Optimale Kontrolle

TECHNISCHE UNIVERSITÄT DARMSTADT

Vorab: Einordungen weiterer Begriffe

diskrete Zustände	vs.	kontinuierliche Zustände
endliche Zustandsmenge	vs.	unendliche Zustandsmenge
diskrete Zeit	vs.	kontinuierliche Zeit
endlicher Horizont	vs.	unendlicher Horizont
open-loop	vs.	closed-loop Optimierung

Optimale Kontrolle: Einführung

nach Dimitri P. Bertsekas

TECHNISCHE UNIVERSITÄT DARMSTADT

Basissystem

- zeitdiskretes dynamisches System
- Kostenfunktion ist additiv über die Zeit
- das System besitzt Zustandsvariablen, die sich im Lauf der Zeit verändern, unter dem Einfluss eigener Entscheidungen und des Zufalls

- das System hat die Form

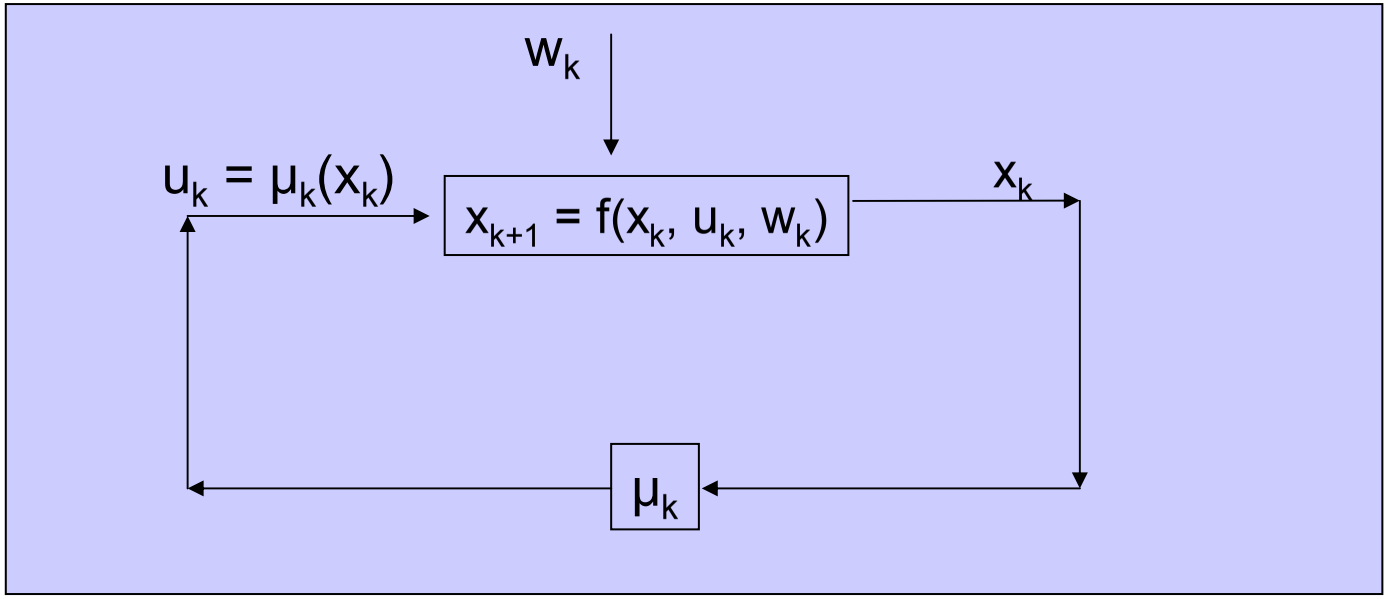
$$x_{k+1} = f_k(x_k, u_k, w_k) \quad \text{mit} \quad k = 0,\ldots,N-1, \text{ wobei}$$

