Robust Optimization

Fabrice Poirion

ONERA

October 23, 2019

Some general facts

➤ Robust optimization is a subdomain of optimization which allows to consider problems involving uncertain parameters.

Some general facts

- Robust optimization is a subdomain of optimization which allows to consider problems involving uncertain parameters.
- A first ingredient of robust optimization is to determine the "uncertainty sets" which contain the most plausible values for the uncertain coefficients.

Some general facts

- Robust optimization is a subdomain of optimization which allows to consider problems involving uncertain parameters.
- A first ingredient of robust optimization is to determine the "uncertainty sets" which contain the most plausible values for the uncertain coefficients.
- A solution of an optimization problem is said to be robust if it fulfills the constraints for all values of the parameter in the uncertainty set : it is the worst case scenario.

Definition

A vector $x \in \mathbb{R}^n$ is a robust solution if it fulfills the constraints for all uncertain parameters values.

Some general facts

- Robust optimization is a subdomain of optimization which allows to consider problems involving uncertain parameters.
- A first ingredient of robust optimization is to determine the "uncertainty sets" which contain the most plausible values for the uncertain coefficients.
- A solution of an optimization problem is said to be robust if it fulfills the constraints for all values of the parameter in the uncertainty set : it is the worst case scenario.

Definition

A vector $x \in \mathbb{R}^n$ is a robust solution if it fulfills the constraints for all uncertain parameters values.

- A robust solution is therefore solution of a (finite or infinite) set of optimization problems: how to deal with the infinite case?

Some general facts

- Robust optimization is a subdomain of optimization which allows to consider problems involving uncertain parameters.
- A first ingredient of robust optimization is to determine the "uncertainty sets" which contain the most plausible values for the uncertain coefficients.
- A solution of an optimization problem is said to be robust if it fulfills the constraints for all values of the parameter in the uncertainty set : it is the worst case scenario.

Definition

A vector $x \in \mathbb{R}^n$ is a robust solution if it fulfills the constraints for all uncertain parameters values.

- A robust solution is therefore solution of a (finite or infinite) set of optimization problems: how to deal with the infinite case?
- In practice, the uncertainty sets should neither be too big (containing improbable values for the coefficients) nor too small.

$$\min_x \{c^T x + d \mid Ax \leq b\} \tag{1}$$

where $x \in \mathbb{R}^n$ is a design variable, $c \in \mathbb{R}^n$ et $d \in \mathbb{R}$ define the linear objective function, $A \in \mathrm{Mat}_{\mathbb{R}}(m, n)$ and $b \in \mathbb{R}^m$ define the constraints.
Assume that all coefficients are uncertain: they take their values in an uncertainty set $\mathcal{I} \subset \mathbb{R}^{(m+1)\times(n+1)}$.
The optimization problem with uncertainties can be written as the family of optimization problem

$$\left\{\min_x \{c^T x + d \mid Ax \leq b\}\right\}_{(c,d,A,b)\in\mathcal{I}}. \tag{2}$$

x is a robust solution if:

$$Ax \leq b \; ; \; \forall (c, d, A, b) \in \mathcal{I} \tag{3}$$

A simple example II

Definition

Let x_0 be fixed, the robust value $J(x_0)$ of the objective function is the maximal value of the objective function over the uncertainty space:

$$J(x_0) = \sup_{(c,d,A,b)\in\mathcal{I}} (c^T x_0 + d) \tag{4}$$

Definition

the robust counterpart of the optimization problem with uncertainty is the following minmax problem:

$$\min_x \left\{ J(x) = \sup_{(c,d,A,b)\in\mathcal{I}} (c^T x + d) \mid \; Ax \leq b \;\; \forall (c,d,A,b) \in \mathcal{I} \right\} \tag{5}$$

An optimal solution of the robust counterpart problem is called an optimal robust solution.

A simple example III

- Illustration

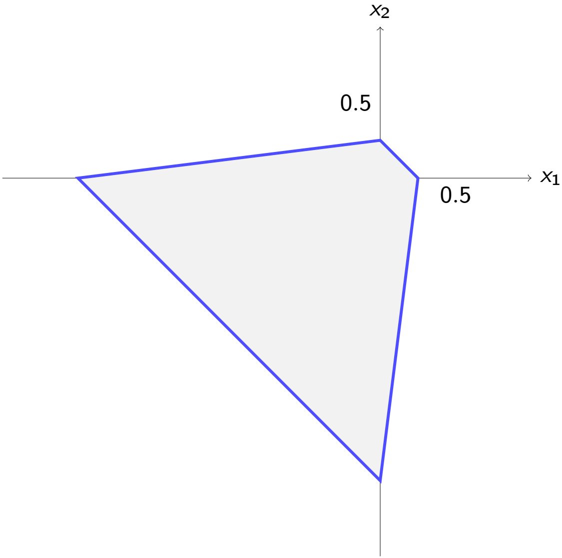
$$a_1 x_1 + a_2 x_2 \leq b \tag{6}$$

where the parameters a_i take their values in the intervals $[a_{i,0} - \alpha_i, a_{i,0} + \alpha_i]$; $i = 1, 2$.
If constraints(6) have to be fulfilled for all values of parameters a_i, it is sufficient to check that they are fulfilled for the worst case :

$$a_{1,0} x_1 + |\alpha_1 x_1| + a_{2,0} x_2 + |\alpha_2 x_2| \leq b \tag{7}$$

This last relation describes a simple polyhedral defined by four lines: (in the example $a_{1,0} = a_{2,0} = 1$; $\alpha_1 = \alpha_2 = 1.5$; $b = 1$):

A simple example IV



A simple example V

- The optimization problem with uncertainty, written in the robust counterpart, is a standard optimization problem with simple, convex constraints.

- Where and when does the probability theory appear?

The probability context

- Often (most of the time) uncertainty is introduced through random variable models.
- What does that change for the formulation ?
- There exists 2 directions
  - construct an adapted uncertainty space : the expression most plausible values becomes now comprehensive
  - change the problem by taking into account the probability information on the random model introduced.

From now on the various uncertain parameters will be modeled as a random variable $\xi(\omega)$ defined on the probability space $(\Omega, \mathcal{T}, P)$ $\xi : \Omega \to \Xi$; $\omega \mapsto \xi(\omega)$.

We are looking now at optimization problems written informally as

$$\min_x f(x, \xi(\omega)) \ ; \ g(x, \xi(\omega) \leq 0$$

Construction of the uncertainty set in the probability context

The uncertainty set is constructed by allowing the constraint to be violated with a given probability:

$$\mathcal{I} = \{\xi \mid \mathrm{Prob}\{g(x, \xi(\omega)) \leq 0\} \geq 1 - \alpha\}$$

Corresponding robust optimization problem with uncertainties:

$$\min_{x} \left\{ J(x) = \sup_{\xi \in \mathcal{I}} (f(x, \xi) \mid \; g(x, \xi) \leq 0 \;\; \forall \xi \in \mathcal{I} \right\} \tag{8}$$

- Technical problem:
  - the set $\mathcal{I}$ may be very difficult to construct
  - may have not any good properties : convexity, etc

Back to the simple example

Consider the simple constraint

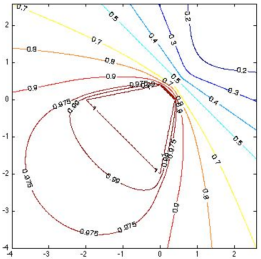
$$g(x, \xi(\omega)) = a_1(\xi_1(\omega))x_1 + a_2(\xi_1(\omega))x_2 \leq b \quad (9)$$

with

$$a_i(\xi_i(\omega)) = a_{i,0} + a_{i,1} \times \xi_i(\omega)$$

$\xi_i(\omega)$ are independent uniform random variables on $[-1, 1]$.

