12th workshop on Markov processes and related topics

Minimum average value-at-risk for finite horizon semi-Markov decision processes

Xianping Guo (with Y.H. HUANG)
Sun Yat-Sen University
Email: mcsgxp@mail.sysu.edu.cn

13-17 July 2016, Xuzhou

Outline

- The model of SMDPs
- The AVaR-optimality problem
- Expected-positive-deviation problems
- The existence of AVaR-optimal policies
- Algorithm Aspects
- Applied examples

1. The model of SMDPs

$$\{E, (A(x), x \in E), Q(\cdot, \cdot | x, a), c(x, a)\}$$

- E: State space, endowed with the Borel σ–algebras $\mathcal{B}(E)$.
- $A(x)$: The finite set of available actions at state $x \in E$.
- $Q(dt, dy|x, a)$: Semi-Markov kernel depending on the current states x and the taken action $a \in A(x)$. According to the Radon-Nikodym theorem, the Q can be partitioned as

$$Q(dt, dy|x, a) = \int_{dy} F(dt|x, a, z)p(dz|x, a) \qquad (1)$$

- $c(x, a)$: Cost function of states x and actions a

SMDPS: The meaning of the model data above: If the system occupies state x_0 at the initial time $t_0 \geq 0$, a controller chooses an action $a_0 \in A(x_0)$ according to some decision rule. As a consequence of this action choice, two things occur:

First, the system jumps to state $x_1 \in E$ after a sojourn time $\theta_1 \in (0, \infty)$ in x_0, with the distribution $F(\cdot|x_0, a_0, x_1)$;

Second, costs are continuously accumulated at rate $c(x_0, a_0)$ for a period of time θ_1.

At time $(t_0 + \theta_1)$, the controller chooses an action $a_1 \in A(x_1)$ according to some decision rule, and the same sequence of events occur.

From the evolving of a SMDP, we obtain an admissible history $h_n := (t_0, x_0, a_0, \theta_1, x_1, a_1, \ldots, \theta_n, x_n)$. Let $t_n := t_{n-1} + \theta_n$.

- Randomized history-dependent policy π: $\pi := \{\pi_n\}$ of stochastic kernels $\{\pi_n(da|h_n)\}$ on A s.t. $\pi_n(A(x_n)|h_n) = 1$

- Markov policy $\pi := \{\pi_n\}$: $\pi_n(da|t, x)$ depending on (n, t, x)

- stationary policy f : Measurable map f, $f(t, x) \in A(x)$

- Π: The class of all randomized history-dependent policies.

- Π_{RM}: The class of all randomized Markov policies.

- F : The class of all stationary policies.

2. The AVaR-optimality problem

Given the semi-Markov kernel Q, an initial time-state pair $(t,x) \in [0,\infty) \times E$, and a policy $\pi \in \Pi$, the Ionescu Tulcea theorem ensures the a unique probability measure space $(P^{\pi}_{(t,x)}, \Omega, \mathcal{F})$ and a process $\{T_n, X_n, A_n\}$ such that

$$P^{\pi}_{(t,x)}(T_0 = t, X_0 = x) = 1,$$

$$P^{\pi}_{(t,x)}(A_n \in da|h_n) = \pi_n(da|h_n),$$

$$P^{\pi}_{(t,x)}(T_{n+1} - T_n \in dt, X_{n+1} \in dy|h_n, a_n) = Q(dt, dy|x_n, a_n),$$

$E^{\pi}_{(t,x)}$: the expectation operator with respect to $P^{\pi}_{(t,x)}$.

Let $T_\infty := \lim_{k\to\infty} T_k$ be the explosive time of the system. Although T_∞ may be finite, we do not intend to consider the controlled process after the moment T_∞. For $t < T_\infty$, let

$$Z(t) := \sum_{n\geq 0} I_{\{T_n \leq t < T_{n+1}\}} X_n, \; U(t) := \sum_{n\geq 0} I_{\{T_n \leq t < T_{n+1}\}} A_n$$

denote the underlying state and action processes, respectively, where I_D stands for the indicator function on a set D.

In the following, we consider a T-horizon SMDP (with $T > 0$). To make the T-horizon SMDP sensible, we need to avoid the possibility of infinitely many jumps during the interval $[0, T]$, and thus the condition below is introduced.

