Optimization Of Quasi-convex Function Over Product Measure Sets

Jerome Stenger, Fabrice Gamboa, Merlin Keller

OPTIMIZATION OF QUASI-CONVEX FUNCTION OVER PRODUCT MEASURE SETS

JÉRÔME STENGER[‡]*, FABRICE GAMBOA*[†], AND MERLIN KELLER[‡]

Abstract. We consider a generalization of the Bauer maximum principle. We work with tensorial products of convex measures sets, that are non necessarily compact but generated by their extreme points. We show that the maximum of a quasi-convex lower semicontinuous function on this product space is reached on the tensorial product of finite mixtures of extreme points. Our work is an extension of the Bauer maximum principle in three different aspects. First, we only assume that the objective functional is quasi-convex. Secondly, the optimization is performed over a space built as a product of measures sets. Finally, the usual compactness assumption is replaced with the existence of an integral representation on the extreme points. We focus on the product of two different types of measure sets, called the moment class and the unimodal moment classes. The elements of these classes are probability measures (respectively unimodal probability measures) satisfying *generalized* moment constraints. We show that an integral representation on the extreme points is available for such spaces and that it extends to their tensorial product. We give several applications of the Theorem, going from robust Bayesian analysis to the optimization of a quantile of a computer code output.

Key words. Quasi-convexity, Lower semicontinuity, Optimization, Product space, Measure space

AMS subject classifications. 46E27, 60B11, 62P30, 52A40

1. Introduction. Optimization of convex functions is one of the most studied topic in optimization theory. Indeed, their properties are really interesting especially for minimum search. However, convex functions are also attractive for maximum search. On this matter, the famous Bauer maximum principle [8, 22, 1] states that a convex upper semicontinuous function, defined on a compact convex subset of a locally convex topological vector space, reaches its maximum on some extreme points. However, in practice the functions to optimize are rarely convex, and quasi-convex functions are a well-tailored generalization for optimization. In this paper, we are interested in such quasi-convex functions [27]. They are defined on a convex subset $\mathcal{A}$ of a topological vector space, as functions satisfying the inequality $f(\lambda x+(1-\lambda)y) \leq \max\{f(x), f(y)\}$ for all $x, y \in \mathcal{A}$ and $\lambda \in [0,1]$. Many properties of convex functions have equivalent for quasi-convex functions, we refer to the excellent review on quasi-convex function in [15] where a non-exhaustive list is provided. For instance, the Bauer maximum principle remains true for quasi-convex functions. What is more surprising is that the proof of this claim available in [8, p.102] is similar in all respects to the one for convex functions.

In this paper, we study a quasi-convex lower semicontinuous function (meaning that $\{x \in \mathcal{A} : f(x) \leq \alpha\}$ is a closed and convex set for all $\alpha \in \mathbb{R}$) and its optimization on a product space. To our knowledge, this has not been addressed before. Our work is an extension of [25], where the authors study the optimization of an affine function on a product of measure spaces with moment constraints. Although our theoretical approach deals with general topological spaces, we focus in our applications on convex sets $\mathcal{A}_1, \dots, \mathcal{A}_d$ of probability measures. We aim to optimize a quasi-convex lower semicontinuous (lsc) functional over the product space $\prod_{i=1}^d \mathcal{A}_i$. The product of

*IMT, Toulouse (fabrice.gamboa@math.univ-toulouse.fr).
[†]ANITI - Artificial and Natural Intelligence Toulouse Institute.
[‡]EDF R&D, Chatou (jerome.stenger@edf.fr, merlin.keller@edf.fr).

measure sets, also called tensorization, is one of the key elements of this framework. As we will see, it suits numerous industrial optimization problems that will be discussed in this paper.

The strength of our approach is that contrary to the Bauer maximum principle, we do not assume compactness. Instead of the compactness assumption of the optimization sets $\mathcal{A}_i$ for $1 \leq i \leq d$, we assume the existence of an integral representation [24] by a subset Δ_i of $\mathcal{A}_i$. This means that every measure μ in $\mathcal{A}_i$ is the barycenter of a measure $\boldsymbol{\nu}$ supported on Δ_i, such that for any linear function ϕ of the topological dual of $\mathcal{A}_i$ it holds

$$\langle \phi, \mu \rangle := \int_X \phi(x) \; \mu(dx) = \int_{\Delta_i} \langle \phi, s \rangle \, d\boldsymbol{\nu}(s) \; ,$$

where X denotes the support of μ. The bold type indicates that $\boldsymbol{\nu}$ is a probability measure supported by a set of probability measures. The compact case is also included in this framework as the Choquet representation holds. More precisely, every point of a compact convex set is the barycenter of a probability measure carried by every bordering set (see [8, Theorem 27.6] for details, note that the representation is supported on the set of extreme points if the space is also metrizable [8, p.140]). The existence of the integral representation is a strong assumption. However, we will provide two different measure spaces for which the integral representation holds; namely the moment class [36, 34] and the unimodal moment class [5].

Our main theorem is therefore an extension of the Bauer maximum principle in three directions: quasi-convexity of the optimization function replaces convexity, tensorization generalizes the structure of the optimization space, and the existence of an integral representation on the marginal sets covers the compact case. By doing so, we build a framework that includes many optimization procedures developed earlier. We refer for example to robust Bayesian analysis [29, 4, 30], that studies the sensitivity of the Bayesian analysis to the choice of an uncertain prior distribution. Another example is the work in [25] called Optimal Uncertainty Quantification. Further, we present new applications, all illustrated on a toy case. The theoretical approach is made as general as possible, while the proofs of our claims only rely on simple topological arguments.

The paper is organized as follows. Section 2 provides the framework basis before introducing our main result in Section 3. Section 4 is dedicated to the presentation of some applications that are illustrated in Section 5 on a use case. Section 6 provides the theoretical formulation and proofs of our results. We discuss in the last section conclusion and perspectives of this work.

2. Measure spaces. We will work with a subset of $\mathcal{P}(X)$, the set of all Borel probability measures on a topological space X (specified in the following). Let $C_b(X)$ denote the set of all continuous bounded real valued functions on X. We deal with a convex subset of $\mathcal{P}(X)$ satisfying the integral representation property. Note that generally speaking, $\mathcal{P}(X)$ can be considered as a subset of the closed unit ball of the topological dual of $C_b(X)$ and it inherits its topology, which is the topology of weak* convergence. Moreover, the weak* topology is always locally convex, since it is induced by the seminorms $\mu \mapsto |\langle \phi, \mu \rangle|$, where $\phi \in C_b(X)$, and μ belongs to the topological dual of $C_b(X)$.

2.1. Moment class. Assume now that X is a Suslin space [7], i.e. the image of a Polish space under a continuous mapping. We study a convex subspace of $\mathcal{P}(X)$, called the moment class. All measures in the moment class $\mathcal{A}^*$ satisfy *generalized* moment constraints. That is, a measure $\mu \in \mathcal{A}^*$ verifies $\mathbb{E}_\mu[\varphi_i] \leq 0$, where $\mathbb{E}_\mu[\varphi_i]$ denotes the expectation of φ_i with respect to μ, for given measurable functions $\varphi_1, \dots, \varphi_n \in C_b(X)$. Because X is Suslin, all measures $\mu \in \mathcal{P}(X)$ are regular. Hence, the following Theorem 2.1 due to Winkler [36, p.586] holds.

Theorem 2.1 (Extreme points of moment class). *Consider the space $\mathcal{P}(X)$ of Borel measures on a Suslin space X, and measurable functions $\varphi_1, \dots, \varphi_n$ on X. Then, for any measure μ in the moment class $\mathcal{A}^* = \{\mu \in \mathcal{P}(X) \mid \mathbb{E}_\mu[\varphi_i] \leq 0, 1 \leq i \leq n\}$, there exists a probability measure $\boldsymbol{\nu}$ supported on $\Delta^*(n)$ such that μ is the barycenter of $\boldsymbol{\nu}$. Where*

$$\Delta^*(n) = \left\{ \mu \in \mathcal{A}^* \;:\; \mu = \sum_{i=1}^{n+1} \omega_i \delta_{x_i},\, \omega_i \geq 0,\, x_i \in X \right\} .$$

is the set of discrete probability measures of $\mathcal{A}^*$ supported on at most $n+1$ points.

The case of equality in the constraints defining $\mathcal{A}^*$ is covered by this result [36, p.586]. Theorem 2.1 states that the extreme points of a class of measures with n *generalized* moment constraints are discrete measures supported on at most $n+1$ points of X.

2.2. Unimodal moment class. In this section, X denotes an interval of the real line $\mathbb{R}$. Let μ be a probability distribution on X, and let F_μ be its distribution function, that is $F_\mu(h) := \mu(]-\infty, h])$. The measure μ is said to be unimodal with mode at a, if F_μ is convex on $]-\infty, a[$ and concave on $]a, +\infty[$. We denote $\mathcal{H}_a(X)$ the set of all probability measures on X which are unimodal at a. The set $\mathcal{H}_a(X)$ is closed but not necessarily compact ($\mathcal{H}_a(\mathbb{R})$ is not compact, see [5, p.19]). Clearly, any uniform probability measure on an interval of the form $\mathrm{co}(a, z)$, $z \in X$ (co is the convex hull) including the Dirac mass in a, is unimodal at a. The set $\mathcal{U}_a(X) = \{\mathfrak{u} \text{ is uniformly distributed on } \mathrm{co}(a, z), z \in X\}$ of these uniform probability measures is closed in $\mathcal{P}(X)$ [5, p.19]. In this section, we are interested by the convex subset $\mathcal{A}^\dagger$ of unimodal measures satisfying *generalized* moment constraints $\mathbb{E}_\mu[\varphi_i] \leq 0$, for measurable functions $\varphi_1, \dots, \varphi_n$. This subspace is called a unimodal moment class and an equivalent of Theorem 2.1 holds:

Theorem 2.2 (Extreme points of unimodal class). *Consider the space $\mathcal{H}_a(X)$ of unimodal measures on an interval X with mode a, and measurable functions $\varphi_1, \dots, \varphi_n$ on X. Then, for any measure μ in the unimodal moment class $\mathcal{A}^\dagger = \{\mu \in \mathcal{H}_a(X) \mid \mathbb{E}_\mu[\varphi_i] \leq 0, 1 \leq i \leq n\}$, there exists a probability measure $\boldsymbol{\nu}$ supported on $\Delta^\dagger(n)$ such that μ is the barycenter of $\boldsymbol{\nu}$. Here*

$$\Delta^\dagger(n) = \left\{ \mu \in \mathcal{A}^\dagger \mid \mu = \sum_{i=1}^{n+1} \omega_i \mathfrak{u}_i,\, \omega_i \geq 0,\, \mathfrak{u}_i \in \mathcal{U}_a(X) \right\} .$$

Elements of $\Delta^\dagger(n)$ are mixtures of at most $n+1$ uniform distributions supported on $\mathrm{co}(a, z)$ for some $z \in X$.

The proof of this theorem is postponed to the Appendix. The unimodal class was first explored by Khinchin [21] who revealed the fundamental relationship between the set of unimodal probability distributions and uniform probability densities. It was later demonstrated in [6] that the Khinchin Theorem may be considered as a non-compact form of the Krein-Milman Theorem [8, p.105]. In [30], one can find a first application of this result to the context of robust Bayesian analysis. In the same paper, the class of *symmetric* unimodal distributions with mode a is also considered. They show that the extreme points of this set are mixture of uniform probability measures on an interval of the form $[\theta - z, \theta + z]$, $z \geq 0$.