Uncertainty sets (α)



Back to the simple example II

Now if the uncertainty is modeled through Gaussian random variables:

$$a_i(\xi_i(\omega)) = a_{i,0} + a_{i,1} \times g_i(\omega)$$

$g_i(\omega)$ are independent standard Gaussian random variables.

Then the uncertainty set are described by the algebraic expression

$$\boxed{a_{1,0}x_1 + a_{2,0}x_2 + \Phi^{-1}(\alpha)\sqrt{a_{1,1}^2 x_1^2 + a_{2,1}^2 x2_2^2} \leq b}$$

Φ : cumulative function of the normal distribution.

Knowing the probability distributions of the random variables modeling the uncertain parameters it is possible to express the uncertainty in terms of a deterministic function : the risk measure.
The robust optimization problem will then be written as a deterministic optimization problem.
The difficulty brought by the uncertain parameters is replaced by the (numerical) difficulty of evaluating the risk measure.

Definition (First order stochastic dominance)

Let $X(\omega)$ and $Y(\omega)$ two real valued random variables. X is said to stochastically dominate Y at the first order if for any increasing continuous positive function f we have $E[f(X)] \geq E[f(Y)]$

An equivalent definition is that

$$F_X(u) \leq F_Y(u) \ ; \ \forall u \in \mathbb{R}$$

We denote then $X \succcurlyeq Y$.

Definition (Risk measure)

A risk measure associates to each random variable $Z(\omega)$ a real number $\mathcal{R}(Z) \in \mathbb{R} \cup -\infty \cup +\infty$.
A risk measure has to fulfill the following properties:

Convexity

$$\mathcal{R}(\alpha Z_1 + (1-\alpha)Z_2) \leq \alpha\mathcal{R}(Z_1) + (1-\alpha)\mathcal{R}(Z_2) \; ; \; \alpha \in [0,1]$$

Monotony

$$Z_2 \succcurlyeq Z_2 \Rightarrow \mathcal{R}(Z_2) \geq \mathcal{R}(Z_1)$$

Translation invariance

$$\mathcal{R}(Z+a) = \mathcal{R}(Z) + a \; ; \; \forall a \in \mathbb{R}$$

Positive homogeneity

$$\mathcal{R}(aZ) = a \times \mathcal{R}(Z) \; ; \; \forall a > 0$$

We consider now the class of risk measures $\mathcal{R}$ defined by the following construction :

$$\mathcal{R} = \inf_{x \in \mathbb{R}} E[\rho(Z(\omega), x)] \tag{10}$$

Classical risk measures I

Let $Z(\omega)$ a random variable.

Mean value	$\mathcal{R}(Z) = E(Z)$
	$\rho(Z,x)=Z$
Variance	$\mathcal{R}(Z) = E[(Z - E(Z))^2]$
	$\rho(Z,x)=(Z-x)^2$; $E[(Z-E(Z))^2]=\min_{x\in\mathbb{R}} E[(Z-x)^2]$
Mean variance compromise	$\mathcal{R}(Z) = E(Z) + \alpha \times E[(Z - E(Z))^2]$; $\alpha \geq 0$
	$\rho(Z,x)=Z+\alpha(Z-x)^2$
Value at risk	$\mathcal{R}(Z) = VaR_p(Z) = \min\{x \in \mathbb{R} \mid F_Z(x) \geq p\}$
	$\rho(Z,x)=\mathrm{Prob}(Z\leq x)\geq p$
Conditional value at risk	$\mathcal{R}(Z) = CVaR_p(Z) = E[Z \mid Z \geq VaR_p(Z)]$
	$\rho(Z,x)=E[x+(Z-x)_+/(1-p)]$; $x_+=\max(x,0)$

Classical risk measures II

The robust counterpart of the optimization problem with uncertainty :

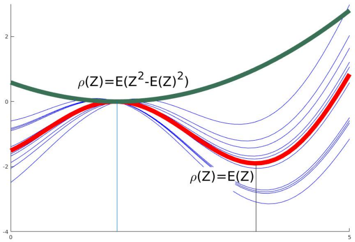
$$\min_x \left\{ J(x) = \sup_{\xi \in \mathcal{I}} (f(x,\xi) \mid \; g(x,\xi) \leq 0 \;\; \forall \xi \in \mathcal{I} \right\} \tag{11}$$

is replaced by the optimization of the risk measure:

$$\min_x \mathcal{R}(f(x,\xi(\omega)) \mid \mathcal{R}(g(x,\xi(\omega)) \leq 0 \tag{12}$$

- To each risk measure corresponds a distinct problem and different solutions.

Different solutions for 2 risk measures



Solving robust optimization problem I

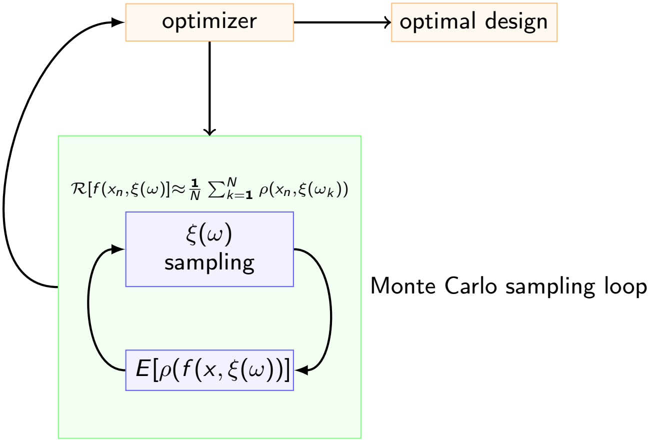
- Recall that RO problems can be written, (without any constraint for the moment)

$$\min_x \mathcal{R}(f(x,\xi(\omega)) =: \min_x \min_u E[\rho(f(x,\xi(\omega)),u)]$$

- The optimization problem is deterministic.
- The numerical difficulty is to evaluate the expectancy: no analytic expression.
- Construct an estimator ? Very expensive because it has to be done at each optimization step x: double loop issue. But it is often done together with the use of deterministic genetic algorithms
- Solution : use algorithms adapted to the probability context : stochastic algorithms.