  - k indiziert die diskrete Zeit
  - x_k ist der Zustand des Systems (mit Zustandsmenge S_k) und summiert vergangene Informationen, die für die Zukunft relevant sind
  - $u_k \in U_k(x_k) \subseteq C_k$ sind so genannte Kontrollvariablen. Die Menge $U_k(x_k)$ der möglichen Aktionen („Züge“) hängt vom aktuellen Zustand ab.
  - $w_k \in D_k$ ist ein Zufallsparameter, eine „Störung“. w_k gehorcht einer Wahrscheinlichkeitsverteilung $P_k(\cdot \mid x_k, u_k)$, die von x_k und u_k abhängen kann, jedoch nicht von früheren Realisierungen $w_{k-1}, \ldots, w_0$.
  - N ist der Horizont, die Anzahl der Zeitschritte, die wir untersuchen

Optimale Kontrolle: Einführung

TECHNISCHE UNIVERSITÄT DARMSTADT

Basissystem



Optimale Kontrolle: Einführung

TECHNISCHE UNIVERSITÄT DARMSTADT

Basissystem I

- das System hat die Form

  ...

  - f_k ist eine Funktion, die das System und den Mechanismus, mit dem ein Zustand in einen nächsten überführt wird, beschreibt.
  - die dazugehörige Kostenfunktion ist additiv, d.h. die Kosten $g_k(x_k,u_k,w_k)$ werden akkumuliert. Zusätzlich gibt es Abschlußkosten $g_N(x_N)$ zum Zeitpunkt N. Die Gesamtkosten werden beschrieben als

$$g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)$$

  - Eine **Politik** ist eine Funktionenfolge

    $\pi = \{\mu_0,\ldots, \mu_{N-1}\}$

    wobei μ_k Zustände x_k auf Kontrollvariablen $u_k = \mu_k(x_k)$ so abbildet, dass für alle $x_k \in S_k$ gilt: $\mu_k(x_k) \in U_k(x_k)$.

Optimale Kontrolle: Einführung

TECHNISCHE UNIVERSITÄT DARMSTADT

Basissystem

- Wegen des Zufallseinflusses w_k sind die Kosten im Allgemeinen eine Zufallsvariable und können deshalb nicht sinnvoll optimiert werden. Wir betrachten deshalb das Problem der erwarteten Kosten für eine bei x_0 startende Politik π:

$$J_\pi(x_0) = E\left\{ g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k) \right\}$$

  wobei die Erwartung sich auf die (möglicherweise implizite) Gesamtverteilung der Zufallsvariablen bezieht.
- Eine optimale Politik π* ist eine, die die optimalen Kosten $J^*(x_0)$ minimiert:

$$J^*(x_0) = J_{\pi^*}(x_0) = \min_{\pi \in \Pi} \quad J_\pi(x_0)$$

Optimale Kontrolle: Einführung

TECHNISCHE UNIVERSITÄT DARMSTADT

Beispiel I, Optimierung einer Schach-Turnier-Strategie

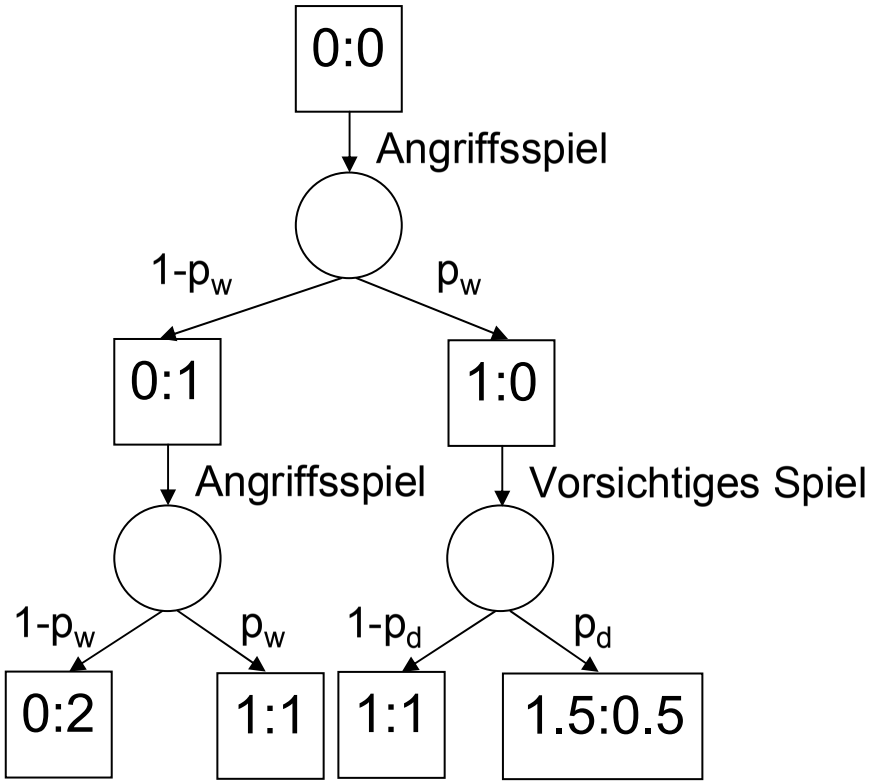
- Ein Spieler muss ein Match mit 2 Partien spielen. Jede Partie hat einen der 3 Ausgänge „win“, „loss“, and „draw“.
- Beim Spielstand von 1:1 wird so lange gespielt, bis einer eine Partie gewinnt (sudden-death)
- Unser Spieler habe 2 Spielmodi:
  - „Vorsichtiges Spiel“. Hier bekommt er ein Remis (draw) mit Wahrscheinlichkeit $p_d > 0$ und verliert mit Wahrscheinlichkeit $1-p_d$. Kein Gewinn.
  - „Angriffsspiel“. Gewinnwahrscheinlichkeit p_w und Verlustwahrscheinlichkeit $1-p_w$. Nie Remis.
- Sobald der Sudden-death beginnt, sollte der Spieler Angriffsspiel zeigen.

Frage: Wie sollte er sich in den ersten beiden Runden verhalten?
Frage: Wie groß sind seine Gewinnchancen, wenn z.B. $p_w = 0.45$ und $p_d = 0.9$?

Optimale Kontrolle: Einführung

TECHNISCHE UNIVERSITÄT DARMSTADT

Beispiel I, Optimierung einer Schach-Turnier-Strategie



Nach 2 Spielen:

- W. für Matchwin ist $p_w \cdot p_d$
- W. für Matchverlust ist $(1-p_w)^2$
- W. für Gleichstand ist $p_w(1-p_d)+(1-p_w)p_w$. Danach ist die Gewinnwahrscheinlichkeit p_w.
- Somit: Gewinnw. dieser Strategie:
  $p_w p_d + p_w(p_w(1-p_d)+(1-p_w)p_w)$

- Für $p_w = 0.45$ und $p_d = 0.9$ ergibt sich eine Gesamtgewinnwahrscheinlichkeit von ca. 0.53.

Optimale Kontrolle: Einführung

TECHNISCHE UNIVERSITÄT DARMSTADT

Beispiel I, Optimierung einer Schach-Turnier-Strategie

´Betrachte nun alle open-loop-Politiken (sind nur 4):

Bezeichne W die Wahrscheinlichkeit, das Match zu gewinnen.

1. Spiele vorsichtig in beiden Spielen. → $W = p_d^2 p_w$.
2. Zeige beide Male Angriffsspiel → $W = p_w^2 + 2\, p_w^2(1-p_w) = p_w^2(3 - 2p_w^2)$
3. Spiele erst auf Angriff, dann vorsichtig → $W = p_w\, p_d + p_w^2(1-p_d)$
4. Spiele erst vorsichtig, dann auf Angriff → $W = p_w\, p_d + p_w^2(1-p_d)$

mit ein bisschen „Herumrechnen“ ergibt sich:

$$W = p_w^2 + p_w(1-p_w)\, \max(2p_w\, p_d)$$

Für $p_w = 0.45$ und $p_d = 0.9$ ergibt sich eine Gesamtgewinnwahrscheinlichkeit von ca. 0.425.

