Kapitel D

Markov–Spiele und Bellmans Optimalitätsprinzip

So you're telling me it is a matter of probability and odds;
I was worried there was some chance involved.
Vesper Lynd zu James Bond im Film *Casino Royale* (2006)

While in theory randomness is an intrinsic property,
in practice, randomness is incomplete information.
Nassim Nicholas Taleb (1960–)

Inhalt dieses Kapitels D

Motivation und Überblick

In diesem Kapitel betrachten wir Markov–Spiele, in denen ein Akteur gegen den Zufall spielt und hierzu eine **optimale Strategie** sucht.

Ich beginne simpel und möglichst **problemorientiert** [*problem driven*]: Wie können wir uns in sehr einfachen Bei-Spielen optimal verhalten? Daraus ergeben sich die richtigen Fragen, oft direkt von Studierenden, die wir dann durch Aufbau einer **passenden Theorie** zu lösen suchen [*method driven*]. Beide Arbeitsweisen ergänzen sich bestens.

Je considère comme complètement inutile la lecture de gros traités d'analyse pure: un trop grand nombre de méthodes passent en même temps devant les yeux. C'est dans les travaux d'application qu'on doit les étudier; c'est là qu'on juge leurs capacités et qu'on apprend la manière de les utiliser.
Joseph-Louis Lagrange (1736–1813)

Monsieur Cauchy annonce, que, pour se conformer au voeu du Conseil, il ne s'attachera plus à donner, comme il a fait jusqu'à présent, des démonstrations parfaitement rigoureuse.
Conseil d'instruction de l'École Polytechnique (1825)

Motivation und Überblick

In den vorigen Kapiteln haben wir die **Rekursion** / Rückwärtsinduktion kennen und nutzen und schätzen gelernt. In Anwendungen sind die zu lösenden Gleichungen oft nicht rekursiv aufgebaut, also nicht von klein nach groß „topologisch" sortierbar, sondern selbstbezüglich / zyklisch. Dies sind also allgemeine **Fixpunktgleichungen**. Ihre wunderschöne Theorie und praktische Anwendung sind überall von großer Bedeutung.

Nancy L. Stokey, Robert E. Lucas:
Recursive Methods in Economic Dynamics. Harvard Univ. Press 1989

Lars Ljungqvist, Thomas J. Sargent:
Recursive Macroeconomic Theory. The MIT Press (3rd ed.) 2012

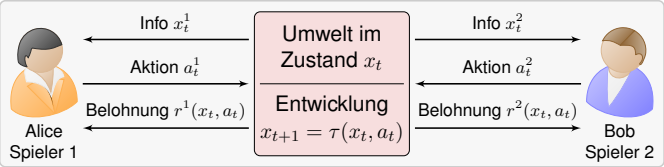
Die hier gesuchte Optimierung des strategischen Handelns führt aus algorithmischer Sicht zum **maschinellen Lernen** [*machine learning*].

Stuart Russell, Peter Norvig: *Artificial Intelligence: A Modern Approach.* Addison Wesley (3rd ed.) 2016 (Kapitel 17: Making complex decisions)

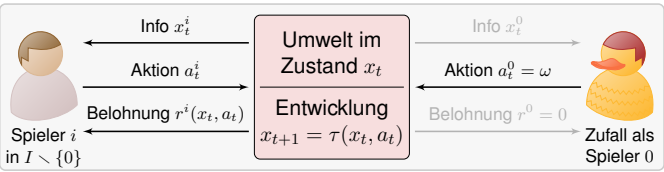
Richard S. Sutton, Andrew G. Barto: *Reinforcement Learning.* The MIT Press (2nd ed.) 2018 (hier speziell §6.5: Q–learning).

Dynamische Spiele und strategische Interaktion

Dynamisches System mit Steuerung durch zwei Spieler: Markov–Spiel



Steuerung durch mehrere Spieler und Zufall: Markov–Spiel (MMDP)



Jeder Spieler will seinen individuellen Gesamtnutzen maximieren.

Dynamische Spiele und strategische Interaktion

In Kapitel C haben wir **neutrale kombinatorische Spiele** betrachtet: Spieler ziehen abwechselnd und haben dieselben Zugmöglichkeiten. Das vereinfacht die Analyse und enthüllt besonders klare Strukturen: Der Satz von Sprague–Grundy überführt jedes neutrale Spiel in ein äquivalentes Nim-Spiel und erlaubt (oft genug) eine effiziente Lösung.

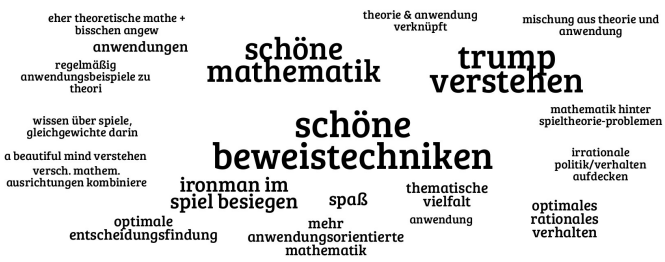
Unser Ziel sind **allgemeine Spiele** für zwei oder mehr Personen, bei denen Personen abwechselnd oder auch gleichzeitig ziehen. Bevor wir solche Spiele erklären und lösen, möchte ich noch etwas genauer einige spezielle und besonders einfache Fälle beleuchten.

In diesem Kapitel untersuchen wir Markov–Spiele, mit nur einem Spieler, aber Zufallszügen. Es handelt sich also um **Spiele gegen den Zufall**. Das Verhalten des Zufalls ist (in unseren Modellen) fest vorgegeben, daher handelt es sich um ein (stochastisches) **Optimierungsproblem**.

Der übliche Name hierfür ist **Markov–Entscheidungsprozess** [*Markov decision process*, MDP], doch auch das ist nichts anderes als ein Spiel. Hierzu erkläre ich erste Beispiele und Lösungsmethoden.

Schnappschuss: Erwartungen der Studierenden

Erwartungen der Studierenden zu Beginn des Semesters (2019):



Schnappschuss: Erwartungen der Studierenden

Erwartungen der Studierenden zu Beginn des Semesters (2022):

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D101

0	1	2	3	4	5	6	7	8
7€								23€

Aufgabe: Ihre Spielfigur startet auf einem gelben Spielfeld im Inneren. In jedem Zug rückt sie auf ein Nachbarfeld, zufällig und gleichverteilt. Das Spiel endet am Rand $\partial X = \{0, 8\}$ mit dem gezeigten Gewinn. Wie viel würden Sie als Teilnehmer zahlen / als Anbieter verlangen?

(0) Was ist die Gewinnerwartung $u(x)$ für jedes Startfeld $x \in X$?
(1) Jeder Zug kostet, $c = -1$€, und Sie müssen zu Ende spielen.
(2) Jeder Zug kostet, $c = -1$€, und Sie dürfen jederzeit aufgeben.

Schätzen Sie zunächst! Wie treffsicher ist Ihre intuitive Erwartung? Formulieren Sie allgemeine Gleichungen und Lösungsmethoden: Wie *prüfen* Sie eine Lösung? Wie *finden* Sie eine/alle Lösungen?

7	9	11	13	15	17	19	21	23
7	2	−1	−2	−1	2	7	14	23
7.00	2.67	0.33	0.00	0.60	3.20	7.80	14.40	23.00

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D102 Erläuterung

Hier ist (0) ein extrem einfaches Spiel, schon (1) dürfte überraschen: Ungeschult haben wir herzlich wenig Erfahrung mit zufälligen Irrfahrten. Allgemein fällt Menschen rekursives Denken erfahrungsgemäß schwer, doch gerade dies ist für rationale Entscheidungen (2) wesentlich.

Bevor wir die Lösung diskutieren, schätzen Sie bitte die Erwartung. Ist Ihre Intuition präzise und treffsicher, oder allzu vage und irrig?

Diese quantitativen Schätzfragen sind ein aufschlussreicher Test der vielzitierten Schwarmintelligenz und mahnen zur Vorsicht: Betrügerische Geschäftspraktiken beruhen gerade darauf, dass das Gegenüber die Situation schlecht einschätzen kann und daher Fehlentscheidungen trifft.

Es ist schön und gut, die eigene Intuition zu nutzen und zu entwickeln. Leider hilft es wenig, eine Antwort ohne Begründung anzugeben. Wir wollen begründete, nachvollziehbare, tragfähige Argumente! Auch das ist ein Qualitätsmerkmal rationalen Handelns.

Übung: Angenommen, alle Spieldaten sind rational, in $\mathbb{Q}$, so wie hier. Sind dann ebenfalls auch alle Ergebnisse rational? Hier scheint es so!

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D103 Übung

0	1	2	3	4	5	6	7	8
7	9	11	13	15	17	19	21	23

Aufgabe: (0) Wie berechnen Sie die Gewinnerwartung?

Lösung: Für $x \in X = \{0, 1, \dots, 8\}$ suchen wir $u_x = u(x)$. Wir haben:

$$\begin{aligned}
u_0 &= 7 \\
-\tfrac{1}{2}u_0 + u_1 - \tfrac{1}{2}u_2 &= 0 \\
-\tfrac{1}{2}u_1 + u_2 - \tfrac{1}{2}u_3 &= 0 \\
-\tfrac{1}{2}u_2 + u_3 - \tfrac{1}{2}u_4 &= 0 \\
-\tfrac{1}{2}u_3 + u_4 - \tfrac{1}{2}u_5 &= 0 \\
-\tfrac{1}{2}u_4 + u_5 - \tfrac{1}{2}u_6 &= 0 \\
-\tfrac{1}{2}u_5 + u_6 - \tfrac{1}{2}u_7 &= 0 \\
-\tfrac{1}{2}u_6 + u_7 - \tfrac{1}{2}u_8 &= 0 \\
u_8 &= 23
\end{aligned}$$

Die Koeffizienten bilden eine Bandmatrix: tridiagonal, dünn besetzt

☺ Lineare Gleichungssysteme können Sie lösen: Gauß gelingt immer! Vereinfachung: Für $d_x = u_x - u_{x-1}$ erhalten wir $d_1 = d_2 = \cdots = d_8 = d$ und $8d = 23 - 7 = 16$, also $d = 2$ und $u_x = 7 + d \cdot x$ für alle $x \in X$.

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D104 Erläuterung

Dies ist ein **lineares Gleichungssystem**, homogen im Inneren X°, inhomogen am Rand ∂X aufgrund der Dirichlet–Randbedingung.

Es ist klein genug, um noch von Hand gelöst zu werden. Versuchen Sie es selbst: Rechnen reinigt die Seele!

Es ist zudem dünn besetzt und hat eine hohe Symmetrie. Das nutzen wir gerne, etwa durch wiederholte Halbierung.

Die hier gezeigte simple Lösung funktioniert ganz allgemein: Die Lösung ist die eindeutige Gerade durch die beiden Randwerte.

Die Gleichung $u_x = \frac{1}{2}u_{x-1} + \frac{1}{2}u_{x+1}$ nennen wir **Mittelwerteigenschaft**. Eine solche Funktion $u : X \to \mathbb{R}$ heißt **harmonisch**, in Anlehnung an die klassische partielle Differentialgleichung $\Delta u = 0$ auf $\Omega \subseteq \mathbb{R}^n$, wobei $\Delta = \partial_1^2 + \partial_2^2 + \cdots + \partial_n^2$ der Laplace–Operator ist.

Die Differenz $-\frac{1}{2}u_{x-1} + u_x - \frac{1}{2}u_{x+1}$ entspricht der (negativen) **zweiten Ableitung** an der Stelle x. Ist sie gleich Null, so handelt es sich um eine Gerade. Ist sie negativ, so sehen wir eine nach oben geöffnete Parabel. Die Zugkosten c_x im Punkt x entsprechen der Krümmung im Punkt x!

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D105 Übung

0	1	2	3	4	5	6	7	8
7	2	−1	−2	−1	2	7	14	23

Aufgabe: (1) Wie berechnen Sie die Gewinnerwartung bei Zugkosten?

Lösung: Für $x \in X = \{0, 1, \dots, 8\}$ suchen wir $u_x = u(x)$. Wir haben:

$$\begin{aligned}
u_0 &= 7 \\
-\tfrac{1}{2}u_0 + u_1 - \tfrac{1}{2}u_2 &= c_1 \\
-\tfrac{1}{2}u_1 + u_2 - \tfrac{1}{2}u_3 &= c_2 \\
-\tfrac{1}{2}u_2 + u_3 - \tfrac{1}{2}u_4 &= c_3 \\
-\tfrac{1}{2}u_3 + u_4 - \tfrac{1}{2}u_5 &= c_4 \\
-\tfrac{1}{2}u_4 + u_5 - \tfrac{1}{2}u_6 &= c_5 \\
-\tfrac{1}{2}u_5 + u_6 - \tfrac{1}{2}u_7 &= c_6 \\
-\tfrac{1}{2}u_6 + u_7 - \tfrac{1}{2}u_8 &= c_7 \\
u_8 &= 23
\end{aligned}$$

Die Koeffizienten bilden eine Bandmatrix: tridiagonal, dünn besetzt

☺ Allgemeine Faustregel: Ausrechnen ist mühsam. Prüfen ist leicht! Wir vermuten, dass die Lösung eindeutig ist. Gilt das? Satz D1A! Negative Gewinnerwartung bedeutet: Ab hier besser nicht spielen!

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D106 Übung

0	1	2	3	4	5	6	7	8
7.00	2.67	0.33	0.00	0.60	3.20	7.80	14.40	23.00

Aufgabe: (2) Wie berechnen Sie die Gewinnerwartung bei Zugkosten? Als zusätzliche Option darf der Spieler nun auch aufgeben / abbrechen.

Lösung: Wir suchen $u : X \to \mathbb{R}$. Am Rand gilt $u(0) = 7$ und $u(8) = 23$. In jedem aktiven Zustand $x \in \{1, 2, \dots, 7\}$ sind zwei Aktionen möglich:

Weiterspielen: $u(x) = c(x) + \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$
Aufgeben: $u(x) = 0$

Ein rationaler Spieler wählt jeweils den besten Zug, also das Maximum:

$$u(x) = \max\{\ 0,\ c(x) + \tfrac{1}{2}u(x-1) + \tfrac{1}{2}u(x+1)\ \}$$

Dieses Gleichungssystem ist… nicht-linear… und selbstbezüglich. Wie immer gilt auch hier: Ausrechnen ist mühsam. Prüfen ist leicht! Existiert immer eine Lösung u? Gibt es mehrere oder genau eine? Wie können wir sie berechnen? Zudem möglichst effizient? Übung!

☺ Hier hilft Ihnen ein Python-Skript oder eine Tabellenkalkulation!

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D107

(2) Eine Näherung gelingt einfach und effizient durch Iteration:

$$u_{t+1}(x) = \max\{\ 0,\ c(x) + \tfrac{1}{2}u_t(x-1) + \tfrac{1}{2}u_t(x+1)\ \}$$

t	$u_t(0)$	$u_t(1)$	$u_t(2)$	$u_t(3)$	$u_t(4)$	$u_t(5)$	$u_t(6)$	$u_t(7)$	$u_t(8)$
0	7.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	23.00
1	7.00	2.50	0.00	0.00	0.00	0.00	0.00	10.50	23.00
2	7.00	2.50	0.25	0.00	0.00	0.00	4.25	10.50	23.00
3	7.00	2.63	0.25	0.00	0.00	1.13	4.25	12.63	23.00
4	7.00	2.63	0.31	0.00	0.00	1.13	5.88	12.63	23.00
5	7.00	2.66	0.31	0.00	0.00	1.94	5.88	13.44	23.00
6	7.00	2.66	0.33	0.00	0.00	1.94	6.69	13.44	23.00
7	7.00	2.66	0.33	0.00	0.00	2.34	6.69	13.84	23.00
8	7.00	2.66	0.33	0.00	0.17	2.34	7.09	13.84	23.00
9	7.00	2.67	0.33	0.00	0.17	2.63	7.09	14.05	23.00
…									
39	7.00	2.67	0.33	0.00	0.60	3.20	7.80	14.40	23.00
40	7.00	2.67	0.33	0.00	0.60	3.20	7.80	14.40	23.00

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D108 Erläuterung

☺ Vergleichen Sie die hier berechneten Werte mit Ihren intuitiven Schätzungen bei der ersten, informellen Annäherung an diese Frage. Unser kleines Markov–Spiel ist eine sehr einfache, aber durchaus realistische Illustration für einen Markov–Entscheidungsprozess.

☺ Die Lösung u_0 im linearen Fall, ohne Entscheidung, ist leicht zu berechnen durch das lineare Gleichungssystem, also als Fixpunkt: Wir nutzen auf $E = \{\ u : X \to \mathbb{R} \mid u(0) = 7,\ u(8) = 23\ \}$ den linearen Operator $\Phi_0 : E \to E : u \mapsto \bar{u}$ mit $\bar{u}(x) = c(x) + \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$.

⚠ Die Lösung mit Entscheidungsmöglichkeit ist nicht $\max\{0, u_0\}$! Diese naive Fehlannahme führt tatsächlich zu Fehlentscheidungen.

☺ Wir erhalten sie vielmehr als Fixpunkt von $\Phi(u) = \max\{0, \Phi_0(u)\}$. Diesen nicht-linearen Operator können wir zur Iteration nutzen, wie nachfolgend gezeigt. Ist die beobachtete Konvergenz nicht wunderbar?

☺ In der Übung zeigen Sie allgemein, dass $\Phi : E \to E$ kontraktiv ist, so dass Sie Banachs Fixpunktsatz anwenden können. Alles wird gut. Damit können Sie u berechnen und Ihre optimale Strategie ablesen!

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D109

0	1	2	3	4	5	6	7	8	9	10	
100€											...

Aufgabe: Selbes Spiel wie zuvor, aber nach rechts unbegrenzt.
Wie viel würden Sie als Teilnehmer zahlen / als Anbieter verlangen?

(0) Was ist die Gewinnerwartung $u(x)$ für jedes Startfeld $x \in \mathbb{N}$?
(1) Jeder Zug kostet, $c = -1$€, und Sie müssen zu Ende spielen.
(2) Jeder Zug kostet, $c = -1$€, und Sie dürfen jederzeit aufgeben.

Schätzen Sie zunächst! Wie treffsicher ist Ihre intuitive Erwartung?
Formulieren Sie allgemeine Gleichungen und Lösungsmethoden:
Existiert eine Lösung $u : X \to \mathbb{R}$? Gibt es mehrere oder genau eine?
Wie können wir sie berechnen? Zudem möglichst effizient?

100	100	100	100	100	100	100	100	100	100	100	...
100	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	...
100	81	64	49	36	25	16	9	4	1	0	...

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D110
Übung

Lösung: (0) Die Mittelwerteigenschaft $u(x) = \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$ hat als mögliche Lösungen $u_m(x) = 100 + m$ mit $m \in \mathbb{R} \cup \{\pm\infty\}$.

⚠ Allein der eine Punkt $u(0) = 100$ legt die Gerade noch nicht fest!
Die simple Bilanzgleichung allein genügt hier also nicht zur Lösung.

Wir müssen etwas tiefer graben und das zu Grunde liegende **Modell** genauer präzisieren und auswerten: die Irrfahrt auf $\mathbb{Z}$ [*random walk*].
Der Übergang zu einem feineren Modell ähnelt der **Mikrofundierung**, um globale Bilanzgleichungen zu erklären oder notfalls zu ergänzen

🙂 Wir vollenden die Rechnung, indem wir **weitere Bedingungen** einführen und nutzen: Der Erwartungswert $u(x) \in \mathbb{R}$ existiert, erfüllt die Bilanzgleichung $u(x) = \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$ und ist **beschränkt** durch $u(x) \in [0, 100]$ für alle $x \in \mathbb{N}$. Dann bleibt nur $u(x) = 100$ für alle $x \in \mathbb{N}$.

⚠ Wir sehen hier im Miniaturbeispiel den entscheidenden Unterschied zwischen einem endlichem und einem unendlichem Spielgraphen, zwischen einem kompakten und einem nicht-kompakten Gebiet.
Im Allgemeinen ist nur der endliche / kompakte Fall gutartig.

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D111
Übung

(1) Die Lösungen sind $u_m(x) = 100 + mx + x^2$ mit $m \in \mathbb{R} \cup \{\pm\infty\}$.
Jede erfüllt $u_m(x) = \frac{1}{2}u_m(x-1) + \frac{1}{2}u_m(x+1) - 1$ für alle $x \in \mathbb{N}_{>0}$.
⚠ Die Lösung $u_{-\infty}$ ist auch auf $X = \{0, 1, 2, \dots, n\}$ mit $n \geq 3$ möglich.
Die Frage ist also: Warum ist dies auf $\mathbb{N}$ die einzig richtige Antwort?
🙂 Wir vollenden die Rechnung, indem wir **weitere Bedingungen** einführen und nutzen: Nur $u_{-\infty}$ erfüllt die Beschränkung $u \leq 100$.

Jedes mathematische Phänomen lässt sich finanziell ausnutzen (in gut gemeinten Anwendungen) bzw. ausbeuten (in betrügerischer Absicht).
Letzteres gilt besonders für wenig bekannte Regelmäßigkeiten (Sätze), noch besser eignen sich Unverständnis und weit verbreitete Fehler.
Ausgenutzt wird hierbei nicht direkt der mathematische Sachverhalt, sondern vor allem die ungleich verteilte Information darüber.

⚠ Ein Spiel oder Geschäft wie in (1) ist ein **Knebelvertrag** und zielt darauf, eine Vertragspartei möglichst langfristig zu binden und finanziell auszubeuten, meist wie hier durch Kündigungsfristen und -bedingungen.
In klaren Fällen sind solche Verträge sittenwidrig und somit ungültig.