Solving robust optimization problem II

Optimization loop



Stochastic approximation

- Problem : find x^* such that

$$E[F(x^*, \xi(\omega))] = a \; ; \; x^* = \operatorname*{Argmin}_{x \in U} E[F(x, \xi(\omega))] \tag{13}$$

- Example of application in reliability : find the value of a structural parameter q such that an uncertain structure is safe with a probability greater than a given level p:

$$\text{Find} \quad q \;/\; P(A(q,\omega) \geq 0) =: E[\mathbb{I}_{\mathbb{R}+}(A(q,\omega))] \geq p$$

The Robbins Monro algorithm

In order to find x^* such that

$$g(x^*) =: E(F(x^*, \xi(\omega))) = a$$

one can use the following algorithm

$$X_{n+1} = X_n + \gamma_n \left(F(X_n, \xi(\omega_{n+1})) - a\right) \; ; \; X_0 = x_0 \tag{14}$$

The Robbins Monro algorithm

In order to find x^* such that

$$g(x^*) =: E(F(x^*, \xi(\omega))) = a$$

one can use the following algorithm

$$X_{n+1} = X_n + \gamma_n \left(F(X_n, \xi(\omega_{n+1})) - a\right) \; ; \; X_0 = x_0 \tag{14}$$

Assuming that

- $(g(x) - a) \times (x - x^*) < 0$
- $|g(x)| \leq K \times (1 + |x|)$,

then the sequence X_n converges almost surely towards x^* for any initial value x_0 and for any sequence of positive real numbers (γ_n) such that

$$\sum \gamma_n = +\infty \; ; \; \sum \gamma_n^2 < +\infty, \qquad \gamma_n \text{ is called a } \sigma\text{-sequence.} \tag{15}$$

For example $\gamma_n = a/(n^\alpha + b)$, $\alpha \in]0.5, 1]$.

(X_n) defines a Markov chain.

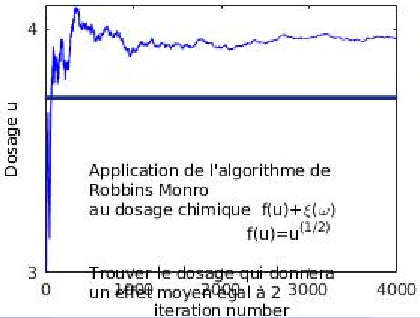
Robbins-Monro example

Chemical dosage

- A dose u of a chemical component induce a random effect X of unknown mean $f(u)$

$$X(u,\omega) = f(u) + \xi(\omega) \; ; \quad \xi(\omega) \text{ is a zero-mean random variable}$$

- find the dose u^* such that $f(u^*) = a$
- Robbins Monro : $U_{n+1} = U_n - \gamma_n(X_{n+1} - a) \; ; \; U_0 = u_0$



Back to robust optimization : stochastic gradient algorithm

➤ Apply Robbins Monro algorithm to find zeros of the objective function gradient.

$$x^* = \underset{x \in X^{\mathbf{ad}}}{\text{Argmin}}\, J(x) \; ; \; J(x) = E[f(x, \xi(\omega))]. \tag{16}$$

➤ The algorithm of the stochastic gradient method uses the optimization iteration in order to build an estimate of the gradient expectation:

- Choose $x_0 \in X$ and $\gamma_k > 0$ for $k \in \mathbb{N}$.
- Draw ξ_{n+1} under the law of ξ independently of ξ_k for $k \leq n$.
- Update

$$X_{n+1} = x_n - \gamma_n \left(f'_x(x_n, \xi_{n+1})\right). \tag{17}$$

- Project over the feasible space $X^{\mathbf{ad}}$

$$x_{n+1} = \Pi_{X^{\mathbf{ad}}}(X_{n+1}). \tag{18}$$

➤ The sequence (x_n) converges to the solution x^* of the problem under the Robbins Monroe assumptions applied to f'_x .

Stopping criteria

- The sequence of differences $|X_{n+1} - X_n|$ is not decreasing : how to know when to stop the algorithm?
- Fake convergences may occur according to the choice of the sequence γ_n used.
- Dvoretzky criteria: necessitates to know the solution in order to construct the sequence γ_n.
- There exists results of the type "central limit theorem":

$$\frac{1}{\sqrt{\gamma_n}}(x_n - x^*) \xrightarrow{\text{Loi}} \mathcal{N}(0, \sigma)$$

but necessitates restrictive assumptions.

- In practice : visual criteria!
- The expectancy of the random variable $f'_x(x_k, \xi_{k+1})$ converges towards the expectancy of $J'(x^*)$, which is equal to 0.

An estimator of this quantity can be constructed

$$(\sum_{l=1}^{k} \gamma_l f'_x(x_l, \xi_{l+1})) \ / \ (\sum_{l=1}^{k} \gamma_l)$$

and may be used in order to check visually that it converges towards 0.

Extensions

As in the case of the classical gradient algorithm there exist several extensions of the stochastic version. For instance

- Stochastic Newton algorithm
- Averaged stochastic gradient descent
- Extension to nonderivable objective objective functions
- Probabilistic constraints
- ...

➤ Does it work for industrial problems ?

Illustration in aeroelasticity I

- Flutter equation with uncertain parameters ξ in the frequency domain:

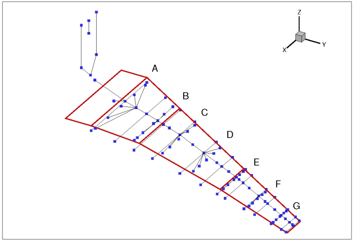
$$L(\xi)^T \left[p^2 \Phi^T(\xi) M(\xi) \Phi(\xi) + \Phi^T(\xi) K(\xi) \Phi(\xi) + \frac{1}{2} \rho V^2 \Phi^T(\xi) A(p/V) \Phi(\xi) \right] R(\xi) = 0$$

- The solution is $p(q, \xi) \in \mathbb{C}$ and depends of the uncertain parameters ξ and of the airflow pressure $q = 1/2 \rho V^2$. The sign of the real part $\Re(p)$ specifies the stability of the coupled system.

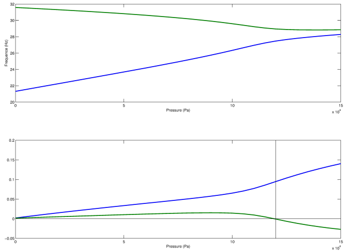
- The critical pressure q_c is the smallest pressure value q such that $\Re(p(q)) = 0$, if any. The critical pressure depends of the uncertain parameters ξ and therefore is itself a random variable.

Illustration in aeroelasticity II

Wing stick model



Flutter diagram



➜ 7 regions of the stick model have been considered for which a random stiffness coefficient ξ_i is introduced in order to model the stiffness uncertainty of each region.

➜ The uncertain parameters are modeled as uniform random variable

➜ 89 grid mass points m_j are chosen as optimization parameters

Illustration in aeroelasticity III

- Optimization problem

Problem I

$$\text{Argmax}\, J(m) = E[q_c(m, \xi(\omega))]$$

$$s.c. \begin{cases} m_i \in [a_i, b_i], \forall i \\ \sum_i m_i = c \end{cases}$$

Gradient expression

$$\frac{\partial q_c}{\partial m_i}(m^0) = -\frac{\Re(\frac{\partial p}{\partial m_i}(m^0, q_c^0))}{\Re(\frac{\partial p}{\partial q}(m^0, q_c^0))}. \tag{19}$$

Optimization parameters : : mass points m_i.
Uncertain parameters: stiffness of the finite element bars, independent uniform distributions.

