CVaR Approximation to Chance-Constrained Programs: What Is Lost and How to Find It Back?

L. Jeff Hong[1] Liwei Zhang[2]

[1]Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology

[2]School of Mathematical Science
Dalian University of Technology, China

April 2010

Constrained Convex Program

Basic formulation:

$$\begin{aligned} \min \quad & h(x) \\ \text{s.t.} \quad & c_1(x,\xi) \le 0, \cdots, c_m(x,\xi) \le 0, \\ & x \in X. \end{aligned}$$

- x is the vector of decision variables and ξ is the vector of parameters.
- $h(x)$ and $c_i(x,\xi), i = 1, \cdots, m$, are differentiable and convex in x, and X is a compact convex set.
- can be solved efficiently (see, for instance, Boyd and Vandenberghe (2004) for a comprehensive introduction).

Parameter Uncertainty

In many practical applications, there is parameter uncertainty in ξ. How should we handle that?

- A often used method: using $\mathrm{E}(\xi)$ to substitute ξ.

- A very simple example:

$$\min\ x, \quad \text{s.t.} \quad x \geq \xi_i,\ i = 1, 2, \cdots, m,$$

  where $\xi_1, \cdots, \xi_m$ are independent $N(0, 0.01)$ observations.
  Suppose that we use the mean of ξ in the optimization, then $x^* = 0$.
  However, $\Pr\{x^* \geq \xi_i,\ i = 1, 2, \cdots, m\} = 0.5^m$.

- The optimal solution of a constrained optimization problem is often at the corner of the feasible region, where multiple constraints are tight. Small perturbations to the constraints can easily make the solution infeasible.

Chance-Constrained Program

A way to handle the parameter uncertainty is to reformulate the problem to

$$\begin{aligned} \min \quad & h(x) \\ \text{s.t.} \quad & \Pr\{c_1(x,\xi) \le 0, \cdots, c_m(x,\xi) \le 0\} \ge 1-\alpha, \\ & x \in X. \end{aligned}$$

- The formulation is intuitive and easy to understand, although other formulations are possible.
- There are abundant applications, e.g., power network design, reservoir design, cash matching problem.
- When $m = 1$, the constraint is called a single chance constraint; when $m \ge 2$, it is called a joint chance constraint.

The Difficulties

Let $p(x) = 1 - \Pr\{c_1(x, \xi) \leq 0, \cdots, c_m(x, \xi) \leq 0\}$ and $p(x)$ is the probability that at least a constraint is violated. Then, the joint chance constraint becomes $p(x) \leq \alpha$.

There are two major difficulties in solving JCCPs:

- $p(x)$ may not be convex (or quasi-convex).
- $p(x)$ typically has no closed forms and is difficult to evaluate.

Literature Review

Convexity of the chance constraint:

- If $c_1(x,\xi), \cdots, c_m(x,\xi)$ are quasiconvex functions of (x,ξ), and ξ has a logconcave probability distribution, then $p(x)$ is quasiconvex and hence the JCCP is convex (Prékopa 2003).
- An individual chance constraint in the form of $\Pr\{a^T x \le b\} \ge 1-\alpha$ defines a convex set provided that the vector $(a^T, b)^T$ has a symmetric logconcave density with $\alpha < 1/2$ (Lagoa et al. 2005).
- A joint chance constraint in the form of $\Pr\{g_i(x) \ge \xi_i, i = 1, \ldots, m\} \ge 1-\alpha$ defines a convex set if $g_i(x)$ is $(-r_i)$-concave and $\xi_i, i = 1, \ldots, m$, are independent random variables with (r_i+1)-decreasing densities for some $r_i > 0$ for sufficient small α values (Henrion and Strugarek 2008).

Literature Review (cont'd)

Convex conservative approximations:

- Find a convex function $\tilde{p}(x)$ such that $\tilde{p}(x) \geq p(x)$ for all $x \in X$, and

$$\min \; h(x) \quad \text{s.t.} \quad \tilde{p}(x) \leq \alpha, \quad x \in X.$$

- Conditional value at risk (CVaR) approximation of Rockafellar and Uryasev (2000), it is known as the "best" convex conservative approximation.

- Quadratic approximation of Ben-Tal and Nemirovski (2000).

- Bernstein approximation of Nemirovski and Shapiro (2006)

- Bonferroni inequality is often used to break a joint chance constraint into m single chance constraints. It makes the solution more conservative.

