Robust Formulation

Romain Couderc

August 26, 2020

Contents

In this document, we discuss the subject of robustness regarding the objective function. First, we define robustness and the different types of uncertainties. Secondly, we present theoretical aspects about a new measure used to take into account the robustness of the objective function. Then, we present numerical experiments which allow to understand the interests of this new measure. Finally, we study theoretically the convexification phenomenon appeared during our tests with the new measure.

1 Robustness in design optimization

Robust optimization is a part of mathematical optimization which aims to solve an optimization problem in taking into account the different sources of uncertainties. However, several questions arise:

- why are there some uncertainties in a design model and why is it necessary to take them into account?
- How can we classify these different uncertainties ?
- Which are the current methods in robust optimization ?

We will answer these questions in the next subsections.

1.1 Necessity of robust optimization

When engineers need to optimize a problem, they search to find the right design parameters with the highest possible precision. However, even if they make great effort in the modelling and the conception of the product, there will be always some small changes between the behavior of the modelling product and the real one. Here is a list of the main factors of this difference [13]:

- The desired design variables are the solution of an optimization problem having an objective function and constraints. Nevertheless, these objective and constraint functions are numerical modelling of reality. Even if these models are as precise as possible, it is possible that the optimum obtained by solving the optimization problems does not match with the optimum in the reality.
- In the case where the optimum in the reality matches with the numerical model optimum, there is still a difficulty. It is possible that the optimal design found by the numerical modelling is unreachable because of manufacturing uncertainties or the design costs to reach the desired accuracy.
- Numerous parameters in the model are fluctuating during time, like the temperature or the humidity rate but are considered constant in the model. Moreover, even when designers take into account this fluctuation, there is always some parameters, sometimes even unknown by the designer, which are not taken into account by the model.

Their influence on the design is often very limited but may induce variations in the final product.

- Finally, a product is never completely finished, each component may be replaced or repaired during the lifetime of the final product. That can bring some variations in the behavior of the product which are not forecasted by the numerical model.

Therefore, there are numerous factors which may alter the behavior of a final product. Using a robust optimization method in the early stage of the design of a product may avoid flaws in the final product behavior and may reduce the cost induced to repaired them.

1.2 Definition of robustness

1.3 Classes of uncertainties

There are different ways to classify the sources of uncertainties. The terminologies differ according to the authors or the field in which the concept of design uncertainties is tackled. Consider the robust scenario of figure 1 taken from [13]. Three main classes of uncertainties are commonly used in the mathematical community:

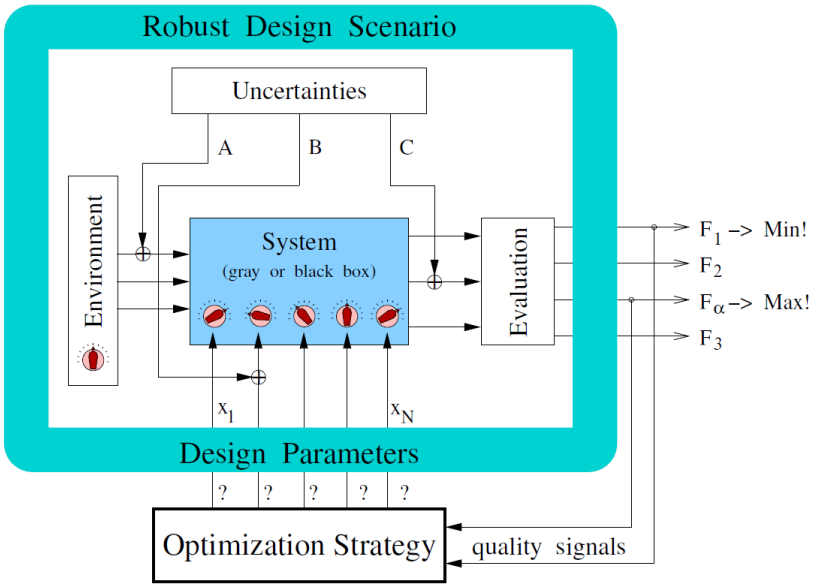

Figure 1: A robust design scenario with three classes of uncertainties A, B and C showing on which part of the problem they appear

- **Class A uncertainties**
  These uncertainties represent the uncertainties due to the environmental parameters of the system as the temperature, the pressure or the humidity rate. Let f denote the

behavior of the system and x the design variables. To take into account these parameters,random variables ν are added, the system is modelled by:

$$f = f(x, \nu). \tag{1}$$

This class of uncertainties are also called "Type I variations" in [4].

- **Class B uncertainties**
  These uncertainties correspond to the ability of the manufacturer to build the desire design at a specified degree of accuracy. Indeed, designing a product with a very high precision may be impossible or very expensive. That is why, there often exists a tolerance in the manufacturing of a product. These uncertainties may be modelled in the function f as small variations δ in the design variables:

  $$f = f(x + \delta) \tag{2}$$

  These uncertainties are called "Type II variations" in [4] and "Precision Error" in [12].

- **Class C uncertainties**
  These uncertainties concern the imprecision in the system output. In practice, it is impossible to have an exact measure of the output of a system. Moreover, these uncertainties group also the measuring error and all the approximation error due to the use of models. We can model them as a random function $\tilde{f}$:

  $$\tilde{f} = \tilde{f}(f(x)) \tag{3}$$

  These errors are also called "Bias error" in [12].

Unlike in [13], we do not create another category of uncertainties for the feasibility field of the function. In fact, each of the previous uncertainties may also be applied to constraint functions. The three classes of uncertainties are not mutually exclusive and may be combined. These classes present here are mainly used in the mathematical community because they differ by their mathematical formulation. In the engineering community, they divide the different uncertainties in two classes:

- The aleatory or random uncertainties [16] which are irreducible due to their natural origin. Indeed, these uncertainties have an environmental source like the temperature, the humidity rate or the stiffness of a material. Even if the manufacturer make a great effort to contain them, these uncertainties will be still present.

- The epistemic uncertainties represent the lack of knowledge of the manufacturer. These uncertainties could be reduced but at a great financial or technological cost. These uncertainties are often due to the model used to represent the reality or due to the numerical errors made during the problem solving.

In this paper, the mathematical classification is often used in order to be more formal. Nevertheless, it is possible that in the practice case, we use also the terminology of the engineering community.

1.4 Uncertainties frameworks

Once we have classify the different classes of uncertainties, we give some mathematical uncertainties frameworks. In the literature, there are basically three:

- Using a probabilistic framework [19], for each uncertain variable, a probabilistic law is defined. This framework is particularly adapt for aleatory uncertainties. Indeed, the source of this class of uncertainties is often known for a long time. Thus, their behavior may be modelled thanks to a probabilistic law for instance.
- Using a deterministic interval framework [22], for each uncertain variable, a domain is stated in which the variable can vary.
- Using the possibility framework [9] [19] , for each uncertain variable, a fuzzy set is defined. These two later frameworks are often used to model the epistemic uncertainties because these uncertainties vary from a model to another and it makes it difficult to model with a probabilistic framework.

In this paper, the uncertainties take into account are those of class A and B. Their behaviors are assumed known, thus the framework develop in the next section is the probabilistic one.

1.5 Current approaches in robust optimization

In this Section, we present three different methods to treat uncertainties in robust optimization based on the three different models from the previous Section.

1.5.1 The Robust Counterpart Approach

This approach was developed mainly by the authors of [1]. The concept is simple, given an objective function $f(x)$ to optimize, the Robust Counterpart of this function is defined as:

$$F_{RC}(x;\epsilon) = \sup_{\xi \in \mathcal{B}(x,\epsilon)} f(\xi) \tag{4}$$

where $\mathcal{B}(x,\epsilon)$ is a neighborhood on the design variable x whose the size depends of the regularization parameter ϵ. The link between the function f and its robust counterpart F_{RC} is given by the size ϵ, by the following relation:

$$\lim_{\epsilon \to 0} F_{RC}(x;\epsilon) = f(x). \tag{5}$$

In the previous equations we show the application of Robust Counterpart in case of class B uncertainties. However, we can also easily apply it with uncertainties of class A:

$$F_{RC}(x;\epsilon) = \sup_{\alpha \in \mathcal{B}(\epsilon)} f(x,\alpha) \tag{6}$$

where $\mathcal{B}(\epsilon)$ is a neighborhood on the random parameters of the system. The main strength of the Robust Counterpart approach is, in addition to handle class A or B uncertainties, to take into account the constraints also. For instance, if we consider the following linear programming (LP) problem:

$$\min_x \{c^T x : Ax \leq b\} \tag{7}$$

Then, in presence of uncertainties of class A (on the data of the problems), the uncertain LP problems may be defined as a collection of LP problems:

$$\{\min_x \{c^T x : Ax \leq b\} : (c, A, b) \in \mathcal{U}\} \tag{8}$$

where the data (c, A, b) varying in a given uncertainty set $\mathcal{U}$. Solve this uncertain LP problem, is searching a solution which is a fixed vector remaining feasible for the constraints, whatever the realization of the data within $\mathcal{U}$. That leads to consider a worst case scenario mathematically modelled by the resolution of a min-max problem:

$$\min_x \left\{ \sup_{(c,A,b)\in\mathcal{U}} c^T : Ax \leq b \; \forall (c, A, b) \in \mathcal{U} \right\} \tag{9}$$

or equivalently what is called the Robust Counterpart of the original uncertain problem:

$$\min_{x,t} \{t : c^T x \leq t, Ax \leq b \; \forall (c, A, b) \in \mathcal{U}\}. \tag{10}$$

The methodology of the Robust Counterpart is fully applicable in context of linear, conic and semidefinite programming. There are advantages to the Robust Counterpart approach, however in our context of blackbox optimization there is a great limitation: the Robust Counterpart approach necessitates to have a problem with a known fixed structure. It is not the case in the blackbox optimization where the objective function and the constraints are very often the results of numerical simulations.

1.5.2 The possibilistic uncertainties approach

In the previous section and in the next section, the uncertainties are modelled using complete information on the uncertainties, for instance their bounds or their distributions. However, there are numerous cases where these information are not available, especially if epistemic uncertainties are treated. To treat this type of uncertainties, there are commonly two ways:

- Either, a probabilistic model is used and improved iteratively with new knowledge, using bayesian statistic for instance.
- Or fuzzy sets and evidence theory are used.

The idea behind the fuzzy set methodology is to associate at each value of the objective or constraints functions a measure belonging to a certain set. For example if an optimum of the unconstrained objective function is denoted $f(x^*)$, then we associate to this x^* a measure

$\mu_f(x^*) = 1$. In the contrary the worst design is assigned a measure of 0. Between the two design, we can assign intermediate value. In the constrained case, it is almost the same. The measure $\mu_c(x) \in [0, 1]$ indicates if the design is certainly feasible $\mu_c(x) = 1$, certainly infeasible $\mu_c(x) = 0$ or intermediate $\mu_c(x) \in]0, 1[$. The problem is then to find the design x^* in:

$$\underset{x}{\text{argmax}}[\min(\mu_f(x), \mu_c(x))]. \tag{11}$$

Although, there is a mathematical background quite large on the possibility theory, the fuzzyfication of a problem stay complex:

- Determing the concrete forme of the measure μ may be quite difficult, especially when the problem combines different fuzzy constraints.
- The numerical effort to solve the min-max problem may be computationally high in case of real application.