Assumption 1: $P^{\pi}_{(t,x)}(\{T_\infty > T\}) \equiv 1.$

Assumption 1 above is trivially fulfilled in discrete-time MDPs with $T_\infty = \infty$, and also holds under many conditions (Ref., Huang & Guo, European. J. Oper. Res.,2011; Puterman, John Wiley & Sons Inc., New York, 1994).

We suppose Assumption 1 is satisfied *throughout the paper*.

Define the *value-at-risk* (VaR) of finite horizon total cost at level $\gamma \in (0,1)$ under a policy $\pi \in \Pi$ by

$$\zeta_\gamma^\pi(t,x) := \inf\Big\{\lambda \mid P^\pi_{(t,x)}\Big(\int_t^T c(Z(s),U(s))ds \le \lambda\Big) \ge \gamma\Big\},$$

which denotes the maximum cost over the time horizon $[t,T]$ that might be incurred with probability at least γ.

$\eta_\gamma^\pi(t,x)$: The *average value-at-risk* (AVaR) of finite horizon total cost at level γ under a policy $\pi \in \Pi$ is given

$$\eta_\gamma^\pi(t,x) := \frac{1}{1-\gamma}\int_\gamma^1 \zeta_s^\pi(t,x)ds$$

$$= E^\pi_{(t,x)}\Big[\int_t^T c(Z(s),U(s))ds \Big| \int_t^T c(Z(s),U(s))ds \ge \zeta_\gamma^\pi(t,x)\Big]$$

Our AVaR minimization problem (Prob-1): minimizing η_γ^π over $\pi \in \Pi$, that is, we aim to find $\pi^* \in \Pi$ such that

$$\eta_\gamma^{\pi^*}(t,x) = \inf_{\pi \in \Pi} \eta_\gamma^\pi(t,x) =: \eta_\gamma^*(t,x),$$

which is the value function (or minimum AVaR).

Such a policy π^*, when it exists, is called AVaR optimal.

Our goal is to

- prove the existence of an optimal policy,
- present an algorithm for optimal policies, the value function
- give computable examples to show the application.

3. Expected-positive-deviation problems

Lemma 1: Let $\pi \in \Pi$ and $\gamma \in (0,1)$. Then, for every $(t,x) \in [0,T] \times E$, we have:

$$\eta_\gamma^\pi(t,x) = \min_\lambda \left\{ \lambda + \frac{1}{1-\gamma} E_{(t,x)}^\pi \left[\int_t^T c(Z(s),U(s))ds - \lambda \right]^+ \right\}$$

and the minimum-point is given by $\lambda^*(t,x) = \zeta_\gamma^\pi(t,x)$.

By Lemma 1, the value function can be rewritten as follows:

$$\eta_\gamma^*(t,x) = \inf_\lambda \left\{ \lambda + \frac{1}{1-\gamma} \inf_{\pi \in \Pi} E_{(t,x)}^\pi \left[\int_t^T c(Z(s),U(s))ds - \lambda \right]^+ \right\}$$

Hence, to solve our original problem, we define the expected-positive-deviation (EPD) from a level λ under $\pi \in \Pi$ by

$$J^\pi(t,x,\lambda) := E^\pi_{(t,x)}\left[\int_t^T c(Z(s),U(s))ds - \lambda\right]^+$$

where, λ can be interpreted as the acceptable cost/loss.

Fixed λ. Our goal now is to minimize $J^\pi(\cdot,\cdot,\lambda)$ over $\pi \in \Pi$.

The EPD-minimization problem (Prob-2): An EPD-optimal policy $\pi^*_\lambda \in \Pi$ (depending on λ) satisfying

$$J^{\pi^*_\lambda}(t,x,\lambda) = \inf_{\pi\in\Pi} J^\pi(t,x,\lambda) =: J^*(t,x,\lambda),$$

which denotes the value function for the EPD criterion.

To solve Prob-2 depending on the cost level λ, we introduce some new notation.

