reverse implication is also true, but not needed in the paper.

Proof. Assume without loss of generality that U, V and W are centered. Assume further that U is not orthogonal to $L^2(V)$. Then, there exists a smallest integer k_0 such that $\mathrm{cov}(U, V^{k_0}) \neq 0$. Indeed, we would have otherwise $\mathrm{cov}(U, H_k(V)) = 0$ for all k, where H_k denotes the k^{th} Hermite polynomial, and thus U would be orthogonal to $L^2(V)$ since $(H_k(V))_{k\in\mathbb{N}}$ is an orthonormal basis of $L^2(V)$. Using that $\mathrm{cov}(U, V^k) = 0$ for all $k < k_0$, we have:

$$\mathrm{cov}\left(U, (V+W)^{k_0}\right) \;=\; \sum_{k=0}^{k_0} \binom{k_0}{k} \mathrm{E}\left(UV^k\right) \mathrm{E}\left(W^{k-k_0}\right) \;=\; \mathrm{E}\left(UV^{k_0}\right) \;\neq\; 0. \tag{A.7}$$

Therefore U is not orthogonal to $L^2(V+W)$, which concludes the proof by contraposition. □

Acknowledgements

Part of David Ginsbourger's contribution took place within framework of the "Bayesian set estimation under random field priors" project (Number 146354) funded by the Swiss National Science Foundation. The authors would like to thank Luc Pronzato for pointing out the important connection with De Groot's earlier work on uncertainty functionals, and Dario Azzimonti for proofreading.

References

[1] Azais, J.-M. and Wschebor, M. (2009). *Level sets and extrema of random processes and fields*. Wiley.

[2] Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2012). A framework for validation of computer models. *Technometrics*.

[3] Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22 (3):773–793.

[4] Billingsley, P. (1999). *Convergence of Probability Measures. Second edition*. John Wiley & Sons.

[5] Binois, M. (2015). *Uncertainty quantification on pareto fronts and high-dimensional strategies in bayesian optimization, with applications in multi-objective automotive design*. PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne.

[6] Bogachev, V. I. (1998). *Gaussian measures*. American Mathematical Society.

[7] Brezis, H. (2010). *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media.

[8] Bull, A. D. (2011). Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, 12:2879–2904.

[9] Chevalier, C. (2013). *Fast uncertainty reduction strategies relying on Gaussian process models*. PhD thesis, University of Bern.

[10] Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.

[11] Cohn, D. A., Ghahramani, Z., and Jordan, M. I. (1996). Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145.

[12] DeGroot, M. H. (1962). Uncertainty, information, and sequential experiments. *The Annals of Mathematical Statistics*, 33(2):404–419.

[13] Diaconis, P. (1988). Bayesian numerical analysis. *Statistical Decision Theory and Related Topics IV*, 1:163–175.

[14] Emmerich, M., Giannakoglou, K., and Naujoks, B. (2006). Single-and multiobjective optimization assisted by Gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10 (4).

[15] Feliot, P., Bect, J., and Vazquez, E. (2016). A Bayesian approach to constrained single- and multi-objective optimization. *Journal of Global Optimization (in press)*.

[16] Forrester, A., Sobester, A., and Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*. Wiley.

[17] Frazier, P. (2009). *Knowledge-Gradient Methods for Statistical Learning*. PhD thesis, Princeton University.

[18] Frazier, P. I., Powell, W. B., and Dayanik, S. (2008). A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439.

[19] Frazier, P. I., Powell, W. B., and Dayanik, S. (2009). The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21(4):599–613.

[20] Geman, D. and Jedynak, B. (1996). An active testing model for tracking roads in satellite images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1):1–14.

[21] Ginsbourger, D., Baccou, J., Chevalier, C., Garland, N., Perales, F., and Monerie, Y. (2014). Bayesian adaptive reconstruction of profile optima and optimizers. *SIAM/ASA Journal on Un-*

certainty Quantification, 2(1):490–510.

[22] Gramacy, R., Gray, G., Le Digabel, S., Lee, H., Ranjan, P., Wells, G., and Wild, S. (2016). Modeling an augmented Lagrangian for blackbox constrained optimization. *Technometrics (with discussion)*, 58(1):1–11.

[23] Grünewälder, S., Audibert, J.-Y., Opper, M., and Shawe-Taylor, J. (2010). Regret bounds for gaussian process bandit problems. In *International Conference on Artificial Intelligence and Statistics*.

[24] Hennig, P., Osborne, M. A., and Girolami, M. (2015). Probabilistic numerics and uncertainty in computations. In *Proc. R. Soc. A*, volume 471, page 20150142. The Royal Society.

[25] Hiai, F. and Umegaki, H. (1977). Integrals, conditional expectations, and martingales of multivalued functions. *Journal of Multivariate Analysis*, 7(1):149–182.

