Canonical Moments for Optimal Uncertainty Quantification on a Variety

Jérôme Stenger[1,2(✉)], Fabrice Gamboa[1], Merlin Keller[2], and Bertrand Iooss[1,2]

[1] Institut Mathématiques de Toulouse, 118 route de Narbonne, Toulouse, France
fabrice.gamboa@math.univ-toulouse.fr
[2] EDF R&D, 6 quai Watier, Chatou, France
{jerome.stenger,merlin.keller,bertrand.iooss}@edf.fr

Abstract. The purpose of this work is to optimize an affine functional over positive measures. More precisely, we deal with a probability of failure (P.O.F). The optimization is realized over a set of distributions satisfying moment constraints, called moment set. The optimum is to be found on an extreme point of this moment set. Winkler's classification of those extreme points states they are finite discrete measures. The set of the support points of all discrete measures in the moment set is a manifold over which the P.O.F is optimized. We characterize the manifold's structure by proving it is an algebraic variety. It is the zero locus of polynomials defined thanks to the canonical moments. This reduces a highly constrained optimization over the moment set onto a constraint free manifold.

Keywords: Canonical moments · Optimal uncertainty quantification · Robustness

1 Introduction

1.1 Probability of Failure Inference

Computer codes are increasingly used to measure safety margins, especially in nuclear accident management analysis. In this context, it is essential to evaluate the accuracy of the numerical model results, whose uncertainties come mainly from the lack of knowledge of the underlying physic and the model input parameters. Methods were developed in safety analyses to quantify those uncertainties [6]. Their common principle relies mainly on a probabilistic modeling of the model input uncertainties, on Monte Carlo sampling for running the computer code on sets of input, and on the application of statistical tools to infer probabilities of failure (P.O.F) of the scalar output variables of interest [13].

This takes place in a more general setting, known as Uncertainty Quantification (UQ) methods [10]. Quantitative assessment of the uncertainties tainting

The authors are grateful to the anonymous reviewers for their helpful comments.

F. Nielsen and F. Barbaresco (Eds.): GSI 2019, LNCS 11712, pp. 571–578, 2019.
https://doi.org/10.1007/978-3-030-26980-7_59

the results of computer simulations is a major topic of interest in both industrial and scientific communities.

P.O.F inference is tainted by the uncertainty of the input modeling. More specifically, the inputs probability densities are usually chosen in parametric families (uniform, normal, log-normal, etc.), and their parameters are estimated using available datas and/or the opinion of an expert. However, they may differ from the reality. This uncertainty on the input probability densities is propagated to the P.O.F. As a consequence, different choices of distributions will lead to different P.O.F values, thus different safety margins.

1.2 Optimal Uncertainty Quantification

In this work, we propose to gain robustness on the quantification of this measure of risk. We aim to account for the uncertainty on the input distributions by evaluating the minimal P.O.F over a class of probability measures $\mathcal{A}$. In this optimization problem, the set $\mathcal{A}$ must be large enough to effectively represent our uncertainty on the inputs, but not too large in order to keep the estimation of the quantile representative of the physical phenomena. For example, the minimal P.O.F over the very large class $\mathcal{A} = \{\text{all distributions}\}$, proposed in [5], will certainly be too conservative to remain physically meaningful. Several articles which discuss possible choices of classes of distributions can be found in the literature of Bayesian robustness (see [11]). Deroberts et al. [2] consider a class of measures specified by a type of upper and lower envelope on their density. Sivaganesan et al. [12] study the class of unimodal distributions. In more recent work, Owhadi et al. [9] propose to optimize the measure of risk over a class of distributions specified by constraints on their *generalized* moments. They call their work Optimal Uncertainty Quantification (OUQ). However, in practical engineering cases, the available information on an input distribution is often reduced to the knowledge of its mean and/or variance. This is why in this paper, we are interested in a specific case of the framework introduced in [9]. We consider the class of measures known by some of their *classical* moments, we refer to it as the moment class:

$$\mathcal{A} = \Big\{ \mu = \otimes\mu_i \in \bigotimes_{i=1}^{d} \mathcal{M}_1([l_i, u_i]) \mid \mathbb{E}_{\mu_i}[x^j] = c_j^{(i)}, \qquad (1)$$
$$c_j^{(i)} \in \mathbb{R},\ \text{for } 1 \leq j \leq N_i \text{ and } 1 \leq i \leq d \Big\},$$

where $\mathcal{M}_1([l_i, u_i])$ denotes the set of scalar probability measures on the interval $[l_i, u_i]$. The tensorial product of measure set traduces the mutual independence of the d-components of the input vector μ.

1.3 Reduction Theorem

The solution of our optimization problem is numerically computed thanks to the OUQ reduction theorem [9,14]. This theorem states that the measure corresponding to the minimal P.O.F is located on the extreme points of the distribution set. In the context of the moment class, the extreme distributions are located on the d-fold product of finite convex combinations of Dirac masses:

$$\mathcal{A}_{\Delta} = \left\{ \mu \in \mathcal{A} \mid \mu_i = \sum_{k=1}^{N_i+1} w_k^{(i)} \delta_{x_k^{(i)}} \quad \text{for} \quad 1 \leq i \leq d \right\}, \tag{2}$$

To be more specific it holds that when n pieces of information are available on the moments of a scalar measure μ, it is enough to pretend that the measure is supported on at most $n+1$ points. This powerful theorem gives the basis for practical optimization of our optimal quantity of interest. In this matter, Semi-Definite-Programming [4] has been already explored in [1] and [7], but the deterministic solver used rapidly reaches its limitation as the dimension of the problem increases. One can also find in the literature a Python toolbox developed by McKerns [8] called Mystic framework that fully integrates the OUQ framework. However, it was built as a generic tool for generalized moment problems and the enforcement of the moment constraints is not optimal.

By restricting the work to our moment class, we propose an original and practical approach based on the theory of canonical moments [3]. Canonical moments of a measure can be seen as the relative position of its moment sequence in the moment space. They are inherent to the measure and therefore present many interesting properties. Our main contribution is in the proof that the optimization set $\mathcal{A}_{\Delta}$ is an algebraic manifold, more specifically it is the zero locus of polynomials whose coefficients are function of canonical moments. This geometric approach replaces the optimization on the constrained space in Eq. (2) into a constraint free optimization.

This paper proceeds as follows. In Sect. 2 we develop the reduction theorem and the parameterization of the optimization space, we present the manifold over which the optimization takes place. Section 3 is dedicated to the canonical moments and the construction of the polynomials of interest. We show that the zero locus of those polynomials constitute the optimization space. Section 4 gives some conclusions and perspectives.

2 Problem Reduction

2.1 OUQ Theorem

In this work, we consider a P.O.F on the output of a computer code $G : \mathbb{R}^d \to \mathbb{R}$, seen as a black box function. In order to gain robustness on our safety margin choice, our goal is to find the minimal P.O.F over the moment set $\mathcal{A}$ described in Eq. (1). For a given threshold h, it reads:

$$\inf_{\mu \in \mathcal{A}} P_{\mu}(G(X) \leq h) \tag{3}$$

The OUQ reduction theorem applies (Theorem 1). It states that the optimal solution of the optimization problem (3) is a product of discrete measures. A general form of the theorem reads as follows:

Theorem 1 (OUQ reduction [9]**).** *Suppose that $\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_d$ is a product of Radon spaces. Let*

$$\mathcal{A} := \left\{ (G,\mu) \left| \begin{array}{l} G : \mathcal{X} \to \mathcal{Y}, \text{ is a real valued measurable function,} \\ \mu = \mu_1 \otimes \cdots \otimes \mu_d \in \bigotimes_{i=1}^{d} \mathcal{M}_1(\mathcal{X}_i), \\ \text{for some integers } N_0, \ldots, N_d, \text{ and measurable functions} \\ \varphi_l : \mathcal{X} \to \mathbb{R} \text{ and } \varphi_j^{(i)} : \mathcal{X}_i \to \mathbb{R}, \\ \qquad \bullet\ \mathbb{E}_\mu[\varphi_l] \leq 0 \text{ for } l = 1, \ldots, N_0, \\ \qquad \bullet\ \mathbb{E}_{\mu_i}[\varphi_j^{(i)}] \leq 0 \text{ for } j = 1, \ldots, N_i \text{ and } i = 1, \ldots, d \end{array} \right. \right\}$$

Let $\Delta_n(\mathcal{X})$ be the set of all discrete measures supported on at most $n+1$ points of $\mathcal{X}$, and

$$\mathcal{A}_\Delta := \left\{ (G,\mu) \in \mathcal{A} \mid \mu_i \in \Delta_{N_0+N_i}(\mathcal{X}_i) \right\}.$$

Let q be a measurable real function on $\mathcal{X} \times \mathcal{Y}$. Then

$$\sup_{(G,\mu)\in\mathcal{A}} \mathbb{E}_\mu[q(X, G(X))] = \sup_{(G,\mu)\in\mathcal{A}_\Delta} \mathbb{E}_\mu[q(X, G(X))].$$

This theorem derives from the work of Winkler [14], who has shown that the extreme measures of moment class $\{\mu \in \mathcal{M}_1(\mathcal{X}) \mid \mathbb{E}_\mu[\varphi_1] \leq 0, \ldots, \mathbb{E}_\mu[\varphi_n] \leq 0\}$ are the discrete measures that are supported on at most $n+1$ points. The strength of Theorem 1 is that it extends the result to a tensorial product of moment sets. The proof relies on a recursive argument using Winkler's classification on every set $\mathcal{X}_i$. A remarkable fact is that, as long as the quantity to be optimized is an affine function of the underlying measure μ, this theorem remains true whatever the function G and the quantity of interest q are.

Now, by taking $\varphi_j^{(i)}(x) = x^j$ for $1 \leq i \leq N_i$, we enforced N_i moment constraints to μ_i, as in Eq. (1). Applying Theorem 1 to the function $q(X, G(X)) = -\mathbb{1}_{\{G(X)\leq h\}}$, the P.O.F reaches its optimum on the reduced set $\mathcal{A}_\Delta$, such that for a fixed threshold h we have:

$$\begin{aligned} \inf_{\mu\in\mathcal{A}} F_\mu(h) &= \inf_{\mu\in\mathcal{A}_\Delta} \mathbb{E}_\mu[\mathbb{1}_{\{G(X)\leq h\}}], \\ &= \inf_{\mu\in\mathcal{A}_\Delta} P_\mu(G(X) \leq h), \\ &= \inf_{\mu\in\mathcal{A}_\Delta} \sum_{i_1=1}^{N_1+1} \cdots \sum_{i_d=1}^{N_d+1} \omega_{i_1}^{(1)} \ldots \omega_{i_d}^{(d)}\ \mathbb{1}_{\{G(x_{i_1}^{(1)}, \ldots, x_{i_d}^{(d)})\leq h\}}, \end{aligned} \tag{4}$$

2.2 Parameterization Simplification

The optimization problem in Eq. (4) shows that the weights and the positions of the input distributions provide a natural parameterization for the computation of the P.O.F. However, we now highlight the fact that the knowledge of the support points of a discrete measure (Eq. (5)) fully determines the corresponding weights. Hence, the support points are sufficient to compute the P.O.F (Eq. (4)). Indeed, we recall that in the optimization set $\mathcal{A}_\Delta$ (Eq. (2)), N_i constraints are enforced on the moment of the ith input. The measure μ_i is therefore supported on at most $N_i + 1$ points, which reads:

$$\mu_i = \sum_{i=1}^{N_i} w_j^{(i)} \delta_{x_j^{(i)}} \tag{5}$$

The $N_i + 1$ Vandermonde system holds

$$\begin{cases} \omega_1^{(i)} & + \ldots + \omega_{N_i+1}^{(i)} & = 1 \\ \omega_1^{(i)} x_1^{(i)} & + \ldots + \omega_{N_i+1}^{(i)} x_{N_i+1}^{(i)} & = c_1^{(i)} \\ \vdots & \vdots & \vdots \\ \omega_1^{(i)} {x_1^{(i)}}^{N_i} & + \ldots + \omega_{N_i+1}^{(i)} {x_{N_i+1}^{(i)}}^{N_i} & = c_{N_i}^{(i)} \end{cases} \tag{6}$$

where the N_i last equations derive from the constraints and the first one is the expression of the measure mass equals to one. Because every support points $(x_j^{(i)})_j$ are distinct, when they are set, the corresponding weights are uniquely determined.

The optimization problem in Eq. (4) is therefore parameterized only with the position of the support points of every input, so that the optimization takes place on the following manifold:

$$\begin{aligned} \mathcal{V} &= \prod_{i=1}^{d} \mathcal{V}_i, \\ &= \prod_{i=1}^{d} \left\{ \mathbf{x}_i = \left(x_1^{(i)}, \ldots, x_{N_i+1}^{(i)} \right) \in \mathbb{R}^{N_i+1}, \text{ s.t } \mu_i = \sum_{j=1}^{N_i+1} \omega_j^{(i)} \delta_{x_j^{(i)}} \in \mathcal{A}_\Delta^{(i)} \right\}, \end{aligned} \tag{7}$$

where $\mathcal{A}_\Delta^{(i)}$ is such that $\mathcal{A}_\Delta = \bigotimes_{i=1}^{d} \mathcal{A}_\Delta^{(i)}$, this reads

$$\mathcal{A}_\Delta^{(i)} = \left\{ \mu_i = \sum_{k=1}^{N_i+1} \omega_k^{(i)} \delta_{x_k^{(i)}} \mid \mathbb{E}_{\mu_i}[x^j] = c_j^{(i)}, \text{ for } 1 \leq \mathrm{j} \leq \mathrm{N_i} \right\}. \tag{8}$$

$\mathcal{V}_i$ is simply the set of support points of all measures in $\mathcal{A}_\Delta^{(i)}$ respecting the constraints. Our main contribution in this work is to show that $\mathcal{V}_i$ is an algebraic manifold, meaning it is the zero locus of some well defined polynomials.

3 Optimization Space Seen as a Variety

3.1 Canonical Moments

We define the moment space $M := M(a, b) = \{\mathbf{c}(\mu) \mid \mu \in \mathcal{M}_1([a, b])\}$ where $\mathbf{c}(\mu)$ denotes the sequence of all moments of some measure μ. The nth moment space M_n is defined by projecting M onto its first n coordinates, $M_n = \{\mathbf{c}_n(\mu) = (c_1, \ldots, c_n) \mid \mu \in \mathcal{M}_1([a, b])\}$. We now define the extreme values,

$$c_{n+1}^+ = \max\left\{c \in \mathbb{R} : (c_1, \ldots, c_n, c) \in M_{n+1}\right\},$$
$$c_{n+1}^- = \min\left\{c \in \mathbb{R} : (c_1, \ldots, c_n, c) \in M_{n+1}\right\},$$

which represent the maximum and minimum values of the $(n + 1)$th moment that a measure can have, when its moments up to order n are fixed. The nth canonical moment is then defined recursively as

$$p_n = p_n(\mathbf{c}) = \frac{c_n - c_n^-}{c_n^+ - c_n^-}. \tag{9}$$

Note that the canonical moments are defined up to the degree $N = N(\mathbf{c}) = \min\{n \in \mathbb{N} \mid \mathbf{c}_n \in \partial M_n\}$, and p_N is either 0 or 1. Indeed, we know from [3, Theorem 1.2.5] that $\mathbf{c}_n \in \partial M_n$ implies that the underlying μ is uniquely determined, so that, $c_n^+ = c_n^-$. We also introduce the quantity $\zeta_n = (1 - p_{n-1})p_n$ that will be of some importance in the following. The very nice properties of canonical moments are that, by construction, they belong to $[0, 1]$ and are invariant under linear transformation of the measures, $y = a + (b - a)x$. Hence, we may restrict ourselves to the case $a = 0$, $b = 1$.

