HAL
open science

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. Bernoulli, 2019, 25 (4A), pp.2883-2919. 10.3150/18-bej1074 . hal-01351088v4

HAL Id: hal-01351088

https://hal.science/hal-01351088v4

Submitted on 13 Sep 2018

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: 1608.01118

A supermartingale approach to Gaussian process based sequential design of experiments

Julien Bect[*,1], **François Bachoc**[2] **and David Ginsbourger**[3,4]

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

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

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

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

[*] *Corresponding author.*

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 study the consistency of an important class of sequential designs, 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. This observation enables us to establish generic consistency results for a broad class of SUR strategies. The consistency of several popular sequential design strategies is then obtained by means of this general result. Notably, we establish the consistency of two SUR strategies proposed by Bect, Ginsbourger, Li, Picheny and Vazquez (Stat. Comp., 2012)—to the best of our knowledge, these are the first proofs of consistency for GP-based sequential design algorithms dedicated to the estimation of excursion sets and their measure. We also establish a new, more general proof of consistency for the expected improvement algorithm for global optimization which, unlike previous results in the literature, 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* [39, 52, 53], where Gaussian process

models are massively used to define efficient sequential designs 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. The definition, calculation and optimization of sampling criteria tailored to various application goals have inspired a significant number of research contributions in the last decades [see, e.g., 3, 5, 12–14, 22, 24–26, 28, 30, 48–50, 55, 56, 66]. Yet, available convergence results for the associated sequential designs are quite heterogeneous in terms of their respective extent and underlying hypotheses [11, 31, 55, 57, 63, 65]. Here we develop a probabilistic approach to the analysis of a large class of strategies. This enables us to establish generic consistency 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 the key property—which will be referred to as the *supermartingale property*—that, for any sequential design, the sequence of random variables produced by these functionals is a supermartingale with respect to the filtration generated by the observations.

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. [43] and the considerable impact of the work of Jones et al. [36], EI and other Bayesian optimization strategies have spread in a variety of application fields. They are now commonly used in engineering design [23] and, in the field of machine learning, for automatic configuration algorithms (see [56] and references therein). Extensions to constrained, multi-objective and/or robust optimization constitute an active field of research [see, e.g., 8, 21, 22, 30, 48, 68]. 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, 5, 6, 13, 28, 49, 50, 62, 67, 71].

More specifically, we consider in this paper sequential design strategies built according to the *stepwise uncertainty reduction* (SUR) paradigm [see 5, 12, 66, and references therein]. Our main focus is the consistency of these algorithms under the assumption that the function of interest is a sample path of the Gaussian process model that is used to construct the sequential design. Almost sure consistency has been proved for the EI algorithm in [63], 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 consistency 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. [5]. The scheme of proof developed in this work allows us to address the excursion case and also to revisit the consistency of the knowledge gradient algorithm [24–26], as well as that of the EI algorithm—which can also be seen as a particular case of a SUR strategy [12]—without requiring the NEB assumption. Before outlining the paper in more detail, let us briefly introduce the 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

[1] On a related note, Bull [11] proves an upper-bound for the convergence rate of the expected improvement algorithm under the assumption that the covariance function is Hölder, but his result only holds for functions that belong to the reproducing kernel Hilbert space (RKHS) of the covariance—a condition which, under appropriate assumptions, is almost surely not satisfied by sample paths of the Gaussian process according to Driscoll's theorem [41]. Another result in the same vein is provided by Yarotsky [70] for the squared exponential covariance in the univariate case, assuming that the objective function is analytic in a sufficiently large complex domain around its interval of definition.

used with other types of models [see, e.g., 14, 27, 35, 38, 42, 45].

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 (observations) $Z_n = \xi(X_n) + \epsilon_n$ are to be made, sequentially, in order to estimate some quantity of interest (e.g., the maximum of ξ, or the volume of its excursion above some given threshold). 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 centered 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. Assuming that the H_n's are $\mathcal{F}_n$-measurable random variables, a SUR sampling criterion is 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 quantifies 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 greedily choosing at each step a point that provides the smallest expected residual uncertainty—equivalently, the largest expected uncertainty reduction—that is,

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

Such a greedy strategy is sometimes called myopic or one-step look-ahead (as opposed to a Bayes-optimal strategy, which would consider the reduction of uncertainty achieved at the end of the sequential design, i.e., when the entire experimental budget has been spent). Our goal is to establish the consistency of these strategies, where consistency means that the residual uncertainty H_n goes almost surely to zero.

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 [5] is the integrated indicator variance $H_n = \mathcal{H}\left(\mathrm{P}_n^\xi\right) = \int_{\mathbb{X}} p_n\,(1-p_n)\,\mathrm{d}\mu$ (also called *integrated Bernoulli variance* in what follows) where $p_n(x) = \mathrm{P}_n\left(\xi(x) \geq T\right)$ and P_n denotes the conditional probability with respect to $\mathcal{F}_n$. Note that $p_n(x)(1-p_n(x)) = \mathrm{var}_n(\mathbb{1}_{\xi(x)\geq T})$, where var_n denotes the conditional variance with respect to $\mathcal{F}_n$. Recalling the definition of J_n from (1.2), and using the law of total variance, we obtain that

$$J_n(x) = \int_{\mathbb{X}} \mathrm{E}_{n,x}\left(\mathrm{var}_{n+1}\left(\mathbb{1}_{\xi(u)\geq T}\right)\right)\mu(\mathrm{d}u) \leq \int_{\mathbb{X}} \mathrm{var}_n\left(\mathbb{1}_{\xi(u)\geq T}\right)\mu(\mathrm{d}u) = H_n, \tag{1.4}$$

which shows that (H_n) is an $(\mathcal{F}_n)$-supermartingale. Another related measure of uncertainty, for which a semi-analytical formula is provided in [13], is the *variance of the excursion volume*, $H_n = \mathrm{var}_n(\mu(\{x \in \mathbb{X} : \xi(x) \geq T\}))$. The supermartingale property follows again in this case from the law of total variance. In the optimization case, on the other hand, it turns out [see, e.g., 12, Section 3.3] that the EI criterion is underlaid by the measure of residual uncertainty $H_n = \mathrm{E}_n\left(\max\xi - M_n\right)$, where M_n is defined as $M_n = \max_{i\leq n}\xi(X_i)$ for non-degenerate Gaussian processes (see Remark 4.10), and E_n denotes the conditional expectation with respect to $\mathcal{F}_n$. A similar construction can be obtained for the *knowledge gradient*, as developed later. It turns out as shown later in the paper that,

for both criteria, the associated measures of residual uncertainty also possess the aforementioned supermartingale property.

It must be pointed out here that, under very weak assumptions about the Gaussian process model and the uncertainty functional, consistency can also be achieved more simply using any dense ("space filling") deterministic sequence of design points. It has been largely demonstrated, however, that SUR strategies typically outperform in practice these simple deterministic designs (see references above). Hence there remains a gap between theory and practice that is not filled by our consistency results since, by themselves, they do not provide a very strong theoretical support for the choice of SUR sequential designs over other types of designs, and in particular over non-sequential designs.

The main practical interest of our consistency results is rather to answer the natural concern that SUR strategies, because of their greedy nature, might fail to be consistent in some situations. Such a concern is justified for instance by the explicit counterexample, provided by Yarotsky [69], of a particular function for which the sequence of points generated by the EI strategy fails to produce a consistent estimate of the optimum of the function. Our results show that such functions are negligible under the distribution of the Gaussian process used to construct the sequential design. Further study of the convergence *rate* of SUR sequential designs is nevertheless needed to provide a full theoretical support for their practical effectiveness, and will be the subject of future work. Understanding their consistency, in relation with the properties of the uncertainty functionals that defined them, is an important first step in this direction.

The rest of the paper is structured as follows. Section 2 defines more precisely the statistical model and design problem considered in the paper, and addresses properties of conditioning and convergence of Gaussian measures that are instrumental in proving the main results of the paper. Section 3 discusses uncertainty functionals and their properties, and formulates general sufficient conditions for the consistency of SUR sequential designs in terms of properties of the associated uncertainty functionals. Finally, Section 4 presents applications of the general result to four popular examples of SUR sequential designs, establishing in each case 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 processes and sequential design

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 heteroscedastic Gaussian noise:

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

where $\tau : \mathbb{X} \to [0, +\infty)$ gives the (known) standard deviation $\tau(x)$ of an observation at $x \in \mathbb{X}$, and $(U_i)_{i\geq 1}$ denotes a sequence of independent and identically distributed $\mathcal{N}(0, 1)$ variables, independent of 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) τ *is continuous.*

Remark 2.3. Note 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. 20, 33, 46, 51] and in the design and analysis of deterministic computer experiments [see, e.g., 4, 52, 53].

Remark 2.4. A Gaussian process with continuous sample paths automatically has continuous mean and covariance functions (see, e.g., Lemma 1 in [34]). Conversely, assuming continuity of the mean function, let us recall a classical sufficient condition for sample path continuity on $\mathbb{X} \subset \mathbb{R}^d$ [see, e.g., 1, Theorem 3.4.1]: if there exist $C > 0$ and $\eta > 0$ such that

$$k(x,x) + k(y,y) - 2k(x,y) \leq \frac{C}{|\log\|x-y\||^{1+\eta}}, \qquad \forall x, y \in \mathbb{X},$$

then there exists a version of ξ with continuous sample paths. This is a very weak condition, which is satisfied by all commonly used continuous covariance functions on $\mathbb{R}^d$ (e.g., geometrically anisotropic or tensor-product Matérn covariance functions).

Remark 2.5. The setting described in this section arises, notably, when considering from a Bayesian point of view the following nonparametric 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., 2, 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 $\mathbb{S}$. The reader is referred to Vakhania et al. [60] and Ledoux and Talagrand [40] for background information concerning random elements and measures in Banach spaces, and to van der Vaart et al. [61] and Bogachev [9] 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 $\mathbb{S}$. The mean function m_ν and covariance function k_ν of this Gaussian process are continuous (see Remark 2.4) 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 $\mathbb{S}$—seen as transition kernels from Ω to $\mathbb{S}$—can be conveniently identified to random elements in $\mathbb{M}$ [see, e.g. 37, 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 Gaussian measures $\boldsymbol{\nu}$ such that $\boldsymbol{\nu} = \mathrm{P}\big(\xi \in \cdot \mid \mathcal{F}'\big)$ for some σ-algebra $\mathcal{F}' \subset \mathcal{F}$. Note that we use a bold letter $\boldsymbol{\nu}$ to denote a random element in $\mathbb{M}$ (i.e., a random Gaussian measure), and a normal letter ν to denote a point in the same space (i.e., a Gaussian measure). Not all conditional distributions of the form $\boldsymbol{\nu} = \mathrm{P}\big(\xi \in \cdot \mid \mathcal{F}'\big)$ are Gaussian, but an important class of such Gaussian conditional distributions is discussed in the following section and in Proposition 2.9.

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 combinations of pointwise evaluations, possibly corrupted by independent additive Gaussian noise (explicit expressions of the conditional mean and covariance functions are recalled in Appendix A.2). In the language of nonparametric Bayesian statistics (see Remark 2.5), 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 2.6. *For all $n \geq 1$, there exists a measurable mapping*

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

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

A proof of this result is provided in Appendix A.2. In the rest of the paper, we will denote by $\mathrm{P}_n^\xi = \mathcal{GP}(m_n, k_n)$ the conditional distribution $\mathrm{Cond}_{X_1,Z_1,\dots,X_n,Z_n}(\mathrm{P}^\xi)$ of ξ given $\mathcal{F}_n$, which can be seen as a random element in $(\mathbb{M}, \mathcal{M})$. The posterior mean m_n is also referred to as the *kriging predictor* [see, e.g., 16, 23, 58]. 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 $s_n^2(x) = k_n(x,x) + \tau^2(x)$.