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D112
Übung

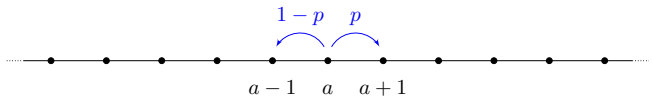
(2) Ein rationaler Spieler wählt jeweils den besten Zug:

$$u(x) = \max\{\, 0,\ \tfrac{1}{2}u(x-1) + \tfrac{1}{2}u(x+1) - 1 \,\}$$

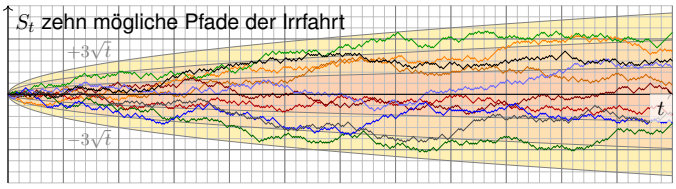
Diese Problemstellung auf $\mathbb{N}$ ist potentiell unendlich, lässt sich aber leicht auf das endliche Problem auf $X = \{0, 1, 2 \dots, n\}$ zurückführen:
Wir wählen den rechten Rand n hinreichend groß und setzen $u(n) = 0$.
Aufgrund der Beschränkung $u(x) \in [0, 100]$ für $x \geq 0$ und der Zugkosten $c = -1$ finden wir $u(x) \in [0, 99]$ für $x \geq 1$, sodann $u(x) \in [0, 98]$ für $x \geq 2$ und so weiter, bis schließlich $u(x) = 0$ für $x \geq 100$. Also genügt $n = 100$.
Das lineare Gleichungssystem ergibt unsere obigen Quadratzahlen!
Wenn wir erst einmal $u(10) = 0$ wissen, dann ist dies leicht zu sehen:
Von der Position $x \in \{0, 1, \dots, 10\}$ bis zum Rand $\{0, 10\}$ ist die lineare Erwartung $100 - 10x$ und zudem die erwartete Reisezeit $x(10 - x)$.
Die Gewinnerwartung ist demnach $u(x) = 100 - 20x + x^2 = (10 - x)^2$.
🙂 Wie immer gilt auch hier: Ausrechnen ist mühsam. Prüfen ist leicht!
🙂 Im Gegensatz zum Knebelvertrag (1) scheint mir dieses Spiel (2) überschaubar und sowohl moralisch wie juristisch akzeptabel.

Zufällige Irrfahrt / Random Walk

D113
Erläuterung



Zufällige Irrfahrt in $X = \mathbb{Z}$: Zur Zeit $t = 0$ starten Sie im Punkt $S_0 = a$.
Im Schritt von S_t nach S_{t+1} gehen Sie mit Wahrscheinlichkeit $p \in [0, 1]$ nach rechts und entsprechend mit Wahrscheinlichkeit $(1 - p)$ nach links.
Das heißt, $S_t : \Omega \to \mathbb{Z}$ ist gegeben durch $S_t = a + X_1 + \cdots + X_t$ mit unabhängigen Zuwächsen, $\mathbf{P}_a(X_t{=}{+1}) = p$ und $\mathbf{P}_a(X_t{=}{-1}) = 1 - p$.



Zufällige Irrfahrt / Random Walk

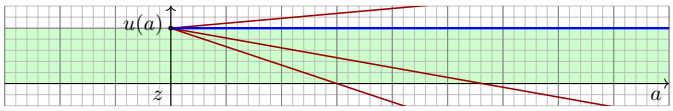
D114
Erläuterung

Aufgabe: (0) Bestimmen Sie zu S_t die Verteilung, Erwartung, Streuung.
Wir untersuchen speziell den symmetrischen Fall $p = 1/2$, ohne Drift.

(1) Sie beginnen im Startpunkt $a \in \mathbb{Z}$ und fixieren einen Zielpunkt $z \in \mathbb{Z}$.
Wie groß ist die Wkt $u(a) \in [0, 1]$, das Ziel z irgendwann zu erreichen?

(2) Wie groß ist hierbei die erwartete Reisezeit von a nach z?

Lösung: (1) Wir erhalten eine Binomialverteilung, affin transformiert:
$\mathbf{P}_a(S_t = a + 2k - t) = \binom{t}{k} p^k (1-p)^{t-k} = \binom{t}{k} \frac{1}{2^t}$ für $t, k \in \mathbb{N}$ und $p = \frac{1}{2}$.
Somit gilt $\mathbf{E}(S_t) = a$ und $\mathbf{V}(S_t) = t$, also $\sigma(S_t) = \sqrt{t}$ und $S_t \approx N(a, t)$.



(1) Offensichtlich gilt $u(z) = 1$, denn hier ist der Start auch das Ziel.
Für $a > z$ gilt die Mittelwerteigenschaft $u(a) = \frac{1}{2}u(a+1) + \frac{1}{2}u(a-1)$.
Somit ist $u : \mathbb{Z}_{\geq z} \to [0, 1]$ eine Gerade, $u(a) = 1 + m(a - z)$. (Warum?)
Zudem ist u beschränkt, $0 \leq u \leq 1$, daher folgt $m = 0$. Ebenso auf $\mathbb{Z}_{\leq z}$.

Zufällige Irrfahrt / Random Walk

D115
Erläuterung

(2) Sei $T_z \in \mathbb{N}$ die Zeit des ersten Besuchs im Zielpunkt z und $\mathbf{E}_a(T_z)$ die erwartete Reisezeit vom Startpunkt a zum Zielpunkt z.
Dies ist invariant unter Verschiebungen, also $\mathbf{E}_{a+k}(T_{z+k}) = \mathbf{E}_a(T_z)$.
Zunächst gilt $\mathbf{E}_a(T_a) = 0$. Für $a \neq z$ zeigen wir nun $\mathbf{E}_a(T_z) = \infty$:

$$w := \mathbf{E}_0(T_1) \ = 1 + \tfrac{1}{2}\big[\underbrace{\mathbf{E}_1(T_1)}_{=0} + \mathbf{E}_{-1}(T_1)\big]$$

$$\mathbf{E}_{-1}(T_1) = \underbrace{\mathbf{E}_{-1}(T_0)}_{=w} + \underbrace{\mathbf{E}_0(T_1)}_{=w} = 2w$$

Hieraus folgt $w = 1 + w$. Das ist für $w \in \mathbb{R}$ unmöglich.
Es bleibt nur $\mathbf{E}_0(T_1) = w = \infty$, also $\mathbf{E}_0(T_z) = \infty$ für alle $z \neq 0$.

🙂 Bei einer symmetrischen Irrfahrt ($p = 1/2$, ohne Drift) erreichen wir jeden Punkt mit Wkt 100%, aber die erwartete Reisezeit ist unendlich!
Die Rechnung ist einfach, dank unserer geschickten Formalisierung.
Die Interpretation hingegen muss man erst einmal verarbeiten.
Die naive Anschauung kann einen hier leicht narren.

Zufällige Irrfahrt / Random Walk

D116
Erläuterung

Die Irrfahrt ist ein einfaches, aber wichtiges Modell. Mögliche Anwendung: Kontostand bei zufälligen Gewinnen und Verlusten. Daher werden solche Modelle für **Aktienkurse** genutzt.

Ähnlich entsteht die **Brownsche Bewegung** durch Wärmebewegung. Der schottische Botaniker Robert Brown (1773–1858) entdeckte 1827 unter dem Mikroskop das unregelmäßige Zittern von Pollen in Wasser. Anfangs hielt er Pollen für belebt, doch er fand dasselbe bei Staubteilchen.

Albert Einstein erklärte die Zitterbewegung 1905 durch die ungeordnete Wärmebewegung der Wassermoleküle, die aus allen Richtungen in großer Zahl gegen die Pollen stoßen. Quantitativ konnte er so die Größe von Atomen bestimmen und die Anzahl pro Mol, die **Avogadro–Zahl**. Die präzisen quantitativen Vorhersagen wurden in den Folgejahren experimentell bestätigt.

Die Gerade finden wir durch vollständige Induktion: Aus $u(z) = 1$ und $u(z+1) = 1 + m$ folgt $u(a+1) = 2u(a) - u(a-1) = [2 + 2m(a-z)] - [1 + m(a-1-z)] = 1 + m(a+1-z)$.

🙂 Bei einer symmetrischen Irrfahrt ($p = 1/2$, ohne Drift) erreichen wir jeden Punkt mit Wkt 1! George Pólya (1887–1985) zeigte 1921: Jeden Punkt in $\mathbb{Z}$ besuchen wir mit Wkt 1 unendlich oft. Dies gilt ebenso in Dimension 2 bei Irrfahrt auf dem ebenen Gitter $\mathbb{Z}^2$. Erstaunlicherweise gilt es nicht mehr in Dimension $n \geq 3$ bei Irrfahrt auf dem Gitter $\mathbb{Z}^n$. Anschaulich bedeutet das: Ein betrunkener Mensch findet sicher irgendwann nach Hause, ein betrunkener Vogel hingegen nicht!

🕮 Ausführung bei Feller, *Introduction to Probability*, vol. 1 (1968), §XIV.7: Das Sprichwort „Alle Wege führen nach Rom.“ stimmt zumindest zweidimensional. Dreidimensional ist die Rückkehrwahrscheinlichkeit nur etwa 34%, siehe `en.wikipedia.org/wiki/Random_walk`.

Erwartete Reisezeit bis zum Rand

D117 Übung

Aufgabe: Wie lange ist die erwartete Reisezeit bis zum Rand?

0	1	0								
0	2	2	0							
0	3	4	3	0						
0	4	6	6	4	0					
0	5	8	9	8	5	0				
0	6	10	12	12	10	6	0			
0	7	12	15	16	15	12	7	0		
0	8	14	18	20	20	18	14	8	0	
0	9	16	21	24	25	24	21	16	9	0

Auf $X = \{0, 1, \dots, n\}$ vermuten wir $u(x) = x(n - x)$. Beweisen Sie dies!

Erwartete Reisezeit bis zum Rand

D118 Übung

Wir betrachten den Graphen $X = \{0, 1, \dots, n\}$ mit Rand $\partial X = \{0, n\}$ und lösen $u(x) = 1 + \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$ mit $u(0) = u(n) = 0$.
Wir vermuten, dass die Lösung eindeutig ist. Gilt das? Satz D1A!

☺ Im endlichen Fall genügt unsere einfache Bilanzgleichung.

Wie immer ist es hilfreich, zunächst kleine Beispiele zu betrachten.
Die kleinen Fälle für $n = 2, 3, \dots, 10$ lösen Sie leicht per Hand,
als lineares Gleichungssystem oder Probieren & Prüfen.

☺ Steht das Ergebnis erst einmal da, so ist es leicht zu prüfen!

Für jedes n beschreiben die Werte $x \mapsto u(x)$ eine Parabel:
Dies sehen Sie am leichtesten, wenn sie die Differenzen betrachten
(diskrete erste Ableitung) und diese als lineare Funktionen erkennen.

☺ Damit haben Sie die Formel gefunden, die Sie nun beweisen wollen.

Damit lösen Sie die eindimensionalen Spiele auf dem diskreten Intervall $X = \{0, 1, \dots, n\}$: (0) lineare Erwartung plus (1) quadratische Reisezeit.

Erwartete Reisezeit bis zum Rand

D119 Übung

Beweis: Auf $X = \{0, 1, \dots, n\}$ betrachten wir die Funktion

$$u : X \to \mathbb{R} : x \mapsto x(n - x).$$

Wir finden:

$$u(x-1) = xn - n - x^2 + 2x - 1$$
$$u(x+1) = xn + n - x^2 - 2x - 1$$

Also erfüllt u die geforderte Differenzengleichung:

$$u(x) = 1 + \tfrac{1}{2}u(x-1) + \tfrac{1}{2}u(x+1)$$

Dies ist also *eine* Lösung, und dank Satz D1A zudem die einzige.

Beispiele: Für $r, s \in \mathbb{N}$ und $n = r + s$ gelten folgende schöne Formeln:

☺ Die erwartete Reisezeit von x bis $\{x - r, x + s\}$ beträgt rs.
☺ Diese Formel besteht weiter sogar für $r = \infty$ oder $s = \infty$.
☺ Die erwartete Reisezeit von x bis $\{x - r, x + r\}$ beträgt r^2.

Erwartete Reisezeit bis zum Rand

D120 Übung

Diese Rechnung liefert uns einen zweiten und unabhängigen Beweis für die erwartete Reisezeit $u(x)$ in $X = \mathbb{N}$ bis zum Rand $\partial X = \{0\}$.
Wir haben $u(x) \geq x(n - x)$ für jedes $n \geq x$, also

$$u(x) = \begin{cases} 0 & \text{falls } x = 0, \\ \infty & \text{falls } x > 0. \end{cases}$$

Es ist hilfreich, neben Funktionen $u : X \to \mathbb{R}$ auch $u : X \to \mathbb{R} \cup \{+\infty\}$ oder $u : X \to \mathbb{R} \cup \{-\infty\}$ zuzulassen. Diese treten natürlich auf, wie hier.

Mittelwertgleichungen wir $u(x) = 1 + \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1)$ bleiben sinnvoll, solange niemals $+\infty$ und $-\infty$ addiert werden müssen.

Die Erweiterung von $\mathbb{R}$ zu $\mathbb{R} \cup \{+\infty\}$ bzw. $\mathbb{R} \cup \{-\infty\}$ nützt für monotone Grenzwerte wie $u_0 \leq u_1 \leq u_2 \leq \dots \nearrow u$ oder $u_0 \geq u_1 \geq u_2 \geq \dots \searrow u$, und sei es nur als technisches Hilfsmittel für Zwischenrechnungen.

☺ Bilanzgleichungen funktionieren wunderbar auf *endlichen* Graphen.
Auf unendlichen Graphen benötigen wir zusätzliche Bedingungen.
Manchmal können wir durch endliche Teilgraphen ausschöpfen.

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D121 Übung

0	1	2	3	4	5	6	7	8
7€								23€

9	
10	
11	5€

Aufgabe: Selbes Spiel wie zuvor, aber auf einem neuem Spielbrett.
Wie viel würden Sie als Teilnehmer zahlen / als Anbieter verlangen?
(0) Was ist die Gewinnerwartung $u(x)$ für jedes Startfeld $x \in X$?
(1) Jeder Zug kostet, $c = -1$€, und Sie müssen zu Ende spielen.
(2) Jeder Zug kostet, $c = -1$€, und Sie dürfen jederzeit aufgeben.
Schätzen Sie zunächst! Wie treffsicher ist Ihre intuitive Erwartung?
Formulieren Sie allgemeine Gleichungen und Lösungsmethoden!

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D122 Übung

Lösung: (0) Gewinnerwartung ohne Zugkosten:

7	8	9	10	11	14	17	20	23

9
7
5

Wir setzen hierzu $E = \{\, u : X \to \mathbb{R} \mid u(0) = 7,\ u(8) = 23,\ u(11) = 5 \,\}$ und $\Phi_0 : E \to E : u \mapsto \bar{u}$ mit $\bar{u}(x) = \sum_y p(x, y)u(y)$ und Übergangswkt $p(x, y) \in [0, 1]$ von Feld x auf das Nachbarfeld y, wobei $\sum_y p(x, y) = 1$.

Die obigen Werte sind der eindeutige Fixpunkt von Φ_0, also die Lösung der Gleichung $\Phi_0(u) = u$. Dieses lineare Gleichungssystem können Sie exakt lösen, mit den Mitteln der Linearen Algebra, oder iterativ annähern durch Banachs Fixpunktsatz, mit den Mitteln der Analysis.

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D123 Übung

(1) Erwartete Reisezeit bis zum Rand:

0.00	6.30	10.60	12.90	13.20	12.90	10.60	6.30	0.00

10.80
6.40
0.00

Daraus erhalten wir folgende Gewinnerwartung:

7.00	1.70	−1.60	−2.90	−2.20	1.10	6.40	13.70	23.00

−1.80
0.60
5.00

Wir setzen $E = \{\, u : X \to \mathbb{R} \mid u(0) = 7,\ u(8) = 23,\ u(11) = 5 \,\}$
und $\Phi_0 : E \to E : u \mapsto \bar{u}$ mit $\bar{u}(x) = c(x) + \sum_y p(x, y)u(y)$
und lösen die Fixpunktgleichung $\Phi_0(u_0) = u_0$.

Irrfahrt, Gewinnerwartung und optimale Entscheidung

D124 Übung

(2) Gewinnerwartung mit Zugkosten und Abbruchmöglichkeit:

7.00	2.67	0.33	0.00	0.00	2.20	6.40	13.70	23.00

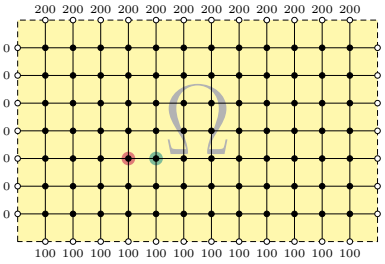
0.00
1.50
5.00

☺ Die Lösung u_0 im linearen Fall, ohne Entscheidung, ist leicht.
⚠ Die Lösung mit Entscheidungsmöglichkeit ist nicht $\max\{0, u_0\}$!
Diese naive Fehlannahme führt tatsächlich zu Fehlentscheidungen.
☺ Wir erhalten sie vielmehr als Fixpunkt von $\Phi(u) = \max\{0, \Phi_0(u)\}$.
Diesen Operator können wir zur Iteration nutzen, wie hier gezeigt.
☺ In der Übung zeigen Sie allgemein, dass $\Phi : E \to E$ kontraktiv ist, so dass Sie Banachs Fixpunktsatz anwenden können. Alles wird gut.
Damit können Sie u berechnen und Ihre optimale Strategie ablesen!

Irrfahrten und Potentiale: das Dirichlet–Problem

D125 Erläuterung

Zufällige Irrfahrt auf einem Spielfeld $\Omega \subseteq \mathbb{Z}^2$:



Sie ziehen mit Wkt $1/4$ nach links / rechts / oben / unten.

Das Spiel endet am Rand mit dem gezeigten Gewinn.

Für welche Startpunkte würden Sie 100 zahlen?

Aufgabe: (1) Wie groß ist die Gewinnerwartung $u(x,y)$ auf jedem Feld? Wo ist sie maximal? Ist die gesuchte Lösung $u : \Omega \to \mathbb{R}$ eindeutig? Wie berechnet man sie? möglichst effizient? näherungsweise?

☺ Kontext und Anwendung ändern sich, die Rechnung bleibt dieselbe! (2) Hooke: Netz aus Massen und Federn. (3) Kirchhoff: Spannung einer elektrischen Schaltung. (4) Fourier: diskrete Wärmeleitung / Diffusion.

Irrfahrten und Potentiale: das Dirichlet–Problem

D126 Erläuterung

Lösung: (1) Sei $u(x,y)$ die Gewinnerwartung auf dem Feld $(x,y) \in \Omega$. In jedem Randpunkt $(x,y) \in \partial\Omega$ ist der Gewinn $u(x,y)$ fest vorgegeben. In jedem inneren Punkt $(x,y) \in \Omega^\circ$ gilt die **Mittelwerteigenschaft**:

$$u(x,y) = \tfrac{1}{4}u(x-1,y) + \tfrac{1}{4}u(x+1,y) + \tfrac{1}{4}u(x,y-1) + \tfrac{1}{4}u(x,y+1)$$

Eine solche diskrete Funktion $u : \mathbb{Z}^2 \supset \Omega \to \mathbb{R}$ nennen wir **harmonisch**.

☺ Die folgende Tabelle und Graphik zeigen die / eine Lösung u.

Wir betrachten hier das ebene Gitter $\mathbb{Z}^2$ und darin eine Teilmenge $\Omega \subseteq \mathbb{Z}^2$. Innere Punkte $z \in \Omega$ sind solche, deren vier direkte Nachbarn ebenfalls in Ω liegen. Bei einem Randpunkt $z \in \Omega$ liegt mindestens ein Nachbar außerhalb von Ω. **Dirichlet–Problem**: In jedem Randpunkt $z \in \partial\Omega$ ist der Wert $u(z)$ festgelegt durch die vorgegebene Randfunktion $v = u|_{\partial\Omega} : \partial\Omega \to \mathbb{R}$. Gesucht ist eine harmonische Funktion $u : \Omega \to \mathbb{R}$ mit $u|_{\partial\Omega} = v$. Existiert eine Lösung? Ist sie eindeutig? Wie können wir sie berechnen bzw. annähern? Kurzum: Ist das Problem **gut gestellt**?

Diese Anwendung ist recht faszinierend: Sie ist physikalisch-anschaulich motiviert und fördert sowohl Intuition als auch Methodik. Hier gilt das Minimum-Maximum-Prinzip (Satz D1A) und daraus können wir nicht nur die Eindeutigkeit, sondern sogar die Existenz einer Lösung ableiten. Diese einfache Differenzengleichung diskretisiert partielle Differentialgleichungen und enthüllt so unerwartete Zusammenhänge, als Modell / Näherung / Anschauung für harmonische Funktionen $\Delta u = 0$ und die Wärmeleitungsgleichung $\partial_t u = \kappa\,\Delta u$.

Irrfahrten und Potentiale: das Dirichlet–Problem

D127 Erläuterung

Dieses einfache Beispiel illustriert das allgemeine und überall wichtige **Dirichlet–Problem**.

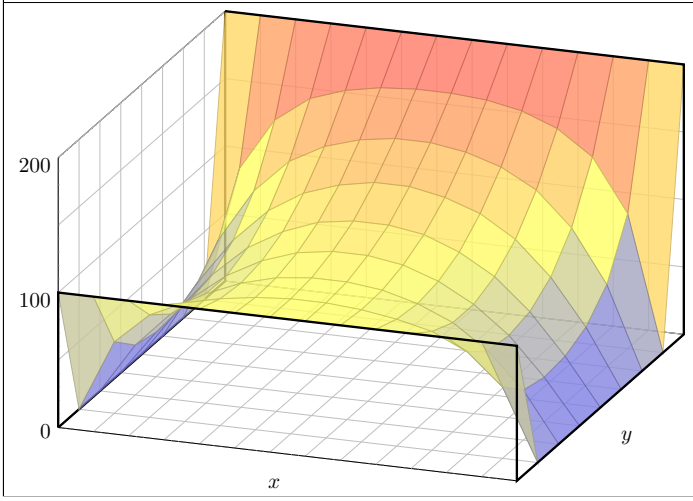
	200	200	200	200	200	200	200	200	200	200	200	200	
000	100	139	158	167	172	174	174	172	167	157	139	100	000
000	061	100	125	139	147	151	151	147	139	125	100	061	000
000	043	077	102	118	127	132	132	127	118	102	077	043	000
000	035	065	088	103	113	117	117	113	103	088	065	035	000
000	033	061	081	095	104	108	108	104	095	081	061	033	000
000	037	063	081	092	099	102	102	099	092	081	063	037	000
000	053	076	088	094	098	100	100	098	094	088	076	053	000
	100	100	100	100	100	100	100	100	100	100	100	100	

Die Aufgabe führt uns zu einem **linearen Gleichungssystem** mit $7 \times 12 = 84$ Unbekannten. Für diese haben wir genau 84 Gleichungen. Das sieht vernünftig aus, bedeutet aber noch nicht, dass es genau eine Lösung gibt. Hierzu müssen wir genauer hinschauen und präzise begründen!