The sets $\mathcal{A}^\dagger$ and $\mathcal{A}^*$ are very interesting. Indeed, measure spaces are non-obvious sets and it is generally not straightforward to exhibit their extreme points.

3. Main results.

3.1. Construction of the product measure spaces. We now give our main Theorem. The measure sets introduced in Section 2 have very similar properties, so that they are gathered under the same notation. Indeed, we enforce *generalized* moment constraints in both cases. The difference lies in the unimodality of the measures of $\mathcal{A}^\dagger$, while $\mathcal{A}^*$ can contain any Radon measure. Consequently, the difference between Theorem 2.1 and 2.2 lies in the nature of the extreme points. Indeed, the generator of the unimodal moment class $\mathcal{A}^\dagger$ is the set of finite convex combination of uniform distributions. Whereas, the generator of the moment class $\mathcal{A}^*$ is the set of finite convex combinations of Dirac masses.

To begin with, we detail the construction of the product space. Let $\mathcal{X} := X_1 \times \cdots \times X_d$ be a product of p Suslin spaces $X_1, \ldots, X_p$, and $d-p$ real intervals $X_{p+1}, \ldots, X_d$. Given real numbers $a_i \in X_i$ for $p < i \leq d$ and measurable functions $\varphi_i^{(j)} : X_i \to \mathbb{R}$ for $1 \leq j \leq N_i$ and $1 \leq i \leq d$, we construct d measure spaces with the integral representation property

$$\mathcal{A}_i = \mathcal{A}_i^* = \left\{ \mu_i \in \mathcal{P}(X_i) \mid \mathbb{E}_{\mu_i}[\varphi_i^{(j)}] \leq 0 \text{ for } j = 1, \ldots, N_i \right\} \quad \text{for } 1 \leq i \leq p \ ,$$
$$\mathcal{A}_i = \mathcal{A}_i^\dagger = \left\{ \mu_i \in \mathcal{H}_{a_i}(X_i) \mid \mathbb{E}_{\mu_i}[\varphi_i^{(j)}] \leq 0 \text{ for } j = 1, \ldots, N_i \right\} \text{ for } p < i \leq d \ .$$

Therefore, the space $\mathcal{A}_i$ is either a moment space on a Suslin space, or a unimodal moment space on an interval as presented in Section 2. We denote by $\Delta_{N_i} \subset \mathcal{A}_i$, the generator of the space $\mathcal{A}_i$, as defined in Section 2. Summarizing, we have

$$\Delta_i(N_i) = \Delta_i^*(N_i) = \left\{ \mu_i \in \mathcal{A}_i \mid \mu_i = \sum_{k=1}^{N_i+1} \omega_k \delta_{x_k},\, x_k \in X_i \right\} \text{ for } 1 \leq i \leq p \ , \tag{3.1}$$
$$\Delta_i(N_i) = \Delta_i^\dagger(N_i) = \left\{ \mu_i \in \mathcal{A}_i \mid \mu_i = \sum_{k=1}^{N_i+1} \omega_k \mathfrak{u}_k,\, \mathfrak{u}_k \in \mathcal{U}_{a_i}(X_i), \right\} \text{ for } p < i \leq d \ .$$

With these definitions and as discussed in the previous section, any measure $\mu_i \in \mathcal{A}_i$ is the barycenter of a probability measure supported on $\Delta_i(N_i)$, the set of convex combination of at most $N_i + 1$ Dirac masses or uniform distributions.

In the remaining of the paper, the product spaces $\mathcal{A} = \prod_{i=1}^d \mathcal{A}_i$ and $\Delta = \prod_{i=1}^d \Delta_i(N_i)$ are equipped with the product σ-algebra (not to be confused with the Borel σ-algebra of the product).

The following definition specifies the meaning of quasi-convexity and lower semicontinuity of a function on a product space.

Definition 3.1. *A function $f : \mathcal{A} \to \mathbb{R}$ is said to be marginally quasi-convex (marginally lsc) if for all $\{\mu_k \in \mathcal{A}_k,\, k \neq i\}$, the function $\mu_i \mapsto f(\mu_1, \ldots, \mu_d)$ is quasi convex (respectively lsc for the topology of $\mathcal{A}_i$).*

Notice that if f is globally quasi-convex (*lsc* for the product topology) then it is marginally quasi-convex (respectively marginally *lsc*). Indeed, if f is globally *lsc*, then $\{\mu \in \mathcal{A} \mid f(\mu_1, \ldots, \mu_d) > \alpha\}$ is open for any α and as the canonical projections are open maps, $\{\mu_i \in \mathcal{A}_i \mid f(\mu_1, \ldots, \mu_d) > \alpha\}$ is also open. It is clear for quasi-convexity. Having defined properly the product spaces, we give our main result in the next section.

3.2. Reduction Theorem. We assume that any measure $\mu_i \in \mathcal{A}_i$ belongs to either a moment space or a unimodal moment space. Hence, μ_i always satisfies N_i moment constraints. In the following theorem (Theorem 3.2), we also enforce constraints on the product measure $\mu = \mu_1 \otimes \cdots \otimes \mu_d \in \mathcal{A}$, such that, for N measurable functions $\varphi^{(j)} : \mathcal{X} \to \mathbb{R}$, $1 \leq j \leq N$, we have $\mathbb{E}_\mu[\varphi^{(j)}] \leq 0$. In other words, we investigate a subspace of $\mathcal{A}$ defined as $\{\mu \in \mathcal{A} \mid \mathbb{E}_\mu[\varphi^{(j)}] \leq 0,\, 1 \leq j \leq N\}$.

Theorem 3.2. *Suppose that $\mathcal{X}, \mathcal{A}$ and Δ are defined as in Subsection 3.1. Let $\varphi^{(j)} : \mathcal{X} \to \mathbb{R}$, $1 \leq j \leq N$ be measurable functions. Let $f : \mathcal{A} \to \mathbb{R}$ be a marginally quasi-convex lower semicontinuous function. Then,*

$$\sup_{\substack{\mu_1,\ldots,\mu_d \in \otimes \mathcal{A}_i \\ \mathbb{E}_{\mu_1,\ldots,\mu_d}[\varphi^{(j)}] \leq 0 \\ 1 \leq j \leq N}} f(\mu) \quad = \quad \sup_{\substack{\mu_1,\ldots,\mu_d \in \otimes \Delta_i(N_i+N) \\ \mathbb{E}_{\mu_1,\ldots,\mu_d}[\varphi^{(j)}] \leq 0 \\ 1 \leq j \leq N}} f(\mu)$$

In other words, the supremum of a quasi-convex function on a product space can be computed considering only the d-fold product of finite convex combinations of extreme points of the marginal spaces. That is, finite convex combinations of either Dirac masses or uniform distributions.

The underlying assumption is that there exists an integral representation on $\mathcal{A}_i$ for $1 \leq i \leq d$. We take two examples of measure spaces whose generators are known, but the proofs in Section 6 are given in a much more general framework. Notice that this assumption is somehow different from the one used in the Bauer maximum principle, wherein the compactness of the convex space is assumed. However, from the Krein-Millman Theorem [8, p.105], the compactness assumption implies the integral representation. Further, since we extend the result to product spaces, our framework is more general. The next proposition highlights that the existence of an integral representation on each marginal space, also implies the existence of an integral representation on the product space.

Proposition 3.3. *Let $\mathcal{X}, \mathcal{A}$ and Δ be defined as in Subsection 3.1, such that the integral representation property holds on every marginal space $\mathcal{A}_i$. Then any measure in $\mathcal{A}$ is also the barycenter of a probability measure supported by Δ.*

Proof. let μ be in $\mathcal{A}$, so that $\mu = \mu_1 \otimes \cdots \otimes \mu_d$, with $\mu_i \in \mathcal{A}_i$. Because of the integral representation property of $\mathcal{A}_i$, there exists a probability measure $\boldsymbol{\nu_i}$ supported by $\Delta_i(N_i)$ such that μ_i is the barycenter of $\boldsymbol{\nu_i}$, i.e. $\mu_i = \int_{\Delta_i(N_i)} s_i \, d\boldsymbol{\nu}(s_i)$. Therefore, for any function $\phi \in C_b(\mathcal{X})$, using Fubini's Theorem, we have

$$\begin{aligned}
\int_{\mathcal{X}} \phi(\mathbf{x}) \, \mu(\mathrm{d}\mathbf{x}) &= \int_{\mathcal{X}} \phi(x_1, \ldots, x_d) \, \mu_1(\mathrm{d}x_1) \ldots \mu_d(\mathrm{d}x_d) \ , \\
&= \int_{\mathcal{X}} \phi(x_1, \ldots, x_d) \int_{\Delta_1(N_1)} s_1(\mathrm{d}x_1) \, \mathrm{d}\boldsymbol{\nu}_1(s_1) \cdots \int_{\Delta_d(N_d)} s_d(\mathrm{d}x_d) \, \mathrm{d}\boldsymbol{\nu}_d(s_d) \ , \\
&= \int_{\Delta_1(N_1)} \cdots \int_{\Delta_d(N_d)} \int_{\mathcal{X}} \phi(x_1, \ldots, x_d) \, s_1(\mathrm{d}x_1) \cdots s_d(\mathrm{d}x_d) \, \mathrm{d}\boldsymbol{\nu}_1(s_1) \ldots \mathrm{d}\boldsymbol{\nu}_d(s_d) \ , \\
&= \int_{\Delta} \int_{\mathcal{X}} \phi(\mathbf{x}) \, s(\mathrm{d}\mathbf{x}) \, \mathrm{d}\boldsymbol{\nu}(s) \ ,
\end{aligned}$$

where $\boldsymbol{\nu} = \boldsymbol{\nu}_1 \otimes \cdots \otimes \boldsymbol{\nu}_d$ is a probability measure supported on Δ. This means that μ is the barycenter of $\boldsymbol{\nu}$ in the product space $\mathcal{A}$. ∎

The proposition above guarantees the existence of the integral representation on the product space $\mathcal{A}$ whenever each marginal space $\mathcal{A}_i$ possesses itself an integral representation property. However, the subspace of $\mathcal{A}$ restricted by moment constraints on the joint distribution $\{\mu \in \mathcal{A} \mid \mathbb{E}_\mu[\varphi^{(j)}] \leq 0, 1 \leq j \leq N\}$ has a more complex structure than the marginal spaces $\mathcal{A}_i$ from Section 2. Indeed, its extreme points are not convex combinations of $N+1$ elements of the generator of $\mathcal{A}$: $\Delta = \bigotimes_{i=1}^d \Delta_i(N_i)$. As stated in reduction Theorem 3.2, its extreme points are elements of the d-fold product of finite convex combinations of extreme points of $\mathcal{A}_i$, that is $\bigotimes_{i=1}^d \Delta_i(N_i + N)$.

Notice also that Theorem 3.2 extends the work of [25]. Indeed, in this paper the authors were the first to propose the reduction Theorem on a product space. Nevertheless, the optimization considered therein is restricted only to products of moment classes and did not include unimodal moment classes. Moreover, the optimized functional in [25] is an affine function of the measure. This is a very particular case of our framework. We emphasize that measure affine functions are useful; some of their properties are discussed in the next section.

