Robust Formulation

Romain Couderc

June 22, 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 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 A new risk measure: Average Value-at-Risk

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a sample space, equipped with the sigma algebra $\mathcal{F}$, 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 [1]. Nevertheless, this measure has been called Conditional Value-at-Risk in [4] and is equivalent to Expected Shortfall in [2]. 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 [3] 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.

1.1 Average Value-at-Risk formulations

First, we are going to define the Average Value-at-Risk in the terminology of [4].

Definition 1.1. *The Conditional Value-at-Risk of $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathcal{P})$ at level $\alpha \in [0,1]$ is the risk measure define by:*

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

In [1], 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 1.2. *Let $\beta_1 \in [0,1]$ and $\beta_2 \in [0,+\infty]$, the average value-at-risk of $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathcal{P})$ at level $\alpha = \frac{\beta_1}{\beta_1+\beta_2}$ is defined by:*

$$AVAR_\alpha(Z) := \mathbb{E}[Z] + \inf_{t\in\mathbb{R}} \mathbb{E}\{\beta_1[t-Z]_+ + \beta_2[Z-t]_+\}. \tag{2}$$

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 [1].

Lemma 1.3. *If $\beta_1 = 1$ and $\beta_2 > 0$ then we have $\forall Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathcal{P})$:*

$$AVAR_\alpha(Z) = CVAR_\alpha(Z). \tag{3}$$

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

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

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 [2]: 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 1.4. *We define the Value-at-Risk of level α as:*

$$VAR_\alpha(X) = \inf\{t : \mathbb{P}(X \leq t) \geq 1-\alpha\} \tag{4}$$

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 [2]).

Definition 1.5. *We call the Expected Shortfall of $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathcal{P})$ at level α the following value:*

$$ES_\alpha(Z) = \frac{1}{\alpha}\int_{1-\alpha}^{1} VAR_{1-\tau}(Z)d\tau. \tag{5}$$

In fact, the Expected Shortfall is equivalent to the two previous formulation. We will prove it in 1.10 after having stated some other theoretical results.

1.2 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 $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ 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 $\mathcal{P}$.

Definition 1.6. *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. $\omega \in \Omega$, 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 [1] but it is an application of the proof they made in a more general case.

Lemma 1.7. *The average value-at-risk measure defined with the specified β_1 and β_2 in (2) is a coherent risk measure.*

Proof. First, let introduce the functions:

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

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 (2) as follows:

$$AVAR_\alpha(Z) = \inf_{t \in \mathbb{R}} \mathbb{E}[g(Z)] \tag{6}$$

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 $AVAR_\alpha(\cdot)$ which is convex.
- With the same arguments, since g is non decreasing then $AVAR_\alpha(\cdot)$ is monotone.
- 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 risk measure or 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.

1.3 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 1.8. *The measure $AVAR_\alpha(\cdot)$ defined in (2) attains its minimum with respect to t at:*

$$t^* = \inf\{t : \mathbb{P}(Z \le t) \ge 1-\alpha\} \quad \textit{with} \quad \alpha = \frac{\beta_1}{\beta_1+\beta_2}. \tag{7}$$

In order to prove the above proposition, we will need the following result from theorem 7.46 of [1]. In this theorem, we consider a random function $F : \mathbb{R}^n \times \Omega \to \bar{\mathbb{R}}$ and the corresponding expected value function $f(x) = \mathbb{E}[F(x,\omega)]$.

Theorem 1.9. *Suppose that the random function $F(x,\omega)$ 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:*

$$f'(x_0, d) = \mathbb{E}[F'_\omega(x_0, d)] \quad \textit{for all direction } d. \tag{8}$$

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

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

Proof. See the proof of theorem 7.46 of [1]. □

Henceforth, we have all the results necessary to prove the proposition 1.8.