Meine Tabellenkalkulation *LibreOffice* weigert sich zunächst mit „Fehler 522: zirkulärer Bezug“. Mit Extras > Optionen > Calc > Iterationen rechnet sie dann iterativ obige Näherungslösung aus.

Irrfahrten und Potentiale: das Dirichlet–Problem

D128 Erläuterung



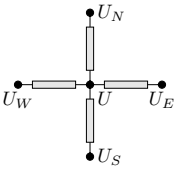
Harmonische Funktionen: Gleichgewicht!

D129 Erläuterung

☺ Kontext und Anwendung ändern sich, die Rechnung bleibt dieselbe!

(2) Wir betrachten Massenpunkte in $(x,y,u(x,y)) \in \mathbb{R}^3$ in Ruhelage. Jeder ist durch gleich starke Federn mit seinen Nachbarn verbunden. Somit gilt: Ruhelage $=$ Kräftegleichgewicht $\approx$ Mittelwerteigenschaft!

Sie können es nachrechnen! Genauer gesagt ist dies die Näherung bei geringer Krümmung.



(3) Wir betrachten die gezeigte Schaltung mit vier gleichen Widerständen. An den Nachbarpunkten liegen die Potentiale U_E, U_N, U_W, U_S an.

Ohmsches Gesetz und Kirchhoffsche Regel:

$$U = \frac{U_E + U_N + U_W + U_S}{4}$$

Ausführlich: Es gilt das Ohmsche Gesetz $I_E = (U_E - U)/R$. Die Kirchhoffsche Regel besagt hier $I_E + I_N + I_W + I_S = 0$. Einsetzen und Auflösen nach U ergibt die Mittelwerteigenschaft!

☺ Wir können Ω als Schaltung realisieren und am Rand die genannten Spannungen anlegen. Mit einem Voltmeter messen wir das Potential $u(x,y)$ im Inneren und finden obige Lösung. Physikalische Intuition suggeriert Existenz und Eindeutigkeit der Lösung, siehe Satz D1A.

Wärmeleitungsgleichung: Alles fließt!

D130 Erläuterung

(4) Wir betrachten Ω als ein Bauteil aus wärmeleitendem Material. Am Rand liegen die konstanten Temperaturen $0°\mathrm{C}$, $100°\mathrm{C}$, $200°\mathrm{C}$ an. Wärme fließt von warm nach kalt proportional zur Temperaturdifferenz. Stationäre Verteilung $=$ Fließgleichgewicht $=$ Mittelwerteigenschaft!

☺ Unser LGS ist die diskrete Version der Potentialgleichung $\Delta u = 0$. Für $u : \mathbb{R}^2 \to \mathbb{R}$ zweimal stetig differenzierbar gilt die Taylor–Formel:

$$0 = \underbrace{u(x-1,y) - 2u(x,y) + u(x+1,y)}_{= \partial_x^2 u(x,y) \text{ + h.o.t.}} + \underbrace{u(x,y-1) - 2u(x,y) + u(x,y+1)}_{= \partial_y^2 u(x,y) \text{ + h.o.t.}}$$

Wir diskretisieren die Wärmeleitungsgleichung $\partial_t u = \kappa\,\Delta u$ für $\kappa = \frac{1}{4}$:

$$u(t+1,x,y) = \frac{u(t,x-1,y) + u(t,x+1,y) + u(t,x,y-1) + u(t,x,y+1)}{4}$$

☺ Diese Iteration nähert sich der stationären Lösung u mit $\Delta u = 0$.

Diese Beobachtung führt uns schließlich zur Methode der iterativen Näherung. Dieses Verfahren beginnt mit einer (beliebigen!) Startverteilung und konvergiert schnell gegen die (eindeutige!) stationäre Lösung. Das ist für den Computer einfach zu rechnen, siehe obige Tabellenkalkulation.

Harmonische Funktionen: Minimum-Maximum-Prinzip

D131 Erläuterung

Wie zuvor sei $\Omega \subset \mathbb{Z}^n$ ein endlicher Graph. Wir zerlegen $\Omega = \Omega^\circ \sqcup \partial\Omega$: Jeder innere Punkt $z \in \Omega^\circ$ hat alle seine $2n$ Nachbarn $z \pm e_k$ in Ω; jeder Randpunkt $z \in \partial\Omega$ hat mindestens einen Nachbarn außerhalb von Ω. Wir nennen $u : \Omega \to \mathbb{R}$ harmonisch, wenn die Mittelwerteigenschaft gilt.

Satz D1A: Minimum-Maximum-Prinzip, Eindeutigkeit & Existenz

(1) Jede harmonische Funktion $u : \Omega \to \mathbb{R}$ nimmt Minimum und Maximum am Rand $\partial\Omega$ an: Es gilt also $\min_\Omega u = \min_{\partial\Omega} u$ und $\max_\Omega u = \max_{\partial\Omega} u$.

Für je zwei harmonische Funktionen $u, \tilde{u} : \Omega \to \mathbb{R}$ gilt demnach:

(2) Monotonie: Aus $u \le \tilde{u}$ auf dem Rand $\partial\Omega$ folgt $u \le \tilde{u}$ auf ganz Ω.

(3) Eindeutigkeit: Aus $u = \tilde{u}$ auf dem Rand $\partial\Omega$ folgt $u = \tilde{u}$ auf ganz Ω.

Hieraus gewinnen wir die Existenz und sogar eine Konstruktion:

(4) Existenz: Zu jeder vorgegebenen Randfunktion $v : \partial\Omega \to \mathbb{R}$ existiert genau eine harmonische Funktion $u : \Omega \to \mathbb{R}$ mit $u|_{\partial\Omega} = v$.

(5) Konstruktion: Die diskrete Wärmeleitung liefert zu v Näherungen $u_0, u_1, u_2, \ldots \to u$, die gegen die eindeutige Lösung u konvergieren.

Harmonische Funktionen: Eindeutigkeit und Existenz

D132 Erläuterung

Aufgabe: Beweisen Sie (1-4) mit Linearer Algebra. **Lösung:** Die Mittelwerteigenschaft (MWE) ist ein lineares Gleichungssystem: $u(z) = \frac{1}{2n}\sum_{k=1}^n u(z+e_k) + u(z-e_k)$ für alle $z \in \Omega^\circ$.

(1) Sei $u : \Omega \to \mathbb{R}$ harmonisch. Da die Menge Ω endlich ist, existiert eine Minimalstelle $z \in \Omega$, das heißt $u(z) \le u$. Liegt z im Inneren, so sind alle Nachbarn $z \pm e_k$ ebenfalls Minimalstellen. Es gibt einen Weg von z zu einem Randpunkt $z' \in \partial\Omega$, also ist auch dieser eine Minimalstelle.

(2) Die Differenz $v = \tilde{u} - u$ ist harmonisch und erfüllt $v \ge 0$ auf $\partial\Omega$, nach (1) also $v \ge 0$ auf Ω. Daraus folgt sofort (3). Alternativ: $v = \tilde{u} - u$ erfüllt $v = 0$ auf $\partial\Omega$, nach (1) also $v = 0$ auf Ω.

(4) Kurze Antwort: Wir haben ein Gleichungssystem $Sw = b$ mit $N = |\Omega^\circ|$ Unbestimmten und ebenso vielen Gleichungen: Die Matrix S ist quadratisch und dank (3) invertierbar. Ausführlich:

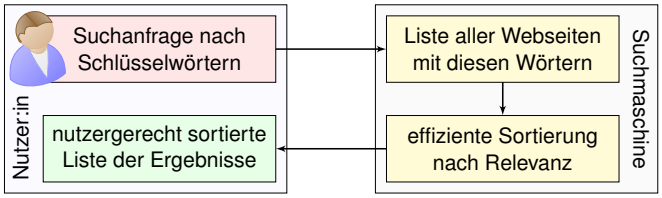
Harmonische Funktionen $u : \Omega \to \mathbb{R}$ sind der Kern der linearen Abbildung $\Lambda : \mathbb{R}^\Omega \to \mathbb{R}^{\Omega^\circ}$ mit $\Lambda(u)(x) := u(x) - \frac{1}{2n}\sum_{k=1}^n u(x+e_k) + u(x-e_k)$. Vorgegeben ist die Randfunktion $v : \partial\Omega \to \mathbb{R}$ als Bedingung $u|_{\partial\Omega} = v$. Zu festem v suchen wir alle harmonischen Ergänzungen $w = u|_{\Omega^\circ} : \Omega^\circ \to \mathbb{R}$. Dazu zerlegen wir $\mathbb{R}^\Omega \cong \mathbb{R}^{\partial\Omega} \times \mathbb{R}^{\Omega^\circ} : u \rightleftarrows (v,w)$, also $v = u|_{\partial\Omega}$ und $w = u|_{\Omega^\circ}$ sowie $u = v \sqcup w$, und betrachten $\Lambda(v,w) := \Lambda(v \sqcup w)$. Dank Linearität gilt dann $\Lambda(v,w) = \Lambda(v,0) + \Lambda(0,w)$. Die Abbildungen $R = \Lambda(-,0) : \mathbb{R}^{\partial\Omega} \to \mathbb{R}^{\Omega^\circ} : v \mapsto \Lambda(v,0)$ und $S = \Lambda(0,-) : \mathbb{R}^{\Omega^\circ} \to \mathbb{R}^{\Omega^\circ} : w \mapsto \Lambda(0,w)$ sind linear. Zu $v \in \mathbb{R}^{\partial\Omega}$ suchen wir $w \in \mathbb{R}^{\Omega^\circ}$ mit $Rv + Sw = 0$. Dank (3) ist S injektiv, also bijektiv. Daher gilt $w = -S^{-1}Rv$.

(5) Hierzu liefert die Analysis den Fixpunktsatz von Banach (D2A). All unsere Argumente gelten allgemein für jeden endlichen Graphen mit Übergangswkten auf den Kanten, also eine endliche Markov–Kette. Genau dies nutzt Googles PageRank, wie im Folgenden erläutert.

Die Mathematik hinter Google

D133 Erläuterung

„Wo simmer denn dran? Aha, heute krieje mer de Suchmaschin.
Wat is en Suchmaschin? Da stelle mer uns janz dumm. . . . “



Mathematik: Wie misst man Relevanz von Informationen?
Artificial Intelligence (AI), Machine Learning (ML), . . .

Informatik: Wie verarbeitet man enorm große Datenmengen?
Big Data, Data Mining, Data Science, . . . „Data is the new oil.“

Finanzstrategie: Wie verdient man Geld mit einem Gratisprodukt?
„If you’re not paying for it, you’re not the customer, you are the product.“

Die Mathematik hinter Google

D134 Erläuterung

Als das World Wide Web Mitte der 1990er noch klein war, da genügte es, zu einer Suchanfrage einfach alle Treffer aufzulisten. Die Liste war noch kurz, jede:r Nutzer:in konnte sie leicht selbst überblicken. Das Internet blieb jedoch nicht lange so überschaubar. . . . Das Volumen explodierte! Als Versuch einer Lösung ging 1998 die Suchmaschine Google in Betrieb und dominiert seither den Markt. Sie wird ständig weiterentwickelt. Die meisten Optimierungen hütet Google streng als Firmengeheimnis, doch das ursprüngliche Grundprinzip ist veröffentlicht und genial einfach:

📖 Sergey Brin, Larry Page: *The anatomy of a large-scale hypertextual web search engine.* Stanford University 1998, `infolab.stanford.edu/pub/papers/google.pdf`

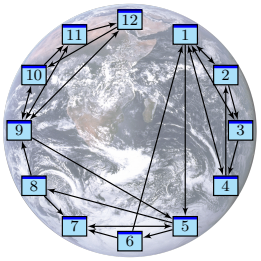
Bei vorherigen Suchmaschinen musste man endlose Trefferlisten durchforsten, bis man auf die ersten interessanten Ergebnisse stieß. Bei Google stehen sie auf wundersame Weise ganz oben. Wie ist das möglich? Die Antwort liegt (zu einem großen Teil) in folgender genial-einfachen Idee. Google misst die Popularität p_i (*PageRank*) jeder Seite i durch folgendes Gleichungssystem:

$$\text{PageRank}\quad p_i = \frac{q}{N} + \sum_{j\to i} \frac{1-q}{\ell_j}\, p_j$$

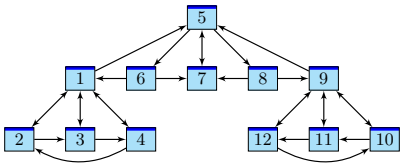
Keine Angst, die Formel sieht nur auf den ersten Blick kompliziert aus. Ich werde sie anhand von Beispielen Schritt für Schritt erläutern. Wer sowas schon gesehen hat, weiß, dass es sich um eine besonders einfache Formel handelt, nämlich ein *lineares Gleichungssystem*, das keine Quadrate oder komplizierteres enthält. Schon die Formel von Pythagoras $a^2 + b^2 = c^2$ ist komplizierter.

Chaos und Struktur im Web

D135 Erläuterung



Miniaturbeispiel des Web als ein Graph aus Seiten $i = 1, \dots, N$ und Links $i \to j$.
Versuch einer hierarchischen Anordnung:



Eine Seite ist populär, wenn viele Seiten auf sie verweisen? Zu naiv!
Eine Seite ist populär, wenn viele populäre Seiten auf sie verweisen.

Ein zufälliger Surfer folgt von der aktuellen Seite zufällig einem der Links.

Aufgabe: Berechnen Sie die Aufenthaltswkten. Konvergieren sie gegen ein Gleichgewicht? Wie schnell? Immer dasselbe, d.h. ist es eindeutig?

☺ Im Rückblick ist die abstrakt-mathematische Idee genial einfach.
Wer diese Aufgabe bis 1998 professionell löste, ist heute Milliardär.

Chaos und Struktur im Web

D136 Erläuterung

Klassische Texte sind von einer Person geschrieben und linear: Ein Buch hat einen Anfang und ein Ende, typischerweise liest man es von vorne nach hinten in der Reihenfolge der Seiten. Meist gibt es zudem ein Inhaltsverzeichnis oder einen Index zum leichteren Nachschlagen. (Ja, liebe Kinder, unsere Vorfahren konnten Texte mit hunderttausend Buchstaben am Stück lesen, ohne Clicks und ohne Werbung. Man nannte das „Buch“ und speicherte es auf „Papier“. Damals!)

Webseiten bilden hingegen eine gänzlich andere Struktur. Niemand käme auf die Idee, das Internet von Anfang bis Ende durchzulesen: Es hat keine lineare Struktur, keine erste und keine letzte Seite, es ist zudem viel zu groß, und das meiste ist ohnehin uninteressant.

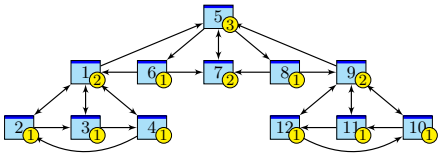
Die Webseiten verweisen gegenseitig aufeinander und bilden einen *Hypertext*. Zur Illustration betrachten wir ein Miniaturbeispiel bestehend aus 12 Webseiten. Unter den Seiten $1, 2, 3, 4$ wird 1 am häufigsten zitiert. Die Seite 1 scheint daher besonders relevant oder populär. Gleiches gilt für $9, 10, 11, 12$ mit 9 an der Spitze. Die Struktur von $5, 6, 7, 8$ ist ähnlich mit 7 an der Spitze. Aber die Seiten $1, 7, 9$, die wir schon als relevant erkannt haben, verweisen alle auf die Seite 5. Diese scheint daher populär / wichtig / zentral und für eine spätere Suche besonders relevant.

Diese Anordnung war Handarbeit. Lässt sie sich automatisieren? Nach welchen Regeln?
Erster Versuch einer Bewertung: Eine Seite ist populär, wenn viele Seiten auf sie verweisen.
Nachteil: Die simple Linkzählung ist zu naiv und anfällig für Manipulationen! (Linkfarmen)

Zweiter Versuch: Eine Seite ist populär, wenn viele populäre Seiten auf sie verweisen.
Das klingt zunächst zirkulär, lässt sich aber in eine einfache Gleichung fassen und lösen.
Ich erläutere dazu die besonders anschauliche Betrachtungsweise des zufälligen Surfers.

Irrfahrt eines zufälligen Surfers

D137 Erläuterung

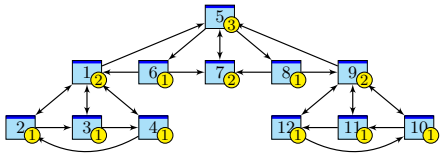


Googles Heuristik: Aufenthaltswkt ∼ Popularität ∼ Relevanz
Aufgabe: Berechnen Sie die Aufenthaltswkten bei Start auf Seite 7.

	1	2	3	4	5	6	7	8	9	10	11	12
$t=0$	.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t=1$	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
$t=2$	.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000
$t=3$	.167	.000	.000	.000	.333	.000	.333	.000	.167	.000	.000	.000
$t=4$	.000	.042	.042	.042	.417	.111	.111	.111	.000	.042	.042	.042
$t=5$	.118	.021	.021	.021	.111	.139	.250	.139	.118	.021	.021	.021
...												
$t=29$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t=30$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Irrfahrt eines zufälligen Surfers

D138 Erläuterung



Googles Heuristik: Aufenthaltswkt ∼ Popularität ∼ Relevanz
Aufgabe: Berechnen Sie die Aufenthaltswkten bei Start auf Seite 1.

	1	2	3	4	5	6	7	8	9	10	11	12
$t=0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t=1$	.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
$t=2$	.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000
$t=3$	.229	.156	.156	.156	.177	.000	.083	.000	.042	.000	.000	.000
$t=4$	.234	.135	.135	.135	.151	.059	.059	.059	.000	.010	.010	.010
$t=5$	.233	.126	.126	.126	.118	.050	.109	.050	.045	.005	.005	.005
...												
$t=69$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t=70$	.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Zufall und Notwendigkeit: Googles PageRank

D139 Erläuterung

Wir beobachten eine Diffusion: Sie konvergiert gegen eine stationäre Gleichgewichtsverteilung! Ebenso beim Start in 1; sie konvergiert langsamer, aber schließlich zum selben Gleichgewicht! Dank dieser Betrachtungsweise löst sich unser LGS sozusagen von allein! Verfeinertes Modell:

Sprung: Mit Wkt q startet unser Surfer neu auf irgendeiner zufälligen Seite $i \in \{1, \dots, N\}$.
Fluss: Mit Wkt $(1-q)$ folgt er von der aktuellen Seite j zufällig irgendeinem der ℓ_j Links.
Dies führt zu folgenden Gleichungen, analog zur Wärmeleitung bzw. Potentialgleichung: D125

$$\text{Diffusion}\qquad p_i(t+1) = \frac{q}{N} + \sum_{j\to i} \frac{1-q}{\ell_j}\, p_j(t)$$

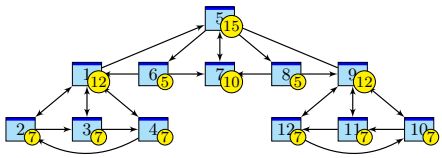
$$\text{Gleichgewicht}\qquad p_i = \frac{q}{N} + \sum_{j\to i} \frac{1-q}{\ell_j}\, p_j$$

Dieses verfeinerte Modell mit Teleportation lässt sich ebenso leicht berechnen. Für $q = 0.15$ entspricht es dem typischen Verhalten, sechs bis sieben Links zu folgen, bevor man neu anfängt.
☺ Die Ergebnisse entsprechen der Nutzererwartung und sind recht robust gegen Manipulationen.
☺ Unsere obigen Beobachtungen zur Konvergenz sind nicht bloß zufällig, sondern beruhen auf mathematischen Gesetzmäßigkeiten. Diese kann man beweisen und darf sich darauf verlassen:

Der **Fixpunktsatz von Banach** garantiert bei positiver Sprunghaftigkeit $0 < q \le 1$ Folgendes:
(1) Es gibt genau ein Gleichgewicht p. Dieses erfüllt $p_1, \dots, p_N > 0$ und $p_1 + \cdots + p_N = 1$.
(2) Für jede Anfangsverteilung konvergiert die Diffusion gegen die Gleichgewichtsverteilung p.
(3) Die Konvergenz ist mindestens so schnell wie die der geometrischen Folge $(1-q)^n \searrow 0$.

Zufall und Notwendigkeit: Googles PageRank

D140 Erläuterung



Googles Heuristik: Aufenthaltswkt ∼ Popularität ∼ Relevanz
Aufgabe: Aufenthaltswkten bei Sprunghaftigkeit $q = 0.15$:

	1	2	3	4	5	6	7	8	9	10	11	12
$t=0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t=1$	.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013
$t=2$	.305	.111	.111	.111	.028	.076	.087	.076	.034	.020	.020	.020
$t=3$	.186	.124	.124	.124	.158	.021	.085	.021	.071	.028	.028	.028
$t=4$	.180	.105	.105	.105	.140	.057	.075	.057	.057	.040	.040	.040
$t=5$	.171	.095	.095	.095	.126	.052	.101	.052	.087	.042	.042	.042
...												
$t=29$	.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066
$t=30$	.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066

Banachs Fixpunktsatz ist ein universelles Werkzeug.

