Computergestützte diskrete stochastische Optimierung

Anjana Buckow[1], **Michaela Dreike**[1], **Marc Wagner**[2]

[1] Deutsche Forschungsgemeinschaft (DFG), Graduiertenkollegs, Graduiertenschulen, Nachwuchsförderung, D-53170 Bonn, Germany

[2] Goethe-Universität Frankfurt am Main, Institut für Theoretische Physik, Emmy Noether-Gruppe "Lattice QCD with 2+1+1 dynamical quark flavors", Max-von-Laue-Straße 1, D-60438 Frankfurt am Main, Germany

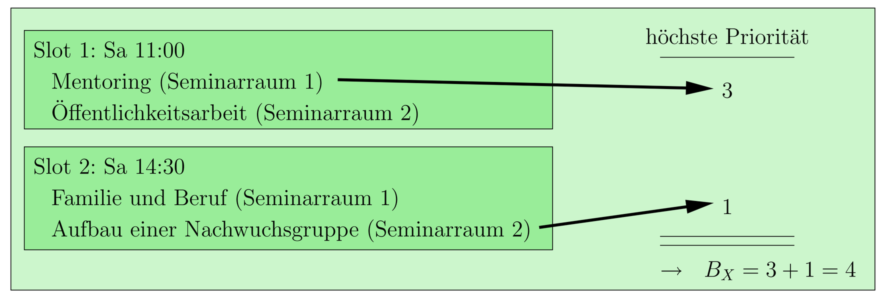
11. Emmy Noether-Jahrestreffen, 13. bis 15. Juli 2012, Potsdam

Organisation einer Tagung, optimales Verteilen von Vorträgen auf Slots

- Organisation einer Tagung ... N Vorträge, $M < N$ Slots (d.h. parallele Vorträge) ... Vortragswunschliste der Teilnehmer ... **Optimale Verteilung der Vorträge auf die Slots?**
- Emmy Noether-Treffen 2012: $N = 10$ Vorträge, $M = 4$ Slots.
- Gesucht: Optimale Verteilung der Vorträge auf die Slots.
- Was bedeutet optimale Verteilung?
  - Grundgedanke: "Das Wohl Vieler ist wichtiger als das Wohl Weniger, oder eines Einzelnen." (Spock, Sternzeit 8130.3).
  - Definiere eine Bewertungsfunktion B, die jeder Verteilung der Vorträge auf die Slots eine Zahl zuweist.
    - Jeder Teilnehmer gibt im Vorfeld der Tagung eine Wunschliste für die Vorträge ab, d.h. er sortiert sie nach ihrer Wichtigkeit, z.B. bei $N = 4$ Vorträgen

Wunschliste von Teilnehmer X	Priorität	Vortrag
	1	Aufbau einer Nachwuchsgruppe
	2	Familie und Beruf
	3	Mentoring
	4	Öffentlichkeitsarbeit

    - Bewertung einer Verteilung für einen Teilnehmer X, B_X: Addiere die jeweils niedrigste Priorität in jedem Slot, z.B. bei $M = 2$ Slots



    - Gesamtbewertung: Addiere die Bewertungen aller Teilnehmer für eine Verteilung, um die Bewertungsfunktion für diese Verteilung zu erhalten; mathematisch $B = \sum_X B_X$.
    - Je kleiner die Bewertungsfunktion B für eine Verteilung ausfällt, desto besser ist diese Verteilung.
  - Zu bestimmen ist diejenige Verteilung der Vorträge auf die Slots, die minimales B liefert; diese Verteilung ist die im Mittel beste Verteilung für alle.
  - Wie findet man diese optimale Verteilung?

Akademisches "Minibeispiel": 4 Vorträge, 2 Slots, 12 Tagungsteilnehmer → Lösung sehr einfach

- Lösung kann "von Hand" bestimmt werden, indem man alle Verteilungen durchprobiert und dafür jeweils B berechnet.
  - Bewertung einer Verteilung für einen Teilnehmer: $\approx 15\,\text{sec}$.
  - Bewertung einer Verteilung für alle Teilnehmer, d.h. Berechnen von B: $\approx 12 \times 15\,\text{sec} = 3\,\text{min}$.
  - Anzahl der möglichen Verteilungen:
    - 2 Möglichkeiten bzw. Slots für jeden Vortrag.
    - Damit $2 \times 2 \times 2 \times 2 = 2^4 = 16$ mögliche Verteilungen.
  - Insgesamt benötigte Zeit zum Finden der optimalen Verteilung: $\approx 16 \times 3\,\text{min} = 48\,\text{min} < 1\,\text{h}$.

