Robust Formulation

Romain Couderc

July 7, 2020

In this document, we discuss the subject of the robustness regarding the objective function. It was the first purpose of my research project. First, we present what is robustness and the different type of uncertainties. Secondly, we present some theoretical aspects about the new measure used to take care of the robustness of the objective function. Then, we present some 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

The 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 question arise:

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

We will answer this questions in the next subsections.

1.1 Necessity of robust optimization

When you optimize a problem, you search to find the right design parameters with the highest possible precision. However, even if you make great effort in the modelling and the conception of your 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 factor of this difference [9]:

- The desire design parameters are the solution of an optimization problem having an objective function and constraints. Nevertheless, these objective function and constraints are numerical modelling of the 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 match 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 too great cost to design with the desire 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 you take into account this fluctuation, there is always some parameters, sometimes even unknown by the designer, which are not take 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 has a life time and may be replaced or repaired during the life time of the final product. That can bring some variations in the behavior of the product which are not forecast by the numerical model.

So, as we can see, there are numerous factors which may altered the behavior of a final product. Used a robust optimization in the early stage of the design of a product may avoid bad surprises in the final product behavior and the cost induced to repaired them.

1.2 Class of uncertainties

There are different ways to class the source 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 [9], we can find the three main class of uncertainties commonly used:

- **The uncertainties of type A**
  These uncertainties represent the uncertainties due to the environmental parameters of the system as the temperature, the pressure or the humidity rate. If we call f the behavior of the system and x the design variables, then to take into account this parameter we must add random variables ν, the system is than modelled by:
  $$f = f(x, \nu). \tag{1}$$
  This type of incertitude are also called "Type I variations" in [6].
- **The uncertainties of type B**
  These uncertainties correspond to the ability of the manufacturer to build the desire design at a certain degree of accuracy. Indeed, design a product with a very high precision may be impossible or very expensive. That is why, there exists often a tolerance in the

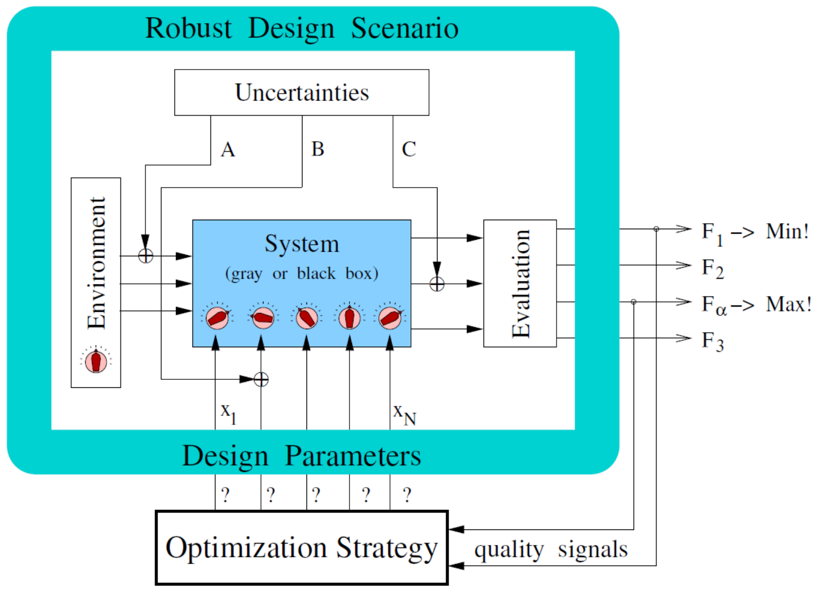

Figure 1: A robust design scenario with three types of uncertainties A, B and C

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 [6] and "Precision Error" in [8].

- **The uncertainties of type C**
  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 [8].

Unlike in [9], we do not create another category of uncertainties for the feasibility field of the function. In fact, each of the previous uncertainties may be apply to constraints g instead of objective function f. Besides, we present the three kinds of uncertainties separately, but it is obvious that in practice a system may combined the three kinds of uncertainties at the same time.

1.3 Models of uncertainties

Once we have classify the different kinds of uncertainties, we may now interest us how models these uncertainties. In the literature, there are basically three ways to model the uncertainties:

- Using deterministic interval [14], for each uncertain variable, we state a domain in which the variable can vary.
- Using a probabilistic framework [11], for each uncertain variable, we define a probabilistic law.
- Using the possibility framework [11] [1], for each uncertain variable, we define a fuzzy set.