D201 Erläuterung

Aufgabe: (1) Illustrieren Sie Iterationen und Banachs Fixpunktsatz.
(2) Sie kennen bereits spektakuläre Anwendungen: Nennen Sie einige!
(3) Wiederholung: Formulieren und beweisen Sie Banachs Fixpunktsatz.

Lösung: (1) Bildhaftes Beispiel: Wenn Sie von Ihrer geographischen Umgebung X eine Landkarte im Maßstab $k = 1 : n$ (mit $n > 1$) vor sich auf den Tisch legen, dann definiert die Zuordnung jedes realen Punktes zu seinem Bildpunkt eine k–kontraktive Abbildung $f : X \to X$. Genau ein Punkt der Karte liegt auf dem geographischen Punkt, den er bezeichnet.

(2) Mit dem Iterationsverfahren können Sie viele Gleichungen numerisch lösen wie $x = \cos(x)$ für $x \in [0, 1]$. Das **Newton–Verfahren** baut darauf auf und verbessert ganz wesentlich die Konvergenzgeschwindigkeit.

Weitere Anwendungen sind der **Satz von Picard–Lindelöf** zur Lösung von Differentialgleichungen $y' = f(x, y)$ und der **lokale Umkehrsatz** zur Konstruktion lokaler Diffeomorphismen $(f, g) : \mathbb{R}^n \supseteq U \cong V \subseteq \mathbb{R}^n$.

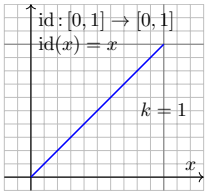
Das ist schon spektakulär – und doch nur die ersten Anwendungen im Grundstudium, als Spitze des Eisbergs! Wir fügen noch weitere hinzu.

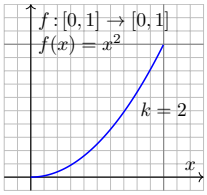
Illustration zu Banachs Fixpunktsatz

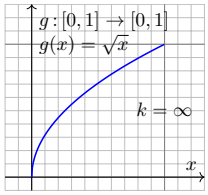
D202

Dehnungsschranken und Kontraktionen

D203 Erläuterung







Gegeben seien metrische Räume (X, d_X) und (Y, d_Y) und $k \in \mathbb{R}_{\geq 0}$. Eine Abbildung $f : X \to Y$ nennen wir **k–lipschitz–stetig**, falls gilt:

$$d_Y(f(a), f(b)) \leq k\, d_X(a, b) \qquad \text{für alle } a, b \in X$$

Im Falle $0 \leq k < 1$ nennen wir f **kontraktiv** oder eine **k–Kontraktion**.

Beispiel: Für $f : \mathbb{R}^n \to \mathbb{R}^m : x \mapsto Ax + b$ mit $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^n$ gilt $|f(x) - f(y)| = |A(x - y)| \leq \|A\| \cdot |x - y|$ bezüglich der Operatornorm.

Beispiel: Sei $X \subseteq \mathbb{R}^n$ konvex und $f : X \to \mathbb{R}^m$ diff'bar mit $\|f'\| \leq k$. Für alle $x, y \in X$ existiert $z \in [x, y]$, sodass $f(x) - f(y) = f'(z)(x - y)$. Demnach gilt $|f(x) - f(y)| \leq \|f'(z)\| \cdot |x - y| \leq k|x - y|$.

Dehnungsschranken und Kontraktionen

D204 Erläuterung

Wir nennen eine solche Funktion f auch **dehnungsbeschränkt**. Der **metrische Differenzenquotient** ist hier beschränkt gemäß

$$\frac{d_Y(f(a), f(b))}{d_X(a, b)} \leq k \qquad \text{für alle } a \neq b \text{ in } X.$$

Für $f : (X, d_X) \to (Y, d_Y)$ definieren wir daher die **Lipschitz–Norm**

$$\|f\| = \|f\|_{\mathrm{Lip}} = \mathrm{Lip}(f) := \sup\left\{ \frac{d_Y(f(a), f(b))}{d_X(a, b)} \;\middle|\; a \neq b \text{ in } X \right\}.$$

Die folgende Formulierung vereinheitlicht Ausnahmen und Sonderfälle:

$$\|f\| := \inf\left\{ k \in \mathbb{R}_{\geq 0} \;\middle|\; \forall a, b \in X : d_Y(f(a), f(b)) \leq k\, d_X(a, b) \right\}.$$

Dies entspricht der **Operatornorm** von linearen Abbildungen normierter Vektorräume. Genau dann gilt $\|f\| < \infty$, wenn f lipschitz–stetig ist.

Genau dann gilt $\|f\| = 0$, wenn f konstant ist. Speziell für die Identität $\mathrm{id}_X : X \to X$ gilt $\|\mathrm{id}_X\| = 1$, im Sonderfall $X = \{x\}$ jedoch nur $\|f\| = 0$.

Für die Komposition von Abbildungen gilt $\|g \circ f\| \leq \|g\| \cdot \|f\|$.

Banachs Fixpunktsatz

D205

Satz D2A: Fixpunktsatz von Banach, 1922

Sei (X, d) ein metrischer Raum, nicht-leer und vollständig. Hierauf sei $f : X \to X$ eine k–Kontraktion: Wir haben eine Kontraktionskonstante $k \in [0, 1[$, und für alle $x, y \in X$ gilt $d(f(x), f(y)) \leq k\, d(x, y)$. Dann folgt:

(1) Zur Abbildung f existiert genau ein Fixpunkt $a \in X$, mit $f(a) = a$.

(2) Jede Iteration mit $x_0 \in X$ und $x_{n+1} = f(x_n)$ konvergiert gegen a.

(3) Für alle $n \in \mathbb{N}_{\geq 1}$ gelten dabei die beiden Fehlerschranken

$$d(a, x_n) \;\leq\; \underbrace{\frac{k}{1-k} d(x_n, x_{n-1})}_{\text{a posteriori}} \;\leq\; \underbrace{\frac{k^n}{1-k} d(x_1, x_0)}_{\text{a priori}} \;\searrow\; 0.$$

(4) Allgemeiner genügt $d(f^n(x), f^n(y)) \leq k_n\, d(x, y)$ für alle $x, y \in X$ und alle $n \in \mathbb{N}$ sowie $\sum_{n=0}^{\infty} k_n < \infty$. Damit gilt die feinere Fehlerschranke

$$d(a, x_n) \;\leq\; d(x_1, x_0) \sum_{j=n}^{\infty} k_j \;\searrow\; 0.$$

Banachs Fixpunktsatz: Erläuterung

D206 Erläuterung

📖 S. Banach: *Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales.* Fund. Math. 3 (1922) 133–181

Die Aussage (1) garantiert Existenz und Eindeutigkeit des Fixpunktes. Die Konstruktion (2) liefert zudem eine extrem praktische Approximation.

Gemäß (3) ist die Konvergenz $x_n \to a$ hierbei mindestens so schnell wie die Konvergenz der geometrischen Folge $k^n \searrow 0$: Dies nutzen wir bei iterativen Berechnungen als *a priori* Abschätzung des Zeitaufwandes.

Dieser wunderbare Satz geht auf Stefan Banach (1892–1945) zurück, der nützliche Zusatz (4) stammt von Johannes Weissinger (1913–1995).

Zum Beispiel genügt es, dass eine gewisse Iteration f^m kontraktiv ist, also $d(f^m(x), f^m(y)) \leq k\, d(x, y)$ für eine Konstante $k \in [0, 1[$ gilt.

Ebenso kommt es vor, dass höhere Iterationen f^n stärker kontrahieren, und die Reihe $\sum k_n$ somit viel kleiner ausfällt als die geometrische $\sum k^n$.

Für die qualitative Konvergenzaussage ist dies zunächst unwesentlich, doch für die praktische Fehlerabschätzung ist es ungemein hilfreich.

Banachs Fixpunktsatz: Beweis der Konvergenz

D207 Erläuterung

Beweis: *Eindeutigkeit:* Für je zwei Fixpunkte $a = f(a)$ und $b = f(b)$ gilt $d(a, b) = d(f(a), f(b)) \leq k\, d(a, b)$ mit $k < 1$, also $d(a, b) = 0$, somit $a = b$.

Die *Existenz* beweisen wir durch die iterative Folge $(x_n)_{n \in \mathbb{N}}$: Wir wählen $x_0 \in X \neq \emptyset$ und setzen $x_{n+1} = f(x_n)$ für alle $n \in \mathbb{N}$.

Per Induktion gilt $d(x_{n+1}, x_n) \leq k^n d(x_1, x_0)$: Für $n = 0$ ist dies trivial, für $n \geq 1$ gilt $d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq k\, d(x_n, x_{n-1}) \leq k^n d(x_1, x_0)$.

Dank Dreiecksungleichung erhalten wir für alle $n \leq p < q$:

$$\begin{aligned} d(x_q, x_p) &\leq d(x_q, x_{q-1}) + \cdots + d(x_{p+2}, x_{p+1}) + d(x_{p+1}, x_p) \\ &\leq (k^{q-p-1} + \cdots + k + 1)\, d(x_{p+1}, x_p) = \frac{1 - k^{q-p}}{1 - k} d(x_{p+1}, x_p) \\ &\leq \frac{k^p - k^q}{1 - k} d(x_1, x_0) \leq \frac{k^n}{1 - k} d(x_1, x_0) \;\searrow\; 0 \quad \text{für } n \to \infty. \end{aligned}$$

Demnach ist $(x_n)_{n \in \mathbb{N}}$ eine Cauchy–Folge im metrischen Raum (X, d). Da (X, d) vollständig ist, existiert ein Grenzwert $a \in X$ mit $x_n \to a$.

Da f kontraktiv und somit stetig ist, folgt aus $x_{n+1} = f(x_n)$ für alle $n \in \mathbb{N}$ per Grenzübergang $a = \lim x_{n+1} = \lim f(x_n) = f(\lim x_n) = f(a)$.

Banachs Fixpunktsatz: Beweis der Fehlerschranke

D208 Erläuterung

Die Ungleichung für $n = p$ und $q \to \infty$ ergibt die Fehlerabschätzung (3):

$$d(a, x_n) \;\leq\; \underbrace{\frac{k}{1-k} d(x_n, x_{n-1})}_{\text{a posteriori}} \;\leq\; \underbrace{\frac{k^n}{1-k} d(x_1, x_0)}_{\text{a priori}} \;\searrow\; 0$$

Aussage (4) beweisen wir genauso: Wegen $\sum_{n=0}^{\infty} k_n < \infty$ gilt $k_n \to 0$ für $n \to \infty$. Insbesondere existiert $m \in \mathbb{N}$ sodass $k_n \leq 1/2$ für alle $n \geq m$, das heißt f^n ist kontraktiv für alle $n \geq m$. Sind $a = f(a)$ und $b = f(b)$ Fixpunkte von f, so auch von f^n, also folgt $a = b$ wie oben.

Für jede iterative Folge mit $x_0 \in X$ und $x_{n+1} = f(x_n)$ erhalten wir für $n \leq p < q$ wie oben $d(x_q, x_p) \leq d(x_1, x_0) \sum_{j=n}^{\infty} k_j$. Somit ist $(x_n)_{n \in \mathbb{N}}$ eine Cauchy–Folge in (X, d), also existiert ein Grenzwert $a \in X$ mit $x_n \to a$.

Für $n = p$ und $q \to \infty$ erhalten wir die feinere Fehlerabschätzung

$$d(a, x_n) \leq d(x_1, x_0) \sum_{j=n}^{\infty} k_j \;\searrow\; 0.$$

Wie immer bei Konvergenz gilt auch hier: Ende gut, alles gut. QED

Erinnerung: normierte Vektorräume

D209 Erläuterung

Zur Erinnerung: Auf dem Raum $X = \mathbb{R}^n$ definieren wir für jedes $x \in \mathbb{R}^n$

die euklidische Norm $\quad |x|_2 := \sqrt{{x_1}^2 + {x_2}^2 + \cdots + {x_n}^2}$,

die Maximumsnorm $\quad |x|_\infty := \max\{|x_1|, |x_2|, \ldots, |x_n|\}$,

die Taxinorm $\quad |x|_1 := |x_1| + |x_2| + \cdots + |x_n|$.

Übung: Jede dieser Normen $x \mapsto |x|$ erfreuen sich folgender Eigenschaften für alle Vektoren $x, y \in X$ und Skalare $\lambda \in \mathbb{R}$:

N0: $|x| \geq 0 = |0|$ (Positivität)

N1: $|x| > 0$ für $x \neq 0$ (Definitheit)

N2: $|\lambda x| = |\lambda| \cdot |x|$ (Homogenität)

N3: $|x + y| \leq |x| + |y|$ (Dreiecksungleichung)

Definition D2B: Norm auf einem Vektorram

Eine **Norm** auf einem Vektorraum X ist eine Abbildung $|-| : X \to \mathbb{R}_{\geq 0}$, die (N0–3) erfüllt. Das Paar $(V, |-|)$ heißt dann **normierter Raum** oder **Prä–Banach–Raum**, und bei Vollständigkeit **Banach–Raum** (D2D).

Erinnerung: metrische Räume

D210 Erläuterung

Sei $(X, |-|)$ ein **normierter $\mathbb{R}$–Vektorraum**, etwa $X = \mathbb{R}^n$ mit einer der obigen Normen. Allgemeiner genügt eine **Pseudonorm** $|-| : X \to [0, \infty]$ mit Eigenschaften (N0–3). Die zugehörige **Metrik** misst den Abstand:

$$d \,:\, X \times X \to [0, \infty] \,:\, (x, y) \mapsto |x - y|$$

Übung: Für alle Punkte $x, y, z \in X$ gilt dann:

M0: $d(x, y) \geq 0 = d(x, x)$ (Positivität)

M1: $d(x, y) > 0$ für $x \neq y$ (Definitheit)

M2: $d(x, y) = d(y, x)$ (Symmetrie)

M3: $d(x, z) \leq d(x, y) + d(y, z)$ (Dreiecksungleichung)

Definition D2C: Metrik und metrischer Raum

Eine **Metrik** auf einer Menge X ist eine Abbildung $d : X \times X \to [0, \infty]$, die (M0–3) erfüllt. Das Paar (X, d) heißt dann ein **metrischer Raum**.

Für Metriken ist es bequem, die Möglichkeit $d(x, y) = \infty$ zuzulassen. Für Normen hingegen verlangen wir stets Endlichkeit, wie oben erklärt.

Vollständige metrische Räume

D211 Erläuterung

Eine Folge $(x_n)_{n\in\mathbb{N}}$ in (X, d) **konvergiert** gegen einen Punkt $a \in X$, wenn der Abstand $d(x_n, a)$ schließlich beliebig klein wird. Ausführlich:

$$(x_n)_{n\in\mathbb{N}} \to a \text{ in } (X, d) \quad :\Leftrightarrow \quad d(x_n, a) \to 0 \text{ für } n \to \infty$$
$$:\Leftrightarrow \quad \forall \varepsilon \in \mathbb{R}_{>0} \;\; \exists m \in \mathbb{N} \;\; \forall n \in \mathbb{N}_{\geq m} \,:\, d(x_n, a) < \varepsilon$$

Wir nennen dann die Folge $(x_n)_{n\in\mathbb{N}}$ **konvergent** und a ihren **Grenzwert**. Konvergenz in (X, d) ist eine **zweistellige Relation** zwischen Folgen $(x_n)_{n\in\mathbb{N}}$ und Punkten a in X. Wir schreiben hierfür kurz $x_n \to a$.

Wir nennen $(x_n)_{n\in\mathbb{N}}$ **Cauchy–Folge** in (X, d), wenn der Durchmesser $\delta_n := \sup_{p,q\geq n} d(x_p, x_q)$ eine Nullfolge ist. In Quantorenschreibweise:

$$\forall \varepsilon \in \mathbb{R}_{>0} \;\; \exists n \in \mathbb{N} \;\; \forall p, q \geq n \,:\, d(x_p, x_q) \leq \varepsilon$$

Jede konvergente Folge ist eine Cauchy–Folge, aber nicht umgekehrt.

Definition D2D: vollständiger metrischer Raum

Ein metrischer Raum (X, d) heißt **vollständig**, wenn jede Cauchy–Folge $(x_n)_{n\in\mathbb{N}}$ in (X, d) konvergiert, also ein Grenzwert $x_n \to a \in X$ existiert.

Vollständige metrische Räume

D212 Erläuterung

Beispiel: Im Raum $X =]0, 1]$ mit euklidischer Metrik ist $x_n = 2^{-n}$ eine Cauchy–Folge, aber nicht konvergent. (Im Raum $[0, 1]$ gilt $x_n \to 0$.)

Beispiel: Im Raum $\mathbb{Q}$ mit euklidischer Metrik ist $x_n = \sum_{k=0}^{n} 1/k!$ eine Cauchy–Folge, hat aber in $\mathbb{Q}$ keinen Grenzwert. (In $\mathbb{R}$ hingegen gilt $x_n \to \mathrm{e} = 2.71828\ldots$, aber dieser Grenzwert liegt nicht in $\mathbb{Q}$.)

Beispiel: Das Newton–Verfahren liefert eine effiziente Approximation von $\sqrt{5}$ durch eine rasch konvergente Folge $x_n \to \sqrt{5}$ *rationaler Zahlen*: Wir definieren $(x_n)_{n\in\mathbb{N}}$ rekursiv durch $x_0 = 3$ und $x_{n+1} = \frac{1}{2}(x_n + 5/x_n)$. Diese Folge ist eine Cauchy–Folge in $\mathbb{Q}$, sie konvergiert aber nicht in $\mathbb{Q}$. (In $\mathbb{R}$ gilt wie gewünscht $x_n \to \sqrt{5}$, aber dieser Wert liegt nicht in $\mathbb{Q}$.)

Der Raum $\mathbb{Q}$ ist unvollständig, hat also ganz anschaulich noch Lücken. Jede Cauchy–Folge möchte konvergieren, doch oft fehlt der Grenzwert. In einem vollständigen Raum kann dieses Problem niemals auftreten!

Beispiel: Die reellen Zahlen $\mathbb{R}$ sind vollständig bezüglich der Metrik $d(x, y) = |x - y|$, ebenso $\mathbb{R}^n$ bezüglich jeder beliebigen Norm und jede abgeschlossene Menge $X \subseteq \mathbb{R}^n$ bezüglich der eingeschränkten Metrik.

Vollständigkeit des Raumes $B(X, \mathbb{R})$

D213 Erläuterung

Sei X eine Menge. Für $u : X \to \mathbb{R}$ definieren wir die Supremumsnorm:

$$\|u\| = |u|_X := \sup\{\, |u(x)| \mid x \in X \,\}$$

Dies ist eine Norm auf dem Vektorraum der beschränkten Funktionen:

$$B(X, \mathbb{R}) := \{\, u : X \to \mathbb{R} \mid \|u\| < \infty \,\}$$

Hier steht „B“ für beschränkt [engl. *bounded*, frz. *borné*].

Satz D2E: Vollständigkeit von $B(X, \mathbb{R})$

Der $\mathbb{R}$–Vektorraum $B(X, \mathbb{R})$ ist vollständig, also ein Banach–Raum.

Beweis: Gegeben sei eine Cauchy–Folge $(u_n)_{n\in\mathbb{N}}$ in $B(X, \mathbb{R})$:
Zu $\varepsilon \in \mathbb{R}_{>0}$ existiert $m \in \mathbb{N}$ sodass für alle $p, q \geq m$ gilt $\|u_p - u_q\| \leq \varepsilon$.
Zu jedem Punkt $x \in X$ ist dann $u_n(x)_{n\in\mathbb{N}}$ eine Cauchy–Folge in $\mathbb{R}$.
Da $\mathbb{R}$ vollständig ist, existiert $u(x) \in \mathbb{R}$ als Grenzwert $u_n(x) \to u(x)$.
Für $p = m$ und $q \to \infty$ folgt $|u_m(x) - u(x)| \leq \varepsilon$, somit $\|u_m - u\| \leq \varepsilon$.
Also ist u beschränkt, und es gilt $u_n \to u$ in $B(X, \mathbb{R})$. QED

🙂 Somit ist jede abgeschlossene Teilmenge $E \subseteq B(X, \mathbb{R})$ vollständig.

Äquivalenz aller Normen auf $\mathbb{R}^n$

D214 Erläuterung

Im Spezialfall einer endlichen Menge X, etwa $X = n = \{0, 1, \ldots, n-1\}$, ist $B(X, \mathbb{R})$ unser Modellraum $\mathbb{R}^n$ mit der Maximumsnorm $\|-\| = |-|_\infty$. Zudem haben wir die Norm $|x|_p = \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p}$ für $1 \leq p < \infty$. Diese Normen sind äquivalent gemäß $|x|_\infty \leq |x|_p \leq |x|_1 \leq n|x|_\infty$, daher definieren sie dieselbe Topologie, Konvergenz, Cauchy–Folgen, etc.

Satz D2F: Äquivalenz aller Normen auf $\mathbb{R}^n$

Auf jedem $\mathbb{R}$–Vektorraum X endlicher Dimension sind je zwei Normen $|-|$ und $\|-\|$ äquivalent. Ausführlich bedeutet das: Es gibt positive Konstanten $\ell, L \in \mathbb{R}_{>0}$, sodass $\ell\,|x| \leq \|x\| \leq L\,|x|$ für alle $x \in X$ gilt.

Speziell für unseren Modellraum $X = \mathbb{R}^n$ mit euklidischer Norm $|-|$ ist die Sphäre $\mathbb{S}^{n-1} = \{\, x \in \mathbb{R}^n \mid |x| = 1 \,\}$ kompakt, und es genügen

$$\ell = \min\{\, \|x\| \mid x \in \mathbb{S}^{n-1} \,\} \quad \text{und} \quad L = \max\{\, \|x\| \mid x \in \mathbb{S}^{n-1} \,\}.$$

Übung: Beweisen Sie diese Schranken zur Wiederholung. Mahnendes Gegenbeispiel: Für die ℓ^p–Normen auf $\mathbb{R}^{(\mathbb{N})} \subset \ell^p(\mathbb{N}, \mathbb{R})$ gilt dies nicht!

