Risk Sensitive, Risk Constrained Stochastic Optimal Control with Time Consistent, Dynamic Risk Metrics

Yinlam Chow[1], M. Pavone (PI)[1]

[1] Autonomous Systems Laboratory, Stanford University, Stanford, CA

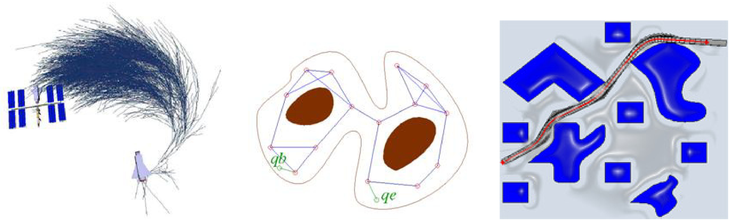
Objective

Develop a novel theory for risk-sensitive constrained stochastic optimal control and provide closed loop controller synthesis methods.

Key questions:

1. Can we formulate a risk constrained stochastic optimal control problem using dynamic programming?;
2. Do we need to update our dynamic risk? What is the "right" form of risk updates ?
3. How can we compute optimal closed loop control policies?

Areas of applications: Control of power systems and robotic motion planning



Problem Formulation

A **Markov Decision Process (MDP)** can be defined as a four-tuple $(S, U, P, U(\cdot))$:

- S is the state space, U is the control space.
- $U(x)$ is a set of admissible controls when the system state is x.
- $P(\cdot|x,u)$ is a conditional probability, given a state-control pair (x,u).

Closed loop (state feedback) control policies:

$$\pi := (\pi_0(x_0), \dots, \pi_{N-1}(x_{N-1})),\ u_k \in U(x_k),\ u_k = \pi_k(x_k),\ k \in \{0,1,\dots\}.$$

Problem $\mathcal{OPT}$:
Let $c : S \times U \to \mathbb{R}$ be the stage-wise cost and $d : S \times U \to \mathbb{R}$ be constraint cost. Solve:

$$\min_{\pi}\ \mathbb{E}\left[\sum_{k=0}^{N-1} c(x_k,u_k) + c_N(x_N)\right]$$
$$\text{subject to } \rho_{0,N}\Big(d(x_0,u_0), \dots, d(x_{N-1},u_{N-1}), 0\Big) \le r_0$$

Time Consistent Risk Metrics

Key property: A crucial property for a well-posed risk measure is time consistency.

Under a time consistent risk measure, if asset A is more risky than asset B at some time in the future, then A is more risky than B at any time prior to that point.

Time inconsistent risk measures often lead to inconsistent behavior in risk management!

If one chooses to optimize using a time inconsistent multi-period risk measure, then there is a possible scenario that at this current stage, he/she prefers an optimal policy for which he/she will regret that in the next time step.

An example with time inconsistent risk metrics:
Variance-constrained planning

$$\min_{\pi}\quad \mathbb{E}\left[\sum_{k=0}^{N-1} c(x_k,u_k) + c_N(x_N)\right]$$
$$\text{subject to}\quad \operatorname{var}\left(\sum_{k=0}^{N-1} d(x_k,u_k) + d_N(x_N)\right) \le r.$$

Basic Properties of Risk Measures

First, we have some basic assumptions on risk metrics:

- Convexity, Monotonicity, Translation invariance, Positive homogeneity, Markov risk.

Equipped with these assumptions, multi-period time consistent risk metrics can be constructed by compounding single period risk metrics:

$$\rho_{k,N} = Z_k + \rho_k(Z_{k+1} + \rho_{k+1}(Z_{k+2} + \dots + \rho_{N-2}(Z_{N-1} + \rho_{N-1}(Z_N))\dots)),$$

At the same time, any time consistent risk metrics can be decomposed into the above compound/iterative form.

Examples:

- Expectation (Risk neutral)
- Essential supremum (Worst case)
- Mean semi-deviation (Second order moments)
- Conditional value-at-risk (A convex approx. of chance constraints)
- Spectral risk measure (Capture risk aversion with utility functions, CARA, HARA etc.)
- Entropic value-at-risk (Risk measure with relative entropy)
- Super-hedging price

Dynamic Programming (DP) Result

For $\Phi_k(x_k) := [\underline{R}_N(x_k), \overline{R}_N]$, $\Phi_N(x_N) := \{0\}$, define value function:

- If $k < N$ and $r_k \in \Phi_k(x_k)$:

$$V_k(x_k,r_k) = \min_{\pi\in\Pi_k}\ J_N^{\pi}(x_k)$$
$$\text{subject to}\quad R_N^{\pi}(x_k) \le r_k(x_k);$$

the minimum is well-defined since the state and control spaces are finite.

- If $k = N$ and $r_N = 0$: $V_N(x_N, r_N) = 0$.
- If $k \le N$ and $r_k \notin \Phi_k(x_k)$: $V_k(x_k,r_k) = \infty$.