2.4. Convergence in $\mathbb{M}$

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

Definition 2.7. *Let $\nu_n = \mathcal{GP}\big(m_n, k_n\big) \in \mathbb{M}$, $n \in \mathbb{N} \cup \{+\infty\}$. 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 2.8. 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 of this space through the injection $\nu \mapsto (m_\nu, k_\nu)$.

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

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 2.9. *For any Gaussian random element ξ in $\mathbb{S}$, defined on any probability space, and for any sequential design $(X_n)_{n\geq 1}$, the conditional distribution of ξ given $\mathcal{F}_\infty$ admits a version P^ξ_∞ which is an $\mathcal{F}_\infty$-measurable random element in $\mathbb{M}$, and $\mathrm{P}^\xi_n \to \mathrm{P}^\xi_\infty$ almost surely.*

Proposition 2.10. *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.3 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 conditional distribution P^ξ_n to a measure H_n of residual uncertainty for the quantity of interest. The minimal value of the uncertainty functional represents an absence of uncertainty on the quantity of interest: for clarity, and without loss of generality as long as $\mathcal{H}$ is bounded from below and attains its minimum, we will assume in the rest of this section that $\min \mathcal{H} = 0$, thus restricting our attention to non-negative uncertainty functionals.

More formally, let $\mathcal{H}$ denote a measurable functional from $\mathbb{M}$ to $[0, +\infty)$. 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. 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 Gaussian random element ξ in $\mathbb{S}$, defined on any probability space, and for any sequential design $(X_n)_{n\geq 1}$, the sequence $\left(\mathcal{H}(\mathrm{P}^\xi_n)\right)_{n\geq 0}$ is an $(\mathcal{F}_n)$-supermartingale.*

The supermartingale property echoes DeGroot's observation that "reasonable" measures of uncertainty should be decreasing on average for any possible experiment [17]. To discuss this connection more precisely in our particular setting, let us consider the following definition.

Definition 3.2. *Let $\mathbb{M}_0$ denote a set of probability measures on a measurable space $(\mathbb{E}, \mathcal{E})$, and let $\mathcal{M}_0$ denote the σ-algebra generated on $\mathbb{M}_0$ by the evaluation maps. For any random element $\boldsymbol{\nu}$ in $(\mathbb{M}_0, \mathcal{M}_0)$, defined on any probability space, let $\overline{\boldsymbol{\nu}}$ denote the probability measure defined by $\overline{\boldsymbol{\nu}}(A) = \mathrm{E}\left(\boldsymbol{\nu}(A)\right)$, $A \in \mathcal{E}$. We will say that a non-negative measurable functional $\mathcal{H}$ 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 $\overline{\boldsymbol{\nu}} \in \mathbb{M}_0$, $\mathrm{E}\left(\mathcal{H}(\boldsymbol{\nu})\right) \leq \mathcal{H}(\overline{\boldsymbol{\nu}})$.*

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 47, 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 to DoA—hence concave—functionals defined on some larger convex set of probability measures[3].

Remark 3.3. Let $\beta : \mathbb{S} \to \mathbb{R}^p$ denote a measurable function, and let ν^β denote the image of ν by β. Then it is easy to see that any functional of the form $\mathcal{H}(\nu) = \mathcal{H}'(\nu^\beta)$ is DoA, where $\mathcal{H}'$ denotes a DoA functional defined on some appropriate subset of the set of all probability measures on $\mathbb{R}^p$; the reader is referred to [32] for a variety of examples of such functionals. Section 4.2 provides an example of this construction, with $p = 1$ and $\mathcal{H}'$ the variance functional.

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

Proposition 3.4. *If $\mathcal{H}$ is DoA on $\mathbb{M}$, then it has the supermartingale property.*

Let us conclude this section with another useful property of functionals. Recall that $\mathfrak{P}(\xi)$ denotes the set of all Gaussian conditional distributions of ξ (see Section 2.2).

Definition 3.5. *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 $\mathbb{S}$, defined on any probability space, the family $(\mathcal{H}(\boldsymbol{\nu}))_{\boldsymbol{\nu}\in\mathfrak{P}(\xi)}$ is uniformly integrable.*

Proposition 3.6. *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 $|\mathcal{H}(\nu)| \le \int_{\mathbb{S}} L^+ \mathrm{d}\nu$ 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 $\boldsymbol{\nu} = \mathrm{P}\big(\xi \in \cdot \mid \mathcal{F}'\big) \in \mathfrak{P}(\xi)$. Then we have $|\mathcal{H}(\boldsymbol{\nu})| \le \mathrm{E}\left(L^+(\xi) \mid \mathcal{F}'\right)$, and the result follows from the uniform integrability of conditional expectations [see, e.g., 37, Lemma 5.5]. □

Remark 3.7. 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 $\{\mathrm{P}_n^\xi\} \subset \mathfrak{P}(\xi)$), and thus converges almost surely and in L^1.

3.2. SUR sequential designs and associated functionals

The SUR sampling criterion introduced informally as $J_n(x) = \mathrm{E}_{n,x}(H_{n+1})$ in (1.2) can now be more precisely defined as

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

where $Z_{n+1}(x) = \xi(x) + U_{n+1}\,\tau(x)$. A SUR sequential design is then 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:

Definition 3.8. *Let $\mathcal{H}$ denote a non-negative measurable functional on $\mathbb{M}$.*

i) We will say that (X_n) is a SUR sequential design *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 $n \ge n_0$, for some integer n_0.*

*ii) Given a sequence $\varepsilon = (\varepsilon_n)$ of non-negative real numbers such that $\varepsilon_n \to 0$, we will say that (X_n) is an ε-*quasi-SUR *sequential design if it is a sequential design such that $J_n(X_{n+1}) \le \inf J_n + \varepsilon_n$ for all $n \ge n_0$, for some integer n_0.*

[3] More precisely, the functionals discussed in Sections 4.1 and 4.2 can be extended to DoA functionals on the set of all probability measures on $\mathbb{S}$, and those of Sections 4.3 and 4.4 to DoA functionals on the set of all probability measures ν on $\mathbb{S}$ such that $\mathrm{E}(\max\xi - \min\xi) < +\infty$ for $\xi \sim \nu$.

Remark 3.9. In practice it is not always easy to guarantee that, for a given uncertainty functional $\mathcal{H}$, the sampling criteria J_n attain their infimum over $\mathbb{X}$. Moreover, the actual minimization of J_n is typically carried out by means of a numerical optimization algorithm, which cannot be expected to provide the exact minimizer. For these reasons, it seems important to study the convergence of quasi-SUR designs, as introduced by Definition 3.8.ii, instead of the more restrictive case of (exact) SUR designs. General existence results for SUR and quasi-SUR designs, based on the measurable selection theorem for random closed sets, are provided in Appendix A.4.

Let us now introduce some useful functionals associated to a given (non-negative) uncertainty functional $\mathcal{H}$. First, observe that $J_n(x) = \mathcal{J}_x(\mathrm{P}_n^\xi)$, 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 $s_\nu^2(x) = k_\nu(x,x) + \tau^2(x)$ and ϕ 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.6), and it is easy to see that $\mathcal{H}$ has the supermartingale property if, and only if,

$$\mathcal{J}_x(\nu) \le \mathcal{H}(\nu), \quad \text{for all } x \in \mathbb{X} \text{ and } \nu \in \mathbb{M}. \tag{3.4}$$

Assuming that $\mathcal{H}$ has the supermartingale property, we will then denote by $\mathcal{G}_x : \mathbb{M} \to [0, +\infty)$ the corresponding *expected gain functional* at x:

$$\mathcal{G}_x(\nu) = \mathcal{H}(\nu) - \mathcal{J}_x(\nu), \tag{3.5}$$

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

$$\mathcal{G}(\nu) = \sup_{x\in\mathbb{X}} \mathcal{G}_x(\nu). \tag{3.6}$$

Remark 3.10. Following [17], $\mathcal{G}_x$ could be called the "information" brought by an evaluation at x about the quantity of interest. This would be consistent with the usual definition of mutual information, when $\mathcal{H}$ is taken to be the Shannon entropy of some discrete quantity of interest [see, e.g., 15]. Note that DeGroot renamed it "expected information" in some of his subsequent work on this topic [see, e.g., 18, 19].

Remark 3.11. Alternatively, SUR sequential designs can be defined by the relation $X_{n+1} \in \operatorname{argmax} G_n$, where G_n denotes the sampling criterion $x \mapsto G_n(x) \coloneqq \mathcal{G}_x(\mathrm{P}_n^\xi) = H_n - J_n(x)$. In the particular cases discussed in Sections 4.3 and 4.4, G_n corresponds to the *knowledge gradient* and *expected improvement* criteria, respectively.

3.3. General convergence results

Denote by $\mathbb{Z}_\mathcal{H}$ and $\mathbb{Z}_\mathcal{G}$ the subsets of $\mathbb{M}$ where the functionals $\mathcal{H}$ and $\mathcal{G}$ vanish, respectively. The inclusion $\mathbb{Z}_\mathcal{H} \subset \mathbb{Z}_\mathcal{G}$ always hold: indeed, $0 \le \mathcal{J}_x \le \mathcal{H}$ for all x by (3.4), thus $0 \le \mathcal{G}_x \le \mathcal{H}$, and therefore $0 \le \mathcal{G} \le \mathcal{H}$. The reverse inclusion plays a capital role in the following result, which provides sufficient conditions for the almost sure convergence of quasi-SUR sequential designs associated with uncertainty functionals that enjoy the supermartingale property.

Theorem 3.12. *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) $H_n = \mathcal{H}(\mathrm{P}_n^\xi) \to \mathcal{H}(\mathrm{P}_\infty^\xi)$ *almost surely,*
ii) $\mathcal{G}(\mathrm{P}_n^\xi) \to \mathcal{G}(\mathrm{P}_\infty^\xi)$ *almost surely,*
iii) $\mathbb{Z}_\mathcal{H} = \mathbb{Z}_\mathcal{G}$,

then $H_n \to 0$ almost surely.

The proof of Theorem 3.12 relies on two main ideas. First, because the sequence (H_n) is a non-negative supermartingale, the conditional mean of its increments goes to zero almost surely, which implies by the quasi-SUR assumption that the maximal expected gain goes to zero as well. Second, using Assumptions i and ii, it is enough to study the limiting distribution P_∞^ξ: this is where the reverse inclusion $\mathbb{Z}_\mathcal{G} \subset \mathbb{Z}_\mathcal{H}$ is used to conclude that the uncertainty in the limiting distribution is zero.

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

$$\begin{aligned} J_n(X_{n+1}) &= \mathrm{E}_n\left(\mathcal{H}\left(\mathrm{Cond}_{x,Z_{n+1}(x)}(\mathrm{P}_n^\xi)\right)\right)_{|x=X_{n+1}} \\ &= \mathrm{E}_n\left(\mathcal{H}\left(\mathrm{Cond}_{X_{n+1},Z_{n+1}}(\mathrm{P}_n^\xi)\right)\right) = \mathrm{E}_n(H_{n+1}). \end{aligned} \tag{3.7}$$

Set $\Delta_{n+1} = H_n - H_{n+1}$ and $\overline{\Delta}_{n+1} = \mathrm{E}_n(\Delta_{n+1}) = H_n - \mathrm{E}_n(H_{n+1})$. 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 $n \geq n_0$:

$$\overline{\Delta}_{n+1} = H_n - \mathrm{E}_n(H_{n+1}) = H_n - J_n(X_{n+1}) \geq H_n - \inf_{x\in\mathbb{X}} J_n(x) - \varepsilon_n, \tag{3.8}$$

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

$$\overline{\Delta}_{n+1} \geq \sup_{x\in\mathbb{X}} \mathcal{G}_x(\mathrm{P}_n^\xi) - \varepsilon_n = \mathcal{G}(\mathrm{P}_n^\xi) - \varepsilon_n. \tag{3.9}$$

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

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

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, $\mathcal{G}(\mathrm{P}_n^\xi) \to 0$ almost surely, since $0 \leq \mathcal{G}(\mathrm{P}_n^\xi) \leq \overline{\Delta}_{n+1} + \varepsilon_n$.

