Grundlagen adaptiver Wissenssysteme

Übungen zur Vorlesung

Prof. Dr. Thomas Gabel
Frankfurt University of Applied Sciences
Faculty of Computer Science and Engineering
tgabel@fb2.fra-uas.de

Aufgabenblatt 2

1. **Aufgabenblatt 2 – Übung 5**
2. Aufgabenblatt 2 – Übung 6

Aufgabe 5: Absorbierende Zustände

(a) Was ist ein absorbierender Zustand?

FRANKFURT UNIVERSITY OF APPLIED SCIENCES

Aufgabe 5: Absorbierende Zustände

(a) Was ist ein absorbierender Zustand?

- Zustand, der nicht mehr verlassen werden kann
- Terminalzustand
- für solch ein $i \in S$ gilt: $p_{sj}(a) = 0, \forall a, \forall j \neq i$
- Bezeichnung oftmals mittels '0'

Aufgabe 5: Absorbierende Zustände

(b) Sie wollen ein Pfadplanungsproblem mittels eines MDP modellieren. Der Zielzustand muss hierbei ein absorbierender Zustand sein. Welche Form direkter Kosten sollten hierbei dem Zielzustand und den Nicht-Zielzuständen zugeordnet sein, wenn das Ziel darin besteht, die Pfadlänge zu minimieren.

Aufgabe 5: Absorbierende Zustände

(b) Sie wollen ein Pfadplanungsproblem mittels eines MDP modellieren. Der Zielzustand muss hierbei ein absorbierender Zustand sein. Welche Form direkter Kosten sollten hierbei dem Zielzustand und den Nicht-Zielzuständen zugeordnet sein, wenn das Ziel darin besteht, die Pfadlänge zu minimieren.

- Zielzustand kann nicht mehr verlassen werden $\rightarrow$ absorbierender Zustand
- Variante 1: SKP-Problem
  - keine Diskontierung
  - Nullkosten im (absorbierenden) Terminalzustand
  - konstante Kosten in allen anderen Zuständen

Aufgabe 5: Absorbierende Zustände

(b) Sie wollen ein Pfadplanungsproblem mittels eines MDP modellieren. Der Zielzustand muss hierbei ein absorbierender Zustand sein. Welche Form direkter Kosten sollten hierbei dem Zielzustand und den Nicht-Zielzuständen zugeordnet sein, wenn das Ziel darin besteht, die Pfadlänge zu minimieren.

- Zielzustand kann nicht mehr verlassen werden $\rightarrow$ absorbierender Zustand
- Variante 1: SKP-Problem
  - keine Diskontierung
  - Nullkosten im (absorbierenden) Terminalzustand
  - konstante Kosten in allen anderen Zuständen
- Variante 2: diskontiertes Problem

Aufgabe 5: Absorbierende Zustände

(b) Sie wollen ein Pfadplanungsproblem mittels eines MDP modellieren. Der Zielzustand muss hierbei ein absorbierender Zustand sein. Welche Form direkter Kosten sollten hierbei dem Zielzustand und den Nicht-Zielzuständen zugeordnet sein, wenn das Ziel darin besteht, die Pfadlänge zu minimieren.

- Variante 2: diskontiertes Problem

Aufgabe 5: Absorbierende Zustände

(b) Sie wollen ein Pfadplanungsproblem mittels eines MDP modellieren. Der Zielzustand muss hierbei ein absorbierender Zustand sein. Welche Form direkter Kosten sollten hierbei dem Zielzustand und den Nicht-Zielzuständen zugeordnet sein, wenn das Ziel darin besteht, die Pfadlänge zu minimieren.

- Variante 2: diskontiertes Problem
  - Kosten ungleich null im (absorbierenden) Terminalzustand, z.B. negative Kosten ("positive Belohnungen")
  - Diskontierung erforderlich (damit hat eine Strategie, die den Terminalzustand erreicht endliche Kosten)

Aufgabe 5: Absorbierende Zustände

(b) Sie wollen ein Pfadplanungsproblem mittels eines MDP modellieren. Der Zielzustand muss hierbei ein absorbierender Zustand sein. Welche Form direkter Kosten sollten hierbei dem Zielzustand und den Nicht-Zielzuständen zugeordnet sein, wenn das Ziel darin besteht, die Pfadlänge zu minimieren.

- Variante 2: diskontiertes Problem
  - Kosten ungleich null im (absorbierenden) Terminalzustand, z.B. negative Kosten ("positive Belohnungen")
  - Diskontierung erforderlich (damit hat eine Strategie, die den Terminalzustand erreicht endliche Kosten)
  - konstante Kosten in allen anderen Zuständen, die höher als Terminalkosten sind

    Frage: Was passiert, wenn die Schrittkosten niedriger als die Terminalkosten gewählt werden?

Aufgabe 5: Absorbierende Zustände

(b) Sie wollen ein Pfadplanungsproblem mittels eines MDP modellieren. Der Zielzustand muss hierbei ein absorbierender Zustand sein. Welche Form direkter Kosten sollten hierbei dem Zielzustand und den Nicht-Zielzuständen zugeordnet sein, wenn das Ziel darin besteht, die Pfadlänge zu minimieren.

- Variante 2: diskontiertes Problem
  - Kosten ungleich null im (absorbierenden) Terminalzustand, z.B. negative Kosten ("positive Belohnungen")
  - Diskontierung erforderlich (damit hat eine Strategie, die den Terminalzustand erreicht endliche Kosten)
  - konstante Kosten in allen anderen Zuständen, die höher als Terminalkosten sind

    Frage: Was passiert, wenn die Schrittkosten niedriger als die Terminalkosten gewählt werden?

    $\rightarrow$ Tafel (nach Teilaufgabe (c))

Aufgabe 5: Absorbierende Zustände

(c) Betrachten Sie nun ein Diskontierungsproblem mit Diskontierungsfaktor $\gamma < 1$ das einen absorbierenden Zustand s enthält, in dem direkte Kosten c entstehen. Was sind die Pfadkosten $V(s)$ jenes Zustandes?

Aufgabe 5: Absorbierende Zustände

(c) Betrachten Sie nun ein Diskontierungsproblem mit Diskontierungsfaktor $\gamma < 1$ das einen absorbierenden Zustand s enthält, in dem direkte Kosten c entstehen. Was sind die Pfadkosten $V(s)$ jenes Zustandes?

- Es gilt $p_{sj} = 0 \forall j \neq s$ und $p_{ss} = 1$.

Aufgabe 5: Absorbierende Zustände

(c) Betrachten Sie nun ein Diskontierungsproblem mit Diskontierungsfaktor $\gamma < 1$ das einen absorbierenden Zustand s enthält, in dem direkte Kosten c entstehen. Was sind die Pfadkosten $V(s)$ jenes Zustandes?