Illustration in aeroelasticity IV

Convex feasible space:

$$X^{\mathbf{ad}} = \{m = (m_1, ..., m_N) \in \mathbb{R}^N \mid \sum_{j=1}^{N} m_j = c \ ; \ m_i \in [a_i, b_i], \forall i\}$$

.

➤ algorithm

$$m^{n+1} = \Pi_{X^{\mathbf{ad}}}(m^n + \gamma_n \frac{\partial q_c}{\partial m}(m^n, \xi_{n+1})). \tag{20}$$

- initialize m^0 with model original mass;
- draw model random variables;
- calculate the gradient: $U_n = \left(\frac{\partial q_c}{\partial m_i}(m^n)\right)_i$;
- set $u_{n+1} = m^n + \rho_n U_n$;
- project u_{n+1} on feasible space: $m^{n+1} = \Pi_{X^{\mathbf{ad}}}(u_{n+1})$.

Illustration in aeroelasticity V

Almost sure Convergence

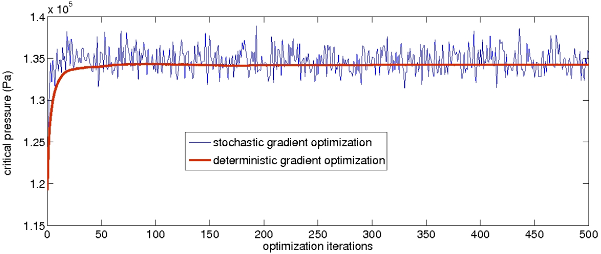
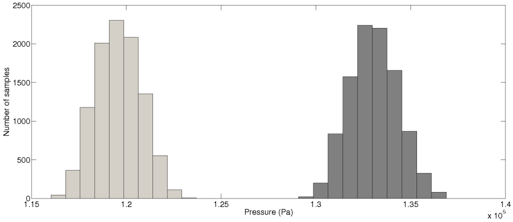


Illustration in aeroelasticity VI

Random critical pressure distribution before and after optimization



Chance constraint optimization : the Arrow Hurwicz algorithm

- The deterministic case

$$\underset{x\in X^{\mathbf{ad}}}{\mathrm{Argmin}}\{J(x)\,|\,G(x)\leq 0\}. \tag{21}$$

- Lagrangian formulation:

$$L(x,\lambda) = J(x) + \lambda G(x)$$

Arrow Hurwicz algorithm

$$x_{n+1} = \Pi_{X^{\mathbf{ad}}}(x_n - \gamma\,(J'_x(x_n) - \lambda_k\; G'_x(x_n))), \gamma > 0 \tag{22}$$

$$\lambda_{n+1} = \Pi_{\mathbb{R}+}(\lambda_n + \sigma\,(G(x_n))), \sigma > 0 \tag{23}$$

Stochastic Arrow Hurwicz algorithm

$$\underset{x \in X^{\mathbf{ad}}}{\mathrm{Argmin}} \{\mathbb{E}[f(x,\xi(\omega))] \mid P[g(x,\xi(\omega)) \geq 0] \geq p_0\}. \tag{24}$$

$g : X \times \mathbb{R}^d \to \mathbb{R}$: physical or mechanical quantity.

$$L(x,\lambda) = \mathbb{E}[f(x,\xi(\omega))] + \lambda\big[p_0 - P[g(x,\xi(\omega)) \geq 0]\big], \lambda \in \mathbb{R}$$

Remark

$P[g(x,\xi(\omega)) \geq 0] = E[G(x,\xi(\omega))]$ *avec* $G(x,\xi(\omega)) = \mathbb{I}_{\mathbb{R}_+}(g(x,\xi(\omega)))$

Stochastic Arrow Hurwicz algorithm

$$x_{n+1} = \Pi_{X^{\mathbf{ad}}}(x_n - \gamma_n\,(f'_x(x_n,\xi_{n+1}) - \lambda_n\,G'_x(x_n,\xi_{n+1}))),$$

$$\lambda_{n+1} = \Pi_{\mathbb{R}^+}(\lambda_n + \sigma_n\,(p_0 - G(x_n,\xi_{n+1}))),$$

$$\sum \gamma_n = +\infty \; ; \; \sum \sigma_n = +\infty \; ; \; \sum \gamma_n^2 < +\infty \; ; \; \sum \sigma_n^2 < +\infty$$

$$\underset{x\in X^{\mathbf{ad}}}{\operatorname{Argmin}}\{\mathbb{E}[f(x,\xi(\omega))]\,|\,P[g(x,\xi(\omega))\geq 0]\geq p_0\}. \tag{24}$$

$g : X \times \mathbb{R}^d \to \mathbb{R}$: physical or mechanical quantity.

$$L(x,\lambda) = \mathbb{E}[f(x,\xi(\omega))] + \lambda\big[p_0 - P[g(x,\xi(\omega))\geq 0]\big], \lambda\in\mathbb{R}$$

Remark

$P[g(x,\xi(\omega))\geq 0] = E[G(x,\xi(\omega))]$ *avec* $G(x,\xi(\omega)) = \mathbb{I}_{\mathbb{R}_+}(g(x,\xi(\omega)))$

Stochastic Arrow Hurwicz algorithm

$$x_{n+1} = \Pi_{X^{\mathbf{ad}}}(x_n - \gamma_n\,(f'_x(x_n,\xi_{n+1}) - \lambda_n\,G'_x(x_n,\xi_{n+1}))),$$

$$\lambda_{n+1} = \Pi_{\mathbb{R}^+}(\lambda_n + \sigma_n\,(p_0 - G(x_n,\xi_{n+1}))),$$

$$\sum\gamma_n = +\infty\ ;\ \sum\sigma_n = +\infty\ ;\ \sum\gamma_n^2 < +\infty\ ;\ \sum\sigma_n^2 < +\infty$$

Problem: $G(x,\xi(\omega))$ non derivable.

Constraint regularization

The non-derivable function G is approximated by a smooth function G_r :

Convolution regularization

$$G_r(x,\xi) = \frac{1}{r}\int_0^{+\infty} h\left(\frac{y - G(x,\xi)}{r}\right) dy \ ; \ r > 0$$

Constraint $P[g(x,\xi(\omega)) \geq 0]$ is replaced by the constraint

$$P_r(x) = E[G_r(x,\xi)]$$

with r small and where h is an even distribution.

For instance $h(x) = \frac{3(1-x^2)}{4}\mathbb{I}_{[-1,1]}(x)/4$

Theorem

The solution of the smooth problem with constraint P_r converges towards the solution of the original problem when $r \rightarrow 0$.