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

Remark 3.13. Note that the conclusions of Theorem 3.12 still hold partially if it is only assumed that the condition $J_n(X_{n+1}) \leq \inf J_n + \varepsilon_n$ holds infinitely often, almost surely: in this case the conclusion of the first part of the theorem is weakened to $\liminf \mathcal{G}(\mathrm{P}_n^\xi) = 0$, but the final conclusion ($H_n \to 0$ a.s.) remains the same.

Since $\mathrm{P}_n^\xi \to \mathrm{P}_\infty^\xi$ almost surely by Proposition 2.9, Assumptions i and ii of Theorem 3.12 hold if $\mathcal{H}$ and $\mathcal{G}$, respectively, are continuous. Assuming $\mathcal{H}$ to be continuous, however, would be too strong a requirement, that some important examples would fail to satisfy. For instance, the uncertainty functional

$$\mathcal{H} : \nu \mapsto \int_{\mathbb{X}} p_\nu(1 - p_\nu)\,\mathrm{d}\mu \tag{3.10}$$

studied in Section 4.1, where $p_\nu(u) = \int_{\mathbb{S}} \mathbb{1}_{f(u)\geq T}\,\nu(\mathrm{d}f)$ for some threshold $T \in \mathbb{R}$, is clearly discontinuous at the degenerate measure $\nu = \mathcal{GP}\left(T\mathbb{1}_{\mathbb{X}}, 0\right)$. The following weaker notion of continuity will turn out to be suitable for our needs:

Definition 3.14. *A measurable functional $\mathcal{H}$ on $\mathbb{M}$ will be said to be* $\mathfrak{P}$-continuous *if, for any Gaussian random element ξ in $\mathbb{S}$, defined on any probability space, and any sequence of random measures $\boldsymbol{\nu}_n \in \mathfrak{P}(\xi)$ such that $\boldsymbol{\nu}_n \xrightarrow{\text{a.s.}} \boldsymbol{\nu}_\infty \in \mathfrak{P}(\xi)$, the convergence $\mathcal{H}(\boldsymbol{\nu}_n) \xrightarrow{\text{a.s.}} \mathcal{H}(\boldsymbol{\nu}_\infty)$ holds.*

Remark 3.15. The uncertainty functional (3.10) provides an explicit example of a functional which is $\mathfrak{P}$-continuous (cf. the proof of Theorem 4.1) but not continuous. The expected improvement functional, discussed in Section 4.4, provides an example of a functional which is not even $\mathfrak{P}$-continuous (see Proposition 4.11), but for which consistency can nonetheless be proved by a direct application of Theorem 3.12.

Checking that $\mathcal{G}$ is $\mathfrak{P}$-continuous, however, is not easy in practice. The following results provides sufficient conditions for Assumption 3.12.ii that are easier to check.

Theorem 3.16. *Let $\mathcal{H}$ denote a non-negative, measurable uncertainty functional on $\mathbb{M}$, and let $\mathcal{G}$ denote the associated maximal expected gain functional. Assume that $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$, where*

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

Then, for any quasi-SUR sequential design associated with $\mathcal{H}$, $\mathcal{G}\big(\mathrm{P}_\infty^\xi\big) = 0$ almost surely.

Proof. First, note that $\mathcal{H}_0\big(\mathrm{P}_n^\xi\big) = \mathrm{E}_n\left(L_0(\xi)\right)$. Thus, since $L_0 \in \mathcal{L}^1\left(\mathbb{S}, \mathcal{S}, \mathrm{P}^\xi\right)$, the sequence $\big(\mathcal{H}_0(\mathrm{P}_n^\xi)\big)$ is a uniformly integrable martingale (see, e.g., Kallenberg [37], Theorem 6.23), which converges almost surely and in L^1 to $\mathrm{E}_\infty\left(L_0(\xi)\right) = \mathcal{H}_0\big(\mathrm{P}_\infty^\xi\big)$. As a consequence, $\mathcal{H}\big(\mathrm{P}_n^\xi\big) \xrightarrow{\text{a.s.}} \mathcal{H}\big(\mathrm{P}_\infty^\xi\big)$ since $\mathcal{H}_1$ is $\mathfrak{P}$-continuous and $\mathrm{P}_n^\xi \xrightarrow{\text{a.s.}} \mathrm{P}_\infty^\xi$ by Proposition 2.9.

Let $x \in \mathbb{X}$. The functional $\mathcal{H}_0$ has the supermartingale property by the preceding argument, and therefore $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$ also has the supermartingale property. Then, it follows from the first part of Theorem 3.12 that $\mathcal{G}_x\big(\mathrm{P}_n^\xi\big) \xrightarrow{\text{a.s.}} 0$, and thus

$$\mathcal{J}_x\big(\mathrm{P}_n^\xi\big) = \mathcal{H}\big(\mathrm{P}_n^\xi\big) - \mathcal{G}_x\big(\mathrm{P}_n^\xi\big) \xrightarrow{\text{a.s.}} \mathcal{H}\big(\mathrm{P}_\infty^\xi\big). \tag{3.11}$$

Let $\mathrm{P}_{n,x}^\xi = \mathrm{Cond}_{x,Z(x)}\left(\mathrm{P}_n^\xi\right)$, 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:

$$\begin{aligned}\mathcal{J}_x\big(\mathrm{P}_n^\xi\big) &= \mathrm{E}_n\left(\mathcal{H}\big(\mathrm{P}_{n,x}^\xi\big) - \mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\right) + \mathrm{E}_n\left(\mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\right)\\ &= \mathrm{E}_n\left(\mathcal{H}_1\big(\mathrm{P}_{n,x}^\xi\big) - \mathcal{H}_1\big(\mathrm{P}_{\infty,x}^\xi\big)\right) + \mathrm{E}_n\left(\mathcal{H}\big(\mathrm{P}_{\infty,x}^\xi\big)\right)\end{aligned} \tag{3.12}$$

where the second equality simply follows from the fact that $\mathrm{E}_n\left(\mathcal{H}_0\big(\mathrm{P}_{n,x}^\xi\big)\right) = \mathrm{E}_n\left(\mathcal{H}_0\big(\mathrm{P}_{\infty,x}^\xi\big)\right) = \mathrm{E}_n\left(L_0(\xi)\right)$ by the law of total expectation. The second conditional expectation in (3.12) is, again,

a uniformly integrable martingale that converges almost surely and in L^1:

$$\mathrm{E}_n\left(\mathcal{H}\big(\mathrm{P}^\xi_{\infty,x}\big)\right) \xrightarrow[n\to\infty]{\text{a.s.},\,L^1} \mathrm{E}_\infty\left(\mathcal{H}\big(\mathrm{P}^\xi_{\infty,x}\big)\right). \tag{3.13}$$

Moreover, note that

$$\begin{aligned}\mathrm{P}^\xi_{n,x} &= \mathrm{Cond}_{X_1,\,Z_1,\,\ldots,\,X_n,\,Z_n,\,x,\,Z(x)}\left(\mathrm{P}^\xi_0\right)\\ &= \mathrm{Cond}_{x,\,Z(x),\,X_1,\,Z_1,\,\ldots,\,X_n,\,Z_n}\left(\mathrm{P}^\xi_0\right)\end{aligned}$$

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 $\mathrm{P}^\xi_{\infty,x}$ 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,

$$\mathrm{E}_n\left(\mathcal{H}_1\big(\mathrm{P}^\xi_{n,x}\big) - \mathcal{H}_1\big(\mathrm{P}^\xi_{\infty,x}\big)\right) \xrightarrow[n\to\infty]{L^1} 0 \tag{3.14}$$

since $\mathcal{H}_1$ is $\mathfrak{P}$-continuous and $\mathfrak{P}$-uniformly integrable. Combine (3.12), (3.13) and (3.14) to prove that $\mathcal{J}_x\big(\mathrm{P}^\xi_n\big) \to \mathrm{E}_\infty\big(\mathcal{H}\big(\mathrm{P}^\xi_{\infty,x}\big)\big)$ in L^1. Then, it follows from a comparison with (3.11) that $\mathcal{H}\big(\mathrm{P}^\xi_\infty\big) = \mathrm{E}_\infty\big(\mathcal{H}\big(\mathrm{P}^\xi_{\infty,x}\big)\big)$ almost surely, and therefore

$$\mathcal{G}_x\big(\mathrm{P}^\xi_\infty\big) = \mathcal{H}\big(\mathrm{P}^\xi_\infty\big) - \mathrm{E}_\infty\left(\mathcal{H}\big(\mathrm{P}^\xi_{\infty,x}\big)\right) = 0 \quad \text{almost surely.} \tag{3.15}$$

To conclude, note that by Assertion (a) of Theorem A.8 the sample paths of $J_\infty : x \mapsto \mathrm{E}_\infty\big(\mathcal{H}\big(\mathrm{P}^\xi_{\infty,x}\big)\big)$ are continuous on $\big\{x \in \mathbb{X} : s^2_\infty(x) > 0\big\}$. Let $\{x_j\}$ denote a countable dense subset of $\mathbb{X}$. We have proved that, almost surely, $\mathcal{G}_{x_j}\big(\mathrm{P}^\xi_\infty\big) = 0$ for all j. Using the continuity of J_∞ on $\big\{s^2_\infty > 0\big\}$, and the fact that $\mathcal{G}_x = 0$ on $\big\{s^2_\infty = 0\big\}$, we conclude that, almost surely, $\mathcal{G}_x\big(\mathrm{P}^\xi_\infty\big) = 0$ for all x, and therefore $\mathcal{G}\big(\mathrm{P}^\xi_\infty\big) = 0$, which concludes the proof. □

3.4. Uncertainty functionals based on a loss function

Let us now consider, more specifically, uncertainty functionals $\mathcal{H}$ defined in the form of a risk:

$$\mathcal{H}(\nu) = \inf_{d\in\mathbb{D}} \int_{\mathbb{S}} L(f,d)\,\nu(\mathrm{d}f) = \inf_{d\in\mathbb{D}} \overline{L}_\nu(d), \tag{3.16}$$

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 particular form.

The following result formalizes an important observation of DeGroot [17, 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 3.12).

Proposition 3.17. *Let $\mathcal{H}$ denote a measurable functional on $\mathbb{M}$. If $\mathcal{H}$ is of the form (3.16), 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 $d \in \mathbb{D}$) that commute with expectations: for any random element $\boldsymbol{\nu}$ in $\mathbb{M}$ and any $d \in \mathbb{D}$,

$$\mathrm{E}\left(\overline{L}_{\boldsymbol{\nu}}(d)\right) = \mathrm{E}\left(\int_{\mathbb{S}} L(f,d)\,\boldsymbol{\nu}(\mathrm{d}f)\right) = \overline{L}_{\overline{\boldsymbol{\nu}}}(d), \tag{3.17}$$

where $\overline{\boldsymbol{\nu}}$ is defined as in Definition 3.2. (In other words, the linear functionals $\nu \mapsto L_\nu(d)$ are DoA themselves, with an equality in (3.17) instead of the inequality in Definition 3.2.) □

An uncertainty functional of the form (3.16) 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.1). This is true, for instance, if $\mathbb{D}$ is separable and $d \mapsto \overline{L}_\nu(d)$ is continuous for all ν. See Proposition A.4 for an example where L is discontinuous.

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 3.18. *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 corollary is provided as a convenient summary of the results that hold for uncertainty functionals based on regular non-negative loss functions.