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 $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathcal{P})$ and $t \in \mathbb{R}$:

$$f(t) = \mathbb{E}\{h(t,w)\} = \mathbb{E}\{\beta_1[t - Z(\omega)]_+ + \beta_2[Z(\omega) - t]_+\}$$

We know that f is well defined and finite valued for $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathcal{P})$. Moreover, since what we have proved in lemma 1.7, we know that $h(\cdot, \omega)$ is convex and differentiable everywhere except at $x = Z(\omega)$. The corresponding derivative is given by:

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

And we have $\forall t \neq Z(\omega)$ by theorem 1.9:

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

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 1.9 (we can see that in a way as an extension by continuity):

$$f'(t_0) = \mathbb{E}\left[\frac{\partial h(t,\omega)}{\partial t}\right] = \beta_1 \mathbb{P}(t_0 > Z) - \beta_2 \mathbb{P}(t_0 < Z) \tag{9}$$

  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 1.9 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:

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

because $h(t_0, \omega) = 0$. And then by the Monotone Convergence Theoreme (using the same arguments that in the proof of theorem 7.46 of [1]), 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:

$$\lim_{h\to 0} \mathbb{P}(Z \leq t_0 + h) = 1 - \alpha \quad \text{i.e.} \quad t^* = \inf\{t : \mathbb{P}(Z \leq t) \geq 1 - \alpha\} \tag{10}$$

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:

$$\lim_{h\to 0} \mathbb{P}(Z \leq t_0 - h) = 1 - \alpha \quad \text{i.e.} \quad t^{**} = \sup\{t : \mathbb{P}(Z \leq t) \leq 1 - \alpha\} \tag{11}$$

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 [1].

Proposition 1.10. *Let $Z \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathcal{P})$ and $H(z)$ its cumulative distribution functin, with $H(z)$ continuous at $z = VAR_\alpha(Z)$, we have the following identities:*

$$CVAR_\alpha(Z) = \frac{1}{\alpha} \int_{1-\alpha}^{1} VAR_{1-\tau}(Z) d\tau = \mathbb{E}(Z | Z \geq VAR_\alpha(Z)) \tag{12}$$

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:

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

By the substitution $\tau = H(z)$, this shows the first equality in 1.10 and then the second equality in 1.10 follows provided that $\mathbb{P}(Z = t^*) = 0$. □

2 Numerical Experiments

In order to see the effect of the previous theoretical results, we are going to test different risk measures. Since the mean risk measure and the mean-variance risk measure are the most used measure in context of robust design optimization, it will be them which will be used to draw a comparison with the Average Value-at-Risk measure.

2.1 Test function and comparison between the measures

The function on which we are going to test the different risk measures is drawn on Figure 1 and its formulation is:

$$f(x) = 1 - \exp\left(-\frac{1}{2}x^2\right) - \exp\left(-\frac{1}{50}(x-7)^2\right) - \exp\left(-5(x+5)^2\right). \tag{13}$$

This function has two local minima at $x = -5$ and $x = 7$ and one global minimum at $x = 0$.

The principle to test the different risk measure is to assume that the variable x is uncertain, in the following example $X \sim x + \mathcal{N}(0, 1)$. Then, we are going to draw the different function to optimize according to the risk measure used. For example, if we want to test the mean risk measure, then we draw $\mathbb{E}[f(X)]$. That allow us to understand the behavior of the function to optimize. To trace the curve of the different functions, for each x, we draw 1000 random samples X and then we apply the chosen measure ρ on the samples $f(X)$.