🙂 Je nach Anwendung wählen wir eine bequem passende Norm.

Blackwells hinreichendes Kriterium

D215

Sei X eine Menge und sowie $E \subseteq B(X, \mathbb{R})$ mit Supremumsnorm $\|-\|$.

Für $u, \tilde{u} : X \to \mathbb{R}$ schreiben wir $u \leq \tilde{u}$ falls $u(x) \leq \tilde{u}(x)$ für alle $x \in X$ gilt.
Ein Operator $\Phi : E \to E$ ist **isoton** falls gilt: Aus $u \leq \tilde{u}$ folgt $\Phi(u) \leq \Phi(\tilde{u})$.

Sei $\delta \in [0, 1]$. Wir nennen $\Phi : E \to E$ **isoton δ–diskontiert**, falls für alle $u, \tilde{u} \in E$ und $c \in \mathbb{R}_{\geq 0}$ gilt: Aus $u \leq \tilde{u} + c$ folgt $\Phi(u) \leq \Phi(\tilde{u}) + \delta c$.

Satz D2G: Blackwells hinreichendes Kriterium

Ist $\Phi : E \to E$ isoton δ–diskontiert, so auch δ–lipschitz–stetig.

Beweis: Für alle $u, \tilde{u} \in E$ gilt $u - \tilde{u} \leq \|u - \tilde{u}\|$, also $u \leq \tilde{u} + \|u - \tilde{u}\|$.
Dank isotoner δ–Diskontierung folgt daraus $\Phi(u) \leq \Phi(\tilde{u}) + \delta\|u - \tilde{u}\|$.
Wir erhalten $\Phi(u) - \Phi(\tilde{u}) \leq \delta\|u - \tilde{u}\|$, ebenso $\Phi(\tilde{u}) - \Phi(u) \leq \delta\|u - \tilde{u}\|$.
Das bedeutet $\|\Phi(u) - \Phi(\tilde{u})\| \leq \delta\|u - \tilde{u}\|$, wie behauptet. QED

🙂 Dieses Kontraktionskriterium ist nicht *notwendig*, aber *hinreichend*: Wir nutzen die partielle Ordnung auf E. Mehr Struktur vereinfacht.

🙂 Ist der betrachtete Teilraum E zudem abgeschlossen in $B(X, \mathbb{R})$, so ist E vollständig, und wir können Banachs Fixpunktsatz anwenden.

Blackwells hinreichendes Kriterium

D216 Erläuterung

Jede Grundvorlesung zur Analysis behandelt Banachs Fixpunktsatz im Laufe des ersten Jahres. Dazu notwendig sind Konvergenz von Folgen, Cauchy–Kriterium und Vollständigkeit. Zum allgemeinen Aufbau gehören ebenso Skalarprodukte, Normen, sowie Metriken und Topologien.

Das Kriterium von Blackwell wird in der Analysis meist nicht erwähnt, da es für die dortigen Anwendungen nicht unmittelbar benötigt wird. Im Kontext der Ökonomie und Optimierung hilft es jedoch ungemein, und somit rückt es nun in den Mittelpunkt unseres Interesses.

🙂 Für Blackwell genügt eine beliebige Teilmenge $E \subseteq B(X, \mathbb{R})$.
Für Banach sollte E zudem abgeschlossen in $B(X, \mathbb{R})$ sein.

Gegeben sei eine Zerlegung $X = X^\circ \sqcup \partial X$ sowie $v \in B(\partial X, \mathbb{R})$ als Randbedingung. Dies definiert in $B(X, \mathbb{R})$ den affinen Teilraum

$$E_v := \{\, u \in B(X, \mathbb{R}) \mid u|_{\partial X} = v \,\}.$$

🙂 Der $\mathbb{R}$–Vektorraum $B(X, \mathbb{R})$ ist vollständig, also ein Banach–Raum. Hierin ist $E_v \subseteq B(X, \mathbb{R})$ abgeschlossen, also ebenfalls vollständig.

Von Graphen und Markov–Ketten zu Markov–Graphen

D217 Erläuterung

Wir betrachten einen **Graphen** $\Gamma = (X, A, \sigma, \tau)$ wie in B1D erklärt. Vereinfachend gelte $A = \bigcup_{x\in X}\{x\} \times A_x$ mit $\sigma = \mathrm{pr}_1 : A \to X : (x, a) \mapsto x$ und der **Transition** $\tau : A \to X : (x, a) \mapsto y$ bzw. lokal $\tau_x : A_x \to X : a \mapsto y$.

Diese deterministische Sichtweise als Graph verallgemeinern wir nun zu einem stochastischen Modell mit Zufall / Unsicherheit. Zur Erinnerung:

Definition D2H: (diskrete) Markov–Kette

Eine (diskrete) **Markov–Kette** (X, τ) besteht aus einer (abzählbaren) Zustandsmenge X und Transition $\tau : X \to [X] : x \mapsto \sum_{y\in X} p(x, y)\, y$ mit Wkten $p(x, y) \geq 0$ und $\sum_{y\in X} p(x, y) = 1$ für alle $x \in X$.

Dies entspricht einer **stochastischen Matrix** $P = (p(x, y))_{(x,y)\in X\times X}$. Aus dem Zustand x entsteht mit Übergangswkt $p(x, y)$ der Zustand y. Die Verteilung $\mu \in [X]$ geht über in die Verteilung $\bar\mu = \mu\, P \in [X]$ mit

$$\bar\mu(y) = \sum_{x\in X} \mu(x)\, p(x, y).$$

Dies entspricht dem **Matrixprodukt**: Zeilenvektor μ mal Matrix P.

Von Graphen und Markov–Ketten zu Markov–Graphen

D218 Erläuterung

Angenommen, auf X sind nach dem nächsten Schritt $x \to y$ die **Gewinnerwartungen** gegeben durch $u : X \to \mathbb{R} : y \mapsto u(y)$. Dann ist die Gewinnerwartung vor dem Schritt gegeben durch

$$\bar u(x) = \sum_{y\in X} p(x, y)u(y).$$

Dies entspricht dem **Matrixprodukt**: Matrix P mal Spaltenvektor u.
Die Wkten μ werden nach vorne geschoben gemäß $\mu \mapsto \bar\mu = \mu P$.
Die Erwartungen u werden zurück gezogen gemäß $u \mapsto \bar u = Pu$.

Das erinnert uns an das Prinzip der Rekursion / Rückwärtsinduktion: Gespielt wird immer vorwärts, aber optimiert wird leichter rückwärts.

Wir verbinden nun deterministische Spiele (Zustände und Aktionen) mit stochastischen Prozessen (aus Zuständen und Übergangswkten): Der Spieler wählt in jedem aktiven Zustand $x \in X^\circ$ eine Aktion $a \in A_x$. Das System geht dann vom Zustand x über nach y mit Wkt $p(x, a, y)$.

Damit gelangen wir also zu den extrem vielseitigen **Markov–Graphen**, die wir schon aus unseren bisherigen Beispielen kennen und schätzen.

Markov–Belohnungsprozess / MRP

D219 Erläuterung

Bevor wir Markov–Spiele [*Markov decision processes* / MDP] erklären, ist es als Zwischenschritt vielleicht hilfreich, zunächst vereinfachend Belohnugsprozesse [*Markov reward processes* / MRP] zu betrachten: Wir verfeinern Markov–Ketten durch die Zerlegung in aktive Zustände X° und terminale Zustände ∂X und erklären zudem Belohnungen.

Definition D2I: Markov–Belohnungsprozess / MRP

Ein **Markov–Belohnungsprozess** (X, τ, r, v) besteht aus einer Menge $X = X^\circ \sqcup \partial X$ mit einer Transition $\tau : X^\circ \to [X] : x \mapsto \sum_{y\in X} p(x, y)\, y$ sowie einer sofortigen Belohnung $r : X^\circ \times X \to \mathbb{R} : (x, y) \mapsto r(x, y)$ und einer terminalen Auszahlung $v : \partial X \to \mathbb{R} : x \mapsto v(x)$. Zum Diskontfaktor $\delta \in [0, 1]$ erklären wir die **Erwartungsgleichung** für $u : X \to \mathbb{R}$ durch

$$u(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ \sum_{y\in X} p(x, y)\big[r(x, y) + \delta u(y)\big] & \text{für } x \in X^\circ. \end{cases}$$

Wir nennen den Prozess **lösbar**, falls diese Gleichung eine eindeutige Lösung u besitzt. (Insbesondere muss jede Reihe absolut konvergieren.)

Markov–Belohnungsprozess / MRP

D220 Erläuterung

Bei einem Belohnungsprozess gibt es noch nichts zu entscheiden: Er beschreibt eine Markov–Kette, eventuell mit terminalen Zuständen, und einen Strom von Zahlungen, die wir diskontiert aufsummieren.

Ist jede Trajektorie endlich, $x_0, x_1, \dots, x_n \in X$, so führt sie zur endlichen Summe $r(x_0, x_1) + \delta r(x_1, x_2) + \cdots + \delta^{n-1} r(x_{n-1}, x_n) + \delta^n v(x_n) \in \mathbb{R}$. Die Berechnung der erwarteten Auszahlung $u : X \to \mathbb{R}$ gelingt dann per Rückwärtsinduktion, wie in B1F erklärt. Wir müssen lediglich absolute Konvergenz sicherstellen, so dass die Erwartung wohldefiniert ist. Dies gilt zum Beispiel, falls $y \mapsto r(x, y) + \delta u(y)$ beschränkt ist.

Im Allgemeinen gibt es unendliche Trajektorien, insbesondere Zykel. Wir interpretieren die Erwartungsgleichung D2I als Fixpunktgleichung und nutzen den Banachschen Fixpunktsatz D2A: Sind $r : X^\circ \times X \to \mathbb{R}$ und $v : \partial X \to \mathbb{R}$ beschränkt und $\delta \in [0, 1[$, so existiert genau eine Lösung $u : X \to \mathbb{R}$, und diese lässt sich iterativ berechnen.

Übung: Formulieren Sie dies als Satz und beweisen Sie ihn. (Wir führen die Rechnungen in Satz D2K allgemein aus.)

Markov–Spiel / MDP und Bellman–Gleichung

D221

Definition D2J: Markov–Spiel und Bellman–Gleichung

Ein **Markov–Graph** $\Gamma = (X, A, \tau)$ besteht aus einer Zustandsmenge X und einer Aktionsmenge $A = \bigcup_{x\in X}\{x\} \times A_x$ zusammen mit Projektion $\sigma : (x, a) \mapsto x$ und Transition $\tau : A \to [X] : (x, a) \mapsto \sum_{y\in X} p(x, a, y)\, y$.

Für alle $(x, a) \in A$ und $y \in X$ gelte $p(x, a, y) \geq 0$ und $\sum_y p(x, a, y) = 1$. Wie üblich zerlegen wir die Zustandsmenge $X = X^\circ \sqcup \partial X$ in aktive Zustände $X^\circ = \mathrm{Bild}(\sigma)$ und terminale Zustände $\partial X = X \setminus X^\circ$.

Die **Strategiemenge** ist $S = S(\Gamma) := \{\, s : X^\circ \to A \mid \sigma \circ s = \mathrm{id}_{X^\circ} \,\}$. Auszahlungen seien terminal $v : \partial X \to \mathbb{R} : x \mapsto v(x)$ oder instantan $r : A \times X \to \mathbb{R} : (x, a, y) \mapsto r(x, a, y)$ mit dem Diskontfaktor $\delta \in [0, 1]$.

(0) Dieses **Markov–Spiel** (Γ, r, v) definiert die **Bellman–Gleichung**

$$u(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ \sup\limits_{a\in A_x} \underbrace{\sum_{y\in X} p(x, a, y)\big[r(x, a, y) + \delta u(y)\big]}_{\text{Hamilton–Funktion } H(x,\, a,\, u)} & \text{für } x \in X^\circ. \end{cases}$$

Markov–Spiel / MDP und Bellman–Gleichung

D222 Erläuterung

⚠ Für die Reihe über $y \in X$ verlangen wir absolute Konvergenz, etwa $y \mapsto p(x, a, y)$ endlich getragen oder $y \mapsto r(x, a, y) + \delta u(y)$ beschränkt.

🙂 Im Folgenden verlangen wir, dass r und v sowie u beschränkt sind. Damit lösen sich alle Fragen zur Konvergenz in Wohlgefallen auf.

🙂 Der deterministische Spezialfall entspricht $\tau : (x, a) \mapsto y$ wie zuvor, also $p(x, a, y) = 1$ für das Ziel $y \in X$ und $p(x, a, y') = 0$ für alle $y' \neq y$.

🙂 Ist der (deterministische) Graph $\Gamma = (X, A, \tau)$ zudem artinsch, so löst Rekursion / Rückwärtsinduktion B1F die Bellman–Gleichung.

🙂 Jede Aktion $a : x \to y$ hat zwei Auswirkungen, die wir ausbalancieren: die sofortige Belohnung $r(x, a, y)$ und der langfristige Nutzen $u(y)$.

🙂 Ist Γ lokal-endlich, so wird das Supremum jeweils angenommen. Im lösbaren Falle gilt also $u(x) = \max_{a\in A_x} H(x, a, u)$ für alle $x \in X^\circ$.

🙂 Die Funktion u liefert uns die optimale Auszahlung $x \mapsto u(x)$ sowie Aktionen $s(x) \in \operatorname{Arg\,max}_{a\in A_x} H(x, a, u)$, also eine optimale Strategie s!

🙂 Optimale Züge erkennen Sie leicht sobald Sie das Optimum kennen.

Bellman–Operator und Erwartungsoperator

D223

Definition D2J: Gewinnerwartung und Optimalität

(1) Sei $E = \{\, u \in B(X, \mathbb{R}) \mid u|_{\partial X} = v \,\}$. Auf diesem affinen Teilraum definieren wir den **Bellman–Operator** $\Phi : E \to E : u \mapsto \bar u$ durch

$$\bar u(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ \sup\limits_{a\in A_x} \underbrace{\sum_{y\in X} p(x, a, y)\big[r(x, a, y) + \delta u(y)\big]}_{\text{Hamilton–Funktion } H(x,\, a,\, u)} & \text{für } x \in X^\circ. \end{cases}$$

Die Fixpunkte von Φ sind genau die Lösungen der Bellman–Gleichung. Wir nennen Φ **eindeutig lösbar**, falls genau ein Fixpunkt $u \in E$ existiert, und **konvergent**, falls zudem $\Phi^n(\tilde u) \to u$ für $n \to \infty$ gilt für alle $\tilde u \in E$.

(2) Zu $s \in S(\Gamma)$ definieren wir den **Erwartungsoperator** Φ_s durch

$$\bar u(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ \underbrace{\sum_{y\in X} p(x, s(x), y)\big[r(x, s(x), y) + \delta u(y)\big]}_{\text{Hamilton–Funktion } H(x,\, s(x),\, u)} & \text{für } x \in X^\circ. \end{cases}$$

Fixpunkte $u \in E$ von Φ_s sind Lösungen der **Erwartungsgleichung**.

Bellman–Operator und Erwartungsoperator

D224 Erläuterung

Aus unserem Markov–Spiel (Γ, r, v), also den Entscheidungsprozess (X, A, τ, r, v) gemäß D2J, wird durch die Festlegung einer Strategie $s \in S(\Gamma)$ ein Belohnungprozess $(X, \tilde\tau, \tilde r, v)$ gemäß D2I. Ausführlich:

$$\tilde\tau \;:\; X \to [X] \;:\; x \mapsto \textstyle\sum_{y\in X} p(x, s(x), y)\, y$$
$$\tilde r \;:\; X \times X \to \mathbb{R} \;:\; (x, y) \mapsto r(x, s(a), y)$$

Die zugehörige Erwartungsgleichung haben wir in (2) wiederholt, als Fixpunktgleichung für den Erwartungsoperator Φ_s. Umgekehrt ist jeder Belohnungsprozess $(X, \tilde\tau, \tilde r, v)$ ein Entscheidungsprozess (X, A, τ, r, v) mit $A_x = \{a_x\}$, $p(x, a_x, y) = \tilde p(x, y)$, $r(x, a_x, y) = \tilde r(x, y)$ für alle $x \in X^\circ$

Um einen direkten, raschen Zugang anzubieten, habe ich hier beide Aspekte in einer gemeinsamen Definition D2J zusammengefasst.

Alternativ kann man die Theorie kleinschrittiger aufbauen und zunächst Markov–Ketten D2H und Belohnugsprozesse D2I als Zwischenetappen untersuchen und ihre Konvergenzfragen klären. Ich formuliere dies hier als Übung und konzentriere mich auf allgemeine Markov–Spiele (D2K).

Bellman–Operator und Erwartungsoperator

D225 Erläuterung

Aufgabe: Formalisieren Sie das Spiel vom Kapitelanfang explizit als ein Markov–Spiel und lösen Sie es mit Hilfe der Bellman–Gleichung.

7€								23€

Lösung: Die Zustandsmenge ist $X = \{*, 0, 1, \dots, 8\}$ mit $\partial X = \{*, 0, 8\}$ und den Auszahlungen $v(*) = 0$ sowie $v(0) = 7$ und $v(8) = 23$. Jeder aktive Zustand $x \in X^\circ = \{1, 2, \dots, 7\}$ bietet zwei Züge, $A_x = \{\text{go}, \text{stop}\}$, mit Wkten $p(x, \text{go}, x \pm 1) = 1/2$ und Belohnung $r(x, \text{go}, x \pm 1) = c := -1$ sowie den Spielabbruch mit $p(x, \text{stop}, *) = 1$ und $r(x, \text{stop}, *) = 0$.

(1) Die Strategie $s' : x \mapsto \text{go}$ für alle $x \in X^\circ$ ergibt $u_{s'} = \Phi_{s'}(u_{s'})$ wie folgt:

7	2	−1	−2	−1	2	7	14	23

(2) Wechsel zur Strategie s mit $s(3) = \text{stop}$ ergibt $u = \Phi_s(u)$ wie folgt:

7	8/3	1/3	0	3/5	16/5	39/5	72/5	23

🙂 Diese Funktion u erfüllt zugleich die Bellman–Gleichung $u = \Phi(u)$. Dank Bellmans Optimalitätsprinzip D2M ist dies die optimale Strategie!

Bellman–Operator und Erwartungsoperator

D226 Erläuterung

🙂 Markov–Spiele und Bellman–Gleichung sind natürlich und einfach. Das erste Anwendungsbeispiel zeigt: Unser Modell passt wunderbar!

🙂 Die gezeigten Funktionen sind jeweils die einzigen Lösungen. Jede Strategie $s \in S$ ist randverbunden und somit Φ_s kontraktiv. (D2L)

Insgesamt gibt es $|S| = 2^7 = 128$ Strategien, also exponentiell in $|X^\circ|$. Wir können jede dieser 128 Strategien $s \in S$ untersuchen und jeweils die Gewinnerwarung $u_s : X \to \mathbb{R}$ berechnen. Dann können wir daraus die optimale Strategie auswählen, genauer die optimale Gewinnerwartung $u_* = \max u_s$ und dazu eine optimale Strategie $s \in S$, sodass $u_* = u_s$.

☹ Diese *globale* Optimierung durch *brute force* ist jedoch aufwändig, da die Strategiemenge $S = S(\Gamma)$ sehr groß und unübersichtlich ist.

🙂 Die Bellman–Gleichung bietet dagegen eine *lokale* Optimierung, und diese gelingt wesentlich effizienter: Das ist ihr großer Nutzen! Dass beide Rechenwege zum selben Ergebnis führen, ist die Aussage von Bellmans Optimalitätsprinzip D2M, das wir als nächstes beweisen.

Gewinnerwartung und Optimalität

D227 Erläuterung

Definition D2J erklärt das grundlegende Modell und alle relevanten Daten: der Markov–Graph $\Gamma = (X, A, \tau)$, das Markov–Spiel (Γ, r, v), die Strategiemenge $S(\Gamma)$ und die Bellman–Gleichung für $u : X \to \mathbb{R}$.

Darauf bauend erklären wir in Definition D2J die Operatoren Φ und Φ_s. Wir wünschen uns, dass sie kontrahieren, oder wenigstens konvergieren, oder jeder einen eindeutigen Fixpunkt hat: $u = \Phi(u)$ bzw. $u_s = \Phi_s(u_s)$.

Das sind zunächst einmal Hoffnungen / Wünsche / Annahmen / Axiome. Für praktische Anwendungen benötigen wir jeweils handfeste Kriterien. Immerhin können wir mit den Begriffen Phänomene präzise benennen.

🙂 Der Operator Φ_s beschreibt die **Erwartung** der Strategie $s \in S(\Gamma)$, die zugehörige Fixpunktgleichung $u_s = \Phi_s(u_s)$ heißt dementsprechend **Erwartungsgleichung**. Ihre (hoffentlich eindeutige) Lösung $u_s : X \to \mathbb{R}$ heißt **Erwartungsfunktion** der hier vorgegebenen Strategie $s \in S(\Gamma)$: Vom Zustand $x \in X$ erwarten wir mit Strategie s den Gewinn $u_s(x)$.

🙂 Die **Erwartungsoperatoren** Φ_s sind affin-linear, haben bessere Eigenschaften, die stärkere Theorie und sind leichter zu behandeln.

Gewinnerwartung und Optimalität

D228 Erläuterung

Die Abbildung $\Phi : E \to E$ heißt **Bellman–(Optimalitäts–)Operator**, die Fixpunktgleichung $u = \Phi(u)$ heißt **Bellman–(Optimalitäts–)Gleichung**. Ihre (hoffentlich eindeutige) Lösung $u : X \to \mathbb{R}$ heißt **Gewinnfunktion**: Vom Zustand $x \in X$ erwarten wir bei optimalem Spiel den Gewinn $u(x)$.

Die Bellman–Gleichung ist vielseitig einsetzbar und daher berühmt. Sie ist eine Funktionalgleichung, denn die hierbei gesuchte Größe ist eine Funktion $u : X \to \mathbb{R}$. (Nun ja, eigentlich ist u auch nur ein Vektor, doch wir erlauben vorsorglich auch unendliche Zustandsmengen X.)