Die Differenz 0.53 – 0.425 nennt man den “Value of Information”.

Optimale Kontrolle: Einführung

TECHNISCHE UNIVERSITÄT DARMSTADT

Dynamic Programming (DP)

Grundlage für DP ist das **Optimalitätsprinzip**

Sei $\pi^* = \{\mu^*_0, \ldots, \mu^*_{N-1}\}$ eine optimale Strategie. Nehmen wir an, aufgrund der ersten i Schritte wird Zustand x_i erreicht. Betrachten wir nun das Teilproblem von Zeitpunkt i bis N. Die restlichen Kosten bis zum Zeitpunkt N sind:

$$E\left\{ g_N(x_N) + \sum_{k=i}^{N-1} g_k(x_k, \mu_k(x_k), w_k) \right\}$$

Dann ist die abgeschnittene Politik $\{\mu^*_i, \ldots, \mu^*_{N-1}\}$ optimal für das Restproblem.

Optimale Kontrolle: Einführung

TECHNISCHE UNIVERSITÄT DARMSTADT

Der DP-Algorithmus (Dynamic Programming Algorithmus)

Für jeden Startzustand x_0 sind die optimalen Kosten $J^*(x_0)$ gleich $J_0(x_0)$, welches durch den letzten Schritt des folgenden Algorithmus berechnet wird, der sich rückwärts von Periode N-1 zu Periode 0 in der Zeit bewegt:

$$J_N(x_N) = g_N(x_N),$$

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} E_{w_k}\{g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k))\},$$

$$k = 0,1,\ldots,N-1$$

Die Erwartung bezieht sich hier auf Wahrscheinlichkeitsverteilung von w_k, die von x_k und u_k abhängt. Wenn $u^*_k = \mu^*_k$ die rechte Seite von (*) für alle x_k und k minimiert, ist die Politik $\pi^* = \{\mu^*_0,\ldots, \mu^*_{N-1}\}$ eine optimale Strategie.

Der DP-Algorithmus (Dynamic Programming Algorithmus)

Unter der Annahme, dass alle involvierten Wahrscheinlichkeitsverteilungen endlich und diskret sind, ergibt sich folgender einfach Korrektheitsbeweis über Induktion.

Bezeichne dafür für jede gültige Politik π und jedes k

$$\pi^k = \{\mu_k, \ldots, \mu_{N-1}\}$$

die Restpolitik von π der letzten Perioden.
Für k = 0,..,N-1 seien

$$J^*_k(x_k) = \min_{\pi_k} \underset{w_k,\ldots,w_{N-1}}{E} \left\{ g_N(x_N) + \sum_{i=k+1}^{N-1} g_i(x_i, \mu_i(x_i), w_i) \right\}$$

die optimalen Kosten für das (N-k)-stufige Restproblem, welches in x_k zum Zeitpunkt k startet und zum Zeitpunkt N endet.

Der DP-Algorithmus (Dynamic Programming Algorithmus)

$$J^*_N(x_N) = g_N(x_N),$$

$$J^*_k(x_k) = \min_{(\mu_k, \pi^{k+1})} \underset{w_k, \ldots, w_{N-1}}{E} \left\{ g_k(x_k, \mu_k(x_k), w_k) + g_N(x_N) + \sum_{i=k+1}^{N-1} g_i(x_i, \mu_i(x_i), w_i) \right\}$$

$$= \min_{\mu_k} \underset{w_k}{E} \left\{ g_k(x_k, \mu_k(x_k), w_k) + \min_{\pi^{k+1}} \left[\underset{w_{k+1}, \ldots, w_{N-1}}{E} \left\{ g_N(x_N) + \sum_{i=k+1}^{N-1} g_i(x_i, \mu_i(x_i), w_i) \right\} \right] \right\}$$

