APPROXIMATION OF LIMIT STATE SURFACES IN MONOTONIC MONTE CARLO SETTINGS

Nicolas Bousquet, Thierry Klein, Vincent Moutoussamy

Applied Probability Trust (18 March 2015)

APPROXIMATION OF LIMIT STATE SURFACES IN MONOTONIC MONTE CARLO SETTINGS

NICOLAS BOUSQUET,* *EDF R&D & Institut de Mathématiques de Toulouse*

THIERRY KLEIN,** *Institut de Mathématique de Toulouse*

VINCENT MOUTOUSSAMY,*** *EDF R&D & Institut de Mathématiques de Toulouse*

Abstract

This article investigates the theoretical convergence properties of the estimators produced by a numerical exploration of a monotonic function with multivariate random inputs in a structural reliability framework. The quantity to be estimated is a probability typically associated to an undesirable (unsafe) event and the function is usually implemented as a computer model. The estimators produced by a Monte Carlo numerical design are two subsets of inputs leading to safe and unsafe situations, the measures of which can be traduced as deterministic bounds for the probability. Several situations are considered, when the design is independent, identically distributed or not, or sequential. As a major consequence, a consistent estimator of the (limit state) surface separating the subsets under isotonicity and regularity arguments can be built, and its convergence speed can be exhibited. This estimator is built by aggregating semi-supervized binary classifiers chosen as constrained Support Vector Machines. Numerical experiments conducted on toy examples highlight that they work faster than recently developed monotonic neural networks with comparable predictable power. They are therefore more adapted when the computational time is a key issue.

* Postal address: EDF Lab, 6 quai Watier, 78401 Chatou, France
** Postal address: Institut de Mathématique de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France
*** Postal address: EDF Lab, 6 quai Watier, 78401 Chatou, France

Keywords: limit state surface ; computer model ; monotonicity ; convexity ; structural reliability ; rare event probability ; random sets ; support vector machines

2010 Mathematics Subject Classification: Primary 65C05
Secondary 62L12;60K10

1. Introduction

Due to the increasingly powerful computational tools, so-called computer models g are developed to mimic complex processes (e.g., technical, biological, socio-economical, etc. [7, 10]). The exploration of multivariate deterministic *black-box* functions $x \mapsto g(\mathbf{x})$ by numerical designs of experiments has become, over the last years, a major theme of research in the interacting fields of engineering, numerical analysis, probability and statistics [30]. Numerical investigations based on Monte Carlo variance reduction techniques [27] are often needed since, on the one hand, the complexity of g restricts the use of intrusive explorations, and in the other hand each run of the computer model can be very costly. A number of studies have for common aim to delimit the set of input situations $\mathbf{x}$ dominated by a limit state surface Γ defined by the equation $g(\mathbf{x}) = 0$, where the dominance rule is defined by a partial order. In multi-objective optimization frameworks, where g is assimilated to a decision rule, Γ can be viewed as a Pareto frontier delimiting a performance space [15]. In structural reliability it is often wanted to assess the measure of the set $\mathbb{U}^- = \{x; g(\mathbf{x}) \leq 0\}$, which can be interpreted as the probability p of an undesirable event when the input vector $\mathbf{x}$ is randomized [21].

The Pareto dominance between the inputs $\mathbf{x}$ is the natural partial order needed to formalize an assumption of monotonicity placed on g. This hypothesis corresponds to a technical reality in numerous situations encountered in engineering [25] or decision-making [4], or a conservative approximation to reality in reliability studies. Therefore the exploration of monotonic time-consuming computer models by numerical means has been investigated by a growing number of researchers [6, 9, 17]. The most immediate benefit of monotonicity is surrounding the isotonic limit state surface Γ by two Pareto-dominated sets delimited by the elements of the numerical design, the measures of which defining deterministic bounds for the probability p [25]. Recent works focused

on the estimation of these bounds and the computation of statistical estimators of p based on Monte Carlo numerical designs [6].

While random designs appear as powerful tools for surrounding Γ and building estimators of p, the stochastic behavior of the associated random sets and their measures has not been investigated yet. The aim of this paper is to fill in this gap by establishing first convergence results for these objects from random set theory. A corollary of these results, under a convexity argument, is the consistency of an estimator of Γ built from the hull of the numerical design, accompanied with its convergence rate. Such an estimator is proposed by a combination of Support Vector Machines (SVM), which appears as a faster alternative in large dimensions to neural networks specifically developed for isotonic situations [32], while both techniques are usual in structural reliability frameworks [20].

This article is organized as follows. Section 2 details the main concepts and notations. The convergence of a dominated set is examined in Section 3. First we study conditions leading to convergence. Then, under some regularity assumption on the limit state, a rate of convergence for an estimator of Γ is provided in term of Hausdorff distance. Lastly, more precise results are given for two particular cases. Taking into account monotonicity properties, Section 4 focuses on a sequential sampling strategy. A non naive acceptance-rejection method to simulate in the so-called non-dominated set is provided . Then the convergence rates of bounds is compared numerically with a standard Monte Carlo and a sequential sampling. These theoretical results are derived in Section 5 to build a consistent semi-adaptive SVM-based classifier of the limit state surface Γ respecting isotonic and convexity constraints. Numerical experiments conducted in Section 5.2 complete this analysis by comparing the properties of the classifier with the recently developed monotonic neural networks. The proofs of the new technical results are postponed to Section 7.

2. Framework

Let consider as in [6] the space $\mathbb{U} = [0,1]^d$. Let X be a random vector uniformly distributed on $\mathbb{U}$ and g be a measurable application from $\mathbb{U}$ to $\mathbb{R}$. The function g discriminates two sets of input situations, leading to undesirable and safe events,

respectively. These classes are denoted $\mathbb{U}^- = \{\mathbf{x} \in [0,1]^d : g(\mathbf{x}) \leq 0\}$ and $\mathbb{U}^+ = \{\mathbf{x} \in [0,1]^d : g(\mathbf{x}) > 0\}$, with $[0,1]^d = \mathbb{U}^- \cup \mathbb{U}^+$. Hence the probability p that an undesirable event may occur is

$$p = \mathbb{P}(g(\mathbf{X}) \leq 0) = \mathbb{P}(\mathbf{X} \in \mathbb{U}^-).$$

Estimating p by a minimum number of calls to g is a prominent subject of interest in engineering. This number can be drastically diminished when g is assumed to be monotonous. Characterizing the monotonicity of g requires to use the Pareto dominance between two elements of $\mathbb{R}^d$, recalled in next definition.

Definition 1. Let $\mathbf{u} = (u_1, \ldots, u_d)$ and $\mathbf{v} = (v_1, \ldots, v_d) \in \mathbb{R}^d$. We say that $\mathbf{u}$ is (*resp.* strictly) dominated by $\mathbf{v}$, denoting $\mathbf{u} \preceq \mathbf{v}$ (*resp.* $\mathbf{u} \prec \mathbf{v}$), if for all $i = 1, \ldots, d$, $u_i \leq v_i$ (*resp.* $u_i < v_i$).

Assumption 1. *Let $g : \mathbb{U} \subset \mathbb{R}^d \to \mathbb{R}$. It is assumed that g is globally increasing, i.e. for all $\mathbf{u}, \mathbf{v} \in \mathbb{U}$ such that $\mathbf{u} \preceq \mathbf{v}$, then $g(\mathbf{u}) \leq g(\mathbf{v})$.*

Remark 1. Obviously as pointed in [6], any monotonic function can be reparametrized to be increasing with respect to all its inputs.

This property has for consequence to simplify the exploration of $\mathbb{U}^-$ and $\mathbb{U}^+$ and provide deterministic bounds on p. Let $\mathbf{x} \in \mathbb{U}^-$ (*resp.* $\mathbb{U}^+$) and $\mathbf{y} \in \mathbb{U}$. Since g is increasing, if $\mathbf{y} \preceq \mathbf{x}$ (*resp.* $\mathbf{y} \succeq \mathbf{x}$) then $\mathbf{y} \in \mathbb{U}^-$ (*resp.* $\mathbb{U}^+$). More generally, consider a set A of elements on $[0,1]^d$ (for instance a design of numerical experiments) and define

$$\mathbb{U}^-(A) = \bigcup_{\mathbf{x} \in A \cap \mathbb{U}^-} \{\mathbf{u} \in [0,1]^d : \ \mathbf{u} \preceq \mathbf{x}\},$$
$$\mathbb{U}^+(A) = \bigcup_{\mathbf{x} \in A \cap \mathbb{U}^+} \{\mathbf{u} \in [0,1]^d : \ \mathbf{u} \succeq \mathbf{x}\}.$$

Then the monotonicity property implies the following result, since $\mathbb{U}^-(A) \subset \mathbb{U}^- \subset [0,1]^d \backslash \mathbb{U}^+(A)$.

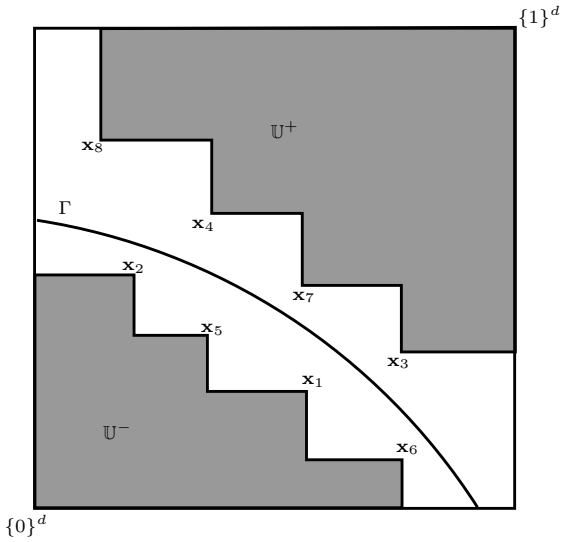

FIGURE 1: Illustration for $d = 2$ of the designs $\mathbb{U}^-(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_6)$ and $\mathbb{U}^+(\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_7, \mathbf{x}_8)$.

Proposition 1. *Denote μ the Lebesgue measure on $\mathbb{U}$. For any countable set A,*

$$\mu(\mathbb{U}^-(A)) \leq p \leq 1 - \mu(\mathbb{U}^+(A)).$$

Remark 2. In general, the set A will be $\{X_1, \ldots, X_n\}$, the sets $\mathbb{U}^-(A)$ and $\mathbb{U}^+(A)$ would then be random sets and the inequality becomes an almost sure inequality.

The *non-dominated* subset $\mathbb{U}(A) = [0,1]^d \backslash (\mathbb{U}^-(A) \cup \mathbb{U}^+(A))$ contains necessarily Γ and is the only subset of $[0,1]^d$ for which a deeper numerical exploration is needed in view of estimating Γ and p. Formally

$$\Gamma = \left(\overline{\mathbb{U}}^- \backslash \mathring{\mathbb{U}}^-\right) \cap \left(\overline{\mathbb{U}}^+ \backslash \mathring{\mathbb{U}}^+\right)$$

where $\overline{E}$ and $\mathring{E}$ are respectively the closure and the interior of a set E. Next (mild) assumption is required to view the problem of estimating Γ as a binary classification problem with perfectly separable classes, and is needed to study the convergence of the dominated sets to $\mathbb{U}^-$ and $\mathbb{U}^+$ when the number of elements of the design increases. A two-dimensional illustration is provided on Figure 1.

Assumption 2. *Assume that $\mu(\Gamma) = 0$ and Γ is simply connex.*

In the remainder of the article, a design of numerical experiments $\{X_i, g(X_i)\}_{i=1,\dots,n}$ is considered. To alleviate the notations denote for all $n \geq 1$,

$$\mathbb{U}_n^- = \mathbb{U}^- \{\mathbf{X}_1, \dots, \mathbf{X}_n\},$$
$$\mathbb{U}_n^+ = \mathbb{U}^+ \{\mathbf{X}_1, \dots, \mathbf{X}_n\},$$

and the non-dominated set $\mathbb{U}_n = [0,1]^d \backslash (\mathbb{U}_n^- \cup \mathbb{U}_n^+)$. The Lebesgue measures (or hypervolumes) of sets $(\mathbb{U}_n^-, \mathbb{U}_n^+)$ are then denoted p_n^- and $1 - p_n^+$, such that, from Proposition 1, we have with probability one

$$p_n^- \leq p \leq p_n^+. \tag{1}$$

These two bounds can be computed using a sweepline algorithm described in [6] at an exponential cost. A faster approximation method is provided in Section 4. Two situations are further investigated from the point of view of the convergence of dominated sets and bounds. First, the design is chosen static in $[0,1]^d$ (usual Monte Carlo sampling). Then a sequential Monte Carlo approach is explored, assuming $\mathbb{U}_0 = [0,1]^d$ and $\mathbf{X}_i$ being sampled within the current non-dominated set $\mathbb{U}_{i-1}$ for all $i > 0$. The convergence of the sets $\mathbb{U}_n^-$ and $\mathbb{U}_n^+$ will be studying via the Hausdorff distance defined as follows.

Definition 2. Let $\|\cdot\|_q$ be the $\mathcal{L}^q$ norm on $\mathbb{R}^d$, that is for $0 < q < +\infty$, $\|\mathbf{x}\|_q = (\sum_{i=1}^d |x_i|^q)^{1/q}$, and for $q = +\infty$, $\|\mathbf{x}\|_\infty = \max_{i=1,\dots,d} x_i$. Let (A, B) be two non-empty subsets of the normed vector space $([0,1]^d, \|\cdot\|_q)$. The Hausdorff distance $d_{H,q}$ is defined by

$$d_{H,q}(A,B) = \max(\sup_{\mathbf{y}\in A} \inf_{\mathbf{x}\in B} \|\mathbf{x}-\mathbf{y}\|_q;\ \sup_{\mathbf{x}\in B} \inf_{\mathbf{y}\in A} \|\mathbf{x}-\mathbf{y}\|_q).$$

It is always finite since $(A, B) \subset \left([0,1]^d\right)^2$ are bounded. The usual form of the Hausdorff distance is defined for $q = 2$. To alleviate notations, it is now denoted $d_H(.,.) = d_{H,2}(.,.)$ and $\|.\| = \|.\|_2$.