⚠ Hierzu ist noch keine Strategie vorgegeben. Im Gegenteil nutzen wir die Funktion u, um daraus schließlich eine optimale Strategie abzulesen!

Zur Berechnung bzw. Approximation der Funktion u wurden Dutzende Verfahren vorschlagen; uns geht es hier zunächst um ihre Definition, dann um grundlegende Eigenschaften und schließlich die Berechnung.

⚠ Anders als Φ_s ist der Operator Φ nicht-linear und daher schwieriger. Zur Definition der Funktion u nutzen wir zunächst die Fixpunktgleichung $u = \Phi(u)$; anschließend zeigen wir das Optimalitätsprinzip $u = \sup_s u_s$.

Iteration für Gewinnerwartung und Optimalität

D229

Sei (Γ, r, v) ein Markov–Spiel mit beschränkten Belohnungen r und v. Letzteres gilt automatisch auf jedem endlichen Markov–Graphen Γ.

Auf $E = E_v = \{\, u \in B(X, \mathbb{R}) \mid u|_{\partial X} = v \,\}$ nutzen wir die Operatoren

$$\Phi : u \mapsto \bar{u} : \bar{u}(x) = \sup_{a \in A_x} \underbrace{\textstyle\sum_{y \in X} p(x,a,y)\big[r(x,a,y) + \delta u(y)\big]}_{\text{Hamilton–Funktion } H(x,a,u)} \quad \text{und}$$

$$\Phi_s : u \mapsto \bar{u} : \bar{u}(x) = H(x, s(x), u) \quad \text{für jede Strategie } s \in S(\Gamma).$$

Satz D2K: Kontraktion, somit Existenz und Eindeutigkeit

Für jeden Diskontfaktor $\delta \in [0, 1]$ gilt: Alle Erwartungsoperatoren Φ_s und der Bellman–Operator Φ sind isoton δ–diskontiert, somit δ–lipschitz.

Speziell für $\delta \in [0, 1[$ können wir Banachs Fixpunktsatz D2A anwenden, wie in unseren Beispielen motiviert: Es gibt genau einen Fixpunkt und diesen können wir durch das Iterationsverfahren effizient annähern.

Beweis: Dies folgt aus den Definitionen durch geduldiges Nachrechnen. Führen Sie dies sorgsam aus, es ist eine gute Übung zur Wiederholung!

Iteration für Gewinnerwartung und Optimalität

D230 Erläuterung

Ausführlich: Sei $u \le \tilde{u} + c$. Für alle $x \in X^\circ$, $a \in A_x$ und $y \in X$ gilt:

$$\begin{aligned}
u(y) &\le \tilde{u}(y) + c \\
r(x,a,y) + \delta u(y) &\le r(x,a,y) + \delta\tilde{u}(y) + \delta c \\
p(x,a,y)\big[r(x,a,y) + \delta u(y)\big] &\le p(x,a,y)\big[r(x,a,y) + \delta\tilde{u}(y) + \delta c\big] \\
\textstyle\sum_y p(x,a,y)\big[r(x,a,y) + \delta u(y)\big] &\le \textstyle\sum_y p(x,a,y)\big[r(x,a,y) + \delta\tilde{u}(y)\big] + \delta c \\
H(x,a,u) &\le H(x,a,\tilde{u}) + \delta c
\end{aligned}$$

Zur Definition der Erwartung $H(x, a, u)$ fordern wir absolute Konvergenz. Das ist garantiert, falls neben u und $\tilde{u}$ auch $y \mapsto r(x, a, y)$ beschränkt ist.

Ist die Strategie $s \in S(\Gamma)$ vorgegeben, so wählen wir im Zustand $x \in X^\circ$ die Aktion $a = s(x)$. Obige Rechnung garantiert $\Phi_s(u) \le \Phi_s(\tilde{u}) + \delta c$.

Für den Operator Φ wird optimiert: Wir bilden das Supremum über alle Aktionen $a \in A_x$, erst rechts dann links, und erhalten $\Phi(u) \le \Phi(\tilde{u}) + \delta c$.

Endlichkeit ist garantiert, falls neben u und $\tilde{u}$ auch für jeden Zustand x die Funktion $a \mapsto \sum_{y \in X} p(x, a, y) r(x, a, y)$ beschränkt ist. Ist zudem A_x endlich, so wird das Supremum angenommen, ist also ein Maximum.

Iteration für Gewinnerwartung und Optimalität

D231 Erläuterung

Sei (Γ, r, v) ein Markov–Spiel mit r, v beschränkt und $\delta = 1$. Wir fixieren $s \in S(\Gamma)$ mit Übergangswkten $p : X^\circ \times X \to [0, 1] : (x, y) \mapsto p(x, s(x), y)$. Wir schreiben $x \to y$ falls $p(x, y) > 0$. Für $x \in \partial X$ setzen wir $p(x, x) = 1$. Somit ist (X, p) eine Markov–Kette mit endlichem Zustandsraum X.

Wir nennen s **randverbunden**, wenn es zu jedem $x_0 \in X$ einen Weg $x_0 \to x_1 \to \dots \to x_\ell$ mit $x_\ell \in \partial X$ gibt. Falls X endlich ist, so ist s sogar **stark (ℓ, ε)–randverbunden** für ein geeignetes Paar $(\ell, \varepsilon) \in \mathbb{N} \times \mathbb{R}_{>0}$: Jeder Startzustand führt nach ℓ Schritten mit Wkt $\ge \varepsilon$ in den Rand ∂X.

🙂 Anschaulich: Die Wkt diffundiert in den Rand, langsam aber sicher. Nach ℓ Schritten ist die Gesamtwkt aller aktiven Zustände $\le k = 1 - \varepsilon$. Auch für $\delta = 1$ kann also Kontraktion vorliegen; wir müssen hinschauen:

Satz D2L: randverbunden impliziert konvergent

Sei (Γ, r, v) ein Markov–Spiel und $\delta = 1$. Die Strategie $s \in S(\Gamma)$ sei stark (ℓ, ε)–randverbunden. Dann ist Φ_s^ℓ kontraktiv mit Konstante $k = 1 - \varepsilon$.

Somit ist $\Phi_s : E \to E$ konvergent: Es existiert genau ein Fixpunkt $u_s \in E$ und zudem konvergiert $\Phi_s^n(\tilde{u}) \to u$ für jeden Startwert $\tilde{u} \in E$.

Iteration für Gewinnerwartung und Optimalität

D232 Erläuterung

Beweis: Wir untersuchen $\Phi_s(u)(x) = \sum_{y \in X} p(x, y)[r(x, y) + u(y)]$.

Wir betrachten $P = (p(x, y))_{(x,y) \in X \times X}$ als zeilen-stochastische Matrix und $u = (u(y))_{y \in X}$ als Spaltenvektor. Damit gilt $\Phi_s(u) = c + Pu$ mit additiver Belohnung $c = (c(x))_{x \in X}$ und $c(x) = \sum_{y \in X} p(x, y) r(x, y)$.

Per Induktion erhalten wir $\Phi_s^n(u) = c + Pc + P^2c + \dots P^{n-1}c + P^n u$. Die Potenz $P^n = (p_n(x, y))_{(x,y) \in X \times X}$ berechnet die Wkt $p_n(x, y)$, in n Schritten von x nach y zu gehen, als Summe aller Wege der Länge n.

Für alle $u, \tilde{u} \in E$ und $n \in \mathbb{N}$ gilt somit $\Phi_s^n(u) - \Phi_s^n(\tilde{u}) = P^n(u - \tilde{u})$. Wir zeigen nun $|P^\ell(u - \tilde{u})| \le k|u - \tilde{u}|$. Wir untersuchen $\hat{u} = P^\ell(u - \tilde{u})$: Für $x \in \partial X$ gilt $p_\ell(x, x) = 1$, also $\hat{u}(x) = u(x) - \tilde{u}(x) = v(x) - v(x) = 0$. Für jeden aktiven Zustand $x \in X^\circ$ hingegen finden wir:

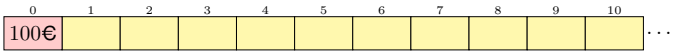
$$\hat{u}(x) = \sum_{y \in X} p_\ell(x, y)\big[u(y) - \tilde{u}(y)\big] = \sum_{y \in X^\circ} p_\ell(x, y)\big[u(y) - \tilde{u}(y)\big]$$

$$|\hat{u}(x)| \le \sum_{y \in X^\circ} p_\ell(x, y)\big|u(y) - \tilde{u}(y)\big| \le \sum_{y \in X^\circ} p_\ell(x, y)\big|u - \tilde{u}\big| \le k\big|u - \tilde{u}\big|$$

Somit ist Φ_s^ℓ kontraktiv und Banachs Fixpunktsatz D2A anwendbar. QED

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

D233 Erläuterung



Aufgabe: Wir untersuchen die Irrfahrt auf $X = \mathbb{N}$ mit Rand $\partial X = \{0\}$; mit Wkt $p \in]0,1[$ geht es nach rechts, mit Wkt $q = 1-p$ nach links. Die Belohnung sei $r \in \mathbb{R}$ in jedem Schritt mit Diskontfaktor $\delta \in]0,1[$.

(1) Finden Sie alle $u : \mathbb{N} \to \mathbb{R}$ mit $\Phi(u) = u$. Welche sind beschränkt?
(2) Wie verhalten sich beschränkte Lösungen für $r = 0$ und $\delta \nearrow 1$?
(3) Lösen Sie den Fall $\delta = 1$. (4) Ist Φ kontraktiv? (5) konvergent?

Lösung: (1) Für $x \in \mathbb{N}_{\geq 1}$ lösen wir die Fixpunktgleichung $u = \Phi(u)$:

$$\boxed{u(x) = r + \delta\big[pu(x+1) + qu(x-1)\big]}$$

Durch $u(0)$ und $u(1)$ sind alle Werte $u(2), u(3), \dots$ rekursiv festgelegt.
Der Ansatz $u(x) = ab^x + r/(1-\delta)$ führt zu $b^x = \delta[pb^{x+1} + qb^{x-1}]$, also

$$b \in \{b_1, b_2\} \quad \text{mit} \quad b_{1/2} = \frac{1 \mp \sqrt{1 - 4pq\delta^2}}{2p\delta} \quad \text{und} \quad 0 < b_1 < 1 < b_2.$$

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

D234 Erläuterung

Damit erhalten wir alle Lösungen der Erwartungsgleichung $u = \Phi(u)$:

$$u(x) = a_1 b_1^x + a_2 b_2^x + \frac{r}{1-\delta} \qquad \text{mit} \qquad a_1, a_2 \in \mathbb{R}$$

Die beschränkten Lösungen sind demnach $u(x) = a_1 b_1^x + r/(1-\delta)$.
Der Startwert $u(0) = v(0)$ bestimmt die Konstante $a_1 = u(0) - r/(1-\delta)$.

$$\boxed{u(x) = \left[u(0) - \frac{r}{1-\delta}\right]\left[\frac{1 - \sqrt{1-4pq\delta^2}}{2p\delta}\right]^x + \frac{r}{1-\delta}}$$

☺ Dank Diskont $\delta < 1$ erhalten wir in jedem Falle genau eine Lösung!
Das entspricht dem Kontraktionssatz D2K, hier nun wunderbar konkret.
Obwohl der Graph unendlich ist, greifen unsere Werkzeuge bestens.

☺ Der gegebene Anfangswert klingt exponentiell ab gegen $r/(1-\delta)$.
Konkretes Beispiel: Für $p = q = 1/2$ und $\delta = 0.8$ finden wir $b_1 = 1/2$,
und somit die Gewinnfunktion $u : \mathbb{N} \to \mathbb{R} : x \mapsto [u(0) - 5r]2^{-x} + 5r$.

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

D235 Erläuterung

(2) Für $r = 0$ finden wir den Grenzwert $\lim_{\delta \nearrow 1} u(x) = u(0)\, b^x$ mit

$$b = \frac{1 \mp \sqrt{1-4p(1-p)}}{2p} = \frac{1 \mp \sqrt{(1-2p)^2}}{2p}$$
$$= \frac{1 - |1-2p|}{2p} = \begin{cases} 1 & \text{für } 0 < p \leq 1/2, \\ (1-p)/p & \text{für } 1/2 < p < 1. \end{cases}$$

☺ Das ist tatsächlich eine beschränkte Lösung des Falls $\delta = 1$.
Für $1/2 < p < 1$ gilt $b < 1$ und somit $u(x) \to 0$ für $x \to \infty$.
Unten in (3) finden wir weitere beschränkte Lösungen,
die diese beiden Eigenschaften nicht haben.

Nochmal anders gesagt: Für $\delta = 1$ sind selbst beschränkte Lösungen nicht eindeutig. Unter den vielen „mathematischen“ Lösungen finden wir genau eine „natürliche“ oder „physikalisch-plausible“ Lösung wie oben. Allein die Erwartungsgleichung / Bilanzgleichung genügt hier also noch nicht zur Charakterisierung der „richtigen“ Lösung.

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

D236 Erläuterung

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

D237 Erläuterung

(3) Wir setzen $\delta = 1$ und lösen die Fixpunktgleichung $u = \Phi(u)$:

$$\boxed{u(x) = r + pu(x+1) + qu(x-1).}$$

Für $p = q = 1/2$ haben wir bereits zuvor alle Lösungen bestimmt:
Diese sind Parabeln der Form $u(x) = u(0) + ax - rx^2$ mit $a \in \mathbb{R}$.
Im Folgenden sei daher $p \neq 1/2$ und weiterhin $0 < p < 1$.

Der Ansatz $u(x) = ab^x + rx/(1-2p)$ führt zu $b^x = pb^{x+1} + qb^{x-1}$, also

$$b \in \left\{\frac{1 \mp \sqrt{1-4pq}}{2p}\right\} = \left\{\frac{1 \mp |1-2p|}{2p}\right\} = \left\{1, \frac{1-p}{p}\right\}.$$

Für $b = (1-p)/p$ und $a \in \mathbb{R}$ erhalten wir also die allgemeine Lösung

$$\boxed{u(x) = u(0) - a + ab^x + \frac{rx}{1-2p}}$$

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

D238 Erläuterung

Für $0 < p < 1/2$ gilt $b > 1$: Beschränkte Lösungen existieren nur für $a = r = 0$, also nur die konstante Lösung $u(x) = u(0)$ für alle $x \in \mathbb{N}$.

Für $1/2 < p < 1$ gilt $0 < b < 1$: Beschränkte Lösungen existieren nur für $r = 0$; dann gilt $u(x) = u(0) - a + ab^x$, wobei $a \in \mathbb{R}$ beliebig wählbar ist.

Für $x \to \infty$ erfüllt diese Lösung $u(x) \to u(0) - a$. Nur für $a = u(0)$ erfüllt unsere Lösung $u(x) = u(0)\, b^x$ zudem $u(x) \to 0$ für $x \to \infty$.

☺ Das ist die „natürliche“ oder „physikalisch-plausible“ Lösung,
die wir oben in (2) als den Grenzwert für $\delta \nearrow 1$ gefunden haben.
Alle anderen Lösungen erfüllen ebenfalls die Erwartungsgleichung,
doch diese alleine genügt hier noch nicht zur Eindeutigkeit.

☺ Diese schöne Illustration ist ein heilsames Gegenbeispiel:
Für $\delta = 1$ ist Eindeutigkeit keinesfalls selbstverständlich!

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

D239 Erläuterung

(4) Weiter sei $\delta = 1$ und $r = 0$. Für $0 < p \leq 1/2$ hat $\Phi(u) = u$ genau eine beschränkte Lösung $u \in E$, nämlich $u(x) = u(0)$ für alle $x \in \mathbb{N}$.

Dennoch ist $\Phi : E \to E$ **nicht kontraktiv**, auch keine Potenz Φ^n:
Wir vergleichen $u, \tilde{u} : \mathbb{N} \to \mathbb{R}$ mit $u(0) = \tilde{u}(0) = 0$ sowie $u(x) = 0$ und $\tilde{u}(x) = 1$ für $x \in \mathbb{N}_{\geq 1}$. Dann gilt $\Phi^n(u) = u$ und $\Phi^n(\tilde{u})(x) = 1$ für $x > n$.

Dies zeigt, dass $\Phi : E \to E$ **nicht gleichmäßig konvergent** ist,
denn es gilt nicht $\Phi^n(\tilde{u}) \to u$ bezüglich der Supremumsnorm auf E.

☺ Das ist ein weiteres schönes Beispiel für unser Repertoire:
Der Bellman–Operator $\Phi : E \to E$ ist hier nicht kontraktiv,
dennoch existiert genau eine Lösung $u = \Phi(u)$.
Die vorsichtige Begriffsbildung in Definition D2J nimmt so langsam konkrete Gestalt an und füllt sich nachträglich mit Leben.

Gewinnerwartung auf $\mathbb{N}$ als unendlicher Graph

D240 Erläuterung

(5) Für $0 < p \leq 1/2$ ist Φ immerhin noch **punktweise konvergent**:
Es existiert genau ein Fixpunkt $u \in E$ und für jeden Startwert $\tilde{u} \in E$ gilt Konvergenz $\Phi^n(\tilde{u})(x) \to u(x)$ für $n \to \infty$ in jedem Punkt $x \in X$.

Wir zeigen dies für $v(0) = 0$. Sei $u_0 \in E$ gegeben durch $u_0(0) = 0$ und $u_0(x) = 1$ für $x \in \mathbb{N}$. Für $u_n = \Phi^n(u_0)$ gilt $u_0 \geq u_1 \geq u_2 \geq \cdots \searrow u \geq 0$.
Aus $u_{n+1}(x) = pu_n(x+1) + qu_n(x-1)$ wird $u(x) = pu(x+1) + qu(x-1)$ für $n \to \infty$, also $u = \Phi(u)$. Somit gilt $u = 0$ dank Eindeutigkeit.

Für jeden Startwert $\tilde{u} \in E$ gilt $-cu_0 \leq \tilde{u} \leq cu_0$ mit $c = |\tilde{u}|$. Daraus folgt $-cu_n \leq \Phi^n(\tilde{u}) \leq cu_n$ für alle $n \in \mathbb{N}$, also punktweise $\Phi^n(\tilde{u})(x) \to 0$.

☺ Auch dies ist ein weiteres schönes Beispiel für unser Repertoire.
Satz D2O erklärt ein allgemeines Kriterium für punktweise Konvergenz.

Bellmans Optimalitätsprinzip $u_* = u$

D241

Gegeben sei ein Markov–Spiel (Γ, r, v) und $u : X \to \mathbb{R}$ mit $u = \Phi(u)$.
Wie kann die Auszahlung u realisiert werden, oder gar noch höhere?
Wir vergleichen mit $u_* := \sup\{\, u_s \in E \mid s \in S(\Gamma),\ \Phi_s(u_s) = u_s \,\}$.

Satz D2M: Bellmans Optimalitätsprinzip $u_* = u$

(1) Wenn Φ für jeden Startwert gegen u konvergiert, so gilt $u_* \leq u$.
Das bedeutet, lokale Optimierung ist mindestens so gut wie globale.
Das gilt insbesondere, wenn Φ isoton δ–diskontiert ist, also δ–kontraktiv.

(2) Wird in der Bellman–Gleichung überall das Supremum angenommen, etwa weil Γ lokal-endlich ist, so existieren optimale Strategien $s \in S(\Gamma)$ mit $s(x) \in \operatorname{Arg\,max}_{a \in A_x} H(x, a, u)$, also $\Phi_s(u) = u$, und es folgt $u_* \geq u$.

Beweis: (1) Für jede Strategie $s \in S$ und jeden Fixpunkt $u_s = \Phi_s(u_s)$ gilt $\Phi(u_s)(x) = \sup_{a \in A_x} H(x, a, u_s) \geq H(x, s(x), u_s) = u_s(x)$ in $x \in X^\circ$.
Zudem ist Φ isoton: $u_s \leq \Phi(u_s) \leq \Phi^2(u_s) \leq \ldots \nearrow u$. Somit gilt $u_* \leq u$.
(2) Für s gilt $u(x) = H(x, s(x), u)$, also $u = \Phi_s(u)$, somit $u_* \geq u$. QED

Bellmans Optimalitätsprinzip $u_* = u$

D242 Erläuterung

Bellmans Optimalitätsprinzip $u_* = u$ kommt recht unscheinbar daher, daher möchte ich seine anschaulich-praktische Bedeutung erläutern.

Lokale Endlichkeit des Markov–Graphen Γ vereinfacht alle Argumente. Den allgemeinen Fall behandeln wir gleich anschließend in Satz D2N.

Rückwärts / lokale Optimierung: Ein rationaler Spieler wählt in jedem aktiven Zustand $x \in X^\circ$ eine optimale Aktion $a \in A_x$, die seine erwartete Auszahlung $H(x, a, u)$ maximiert. Dies führt zur Bellman–Gleichung:

$$u(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ \max\limits_{a \in A_x} \underbrace{\sum_{y \in X} p(x,a,y)\big[r(x,a,y) + \delta u(y)\big]}_{\text{Hamilton–Funktion } H(x,\,a,\,u)} & \text{für } x \in X^\circ. \end{cases}$$

In jedem Zustand $x \in X^\circ$ stellen wir uns einen **lokalen Optimierer** vor. Dieser handelt lokal, kurzsichtig, egoistisch: Jeder optimiert für sich!

Das erinnert uns an das Prinzip der Rekursion / Rückwärtsinduktion: Optimiert wird rückwärts, aber gespielt wird immer vorwärts.

Bellmans Optimalitätsprinzip $u_* = u$

D243 Erläuterung

Vorwärts / globale Optimierung: Jede Strategie $s \in S = S(\Gamma)$ definiert eine Gewinnerwartung $u_s : X \to \mathbb{R}$ durch die Gleichung $u_s = \Phi_s(u_s)$:

$$u_s(x) = \begin{cases} v(x) & \text{für } x \in \partial X, \\ H(x, s(x), u_s) & \text{für } x \in X^\circ. \end{cases}$$

Eine **globale Optimiererin** kann somit zu jeder Strategie $s \in S(\Gamma)$ ihre Auszahlung u_s berechnen und das Supremum $u_* = \sup u_s$ bilden.
Dies geschieht global, weitblickend, kooperativ: alle $x \in X^\circ$ gemeinsam!