Second aeroelastic illustration

Weight minimization

$$\min \sum m_i$$

$$s.c. \begin{cases} m_i \in [a_i, b_i], \forall i \\ P(q_c(m) > q_0) > \alpha \end{cases}$$

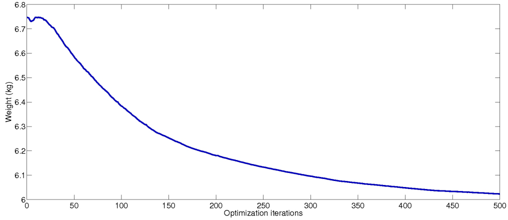
Constraint regularization :

$$G_r(m, \xi(\omega)) = \frac{1}{r} \int_0^{+\infty} h(\frac{y + \mathbb{I}_{]\alpha,\infty[}(q_0 - q_c(m, \xi(\omega)))}{r}) dy \ ; \ r \rightsquigarrow r_n = \frac{K_3}{n^{1/5}}$$

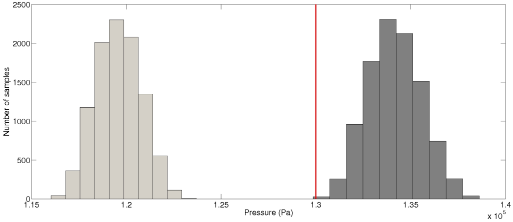
Algorithm

- initialization of m^0 and λ^0;
- uncertain parameters random sampling;Random critical pressure distribution
- gradient calculation: $U_n = \left(\frac{\partial q_c}{\partial m_i}(m^n, \xi_{n+1})\right)_i$;
- iteration $m_{n+1} = \Pi_{U^{ad}}(m_n + \gamma_n(U_n - G'_{r_n}(m, \xi_{n+1})\lambda_n)) \ ; \ \gamma_n = \frac{K_1}{K_2+n}$;
- iteration $\lambda_{n+1} = \Pi_{\mathbb{R}_+}(\lambda_n - \sigma_n G_{r_n}(m, \xi_{n+1})) \ ; \ \sigma_n = \frac{K_3}{n}$.

Almost sure convergence : 10% gain



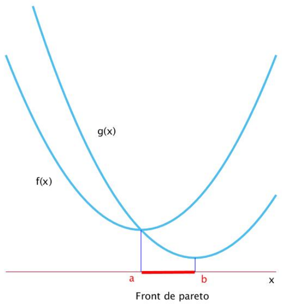
Random critical pressure distribution comparison



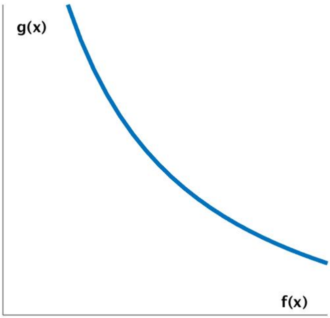
Some introductory facts on multiobjective optimization

$$\min_{x} \{f(x), g(x)\}$$

Pareto front of solutions



Pareto front of objectives



Pareto dominance

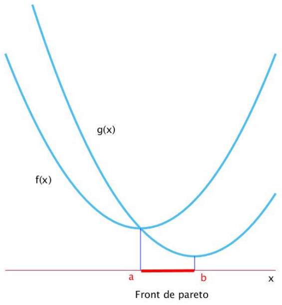
Let $J_i(x)$; $i = 1, n$ be n objectives that have to be minimized:

$$\min_x \{J_1(x), ..., J_n(x)\}$$

$x \in V$ Pareto dominates $y \in V$, $x \succ y$ if

$$\forall i \in [1,p], J_i(x) \leq J_i(y) \ ; \ \exists j \in [1,p], J_j(x) < J_j(y)$$

Optimal Pareto points



Common descent vector

Definition

A common descent vector for functions $x \mapsto f(x)$ and $x \mapsto g(x)$ is any vector d such that:

$$\exists t_0 > 0 \quad : f(x+td) < f(x) \text{ et } g(x+td) < g(x) \ \forall \ t \in]0, t_0]$$

- When does a common vector exist?

First case: smooth functions

Multiobjective optimization

$$\min_x \{J_1(x), J_2(x), ..., J_n(x)\} \; ; \; x \in \mathcal{K}$$

- $x = (x_1, ..., x_N) \in \mathbb{R}^N$ are the optimization parameters.

Definition

A function f d $\mathbb{R}^n \to \mathbb{R}$ is convex if for any $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

Assumptionss:

- $J_k \; ; \; k = 1, n$ are convex
- $J_k \; ; \; k = 1, n$ are derivable, $\nabla J_k(x)$ denotes their gradient at point x.

Convex hull

Definition

let X_i ; $i = 1, n$ be a family of n vectors in $\mathbb{R}^N$. The convex hull of this vector family is :

$$\left\{ w \in \mathbb{R}^N : w = \sum_{i=1}^{n} \alpha_i X_i \;;\; \alpha_i \geq 0 \;;\; \sum_{i=1}^{n} \alpha_i = 1 \right\}.$$

➤ In the following we consider the convex hull $\mathcal{U}$ generated by the objective function gradients::

Convex hull construction

$$\mathcal{U} = \left\{ w \in \mathbb{R}^N : w = \sum_{i=1}^{n} \alpha_i \nabla J_i(x) \;;\; \alpha_i \geq 0 \;;\; \sum_{i=1}^{n} \alpha_i = 1 \right\}.$$

Definition

A point $x \in \mathbb{R}^N$ is said to be Pareto stationary if:

$$\exists \lambda_i \; ; \; \lambda_i \geq 0 \; ; \; \sum_{i=1}^{m} \lambda_i = 1 \; ; \; \mid \sum_{i=1}^{m} \lambda_i \nabla J_i(x) = 0.$$

Lemma

There exists in the convex hull $\mathcal{U}$ an unique vector $p^* = \mathrm{Argmin}_{p \in C} ||p||$ such that

$$\forall p \in \mathcal{U} \; : \; p^T p^* \geq ||p^*||^2.$$

Theorem

Vector p^* has the following properties:

1. *either $p^* = 0$ which implies that x is Pareto stationary*
2. *or $p^* \neq 0$ which implies that $-p^*$ is a common descent vector for the objectives $J_k \; ; \; k = 1, n$.*

Convex hull construction

$$\mathcal{U} = \Big\{ w \in \mathbb{R}^N : w = \sum_{i=1}^{n} \alpha_i \nabla J_i(x) \; ; \; \alpha_i \geq 0 \; ; \; \sum_{i=1}^{n} \alpha_i = 1 \Big\}.$$

Theorem

A necessary condition for x^* to be a Pareto optimal solution is:

$$\exists \lambda_i \geq 0 \; : \; \sum_{i=1}^{m} \lambda_i \nabla J_i(x) = 0.$$

The condition is sufficient when $\lambda_i > 0, \; i = 1, ..., m$.

Algorithm

MGDA (Désidéri 2012)

1. At step n calculate $\nabla J_i(x_n^0)$; $i = 1, n$.
2. Construct the common descent direction p_n^* by solving the quadratic problem $p_n^* = \mathrm{Argmin}_{p \in C} ||p||$
   - If $p_n^* = 0$: STOP.
   - In the contrary calculate the step h_n.
3. Actualize the current point $x_{n+1}^0 = x_n^0 - h_n \times p_n^*$.

