Advances in Risk-Averse Learning[1]

Andrzej Ruszczyński

Erice 2022

[1]Supported by the NSF Award DMS-1907522 and the ONR Award N00014-21-1-2161

Distributionally Robust Learning

A supervised learning problem: $\min_{x \in X} \mathbb{E}_{D \sim \mathbb{P}} [\ell(x, D)]$

where $\ell : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ is the loss function of the predictor (controlled by x) on the random data D, and $X \subset \mathbb{R}^n$ is the feasible set. This framework includes a large class of problems, including deep learning and classification.

Key Issues

- generalization (good performance on unseen data)
- robustness w.r.t to the distribution of the data

Distributionally robust version: $\min_{x \in X} \max_{\mathbb{Q} \in \mathcal{M}(\mathbb{P})} \mathbb{E}_{D \sim \mathbb{Q}} [\ell(x, D)]$

where $\mathcal{M}(\mathbb{P})$ is a closed convex set of probability measures, defined by f-divergence, Monge distance, etc.

Difficulties of explicit robust formulations

- Restriction to convex and smooth min-max problems
- High cost when the sample size is very large
- No sequential (learning) forms when new data arrive

Modeling the uncertainty set $\mathcal{M}(\mathbb{P})$ with a risk measure

The risk minimization problem

$$\min_{x \in X} \rho\big[\ell(x, D)\big]$$

with a coherent measure of risk $\rho[\cdot]$; Coherence means convexity, monotonicity, translation equivariance, and positive homogeneity of $\rho[\cdot]$.

The dual representation

$$\rho[Z] = \max_{\mathbb{Q} \,:\, \frac{d\mathbb{Q}}{d\mathbb{P}} \in \mathcal{A}} \mathbb{E}_{\mathbb{Q}}[Z],$$

where $\mathcal{A}(\mathbb{P})$ is a convex and closed set of measures $\mathbb{Q} \ll \mathbb{P}$

The implicit min-max formulation

$$\min_{x \in X} \max_{\mathbb{Q} \,:\, \frac{d\mathbb{Q}}{d\mathbb{P}} \in \mathcal{A}} \mathbb{E}_{\mathbb{Q}}\big[\ell(x, D)\big]$$

Challenges

- We want to cover nonsmooth and nonconvex $\ell(\cdot, D)$
- Statistical estimates of $\rho[\cdot]$ and its subgradients are needed for learning

The Mean–Semideviation Risk

The first-order mean–semideviation risk measure:

$$\rho[Z] = \mathbb{E}[Z] + \varkappa\, \mathbb{E}\big[\max\big(0, Z - \mathbb{E}[Z]\big)\big], \qquad \varkappa \in [0,1]$$

The measure has the set $\mathcal{A}(\mathbb{P})$ defined as follows:

$$\mathcal{A}(\mathbb{P}) = \big\{\mu = \mathbb{1} + \xi - \mathbb{E}[\xi] : \ \xi \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P}),\ \|\xi\|_\infty \le \varkappa,\ \xi \ge 0\big\}$$

Equivalent composition optimization problem (for $Z = \ell(x, D)$)

$$\min_{x \in X}\ f(x, h(x))$$

with the functions

$$f(x,u) = \mathbb{E}\Big[\ell(x,D) + \varkappa \max\big(0, \ell(x,D) - u\big)\Big]$$
$$h(x) = \mathbb{E}[\ell(x,D)]$$

Advantage: The expected values allow for statistical estimates
Challenge: Composition implies bias

The Single Timescale Method

Three random sequences: approximate solutions $\{x^k\}$, path-averaged stochastic subgradients $\{z^k\}$, and inner function estimates $\{u^k\}$