Corollary 3.19. *Let $\mathcal{H}$ denote a functional of the form (3.16) for some non-negative loss function L. If L is regular, then $\mathcal{H}$ is a measurable functional that satisfies the assumptions of Theorems A.8, 3.12 and 3.16. In particular, for any quasi-SUR design associated with $\mathcal{H}$, $H_n = \mathcal{H}\big(\mathrm{P}_n^\xi\big) \to 0$ almost surely.*

4. Applications to popular sequential design strategies

This section presents applications of our results to four popular sequential design strategies, two of them addressing the excursion case (Sections 4.1 and 4.2), and the other two addressing 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

Assume that $\mathbb{X}$ is endowed with a finite measure μ and let $T \in \mathbb{R}$ be a given excursion threshold. For any measurable function $f : \mathbb{X} \mapsto \mathbb{R}$, 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 [5]:

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

which corresponds to the uncertainty functional

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

where $p_\nu(u) = \int_{\mathbb{S}} \mathbb{1}_{f(u)\geq T}\, \nu(\mathrm{d}f)$. See [13] for more information on the computation of the corresponding SUR sampling criterion

$$J_n(x) = \mathrm{E}_{n,x}\left(\int_{\mathbb{X}} p_{n+1}\left(1 - p_{n+1}\right) \mathrm{d}\mu\right).$$

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

$$\begin{aligned} L : \mathbb{S} \times \mathbb{D} &\to [0, +\infty), \\ (f, d) &\mapsto \|\mathbb{1}_{\Gamma(f)} - d\|^2_{L^2(\mathbb{X})}, \end{aligned} \tag{4.3}$$

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 $\nu \in \mathbb{M}$ and $\xi \sim \nu$,

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

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 3.18. As a consequence, all the conclusions of Corollary 3.19 hold, and 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 [10]). Hence $\mathbb{D}$ is also separable.

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

Indeed, $f \mapsto \int_{\mathbb{X}} \left(\mathbb{1}_{f(x)\geq T} - d(x)\right)^2 \mu(\mathrm{d}x)$ 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 $\xi \sim \mathcal{GP}(m, k)$ and let $(\boldsymbol{\nu}_n)$ be a sequence of random measures $\boldsymbol{\nu}_n \in \mathfrak{P}(\xi)$ such that a.s. $\boldsymbol{\nu}_n \to \boldsymbol{\nu}_\infty \in \mathfrak{P}(\xi)$. For $n \in \mathbb{N}\cup\{\infty\}$, let m_n and k_n be the (random) mean and covariance functions of $\boldsymbol{\nu}_n$. For $u \in \mathbb{X}$ and $n \in \mathbb{N}\cup\{\infty\}$, let also $\sigma^2(u) = k(u,u)$, $\sigma_n^2(u) = k_n(u,u)$, and

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

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 $\bar{\Phi}(0/0) = 1$. We will prove below that, for all $n \in \mathbb{N}\cup\{+\infty\}$,

$$\mathcal{H}(\boldsymbol{\nu}_n) = \int_{\mathbb{X}} g_n(u)\,\mu(\mathrm{d}u) \overset{\text{a.s.}}{=} \int_A g_n(u)\,\mu(\mathrm{d}u), \tag{4.4}$$

where A denotes the random subset of $\mathbb{X}$ defined by

$$A(\omega) = \{u \in \mathbb{X} : \sigma(u) > 0, \sigma_\infty(\omega,u) = 0, m_\infty(\omega,u) \neq T\} \cup \{u \in \mathbb{X} :\ \sigma(u) > 0, \sigma_\infty(\omega,u) > 0\}.$$

The motivation for using (4.4) is that it is easy to prove the convergence of $g_n(\omega,u)$ for $u \in A(\omega)$ and that the set $A(\omega)$ does not depend on n, which makes it possible to conclude using the dominated convergence theorem on the set $A(\omega)$ for almost all ω. In more detail: since $\boldsymbol{\nu}_n \to \boldsymbol{\nu}_\infty$ almost surely, it holds for almost all $\omega \in \Omega$ that $m_n(\omega,\cdot) \to m_\infty(\omega,\cdot)$ and $\sigma_n(\omega,\cdot) \to \sigma_\infty(\omega,\cdot)$ uniformly on $\mathbb{X}$. Furthermore, for each $u \in A(\omega)$, either $\sigma_\infty(\omega,u) > 0$ or $\sigma_\infty(\omega,u) = 0, m_\infty(\omega,u) \neq T$. In both cases, we have that $g\left(\bar{\Phi}([m_n(\omega,u) - T]/\sigma_n(\omega,u))\right) \to g\left(\bar{\Phi}([m_\infty(\omega,u) - T]/\sigma_\infty(\omega,u))\right)$. So, for almost all $\omega \in \Omega$ we can apply the dominated convergence theorem on $A(\omega)$ and thus obtain that

$$\mathcal{H}(\boldsymbol{\nu}_n) = \int_A g_n(u)\,\mu(\mathrm{d}u) \xrightarrow[n\to\infty]{\text{a.s.}} \mathcal{H}(\boldsymbol{\nu}_\infty) = \int_A g_\infty(u)\,\mu(\mathrm{d}u),$$

which proves the claim.

Let us now prove (4.4). Observe first that, for any u such that $\sigma(u) = 0$, we have $\sigma_n(u) \overset{\text{a.s.}}{=} 0$ for all $n \in \mathbb{N}\cup\{\infty\}$ since $\boldsymbol{\nu}_n \in \mathfrak{P}(\xi)$. Hence, $g_n(u) \overset{\text{a.s.}}{=} 0$ when $\sigma(u) = 0$. [This is because of the convention $\bar{\Phi}(0/0) = 1$, which yields $g_n(u) = g(\mathbf{1}_{m_n(u)\geq T}) = 0$ when $\sigma_n(u) = 0$, regardless of whether $m_n(u) = T$ or not.] Thus, setting $B(\omega) = \{u \in \mathbb{X};\ \sigma(u) > 0,\ \sigma_\infty(\omega,u) = 0,\ m_\infty(\omega,u) = T\}$, we have for all $\omega \in \Omega$

$$\mathbb{X} = \{u \in \mathbb{X}, \sigma(u) = 0\} \cup A(\omega) \cup B(\omega),$$

and the three sets of the right-hand side of the previous display are disjoint. Since, as discussed above, $g_n(u) \overset{\text{a.s.}}{=} 0$ for any u such that $\sigma(u) = 0$, we obtain

$$\mathcal{H}(\boldsymbol{\nu}_n) \overset{\text{a.s.}}{=} \int_A g_n(u)\,\mu(\mathrm{d}u) + \int_B g_n(u)\,\mu(\mathrm{d}u).$$

Thus, in order to prove (4.4), it is sufficient to show that

$$\int_B g_n(u)\,\mu(\mathrm{d}u) \overset{\text{a.s.}}{=} 0. \tag{4.5}$$

We will now establish (4.5) by proving that, in fact, $\mu(B) = 0$ almost surely. First, since $(\omega, u) \mapsto m_\infty(\omega, u)$ and $(\omega, u) \mapsto \sigma_\infty(\omega, u)$ are jointly measurable (by continuity of $m_\infty(\omega, \cdot)$ and $\sigma_\infty(\omega, \cdot)$ for all $\omega \in \Omega$), it follows from the Fubini-Tonelli theorem that

$$\mathrm{E}(\mu(B)) = \int_{\mathbb{X}} \mathbb{1}_{\sigma(u)>0}\, \mathrm{E}(\mathbb{1}_{\sigma_\infty(u)=0} \mathbb{1}_{m_\infty(u)=T})\, \mu(\mathrm{d}u). \tag{4.6}$$

Then we have, for any $u \in \mathbb{X}$,

$$\begin{aligned}
\mathrm{E}\left(\mathbb{1}_{\sigma_\infty(u)=0}(\xi(u) - m_\infty(u))^2\right) &= \mathrm{E}\left(\mathrm{E}\left[\mathbb{1}_{\sigma_\infty(u)=0}(\xi(u) - m_\infty(u))^2 \middle| \mathcal{F}'_\infty\right]\right) \\
&= \mathrm{E}\left(\mathbb{1}_{\sigma_\infty(u)=0}\, \mathrm{E}\left[(\xi(u) - m_\infty(u))^2 \middle| \mathcal{F}'_\infty\right]\right) \\
&= \mathrm{E}\left(\mathbb{1}_{\sigma_\infty(u)=0}\, \sigma_\infty^2(u)\right) \\
&= 0,
\end{aligned}$$

where $\mathcal{F}'_\infty$ denotes the σ-algebra such that $\boldsymbol{\nu}_\infty = \mathrm{P}\left(\xi \in \cdot \mid \mathcal{F}'_\infty\right)$. Hence, the random variable $\mathbb{1}_{\sigma_\infty(u)=0}(\xi(u) - m_\infty(u))^2$ is almost surely zero, since it is non-negative and has a zero expectation. Thus, for any $u \in \mathbb{X}$, the implication

$$\sigma_\infty(u) = 0 \Longrightarrow \xi(u) = m_\infty(u)$$

holds almost surely. As a consequence, we have

$$\begin{aligned}
\mathbb{1}_{\sigma(u)>0}\, \mathbb{1}_{\sigma_\infty(u)=0} \mathbb{1}_{m_\infty(u)=T} &= \mathbb{1}_{\sigma(u)>0} \mathbb{1}_{\sigma_\infty(u)=0} \mathbb{1}_{\xi(u)=T} \\
&\leq \mathbb{1}_{\sigma(u)>0} \mathbb{1}_{\xi(u)=T} \\
&= 0
\end{aligned}$$

almost surely, since $\xi(u) \sim \mathcal{N}(0, \sigma(u)^2)$. Hence, the integrand in the right-hand side of (4.6) is zero, which implies that $\mathrm{E}\left(\mu(B)\right) = 0$, and therefore $\mu(B) \overset{\text{a.s.}}{=} 0$ since $\mu(B)$ is a non-negative random variable. Thus (4.4) holds and the proof of **e)** is complete.

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 the law of total variance

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

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 proved without difficulty by separating the cases of nullity and non-nullity of the denominator). Thus, 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 neighborhood 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,*

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

and

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

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

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

Also, for almost all $\omega \in \Omega$ and all $u \in A(\omega)$, $p_n(\omega, u) \to \mathbb{1}_{\xi(\omega,u)\geq T}$ 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 example of Section 4.1, we consider now the alternative measure of residual uncertainty $H_n = \mathrm{var}_n(\alpha(\xi))$ from [5, 13]; in other words, we consider the uncertainty functional

$$\mathcal{H}(\nu) = \int_{\mathbb{S}} \left(\alpha(f) - \overline{\alpha}_\nu\right)^2 \nu(\mathrm{d}f), \tag{4.7}$$

where $\overline{\alpha}_\nu = \int_{\mathbb{S}} \alpha \, \mathrm{d}\nu$. The corresponding sampling criterion is

$$J_n(x) = \mathrm{E}_{n,x}\big(\mathrm{var}_{n+1}\left(\alpha(\xi)\right)\big).$$

This uncertainty functional again derives from a loss function: indeed, $L(f, d) = (\alpha(f) - d)^2$ with $\mathbb{D} = \mathbb{R}$ leads to

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

where $\overline{L}_{\mathrm{P}_n^\xi}$ reaches its infimum for $d = \mathrm{E}_n(\alpha(\xi))$, and therefore $H_n = \inf_{d\in\mathbb{D}} \overline{L}_{\mathrm{P}_n^\xi}(d)$. As in the previous section, consistency is established in the following theorem by proving that the loss function L is regular.

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 3.18. As a consequence, all the conclusions of Corollary 3.19 hold, and 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 of the same six points as the proof of Theorem 4.1. Here **a)** and **c)** are obvious. Let us now prove the four remaining points.