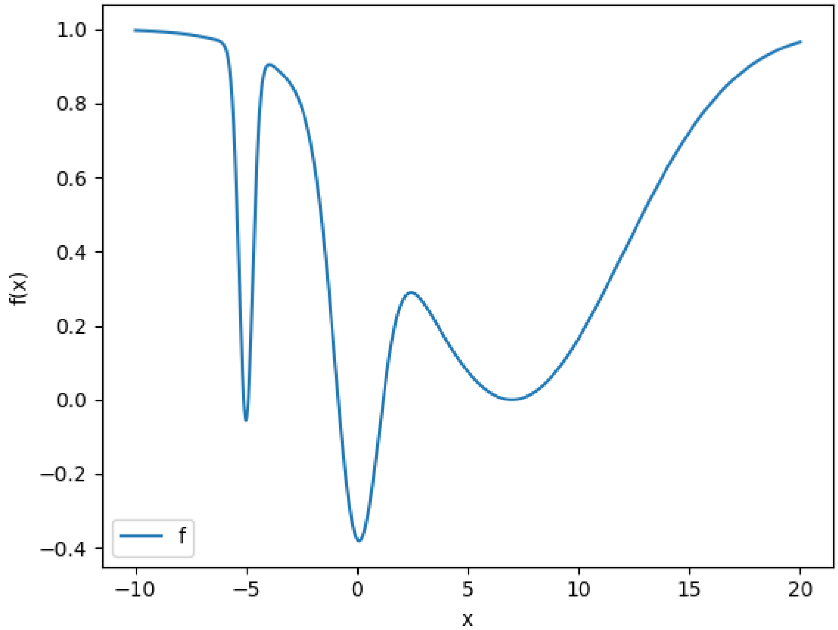

Figure 1: Test function

2.1.1 Mean risk measure

The result of the process described above and applied to mean risk measure is presented on Figure 2. The mean risk measure allow to remove the local minimum the least robust. Nevertheless, the global optimum is still the one at $x = 0$ while we could think that the robust optimum is rather at $x = 7$.

2.1.2 Mean-Variance risk measure

The result of the process described above and applied to mean-variance risk measure is presented on Figure 3.

Definition 2.1. *We recall that the mean variance risk measure is defined by:*

$$\rho(f(X)) = \mathbb{E}[f(X)] + c \times Var[f(X)] \textit{ with } c \in \mathbb{R} \tag{14}$$

I draw this measure with $c = 1$, $c = 2$ and $c = 3$. The mean variance risk measure allow to determine the robust minimum in any cases, even if for $c = 1$ the solution obtained at $x = 0$ is close to the one obtained at $x = 7$. However, the main drawback is that the function is totally different from the original function f. For instance, some maxima are appeared around the solution the least robust. Morevover, some a minimum is appeared at $x = -3$, which is a maximum of the function. This can make the function more difficult to optimize.

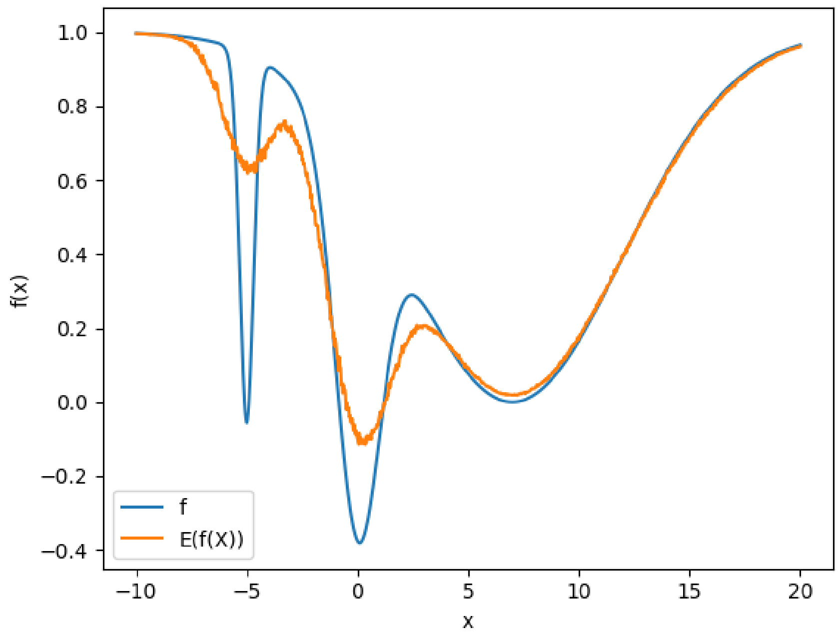

Figure 2: Mean risk measure

2.1.3 Average Value-at-Risk measure

The result of the process described above and applied to Average Value-at-Risk measure is presented on Figure 4. I draw it for different values of β but I try to keep $\alpha = \frac{\beta_1}{\beta_1+\beta_2}$ around 0.1 (i.e. an index of robustness close to 90 percent). The result shows several things:

- First, according to the values of β, the behavior of the measure changes. For $\beta_1 = 0.1$ and $\beta_2 = 1$, it is close to the behavior of mean risk measure. While for $\beta_1 = 0.5$ and $\beta_2 = 5$, it is rather close to the behavior of the mean variance risk measure with $c = 1$. Finally for $\beta_1 = 0.1$ and $\beta_2 = 1$, the behavior is quite interesting because it seems to "erase" the local minima non robust.