At each iteration $k = 0, 1, 2, \ldots$, we compute

$$y^k = \operatorname*{argmin}_{y \in X} \left\{ \langle z^k, y - x^k \rangle + \frac{c}{2} \|y - x^k\|^2 \right\}$$

$$x^{k+1} = x^k + \tau_k (y^k - x^k).$$

New statistical estimates

- $\tilde{g}^{k+1} = \begin{bmatrix} \tilde{g}_x^{k+1} \\ \tilde{g}_u^{k+1} \end{bmatrix}$ of an element $g^{k+1} = \begin{bmatrix} g_x^{k+1} \\ g_u^{k+1} \end{bmatrix} \in \hat{\partial} f(x^{k+1}, u^k)$,
- $\tilde{h}^{k+1}$ of $h(x^{k+1})$, and $\tilde{J}^{k+1}$ of an element $J^{k+1} \in \hat{\partial} h(x^{k+1})$ (a row vector)

Update of the running averages

$$z^{k+1} = z^k + a\tau_k \left(\tilde{g}_x^{k+1} + \left[\tilde{J}^{k+1} \right]^\top \tilde{g}_u^{k+1} - z^k \right),$$

$$u^{k+1} = u^k + \tau_k \tilde{J}^{k+1} (y^k - x^k) + b\tau_k \left(\tilde{h}^{k+1} - u^k \right).$$

Assumptions

- The set $X \subset \mathbb{R}^n$ is convex and compact;
- For almost every (a.e.) $\omega \in \Omega$, the function $\ell(\cdot, D(\omega))$ is differentiable in a generalized (Norkin) sense with the generalized subdifferential $\hat{\partial}\ell(x, D(\omega))$. Moreover, for every compact set K an integrable function $L_K : \Omega \to \mathbb{R}$ exists, satisfying $\sup_{x\in K} \sup_{g\in\hat{\partial}\ell(x,D(\omega))} \|g\| \leq L_K(\omega)$.
- $\tau_k \in \left(0, \min(1, 1/a)\right]$ for all k, $\sum_{k=0}^{\infty} \tau_k = \infty$, $\sum_{k=0}^{\infty} \mathbb{E}\left[\tau_k^2\right] < \infty$;
- For all k,

  (i) $\tilde{g}^{k+1} = g^{k+1} + e_g^{k+1} + \delta_g^{k+1}$, with
  $g^{k+1} \in \hat{\partial} f(x^{k+1}, u^k)$, $\mathbb{E}\left\{e_g^{k+1} \middle| \mathcal{F}_k\right\} = 0$, $\mathbb{E}\left\{\|e_g^{k+1}\|^2 | \mathcal{F}_k\right\} \leq \sigma_g^2$,
  $\lim_{k\to\infty} \delta_g^{k+1} = 0$,

  (ii) $\tilde{h}^{k+1} = h(x^{k+1}) + e_h^{k+1} + \delta_h^{k+1}$, with
  $\mathbb{E}\left\{e_h^{k+1} \middle| \mathcal{F}_k\right\} = 0$, $\mathbb{E}\left\{[e_h^{k+1}]^2 | \mathcal{F}_k\right\} \leq \sigma_h^2$, $\lim_{k\to\infty} \delta_h^{k+1} = 0$,

  (iii) $\tilde{J}^{k+1} = J^{k+1} + E^{k+1} + \Delta^{k+1}$, with
  $J^{k+1} \in \hat{\partial} h(x^{k+1})$, $\mathbb{E}\left\{E^{k+1} \middle| \mathcal{F}_k\right\} = 0$, $\mathbb{E}\left\{\|E^{k+1}\|^2 | \mathcal{F}_k\right\} \leq \sigma_E^2$,
  $\lim_{k\to\infty} \Delta^{k+1} = 0$, and $\mathbb{E}\left[(E^{k+1})^\top e_{gu}^{k+1} \,\middle|\, \mathcal{F}_k\right] = 0$