3. Convergence results

3.1. Almost sure convergence for general sample strategy

The sequences $(p_n^-)_n$ and $(p_n^+)_n$ are respectively increasing bounded from above and decreasing bounded from below. Hence, there exists two random variables p_∞^- and p_∞^+ such that almost surely $p_n^- \longrightarrow p_\infty^-$, $p_n^+ \longrightarrow p_\infty^+$ and $p_\infty^- \leq p \leq p_\infty^+$. In this section we provide general conditions on the design of experiments that ensures that $p_\infty^- = p_\infty^+ = p$.

Proposition 2. *[Independant Sampling] Let $(\mathbf{X}_k)_{k\geq 1}$ be a sequence of independent random vectors on $[0,1]^d$. If there exists $\varepsilon_1 > 0$ such that for all $\mathbf{x} \in \Gamma^{\varepsilon_1} = \{\mathbf{u} \in [0,1]^d,\ d(\mathbf{u},\Gamma) < \varepsilon_1\}$ there exists $\varepsilon_2 > 0$ such that*

$$\sum_{n\geq 1} \mathbb{P}(\mathbf{X}_n \in B(\mathbf{x},\varepsilon_2)) = +\infty,$$

then $(p_n^-, p_n^+) \overset{a.s.}{\underset{n\to+\infty}{\to}} (p,p)$.

Corollary 1. *Under Assumption 2 and the conditions of Proposition 2, then*

$$\big(d_H(\mathbb{U}_n^-, \mathbb{U}^-),\ d_H(\mathbb{U}_n^+, \mathbb{U}^+)\big) \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} (0,0).$$

Example 1. If a sequence $(X_n)_{n\in I\!N^*}$ that is i.i.d. and uniformly distributed on $[0,1]^d$ then it satisfies the assumption of Proposition 2. More generally any sequence of iid random variables with bounded from below densities on $[0,1]^d$ satisfies the assumption of Proposition 2.

Since the only useful experiments are those which fall into the non-dominated set $\mathbb{U}_n$, a sequential Monte Carlo strategy can be carried out, as detailed in next proposition.

Proposition 3. *[Sequential sampling] Let $(\mathbf{Y}_n)_{n\geq 1}$ be a sequence of iid random variables uniformly distributed on $[0,1]^d$. Let $(T_n)_{n\geq 1}$ be a sequence of random variables on $\mathbb{N}$ defined by $T_1 = 1$ and for all $n \geq 1$*

$$T_{n+1} = \min\{j > T_n,\ \mathbf{Y}_j \in \mathbb{U}_{T_n}\}.$$

Let $(\mathbf{X}_n)_{n\geq 1}$ *be the sequence of random variables defined for all* $n \geq 1$ *by* $\mathbf{X}_n = \mathbf{Y}_{T_n}$. *Then, conditionally to* $\mathbf{X}_1, \ldots, \mathbf{X}_{n-1}$, $\mathbf{X}_n \sim \mathcal{U}(\mathbb{U}_{T_{n-1}})$, *and* $(p_n^-, p_n^+) \overset{a.s.}{\to} (p, p)$, *where* (p_n^-, p_n^+) *are obtained from the sequence* $(\mathbf{X}_n)_{n\geq 1}$.

It is now natural to quantify the rate of convergence. While it seems difficult to provide an explicit rate in the general case. We will then provide some explicit rate in particular cases.

3.2. Rate of convergence under some regularity assumptions

In this section convergence in Haussdorf distance sense of $(\mathbb{U}_n^-)_n$ towards $\mathbb{U}^-$ is proven, and under regularity assumption for the set $\mathbb{U}^-$ a rate of convergence is provided. The regularity considered here is the (α, γ)-regularity introduced in [1].

Definition 3. A set K is said (α, γ)-regular if there exist $\alpha, \gamma > 0$ such that for all $0 < \varepsilon \leq \gamma$ and for all $\mathbf{x} \in K$ one has

$$\mu(\mathbf{B}(\mathbf{x}, \varepsilon) \cap K) \geq \alpha \mu(\mathbf{B}(\mathbf{x}, \varepsilon))$$

where $\mathbf{B}(\mathbf{x}, r)$ is the open ball with center $\mathbf{x}$ and radius r.

This notion was introduced in [1] to prove minimax rates in classification problems. It has also been considered in [8, 26] in order to state convergence rate when estimating sets. In particular Proposition 4 below can be seen as an application of Theorem 2 in [26].

Roughly speaking, this condition excludes pathological sets $\mathbb{U}^-$, as those having infinitely many sharper and sharper peaks or presenting a jagged border Γ. For instance, it is indirectly proved in [12] that if $\mathbb{U}^-$ is convex and has non-empty interior then it is regular. In a structural reliability framework this condition is conservative since it traduces the hypothesis that any combination of inputs located between two situations leading to an undesirable event leads also to such an event. Described in [34], so-called $r-$convex sets that generalize the notion of convexity are also regular [26]. It is traduced in $\mathbb{U}^-$ by the assumption (or "rolling condition", cf. [34]) that for each boundary point $\mathbf{x} \in \Gamma$, there exists $\mathbf{y} \in \mathbb{U}^-$ such that $\mathbf{x} \in B(\mathbf{y}, r)$. In practice the assumption of $r-$convexity appears mild, as small values of r can produce a large

variety of limit state surfaces presenting irregularities [26].

Proposition 4. *Let $(\mathbf{X}_k)_{k\geq 1}$ be a sequence of iid random variables uniformly distributed on $[0,1]^d$ and $(\tilde{\mathbf{X}}_k)_{k\geq 1}$ be a sequence of iid random variables uniformly distributed on $\mathbb{U}^-$. Denote $\tilde{\mathbb{U}}_n^- = \mathbb{U}^-(\tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_n)$. Let $(F_n)_{n\geq 1}$ be a sequence of measurable subsets of $[0,1]^d$ such that for all $n \geq 1, \mathbb{U}_n^- \subset F_n \subset [0,1]^d \backslash \mathbb{U}_n^+$. Then*

(1) *$d_H(F_n, \mathbb{U}^-) \xrightarrow[n\to+\infty]{a.s.} 0$ and $\mu(F_n) \xrightarrow[n\to+\infty]{a.s.} p$.*

(2) *If $\mathbb{U}^-$ is regular, then almost surely*

$$d_H(\tilde{\mathbb{U}}_n^-, \mathbb{U}^-) = O\left((\log n/n)^{1/d}\right).$$

(3) *Furthermore, if $\mathbb{U}^+$ is also regular, and if g is continuous, then almost surely*

$$d_H(F_n, \mathbb{U}^-) = O\left((\log n/n)^{1/d}\right). \tag{2}$$

Remark 3. Note, that if in **(2)**, we replace $(\tilde{\mathbb{U}}_n^-)$ by $(\mathbb{U}_n^-)$, the a.s. bound becomes $O\left((\log N_1/N_1)^{1/d}\right)$, where N_1 is the number of points X_i such that $g(X_i)) < 0$ and follows the binomial distribution with parameter n and p. Moreover the continuity assumption in **(3)** can be weakened, it is enough (see the proof) that one can cut $[0,1]^d$ following $g^{-1}(\{0\})$ and glue it.

3.3. Convergence results in dimension 1

In dimension 1, we provide a complete description of the rate of convergence, in particular we prove that $n(p - p_n^-)$ and $n(p_n^+ - p)$ are asymptotically exponentially distributed.

Proposition 5. *Let $(X_n)_{n\geq 1}$ be an iid sequence of random variables uniformly distributed on $[0,1]$ and $p \in [0,1]$. Define $p_n^- = \max_{i=1,\ldots,n}(X_i \cdot \mathbb{1}_{\{X_i \leq p\}})$ and $p_n^+ =$*

$\mathbb{1}_{\{\max_{i=1,\dots,n}(X_i)\leq p\}} + \min_{i=1,\dots,n}(X_i \cdot \mathbb{1}_{\{X_i>p\}})$. *Then*

$$\begin{aligned} n(p - p_n^-) &\xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}xp(1), \\ n(p_n^+ - p) &\xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}xp(1), \\ \mathbb{E}[p_n^+ - p_n^-] &= \frac{2}{n+1} - \frac{1}{n+1}\left(p^{n+1} + (1-p)^{n+1}\right), \end{aligned} \tag{3}$$

where $\mathcal{E}xp(\lambda)$ *is the exponential distribution with density* $f_\lambda(x) = \lambda \exp(-\lambda x)\mathbb{1}_{\{x\geq 0\}}$.

The cost of adopting a naive sequential strategy to decrease the volume of the non-dominated set, by sampling a new design element within $[0,1]^d$, can be appreciated by coming back to the unidimensional case ($d = 1$). In this framework, denote $W_n = \min\{j \geq 1 :\ X_{n+j} \in]p_n^-, p_n^+[\}$ where $(X_{n+j})_{j\geq 1}$ is an iid sequence of random variables uniformly sampled on $[0,1]$. For all $r \geq 1$, one has

$$\begin{aligned} \mathbb{P}(W_n = r) &= \mathbb{P}(X_{n+1} \notin]p_n^-, p_n^+[, \dots, X_{n+r-1} \notin]p_n^-, p_n^+[, X_{n+r} \in]p_n^-, p_n^+[), \\ &= \mathbb{E}[\mathbb{P}(X_{n+1} \notin]p_n^-, p_n^+[, \dots, X_{n+r-1} \notin]p_n^-, p_n^+[, X_{n+r} \in]p_n^-, p_n^+[| X_1, \dots, X_n)], \\ &= \mathbb{E}[(1 - (p_n^+ - p_n^-))^{r-1}(p_n^+ - p_n^-)]. \end{aligned}$$

From linearity of the expectation and Jensen inequality, it comes

$$\begin{aligned} \mathbb{E}[W_n] &= \mathbb{E}[(p_n^+ - p_n^-)^{-1}], \\ &\geq \mathbb{E}[p_n^+ - p_n^-]^{-1}, \\ &\geq \frac{n+1}{2} \quad \text{from (3)} \end{aligned} \tag{4}$$

which is (as expected) a prohibitive cost for large values of n.

More accurate results can be only obtained in some particular case. The following proposition provides a result of the expected Lebesgue measure of the non-dominated set for $d = 1$.

Proposition 6. *Let* $p_0^- = 0$, $p_0^+ = 1$, $X_1 \sim \mathcal{U}([p_0^-, p_0^+])$, $\xi_1 = \mathbb{1}_{\{X_1\leq p\}}$ *and* $\mathcal{F}_1 =$

$\sigma\{X_1\}$. *For all* $n \geq 1$ *define conditionally to* $\mathcal{F}_n = \sigma\{X_k, 1 \leq k \leq n\}$*:*

$$\begin{aligned} X_{n+1} &\sim \mathcal{U}([p_n^-, p_n^+]), \\ \xi_{n+1} &= \mathbb{1}_{\{X_{n+1} \leq p\}} \\ p_{n+1}^- &= p_n^- + (X_{n+1} - p_n^-)\xi_{n+1}, \\ p_{n+1}^+ &= p_n^+ - (p_n^+ - X_{n+1})(1 - \xi_{n+1}). \end{aligned}$$

Then, for $n \geq 1$*,*

$$\frac{1}{2^n} \leq \mathbb{E}[p_n^+ - p_n^-] \leq \left(\frac{3}{4}\right)^n.$$

Corollary 2. *If* $p \in \{0, 1\}$, $\mathbb{E}[p_n^+ - p_n^-] = 2^{-n}$.

Remark 4. The shrinking convergence rate, which is inversely linear in n in a static Monte Carlo strategy, is significantly improved by becoming exponential when opting for a sequential strategy.

3.4. Asymptotic results when $\Gamma = \{1\}^d$

Equivalent results are difficult to obtain for greater dimension and for a general limit state Γ. The followings proposition provides asymptotic results in the particulare case $\Gamma = \{1\}^d$.

Proposition 7. *Assume* $\Gamma = \{1\}^d$*. Let* $(\mathbf{X}_k)_{k \geq 1}$ *be a sequence of iid random variables uniformly distributed on* $[0, 1]^d$*. For* $0 < q < +\infty$ *denote* $A(1, q) = 1$ *and for* $d \geq 2$*,* $A_{d,q} = \frac{1}{dq^{d-1}} \prod_{i=1}^{d-1} B(i/q, 1/q)$ *with* $B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$*. For all* $n \geq 1$*, let* $\mathbb{U}_n^- = \mathbb{U}^-(\mathbf{X}_1, \ldots, \mathbf{X}_n)$.

(1) *If* $0 < q < +\infty$ *then*

$$(A_{d,q}n)^{1/d} d_{H,q}\left(\mathbb{U}_n^-, [0, 1]^d\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{W}(1, d).$$

(2) *If* $q = +\infty$ *then*

$$n^{1/d} d_{H,\infty}\left(\mathbb{U}_n^-, [0, 1]^d\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{W}(1, d),$$

where $\mathcal{W}(1, d)$ *is the Weibull distribution with scale parameter* 1 *and shape parameter*

d having cumulative density function $F(t) = 1 - e^{-t^d}$ *for all* $t \geq 0$.

Proposition 8. *Under the assumptions and notations of Proposition 7 we have*

$$\mathbb{E}[\mu([0,1]^d \backslash \mathbb{U}_n^-)] \underset{n \to +\infty}{\sim} \frac{\log(n)^{d-1}}{n(d-1)!}.$$

Remark 5. Denoting $C_\infty(K)$ the convex hull of a set K, the result of Proposition 8 is tantamount to the following result provided in [3]:

$$\mathbb{E}[\mu([0,1]^d \backslash C_\infty(\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}))] \underset{n \to +\infty}{\sim} \frac{\log(n)^{d-1}}{n}.$$

When $d = 1$, the convergence order obtained in Proposition 5 is retreived.