- λ_0: the initial cost level,
- $\lambda_{m+1} := \lambda_m - c(x_m, a_m)(t_{m+1} - t_m)$: the cost level at the $(m+1)$th jump time. (This is because there is a cost $c(x_m, a_m)(t_{m+1} - t_m)$ incurred between the two jumps.)

Since the levels $\{\lambda_m\}$ usually affect the behavior of the controller, we imbed them into histories of the form:

$$\tilde{h}_n := (t_0, x_0, \lambda_0, a_0, \dots, t_{n-1}, x_{n-1}, \lambda_{n-1}, a_{n-1}, t_n, x_n, \lambda_n).$$

For the general state space $\widetilde{E} := [0, \infty) \times E \times (-\infty, \infty)$,

- A randomized history-dependent general policy $\widetilde{\pi} = \{\widetilde{\pi}_n\}$: stochastic kernels $\widetilde{\pi}_n$ on A satisfying $\widetilde{\pi}_n(A(x_n) \mid \tilde{h}_n) \equiv 1$.
- $\widetilde{\Pi}$: class of randomized history-dependent general policies
- $\widetilde{\Pi}_{RM}$: class of all randomized general Markov policies
- $\widetilde{\mathbb{F}}$: class of all stationary general policies.

Accordingly, for each $(t, x, \lambda) \in [0, T] \times E \times R$, we define the expected-positive-deviation of finite horizon cost from the level λ under a policy $\tilde{\pi} \in \tilde{\Pi}$ by

$$V^{\tilde{\pi}}(t, x, \lambda) := E^{\tilde{\pi}}_{(t,x,\lambda)} \left[\int_t^T c(Z(s), U(s)) ds - \lambda \right]^+$$

Lemma 2. Fix any λ. Then, for each $\tilde{\pi} \in \widetilde{\Pi}$, there exists a λ-depending policy $\pi^\lambda = \{\pi_0^\lambda, \pi_1^\lambda, \ldots\} \in \Pi$ such that

$$J^{\pi^\lambda}(t, x, \lambda) = V^{\tilde{\pi}}(t, x, \lambda)$$

where $\pi_0^\lambda(\cdot|t_0, x_0) := \tilde{\pi}_0(\cdot|t_0, x_0, \lambda)$, $\pi_1^\lambda(\cdot|t_0, x_0, a_0, t_1, x_1) = \tilde{\pi}_1(\cdot|t_0, x_0, \lambda, a_0, t_1, x_1, \lambda - c(x_0, a_0)(t_1 - t_0)), \ldots\ldots$

Lemma 2 shows that Prob-2 is equivalent the following one

Prob-3: Find a so called EPD-optimal policy $\tilde{\pi}^* \in \tilde{\Pi}$ such that

$$V^{\tilde{\pi}^*}(t, x, \lambda) = V^*(t, x, \lambda)$$

where

$$V^*(t, x, \lambda) = \inf_{\tilde{\pi} \in \tilde{\Pi}} V^{\tilde{\pi}}(t, x, \lambda),$$

is also called the value function.

To analyze Prob-3, we introduce some notation.

Let $\mathbb{M} := \{$ measurable $v \geq 0$ on $[0,T] \times E \times \mathbb{R}\}$.

Define operators H and $H^{\tilde{\varphi}}$ ($\tilde{\varphi}(da|t,x,\lambda)$) as follows:

$$\begin{aligned} H^{\tilde{\varphi}} v(t,x,\lambda) &:= \int_{A(x)} \tilde{\varphi}(da|t,x,\lambda) H^a v(t,x,\lambda) \\ Hv(t,x,\lambda) &:= \inf_{A(x)} H^a v(t,x,\lambda) \end{aligned}$$

for all $v \in \mathbb{M}$, where, for each $a \in A(x)$,

$$\begin{aligned} H^a v(t,x,\lambda) := {} & (1 - Q(T-t, E \mid x,a))(\lambda - c(x,a)(T-t))^- \\ & + \int_E \int_0^{T-t} Q(ds, dy|x,a) v(t+s, y, \lambda - c(x,a)s) \end{aligned}$$

Moreover, define $V^{\tilde{\pi}}_{-1}(t,x,\lambda) := (0-\lambda)^+ = \lambda^-$, and

$$V^{\tilde{\pi}}_n(t,x,\lambda) := E^{\tilde{\pi}}_{(t,x,\lambda)}\Big[\sum_{m=0}^{n} c(X_m, A_m)((T - T_m)^+ \wedge \Theta_{m+1}) - \lambda\Big]^+$$

for every $(t,x,\lambda) \in [0,T] \times E \times \mathbb{R}$ and $n \geq 0$.