- It seems for β_1 too close of 0, the measure does not show a real robust minimum, it is more useful to fix $\beta_1 = 1$ and to vary β_2 for obtaining the index of robustness desired.

- Finally, we would add that the time to compute the Average Value-at-Risk measure is far longer than the one to calculate the two others measure. That is due to the optimization process realize to calculate this measure. Nevertheless, if we consider that the computation of the objective function is the process the most time consuming, then it is not a big matter.

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

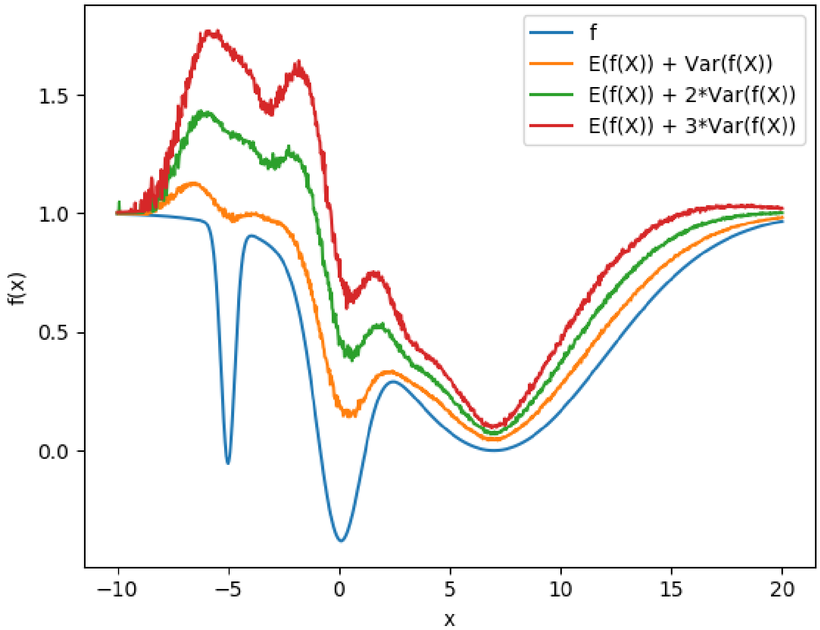

Figure 3: Mean Variance risk measure

test function. In addition, it would seem that the non robust minima may be transformed in kind of ”levels”. That could be interesting to determine a process giving some non optimal intermediate results.

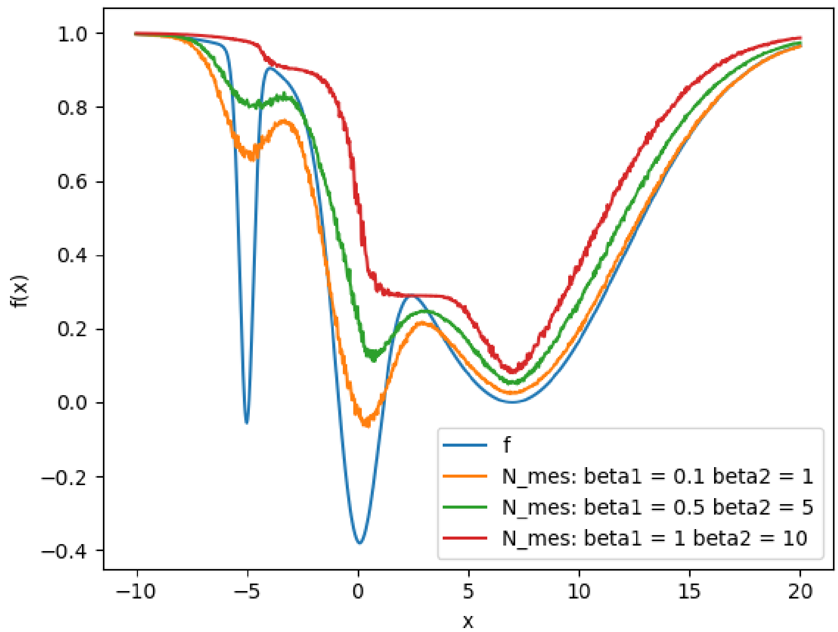

Figure 4: Average Value-at-Risk measure

2.2 Others results with Average Value-at-Risk measure

In the previous section, we have drawn a comparison between the different kind of measure and we have highlighted the different advantages of the Average Value-at-Risk measure. In this section, we want to explore further two points:

- In a real case, we want to reduce as much as possible the number of function calls: what is the impact on the measure ? May we still observe the different levels ?
- We have chosen an incertitude on x arbitrarily, indeed the standard deviation around x is equal to 1. What is the impact if we increase or decrease this standard deviation ?