1.5.3 The probabilistic risk measure approach

A third approach to handle uncertainties is the probabilistic one. A strong hypothesis in this approach is to suppose that the distribution of the random variables is known. In the two previous approaches less information is needed on the random variables, the bound in the first case and nothing in the second one. Nevertheless, in context of blackbox optimization where the structure of the problem is not known, these two approaches are difficult to enforce. In case of probabilistic approach, we regard δ, ν and $\tilde{f}$ of equations 1, 2 and 3 as random variables. That is, the function f and the constraints themselves become random functions. We will expose the different manner to treat the uncertainties in this case, first for the objective function then for the constraints.

Case 1: how to handle a random objective function?

The search for an optimal design of a random objective function is not obvious. Indeed, the random aspect of the function introduces a new paradigm: what is called a minimum ? However, a first answer to this question is to obtain the optimal design in "mean". Nevertheless, consider the optimal design, just in term of mean may be hazardous. Indeed, the best mean result may lead to a design with a great deviation in its behavior. That is why, the goal of the designer is to find a trade-off between an optimal solution in term of mean performance and the deviation of this optimal performance. The original robust problem becomes a bi-objective problem and there are two commonly ways to solve it:

- Optimize a weighted sum between the expectation of the objective function $\mathbb{E}[f(x+\delta)]$ and its standard deviation $\sigma[f(x+\delta)]$, i.e.:

$$\min(1-\beta)\mathbb{E}[f(x+\delta)] + \beta\sigma[f(x+\delta)] \tag{12}$$

  Examples of this solution is present in [21].

- See that a multi-objective problem and search the Pareto optimal solutions of the problem:

$$\begin{cases} \min \mathbb{E}[f(x+\delta)] \\ \min \sigma[f(x+\delta)] \end{cases}$$

  An application of this method may be found in [25].

We present these previous cases with the class B uncertainties but it works for any classes of uncertainties. Each of this method rises a technical matter: the choice of β in the first case and the numerical cost of searching a Pareto front in the second one.

Case 2: how to handle the random constraints ? A common way of handling the random constraints $g(x+\delta) \leq 0$ is to consider the inequalities probabilistically, called a chance constraints:

$$\mathbb{P}(g(x+\delta) \leq 0) \geq \alpha \tag{13}$$

where α is the confidence index and δ a random variable. This field is called reliability based design optimization (RBDO) see [27] for a review of the main methods. The difficulty in RBDO is the fact that the constraint 13 is an integral. Let consider a unique constraint, equation 13 may be written as:

$$\mathbb{P}(g(x+\delta) \leq 0) = \int_{g(x+\delta)\leq 0} p(\delta)d\delta \geq \alpha \tag{14}$$

with p the probability density function of δ. Since the chance constraint is non linear and involve integral, it numerical calculation is often costly. In practice, designers use approximation as First-Order Reliability Methods(FORM) or Second Order Reliability Methods (SORM) [6] to handle it (see section on the handling of constraints).

2 A new risk measure: Average Value-at-Risk

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a sample space, equipped with the sigma algebra $\mathcal{F}$ and a measure on this sigma algebra $\mathbb{P}$, on which considered uncertain outcomes (random function $Z = Z(\omega) \in \mathbb{R}$) are defined. By a risk measure we understand a function $\rho(Z)$ which maps Z into the extended real line $\bar{\mathbb{R}}$. The specific measure introduced below is the subject of many works. The terminology used in this paper is the one used in [2]. Nevertheless, this measure has been called Conditional Value-at-Risk in [23] and is equivalent to Expected Shortfall in [3]. For each terminology a corresponding formulation exists. That is why, in this section, we describes first the different formulation, then we speak about the coherence of a risk measure [11] which is a set of properties of a risk measure. Finally, we will show a result about the minimization of the Average Value-at-Risk.

2.1 Average Value-at-Risk formulations

First, we are going to define the Average Value-at-Risk in the terminology of [23].

Definition 2.1. *The Conditional Value-at-Risk of* $Z \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ *at level* $\alpha \in [0,1]$ *is the risk measure define by:*

$$CVAR_\alpha(Z) = \inf_{t\in\mathbb{R}}\{t + \alpha^{-1}\mathbb{E}[(Z-t)_+]\} \text{ with } (Z-t)_+ = \max(Z-t, 0). \tag{15}$$

In [2], they use another formulation with two different parameters, in the following of the paper, we use this formulation because it allows more flexibility (due to the two parameters).

Definition 2.2. *Let* $\beta_1 \in [0,1]$ *and* $\beta_2 \in [0, +\infty]$*, the average value-at-risk of* $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ *at level* $\alpha = \frac{\beta_1}{\beta_1+\beta_2}$ *is defined by:*

$$AVAR_\alpha(Z) := \mathbb{E}[Z] + \inf_{t\in\mathbb{R}} \mathbb{E}[\beta_1(t-Z)_+ + \beta_2(Z-t)_+]. \tag{16}$$

In the next lemma, we are going to see that the two formulations are equivalent in the case where $\beta_1 = 1$ and $\beta_2 > 0$. This result is p. 276 of [2].

Lemma 2.3. *If* $\beta_1 = 1$ *and* $\beta_2 > 0$ *then we have* $\forall Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$*:*

$$AVAR_\alpha(Z) = CVAR_\alpha(Z). \tag{17}$$

Proof. Let $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$, $\beta_1 = 1$ and $\beta_2 > 0$, we have:

$$\begin{aligned}
AVAR_\alpha(Z) &= \mathbb{E}[Z] + \inf_{t\in\mathbb{R}} \mathbb{E}[\beta_1(t-Z)_+ + \beta_2(Z-t)_+] \\
&= \mathbb{E}[Z] + \inf_{t\in\mathbb{R}} \mathbb{E}[\beta_1(t-Z) + \beta_1(Z-t)_+ + \beta_2(Z-t)_+] \\
&= \mathbb{E}[Z] - \beta_1\mathbb{E}[Z] + \inf_{t\in\mathbb{R}} \mathbb{E}[\beta_1 t + (\beta_1+\beta_2)(Z-t)_+] \\
&= (1-\beta_1)\mathbb{E}[Z] + \beta_1 \inf_{t\in\mathbb{R}}\{t + \mathbb{E}[\frac{\beta_1+\beta_2}{\beta_1}(Z-t)_+]\} \\
&= (1-\beta_1)\mathbb{E}[Z] + \beta_1 CVAR_\alpha(Z)
\end{aligned}$$

And replacing β_1 by 1, we obtain:

$$AVAR_\alpha(Z) = CVAR_\alpha(Z).$$

□

Another terminology for this risk measure exists and is due to [3]: Expected Shortfall. In order to formulate this notion, we must first introduce the Value-at-Risk risk measure which the based of all the others measure described in this section (even if it does not appear directly in the previous formulation).

Definition 2.4. *We define the Value-at-Risk of level α as:*

$$VAR_\alpha(X) = \inf\{t : \mathbb{P}(X \leq t) \geq 1 - \alpha\} \tag{18}$$

Once, we have define this measure,then we can also define the Expected Shortfall in case where Z is a real valued random variable (see proposition 3.2 of [3]).

Definition 2.5. *We call the Expected Shortfall of $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ at level α the following value:*

$$ES_\alpha(Z) = \frac{1}{\alpha} \int_{1-\alpha}^{1} VAR_{1-\tau}(Z) d\tau. \tag{19}$$

In fact, the Expected Shortfall is equivalent to the formulations 2.1. We will prove it in 2.10 after having stated some other theoretical results.

2.2 Average Value-at-Risk: a coherent risk measure

The different measure of risk are not all equivalent. Indeed, some risk measures have some good properties. We say then that the measure is coherent. In the following, we note $\mathcal{Z} = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ with $p \in [1, +\infty)$ the spaces of variables Z. By assuming that $Z \in \mathcal{Z}$, we assume that the random variable $Z(\omega)$ has a finite p-th order moment with respect to the reference probability measure $\mathbb{P}$.

Definition 2.6. *We say that a measure ρ is a coherent risk measure if it satisfies the four conditions below:*

- *Convexity:*
$$\rho(tZ + (1-t)Z') \leq t\rho(Z) + (1-t)\rho(Z')$$
for all $Z, Z' \in \mathcal{Z}$ and all $t \in [0, 1]$.

- *Monotonicity: If $Z, Z' \in \mathcal{Z}$ and $Z(\omega) \geq Z'(\omega)$ for a.e. (almost every) $\omega \in \Omega$, then:*
$$\rho(Z) \geq \rho(Z').$$

- *Translation equivariance: If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(Z + a) = \rho(Z) + a$.*

- *Positive homogeneity: If $t > 0$ and $Z \in \mathcal{Z}$, then $\rho(tZ) = t\rho(Z)$.*

These properties will be able to be useful in further works particularly if we decide to work on differentiability of risk measures. The following lemma is not explicitly prove in [2] but it is an application of the proof they made in a more general case.

Lemma 2.7. *The average value-at-risk measure defined with the specified β_1 and β_2 in (16) is a coherent risk measure.*

Proof. First, let introduce the functions:

$$\forall x \in \mathbb{R} \quad h(x) = \beta_1(-x)_+ + \beta_2(x)_+ \text{ and } g(x) = x + h(x).$$

The derivatives of g is:

$$g'(z) = \begin{cases} 1 - \beta_1 & \text{if } z < 0 \\ 1 + \beta_2 & \text{if } z > 0 \end{cases}$$

Thus, for the specified β_1 and β_2, g is non decreasing and is convex. Moreover, we can see (16) as follows:

$$AVAR_\alpha(Z) = \inf_{t \in \mathbb{R}} \mathbb{E}[t + g(Z - t)] \tag{20}$$

where g is the function defined with $\beta_1 \in [0, 1]$ and $\beta_2 \in [0, +\infty)$. Thus, we can prove all the points above:

- Since g is convex with the specified β_1, β_2, then by linearity of the expectation and convexity of $\inf$, we have $AVAR_\alpha$ which is convex.
- With the same arguments, since g is non decreasing then $AVAR_\alpha$ is monotone and non decreasing.
- We have for any $a \in \mathbb{R}$:

$$\begin{aligned} AVAR_\alpha(Z + a) &= \inf_{t \in \mathbb{R}} \mathbb{E}[Z + a + h(Z + a - t)] \\ &= \mathbb{E}[Z] + a + \inf_{t \in \mathbb{R}} \mathbb{E}[h(Z + a - t)] \\ &= \mathbb{E}[Z] + a + \inf_{t \in \mathbb{R}} \mathbb{E}[h(Z - t)] = AVAR_\alpha(Z) + a \end{aligned}$$

- Finally, for all $u > 0$, we have:

$$\begin{aligned} AVAR_\alpha(uZ) &= \inf_{t \in \mathbb{R}} \mathbb{E}[uZ + h(uZ - t)] \\ &= u \times \inf_{t \in \mathbb{R}} \mathbb{E}[Z + h(Z - \frac{t}{u})] \\ &= u \times \inf_{t \in \mathbb{R}} \mathbb{E}[Z + h(Z - t)] = u \times AVAR_\alpha(Z). \end{aligned}$$

□

We have seen that the Average Value-at-Risk measure is a coherent risk measure. Fulfilling the four conditions are not obvious, for instance the mean-variance risk measure are not coherent risk measure. In the next Section, we are going to use a part of this result to prove others results.