Lemma 3. $\lim_{n\to\infty} V^{\tilde{\pi}}_n = V^{\tilde{\pi}}$.

Hence, we shall calculate $V^{\tilde{\pi}}_n$ so as to compute $V^{\tilde{\pi}}$. A basic lemma is now given.

Lemma 4: Suppose Assumption 1 holds. For each $\widetilde{\pi} = \{\tilde{\pi}_0, \tilde{\pi}_1, \ldots\} \in \widetilde{\Pi}$, and $n \geq -1$, we have

$$V_{n+1}^{\tilde{\pi}}(t,x,\lambda) = \int_{A(x)} \tilde{\pi}_0(da|t,x,\lambda) H^a V_n^{{}^{(1)}\tilde{\pi}^{(t,x,\lambda,a)}}(t,x,\lambda),$$

$$V^{\tilde{\pi}}(t,x,\lambda) = \int_{A(x)} \tilde{\pi}_0(da|t,x,\lambda) H^a V^{{}^{(1)}\tilde{\pi}^{(t,x,\lambda,a)}}(t,x,\lambda),$$

where ${}^{(1)}\tilde{\pi}^{(t,x,\lambda,a)} = \{{}^{(1)}\tilde{\pi}_0^{(t,x,\lambda,a)}, {}^{(1)}\tilde{\pi}_1^{(t,x,\lambda,a)}, \ldots\}$ is a shift-policy defined by

$$\begin{aligned} &{}^{(1)}\tilde{\pi}_k^{(t,x,\lambda,a)}(\cdot|t_1,x_1,\lambda_1,a_1,\ldots,t_{k+1},x_{k+1},\lambda_{k+1}) \\ &:= \tilde{\pi}_{k+1}(\cdot|t,x,\lambda,a,t_1,x_1,\lambda_1,a_1,\ldots,t_{k+1},x_{k+1},\lambda_{k+1}) \end{aligned}$$

Assumption 2. $0 \le c(x,a) \le \bar{C}$ for all $(x,a) \in K$, and some constant $\bar{C} > 0$.

Inspired by the definition of $V^{\tilde{\pi}}$, we denote by $\mathbb{M}_1$ the set

$$\mathbb{M}_1 := \{v \in \mathbb{M} \mid \max\{0, -\lambda\} \le v(t,x,\lambda) \le (\bar{C}(T-t)-\lambda)^+\}$$

Lemma 5: Suppose Assumptions 1 and 2. hold. Then:

(a) $V_n^{\tilde{\pi}} \uparrow V^{\tilde{\pi}}$ as $n \to \infty$, and $V^{\tilde{\pi}} \in \mathbb{M}_1$ for each $\tilde{\pi}$.

(b) For any $\tilde{f} \in \widetilde{\mathbb{F}}$, $V^{\tilde{f}}$ is a minimum solution in $\mathbb{M}_1$ to the equation $v = H^{\tilde{f}} v$.

Theorem 1 (Solvability of Prob-3). Under Assumptions 1 and 2, the following assertions are true.

(a) For each $(t, x, \lambda) \in [0, T] \times E \times \mathbb{R}$, let

$$V^*_{-1}(t, x, \lambda) := \lambda^-, \ V^*_{n+1}(t, x, \lambda) := HV^*_n(t, x, \lambda), \ n \geq -1.$$

Then, the V^*_n increase in n, and $\lim_{n\to\infty} V^*_n = V^* \in \mathbb{M}_2$.

(b) V^* is a minimum solution in $\mathbb{M}_2$ to the optimality equation $v = Hv$.

(c) There exists an $\tilde{f} \in \widetilde{\mathbb{F}}$ such that $V^* = H^{\tilde{f}} V^*$, and such a policy is EPD-optimal for Prob-3.

Theorem 1 proposes a value iteration algorithm for computing the value function V^* and an optimal policy for Prob-3, which we discuss in more detail below.