4. Sequential sampling

Using a naive methods to simulate in a constrained space can be time consuming. Indeed, Equation 4 gives the expected number of simulation needed to simulate in the non-dominated set in dimension 1. A method to simulate in the non-dominated set is provided in the first part of this section. In a second part, the measure of the non-dominated set is compared in function of the strategies of simulation adopted: standard Monte Carlo or sequential sampling.

4.1. Fast uniform sampling within the non-dominated set

Simulating uniformly on $[0,1]^d$ can be time consuming when the aim is to simulate in the non-dominated set. The strategy provided in this section involves to build a subset of $[0,1]^d$ containing the non-dominated set then to simulate uniformly within this subset. After n calls to g let $\mathbf{x}$ in $\mathbb{U}_n^+$. From construction, the non-dominated set is in $[0,1]^d \backslash \mathbb{U}_n^+(\mathbf{x})$, it remains to simulate uniformly on this subset of $[0,1]^d$. The choice of $\mathbf{x}$ is described further in this section. The proposed method use the following lemma coming from a simple acceptance-rejection method.

Lemma 1. *Let* A *be a measurable subset of* $[0,1]^d$ *with* $\mu(A) > 0$ *and such that* $A = A_1 \cup \cdots \cup A_m$ *with* $A_i \cap A_j = \emptyset$ *if* $i \neq j$. *Let* $(\mathbf{X}_k)_{k \geq 1}$ *be an iid sequence of random vectors satisfying for all* i, $\mathbb{P}(\mathbf{X}_k \in A_i) = \mu(A_i) \backslash \mu(A)$. *Let* C *be a measurable subset*

of A, and $T = \inf\{k \geq 1, \mathbf{X}_k \in C\}$ then $\mathbf{X}_T$ is uniformly distributed on C.

The construction involves to split up $[0,1]^d$ in hyperrectangles (see Figure 2) which are defined for all $i = 0, \ldots, 2^d - 1$ by

$$Q_i(\mathbf{x}) = I_i^1 \times I_i^2 \times \cdots \times I_i^d,$$

with

$$I_i^j = \begin{cases} [x^j, 1] & \text{if } b_j^i = 0, \\ [0, x^j] & \text{if } b_j^i = 1, \end{cases}$$

where $b^i = (b_1^i, \ldots, b_d^i)$ is equal to i coded in base 2. Since 0 is coded as $00\cdots0$ in base 2 with d numbers, one deduce that $Q_0(\mathbf{x}) = [x^1, 1] \times [x^2, 1] \times \cdots \times [x^d, 1] = \mathbb{U}^+(\mathbf{x})$. Let $(\mathbf{X}_k)_{k\geq 1}$ be an iid sequence of random vectors satisfying for all $i = 1, \ldots, 2^d - 1$

$$\mathbb{P}(\mathbf{X} \in Q_i(\mathbf{x})) = \frac{\mu(Q_i(\mathbf{x}))}{1 - \mu(Q_0(\mathbf{x}))}. \tag{5}$$

Define $T = \inf\{k \geq 1, \mathbf{X}_k \in \mathbb{U}_n\}$.Applying Lemma 1 with $A = Q_1(\mathbf{x}) \cup \cdots \cup Q_{2^d-1}(\mathbf{x})$ and C the non-dominated set $\mathbb{U}_n$, one deduce that $\mathbf{X}_T$ is uniformly distributed on the non-dominated set. The point $\mathbf{x}$ is chosen in order to maximise the quantity in Equation (5). It must be noticed that more p is close to 0 more the points of Ξ_n^+ are close to $\{0\}^d$ then the probability to have a simulation in the non-dominated set increases.

4.2. Numerical study

To complete the previous theoretical studies, a brief numerical comparison of convergence rates obtained for some static and sequential strategies is conducted in this paragraph. The impact of design strategies on the mean measure of the non-dominated set $\mathbb{U}_n$ is examined through two repeated experiments conducted over the following model, defined such that $\mathbb{U}^- = [0,1]^d$, $p = 1$ and $\Gamma = \{1\}^d$ for various dimensions

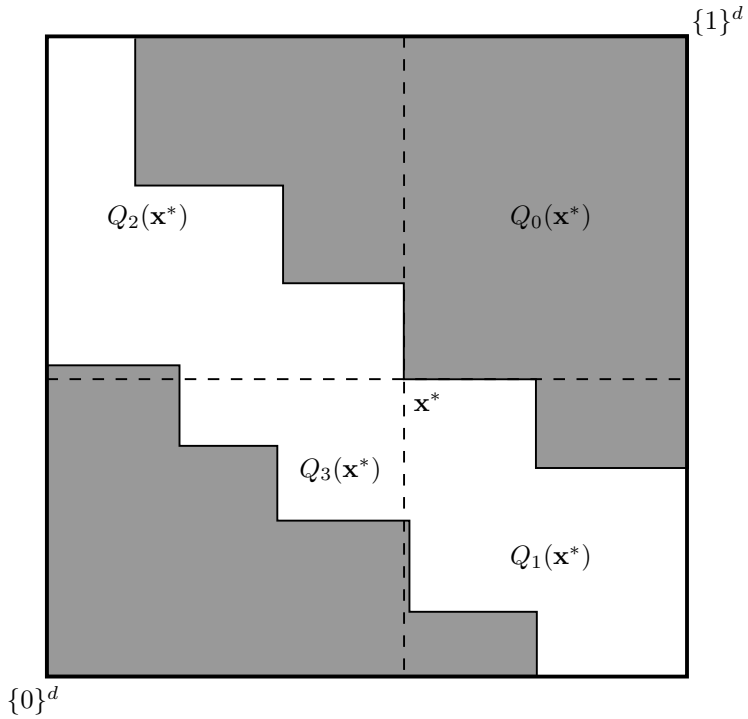

FIGURE 2: Splitting up $[0,1]^d$ from $\mathbf{x}^*$ in dimension 2. The hyperrectangle $Q_0(\mathbf{x}^*)$ is in $\mathbb{U}^+$.

$d \in \{1, 2, 3, 4, 5, 10\}$. That means, after n simulations, the volume of

$$\bigcup_{\mathbf{x} \in \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}} \{\mathbf{u} \in [0,1]^d : \ \mathbf{u} \preceq \mathbf{x}\},$$

is compared, where $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are respectively obtained from a uniform and a sequential sampling strategy. The model is run $n = 100$ times per experiment and repeated 100 times. The results of these experiments are summarized on Table 1. As expected, the remaining volume is lower with a sequential Monte Carlo than a standard Monte Carlo. Nonetheless, this difference decreases when the dimension increases and the remaining volume is nearly equal to 1 in dimension 10 for the two Monte Carlo simulations.

Dimension	1	2	3	4	5	10
Monte Carlo strategy	9.9×10^{-3}	0.052	0.145	0.28	0.43	0.937
Sequential strategy	7.89×10^{-31}	4.17×10^{-5}	0.017	0.14	0.33	0.935

TABLE 1: Mean measures (volumes) of non-dominated set $\mathbb{U}_n$ after $n = 100$ runs.

Remark 6. Table 1 shows that the gain provided by a sequential sampling strategy decrease as the dimension increase. For $d = 10$, the remaining volume obtain with the two strategies of simulations are close.

5. Application : estimating Γ using Support Vector Machines (SVM)

The previous sections provide convergence results for different strategies of simulations without the estimation of the limit state Γ. This section provide an estimator of the sign of g under some convexity assumption on $\mathbb{U}^-$. The proposed estimator is in fact a classifier based on linear SVMs. It will then be compared on a toy example, with the monotonic neural networks recently developed in [32].

5.1. Theoretical study

Dominated sets $(\mathbb{U}_n^-, \mathbb{U}_n^+)$ surrounding Γ can be improved by sampling within the non-dominated set $\mathbb{U}_n$. In view of improving a naive Monte Carlo approach to estimate p, as in [6], usual techniques like importance sampling or subset simulation should aim at targeting input situations close to Γ [5]. Such approaches can be guided by a consistent estimation of Γ, under the form of a supervised binary classification rule calibrated from $(\mathbb{U}_n^-, \mathbb{U}_n^+)$. This classifier has to agree with the following (isotonic) ordinal property of the limit state surface Γ.

Proposition 9. *Under Assumption 2 For all* $\mathbf{u}, \mathbf{v} \in \Gamma$ *such that* $\mathbf{u} \neq \mathbf{v}$, $\mathbf{u}$ *is not strictly dominated by* $\mathbf{v}$.

Respecting this constraint a monotonic neural network classifier was recently proposed in [32] and applied in [6] to structural reliability frameworks. While consistent, its computational cost remains high or even prohibitive when the size of the design $\mathbf{X}_1, \ldots, \mathbf{X}_n$ defining $(\mathbb{U}_n^-, \mathbb{U}_n^+)$ increases. Benefiting from a clear geometric interpretation, Support Vector Machines (SVM) offer an alternative to neural networks by their

robustness to the curse of dimensionality [20]. A semi adaptive solution can be build from a combination of SVM when $\mathbb{U}^-$ is convex. Conversely, it can be easily adapted when $\mathbb{U}^+$ is a convex set.

Assuming $\mathbb{U}^-$ is convex, any points $\mathbf{x}$ of $\mathbb{U}^+$ can be separated from $\mathbb{U}^-$ by a hyperplane $h_{\mathbf{x}}(\mathbf{u}) = \alpha + \boldsymbol{\beta}_{\mathbf{x}}^T\mathbf{u}$ (see [24] Theorem 11.5) that maximises the minimal distance of $h_{\mathbf{x}}$ to $\mathbf{x}$ and $\mathbb{U}^-$. It is also possible to construct h satisfying the ordinal property on Assumption 1 (see Appendix B for more details).

Given the numerical experiment $D_n = (\mathbf{X}_i, y_i)_{1\leq i\leq n} \in \mathbb{U}\times\{-1,1\}\}$ where $y_i = 1$ if $g(\mathbf{X}_i) > 0$ and -1 otherwise. Let $\Xi_n^+ = \{\mathbf{X}_1,\ldots,\mathbf{X}_n\}\cap\mathbb{U}^+$ and $\Xi_n^- = \{\mathbf{X}_1,\ldots,\mathbf{X}_n\}\cap\mathbb{U}^-$ and for any $\mathbf{x}\in\Xi_n^+$ define $h_{\mathbf{x}}$ as the hyperplane separating $\mathbf{x}$ from Ξ_n^-. The proposed classifier f_n is defined by

$$\begin{aligned} f_n : [0,1]^d &\to \{-1,+1\} \\ \mathbf{y} &\mapsto \begin{cases} -1 & \text{if for all } \mathbf{X}\in\Xi_n^+,\ h_{\mathbf{X}}(\mathbf{y})\leq 0 \\ +1 & \text{otherwise.} \end{cases} \end{aligned}$$

Denote

$$F_n = \{\mathbf{x}\in[0,1]^d :\ f_n(\mathbf{x}) = -1\}$$

the set of all inputs classified as leading to undesirable events by f_n.

Theorem 1. *Assume $\mathbb{U}^-$ is convex, then*

(1) *f_n is globally increasing.*

(2) *For all $\mathbf{X}\in\{\mathbf{X}_1,\ldots,\mathbf{X}_n\}$, $sign(g(\mathbf{X})) = f_n(\mathbf{X})$.*

(3) *The set F_n is a convex polyhedron.*

(4) *Furthermore if $(\mathbf{X}_k)_{k\geq 1}$ is a sequence of independent random vectors uniformly distributed on $[0,1]^d$, then*

$$d_H(F_n,\mathbb{U}^-) \xrightarrow[n\to+\infty]{a.s.} 0,$$

and almost surely,

$$d_H(F_n, \mathbb{U}^-) = O\left((\log n/n)^{1/d}\right).$$

Updating the classifier given a new design element $\mathbf{X}_{n+1}$ found in $\mathbb{U}^+$ can be done by a partially (semi) adaptive principle, illustrated on Figure 3 in dimension 2. To get f_{n+1} it is enough to build the hyperplane $h_{\mathbf{X}_{n+1}}$ which separates $\mathbf{X}_{n+1}$ from $\Xi^-_{n+1} = \Xi^-_n$ (unfortunately, if $\mathbf{X}_{n+1} \in \mathbb{U}^-$ all hyperplanes must be rebuild, but this occurs rarely with low probability $p - p^-_n$ when the design is sampled uniformly on $[0,1]^d$). If $\mathbf{X}_{n+1} \not\preceq \mathbf{x}$ for all $\mathbf{x} \in \Xi^-_n$, the support vectors and the hyperplanes will not be modified.

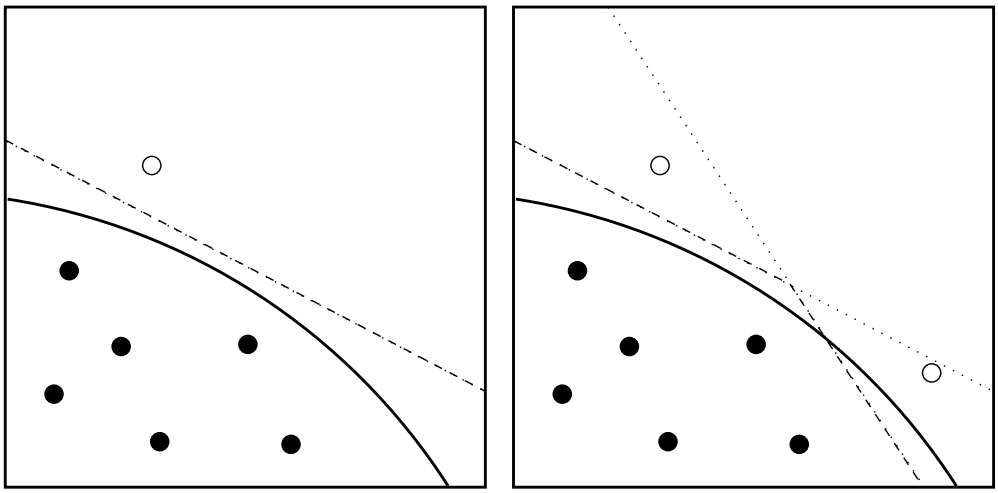

FIGURE 3: Construction of the classifier based on SVM. The plain line represent Γ and black (resp. white) points are in $\mathbb{U}^-$ (resp. $\mathbb{U}^+$). The dotted lines represents $\{\mathbf{x} : h_{\mathbf{X}}(\mathbf{x}) = 0\}$ for some $\mathbf{X}$ in $\mathbb{U}^+$. **Left**: there is one point in $\mathbb{U}^+$ then the dashed line is both $\{\mathbf{x} : h_{\mathbf{X}}(\mathbf{x}) = 0\}$ and the frontier of the classifier. **Right**: Dotted lines represent the two sets $\{\mathbf{x} : h_{\mathbf{X}}(\mathbf{x}) = 0\}$ for $\mathbf{X}$ represented by the two white points. Dashed line represent the frontier of the classifier.