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

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

e) Let us now show that $\mathcal{H}_1 = \mathcal{H}$ is $\mathfrak{P}$-continuous. Let ξ denote a random element in $\mathbb{S}$, $\xi \sim \mathcal{GP}(m,k)$, and let $(\boldsymbol{\nu}_n)$ be a sequence of random measures $\boldsymbol{\nu}_n \in \mathfrak{P}(\xi)$ such that $\boldsymbol{\nu}_n \to \boldsymbol{\nu}_\infty \in \mathfrak{P}(\xi)$ almost surely. For all $n \in \mathbb{N} \cup \{+\infty\}$, let $\mathcal{F}'_n$ denote a σ-algebra such that $\boldsymbol{\nu}_n = \mathrm{P}\left(\xi \in \cdot \mid \mathcal{F}'_n\right)$. Then, using Fubini's theorem, we can rewrite $\mathcal{H}(\boldsymbol{\nu}_n)$ as

$$\mathcal{H}(\boldsymbol{\nu}_n) = \int_{\mathbb{X}} \int_{\mathbb{X}} c_n(u_1, u_2)\, \mu(\mathrm{d}u_1)\, \mu(\mathrm{d}u_2),$$

where $c_n(u_1,u_2) = \mathrm{cov}\left(\mathbb{1}_{\xi(u_1)\geq T}, \mathbb{1}_{\xi(u_2)\geq T} \mid \mathcal{F}'_n\right)$ for $n \in \mathbb{N} \cup \{+\infty\}$. Consider the partition

$$\mathbb{X} = \{u \in \mathbb{X}, \sigma(u) = 0\} \cup A(\omega) \cup B(\omega), \qquad \omega \in \Omega,$$

where σ, A and B are defined as in the proof of Theorem 4.1. Recalling from step **e)** of this proof that $\mathrm{var}\left(\mathbb{1}_{\xi(u)\geq T} \mid \mathcal{F}'_n\right) \overset{\text{a.s.}}{=} 0$ when $\sigma(u) = 0$, and that $\mu(B) \overset{\text{a.s.}}{=} 0$, we obtain that, for all $n \in \mathbb{N} \cup \{+\infty\}$,

$$\mathcal{H}(\boldsymbol{\nu}_n) \overset{\text{a.s.}}{=} \int_A \int_A c_n(u_1, u_2)\, \mu(\mathrm{d}u_1)\, \mu(\mathrm{d}u_2).$$

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 $c_n(u_1,u_2) \to c_\infty(u_1,u_2)$ by the following lemma (proved 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 such that*

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

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 $n \to \infty$, $\mathrm{cov}(\mathbb{1}_{\{Z_{n1}\geq T\}}, \mathbb{1}_{\{Z_{n2}\geq T\}}) \to \mathrm{cov}(\mathbb{1}_{\{Z_1\geq T\}}, \mathbb{1}_{\{Z_2\geq T\}})$.

Finally, using the dominated convergence theorem on $A(\omega) \times A(\omega)$ for almost all $\omega \in \Omega$, we conclude that $\mathcal{H}(\boldsymbol{\nu}_n) \to \mathcal{H}(\boldsymbol{\nu}_\infty)$ almost surely, which proves that $\mathcal{H}$ is $\mathfrak{P}$-continuous.

f) Let $\nu \in \mathbb{M}$ and let $\xi \sim \nu$. Let also $Z_x = \xi(x) + \tau(x)\, U$, with $U \sim \mathcal{N}(0,1)$ independent of ξ, so that $\mathcal{J}_x(\nu) = \mathrm{E}\left(\mathrm{var}\left(\alpha(\xi) \mid Z_x\right)\right)$. We first remark that, from (3.5) and the law of total variance, for any $x \in \mathbb{X}$,

$$\mathcal{G}_x(\nu) = \mathrm{var}(\alpha(\xi)) - \mathrm{E}(\mathrm{var}(\alpha(\xi)|Z_x)) = \mathrm{var}(\mathrm{E}(\alpha(\xi)|Z_x)). \tag{4.8}$$

Then we have the following sequence of equivalences:

$$\begin{aligned} \mathcal{G}(\nu) = 0 &\iff \forall x \in \mathbb{X},\ \mathcal{G}_x(\nu) = 0 \\ &\iff \forall x \in \mathbb{X},\ \mathrm{var}\left(\mathrm{E}(\alpha(\xi) \mid Z_x)\right) = 0 \\ &\iff \forall x \in \mathbb{X},\ \alpha(\xi) - \mathrm{E}(\alpha(\xi)) \perp L^2(Z_x). \end{aligned} \tag{4.9}$$

The first equivalence follows directly from the definition of $\mathcal{G}$: $\mathcal{G}(\nu) = \sup_{x\in\mathbb{X}} \mathcal{G}_x(\nu)$, since $\mathcal{G}_x(\nu)$ is non-negative for all $x \in \mathbb{X}$, the second one from (4.8), and the third one from the fact that $\mathrm{E}(\alpha(\xi) \mid Z_x)$ is the orthogonal projection of $\alpha(\xi)$ onto $L^2(Z_x)$.

Let now $\nu \in \mathbb{Z}_{\mathcal{G}}$. Using Lemma A.10, it follows from (4.9) 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}) converges in distribution to (Z_1, Z_2). Furthermore, from the assumptions the cumulative distribution functions of Z_1 and Z_2 are continuous at T, which implies that, by the Portemanteau theorem, $P(Z_{ni} \geq T) \to P(Z_i \geq T)$. In addition, $Y := \min(Z_1, Z_2)$ 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 get similarly that $\mathrm{E}(\mathbb{1}_{\{Z_{n1}\geq T\}}\mathbb{1}_{\{Z_{n2}\geq T\}}) \to \mathrm{E}(\mathbb{1}_{\{Z_1\geq T\}}\mathbb{1}_{\{Z_2\geq T\}})$, which completes the proof. □

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 $\alpha = \alpha(\xi)$. We know from, e.g., Theorem 6.23 in [37], that $\mathrm{E}_n(\alpha) \to \mathrm{E}_\infty(\alpha)$ almost surely and in L^1. Moreover, it follows from Theorem 4.3 that $\mathrm{var}_n(\alpha) \to 0$ almost surely, and therefore $\mathrm{E}(\mathrm{var}_n(\alpha)) \to 0$ by dominated convergence. Hence, $\mathrm{E}(\mathrm{E}_n[(\mathrm{E}_n(\alpha) - \alpha)^2]) \to 0$, which shows that $\mathrm{E}_n(\alpha)$ converges in L^1, and thus almost surely as well, to $\mathrm{E}_\infty(\alpha) \overset{\text{a.s.}}{=} \alpha$ □

Remark 4.6. Contrary to the case of the integrated Bernoulli variance functional in Proposition 4.2, it is not possible to prove that a SUR sequential design associated with the uncertainty functional (4.7) results in a consistent estimation of the set $\Gamma(\xi) = \{u \in \mathbb{X} : \xi(u) \geq T\}$. Indeed, for instance, let $\mathbb{X} = [-1, 1]$, let μ be Lebesgue measure, let $T = 0$ and let ξ have zero mean function and covariance function k defined by $k(u, v) = uv$ for $u, v \in [-1, 1]$. Then, almost surely, the set $\Gamma(\xi)$ is equal to $[-1, 0]$ or to $[0, 1]$ (with probabilities $1/2$ for both cases). Hence, we have, with ν the distribution of ξ, $\mathcal{H}(\nu) = 0$ because $\alpha(\xi) = 1$ almost surely. Thus, any sequential design $(X_n)_{n\geq 1}$ is a SUR sequential design associated with (4.7), since we have $J_n(x) = 0$ for any $x \in \mathbb{X}$ and $n \geq 1$. However, the conclusions of Proposition 4.2 clearly do not hold for any sequential design. For instance, if $X_n = 0$ for all $n \geq 1$, we have $p_n(u) = 1/2$ for all $u \in \mathbb{X}$.

As a conclusion, for the SUR strategy associated with the uncertainty functional (4.7), and thus based on $\mu(\Gamma(\xi))$, it can only be guaranteed that $\mu(\Gamma(\xi))$, but not $\Gamma(\xi)$ in general, is estimated consistently.

4.3. The knowledge gradient functional

Coming to the topic of sequential design for global optimization, we now focus on the knowledge gradient criterion [25, 26, 55], which is an extension to the general (noisy) case of the strategy proposed in the 70's by [43] for the noiseless case. We shall consider, here and in the next section, the case of a maximization problem. The knowledge gradient sampling criterion, to be maximized, is then defined by

$$G_n(x) = \mathrm{E}_{n,x}\left(\max m_{n+1}\right) - \max m_n, \tag{4.10}$$

with maxima taken over the whole domain $\mathbb{X}$ as in [25, 26] for the case of a discrete $\mathbb{X}$, and in Section 3 of [55] for the case of a "continuous" $\mathbb{X}$; we do not consider the KGCP approximation introduced in Section 4 of [55].

Remark 4.7. The quantity $\max m_n$ in (4.10) does not depend on the sampling point x, and thus plays no part in the selection of the next observation point. The motivation for writing $G_n(x)$ in this form is that the sampling criterion thus defined is non-negative, and becomes equal to zero when $\sigma_n \equiv 0$. The first term in the right-hand side of (4.10) is exactly the sampling criterion proposed by [43] in the noiseless case.

The following criterion, to be minimized, clearly defines the same strategy as (4.10):

$$\begin{aligned} J_n(x) &= \mathrm{E}_n\left(\max \xi\right) - \mathrm{E}_{n,x}\left(\max m_{n+1}\right) \\ &= \mathrm{E}_{n,x}\left(\mathrm{E}_{n+1}(\max \xi) - \max m_{n+1}\right), \end{aligned}$$

and clearly appears, under the second form, as the SUR sampling criterion corresponding to the uncertainty functional

$$\mathcal{H}(\nu) = \int_{\mathbb{S}} \max f \, \nu(\mathrm{d}f) - \max m_\nu. \tag{4.11}$$

Moreover, the original sampling criterion (4.10) is easily seen to be the value $G_n(x) = \mathcal{G}_x\left(\mathrm{P}_n^\xi\right)$ of the associated expected gain functional.

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

$$\overline{L}_{\mathrm{P}_n^\xi}(d) = \mathrm{E}_n\left(\max \xi\right) - m_n(d).$$

The average loss $\overline{L}_{\mathrm{P}_n^\xi}$ reaches its infimum for $d \in \operatorname{argmax} m_n$, and so $H_n = \inf_{d\in\mathbb{D}} \overline{L}_{\mathrm{P}_n^\xi}(d)$. Following the same route as in the last two sections, we have:

Theorem 4.8. *The loss function $L(f, d) = \max f - f(d)$, where $d \in \mathbb{D} = \mathbb{X}$, is regular in the sense of Definition 3.18. As a consequence, all the conclusions of Corollary 3.19 hold, and 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.

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

d) Let $L_0(f) = \max f$. Since $\mathbb{X}$ is compact, it holds for any Gaussian measure $\nu \in \mathbb{M}$ and any $\xi \sim \nu$ that $\mathrm{E}(\max_{\mathbb{X}} |\xi|) < \infty$ (see Section 2.2), and thus we have $L_0 \in \cap_{\nu\in\mathbb{M}} \mathcal{L}^1\left(\mathbb{S}, \mathcal{S}, \nu\right)$. Moreover, it follows from Proposition 3.6 that $\mathcal{H}_1 : \nu \mapsto -\max m_\nu$ is $\mathfrak{P}$-uniformly integrable, since $|\mathcal{H}_1(\nu)| \leq \int L^+ \mathrm{d}\nu$ with $L^+(f) \coloneqq \max |f|$, and $L^+ \in \cap_{\nu\in\mathbb{M}} \mathcal{L}^1\left(\mathbb{S}, \mathcal{S}, \nu\right)$.

e) $\mathcal{H}_1 : \nu \mapsto -\max m_\nu$ is continuous, hence $\mathfrak{P}$-continuous. Indeed, consider a sequence of measures $\nu_n \in \mathbb{M}$ converging to a limit $\nu_\infty \in \mathbb{M}$ in the sense of Definition 2.7. Then m_{ν_n} converges uniformly

to m_{ν_∞} as $n \to \infty$, and therefore $\mathcal{H}_1(\nu_n) = -\max m_{\nu_n}$ converges to $\mathcal{H}_1(\nu_\infty) = -\max m_{\nu_\infty}$ by continuity of $f \mapsto \max f$ on $\mathbb{S}$.