$$= \min_{\mu_k} \underset{w_k}{E} \left\{ g_k(x_k, \mu_k(x_k), w_k) + J^*_{k+1}(g_k(x_k, \mu_k(x_k), w_k)) \right\}$$

$$= \min_{u_k \in U_k(x_k)} \underset{w_k}{E} \left\{ g_k(x_k, u_k, w_k) + J_{k+1}(g_k(x_k, u_k, w_k)) \right\}$$

$$= J_k(x_k)$$

Optimale Kontrolle: Einführung

TECHNISCHE UNIVERSITÄT DARMSTADT

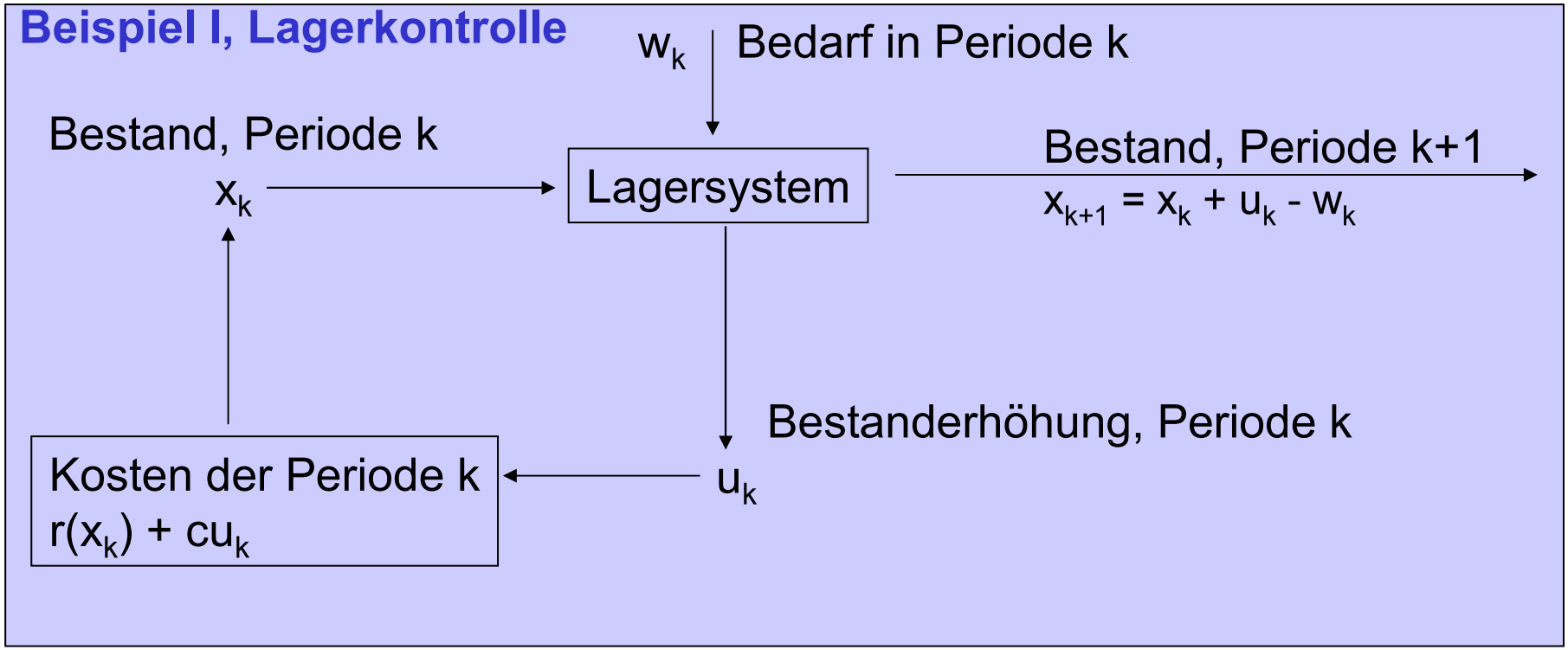
Beispiel I, Lagerkontrolle

- Im Verlauf von N Zeitschritten wird zu jedem Zeitschritt eine bestimmte Anzahl eines bestimmten Gutes von Außen geordert. Wir müssen den Lagerbestand möglichst klein halten, gleichzeitig aber verhindern, dass Anforderungen nicht erfüllt werden können.
  - x_k ist der Lagerbestand zu Beginn von Periode k
  - u_k ist die Menge des Gutes, mit dem wir das Lager nach Periode k auffüllen.
  - w_k Bedarf während der k-ten Periode mit gegebener Zufallsverteilung, $w_0,...,w_{N-1}$ seien unabhängige Zufallsvariablen.
  - Bedarf, den wir nicht decken können wird als negativer Lagerbestand fortgeführt, und wird so bald wie möglich bedient.
  - - Kosten $r(x_k)$ repräsentieren Strafen für positiven und negativen Bestand.
    - $R(x_N)$ sind Endkosten für Lagerbestand am Ende
    - cu_k sind Bestellkosten, wobei c die Kosten pro Einheit des Gutes sind.
  - Der Lagerbestand entwickelt sich also wie folgt:

$$x_{k+1} = x_k + u_k - w_k$$

Einführung, Beispiele

TECHNISCHE UNIVERSITÄT DARMSTADT



TECHNISCHE UNIVERSITÄT DARMSTADT