5.2. Numerical experiments

This section examines first the gain of including the constraint of monotonicity in the calibration of SVM (detailed in Appendix B), with respect to usual SVM and the constrained neural networks proposed in [32] when the associated code is globally monotonic.

First, the case of a linear limit state is studied. Next, the comparison will be made on a function which verifies the monotonicity and convexity properties. The rate of good classification obtained with the monotonic classifier and the monotonic neural networks are compared at each step of the algorithm.

5.2.1. *Linear case.* Investigating the efficiency of the new classifier, let start the numerical experiments with the linear case. The function g is a hyperplane defined by

$$h(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^T \mathbf{x},$$

with positive parameters which are, in dimension d, uniformly distributed on the unit sphere

$$\mathbb{S}_d^+ = \{\mathbf{x} \in I\!R_+^d : \ \|\mathbf{x}\| = 1\}.$$

Various sample size N are used to examine the rate of good classification when the monotonicity is taking account or not. For each ordered pair (d, N), a hyperplane is build with parameters uniformly distributed on $\mathbb{S}_d^+$. Then, 200 data are simulated on the unit sphere to be predicted. The rate of good classification is compared for linear SVM and linear monotonic SVM. That step is repeated 100 times for each hyperplane, and also repeated with 100 differents hyperplane.

The results are summarised by Table 2 where there is respectively a proportion 0.1, 0.2, 0.3, 0.4 and 0.5 among N which are in $\{\mathbf{x} : \ g(\mathbf{x}) \leq 0\}$. In general, taking account of the monotonicity provides better results. As expected, more N is great and d is low more the rate of good classification increase for the two methods. Results are comparable for $d = 2$ but for a fixed N the difference grows when d increase. Finally, the constrained SVM provides significantly better results for small N.

5.2.2. *Convex case.* In this section section, the classifier given in Theorem 1 is tested on a toy example. Let $\mathbf{U} = (U^1 \ldots, U^d)$ be uniformly distributed on $[0, 1]^d$, and denote

$$Z_d = (U^1 U^2 \cdots U^d)^2.$$

	$d=2$	$d=10$	$d=20$	$d=40$	$d=50$	$d=100$
$N=10$	78.24/**78.40**	58.14/**62.72**	53.05/**57.65**	50.80/**53.53**	50.49/**52.32**	50.09/**50.88**
$N=20$	84.04/**84.19**	63.77/**68.99**	56.66/**62.20**	52.37/**56.83**	51.80/**55.85**	50.40/**52.83**
$N=40$	91.22/**91.34**	71.65/**74.13**	62.64/**67.38**	56.25/**61.44**	54.59/**59.91**	51.68/**55.75**
$N=50$	93.04/**93.07**	73.92/**76.41**	65.04/**70.05**	57.43/**63.04**	56.16/**61.65**	52.36/**56.87**
$N=75$	92.57/**92.74**	78.72/**80.34**	68.95/**72.59**	60.24/**66.02**	58.71/**63.99**	53.73/**58.58**
$N=100$	94.41/94.41	83.36/**84.43**	73.66/**76.95**	64.50/**69.03**	62.11/**67.26**	55.54/**60.96**
$N=200$	97.09/97.09	88.87/**89.30**	82.09/**83.55**	72.83/**76.07**	69.55/**73.22**	61.24/**66.50**
$N=10$	84.86/**85.73**	62.79/**68.05**	57.78/**63.67**	53.08/**58.46**	52.75/**57.92**	50.95/**54.41**
$N=20$	90.41/**90.44**	70.27/**74.69**	61.98/**68.49**	57.64/**63.62**	55.88/**62.30**	52.56/**57.87**
$N=40$	93.65/**93.67**	78.89/**80.90**	70.51/**74.88**	62.48/**68.38**	60.23/**66.30**	55.44/**61.93**
$N=50$	95.51/95.51	82.06/**83.78**	72.30/**76.28**	64.84/**70.50**	62.62/**68.35**	57.01/**63.39**
$N=75$	96.28/**96.29**	85.36/**86.99**	77.63/**80.25**	69.12/**73.65**	66.41/**71.86**	59.88/**66.15**
$N=100$	96.87/96.87	88.44/**89.18**	81.04/**82.94**	72.39/**76.21**	69.71/**74.30**	62.02/**67.80**
$N=200$	98.48/**98.49**	93.37/**93.63**	88.37/**89.19**	81.10/**83.18**	78.02/**80.40**	69.65/**74.13**
$N=10$	86.55/**87.42**	67.76/**73.2**	60.98/**68.26**	56.49/**63.74**	55.67/**62.95**	52.86/**59.30**
$N=20$	92.39/**92.68**	74.94/**78.36**	67.68/**73.32**	61.13/**68.43**	58.87/**66.38**	55.40/**62.69**
$N=40$	95.27/**95.40**	83.20/**85**	75.22/**78.75**	66.73/**72.70**	65.15/**71.56**	59.24/**66.67**
$N=50$	95.35/**95.37**	85.12/**86.46**	77.81/**81.04**	69.39/**74.81**	66.72/**72.89**	61.00/**68.37**
$N=75$	97.06/97.06	88.88/**89.42**	82.24/**84.51**	73.64/**78.11**	71.40/**76.19**	63.80/**70.19**
$N=100$	97.90/97.90	91.55/**91.79**	85.68/**86.71**	77.47/**80.77**	74.69/**78.78**	66.59/**72.88**
$N=200$	98.73/98.73	95.16/**95.26**	91.53/**91.94**	85.40/**86.99**	82.95/**85.09**	74.62/**79.08**
$N=10$	88.53/**89.14**	68.99/**74.31**	64.20/**71.35**	59.36/**67.83**	58.06/**66.78**	55.23/**64.17**
$N=20$	92.10/**92.22**	77.54/**81.02**	70.60/**75.69**	63.92/**71.54**	62.88/**70.56**	58.04/**66.88**
$N=40$	95.25/**95.26**	85.06/**86.52**	77.75/**81.31**	70.62/**76.28**	68.48/**74.74**	62.33/**70.43**
$N=50$	97.20/**97.20**	87.02/**87.74**	80.91/**83.52**	72.58/**77.60**	70.60/**75.88**	63.98/**71.50**
$N=75$	97.51/**97.51**	90.99/**91.54**	84.84/**86.84**	76.90/**80.92**	74.86/**79.32**	67.58/**74.37**
$N=100$	97.98/**97.99**	92.86/**93.06**	87.65/**88.57**	80.11/**83.03**	77.88/**81.74**	70.57/**76.39**
$N=200$	98.81/**98.81**	95.95/**96.07**	93.50/**93.83**	88.27/**89.44**	86.55/**88.18**	78.6/**82.36**
$N=10$	88.91/**89.52**	70.12/**75.03**	64.23/**71.39**	59.93/**68.41**	59.17/**67.97**	56.88/**66.48**
$N=20$	93.44/**93.59**	77.65/**80.22**	71.63/**76.71**	65.03/**72.34**	63.77/**71.59**	59.74/**68.46**
$N=40$	95.85/95.85	86.33/**87.23**	78.95/**81.88**	71.58/**77.11**	69.16/**75.35**	64.00/**71.87**
$N=50$	96.28/**96.36**	88.25/**89.00**	81.36/**84.56**	74.22/**79.23**	71.72/**77.56**	65.78/**73.53**
$N=75$	97.51/**97.51**	91.01/**91.61**	85.96/**87.60**	78.60/**82.51**	75.85/**80.09**	69.66/**76.18**
$N=100$	98.04/**98.04**	93.18/**93.46**	89.03/**89.81**	81.92/**84.43**	79.55/**82.73**	72.08/**78.06**
$N=200$	99.07/**99.08**	96.18/**96.27**	93.57/**94.12**	89.05/**90.12**	87.16/**88.70**	80.09/**83.76**

TABLE 2: Rate of good classification for usual SVM (left) and monotonic SVM (right) in function of d and N. From up to down, there is respectively a proportion of 0.1, 0.2, 0.3, 0.4, 0.5 in the set $\{\mathbf{x} : \ g(\mathbf{x}) \leq 0\}$.

Let $q_{d,p}$ be the p-quantile of Z_d and define the function $g(\mathbf{U}) = Z_d - q_{d,p}$. This quantile can be deduced from Equation (13) in Section 7. Indeed, let $t \in [0,1]$, then

$$\mathbb{P}(Z_d \leq t) = \mathbb{P}(U^1 U^2 \cdots U^d \leq \sqrt{t}) = \int_0^{\sqrt{t}} \frac{(-\log(x))^{d-1}}{(d-1)!} dx.$$

The function g is globally increasing and the set $\{\mathbf{u} \in [0,1]^d, g(\mathbf{u}) > 0\}$ is convex. The SVM-based classifier and the constrained neural networks are built from the points of a sequential design used to delimit the non-dominated set. At step n, the number of points which delimits this non-dominated set is denoted N_n.

The comparison is conducted as follows. At each step $n \geq 1$, $M = 500$ random vectors are uniformly distributed on the non-dominated set $\mathbb{U}_{n-1}$. The rate of good classification on these M random vectors and the time needed to build the two classifiers are stored. Then non-dominated set is updated with a random vector uniformly distributed on it. The comparison is conducted for different dimensions: 2, 3 , 4 and 5 and for $p = 0.01$ and have been averaged on 40 independent experiments.

In Figure 4 the rate of good classification is compared in function of N_n. The red values represent the number of times the situation where the non-dominated set is delimited by N_n points. For some $n \geq 1$, this number is greater than 40. Indeed, the sequence $(N_n)_{n\geq 1}$ is not monotonic: at any step n, a new simulation employed to update the non-dominated set $\mathbb{U}_{n-1}$ can dominate one of the point of its frontier. Then the value of N_n can be lower than N_{n-1}. At step n, having $N_n = n$ means no points of the sequential design is dominated by one another. The results show that the mean rate of good classification is equivalent for the two classifier and for the tested dimension. But, when the dimension increases the SVM-based classifier is more stable than the constrained neural networks.

Denote $(T_n^{SVM})_{n\geq 1}$ and $(T_n^{NN})_{n\geq 1}$ respectively the cumulative time needed to build the SVM-based classifier and the constrained neural networks until step n. The ratio T_n^{NN}/T_n^{SVM} of time needed to construct these estimator are compared on Figure 5. It is shown in this example that the semi-adaptive SVM-based classifier is less time-consuming than the non-adaptive constrained neural networks for different dimensions. Nonetheless, when the dimension increases the gain of time decrease.

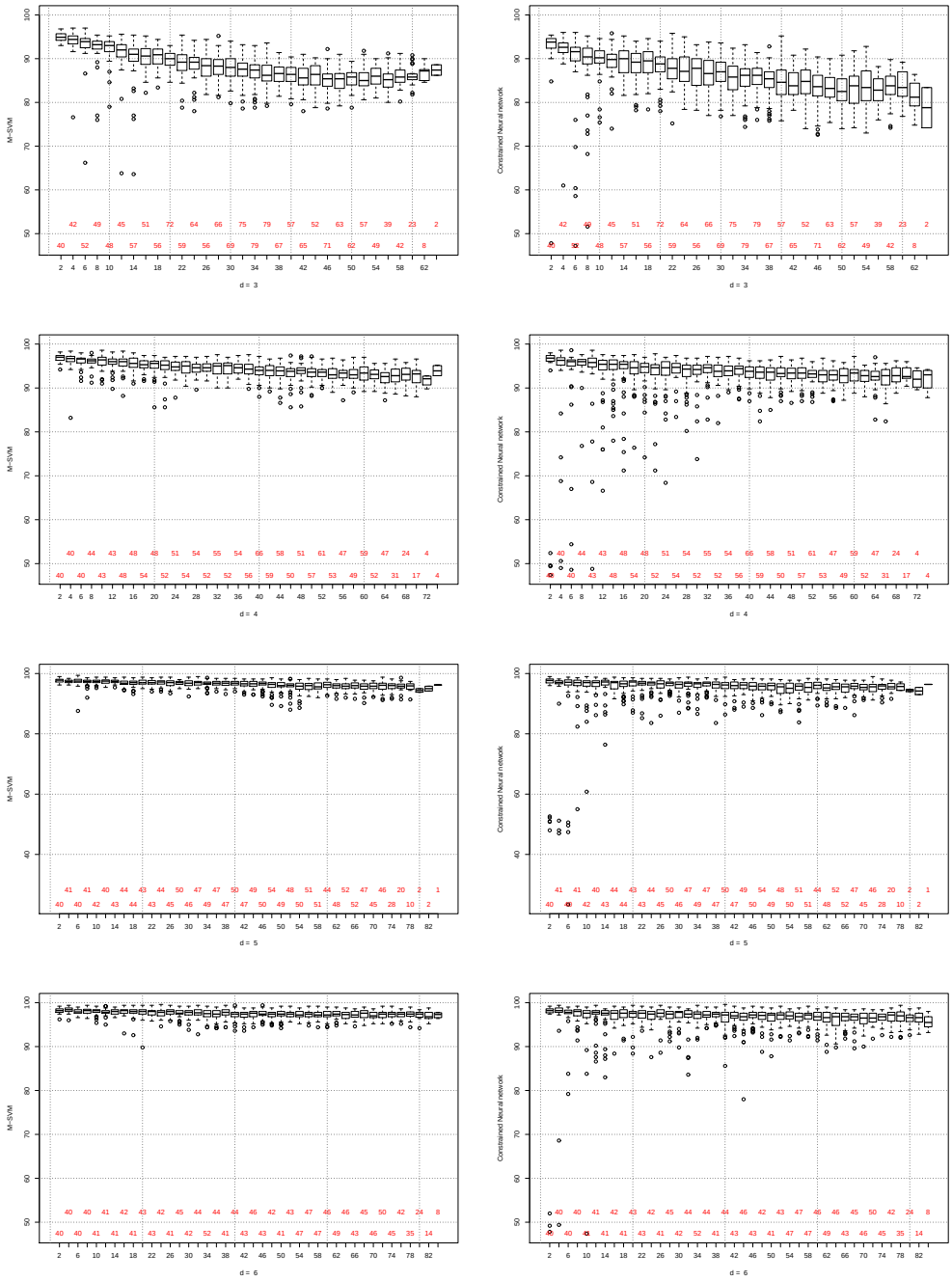

FIGURE 4: Boxplots of the rate of good classification for $d \in \{3, 4, 5, 6\}$, in function of n. In red are the sample size used for each boxplot. Left: Monotonic SVM. Right: Constrained neural networks. The results have been averaged on 40 independent experiments.