In jedem Zustand $x \in X$ können wir demnach bestenfalls den Gewinn $u_*(x) = \sup_{s \in S} u_s(x)$ erwarten bzw. als Auszahlung realisieren.

Das Optimalitätsprinzip $u_* = u$ besagt, dass die Bellman–Gleichung tatsächlich tut, was sie soll: Die globale Optimierung u_* und die lokale Optimierung u stimmen überein! Beide Sichtweisen werden versöhnt.

⚠ Das Optimalitätsprinzip $u_* = u$ erfordert gewisse Voraussetzungen; diese versuche ich hier möglichst schwach und allgemein zu halten.

Bellmans Optimalitätsprinzip $u_* = u$

D244 Erläuterung

Vielleicht scheint Ihnen auf den ersten Blick das Optimalitätsprinzip $u_* = u$ intuitiv-selbstverständlich und daher kaum der Rede wert.
Das wäre ein Jammer und ein Irrtum! Ist hier etwas zu beweisen?
Ja, sicher, und zur Illustration helfen Gegenbeispiele wie B1I.

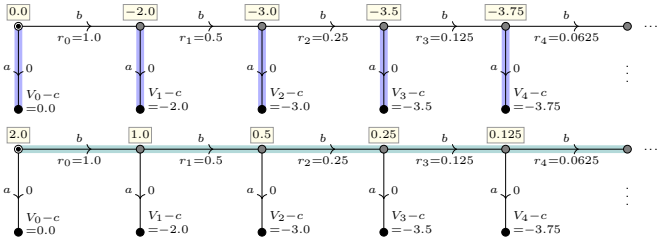
Drastische Gegenbeispiele wie das folgende sind überaus lehrreich, und ein großes Beispielrepertoire bewahrt Sie vor voreiligen Trugschlüssen.
Die gute mathematische Vorgehensweise ist wie immer komplementär: Sätze und Gegen/Beispiele ergänzen und erklären sich gegenseitig.

In §D3 skizzieren wir eine wunderschöne Anwendung aus dem Bereich des Maschinellen Lernens: Robi, der Saugroboter, versucht seine Route optimal zu planen. Meist wählt man dort einen Diskontfaktor $\delta \in [0, 1[$, da die theoretische Grundlage dann besonders einfach und solide ist.

Die Robustheit der Rechnungen ist für die Anwendung extrem wichtig, gerade deshalb diskutieren wir auch den kritischen Randfall $\delta = 1$.
Gute Beispiele bewahren Sie vor naivem Irrglauben und illustrieren eindrücklich den Nutzen möglichst starker theoretischer Werkzeuge.

Exotische Lösungen der Bellman–Gleichung

D245 Erläuterung



♦ Beispiel B1I: exotische Lösungen der Bellman–Gleichung

Unser Graph Γ habe die aktiven Zustände $x \in X^\circ = \{\, b^k \mid k \in \mathbb{N} \,\}$ mit Aktionen $A_x = \{a, b\}$ und Belohnungen $r(b^k, a) = 0$ und $r(b^k, b) = r_k$.
Terminal sind $\partial X = \{\, b^k a \mid k \in \mathbb{N} \,\}$ mit Auszahlungen $v(b^k a) = V_k - c$.

Gegeben seien hierzu $0 < r_k < v_k$ in $\mathbb{R}$ für $k \in \mathbb{N}$ mit $\sum_{k=0}^{\infty} v_k < \infty$.
Wir setzen $R_k = \sum_{i=k}^{\infty} r_i$ und $V_k = \sum_{i=k}^{\infty} v_i$ und wählen $c > V_0 - R_0 > 0$.
Die Skizze zeigt $r_k = 2^{-k}$, $R_k = 2^{1-k}$, $v_k = 2^{1-k}$, $V_k = 2^{2-k}$ und $c = -4$.

Exotische Lösungen der Bellman–Gleichung

D246 Erläuterung

Aufgabe: Zeigen Sie, dass die abgebildeten Funktionen $u_0, u_1 : X \to \mathbb{R}$ die Bellman–Gleichung lösen: Einerseits die pessimistische Lösung $u_0(b^k) = V_k - c$, andererseits die optimistische Lösung $u_1(b^k) = R_k$.

Lösung: Jede der beiden gezeigten Strategien ist lokal optimal:
(1) Für u_0 wird immer $s_0(x) = a$ gespielt, und dies ist lokal optimal.
(2) Für u_1 wird immer $s_1(x) = b$ gespielt, und dies ist lokal optimal.

Sie kennen das Problem *lokale vs globale Optimierung* aus der Analysis: Hier gilt $u_0 < u_1$; bei **globaler Optimierung** würde man also u_1 wählen.
Bei **lokaler Optimierung** erlaubt s_0 keine Möglichkeit der Verbesserung: Dies geschieht lokal, kurzsichtig, egoistisch: Jeder optimiert für sich!
Das globale Optimum erreichen wir hier nur weitblickend, kooperativ.

⚠ In Fällen wie diesem gilt Bellmans Optimalitätsprinzip nicht! Unsere allgemeinen Werkzeuge versagen hier, wir müssen genauer hinschauen.

Frustration: Das obige Beispiel kennen Sie aus dem Alltag: „Ich würde gerne etwas verbessern, doch alleine kann ich gar nichts ausrichten.” So auch hier, wenn jeder lokale Optimierer in $x \in X^\circ$ allein handelt.

Exotische Lösungen der Bellman–Gleichung

D247 Erläuterung

Bemerkung: Weitere Bellman–Lösungen sind $u(b^k) = R_k + v(b^\infty)$.
Die Konstante $v(b^\infty) \in \mathbb{R}$ ist dabei die (fiktive) Auszahlung im „unendlich fernen“ Randpunkt b^∞. Wir vereinbaren hier kurzerhand $v(b^\infty) = 0$.

Aufgabe: Gibt es weitere Lösungen neben (u_0, s_0) und (u_1, s_1)?

Lösung: Nein. Sei $u : X \to \mathbb{R}$ eine Lösung der Bellman–Gleichung und $s : X^\circ \to \{a, b\}$ eine (lokal-optimale) Strategie mit $u(x) = H(x, s(x), u)$.
Gilt $s(b^j) = a$, so folgt $s(b^i) = a$ für alle $i \leq j$. Somit wird entweder $s = s_0$ gespielt oder es existiert ein $k \in \mathbb{N}$ mit $s(b^i) = a$ für $i < k$ und $s(b^i) = b$ für alle $i \geq k$. Der Fall $k > 0$ ist nicht lokal-optimal, da die Alternative $s(b^{k-1}) = b$ lukrativer ist. Also muss $k = 0$ gelten und somit $s = s_1$.

Das Beispiel B1I ist raffiniert gebaut und notgedrungen nicht-artinsch.
Für jeden artinschen Graphen können wir dank Rückwärtsinduktion B1F die Bellman–Gleichung rekursiv lösen, und diese Lösung ist eindeutig!
Ist unser artinscher Graph Γ zudem lokal-endlich, so ist diese Lösung tatsächlich optimal dank B1H. In diesen günstigen Fällen geht alles gut.
Es gibt einen zweiten Ausweg aus unserer Misslage: Diskontierung!

Exotische Lösungen der Bellman–Gleichung

D248 Erläuterung

Aufgabe: Analysieren Sie das obige Beispiel B1I mit Diskont $\delta \in [0, 1[$.
Bestimmen Sie die (dank Satz D2K) eindeutige beschränkte Lösung $u_\delta : X \to \mathbb{R}$ der Bellman–Gleichung $u_\delta = \Phi_\delta(u_\delta)$. Was gilt für $\delta \nearrow 1$?

Lösung: (1) Ist die Strategie $s_0(b^k) = a$ weiterhin lokal optimal? Nein!
Wir vergleichen die Auszahlung $\alpha_k := r(b^k, a) + \delta v(b^k a) = \delta(V_k - c)$
mit $\beta_k := r(b^k, b) + \delta[r(b^{k+1}, a) + \delta v(b^{k+1} a)] = r_k + \delta^2(V_{k+1} - c)$.
Die Differenz $\beta_k - \alpha_k = r_k + \delta(1 - \delta)c - \delta(V_k - \delta V_{k+1})$ ist positiv für hinreichend große $k \in \mathbb{N}$, denn $r_k, c > 0$ und $V_k - \delta V_{k+1} \searrow 0$.
Es lohnt sich also (lokal!), von $s_0(b^k) = a$ auf b zu wechseln.

Im Extremfall $\delta = 1$ gilt $\beta_k - \alpha_k = r_k - v_k < 0$ für alle $k \in \mathbb{N}$, daher ist ausgehend von der Strategie s_0 keine lokale Verbesserung möglich.

(2) Ist die Strategie $s_1(b^k) = b$ weiterhin lokal optimal? Ja!
Das Argument ist für alle Diskontfaktoren $\delta \in [0, 1]$ gleich:
Jede lokale Abweichung verschlechtert die Auszahlung.
Für alle $\delta \in [0, 1[$ erhalten wir also dieselbe Strategie $s_\delta = s_1$
mit der optimistischen Gewinnfunktion $u_\delta \nearrow u_1$ für $\delta \nearrow 1$.

Bellmans Optimalitätsprinzip $u_* = u$

☺ Bei Diskontierung $\delta < 1$ gilt das Optimalitätsprinzip ganz allgemein für alle beschränkten Markov–Spiele auf beliebigen Markov–Graphen:

Satz D2N: Optimalitätsprinzip, allgemeiner diskontierter Fall

Sei (Γ, r, v) ein Markov–Spiel mit r, v beschränkt und $0 \leq \delta < 1$.
Auf $E = E_v = \{\, u \in B(X, \mathbb{R}) \mid u|_{\partial X} = v \,\}$ nutzen wir die Operatoren $\Phi(u)(x) = \sup_{a \in A_x} H(x, a, u)$ und $\Phi_s(u)(x) = H(x, s(x), u)$ für $x \in X^\circ$.
Diese konvergieren gegen $u = \Phi(u)$ bzw. $u_s = \Phi_s(u_s)$ für $s \in S(\Gamma)$.
Wir setzen punktweise $u_*(x) = \sup_{s \in S} u_s(x)$. Dann gilt $u_* = u$.

Beweis: (1) Wie in Satz D2M gilt $u_* \leq u$. (2) Wir zeigen noch $u \leq u_*$:
Sei $\varepsilon \in \mathbb{R}_{>0}$ vorgegeben. Zu jedem aktiven Zustand $x \in X^\circ$ wählen wir eine ε–optimale Aktion $s(x) \in A_x$, sodass $H(x, s(x), u) \geq u(x) - \varepsilon$ gilt.
Das bedeutet $u \leq \Phi_s(u) + \varepsilon$, dank Diskontierung $\Phi_s(u) \leq \Phi_s^2(u) + \delta\varepsilon$, per Induktion $u \leq \Phi_s^n(u) + \varepsilon \sum_{k=0}^{n-1} \delta^k = \Phi_s^n(u) + \varepsilon(1 - \delta^n)/(1 - \delta)$.
Für $n \to \infty$ erhalten wir $u \leq u_s + \varepsilon/(1 - \delta) \leq u_* + \varepsilon/(1 - \delta)$.
Letzteres gilt für alle $\varepsilon \in \mathbb{R}_{>0}$. Daraus folgt $u \leq u_*$ QED

Bellmans Optimalitätsprinzip $u_* = u$

Im diskontierten Fall $0 \leq \delta < 1$ sind alle Operatoren Φ_s und Φ kontraktiv.
Für jede Strategie $s \in S = S(\Gamma)$ können wir Φ_s iterieren und erhalten:

$$\Phi_s^n(\tilde{u}) \to u_s$$

Dies können wir anschließend global optimieren und gewinnen so:

$$u_* = \sup_{s \in S} u_s$$

Umgekehrt können wir auch erst lokal optimieren und dann iterieren.
Nach Definition gilt $\Phi = \sup_{s \in S} \Phi_s$, das heißt punktweise für alle $x \in X$:

$$\Phi(\tilde{u})(x) = \sup_{s \in S} \Phi_s(\tilde{u})(x) \overset{x \in X^\circ}{=} \sup_{a \in A_x} H(x, a, \tilde{u})$$

Wenn wir nun diesen Bellman–Operator Φ iterieren, so erhalten wir:

$$\Phi^n(\tilde{u}) \to u$$

Bellmans Optimalitätsprinzip besagt, unter der Voraussetzung $0 \leq \delta < 1$, dass wir auf beiden Wegen dasselbe Ergebnis erhalten, kurz $u_* = u$.

Zusammenfassung zum Optimalitätsprinzip

Bellmans Optimalitätsprinzip betrachten wir rückblickend als

0 Vertauschung von Iteration/Kontraktion und Maximum/Supremum.

Aus der Analysis kennen Sie viele wichtige Sätze zur Vertauschung:

1 Folgengrenzwerte $\lim_{k\to\infty} \lim_{n\to\infty} a_{k,n} \overset{?}{=} \lim_{n\to\infty} \lim_{k\to\infty} a_{k,n}$
2 Vertauschung von Ableitungen $\partial_x \partial_y f(x, y) \overset{?}{=} \partial_y \partial_x f(x, y)$
3 Ableitung und Grenzwert $\lim_{n\to\infty} \partial_x f_n(x) \overset{?}{=} \partial_x \lim_{n\to\infty} f_n(x)$
4 Vertauschung von Reihen $\sum_{k=0}^{\infty} \sum_{n=0}^{\infty} a_{k,n} \overset{?}{=} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} a_{k,n}$
5 Reihen und Grenzwert $\lim_k \sum_{n=0}^{\infty} a_{k,n} \overset{?}{=} \sum_{n=0}^{\infty} \lim_k a_{k,n}$
6 Reihen und Ableitung $\partial_x \sum_{n=0}^{\infty} f_n(x) \overset{?}{=} \sum_{n=0}^{\infty} \partial_x f_n(x)$
7 Integrale $\int_{x\in X} \int_{y\in Y} f(x, y) \,\mathrm{d}y \,\mathrm{d}x \overset{?}{=} \int_{y\in Y} \int_{x\in X} f(x, y) \,\mathrm{d}x \,\mathrm{d}y$
8 Integral und Ableitung $\partial_x \int_{y\in Y} f(x, y) \,\mathrm{d}y \overset{?}{=} \int_{y\in Y} \partial_x f(x, y) \,\mathrm{d}y$
9 Integral und Grenzwert $\lim_n \int_{x\in X} f_n(x) \,\mathrm{d}x \overset{?}{=} \int_{x\in X} \lim_n f_n(x) \,\mathrm{d}x$

Übung: Geben Sie jeweils zwei interessante Gegenbeispiele.
Anschließend wiederholen Sie Voraussetzungen, Satz und Beweis.

Konvergenz des Bellman–Operators

Wir nutzen hier möglichst einfache, zumeist hinreichende Bedingungen:
r, v beschränkt und $\delta \in [0, 1[$. Für $\delta = 1$ müssen wir genauer hinschauen!
Die Übungen geben konkretes und motivierendes Anschauungsmaterial.
Dort können Sie in konkreten Beispielen auch $\delta = 1$ durchrechnen.
Auch der folgende Satz D2O kann im Falle $\delta = 1$ genutzt werden.

Die Konvergenzfrage ist ähnlich zur Theorie komplexer **Potenzreihen**:
Für jede Potenzreihe $\sum_{n=0}^{\infty} a_n z^n$ definieren wir den **Konvergenzradius**

$$\rho := 1/\limsup \sqrt[n]{|a_n|}.$$

Für $|z| < \rho$ konvergiert die Reihe absolut, für $|z| > \rho$ divergiert sie.
Auf dem Rand $|z| = \rho$ ist alles möglich; dazu gibt es keine allgemeine Aussage. Antworten über das subtile Verhalten auf dem Kreisrand gibt die Theorie der Fourier–Reihen. Allgemein gehört die Konvergenz solcher Reihen zu den schwierigsten Fragen der Analysis.

Auf hoher See und am Konvergenzkreisrand sind wir alle in Gottes Hand.

Punktweise Konvergenz des Bellman–Operators

Satz D2O: punktweise Konvergenz des Bellman–Operators

Sei (Γ, r, v) ein Markov–Spiel, Γ lokal-endlich, r, v beschränkt und $r \leq 0$.
(1) Sei $F = \{\, u_s \in E \mid s \in S,\ \Phi_s(u_s) = u_s \,\}$. Ist $u_* = \sup F$ beschränkt, dann ist u_* der größte Fixpunkt des Bellman–Operators $\Phi : E \to E$.
(2) Sei $s \in S(\Gamma)$ eine Strategie mit $\Phi_s(u_*) = u_*$. Konvergiert $\Phi_s^n(u) \to u_*$ für jeden Startwert $u \in E$, dann konvergiert auch Φ gegen u_*.

☺ Anders als im vorigen Satz D2N verlangen wir keine Diskontierung mit $\delta \in [0, 1[$. Der Satz D2O behandelt vielmehr den Randfall $\delta = 1$.
☺ Aussage (1) ist eine schwache Fassung des Optimalitätsprinzips. Hat der Operator $\Phi : E \to E$ zudem nur einen Fixpunkt u, so gilt $u = u_*$.
☺ Aussage (2) ist eine starke Fassung analog zum obigen Satz D2M: Φ konvergiert für jeden Startwert gegen den einzigen Fixpunkt u_*.
☺ Die Einschränkung $r \leq 0$ scheint zunächst nicht besonders elegant, doch in manchen Anwendungen wie §D3 ist sie recht natürlich.

Punktweise Konvergenz des Bellman–Operators

Beweis: (0) Zu jedem Fixpunkt $u = \Phi(u)$ in E existiert eine Strategie $s \in S = S(\Gamma)$ mit $\Phi_s(u) = u$. Demnach gilt $u \in F$ und somit $u \leq u_*$

(1) Für jedes $u_s \in F$ gilt $u_* \geq u_s$, also $\Phi(u_*) \geq \Phi(u_s) \geq \Phi_s(u_s) = u_s$.
Daraus folgt $\Phi(u_*) \geq u_*$, also $u_* \leq \Phi(u_*) \leq \Phi^2(u_*) \leq \ldots \nearrow u : X \to \bar{\mathbb{R}}$.
Wir zeigen nun, dass $u : X \to \mathbb{R} \cup \{\infty\}$ beschränkt ist, also $u \in E$ gilt.
Für $c \geq \max u_*$ und $\hat{u} = v\,\mathbf{I}_{\partial X} + c\,\mathbf{I}_{X^\circ}$ gilt $u_* \leq \hat{u}$, also $\Phi(u_*) \leq \Phi(\hat{u}) \leq \hat{u}$, letzteres dank $r \leq 0$. Per Induktion erhalten wir $\Phi^n(u_*) \leq \hat{u}$, also $u \leq \hat{u}$.
Dank Stetigkeit folgt $\Phi(u) = u$. Dank (0) gilt $u \leq u_*$, also $u = u_* = \Phi(u_*)$.

(2a) Aus $\hat{u} \geq \Phi(\hat{u}) \geq \Phi(u_*) = u_*$ folgt $\hat{u} \geq \Phi(\hat{u}) \geq \Phi^2(\hat{u}) \geq \ldots \searrow \bar{u} \geq u_*$.
Dank Stetigkeit folgt $\Phi(\bar{u}) = \bar{u}$. Dank (1) folgt $\bar{u} = u_*$.
(2b) Gegeben sei $u \in E$. Wir wählen $\hat{u} = v\,\mathbf{I}_{\partial X} + c\,\mathbf{I}_{X^\circ} \geq u, u_*$ wie oben.
Dann gilt $\Phi_s(u) \leq \Phi(u) \leq \Phi(\hat{u})$, per Induktion $\Phi_s^n(u) \leq \Phi^n(u) \leq \Phi^n(\hat{u})$.
Nach Voraussetzung gilt $\Phi_s^n(u) \to u_*$. Dank (2a) gilt auch $\Phi^n(\hat{u}) \to u_*$.
Daraus folgt die behauptete Konvergenz $\Phi^n(u) \to u_*$. QED

Punktweise Konvergenz des Bellman–Operators

☺ Der Fixpunktsatz von Banach fordert eine δ–Kontraktion, also eine starke Voraussetzung, garantiert dafür aber gleichmäßige Konvergenz mit expliziter Fehlerschranke als extrem nützliche Schlussfolgerung.
Das ist der Idealfall, an dem wir uns orientieren möchten.

Bellmans Optimalitätsprinzip formulieren wir entsprechend, je nach Anwendung, unter stärkeren und schwächeren Voraussetzungen.
Der bequemste Fall ist die Diskontierung mit $\delta \in [0, 1[$ (Satz D2N).
Diese strenge Voraussetzung ist leider nicht immer gegeben.

Für Bellmans Optimalitätsprinzip D2M, genauer für die Ungleichung $u_* \leq u$, genügt uns bescheidener bereits die punktweise Konvergenz $\Phi^n(\tilde{u}) \to u$. Genau hierfür bietet Satz D2O ein hinreichendes Kriterium.
Das kann in Anwendungen mit $\delta = 1$ ein praktisches Hilfsmittel sein.

Aus Erfahrung empfehle ich, diese schönen Sätze zunächst einmal zur Kenntnis zu nehmen und dann konkrete Anwendungen zu untersuchen.
In einem zweiten Durchgang möchten Sie dann stärkere Werkzeuge und werden gerne auf hilfreiche Sätze und Beweise zurückkommen.

Ausblick auf erfolgreiche Anwendungen

☺ In Markov–Spielen steckt viel schöne Mathematik!
Sie sind sehr einfach gebaut und doch vielseitig einsetzbar.
Zur Programmierung nutzen wir vor allem endliche Markov–Graphen, doch auch der unendliche Fall ist mathematisch reizvoll und nützlich.

Markov–Spiele (aka Markov–Entscheidungsprozesse) finden Sie daher in zahlreichen Anwendungen der Ökonomik und Finanzmathematik:

Investition vs Konsum optimieren: Geld sparen oder ausgeben?
Anbau vs Abbau optimieren: Bäume pflanzen oder fällen?
Anlage optimieren: Aktienportfolio optimal steuern?
Allgemein: kurzfristiger oder langfristiger Nutzen?