3.3. Relaxation of the lower semicontinuity assumption.

3.3.1. Measure affine functions. The function to be optimized is assumed both lower semicontinuous and quasi-convex. It appears that quasi-convexity covers a large class of functionals fitting most of our application cases. Nevertheless, lower semicontinuity is not always satisfied. So that, it is very interesting to relax this assumption.

In this section, we study specific class of functions that are called measure affine [36]. These functions and their optimization on product measure spaces have been already studied in [25]. We recall that $\mathcal{X}, \mathcal{A}$ and Δ are the product spaces constructed in Subsection 3.1.

Definition 3.4. *A function F is called measure affine whenever F is integrable with respect to any probability measure $\boldsymbol{\nu}$ on Δ with barycenter $\mu \in \mathcal{A}$ and F fulfills the following barycentrical formula*

$$F(\mu) = \int_\Delta F(s)\, d\boldsymbol{\nu}(s) \ .$$

Notice that any measure affine function satisfies $F(\lambda\mu + (1-\lambda)\pi) = \lambda F(\mu) + (1-\lambda)F(\pi)$, for $\mu, \pi \in \mathcal{A}$ and $\lambda \in [0,1]$. Hence, it is both quasi-convex and quasi-concave. In the following, we show that the optimum of a measure affine function can be computed only on the extreme points of the optimization set, independently of the regularity of F. For an extended version enforcing moment constraints on the product measure as in Theorem 3.2, we refer to [25, p.71].

Theorem 3.5. *Let $\mathcal{A}$ be a convex subset of a locally convex topological vector space satisfying barycentric property. For any measure affine functional F we have*

$$\sup_{\mu \in \mathcal{A}} F(\mu) = \sup_{\mu \in \Delta} F(\mu) \ ,$$

and,

$$\inf_{\mu \in \mathcal{A}} F(\mu) = \inf_{\mu \in \Delta} F(\mu) \ .$$

Proof. The proof is given for the supremum, but it is similar for the infimum. Given $\mu \in \mathcal{A}$, the integral representation property states that there exists a probability measure $\boldsymbol{\nu}$ supported on Δ such that μ is the barycenter of $\boldsymbol{\nu}$. Therefore,

$$F(\mu) = \int_\Delta F(s) d\boldsymbol{\nu}(s) \leq \sup_{s\in\Delta} F(s)$$

for any $\mu \in \mathcal{A}$. Hence, $\sup_{\mu\in\mathcal{A}} F(\mu) \leq \sup_{\mu\in\Delta} F(\mu)$, the converse is obvious as $\Delta \subset \mathcal{A}$. ∎

3.3.2. Ratio of measure affine functions. From the previous theorem, the supremum of a measure affine functional can be searched for only on the generator of the measure space $\mathcal{A}$. We examine some transformation of measure affine functions under which the reduction Theorem 3.2 still holds and for which the lower semi-continuity assumption remains a non-necessary condition. Ratios of measure affine functions are particularly interesting as it appears in many practical quantities of interest (see for instance Subsection 4.4 and Subsection 4.5).

PROPOSITION 3.6. *Let $\mathcal{A}$ be a convex set of measures with generator Δ. Let ϕ and ψ be two measure affine functionals, $\psi > 0$. Then*

$$\sup_{\mu\in\mathcal{A}} \frac{\phi}{\psi} = \sup_{\mu\in\Delta} \frac{\phi}{\psi} \ ,$$

and,

$$\inf_{\mu\in\mathcal{A}} \frac{\phi}{\psi} = \inf_{\mu\in\Delta} \frac{\phi}{\psi} \ .$$

Proof. The proof is given for the supremum, but it is similar for the infimum. Given $\mu \in \mathcal{A}$, the integral representation property states that there exists a probability measure $\boldsymbol{\nu}$ supported on Δ with barycenter μ. Therefore,

$$\begin{aligned}
\phi(\mu) &= \int_\Delta \phi(s)\, d\boldsymbol{\nu}(s) \ , \\
&= \int_\Delta \frac{\phi(s)}{\psi(s)} \psi(s)\, d\boldsymbol{\nu}(s) \ , \\
&\leq \sup_\Delta \frac{\phi}{\psi} \int_\Delta \psi(s)\, d\boldsymbol{\nu}(s) \ , \\
&= \sup_\Delta \frac{\phi}{\psi}\ \psi(\mu) \ .
\end{aligned}$$

So that, $\phi(\mu)/\psi(\mu) \leq \sup_\Delta \phi/\psi$ for all $\mu \in \mathcal{A}$, hence $\sup_\mathcal{A} \phi/\psi \leq \sup_\Delta \phi/\psi$. The other inequality is obvious as $\Delta \subset \mathcal{A}$. ∎

Notice that the ratio of a convex function by a positive concave function is quasi-convex [9, p.51]. Thus, in the previous Proposition the ratio ϕ/ψ is quasi-convex.

4. Applications. In this section, we study some practical applications of Theorem 3.2, based on real life engineering problems. In the following, we consider a computer code G, that can be seen as a black box function. The code G takes d scalar input parameters, that may represent for instance physical quantities. In uncertainty quantification (UQ) methods [14], we aim to assess the uncertainty tainting the result of the computer simulation, whose input values are uncertain and modeled as random variables $X_i \sim \mu_i$. They are all assumed to be independent for sake of simplicity. The output of the code $Y = G(X_1, \dots, X_d)$ is therefore also a random variable. Generally, one is interested by the computation of some quantity of interest on the output of the code. However, the choice of the input distributions $(\mu_i)_{\{1\leq i\leq d\}}$ is often itself uncertain. So that, the distributions are often restricted for simplicity to some parametric family, such as Gaussian or uniform. The distribution parameters are then generally estimated with the available information coming from data and/or expert opinion. In practice, this information is often reduced to an input mean value or a variance. We aim to account for the uncertainty on the input distribution choice. So that, we wish to evaluate the maximal quantity of interest over a class of probability distributions.

In this section, $\mathcal{X}, \mathcal{A}$ and Δ are constructed as in Subsection 3.1. Therefore, the input distribution $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ is an element of $\mathcal{A}$, i.e. a product of d independent input measures $\mu_i \in \mathcal{A}_i$. Thus, any input distribution is imprecisely defined and supposed to belong to a moment class or a unimodal moment class.

4.1. Example of measure affine functions. It was shown in Subsection 3.3.1, that measure affine functions were particularly interesting because of the relaxation of the lower semicontinuity assumption. Moreover, we have also seen that an affine functional is both quasi-convex and quasi-concave. Hence, it is possible to minimize or maximize the quantity of interest (Theorem 3.5). We study here specific measure affine functions.

PROPOSITION 4.1. *Let $\mathcal{A}$ be a convex set of probability measure with generator Δ, and let q_G be integrable on $\mathcal{X}$ with respect to any measure in Δ. Then the functional $\mu \mapsto \mathbb{E}_\mu[q_G] = \int_{\mathcal{X}} q_G \, d\mu$ is measure affine.*

Proof.

$$\begin{aligned}
\mathbb{E}_\mu[q_G] &= \int_{\mathcal{X}} q_G(\mathbf{x})\, \mu(d\mathbf{x}) \ , \\
&= \int_{\mathcal{X}} q_G(\mathbf{x}) \int_{\Delta} s(d\mathbf{x})\, d\boldsymbol{\nu}(s) \ , \\
&= \int_{\Delta} \left(\int_{\mathcal{X}} q_G(\mathbf{x})\, s(d\mathbf{x}) \right) d\boldsymbol{\nu}(s) \ , \\
&= \int_{\Delta} \mathbb{E}_s[q_G]\, d\boldsymbol{\nu}(s) \ .
\end{aligned}$$

The measure affine functional $\mu \mapsto \mathbb{E}_\mu[q_G]$ covers a large range of interest quantities. For instance, the choice $q_G(\mathbf{x}) = G(\mathbf{x})$ leads to the expectation of the computer code G. Further, any moment can be studied using $q_G(\mathbf{x}) = G(\mathbf{x})^n$. The choice $q_G(\mathbf{x}) = \mathbb{1}_{C_G}$, the indicator function on a set C_G, yields a probability. An important example would be $q_G(\mathbf{x}) = \mathbb{1}_{\{G(\mathbf{x})\leq h\}}$, which yields the failure probability at threshold $h \in \mathbb{R}$. The choice of a loss function $q_G(\mathbf{x}) = L(G(\mathbf{x}), a)$ where a is some decision, yields the expected loss associated to the decision a.

The interested reader will remark that the lower semicontinuity of the previous affine functional relies on that of q_G. More precisely, lower semicontinuity of q_G (respectively upper semicontinuous) implies lower semicontinuity (respectively upper semicontinuity) of the mapping $\mu \mapsto \int q_G \, d\mu$ [2, Theorem 15.5].

4.2. Non-Linear Quantities. We briefly extend the function presented in Subsection 4.1 to deal with more general quantities of the form proposed in [4]. That is

$$\mu \mapsto F(\mu) = \int q_G(\mathbf{x}, \varphi(\mu))\, \mu(d\mathbf{x}) \ ,$$

where $\varphi(\mu)$ is measurable. The most common example is:

$$q_G(\mathbf{x}, \varphi(\mu)) = (G(\mathbf{x}) - \mathbb{E}_\mu[G(\mathbf{x})])^2 \ ,$$

that yields the variance of the distribution μ. In order to compute this quantity, one needs to linearize the problem. The idea is to rewrite the optimization set $\mathcal{A}$ as

$$\sup_{\mu \in \mathcal{A}} \int q_G(\mathbf{x}, \varphi(\mu))\, \mu(d\mathbf{x}) = \sup_{\varphi_0} \sup_{\substack{\mu \in \mathcal{A} \\ \varphi(\mu) = \varphi_0}} \int q_G(\mathbf{x}, \varphi_0)\, \mu(d\mathbf{x}) \ .$$

The reduction Theorem 3.2 applies to the measure affine function $\mu \mapsto \int q_G(\mathbf{x}, \varphi_0)$ $\mu(d\mathbf{x})$ on the set $\{\mu \in \mathcal{A} \mid \varphi(\mu) = \varphi_0\}$, which is the set $\mathcal{A}$ with additional constraints.

4.3. Quantile Function.

4.3.1. Lower Quantile Function. A classical measure of risk widely used in industrial applications [33, 31], is the quantile of the output. It is a critical criterion for evaluating safety margins [19]. In the following, F_μ denotes the cumulative distribution function of the output of the code, i.e. $F_\mu(\alpha) = \mathbb{P}(G(X) \le \alpha)$.