- Es gilt $p_{sj} = 0 \forall j \neq s$ und $p_{ss} = 1$.
- Damit ergibt sich $V(s) = \min_{a \in A(s)} \sum_{j=1}^{n} p_{sj}(a)(c(s,a) + \gamma V(j)) = c + \gamma V(s)$

Aufgabe 5: Absorbierende Zustände

(c) Betrachten Sie nun ein Diskontierungsproblem mit Diskontierungsfaktor $\gamma < 1$ das einen absorbierenden Zustand s enthält, in dem direkte Kosten c entstehen. Was sind die Pfadkosten $V(s)$ jenes Zustandes?

- Es gilt $p_{sj} = 0 \forall j \neq s$ und $p_{ss} = 1$.
- Damit ergibt sich $V(s) = \min_{a \in A(s)} \sum_{j=1}^{n} p_{sj}(a)(c(s,a) + \gamma V(j)) = c + \gamma V(s)$
- Also: $(1-\gamma)V(s) = c \Rightarrow V(s) = \frac{c}{1-\gamma}$

Aufgabenblatt 2

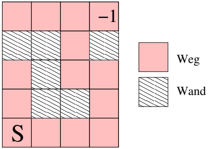
1. Aufgabenblatt 2 – Übung 5
2. **Aufgabenblatt 2 – Übung 6**

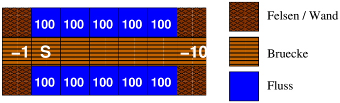
Aufgabe 6: Labyrinth und Brücke

In der folgenden Abbildung sind zwei MDPs dargestellt. Der Agent startet stets im Startzustand S. Im Brücken-MDP sind Aktionen, die den Agenten in das Wasser (N, S) bewegen, deterministisch. Alle anderen Aktionen (Bewegung nach W oder O) gelingen mit einer Wahrscheinlichkeit von $1 - \varrho$. Mit einer Wahrscheinlichkeit von $\frac{\varrho}{3}$ (ϱ wird im Folgenden als Rauschintensität bezeichnet) hingegen verharrt der Agent auf der Stelle und mit ebenfalls jeweils $\frac{\varrho}{3}$ bewegt er sich nach N oder S (und stürzt damit von der Brücke).
Im Labyrinth-MDP sind alle Zustandsübergänge deterministisch.

Aufgabe 6: Labyrinth und Brücke

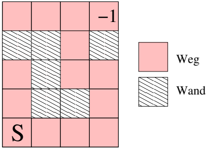
In der folgenden Abbildung sind zwei MDPs dargestellt. Der Agent startet stets im Startzustand S. Im Brücken-MDP sind Aktionen, die den Agenten in das Wasser (N, S) bewegen, deterministisch. Alle anderen Aktionen (Bewegung nach W oder O) gelingen mit einer Wahrscheinlichkeit von $1 - \varrho$. Mit einer Wahrscheinlichkeit von $\frac{\varrho}{3}$ (ϱ wird im Folgenden als Rauschintensität bezeichnet) hingegen verharrt der Agent auf der Stelle und mit ebenfalls jeweils $\frac{\varrho}{3}$ bewegt er sich nach N oder S (und stürzt damit von der Brücke).
Im Labyrinth-MDP sind alle Zustandsübergänge deterministisch.

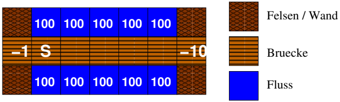




Aufgabe 6: Labyrinth und Brücke

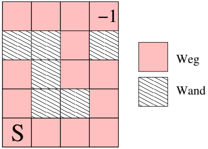
Felder, in denen eine Zahl steht, sind Terminalzustände. Für Übergange in Terminalzustände sind stets die zugehörigen direkten Kosten angegeben (z.B. Kosten von 100 dafür, von der Brücke zu fallen; Kosten von -10 dafür, die Brücke überquert zu haben, Kosten von -1 dafür, das Labyrinth zu verlassen). Für Aktionen, die den Agenten in eine Wand bewegen würden, werden Kosten von 1 vergeben (Agent bewegt sich nicht); alle anderen Transitionen sind kostenfrei.

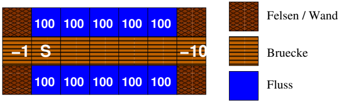




Aufgabe 6: Labyrinth und Brücke

Felder, in denen eine Zahl steht, sind Terminalzustände. Für Übergange in Terminalzustände sind stets die zugehörigen direkten Kosten angegeben (z.B. Kosten von 100 dafür, von der Brücke zu fallen; Kosten von -10 dafür, die Brücke überquert zu haben, Kosten von -1 dafür, das Labyrinth zu verlassen). Für Aktionen, die den Agenten in eine Wand bewegen würden, werden Kosten von 1 vergeben (Agent bewegt sich nicht); alle anderen Transitionen sind kostenfrei.





(a) Wie viele Iterationen des Wertiterationsverfahrens sind für den Labyrinth-MDP notwendig, bis dieser in seiner Wertfunktion (Pfadkostenvektor) für den Zustand S (Start) erstmalig einen Wert ungleich null einträgt? Warum?

Aufgabe 6: Labyrinth und Brücke

(a) Wie viele Iterationen des Wertiterationsverfahrens sind für den Labyrinth-MDP notwendig, bis dieser in seiner Wertfunktion (Pfadkostenvektor) für den Zustand S (Start) erstmalig einen Wert ungleich null einträgt? Warum?

Aufgabe 6: Labyrinth und Brücke

(a) Wie viele Iterationen des Wertiterationsverfahrens sind für den Labyrinth-MDP notwendig, bis dieser in seiner Wertfunktion (Pfadkostenvektor) für den Zustand S (Start) erstmalig einen Wert ungleich null einträgt? Warum?

- Im Wertiterationsverfahren wird das Bellman-Update wiederholt auf die Kostenfunktion V_k angewandt, um V_{k+1} zu ermitteln.
- Wir gehen hier davon aus, dass V_0 mit null initialisiert worden ist, d.h. $V_0(i) = 0$ für $i = 1, \ldots, n$.

Aufgabe 6: Labyrinth und Brücke

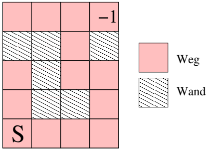
(a) Wie viele Iterationen des Wertiterationsverfahrens sind für den Labyrinth-MDP notwendig, bis dieser in seiner Wertfunktion (Pfadkostenvektor) für den Zustand S (Start) erstmalig einen Wert ungleich null einträgt? Warum?