2.2.1 Impact of number of points

In this section, we fix the values of $\beta_1 = 1$ and $\beta_2 = 10$. Then, we trace the Average Value-at-Risk function with a number of sample point $N_s \in \{2, 10, 100, 1000\}$. We recall that in the previous result N_s was fixed to 1000. Results are on Figure 5.

We can see on the right Figure that the difference between 100 and 1000 sample points is low. It is a little bit more noised for $N_s = 100$. For $N_s = 10$, the curve is noised, but the different levels are easily identifiable and the robust minimum is visible. Finally, for $N_s = 2$ we can guess the different level and the minimum the least robust is identifiable. However, we can not determine where is the robust minimum.

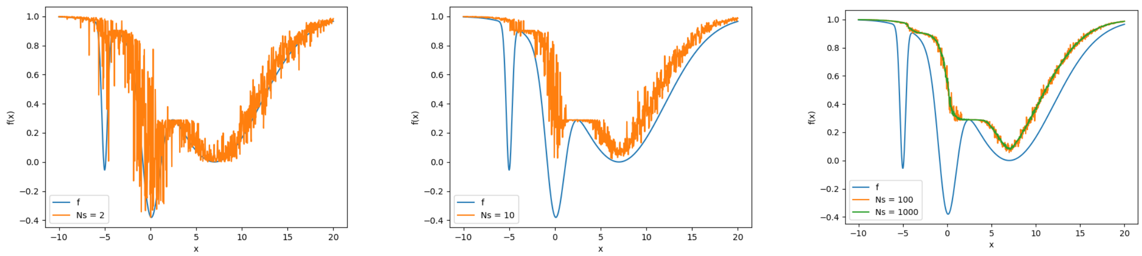

Figure 5: Results for different number of samples: $N_s = 2$ (left), $N_s = 10$ (middle) and $N_s = 100$ and $N_s = 1000$ (right)

2.2.2 Impact of the Standard deviation

In this section, we compare fix $\beta_1 = 1$, $\beta_2 = 10$ and $N_s = 1000$. Then, we draw the Average Value-at-Risk measure for $X \sim \mathcal{N}(x, \sigma)$ with $\sigma \in \{0.5, 1, 2\}$. In the previous section, we recall that $\sigma = 1$. The result is presented on Figure (6).