Emmy Noether-Treffen 2012: 10 Vorträge, 4 Slots, 169 Tagungsteilnehmer → Lösung mit Hilfe eines Computers einfach

- Ist es sinnvoll auch hier die Lösung "von Hand" zu bestimmen?
  - Bewertung einer Verteilung für einen Teilnehmer: $\approx 1\,\text{min}$.
  - Bewertung einer Verteilung für alle Teilnehmer, d.h. Berechnen von B: $\approx 169 \times 1\,\text{min} \approx 3\,\text{h}$.
  - Anzahl der möglichen Verteilungen:
    - 4 Möglichkeiten bzw. Slots für jeden Vortrag.
    - Damit $\underbrace{4 \times 4 \times \ldots \times 4}_{\text{10 Faktoren}} = 4^{10} = 1\,048\,576 \approx 10^6$ mögliche Verteilungen.
  - Insgesamt benötigte Zeit zum Finden der optimalen Verteilung: $\approx 10^6 \times 3\,\text{h} \approx 342$ Jahre.
- Ein Computer (mein 7 Jahre alter Laptop) benötigt zum Durchprobieren aller Verteilungen nur 10 sec ... ist also etwa um den Faktor 10^9 schneller als ein Mensch.

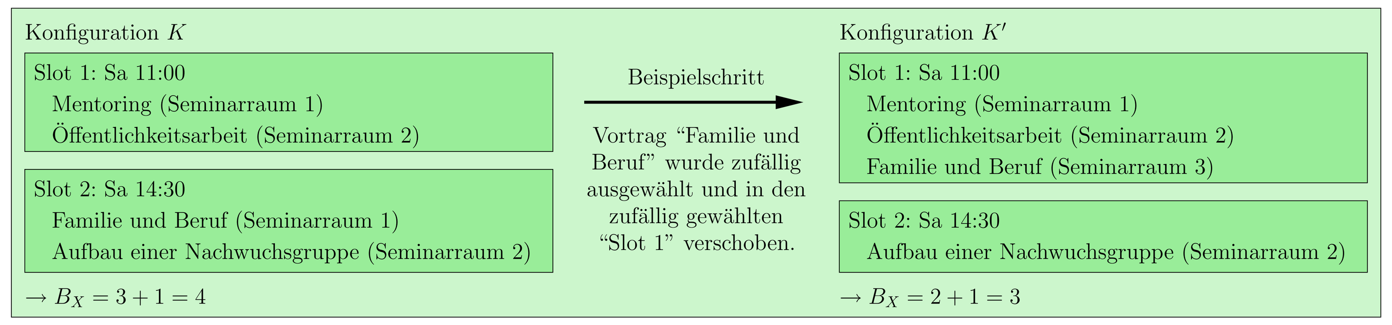
```
Optimale Kombination 1 (Bewertung = 1759, Bewertung / Anzahl aller Teilnehmer = 10.41):
    Sa 11:00
        Vom_Ruf_zur_Professur (erwartete Teilnehmer 81.83)
        Familie_und_Beruf (erwartete Teilnehmer 37.83)
        Mentoring (erwartete Teilnehmer 49.33)
    Sa 14:30
        Oeffentlichkeitsarbeit (erwartete Teilnehmer 56.00)
        EU_Foerderung (erwartete Teilnehmer 113.00)
    Sa 16:30
        Heisenberg_Professur (erwartete Teilnehmer 83.50)
        Gutachtertaetigkeit (erwartete Teilnehmer 85.50)
    So 09:00
        Aufbau_Nachwuchsgruppe (erwartete Teilnehmer 45.00)
        Wie_erhalte_ich_einen_Ruf (erwartete Teilnehmer 95.00)
        Frauenquote (erwartete Teilnehmer 29.00)


Optimale Kombination 2 (Bewertung = 1759, ...
```

Große internationale Fachtagung: 600 Vorträge, 100 Slots, 400 Tagungsteilnehmer → Durchprobieren überfordert jeden Computer

- Wieviel Zeit benötigt ein Computer zum Durchprobieren aller Verteilungen?
  - Bewertung einer Verteilung für alle Teilnehmer, d.h. Berechnen von B: $\approx 10^{-9}\,\text{sec}$.
  - Anzahl der möglichen Verteilungen:
    - 100 Möglichkeiten bzw. Slots für jeden Vortrag.
    - Damit $\underbrace{100 \times 100 \times \ldots \times 100}_{\text{600 Faktoren}} = 100^{600} = 10^{1200}$ mögliche Verteilungen.
  - Insgesamt benötigte Zeit zum Finden der optimalen Verteilung: $\approx 10^{1200} \times 10^{-9}\,\text{sec} = 10^{1191}\,\text{sec} \approx 10^{1173}$ Weltalter (Alter des Universums $\approx 13.8 \times 10^9$ Jahre $\approx 4.4 \times 10^{17}\,\text{sec}$).