Theorem 4.2. *We suppose that the code G is continuous. Then the quantile function $\mu \mapsto Q_p^L(\mu) = \inf\{x : F_\mu(x) \ge p\}$ is quasi-convex and lower semicontinuous on $\mathcal{A}$.*

Proof. A function is quasi-convex if all lower level sets are convex. Further, it is lower semicontinuous if they are closed. Hence, we consider for $\alpha \in \mathbb{R}$ the lower level set for $\alpha \in \mathbb{R}$:

$$\begin{aligned} L_\alpha &= \left\{\mu \in \mathcal{A} \mid Q_p^L(\mu) \le \alpha\right\} \ , \\ &= \left\{\mu \in \mathcal{A} \mid F_\mu(\alpha) \ge p\right\} \ . \end{aligned}$$

Indeed, the quantile is the unique function satisfying the Galois inequalities. Therefore,

$$L_\alpha = \left\{\mu \in \mathcal{A} \mid \mu\left(G^{-1}(]-\infty, \alpha])\right) \ge p\right\} \ .$$

L_α is obviously convex and applying Corollary 15.6 in [2], L_α is also closed (for the weak topology), as $G^{-1}(]-\infty, \alpha])$ is closed . □

Notice that, in this work, the quantile is a function of the measure μ. However, the quantile seen as a function of random variables is not quasi-convex, this subtle point is explained in [13].

4.3.2. Upper Quantile Function. In this paragraph we investigate the minimal value of the quantile of the computer model G. In that way, we dispose of bounds around the quantile, which quantifies the range of variation of this QoI over the measure space. However, in order to minimize the quantile, both the upper semicontinuity and quasi-concavity of the optimization function are required. A modified quantity of interest called the upper quantile function is proposed hereunder [16].

Theorem 4.3. *We suppose that the code G is continuous. Then, the upper quantile function $\mu \mapsto Q_p^R(\mu) = \inf\{x : F_\mu(x) > p\}$ is quasi-concave upper semicontinuous on $\mathcal{A}$.*

Proof. A function is quasi-concave if all upper level sets are convex. It is upper semicontinuous if any upper level set is closed. For $\alpha \in \mathbb{R}$, the upper level set is

$$\begin{aligned}
U_\alpha &= \left\{ \mu \in \mathcal{A} \mid Q_p^R(\mu) \geq \alpha \right\} \ , \\
&= \{ \mu \in \mathcal{A} \mid \forall \varepsilon > 0 : F_\mu(\alpha - \varepsilon) \leq p \} \ , \\
&= \bigcap_{\varepsilon > 0} \{ \mu \in \mathcal{A} \mid F_\mu(\alpha - \varepsilon) \leq p \} \ , \\
&= \bigcap_{\varepsilon > 0} \left\{ \mu \in \mathcal{A} \mid \mu(G^{-1}(] - \infty, \alpha - \varepsilon])) \leq p \right\} \ , \\
&= \bigcap_{\varepsilon > 0} \left\{ \mu \in \mathcal{A} \mid \mu(G^{-1}(] - \infty, \alpha - \varepsilon[)) \leq p \right\} \ ,
\end{aligned}$$

The last two equations differ from the measured interval that is succesively semi-closed then open in the last equation. We prove that this equality holds in two times. For $\varepsilon > 0$, we denote

$$F_c(\varepsilon) = \left\{ \mu \in \mathcal{A} \mid \mu(G^{-1}(] - \infty, \alpha - \varepsilon])) \leq p \right\} \ ,$$

which accounts for the semi-closed interval, and

$$F_o(\varepsilon) = \left\{ \mu \in \mathcal{A} \mid \mu(G^{-1}(] - \infty, \alpha - \varepsilon[)) \leq p \right\} \ ,$$

which is based on the open interval. Clearly, $F_c(\varepsilon) \subset F_o(\varepsilon)$ for all $\varepsilon > 0$, so that

$$\cap_{\varepsilon > 0} F_c(\varepsilon) \subset \cap_{\varepsilon > 0} F_o(\varepsilon) \ .$$

For the reverse inclusion, let μ be an element of $\cap_{\varepsilon>0} F_o(\varepsilon)$. Suppose that μ is not in $\cap_{\varepsilon>0} F_c(\varepsilon)$. Then, there exists an $\varepsilon_0 > 0$ such that $\mu(G^{-1}(] - \infty, \alpha - \varepsilon_0])) > p$. But $\mu(G^{-1}(] - \infty, \alpha - \varepsilon_0])) \leq \mu(G^{-1}(] - \infty, \alpha - \frac{\varepsilon_0}{2}[)) \leq p$, because μ is in $\cap_{\varepsilon>0} F_o(\varepsilon)$ by construction, leading to a contradiction.

To conclude, [2, Corollary 15.6] proves that $F_o(\varepsilon)$ is closed because $G^{-1}(] - \infty, \alpha - \varepsilon[)$ is open as G is continuous. Hence, U_α is closed as an intersection of closed sets. U_α is also obviously convex. ∎

4.4. Sensitivity index. Global sensitivity analysis aims at determining which uncertain parameters of a computer code mainly drive the output. In that matter, Sobol' indices are widely used as they quantify the contribution of each input to the variance of the output of the model [18]. However, because the probability distributions modeling the uncertain parameters are themselves uncertain, we propose to evaluate bounds on the Sobol' indices over a class of probability measures. We will focus for simplicity on the well-known first order sensitivity index:

$$S_i(\mu_i) = \frac{\mathrm{Var}_{\mu_i}(\mathbb{E}_{\sim i}[Y|X_i])}{\mathrm{Var}(Y)} \ ,$$

where $\mathbb{E}_{\sim i}$ denotes the expectation over all but the ith input variable. The total-effect index S_{Ti} [18] could be processed in the same way (see Equation (5.4)).

THEOREM 4.4. *Let $\mathcal{X}, \mathcal{A}$ and Δ be defined as in Subsection 3.1. Then*

$$\begin{aligned}
\sup_{\mu_i \in \mathcal{A}_i} S_i(\mu_i) &= \sup_{\mu_i \in \Delta_i(N_i+1)} S_i(\mu_i) \ , \\
\inf_{\mu_i \in \mathcal{A}_i} S_i(\mu_i) &= \inf_{\mu_i \in \Delta_i(N_i+1)} S_i(\mu_i) \ .
\end{aligned}$$

Proof. The proof is made for the supremum but is similar for the infimum

$$\begin{aligned}
\sup_{\mu_i \in \mathcal{A}_i} S_i(\mu_i) &= \sup_{\mu_i \in \mathcal{A}_i} \frac{\mathbb{E}_{\mu_i}\left[(\mathbb{E}_{\sim i}[Y|X_i])^2\right] - \left(\mathbb{E}_{\mu_i}\left[\mathbb{E}_{\sim i}[Y|X_i]\right]\right)^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{\sim i}[Y^2]\right] - (\mathbb{E}[Y])^2}\ , \\
&= \sup_{\mu_i \in \mathcal{A}_i} \frac{\mathbb{E}_{\mu_i}\left[(\mathbb{E}_{\sim i}[Y|X_i])^2\right] - (\mathbb{E}[Y])^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{\sim i}[Y^2]\right] - (\mathbb{E}[Y])^2}\ , \\
&= \sup_{\overline{y}_0} \sup_{\substack{\mu_i \in \mathcal{A}_i \\ \mathbb{E}[Y] = \overline{y}_0}} \frac{\mathbb{E}_{\mu_i}\left[(\mathbb{E}_{\sim i}[Y|X_i])^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{\sim i}[Y^2]\right] - \overline{y}_0^2}\ ,
\end{aligned}$$

where $\overline{y}_0$ is a real. Now, the function

$$\mu_i \longmapsto \frac{\mathbb{E}_{\mu_i}\left[(\mathbb{E}_{\sim i}[Y|X_i])^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{\sim i}[Y^2]\right] - \overline{y}_0^2}\ ,$$

is a ratio of two measure affine functionals. Proposition 3.6 states that the reduction Theorem 3.2 applies so that

$$\sup_{\substack{\mu_i \in \mathcal{A}_i \\ \mathbb{E}[Y] = \overline{y}_0}} \frac{\mathbb{E}_{\mu_i}\left[(\mathbb{E}_{\sim i}[Y|X_i])^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{\sim i}[Y^2]\right] - \overline{y}_0^2} = \sup_{\substack{\mu_i \in \Delta_i(N_i+1) \\ \mathbb{E}[Y] = \overline{y}_0}} \frac{\mathbb{E}_{\mu_i}\left[(\mathbb{E}_{\sim i}[Y|X_i])^2\right] - \overline{y}_0^2}{\mathbb{E}_{\mu_i}\left[\mathbb{E}_{\sim i}[Y^2]\right] - \overline{y}_0^2}\ ,$$

and the result follows. ◻

4.5. Robust Bayesian framework. Robust Bayesian analysis [29] studies the influence of the choice of an uncertain prior distribution. The answer is robust if it does not depend significantly on the choice of the inputs prior distributions. Therefore, a Bayesian analysis is applied to all possible prior distributions in a given class of measures.

The posterior probability distribution can be calculated with Bayes' Theorem by multiplying the prior probability distribution π by the likelihood function $\theta \mapsto l(x \,|\, \theta)$, and then dividing by the normalizing constant, as follows:

$$l(\theta \,|\, x) = \frac{l(x \,|\, \theta)\pi(\theta)}{\int l(x \,|\, \theta)\, \pi(d\theta)}$$

Thus, it is natural to define Ψ that maps the prior probability measure to the posterior probability measure. In what follows, X denotes a Polish space

$$\begin{aligned}
\Psi :\ \ \mathcal{P}(X) &\longrightarrow \mathcal{P}(X) \\
\pi &\longmapsto \Psi(\pi) :\ \ C_b(X) \longrightarrow \mathbb{R} \\
&\qquad\qquad\qquad\quad q \longmapsto \Psi(\pi)(q) = \frac{\int_X q(\theta) l(x|\theta)\pi(d\theta)}{\int_X l(x|\theta)\pi(d\theta)}
\end{aligned}$$

The functional Ψ has very useful properties:

LEMMA 4.5. *If the likelihood function $l(x|\cdot) : \theta \mapsto l(x|\theta)$ is continuous, then Ψ is continuous for the weak* topology in $\mathcal{P}(\mathcal{X})$.*

Proof. let (π_n) be a sequence of probability measure in $\mathcal{P}(X)$ converging in weak* topology towards some probability measure π. The convergence in weak* topology means that for every $q \in C_b(X)$, $\langle \pi_n | q \rangle \to \langle \pi | q \rangle$. But because $l(x|\cdot)$ is continuous the function $q \times l(x|\cdot)$ is also an element of $C_b(X)$, therefore

$$\int_X q(\theta) l(x|\theta)\, \pi_n(d\theta) = \langle q \times l(x|\cdot) | \pi_n \rangle \longrightarrow \langle q \times l(x|\cdot) | \pi \rangle = \int_X q(\theta) l(x|\theta)\, \pi(d\theta) \ ,$$

This exactly means that $\Psi(\pi_n)$ converges to $\Psi(\pi)$ in the weak* topology. This implies the sequential continuity of Ψ, thus its continuity. Indeed, as X is polish it is separable and metrizable. So that, $\mathcal{P}(X)$ is also metrizable [2, Theorem 15.12]. Hence, it is first-countable [11, Theorem 4.7] which implies it is also sequential. This means that sequential continuity is equivalent to continuity. ◻

The function Ψ can be decomposed as a ratio $\Psi = \Psi_1/\Psi_2$, with

$$\begin{aligned} \Psi_1 : \quad & \mathcal{P}(X) \longrightarrow \mathcal{M}(X) \\ & \pi \longmapsto \Psi_1(\pi) : \quad C_b(X) \longrightarrow \mathbb{R} \\ & \qquad\qquad q \longmapsto \Psi_1(\pi)(q) = \int_X q(\theta) l(x|\theta) \pi(d\theta) \ . \end{aligned}$$

where $\mathcal{M}(X)$ denotes the set of all Borel measures on X, and

$$\begin{aligned} \Psi_2 : \quad & \mathcal{P}(X) \longrightarrow \mathbb{R}_+^* \\ & \pi \longmapsto \int_X l(x|\theta) \pi(d\theta) \end{aligned}$$

The main property of Ψ_1 and Ψ_2 is that they are linear maps. The posterior distribution is therefore the ratio of two linear functions of the prior density. This is particularly interesting due to the following Proposition, which states that the composition of a quasi-convex function with the ratio of two linear mapping is also quasi-convex.