Generation of Random Estimates

At each iteration, we sample an independent Bernoulli random variable β with $\mathbb{P}[\beta = 1] = \varkappa$ and $\mathbb{P}[\beta = 0] = 1 - \varkappa$, and we set

$$\tilde{g}_x^{k+1} \in \begin{cases} \hat{\partial}\ell(x^{k+1}, D_1^{k+1}) & \text{if } \beta = 0 \text{ or } \ell(x^{k+1}, D_1^{k+1}) < u^k, \\ 2\hat{\partial}\ell(x^{k+1}, D_1^{k+1}) & \text{if } \beta = 1 \text{ and } \ell(x^{k+1}, D_1^{k+1}) \geq u^k, \end{cases}$$

$$\tilde{g}_u^{k+1} = \begin{cases} 0 & \text{if } \beta = 0 \text{ or } \ell(x^{k+1}, D_1^{k+1}) < u^k, \\ -1 & \text{if } \beta = 1 \text{ and } \ell(x^{k+1}, D_1^{k+1}) \geq u^k, \end{cases}$$

$$\tilde{h}^{k+1} = \ell(x^{k+1}, D_1^{k+1}),$$

$$\tilde{J}^{k+1} \in \begin{cases} \hat{\partial}\ell(x^{k+1}, D_1^{k+1}) & \text{if } \beta = 0, \\ \hat{\partial}\ell(x^{k+1}, D_2^{k+1}) & \text{if } \beta = 1. \end{cases}$$

Here D_1^{k+1} and D_2^{k+1} are independent samples from the distribution of D.

The need for the second sample from the data, D_2^{k+1}, occurs only if $\beta = 1$, that is, with probability $\varkappa$. Therefore, on average $1 + \varkappa$ samples are needed per iteration.

Convergence

Additional assumption:

- The set $F(X^*)$ does not contain an interval of nonzero length.

Theorem

With probability 1 every accumulation point $\hat{x}$ of the sequence $\{x^k\}$ is stationary, $\lim_{k\to\infty}(u^k - h(x^k)) = 0$, and the sequence $\{F(x^k)\}$ is convergent.

The analysis uses the differential inclusion method, relating the interpolated trajectories of the method to a solution to the system

$$\left(\dot{x}(t), \dot{z}(t), \dot{u}(t)\right) \in \Gamma\left(x(t), z(t), u(t)\right)$$

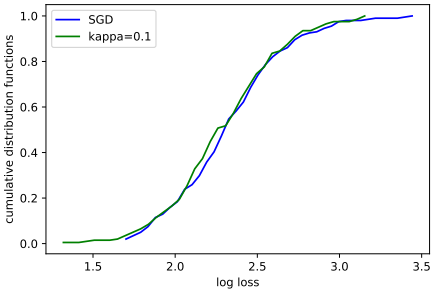
with a convex and compact valued multifunction $\Gamma(\cdot,\cdot,\cdot)$.

The Lyapunov function:

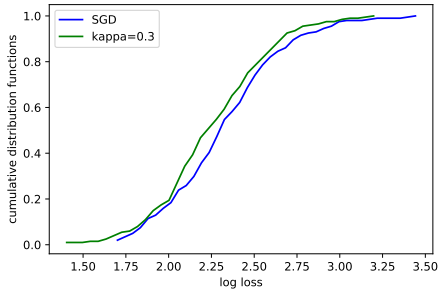
$$W(x,z,u) = \underbrace{af(x,u) - \min_{y\in X}\left\{\langle z, y-x\rangle + \frac{c}{2}\|y-x\|^2\right\}}_{\text{gap (if } z\in\hat{\partial}F(x) \text{ and } u = h(x))} + \gamma \underbrace{\|h(x)-u\|}_{\text{tracking error}}$$

Example: Deep Learning

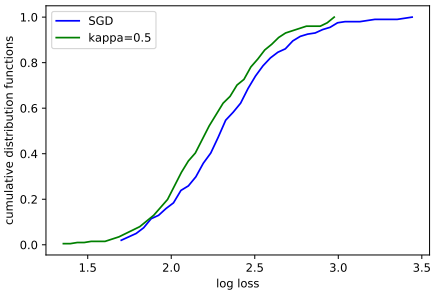
CIFAR10 Dataset: 60 000 color images of size 32×32 in 10 different classes
Model: 3-layer fully connected neural network with 328 510 parameters