[26] Ibragimov, I. and Rozanov, Y. (1978). *Gaussian Random Processes*. Springer-Verlag, New York.

[27] Jones, D., Schonlau, M., and Welch, W. (1998). Efficient global optimization of expensive black box functions. *Journal of Global Optimization*, 13:455–492.

[28] Kallenberg, O. (2002). *Foundations of Modern Probability. Second edition*. Springer-Verlag.

[29] Koehler, J., Puhalskii, A. A., and Simon, B. (1998). Estimating functions evaluated by simulation: A bayesian/analytic approach. *Annals of Applied Probability*, 8(4):1184–1215.

[30] Ledoux, M. and Talagrand, M. (2013). *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media.

[31] Lukić, M. and Beder, J. (2001). Stochastic processes with sample paths in reproducing kernel hilbert spaces. *Transactions of the American Mathematical Society*, 353(10):3945–3969.

[32] MacKay, D. J. C. (1992). Information-based objective functions for active data selection. *Neural computation*, 4(4):590–604.

[33] Mockus, J. B., Tiesis, V., and Žilinskas, A. (1978). The application of Bayesian methods for seeking the extremum. In Dixon, L. C. W. and Szegö, G. P., editors, *Towards Global Optimization*, volume 2, pages 117–129, North Holland, New York.

[34] Molchanov, I. (2006). *Theory of random sets*. Springer Science & Business Media.

[35] O'Hagan, A. (1991). Bayes–hermite quadrature. *Journal of statistical planning and inference*, 29(3):245–260.

[36] Perlman, M. D. (1974). Jensen's inequality for a convex vector-valued function on an infinite-dimensional space. *Journal of Multivariate Analysis*, 4(1):52–65.

[37] Picheny, V. (2014). A stepwise uncertainty reduction approach to constrained global optimization. In *Proceedings of the* 17^{th} *International Conference on Artificial Intelligence and Statistics (AISTATS)*.

[38] Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R., and Kim, N.-H. (2010). Adaptive designs of experiments for accurate approximation of target regions. *Journal of Mechanical Design*, 132(7).

[39] Ranjan, P., Bingham, D., and Michailidis, G. (2008). Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541.

[40] Ritter, K. (2000). *Average-case analysis of numerical problems*, volume 1733 of *Lecture Notes in Mathematics*. Springer Verlag.

[41] Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, 4:409–423.

[42] Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.

[43] Schönfeld, P. (1973). A note on the measurability of the pseudo-inverse. *Journal of Econometrics*, 1(3):313–314.

[44] Scott, W., Frazier, P., and Powell, W. (2011). The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression. *SIAM Journal on Optimization*, 21:996.

[45] Shahriari, B., Swersky, K., Wang, Z., Adams, R., and de Freitas, N. (2016). Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175.

[46] Srinivas, N., A., K., Kakade, S., and Seeger, M. (2012). Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58:3250–3265.

[47] Stroock, D. W. (2010). *Probability Theory: An Analytic View. Second Edition.* Cambridge University Press.

[48] Vakhania, N., Tarieladze, V., and Chobanyan, S. (1987). *Probability distributions on Banach spaces*, volume 14. Springer Science & Business Media.

[49] van der Vaart, A. W., van Zanten, J. H., et al. (2008). Reproducing kernel hilbert spaces of gaussian priors. In *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, pages 200–222. Institute of Mathematical Statistics.

[50] Vazquez, E. and Bect, J. (2009). A sequential Bayesian algorithm to estimate a probability of failure. In *15th IFAC Symposium on System Identification (SYSID 2009)*.

[51] Vazquez, E. and Bect, J. (2010a). Convergence properties of the expected improvement algorithm with fixed mean and covariance functions. *Journal of Statistical Planning and inference*, 140(11):3088–3095.

[52] Vazquez, E. and Bect, J. (2010b). Pointwise consistency of the kriging predictor with known mean and covariance functions. In *mODa 9–Advances in Model-Oriented Design and Analysis*, pages 221–228. Springer.

[53] Villemonteix, J., Vazquez, E., and Walter, E. (2009). An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534.

[54] Wang, H., Lin, G., and Li, J. (2016). Gaussian process surrogates for failure detection: a Bayesian experimental design approach. *Journal of Computational Physics*, 313:247–259.

[55] Williams, B. J., Santner, T. J., and Notz, W. I. (2000). Sequential design of computer experiments to minimize integrated response functions. *Statistica Sinica*, 10:1133–1152.

[56] Yarotsky, D. (2013). Univariate interpolation by exponential functions and gaussian RBFs for generic sets of nodes. *Journal of Approximation Theory*, 166:163–175.

[57] Zuluaga, M., Krause, A., Sergent, G., and Püschel, M. (2013). Active learning for level set estimation. In *International Joint Conference on Artificial Intelligence (IJCAI)*.