Ausweg: Stochastische Optimierung, "Simulated Annealing"

- Grundidee:
  - Bewerte nicht alle Verteilungen (im Folgenden als Konfigurationen bezeichnet), sondern vollführe kleine Schritte von einer Konfiguration zur einer ähnlichen anderen Konfiguration, so dass nach einer "überschaubaren" Anzahl von Schritten die beste Konfiguration erreicht wird.
  - Gehe überwiegend zu besseren Konfigurationen.
  - Gehe manchmal (insbesondere zu Beginn) auch zu schlechteren Konfigurationen.
- Startkonfiguration: Beliebig, z.B. zufällige Verteilung der Vorträge auf die Slots.
- Schritt: Zufälliger Übergang von der aktuellen Konfiguration K zu einer ähnlichen neuen Konfiguration K' (viele Freiheiten bei der Definition erlaubter Übergänge), z.B. wähle zufällig einen Vortrag aus und verschiebe diesen in einen zufällig gewählten Slot,



  - Wenn K' besser als K, akzeptiere K' als aktuelle Konfiguration.
  - Wenn K' schlechter als K, akzeptiere K' nur mit Wahrscheinlichkeit $e^{-(B(K')-B(K))/T}$ als aktuelle Konfiguration, ansonsten behalte K als aktuelle Konfiguration (je schlechter eine Konfiguration, desto unwahrscheinlicher, dass sie akzeptiert wird).
  - T ist die sogenannte Temperatur, die im Lauf der Optimierung langsam gesenkt wird, d.h. Übergänge zu schlechteren Konfigurationen werden immer unwahrscheinlicher.
- Interessant für Mathematiker sind beweisbare Aussagen der Art "Wenn man unendlich lang optimiert und T unendlich langsam absenkt, endet man mit 100%iger Wahrscheinlichkeit bei der besten Konfiguration".
- Interessant in der Praxis: Nach längerer Laufzeit (einige Minuten bzw. Stunden) endet man mit großer Wahrscheinlichkeit bei der besten Konfiguration.
- Durch solche stochastischen Optimierungsverfahren kann man heutzutage umfangreiche Probleme in kurzer Zeit (Zeitaufwand: Minuten/Stunden) lösen, die früher als unlösbar galten (Zeitaufwand: Millionen von Millionen von Millionen ... Jahren, alle Konfigurationen auszuprobieren).

Rucksack-Problem

- Überfall auf einen Juwelierladen ... ein kleiner Rucksack zum Verstauen der Beute ... **Welche Schmuckstücke nehme ich mit?**
- Übungsaufgabe, Vorlesung "Numerische Methoden der Physik", Goethe-Universität Frankfurt am Main, WS 2011/12:

Um Ihren bahnbrechenden theoretischen Arbeiten eine noch größere Tragweite zu verleihen, planen Sie eine Expedition zum Nordpol mit dem Ziel, magnetische Monopole experimentell nachzuweisen. Bedauerlicher Weise besteht an Bord Ihres Transportflugzeugs ein striktes Gewichtslimit von 20 kg für Ihren Expeditionsrucksack. Dies sprengt das Gewicht Ihrer ursprünglich zur Mitnahme geplanten Ausrüstung bei weitem. Um eine optimale Untermenge an Ausrüstungsgegenständen mitzuführen, weisen Sie jedem Gegenstand I_j einen Nutzwert v_j zu. Die optimale Untermenge entspricht dann dem Maximum von

$$F = \sum_{I_j \in \mathit{Expeditionsrucksack}} v_j .$$

wobei gleichzeitig

$$\sum_{I_j \in \mathit{Expeditionsrucksack}} w_j \leq 20\,kg$$

erfüllt sein muss (w_j ist das Gewicht von I_j). Eine Liste Ihrer Ausrüstungsgegenstände finden Sie unter
`http://th.physik.uni-frankfurt.de/~mwagner/teaching/numerik/items.dat`
, wobei in der ersten Spalte die Bezeichnung des Gegenstands aufgelistet ist, in der zweiten Spalte sein Gewicht w_j (in kg) und in der dritten Zeile sein Wert v_j.

(a) Wieviele verschiedene Untermengen von Gegenständen existieren? Wie lange würde ein Computer etwa benötigen, um für jede dieser Mengen das Gewichtslimit von 20 kg zu testen und den Gesamtwert F zu berechnen, wenn dies für eine Untermenge in $\approx 10^{-9}$s möglich ist?

(b) Entwickeln Sie eine oder mehrere Strategien, Simulated-Annealing auf die Optimierung Ihrer Expeditionsausrüstung anzuwenden. Diskutieren Sie insbesondere, welche Update-Schritte Ihnen geeignet erscheinen.

(c) Implementieren Sie Ihre in (b) entwickelten Strategien und wenden Sie sie zur Ausrüstungsoptimierung an. Geben Sie die Liste von Gegenständen an, mit denen Sie zum Nordpol fliegen. Was ist deren Gesamtgewicht sowie deren Gesamtwert F?

Traveling-Salesperson-Problem

- Urlaub geplant, Rundreise durch N Städte ... **Kürzester Weg, der die Städte verbindet?**