Convex Analytic Approach for Constrained Stochastic Optimal Control

Yin-Lam Chow

Stanford University

ychow@stanford.edu

November 7, 2013

A finite Markov Decision Process (MDP) is a five-tuple $(S, U, Q, U(\cdot), \beta)$,

- S, the state space, is a finite set;
- U, the control space, is a finite set; for every $x \in S$,
- $U(x) \subseteq U$ is a nonempty set which represents the set of admissible controls when the system state is x
- $Q(\cdot|x, u)$ (the transition probability) is a conditional probability on S given the set of admissible state-control pairs, i.e., the sets of pairs (x, u) where $x \in S$ and $u \in U(x)$
- $\beta \in \mathcal{P}(S)$ is the initial distribution of state x

Let Π be the set of all stationary state feedback deterministic control policies,

$$\Pi := \Big\{\pi : S \to U\}$$

and Π_R be the set of all stationary state feedback randomized control policies,

$$\Pi_R := \Big\{\pi : S \to \mathcal{P}(U(S))\}.$$

Problem

Given a policy $\pi \in \Pi_R$, an initial distribution β, and a discounted factor $\alpha \in (0, 1)$, the cost function is defined as

$$J^\pi(\beta) := \lim_{N \to \infty} (1 - \alpha) \mathbb{E} \sum_{k=0}^{N-1} \alpha^k c(x_k, u_k),$$

and the risk constraint is defined as

$$R^\pi(\beta) := (1 - \alpha) \rho \left(\lim_{N \to \infty} \sum_{k=0}^{N-1} \alpha^k d(x_k, u_k) \right),$$

The problem we wish to solve is then as follows:

Optimization problem $\mathcal{OPT}$ — *Given a risk threshold* $r_0 \in \mathbb{R}$, *solve*

$$\begin{aligned} \min_{\pi \in \Pi_R, \beta \in \mathcal{P}(S)} \quad & J^\pi(\beta) \\ \textit{subject to} \quad & R^\pi(\beta) \leq r_0. \end{aligned}$$

Now define the feasible region of Problem $\mathcal{OPT}$:

$$\Delta = \{(\pi, \beta) \in \Pi_R \times \mathcal{P}(S) : J^\pi(\beta) > -\infty,\ R^\pi(\beta) \leq r_0\}$$

This allows us to re-write Problem $\mathcal{OPT}$ as

$$\min\{J^\pi(\beta) : (\pi, \beta) \in \Delta\}. \tag{1}$$

Polyhedral Risk

We first introduce the notion of polyhedral risk measure.

Definition

(Polyhedral Risk Measures) A risk measure $\rho : \mathcal{Z} \to \mathbb{R}$ is called a polyhedral risk if there exists matrices A_E, B_E, A_I, B_I and vectors a_E, a_I, c such that

$$\begin{aligned} \rho(Z) = \inf_{y \in \mathcal{Z}} \quad & \mathbb{E}\left[c^T y\right] \\ \text{s.t.} \quad & A_I y \leq B_I, \ A_E y = B_E, \ \text{a.s.} \\ & a_E^T y = Z, \ a_I^T y \leq Z, \ \text{a.s.} \end{aligned} \tag{2}$$

Assumption: The feasible region of the optimization problem in $\rho(Z)$ is non-empty.

Remark:
We will restrict our attention to disturbance models characterized by probability mass functions, and finite probability spaces (i.e., Ω has a finite number of elements (for example L elements) or, equivalently, $\mathcal{F}$ is a finitely generated algebra). Accordingly, by definition of expectation,

$$\mathbb{E}\left[c^T y\right] = \sum_{j=1}^{L} p(w_j) c^T y(w_j).$$

where p is the reference probability measure:

$$\sum_{j=1}^{L} p(w_j) = 1, \;\; p(w_j) \geq 0, \forall j \in \{1, \ldots, L\}.$$

Examples

The expected value of a random variable Z can be represented according to definition 1 with

$$\begin{aligned} \mathbb{E}[Z] = \inf_{y \in \mathcal{Z}} \quad & \mathbb{E}\left[c^T y\right] \\ \text{s.t.} \quad & c^T y = Z \text{ a.s.} \end{aligned}$$