Note that V^* is a minimum (rather than *the unique*) solution in $\mathbb{M}_2$ to the optimality equation $v = Hv$. To further ensure the uniqueness for the requirement of the policy improvement algorithms, we need the following condition.

Assumption 3. There exist constants $\sigma > 0$ and $0 < \rho < 1$ such that

$$F(\sigma|x, a, y) \leq 1 - \rho$$

for all (x, a, y), where $F(\cdot|x, a, y)$ is as in (1).

Theorem 2. Under Assumptions 1-3, we have the following statements.

(a) $\lim_{n\to\infty} \sup_{(t,x,\lambda)} |V_n^*(t,x,\lambda) - V^*(t,x,\lambda)| = 0$

(b) V^* is the unique solution in $\mathbb{M}_1$ to the equation $v = Hv$.

(b) There exists an $\tilde{f} \in \widetilde{\mathbb{F}}$ such that $V^* = H^{\tilde{f}} V^*$, and such a policy is EPD-optimal for the Prob-3.

Remark 1: Theorems 1 and 2, together with Lemma 2, show that Prob-2 is also solvable.

4. The existence of AVaR-optimal policies

We can now solve the original Prob-1. Let $w(t,x,\lambda) := \lambda + \dfrac{1}{1-\gamma} V^*(t,x,\lambda)$, and consider the problem

$$\inf_{\lambda\in\mathbb{R}} w(t,x,\lambda) = \inf_{\lambda\in\mathbb{R}} \left[\lambda + \frac{1}{1-\gamma} V^*(t,x,\lambda)\right]. \tag{2}$$

Theorem 3. Under Assumptions 1–3, there exists a minimum point λ^* (depending on (t,x)) in (2), and the policy $f^*(\cdot,\cdot) := \widetilde{f}^*(\cdot,\cdot,\lambda^*(\cdot,\cdot)) \in \mathbb{F}$ is AVaR-optimal for Prob–1, where $\widetilde{f}^* \in \widetilde{\mathbb{F}}$ is an EPD-optimal policy for Prob–3.

5. Algorithm Aspects

Under Assumptions 1 and 2, the algorithm is stated as follows:

Step 1. Choose $\tilde{f}_0 \in \widetilde{\mathbb{F}}$ arbitrarily, and set $k = 0$;

Step 2. Solve $V^{\tilde{f}_k}$ from the equation $v = H^{\tilde{f}_k} v$;

Step 3. Obtain $\tilde{f}_{k+1}$ such that $H^{\tilde{f}_{k+1}} V^{\tilde{f}_k} = HV^{\tilde{f}_k}$;

Step 4. If $\tilde{f}_{k+1} = \tilde{f}_k$, then $\tilde{f}_{k+1}$ is EPD-optimal, and go to step 5; Else, set $k = k + 1$ and go to step 2;

Step 5. Find a minimum $\lambda^*(t, x)$ of $\lambda + \dfrac{1}{1-\gamma} V^{\tilde{f}_{k+1}}(t, x, \lambda)$, $f_{k+1}(\cdot, \cdot) := \tilde{f}_{k+1}(\cdot, \cdot, \lambda^*(\cdot, \cdot))$ is AVaR optimal, and stop.

Value iteration algorithm:

Step 1. Specify an accuracy $\epsilon > 0$, and set $n = 0$. Let $v_0(t, x, \lambda) := \lambda^-$;

Step 2. Compute $v_{n+1}(t, x, \lambda)$ by $v_{n+1}(t, x, \lambda) = Hv_n(tx, \lambda$

Step 3. If $\|v_{n+1} - v_n\| < \epsilon$, go to Step 4. Otherwise, increment n by 1 and return to Step 2;

Step 4. choose f_ϵ^* such that $H^{f_\epsilon^*} V_{n+1}(t, x, \lambda) = HV_{n+1}(t, x, \lambda)$

Step 5. Find the minimum $\lambda^*(t, x)$ of $\lambda + \dfrac{1}{1-\gamma} v_{n+1}(t, x, \lambda)$, and stop.