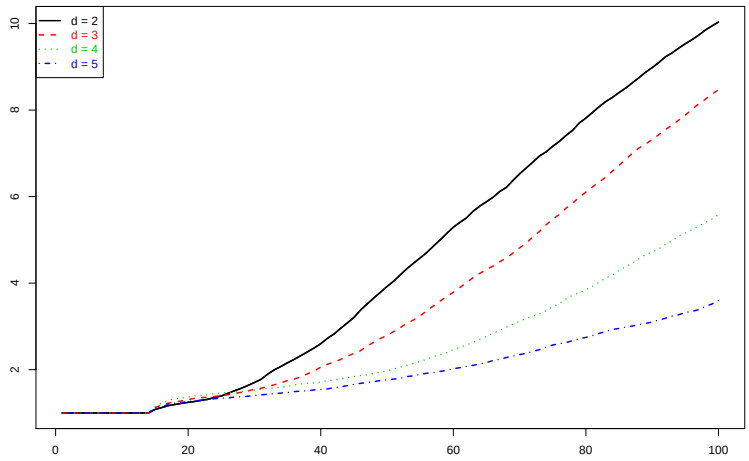

FIGURE 5: Ratio of time construction of the SVM-based classifier and the constrained neural network in function of n. The results have been averaged on 40 independent experiments.

6. Discussion

This articles initiates a theoretical counterpart to the increasing developments linked to the exploration of computer models that make use of their geometrical properties. The original framework is related to structural reliability problems, but the obtained results can be adapted to multi-objective optimisation contexts, characterized by strong constraints of monotonicity. The main results presented here are the consistency and law convergence of the main ingredients (random sets, deterministic bounds, classification tools) used to solve such problems in the common case where the numerical design is generated by a Monte Carlo approach. Therefore such results should be understood as benchmark objectives for more elaborated approaches. For example, use the monotonic estimation of the limit state surface proposed in this article which presents good theoretical and practical properties.

These more elaborated approaches are, obviously, based on sequential sampling within the current non-dominated set. Importance sampling (or more generally variance reduction) techniques could directly benefit from the current estimation of the limit state surface by the SVM-based tool : the importance distribution must be

calibrated to optimize the sampling according to some criterion of volumic reduction. A first step was done in this direction in [22]. Such approaches appear more difficult to study theoretically, especially in multidimensional settings. However, a straightforward argument consists in selecting the sequence of importance samplings as a (finite or infinite) subsequence of an infinite nested uniform sampling, such that the consistency properties be kept. The extension of the results presented here to sequential importance sampling, which was already suggested in [6], is the subject of a future work. Developing specific techniques for fast sampling in a tortured set $\mathbb{U}_n$ should also be part of this work.

Numerous other avenues for future research can be highlighted. Especially, monotonicity of computer models can be exploited to get deterministic bounds of output quantiles, instead of probabilities (associated to undesirable events too). Quantiles provided for a limit tail value are typical tools in risk analysis. A preliminary result is given beneath in dimension 1, in corollary of Proposition 5, to introduce this theme of research, which will be investigated in another future work.

Corollary 3. *Assume that g is continuous, strictly increasing and differentiable at p. Denote $q_p = \inf_{q\in\mathbb{R}} (\mathbb{P}(g(\mathbf{X}) \leq q) = p)$ the p-order quantile of $g(\mathbf{X})$ and g' the derivative of g. Then*

$$n(q_p - g(p_n^-)) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}xp(1/g'(p)),$$
$$n(g(p_n^+) - q_p) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}xp(1/g'(p)).$$

Acknowledgement

The authors express their grateful thanks to Fabrice Gamboa (Institut de Mathématiques de Toulouse) and Bertrand Iooss (EDF R&D and Institut de Mathématique de Toulouse) for their great comments, technical advices and support during the redaction of this article.

7. Proofs

Proof of Proposition 2. Set $\mathbb{U}_\infty^- = \bigcup_{k\geq 1} \mathbb{U}_k^-$, $\mathbb{U}_\infty^+ = \bigcup_{k\geq 1} \mathbb{U}_k^+$ and the non-dominated set $\mathbb{U}_\infty = [0,1]^d \setminus (\mathbb{U}_\infty^- \cup \mathbb{U}_\infty^+)$. We define $p_\infty^- = \mu(\mathbb{U}_\infty^-)$ and $p_\infty^+ = \mu([0,1]^d \setminus \mathbb{U}_\infty^+)$. By inclusion and closure, the bounded sequences $(p_n^-)_{n\geq 1}$ and $(p_n^+)_{n\geq 1}$ are respectively increasing and decreasing, then $p_n^- \overset{a.s.}{\to} p_\infty^-$ and $p_n^+ \overset{a.s.}{\to} p_\infty^+$. Assume $\mathbb{U}_\infty$ is a non-empty open set such that $\mu(\mathbb{U}_\infty) = p_\infty^+ - p_\infty^- > 0$. There exist $\varepsilon_1 > 0$ and $\mathbf{x}_0 \in [0,1]^d$ such that

$$\mathbf{B}(\mathbf{x}_0, \varepsilon_1) \quad \subset \quad \mathbb{U}_\infty. \tag{6}$$

Hence no element of the sequence $(X_k)_{k\geq 1}$ belongs to $\mathbf{B}(\mathbf{x}_0, \varepsilon_1)$. Now, we introduce the events $A_n = \{\mathbf{X}_n \in \mathbf{B}(\mathbf{x}_0, \varepsilon_1)\}$, the are independent by construction and

$$\sum_{n\geq 1} \mathbb{P}(A_N) = \sum_{n\geq 1} \frac{\pi^{d/2}\varepsilon^d}{\Gamma(d/2+1)} = +\infty.$$

We can then apply to Borel-Cantelli's lemma that ensures that $\mathbb{P}(\limsup_n A_n) = 1$. Therefore it exists almost surely at least one $X_k \in \mathbf{B}(\mathbf{x}_0, \varepsilon_1)$, which is contradictory with (6). Hence $\mu(\mathbb{U}_\infty) = 0$ and necessarily $p_\infty^- = p_\infty^+ = p$ almost surely. □

Proof of Corollary 1. From construction, one has $\mathbb{U}_n^- \subset \mathbb{U}^-$ then from Proposition 2

$$d_H(\mathbb{U}_n^-, \mathbb{U}^-) = \sup_{\mathbf{x}\in\mathbb{U}_n^-} \inf_{\mathbf{y}\in\mathbb{U}^-} \|\mathbf{x}-\mathbf{y}\| \leq \mu(\mathbb{U}^-) - \mu(\mathbb{U}_n^-) = p - p_n^-, \underset{n\to+\infty}{\longrightarrow} 0.$$

□

Proof of Proposition 3. It is an alternative proof to the consistency result given in [6]. For any measurable set $A \subset [0,1]^d$. If $(Y_n)_n$ is an i.i.d. sequence of variable uniformly distributed on $[0,1]^d$ and $T = \inf\{n, X_n \in A\}$. Then, it is a well known fact that Y_T is uniformly distributed on A. Hence conditionnaly to Conditionally to

$\mathbf{X}_1, \ldots, \mathbf{X}_{n-1}$ one has

$$\mathbf{X}_n \sim \mathcal{U}(\mathbb{U}_{T_{n-1}}).$$

Denote $(q_n^-)_{n\geq 1}$ and $(q_n^+)_{n\geq 1}$ the sequences of bounds obtained from $(\mathbf{Y}_n)_{n\geq 1}$. By construction, the sequences $(p_n^-)_{n\geq 1}$ and $(p_n^+)_{n\geq 1}$ are subsequences of $(q_n^-)_{n\geq 1}$ and $(q_n^+)_{n\geq 1}$. Then

$$p_n^+ - p_n^- \leq q_n^+ - q_n^- \underset{n\to+\infty}{\overset{a.s.}{\to}} 0.$$

□

Proof of Proposition 4. Proof of **(1)**. By triangle inequality, one has

$$\begin{aligned} d_H(F_n, \mathbb{U}^-) &\leq d_H(F_n, \mathbb{U}_n^-) + d_H(\mathbb{U}_n^-, \mathbb{U}^-), \\ &\leq d_H([0,1]^d \backslash \mathbb{U}_n^+, \mathbb{U}_n^-) + d_H(\mathbb{U}_n^-, \mathbb{U}^-), \text{ since } \mathbb{U}_n^- \subset F_n \subset [0,1]^d \backslash \mathbb{U}_n^+ \\ &\leq d_H([0,1]^d \backslash \mathbb{U}_n^+, \mathbb{U}^-) + 2d_H(\mathbb{U}_n^-, \mathbb{U}^-), \text{ by a second triangle inequality.} \end{aligned}$$

We know from Corollary 1, that

$$\begin{aligned} d_H([0,1]^d \backslash \mathbb{U}_n^+, \mathbb{U}_n^-) &\underset{n\to+\infty}{\overset{a.s.}{\to}} 0, \\ d_H(\mathbb{U}_n^-, \mathbb{U}^-) &\underset{n\to+\infty}{\overset{a.s.}{\to}} 0, \end{aligned}$$

hence $d_H(F_n, \mathbb{U}^-) \underset{n\to+\infty}{\overset{a.s.}{\to}} 0$. Besides $\mu(\mathbb{U}_n^-) \leq \mu(F_n) \leq \mu([0,1]^d \backslash \mathbb{U}_n^+)$ by construction, then $p_n^- \leq \mu(F_n) \leq p_n^+$, and from Corollary 1, it can be deduced that

$$\mu(F_n) \underset{n\to+\infty}{\overset{a.s.}{\to}} p,$$

that concludes the proof of **(1)**. The proof of **(2)** and **(3)** are based on the following Theorem due to [26].

Theorem 2. *([26], Theorem 2). Let K be a compact set on $\mathbb{R}^d$ and standard with respect to a measure ν. Let $(\mathbf{X}_k)_{k\geq 1}$ be a sequence of i.i.d. random variables uniformly distributed on K, and K_n be a set such that for all n large enough one has almost*

surely $(\mathbf{X}_1, \ldots, \mathbf{X}_n) \subset K_n \subset K$. *Then almost surely*

$$d_H(K_n, K) = O\left(\left(\frac{\log n}{n}\right)^{1/d}\right).$$

The notion of standard set generalizes the notion of (α, γ)-regularity that is used here. Hence any (α, γ)- regular set is also standard in the sense of Theorem 2.

Proof of **(2)**. It is a direct consequence of Theorem 2.