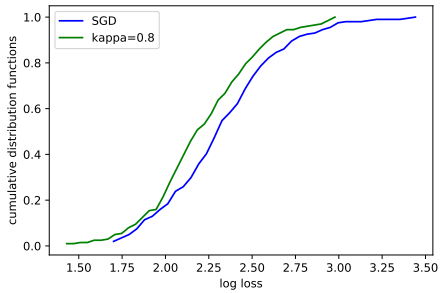
(a) $\varkappa = 0.1$



(b) $\varkappa = 0.3$



(c) $\varkappa = 0.5$



(d) $\varkappa = 0.8$

Figure: The CDFs of the loss of the SGD solution and the STS solution on the original data. The models are trained with contaminated data.

Markov Decision Processes

- State space $\mathcal{X}$ (finite but large)
- Control space $\mathcal{U}$ (finite)
- Feasible control set $U : \mathcal{X} \rightrightarrows \mathcal{U}$,
- Controlled transition kernel $Q : \text{graph}(U) \to \mathcal{P}(\mathcal{X})$,
  $\mathcal{P}(\mathcal{X})$ - set of probability measures on $\mathcal{X}$
- Cost functions $c : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$, $t = 1, 2, \ldots$
- Stationary Markov policy $\Pi = \{\pi, \pi, \ldots\}$ with $\pi : \mathcal{X} \to \mathcal{U}$,

$$x_t \longrightarrow u_t = \pi(x_t)$$

$$(x_t, u_t) \longrightarrow x_{t+1} \sim Q(x_t, u_t)$$

Markov Risk Measures

For a given policy Π, we observe a random sequence of costs

$$c(X_t, \pi(X_t)), \quad t = 1, 2, \dots$$

where the process $\{X_t\}$ is generated by the Markov chain with the transition probability matrix P^{Π}:

$$P^{\Pi}(i, j) = Q\big(j \,\big|\, i, \pi(i)\big), \quad i, j \in \mathcal{X},$$

We introduce the notation:

$$c_i^{\Pi} = c(i, \pi(i)), \quad i \in \mathcal{X}.$$

Markov risk measures evaluate the risk of discounted future costs as a function of the current state:

$$v_i^{\Pi} = \rho_{1,\infty}\big(c(X_1, \pi(X_1)), \alpha c(X_2, \pi(X_2)), \alpha^2 c(X_3, \pi(X_3)), \dots\big)$$

with $X_1 = i$ and a Markov policy $\Pi = (\pi, \pi, \dots)$.

The Structure of Markov Risk Measures. Policy Evaluation

Under the conditions of time consistency, translation, monotonicity, and normalization of the risk measure, we have the policy evaluation equation

$$v_i^\Pi = c_i^\Pi + \alpha\sigma\big(i, P_i^\Pi, v^\Pi\big), \quad i \in \mathcal{X}, \quad t = 0, 1, 2, \ldots.$$

Here, $\sigma : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \times \mathcal{V} \to \mathbb{R}$ is a transition risk mapping: a generalization of the usual conditional expected value.

Classical Case: Expectation

$$\sigma\big(i, P_i^\Pi, v^\Pi\big) = \sum_{j\in\mathcal{X}} P_{ij}^\Pi v_j^\Pi.$$

Risk-Averse Example: Mean–Semideviation

$$\sigma\big(i, P_i^\Pi, v^\Pi\big) = \underbrace{\sum_{j\in\mathcal{X}} P_{ij}^\Pi v_j^\Pi}_{\mu_i} + \kappa \sum_{j\in\mathcal{X}} P_{ij}^\Pi \big(v_j^\Pi - \mu_i\big)_+, \quad \kappa \in [0, 1]$$