Literature Review (cont'd)

Scenario analysis:

- Generate an i.i.d. sample $\{\xi_1, \cdots, \xi_n\}$ and solve

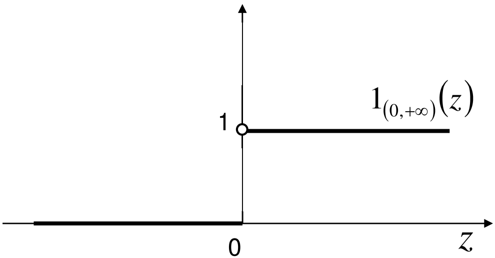
$$\begin{array}{ll} \min & h(x) \\ \text{s.t.} & c_i(x, \xi_\ell) \le 0, i = 1, \cdots, m, \ell = 1, \cdots, n, \\ & x \in X. \end{array}$$

- The key is to determine the sample size n to ensure that $p(x) \le \alpha$ with high probability.
- It is studied by Calafiore and Campi (2005 and 2006) and De Farias and Van Roy (2004).
- It is very fast to solve even when n is large, but typically very conservative and volatile.

CVaR Approximation

We start with a single chance constraint. A little transformation...

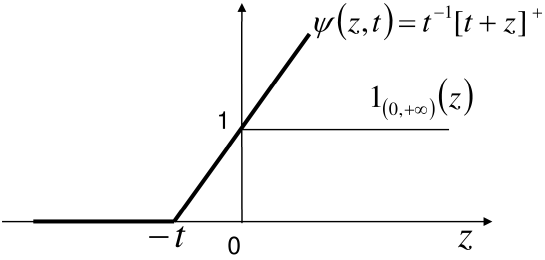
$$\begin{aligned} p(x) &= 1 - \Pr\{c(x,\xi) \le 0\} = \Pr\{c(x,\xi) > 0\} \\ &= \mathrm{E}\left[1_{(0,+\infty)}(c(x,\xi))\right]. \end{aligned}$$



The indicator function $1_{(0,+\infty)}(z)$

CVaR Approximation (cont'd)

"Best" convex conservative approximation of $1_{(0,+\infty)}(z)$ is in the following form:



The CVaR approximation $\psi(z,t)$

CVaR Approximation (cont'd)

Then, the "best" convex conservative approximation of $p(x)$ is

$$\tilde{p}(x) = \inf_{t>0} \mathrm{E}\left[\psi(c(x,\xi),t)\right] = \inf_{t>0} \frac{1}{t}\mathrm{E}\left[t + c(x,\xi)\right]^{+}.$$

This approximation proposes to solve

$$\begin{aligned} \text{minimize} \quad & h(x) \\ \text{s.t.} \quad & \inf_{t>0} \frac{1}{t}\mathrm{E}\left[t + c(x,\xi)\right]^{+} \le \alpha \\ & x \in X. \end{aligned}$$

CVaR Approximation (cont'd)

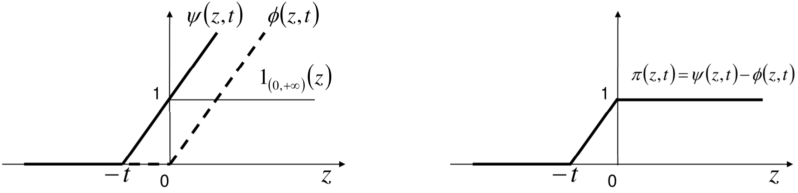
Note that

$$
\begin{aligned}
& \inf_{t>0} \frac{1}{t} \mathrm{E}\left[t + c(x,\xi)\right]^{+} \le \alpha \\
\Leftrightarrow \quad & \inf_{t>0} \left\{ \frac{1}{t} \mathrm{E}\left[t + c(x,\xi)\right]^{+} - \alpha \right\} \le 0 \\
\Leftrightarrow \quad & \inf_{\tau<0} \left\{ \tau + \frac{1}{\alpha} \mathrm{E}\left[c(x,\xi) - \tau\right]^{+} \right\} \le 0 \\
\Leftrightarrow \quad & \inf_{\tau\in\Re} \left\{ \tau + \frac{1}{\alpha} \mathrm{E}\left[c(x,\xi) - \tau\right]^{+} \right\} \le 0 \\
\Leftrightarrow \quad & \mathrm{CVaR}_{1-\alpha}(c(x,\xi)) \le 0
\end{aligned}
$$

This is the reason that the approximation is known as the **CVaR approximation**.

DC Approximation

Hong, Yang and Zhang (*OR*, forthcoming) propose a DC (difference of convex functions) approximation of $1_{(0,+\infty)}(z)$.