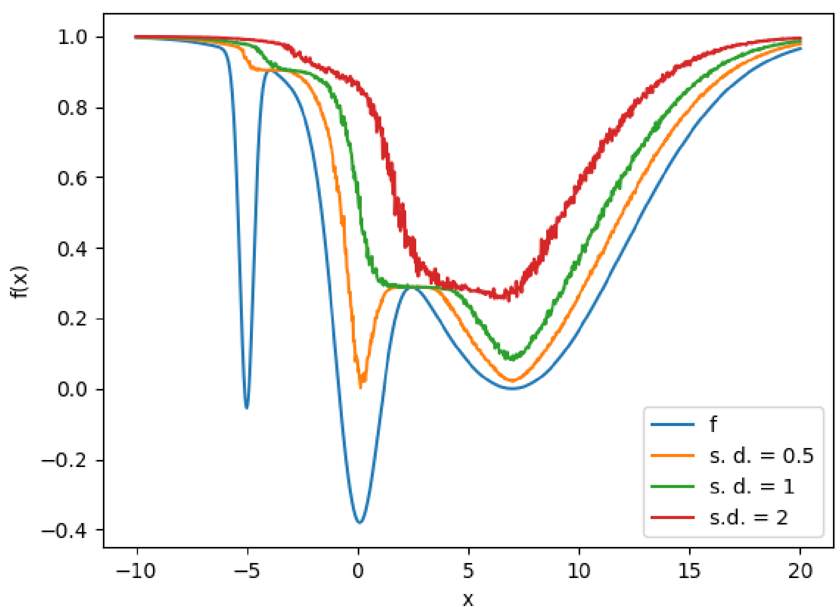

Figure 6: Impact of the variance on the Average Value-at-Risk measure

The fact of modify the level of uncertainty changes a bit our analyse. With a smaller standard deviation, the robust minimum is not necessarily at $x = 7$. We can be interested by some smaller values of the objective function (for instance the minimum at $x = 0$). On Figure (6), we can see that this trade-off is taken into account by the Average Value-at-Risk risk measure. More we are precise (i.e. smaller is the standard deviation), more the global minimum at $x = 0$ is considered as a potential robust solution. On the contrary, bigger is the standard deviation, more the new measure indicates the minimum at $x = 7$ as the robust minimum. In any cases, the measure function do not consider the minimum at $x = -5$ because of this narrowness. Thus, the new measure function seems to be an adequate measure for different level of uncertainty.

2.3 With others test functions

In the previous sections, we have seen a behavior of the Average Value-at-Risk risk measure which is interesting: it seems to convexify the function. In this section, we are going to see whether this behavior may be generalized with any others test functions. In this section, the default parameter are $\beta_1 = 1$, $\beta_2 = 10$, $N_s = 1000$ and the standard deviation of X called σ is equal to 1.

2.3.1 Another test function with an interesting behavior

The test function used here to test the behavior of the Average Value-at-Risk risk measure is a small modification of the original test function. Its formula is:

$$f(x) = 1 - \exp\left(-\frac{1}{2}x^2\right) - \exp\left(-\frac{1}{50}(x-7)^2\right) - \exp\left(-5(x+5)^2\right) + \frac{\sin\left(\frac{x}{2}\right)}{2}. \quad (15)$$

Applying the Average Value-at-Risk risk measure on this function with $X \sim \mathcal{N}(x, \sigma)$ with $\sigma \in \{1, 2, 2.5\}$ and others parameters at their default values, we obtain the Figure 7. On

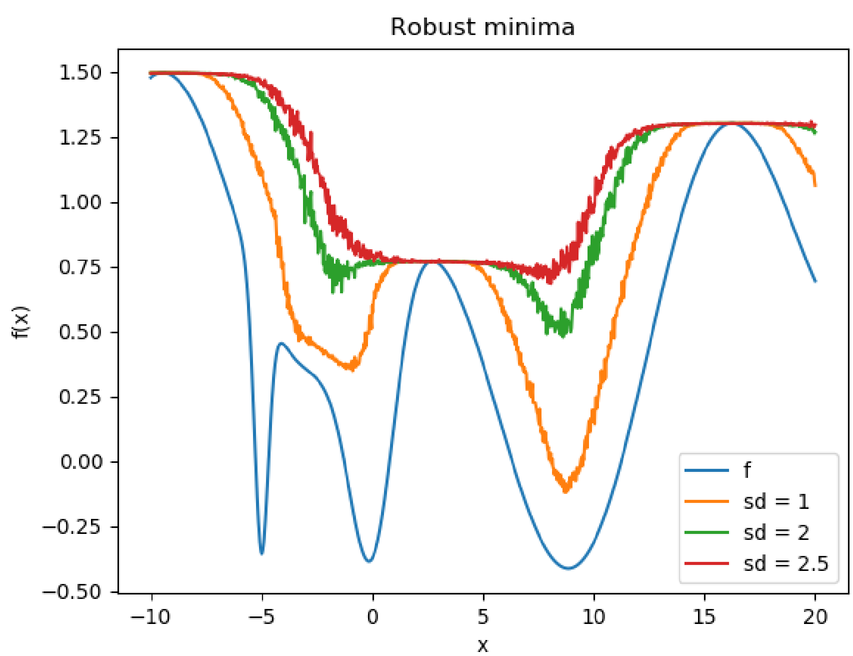

Figure 7: With another test function the result is not directly convexified, we must increase the standard deviation of X.

Figure 7, the behavior of the Average Value-at-Risk is a bit different. In fact, with the default values, there is no convexification of the original function. However, increasing the standard deviation allow to have this behavior of convexification. The influencing parametes of this behavior are the standard deviation of X, the value of α (and thus of β_2). Moreover, its influence depends on the "size of the gap" above the minimum concerned.