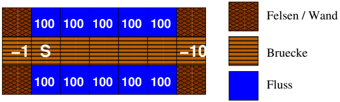
- Im Wertiterationsverfahren wird das Bellman-Update wiederholt auf die Kostenfunktion V_k angewandt, um V_{k+1} zu ermitteln.
- Wir gehen hier davon aus, dass V_0 mit null initialisiert worden ist, d.h. $V_0(i) = 0$ für $i = 1, \dots, n$.
- Der Aktualisierungsschritt ist definiert als:

$$V_{k+1}(i) := \min_{a \in A(i)} \sum_{j=0}^{n} p_{ij}(a) \left(c(i, a, j) + \gamma V_k(j) \right)$$

- Die Aktualisierung liefert die verbesserte Wertfunktion für das einstufige (diskontierte) Entscheidungsproblem mit Übergangskosten $c(\cdot)$ und Terminalkosten V_k.

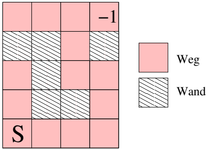
Aufgabe 6: Labyrinth und Brücke

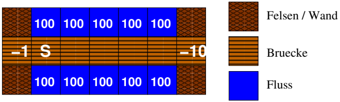




- Welche Zustände werden bei Null-Initialisierung von V und einmaliger Anwendung der Aktualisierung Einträge ungleich null erhalten?

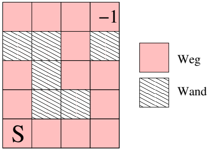
Aufgabe 6: Labyrinth und Brücke

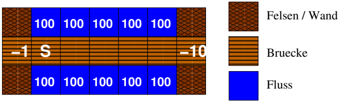




- Welche Zustände werden bei Null-Initialisierung von V und einmaliger Anwendung der Aktualisierung Einträge ungleich null erhalten?
- Wände sind Zustände, die nach Aufgabenstellung nicht betreten werden können.

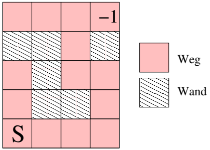
Aufgabe 6: Labyrinth und Brücke

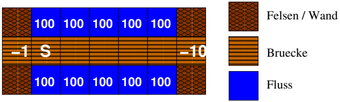




- Welche Zustände werden bei Null-Initialisierung von V und einmaliger Anwendung der Aktualisierung Einträge ungleich null erhalten?
- Wände sind Zustände, die nach Aufgabenstellung nicht betreten werden können.
- Da die direkten Kosten null sind, werden lediglich die wandfreien direkten Nachbarzustände vom Zielzustand (oben rechts) einen Wert ungleich null annehmen, nämlich für $i = s_{1,3}$ (beste Aktion ist O):

Aufgabe 6: Labyrinth und Brücke

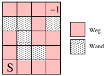


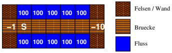


- Welche Zustände werden bei Null-Initialisierung von V und einmaliger Anwendung der Aktualisierung Einträge ungleich null erhalten?
- Wände sind Zustände, die nach Aufgabenstellung nicht betreten werden können.
- Da die direkten Kosten null sind, werden lediglich die wandfreien direkten Nachbarzustände vom Zielzustand (oben rechts) einen Wert ungleich null annehmen, nämlich für $i = s_{1,3}$ (beste Aktion ist O):

  $-1 + \gamma \cdot 0 = -1$

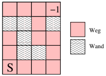
Aufgabe 6: Labyrinth und Brücke

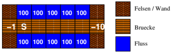




- In allen anderen Zuständen bewirkt die Minimumbildung der Aktualisierungsvorschrift, dass die positiven Kosten, die entstehen, wenn der Agent in eine Wand zu laufen versucht, nicht in V einfließen.
- Beispiel: $i = s_{5,2}$

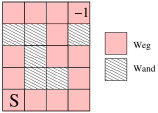
Aufgabe 6: Labyrinth und Brücke

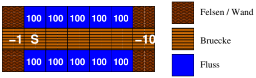




- In allen anderen Zuständen bewirkt die Minimumbildung der Aktualisierungsvorschrift, dass die positiven Kosten, die entstehen, wenn der Agent in eine Wand zu laufen versucht, nicht in V einfließen.
- Beispiel: $i = s_{5,2}$
  - Aktionen sind $a \in \{N, S, O, W\}$
  - für $a = O$ und $a = W$: $c(i, a) = 0$, Folgezustand ist $j = s_{5,3}$ bzw. $j = s_{5,1}$ jeweils mit $V(j) = 0$
  - für $a = N$ und $a = S$: $c(i, a) = 1$, Folgezustand ist $j = s_{5,2}$ bzw. $j = s_{5,2}$ jeweils mit $V(j) = 0$

Aufgabe 6: Labyrinth und Brücke



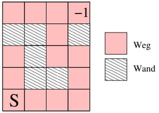


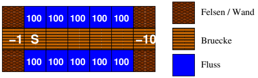
- Mit

$$V_{k+1}(i) := \min_{a \in A(i)} \sum_{j=0}^{n} p_{ij}(a)\,(c(i,a) + \gamma\, V_k(j))$$

  ergibt sich daher $V_1(s_{5,2}) = min\{0, 0, 1, 1\} = 0$
- Nach analogem Schema setzen sich die Aktualisierungen der Pfadkosten in den folgenden Schritten des Wertiterationsverfahrens fort.

Aufgabe 6: Labyrinth und Brücke



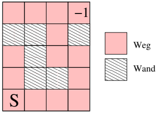


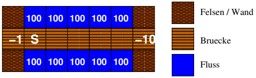
- Mit

$$V_{k+1}(i) := \min_{a \in A(i)} \sum_{j=0}^{n} p_{ij}(a)\,(c(i,a) + \gamma V_k(j))$$

  ergibt sich daher $V_1(s_{5,2}) = min\{0, 0, 1, 1\} = 0$
- Nach analogem Schema setzen sich die Aktualisierungen der Pfadkosten in den folgenden Schritten des Wertiterationsverfahrens fort.
- Nach dem zweiten Update haben auch $s_{1,2}$ und $s_{2,3}$ Einträge ungleich null.

Aufgabe 6: Labyrinth und Brücke





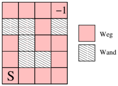
- Mit

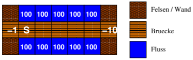
$$V_{k+1}(i) := \min_{a \in A(i)} \sum_{j=0}^{n} p_{ij}(a)\,(c(i,a) + \gamma\, V_k(j))$$

  ergibt sich daher $V_1(s_{5,2}) = min\{0, 0, 1, 1\} = 0$
- Nach analogem Schema setzen sich die Aktualisierungen der Pfadkosten in den folgenden Schritten des Wertiterationsverfahrens fort.
- Nach dem zweiten Update haben auch $s_{1,2}$ und $s_{2,3}$ Einträge ungleich null.
- Nach 9 Iterationen erhält schließlich auch der Zustand S erstmalig einen Eintrag ungleich null.