Policy Evaluation as Very Large System of Nonsmooth Equations

The risk-averse policy evaluation equation:

$$v_i^{\Pi} = c_i^{\Pi} + \alpha\sigma\big(i, P_i^{\Pi}, v^{\Pi}\big), \quad i \in \mathcal{X}.$$

We introduce the space $\mathcal{Q}$ of transition kernels on $\mathcal{X}$, define a vector-valued transition risk operator $S : \mathcal{Q} \times \mathcal{V} \to \mathcal{V}$, with components

$$S_i(P^{\Pi}, v) \stackrel{\Delta}{=} \sigma(i, P_i^{\Pi}, v), \quad i \in \mathcal{X},$$

and rewrite the last equation as a nonsmooth equation:

$$v^{\Pi} = c^{\Pi} + \alpha S(P^{\Pi}, v^{\Pi})$$

While it can be solved by a nonsmooth Newton's method and the resulting evaluation used in a policy iteration method, all these techniques require solving linear equations with the full transition probability matrix P^{Π} and become impractical, when the size of the state space is very large.

The Projected Policy Evaluation Equation

We assume that each state $i \in \mathcal{X}$ has a number of relevant features $\varphi_j(i) \in \mathbb{R}$, $j = 1, \ldots, m$, where $m \ll |\mathcal{X}|$, and that the value v_i^Π of a state can be approximated by a linear combination of its features:

$$v_i^\Pi \approx \tilde{v}_i^\Pi = \sum_{j=1}^{m} r_j \varphi_j(i), \quad i \in \mathcal{X}, \qquad v^\Pi \approx \tilde{v}^\Pi = \Phi r$$

with the feature matrix $\Phi = \begin{bmatrix} \varphi^\top(1) \\ \varphi^\top(2) \\ \vdots \\ \varphi^\top(n) \end{bmatrix}$.

With a projection operator $L : \mathcal{V} \to \text{range}(\Phi)$, we formulate the equation

$$\Phi r = L\big(c^\Pi + \alpha S(P^\Pi, \Phi r)\big)$$

Projection and Distortion

Assumption

The system under policy Π is ergodic with stationary probabilities q.

The "orthogonal" projection:

$$L(w) = \operatorname*{argmin}_{z \in \text{range}(\Phi)} \|z - w\|_q, \quad w \in \mathcal{V}.$$

with

$$\langle v, w \rangle_q = \sum_{i=1}^n q_i v_i w_i, \quad \|w\|_q^2 = \langle w, w \rangle_q.$$

The dual representation of each component of a coherent S:

$$S_i(P_i, v) = \max_{\zeta_i \in \mathcal{A}(i, P_i)} \sum_{j \in \mathcal{X}} \zeta_{ij} P_{ij} v(j), \quad i \in \mathcal{X}.$$

The distortion coefficient (risk premium) of the operator S

$$\kappa = \max \left\{ |\zeta_{ij} - 1| : \zeta_i \in \mathcal{A}(i, P_i), \ p_{ij} > 0, \ i, j \in \mathcal{X} \right\}.$$

Contraction of The Policy Evaluation Operator

The usual subgradients of $S_i(P_i, \cdot)$:

$$\partial S_i(P_i, 0) = \left\{ m_i : \exists \left(\zeta_i \in \mathcal{A}(i, P_i) \right) \; m_{ij} = \zeta_{ij} p_{ij}, \; j \in \mathcal{X} \right\}, \quad i \in \mathcal{X}.$$

The transition risk operator satisfies for all $w, v \in \mathcal{V}$ the inequality:

$$\| S(P, w) - S(P, v) \|_q \leq \sqrt{1 + \kappa} \, \| w - v \|_q.$$

Consider the operator

$$\widetilde{\mathfrak{D}}_\pi(v) = L\big(c + \alpha S(P, v)\big), \quad v \in \mathcal{V},$$

The policy evaluation equation:

$$v = \widetilde{\mathfrak{D}}_\pi v.$$

If $\alpha \sqrt{1 + \kappa} < 1$ then the equation has a unique solution v^Π.

