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

L. Jeff Hong\textsuperscript{1} Liwei Zhang\textsuperscript{2}

\textsuperscript{1}Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology

\textsuperscript{2}School of Mathematical Science
Dalian University of Technology, China

April 2010
Constrained Convex Program

Basic formulation:

$$\min \ h(x)$$

s.t. \ \ \ \ \ \ \ \ c_1(x, \xi) \leq 0, \cdots, c_m(x, \xi) \leq 0, \ \ \ \ x \in X.$$

- $x$ is the vector of decision variables and $\xi$ 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 $\xi$. How should we handle that?

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

- A very simple example:

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

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

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

$$\begin{align*}
\text{min} & \quad h(x) \\
\text{s.t.} & \quad \Pr \{c_1(x, \xi) \leq 0, \ldots, c_m(x, \xi) \leq 0\} \geq 1 - \alpha, \\
& \quad x \in X.
\end{align*}$$

- 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 $m \geq 2$, 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.
Convexity of the chance constraint:

- If $c_1(x, \xi), \cdots, c_m(x, \xi)$ are quasiconvex functions of $(x, \xi)$, and $\xi$ 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 $\Pr\{a^T x \leq b\} \geq 1 - \alpha$ 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 $\Pr\{g_i(x) \geq \xi_i, i = 1, \ldots, m\} \geq 1 - \alpha$ 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 $\alpha$ values (Henrion and Strugarek 2008).
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.


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

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

\[
\begin{align*}
\min & \quad h(x) \\
\text{s.t.} & \quad c_i(x, \xi_\ell) \leq 0, \ i = 1, \cdots, m, \ \ell = 1, \cdots, n,
\end{align*}
\]

- \( x \in X \).

- The key is to determine the sample size \( n \) to ensure that \( p(x) \leq \alpha \) 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.
We start with a single chance constraint. A little transformation...

\[ p(x) = 1 - \Pr\{c(x, \xi) \leq 0\} = \Pr\{c(x, \xi) > 0\} \]

\[ = \mathbb{E}\left[1_{(0, +\infty)}(c(x, \xi))\right]. \]

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

\[
\psi(z, t) = t^{-1}[t + z]^+
\]

The CVaR approximation $\psi(z, t)$
CVaR Approximation (cont’d)

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

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

This approximation proposes to solve

$$\begin{align*}
\text{minimize} & \quad h(x) \\
\text{s.t.} & \quad \inf_{t > 0} \frac{1}{t} E [t + c(x, \xi)]^+ \leq \alpha \\
& \quad x \in X.
\end{align*}$$
Note that

$$\inf_{t>0} \frac{1}{t} E [t + c(x, \xi)]^+ \leq \alpha$$

$$\Leftrightarrow \inf_{t>0} \left\{ \frac{1}{t} E [t + c(x, \xi)]^+ - \alpha \right\} \leq 0$$

$$\Leftrightarrow \inf_{\tau<0} \left\{ \tau + \alpha E [c(x, \xi) - \tau]^+ \right\} \leq 0$$

$$\Leftrightarrow \inf_{\tau \in \mathbb{R}} \left\{ \tau + \frac{1}{\alpha} E [c(x, \xi) - \tau]^+ \right\} \leq 0$$

$$\Leftrightarrow \text{CVaR}_{1-\alpha}(c(x, \xi)) \leq 0$$

This is the reason that the approximation is known as the CVaR approximation.
Hong, Yang and Zhang (OR, forthcoming) propose a DC (difference of convex functions) approximation of $1_{(0, +\infty)}(z)$. Let $\tilde{p}(x, t) = \mathbb{E}[\pi(c(x, \xi), t)] = \frac{1}{t} \left\{ \mathbb{E}[t + c(x, \xi)]^+ - \mathbb{E}[c(x, \xi)]^+ \right\}$. We propose to solve

$$\min h(x), \quad \text{s.t.} \quad \inf_{t > 0} \tilde{p}(x, t) \leq \alpha, \quad x \in X.$$
The equivalence between the DC approximation and the CCP:

- Note that \( \inf_{t>0} \tilde{p}(x, t) = \lim_{t \to 0} \tilde{p}(x, t) = p(x). \)

- Solving the CCP is equivalent to solving the DC approximation.