In the value iteration algorithm, since $(t, x, \lambda) \in [0, T] \times E \times \mathbb{R}$ and $A(x)$ are all uncountable variables, for practical implementation in computers, we assume the state space E and the action set A are partitioned into n_0 and m_0 parts with suitable scales, respectively. Moreover, we choose suitable step-lengths of the time and level, say, $\delta_1 > 0$ and $\delta_2 > 0$, respectively.

Theorem 4. Under Assumptions 1–3, the value iteration algorithm has complexity:

$$O(m_0 n_0^2 N \rho^{-N} \lfloor T/\delta_1 \rfloor^2 \lfloor \bar{C} T/\delta_2 \rfloor^2 \log(\bar{C} T/\epsilon)),$$

with $N := \lfloor T/\sigma \rfloor + 1$.

Monte Carlo Simulation: As shown in [Boda & Filar, Math. Methods Oper. Res., 63 (2006)] for multi-period loss, Monte Carlo simulation is an elegant algorithm for producing an AVaR optimal control or policy. In the context of finite horizon SMDPs, we can also develop a Monte Carlo simulation algorithm for calculating an AVaR optimal policy.

The details are omitted.

6. Applied examples

- A repaired system with two states, say 1 and 2.
- $A(1) := \{a_{11}, a_{12}\}, A(2) := \{a_{21}, a_{22}\}$
- The system remains at state 1 (under action a_{1j}) for a random period of time uniformly-distributed in the region $[0, \mu(1, a_{1j})]$, and then transitions to state 2 with probability $p(2|1, a_{1j})$; The system remains at state 2 (under action a_{2j}) for a random period of time exponential-distributed with parameter $\mu(2, a_{2j}) > 0$; and then transitions to state 1 with probability $p(1|2, a_{2j})$.

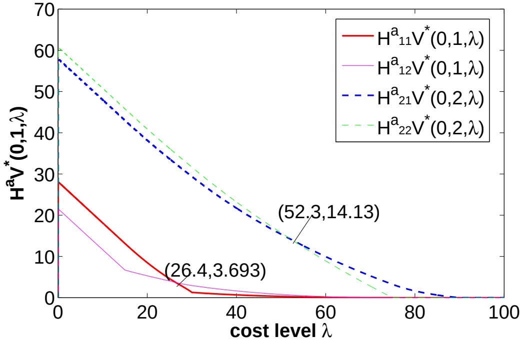
To conduct the computation, we use the following data:

<table>
<tr><th rowspan="2">State
x</th><th rowspan="2">Action
a</th><th rowspan="2">Parameter for sojourn time
$\mu(x,a)$</th><th colspan="2">Transition probability
$p(y \mid x,a)$</th><th rowspan="2">Cost rate
$c(x,a)$</th><th rowspan="2">Horizon
T</th><th rowspan="2">Confidence level γ</th></tr>
<tr><th>$y=1$</th><th>$y=2$</th></tr>
<tr><td rowspan="2">1</td><td>a_{11}</td><td>25</td><td>0.9</td><td>0.1</td><td>2</td><td rowspan="4">15</td><td rowspan="4">0.95</td></tr>
<tr><td>a_{12}</td><td>20</td><td>0.7</td><td>0.3</td><td>1</td></tr>
<tr><td rowspan="2">2</td><td>a_{21}</td><td>0.15</td><td>0.6</td><td>0.4</td><td>6</td></tr>
<tr><td>a_{22}</td><td>0.10</td><td>0.4</td><td>0.6</td><td>5</td></tr>
</table>

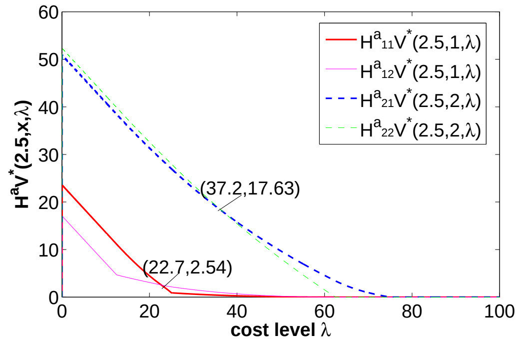
Table 4.1. The data of the model

Under the data, Assumptions 1–3 obviously hold. Therefore, the VI algorithm is valid, and an AVaR-optimal policy exists.