Aufgabe 6: Labyrinth und Brücke

Betrachten Sie nun den Brücken-MDP, in dem ein Diskontierungsfaktor von $\gamma = 0.9$ sowie eine Rauschintensität von $\varrho = 0.3$ zur Anwendung kommen.

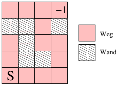


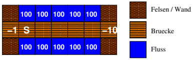


(b) Ermitteln Sie die optimalen Pfadkosten im Zustand S sowie in allen Zuständen östlich und westlich von S.

Aufgabe 6: Labyrinth und Brücke

Betrachten Sie nun den Brücken-MDP, in dem ein Diskontierungsfaktor von $\gamma = 0.9$ sowie eine Rauschintensität von $\varrho = 0.3$ zur Anwendung kommen.



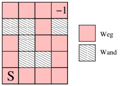


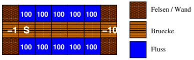
(b) Ermitteln Sie die optimalen Pfadkosten im Zustand S sowie in allen Zuständen östlich und westlich von S.

- Wir bezeichnen die Zustände auf der Brücke im Folgenden von Westen nach Osten als $s_0, \ldots, s_6$, so dass der Startzustand S dem Zustand s_1 entspricht.

Aufgabe 6: Labyrinth und Brücke

Betrachten Sie nun den Brücken-MDP, in dem ein Diskontierungsfaktor von $\gamma = 0.9$ sowie eine Rauschintensität von $\varrho = 0.3$ zur Anwendung kommen.



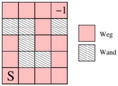


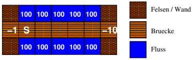
(b) Ermitteln Sie die optimalen Pfadkosten im Zustand S sowie in allen Zuständen östlich und westlich von S.

- Wir bezeichnen die Zustände auf der Brücke im Folgenden von Westen nach Osten als $s_0, \ldots, s_6$, so dass der Startzustand S dem Zustand s_1 entspricht.
- Für s_0 als absorbierenden Terminalzustand, in dem keine Aktionen Kosten verursachen gilt per Definition:

Aufgabe 6: Labyrinth und Brücke

Betrachten Sie nun den Brücken-MDP, in dem ein Diskontierungsfaktor von $\gamma = 0.9$ sowie eine Rauschintensität von $\varrho = 0.3$ zur Anwendung kommen.





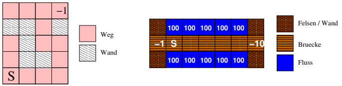
(b) Ermitteln Sie die optimalen Pfadkosten im Zustand S sowie in allen Zuständen östlich und westlich von S.

- Wir bezeichnen die Zustände auf der Brücke im Folgenden von Westen nach Osten als $s_0, \dots, s_6$, so dass der Startzustand S dem Zustand s_1 entspricht.
- Für s_0 als absorbierenden Terminalzustand, in dem keine Aktionen Kosten verursachen gilt per Definition: $V^\star(s_0) = 0$.

FRANKFURT UNIVERSITY OF APPLIED SCIENCES

Aufgabe 6: Labyrinth und Brücke

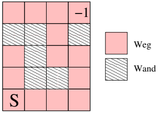
Betrachten Sie nun den Brücken-MDP, in dem ein Diskontierungsfaktor von $\gamma = 0.9$ sowie eine Rauschintensität von $\varrho = 0.3$ zur Anwendung kommen.

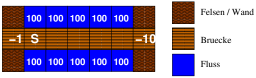


(b) Ermitteln Sie die optimalen Pfadkosten im Zustand S sowie in allen Zuständen östlich und westlich von S.

- Wir bezeichnen die Zustände auf der Brücke im Folgenden von Westen nach Osten als $s_0, \ldots, s_6$, so dass der Startzustand S dem Zustand s_1 entspricht.
- Für s_0 als absorbierenden Terminalzustand, in dem keine Aktionen Kosten verursachen gilt per Definition: $V^\star(s_0) = 0$.
- Gleiches gilt für den Zustand s_6, d.h. $V^\star(s_6) = 0$, sowie für alle Zustände im Fluss.

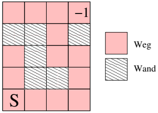
Aufgabe 6: Labyrinth und Brücke

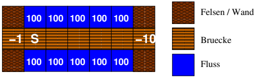




- Im Folgenden betrachten wir zwei Strategien, π_W und π_O, die den Agenten nach links bzw. rechts bewegen wollen, d.h. $\pi_W(i) = W$ und $\pi_O(i) = O$ für alle i.

Aufgabe 6: Labyrinth und Brücke

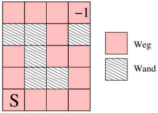


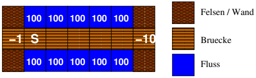


- Im Folgenden betrachten wir zwei Strategien, π_W und π_O, die den Agenten nach links bzw. rechts bewegen wollen, d.h. $\pi_W(i) = W$ und $\pi_O(i) = O$ für alle i.
- Für den Startzustand $S = s_1$ und Strategie π_W gilt daher:

$$V^{\pi_W}(s_1) \quad = \quad (1 - \varrho) \cdot$$

Aufgabe 6: Labyrinth und Brücke

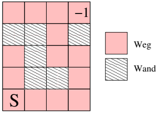


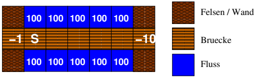


- Im Folgenden betrachten wir zwei Strategien, π_W und π_O, die den Agenten nach links bzw. rechts bewegen wollen, d.h. $\pi_W(i) = W$ und $\pi_O(i) = O$ für alle i.
- Für den Startzustand $S = s_1$ und Strategie π_W gilt daher:

$$V^{\pi_W}(s_1) \quad = \quad (1 - \varrho) \cdot (-1 + \gamma \cdot 0) +$$

Aufgabe 6: Labyrinth und Brücke

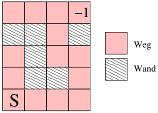


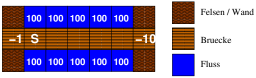


- Im Folgenden betrachten wir zwei Strategien, π_W und π_O, die den Agenten nach links bzw. rechts bewegen wollen, d.h. $\pi_W(i) = W$ und $\pi_O(i) = O$ für alle i.
- Für den Startzustand $S = s_1$ und Strategie π_W gilt daher:

$$V^{\pi_W}(s_1) \quad = \quad (1 - \varrho) \cdot (-1 + \gamma \cdot 0) + \frac{\varrho}{3}(0 + \gamma V^{\pi_W}(s_1))$$