The construction of $\phi(z,t)$ — The difference function $\pi(z,t)$

Let $\tilde{p}(x,t) = \mathrm{E}\left[\pi(c(x,\xi),t)\right] = \frac{1}{t}\left\{\mathrm{E}\left[t + c(x,\xi)\right]^+ - \mathrm{E}\left[c(x,\xi)\right]^+\right\}$. We propose to solve

$$\min \; h(x), \quad \text{s.t.} \quad \inf_{t>0} \; \tilde{p}(x,t) \le \alpha, \quad x \in X.$$

DC Approximation (cont'd)

The equivalence between the DC approximation and the CCP:

- Note that $\inf_{t>0} \tilde{p}(x,t) = \lim_{t \searrow 0} \tilde{p}(x,t) = p(x)$.
- Solving the CCP is equivalent to solving the DC approximation.
- Hong, Yang and Zhang (forthcoming) proposed to solve an ε-approximation:

$$\min \; h(x), \quad \text{s.t.} \quad \tilde{p}(x,\varepsilon) \leq \alpha, \quad x \in X$$

for some small $\varepsilon > 0$.

A Different DC Approximation

Note that

$$
\begin{aligned}
& \inf_{t>0} \frac{1}{t} \left\{ \mathrm{E}\left[t + c(x,\xi)\right]^+ - \mathrm{E}\left[c(x,\xi)\right]^+ \right\} \le \alpha \\
\Leftrightarrow \quad & \inf_{t>0} \left\{ \frac{1}{\alpha}\mathrm{E}\left[t + c(x,\xi)\right]^+ - t \right\} - \frac{1}{\alpha}\mathrm{E}\left[c(x,\xi)\right]^+ \le 0 \\
\Leftrightarrow \quad & \inf_{\tau<0} \left\{ \tau + \frac{1}{\alpha}\mathrm{E}\left[c(x,\xi) - \tau\right]^+ \right\} - \frac{1}{\alpha}\mathrm{E}\left[c(x,\xi)\right]^+ \le 0 \\
\Leftrightarrow \quad & \inf_{\tau\in\Re} \left\{ \tau + \frac{1}{\alpha}\mathrm{E}\left[c(x,\xi) - \tau\right]^+ \right\} - \frac{1}{\alpha}\mathrm{E}\left[c(x,\xi)\right]^+ \le 0 \\
\Leftrightarrow \quad & \mathrm{CVaR}_{1-\alpha}(c(x,\xi)) - \frac{1}{\alpha}\mathrm{E}\left[c(x,\xi)\right]^+ \le 0
\end{aligned}
$$

A Different DC Approximation (cont'd)

Then, the original CCP is equivalent to

$$\begin{array}{ll} \min & h(x) \\ \text{s.t.} & \text{CVaR}_{1-\alpha}(c(x,\xi)) - \frac{1}{\alpha}\mathrm{E}\left[c(x,\xi)\right]^{+} \leq 0 \\ & x \in X. \end{array}$$

The CVaR approximation is

$$\begin{array}{ll} \min & h(x) \\ \text{s.t.} & \text{CVaR}_{1-\alpha}(c(x,\xi)) \leq 0 \\ & x \in X. \end{array}$$

It is clear that the CVaR approximation is a convex conservative approximation and the lost term is $\frac{1}{\alpha}\mathrm{E}\left[c(x,\xi)\right]^{+}$.

Sequential Convex Approximation

Suppose that x_0 is the optimal solution to the CVaR approximation.

- To simplify the notation, let $g(x) = \frac{1}{\alpha}\mathrm{E}\left[c(x,\xi)\right]^+$.
- Suppose that $g(x)$ is differentiable. We let

$$\tilde{g}(x) = g(x_0) + \nabla g(x_0)^T(x - x_0).$$

- Note $\tilde{g}(x)$ is tangent plane of $g(x)$. Because $g(x)$ is a convex function, $g(x) \geq \tilde{g}(x)$ for all $x \in \Re^d$.