Each of this model is quite different, to know which model use in which situation is not obvious. In mechanical engineering, they made another classification of the uncertainties allow to choose between the different models. They distinguish:

- The aleatory or random uncertainties [10] 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. However, the fact that these uncertainties are known for a long time allow to know their behavior thanks to a probability law for instance. The probabilistic framework is often used in that case.
- The epistemic uncertainties in the contrary represent the lack of knowledge of the manufacturer. These uncertainties could be reduce almost totally 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 case, to model the uncertainties with a possibility or deterministic framework seems to be the most appropriate.

In our case, we are going to use the probabilistic framework because of the strong mathematical background it offers.

1.4 Current methods in robust optimization

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

1.4.1 The Robust Counterpart Approach

This approach was developed mainly by the authors of [2] and is very popular in the field of robust optimization thanks to its serious mathematical background. The concept is simple, given an objective function $f(x)$ we want to optimize, then we define the Robust Counterpart of this function:

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

where $\mathcal{B}(x, \epsilon)$ is a neighborhood of 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, indeed we have 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 type B uncertainties. However, we can also easily apply it with uncertainties of type 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 type 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 type 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 that is 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 a lot advantages to use 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 black box optimization where the objective function and the constraints are very often the results of numerical simulations.

1.4.2 The possibilistic uncertainties approach

In the previous section and even more in the next section, the uncertainties are modelled on base of 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 you treat epistemic uncertainties. To treat this type of uncertainties, there are commonly two ways:

- Either, you use a probabilistic model that you improve iteratively with new knowledge, using bayesian statistic for instance.
- Or you use fuzzy sets and evidence theory.

The idea behind the fuzzy set methodology is to associate at each value of the objective or constraints functions a measure of a 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 associated with a measure of 0. Between the two design, we can assign intermediate value. In 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 unfeasible $\mu_c(x) = 0$ or intermediate $\mu_c(x) \in]0, 1[$. The problem is then to find the design x^* such that:

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

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.

1.4.3 The probabilistic risk measure approach

A last approach to handle the uncertainties is the probabilistic one. A strong hypothesis in this approach is you suppose that the distribution of the random variables is known. In the two previous approaches you need less information on the random variables, the bound in the first case and nothing in the second one. Nevertheless, in context of black box 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: there are an infinity of minima. However, in our mind we know that we will optimize in sens 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:

- We can 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 [13].

- We can see that a multi-objectif 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 [16].

We present these previous cases with the uncertainties of type B but it works for any types 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 [17] 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) [7] to handle it.

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 [3]. Nevertheless, this measure has been called Conditional Value-at-Risk in [15] and is equivalent to Expected Shortfall in [4]. 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 [5] 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 [15].

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:*

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

In [3], 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:*

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

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 [3].

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

$$AVAR_\alpha(Z) = CVAR_\alpha(Z). \quad (17)$$

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 [4]: 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 [4]).

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 [3] 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 [3]. 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 [3]. □

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 [3]), 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 [3].

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 see the effect of the previous theoretical results, we are going to test different risk measures. We compare three different measures of uncertainties on test functions mainly from [9]. These measures are Robust Counterpart, mean-variance measure and the Average Value-at-Risk measure. We test for different class of uncertainties. When it is possible we give the robust minimum according to the measure use and in case where it is relevant we provide some figure in 1 and 2 dimensions to show the result.

3.1 Type 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 type B, we have:

$$\begin{aligned}
F_{RC}(x;\epsilon) &= \sup_{||\delta||\leq\epsilon} (x+\delta)^2 && (28)\\
&= \sup_{||\delta||\leq\epsilon} \left(||x||^2 + 2 < x,\delta > + ||\delta||^2\right) \text{ with } < \cdot,\cdot > \text{ the scalar product.} && (29)\\
&= (||x|| + \epsilon)^2 && (30)
\end{aligned}$$

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:

$$\begin{aligned}
\mathbb{E}[f(x+\delta)] &= \mathbb{E}[\sum_{i+1}^{n}(x_i+\delta_i)^2] && (31)\\
&= \sum_{i=1}^{n} \mathbb{E}[x_i^2 + 2\delta_i x_i + \delta_i^2] \text{ by linearity of the expectation} && (32)\\
&= \sum_{i=1}^{n} x_i^2 + 1 && (33)\\
&= ||x||^2 + n && (34)
\end{aligned}$$