PROPOSITION 4.6. *Let $\mathcal{A}$ be a convex subsets of a topological vector space, and f be a quasi-convex lower semicontinuous functional on $\mathcal{A}$. If $\Psi_1 : \mathcal{A} \mapsto \mathcal{A}$ is a linear mapping and $\Psi_2 : \mathcal{A} \mapsto \mathbb{R}_+^*$ is a linear functional. Then, $f \circ (\Psi_1/\Psi_2) : \mathcal{A} \mapsto \mathbb{R}$ is also a quasi-convex lower semicontinuous functional.*

Proof. Let π, μ be in $\mathcal{A}$. Given $\lambda \in [0,1]$, notice that

$$\begin{aligned} f\left(\frac{\Psi_1(\lambda\pi + (1-\lambda)\mu)}{\Psi_2(\lambda\pi + (1-\lambda)\mu)}\right) &= f\left(\frac{\lambda\Psi_1(\pi) + (1-\lambda)\Psi_1(\mu))}{\lambda\Psi_2(\pi) + (1-\lambda)\Psi_2(\mu)}\right) \ , \\ &= f\left(\beta\frac{\Psi_1(\pi)}{\Psi_2(\pi)} + (1-\beta)\frac{\Psi_1(\mu)}{\Psi_2(\mu)}\right) \ , \end{aligned}$$

with $\beta = \dfrac{\lambda\Psi_2(\pi)}{\lambda\Psi_2(\pi) + (1-\lambda)\Psi_2(\mu)}$ in $[0,1]$. Hence,

$$f\left(\frac{\Psi_1(\lambda\pi + (1-\lambda)\mu)}{\Psi_2(\lambda\pi + (1-\lambda)\mu)}\right) \leq \max\left\{ f\left(\frac{\Psi_1(\pi)}{\Psi_2(\pi)}\right) \ ; \ f\left(\frac{\Psi_1(\mu)}{\Psi_2(\mu)}\right)\right\} \ .$$

This proves the quasi-convexity of $f \circ (\Psi_1/\Psi_2)$. The lower semicontinuity stands because for $\alpha \in \mathbb{R}$, the lower level set

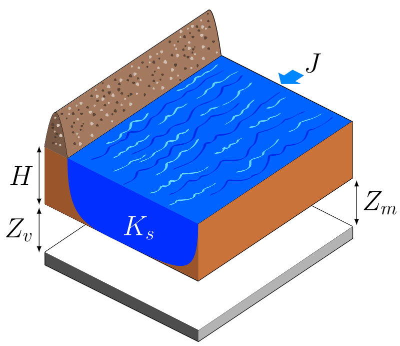
$$\Gamma_\alpha = \left\{ \mu \in \mathcal{A} \mid f\left(\frac{\Psi_1(\mu)}{\Psi_2(\mu)}\right) \leq \alpha \right\} = \left\{ \mu \mid \frac{\Psi_1(\mu)}{\Psi_2(\mu)} \in f^{-1}\left(]-\infty, \alpha]\right) \right\} \ ,$$

is the inverse image of the lower level set α under the continuous map $\mu \mapsto \Psi_1(\mu)/\Psi_2(\mu)$ according to Lemma 4.5. Therefore, Γ_α is closed. □

Proposition 4.6 proves that any lower semicontinuous quasi-convex function presented above is well suited for robust Bayesian analysis. Therefore, the input distribution of a computer model μ can derive from a Bayesian inference such that $\mu = \Psi_1(\pi)/\Psi_2(\pi)$, where π is an imprecise distribution modeled in a reasonable class of prior. For instance, the optimization of the quantile of posterior distributions inferred from priors in a moment class can be reduced to the extreme points of this class.

Moreover, one can easily see that if the functional f is measure affine then $f \circ (\Psi_1/\Psi_2)$ is also the ratio of two measure affine functionals. From Proposition 3.6, it then holds that lower semicontinuity is not necessary to apply the reduction theorem. This means that we can optimize moments or probabilities of the posterior distribution over a class of prior distributions [30].

5. Application to a use case. To illustrate our theoretical optimization results, we address a simplified hydraulic model [26]. This code calculates the water height H of a river subject to a flood event. The height of the river H is calculated through the analytical model given in Equation (5.1).



$$H = \left(\frac{J}{300 K_s \sqrt{\frac{Z_m - Z_v}{5000}}} \right)^{3/5}. \tag{5.1}$$

Fig. 1: Simplified river cross section.

The code takes four inputs whose initial joint distributions are detailed in Table 1. The choice of uniform distributions for Z_v and Z_m comes from expert opinions. The normal distribution for K_s models the uncertainty of the value of the empirical Manning-Strickler coefficient. At last, the choice of a Gumbel distribution is due to the extreme nature of the flooding event. We compute the maximum likelihood parameters of the Gumbel distribution based on a sample of 47 annual maximal flow rates.

Table 1: Initial distribution of the 4 inputs of the hydraulic model.

Variable	Description	Distribution
J	annual maximum flow rate	$Gumbel(\rho = 626, \beta = 190)$
K_s	Manning-Strickler coefficient	$\mathcal{N}(\overline{x} = 30, \sigma = 7.5)$
Z_v	Depth measure of the river downstream	$\mathcal{U}(49, 51)$
Z_m	Depth measure of the river upstream	$\mathcal{U}(54, 55)$

Notice that the modelization of the parameters through the distribution given in Table 1 is questionable. Therefore, as we desire to relax the choice of a specific distribution, we evaluate robust bounds over a measure space taken as a moment class. We display in Table 2 the corresponding moment constraints that the variables must satisfy. These constraints are calculated based on a sample of 47 annual flow rates and expert opinions. The bounds are taken in order to match the most acceptable values of the parameters. Notice that the distribution of K_s belongs to a unimodal moment class because we consider that there is a most significant value for this empirical constant. Using the previous notations, the input distribution $\mu \sim (J, K_s, Z_v, Z_m)$ belongs to $\mathcal{A} = \mathcal{A}_1^* \otimes \mathcal{A}_2^\dagger \otimes \mathcal{A}_3^* \otimes \mathcal{A}_3^*$ with

$$\begin{aligned}
\mathcal{A}_1^* &= \left\{\mu_1 \in \mathcal{P}([160, 3580]) \mid \mathbb{E}_{\mu_1}[X] = 736,\ \mathbb{E}_{\mu_1}[X^2] = 602043\right\}\ , \\
\mathcal{A}_2^\dagger &= \left\{\mu_2 \in \mathcal{H}_{30}([12.55, 47.45]) \mid \mathbb{E}_{\mu_2}[X] = 30,\ \mathbb{E}_{\mu_2}[X^2] = 949\right\}\ , \\
\mathcal{A}_3^* &= \{\mu_3 \in \mathcal{P}([49, 51]) \mid \mathbb{E}_{\mu_3}[X] = 50\}\ , \\
\mathcal{A}_4^* &= \{\mu_4 \in \mathcal{P}([54, 55]) \mid \mathbb{E}_{\mu_4}[X] = 54.5\}\ .
\end{aligned} \tag{5.2}$$

and respective extreme point sets $\Delta_1^*(2), \Delta_2^\dagger(2), \Delta_3^*(1)$, and $\Delta_4^*(1)$.

Table 2: Corresponding moment constraints of the 4 inputs of the hydraulic model.

Variable	Bounds	Mean	Second order moment	Mode
n°1 : J	$[160, 3580]$	736	602043	–
n°2 : K_s	$[12.55, 47.45]$	30	949	$a = 30$
n°3 : Z_v	$[49, 51]$	50	–	–
n°4 : Z_m	$[54, 55]$	54.5	–	–

The following results have been computed in Python. The quantity of interest is optimized globally using a differential evolutionary algorithm [28] over the set of extreme points $\Delta = \Delta_1^*(2) \otimes \Delta_2^\dagger(2) \otimes \Delta_3^*(1) \otimes \Delta_4^*(1)$. This set is parametric as each element is a probability measure written as a product of mixture of Dirac masses and mixture of uniform distributions. In order to explore this space, we rely on a well suited parameterization based on canonical moments [12]. Indeed, it has been proven in [31] that there exists a bijection between the set of discrete probability measures supported on at most $n+1$ points satisfying n moment constraints and a pavement of the form $[0, 1]^{n+1}$. This improves the exploration efficiency of the global optimizer and the overall performance of the optimization.

5.1. Computation of failure probabilities and quantiles. The study addresses the height choice of the protection dike in terms of cost and security. In order to provide safety margins that are optimal with respect to the uncertainty tainting the inputs distributions, we compute the maximal p-quantile over the moment class defined through the constraints in Table 2. The computation of the maximum quantile is equivalent to the computation of the lowest failure probability $\inf_{\mu \in \mathcal{A}} F_\mu$ over the same measure space. Indeed, we have the following duality transformation [31]:

$$\sup_{\mu \in \mathcal{A}} Q_p^L(\mu) = \inf \left\{ h \in \mathbb{R} \mid \inf_{\mu \in \mathcal{A}} F_\mu(h) \geq p \right\}\ .$$

The results are depicted in Figure 2. The quantile of order 0.95 is equal to $2.75m$ for the initial distribution, which gives the appropriate safety margins needed to build a protection dike. However, by considering the uncertainty tainting the input distribution contained in the class $\mathcal{A}$ (the dashed line), the maximum 0.95-quantile over this class is equal to $3.05m$.

5.2. Computation of a Bayesian quantity of interest. We now consider that J is modeled as initially with a Gumbel distribution (see Table 1). Indeed, extreme value theory [10] justifies the choice of a Gumbel distribution for the maximal annual flow rate. However, in a Bayesian setting, the location parameter ρ, and the scale parameter β of the Gumbel distribution are associated with a prior distribution $\pi(\rho, \beta)$. In [26], the prior distribution was taken to be low informative using $\rho \sim \mathcal{G}(1, 500)$ and $1/\beta \sim \mathcal{G}(1, 200)$, where $\mathcal{G}(\alpha, \tau)$ is the Gamma distribution with shape parameter α and scale parameter τ.

This choice of prior is questionable. Instead, we used the previously computed maximum likelihood estimation as a mean constraint. The bounds are chosen as reasonable values.

Table 3: Corresponding moment constraints of the parameters ρ, β of the Gumbel distribution of J.