2.3 Minimum of Average Value-at-Risk and equivalence with Expected Shortfall

In this section, we are going to state some interesting results. The first one shows the link between the Average Value-at-Risk and the Value-at-Risk. In fact, the first risk measure attains its minimum at the second one. The second result shows the link between Expected Shortfall and Conditional Value-at-Risk (and thus with Average Value-at-Risk). It allows also to understand why this name have been chosen. Let begin by the first result:

Proposition 2.8. *The measure $AVAR_\alpha$ defined in (16) attains its minimum with respect to t at:*

$$t^* = \inf\{t : \mathbb{P}(Z \leq t) \geq 1 - \alpha\} \quad \textit{with} \quad \alpha = \frac{\beta_1}{\beta_1 + \beta_2}. \tag{21}$$

In order to prove the above proposition, we will need the following result from theorem 7.46 of [2]. In this theorem, we consider a random function $F_\omega : \mathbb{R}^n \times \Omega \to \bar{\mathbb{R}}$ and the corresponding expected value function $f(x) = \mathbb{E}[F_\omega(x)]$.

Theorem 2.9. *Suppose that the random function $F_\omega(x)$ is convex and the expected value function $f(x)$ is well defined and finite valued in a neighborhood of a point x_0. Then, $f(x)$ is convex, directionally differentiable at x_0 and we have:*

$$f'(x_0, d) = \mathbb{E}[F'_\omega(x_0, d)] \quad \textit{for all direction } d. \tag{22}$$

Moreover, $f(x)$ is differentiable at x_0 with probability 1, in which case:

$$\nabla f(x_0) = \mathbb{E}[\nabla_x F_\omega(x_0)].$$

Proof. See the proof of theorem 7.46 of [2]. □

Henceforth, we have all the results necessary to prove the proposition 2.8.

Proof. Let take $\beta_1 \in [0, 1]$ and $\beta_2 \geq 0$, since we search the minimum of $AVAR_\alpha(\cdot)$ with respect to t, it is sufficient to study, for $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ and $t \in \mathbb{R}$:

$$f(t) = \mathbb{E}\{h_\omega(Z - t)\} = \mathbb{E}[\beta_1(t - Z(\omega))_+ + \beta_2(Z(\omega) - t)_+\}$$

We know that f is well defined and finite valued for $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, since what we have proved in lemma 2.7, we know that $h_\omega(\cdot)$ is convex and differentiable everywhere except at $x = Z(\omega)$. The corresponding derivative is given by:

$$\frac{\partial h_\omega(Z - t)}{\partial t} = \begin{cases} \beta_1 & \text{if } t > Z(\omega) \\ -\beta_2 & \text{if } t < Z(\omega) \end{cases}$$

And we have $\forall t \neq Z(\omega)$ by theorem 2.9:

$$f'(t) = \mathbb{E}\left[\frac{\partial h_\omega(Z-t)}{\partial t}\right] = \beta_1 \mathbb{P}(t > Z) - \beta_2 \mathbb{P}(t < Z).$$

For the case where $t = Z(\omega)$, there are two ways:

- Either, the cumulative distribution function of Z is continuous at $t_0 = Z(\omega)$, then the event $\{t_0 = Z(\omega)\}$ has zero probability, thus we have by the second point of theorem 2.9 (we can see that in a way as an extension by continuity):

$$f'(t_0) = \mathbb{E}\left[\frac{\partial h_\omega(Z-t_0)}{\partial t}\right] = \beta_1 \mathbb{P}(t_0 > Z) - \beta_2 \mathbb{P}(t_0 < Z) \tag{23}$$

  So, this formula is true $\forall t \in \mathbb{R}$. We can conclude since f is convex and differentiable everywhere that the minimum is reached at $t^* \in \mathbb{R}$ such that:

$$f'(t^*) = 0 \Leftrightarrow \mathbb{P}(Z \leq t^*) = \frac{\beta_2}{\beta_2 + \beta_1} = 1 - \alpha.$$

- Or the cumulative distribution function is not continuous at $t_0 = Z(\omega)$ but by the first point of theorem 2.9 the directional derivatives exist. Thus, we must study the derivative of f on the right of t_0 and on the left. On the right side of t_0, we have:

$$\begin{aligned} f'(t_0, 1) &= \mathbb{E}[h'_\omega(Z - t_0, 1)] \\ &= \mathbb{E}\left[\lim_{h \to 0} \frac{h_\omega(Z - (t_0 + h)) - h_\omega(Z - t_0)}{h}\right] \\ &= \mathbb{E}\left[\lim_{h \to 0} \frac{\beta_1(t_0 + h - Z(\omega))_+ + \beta_2(Z(\omega) - t_0 - h)_+}{h}\right] \end{aligned}$$

  because $h_\omega(Z - t_0) = 0$. And then by the Monotone Convergence Theoreme (using the same arguments that in the proof of theorem 7.46 of [2]), we can switch the $\lim$ and $\mathbb{E}[\cdot]$:

$$\begin{aligned} f'(t_0, 1) &= \lim_{h \to 0} \mathbb{E}\left[\frac{\beta_1(t_0 + h - Z(\omega))_+ + \beta_2(Z(\omega) - t_0 - h)_+}{h}\right] \\ &= \lim_{h \to 0} \beta_1 \mathbb{P}(t_0 + h \geq Z(\omega)) - \beta_2 \mathbb{P}(t_0 + h \leq Z(\omega)) \\ &= \lim_{h \to 0} \beta_1 \mathbb{P}(t_0 + h \geq Z(\omega)) - \beta_2 + \beta_2 \mathbb{P}(t_0 + h \geq Z(\omega)) \end{aligned}$$

  So, a first condition on the minimum of f on the right side of $t_0 = Z(\omega)$ is:

$$\lim_{h \to 0} \mathbb{P}(Z \leq t_0 + h) = 1 - \alpha \quad \text{i.e.} \quad t^* = \inf\{t : \mathbb{P}(Z \leq t) \geq 1 - \alpha\} \tag{24}$$

On the left side of $t_0 = Z(\omega)$, we have by symmetry:

$$f'(t_0, -1) = \lim_{h \to 0} \beta_1 \mathbb{P}(t_0 - h \geq Z(\omega)) - \beta_2 + \beta_2 \mathbb{P}(t_0 - h \geq Z(\omega))$$

So, a second condition on the minimum of f on the left side of $t_0 = Z(\omega)$ is:

$$\lim_{h \to 0} \mathbb{P}(Z \leq t_0 - h) = 1 - \alpha \quad \text{i.e.} \quad t^{**} = \sup\{t : \mathbb{P}(Z \leq t) \leq 1 - \alpha\} \tag{25}$$

And finally, the minimum of f in this case is attained on the interval $[t^*, t^{**}]$.

□

Once we have proved that, we can use this result to prove the other important result: the link between the Expected Shortfall and the Conditional Value-at-Risk. The result is the following it is a part of the theorem 6.2 of [2].

Proposition 2.10. *Let $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ and $H(z)$ its cumulative distribution function, with $H(z)$ continuous at $z = VAR_\alpha(Z)$, we have the following identities:*

$$CVAR_\alpha(Z) = \frac{1}{\alpha} \int_{1-\alpha}^{1} VAR_{1-\tau}(Z) d\tau = \mathbb{E}(Z | Z \geq VAR_\alpha(Z)) \tag{26}$$

Proof. As we have just proved, the minimum of $CVAR_\alpha(\cdot)$ is attained at $t^* = VAR_\alpha(Z)$. Therefore, we have:

$$CVAR_\alpha(Z) = t^* + \alpha^{-1} \mathbb{E}[(Z - t^*)_+] = t^* + \alpha^{-1} \int_{t^*}^{+\infty} (z - t^*) dH(z)$$

Moreover, provided that $\mathbb{P}(Z = t^*) = 0$, i.e. $H(z)$ is continuous at $z = VAR_\alpha(Z)$, we have:

$$\begin{aligned}
\alpha^{-1} \int_{t^*}^{+\infty} (z - t^*) dH(z) &= \alpha^{-1} \int_{t^*}^{+\infty} z dH(z) - \alpha^{-1} \int_{t^*}^{+\infty} t^* dH(z) \\
&= \alpha^{-1} \int_{t^*}^{+\infty} z dH(z) - \alpha^{-1} t^* \mathbb{P}(Z \geq t^*) \\
&= \alpha^{-1} \int_{t^*}^{+\infty} z dH(z) - \alpha^{-1} t^* (1 - \mathbb{P}(Z \leq t^*)) \\
&= \alpha^{-1} \int_{t^*}^{+\infty} z dH(z) - \alpha^{-1} t^* \alpha \\
&= \alpha^{-1} \int_{t^*}^{+\infty} z dH(z) - t^*
\end{aligned}$$

By the substitution $\tau = H(z)$, this shows the first equality in 2.10 and then the second equality in 2.10 follows provided that $\mathbb{P}(Z = t^*) = 0$. □

3 Numerical Experiments

In order to quantify the effect of the previous theoretical results, different risk measures are tested. Three different measures of uncertainties are tested on test functions mainly from [13]. These measures are Robust Counterpart, mean-variance measure and the Average Value-at-Risk measure described in Section 1.5. Tests are made for different class of uncertainties. When it is possible the robust minimum is given according to the measure used and in case where it is relevant some figures in 1 and 2 dimensions are provided to show the result.

3.1 Class B uncertainties: the quadratic sphere

The test function is really simple:

$$f(x) = ||x||^2 = \sum_{i=1}^{n} x_i^2 \tag{27}$$

3.1.1 The Robust Counterpart approach

In this case of uncertainties of class B, we have:

$$F_{RC}(x;\epsilon) = \sup_{||\delta|| \leq \epsilon} (x+\delta)^2 \tag{28}$$

$$= \sup_{||\delta|| \leq \epsilon} \left(||x||^2 + 2x^T\delta + ||\delta||^2 \right) \qquad = (||x|| + \epsilon)^2 \tag{29}$$

The robust minimizer is attained at $x^* = 0$.

3.1.2 The mean-variance approach

Assuming that $\delta \sim \mathcal{N}(0, I)$ (or equivalently that each δ_i follows a standard normal law), we have for the expectation:

$$\mathbb{E}[f(x+\delta)] = \mathbb{E}[\sum_{i+1}^{n}(x_i + \delta_i)^2] \tag{30}$$

$$= \sum_{i=1}^{n} \mathbb{E}[x_i^2 + 2\delta_i x_i + \delta_i^2] \text{ by linearity of the expectation} \tag{31}$$

$$= \sum_{i=1}^{n} x_i^2 + 1 \tag{32}$$

$$= ||x||^2 + n \tag{33}$$

And for the variance we have:

$$Var[f(x+\delta)] = 4||x||^2 + 2n \tag{34}$$

The minimizers for the expectation and the variance are exceptionally the same : $x^* = 0$. Thus, in this special case, there is no need of weighted sum or pareto front. We can see that is was also the same that for the robust counterpart approach.