Sequential Convex Approximation (cont'd)

Then, we propose to solve

$$\min \; h(x), \quad \text{s.t. } \mathrm{CVaR}_{1-\alpha}(c(x,\xi)) - \tilde{g}(x) \le 0, \; x \in X.$$

Let x_1 denote its optimal solution. This problem has several properties:

- It is convex.
- It is a conservative approximation, i.e., x_1 is a feasible solution to the original CCP, because

$$\mathrm{CVaR}_{1-\alpha}(c(x_1,\xi)) - g(x_1) \le \mathrm{CVaR}_{1-\alpha}(c(x_1,\xi)) - \tilde{g}(x_1) \le 0.$$

- x_1 is at least as good as x_0, i.e., $h(x_1) \le h(x_0)$, because x_0 is a feasible solution to the problem as

$$\begin{aligned} \mathrm{CVaR}_{1-\alpha}(c(x_0,\xi)) - \tilde{g}(x_0) &= \mathrm{CVaR}_{1-\alpha}(c(x_0,\xi)) - g(x_0) \\ &\le \mathrm{CVaR}_{1-\alpha}(c(x_0,\xi)) \le 0. \end{aligned}$$

Sequential Convex Approximation (cont'd)

A different way of understanding:

- Note that x_0 satisfies

$$\text{CVaR}_{1-\alpha}(c(x_0, \xi)) \le 0.$$

- If $h(x_1) < h(x_0)$, i.e., the new solution is better than the CVaR approximation, then

$$\begin{aligned} &\text{CVaR}_{1-\alpha}(c(x_1, \xi)) > 0 \\ &\text{CVaR}_{1-\alpha}(c(x_1, \xi)) - g(x_1) < 0. \end{aligned}$$

- The lost term $g(x)$ in the CVaR approximation is now partially recovered and becomes effective.

Sequential Convex Approximation (cont'd)

We don't have to stop at x_1. Given x_1, we can continue:

- Let $\tilde{g}(x) = g(x_1) + \nabla g(x_1)^T (x - x_1)$.
- Then, find x_1 that solves

$$\min \; h(x), \quad \text{s.t. } \mathrm{CVaR}_{1-\alpha}(c(x,\xi)) - \tilde{g}(x) \le 0, \; x \in X.$$

- Again, x_2 is a feasible solution of the original CCP and $h(x_2) \le h(x_1)$.

We can continue and obtain an infinite sequence $\{x_0, x_1, \ldots\}$. We call this approach "Sequential Convex Approximation".

A critical question: What are the limiting points of $\{x_0, x_1, \ldots\}$?

Under some technical conditions,

- all cluster points of $\{x_0, x_1, \ldots\}$ are KKT points of the original CCP;
- if $h(x)$ is strictly convex, $\{x_0, x_1, \ldots\}$ converges to a KKT point.

As the chance constraint is typically tight at the KKT points, we have

$$\mathrm{CVaR}_{1-\alpha}(c(\bar{x}, \xi)) - g(\bar{x}) = 0,$$

where $\bar{x}$ is a cluster point of $\{x_0, x_1, \ldots\}$. Then, we completely find back the lost term $g(x)$.

In summary, the Sequential Convex Approximation finds a KKT point that is better than the CVaR approximation (which is the "best" convex conservative approximation).

Handling Joint Chance Constraint

Note that

$$p(x) = 1 - \Pr\{c_1(x,\xi) \le 0, \ldots, c_m(x,\xi) \le 0\} = \Pr\{\max\ c_i(x,\xi) > 0\}.$$

- If we let $c(x,\xi) = \max\ c_i(x,\xi)$, it becomes a single chance constraint.
- However, a critical problem is that $\mathrm{CVaR}(c(x,\xi))$ and $g(x) = \frac{1}{\alpha}\mathrm{E}\left[c(x,\xi)\right]^+$ may not be differentiable with respect to x.
- A differentiable case: If $\Pr\{c_i(x,\xi) = c_j(x,\xi)\} = 0$ for all $x \in X$ and all $i \neq j$, then both $\mathrm{CVaR}(c(x,\xi))$ and $g(x)$ are differentiable with respect to x.

Handling Joint Chance Constraint (cont'd)

What if CVaR$(c(x,\xi))$ and $g(x)$ are not differentiable?

- We have to redefine a stationary point, as the KKT conditions are no longer applicable.
- It becomes a non-convex and non-smooth optimization problem.
- If we substitute $\nabla g(x)$ by any subgradient of $g(x)$, the Sequential Convex Approximation approach still works (in theory).
- However, efficient non-smooth optimization algorithms and solvers are difficult to find.