Aufgabe 6: Labyrinth und Brücke

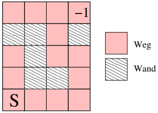


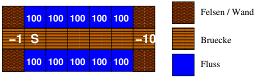


- Im Folgenden betrachten wir zwei Strategien, π_W und π_O, die den Agenten nach links bzw. rechts bewegen wollen, d.h. $\pi_W(i) = W$ und $\pi_O(i) = O$ für alle i.
- Für den Startzustand $S = s_1$ und Strategie π_W gilt daher:

$$\begin{aligned} V^{\pi_W}(s_1) &= (1-\varrho)\cdot(-1+\gamma\cdot 0) + \frac{\varrho}{3}(0+\gamma V^{\pi_W}(s_1)) \\ &\quad +2\frac{\varrho}{3}(100+\gamma\cdot 0) \\ &= \end{aligned}$$

Aufgabe 6: Labyrinth und Brücke





- Im Folgenden betrachten wir zwei Strategien, π_W und π_O, die den Agenten nach links bzw. rechts bewegen wollen, d.h. $\pi_W(i) = W$ und $\pi_O(i) = O$ für alle i.
- Für den Startzustand $S = s_1$ und Strategie π_W gilt daher:

$$
\begin{aligned}
V^{\pi_W}(s_1) &= (1-\varrho)\cdot(-1+\gamma\cdot 0) + \frac{\varrho}{3}(0+\gamma V^{\pi_W}(s_1)) \\
&\quad +2\frac{\varrho}{3}(100+\gamma\cdot 0) \\
&= \varrho - 1 + \frac{\varrho}{3}\gamma V^{\pi_W}(s_1) - 200\frac{\varrho}{3}
\end{aligned}
$$

Aufgabe 6: Labyrinth und Brücke

- Für den Startzustand $S = s_1$ und Strategie π_W gilt daher:

Aufgabe 6: Labyrinth und Brücke

- Für den Startzustand $S = s_1$ und Strategie π_W gilt daher:

$$
\begin{aligned}
V^{\pi_W}(s_1) &= (1-\varrho)\cdot(-1+\gamma\cdot 0) + \frac{\varrho}{3}(0+\gamma V^{\pi_W}(s_1)) \\
&\quad +2\frac{\varrho}{3}(100+\gamma\cdot 0) \\
&= \varrho - 1 + \frac{\varrho}{3}\gamma V^{\pi_W}(s_1) + 200\frac{\varrho}{3} \\
V^{\pi_W}(s_1)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{203\varrho}{3} - 1
\end{aligned}
$$

Aufgabe 6: Labyrinth und Brücke

- Für den Startzustand $S = s_1$ und Strategie π_W gilt daher:

$$\begin{aligned}
V^{\pi_W}(s_1) &= (1-\varrho)\cdot(-1+\gamma\cdot 0) + \frac{\varrho}{3}(0+\gamma V^{\pi_W}(s_1)) \\
&\quad +2\frac{\varrho}{3}(100+\gamma\cdot 0) \\
&= \varrho - 1 + \frac{\varrho}{3}\gamma V^{\pi_W}(s_1) + 200\frac{\varrho}{3} \\
V^{\pi_W}(s_1)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{203\varrho}{3} - 1
\end{aligned}$$

- Damit:

$$V^{\pi_W}(s_1) = \frac{\frac{203\varrho}{3} - 1}{1 - \frac{\varrho}{3}\gamma} = \frac{203\varrho - 3}{3 - \varrho\gamma}$$

Aufgabe 6: Labyrinth und Brücke

- Für den Startzustand $S = s_1$ und Strategie π_W gilt daher:

$$\begin{aligned} V^{\pi_W}(s_1) &= (1-\varrho)\cdot(-1+\gamma\cdot 0) + \frac{\varrho}{3}(0+\gamma V^{\pi_W}(s_1)) \\ &\quad +2\frac{\varrho}{3}(100+\gamma\cdot 0) \\ &= \varrho - 1 + \frac{\varrho}{3}\gamma V^{\pi_W}(s_1) + 200\frac{\varrho}{3} \\ V^{\pi_W}(s_1)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{203\varrho}{3} - 1 \end{aligned}$$

- Damit:

$$V^{\pi_W}(s_1) = \frac{\frac{203\varrho}{3}-1}{1-\frac{\varrho}{3}\gamma} = \frac{203\varrho - 3}{3-\varrho\gamma}$$

- Konkret für $\gamma = 0.9$ und $\varrho = 0.3$: $V^{\pi_W}(s_1) = 21.21$

Aufgabe 6: Labyrinth und Brücke

- In Analogie erhalten wir für den Zustand s_5 und Strategie π_O:
  - $\rightarrow$ Tafel / Papier, Rechnung selber durchführen

Aufgabe 6: Labyrinth und Brücke

- In Analogie erhalten wir für den Zustand s_5 und Strategie π_O:
  - $\rightarrow$ Tafel / Papier, Rechnung selber durchführen

$$
\begin{aligned}
V^{\pi_O}(s_5) &= (1-\varrho)\cdot(-10+\gamma\cdot 0)+\frac{\varrho}{3}(0+\gamma V^{\pi_O}(s_5)) \\
&\quad +2\frac{\varrho}{3}(100+\gamma\cdot 0) \\
&= 10\varrho-10+\frac{\varrho}{3}\gamma V^{\pi_O}(s_5)+200\frac{\varrho}{3}
\end{aligned}
$$

Aufgabe 6: Labyrinth und Brücke

- In Analogie erhalten wir für den Zustand s_5 und Strategie π_O:
  - $\rightarrow$ Tafel / Papier, Rechnung selber durchführen

$$
\begin{aligned}
V^{\pi_O}(s_5) &= (1-\varrho)\cdot(-10+\gamma\cdot 0)+\frac{\varrho}{3}(0+\gamma V^{\pi_O}(s_5)) \\
&\quad +2\frac{\varrho}{3}(100+\gamma\cdot 0) \\
&= 10\varrho - 10 + \frac{\varrho}{3}\gamma V^{\pi_O}(s_5) + 200\frac{\varrho}{3} \\
V^{\pi_O}(s_5)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{230\varrho}{3} - 10
\end{aligned}
$$