3.1.3 The Average Value-at-Risk approach

Assuming that $\delta \sim \mathcal{N}(0, I)$ (or equivalently that each δ_i follows a standard normal law), we have:

$$VAR_\alpha(f(x+\delta)) = \inf\left\{t : \mathcal{P}\left(\sum_{i=1}^{n}(x_i+\delta_i)^2 \leq t\right) \geq 1-\alpha\right\} \tag{35}$$

Since [13], we know also that:

$$Z = \sum_{i=1}^{n}(x_i+\delta_i)^2 \sim \mathcal{X}_n'^2(||x||^2) \tag{36}$$

where $\mathcal{X}_n'^2(||x||^2)$ is a non central chi-squared distribution with n degrees of freedom and a non centrality parameter equal to $||x||^2$. The cumulative distribution function of $\mathcal{X}_n'^2(||x||^2)$ is equal to [20]:

$$\mathcal{P}(Z \leq t) = 1 - Q_{\frac{n}{2}}(||x||, \sqrt{t}) \tag{37}$$

with Q the Marcum-Q function. Moreover, in [28], the authors prove in Theorem 1 that this function is strictly increasing in $||x||$ for all $t > 0$ and $n > 0$. Then, we can deduce that:

$$\mathcal{P}(\sum_{i=1}^{n}(x_i+\delta_i)^2 \leq t) < \mathcal{P}(\sum_{i=1}^{n}(\delta_i)^2 \leq t) \tag{38}$$

and thus that:

$$\inf\{t : \mathcal{P}(\sum_{i=1}^{n}(x_i+\delta_i)^2 \leq t) \geq 1-\alpha)\} \geq \inf\{t : \mathcal{P}(\sum_{i=1}^{n}(\delta_i)^2 \leq t) \geq 1-\alpha\} \tag{39}$$

By integration we obtain finally that $\forall x \in \mathbb{R}^n$:

$$CVAR_\alpha(f(x+\delta)) \geq CVAR_\alpha(f(\delta)) \tag{40}$$

So, with this measure the robust minimizer is attained at $x^* = 0$.

3.2 Class B uncertainties: Function with multiple minima

The function on which we are going to test the different risk measures is drawn on Figure 2 and its formulation is:

$$f(x) = 1 - \exp\left(-\frac{1}{2}x^2\right) - \exp\left(-\frac{1}{50}(x-7)^2\right) - \exp\left(-5(x+5)^2\right). \tag{41}$$

This function has two local minima at $x = -5$ and $x = 7$ and one global minimum at $x = 0$. In this example, we draw the different function to optimize according to the risk measure used. That allow us to understand the behavior of the function to optimize. The robust minimum of this function is considered to be at $x = 7$.

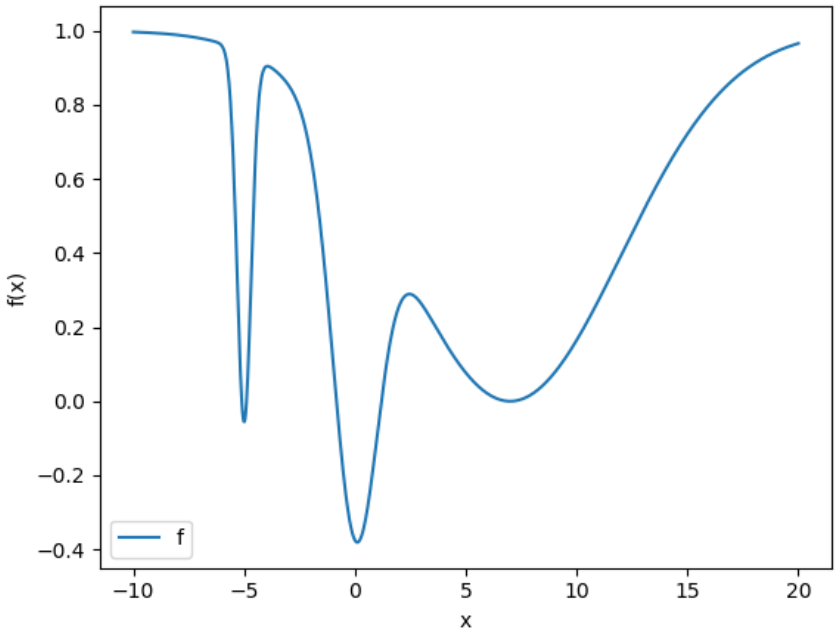

Figure 2: Test function

3.2.1 The Robust Counterpart approach

In this case, we search to optimize the following function:

$$F_{RC}(x;\epsilon) = \sup_{|\delta|\leq\epsilon}(f(x+\delta)) \tag{42}$$

We draw the result function for $\epsilon = 1$ on the figure 3. As we can see on the figure, the robust counterpart approach does not allow to choose really which minimum is the robust minimum.

3.2.2 The mean-variance approach

In this case, we have to calculate the following function for different values of β:

$$F_{MV}(x) = (1-\beta)\mathbb{E}[f(x+\delta)] + \beta Var[f(x+\delta)] \tag{43}$$

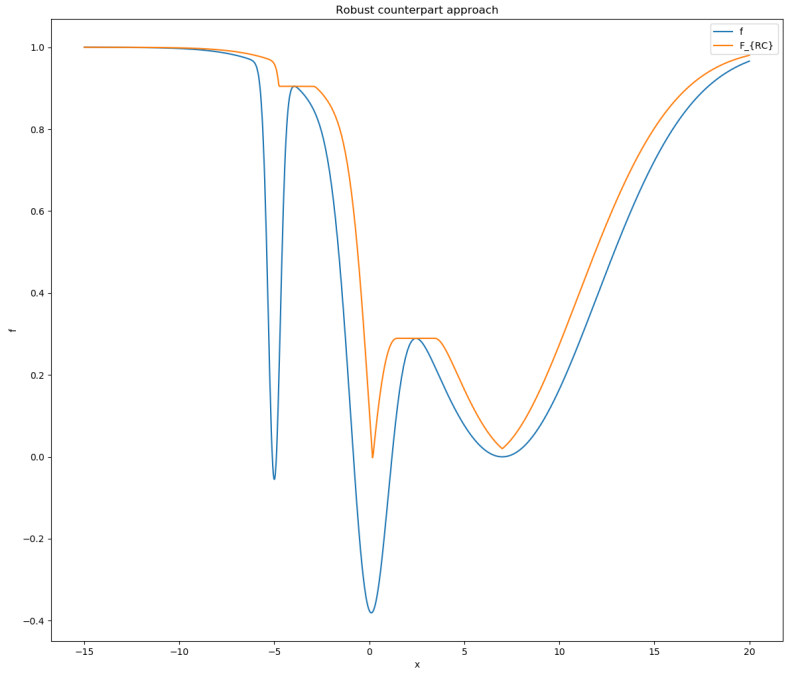

Figure 3: Robust Counterpart Approach

We recall here that $\beta = 0$ correspond to the case where the variance is not taken into account. In the contrary, take $\beta = 1$ corresponds to take into account only the variance. In order to present the different curves, we assume that $\delta \sim \mathcal{N}(0, 1)$. For each point, we draw 1000 samples of $f(x+\delta)$ and then we compute the mean and the variance to draw the curves.

On the figure 4, the result is quite different according to the value of β. There are three main matters:

- In case where β is quite big, the addition of the variance create some maxima in the function to optimize. This is a big problem because, these maxima are create near local minima of the deterministic function.
- In case where β is too small, the minimum of the mean variance function is still located at $x = 0$, we would like to locate it at $x = 7$.
- The great differences in the behavior of the robust mean variance function are confusing. It is quite difficult to know which β we must choose.

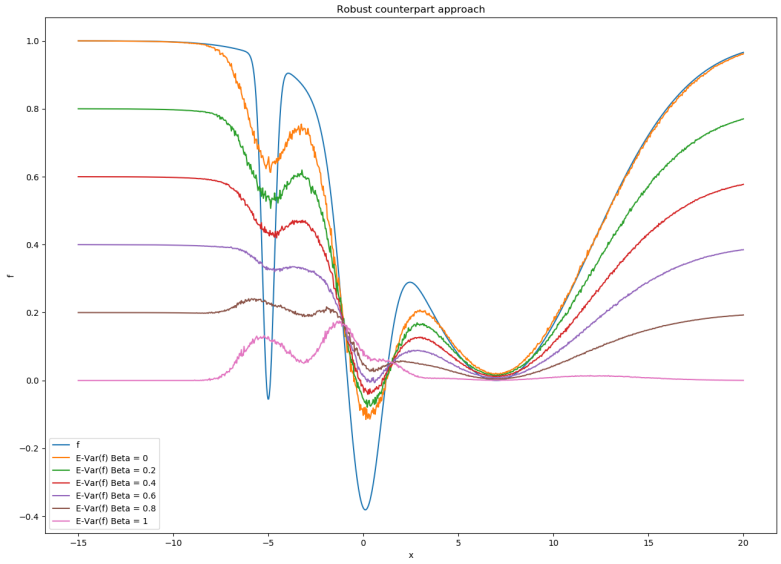

Figure 4: Mean-Variance approach for different values of beta

3.2.3 The Average Value-at-Risk approach

The result of the process described above and applied to Average Value-at-Risk measure is presented on Figure 5. We draw it for different values of $1 - \alpha$ which we call the robustness indices. This index has the same signification for the objective function that the reliability index used in the reliability based design optimization for the constraints.

The result shows several things:

- First, for all the values of $1 - \alpha$ considered, the $AVAR_\alpha$ approach leads to the robust minimum.
- In contrary to the mean variance approach, the behavior of the Average Value-at-Risk measure does not introduce some maxima. The robust function stays quite close to the original one in any cases.
- Finally, there is a phenonenon very interesting which appears. For $\beta = 0.9$, the result of the Average Value-at-Risk measure applied to the test function is quasi convex. It is obvious this phenomenon appears for some particular value of σ and β. However, we come back on this phenomenon in the next section.

To sum up, if we take some proper values for β, then the Average Value-at-Risk measure may identify the true robust minimum without having a behavior too far away of the original test function. In addition, it would seem that the non robust minima may be transformed in kind of "levels", and the function to optimize becomes then quasi convex.

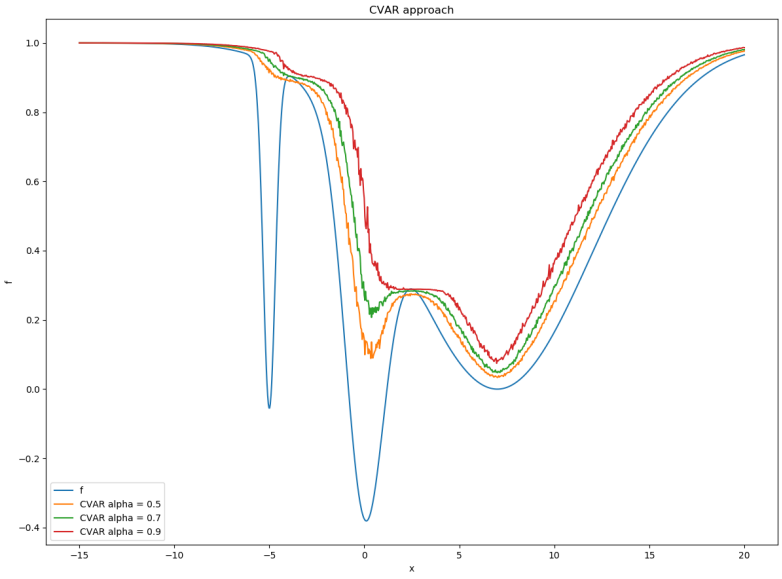

Figure 5: Average Value-at-Risk measure

3.3 Class A uncertainties: the sphere model

In this section, we test the different measure on the uncertainties of class A. We consider the following function:

$$f(x,\nu) = -a + (\nu+1)||x||^{\beta} - b\nu,\ \beta > 0,\ \nu \in \mathbb{R} \tag{44}$$

In the contrary to [13] where they consider a maximization process. In our case, we are interesting in a minimization, that is why we take the opposite of their original function.