And for the variance we have:

$$Var[f(x+\delta)] = 4||x||^2 + 2n \tag{35}$$

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:

$$VAR_\alpha(f(x+\delta)) = \inf\{t : \mathcal{P}(\sum_{i=1}^{n}(x_i+\delta_i)^2 \leq t) \geq 1-\alpha\} \tag{36}$$

Since [9], we know also that:

$$Z = \sum_{i=1}^{n}(x_i+\delta_i)^2 \sim \mathcal{X}_n^{\prime 2}(||x||^2) \tag{37}$$

where $\mathcal{X}_n^{\prime 2}(||x||^2)$ 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 $\mathcal{X}_n^{\prime 2}(||x||^2)$ is equal to [12]:

$$\mathcal{P}(Z \leq t) = 1 - Q_{\frac{n}{2}}(||x||, \sqrt{t}) \tag{38}$$

with Q the Marcum-Q function. Moreover, in [18], 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:

$$\mathcal{P}(\sum_{i=1}^{n}(x_i+\delta_i)^2 \leq t) < \mathcal{P}(\sum_{i=1}^{n}(\delta_i)^2 \leq t) \tag{39}$$

and thus that:

$$\inf\{t : \mathcal{P}(\sum_{i=1}^{n}(x_i+\delta_i)^2 \leq t) \geq 1-\alpha)\} \geq \inf\{t : \mathcal{P}(\sum_{i=1}^{n}(\delta_i)^2 \leq t) \geq 1-\alpha\} \tag{40}$$

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

$$CVAR_\alpha(f(x+\delta)) \geq CVAR_\alpha(f(\delta)) \tag{41}$$

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

3.2 Type 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:

$$f(x) = 1 - \exp\left(-\frac{1}{2}x^2\right) - \exp\left(-\frac{1}{50}(x-7)^2\right) - \exp\left(-5(x+5)^2\right). \tag{42}$$

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 $x = 7$.

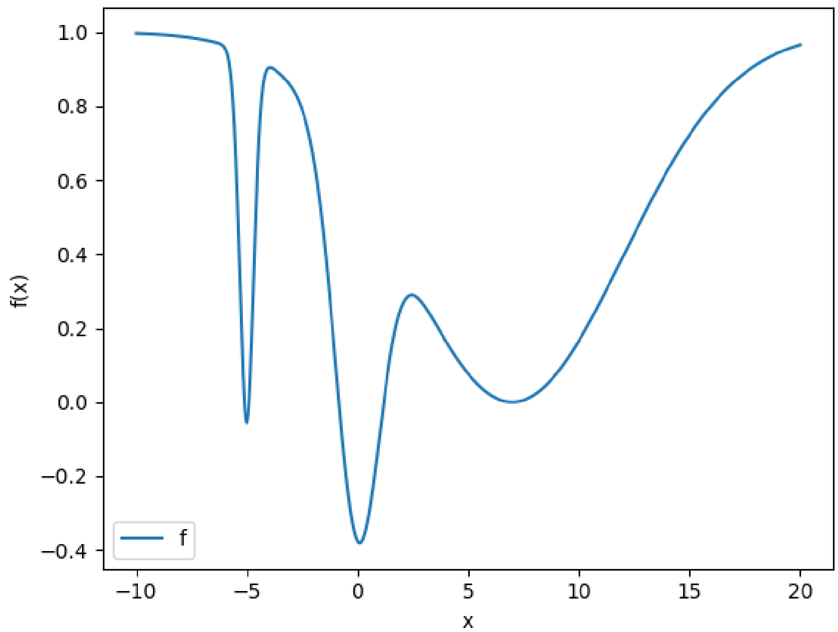

Figure 2: Test function

3.2.1 The Robust Counterpart approach

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

$$F_{RC}(x;\epsilon) = \sup_{|\delta|\leq\epsilon}(f(x+\delta)) \tag{43}$$

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:

$$F_{MV}(x) = (1-\beta)\mathbb{E}[f(x+\delta)] + \beta Var[f(x+\delta)] \tag{44}$$

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:

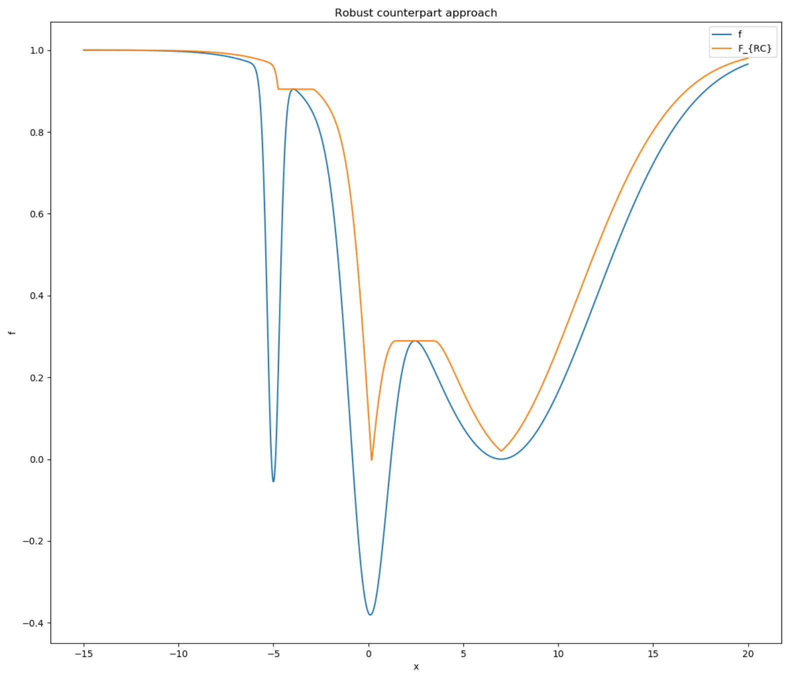

Figure 3: Robust Counterpart Approach

- 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.

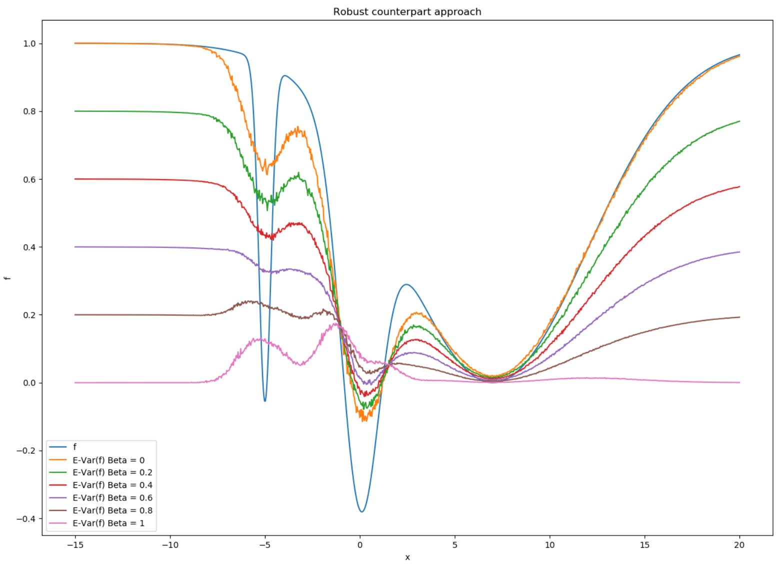

Figure 4: Mean-Variance approach for different values of beta

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.

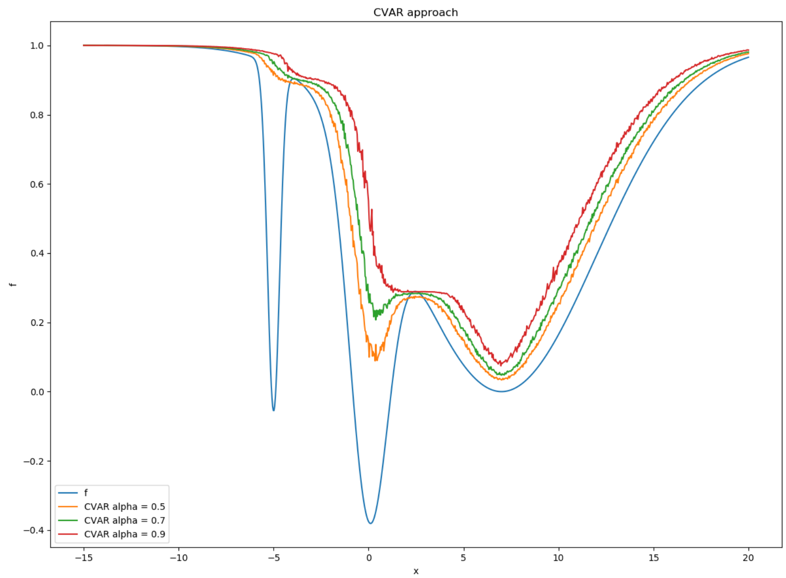

Figure 5: Average Value-at-Risk measure

3.3 Type A uncertainties: the sphere model

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

$$f(x, \nu) = -a + (\nu + 1)||x||^{\beta} - b\nu, \ \beta > 0, \ \nu \in \mathbb{R} \tag{45}$$