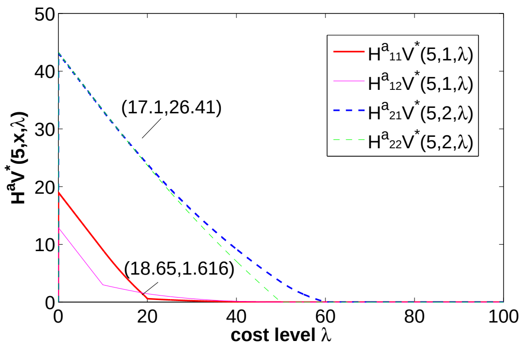
Set $\epsilon = 10^{-12}$, and discretize the time interval $[0, 15]$ and the cost level interval $[0, 100]$ with $\delta_1 = \delta_2 = 0.05$. Then, we implement Steps 1-3 of the VI algorithm in MATLAB software, and obtain data on the functions V^* and $H^a V^*$ (see Fig. 4.1). To execute Step 4 of the VI algorithm, we shall compare the data $H^a V^*(t, x, \lambda)$ under admissible actions a for every $(t, x, \lambda) \in [0, 15] \times \{1, 2\} \times \mathbb{R}$. To be specific, we analyze the data of $H^a V^*(0, x, \lambda)$, $H^a V^*(2.5, x, \lambda)$, $H^a V^*(5, x, \lambda)$, and $H^a V^*(10, x, \lambda)$ as examples, which are shown in Fig. 4.1 below.



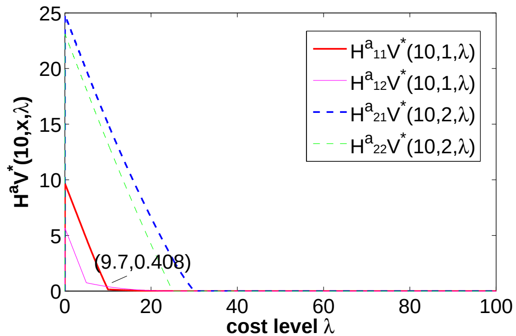
$(a)\ H^aV^*(0, x, \lambda)$



$(b)\ H^aV^*(2.5, x, \lambda)$



$(c)\ H^aV^*(5, x, \lambda)$



$(d)\ H^aV^*(10, x, \lambda)$

Fig. 4.1 The function $H^aV^*(t,x,\lambda)$

Comparing the data $H^aV^*(t,x,\lambda)$ under admissible actions a for every $(t,x,\lambda) \in [0,15] \times \{1,2\} \times \mathbb{R}$, one may obtain an optimal policy $\tilde{f}^*$ for the Prob-3. For example, in the light of Fig. 4.1, we can define $\tilde{f}^*$ by

$$\tilde{f}^*(0,1,\lambda) = \begin{cases} a_{12}, & \lambda < 26.4 \\ a_{11}, & \lambda \geq 26.4 \end{cases} \qquad \tilde{f}^*(0,2,\lambda) = \begin{cases} a_{21}, & \lambda < 52.3 \\ a_{22}, & \lambda \geq 52.3 \end{cases}$$

$$\tilde{f}^*(2.5,1,\lambda) = \begin{cases} a_{12}, & \lambda < 22.7 \\ a_{11}, & \lambda \geq 22.7 \end{cases} \quad \tilde{f}^*(2.5,2,\lambda) = \begin{cases} a_{21}, & \lambda < 37.2 \\ a_{22}, & \lambda \geq 37.2 \end{cases}$$

$$\tilde{f}^*(2.5, 1, \lambda) = \begin{cases} a_{12}, & \lambda < 18.65 \\ a_{11}, & \lambda \geq 18.65 \end{cases} \quad \tilde{f}^*(2.5, 2, \lambda) = \begin{cases} a_{21}, & \lambda < 17.1 \\ a_{22}, & \lambda \geq 17.1 \end{cases}$$

and

$$\tilde{f}^*(10, 1, \lambda) = \begin{cases} a_{12}, & \lambda < 9.7 \\ a_{11}, & \lambda \geq 9.7 \end{cases} \qquad \tilde{f}^*(10, 2, \lambda) = a_{22}, 0 \leq \lambda \leq 90.$$

Now, to obtain an AVaR optimal policies, we seek the minimum-point $\lambda^*(t, x)$ of the function $\lambda \mapsto w(t, x, \lambda)$ with $\gamma = 0.95$. Fig 4.2 below gives the graphs of $w(t, x, \lambda)$ with $t = 0, 2.5, 5, 10$.