Handling Joint Chance Constraint (cont'd)

A log-exponential smoothing approach to the non-smooth JCCP:

- By Rockafellar (1970), when $\delta > 0$,

$$\max c_i(x,\xi) \leq \delta \log \left[\sum_{i=1}^{m} e^{\frac{1}{\delta} c_i(x,\xi)} \right] \leq \max c_i(x,\xi) + \delta \log m.$$

- We propose to solve the problem

$$\min \ h(x), \quad \text{s.t.} \quad \Pr\left\{ \delta \log \left[\sum_{i=1}^{m} e^{\frac{1}{\delta} c_i(x,\xi)} \right] \leq 0 \right\} \geq 1-\alpha, \ x \in X.$$

- Note that the new problem is a smooth CCP. It can be solved by using the Sequential Convex Approximation algorithm.

- The KKT solutions of the problem converges to the set of stationary points of the non-smooth problem as $\delta \searrow 0$.

To apply the Sequential Convex Approximation algorithm, in each iteration, we need to solve

$$\min\ h(x), \quad \text{s.t. } \mathrm{CVaR}_{1-\alpha}(c(x,\xi)) - \tilde{g}(x) \le 0,\ x \in X$$

where

$$\tilde{g}(x) = \frac{1}{\alpha}\mathrm{E}\left[c(x_i,\xi)\right]^+ + \left(\nabla \frac{1}{\alpha}\mathrm{E}\left[c(x_i,\xi)\right]^+\right)^T (x - x_i).$$

By Broadie and Glasserman (1996),

$$\nabla \mathrm{E}\left[c(x,\xi)\right]^+ = \mathrm{E}\left[\nabla_x c(x,\xi) \cdot 1_{\{c(x,\xi)>0\}}\right].$$

By Hong and Liu (2009),

$$\nabla \mathrm{CVaR}_{1-\alpha}(c(x,\xi)) = \mathrm{E}\left[\nabla_x c(x,\xi) \mid c(x,\xi) > \mathrm{VaR}_{1-\alpha}(c(x,\xi))\right].$$

Solving the Subproblem (cont'd)

We use a gradient-based Monte Carlo method:

- Take an i.i.d. sample of ξ, denoted as $\xi_1, \ldots, \xi_n$.
- We may use the sample to estimate $\mathrm{CVaR}_{1-\alpha}(c(x,\xi))$, $\mathrm{E}\left[c(x,\xi)\right]^+$, $\nabla\mathrm{CVaR}_{1-\alpha}(c(x,\xi))$ and $\nabla\mathrm{E}\left[c(x,\xi)\right]^+$.
- Then, we suggest to use a gradient-based algorithm to solve the problem with the estimated values and gradients. We use *fmincon* of Matlab® in all our numerical studies.
- We prove that, as $n \to \infty$, the optimal solution to the sample problem converges to the set of optimal solutions of the original subproblem.

An Illustrative Example

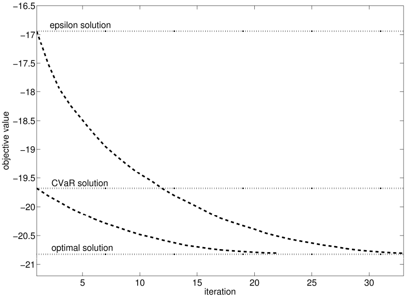
A norm minimization problem:

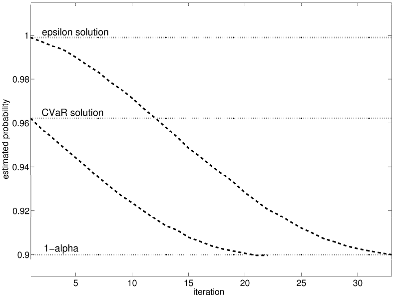
$$\begin{aligned} \text{minimize} \quad & -\|x\|_1 \\ \text{subject to} \quad & \Pr\{\|\xi_i \circ x\| \le 10, \quad i = 1, \ldots, m\} \ge 1 - \alpha, \\ & x_j \ge 0, \quad j = 1, \ldots, d. \end{aligned}$$

- $\|x\|_1 = \sum_{j=1}^d |x_j|$ and $\|x\| = \left(\sum_{j=1}^d x_j^2\right)^{-1/2}$.
- $\xi_i \circ x = (\xi_{i1}x_1, \ldots, \xi_{id}x_d)^T$ denote the Hadamard product (or entrywise product) of ξ_i and x.
- ξ_{ij}, $i = 1, \ldots, m$ and $j = 1, \ldots, d$, are i.i.d. standard normal random variables.
- The optimal solution can be derived by the symmetry and property of normal distributions.

An Illustrative Example

We set $d = 10$, $m = 10$ and $\alpha = 0.1$. Then, the optimal solution is $x_1^* = \cdots = x_d^* = 2.08$ and the optimal objective value is $f^* = -20.82$.





THANK YOU!