3.2 Support Points and Canonical Moments

From a given sequence of canonical moments, one wishes to reconstruct the support of a discrete measure. This link arises through the following theorem

Theorem 2 ([3, Theorem 3.6.1]). *Let μ denote a measure on the interval $[a, b]$ supported on $n + 1$ points with canonical moments $p_1, p_2, \ldots$. Then the support of μ consists of the zeros of $P_{n+1}^*(x)$ where*

$$P_{k+1}^*(x) = (x - a - (b - a)(\zeta_{2k} + \zeta_{2k+1}))P_k^*(x) - (b - a)^2\zeta_{2k-1}\zeta_{2k}P_{k-1}^*, \tag{10}$$

with $P_{-1}^*(x) = 0,\ P_0^*(x) = 1$ and $\zeta_k = (1 - p_{k-1})p_k$

The polynomial P_{n+1}^* is defined with the sequence of canonical moments up to order $2n + 1$. In the following, we consider a fixed sequence of moments $\mathbf{c}_n = (c_1, \ldots, c_n) \in M_n$, let μ be a measure supported on at most $n + 1$ points, with classical moments $\mathbf{c}_n$. Hence, μ has canonical moments equal to $\mathbf{p}_n = (p_1, \ldots, p_n)$ the corresponding sequence of canonical moments related to $\mathbf{c}_n$, as described in Sect. 3.1. We define the set $\Theta_{n+1} = \{\mathbf{x} \in [0, 1]^{n+1} \mid x_i \in \{0, 1\} \Rightarrow x_k = 0, k > i\}$ and the functional:

$$\begin{aligned} \phi_{\mathbf{p}_n} : \qquad \Theta_{n+1} \qquad &\to \mathbb{R}[X] \\ (p_{n+1}, \ldots, p_{2n+1}) &\mapsto P_{n+1}^*, \end{aligned} \tag{11}$$

The function ϕ computes, from a sequence of canonical moments $(p_1, \ldots, p_{2n+1})$, a polynomial P_{n+1}^* in regards of Theorem 2. Therefore, the roots of P_{n+1}^* correspond to the support of a measure with moments $\mathbf{c}_n$. We derive the following Theorem, it is the geometric version of Theorem 2.

Theorem 3. *The set* $\mathcal{V}_i$ *of* (N_i+1)*-tuples corresponding to the support points of a discrete measure with prescribed first* N_i *moments* $(c_1^{(i)}, \ldots, c_{N_i}^{(i)})$ *is an algebraic manifold of* $\mathbb{R}^{N_i+1}$*. It is the zeros locus of the set of polynomials:*

$$\mathcal{S}_i = \left\{ P^*_{N_i+1} = \phi_{\mathbf{p}_{N_i}}(p_{N_i+1}, \ldots, p_{2N_i+1}),\ (p_{N_i+1}, \ldots, p_{2N_i+1}) \in \Theta_{N_i+1} \right\} \quad (12)$$

In order to optimize our quantity of interest in Eq. (3), one need to explore the space of admissible measures $\mathcal{A}_\Delta$. More precisely, the P.O.F in Eq. (4) is computed over the space $\mathcal{V}$. This space corresponds to the support points of all discrete measures respecting the constraints in $\mathcal{A}_\Delta$. What is interesting is that Θ_{n+1} provides a very simple parameterization of $\mathcal{V}$ through the computation of roots of some well defined polynomial.