If Φ has full column rank, only one r satisfies $v^\Pi = \Phi r$.

The Risk-Averse Method of Temporal Differences

The risk-averse temporal difference:

$$d_t = \underbrace{\varphi^\top(i_t) r_t}_{\approx v(i_t)} - c(i_t) - \alpha\sigma(i_t, P_{i_t}, \underbrace{\Phi r_t}_{\approx v(\cdot)}), \quad t = 0, 1, 2, \ldots.$$

We assume that we can observe a random estimate $\widetilde{\sigma}(i_t, P_{i_t}, \cdot)$, such that

$$\widetilde{\sigma}(i_t, P_{i_t}, \Phi r_t) = \sigma(i_t, P_{i_t}, \Phi r_t) + \xi_t, \quad t = 0, 1, 2, \ldots,$$

with some random errors ξ_t. The observed risk-averse temporal differences,

$$\widetilde{d}_t = \varphi^\top(i_t) r_t - c(i_t) - \alpha\widetilde{\sigma}(i_t, P_{i_t}, \Phi r_t), \quad t = 0, 1, 2, \ldots,$$

The Method

For a simulated trajectory $\{i_1, i_2, \ldots, i_t, \ldots\}$ of the system, evaluate

$$r_{t+1} = r_t - \gamma_t \varphi(i_t)\, \widetilde{d}_t, \quad t = 0, 1, 2, \ldots,$$

with stepsizes $\gamma_t > 0$.

A Deterministic Model

- The random errors ξ_t are temporarily ignored
- The updates of $\{r_t\}$ are averaged over all states with the distribution q.

Using the matrix $Q = \mathrm{diag}(q)$, we define the operator:

$$U(r) = \mathbb{E}_{i\sim q}\big[\varphi(i)\big(\varphi^\top(i)r - c(i) - \alpha\sigma(i, P_i, \Phi r)\big)\big] \\ = \Phi^\top Q\big[\Phi r - c - \alpha\, S(P, \Phi r)\big].$$

The deterministic analog of the method:

$$\bar{r}_{t+1} = \bar{r}_t - \gamma\, U(\bar{r}_t), \quad t = 0, 1, 2, \ldots, \quad \gamma > 0.$$

By the definition of the projection, a point r^* is a solution if and only if

$$r^* = \operatorname*{argmin}_r \frac{1}{2}\big\|\Phi r - \big(c + \alpha S(P, \Phi r^*)\big)\big\|_q^2.$$

This occurs if and only if r^* is a zero of $U(\cdot)$.

If $\alpha\sqrt{1+\kappa} < 1$, then for all $\gamma \in (0, \gamma_0)$, with $\gamma_0 > 0$, the algorithm generates a sequence $\{\bar{r}_t\}$ convergent to a point r^* such that $U(r^*) = 0$.

Convergence in the Stochastic Case

Assumptions

The sequence $\{\gamma_t\}$ is adapted to the filtration $\{\mathcal{F}_t\}$ and such that

(i) $\gamma_t > 0,\ t = 0, 1, \ldots,$ a.s.;

(ii) $\sum_{t=0}^{\infty} \gamma_t = \infty$ a.s.;

(iii) $\mathbb{E} \sum_{t=0}^{\infty} \gamma_t^2 < \infty$;

(iv) For any $\varepsilon > 0$, $\lim_{t_0 \to \infty} \sup_{\{T : \sum_{t=t_0}^{T} \gamma_t \le \varepsilon\}} \sum_{t=t_0}^{T} |\gamma_t - \gamma_{t+1}| = 0$ a.s.

The sequence of errors $\{\xi_t\}_{t \ge 1}$ satisfies for $t = 0, 1, 2 \ldots$ the conditions

(v) $\mathbb{E}[\xi_t \,|\, \mathcal{F}_t] = 0$ a.s.;

(vi) $\mathbb{E}[\|\xi_t\|^2 \,|\, \mathcal{F}_t] \le C(1 + \|r_t\|^2)$ a.s., with some constant $C > 0$.