3.3.1 The Robust Counterpart Approach

The robust counterpart function of f is the function:

$$F_{RC}(x,\epsilon) = \sup_{|\nu|\leq\epsilon} f(x,\nu) \tag{45}$$

$$= \sup_{|\nu|\leq\epsilon} \{-a + ||x||^{\beta} + \nu(||x||^{\beta} - b)\} \tag{46}$$

$$= \begin{cases} -a + b\epsilon + (1-\epsilon)||x||^{\beta} & \text{if } ||x||^{\beta} \leq b \\ -a - b\epsilon + (1+\epsilon)||x||^{\beta} & \text{if } ||x||^{\beta} > b \end{cases} \tag{47}$$

In the case of minimization on $\mathbb{R}$, the solution is located when $||x||^{\beta} \leq b$ and its depends on the value of ϵ. We have:

$$\min_x F_{RC} = \begin{cases} -a + b\epsilon & \text{if } 0 \leq \epsilon \leq 1, \\ -a + b & \text{if } \epsilon > 1. \end{cases} \tag{48}$$

The solutions are $x^* = 0$ for the case where $0 \leq \epsilon \leq 1$ and $||x||^* = b^{\frac{1}{\beta}}$ for the case where $\epsilon > 1$.

3.3.2 The mean-variance approach

In this section we suppose that the sochastic uncertainties ν follow a centred distribution with a fixed variance denoted ϵ^2. In this case, we can write:

$$F_{MV} = (1-\beta)\mathbb{E}[f(x,\nu)] + \beta Var[f(x,\nu)] \tag{49}$$

$$= \beta(-a + ||x||^\beta) + (1-\beta)\epsilon^2(-b + ||x||^\beta)^2 \tag{50}$$

Here, as it was the case in the second example of type B uncertainties, the optimum of the variance and of the expectation does not coincide. Moreover, we can see that the optimum of the expectation does not depend on ν, we have always: $x^* = 0$. For the variance, we find the same optimum that in the case where $\epsilon > 1$ in the robust counterpart approach: $||x^*|| = b^{\frac{1}{b}}$.

3.3.3 The Average Value-at-Risk approach

In this case we assume that ν follows a standard normal distribution. In that case, we have, for $||x||^\beta \neq b$:

$$VAR_\alpha(f(x,\nu) = \inf\{t : \mathcal{P}(f(x,\nu) \leq t) \geq 1-\alpha\} \tag{51}$$

$$= \inf\{t : \mathcal{P}\left(\nu \leq \frac{t + a + ||x||^\beta}{|b - ||x||^\beta|}\right) \geq 1-\alpha\} \tag{52}$$

If we denote Φ the cumulative distribution function of the standard normal law, then:

$$VAR_\alpha(f(x,\nu) = \inf\{t : \Phi\left(\frac{t + a - ||x||^\beta}{|b - ||x||^\beta|}\right) \geq 1-\alpha\} \tag{53}$$

$$= -a + ||x||^\beta + \Phi^{-1}(1-\alpha)|b - ||x||^\beta| \tag{54}$$

And thus we obtain by integration:

$$CVAR_\alpha(f(x,\nu)) = -a + ||x||^\beta + \frac{1}{\alpha}|b - ||x||^\beta| \int_{1-\alpha}^{1} \Phi^{-1}(\tau)d\tau \tag{55}$$

If we note:

$$I(\alpha) = \frac{1}{\alpha}\int_{1-\alpha}^{1} \Phi^{-1}(\tau)d\tau \tag{56}$$

then, rearranging the term, we obtain:

$$CVAR_\alpha(f(x,\nu)) = \begin{cases} -a + bI(\alpha) + (1 - I(\alpha))||x||^\beta & \text{if } ||x||^\beta < b \\ -a - bI(\alpha) + (1 + I(\alpha))||x||^\beta & \text{if } ||x||^\beta > b \end{cases} \tag{57}$$

This formulation is very similar to the one in 47, and according to the value of $I(\alpha)$, we can obtain the exact same solutions that in the robust counterpart approach. Moreover, the fact that $1-\alpha$ is our robustness index allows us to understand the link between this index and the minimum that we obtain.

3.4 Class A uncertainties: function "f_2

In this section, we consider the robust minimization of:

$$f_2(x,\nu) = -a + \frac{(x_{n-1}+\nu)^2 + \sum_{i=1}^{n-2} x_i^2}{x_n^2 + b} + x_n^2,\ b > 0,\ x \in \mathbb{R}^n \tag{58}$$

with type A uncertainties.

3.4.1 The Robust Counterpart approach

The robust counterpart function of f_2 is:

$$F_{RC}(x,\epsilon) = -a + x_n^2 + \frac{\sum_{i=1}^{n-2} x_i^2}{x_n^2 + b} + \frac{1}{x_n^2 + b} \max_{\nu\in[-\epsilon,\epsilon]} (x_{n-1}+\nu)^2 \tag{59}$$

$$= -a + x_n^2 + \frac{\sum_{i=1}^{n-2} x_i^2}{x_n^2 + b} + \frac{(|x_{n-1}| + \epsilon)^2}{x_n^2 + b} \tag{60}$$

By calculation [13], the optimum is obtained at

$$x^* = \begin{cases} 0 & \text{if } \epsilon \leq b \\ (0,...,0,\pm\sqrt{\epsilon - b}) & \text{if } \epsilon > b \end{cases} \tag{61}$$

In fact, for $\epsilon \leq b$, the minimum is located at $x = 0$ and otherwise there are two minima located. According to the level of uncertainties the robust counterpart function changes from an unimodal to a bimodal function.

3.4.2 The Mean-Variance approach

In this section, the uncertaities are assumed to follow a centred distribution with a fixed variance denoted ϵ^2. Then:

$$F_{MV} = (1-\beta)\mathbb{E}[f(x,\nu)] + \beta Var[f(x,\nu)] \tag{62}$$

$$= (1-\beta)\left(-a + \frac{\sum_{i=1}^{n-1} x_i^2 + \epsilon^2}{x_n^2 + b} + x_n^2\right) + \beta\left(\frac{4x_{n-1}^2\epsilon^2 + 2\epsilon^4}{(x_n^2+b)^2}\right) \tag{63}$$

In case, where $\beta = 0$, the authors of [26] have shown that the F_{MV} function yields the same result as in the robust counterpart approach.

3.4.3 The Average Value-at-Risk approach

In this section, we suppose that ν follows a standard normal distribution, we have:

$$VAR_\alpha(f(x,\nu)) = \inf\{t : \mathcal{P}(f(x,\nu) \leq t) \geq 1-\alpha\} \tag{64}$$

$$= \inf\left\{t : \mathcal{P}\Big(-a + \frac{(x_{n-1}+\nu)^2 + \sum_{i=1}^{n-2} x_i^2}{x_n^2 + b} + x_n^2 \leq t\Big) \geq 1-\alpha\right\} \tag{65}$$

As in 3.3.3, we can rewrite the probability such that:

$$\mathcal{P}\Big((x_{n-1}+\nu)^2 \leq (t+a-x_n^2)(x_n^2+b)-r\Big) \text{ with } r=\sum_{i=1}^{n-2} x_i^2 \tag{66}$$

In the case where $(t+a-x_n^2)(x_n^2+b)-r < 0$, this probability is null. However, that contradicts the fact that probability must be greater than $1-\alpha$. Then, t will be choosen such that this quantity was non negative, that is why, we can write:

$$\mathcal{P}\Big(|x_{n-1}+\nu| \leq \sqrt{(t+a-x_n^2)(x_n^2+b)-r}\Big) \tag{67}$$

Then, we have two cases:

$$VAR_\alpha(f(x,\nu)) = \begin{cases} \inf\{t : \Phi(\sqrt{(t+a-x_n^2)(x_n^2+b)-r} - x_{n-1}) \geq 1-\alpha\} & \text{if } \nu + x_{n-1} \geq 0 \\ \inf\{t : \Phi(\sqrt{(t+a-x_n^2)(x_n^2+b)-r} + x_{n-1}) \geq 1-\alpha\} & \text{if } \nu + x_{n-1} < 0 \end{cases} \tag{68}$$

$$= \begin{cases} -a + x_n^2 + \frac{(\Phi^{-1}(1-\alpha)+x_{n-1})^2+r}{x_n^2+b} & \text{if } \nu + x_{n-1} \geq 0 \\ -a + x_n^2 + \frac{(\Phi^{-1}(1-\alpha)-x_{n-1})^2+r}{x_n^2+b} & \text{if } \nu + x_{n-1} < 0 \end{cases} \tag{69}$$

$$= -a + x_n^2 + \frac{(|x_{n-1}| + \Phi^{-1}(1-\alpha))^2 + \sum_{i=1}^{n-2} x_i^2}{x_n^2+b} \tag{70}$$

3.5 Convexification phenomenon: the 2D case

In this section, we try to apply the Average Value-at-Risk risk measure on a function of two variables. The test function used in this purpose is the Branin function:

$$f(x,y) = (y - 0.1291845091 \times x^2 + 1.591549431 \times x - 6)^2 + 9.602112642 \times \cos(x) + 10$$

We draw this function on Figure 6.

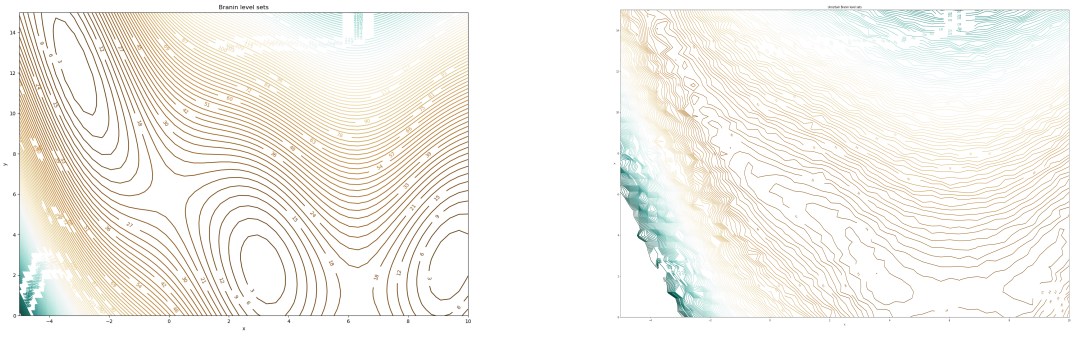

Figure 6: Branin function: certain (right) and uncertain (left) level set

Then, we assume that x and y are random variables following the following distributions: $X \sim \mathcal{N}(x, 1)$ and $Y \sim \mathcal{N}(y, 1)$. We apply the Average Value-at-Risk on this function to obtain the result on Figure 6. The convexification phenomenon appeared on it, nevertheless, the function does not seem become quasi convex as it was the case on the 1-D tests. However, we think that with a greater standard deviation, we could have got a quasi convex function.