In the contrary to [9] 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:

$$F_{RC}(x, \epsilon) = \sup_{|\nu| \leq \epsilon} f(x, \nu) \tag{46}$$

$$= \sup_{|\nu| \leq \epsilon} \{-a + ||x||^{\beta} + \nu(||x||^{\beta} - b)\} \tag{47}$$

$$= \begin{cases} -a + b\epsilon + (1-\epsilon)||x||^{\beta} & \text{if } ||x||^{\beta} \leq b \\ -a - b\epsilon + (1+\epsilon)||x||^{\beta} & \text{if } ||x||^{\beta} > b \end{cases} \tag{48}$$

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:

$$\min_{x} F_{RC} = \begin{cases} -a + b\epsilon & \text{if } 0 \leq \epsilon \leq 1, \\ -a + b & \text{if } \epsilon > 1. \end{cases} \tag{49}$$

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:

$$F_{MV} = (1-\beta)\mathbb{E}[f(x,\nu)] + \beta Var[f(x,\nu)] \tag{50}$$
$$= \beta(-a + ||x||^\beta) + (1-\beta)\epsilon^2(-b + ||x||^\beta)^2 \tag{51}$$

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 $||x||^\beta \neq b$:

$$VAR_\alpha(f(x,\nu) = \inf\{t : \mathcal{P}(f(x,\nu) \leq t) \geq 1-\alpha\} \tag{52}$$
$$= \inf\{t : \mathcal{P}\left(\nu \leq \frac{t + a + ||x||^\beta}{|b - ||x||^\beta|}\right) \geq 1-\alpha\} \tag{53}$$

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

$$VAR_\alpha(f(x,\nu) = \inf\{t : \Phi\left(\frac{t + a - ||x||^\beta}{|b - ||x||^\beta|}\right) \geq 1-\alpha\} \tag{54}$$
$$= -a + ||x||^\beta + \Phi^{-1}(1-\alpha))|b - ||x||^\beta| \tag{55}$$

And thus we obtain by integration:

$$CVAR_\alpha(f(x,\nu)) = -a + ||x||^\beta + \frac{1}{\alpha}|b - ||x||^\beta| \int_{1-\alpha}^{1} \Phi^{-1}(\tau)d\tau \tag{56}$$

If we note:

$$I(\alpha) = \frac{1}{\alpha}\int_{1-\alpha}^{1} \Phi^{-1}(\tau)d\tau \tag{57}$$

then, rearranging the term, we obtain:

$$CVAR_\alpha(f(x,\nu)) = \begin{cases} -a + bI(\alpha) + (1 - I(\alpha))||x||^\beta & \text{if } ||x||^\beta < b \\ -a - bI(\alpha) + (1 + I(\alpha))||x||^\beta & \text{if } ||x||^\beta > b \end{cases} \tag{58}$$

This formulation is very similar to the one in 48, 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 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.

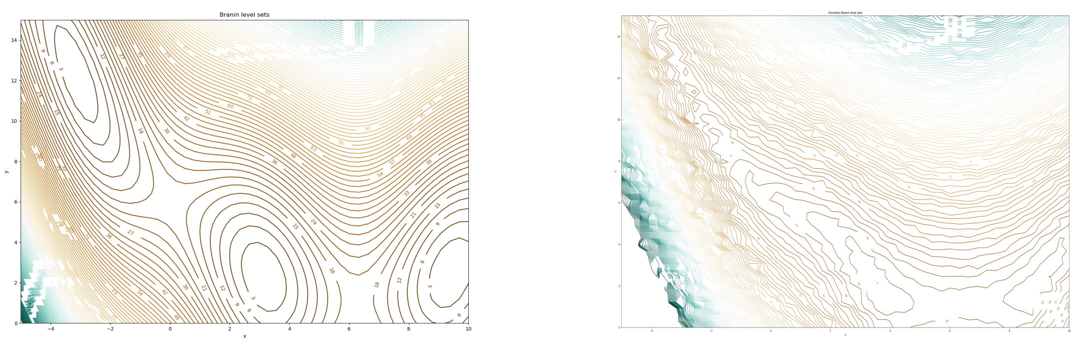

Figure 6: Branin function: certain (right) and uncertain (left) level set

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.