Proof of **(3)**. Application of Theorem 2. In order to do so, we will cut $[0,1]^d$ following $g^{-1}(\{0\})$ and glue it on $[0,1]^{d-1}$ in the following way. Set

$$G : [0,1]^{d-1} \longrightarrow [0,1]$$
$$x \mapsto \begin{cases} \inf\{t \in [0,1], g(x,t) = 0\} & \text{if} \quad \{t \in [0,1], g(x,t) = 0\} \neq \emptyset, \\ 0 & \text{if} \quad \{t \in [0,1], g(x,t) = 0\} = \emptyset, \end{cases}$$

and define $\tilde{\Gamma}$ as the graph of the application $x \mapsto G(x) - 1$. Now let

$$\Psi : [0,1]^d \longrightarrow I\!R^d$$
$$(x,y) \mapsto \begin{cases} (x,y) & \text{if} \quad g(x,y) \leq 0, \\ (x, y-1) & \text{if} \quad g(x,y) > 0. \end{cases}.$$

We define now $\tilde{\mathbb{U}} = \Psi([0,1]^d) \bigcup \tilde{\Gamma}$ (we have cut the space $[0,1]^d$ following $g^{-1}(\{0\})$ and glue together $[0,1]^{d-1} \times \{0\}$ and $[0,1]^{d-1} \times \{1\}$, see Figure 6). Now by assumption $\tilde{\mathbb{U}}$ is a compact set and the random variables $(\Psi(X_i))_i$ are i.i.d uniformly distributed in $\tilde{\mathbb{U}}$. The result follows applying Theorem 2. □

Proof of Proposition 5. Let $x \in [0,p]$. Then

$$\mathbb{P}(p_n^- \leq x) = \mathbb{P}(\max_{i=1,\ldots,n}(X_i \mathbb{1}_{\{X_i \leq p\}}) \leq x) = \mathbb{P}(X_1 \mathbb{1}_{\{X_1 \leq p\}} \leq x)^n$$
$$= \left(1 - \mathbb{P}(X_1 \mathbb{1}_{\{X_1 \leq p\}} > x)\right)^n = (1 - \mathbb{P}(p \geq X_1 > x))^n = (1 - p + x)^n. \quad (7)$$

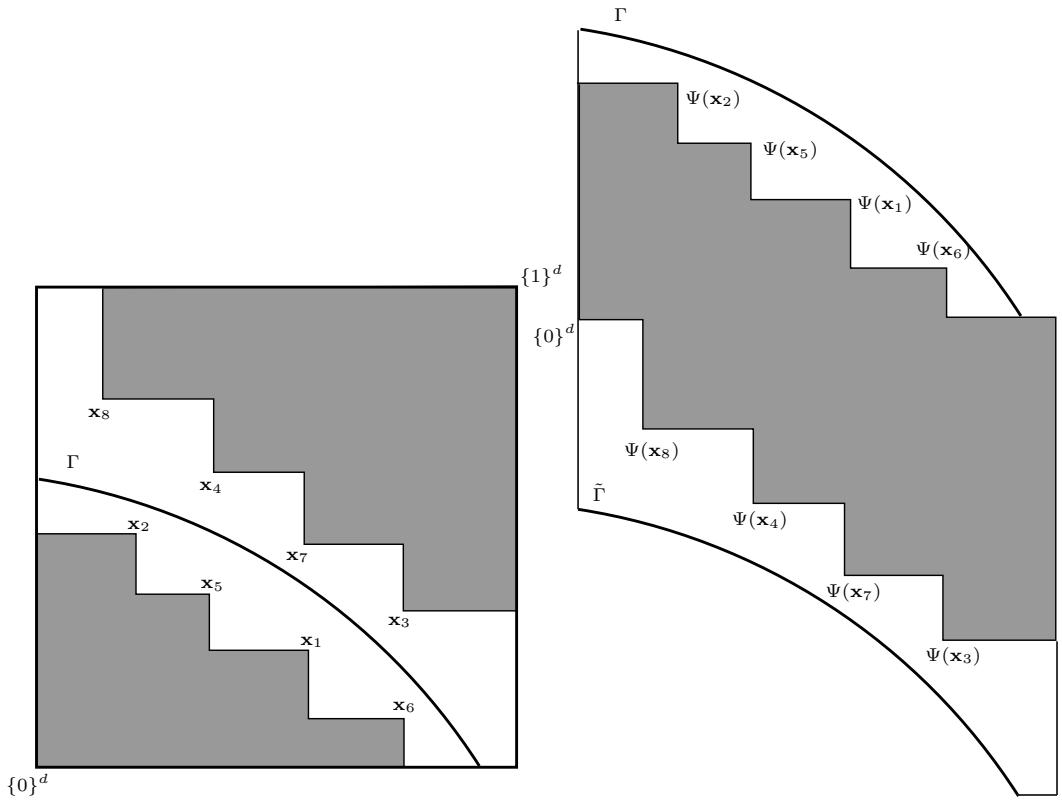

FIGURE 6: Illustration of the cut given in proof of Proposition 4. **Left**: Illustration of $[0,1]^d$ after some simulations. **Right**: Representation of $\tilde{\mathbb{U}}$ delimited by the set in gray, Γ and $\tilde{\Gamma}$. $\Psi(\mathbf{x}_i) = \mathbf{x}_i$ for $i \in \{3,4,7,8\}$.

Then for all $x \in [0,1]$,

$$\mathbb{P}(p_n^- \leq x) = \begin{cases} 0 & \text{if } x < 0, \\ (1-p+x)^n & \text{if } 0 \leq x \leq p, \\ 1 & \text{if } x > p. \end{cases} \tag{8}$$

Let $x > 0$ and n be large enough such that $x \in [0, np]$, then

$$\mathbb{P}(n(p - p_n^-) < x) = 1 - (1 - x/n)^n \xrightarrow[n \to +\infty]{} (1 - e^{-x}),$$

hence $n(p - p_n^-) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}xp(1)$. From (8), it comes:

$$\mathbb{E}[p_n^-] = \int_0^1 \mathbb{P}(p_n^- \geq x)dx = p - \frac{1}{n+1}[1-(1-p)^{n+1}].$$

Denote $p_n^-(p)$ a random variable such that for all $x \in [0,1]$,

$$\mathbb{P}(p_n^-(p) \leq x) = \begin{cases} 0 & \text{if } x < 0, \\ (1-p+x)^n & \text{if } 0 \leq x \leq p, \\ 1 & \text{if } x > p. \end{cases}.$$

Then, p_n^+ have the same distribution as $1 - p_n^-(1-p)$. For all $x \in [0,1]$, one has

$$\mathbb{P}(p_n^+ > x) = \mathbb{P}(p_n^-(1-p) \leq 1-x) = \begin{cases} 1 & \text{if } x < p \\ (1+p-x)^n & \text{if } p < x \leq 1 \\ 0 & \text{if } x > 1. \end{cases},$$

and

$$\mathbb{E}[p_n^+] = \mathbb{E}[1 - p_n^-(1-p)] = p + \frac{1}{n+1}[1-p^{n+1}].$$

It can be deduced that for all n,

$$\mathbb{E}[p_n^+ - p_n^-] = \frac{2}{n+1} - \frac{1}{n+1}\left(p^{n+1} + (1-p)^{n+1}\right).$$

□

Proof of Proposition 6. $\mathcal{F}_n$ will denote the natural filtration o One has $p_{n+1}^- + p_{n-1}^+ = (p_n^+ - p_n^-)\xi_{n+1} + X_{n+1} + p_n^-$. Since for all $k \geq 1, X_k$ is uniformly distributed on $[p_{k-1}^-, p_{k-1}^-]$ if $\mathcal{F}_n$ denotes the natural filtration associated to this sequence, it comes

that

$$\begin{aligned}\mathbb{E}[p_{n+1}^- + p_{n+1}^+|\mathcal{F}_n] &= (p_n^+ - p_n^-)\mathbb{E}[\xi_{n+1}|\mathcal{F}_n] + \mathbb{E}[X_{n+1}|\mathcal{F}_n] + p_n^-,\\ &= p + \mathbb{E}[X_{n+1}|\mathcal{F}_n] = p + (p_n^- + p_n^+)/2,\end{aligned}$$

the last equality holds since $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = (p_n^- + p_n^+)/2$. By recursion, it can be deduced that

$$\begin{aligned}\mathbb{E}[X_{n+1}] &= \frac{1}{2}\mathbb{E}[p_n^- + p_n^+] = \frac{1}{2}(p + \mathbb{E}[X_n]) = \frac{p}{2} + \frac{p}{2^2} + \frac{\mathbb{E}[X_{n-1}]}{2^2},\\ &= p\left(1 - \frac{1}{2^n}\right) + \frac{1}{2^{n+1}}.\end{aligned}$$

Since $X_{n+1} \in [0,1]$, it comes that $\mathrm{Var}[X_{n+1}] \leq \mathbb{E}[X_{n+1}^2] \leq \mathbb{E}[X_{n+1}] = p(1-\frac{1}{2^n}) + \frac{1}{2^{n+1}}$. Besides

$$\begin{aligned}\mathbb{E}[p_{n+1}^+ - p_{n+1}^-|\mathcal{F}_n] &= (p_n^+ + p_n^-)\mathbb{E}[\xi_{n+1}|\mathcal{F}_n] - 2\mathbb{E}[X_{n+1}\xi_{n+1}|\mathcal{F}_n] + \mathbb{E}[X_{n+1}|\mathcal{F}_n] - p_n^-,\\ &= \frac{(p_n^+ + p_n^-)(p - p_n^-)}{p_n^+ - p_n^-} - \frac{p^2 - (p_n^-)^2}{p_n^+ - p_n^-} + \frac{p_n^+ + p_n^-}{2} - p_n^-,\\ &= \frac{p_n^+ - p_n^-}{2} + \frac{(p_n^+ - p)(p - p_n^-)}{p_n^+ - p_n^-}. \qquad (9)\end{aligned}$$

Since $\mathbb{E}[p_{n+1}^+ - p_{n+1}^-] \geq \frac{1}{2}\mathbb{E}[p_n^+ - p_n^-]$, it comes by recursion that

$$\mathbb{E}[p_{n+1}^+ - p_{n+1}^-] \geq \frac{1}{2^{n+1}}.$$

Since $(p_n^+ - p)(p - p_n^-)$ is lower than $(p_n^+ - p_n^-)^2/4$, it can be deduced from (9) that

$$\mathbb{E}[p_{n+1}^+ - p_{n+1}^-|\mathcal{F}_n] \leq \frac{3}{4}(p_n^+ - p_n^-).$$

By recursion, it comes that $\mathbb{E}[p_n^+ - p_n^-] \leq \left(\frac{3}{4}\right)^n$. □

Proof of Corollary 2. Let $p = 0$, then $\mathbb{E}[p_{n+1}^+|\mathcal{F}_n] = \mathbb{E}[X_{n+1}|\mathcal{F}_n]$. Since $X_{n+1} \sim \mathcal{U}([0, p_n^+])$, it comes that

$$\mathbb{E}[p_{n+1}^+|\mathcal{F}_n] = p_n^+/2,$$

by recursion $\mathbb{E}[p_n^+] = 2^{-n}$. For $p = 1$, using $X_{n+1} \sim \mathcal{U}([p_n^-, 1])$ proves the result. □

The proof of Proposition 7 relies on the following Lemma that gives the value of the probability density function of a sum of independent beta distributed random variables.

Lemma 2. *Denote f_d the probability density function (pdf) of $\sum_{i=1}^d B_i$, supported over $[0, d]$, where the B_i are iid random variables following the $\mathcal{B}eta(1/q, 1)$ distribution. Then for all $x \in [0, 1]$*

$$f_d(x) = C_{d,q} x^{d/q-1} \tag{10}$$

with $C_{1,q} = 1$ and for $d \geq 2$, $C_{d,q} = \frac{1}{q^d} \prod_{i=1}^{d-1} B(i/q, 1/q)$.

Proof of Lemma 2. We proceed by induction. For $d = 1$, the density given by (10) is the density of $\mathcal{B}eta(1/q, 1)$ distribution, that is

$$f_1(x) = \frac{1}{q} x^{1/q-1} \mathbb{1}_{\{0 \leq x \leq 1\}}.$$

Assume there exists $k \in \mathbb{N}^*$ such that for all $1 \leq j \leq k$, f_j is the density of $\sum_{i=1}^j B_i$ (only for $x \in [0, 1]$). Denote $f \star g$ the convolution of functions f and g. Then, for $x \in [0, 1]$,

$$\begin{aligned}
f_{k+1}(x) = (f_k \star f_1)(x) &= \int_0^x f_k(t) f_1(x-t) dt, \\
&= \int_0^x \frac{1}{q^k} \left[\prod_{i=1}^{k-1} B(i/q, 1/q) \right] t^{k/q-1} \frac{1}{q} (x-t)^{1/q-1} dt, \\
&= \frac{1}{q^{k+1}} \prod_{i=1}^{k-1} B(i/q, 1/q) \int_0^x x^{1/q-1} t^{k/q-1} (1 - t/x)^{1/q-1} dt, \\
&= C_{k+1,q} x^{(k+1)/q-1}
\end{aligned}$$

which proves the validity of (10). □

Proof of Proposition 7. Set $X_i = (X_i^1, \ldots, X_i^j)$. Since $\mathbb{U}_n^- \subset \mathbb{U}^-$, one has

$$\begin{aligned} d_{H,q}(\mathbb{U}_n^-, \mathbb{U}^-) &= \max(\sup_{\mathbf{y}\in\mathbb{U}_n^-} \inf_{\mathbf{x}\in\mathbb{U}^-} \|\mathbf{x}-\mathbf{y}\|_q; \sup_{\mathbf{x}\in\mathbb{U}^-} \inf_{\mathbf{y}\in\mathbb{U}_n^-} \|\mathbf{x}-\mathbf{y}\|_q), \\ &= \sup_{\mathbf{x}\in\mathbb{U}^-} \inf_{\mathbf{y}\in\mathbb{U}_n^-} \|\mathbf{x}-\mathbf{y}\|_q = \inf_{\mathbf{y}\in\mathbb{U}_n^-} \|\{1\}^d - \mathbf{y}\|_q, \\ &= \inf_{\mathbf{y}\in\{\mathbf{X}_1,\ldots,\mathbf{X}_n\}} \|\{1\}^d - \mathbf{y}\|_q. \end{aligned}$$

(1). Assume $0 < q < \infty$. For $t \in [0, d^{1/q}]$, using the fact that if X is uniformly distributed on $[0,1]$ so is $1-X$ we have

$$\mathbb{P}(d_{H,q}(\mathbb{U}_n^-, \mathbb{U}^-) \leq t) = 1 - \left(1 - \mathbb{P}\left(\sum_{j=1}^{d} (X_1^j)^q \leq t^q\right)\right)^n.$$

Let $(\alpha_n)_{n\geq 1}$ be a sequence of real numbers such that $\alpha_n \to +\infty$ as $n \to +\infty$. Hence,

$$\mathbb{P}(\alpha_n d_{H,q}(\mathbb{U}_n^-, \mathbb{U}^-) \leq t) = 1 - \left(1 - \mathbb{P}\left(\sum_{i=1}^{d} B_i \leq \frac{t^q}{\alpha_n^q}\right)\right)^n \tag{11}$$

where each $B_i = X_i^q \overset{\mathcal{L}}{\sim} \mathcal{B}eta(1/q, 1)$. From Lemma 2, for $u \in [0,1]$,

$$\mathbb{P}\left(\sum_{i=1}^{d} B_i \leq u\right) = \int_0^u C_{d,q} x^{d/q-1} dx = \frac{q}{d} C_{d,q} u^{d/q},$$

where $C_{d,q} = \frac{1}{q^d} \prod_{i=1}^{d-1} B(i/q, 1/q)$. Therefore, for n large enough

$$\mathbb{P}(\alpha_n d_{H,q}(\mathbb{U}_n^-, \mathbb{U}^-) \leq t) = 1 - \left(1 - \frac{q}{d} C_{d,q} \frac{t^d}{\alpha_n^d}\right)^n.$$