Main Result:

$$V_k(x_k,r_k) = T_k[V_{k+1}](x_k,r_k)$$

where

$$T_k[V_{k+1}](x_k,r_k) := \inf_{(u,r')\in F_k(x_k,r_k)} \Big\{ c(x_k,u) + \sum_{x_{k+1}\in S} Q(x_{k+1}|x_k,u)\, V_{k+1}(x_{k+1}, r'(x_{k+1})) \Big\},$$

and $F_k \subset \mathbb{R} \times B(S)$ is the set of control/threshold *functions*:

$$F_k(x_k,r_k) := \Big\{ (u,r') \,\Big|\, u \in U(x_k),\ r'(x') \in \Phi_{k+1}(x') \text{ for all } x' \in S,\ \text{and } d(x_k,u) + \rho_k(r'(x_{k+1})) \le r_k \Big\}.$$

$r'(x_{k+1})$ is the "Risk-to-go" function

A Conceptual Formula for Risk Updates

Time consistency of optimal control policies:
The k^{th} tail-subsequence of optimal control policies for Problem $\mathcal{OPT}$ is an optimal policy for its k^{th}-tail optimization problem.
Guaranteed by the above DP

Solve the DP $\implies$ Construction of history dependent polices.

$$\pi_k(h_k) = \begin{cases} \pi(x_0) = u^*(x_0,r_0) & \text{for } k = 0 \\ u^*(x_k,r_k),\ \text{with } r_k = r'(x_{k-1},r_{k-1})(x_k) & \text{for } k \in \{1,\dots,N-1\} \end{cases}$$

Note, the "Risk-to-go" has a Markovian structure.

The conceptual risk update (finite horizon case):

$$\mathcal{R}^*_{j+1}(x_j,r_j)(x_{j+1}) = \begin{cases} R_N^{\pi^*}(x_{j+1}) - R_N^{\pi^*}(x_j) + r_j & \text{if } j \in \{0,\dots,N-2\} \\ 0 & \text{otherwise} \end{cases}$$

Uniform Multi-grid Discretization Algorithm

The original DP has a continuous state/ control on risk threshold/ update (In general difficult to iterate)!

Idea of our approximation algorithm: We discretize the bounded continuous risk threshold/ update spaces. The step size is known as Δ.

The approximation DP operator is as follows:

$$T^D_{\Delta,k}[V](x_k,r_k) := \bar{T}^D_{\Delta,k}[V](x_k, r_k^{(\tau)})$$

where

$$\bar{T}^D_{\Delta,k}[V](x_k,r_k) := \min_{(u,r^{D,\prime})\in F^D_k(x_k,r_k)} \Big\{ c(x_k,u) + \sum_{x'\in S} Q(x'|x_k,u)\, V(x', r^{D,\prime}(x')) \Big\}$$

where F^D_k is the set of control/threshold *functions*:

$$F^D_k(x_k,r_k) := \Big\{ (u,r') \,\Big|\, u \in U(x_k),\ r^{D,\prime}(x') \in \overline{\Phi}_{k+1}(x'), \forall x' \in S,\ d(x_k,u) + \rho_k(r^{D,\prime}(x_{k+1})) \le r_k \Big\}.$$

Iterate:

$$V^D_k(x_k,r_k) := T^D_{\Delta,k}[V^D_{k+1}](x_k,r_k)$$

Error Analysis

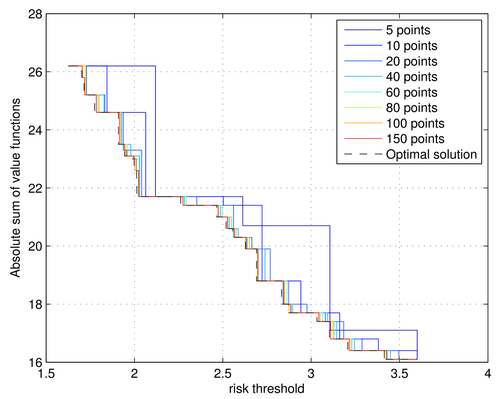
With extra assumptions on Lipschitz-ness on 1) stage-wise costs, 2) constraint costs and 3) transition probabilities, we have a linear error bound in step size Δ:

$$\|V^D_k - V_k\|_\infty \le 2M_{N,k}\Delta, \quad \exists M_{N,k} > 0.$$

When $\Delta \to 0$, $V^D_k(x_k,r_k) \to V_k(x_k,r_k)$.

Example : A 3-state, 3-action example, time Horizon $N = 3$, subjected to a multi-period mean semi-deviation constraint:

$$\rho_k(V) = \mathbb{E}[V] + 0.2\left(\mathbb{E}\left[[V - \mathbb{E}[V]]_+^2\right]\right)^{1/2}.$$



Unfortunately, this discretization scheme is severely affected by the curse of dimensionality.

Current/ Future work

- Develop techniques in Model Predictive Control with multi-period risk sensitive objectives and constraints.
- Extend the current finite horizon settings to infinite horizon settings;
- Improve the discretization method with randomized sampling or sampling with multi-resolution, in order to alleviate the curse of dimensionality;
- Study risk sensitive constrained optimal control using Approximate Dynamic Programming;
- Investigate the cases of uncertain transitions and costs via Reinforcement Learning and Data Driven Optimization;
- Lagrangian formulations of risk sensitive optimal control problems.

Contact: ychow@stanford.edu