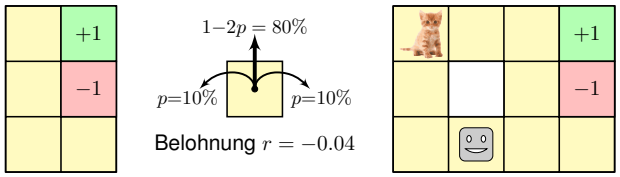
Markov–Spiele finden Sie ebenso in der künstlichen Intelligenz, insbesondere maschinellem Lernen, etwa bestärkendem Lernen:

Logistik / Routenplanung: Wege minimieren, Nutzen maximieren.
Strategischer Akteur / autonomes Fahrzeug in stochastischer Umwelt: Aus Erfahrung lernen, Strategie optimieren, Exploration vs Exploitation.
Aufzugsteuerung: Ja, es soll auch intelligente Aufzüge geben!

Robi the Robot und Machine Learning

D301

Robi, der Saugroboter, lebt in einer 2×3–Wohnung (links), später in einer 4×3–Wohnung mit einem Rundgang (rechts).



Nutzen $\sum_{t=0}^{T-1} \delta^t r_t + \delta^T v(x_T)$ mit $x_T \in \partial X$ und Diskont $\delta \in [0,1]$.

In jedem Zeitschritt wählt Robi eine Richtung; diese fährt er mit Wkt 80%, mit Wkt $p = 10\%$ links oder rechts dazu, weil z.B. die Katze ihn schubst. Falls sich in Fahrtrichtung eine Wand befindet, bleibt er einfach stehen.

Jeder Zeitschritt bringt eine Belohnung $r \in \mathbb{R}$, etwa $r < 0$ für Verbrauch, egal ob oder wohin er fährt. Wenn er die Ladestation oder die Treppe erreicht, endet seine Fahrt mit der Auszahlung $+1$ bzw. -1.

Robi the Robot und Machine Learning

D302 Erläuterung

Dieses beliebte Lehrbeispiel des maschinellen Lernens stammt aus
📖 S. Russell, P. Norvig: *Artificial Intelligence: A Modern Approach.* Addison Wesley (3rd ed.) 2016 (Kapitel 17: Making complex decisions)

In den letzten Jahren hat dieser Ansatz großen Zulauf erhalten: Es gibt einen Massenmarkt, denn Geräte sind preiswert und alltagstauglich. Zudem genügen in vielen einfachen Projekten bereits die Grundlagen der Theorie für beachtliche Anwendungserfolge. Dank Softwarepaketen ist auch die Hürde in der Programmierung inzwischen recht gering.

🙂 Sie können und sollen mit diesem Übungsbeispiel konkret rechnen und numerische Erfahrungen sammeln: Sie finden eine Umsetzung als Tabellenkalkulation unter `eiserm.de/lehre/Spieltheorie/Robi.ods`.

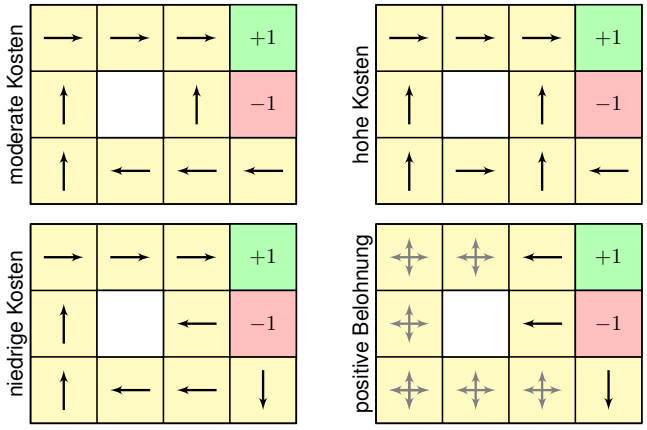
Selbst in diesem einfachen Beispiel sind die Strategien überraschend. Spielen Sie etwas mit den Parametern, Belohnung und Auszahlungen! Wie ändert das Robis Verhalten? Begründung? Interpretation? Wann und warum konvergiert der Bellman–Operator?

Übung: Berechnen / implementieren Sie ebenso die 4×3–Wohnung.

Robi the Robot und Machine Learning

D303

Robis Verhalten ist erstaunlich komplex, abhängig von den Kosten r:



Robi the Robot und Machine Learning

D304 Erläuterung

🙂 Robi handelt strategisch, er sucht die Balance zwischen Belohnung und Risiko. Das ist durchaus typisch für viele realistische Anwendungen.

Bei moderaten Kosten lohnt sich für Robi der Umweg des Rundgangs. Bei hohen Kosten ist der direkte Weg besser. Probieren Sie es aus!

Bei geringen Kosten, r knapp unter Null, kann Robi sogar garantieren, nie die Treppe hinunterzufallen: Er kann ihr ganz vorsichtig ausweichen, auch wenn er sich öfter die Nase an der Wand stößt und länger fährt.

Der Fall $r = 0$ bietet die ersten Überraschungen: Probieren Sie es!

Bei positiver Belohnung $r > 0$ will Robi nie aufhören, auch nie aufladen, da er beliebig Nutzen generieren kann, indem er fröhlich umherfährt. Er ist wie auf Drogen, *high on reward*, als gäbe es kein morgen mehr.

Bei zu hohen Kosten jedoch ist Robis Leben so miserabel, dass er stets den nächsten Ausgang wählt, notfalls stürzt er sich die Treppe hinunter. Das ist zwar traurig, aber unter diesen Bedingungen das beste für ihn.

⚠ Die Wahl des Belohnugssystems entscheidet über das Verhalten. Alle Trainer / Lehrer / Eltern wissen das: *Choose rewards wisely!*

Gewinniteration / *value iteration*

D305

Zu (Γ, r, v) betrachten wir $\Phi : E \to E$ sowie $\Phi_s : E \to E$ für $s \in S(\Gamma)$. Wir nehmen an, dass wir zu Φ eine Kontraktionskonstante k kennen.

Algo D3A: Gewinniteration / *value iteration*

Global: Markov–Spiel (Γ, r, v), Kontraktionskonstante $k \in [0,1[$
Eingabe: Eine initiale Erwartung $u \in E$ und eine Toleranz $\varepsilon \in \mathbb{R}_{>0}$
Ausgabe: Eine ε–optimale Erwartung $u \in E$ und Strategie $s \in S(\Gamma)$

1: repeat
2: kopiere $u' \leftarrow u$ und aktualisiere $u \leftarrow \Phi(u)$
3: until $\frac{k}{1-k}|u - u'| \le \varepsilon$
4: for $x \in X^\circ$ do wähle $s(x) \in \operatorname{Arg\,max}_{a \in A_x} H(x,a,u)$
5: return (u, s)

Aufgabe: Warum endet dieser Algorithmus und ist korrekt?

Lösung: Beides verdanken wir Banachs Fixpunktsatz D2A!

⚠ Die abgelesene Strategie s ist im Allgemeinen noch nicht optimal! Um dies zu garantieren, muss die Näherung hinreichend gut sein.

Strategie vs Gewinniteration

D306 Erläuterung

Aufgabe: Wie klein sollten wir $\varepsilon \in \mathbb{R}_{>0}$ wählen, damit wir aus jeder ε–Näherung $\tilde{u}$ alle optimalen Strategien $s \in S(\Gamma)$ ablesen können?

Lösung: Sei $u = \Phi(u)$ die exakte Lösung. Gegeben ist $|\tilde{u} - u| \le \varepsilon$.
Zu $x \in X^\circ$ sei $\{\, u(x) - H(x,a,u) \mid a \in A_x \,\} = \{0 = \lambda_x^0 < \lambda_x^1 < \dots\}$.
Wir nennen $\lambda := \min\{\, \lambda_x^1 \mid x \in X^\circ \,\}$ die **Spektrallücke** von (Γ, r, v).
Für jede Aktion $a \in A_x$ ist $\tilde{u}(x) - H(x,a,\tilde{u})$ gegeben durch

$$\big[\tilde{u}(x) - u(x)\big] + \big[u(x) - H(x,a,u)\big] + \big[H(x,a,u) - H(x,a,\tilde{u})\big]$$

Für $a \in A_x$ sub/optimal gilt $\tilde{u}(x) - H(x,a,\tilde{u}) \ge \lambda - 2\varepsilon$ bzw. $\le 2\varepsilon$.
Für $4\varepsilon < \lambda$ können wir so aus $\tilde{u}$ alle optimalen Strategien s ablesen.

Aufgabe: Die Schranke λ nutzt die exakte, unbekannte Lösung u. Können Sie eine Schranke finden, die nur die Näherung $\tilde{u}$ nutzt?

Lösung: Zu $x \in X^\circ$ und $A_x = \{a_0, \dots, a_\ell\}$ sei $\mu_x^i = \tilde{u}(x) - H(x, a_i, \tilde{u})$ sortiert gemäß $\mu_x^0 \le \mu_x^1 \le \dots \le \mu_x^\ell$. Wir setzen $\mu := \min\{\, \mu_x^1 \mid x \in X^\circ \,\}$.
Gilt $2\varepsilon < \mu$, so können wir die optimale Strategie s aus $\tilde{u}$ ablesen:
Für $a \in A_x$ sub/optimal gilt $\tilde{u}(x) - H(x,a,\tilde{u}) \ge \mu > 2\varepsilon$ bzw. $\le 2\varepsilon$.

Strategieiteration / *policy iteration*

D307

Algo D3B: Strategieiteration / *policy iteration*

Global: Das endliche Markov–Spiel (Γ, r, v)
Eingabe: Eine initiale Strategie $s \in S(\Gamma)$
Ausgabe: Eine optimale Strategie $s \in S(\Gamma)$ und ihre Erwartung $u \in E$

1: repeat
2: löse $u = \Phi_s(u)$ und setze done $\leftarrow$ true
3: for $x \in X^\circ$ do
4: if $\max_{a \in A_x} H(x,a,u) > H(x, s(x), u)$ then
5: wähle $s(x) \in \operatorname{Arg\,max}_{a \in A_x} H(x,a,u)$ und setze done $\leftarrow$ false
6: until done
7: return (s, u)

Aufgabe: Warum endet dieser Algorithmus und ist korrekt?

Lösung: Die Strategiemenge $S = S(\Gamma)$ ist endlich, und es gilt $u_0 \lneq u_1 \lneq u_2 \lneq u_3 \lneq \dots$ bis schließlich $u = \Phi_s(u) = \Phi(u)$.
Ist Φ zudem kontraktiv, so ist u die eindeutige Lösung.

kombinierte Strategie-und-Gewinniteration

D308 Erläuterung

Oft kombiniert man die Strategieiteration mit der Gewinniteration:

Algo D3C: kombinierte Strategie-und-Gewinniteration

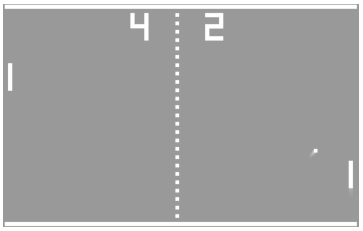
Global: Das Markov–Spiel (Γ, r, v)
Eingabe: Initiale Strategie $s \in S(\Gamma)$ und Erwartung $u \in E$
Ausgabe: Eine verbesserte Strategie s und aktualisierte Erwartung u

1: repeat
2: setze $u \leftarrow \Phi_s^m(u)$ und done $\leftarrow$ true
3: for $x \in X^\circ$ do
4: if $\max_{a \in A_x} H(x,a,u) > H(x, s(x), u)$ then
5: wähle $s(x) \in \operatorname{Arg\,max}_{a \in A_x} H(x,a,u)$; setze done $\leftarrow$ false
6: until done
7: return (s, u)

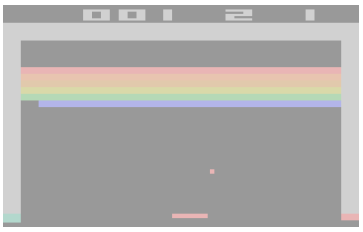
Übung: Wenn Sie möchten, können Sie diese Methode (und Varianten) mathematisch untersuchen, implementieren und experimentell erproben. Wird die Theorie schwächer, so muss eben die Erfahrung ausgleichen.

Pinball Wizard (frei nach *The Who*, 1969)

D309



Pong
(Atari 1972)
zwei Spieler
deterministisch



Breakout
(Atari 1976)
ein Spieler
probabilistisch

Pinball Wizard (frei nach *The Who*, 1969)

D310
Erläuterung

Die **Künstliche Intelligenz** feiert inzwischen beachtliche Erfolge in Alltagsprodukten, von Spracherkennung bis autonomen Fahrzeugen.

Durch **bestärkendes Lernen** [*reinforcement learning*] und ähnlichen Algorithmen kann ein strategischer Akteur (Agent, Computer, Roboter) selbständig aus Erfahrung lernen und immer bessere Strategien finden.

Ein amüsant-spektakuläres Beispiel sind Arcade-Spiele: Das Startup **DeepMind** (`en.wikipedia.org/wiki/DeepMind`) gehört inzwischen zu Google und hat diese grundlegende Idee sehr erfolgreich umgesetzt.

Mnih et al: *Human-level control through deep reinforcement learning.* Nature 518 (2015) 529–533, `www.nature.com/articles/nature14236`

Als zweiminütiges Video `youtu.be/V1eYniJ0Rnk` oder ausführlicher als schön geschriebener Artikel `www.sciencemag.org/news/2015/02/artificial-intelligence-bests-humans-classic-arcade-games`

Arcade-Spiele sind ein gutes Testfeld: weder zu einfach noch zu schwer. Die Belohnungen sind zumeist dicht genug, um das Lernen zu fördern.

Pinball Wizard (frei nach *The Who*, 1969)

D311
Erläuterung

Einige Besonderheiten des **bestärkenden Lernens**:

- **Unsupervised:** Es gibt keine Anleitung und keinen Trainer
  Der Spieler lernt nur aus den Signalen seiner Belohnungen.
- **Exploration:** Aktionen beeinflussen zukünftige Informationen.
  Der Spieler muss aktiv interagieren und erforschen.
- **Delayed Feedback:** Aktionen bewirken spätere Belohnungen.
  Der Zeitablauf des Spiels ist ein wesentlicher Faktor.

Der Spieler muss seine Aktionen wählen und hierzu einen Kompromiss finden zwischen der **Erkundung** [*exploration*] neuer Möglichkeiten und der **Nutzung** [*exploitation*] bewährter Wege. Wie im richtigen Leben!

Die **sofortigen Belohnungen** leiten den Spieler in seinem Lernprozess: Er sucht die Balance zwischen kurzfristigem und langfristigem Nutzen. Nur mit gutem **Belohnungssignal** macht das Lernen Fortschritte.

Bei **allzu seltenen Belohnung** r oder endlastigen Auszahlungen v sind die Fortschritte recht langsam. In großen Umgebungen (Spielgraphen) müssen die Zustände geeignet zusammengefasst / abstrahiert werden.

Pinball Wizard (frei nach *The Who*, 1969)

D312
Erläuterung

Das grundlegende Lehrbuch zu bestärkendem Lernen:

Richard S. Sutton, Andrew G. Barto: *Reinforcement Learning.* The MIT Press (2nd ed.) 2018 (hier speziell §6.5: Q–learning), online verfügbar unter `incompleteideas.net/book/RLbook2018.pdf`

Das folgende Vorlesungsskript ist deutlich kürzer und knapper:

Csaba Szepesvári: *Algorithms for Reinforcement Learning*, online `sites.ualberta.ca/~szepesva/papers/RLAlgsInMDPs-lecture.pdf`

Es gibt exzellente Online-Kurse zu diesem Thema, zum Beispiel von David Silver: *Intro to Reinforcement Learning*, `youtu.be/2pWv7GOvuf0`.

Ich kann der Versuchung nicht widerstehen und will Ihnen einen ganz kurzen Ausblick dieses aktuellen und faszinierenden Gebiets skizzieren: Es verbindet Informatik und Mathematik, Lerntheorie und Spieltheorie, und ist im Ingenieurwesen und seinen Anwendungen angekommen.

Bestärkendes Lernen / *reinforcement learning*

D313
Erläuterung

☺ Angenommen, wir kennen zu (Γ, r, v) die Gewinnfunktion $u = \Phi(u)$. Im Zustand $x \in X^\circ$ bewerten wir die Qualität jeder Aktion $a \in A_x$ durch

$$q(x,a) = H(x,a,u) := \textstyle\sum_{y \in X} p(x,a,y)\big[r(x,a,y) + \delta u(y)\big].$$

Optimalitätsprinzip: Jede optimale Aktion $a \in A_x$ maximiert $q(x,a)$.

Angenommen, der Spieler kennt anfangs nur den Markov–Graphen Γ. Wie kann er die Bewertungen q und u spielend-explorativ erlernen?

☺ Als Näherung für q nutzt er sein bisheriges **Erfahrungswissen**:

$$Q : A \to \mathbb{R} : (x,a) \mapsto Q(x,a)$$

Jeden aktiven Zustand $x \in X^\circ$ bewertet er näherungsweise durch

$$U : X \to \mathbb{R} : x \mapsto U(x) = \max_{a \in A_x} Q(x,a).$$

Im aktiven Zustand $x \in X^\circ$ wählt der Spieler eine Aktion $a \in A_x$ und gelangt zum Zustand y mit Belohnung $R = r(x,a,y)$. Er aktualisiert

$$Q(x,a) \leftarrow (1-\alpha)\underbrace{Q(x,a)}_{\text{alte}} + \alpha\underbrace{\big[R + \delta U(y)\big]}_{\text{neue Erfahrung}} \quad \text{und} \quad U(x) \leftarrow \max_{a \in A_x} Q(x,a).$$

Bestärkendes Lernen / *reinforcement learning*

D314
Erläuterung

Der Spieler lernt durch seine Erfahrung mit der **Lernrate** $\alpha \in]0,1[$. Sein **Erfahrungswissen** Q und U wird dabei wie folgt aktualisiert:

Algo D3D: Aktualisierung von Q und U

Global: Die Bewertungen $Q : A \to \mathbb{R}$ und $U : X \to \mathbb{R}$
Eingabe: Die Aktion $a : x \to y$ mit Belohnung $R \in \mathbb{R}$.

1: Aktualisiere $Q(x,a) \leftarrow (1-\alpha)Q(x,a) + \alpha\big[R + \delta U(y)\big]$
2: Aktualisiere $U(x) \leftarrow \max\{U(x), Q(x,a)\}$

Ist y terminal, so erhält er zudem $V = v(y)$ und aktualisiert $U(y) \leftarrow V$. Damit endet diese **Episode**, die Trajektorie dieses Spieldurchgangs. Zur weiteren Verbesserung werden noch weitere Episoden gespielt.

Der hier beschriebene Algorithmus heißt **Q–Lernen** [*Q–learning*]. Der Buchstabe „Q“ steht für die Funktion $Q : A \to \mathbb{R}$, die die Qualität der Aktionen bewertet und in diesem Algorithmus die Hauptrolle spielt.

☺ Erhofft ist die rasche **Konvergenz** $Q_t \to q$ und $U_t \to u$ für $t \to \infty$. Dazu gibt es mathematische Sätze und eindrückliche praktische Erfolge.

Bestärkendes Lernen / *reinforcement learning*

D315
Erläuterung

Maschinelles Lernen nutzt Algorithmen, die aus Erfahrungen lernen, mit Hilfe statistischer Methoden auf Trainingsdaten, hier Spieldaten.

Zwei grundlegende Algorithmen sind **Gewinniteration** [*value iteration*] und **Strategieiteration** [*policy iteration*], wie oben zusammengefasst. Sie setzen allerdings voraus, dass das ganze Spiel (Γ, r, v) bekannt ist; dann stehen alle genannten mathematischen Werkzeuge zur Verfügung.

Beim **bestärkenden Lernen** [*reinforcement learning*] erlernt der Agent selbständig eine Strategie. Ihm wird anfangs nur der Graph Γ gegeben, aber keine Hinweise, welche Aktion in welcher Situation die beste wäre. Noch realistischer: Er muss auch den Graphen Γ erst selbst erkunden!

Reines Beobachten genügt in diesem Lernprozess nicht, Informationen gewinnt der Spieler nur durch **Interaktion**. Alle seine Aktionen $a : x \to y$ führen zu Belohnungen, aus diesen approximiert er die Qualität $Q(x,a)$ der Aktion und den Nutzen $U(x) = \max_{a \in A_x} Q(x,a)$ des Zustands x.

Der **Algorithmus** stammt von Watkins (1989) und Bozinovski (1981). Seine **Konvergenz** wurde bewiesen von Watkins und Dayan (1992).

Bestärkendes Lernen / *reinforcement learning*

D316
Erläuterung

Die Grundidee stammt aus der **Psychologie** und wurde bereits seit den Anfängen der Kybernetik verwendet, so etwa von Marvin Minsky in seiner Dissertation: *Neural Nets and the Brain Model Problem.* (1954)

Bestärkendes Lernen ist inzwischen ein großes interdisziplinäres Gebiet und verbindet Informatik, Optimierung und Ökonomik mit Psychologie und Neurowissenschaften. Verfolgt werden dabei zwei Ziele:

(1) Bei realen Lebewesen soll das beobachtete (Lern)Verhalten durch geeignete Modelle möglichst gut erklärt werden (deskriptiv, explikativ).

(2) Künstliche Agenten sollen mit ihrer Umwelt strategisch interagieren und daraus möglichst effizient lernen (normativ, konstruktiv).

Übung: Wenn Sie gerne programmieren, dann können Sie unsere Miniaturbeispiele implementieren und durch bestärkendes Lernen lösen.

Allgemein lohnt sich hierbei eine möglichst generische Problemlösung, die allgemein Markov–Spiele behandelt. Sie können dabei viel lernen!

Wenn Sie mit den Grundideen (und auch Problemen) vertraut sind, dann lohnt sich ein Blick auf die umfangreichen Softwarepakete.