- Hong, Yang and Zhang (forthcoming) proposed to solve an \( \varepsilon \)-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

\[
\inf_{t>0} \frac{1}{t} \left\{ E[t + c(x, \xi)]^+ - E[c(x, \xi)]^+ \right\} \leq \alpha \\
\Leftrightarrow \inf_{t>0} \left\{ \frac{1}{\alpha} E[t + c(x, \xi)]^+ - t \right\} - \frac{1}{\alpha} E[c(x, \xi)]^+ \leq 0 \\
\Leftrightarrow \inf_{\tau<0} \left\{ \tau + \frac{1}{\alpha} E[c(x, \xi) - \tau]^+ \right\} - \frac{1}{\alpha} E[c(x, \xi)]^+ \leq 0 \\
\Leftrightarrow \inf_{\tau \in \mathbb{R}} \left\{ \tau + \frac{1}{\alpha} E[c(x, \xi) - \tau]^+ \right\} - \frac{1}{\alpha} E[c(x, \xi)]^+ \leq 0 \\
\Leftrightarrow \text{CVaR}_{1-\alpha}(c(x, \xi)) - \frac{1}{\alpha} E[c(x, \xi)]^+ \leq 0
\]
A Different DC Approximation (cont’d)

Then, the original CCP is equivalent to

\[
\begin{align*}
\min & \quad h(x) \\
\text{s.t.} & \quad \text{CVaR}_{1-\alpha}(c(x, \xi)) - \frac{1}{\alpha} \mathbb{E} [c(x, \xi)]^+ \leq 0 \\
& \quad x \in X.
\end{align*}
\]

The CVaR approximation is

\[
\begin{align*}
\min & \quad h(x) \\
\text{s.t.} & \quad \text{CVaR}_{1-\alpha}(c(x, \xi)) \leq 0 \\
& \quad x \in X.
\end{align*}
\]

It is clear that the CVaR approximation is a convex conservative approximation and the lost term is \( \frac{1}{\alpha} \mathbb{E} [c(x, \xi)]^+ \).
Sequential Convex Approximation

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

- To simplify the notation, let $g(x) = \frac{1}{\alpha} E [c(x, \xi)]^+.$
- 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 $x \in \mathbb{R}^d.$
Sequential Convex Approximation (cont’d)

Then, we propose to solve

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

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

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

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

$$\text{CVaR}_{1-\alpha}(c(x_0, \xi)) - \tilde{g}(x_0) = \text{CVaR}_{1-\alpha}(c(x_0, \xi)) - g(x_0) \leq \text{CVaR}_{1-\alpha}(c(x_0, \xi)) \leq 0.$$
A different way of understanding:

- Note that $x_0$ satisfies

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

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

$$\text{CVaR}_{1-\alpha}(c(x_1, \xi)) > 0$$

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

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

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

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

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 $\{x_0, x_1, \ldots\}$?**
Sequential Convex Approximation (cont’d)

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

\[
\text{CVaR}_{1-\alpha}(c(\bar{x}, \xi)) - g(\bar{x}) = 0,
\]

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

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

- If we let \( c(x, \xi) = \max c_i(x, \xi) \), it becomes a single chance constraint.

- However, a critical problem is that \( \text{CVaR}(c(x, \xi)) \) and \( g(x) = \frac{1}{\alpha} \mathbb{E} [c(x, \xi)]^+ \) 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 \( \text{CVaR}(c(x, \xi)) \) and \( g(x) \) are differentiable with respect to \( x \).
What if $\text{CVaR}(c(x, \xi))$ 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.
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

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

- 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 $\delta \searrow 0$. 

---

Hong/Zhang (HKUST/DLUT)  
Chance Constrained Program  
April 2010 24 / 29
Solving the Subproblem

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

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

where

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

By Broadie and Glasserman (1996),

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

By Hong and Liu (2009),

$$\nabla \text{CVaR}_{1-\alpha}(c(x, \xi)) = \mathbb{E} \left[ \nabla_x c(x, \xi) \mid c(x, \xi) > \text{VaR}_{1-\alpha}(c(x, \xi)) \right].$$
Solving the Subproblem (cont’d)

We use a gradient-based Monte Carlo method:

- Take an i.i.d. sample of $\xi$, denoted as $\xi_1, \ldots, \xi_n$.

- We may use the sample to estimate $\text{CVaR}_{1-\alpha}(c(x, \xi))$, $\mathbb{E}[c(x, \xi)]^+$, $\nabla \text{CVaR}_{1-\alpha}(c(x, \xi))$ and $\nabla \mathbb{E}[c(x, \xi)]^+$.

- 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:

\[
\begin{align*}
\text{minimize} & \quad -\|x\|_1 \\
\text{subject to} & \quad \Pr \left\{ \|\xi \circ x\| \leq 10, \quad i = 1, \ldots, m \right\} \geq 1 - \alpha, \\
& \quad x_j \geq 0, \quad j = 1, \ldots, d.
\end{align*}
\]

- \(\|x\|_1 = \sum_{j=1}^{d} |x_j|\) and \(\|x\| = \left(\sum_{j=1}^{d} x_j^2\right)^{-1/2}\).

- \(\xi \circ x = (\xi_1 x_1, \ldots, \xi_d x_d)^T\) denote the Hadamard product (or entrywise product) of \(\xi\) and \(x\).

- \(\xi_{ij}, \ i = 1, \ldots, m\) 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!