Theorem

Suppose the stepsizes and random estimates $\widetilde{\sigma}_{i_t}(P_{i_t}, \Phi r_t)$ satisfy the general assumptions and $\alpha\sqrt{1+\varkappa} < 1$. If the sequence $\{r_t\}$ is bounded with probability 1, then every accumulation point of the sequence $\{r_t\}$ is a solution with probability 1.

Example (Powell & Topaloglu, 2006)

- $K = 200$ vehicles
- $M = 50$ locations
- Stochastic demand D_{ijt} for transportation from location i to location j at time $t = 1, 2, \ldots$. The demand arrays D_t in different time periods are independent
- Only vehicles available at location i may be used to satisfy the demand
- The vehicles may also be moved empty
- There are costs of moving the vehicles and rewards for moving cargo.

The state x_t of the system at time t is the M-dimensional integer vector containing the numbers of vehicles at each location.

The size of the state space is $\binom{K+M-1}{M-1} \sim 10^{427}$.

Approximate Dynamic Programming

Control u_t - all decisions to move vehicles and load cargo.
They are made after the demand D_t is observed.
Next state:

$$x_{t+1} = x_t - Au_t, \qquad \text{(balances of incoming and outgoing vehicles)}$$

Optimal control:

$$u_t^*(x_t, D_t) = \underset{u \in \mathcal{U}(x_t, D_t)}{\operatorname{argmin}} \Big\{ c^\top u + \alpha \underbrace{v(x_t - Au)}_{\text{value function}} \Big\}.$$

Approximate policy Π:

$$u_t^\pi(x_t, D_t) = \underset{u \in \mathcal{U}(x_t, D_t)}{\operatorname{argmin}} \Big\{ c^\top u + \alpha \underbrace{\pi^\top (x_t - Au)}_{\text{approximate value}} \Big\}.$$

π_j is the assumed "cost" of having a vehicle at location j.
We want to evaluate the policy Π.

Features and the TD Method

Features $\varphi(x_t)$ of the state x_t: the state x_t itself

Value function approximation:

$$\widetilde{v}^{\pi}(x_t) = r^{\top} x_t.$$

The observed temporal difference (calculated by simulation):

$$\widetilde{d}_t = r_t^{\top} x_t - \alpha \widetilde{\sigma}\Big(P, c^{\top} u^{\pi}(x_t, D) + \alpha r_t^{\top}\big(x_t - A u^{\pi}(x_t, D)\big)\Big).$$

The method:

$$r_{t+1} = r_t - \gamma x_t \, \widetilde{d}_t, \qquad \gamma > 0$$

In the policy iteration method, after learning the coefficients r^*,
we set $\pi \leftarrow r^*$ (policy improvement).
In fact, we may put $\pi \leftarrow r_t$ at every iteration (the "optimistic" version), which is not always convergent, but which works well in our case.

Simulation Results for TD(0)