Denoting $A_{1,q} = 1$ and for $d \geq 2$, $A_{d,q} = \frac{1}{dq^{d-1}} \prod_{i=1}^{d-1} B(i/q, 1/q)$, and choosing $\alpha_n = (nA_{d,q})^{1/d}$ it comes

$$\mathbb{P}\left((A_{d,q} n)^{1/d} d_{H,q}(\mathbb{U}_n^-, \mathbb{U}^-) \leq t\right) \underset{n\to+\infty}{\longrightarrow} 1 - e^{-t^d}.$$

(2). Assume now $q = \infty$. Let $(\beta_n)_{n\geq 1}$ be a sequence of real numbers such that $\beta_n \to +\infty$ as $n \to +\infty$. Then for all $t \in]0,1[$

$$\begin{aligned}\mathbb{P}(\beta_n d_H^{\infty}(\mathbb{U}_n^-, \mathbb{U}^-) > t) &= (1 - \mathbb{P}(\|\mathbf{X}_1\|_\infty \leq t/\beta_n))^n = \left(1 - \mathbb{P}(\max_{j=1,\dots,d} X_1^j \leq t/\beta_n)\right)^n, \\ &= \left(1 - t^d/\beta_n^d\right)^n,\end{aligned}$$

choosing $\beta_n = n^{1/d}$, it comes $\mathbb{P}(n^{1/d} d_H^{\infty}(\mathbb{U}_n^-, \mathbb{U}^-) > t) \underset{n\to+\infty}{\longrightarrow} e^{-t^d}$, and

$$n^{1/d} d_H^{\infty}(\mathbb{U}_n^-, \mathbb{U}^-) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{W}(1,d).$$

□

Proof of Proposition 8. Let $(\mathbf{X}_n)_{n\geq 1}$ be a sequence of iid random variables uniformly distributed on $[0,1]^d$. Let $U = (U^1, \dots, U^d)$ be uniformly distributed on $[0,1]^d$ and independent of the sample $(\mathbf{X}_n)_{n\geq 1}$, then

$$\begin{aligned}\mathbb{E}[\mu([0,1]^d \backslash \mathbb{U}_n^-)] &= \mathbb{E}[\mathbb{E}[\mathbb{1}_{\mathbf{U}\in\mathbb{U}_n}|\mathbf{U}]], \\ &= \mathbb{E}[\mathbb{E}[\mathbb{1}_{\mathbf{U}\not\leq\mathbf{X}_1,\dots,\mathbf{U}\not\leq\mathbf{X}_n}|\mathbf{U}]], = \mathbb{E}[\mathbb{E}[\mathbb{1}_{\mathbf{U}\not\leq\mathbf{X}_1}|\mathbf{U}]\cdots\mathbb{E}[\mathbb{1}_{\mathbf{U}\not\leq\mathbf{X}_n}|\mathbf{U}]], \\ &= \mathbb{E}[\mathbb{E}[\mathbb{1}_{\mathbf{U}\not\leq\mathbf{X}_1}|\mathbf{U}]^n] = \mathbb{E}[\left(1 - (1-U^1)\cdots(1-U^d)\right)^n], \\ &= \mathbb{E}[\left(1 - U^1\cdots U^d\right)^n]. \qquad (12)\end{aligned}$$

Using the fact that $-\log(U^i) \overset{\mathcal{L}}{\sim} \mathcal{E}xp(1)$, it is easy to see that $\prod_{i=1}^d U^i \overset{\mathcal{L}}{=} \exp(-G_d)$, where $G_d \sim \mathcal{G}amma(d,1)$ with density function

$$f(x) = \frac{x^{d-1}e^{-x}}{(d-1)!}.$$

We then easily get that the density function of $\prod_{i=1}^d U^i$ is given by

$$f_d(x) = \frac{(-\log(x))^{d-1}}{(d-1)!}\mathbb{1}_{\{x\in[0,1]\}}. \qquad (13)$$

From (12) and (13), it comes

$$\mathbb{E}[\mu([0,1]^d \backslash \mathbb{U}_n)] = \mathbb{E}[\big(1 - U^1 \cdots U^d\big)^n] = \int_0^1 (1-u)^n \frac{(-\log(u))^{d-1}}{(d-1)!} du.$$

One has

$$\frac{n(d-1)!}{\log(n)^{d-1}} \mathbb{E}[\mu([0,1]^d \backslash \mathbb{U}_n)] = \frac{n}{\log(n)^{d-1}} \int_0^1 (1-u)^n (-\log(u))^{d-1} du,$$

and the substitution $u = x/n$ gives

$$\begin{aligned} \frac{n}{\log(n)^{d-1}} \int_0^1 (1-u)^n (-\log(u))^{d-1} du &= \frac{1}{\log(n)^{d-1}} \int_0^{+\infty} (1-x/n)^n (-\log(x/n))^{d-1} \mathbb{1}_{\{x \leq n\}} dx, \\ &= \int_0^{+\infty} (1-x/n)^n (1 - \log(x)/\log(n))^{d-1} \mathbb{1}_{\{x \leq n\}} dx, \\ &\underset{n \to +\infty}{\longrightarrow} 1. \end{aligned}$$

The last equation holds by the dominated convergence theorem apply to $f_n(x) = (1-x/n)^n (1 - \log(x)/\log(n))^{d-1} \mathbb{1}_{\{0 \leq x \leq n\}} \leq \exp(-x) \mathbb{1}_{\{x \geq 0\}}$.

Proof of Proposition 9. Let $\mathbf{u} \in \Gamma$ and $\mathbf{v} \in [0,1]^d$ such that $\mathbf{u} \prec \mathbf{v}$. Assume $\mathbf{v} \in \Gamma$. There exist $\varepsilon > 0$ such that $\mathbf{B}((\mathbf{u}+\mathbf{v})/2, \varepsilon) \subset \Gamma$. That implies $\mu(\Gamma) > 0$, which is impossible by Assumption 2. Then $\mathbf{v} \notin \Gamma$. □

Proof of Theorem 1. **(1)**. Obvious since $h_{\mathbf{X}}(\mathbf{u}) \leq h_{\mathbf{X}}(\mathbf{v})$ for all $\mathbf{X} \in \Xi_n^+$ and for all $\mathbf{u}, \mathbf{v} \in [0,1]^d$ such that $\mathbf{u} \preceq \mathbf{v}$.

(2), (3). By construction $\mathbb{U}_n^- \subset F_n \subset [0,1]^d \backslash \mathbb{U}_n^+$.

(4). Since $\mathbb{U}^-$ is regular (by convexity) and $\mathbb{U}^+$ is regular (by Lemma 3 below), the result is a consequence of Equation (2).

Lemma 3. *Let K be a compact, convex subset of $[0,1]^d$ such that for all $\mathbf{y} \in K$ there is no $\mathbf{x} \in K^c = [0,1]^d \backslash K$ such that $\mathbf{x} \preceq \mathbf{y}$. Then K^c is (α, γ)-regular.*

Proof of Lemma 3. Since K is a compact convex set, there exists $\alpha, \gamma > 0$ such that K is (α, γ)-regular [1]. Given $\mathbf{x} \in K^c$, since K is closed convex the projection of

$\mathbf{x}$ on K, denoted $P_K(\mathbf{x})$, is unique. Let $0 \leq \varepsilon \leq \gamma$ and $\mathbf{x} \in K^c$, then

$$\begin{aligned}
\mu(\mathbf{B}(\mathbf{x},\varepsilon) \cap K^c) &\geq \mu(\mathbf{B}(P_K(\mathbf{x}),\varepsilon) \cap K^c), \\
&\geq \mu(\mathbf{B}(P_K(\mathbf{x}),\varepsilon) \cap K), \text{ since } K \text{ is convex} \\
&\geq \alpha\mu(\mathbf{B}(P_K(\mathbf{x}),\varepsilon)), \text{ since } K \text{ is } (\alpha,\gamma)\text{-regular} \\
&\geq \alpha\mu(\mathbf{B}(\mathbf{x},\varepsilon)),
\end{aligned}$$

then K^c is (α,γ)-regular. □

Proof of Corollary 3. Since g is differentiable at p, the Delta method and Proposition 5 imply that

$$\begin{aligned}
n(g(p) - g(p_n^-)) &\xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}xp(1/g'(p)), \\
n(g(p_n^+) - g(q)) &\xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}xp(1/g'(p)).
\end{aligned}$$

Since q_p is the p-order quantile of $g(\mathbf{X})$, it comes that $\mathbb{P}(g(X) \leq q_p) = p$. Since g is continuous and strictly increasing, then

$$\mathbb{P}(g(X) \leq q_p) = \mathbb{P}(X \leq g^{-1}(q_p)).$$

Moreover $X \sim \mathcal{U}([0,1])$, then $\mathbb{P}(X \leq g^{-1}(q_p)) = g^{-1}(q_p) = p$. It can be deduced that $q_p = g(p)$. Then, the two last equations becomes

$$\begin{aligned}
n(q_p - g(p_n^-)) &\xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}xp(1/g'(p)), \\
n(g(p_n^+) - q_p) &\xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}xp(1/g'(p)).
\end{aligned}$$

□

References

[1] Audibert, J-Y. and Tsybakov, A. (2007). *Fast learning rates for plug-in classifiers*, The Annals of Statistics, Institute of Mathematical Statistics, **35**, 608-633.

[2] Bárány, I. and Larman, D. G. (1988). *Convex bodies, economic cap coverings, random polytopes*, Cambridge Univ. Press.

[3] Bárány, I. (2007). *Random polytopes, convex bodies, and approximation*, The Annals of Statistics, Institute of Mathematical Statistics, **35**, 77-118.

[4] Blaszczynski, J. and Greco, S. and Slowinski, R. (2012). *Inductive discovery of laws using monotonic rules*, Engineering Applications of Artificial Intelligence, **25**, 284-294.

[5] Bourinet, J.-M. and Deheeger, F. and Lemaire, M. (2011). *Assessing small failure probabilities by combined subset simulation and Support Vector Machines*, Structural Safety, **33**, 343-353.

[6] Bousquet, N. (2012). *Accelerated Monte Carlo estimation of exceedance probabilities under monotonicity constraints*, Annales de la Faculté des Sciences de Toulouse, **21**, 557-591.

[7] Casti, J.L. (1997). *Would-be worlds: How simulation is changing the frontiers of science*, Wiley.

[8] Cuevas, A. and Fraiman, R. (1997). *A plug-in approach to support estimation*, The Annals of Statistics, JSTOR,**21**, 2300-2312.

[9] Da Veiga, S. and Marrel, A. (2012). *Gaussian process modelling with inequality constraints*, Annales de la Faculté des Sciences de Toulouse, **21**, 529-555.

[10] de Marchi, S. (2005). *Computational and Mathematical Modeling in the Social Sciences*, Cambridge Press.

[11] Doumpos, M. and Zopounidis, C. (2009). *Monotonic Support Vector Machines For Credit Risk Rating*, New Mathematics and Natural Computation,World Scientific, **5**, 557-570.

[12] Dümbgen, L. and Walther, G. (1996). *Rates of convergence for random approximations of convex sets*, *Adv. Appl. Prob.*, JSTOR, **5**, 384-393.

[13] Durot, C. (2008). *Monotone nonparametric regression with random design*, Mathematical Methods in Statistics, **17**, 327-341.

[14] Edelsbrunner, H. and Mücke, E.P. (1994). *Three dimensional alpha shapes*, ACM Trans. Graph., **13**, 43-72.

[15] Figueira, J. and Greco, S. and Erhgott, M. (2005). *Multiple criteria decision analysis - State of the art - Survey*, Springer.

[16] Gadat, S. and Klein, T. and Marteau, C. (2014). *Classification with the nearest neighbor rule in general finite dimensional spaces* (submitted), arXiv:1411.0894.

[17] Golchi, S. and Bingham, D.R. and Chipman, H. and Campbell, D.A. (2013). *Monotone Function Estimation for Computer Experiments* (submitted), arXiv:1309.3802.

[18] Goldfarb, D. and Idnani, A. (1983). *A numerically stable dual method for solving strictly convex quadratic programs*, Mathematical programming, Springer, **27**, 1-33.

[19] Hastie, T. and Tibshirani, R. and Friedman, J. (2009). *The elements of statistical learning*, Springer, **2**.

[20] Hurtado, J. (2004). *An examination of methods for approximating implicit limit state functions from the viewpoint of statistical learning theory*, Structural Safety, Elsevier, **26**, 271-293.

[21] Lemaire, M. and Chateauneuf, A. and Mitteau, J.-C. (2010). *Structural reliability*, Wiley.

[22] Moutoussamy, V. and Bousquet, N. and Gamboa, B. and Iooss, B. and Klein, T. and Rochet, P. (2013). *Comparing conservative estimations of failure probabilities using sequential designs of experiments in monotone frameworks*, Proceedings of the 1th International Conference on Structural Safety & Reliability (ICOSSAR).

[23] Pelckmans, K. and Espinoza, M. and De Brabanter, J. and Suykens, J. and De Moor, B. (2005). *Primal-dual monotone kernel regression*, Neural Processing Letters, Springer, **22**, 171-182.

[24] Rockafellar, R. (1997). *Convex analysis*, Princeton university press, **28**.

[25] de Rocquigny, E. (2009). *Structural reliability under monotony: A review of properties of FORM and associated simulation methods and a new class of monotonous reliability methods (MRM)*, Structural Safety, **31**: 363-374

[26] Rodríguez Casal, A. (2007). *Set estimation under convexity type assumptions*, Annales de l'Institut Henri Poincare (B) Probability and Statistics, Elsevier, **43**, 763-774.