3.4.1 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 and the phenomenon of convexification is still observed. 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 Proof of convexification

This section demonstrates the convexification phenomenon of the Average Value-at-Risk in the $1-\mathrm{D}$, i.e. the case where the function f is univariate. First, we define a notion that we will use in the following. The first definition concerns what we call a "gap" in a function. Informally, we could say that it is the width of the gap above a minimum.

Definition 4.1. *Consider $a, b \in \mathbb{R}$ and $f : [a, b] \to \mathbb{R}$ a continuous function. We call:*

- *$M = \{x_1, ..., x_n\}$ the set of local maxima of f on $[a, b]$ ordered such that $f(x_1) \geq f(x_2) \geq ... \geq f(x_n)$.*
- *a gap of level ℓ_1 between $[x_1, x_2]$ the interval $I_{f(x_p)}(x_1, x_2) = [x_p, x_2]$ such that x_p be a solution of:*

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & x \\ \text{subject to} \quad & f(x) \geq f(x_2), \\ & x \in [x_1, x_2] \end{aligned} \tag{59}$$

- *L_p the union of the gaps of level ℓ_p.*
- *$L = \{L_1, ..., L_m\}$ all the different level of gaps sorted in descending order (see Figure 8).*
- *ρ the measure function defined in 16.*

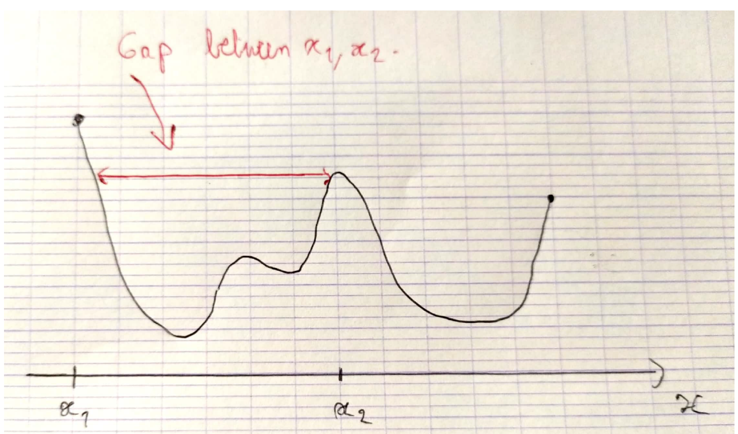

Figure 7: Example of a gap between two local maxima

Let begin by introduce the following lemma that ensures that $\rho \circ f$ is non increasing on a gap:

Lemma 4.2. *Let $\alpha \in [0, 1]$, let suppose that:*

- *X is a unidimensionnal random variable having a mean of 0 and a variance equal to $\sigma^2 < +\infty$.*
- *$a = x_1 \in M$ is a global maximum of f, x_2 is such that $x_2 > x_1$.*
- *$\max_{x \in I(a,x_2)} \{\mathbb{P}[x + X \in I(a, x_2)]\} \leq 1 - \alpha$.*

Then, $\rho \circ f$ is non increasing on $[a, x_2]$.

Proof. We study the following probability:

$$p(x) = \mathbb{P}(x + X \in I(a, x_2)) \text{ for } x \in I(a, x_2) \tag{60}$$

We decompose the interval $[a, x_2]$ as an union of disjoint interval:

$$[a, x_2] = [a, x_p] \cup I(a, x_2) = [a, x_p] \cup [x_p, x_2] \tag{61}$$

and we analyse the behavior of ρ, on each of these intervals.