3.6 Smoothing phenomenon

Another interesting phenomenon appears using the CVAR approach is the smoothing phenomenon. Indeed, let consider the following non smooth function:

$$f(x) = |x| + |x + 3| + |x - 3| \tag{71}$$

Assumiong that $X \sim \mathcal{N}(x, \sigma)$ with $\sigma \in [0.2, 0.1, 0.05, 0.02]$ and using $CVAR_\alpha$ with $\alpha = 0.9$, the graph obtained are presented on the Figure 7.

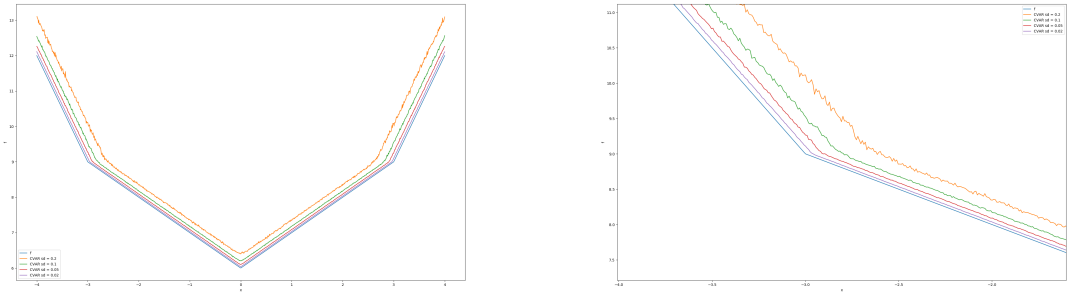

Figure 7: Exemple of non smooth function: full (left) and a zoom (right)

On this graph, it seems that the $CVAR_\alpha(f(X))$ function be smoother than the original function at the different point of non differentiability. This phenomenon is not very noticeable, nevertheless that is due to the small value of the standard deviation used here.

3.7 Concluding remarks on the numerical experiments

Thanks to the previous experiments, we may notice several things:

- The Average Value-at-Risk measure, which is a coherent risk measure (as showed in Section 2), seems to have an interesting behavior in practice compared to the mean measure and the mean-variance risk measure. Indeed, it creates no additional maxima and seems to convexity the function on which it is applied. The last result is particularly interesting in design optimization.
- We apply the Average Value-at-Risk risk measure on non smooth and non convex function, two phenomenons appear: convexification and smoothing. Even if it is less clear, this phenomenon appears also for function with multiple variables.

- Through, the different experiments, we have found three parameters which have an influence on this phenomenon: the value of σ, the value of the index of robustess α (through the β_2 parameter) and the form of the width of the different "gap" of the function.

The next section has for purpose to bring a theoretical base to this phenomenon of convexification.

4 Additional properties of CVAR(f)

During the different numerical experiments, two interesting phenomenon have been observed:

- A convexification of the function f: in all our numerical experiments, the $CVAR_\alpha$ function applied to f is locally quasi convex for a certain α.
- A smoothing of the function f: in the different numerical experiments, the $CVAR_\alpha$ function applied to f seems to be of class C^1.

In this section, this two phenomenon are described formally and proved.

4.1 The convexification phenomenon

We try to show the following result:

Proposition 4.1. *Let assume that:*

- *f is locally Lipschitz.*
- *The random variable X has a continuous probability distribution.*

Then, there exists α such that $CVAR_\alpha \circ f()$ be locally quasi convex.

Proof. The proof will be made in three steps: first prove that $CVAR_\alpha \circ f$ tends to $\sup f$ when $\alpha \to 1$, then prove that $\sup f$ is locally quasi convex and finally prove that a function tending to a quasi convex function is quasi convex from a certain rank.

First, the function $CVAR_\alpha \circ f$ verify the following inequalities, let $x \in \mathcal{D}_f$

$$VAR_\alpha(f(x)) \leq CVAR_\alpha(f(x)) \leq \sup\ f(x) \tag{72}$$

And we know that:

$$\begin{aligned} VAR_\alpha(f(x)) &= \inf\{t : \mathcal{P}(f(x) \leq t) \geq 1\} && (73)\\ &= \inf\{t : \mathcal{P}(f(x) \leq t) = 1\} && (74)\\ &= \sup f(x) && (75) \end{aligned}$$

Thus, by the squeeze theorem, we have :

$$\lim_{\alpha \to 1} CVAR_\alpha(f(x)) = \sup f(x). \tag{76}$$

Then, let $r > 0$, $\bar{x} \in \mathcal{D}_f$ and consider the ball $\mathcal{B}_r(\bar{x}) = \{x \in \mathcal{D}_f, ||x - \bar{x}|| < r\}$. Let prove that $\sup f(x)$ is locally quasi convex. For that, it is sufficient to prove that, $\forall a \geq 0$, the set:

$$A(x) = \{x \in \mathcal{B}_r(\bar{x}); \sup\ f(x) \leq a\} \text{ is convex.} \tag{77}$$

Let $x, y \in A(x)$, let $\lambda \in [0, 1]$, we have by continuity of f:

$$\sup\ f(\lambda x + (1-\lambda)y) \leq \lambda \sup\ f(x) + (1-\lambda) \sup\ f(y) = \sup f(x) \tag{78}$$

Finally, we must prove the following lemma:

Lemma 4.2. *Let f be a quasi convex function and f_α converge simply to f, then f_α is quasi convex from a certain rank.*

Indeed, let assume that $\forall \alpha$, f_α be not quasi convex, then there exist $x, y \in \mathcal{D}_f$ and $\lambda \in [0, 1]$ such that:

$$f_\alpha(\lambda x + (1-\lambda)y) > \max\{f(x), f(y)\} \tag{79}$$

Thus, passing this inequality to the limit, we find that f is not quasi convex, which contradicts the assumption.

□

4.2 The smoothing phenomenon

We try to show the following result:

Proposition 4.3. *Let assume that:*

- *f is locally Lipschitz.*
- *The random variable X has a continuous probability distribution and we denote H the cumulative distribution function of X.*
- *Some assumptions to be able to invert $\mathbb{E}$ and $\frac{\partial}{\partial X}$ operators.*

Then, there exists α such that $CVAR_\alpha \circ f()$ be differentiable.

Proof. This proof is not yet finished, it is just to give some ideas and ask about its correctness.

First, the $CVAR_\alpha \circ f()$ function may be written as:

$$CVAR_\alpha(f(X)) = t^* + \alpha^{-1}\mathbb{E}[(f(X) - t^*)_+]$$

with $t^* = VAR_\alpha(X) = H^{-1}(1-\alpha)$. Now, the differential of $CVAR \circ f()$ may be written as:

$$\frac{\partial}{\partial X} CVAR_\alpha(f(X)) = \alpha^{-1} \frac{\partial}{\partial X} \mathbb{E}[(f(X) - t^*)_+] \tag{80}$$

$$= \mathbb{E}[\frac{\partial}{\partial X}(f(X) - t^*)_+] \tag{81}$$

thanks to the third assumption. Moreover, knowing that the function $(f(\cdot) - t^*)_+$ is locally Lipschitz, it is differentiable almost everywhere in the neighbourhood of $x \in \mathbb{R}^n$. Now, let assume that $(f(\cdot) - t^*)_+$ was differentiable everywhere in the neighborhood of $x_0 \in \mathbb{R}^n$ except at x_0. Then, can we affirm that $\mathbb{E}[(f(x_0) - t^*)_+]$ is differentiable at x_0 if and only if the event $X = x_0$ has zero probability ? (as they do page 376 of [2]) ? □

5 Handling the constraints

Two communities handle constraint under uncertainties with two different names for the same principle at the origin: reliability based design optimization and chance constrained optimization.

5.1 Reliability based design optimization

Definition 5.1. *According to [6], reliability based design optimization (RBDO) is the concept ensuring that a solution of a problem with uncertainty is feasible with a probability greater or equal to a certain rate of safety.*

Mathematically, this amounts to solving the following optimization problem:

$$\begin{aligned} \underset{x,\, d}{\text{minimize}} \quad & f(X, d, P) \\ \text{subject to} \quad & g_j(X, d, P) \leq 0, \quad j = 1, \ldots, J, \\ & x^l \leq x \leq x^u, \\ & d^l \leq d \leq d^u. \end{aligned} \tag{82}$$

Here, X is a set of random vectors to be optimized. In the rest of this section, we will consider that X follows a normal distribution $N(\mu_x, \sigma_x)$ where μ_x is the mean vector and σ_x is the covariance matrix. This may not be the most common case in reality, but it is the easiest way to understand the different methods inherent in reliable optimization problems. Likewise, P is a vector of random parameters (which are not to be optimized) according to the normal distribution $N(\mu_p, \sigma_p)$. The vectors d correspond to the set of deterministic variables which are to be optimized. So, the stochasticity of the problem comes from the two sets of vectors: X and P. The problem is posed in such a way that X and d are the vectors to be optimized, nevertheless in RBDO μ_x and d are often the real decision variables. Finally,

we only consider inequality constraints because if an equality constraint implied X or P the problem might not have a solution. An example of reliable optimization is shown on the Figure 8.

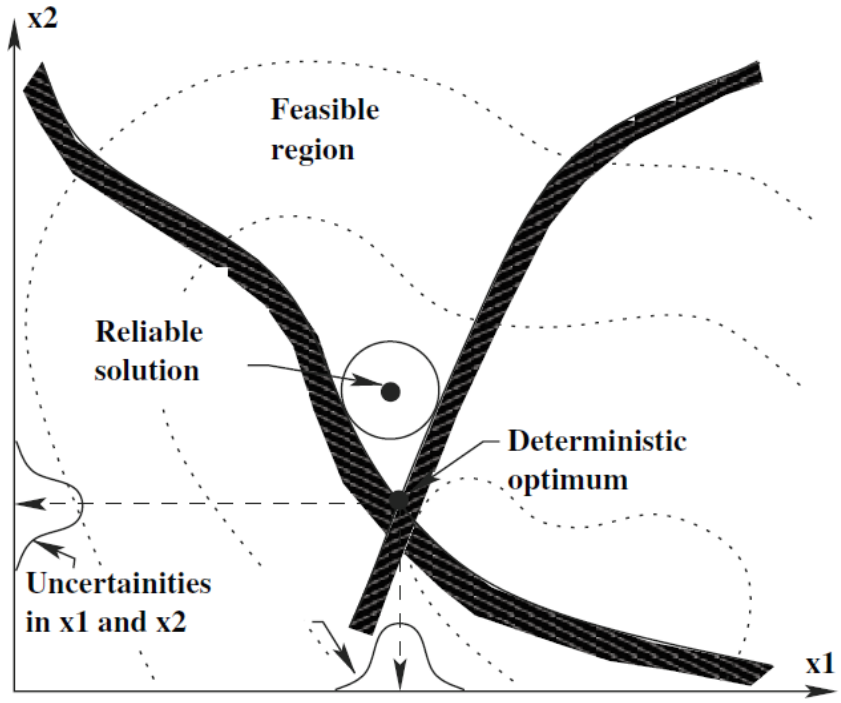

Figure 8: Example of relibility optimization process

To reach this solution, problem (1) can be transformed into a new optimization problem. Since the objective function f and the constraints g_j are probabilistic due to the randomness of the variables X and P, a new deterministic formulation can be written:

$$\begin{aligned} \underset{x,\,d}{\text{minimize}} \quad & f(\mu_x, d, \mu_p) \\ \text{subject to} \quad & P_j(g_j(x,d,p) \leq 0) \leq R_j \quad j = 1, \ldots, J, \\ & x^l \leq \mu_x \leq x^u, \\ & d^l \leq d \leq d^u. \end{aligned} \tag{83}$$

The term $P()$ corresponds to the probability of satisfying the constraint g_j. The quantity R_j is the desired reliability rate, it is fixed between 0 and 1. The transformation of the constraint g_j into a probabilistic constraint is a standard technique. Nevertheless, finding the law of probability for which a constraint is satisfied is a complex mathematical problem. Even more, when it comes to finding the joint probability law allowing to satisfy with a certain degree of reliability the set of probabilistic constraints. The next section shows different ways to effectively approximate this probability.