Non-regular context

- Many practical problems in real life involve non regular objective functions:
  - minimize the maximal force on a structure
  - minimize eigen frequencies of a rotating structure
  - shape optimization when obstacles are taken into account
  - optimization problems involving contacts between object (robot)
- Nonsmooth analysis gives a proper theoretic background to tackle the problem of nonsmooth optimization.
- The notion of derivative or gradient is replaced by the notion of sub-differential and sub-gradient.

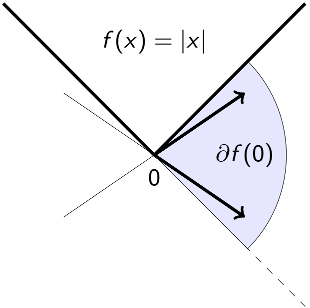
The convex case

➤ A convex function is underestimate at each point by an affine function which defines the equation of an hyperplan. If the function is not differentiable at that point then there exist an infinity of such hyperplane which directions form the sub-differential:

Definition

The sub-differential of function f at point x is the set

$$\partial f(x) = \{s \in \mathbb{R}^n : f(y) \geq f(x) + < s, y - x > \forall y \in \mathbb{R}^n\} \tag{25}$$



Sub-differential calculus example

- Consider the following non-smooth function:

$$f(x) = \max\{f_1(x), f_2(x), ..., f_n(x)\},$$

where the functions f_i ; $i = 1, n$ are derivable. Then the sub-differential of function f is the set

$$\partial f(x) = conv\{\nabla f_k(x) \mid f_k(x) = f(x)\}$$

Optimum characterization

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ a convex function . The following statements are equivalent:

1. *f is minimal at point x^*: $f(y) \geq f(x^*)\ \forall y \in \mathbb{R}^n$,*
2. *$0 \in \partial f(x^*)$.*

Common descent direction for convex functions

Convex hull construction

$$\mathcal{U} = \left\{ w \in \mathbb{R}^N : w = \sum_{i=1}^{n} \alpha_i \xi_i(x) \; ; \; \alpha_i \geq 0 \; ; \; \sum_{i=1}^{M} \alpha_i = 1 \; ; \; \xi_i \in \partial J_i(x) \right\}.$$

Lemma

There exists in the convex hull $\mathcal{U}$ an unique vector $p^* = \mathrm{Argmin}_{p \in C} ||p||$ such that

$$\forall p \in \mathcal{U} \; : \; p^T p^* \geq ||p^*||^2.$$

Theorem

Vector p^* has the following properties:

1. *either $p^* = 0$ which implies that x is Pareto stationary*
2. *or $p^* \neq 0$ which implies that $-p^*$ is a common descent vector for the objectives $J_k \; ; \; k = 1, n$.*

Multiobjective robust optimization

Mltiobjective optimization problem with uncertainty

$$\min_x \{J_1(x, \xi(\omega)), J_2(x, \xi(\omega)), ..., J_m(x, \xi(\omega))\}$$

$x \in \mathcal{K} \subset \mathbb{R}^N$, $\xi(\omega) \in L^0(\Omega, \mathbb{R}^d)$ a random vectore

Robust formulation

$$\min_x \{E[J_1(x, \xi(\omega))], E[J_2(x, \xi(\omega))], ..., E[J_m(x, \xi(\omega))]\}$$

Mercier, Poirion, Désidéri (2018)

- Choose an initial point X_0
- Choose the number of iteration N
- Choose a σ-sequence $\{t_k\}_{k\in\mathbb{N}}$, $\sum t_k = \infty$; $\sum t_k^2 < \infty$
- For $k = 1, N$
  - Generate a sample ξ_k of the random variable $\xi(\omega)$
  - Evaluate the objective functions $(X_{k-1}, \xi_k) \longrightarrow J_i(X_{k-1}, \xi_k)$
  - Construct the common descent vector $d(X_{k-1}, \xi_k)$
  - Actualize: $X_k = X_{k-1} - t_k d(X_{k-1}, w_k)$.
- Repeat the procedure starting from another initial point in order to describe the Pareto set.

Convergences

Theorem

1. *The sequence of random variables $X_0, X_1, ..., X_n$ converges in mean square towards a point X^* located on the Pareto set P:*

$$\lim_{k \to +\infty} E[\|X_k - X_k^\star\|^2] = 0.$$

2. *The sequence converges also almost surely towards X^* :*

$$\mathbb{P}\left(\left\{\lim_{k \to \infty} X_k - X_k^\star = 0\right\}\right) = 1.$$

Numerical illustration

- Fonseca & Fleming benchmark (stochastic version)

$$\min_{x\in[-1,1]^3} \begin{cases} E[f_1] = E\left[1 - \exp\left(-\sum_{i=1}^{3}(x_i - \frac{W_{1i}}{\sqrt{3}})^2\right)\right] \\ E[f_2] = E\left[1 - \exp\left(-\sum_{i=1}^{3}(x_i + \frac{W_{2i}}{\sqrt{3}})^2\right)\right] \end{cases} \tag{26}$$

with $\{W_{1i}\}_{i\in[1,3]}$ et $\{W_{2i}\}_{i\in[1,3]}$ 6 independent uniform random variables $\mathcal{U}(0.3, 1.7)$.

- Initialization using 200 points in $[-1, 1]^3$.
- 300 iterations per points.
- Comparison with the genetic algorithm (NSGAII) using the same number of function calls.
- NSGAII is applied to the following objective estimator:

$$E[f_i(x, \xi(\omega)] \approx 1/N \sum_{k=1}^{N} f_i(x, \xi(\omega_k))$$

- SMGDA is parallelizable.

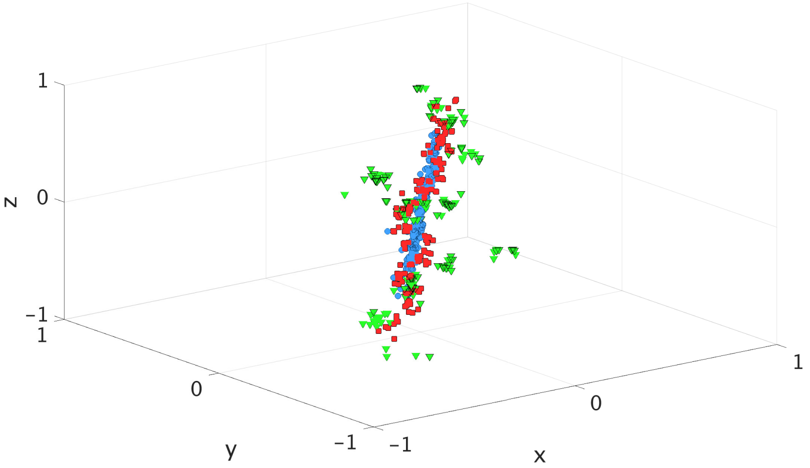


Figure: Design space of the Fonseca & Fleming test case — NSGA-II 300 calls ; NSGA-II 310^5 calls ; SMGDA 300 calls