A second example is the Conditional Value-at-Risk (CVaR):

$$\text{CVaR}_\alpha(Z) := \inf_{x \in \mathbb{R}} \left[x + \frac{1}{\beta} \mathbb{E}\left[(Z - x)^+\right] \right], \tag{3}$$

where $\beta \in (0, 1]$ and $(x)^+ := \max(0, x)$. CVaR can be represented by a polyhedral risk measure as follows:

$$\begin{aligned} \text{CVaR}_\alpha(Z) = \inf_{x \in \mathbb{R},\, y \in \mathcal{Z}} \quad & x + \frac{1}{\beta} \mathbb{E}\left[y\right] \\ \text{s.t.} \quad & y \geq 0,\ y \geq Z - x \text{ a.s.} \end{aligned}$$

A more "restricted" polyhedral risk:

Optimization problem $\mathcal{RISK}$ — *Given constant matrices* A_I, A_E *and vectors* c, a_E, a_I, B_I, B_E *of appropriate dimensions, and* $Z = \sum_{k=0}^{\infty} \alpha^k d(x_k, u_k)$, *solve*

$$\begin{aligned}
\min_{y \in \mathcal{Z}} \quad & \mathbb{E}\left[c^T y\right] \\
s.t. \quad & A_I y(x_0, u_0, x_1, u_1, \ldots) \leq B_I, \ \forall (x_k, u_k) \in S \times A, \\
& A_E y(x_0, u_0, x_1, u_1, \ldots) = B_E, \ \forall (x_k, u_k) \in S \times A, \\
& a_E^T y(x_0, u_0, x_1, u_1, \ldots) = \sum_{k=0}^{\infty} \alpha^k d(x_k, u_k), \ \forall (x_k, u_k) \in S \times A, \\
& a_I^T y(x_0, u_0, x_1, u_1, \ldots) \leq \sum_{k=0}^{\infty} \alpha^k d(x_k, u_k), \ \forall (x_k, u_k) \in S \times A,
\end{aligned}$$

Assumption: There exists a sequence of non-negative random vectors $M_j(x_j, u_j) \in \mathcal{Z} \subseteq \mathbb{R}$, such that

$$\sum_{j \in \mathbb{N}} M_j(x_j, u_j) < \infty$$

for any $(x_j, u_j) \in S \times A$, $j \in \mathbb{N}$, and

$$-\sum_{j \in \mathbb{N}} M_j(x_j, u_j) \leq y(x_0, u_0, x_1, u_1, \ldots) \leq \sum_{j \in \mathbb{N}} M_j(x_j, u_j),$$

for any $(x_j, u_j) \in S \times A$, $j \in \mathbb{N}$ for any feasible solution $y(x_0, u_0, x_1, u_1, \ldots)$ of problem $\mathcal{RISK}$.

Compare problem $\mathcal{RISK}$ with:

$$\min_{d_y(x,u)} \ \min_{\forall (x,u)\in S\times A} \quad \mathbb{E}\left[c^T y\right] \tag{4}$$

$$\begin{aligned}
\text{s.t.} \quad & A_I y(x_0,u_0,x_1,u_1,\ldots) \le B_I, \ \forall (x_k,u_k)\in S\times A,\\
& A_E y(x_0,u_0,x_1,u_1,\ldots) = B_E, \ \forall (x_k,u_k)\in S\times A,\\
& a_E^T y(x_0,u_0,x_1,u_1,\ldots) = \sum_{k=0}^{\infty}\alpha^k d(x_k,u_k),\\
& a_I^T y(x_0,u_0,x_1,u_1,\ldots) \le \sum_{k=0}^{\infty}\alpha^k d(x_k,u_k),\\
& y(x_0,u_0,x_1,u_1,\ldots) := \sum_{k=0}^{\infty}\alpha^k d_y(x_k,u_k) \qquad (5)
\end{aligned}$$

We want to show that we can consider this problem **without loss of optimality.**

Preliminary results:

Lemma

For any given cost function $Z = \sum_{k=0}^{\infty} \alpha^k d(x_k, u_k)$, the feasible region in problem $\mathcal{RISK}$ is convex.

