

Vorlesung

Effiziente Algorithmen und Komplexitätstheorie

Sommersemester 2008

Ingo Wegener

Allgemeine randomisierte Suchheuristiken

Erinnerung

- Einordnung und Abgrenzung
- Anwendungsszenarien
- Möglichkeiten und Grenzen

in der Vorlesung

- Einschränkung auf sehr einfache randomisierte Suchheuristiken
- immer Suchraum $\{0, 1\}^n$
- strukturell sehr ähnliche Suchheuristiken
- immer ohne echtes Stoppkriterium
- immer Analyse der (erwarteten) **Optimierzeit**
= # f -Aufrufe, bis zum Finden eines Maximums

Simulated Annealing

Algorithmus 13.1

Sei $T: \mathbb{N} \rightarrow \mathbb{R}_0^+ \cup \{\infty\}$ eine Funktion, T heißt **Annealingschedule**.
 Sei $N: \{0, 1\}^n \rightarrow \mathcal{P}(\{0, 1\}^n)$ eine **Nachbarschaft**. Die folgende Suchheuristik heißt **Simulated Annealing**.

1. $t := 1$
2. Wähle $x_t \in \{0, 1\}^n$ uniform zufällig.
3. Wähle $y \in N(x_t)$ gemäß einer festen Verteilung.
4. Mit Wahrscheinlichkeit $\min\{1, e^{-(f(y)-f(x_t))/T(t)}\}$,
 setze $x_{t+1} := y$, sonst $x_{t+1} := x_t$.
5. $t := t + 1$
6. Weiter bei 3.

typische Nachbarschaft

$$N(x) = \{y \in \{0, 1\}^n \mid 0 < d(x, y) \leq d^*\}$$

mit $d^* = 1$ oder $d^* = 2$

Metropolis-Algorithmus, RLS und (1+1)-EA

Algorithmus 13.2 (Metropolis-Algorithmus)

Der **Metropolis-Algorithmus** ist Simulated Annealing mit fester **Temperatur** $T \in \mathbb{R}_0^+ \cup \{\infty\}$.

Algorithmus 13.3 (Randomisierte lokale Suche)

RLS ist der Metropolis-Algorithmus mit Temperatur $T = 0$.

Algorithmus 13.4 ((1+1)-EA)

1. $t := 1$
2. Wähle $x_t \in \{0, 1\}^n$ uniform zufällig.
3. Für $i \in \{1, 2, \dots, n\}$
 Mit W'keit p , setze $y[i] := 1 - x_t[i]$, sonst $y[i] := x_t[i]$.
4. If $f(y) \leq f(x_t)$ Then $x_{t+1} := y$ Else $x_{t+1} := x_t$.
5. $t := t + 1$
6. Weiter bei 3.

Minimale Spannbäume

Eingabe ungerichteter, gewichteter, zusammenhängender Graph $(G = (V, E), w)$ mit $w: E \rightarrow \mathbb{N}$, $|E| = m$

Ausgabe ungerichteter, gewichteter, zusammenhängender Graph $(G = (V, E'), w)$ mit $E' \subseteq E$ und $w(E') = \sum_{e \in E'} w(e)$

minimal unter allen solchen Teilgraphen

Entscheidung Modellierung als Kantenauswahlproblem

Suchraum $\{0, 1\}^m$ mit $x[i] = 1 \Leftrightarrow e_i$ gewählt

Notation zu $x \in \{0, 1\}^m$ ist $E_x = \bigcup_{x[i]=1} e_i$

Zielfunktionen

- $f_1(x) := w(E_x)$

Kritik \emptyset ist optimal

- $f_2(x) := \begin{cases} w(E_x) & \text{falls } (V, E_x) \text{ zusammenhängend} \\ \infty & \text{sonst} \end{cases}$

Kritik ohne Hilfe **zu schwer**, überhaupt ST zu finden

- $f_3(x) := (c(x) - 1) \cdot w_b^2 + (|x| - (n - 1)) \cdot w_b + w(E_x)$
mit $c(x) := \#$ Zusammenhangskomponenten in (V, E_x) ,

$$w_b := m \cdot \max\{w(e_1), \dots, w(e_m)\} \text{ und } |x| = \sum_{i=1}^n x[i]$$