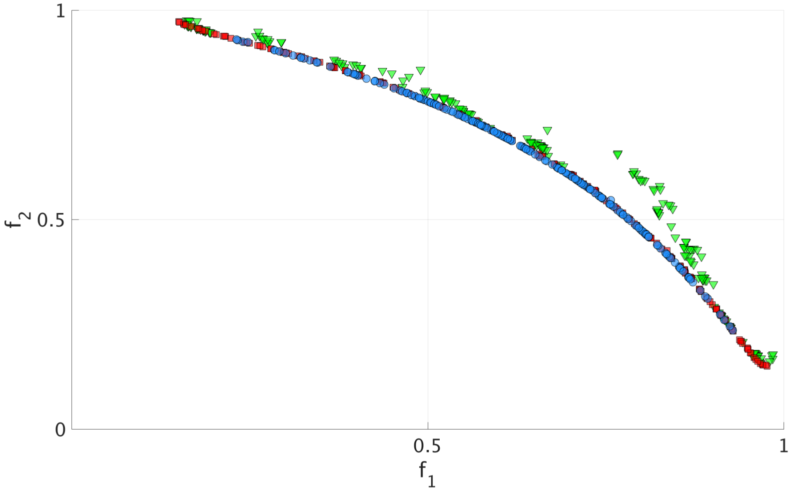
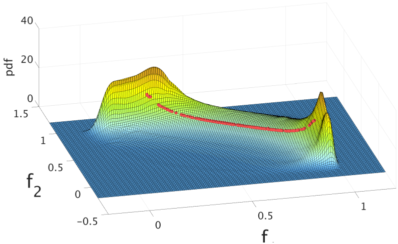


Figure: Objective space of the Fonseca & Fleming test case — NSGA-II 300 calls ; NSGA-II $3 10^5$ calls ; SMGDA 300 calls



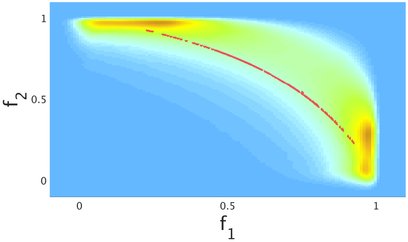


Figure: Probability density of final solutions — Pareto front of the mean problem

$$\left[f_1(x^*, W(\omega), f_2(x^*, W(\omega)\right]$$

Another example

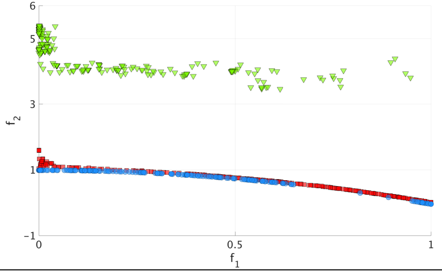
Problem	Results
ZDT2 $$\min_{\mathbf{x}\in[0,1]^{30}} \begin{Bmatrix} \mathbb{E}[x_1] \\ \mathbb{E}[(1+W_2)g(\mathbf{x})h(\mathbf{x})] \\ \mathbb{E}[g(\mathbf{x})] \\ \mathbb{E}[h(\mathbf{x})] \end{Bmatrix}$$ $$\text{With} \begin{cases} W_g \to \mathcal{U}(-.3,.3) \\ W_2 \to \mathcal{U}(-.5,.5) \\ g(\mathbf{x}) = 1 + 9\sum_{i=2}^{30} \frac{x_i}{29} + W_g \\ h(\mathbf{x}) = 1 - \left(\frac{x_1}{g(\mathbf{x})}\right)^2 \end{cases}$$	

Table: Results of the benchmark tests — NSGA-II 300 calls ; NSGA-II $3 10^5$ calls ; SMGDA 300 calls

RBDO formulation

$$\operatorname*{Argmin}_{x \in X^{\mathbf{ad}}} \{ E[f(x,\xi(\omega))] \,|\, P[g(x,\xi(\omega)) \geq 0] \leq p_0 \}.$$

- $P[g(x,\xi(\omega)) \geq 0] \leq p_0 = E[G(x,\xi(\omega))] \leq p_0$ with $G(x,\xi(\omega)) = \mathbb{I}_{\mathbb{R}_+}(g(x,\xi(\omega)))$
- Instead of solving the former problem we solve the following multiobjective problem

Substitute multiobjective problem

$$\operatorname*{Argmin}_{x \in X^{\mathbf{ad}}} \{ E[f(x,\xi(\omega))],\ E[G(x,\xi(\omega))] \}.$$

- Construct the related Pareto set.
- For a given p_0 the RBDO solution $x^\star$ lies on the Pareto set in the solution domain and the point $(\mathbb{E}[f(x^\star,\xi(\omega))], p_0)$ lies on the Pareto front in the objective space.

Application to reliability problems

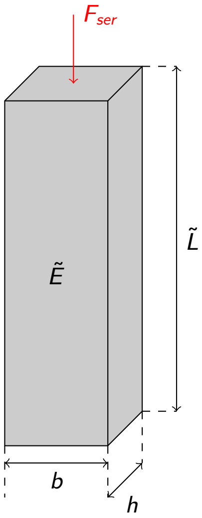
Substitute multicriteria problem

$$\operatorname*{Argmin}_{x \in X^{\mathbf{ad}}} \{E[f(x,\xi(\omega))],\ E[G(x,\xi(\omega))]\}.$$

- The numerically challenging calculation of the constraint $P[g(x,\xi(\omega)) \geq 0] \leq p_0$ is replaced by the calculation of $\partial G(x,\xi(\omega))$ which appears in the SMGDA algorithm.
- The Pareto set gives the best Cost/Reliability compromise for the solutions (efficient frontier).
- Decouple the optimization problem and the evaluation of the failure probability.

Drawbacks:

- Need nevertheless to construct the Pareto set in the objective domain $(\mathbb{E}[f(x,\xi(\omega))],\ P[g(x,\xi(\omega)) \geq 0])$ in order to interpret the Pareto set, which means to evaluate the probability of failure for all the solution x found by the algorithm.
- This can be done using specific classical approaches: surrogate models,subset simulation, etc.
- Need to calculate the objective function derivatives: could be done using adjoint formulation for real life systems



Objectives :

1. Minimize cross section size $b \times h$
2. Minimize probability of failure : $\mathbb{P}[F_{ser} - \tilde{k}\frac{\pi^2 \tilde{E} bh}{12\tilde{L}^2} \leq 0]$

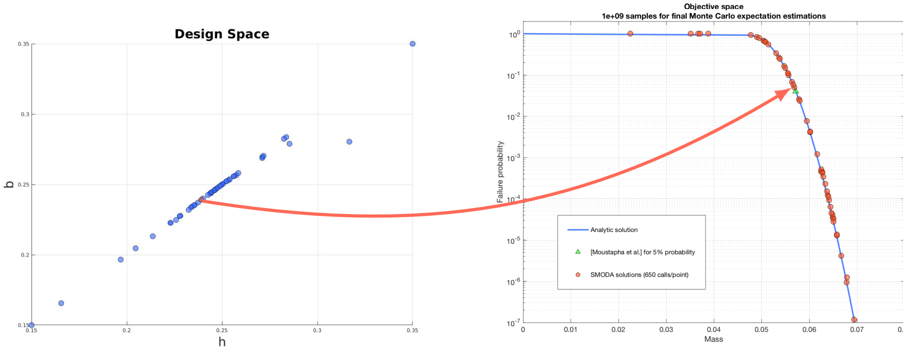
Design variables : $x = (b, h) \in [150, 350]^2$

