Robust Optimization in pSeven

Robust single- and multi-objective constrained optimization in pSeven supports virtually all possible robust formulations, including probabilistic and quantile type constraints. It is based on the original state-of-the-art stochastic approach developed by DATADVANCE. The basis of the method is a careful adjustment of current number of sampled uncertain parameters. Only a small number of random realizations away from optimality is to consider, but their number must increase once optimal solution is approached. The unique feature of this family of algorithms is that they provide both the solution and corresponding uncertainty estimates of objective/constraints values at the solution found. A particular advantage of robust optimization algorithms in pSeven for engineering applications is that explicit distribution law of the uncertain parameters is not required. It is enough to provide their distribution empirically.

I. Overview

Robust optimization tasks emerge when some external to original formulation parameters are not well known. Uncertain model parameters make optimization problem not well defined. Indeed, if we ever admit "random" dependencies of problem data (objectives, constraints), then the problem solution becomes also random and hence has no meaning. One must exclude randomness from the problem. It is common to consider various expectations and moments of relevant distribution. Optimization task reformulated respectively is known as robust counterpart of the original problem. Worst-case scenarios are known to be too restrictive and thus not considered by pSeven. Instead, randomness is assumed to be eliminated by taking appropriate expectation values.

Unfortunately, none of these expectations could be calculated, the only possibility is to approximate them. Then a prime problem is the degree of approximation accuracy. Rough but cheap approximations would be good to consider, but in the same time, high accuracy should be provided to properly represent hypothetical original problem. In this problem expectations are calculated exactly (the term "original" refers to such hypothetical "exact" formulation.) Usually, the optimal strategy is adaptive: the closer optimal solution is, the more accurate approximations must be. Side effect of inability to explicitly consider the original problem is that only approximate (candidate) solutions can be obtained. So every candidate solution must be validated to ensure its closeness to the solution of original problem. This is an example of robust optimization specific features, not present in non-robust cases.

Another issue within the robust design optimization framework is the presence of new type of constraints, so called chance constraints. These are hard to consider, because their accuracy evaluation (even moderate) is extremely expensive, while corresponding approximations lead to strongly discontinuous subproblems. Proposed solution is to construct smooth over-estimators of chance constraints, so that the corresponding accuracy increases as algorithm proceeds.

II. Quantitative Discussion

A. *Generic Formulation of Robust Optimization Problems and its Stochastic Approximation.*

In the majority of practical applications "classical" formulation of optimization problem

$$\min_x \vec{f}(x)$$
$$\vec{c}_L \le \vec{c}(x) \le \vec{c}_U$$

is not adequate. In reality, problem data f, c additionally depend upon external "random" parameters

ξ from some uncertainty set U_ξ:

$$f(x) \to f(x,\xi) c(x) \to c(x,\xi) \xi \in U_\xi$$

In this circumstances, the optimal solution $x^* x_*$ becomes also dependent upon "random" realizations of ξ $x^* \to x^*(\xi)$ and hence is devoid of meaning. Therefore, the dependence upon external parameters must be considered at the level of problem formulation.

$$\min_x \vec{f}(x,\xi) \quad \to x^*(\xi)$$
$$\vec{c}_L \le \vec{c}(x,\xi) \le \vec{c}_U$$

Evidently, one should get rid of ξ-dependence of the solution $x^*(\xi) \longrightarrow x^*$. The decision on what is the optimal solution x_* must be taken as soon as possible, before the specific realization of ξ is known!

Depending on the specific meaning of the reduction $x^*(\xi) \to x^*$, various formulations of robust optimization problems are possible. For instance, if the probability distribution $\rho(\xi)$ of uncertain parameters is known, we could consider

$$f(x,\xi) \to F(x) \equiv E[f(x,\xi)] = \int d\xi \rho(\xi) f(x,\xi)$$

Constraints are more tricky. There are several possibilities:

- Expectation constraints: $\vec{c}_L \le \vec{E}[c(x,\xi)] \le \vec{c}_U$
- Chance constraints: $P\{c_L^i \le c^i(x,\xi) \le c_U^i\} \ge 1 - \alpha \alpha \in (0:1)$

pSeven adapts stochastic approximation to the averages in question. Advantages of stochastic approach:

- Details and specific form of probability densities $\rho(\xi)$ are completely delegated to end-user (Note that in real-life applications determination of appropriate $\rho(\xi)$ might be very difficult)
  * It is enough to know only empirical distribution $\rho(\xi)$ without any analytical equations
  * Uncertainty set U_ξ is never used explicitly
- Flexibility: stochastic formulation is suitable for virtually all cases

However, there are a few notable disadvantages as well:

- Efficient solution of stochastic optimization problems requires specific approaches. Therefore, pSeven considers the following general formulation of robust optimization problems:

$$\min_x E[f(x,\xi)]$$
$$c_L \le E[c(x,\xi)] \le c_U \text{and/or} P\{c_L \le c(x,\xi) \le c_U\} \ge 1 - \alpha$$

The main difficulty is that none of expectations $E\{...\}$ or probabilities $P\{...\}$ could ever be calculated precisely. Then the question arises: what is the meaning of solution of stochastic problem alone? It is not enough to simply solve approximated problem. One must ensure that the approximate solution found is indeed "almost" optimal and close to the solution of original problem. Therefore, the prime difficulty in solving stochastic task is the fact that one has to not simply find the solution, but also validate the proposed candidate.

The idea of stochastic approximation utilized in pSeven is simple:

$$E[f(x,\xi)] \approx \frac{1}{N} \sum_i \quad f(x,\xi_i) \equiv f_N(x)$$

where $\{\xi_i\}$, $i = 1,\ldots,N$ is a finite ξ-sample, given externally and presumably distributed with $\rho(\xi)$.

Key questions to be discussed:

- Since sample size N is always finite -- what might be "the answer" for the original problem?

- How large *N* must be taken? How it depends upon optimality of current approximate solution?

- What should be done with chance constraints since at finite *N* straightforward approximation $P\{g(x,\xi) \geq 0\} \approx \frac{1}{N}\sum_i \quad \Theta(g(x,\xi_i)) \equiv p_N(x)$ is discontinuous (discontinuities being of order ~1/*N*).

B. *The Meaning of Approximated Problem Solution*

Speaking about the meaning of solution approximated with finite sample, it should be noted that it is natural to consider two characteristics for any solution *x* of arbitrary optimization problem:

- Feasibility measure $\psi(x)$
- Optimality measure $\theta(x)$

$$\theta(x) = \min_{\lambda_i \geq 0} |\nabla f(x) + \sum_i \quad \lambda_i \nabla c^i(x)|^2$$

which is the norm of KKT conditions residual.

In the stochastic context, we should also consider corresponding stochastic approximations ψ_N, θ_N.

Therefore, the most complete answer available in stochastic approach reduces to:

- Probability that feasibility measure of approximate solution x_N^* is greater than a given tolerance $\delta\psi$
- is less than predefined threshold Y% $P\{\psi(x_N^*) > \delta\psi\} \leq Y\%$
- Probability that optimality measure of approximate solution x_N^* is greater than a given tolerance $\delta\psi$
- is less than predefined threshold X% $P\{\theta(x_N^*) > \delta\theta\} \leq X\%$

Parameters $\delta\psi$ and $\delta\theta$ are to be given externally. Quantiles Y% and X% might be taken as constant (commonly accepted value is 5%).

Now the generic scheme to solve stochastic optimization problem can be summarized as follows:

1. Approximate original problem for sufficiently large sample size *N*.

2. Solve approximated task to obtain candidate solution x_N^* with usual means. Question: since the solution of approximated subproblem is also approximate - what are the appropriate tolerances ϵ_N^ψ, ϵ_N^θ to be used in finding candidate x_N^*?

3. Check statistics of obtained candidate: ensure that both expectations $((\delta\theta_N, \delta\psi_N)$ and statistical uncertainties $((\delta\theta_N, \delta\psi_N))$ are sufficiently small.

4. If statistical precision of candidate is insufficient, enlarge sample size and continue with next cycle.

C. *Efficient Sample Size Selection*

The central question of this approach is how to select current sample size efficiently. To address it, let us note that there are two different sources of uncertainties at every stage of stochastic problem solution: statistical uncertainties, $\sim 1/\sqrt{N}$ and "systematic" uncertainties (tolerances of optimization at every fixed *N*). Usually, optimal strategy is to keep both statistical and systematic errors of the same order. This simple observation allows to unambiguously fix required tolerances $\epsilon_N^\psi, \epsilon_N^\theta$ to obtain candidate solution and current sample size. Below is the scheme of efficient algorithm to solve stochastic problem:

1. Set initial (relatively small) sample size *N* [heuristics] and *μ*=0

2. Consider the statistics at initial point $x_{N,\mu}^*$:

- calculate statistical uncertainties $(\delta\theta_{N,\mu}, \delta\psi_{N,\mu})$
- set the tolerances $\epsilon_{N,\mu}^\psi \sim \delta\psi_{N,\mu}$, $\epsilon_{N,\mu}^\theta \sim \delta\theta_{N,\mu}$ for optimization at the initial stage

3. Optimization of *N*-approximated problem with tolerances($\epsilon^{\psi}_{N,\mu}$, $\epsilon^{\theta}_{N,\mu}$) obtaining the next candidate solution $x^*_{N,\mu}$.

4. Validation of candidate solution $x^*_{N,\mu}$: statistics estimate in $x_{*N,\mu}$, evaluation of ($\delta\theta_{N,\mu}$, $\delta\psi_{N,\mu}$). finished if statistics is good enough.

5. Estimation of required sample size N_{next} for the next iterate

$N_{next} = N \cdot$ heuristics$\{(\frac{\delta\theta_N}{\delta\theta})^2, (\frac{\delta\psi_N}{\delta\psi})^2, x^*_{\mu-1}, x^*_{N,\mu}\}$

- $\delta\theta$, $\delta\psi$ are the required precisions
- heuristics keeps N_{next} in reasonable limits, depends upon the progress made

6. Rescaling of optimization tolerances with factor $\sim \sqrt{N/N_{next}}$ which expresses "statistics" ≈ "systematics"

7. New iteration (3) for *N*-approximated problem with $N = N_{next}$.

***D**. Chance Constraints Treatment*

The idea of efficient chance constraints treatment is to consider continuous majorant of the original discontinuous stochastic approximations, making approximation more accurate with increasing *N*. In equations:

$$P\{g(x,\xi) \geq 0\} \equiv E[\Theta(g(x,\xi))] \leq \alpha$$

$$E[\Theta(g)] \leq E[R_\varepsilon(g)] \approx \frac{1}{N}\sum_i R_\varepsilon(g_i)$$

where $\Theta(x)$ is a step function, $R_\varepsilon(x)$ is a (smooth) majorant of $\Theta(x)$ and $\varepsilon = \varepsilon(N)$

$$\Theta(x) \leq R_\varepsilon(x) \forall x, \varepsilon$$

$$\lim_{\varepsilon\to 0} R_\varepsilon(x) = \Theta(x) \forall x$$

$$\lim_{N\to\infty} \varepsilon(N) = 0$$

Issue to mention here (it is successfully resolved in pSeven) concerns apparent conformal (scale) non-invariance of proposed approach:

$$\forall \tau > 0: \Theta(\tau x) \equiv \Theta(x) R_\varepsilon(\tau x) \neq R_\varepsilon(x)$$

The point is that regularization revives former zero mode and this is taken into account. Solution is to dynamically adjust the scale parameter:

$$E[R_\varepsilon(g))] \leq \alpha \Longrightarrow \min_{t\geq 0} E[R_\varepsilon(g/t))] \leq \alpha$$

or even better (to avoid singularity at $t = 0$

$$\min_{t\geq 0} [E[tR_\varepsilon(g/t))] - \alpha t] \leq 0$$

Another issue arises when regularized problem appears to be infeasible. One can see that it does not follow from $E[\Theta(g)] > \alpha$ that $E[\Theta(g)] > \alpha$ for any positive $\boldsymbol{\varepsilon}$. Hence, the decision on infeasibility of original problem must be delayed until regularization is completely removed.

***E**. Notes on Statistics Estimation*

The efficiency of the approach to robust optimization crucially depends upon the ability to perform statistics estimation in most economical way. Potentially, it requires a lot of additional model evaluations without any progress in design variables. One of the standard approaches considers auxiliary additionally generated ensemble $\{x^*_N\}_i \;\; i = 1,..,M$, which is extremely expensive (requires $\sim (M-1)\cdot N \gg N$ additional evaluations). Solution implemented in pSeven is to use a well-known jackknife method, which does not require additional evaluations at all.