Lemma

Under the above assumptions, for any given cost function $Z = \sum_{k=0}^{\infty} \alpha^k d(x_k, u_k)$, the feasible region in problem $\mathcal{RISK}$ is closed and compact.

Lemma

The objective function in problem $\mathcal{RISK}$ is continuous over its feasible region.

Theorem

Under the above assumptions, the optimal solution in problem $\mathcal{RISK}$ can be attained by an extreme point solution.

Proof.

Use Bauer's Maximum Principle □

With this result, we want to show that a minimizer in problem $\mathcal{RISK}$ can be characterized by the expression in form of (5)

First, show this result for **"periodic MDP"**.

Optimization problem $\mathcal{RISK}^P$ — *Given constant matrices* A_I, A_E *and vectors* c, a_E, a_I, B_I, B_E *of appropriate dimensions, solve*

$$V_Z(T) = \min_{y \in \mathcal{Z}} \sum_{x_k \in S, u_k \in A, k \in \{0,\ldots,T-1\}} \mathbb{P}(\overline{xu}_T) c^T y(\overline{xu}_T, \overline{xu}_T, \ldots) \tag{6}$$

$$\begin{aligned} \text{s.t.} \quad & A_I y(\overline{xu}_T, \overline{xu}_T, \ldots) \leq B_I, \\ & A_E y(\overline{xu}_T, \overline{xu}_T, \ldots) = B_E, \\ & a_E^T y(\overline{xu}_T, \overline{xu}_T, \ldots) = \sum_{j=0}^{\infty} \sum_{k=0}^{T-1} \alpha^{k+jT} d(x_{k+jT}, x_{k+jT}), \\ & a_I^T y(\overline{xu}_T, \overline{xu}_T, \ldots) \leq \sum_{j=0}^{\infty} \sum_{k=0}^{T-1} \alpha^{k+jT} d(x_{k+jT}, u_{k+jT}) \end{aligned}$$

where $\overline{xu}_T = \{x_0, u_0, \ldots, x_{T-1}, u_{T-1}\}$ *and* $(x_{k+jT}, u_{k+jT})) = (x_k, u_k)$.

Rewritre:

$$\begin{bmatrix} A_I & -A_I & I & 0 \\ A_E & -A_E & 0 & 0 \\ a_E^T & -a_E^T & 0 & 0 \\ a_I^T & -a_I^T & 0 & I \end{bmatrix} = \begin{bmatrix} \mathcal{B} & \mathcal{D} \end{bmatrix} P$$

we can further show that for each of the above basic feasible solution,

$$y(\overline{xu}_T, \overline{xu}_T, \ldots) =$$

$$\sum_{j=0}^{\infty} \sum_{k=0}^{T-1} \alpha^{k+jT} \begin{bmatrix} I & -I & 0 & 0 \end{bmatrix} P^T \begin{bmatrix} \mathcal{B}^{-1} \begin{bmatrix} B_I(1-\alpha) \\ B_E(1-\alpha) \\ d(x_{k+jT}, u_{k+jT}) \\ d(x_{k+jT}, u_{k+jT}) \end{bmatrix} \\ 0 \end{bmatrix}$$

As choosing the optimal basic feasible solution is equivalent to choosing the optimal set of matrices $\mathcal{B}$, $\mathcal{D}$ and P. Without loss of optimality, in problem $\mathcal{RISK}^P$, we can restrict any feasible solution $y \in \mathcal{Z}$ have the desired structure

First, define $F_Z(\infty)$ to be the feasible region of problem $\mathcal{RISK}$ and define $F_Z(T)$ to be the feasible region of problem $\mathcal{RISK}^P$. For cases where general MDPs are concerned, we extend the above results using the following theorem.