2.3.2 Locally Lipschitz continuous case

Until this section, the test functions was differentiable everywhere. To be closer with the reality, we are going to test the Average Value-at-Risk risk measure on the two following test functions which are only locally Lipschitz continuous:

$$f_1(x) = \sqrt{|x|} + \sqrt{|x-3|} + \sqrt{|x+3|}$$
$$f_2(x) = |x|\sin\left(\frac{1}{|x|}\right) + |x-2|\sin\left(\frac{1}{|x-2|}\right) + |x+2|\sin(|x+2|)$$

To be coherent with the others tests, the standard deviation of X is adapted to the definition domain. Thus in this section, $\sigma = 0.25$ for f_1 and $\sigma = 0.05$ and $\beta_2 = 55$ for f_2. The results are presented on the Figure 8. In the two cases, the Average Value-at-Risk risk

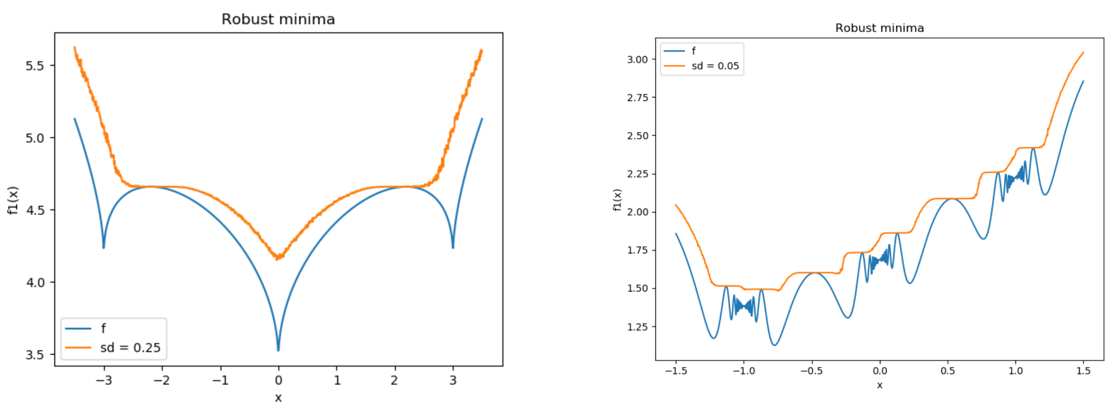

Figure 8: Average Value-at-Risk for test function f_1 on the left and f_2 on the right

measure convexifies the original function. In case of f_2, the Average Value-at-Risk seems to be constant in the vicinity of the minimum of this function even if we can notice a little decreasing above the robust minimum.

2.3.3 Multidimensionnal test function

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 9.

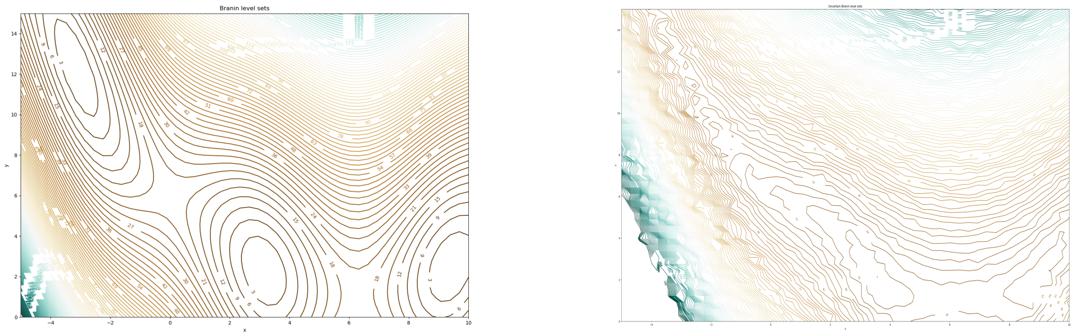

Figure 9: Branin function: certain (right) and uncertain (left) level set

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 9. 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.

2.3.4 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 1.2), 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.