Variable	Bounds	Mean
ρ	$[550, 700]$	626.14
β	$[150, 250]$	190

This corresponds to two moment classes, ρ belongs to $\widetilde{\mathcal{A}}_1^* = \{\mu \in \mathcal{P}([550, 700]) \mid \mathbb{E}_\mu[X] = 626.14\}$ and β to $\widehat{\mathcal{A}}_1^* = \{\mu \in \mathcal{P}([150, 250]) \mid \mathbb{E}_\mu[X] = 190\}$. Distributions of the other parameters K_s, Z_v, Z_m are constrained to their previous classes in Equation (5.2), that is respectively $\mathcal{A}_2^\dagger, \mathcal{A}_3^*$ and $\mathcal{A}_4^*$. Finally, the distribution $\Theta \sim (\rho, \beta, K_s, Z_v, Z_m)$ belongs to the product space $\mathcal{A}' = \widetilde{\mathcal{A}}_1^* \otimes \widehat{\mathcal{A}}_1^* \otimes \mathcal{A}_2^\dagger \otimes \mathcal{A}_3^* \otimes \mathcal{A}_4^*$.

The Gumbel model and the analytic formulation of the code in Equation 5.1 yields the exact calculation of the probability of failure conditional on $(\rho, \beta, K_s, Z_v, Z_m)$:

$$\mathbb{P}(H \leq h \mid \Theta) = \exp\left(-\exp\left\{\beta\left(\rho - 300K_s\sqrt{\frac{Z_m - Z_v}{5000}}(h - Z_v)^{5/3}\right)\right\}\right) .$$

Therefore, the Bayesian probability of failure corresponds to the integrated cost

$$F_\Theta(h) = \mathbb{P}(H \leq h) = \int \mathbb{P}(H \leq h \mid \Theta)\, \pi(\Theta|D)\, d\Theta \ , \tag{5.3}$$

where $\pi(\Theta|D) \propto l(D|\Theta)\pi(\Theta)$ is the posterior distribution of Θ. The quantity in Equation (5.3) is minimized over the product space $\mathcal{A}'$, the results are depicted in Figure 2. The quantile of order 0.95 is equal to $3.19m$ which is slightly higher than for the maximal quantile over the moment class $\mathcal{A}$.

5.3. Computation of Sobol index. In this section, we illustrate the impact of the uncertainty tainting the input distribution on the Sobol indices. We propose different robust computations of the Sobol' indices which lead to different interpretations. Each parameter μ_i belongs to a measure class $\mathcal{A}_i$ presented in Table 2.

The first order indices $(S_i^0)_{1 \leq i \leq 4}$ are classically computed with the nominal input distributions in Table 1. We compute their robust version, corresponding to the

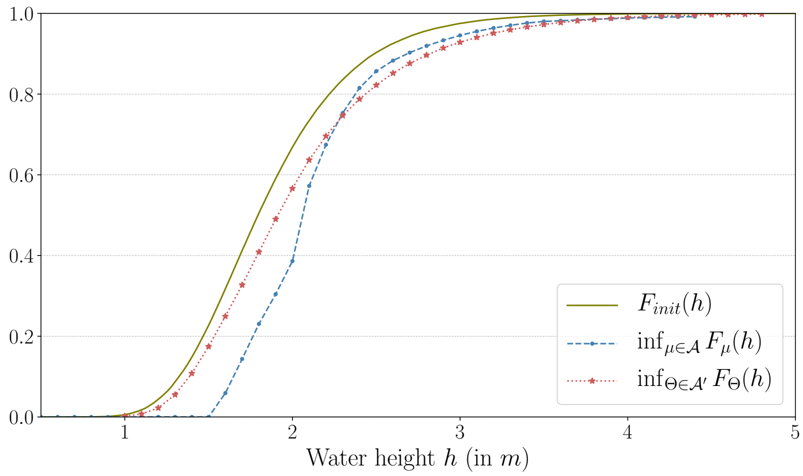


Fig. 2: The solid line represents the CDF of the computer code $h \mapsto \mathbb{P}(H(J, K_s, Z_v, Z_m) \leq h)$ with the initial input distribution depicted in Table 1. The dashed line represents the CDF lowest envelop over the measure set $\mathcal{A}$ from Equation (5.2). The dotted line represents the optimization of the same quantity in a Bayesian framework when J is a Gumbel distribution with prior density on its parameters.

bounds of S_i when μ_i belongs to $\mathcal{A}_i$, that is

$$S_i^+ = \sup_{\mu_i \in \mathcal{A}_i} S_i \quad \text{and} \quad S_i^- = \inf_{\mu_i \in \mathcal{A}_i} S_i \ .$$

Notice that only the ith distribution varies in its moment class when computing bounds on S_i, the other distributions are set to their nominal choice in Table 1. Thus, S_i^+ represents the maximal contribution of the ith input alone onto the output variance, considering the uncertainty of the ith parameter distribution only. Note that $\sum_i S_i^+$ is not necessarily equals to 1. The same interpretation holds for S_i^-. We also define the total indices [18]

$$S_{Ti} = \frac{\mathbb{E}_{\sim i}[\mathrm{Var}_{\mu_i}(Y \mid X_{\sim i})]}{\mathrm{Var}(Y)} \ . \tag{5.4}$$

Bounds on the ith total order index S_{Ti} are computed with respect to the uncertainty of all the input distributions except the ith as follows:

$$S_{Ti}^+ = \sup_{\substack{\mu_j \in \mathcal{A}_j \\ j \neq i}} S_{Ti} \quad \text{and} \quad S_{Ti}^- = \inf_{\substack{\mu_j \in \mathcal{A}_j \\ j \neq i}} S_{Ti} \ .$$

In other words, the bounds obtained are interpreted as the minimal and maximal total order index that the ith input can have, considering the lack of knowledge in all but the ith input distribution. Hence, the lower bound on the ith total effect index corresponds to the minimal variance caused by its interaction between the ith input and the remaining parameters, and the upper bound on S_{Ti} gives the maximal variance caused by its interaction between the ith input and the other parameters.

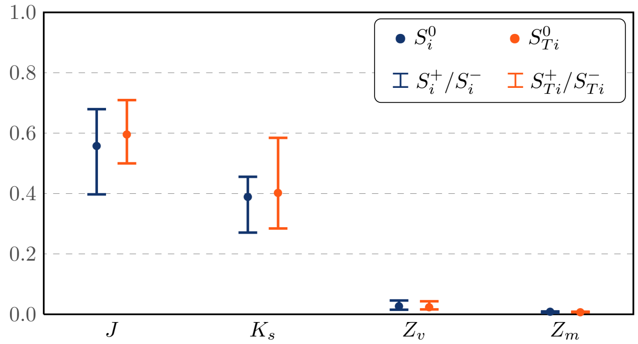


Fig. 3: Different definitions of robustness for the Sobol indices yield different bounds.

The results are displayed in Figure 3. We can see that whatever the input distribution is, the parameters Z_v and Z_m can be considered as weakly influent. The reference Sobol indices S_i^0 and S_{Ti}^0 indicate that J is more influent than K_s. However, the optimization results show that their contribution to the output variance vary when accounting for their distribution uncertainty. The lower bound S_1^- shows that for an imprecise measure μ_1 varying in the moment class $\mathcal{A}_1^*$, the most penalizing distribution corresponds to a first order Sobol' index equals to $S_J^- = 0.39$. On the other hand, the upper bounds S_2^+ shows that there is a distribution in μ_2 in $\mathcal{A}_2^\dagger$ corresponding to a first order Sobol' index $S_2^+ = 0.55$. Therefore, considering the uncertainty affecting the distributions μ_2 and μ_1, the parameter K_s can be in some cases more influent than J. It could be valuable in this situation to refine the information on these parameters in order to reduce the uncertainty tainting their distributions.

6. Reduction theorem proof. In this section, we develop the proof of Theorem 3.2. We make few assumptions on the nature of the optimization space. In this way, the framework stays very general and can be extended to many different spaces, even though measure spaces constitute the main application of this paper. We first develop the reduction theorem for simple topological spaces before extending the result to product spaces. The proofs are quite short and rely only on simple topological arguments. We enlighten the assumption made in this work, in particular we compare our lower semicontinuity assumption with the Bauer maximum principle's upper semicontinuity one.

6.1. Preliminary results. The following two Lemmas are of great importance and gather the main arguments of our demonstration.

Lemma 6.1. *Let $\mathcal{A}$ be a convex subset of a locally convex topological vector space Ω. If any point $x \in \mathcal{A}$ is the barycenter of some probability measure ν supported on $\Delta \subset \mathcal{A}$, then $\mathcal{A} \subset \overline{co}(\Delta)$.*

Proof. Let $K = \overline{\text{co}}(\Delta)$. We suppose that there exists $x_0 \in \mathcal{A} \backslash K$. By applying the Hahn-Banach separation theorem, there exists a continuous linear map $l : \Omega \to \mathbb{R}$, such that $\sup_{x \in K} l(x) < C < l(x_0)$, for some real C. The lower level set

$$Z = \{x \in \mathcal{A} \mid l(x) \leq C\}$$

obviously contains Δ. Let ν_0 be the representative measure of x_0, supported on Δ so that $\nu_0(Z) = 1$. Then,

$$l(x_0) = \int_Z l d\nu_0 \leq C < l(x_0) \ ,$$

leading to a contradiction. ∎

The next Lemma expresses the supremum of a quasi-convex function on the closed convex hull of some subset [3].

LEMMA 6.2. *Let $\mathcal{A}$ be a convex set of a locally convex topological vector space. And let $f : \mathcal{A} \to \mathbb{R}$ be a quasi-convex lower semicontinuous function. Let Y be a subset of $\mathcal{A}$ and $\overline{\text{co}}(Y)$ its closed convex hull. Then*

$$\sup_{\overline{\text{co}}(Y)} f(x) = \sup_Y f(x) \ ,$$

Proof. If $\sup_Y f(x) = \infty$, there is nothing to prove. So, we may assume that $a := \sup_Y f(x)$ is finite. Let $Z_a = \{x \in \mathcal{A} \mid f(x) \leq a\}$.
Obviously, we have $Y \subset Z_a$. But Z_a is convex as f is quasi-convex. Further, it is closed as f is lower semicontinuous. Therefore, we have $\overline{\text{co}}(Y) \subset Z_a$ because of the minimal property of the closed convex hull. Hence,

$$\sup_{\overline{\text{co}}(Y)} f(x) \leq \sup_{Z_a} f(x) \leq a = \sup_Y f(x) \ ,$$

The converse is obvious. ∎

It is remarkable that we assume the lower semicontinuity of the function to maximize. In contrast, the upper semicontinuity required in the Bauer maximum principle is a more standard assumption for function maximization. The proof of Lemma 6.2 clarifies how our assumption acts. Indeed, the lower semicontinuity is used to enforce the closure of the set $Z_a = \{x \in \mathcal{A} \mid f(x) \leq \sup_Y f(x)\}$. This argument differs from Choquet's demonstration of the Bauer maximum principle [8, p.102]. In the latter, the author studies the closure of the set $\{x \in \mathcal{A} \mid f(x) = \sup_Y f(x)\}$ for an upper semicontinuous function f on a compact space. Doing so, the assumptions of compactness and upper semicontinuity in the Bauer maximum principle are used in order to show that the optimum of the function f is reached. This is not needed here.