5.2 Tools to approximate a unique probabilistic constraint

Mathematically, the conversation for a solution X to be reliable with respect to the j th constraint can be written $(1 - P_j)$ with:

$$P_j(\mu_x, d, \mu_p) = \int_{g_j(\mu_x,d,\mu_p)<0} \phi_{\mu_x,\mu_p}(x, p)dxdp \tag{84}$$

where P_j is the probability of failure and ϕ_{μ_x,μ_p} is the joint probability density of (X, P). It is, in most cases, impossible to have an analytical formula of this integral for any constraint. This prompted the researchers to develop two approximation procedures: a statistical approximation by sampling and an optimization procedure to estimate the distance to the constraint.

5.2.1 Sampling approximation

In this procedure [6], given that we assume the joint probability distribution of X and P is known, we create N different samples. For each sample, the constraint g_j is evaluated and it is observed whether there is a violation or not. If in r_j case on N the constraint is not respected then the probability of failure is approximated by $P_j = \frac{r_j}{N}$ and therefore the probability constraint $P(g_j(x, d, p) \geq 0)$ can be replaced by:

$$1 - \frac{r_j}{N} \geq R_j \tag{85}$$

The advantage of this approach is its simplicity of implementation and the fact that it can be extended to several constraints very easily (you just have to see if the samples are feasible with respect to all the constraints). Nevertheless, this procedure is very costly in terms of computation time, in fact to have a precision of the order of 10^{-6} a number of samples of the order of $N = 10^6$ is needed. Techniques have been used to reduce this cost but this may not be sufficient if high reliability is desired. This process may be ipmproved by using Latin Hypercube sampling (LHS) [18], Importance Sampling (IS) [14] or Directional Sampling [8].

5.2.2 Optimization approximation

The idea, on which the various methods of approximation by optimization are based, is common to all the methods: seek to determine a point on the frontier of the constraint which is closest to the solution. This point is called MPP (Most Probable Point) [15]. To find this point, a reliability index β_j is calculated as follows:

$$\beta_j = \phi^{-1}(R_j) \tag{86}$$

with the probability of failure P_j which is approximated by:

$$P_j = \phi(-\beta_j) \tag{87}$$

Where $\phi()$ is the probability density of the standard normal distribution. β_j can be seen as the margin of safety that is wished when the system is in its average state. Formulas 86 and 87 are correct under two conditions:

- That the elements of X are uncorrelated and that they follow a standard normal distribution. To establish this condition, we transform the vector X into a random vector according to a standard normal distribution U by applying the Rosenblatt transformation [24] for example.
- That the boundary of the failure constraint is an hyperplane, which is very rarely the case. To fulfill this condition, we approximate the failure surface by its hyperplane tangent at the MPP point in the U space using its first-order expansion of its Taylor series (this is called a FORM method (First Order Reliability Method)).

Our matter therefore comes down to calculating the MPP, there exists several procedures as it is exposed below.

5.2.2.1 Performance measure approach (PMA)

To find the MPP with the PMA method, the following optimization problem is solved [5]:

$$\begin{aligned} &\text{minimize} \quad G_j(U) \\ &\text{subject to} \quad ||U|| = \beta_j. \end{aligned} \tag{88}$$

with $G_j()$ the constraint g_j translated into the space U (typically by taking $\phi^{-1}(g_j())$). The solution of 88 is a point U^* which is on the circle of radius β_j and which minimizes $G_j(U)$. In the initial problem 83, the probabilistic constraint is then replaced by the following constraint:

$$G_j(U^*) \geq 0. \tag{89}$$

Figure 9 illustrates the operation of PMA in a two-dimensional case.

Although the problem 88 involves an equality constraint, a custom algorithm can be used to consider only the solutions on the $||U|| = \beta_j$ hyper surface. By using this algorithm the search for solutions becomes faster (since they are all feasible).

5.2.2.2 Reliability Index Approach

In this method, the following optimization problem is solved [5]:

$$\begin{aligned} &\text{minimize} \quad ||U|| \\ &\text{subject to} \quad G_j(U) = 0. \end{aligned} \tag{90}$$

Here, the MPP is calculated by finding the point on the surface of the constraint in the space U which is closest to the origin (which is the mean state of the system because all the random

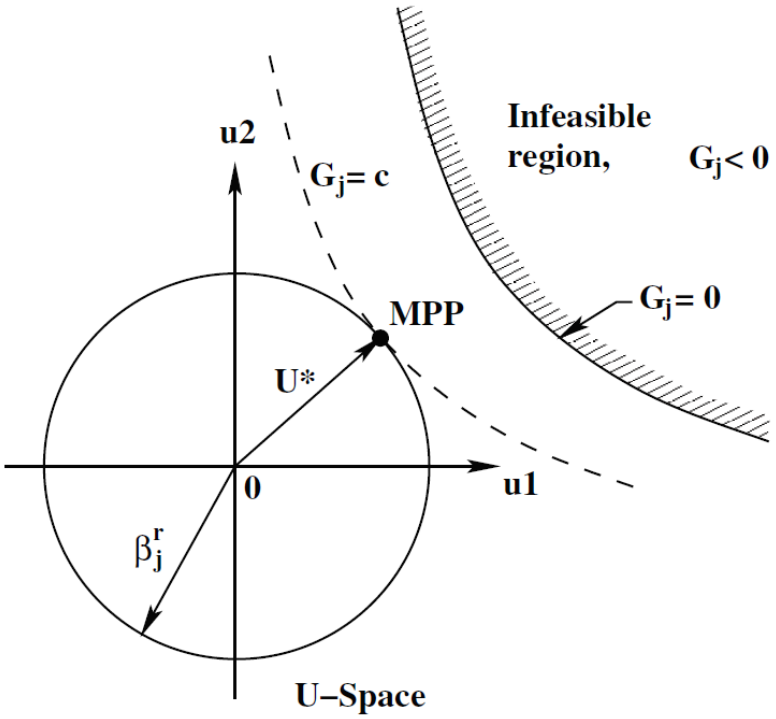

Figure 9: Example of PMA process [6]

variables of x have been translated into variables of u according to standard normal distributions). The optimal point found U^* is used to replace the probabilistic constraint in formula (2) as follows:

$$||U|| \geq \beta_j. \tag{91}$$

An illustration of this method is given in Figure 3. This method is more expensive than the PMA method since it involves an equality constraint that cannot be managed as in the PMA approach because it depends on the problem. But an advantage of this method is that it allows through U^* to have an approximate value of the distance of the solution to the constraint, which can be interesting in some cases.

5.3 Methods for reliability based design optimization

The methods described above for measuring the reliability of a solution have been inserted in several different ways into the optimization algorithm. Here is an overview of the most common methods.

5.3.1 Double-Loop Methods

The double-loop method [17] uses two loops: an outer loop to optimize the original problem given in 83 with (X, d) as decision vectors. Then, for each solution considered by the outer

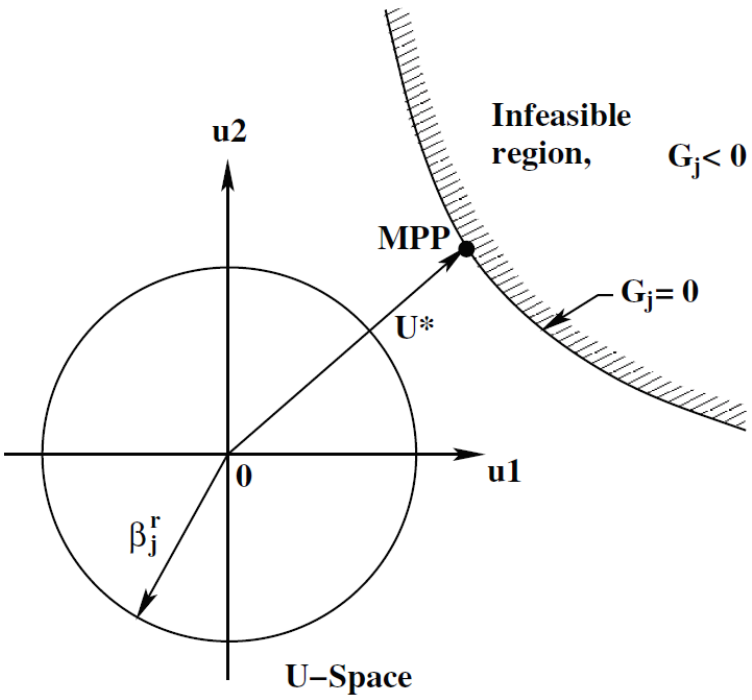

Figure 10: Example of RIA process [6]

loop, a second loop (the inner loop) is launched and the probabilistic constraint is solved thanks to another optimization problem using PMA or RIA. This method is very costly computationally speaking.

5.3.2 Single-Loop Methods

The single-loop [17] method combines the two optimization tasks together. It does not seek to find the optimum for the probabilistic constraint but constitutes a procedure allowing to find approximately the true MPP. For example in [17], the authors replace the original probabilistic constraint by:

$$g_j(\bar{x}, \bar{p}, d) \geq 0 \tag{92}$$

avec :

$$\bar{x} = \mu_x - \beta_j \sigma \frac{\nabla_x g_j}{\sqrt{||\nabla_x g_j||^2 + ||\nabla_p g_j||^2}} \tag{93}$$

$$\bar{p} = \mu_p - \beta_j \sigma \frac{\nabla_p g_j}{\sqrt{||\nabla_x g_j||^2 + ||\nabla_p g_j||^2}} \tag{94}$$

As the method is based only on an approximation, it often has difficulty finding precise results, but it is computationally faster than the double-loop method.

5.3.3 Decoupled Methods

In this method [10], two optimization problems (internal and external) are started one after the other. The decoupled methods have shown a good compromise between the two previous methods mentioned above. The principle is as follows: the algorithm starts by finding the optimal deterministic solution in the search space (without considering any uncertainty on X or on P and using the mean of X as the decision variable). Then, the MPP is searched using the PMA or RIA method. In the next iteration, each constraint is moved according to the MPP point found in the previous iteration and the deterministic optimization of the moved problem is solved. A pseudo code of a decoupled method (SORA) and an illustration of the method can be found in Figure 11.