Aufgabe 6: Labyrinth und Brücke

- In Analogie erhalten wir für den Zustand s_5 und Strategie π_O:
  - $\rightarrow$ Tafel / Papier, Rechnung selber durchführen

$$\begin{aligned} V^{\pi_O}(s_5) &= (1-\varrho)\cdot(-10+\gamma\cdot 0)+\frac{\varrho}{3}(0+\gamma V^{\pi_O}(s_5)) \\ &\quad +2\frac{\varrho}{3}(100+\gamma\cdot 0) \\ &= 10\varrho-10+\frac{\varrho}{3}\gamma V^{\pi_O}(s_5)+200\frac{\varrho}{3} \\ V^{\pi_O}(s_5)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{230\varrho}{3}-10 \end{aligned}$$

- Damit:

$$V^{\pi_O}(s_5)=\frac{\frac{230\varrho}{3}-10}{1-\frac{\varrho}{3}\gamma}=\frac{230\varrho-30}{3-\varrho\gamma}$$

Aufgabe 6: Labyrinth und Brücke

- In Analogie erhalten wir für den Zustand s_5 und Strategie π_O:
  - $\rightarrow$ Tafel / Papier, Rechnung selber durchführen

$$
\begin{aligned}
V^{\pi_O}(s_5) &= (1-\varrho)\cdot(-10+\gamma\cdot 0) + \frac{\varrho}{3}(0+\gamma V^{\pi_O}(s_5)) \\
&\quad +2\frac{\varrho}{3}(100+\gamma\cdot 0) \\
&= 10\varrho - 10 + \frac{\varrho}{3}\gamma V^{\pi_O}(s_5) + 200\frac{\varrho}{3} \\
V^{\pi_O}(s_5)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{230\varrho}{3} - 10
\end{aligned}
$$

- Damit:

$$
V^{\pi_O}(s_5) = \frac{\frac{230\varrho}{3} - 10}{1-\frac{\varrho}{3}\gamma} = \frac{230\varrho - 30}{3-\varrho\gamma}
$$

- Konkret für $\gamma = 0.9$ und $\varrho = 0.3$: $V^{\pi_O}(s_5) = 14.29$

Aufgabe 6: Labyrinth und Brücke

- Weiterbetrachtung von π_O und Fortschreibung für Zustände links von s_5:

Aufgabe 6: Labyrinth und Brücke

- Weiterbetrachtung von π_O und Fortschreibung für Zustände links von s_5:

$$\begin{aligned}
V^{\pi_O}(s_4) &= (1-\varrho)\cdot(0+\gamma V^{\pi_O}(s_5)) + \frac{\varrho}{3}(0+ \\
&\quad \gamma V^{\pi_O}(s_4)) + 2\frac{\varrho}{3}(100+\gamma\cdot 0) \\
V^{\pi_O}(s_4)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{200\varrho}{3} + \gamma V^{\pi_O}(s_5) - \varrho\gamma V^{\pi_O}(s_5)
\end{aligned}$$

Aufgabe 6: Labyrinth und Brücke

- Weiterbetrachtung von π_O und Fortschreibung für Zustände links von s_5:

$$\begin{aligned} V^{\pi_O}(s_4) &= (1-\varrho)\cdot(0+\gamma V^{\pi_O}(s_5)) + \frac{\varrho}{3}(0+ \\ &\quad \gamma V^{\pi_O}(s_4)) + 2\frac{\varrho}{3}(100+\gamma\cdot 0) \\ V^{\pi_O}(s_4)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{200\varrho}{3} + \gamma V^{\pi_O}(s_5) - \varrho\gamma V^{\pi_O}(s_5) \end{aligned}$$

- Damit:

$$V^{\pi_O}(s_4) = \frac{V^{\pi_O}(s_5)\cdot 3\gamma\cdot(1-\varrho)+200\varrho}{3-\varrho\gamma}$$

Aufgabe 6: Labyrinth und Brücke

- Weiterbetrachtung von π_O und Fortschreibung für Zustände links von s_5:

$$
\begin{aligned}
V^{\pi_O}(s_4) &= (1-\varrho)\cdot(0+\gamma V^{\pi_O}(s_5)) + \frac{\varrho}{3}(0+ \\
&\quad \gamma V^{\pi_O}(s_4)) + 2\frac{\varrho}{3}(100+\gamma\cdot 0) \\
V^{\pi_O}(s_4)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{200\varrho}{3} + \gamma V^{\pi_O}(s_5) - \varrho\gamma V^{\pi_O}(s_5)
\end{aligned}
$$

- Damit:

$$V^{\pi_O}(s_4) = \frac{V^{\pi_O}(s_5)\cdot 3\gamma\cdot(1-\varrho)+200\varrho}{3-\varrho\gamma}$$

- Konkret für $\gamma = 0.9$ und $\varrho = 0.3$: $V^{\pi_O}(s_4) = \frac{1.89\,V^{\pi_O}(s_5)+60}{2.73}$

Aufgabe 6: Labyrinth und Brücke

- Weiterbetrachtung von π_O und Fortschreibung für Zustände links von s_5:

$$
\begin{aligned}
V^{\pi_O}(s_4) &= (1-\varrho)\cdot(0+\gamma V^{\pi_O}(s_5)) + \frac{\varrho}{3}(0+ \\
&\quad \gamma V^{\pi_O}(s_4)) + 2\frac{\varrho}{3}(100+\gamma\cdot 0) \\
V^{\pi_O}(s_4)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{200\varrho}{3} + \gamma V^{\pi_O}(s_5) - \varrho\gamma V^{\pi_O}(s_5)
\end{aligned}
$$

- Damit:

$$V^{\pi_O}(s_4) = \frac{V^{\pi_O}(s_5)\cdot 3\gamma\cdot(1-\varrho)+200\varrho}{3-\varrho\gamma}$$

- Konkret für $\gamma = 0.9$ und $\varrho = 0.3$: $V^{\pi_O}(s_4) = \frac{1.89 V^{\pi_O}(s_5)+60}{2.73}$
- Allgemein für $s_1,\ldots,s_4$: $V^{\pi_O}(s_k) = \frac{1.89 V^{\pi_O}(s_{k+1})+60}{2.73}$
- Fortsetzung:

Aufgabe 6: Labyrinth und Brücke

- Weiterbetrachtung von π_O und Fortschreibung für Zustände links von s_5:

$$
\begin{aligned}
V^{\pi_O}(s_4) &= (1-\varrho)\cdot(0+\gamma V^{\pi_O}(s_5)) + \frac{\varrho}{3}(0+ \\
&\quad \gamma V^{\pi_O}(s_4)) + 2\frac{\varrho}{3}(100+\gamma\cdot 0) \\
V^{\pi_O}(s_4)\cdot\left(1-\frac{\varrho}{3}\gamma\right) &= \frac{200\varrho}{3} + \gamma V^{\pi_O}(s_5) - \varrho\gamma V^{\pi_O}(s_5)
\end{aligned}
$$

- Damit:

$$V^{\pi_O}(s_4) = \frac{V^{\pi_O}(s_5)\cdot 3\gamma\cdot(1-\varrho)+200\varrho}{3-\varrho\gamma}$$

- Konkret für $\gamma = 0.9$ und $\varrho = 0.3$: $V^{\pi_O}(s_4) = \frac{1.89\,V^{\pi_O}(s_5)+60}{2.73}$
- Allgemein für $s_1,\ldots,s_4$: $V^{\pi_O}(s_k) = \frac{1.89\,V^{\pi_O}(s_{k+1})+60}{2.73}$
- Fortsetzung: $\rightarrow$ Tafel

Aufgabe 6: Labyrinth und Brücke

(c) Welcher der beiden Parameter γ und ϱ ist abzuändern, damit sich der Agent traut, die Brücke zu überqueren?