From Lemma 6.1 and 6.2, we establish the next Theorem. It is analogous to the Bauer maximum principle, where the assumption of compactness is replaced by that of integral representation. The integral representation is always satisfied on compact sets, thanks to the Choquet representation theorem [8, p.153]. So that the next theorem is analogous to the Bauer maximum principle under a compactness assumption. Hence, our integral representation assumption is, in a way, more general.

THEOREM 6.3. *Let $\mathcal{A}$ be a convex subset of a locally convex topological vector space $\mathcal{X}$. We assume that any point $x \in \mathcal{A}$ is the barycenter of a probability measure ν supported on $\Delta \subset \mathcal{A}$. Let $f : \mathcal{A} \to \mathbb{R}$ be a quasi-convex lower semicontinuous function. Then*

$$\sup_{x \in \mathcal{A}} f(x) = \sup_{x \in \Delta} f(x) \ .$$

Proof. From Lemma 6.1

$$\mathcal{A} \subset \overline{\text{co}}(\Delta) \ .$$

Then applying Lemma 6.2 on the lower semicontinuous quasi-convex function f, we obtain

$$\sup_{\mathcal{A}} f(x) \leq \sup_{\overline{\text{co}}(\Delta)} f(x) = \sup_{\Delta} f(x) \ .$$

The converse inequality holds obviously. ∎

We may relate our result to the one of Vesely [32], who proves that Jensen's integral inequality remains true for a lower semicontinuous convex function on a convex set in a locally convex topological vector space. Indeed, given a probability measure ν supported on Δ with barycenter x_ν and a convex *lsc* function f, we get thanks to Jensen's inequality:

$$f(x_\nu) = f\left(\int_\Delta x d\nu\right) \leq \int_\Delta f(x) d\nu \leq \sup_\Delta f(x) \ .$$

Therefore, Theorem 6.3 can be seen as some extremal version of Jensen's integral inequality. Moreover, it extends to quasi-convex functions which are more general than convex functions.

6.2. Extension to Product spaces. The following Theorem shows that the optimum of a quasi-convex lower semicontinuous function on a product space may be computed on the generator of the optimization set.

THEOREM 6.4. *Let $\mathcal{A}_i$ be a convex subset of a locally convex topological vector space Ω_i, $1 \leq i \leq n$. We assume that any point $x_i \in \mathcal{A}_i$ is the barycenter of some probability measure ν_i supported on $\Delta_i \subset \mathcal{A}_i$. We equip $\mathcal{A} := \prod_{i=1}^n \mathcal{A}_i$ with the product topology, we suppose further that $f : \mathcal{A} \to \mathbb{R}$ is marginally quasi-convex lower semicontinuous. Then*

$$\sup_{\substack{x_i \in \mathcal{A}_i \\ 1 \leq i \leq n}} f((x_1, \dots, x_n)) = \sup_{\substack{x_i \in \Delta_i \\ 1 \leq i \leq n}} f((x_1, \dots, x_n)) \ .$$

Proof. For the sake of simplicity, we only develop the case where $i = 2$. The general case follows the same lines. Then,

$$\sup_{\substack{x_1 \in \mathcal{A}_1 \\ x_2 \in \mathcal{A}_2}} f((x_1, x_2)) = \sup_{x_1 \in \mathcal{A}_1} \sup_{x_2 \in \mathcal{A}_2} f((x_1, x_2)) \ .$$

Now, the map $x_2 \mapsto f((x_1, x_2))$ is a quasi-convex lower semicontinuous function, by applying Theorem 6.4, it follows that

$$\sup_{x_2 \in \mathcal{A}_2} f((x_1, x_2)) = \sup_{x_2 \in \Delta_2} f((x_1, x_2)), \text{ for every } x_1 \in \mathcal{A}_1 \ .$$

Therefore,

$$\begin{aligned}
\sup_{x_1 \in \mathcal{A}_1} \sup_{x_2 \in \mathcal{A}_2} f((x_1, x_2)) &= \sup_{x_1 \in \mathcal{A}_1} \sup_{x_2 \in \Delta_2} f((x_1, x_2)), \\
&= \sup_{x_2 \in \Delta_2} \sup_{x_1 \in \mathcal{A}_1} f((x_1, x_2)), \\
&= \sup_{x_2 \in \Delta_2} \sup_{x_1 \in \Delta_1} f((x_1, x_2)),
\end{aligned}$$

by applying the same reasoning to $x_1 \mapsto f((x_1, x_2))$. ∎

Remark 6.5. If a function is upper semicontinuous and quasi-concave, the same reduction applies to the minimum.

6.3. Proof of the main result. Theorem 3.2 is slightly more complicated than Theorem 6.4, so that we detail its proof. Let $\mathcal{X}, \mathcal{A}$ and Δ be as defined in Subsection 3.1.

Proof of Theorem 3.2. We recall that $\mathcal{A} := \prod_{i=1}^{d} \mathcal{A}_i$ is a product of measure spaces, where $\mathcal{A}_i$ is either a moment space or a unimodal moment space. Therefore, each measure $\mu_i \in \mathcal{A}_i$ satisfies N_i moment constraints. More precisely, for measurable functions $\varphi_i^{(j)} : X_i \to \mathbb{R}$, we have $\mathbb{E}_{\mu_i}[\varphi_i^{(j)}] \le 0$ for $1 \le j \le N_i$ and $1 \le i \le d$.

Moreover, in Theorem 3.2, we also enforce constraints on the product measure $\mu = \mu_1 \otimes \cdots \otimes \mu_d \in \mathcal{A}$, such that, for measurable functions $\varphi^{(j)} : \mathcal{X} \to \mathbb{R}$, $1 \le j \le N$, we have $\mathbb{E}_\mu[\varphi^{(j)}] \le 0$.

Let f be a marginally quasi-convex lower semicontinuous function. Then,

$$\sup_{\substack{\mu_1,\ldots,\mu_d \in \otimes \mathcal{A}_i \\ \mathbb{E}_{\mu_1,\ldots,\mu_d}[\varphi^{(j)}] \le 0 \\ 1 \le j \le N}} f(\mu) = \sup_{\mu_1 \in \mathcal{A}_1} \ldots \sup_{\mu_{d-1} \in \mathcal{A}_{d-1}} \sup_{\substack{\mu_d \in \mathcal{A}_d \\ \mathbb{E}_{\mu_1,\ldots,\mu_d}[\varphi^{(j)}] \le 0 \\ 1 \le j \le N}} f(\mu_1, \ldots, \mu_d) \ .$$

Now, for fixed $\mu_1, \ldots \mu_{d-1}$, we have that

$$\mathbb{E}_{\mu_1,\ldots,\mu_d}[\varphi^{(j)}] = \mathbb{E}_{\mu_d}\left[\mathbb{E}_{\mu_1,\ldots,\mu_{d-1}}[\varphi_i^{(j)}]\right] \le 0 \ ,$$

for $1 \le j \le N$, which are moment constraints on the measure μ_d. This means that there are in total $(N_d + N)$ moment constraints enforced on μ_d. Therefore, μ_d has an integral representation supported on the set of convex combination of $N_d + N + 1$ extreme points (which are either Dirac masses, or uniform distributions). Hence, for fixed $\mu_1, \ldots, \mu_{d-1}$, and because the function $\mu_d \longmapsto f(\mu_1, \ldots, \mu_d)$ is quasi-convex and lower semicontinuous. We have from Theorem 6.3

$$\sup_{\substack{\mu_d \in \mathcal{A}_d \\ \mathbb{E}_{\mu_1,\ldots,\mu_d}[\varphi^{(j)}] \le 0 \\ 1 \le j \le N}} f(\mu_1, \ldots, \mu_d) = \sup_{\substack{\mu_d \in \Delta_{N_d+N} \\ \mathbb{E}_{\mu_1,\ldots,\mu_d}[\varphi^{(j)}] \le 0 \\ 1 \le j \le N}} f(\mu_1, \ldots, \mu_d).$$

So that,

$$\sup_{\substack{\mu_1,\ldots,\mu_d \in \otimes \mathcal{A}_i \\ \mathbb{E}_{\mu_1,\ldots,\mu_d}[\varphi^{(j)}] \le 0 \\ 1 \le j \le N}} f(\mu) = \sup_{\substack{\mu_1,\ldots,\mu_d \in \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_{d-1} \otimes \Delta_d \\ \mathbb{E}_{\mu_1,\ldots,\mu_d}[\varphi^{(j)}] \le 0 \\ 1 \le j \le N}} f(\mu) \ .$$

Consequently, the last component of μ can be replaced by some element of Δ_{N_i+N}. By repeating this argument to the other components, the result follows. ∎

7. Conclusion. We have studied the optimization of a quasi-convex lower semicontinuous function over a set of product of measure spaces $\mathcal{A} = \prod_{i=1}^{d} \mathcal{A}_i$. Specific product measures sets are studied: the product of moment classes or unimodal moment classes. In those classes, we have an integral representation on the extreme points Δ_i, that are either finite mixture of Dirac masses or finite mixture of uniform distributions. This integral representation can be seen as a non-compact form of the Krein-Milmann theorem. We have shown that the optimization of a quasi-convex lower semicontinuous function on the product space $\mathcal{A}$ is reduced to the d-fold product of finite convex combination of extreme points of $\mathcal{A}_i$.

This powerful Theorem provides numerous industrial applications. We develop for example the optimization of the quantile of the output of a computer code whose input distributions belong to measure spaces [31]. We also highlight through several illustrated applications how our framework generalizes both Optimal Uncertainty Quantification [25] and robust Bayesian analysis [17, 4].

Although, we have an explicit representation of the extreme points, the optimization is not obvious because of the high number of *generalized* moment constraints enforced. In [31], the authors propose an original parameterization of the problem in the presence of *classical* moment constraints, allowing fast computation of the quantities of interest presented in Section 4.

The product of measure spaces reflects the mutual independence of the model parameters. In case of a dependence structure, the Lasserre hierarchy of relaxation in semidefinite programming [23] provides an alternative solution for practical optimization.

Appendix A. Proof of Theorem 2.2.