- For $x \in [a, x_p]$, there are two cases:
  - If $f(a) = f(x_2)$, then $[a, x_p] = \{a\}$ thus there is no more study to do.
  - Otherwise, given that $f(x_p) = f(x_2)$ with $f(x_2)$ the second highest maximum, then f is decreasing on $[a, x_p]$ and thus by coherence of ρ, we have $\rho \circ f$ which is also decreasing on $[a, x_p]$.
- For $x \in [x_p, x_2]$, the third assumption of 4.2 ensures that $\forall x \in I(a, x_2), \exists \alpha_x \in [\alpha, 1]$ such that:

  $$\mathbb{P}(x + X \in I(a, x_1)) = 1 - \alpha_x$$

  We can deduce that:

  $$\alpha_x = \mathbb{P}(x + X < x_p) + \mathbb{P}(x + X > x_1)$$

  Let take $x, y \in I(a, x_1)$ such that $x \leq y$, we have directly that:

  $$\mathbb{P}(x + X < x_p) \geq \mathbb{P}(y + X < x_p) \tag{62}$$

  $$\mathbb{P}(x + X > x_1) \leq \mathbb{P}(y + X > x_1) \tag{63}$$

  We use these results to compare the VAR_α of $f(x + X)$ and $f(y + X)$, that will allow us to finally conclude. Even if x and y lay on $[x_p, x_1]$, $x + X$ and $y + X$ lay potentially on $[a, b]$. Once again, we can decompose the space where lay f in different parts and try to compare the value of Value-at-Risk on these different parts, we note:

  $$f(\tilde{x}) = VAR_\alpha(f(x + X)) \text{ and } f(\tilde{y}) = VAR_\alpha(f(y + X))$$

  Then, we have:

– If the distribution of $x+X$ and $y+X$ is such that $f(\tilde{x})$ and $f(\tilde{y})$ are strictly greater than $f(x_2)$. Then, $\tilde{x}$ and $\tilde{y}$ belongs to $[a, x_p]$ (by second assumption of 4.2). On this interval, since 62, we have $f(\tilde{x}) \geq f(\tilde{y})$.

– If the distribution of $x+X$ and $y+X$ is such that $f(\tilde{x})$ and $f(\tilde{y})$ are equal to $f(x_2)$. Then, $\tilde{x}$ and $\tilde{y}$ belongs to $[x_p, x_2]$ or are greater than x_2.

– If the distribution of $x+X$ and $y+X$ is such that $f(\tilde{x})$ and $f(\tilde{y})$ are strictly inferior to $f(x_2)$. Then, $\tilde{x}$ and $\tilde{y}$ are both greater than x_2. In fact, they can not belong to $[x_p, x_2]$ because of third assumption of 4.2 which implies that $\forall x \in [x_p, x_2] p(x) \leq 1-\alpha$, and this implies that $f(\tilde{x}) \geq f(x_2)$.

We do not treat the others cases because they are either trivial (e.g. if $f(\tilde{x}) > f(x_2)$ and $f(\tilde{y}) = f(x_2)$) or impossible because of equations 62 and 63 (e.g. $f(\tilde{x}) = f(x_2)$ and $f(\tilde{y}) > f(x_2)$). Then, we can conclude that on $[x_p, x_2]$:

$$VAR_\alpha(f(x+X)) \geq VAR_\alpha(f(y+X)) \geq f(x_2) \tag{64}$$

Finally, by integration on α (using Equation (9)), we have:

$$\forall x, y \in [x_p, x_2] \text{ such that } AVAR_\alpha(f(x+X)) \geq AVAR_\alpha(f(y+X)). \tag{65}$$

Thus, we have prove that $\rho \circ f$ are non increasing on $[a, x_2]$. □

The following lemma explains in which case, $\rho \circ f$ is quasi convex.

Lemma 4.3. *Let $\alpha \in [0,1]$, let suppose that:*

- *X is a unidimensionnal random variable having a mean of 0 and a variance equal to $\sigma^2 < +\infty$.*
- *$x_1, x_2, x_3 \in M$ be such that : $x_3 \in [x_1, x_2]$*
- $\max\limits_{x \in I(x_1,x_3)} \{\mathbb{P}[x+X \in I(x_1,x_3)]\} \leq 1-\alpha.$
- $\max\limits_{x \in I(x_3,x_2)} \{\mathbb{P}[x+X \in I(x_3,x_2)]\} \leq 1-\alpha$

Then, $\rho \circ f$ is quasi convex on $[x_1, x_2]$.

Proof. It is a consequence of the Lemma 4.2. Indeed, we must:

- Apply the Lemma 4.2 on $[x_1, x_3]$ so $\rho \circ f$ is non increasing on $[x_1, x_3]$.
- Apply the symmetry of the Lemma 4.2 on $[x_3, x_2]$, so $\rho \circ f$ is non decreasing on $[x_3, x_2]$.