Theorem

Under the above assumptions, the optimal value $V_T(Z)$ in problem $\mathcal{RISK}^P$ converges to the optimum value $V_\infty(Z)$ in problem $\mathcal{RISK}$ as $T \to \infty$. Moreover, if $T_n \to \infty$ as $n \to \infty$ and $y_n \in F_{T_n}(Z)$ for each n, then then the sequence $\{y_n\}$ has a limit point in $F_Z(\infty)$.

Proof.

Use Berge's Maximum Theorem. □

First we will discuss occupation measures on the set of state-action pair

$$K = \{(s, u) \in S \times A : a \in U(S)\}.$$

Occupation measure on set K can be interpreted as the total expected number of visits of a stochastic process $\{(s_t, u_t), t \geq 0\}$ to each state-action pair.
We use ρ to denote a probability mass function on K, where $\rho \in \mathcal{P}(K)$.

$$\sum_{x \in S, u \in A} \rho(x, u) = 1.$$

The marginal probability of ρ on S is the probability mass functions $\hat{\rho} \in \mathcal{P}(S)$, where

$$\hat{\rho}(x) = \sum_{u \in A} \rho(x, u).$$

There are two well-known facts in the literature of convex analytic methods for MDPs.

- If ρ is a probability mass function on K, there exists a stationary randomized Markov policy $\pi \in \Pi_R$ such that ρ can be decomposed as $\rho = \hat{\rho} \cdot \pi$. Specifically, $\rho = \hat{\rho} \cdot \pi$ is defined by

$$\rho(x, u) = \pi(u|x)\hat{\rho}(x), \ \forall (x, u) \in S \times A.$$

- For each $\pi \in \Pi_R$ and $\nu \in \mathcal{P}(S)$, the probability measure $\rho = \nu \cdot \pi$ on $S \times A$ satisfies $\sum_{x \in S, u \in A} \rho(x, a) = 1$ and $\hat{\rho} = \nu$. Specifically, $\rho = \nu \cdot \pi$ is defined by

$$\rho(x, u) = \pi(u|x)\nu(x), \ \forall (x, u) \in S \times A.$$

We need to restrict to a certain class of probability mass functions, namely the stable probability mass functions.

Definition

A probability mass function $\rho = \hat{\rho} \cdot \pi$ is well-posed if for any bounded stage-wise cost $r : S \times A \to \mathbb{R}$,

$$\sum_{x \in S, u \in A} r(x,u)\rho(x,u) \in \mathbb{R}.$$

Also, for $\rho = \hat{\rho} \cdot \pi$, the marginal occupation measure $\hat{\rho}$ is invariant if there exists a Markov randomized policy $\pi \in \Pi_R$ such that

$$\hat{\rho}(x) = \sum_{y \in S, u \in A} Q(x|y,u)\pi(u|y)\hat{\rho}(y) + (1-\alpha)\beta(x).$$

It can be easily seen that the above expression is equivalent to

$$\hat{\rho}(x) = \sum_{y \in S, u \in A} Q(x|y,u)\rho(x,y) + (1-\alpha)\beta(x). \tag{7}$$

Define

$$\mu^{\pi,\beta}(x,u) = (1-\alpha)\sum_{k=0}^{\infty}\alpha^k \mathbb{P}^{\pi,\beta}(x_k = x, u_k = u). \tag{8}$$

We can easily verify that $\mu^{\pi,\beta}$ is an occupation measure. Also, the cost function $J_N^{\pi}(\beta)$ can be re-written by

$$J^{\pi}(\beta) = \sum_{x\in S, u\in A} \mu^{\pi,\beta}(x,u)c(x,u)$$

Also, with

$$\pi_c(u|y) = \frac{\mu^{\pi,\beta}(y,u)}{\sum_{u\in A}\mu^{\pi,\beta}(y,u)}.$$

being a randomized Markov stationary policy, From basic properties of transition probability, one obtains

$$\hat{\mu}^{\pi,\beta}(x) = (1-\alpha)\beta(x) + \sum_{y\in S}\hat{\mu}^{\pi,\beta}(y)\sum_{u\in A}\pi_c(u|y)\alpha Q(x|y,a).$$