Aufgabe 6: Labyrinth und Brücke

(c) Welcher der beiden Parameter γ und ϱ ist abzuändern, damit sich der Agent traut, die Brücke zu überqueren?

- Meinungen? Lösungen? Lösungideen? Was muss gelten?

Aufgabe 6: Labyrinth und Brücke

(c) Welcher der beiden Parameter γ und ϱ ist abzuändern, damit sich der Agent traut, die Brücke zu überqueren?

- Meinungen? Lösungen? Lösungideen? Was muss gelten?
- Damit sich der Agent über die Brücke traut, muss offensichtlich $V^{\pi_O}(S) < V^{\pi_W}(S)$ gelten, d.h. der Agent muss unter der Aktion O im Startzustand $S = s_1$ geringere Kosten erwarten als für W.

Aufgabe 6: Labyrinth und Brücke

(c) Welcher der beiden Parameter γ und ϱ ist abzuändern, damit sich der Agent traut, die Brücke zu überqueren?

- Meinungen? Lösungen? Lösungideen? Was muss gelten?
- Damit sich der Agent über die Brücke traut, muss offensichtlich $V^{\pi_O}(S) < V^{\pi_W}(S)$ gelten, d.h. der Agent muss unter der Aktion O im Startzustand $S = s_1$ geringere Kosten erwarten als für W.
- Von der vorherigen Teilaufgabe wissen wir, dass $V^{\pi_W}(s_1) = \frac{203\varrho - 3}{3 - \varrho\gamma}$. $\rightarrow$ Ideen?

Aufgabe 6: Labyrinth und Brücke

(c) Welcher der beiden Parameter γ und ϱ ist abzuändern, damit sich der Agent traut, die Brücke zu überqueren?

- Meinungen? Lösungen? Lösungideen? Was muss gelten?
- Damit sich der Agent über die Brücke traut, muss offensichtlich $V^{\pi_O}(S) < V^{\pi_W}(S)$ gelten, d.h. der Agent muss unter der Aktion O im Startzustand $S = s_1$ geringere Kosten erwarten als für W.
- Von der vorherigen Teilaufgabe wissen wir, dass $V^{\pi_W}(s_1) = \frac{203\varrho-3}{3-\varrho\gamma}$. $\rightarrow$ Ideen?
- Idee: Wir könnten den Wert des Diskontierungsfaktors γ vergrößern, so nimmt der Nenner kleinere Werte an und damit können wir $V^{\pi_W}(s_1)$ vergrößern.
- Im Limit ($\gamma \rightarrow 1$) erhielten wir damit (weil $\varrho = 0.3$): $V^{\pi_W}(s_1) = \frac{60.9-3}{3-0.3} = 21.44$.

Aufgabe 6: Labyrinth und Brücke

(c) Welcher der beiden Parameter γ und ϱ ist abzuändern, damit sich der Agent traut, die Brücke zu überqueren?

- Unter diesem extremen γ würden sich die Werte von $V^{\pi_O}(s_k)$ ebenfalls ändern.

Aufgabe 6: Labyrinth und Brücke

(c) Welcher der beiden Parameter γ und ϱ ist abzuändern, damit sich der Agent traut, die Brücke zu überqueren?

- Unter diesem extremen γ würden sich die Werte von $V^{\pi_O}(s_k)$ ebenfalls ändern.
- Wir erhalten für $V^{\pi_O}(s_5)$ ebenfalls eine Vergrößerung: $V^{\pi_O}(s_5) = \frac{230\varrho-30}{3-\varrho\gamma} = 14.44$

Aufgabe 6: Labyrinth und Brücke

(c) Welcher der beiden Parameter γ und ϱ ist abzuändern, damit sich der Agent traut, die Brücke zu überqueren?

- Unter diesem extremen γ würden sich die Werte von $V^{\pi_O}(s_k)$ ebenfalls ändern.
- Wir erhalten für $V^{\pi_O}(s_5)$ ebenfalls eine Vergrößerung: $V^{\pi_O}(s_5) = \frac{230\varrho - 30}{3 - \varrho\gamma} = 14.44$
- Die Rekursionsvorschrift für die Berechnung von $s_1, \ldots, s_5$ ändert sich zu $V^{\pi_O}(s_k) = \frac{V^{\pi_O}(s_{k+1}) \cdot 3 \cdot (1-\varrho) + 200\varrho}{3-\varrho} = \frac{V^{\pi_O}(s_{k+1}) \cdot 2.1 + 60}{2.7}$.

Aufgabe 6: Labyrinth und Brücke

(c) Welcher der beiden Parameter γ und ϱ ist abzuändern, damit sich der Agent traut, die Brücke zu überqueren?

- Unter diesem extremen γ würden sich die Werte von $V^{\pi_O}(s_k)$ ebenfalls ändern.
- Wir erhalten für $V^{\pi_O}(s_5)$ ebenfalls eine Vergrößerung: $V^{\pi_O}(s_5) = \frac{230\varrho - 30}{3 - \varrho\gamma} = 14.44$
- Die Rekursionsvorschrift für die Berechnung von $s_1, \ldots, s_5$ ändert sich zu $V^{\pi_O}(s_k) = \frac{V^{\pi_O}(s_{k+1}) \cdot 3 \cdot (1-\varrho) + 200\varrho}{3-\varrho} = \frac{V^{\pi_O}(s_{k+1}) \cdot 2.1 + 60}{2.7}$.
- Damit ergeben sich:
  $V^{\pi_O}(s_4) = 33.45$, $V^{\pi_O}(s_3) = 48.24$
  $V^{\pi_O}(s_2) = 59.74$, $V^{\pi_O}(s_1) = 68.69 > V^{\pi_W}(s_1) = 21.44$
- Durch eine Veränderung des Diskontierungsfaktors kann der Agent nicht dazu bewegt werden, die Brücke zu überqueren.