An optimization over the highly constrained space $\mathcal{A}_\Delta$ is therefore simplified into a constraint free optimization program over the space Θ_{n+1}.

4 Conclusion

This work aims to evaluate the maximum quantile over a class of distributions constrained by some of their moments. We used the theory of canonical moments into an improved methodology for solving OUQ problems. The set of optimization corresponds to the support points of the discrete measures in the moment set. We have successfully shown it is the zero locus of a set of polynomials defined with canonical moments. The knowledge of the shape of this manifold allows a computational constraint free optimization program, instead of a highly constrained optimization over the moment set.

References

1. Betrò, B., Guglielmi, A.: Methods for global prior robustness under generalized moment conditions. In: Insua, D.R., Ruggeri, F. (eds.) Robust Bayesian Analysis. Lecture Notes in Statistics. Springer, New York (2000). https://doi.org/10.1007/978-1-4612-1306-2_15. www.springer.com/la/book/9780387988665
2. DeRoberts, L., Hartigan, J.A.: Bayesian inference using intervals of measures. Ann. Stat. **9**(2), 235–244 (1981). https://doi.org/10.1214/aos/1176345391. https://projecteuclid.org/euclid.aos/1176345391
3. Dette, H., Studden, W.J.: The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis. Wiley-Blackwell, New York (1997)
4. Henrion, D., Lasserre, J.B., Löfberg, J.: GloptiPoly 3: moments, optimization and semidefinite programming. Optim. Methods Softw. **24**(4–5), 761–779 (2009). https://doi.org/10.1080/10556780802699201. http://www.tandfonline.com/doi/abs/10.1080/10556780802699201
5. Huber, P.J.: The use of Choquet capacities in statistics. Bull. Int. Stat. Inst. **45**(4), 181–188 (1973)

6. Iooss, B., Marrel, A.: Advanced methodology for uncertainty propagation in computer experiments with large number of inputs. Nuclear Technol. 1–19 (2019). https://doi.org/10.1080/00295450.2019.1573617
7. Lasserre, J.B.: Moments, Positive Polynomials and Their Applications. Imperial College Press Optimization Series, vol. 1. Imperial College Press, London (2010). Distributed by World Scientific Publishing Co., Singapore. oCLC: ocn503631126
8. McKerns, M., Owhadi, H., Scovel, C., Sullivan, T.J., Ortiz, M.: The optimal uncertainty algorithm in the mystic framework. CoRR abs/1202.1055 (2012). http://arxiv.org/abs/1202.1055
9. Owhadi, H., Scovel, C., Sullivan, T.J., McKerns, M., Ortiz, M.: Optimal uncertainty quantification. SIAM Rev. **55**(2), 271–345 (2013). https://doi.org/10.1137/10080782X. arXiv:1009.0679
10. de Rocquigny, E., et al.: Uncertainty in industrial practice: a guide to quantitative uncertainty management (2008). https://doi.org/10.1002/9780470770733
11. Ruggeri, F., Rios-Insua, D., Martin, J.: Robust Bayesian analysis. In: Dey, D.K., Rao, C.R. (eds.) Handbook of Statistics, Bayesian Thinking, vol. 25, pp. 623–667. Elsevier, Amsterdam (2005). https://doi.org/10.1016/S0169-7161(05)25021-6. http://www.sciencedirect.com/science/article/pii/S0169716105250216
12. Sivaganesan, S., Berger, J.O.: Ranges of posterior measures for priors with unimodal contaminations. Ann. Stat. **17**(2), 868–889 (1989). https://doi.org/10.1214/aos/1176347148. https://projecteuclid.org/euclid.aos/1176347148
13. Wallis, G.: Uncertainties and probabilities in nuclear reactor regulation. Nuclear Eng. Des. **237**, 1586–1592 (2004)
14. Winkler, G.: Extreme points of moment sets. Math. Oper. Res. **13**(4), 581–587 (1988). https://doi.org/10.1287/moor.13.4.581