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 $Z_x = \xi(x) + \tau(x)U$ with $U \sim \mathcal{N}(0,1)$ independent of ξ. Let $x^* \in \operatorname{argmax} m$. We have, for all $x \in \mathbb{X}$,

$$\begin{aligned} 0 = \mathcal{G}_x(\nu) &= \mathcal{H}(\nu) - \mathcal{J}_x(\nu) \\ &= (\mathrm{E}\,(\max \xi) - \max m) - \mathrm{E}\Big(\mathrm{E}\,(\max \xi \mid Z_x) - \max_{u \in \mathbb{X}} \mathrm{E}\big(\xi(u) \mid Z_x\big)\Big) \\ &= \mathrm{E}\Big(\max_{u \in \mathbb{X}} \mathrm{E}\big(\xi(u) \mid Z_x\big)\Big) - m(x^*), \end{aligned} \tag{4.12}$$

by the law of total expectation and the optimality property of x^*. For all $x, y \in \mathbb{X}$ it holds that

$$m(x^*) = \mathrm{E}\Big(\mathrm{E}(\xi(x^*) \mid Z_x)\Big) \le \mathrm{E}\Big(\max\big(\mathrm{E}(\xi(y) \mid Z_x),\, \mathrm{E}(\xi(x^*) \mid Z_x)\big)\Big) \le \mathrm{E}\Big(\max_{u \in \mathbb{X}} \mathrm{E}\big(\xi(u) \mid Z_x\big)\Big),$$

and therefore, using (4.12),

$$0 \le \mathrm{E}\Big(\max(\mathrm{E}(\xi(y)|Z_x), \mathrm{E}(\xi(x^*)|Z_x)) - \mathrm{E}(\xi(x^*)|Z_x)\Big) \le \mathcal{G}_x(\nu) = 0.$$

Setting $W_{x,y} \coloneqq \mathrm{E}(\xi(y) \mid Z_x) - \mathrm{E}(\xi(x^*) \mid Z_x)$, we have thus proved that $\mathrm{E}\,(\max\,(0,\, W_{x,y})) = 0$, from which it follows that $\operatorname{var}(W_{x,y}) = 0$ since

$$W_{x,y} = m(y) - m(x^*) + \mathbb{1}_{\sigma^2(x)+\tau^2(x)>0} \frac{k(x,y) - k(x,x^*)}{\sigma^2(x) + \tau^2(x)} (Z_x - m(x))$$

is Gaussian. Observe now that

$$\operatorname{var}(W_{x,y}) = \mathbb{1}_{\sigma^2(x)+\tau^2(x)>0} \, \frac{(k(x,y) - k(x,x^*))^2}{\sigma^2(x) + \tau^2(x)}.$$

Hence it must be the case that either $\sigma^2(x)+\tau^2(x) = 0$ or $k(x,y) = k(x,x^*)$. But, if $\sigma^2(x)+\tau^2(x) = 0$ then $\sigma(x) = 0$ and therefore $k(x,y) = k(x,x^*) = 0$. Summing up, we have proved that

$$k(x,y) = k(x,x^*), \qquad \forall x, y \in \mathbb{X}.$$

As a consequence, for all $x, y \in \mathbb{X}$, we have

$$k(x,x) - k(x,y) = k(x,x^*) - k(x,x^*) = 0,$$

and therefore

$$\operatorname{var}(\xi(x) - \xi(y)) = (k(x,x) - k(x,y)) + (k(y,y) - k(x,y)) = 0.$$

It follows that, almost surely, the sample paths of $\xi - m$ are constant over $\mathbb{X}$, and so $\max \xi = \xi(x^*) - m(x^*) + \max m = \xi(x^*)$. We have thus proved that $\mathcal{H}(\nu) = \mathrm{E}\,(\max \xi) - m(x^*) = 0$ for any $\nu \in \mathbb{Z}_{\mathcal{G}}$, which concludes the proof. □

In the next proposition, we refine Theorem 4.8 by showing that the loss $\max \xi - \xi(X_n^*)$ goes to zero for any sequence of optimal decisions $X_n^* \in \operatorname{argmax} m_n$.

Proposition 4.9. *Let (X_n^*) be any sequence of $\mathcal{F}_n$-measurable $\mathbb{X}$-valued random variables such that $X_n^* \in \operatorname{argmax} m_n$ almost surely for all n. Then, for any quasi-SUR design associated with $\mathcal{H}$, $\xi(X_n^*) \to \max \xi$ almost surely and in L^1.*

Proof. From step **f)** in the proof of Theorem 4.8, and the fact that $\mathcal{H}(P_\infty^\xi) \overset{\text{a.s.}}{=} 0$, it follows that the sample paths of $\xi - m_\infty$ are almost surely constant over $\mathbb{X}$. Let X^* denote an $\mathbb{X}$-valued random variable such that $X^* \in \operatorname{argmax} \xi$. Then, we have

$$\limsup_{n\to\infty} (\xi(X^*) - \xi(X_n^*)) = \limsup_{n\to\infty} (m_\infty(X^*) - m_\infty(X_n^*)) = \limsup_{n\to\infty} (m_n(X^*) - m_n(X_n^*)) \leq 0$$

almost surely. This implies, since $\xi(X^*) - \xi(X_n^*) \geq 0$, that $\xi(X_n^*) \to \xi(X^*) = \max \xi$ almost surely. Convergence in the L^1 sense is finally obtained by the dominated convergence theorem. □

4.4. The expected improvement functional

This section addresses the celebrated expected improvement strategy [36, 43][4]. Assume that exact (noiseless) evaluations can be made, in other words, that $\tau(x) = 0$ for all $x \in \mathbb{X}$: then, we define the expected improvement criterion, to be maximized, as

$$G_n(x) = \mathrm{E}_{n,x} \left(M_{n+1} - M_n \right), \tag{4.13}$$

with $M_n = \max_{x\in\mathbb{X}:\, \sigma_n(x)=0} \xi(x)$ and $\sigma_n^2(x) = k_n(x,x)$. Observe that, on the right-hand side of (4.13), similarly to Remark 4.7 for the knowledge gradient criterion, only M_{n+1} actually depends on the new observation point $X_{n+1} = x$. Note also that we need at least one $x \in \mathbb{X}$ such that $\sigma_n(x) = 0$ for M_n to be well defined, which is always true as soon as $n \geq 1$ (in practice, (4.13) is typically used after an initial design of size $n_0 > 0$).

Remark 4.10. Our definition of the EI strategy, and in particular of the current best value M_n, differs slightly from the usual one [36], which takes $M_n = \max(\xi(X_1), \ldots, \xi(X_n))$. This minor variation is necessary if we want to see the EI strategy as stemming from some uncertainty functional. Remark that, in the case of a non-degenerate Gaussian process (i.e., when $\sigma_n(x) = 0$ if and only if $x \in \{X_1, \ldots, X_n\}$), the two definitions of M_n coincide and the criterion can be written more familiarly as

$$G_n(x) = \mathrm{E}_n \left(\max(0, \xi(x) - M_n) \right). \tag{4.14}$$

The sampling criteria (4.13) and (4.14) no longer agree in general, since it can happen for degenerate Gaussian processes that $M_n > \max(\xi(X_1), \ldots, \xi(X_n))$. Degeneracy occurs, e.g., in the case of finite-dimensional Gaussian processes (i.e., linear models with a Gaussian prior on their coefficients), or for processes with pathwise invariance properties [29] (for instance $\xi(x) = -\xi(-x)$ for all $x \in \mathbb{X}$, almost surely).

It turns out that the sequential design obtained by iteratively maximizing (4.13) can be interpreted as a SUR sequential design. Indeed, we have

$$\mathrm{E}_{n,x} \left(M_{n+1} - M_n \right) = \mathrm{E}_n \left(\max \xi - M_n \right) - \mathrm{E}_{n,x} \left(\max \xi - M_{n+1} \right),$$

[4] As explained in Section 4.3, the strategy originally proposed by [43]—and earlier work in Russian by J. Mockus and A. Žilinskas—is more accurately described, in principle, as a special (noiseless) case of what we have called the "knowledge gradient" strategy. However, for the particular Brownian motion-based Gaussian process prior used in [43], the maximum of m_n always occurs at an observation point, and the criterion of [43] then coincides (see page 21 of the paper) with what is, currently, commonly referred to as the expected improvement criterion.

where the subscript "x" has been dropped from the first expectation since its argument does not depend on the position X_{n+1} of the next evaluation. Thus, using the fact that $\mathrm{E}_{n,x}(\max\xi) = \mathrm{E}_{n,x}(\mathrm{E}_{n+1}(\max\xi))$ by the law of total expectation, we see that maximizing (4.13) is equivalent to minimizing

$$J_n(x) = \mathrm{E}_{n,x}\big(\mathrm{E}_{n+1}(\max\xi) - M_{n+1}\big),$$

which is precisely the SUR strategy associated with the uncertainty functional defined, for any $\nu \in \mathbb{M}$ such that σ_ν vanishes at at least one $x \in \mathbb{X}$, by

$$\mathcal{H}(\nu) = \int_{\mathbb{S}} \max f \; \nu(\mathrm{d}f) - \max_{x\in\mathbb{X}:\sigma_\nu(x)=0} m_\nu(x). \tag{4.15}$$

Indeed, $\max_{x\in\mathbb{X}:\,\sigma_\nu(x)=0}\xi(x) \overset{\text{a.s.}}{=} \max_{x\in\mathbb{X}:\,\sigma_\nu(x)=0} m_\nu(x)$ for any such ν and any $\xi\sim\nu$. Moreover, the expected improvement criterion G_n turns out to be the value $G_n(x) = \mathcal{G}_x\left(\mathrm{P}_n^\xi\right)$ of the associated expected gain functional. In order to have an uncertainty function $\mathcal{H}$ that is well defined on all of $\mathbb{M}$, set

$$\mathcal{H}(\nu) = \int_{\mathbb{S}} (\max f - \min f)\;\nu(\mathrm{d}f) \tag{4.16}$$

for all $\nu \in \mathbb{M}$ such that σ_ν does not vanish.

The uncertainty functional (4.15)–(4.16) is measurable (see Proposition A.4) and can be associated with a certain loss function, as shown in the following result. Contrary to the case of the three previous criteria, however, this loss function is not regular in general.

Proposition 4.11. *The EI uncertainty functional is of the form (3.16), with L the loss function defined on the decision space $\mathbb{D} = \mathbb{X}\times(\mathbb{R}\cup\{-\infty\})$, for any $f\in\mathbb{S}$ and $d=(x^*,z^*)\in\mathbb{D}$, by*

$$L(f,d) = \begin{cases} \max f - z^* & \text{if } f(x^*) = z^* \text{ and } z^* > -\infty,\\ \max f - \min f & \text{if } z^* = -\infty,\\ +\infty & \text{otherwise.}\end{cases} \tag{4.17}$$

Assuming that $\mathbb{X}\subset\mathbb{R}^p$ has a non-empty interior, the loss function (4.17) is not regular, and neither would be any other loss function that could be associated with the EI uncertainty functional.

Proof. Let us first prove that $\mathcal{H}$ is of the form (3.16). Let $\nu\in\mathbb{M}$, $\xi\sim\nu$ and $d=(x^*,z^*)\in\mathbb{D}$. Then the average loss is equal to $\overline{L}_\nu(d) = \mathrm{E}(\max\xi - \min\xi)$ if $z^* = -\infty$ and, using the convention $(+\infty)\cdot 0 = 0$, to

$$\begin{aligned}\overline{L}_\nu(d) &= \mathrm{E}\left(L(\xi,d)\right)\\ &= \mathrm{E}\left((\max\xi - z^*)\,\mathbb{1}_{\xi(x^*)=z^*}\right) \;+\; (+\infty)\cdot\mathrm{P}\left(\xi(x^*)\neq z^*\right)\\ &= \begin{cases}\mathrm{E}(\max\xi) - m_\nu(x^*), & \text{if } \sigma_\nu(x^*)=0 \text{ and } m_\nu(x^*)=z^*,\\ +\infty & \text{otherwise,}\end{cases}\end{aligned} \tag{4.18}$$

if $z^* > -\infty$. The last equality follows from the simple observation that the event $\{\xi(x^*)=z^*\}$ is almost sure if $\sigma_\nu(x^*)=0$ and $m_\nu(x^*)=z^*$, and negligible otherwise.