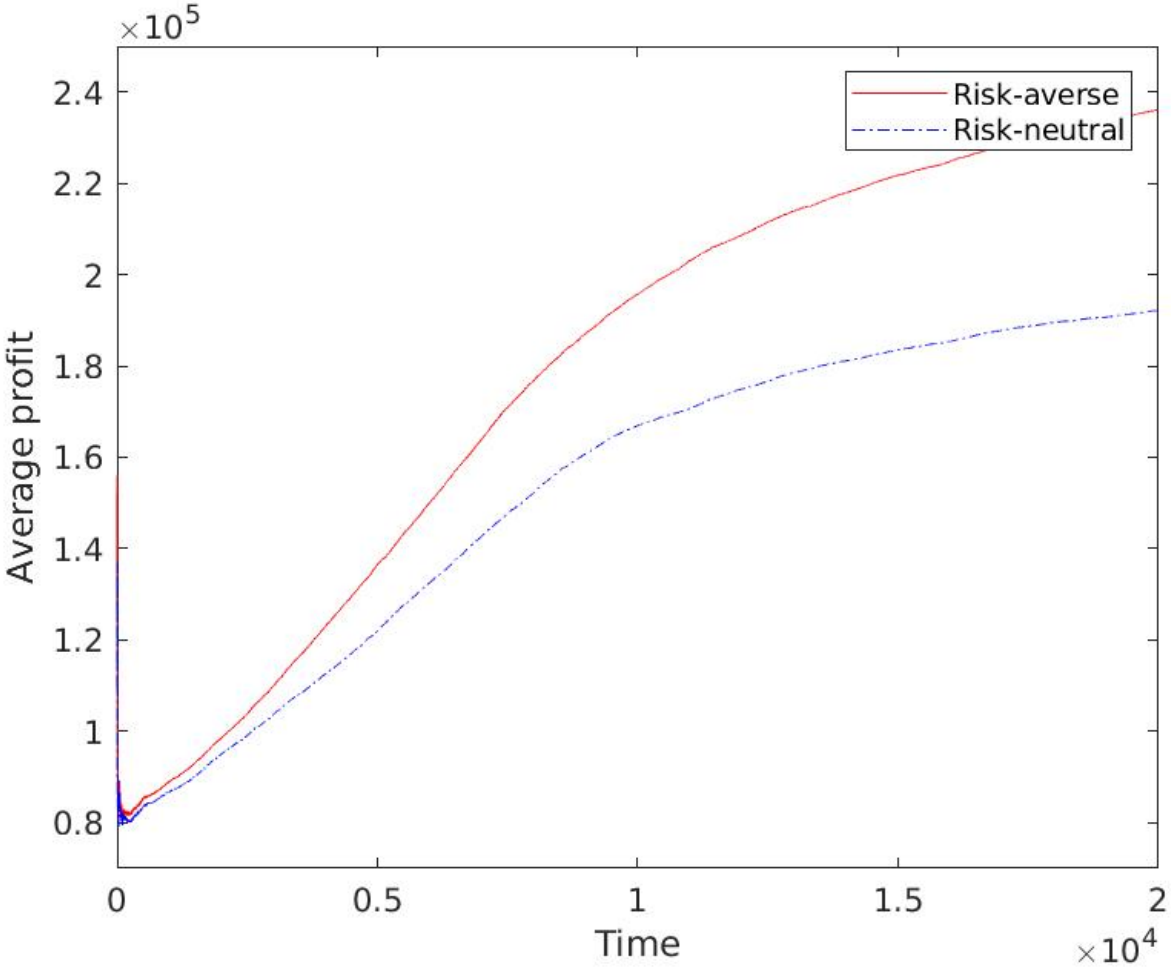


Figure: The average profit per stage in the risk-averse and risk-neutral methods.

Unfortunately TD(λ) does not work well here.

References

- M. Gürbüzbalaban, A. Ruszczyński, L. Zhu, A stochastic subgradient method for distributionally robust non-convex learning, arXiv preprint arXiv:2006.04873, 2020
- S. Ghadimi, A. Ruszczyński, M. Wang, A single timescale stochastic approximation method for nested stochastic optimization, *SIAM Journal on Optimization*, 30 (2020) 960-979
- A. Ruszczyński, A stochastic subgradient method for nonsmooth nonconvex multi-level composition optimization, *SIAM Journal on Control and Optimization*, 59 (2021) 2301-2320
- A. Ruszczyński, Risk-averse dynamic programming for Markov decision processes, *Mathematical Programming*, Series B 125 (2010) 235–261
- D. Dentcheva and A. Ruszczyński, Risk forms: representation, disintegration, and application to partially observable two-stage systems, *Mathematical Programming* 181(2)(2020), 297–317.
- Ü. Köse, A. Ruszczyński, Risk-Averse Learning by Temporal Difference Methods with Markov Risk Measures, *Journal of Machine Learning Research*, 22 (2021) 1-34

. . . and the references therein.