Kritik vielleicht zu sehr in Richtung Lösung formuliert

- $f_4(x) := (c(x) - 1) \cdot w_b + w(E_x)$

Kritik etwas künstliche Verschmelzung zweier Kriterien

- $f_5(x) := (c(x), w(E_x))$

Kritik Algorithmenanpassung nötig

Randomisierte Suchheuristiken auf MST

| Suchheuristik | Modellierung | Resultat |
|---------------|--------------|---|
| RLS | f_3 | $O(m^2(\log(n) + \log(w_{\max})))$ $\Omega(m^2 \log(n))$ |
| | f_4 | $O(m^2(\log(n) + \log(w_{\max})))$ $\Omega(m^2 \log(n))$ |
| (1+1) EA | f_3 | $O(m^2(\log(n) + \log(w_{\max})))$ $\Omega(m^2 \log(n))$ |
| | f_4 | $O(m^2(\log(n) + \log(w_{\max})))$ $\Omega(m^2 \log(n))$ |

Jetzt $f_5(x) := (c(x), w(E_x))$

MST modelliert mit f_5

$f_5(x) = (c(x), w(E_x)) \rightarrow \min$ in beiden Komp.

Kritik Algorithmenanpassung nötig

Problem „ $f_5(y) \leq f_5(x)$ “ sinnvoll definieren

naheliegend

$$f_5(y) = (y_1, y_2) \leq f_5(x) = (x_1, x_2) :\Leftrightarrow (y_1 \leq x_1) \wedge (y_2 \leq x_2)$$

Beobachtung \leq ist Halbordnung
(reflexiv, transitiv, antisymmetrisch)

x **dominiert** $y :\Leftrightarrow (f(x) \leq f(y)) \wedge (f(x) \neq f(y))$

x mit $f(x)$ minimal unter \leq heißt **Pareto-Optimum**.

$\{f(x) \mid x \text{ Pareto-Optimum}\}$ heißt **Pareto-Front**.

Simple Evolutionary Multiobjective Optimizers

Algorithmus 13.16 (SEMO)

1. Wähle $x \in \{0, 1\}^n$ uniform zufällig.
2. $P := \{x\}$
3. Wähle $x \in P$ uniform zufällig.
4. Wähle $y \in N(x)$ gemäß einer festen Verteilung zufällig.
5. Falls nicht $\exists x' \in P: (f(x') \leq f(y)) \wedge (f(x') \neq f(y))$
6. $P := P \setminus \{x' \mid f(y) \leq f(x')\}$
7. $P := P \cup \{y\}$
8. Weiter bei 3.

Beobachtung SEMO ist **lokal** wie RLS.

alternativ Mutation wie beim (1+1)-EA
Ergebnis heißt **global SEMO (GSEMO)**

MST multikriteriell mit f_5

Beobachtung Pareto-Front unter f_5 besteht aus n Elementen, für jedes $k \in \{1, \dots, n\}$ eines mit genau k Zusammenhangskomponenten

Beobachtungen

- MST ist Pareto-Optimum.
- \emptyset ist Pareto-Optimum.
- P ist immer Menge nicht dominierter Punkte.

SEMO und Global SEMO für f_5

Entscheidung $N(x) := \{y \in \{0, 1\}^n \mid d(x, y) = 1\}$
mit uniformer Verteilung auf N

Kann man MST so überhaupt erreichen?

Theorem 13.17

Die erwartete Optimierzeit von SEMO und GSEMO auf MST mit f_5 beträgt $O(mn \cdot (n + \log w_{\max}))$.

Zum Beweis von Theorem 13.17

Betrachte zwei Phasen **flexibler Länge**

- ① Initialisierung bis erstmals $\emptyset \in P$
- ② Anschluss bis erstmals $\text{MST} \in P$

zur ersten Phase

Betrachte $x^* \in P$ mit $w(E_{x^*}) = \min\{w(E_x) \mid x \in P\}$

Beobachtung E_{x^*} hat max. Anzahl ZHK

klar Entfernen von Kante aus E_{x^*} senkt Gewicht

kann #Zusammenhangskomponenten vergrößern

Beobachtung $\exists m$ Hammingnachbarn,

bei denen Gewicht **im Durchschnitt** um $w(x^*)/m$ sinkt

Nachrechnen $\lceil m \ln(2) (\log(mw_{\max}) + 1) \rceil$ Schritte $\rightsquigarrow w(x^*) < 1$

Nicht vergessen Faktor $O(n)$ für Auswahl von x^*

zusammen im Durchschnitt nach

$O(mn (\log(n) + \log(w_{\max})))$ Schritten $\emptyset \in P$ ✓

Zweite Phase

zur zweiten Phase

Voraussetzung $R_n, R_{n-1}, \dots, R_i \in P$, alle Pareto-Optima
 R_j mit genau j Zusammenhangskomponenten