3 Proof of convexification

This section demonstrates the convexification phenomenon of the Average Value-at-Risk in the $1-$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 3.1. *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:*

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & x \\ \text{subject to} \quad & f(x) \geq f(x_2), \\ & x \in [x_1, x_2] \end{aligned} \tag{16}$$

- *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 11).*
- *ρ the measure function defined in 2.*

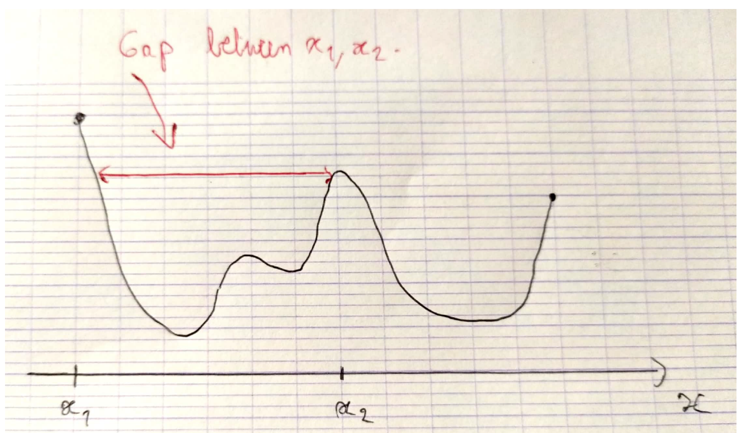

Figure 10: Example of a gap between two local maxima

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

Lemma 3.2. *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:

$$p(x) = \mathbb{P}[x + X \in I(a, x_2)] \text{ for } x \in I(a, x_2) \tag{17}$$

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

$$[a, x_2] = [a, x_p] \cup I(a, x_2) = [a, x_p] \cup [x_p, x_2] \tag{18}$$

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 3.2 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:

$$\mathbb{P}(x + X < x_p) \geq \mathbb{P}(y + X < x_p) \tag{19}$$

$$\mathbb{P}(x + X > x_1) \leq \mathbb{P}(y + X > x_1) \tag{20}$$

  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 3.2). On this interval, since 19, 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 3.2 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 19 and 20 (e.g. $f(\tilde{x}) = f(x_2)$ and $f(\tilde{y}) > f(x_2)$). Then, we can conclude that on $[x_p, x_2]$:

$$VAR_\alpha(f(x + X)) \geq VAR_\alpha(f(y + X)) \geq f(x_2) \tag{21}$$

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

$$\forall x, y \in [x_p, x_2] \text{ such that } AVAR_\alpha(f(x + X)) \geq AVAR_\alpha(f(y + X)). \tag{22}$$

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 3.3. *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 : $x_3 \in [x_1, x_2]$*
- $\max_{x \in I(x_1, x_3)} \{\mathbb{P}[x + X \in I(x_1, x_3)]\} \leq 1 - \alpha$.
- $\max_{x \in I(x_3, x_2)} \{\mathbb{P}[x + X \in I(x_3, x_2)]\} \leq 1 - \alpha$

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

Proof. It is a consequence of the Lemma 3.2. Indeed, we must:

- Apply the Lemma 3.2 on $[x_1, x_3]$ so $\rho \circ f$ is non increasing on $[x_1, x_3]$.
- Apply the symmetry of the Lemma 3.2 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 3.4. *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: $1-\alpha = \mathcal{P}(-z < X < z)$*
- *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 3.4.

Proposition 3.5. *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 11) :

- Case 1: there is no gap of level ℓ_1 respecting the assumption and $\forall i$ $L_i \subset L_1$ by Lemma 3.3, 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 3.2 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 3.2 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.

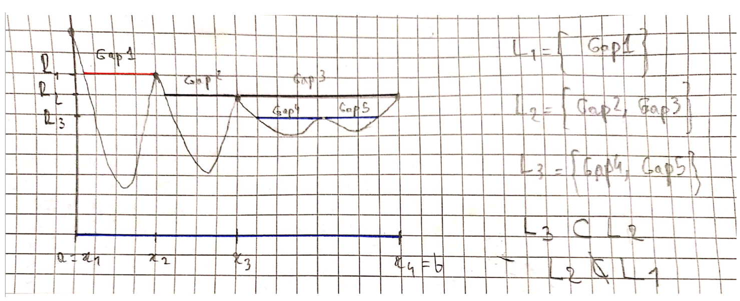

Figure 11: Example of gaps and levels

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 3.3. 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.

□

4 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 3.4 may be very useful in that purpose.

References

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

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

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

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