[27] Rubinstein, R.Y. and Kroese, D.P. (2008). *Simulation and the Monte Carlo Method (2nd edition)*, Wiley.

[28] Sacks, J. and Welch, W.J. and Mitchell, T.J. and Wynn, H.P. (1989). *Design and Analysis of Computer Experiments*, Statistical Science, **4**, 409-435.

[29] Shively, T.S. and Sager, T.W. and Walker, S.G. (2009). *A Bayesian approach to non-parametric monotone function estimation*, Journal of the Royal Statistical Society Series B, **71**, 159-175.

[30] Smith, R.C. (2013). *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM.

[31] Taboga, M. (2012). *Lectures on probability theory and mathematical statistics (2nd edition)*, Amazon CreateSpace, **VI**-65.1, 527-529.

[32] Velikova, M. and Daniels, H. and Feelders, A. (2007). *Solving Partially Monotone Problems with Neural Network*, World Academy of Science, Engineering and Technology, **12**.

[33] Walther, G. and Lüts D. (1996). *Rates of convergence for random approximations of convex sets.*

[34] Walther, G. (1999). *On a generalization of Blaschke's Rolling Theorem and the smoothing of surfaces*, Mathematical methods in the applied sciences, **22**, 301-316.

Appendix A. Computing hypervolumes (deterministic bounds)

The computation of bounds (p_n^-, p_n^+) defining in (1) can be conducted exactly or using simulation, in function of the dimension.

A.1. An exact method in dimension $d = 2$

Consider a design $\mathbf{x}_1, \ldots, \mathbf{x}_n \in [0,1]^2$. The first stage is to order the element according to their first (or second) coordinate, such that $x_1^1 < \cdots < x_n^1$. Since no design element is dominated by another one, then $x_i^2 > x_{i+1}^2$ for $i \in \{1, \ldots, n-1\}$. The point $\mathbf{x}_1$ delimit a first rectangle P_1 with the following vertices:

$$\begin{pmatrix} 0 & 0 \\ x_1^1 & 0 \\ 0 & x_1^2 \\ x_1^1 & x_1^2 \end{pmatrix},$$

such that $\mu(P_1) = x_1^1 x_1^2$. For all $i \in \{2, \ldots, n\}$ a new rectangle P_i can be defined with the following vertices:

$$\begin{pmatrix} x_{i-1}^1 & 0 \\ x_{i-1}^1 & x_i^2 \\ x_i^1 & 0 \\ x_i^1 & x_i^2 \end{pmatrix},$$

such that $\mu(P_i) = (x_i^1 - x_{i-1}^1)x_i^2$. The second stage consist to compute the volume of each rectangles:

$$p_n^- = x_1^1 x_1^2 + \sum_{i=2}^{n} (x_i^1 - x_{i-1}^1) x_i^2.$$

To compute p_n^+ it is enough to transform each $\mathbf{x}_i$ into $1 - \mathbf{x}_i$, to compute $\tilde{p}_n^+$ with the previous approach. Then $p_n^+ = 1 - \tilde{p}_n^+$.

A.2. An accelerated Monte Carlo method in upper dimensions

A sweepline algorithm described in [6] allows to compute the bounds in any dimension, but at an exponential cost. An alternative approach is using a Monte Carlo method. Considering an iid uniform sample $\bar{\mathbf{U}}_N = \mathbf{U}_1, \ldots, \mathbf{U}_N$ over $[0,1]^d$, (p_n^-, p_n^+) can be estimated by

$$\hat{p}_n^- = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{\mathbf{U}_i \in \mathbb{U}_n^-\}},$$

$$\hat{p}_n^+ = 1 - \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{\mathbf{U}_i \in \mathbb{U}_n^+\}}.$$

The computation can be strongly accelerated by adapting the order of evaluation to the monotonic context, using the following algorithm (easily adaptable to estimating p_n^+).

Algorithm 1 Estimation of p_n^- by accelerated Monte Carlo

$\tilde{p}_n^- \leftarrow 0$, $\mathbf{V} = \bar{\mathbf{U}}_n$, $\bar{\mathbf{U}}_N^- = \bar{\mathbf{U}}_n \cap \mathbb{U}_n^-$
for $i : 1$ to $Card(\bar{\mathbf{U}}_N^-)$ **do**
  $\mathbf{u} \in \arg\max_{\mathbf{x} \in \bar{\mathbf{U}}_N^-} \prod_{j=1}^{d} x_j$
  $\Xi_n^- \leftarrow \Xi_n^- \backslash \mathbf{u}$
  $\tilde{p}_n^- \leftarrow \tilde{p}_n^- + Card(\{\mathbf{U} \in \mathbf{V} : \mathbf{U} \preceq \mathbf{u}\})$
  $\mathbf{V} \leftarrow \mathbf{V} \backslash \{\mathbf{U} \in \mathbf{V} : \mathbf{U} \preceq \mathbf{u}\}$
end for
return $\tilde{p}_n^- / N$

Appendix B. Adapted Support Vector Machines

B.1. Usual situations: a reminder

B.1.1. *Linear situation.* Consider a situation where some available data $D_n = (\mathbf{X}_i, y_i)_{1 \le i \le n}$ where $\mathbf{X}_i \in [0,1]^d$ and $y_i = sign(h(\mathbf{X}_i))$, with $sign(x) = 1$ if $x > 0$ and -1 otherwise,

can be perfectly separated according to y by an hyperplane h defined by

$$h(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}, \tag{14}$$

with $(\beta_0, \boldsymbol{\beta}) \in I\!R \times I\!R^d$. In this linear framework, an infinity of hyperplanes can separate perfectly the data. Therefore it is needed to add some constraints to make a unique (optimal) choice of hyperplane. Define the distance of $\mathbf{x}$ from Γ :

$$\Delta(\mathbf{x}, \Gamma) = \inf_{\mathbf{y} \in \Gamma} \|\mathbf{x} - \mathbf{y}\| = |\beta_0 + \boldsymbol{\beta}^T \mathbf{x}| / \|\boldsymbol{\beta}\|,$$

from (14). The chosen hyperplane is the solution of the following problem:

$$\begin{cases} \max\limits_{\beta_0, \boldsymbol{\beta}} & m_{\beta_0, \boldsymbol{\beta}} \\ \min\limits_{i=1,\ldots,n} & \dfrac{|\beta_0 + \boldsymbol{\beta}^T \mathbf{X}_i|}{\|\boldsymbol{\beta}\|} \geq m. \end{cases} \tag{15}$$

After the substitution $\mathbf{w} = \frac{\boldsymbol{\beta}}{m\|\boldsymbol{\beta}\|}$ and $w_0 = \frac{\beta_0}{m\|\boldsymbol{\beta}\|}$ [19], the problem (15) becomes:

$$\begin{cases} \min\limits_{w_0, \mathbf{w}} & \dfrac{1}{2}\|\mathbf{w}\|^2 \\ \text{for } i = 1, \ldots, n & y_i(w_0 + \mathbf{w}^T \mathbf{X}_i) \geq 1. \end{cases} \tag{16}$$

A number $d + 1$ of parameters associated to n linear constraints has to be estimated. Since the problem (16) is quadratic with linear constraints, a unique solution can be found. The dual form of the problem is given by the following optimisation problem

$$\begin{cases} \max\limits_{\mathbf{a}} & \displaystyle\sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \displaystyle\sum_{i=1}^{n} a_i y_i = 0 \\ \mathbf{a} \succeq 0. \end{cases} \tag{17}$$

where $\mathbf{a} = (a_1, \ldots, a_n)$ is a Lagrange multiplier. Let $\mathbf{a} = (a_1, \ldots, a_n)$ be the solution of problem (17). The computer model g can be estimated by

$$\widehat{g}_n(\mathbf{x}) = w_0 + \sum_{i=1}^{n} a_i y_i \mathbf{X}_i^T \mathbf{x}. \tag{18}$$

Table 3 suggest to solve the primal problem if $d \leq n$, and the dual problem in the other case.

	primal	dual
Number of parameters	$d+1$	n
Number of constraints	n	$n+1$

TABLE 3: Numbers of constraints and dimension of parameters to be estimated for the primal and dual problems in the usual SVM framework.

B.1.2. *Nonlinear situation.* Assume now that D_n cannot be separated by a hyperplane but by a nonlinear surface. Let K be a symmetrical positive definite kernel. From Mercer's theorem there exists a transformation $h : [0,1]^d \to \mathcal{H}$ where $\mathcal{H}$ is an Hilbert space with the inner product

$$\langle h(\mathbf{x}), h(\mathbf{y}) \rangle_{\mathcal{H}} = K(\mathbf{x}, \mathbf{y}). \tag{19}$$

such that the data can be linearly separated. Then g can be written as

$$g(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^T h(\mathbf{x}),$$

and using (18) and (19), it can be estimated by:

$$\widehat{g}_n(\mathbf{x}) = w_0 + \sum_{i=1}^{n} a_i y_i K(\mathbf{X}_i, \mathbf{x}).$$

The optimization problem (17) becomes

$$\begin{cases} \max\limits_{\mathbf{a}} & \sum\limits_{i=1}^{n} a_i - \dfrac{1}{2}\sum\limits_{i,j=1}^{n} a_i a_j y_i y_j K(\mathbf{X}_i, \mathbf{x}_j) \\ \sum\limits_{i=1}^{n} a_i y_i = 0 \\ \mathbf{a} \succeq 0 \end{cases}$$

or

$$\begin{cases} \min\limits_{\mathbf{a}} & \frac{1}{2}\mathbf{a}^T \widehat{K}\mathbf{a} - \mathbf{d}^T\mathbf{a} \\ A\mathbf{a} \overset{*}{\geq} \mathbf{c}, \end{cases}$$

where the symbol $\overset{*}{\geq}$ means that the first constraint is an equality constraint and the other are inequalities constraints, and

$$A = \begin{pmatrix} y_1 \cdots y_n \\ I_n \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbf{R}^{n+1}, \quad \mathbf{d} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbf{R}^n$$

and $\widehat{K}_{i,j=1,\ldots,n} = y_i y_j K(\mathbf{X}_i, \mathbf{X}_j)$, with I_n the identity matrix.

B.2. Monotonic SVM

Let g be defined by (14) and assume now that g is globally increasing. Let $\mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d$ and $\mathbf{x}^i = (x_1, \ldots, x_{i-1}, x_i + \eta, x_{i+1}, \ldots, x_d)$ with $\eta > 0$, then $\mathbf{x} \preceq \mathbf{x}^i$ and $g(\mathbf{x}) \leq g(\mathbf{x}^i)$. It comes that

$$g\left(\mathbf{x}^i\right) - g(\mathbf{x}) \geq 0 \Rightarrow \beta_i \geq 0,$$

for all $i = 1, \ldots, d$. The problem (16) becomes:

$$\begin{cases} \min\limits_{w_0, \mathbf{w}} & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{for } i = 1, \ldots, n & y_i(w_0 + \mathbf{w}^T \mathbf{X}_i) \geq 1 \cdot \\ \text{for } i = 1, \ldots, d & w_i \geq 0 \end{cases} \tag{20}$$

The associated Lagrangian $\mathcal{M}(\mathbf{w}, w_0, \mathbf{a}, \mathbf{b})$ can be written as

$$\mathcal{M}(\mathbf{w}, w_0, \mathbf{a}, \mathbf{b}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} a_i[y_i(w_0 + \mathbf{w}^T \mathbf{X}_i) - 1] - \mathbf{b}^T \mathbf{w},$$

where $\mathbf{a}$ and $\mathbf{b}$ are the Lagrangian multipliers. The dual problem becomes

$$\begin{cases} \max\limits_{\mathbf{a}, \mathbf{b}} \sum\limits_{i=1}^{n} a_i - \frac{1}{2} \sum\limits_{i,j=1}^{n} a_i a_j y_i y_j \mathbf{X}_i^T \mathbf{X}_j - \mathbf{b}^T \sum\limits_{i=1}^{n} a_i y_i \mathbf{X}_i - \frac{1}{2}\|\mathbf{b}\|^2 \\ \sum\limits_{i=1}^{n} a_i y_i = 0 \\ \mathbf{a} \succeq 0 \\ \mathbf{b} \succeq 0. \end{cases} .$$

Table 4 shows there is always less constraints and parameters to estimate in the primal problem than in the dual problem. In practice, it is more interesting to solve the primal problem.

	primal	dual
Number of parameters	$d + 1$	$d + n$
Number of constraints	$d + n$	$d + n + 1$

TABLE 4: Numbers of constraints and dimension of parameters to be estimated for the primal and dual problems in the monotonic SVM framework.

The problem (20) can then be rewritten.

$$\begin{cases} \min\limits_{w_0, \mathbf{w}} & \frac{1}{2}\|\mathbf{w}\|^2 \\ & AW \geq \mathbf{c} \end{cases}$$

where

$$A = \begin{pmatrix} A_{1,1} \\ A_{2,1} \end{pmatrix}, \quad W = \begin{pmatrix} \mathbf{w} \\ w_0 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} \mathbf{1}_n \\ \mathbf{0}_d \end{pmatrix} \in \mathbf{R}^{n+d},$$

with

$$A_{1,1} = \begin{pmatrix} y_1 \mathbf{X}_1^T & y_1 \\ \vdots & \vdots \\ y_n \mathbf{X}_n^T & y_n \end{pmatrix} \in \mathcal{M}_{n\times(d+1)}(I\!R), \quad A_{2,1} = \begin{pmatrix} I_d & \mathbf{0}_d \end{pmatrix} \in \mathcal{M}_{d\times(d+1)}(I\!R),$$

and I_d the identity matrix in $I\!R^d$, $\mathbf{1}_n^T = (1, \ldots, 1) \in I\!R^n$ and $\mathbf{0}_d^T = (0, \ldots, 0) \in I\!R^d$.