Then, it is sufficient to conclude that $\rho \circ f$ is quasi convex on $[x_1, x_2]$ □

The following corollary allow to show the link which exists between σ the standard deviation of the random variable X, α the index of robustness and s the size of a gap. Thanks to this result, given a standard deviation and a index of robustness α, we can say if a gap may be avoided with the measure $\rho \circ f$ or not.

Corollary 4.4. *Let suppose that:*

- *X is a Gaussian variable with a 0 mean and a standard deviation σ.*
- *$\alpha \in [0, 1]$ is taken such that: $1 - \alpha = \mathbb{P}(-z < X < z)$*
- *The size s of the interval $[x_1, x_2]$ is such that $s \leq 2z$.*

Then, I is no more a gap for the function $\rho \circ f$.

The following proposition shows how the function f may be convexified. We can have an idea of the convexification by applying the corollary 4.4.

Proposition 4.5. *Let $\alpha \in [0, 1]$, suppose that for each $L_i \in L$ there is at most one gap I such that $\max\limits_{x \in I}\{\mathbb{P}[x + X \in I]\} > 1 - \alpha$ then $\rho \circ f$ is quasi convex on $[a, b]$.*

Proof. We prove that by construction, iterating on the gaps of level ℓ_i. We recall that a global maximum of the function is attained by assumption in $x_1 = a$. Let take $i = 1$, we consider the gaps of level ℓ_1. There are three cases (see Figure 8) :

- Case 1: there is no gap of level ℓ_1 respecting the assumption and $\forall i\ L_i \subset L_1$ by Lemma 4.3, the function $\rho \circ f$ is quasi convex and the proof is over.
- Case 2: there is no gap of level ℓ_1 respecting the assumption and $\exists i, L_i \not\subset L_1$ then by applying Lemma 4.2 we can say that $\rho \circ f$ is non increasing on L_1 and we can iterate the same process on the restriction f_{I_i} of f on the interval I_i.
- Case 3:there is only one gap I_1 respecting the assumption, then by applying Lemma 4.2 we can say that $\rho \circ f$ is non increasing on $[a, u_1]$ and non decreasing on $[u_2, b]$ (by symmetry) with u_1 and u_2 respectively the lower and the upper bound of I_1. Then, we can iterate the same process on the restriction f_{I_1} of f on the interval I_1.

Finally, there are two cases:

- either the process is over before the last level ℓ_m and $\rho \circ f$ and by the previous construction, there is a gap where $\rho \circ f$ is quasi convex by 4.3. On the left of this gap, we have $\rho \circ f$ which is non increasing and the right of this gap it is non decreasing, so we can conclude that $\rho \circ f$ is quasi convex on $[a, b]$.
- or it is over with the last level, i.e there is no more local maxima on this interval then $\rho \circ f$ is non increasing then non decreasing and so quasi convex.

□

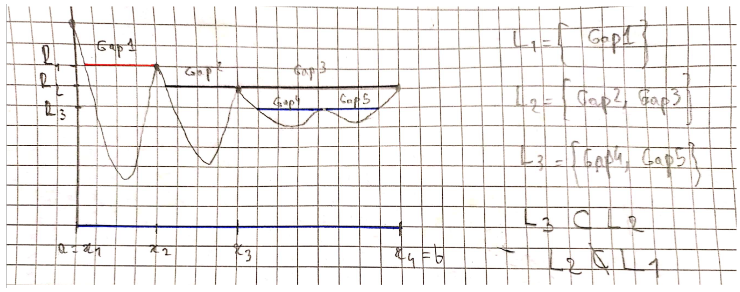

Figure 8: Example of gaps and levels

5 Conclusion

There are several things to do:

- Complete the proof for multidimensional case. Is it necessary ?
- Apply in an optimization context. Indeed, until now we just draw some curves, it is pretty but not very useful. A first step to apply in an optimization context is decided on which test problems. Indeed, we must find some test problems where the robustness taking robustness into account really brings something.
- Still in a goal of optimization, we must find an algorithm adapt to our context. In particular, we think that the corollary 4.4 may be very useful in that purpose.

References

[1] *Robust Design Concept in Possibility Theory and Optimization for System With Both Random and Fuzzy Input Variables*, volume Volume 6: 33rd Design Automation Conference, Parts A and B of *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 09 2007.

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

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

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

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

[6] 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.

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

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

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

[10] 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.

[11] 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.

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

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

[14] 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.

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

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

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

[18] 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.