In the case where there exists at least one $x\in\mathbb{X}$ such that $\sigma_\nu(x)=0$, then it is clear that $\mathrm{E}(\max\xi) - m_\nu(x) < \mathrm{E}(\max\xi - \min\xi)$ for any such x, and thus the expected loss is minimal for $d=(x^*,z^*)$ such that $x^*\in\operatorname{argmax}_{x:\sigma_\nu(x)=0} m_\nu(x)$ and $z^* = m_\nu(x^*)$, which yields (4.15). In the case

where σ_ν does not vanish, on the other hand, then (4.18) is always infinite and thus the expected loss in minimal for any $d = (x^*, -\infty)$, which yields (4.16). In both cases, we have proved that $\mathcal{H}(\nu) = \min_{d\in\mathbb{D}} \overline{L}_\nu(d)$.

We will now prove that there is no regular loss function L such that $\mathcal{H}(\nu) = \min_{d\in\mathbb{D}} \overline{L}_\nu(d)$. To do so, we will show that $\mathcal{H}$ cannot be decomposed as $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$, with $\mathcal{H}_0(\nu) = \int_{\mathbb{S}} L_0 \,\mathrm{d}\nu$ for some $L_0 \in \cap_{\nu\in\mathbb{M}} \mathcal{L}^1(\mathbb{S}, \mathcal{S}, \nu)$, and $\mathcal{H}_1$ a $\mathfrak{P}$-continuous functional.

Assume, for the sake of contradiction, that $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$, with $\mathcal{H}_0(\nu) = \int_{\mathbb{S}} L_0 \,\mathrm{d}\nu$ for some $L_0 \in \cap_{\nu\in\mathbb{M}}\mathcal{L}^1(\mathbb{S}, \mathcal{S}, \nu)$, and $\mathcal{H}_1$ a $\mathfrak{P}$-continuous functional. Then, using the same martingale argument as in the proof of Theorem 3.16, we have $\mathcal{H}(P_n^\xi) \xrightarrow{\text{a.s.}} \mathcal{H}(P_\infty^\xi)$. Also, again by a martingale argument, $\mathrm{E}_n(\max\xi) \xrightarrow{\text{a.s.}} \mathrm{E}_\infty(\max\xi)$, and therefore $\max_{\sigma_n(x)=0} m_n(x) \xrightarrow{\text{a.s.}} \max_{\sigma_\infty(x)=0} m_\infty(x)$.

We will now show that this last convergence does not hold for a certain Gaussian process ξ on $\mathbb{X}$, which yields a contradiction. For simplicity, we assume in the following that $\mathbb{X} = [0, 1]$, but the same argument could be made on any $\mathbb{X} \subset \mathbb{R}^p$ that has a non-empty interior.

Consider a Gaussian process ξ with mean $m(x) = x$ and covariance $k(x, y) = \exp(-(x - y)^2)$. Let (X_n) be a deterministic sequence, dense in $[0, 1/3]$. Then, as follows from the proof of Proposition 1 in [64], we have $\sigma_\infty(x) = 0$ for all $x \in [0, 1]$. Hence, $\max_{\sigma_\infty(x)=0} m_\infty(x) = \max_{x\in[0,1]} \xi(x)$. Also, since $X_k \in [0, 1/3]$ for all $k \in \mathbb{N}$, and since ξ is a non-degenerate Gaussian process, we have $\max_{\sigma_n(x)=0} m_n(x) \le \max_{x\in[0,1/3]} m_n(x)$. This upper bound converges to $\max_{x\in[0,1/3]} \xi(x)$ almost surely, and thus $\max_{x\in[0,1/3]} \xi(x) = \max_{x\in[0,1]} \xi(x)$ almost surely. This last equality cannot hold, however, because

$$\max_{x\in[0,1/3]} \xi(x) \le \frac{1}{3} + \max_{x\in[0,\,1/3]} (\xi(x) - x),$$
$$\max_{x\in[2/3,1]} \xi(x) \ge \frac{2}{3} + \max_{x\in[2/3,\,1]} (\xi(x) - x),$$

and by symmetry $\max_{x\in[0,1/3]}(\xi(x) - x)$ and $\max_{x\in[2/3,1]}(\xi(x) - x)$ have the same distribution. □

Since the EI uncertainty functional does not derive from a regular loss function, consistency cannot be proved using Corollary 3.19 as in the three previous examples. The following result will thus be proved by a direct application of the more general Theorem 3.12.

Theorem 4.12. *For any quasi-SUR sequential design associated with $\mathcal{H}$, as $n \to \infty$, almost surely and in L^1, $H_n \to 0$, $\max m_n \to \max\xi$ and $M_n \to \max\xi$.*

Proof. Since the uncertainty functional $\mathcal{H}$ derives from a loss function by Proposition 4.11, it is DoA by Proposition 3.17 and, consequently, has the supermartingale property.

Consider now a quasi-SUR sequential design associated with $\mathcal{H}$. Theorem 3.12 applies and therefore $\mathcal{G}(P_n^\xi) \to 0$ almost surely. Observe also that, for all $n \ge 1$,

$$M_{n+1} = \max_{\sigma_{n+1}(x)=0} \xi(x) \ge \max\left(\xi\left(X_{n+1}\right), \max_{\sigma_n(x)=0} \xi(x)\right) = \max\left(\xi\left(X_{n+1}\right), M_n\right).$$

Hence $\mathcal{G}(P_n^\xi) = \sup_{x\in\mathbb{X}} \mathrm{E}_{n,x}\left(M_{n+1} - M_n\right) \ge \sup_{x\in\mathbb{X}} \mathrm{E}_n\left(\max\left(0,\, \xi(x) - M_n\right)\right)$, and thus

$$\max_{x\in\mathbb{X}} \gamma\left(m_n(x) - M_n, \sigma_n^2(x)\right) \xrightarrow[n\to\infty]{\text{a.s.}} 0, \tag{4.19}$$

where γ denotes the function defined by $\gamma(a, b) = \mathrm{E}\left(\max\left(0,\, Z_{a,b}\right)\right)$, $Z_{a,b} \sim \mathcal{N}(a, b)$. Recall from Section 3 in Vazquez and Bect [63] that γ is continuous and satisfies

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

Recall also from Proposition 2.9 that, almost surely, $m_n \to m_\infty$ and $\sigma_n \to \sigma_\infty$ uniformly on $\mathbb{X}$. Therefore we have

$$\forall x \in \mathbb{X}, \qquad \gamma\left(m_\infty(x) - M_\infty, \sigma_\infty^2(x)\right) = 0, \tag{4.20}$$

almost surely, where M_∞ denotes the almost sure limit of the increasing sequence (M_n), with $M_\infty \leq \max \xi < +\infty$. (To see that (M_n) is increasing, observe that the set of points $x \in \mathbb{X}$ such that $\sigma_n(x) = 0$ is growing with n, since $(\sigma_n(x))$ is decreasing for any x.) Considering the properties of γ, it follows from (4.20) that almost surely, for all $x \in \mathbb{X}$, $\sigma_\infty(x) = 0$ and $m_\infty(x) - M_\infty \leq 0$. Therefore, almost surely, we have $\xi = m_\infty$ and $M_\infty \geq \max m_\infty$. Since it is clear that $M_n \leq \max m_n$ for all n, we also have $M_\infty \leq \max m_\infty$ in the limit, hence $M_\infty = \max m_\infty = \max \xi$ almost surely.

We have proved so far that $\max m_n \to \max \xi$ and $M_n \to \max \xi$ almost surely. Moreover, $\mathrm{E}_n(\max \xi)$ is a martingale that converges almost surely and in L^1 to $\mathrm{E}_\infty(\max \xi) = \max \xi$ (see, e.g., Theorem 6.23 in [37]), and therefore $H_n = \mathrm{E}_n(\max \xi) - M_n \to 0$ almost surely.

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 4.12 improves the consistency result of [63], since it does not impose the no-empty-ball property on the covariance function k. Hence, Proposition 4.12 also holds with very smooth Gaussian processes, such as Gaussian processes with a Gaussian (a.k.a. squared exponential) covariance function, or with Gaussian processes whose sample paths have symmetry properties [29].

Appendix A: Technical results and proofs

A.1. Measurability results

Lemma A.1. *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 $\mathbb{M} \times \mathbb{E} \to [0, +\infty]$, $(\nu, y) \mapsto \int_{\mathbb{S}} \varphi(f, y)\, \nu(\mathrm{d}f)$, is $\mathcal{M} \otimes \mathcal{E}$-measurable.*

Proof. The result is clear for any $\varphi = \mathbb{1}_{A \times B}$, with $A \in \mathcal{S}$ and $B \in \mathcal{E}$. Indeed, $\int_{\mathbb{S}} \varphi(f, y)\, \nu(\mathrm{d}f) = \pi_A(\nu)\, \mathbb{1}_B(y)$, where π_A denotes the evaluation map $\nu \mapsto \nu(A)$, 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. □

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

Lemma A.2. *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., 60, 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.1. 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 of $\mathbb{S} \times \mathcal{C}(\mathbb{X} \times \mathbb{X})$.

Lemma A.3. *Ψ is a bi-measurable mapping from $(\mathbb{M}, \mathcal{M})$ to $(\Theta, \mathcal{T})$.*

Proof. The measurability of Ψ follows from Lemma A.2. 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., 7]. □

Proposition A.4. *The expected improvement functional (4.15)–(4.16) is $\mathcal{M}$-measurable.*

Proof. Let $\{x_i\}$ denote a countable dense subset of $\mathbb{X}$ and set, for all $k > 0$,

$$\mathcal{H}_k(\nu) = \int_{\mathbb{S}} (\max f - \min f)\, \nu(\mathrm{d}f) - \sup_i \left(m_\nu(x_i) - \int_{\mathbb{S}} \min f\, \nu(\mathrm{d}f) \right) \mathbb{1}_{\sigma_\nu(x_i) \le \frac{1}{k}}.$$

The mappings $\nu \mapsto \int_{\mathbb{S}} \max f\, \nu(\mathrm{d}f)$, $\nu \mapsto \int_{\mathbb{S}} \min f\, \nu(\mathrm{d}f)$, $(\nu, x) \mapsto m_\nu(x)$ and $(\nu, x) \mapsto \sigma_\nu^2(x)$ are measurable by Lemma A.1. As a consequence, for any $k > 0$, the functional $\mathcal{H}_k$ is $\mathcal{M}$-measurable. The result follows from the fact that $\mathcal{H}_k \to \mathcal{H}$ pointwise as $k \to \infty$. □

A.2. 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 the 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

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

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

where $K(\underline{x}_n)^\dagger$ denotes the Moore-Penrose pseudo-inverse of $K(\underline{x}_n) = \left(k(x_i, x_j) + \tau(x_i)^2\, \delta_{i,j}\right)_{1 \le i,j \le n}$, and $k(\underline{x}_n, \cdot\,)$ and the other notations should be self-explanatory.