Recall the risk constraint:

$$\left\{ \begin{array}{ll} \min_{y \in \mathcal{Z}} & (1-\alpha)\mathbb{E}\left[c^T y\right] \\ \text{s.t.} & A_I y \leq B_I, \ A_E y = B_E, \ \text{a.s.} \\ & a_E^T y(x_0, u_0, x_1, u_1, \ldots,) = \lim_{N\to\infty} \sum_{k=0}^{N-1} \alpha^k d(x_k, u_k), \ \text{a.s.} \\ & a_I^T y(x_0, u_0, x_1, u_1, \ldots,) \leq \lim_{N\to\infty} \sum_{k=0}^{N-1} \alpha^k d(x_k, u_k), \ \text{a.s.} \end{array} \right\}$$

$\leq r_0$

We want to re-parametrize this using occupation measures.

By using occupation measures, the above risk constraint is equivalent to

$$\left\{ \begin{array}{ll} \min_{d_y(x,u),\ \forall (x,u)\in S\times A} & \sum_{u\in A, s\in S} \mu^{\pi,\beta}(x,u) c^T d_y(x,u) \\ \text{s.t.} & A_I d_y(x,u) \leq \tilde{B}_I,\ A_E d_y(x,u) = \tilde{B}_E, \\ & a_E^T d_y(x,u) = d(x,u), \\ & a_I^T d_y(x,u) \leq d(x,u), \quad \forall (x,u) \in S \times A \end{array} \right\}$$

$$\leq r_0$$

where $\tilde{B}_I = (1-\alpha) B_I$, $\tilde{B}_E = (1-\alpha) B_E$.

Problem $\mathcal{OPT}^R$ — *Given a risk threshold* $r_0 \in \mathbb{R}$, *solve*

$$\min_{\mu,\beta} \quad \sum_{x\in S, u\in A} \mu(x,u)c(x,u)$$

$$\text{s.t.} \quad \sum_{u\in A}\left[\mu(x,u) - \sum_{y\in S}\mu(y,a)\alpha Q(x|y,u)\right] = (1-\alpha)\beta(x),\ \forall x \in S$$

$$\left\{ \begin{array}{ll} \min_{d_y(x,u),\ \forall (x,u)\in S\times A} & \sum_{u\in A, s\in S}\mu(x,u)c^T d_y(x,u) \\ \text{s.t.} & A_I d_y(x,u) \leq \tilde{B}_I,\ A_E d_y(x,u) = \tilde{B}_E, \\ & a_E^T d_y(x,u) = d(x,u), \\ & a_I^T d_y(x,u) \leq d(x,u), \end{array} \right\}$$

$$\leq r_0$$

$$\mu(x,u) \geq 0$$

$$\sum_{x\in S}\beta(x) = 1,\ \beta(x) \geq 0,$$

By letting

$$d_\mu(x,u) = \mu(x,u) d_y(x,u), \ \forall (x,u) \in S \times A,$$

the above problem is equivalent to

Problem $\mathcal{OPT}^{LP}$ — *Given a risk threshold* $r_0 \in \mathbb{R}$, *solve*

$$\min_{\mu,\beta,d_\mu} \quad \sum_{x\in S, u\in A} \mu(x,u) c(x,u)$$

$$\textit{subject to} \quad \sum_{u\in A} \left[\mu(x,u) - \sum_{y\in S} \mu(y,a) \alpha Q(x|y,u) \right] = (1-\alpha)\beta(x),$$

$$\sum_{u\in A, s\in S} c^T d_\mu(x,u) \leq r_0$$

$$A_I d_\mu(x,u) \leq \mu(x,u)\tilde{B}_I, \ A_E d_\mu(x,u) = \mu(x,u)\tilde{B}_E,$$

$$a_E^T d_\mu(x,u) = \mu(x,u) d(x,u), \ a_I^T d_\mu(x,u) \leq \mu(x,u) d(x,u),$$

$$\mu(x,u) \geq 0, \ \forall (x,u) \in S \times A \ \sum_{x\in S} \beta(x) = 1, \ \beta(x) \geq 0$$