Proof. The proof is quite technical and as it is not the main topic of our paper, details are kept to the bare minimum. We mainly gather different results to prove our point, so that the interested reader might refer to it. Letting X be an interval of $\mathbb{R}$, $\mathcal{H}_a(X)$ is the set of all probability measures on x which are unimodal at $a \in X$. From [5, p.19] we now that $\mathcal{H}_a(X)$ is a simplex, meaning that every probability measure in $\mathcal{H}_a$ is the barycenter of a unique probability measure supported on $\mathcal{U}_a(X)$. Choquet [8, p.160] defines another type of simplex, named Choquet simplex. A convex subset K of a locally convex topological vector space is called a Choquet simplex if and only if the cone $\tilde{K} = \{(\lambda x, \lambda) : x \in K, \lambda > 0\}$ is a lattice cone in its own order (that is, the vector space $\operatorname{span}(\tilde{K})$ is a lattice when its positive cone is taken to be $\tilde{K}$). The important point is that these two definitions are connected, and it holds from [35, p.47] that each simplex is a Choquet simplex. Moreover, in finite dimensional compact sets these two definitions are equivalent. Therefore, $\mathcal{H}_a(X)$ is also a Choquet simplex. We now define

$$\begin{aligned} K &= \{\mu \in \mathcal{H}_a(X) : \varphi_1, \dots, \varphi_n \text{ are } \mu\text{-integrable}\}\ , \\ F(\mu) &= \left(\int \varphi_1 \, d\mu, \dots, \int \varphi_n \, d\mu\right)\ , \\ W &= F[K] \cap \prod_{i=1}^{n}]-\infty, 0] \text{ or } W = F[K] \cap \{(0, \dots, 0)\}\ . \end{aligned}$$

It is already known that the extreme points set of $\mathcal{H}_a(X)$ is precisely $\mathcal{U}_a(X)$ as shown in [5, p.19]. Indeed, a classical result due to Khintchine [21] states that uniform distributions constitute the *elementary units* of the set of all unimodal probability measures. However, we wish to known how the extreme points of $\mathcal{H}_a(X)$ also characterize the extreme points of a convex subset of $\mathcal{H}_a(X)$ defined as $\mathcal{A}^\dagger = F^{-1}[W] = \{\mu \in \mathcal{H}_a(X) \mid \mathbb{E}_\mu[\varphi_i] \le 0,\ 1 \le i \le n\}$. Consequently, we wish to apply [36, Proposition 2.1] which states that the set $\mathcal{A}^\dagger$ satisfies

$$\operatorname{ex}\{\mathcal{A}^\dagger\} \subset \Delta^\dagger(n) = \left\{ \mu \in \mathcal{A}^\dagger \mid \mu = \sum_{i=1}^{n+1} \omega_i \mathfrak{u}_i, \omega_i \ge 0, \mathfrak{u}_i \in \mathcal{U}_a(X) \right\}\ . \tag{A.1}$$

However, in order to apply [36, Proposition 2.1] it remains to check that K is linearly compact, meaning each of its line meets K in a compact interval. By the main theorem in [20], it is sufficient to show that $\mathbb{R}_+ \cdot K$ is a lattice cone in its own order. Indeed, condition (2^0) in the main theorem is an equivalent formulation of linear compactness as shown in the same reference on p.369. Of course, the cone $\mathbb{R}_+ \cdot \mathcal{H}_a(X)$ is a lattice cone in its own order because it is a Choquet simplex. Now, choose $\mu \in \mathbb{R}_+ \cdot K$ and $\nu \in \mathbb{R}_+ \cdot \mathcal{H}_a(X)$ such that $(\mu - \nu) \in \mathbb{R}_+ \cdot K$, then $\nu \in \mathbb{R}_+ \cdot K$ since

$$\int |\varphi_i| \, d\nu \leq \int |\varphi_i| \, d\mu \quad \text{for every } i = 1, \ldots, n.$$

Hence, $\mathbb{R}_+ \cdot K$ is a hereditary subcone of $\mathbb{R}_+ \cdot \mathcal{H}_a(X)$ and consequently a lattice cone in its own order. This proves that K is linearly compact and that [36, Proposition 2.1] applies. As stated, it follows that the set $\mathcal{A}^\dagger = F^{-1}[W] = \{\mu \in \mathcal{H}_a(X) \mid \mathbb{E}_\mu[\varphi_i] \leq 0,\ 1 \leq i \leq n\}$ satisfies Equation (A.1).

Now that the extreme points of $\mathcal{A}^\dagger$ are classified and observing that this set is closed with respect to the weak topology; Corollary 1 in [34] concludes that every measure in $\mathcal{A}^\dagger$ has an integral representation supported on $\Delta^\dagger(n)$. ◻

Acknowledgement. We are grateful to the two reviewers who greatly help to improve the manuscript. Support from the ANR-3IA Artificial and Natural Intelligence Toulouse Institute is gratefully acknowledge.

REFERENCES

[1] E. M. Alfsen, *Compact Convex Sets and Boundary Integrals*, Springer Science & Business Media, Dec. 2012.

[2] C. D. Aliprantis and K. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer-Verlag, Berlin Heidelberg, 3 ed., 2006, //www.springer.com/us/book/9783540295860.

[3] B. Bereanu, *Quasi-convexity, strictly quasi-convexity and pseudo-convexity of composite objective functions*, ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique, 6 (1972), pp. 15–26, http://www.numdam.org/item/M2AN_1972__6_1_15_0/.

[4] J. O. Berger, *Robust Bayesian analysis: sensitivity to the prior*, Journal of Statistical Planning and Inference, 25 (1990), pp. 303–328, https://doi.org/10.1016/0378-3758(90)90079-A.

[5] E. M. J. Bertin, I. Cuculescu, and R. Theodorescu, *Unimodality of Probability Measures*, Mathematics and Its Applications, Springer Netherlands, Southend, 1997, //www.springer.com/la/book/9780792343189.

[6] E. M. J. Bertin and R. Theodorescu, *Hinin spaces and unimodal probability measures*, Mathematische Annalen, 266 (1984), pp. 357–367, https://doi.org/10.1007/BF01475585.

[7] H. Bourls, *3 - topological vector spaces*, in Fundamentals of Advanced Mathematics 2, H. Bourls, ed., Elsevier, 2018, pp. 117 – 196, https://doi.org/https://doi.org/10.1016/B978-1-78548-249-6.50003-1, http://www.sciencedirect.com/science/article/pii/B9781785482496500031.

[8] G. Choquet, J. Marsden, T. Lance, and S. S. Gelbart, *Lectures on analysis. Représentation theory Vol. 2.*, W.A. Benjamin, New York; Amsterdam, 1969. OCLC: 25191467.

[9] A. Clausing, *On Quotients of Lp-Means*, Birkhäuser Basel, Basel, 1983, pp. 43–68, https://doi.org/10.1007/978-3-0348-6290-5_4.

[10] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics, Springer-Verlag, London, 2001, https://www.springer.com/gp/book/9781852334598.

[11] F. H. Croom, *Principles of Topology*, Courier Dover Publications, Feb. 2016.

[12] H. Dette and W. J. Studden, *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*, Wiley-Blackwell, New York, Sept. 1997.

[13] S. Drapeau and M. Kupper, *Risk Preferences and Their Robust Representation*, Mathematics of Operations Research, 38 (2012), pp. 28–62, https://doi.org/10.1287/moor.1120.0560.

[14] *Handbook of Uncertainty Quantification*, Springer International Publishing, 2017, https://www.springer.com/us/book/9783319123844.

[15] H. J. Greenberg and W. P. Pierskalla, *A Review of Quasi-Convex Functions*, Operations Research, 19 (1971), pp. 1553–1570, https://www.jstor.org/stable/169178.

[16] A. A. Gushchin and D. A. Borzykh, *Integrated quantile functions: properties and applications*, Modern Stochastics: Theory and Applications, 4 (2017), pp. 285–314, https://doi.org/10.15559/17-VMSTA88. arXiv: 1801.00977.

[17] *Robust Bayesian Analysis*, Lecture Notes in Statistics, Springer-Verlag, New York, 2000, //www.springer.com/la/book/9780387988665.

[18] B. Iooss and P. Lematre, *A Review on Global Sensitivity Analysis Methods*, 2015, pp. 101–122, https://doi.org/10.1007/978-1-4899-7547-8_5.

[19] B. Iooss and A. Marrel, *Advanced methodology for uncertainty propagation in computer experiments with large number of inputs*, Nuclear Technology, (2019), pp. 1–19, https://doi.org/10.1080/00295450.2019.1573617.

[20] D. G. Kendall, *Simplexes and Vector Lattices*, Journal of the London Mathematical Society, s1-37 (1962), pp. 365–371, https://doi.org/10.1112/jlms/s1-37.1.365.

[21] A. Y. Khintchine, *On unimodal distributions*, Izvestiya Nauchno-Issledovatelskogo Instituta Matematiki i Mekhaniki, 2 (1938), pp. 1–7.

[22] M. Kružík, *Bauer's maximum principle and hulls of sets*, Calculus of Variations and Partial Differential Equations, 11 (2000), pp. 321–332, https://doi.org/10.1007/s005260000047, https://doi.org/10.1007/s005260000047.

[23] J.-B. Lasserre, *Moments, positive polynomials and their applications*, no. v. 1 in Imperial College Press optimization series, Imperial College Press ; Distributed by World Scientific Publishing Co, London : Signapore ; Hackensack, NJ, 2010. OCLC: ocn503631126.

[24] J. Lukeš, J. Malý, I. Netuka, and J. Spurný, *Integral Representation Theory: Applications to Convexity, Banach Spaces and Potential Theory*, De Gruyter, Dec. 2009, https://www.degruyter.com/view/title/32361. Publication Title: Integral Representation Theory.

[25] H. Owhadi, C. Scovel, T. J. Sullivan, M. McKerns, and M. Ortiz, *Optimal Uncertainty Quantification*, SIAM Review, 55 (2013), pp. 271–345, https://doi.org/10.1137/10080782X. arXiv: 1009.0679.

[26] A. Pasanisi, M. Keller, and E. Parent, *Estimation of a quantity of interest in uncertainty analysis: Some help from Bayesian decision theory*, Reliability Engineering & System Safety, 100 (2012), pp. 93–101, https://doi.org/10.1016/j.ress.2012.01.001.

[27] J.-P. Penot and M. Volle, *On quasi-convex duality*, Mathematics of Operations Research, 15 (1990), pp. 597–625, http://www.jstor.org/stable/3689766.

[28] K. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, Natural Computing Series, Springer-Verlag, Berlin Heidelberg, 2005, //www.springer.com/la/book/9783540209508.

[29] F. Ruggeri, D. Ríos Insua, and J. Martín, *Robust Bayesian Analysis*, in Handbook of Statistics, D. K. Dey and C. R. Rao, eds., vol. 25 of Bayesian Thinking, Elsevier, Jan. 2005, pp. 623–667, https://doi.org/10.1016/S0169-7161(05)25021-6.

[30] S. Sivaganesan and J. O. Berger, *Ranges of Posterior Measures for Priors with Unimodal Contaminations*, The Annals of Statistics, 17 (1989), pp. 868–889, https://doi.org/10.1214/aos/1176347148.

[31] J. Stenger, F. Gamboa, M. Kerller, and B. Iooss, *Optimal Uncertainty Quantification of a risk measurement from a thermal-hydraulic code using Canonical Moments*, International Journal for Uncertainty Quantification, 10 (2020), pp. 35–53, https://doi.org/10.1615/Int.J.UncertaintyQuantification.2020030800, http://www.dl.begellhouse.com/journals/52034eb04b657aea,forthcoming,30800.html.

[32] L. Vesely, *Jensen's integral inequality in locally convex spaces*, Annals of the Academy of Romanian Scientists: Series on Mathematics and its Applications, 9 (2017), pp. 136–144.

[33] G. Wallis, *Uncertainties and probabilities in nuclear reactor regulation*, Nuclear Engineering and Design, 237 (2004), pp. 1586–1592.

[34] H. v. Weizscker and G. Winkler, *Integral Representation in the Set of Solutions of a Generalized Moment Problem.*, Mathematische Annalen, 246 (1979), pp. 23–32, https://eudml.org/doc/163325.

[35] G. Winkler, *Choquet order and simplices : with applications in probabilistic models / gerhard winkler*, SERBIULA (sistema Librum 2.0), 1145 (1985), https://doi.org/10.1007/BFb0075051.

[36] G. Winkler, *Extreme Points of Moment Sets*, Math. Oper. Res., 13 (1988), pp. 581–587, https://doi.org/10.1287/moor.13.4.581.