***F**. Couple of Illustrative Examples*

1. Quantile optimization example taken from: B.~Lu, "Theory and practice of uncertain programming" (2009),112.

$$\min\phi$$

$$c_1 \equiv P\{0 \leq \sum_{i=1}^{3} \zeta_i x_i + \phi\} \geq 0.9$$

$$c_2 \equiv P\{\sum_{i=1}^{3} \eta_i x_i^2 \leq 8\} \geq 0.8$$

$$c_3 \equiv P\{\sum_{i=1}^{3} \tau_i x_i^3 \leq 15\} \geq 0.85$$

$$\zeta_1 \sim U(1,2)\eta_1 \sim U(2,3)\tau_1 \sim U(3,4)$$
$$\zeta_2 \sim N(1,1)\eta_2 \sim N(2,1)\tau_2 \sim N(3,1)$$
$$\zeta_3 \sim Exp(1)\eta_3 \sim Exp(2)\tau_3 \sim Exp(3)$$

$U(a,b)$: uniform distribution on [*a*;*b*];

$N(a,b)$: normal distribution with mean *a* and variance *b*;

$Exp(a)$: exponential distribution with parameter *a*.

In the reference cited above the genetic algorithm with meta-modeling was used, the result is:

$$\underline{1.5 \cdot 10^7 \text{evaluations}}$$
$$1x^* \approx (1.404,0.468,0.924) f^* \approx -2.21$$
$$c_1 \approx 0.9 c_2 \approx 0.8 c_3 \approx 0.85$$

Note that the meaning of "≈" was not clarified.

Application of pSeven approach with required tolerances $\delta\theta, \delta\psi = 2.5\%$ gives:

$$x^* = (1.53,0.44,0.61) f^* = -2.22(3)$$
$$c_1 = 0.92(2) c_2 = 0.85(2) c_3 = 0.88(2)$$

with only $\approx 4.2 \cdot 10^4$ evaluations, which by 3 orders of magnitude smaller than the number cited above.

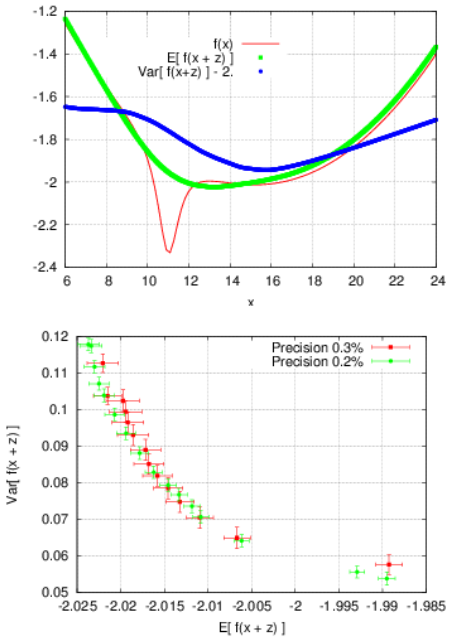
2. Robust bi-objective optimization

Let $f(x) = -\frac{1}{5(11-x)^2+2} + \frac{1}{120}(x-15)(x-16) -$

2. Then the problem is:

$$\min(E[f(x+\xi)], Var[f(x+\xi)])$$
$$x \in [0{:}\,28] \xi \sim N(0,2)$$

Pareto set is known to be located in $x \in [13.2; 15.75]$.



Depending on externally given tolerances, pSeven robust optimization algorithms require $\delta\theta = 0.3\%{:}\,6 \cdot 10^5 \delta\theta = 0.2\%{:}\,1.1 \cdot 10^6$ evaluations, which have to be compared with typical evaluation count of Monte-Carlo based methods, $(5 \div 8) \cdot 10^6$.

III. Notes on Surrogate-Based Robust Optimization in pSeven

It is straightforward to apply pSeven surrogate based optimization capabilities to robust optimization problems. Indeed, one has to repeatedly solve a sequence of *N*-sample approximated subproblems within the stochastic approximation approach, which are expensive to evaluate (because of presumably large *N*) and thus surrogate modeling techniques are required. Surrogate-based robust optimization in pSeven follows this scheme, however, there are a few notable issues:

1. There is no need to perform chance constraints regularization for SBO-type approaches. Surrogate models are almost insensitive to local discontinuities in approximated probabilities until the last stages of optimization, when sample size is hopefully large enough to mitigate the issue completely.

2. Application of surrogate modeling techniques supplemented with constantly growing sample size requires an additional care to properly reuse already evaluated designs. Although this issue is rather technical, it allows in some cases a great reduction of required number of evaluations.