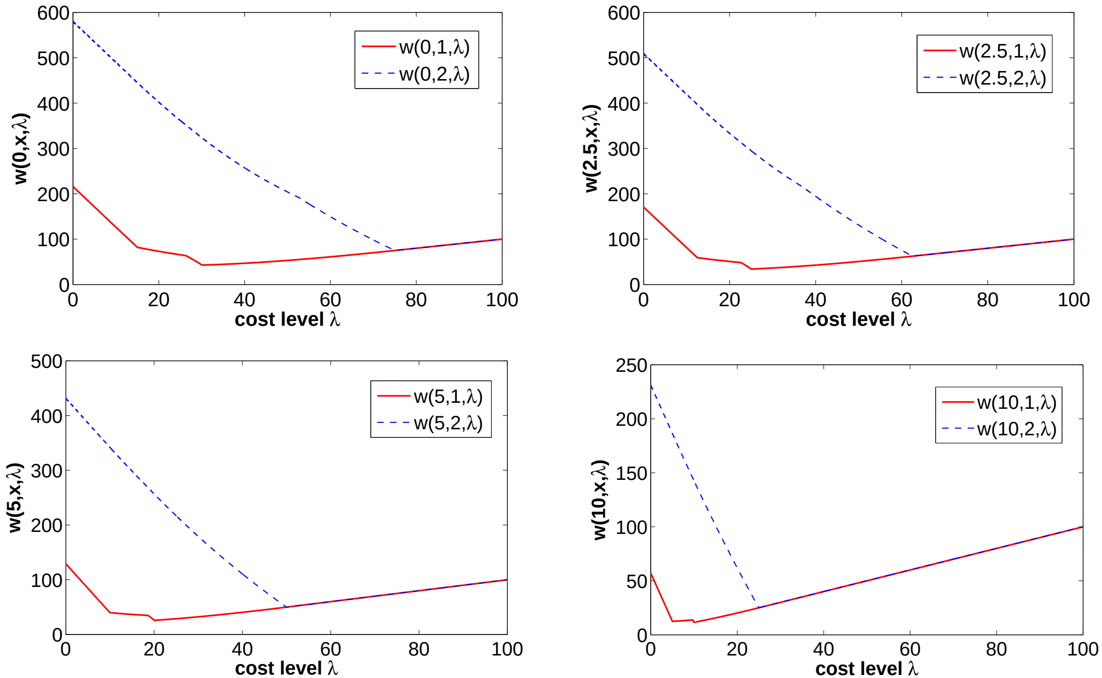


Fig. 4.2 The function $w(t, x, \lambda)$

From Fig. 4.2 above, it is easy to see the minimum-points

$\lambda^*(t,x)$ with $t = 0, 2.5, 5, 10$ and $x = 1, 2$, i.e.,

$$\lambda^*(0,1) = 30, \lambda^*(0,2) = 75, \lambda^*(2.5,1) = 25, \lambda^*(2.5,2) = 62.5,$$
$$\lambda^*(5,1) = 20, \lambda^*(5,2) = 50, \lambda^*(10,1) = 10, \lambda^*(10,2) = 25.$$

For other $t \in [0, 15]$, the minimum-points $\lambda^*(t,x)$ can be similarly calculated.

By Theorem 3, the policy $f^*(t,x) := \tilde{f}^*(t,x,\lambda^*(t,x))$ is AVaR-optimal. For example,

$$f^*(0,1) = a_{11}, f^*(0,2) = a_{22}, f^*(2.5,1) = a_{11}, f^*(2.5,2) = a_{22},$$
$$f^*(5,1) = a_{11}, f^*(5,2) = a_{22}, f^*(10,1) = a_{11}, f^*(10,2) = a_{22}.$$

Thank you very much !!!