Beispiel I, Lagerkontrolle

Rest-Teilproblem der Länge 1:

Zu Beginn der Periode N-1 sei der Lagerbestand x_{N-1}. Klar: Egal was vorher war, sollte man mittels $u_{N-1} \geq 0$ die Orderkosten plus die erwarteten Lager/Strafkosten minimieren:

$$\min_{u_{N-1}} \quad cu_{N-1} + \underset{w_{N-1}}{E}\{R(x_{N-1} + u_{N-1} - w_{N-1})\}$$

Die optimalen Kosten für die letzte Periode sind

$$J_{N-1}(x_{N-1}) = r(x_{N-1}) + \min_{u_{N-1} \geq 0}\left[cu_{N-1} + \underset{w_{N-1}}{E}\{R(x_{N-1} + u_{N-1} - w_{N-1})\} \right]$$

Optimale Kontrolle: Einführung

TECHNISCHE UNIVERSITÄT DARMSTADT

Beispiel I, Lagerkontrolle

Rest-Teilproblem der Länge 2:

Zu Beginn der Periode N-2 sei der Lagerbestand x_{N-2}. Klar: Egal was vorher war, sollte man mittels $u_{N-2} \geq 0$

(erwarteten Kosten der Periode N-2) +
(erwarteten Kosten der Periode N-1, bei optimaler Politik)
→ $r(x_{N-2}) + cu_{N-2} + E\{J_{N-1}(x_{N-1})\}$

minimieren.

$$J_{N-2}(x_{N-2}) = r(x_{N-2}) + \min_{u_{N-2} \geq 0} \left[cu_{N-2} + \underset{w_{N-2}}{E} \{J_{N-1}(x_{N-2} + u_{N-2} - w_{N-2})\} \right]$$

Rest-Teilproblem der Länge k:
minimiere mittel u_{N-k} :

(erwarteten Kosten der Periode N-k) +
(erwarteten Kosten der Periode N-k+1, bei optimaler Politik)

$$J_{N-k}(x_{N-k}) = r(x_{N-k}) + \min_{u_{N-k} \geq 0} \left[cu_{N-k} + \underset{w_k}{E} \{J_{N-k+1}(x_{N-k} + u_{N-k} - w_{N-k})\} \right] \longrightarrow \text{DP}$$