Lemma A.5. *$\tilde{\kappa}_n : (\underline{x}_n, \underline{z}_n, (m,k)) \mapsto (m_n(\,\cdot\,;\, \underline{x}_n, \underline{z}_n),\, k_n(\,\cdot\,;\, \underline{x}_n))$ 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 $K \mapsto K^\dagger$ [54]. □

Proof of Proposition 2.6. 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.3 and A.5 that κ_n is measurable. Standard algebraic manipulations then show that

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

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 2.6 is thus established with $\mathrm{Cond}_{x_1,z_1,\ldots,x_n,z_n} : \nu \mapsto \kappa_n\left(\underline{x}_n, \underline{z}_n, \nu\right)$. □

Proposition A.6. *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 2.6. Using Lemma A.3 and the measurability of κ_1, the integrand in the right-hand side of (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. □

Remark A.7. As a consequence of Proposition A.6, $J_n : x \mapsto \mathcal{J}_x(\mathrm{P}_n^\xi)$ is an $\mathcal{F}_n$-measurable process for all n, and thus $J_n(X)$ is a well-defined $\mathcal{F}_n$-measurable random variable for any $\mathcal{F}_n$-measurable $\mathbb{X}$-valued random variable X.

A.3. Convergence in $\mathbb{M}$

Proof of Proposition 2.9. Recall from Proposition 2.6 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 (see, e.g., [60]) and $\|\xi\|_\infty$ is integrable (see, e.g., Theorem 2.9 in [2]). The conditional expectation $\mathrm{E}\left(\xi \mid \mathcal{F}_n\right)$ of ξ given $\mathcal{F}_n$ is thus well defined as an $\mathbb{S}$-valued random element (since $\mathbb{S} = \mathcal{C}(\mathbb{X})$ is a separable Banach space; see, e.g., Theorem 5.1.12 in [59]) and is easily seen to coincide with m_n. As a consequence, it follows from Theorem 6.1.12 in [59] that m_n converges uniformly, almost surely and in $L^1\left(\Omega, \mathcal{F}, \mathrm{P}\right)$, to $m_\infty := \mathrm{E}\left(\xi \mid \mathcal{F}_\infty\right)$. 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 2.6, 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 $\left(\sigma_n^2(\mu)\right)_{n\ge1}$ 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 $x, y, x', y' \in \mathbb{X}$:

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

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 $\mathbb{X}^2 \to \mathbb{R}$, $(x,y) \mapsto \sigma_n^2(\delta_x + \delta_y)$. This is a decreasing sequence of continuous functions, which converges pointwise to the continuous function $(x,y) \mapsto \sigma_\infty^2(\delta_x + \delta_y)$. Since $\mathbb{X}^2$ is compact, the convergence is uniform by Dini's first theorem. The same argument applies to $(x,y) \mapsto \sigma_n^2(\delta_x - \delta_y)$ 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 [37] 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 $(\mathbb{S}, \mathcal{S}, \mathrm{Q}(\omega,\cdot))$, and thus $\mathrm{Q}(\omega,\cdot)$ 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 2.7, thereby concluding the proof. □

Proof of Proposition 2.10. 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 (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 facts that k is uniformly continuous over $\mathbb{X} \times \mathbb{X}$ (since k is continuous and $\mathbb{X} \times \mathbb{X}$ is a compact metric space) and that $K \mapsto K^\dagger$, where $K^\dagger$ denotes the Moore-Penrose pseudo-inverse of K, is continuous at $K = k_\nu(x,x) + \tau^2(x) > 0$ (the covariance matrix is actually a scalar in this case). □

A.4. Existence of SUR and quasi-SUR sequential designs

This section contains general existence results for ε-quasi-SUR sequential designs. Recall that $\mathbb{X}$ is assumed, throughout the paper, to be a compact metric space (see Standing assumptions 2.2).

Theorem A.8. *Let the assumptions of Theorem 3.16 hold. Then,*

a) for any sequential design, the sample paths of J_n are continuous on $\left\{x \in \mathbb{X} : s_n^2(x) > 0\right\}$;
b) for any sequence $\varepsilon = (\varepsilon_n)$ of strictly 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 $J_n(x) = \mathcal{J}_x\left(\mathrm{P}_n^\xi\right)$, 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 (3.1) that $J_0(x) = \mathcal{J}_x(\mathrm{P}_0^\xi) = \mathrm{E}\big(\mathcal{H}\big(\mathrm{Cond}_{x,\,Z_1(x)}(\mathrm{P}_0^\xi)\big)\big)$. Set $\boldsymbol{\nu}_k = \mathrm{Cond}_{x_k,\,Z_1(x_k)}(\mathrm{P}_0^\xi)$ and $\boldsymbol{\nu}_\infty = \mathrm{Cond}_{x,\,Z_1(x)}(\mathrm{P}_0^\xi)$. 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 2.10. 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 $\mathbb{X}$:

$$B_{n,\gamma}(\omega) = \left\{x \in \mathbb{X} \mid s_n(\omega,x) \geq \gamma^{-1} > 0\right\}, \tag{A.7}$$

$$A_{n,\gamma}(\omega) = B_{n,\gamma}(\omega) \cap \left\{x \in \mathbb{X} \mid J_n(\omega,x) \leq \inf J_n(\omega,x) + \varepsilon_n\right\}. \tag{A.8}$$

Let us prove that, on the event $\{s_n \not\equiv 0\}$, the set $A_{n,\gamma}(\omega)$ is non-empty for large values of γ. Assume that $s_n(\omega,\cdot) \not\equiv 0$, and recall that $J_n(\omega,\cdot) \leq H_n(\omega)$ by (3.4). If $J_n(\omega,\cdot) \equiv H_n$, then for any x such that $s_n(\omega,x) > 0$ and any $\gamma \geq s_n(\omega,x)^{-1}$ we have $x \in A_{n,\gamma}(\omega) = B_{n,\gamma}(\omega)$. If $\inf_x J_n(\omega,x) < H_n$, pick a sequence (x_k) such that $J_n(\omega,x_k) \to \inf_x J_n(\omega,x)$. For some k large enough, $J_n(\omega,x_k) \leq \inf_x J_n(\omega,x) + \varepsilon_n$ and $J_n(\omega,x_k) < H_n$. As a consequence, $s_n(\omega,x_k) > 0$ (this follows from (3.3) and the fact that $\mathrm{Cond}_{x,m_\nu(x)}(\nu) = \nu$ if $s_\nu^2(x) = 0$) and thus $x_k \in A_{n,\gamma}(\omega)$ for any $\gamma \geq s_n(\omega,x_k)^{-1}$. In both cases the claim is proved.

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., 44, 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 $\{A_{n,\gamma} \neq \varnothing\}$. Let $\tilde{x}$ denote an arbitrary fixed point in $\mathbb{X}$. Setting

$$X_{n+1} = \begin{cases} \tilde{x} & \text{if } s_n \equiv 0, \\ X_{n+1}^{(k)} & \text{if } A_{n,k} \neq \varnothing \text{ and } A_{n,l} = \varnothing,\ \forall l < k, \end{cases} \tag{A.9}$$

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

In some situations, it is possible to prove directly the continuity of the sampling criteria J_n on the whole of $\mathbb{X}$ (see Section 4.4 for an example), in which case a stronger existence result can be formulated as in [63], that does not even require the supermartingale property:

Theorem A.9. *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 measurable selection theorem for random closed sets. □

A.5. Miscellaneous

Lemma A.10. *Let U, V and W be real-valued random variables such that*

1. *W is independent of (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.11. 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 kth 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:

$$\mathrm{cov}\left(U, (V+W)^{k_0}\right) \;=\; \sum_{k=0}^{k_0} \binom{k_0}{k} \mathrm{E}\left(UV^k\right) \mathrm{E}\left(W^{k-k_0}\right) \;=\; \mathrm{E}\left(UV^{k_0}\right) \;\neq\; 0. \tag{A.10}$$

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

Acknowledgments

Part of David Ginsbourger's contribution took place within the framework of the "Bayesian set estimation under random field priors" project (Number 146354) funded by the Swiss National Science Foundation. François Bachoc acknowledges support from the "PEPITO" project funded by the French National Agency for Research. The authors would like to thank Luc Pronzato for pointing out the important connection with DeGroot's earlier work on uncertainty functionals, Dario Azzimonti for proofreading, and two anonymous reviewers for their careful reading and numerous suggestions of improvement.

References

[1] Adler, R. J. (1981). *The Geometry of Random Fields*. Wiley, New York.

[2] Azaïs, J.-M. and Wschebor, M. (2009). *Level Sets and Extrema of Random Processes and Fields*. Wiley.

[3] Azzimonti, D., Ginsbourger, D., Chevalier, C., Bect, J., and Richet, Y. (2018). Adaptive design of experiments for conservative estimation of excursion sets. *ArXiv preprint*, arxiv:1611.07256v3.

[4] 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*.

[5] 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.

[6] Bect, J., Li, L., and Vazquez, E. (2017). Bayesian subset simulation. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):762–786.

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

[8] 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.

[9] Bogachev, V. I. (1998). *Gaussian Measures*. American Mathematical Society.

[10] Brezis, H. (2010). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer Science & Business Media.

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

[12] Chevalier, C. (2013). *Fast Uncertainty Reduction Strategies Relying on Gaussian Process Models*. PhD thesis, University of Bern.

[13] 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.

[14] Cohn, D. A., Ghahramani, Z., and Jordan, M. I. (1996). Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4:129–145.

[15] Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. John Wiley & Sons.

[16] Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.

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

[18] DeGroot, M. H. (1986). Concepts of information based on utility. In Daboni, L., Montesano, A., and Lines, M., editors, *Recent Developments in the Foundations of Utility and Risk Theory*, pages 265–275. Springer Netherlands, Dordrecht.

[19] DeGroot, M. H. (1994). Changes in utility as information. *Theory and Decision*, 17(3):287–303.

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

[21] 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).

[22] Feliot, P., Bect, J., and Vazquez, E. (2016). A Bayesian approach to constrained single- and multi-objective optimization. *Journal of Global Optimization*, 67(1):97–133.

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

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

[25] 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.

[26] 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.

[27] 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.

[28] 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 Uncertainty Quantification*, 2(1):490–510.

[29] Ginsbourger, D., Roustant, O., and Durrande, N. (2016). On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *Journal of Statistical Planning and Inference*, 170:117–128.

[30] 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.

[31] 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*.

[32] Hainy, M., Müller, W. G., and Wynn, H. P. (2014). Learning functions and approximate Bayesian computation design: ABCD. *Entropy*, 16(8):4353–4374.

[33] Hennig, P., Osborne, M. A., and Girolami, M. (2015). Probabilistic numerics and uncertainty in computations. *Proceedings of the Royal Society A*, 471:20150142.

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

[35] Johnson, R. (1960). An information theory approach to diagnosis. In *Proceedings of the 6th Symposium on Reliability and Quality Control*, pages 102–109.

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

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

[38] King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., and Supowit, A. (1994). Efficient and unbiased modifications of the quest threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Research*, 34(7):885–912.

[39] 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.

[40] Ledoux, M. and Talagrand, M. (2013). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer Science & Business Media.

[41] 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.

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

[43] 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.

[44] Molchanov, I. (2006). *Theory of Random Sets*. Springer Science & Business Media.

[45] O'Geran, J. H., Wynn, H. P., and Zhiglyavsky, A. A. (1993). Mastermind as a test-bed for search algorithms. *Chance*, 6(1):31–37.

[46] O'Hagan, A. (1991). Bayes–Hermite quadrature. *Journal of Statistical Planning and Inference*, 29(3):245–260.

[47] 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.

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

[49] 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).

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

[51] Ritter, K. (2000). *Average-Case Analysis of Numerical Problems*, volume 1733 of *Lecture Notes in Mathematics*. Springer Verlag.

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

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

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

[55] 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.

[56] 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.

[57] 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.

[58] Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.

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

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

[61] 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.

[62] 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)*.

[63] 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.

[64] 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.

[65] Vazquez, E. and Bect, J. (2012). Sequential search based on kriging: convergence analysis of some algorithms. In *Bulletin of the ISI 58th World Statistics Congress of the International Statistical Institute, 2011*, The Hague, The Netherlands. International Statistical Institute.

[66] 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.

[67] 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.

[68] 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.

[69] Yarotsky, D. (2013a). Examples of inconsistency in optimization by expected improvement. *Journal of Global Optimization*, 56(4):1773–1790.

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

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