Aufgabe 6: Labyrinth und Brücke

(d) Auf welchen Wert muss jener Parameter zu diesem Zweck (näherungsweise) gesetzt werden?

Aufgabe 6: Labyrinth und Brücke

(d) Auf welchen Wert muss jener Parameter zu diesem Zweck (näherungsweise) gesetzt werden?

- Also: Nur durch eine Veränderung der Rauschintensität (naheliegenderweise deren Reduzierung) kann der Agent dazu bewegt werden, die Brücke zu überqueren.
- Der Diskontierungsfaktor belassen wir hierbei konstant bei $\gamma = 0.9$.

Aufgabe 6: Labyrinth und Brücke

(d) Auf welchen Wert muss jener Parameter zu diesem Zweck (näherungsweise) gesetzt werden?

- Also: Nur durch eine Veränderung der Rauschintensität (naheliegenderweise deren Reduzierung) kann der Agent dazu bewegt werden, die Brücke zu überqueren.
- Der Diskontierungsfaktor belassen wir hierbei konstant bei $\gamma = 0.9$.
- Uns interessiert also die Änderung von $V^{\pi_O}(s_k)$ mit $s_k \in \{s_1, \ldots, s_5\}$ im Vergleich zu $V^{\pi_W}(s_1)$ in Abhängigkeit der Rauschintensität ϱ. Insbesondere sind natürlich die Werte im Zustand $S = s_1$ von Interesse, da hier die Entscheidung fällt, ob der Agent sich über die Brücke trauen wird oder nicht.

Aufgabe 6: Labyrinth und Brücke

(d) Auf welchen Wert muss jener Parameter zu diesem Zweck (näherungsweise) gesetzt werden?

- Also: Nur durch eine Veränderung der Rauschintensität (naheliegenderweise deren Reduzierung) kann der Agent dazu bewegt werden, die Brücke zu überqueren.
- Der Diskontierungsfaktor belassen wir hierbei konstant bei $\gamma = 0.9$.
- Uns interessiert also die Änderung von $V^{\pi_O}(s_k)$ mit $s_k \in \{s_1, \ldots, s_5\}$ im Vergleich zu $V^{\pi_W}(s_1)$ in Abhängigkeit der Rauschintensität ϱ. Insbesondere sind natürlich die Werte im Zustand $S = s_1$ von Interesse, da hier die Entscheidung fällt, ob der Agent sich über die Brücke trauen wird oder nicht.
- Zur Erinnerung:
  - Wir hatten

$$V^{\pi_W}(s_1) = \frac{203\varrho - 3}{3 - \varrho\gamma} = \frac{203\varrho - 3}{3 - 0.9\varrho}$$

Aufgabe 6: Labyrinth und Brücke

(d) Auf welchen Wert muss jener Parameter zu diesem Zweck (näherungsweise) gesetzt werden?

- Zur Erinnerung (Forts.):
  - Es gilt

$$V^{\pi_O}(s_5) = \frac{230\varrho - 30}{3 - \varrho\gamma} = \frac{230\varrho - 30}{3 - 0.9\varrho}$$

Aufgabe 6: Labyrinth und Brücke

(d) Auf welchen Wert muss jener Parameter zu diesem Zweck (näherungsweise) gesetzt werden?

- Zur Erinnerung (Forts.):
  - Es gilt

$$V^{\pi_O}(s_5) = \frac{230\varrho - 30}{3 - \varrho\gamma} = \frac{230\varrho - 30}{3 - 0.9\varrho}$$

  - Ferner für $s_k \in \{s_1, \ldots, s_4\}$

$$\begin{aligned} V^{\pi_O}(s_k) &= \frac{200\varrho + V^{\pi_O}(s_{k+1}) \cdot 3\gamma \cdot (1-\varrho)}{3 - \varrho\gamma} \\ &= \frac{200\varrho + 2.7 \cdot V^{\pi_O}(s_{k+1}) \cdot (1-\varrho)}{3 - 0.9\varrho} \end{aligned}$$

Aufgabe 6: Labyrinth und Brücke

(d) Auf welchen Wert muss jener Parameter zu diesem Zweck (näherungsweise) gesetzt werden?

- Zur Erinnerung (Forts.):
  - Es gilt

$$V^{\pi_O}(s_5) = \frac{230\varrho - 30}{3 - \varrho\gamma} = \frac{230\varrho - 30}{3 - 0.9\varrho}$$

  - Ferner für $s_k \in \{s_1, \ldots, s_4\}$

$$\begin{aligned} V^{\pi_O}(s_k) &= \frac{200\varrho + V^{\pi_O}(s_{k+1}) \cdot 3\gamma \cdot (1-\varrho)}{3 - \varrho\gamma} \\ &= \frac{200\varrho + 2.7 \cdot V^{\pi_O}(s_{k+1}) \cdot (1-\varrho)}{3 - 0.9\varrho} \end{aligned}$$

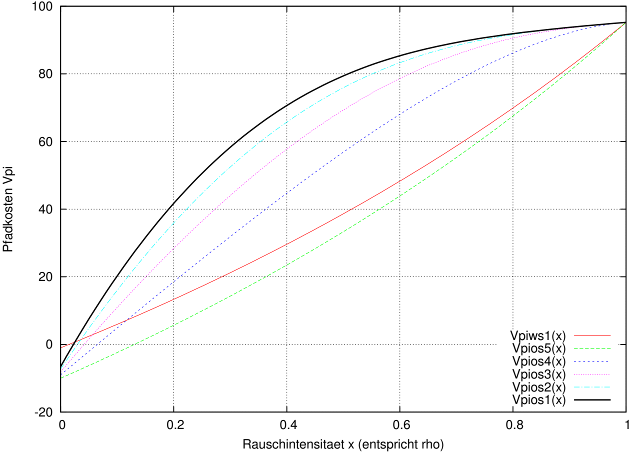
- Ab welchem Wert von ϱ gilt $V^{\pi_W}(s_1) \geq V^{\pi_O}(s_1)$?

Aufgabe 6: Labyrinth und Brücke

- Näherungslösung (Plot der Pfadkosten in Abhängigkeit von ϱ)

Aufgabe 6: Labyrinth und Brücke

- Näherungslösung (Plot der Pfadkosten in Abhängigkeit von ϱ)



Aufgabe 6: Labyrinth und Brücke

- Näherungslösung (Vergrößerung) $\Rightarrow \varrho^* \approx 0.025$

Für Rauschintensitäten $\varrho \leq \varrho^*$ wird sich der Agent gen Osten bewegen und damit über die Brücke trauen.