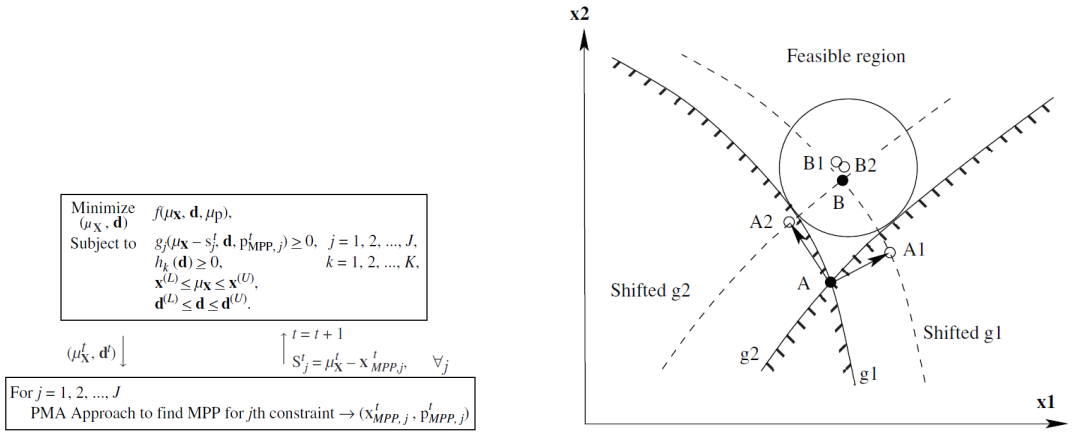

Figure 11: Pseudo code of SORA algorithm (left) and illustration (right) [6]

5.4 Handling several constraints

Ideally, the reliability of a solution should be calculated by considering the effect of all constraints. However, the methods we presented above (RIA and PMA) calculate the MPP for one constraint at a time. There are several ways to extend this approach to multiple constraints, we present two below:

5.4.1 The closest constraint

The simplest way to consider the problem is to determine the probability of failure for each constraint and to calculate the following bounds for the probability of failure according to the set of constraints P_F:

$$\max_j P_j \leq P_F \leq \min(1, \sum_j P_j) \tag{95}$$

Intuitively, the greater the probability of failure for a constraint, the closer the constraint is to the solution. Therefore, the lower bound is the probability of failure of the closest constraint and is an underestimation of the probability of failure under all constraints. The upper bound of P_F is the case where none of the constraints overlap, and it is therefore an overestimation in the other cases of the probability P_F. This solution is a very simple estimate which can lead to quite substantial errors but its advantage is the speed of the computation times.

5.4.2 Ditlevsen bounds

More precise bounds can be given [7], this requires calculating all the failure probabilities individually, the failure probability with respect to all the P_F constraints can then be bounded as follows:

$$P_1 + \sum_{i=1}^{J} \max(0, P_i - \sum_{j=1}^{i-1} P_{ij}) \leq P_F \leq \sum_{i=1}^{J} P_i - \sum_{i=2}^{J} \max_{j|j<i} P_{ji}. \tag{96}$$

The formula depends on the ordering of the constraints but usually, the P_i are sorted in descending order. So P_1 and P_j correspond respectively to the greatest and the smallest probability of failure. The joint probability P_{ij} according to the ith and jth constraint is given by the distribution function of the following bivariable normal distribution:

$$P_{ij} = \Phi(-\beta_j, -\beta_i, \rho_{ij}) \tag{97}$$

and the correlation coefficient rho_{ij} is given by:

$$\rho_{ij} = \frac{< u_j^*, u_i^* >}{||u_j^*||||u_i^*||} \tag{98}$$

with u_j^* the MPP in the space U for the jth constraint computed for the solution X. The inverse cosine of ρ_{ij} indicates the angle between the two u^* vectors. Figure 12 illustrates this procedure.

We make a linear approximation of each constraint at each MPP. Then once the MPPs are calculated for all the stresses, the correlation coefficients ρ_{ij} are the cosines of the angle formed by the tangents to the MPPs of the jth and th constraints. The Ditvelsen bounds are more precise than those given by the nearest constraint method, but this method involves calculating all pairs of joined probabilities of failure. In problem (2), the probabilistic constraint is then replaced by the upper bound of Ditvelsen, we thus obtain with the desired reliability R:

$$1 - \sum_{i=1}^{J} P_i - \sum_{i=2}^{J} \max_{j|j<i} P_{ji} \geq R \tag{99}$$

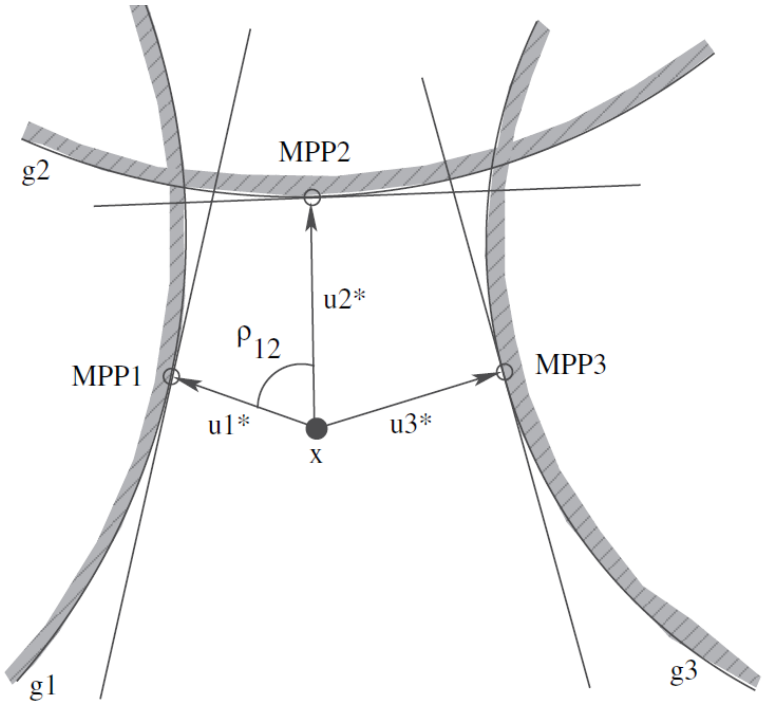

Figure 12: Process allowing to calculate the different correlation coefficients [6]

5.5 Advantages and drawbacks of RBDO

RBDO is a well studied subject and the different parts above are just an exposition of the main methodologies used. However, these sections allow to understand the principle and the assumptions made in the RBDO framework. Here is presented the advantages and the drawbacks of this framework. First, the main advantages are:

- These methods have been developed for design optimization, thus they does not use analytical properties of the constraints. In fact, in these methods the convexity or the differentiability of the constraints are not supposed. The methods are thus adapted to engineering design.
- The methods are intuitive, the different figures in the previous sections allow to understand quickly the different concept and facilitate the implementation.

And the main drawbacks are:

- To my knowledge the Rosenblatt transformation necessitates to know the distribution of the random variables in order to pass in the standard normal space.
- The approximation made using the FORM or SORM process may be really inaccurate in a real engineering problem.
- This subject is very well studied and that lets few ways of improvement.

5.6 Chance constrained optimization

6 Conclusion

There are several things to do:

- Complete the proof of smoothing phenomenon.
- Complete the chance constrained optimization part and explain what we propose to do according to the advantages and drawbacks of the two methods presented.
- Add the third part about the application of our method in the MDO framework.

References

[1] A. Nemirovski A. Ben Tal, L. El Ghaoui. *Robust Optimization*. Princeton University Press, 41 william street, Princeton, New Jersey,USA, 2009.

[2] D. Dentcheva A. Shapiro, A. Ruszczynski. *Lecture on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, USA, 2009.

[3] Carlo Acerbi, Claudio Nordio, and Carlo Sirtori. Expected shortfall as a tool for financial risk management, 2001.

[4] Wei Chen, Janet Allen, Kwok-Leung Tsui, and Farrokh Mistree. A procedure for robust design: Minimizing variations caused by noise factors and control factors. *ASME Journal of Mechanical Design*, 118:478–485, 1996.

[5] Thomas A Cruse. *Reliability-based mechanical design*, volume 108. CRC press, 1997.

[6] K. et al. Deb. Reliability-based optimization using evolutionary algorithms. *Evolutionary Computation, IEEE Transactions on 13.5*, pages 1054–1074, 2009.

[7] Ove Ditlevsen. Narrow reliability bounds for structural systems. *Journal of structural mechanics*, 7(4):453–472, 1979.

[8] Ove Ditlevsen and Peter Bjerager. Plastic reliability analysis by directional simulation. *Journal of engineering mechanics*, 115(6):1347–1362, 1989.

[9] Liu Du, Kyung K Choi, and Ikjin Lee. Robust design concept in possibility theory and optimization for system with both random and fuzzy input variables. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 48078, pages 1183–1193, 2007.

[10] Xiaoping Du and Wei Chen. Sequential optimization and reliability assessment method for efficient probabilistic design. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 36223, pages 871–880, 2002.

[11] Artzner P. Delbaen F. and Heath D. Eber J.M. Thinking coherently. *RISK*, 10(11), 1997.

[12] Batill S. et al. Gu X., Renaud J. Worst case propagated uncertainty of multidisciplinary systems in robust design optimization. *Struct Multidisc Optim*, 20:190–213, 2000.

[13] B. Sendhoff H. G. Beyer. Robust optimization - A comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33-34):3190–3218, 2007.

[14] Alf Harbitz. An efficient sampling method for probability of failure calculation. *Structural safety*, 3(2):109–115, 1986.

[15] Abraham M Hasofer and Niels C Lind. Exact and invariant second-moment code format. *Journal of the Engineering Mechanics division*, 100(1):111–121, 1974.

[16] J. Helton. Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *Journal of Statistical Computation and Simulation - J STAT COMPUT SIM*, 57, 01 1996.

[17] Jinghong Liang, Zissimos P Mourelatos, and Jian Tu. A single-loop method for reliability-based design optimization. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 46946, pages 419–430, 2004.

[18] Daniel H Loughlin and S Ranji Ranjithan. Chance-constrained genetic algorithms. In *Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation-Volume 1*, pages 369–376, 1999.

[19] P. Y. Papalambros M. Kokkolaras, Z. P. Mourelatos. Impact of uncertainty quantification on design: an engine optimisation case study. *International Journal of Reliability and Safety*, 1, 2006.

[20] AH Nuttal. Some integrals involving the qm function. *IEEE Trans. Inform. Theory*, pages 95–96, 1975.

[21] Laura Picheral. *Contribution a la conception preliminaire robuste en ingenierie de produit*. PhD thesis, Laboratoire G-scop, 2013.

[22] A. J. Qureshi, J. Y. Dantan, J. Bruyere, and R. Bigot. Set based robust design of mechanical systems using the quantifier constraint satisfaction algorithm. *Engineering Applications of Artificial Intelligence*, 23(7):1173 – 1186, 2010.

[23] Uryasev S. Rockafellar R. T. Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2000.

[24] Murray Rosenblatt. Remarks on a multivariate transformation. *The annals of mathematical statistics*, 23(3):470–472, 1952.

[25] Chen Liang Saideep Nannapaneni and Sankaran Mahadevan. Bayesian Network Approach to Multidisciplinary, Multi-Objective Design Optimization under Uncertainty. *AIAA Journal*, 2018.

[26] Bernhard Sendhoff, Hans-Georg Beyer, and Markus Olhofer. On noise induced multimodality in evolutionary algorithms. In *Proc. Asia-Pacific Conf. Simulated Evol. Learn*, volume 1, pages 219–224, 2002.

[27] A. Chateauneuf Y. Aoues. Benchmark study of numerical methods for reliability-based design optimization. *Structural and Multidisciplinary Optimization*, 41:277–294, 2010.

[28] Y.Sun, A.Baricz, and S.Zhou. On the monotonicity, log-concavity, and tight bounds of the generalized marcum and nuttall q-functions. *IEEE Transactions on Information Theory*, 56(3):1166–1186, 2010.