Uncertainty: $W = (\tilde{k}, \tilde{E}, \tilde{L})$, lognormal distribution with (10,5,1)% of covariance

Constraint : $b \geq h$

Problem formulation

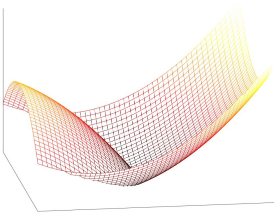
$$\begin{cases} \min_{x \in \mathcal{X}_{ad}} \left\{ \mathbb{E} b \times h, \mathbb{E}\mathbb{I}_{\{F_{ser} - \tilde{k}\frac{\pi^2 \tilde{E} bh}{12\tilde{L}^2} \leq 0\}} \right\} \\ h \leq b \end{cases}$$



Design Space
b
h
0.15
0.2
0.25
0.3
0.35
Objective space
1e+09 samples for final Monte Carlo expectation estimations
Failure probability
Mass
Analytic solution
[Moustapha et al.] for 5% probability
SMODA solutions (650 calls/point)

The nonconvex case I

Nonconvex function

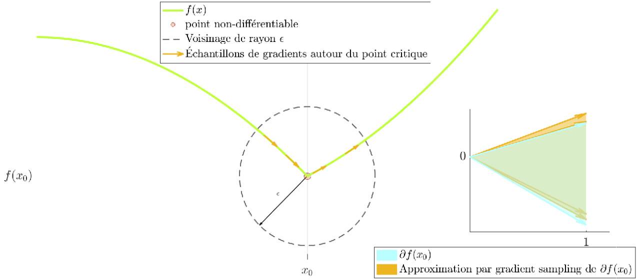


When the function is no longer convex but is locally Lipschitz continuous the notion of subdifferential has to be replaced by the notion of Clarke subdifferential . The Clarke subdifferential at point x is the set containing all the convex combinations of limits of gradients at points located in the neighborhood of x:

$$\partial f(x) = \text{conv}\left\{ \lim_{i \to \infty} \nabla f(x_i) \ ; \ x_i \to x \text{ and } \nabla f(x_i) \text{ exists} \right\}. \tag{27}$$

The nonconvex case II

Clarke subdifferential construction



The nonconvex case III

- Formal definition

Definition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function. The generalized directional derivative of f at x in the direction $v \in \mathbb{R}^n$ is defined by:

$$f^o(x; v) = \limsup_{y \to x; t \downarrow 0} \frac{f(y + tv) - f(y)}{t}.$$

Definition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function. The Clarke subdifferential of f at x is the set $\partial f(x)$ of vectors defined by:

$$\partial f(x) = \{s \in \mathbb{R}^n : f^o(x; v) \geq s^T v \; \forall v \in \mathbb{R}^n\}. \tag{28}$$

The nonconvex case IV

- For smooth functions it is well known that the opposite direction of the gradient is a descent vector.
- For nonsmooth or nonconvex functions an arbitrary subgradient does not necessarily gives a descent direction.
- There exist several techniques for constructing a descent direction: proximal bundle methods (Makela *et al*), quasisecant methods (Bagirov *et al*), or gradient sampling methods (Burke *et al*).
- This last last approach is simple to implement :

$$\partial f(x) \approx C_k = \text{conv}\{\nabla f(x_i)\ ;\ i = 1, 2, ..., k\}\ ;\ x_i \in B(x, \epsilon). \tag{29}$$

where x_i are points chosen randomly in a neighborhood of x.

- The minimum norm element of the set C_k is then used as a descent direction for $f(x)$.

The nonconvex case V

- The stochastic gradient algorithm can be used for solving nonconvex robust optimization problems replacing the gradient by the Clarke's derivative.
- In the same way the SMGDA algorithm can be extended to nonconvex multiobjective robust optimization problem.

References I

R. Arnaud and F. Poirion.
Optimization of an uncertain aeroelastic system using stochastic gradient approaches.
Journal of Aircraft, doi: arc.aiaa.org/doi/abs/10.2514/1.C032142., 2014.

J.A. Désidéri.
Multiple-gradient descent algorithm (MGDA) for multiobjective optimization.
CRAS Paris, Ser. I, 350:313–318, 2012.

Q. Mercier, F. Poirion and J.A. Désidéri
A stochastic multi objective gradient algorithm
European Journal of Operational Research ,2018.

F. Poirion, Q. Mercier, and J.A. Désidéri.
Descent algorithm for nonsmooth stochastic multiobjective optimization.
Computational Optimization and Applications, pages 317-331 68(2), 2017.

Quentin Mercier.
Optimisation multicritère sous incertitudes : un algorithme de descente stochastique.
Université Côte d'Azur, 2018.

References II

Robbins, H. and S. Monro (1951).
A stochastic approximation method.
Ann. Math. Statistics 22, 400–407.

Clarke, F. (1983).
Optimization and Nonsmooth Analysis.
Wiley.

Bagirov, A., N. Karmitsa, and M. Mäkelä (2014).
Introduction to Nonsmooth Optimization: Theory, Practice and Software.
Springer.

M. M. Mäkelä.
New Multiple Subgradient Descent Bundle Method for Nonsmooth Multiobjective Optimization.
Tech. report TUCS, 2014.

Bagirov, A., L. Jin, N. Karmitsa, A. A. Nuimat, and N. Sultanova (2013).
Subgradient method for nonconvex nonsmooth optimization.
Journal of Optimization Theory and Applications 157, 416–435.

References III

Burke, J. V., A. S. Lewis, and M. L. Overton (2002).
Approximating subdifferentials by random sampling of gradients.
Mathematics of Operation Research 27, 567–584.

Burke, J. V., A. S. Lewis, and M. L. Overton (2005).
A robust gradient sampling algorithm for nonsmooth, nonconvex optimization.
SIAM J. Optim. 15, 751–779.

Mäkelä, M., N. Karmitsa, and O. Wilppu (2016).
Mathematical Modeling and Optimization of Complex Structures, Chapter Proximal Bundle Method for Nonsmooth and Nonconvex Multiobjective Optimization, pp. 191–204.
Springer.

Miettinen, K. (1998).
Nonlinear Multiobjective Optimization, Volume 12 of *International Series in Operations Research & Management Science*.
Springer US.

References IV

Moustapha, M., B. Sudret, J. Bourinet, and B. Guillaume (2016).
Quantile-based optimization under uncertainties using adaptive Kriging surrogate models.
Structural and Multidisciplinary Optimization 54, 1403–1421.

Deb, K. and H. Gupta (2005).
Searching for robust Pareto-optimal solutions in multi-objective optimization.
In *Evolutionary Multi-Criterion Optimization*. Springer.

Ray, T., K. Tai, and K. Chyeseow (2007).
Multiobjective design optimization by an evolutionary algorithm,.
Engineering Optimization 33, 399–424.

Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan (2002).
A fast and elitist multiobjective genetic algorithm: NSGA-II.
IEEE Transactions on Evolutionary Computation 6, 182 – 197.

OpenTurns
No references !! :(:(