Beobachtung erfüllt, weil anfangs $\emptyset \in P$

Erinnerung stets passende Kante wählbar $R_i \rightsquigarrow R_{i-1}$
 wie beim Algorithmus von Kruskal

klar W'keit dafür $\Omega\left(\frac{1}{n} \cdot \frac{1}{m}\right)$

also $E(\text{Länge Phase 2}) = O(m \cdot n^2)$

zusammen $E(T) = O(mn (\log(n) + \log(w_{\max}))) + O(m \cdot n^2)$
 $= O(mn \cdot (n + \log(w_{\max})))$



f_4 vs. f_5

bei polynomiellen Kantengewichten

für f_4 RLS und (1+1)-EA $O(m^2 \log n)$

für f_5 SEMO und GSEMO $O(mn^2)$

für dichte Graphen $m = \Theta(n^2) \rightsquigarrow O(n^4 \log n)$ vs. $O(n^4)$

für dünn-besetzte Graphen $m = \Theta(n) \rightsquigarrow O(n^2 \log n)$ vs. $O(n^3)$

Experimente bestätigen die Reihenfolge
der oberen Schranken als **realistisch**

Randomisierte Suchheuristiken auf MST

| Suchheuristik | Modellierung | Resultat |
|---------------|--------------|---|
| RLS | f_3 | $O(m^2(\log(n) + \log(w_{\max})))$ $\Omega(m^2 \log(n))$ |
| | f_4 | $O(m^2(\log(n) + \log(w_{\max})))$ $\Omega(m^2 \log(n))$ |
| (1+1) EA | f_3 | $O(m^2(\log(n) + \log(w_{\max})))$ $\Omega(m^2 \log(n))$ |
| | f_4 | $O(m^2(\log(n) + \log(w_{\max})))$ $\Omega(m^2 \log(n))$ |
| SEMO | f_5 | $O(mn(n + \log(w_{\max})))$ |
| GSEMO | f_5 | $O(mn(n + \log(w_{\max})))$ |

Metropolis-Algorithmus und Simulated Annealing

Erinnerung noch **kein** Ergebnis
für Metropolis-Algorithmus und Simulated Annealing

Doch!

RLS = MA mit $T = 0$

MA = SA mit festem T

Was können MA und SA, was RLS nicht kann?

verschlechternde Schritte \rightsquigarrow Entkommen aus lokalen Optima

Erinnerung RLS braucht 2-Bit-Flips, um f_3/f_4 zu optimieren.

Können MA und SA das auch nur mit 1-Bit-Flips schaffen?

Ist SA signifikant besser als MA?

MA und SA auf f_2

gesehen (V, E_x) zusammenhängend ist leicht erreichbar

Wir sparen uns das und starten deterministisch in 1^n .

Wir betrachten MA und SA

$$\text{auf } f_2(x) = \begin{cases} w(E_x) & \text{falls } (V, E_x) \text{ zusammenhängend} \\ \infty & \text{sonst} \end{cases} \rightarrow \min$$

1. $t := 1; x_t := 1^n$
2. Wähle $y \in N(x_t) = \{x \mid d(x, y) = 1\}$ gleichverteilt.
3. Mit W'keit $\min \{1, e^{-(f(y)-f(x_t))/T(t)}\}$

$$x_{t+1} := y$$

Sonst $x_{t+1} := x_t$

4. $t := t + 1$
5. Weiter bei 2.

mit $T: \mathbb{N} \rightarrow [0; \infty[$

mit $T(t) = T \in [0; \infty[$

Simulated Annealing

Metropolis-Algorithmus

Performanz parametrisierter Algorithmen

Wir beurteilen Performanz auf $\mathcal{F} \subseteq \{f: \{0,1\}^n \rightarrow \mathbb{R}\}$.

Wir sagen

A ist effizient auf \mathcal{F}

$\Leftrightarrow \forall f \in \mathcal{F}: \exists$ Parametrisierung: *A* effizient auf *f*

A ist nicht effizient auf \mathcal{F}

$\Leftrightarrow \exists f \in \mathcal{F}: \forall$ Parametrisierungen: *A* nicht effizient auf *f*

also

SA mindestens so effizient wie **MA**

MA mindestens so effizient wie **RLS**

Gambler's-Ruin-Problem

vorab ein w'keitstheoretisches Tool (auch allgemein **nützlich**)

Theorem 13.19

Seien $s \in \mathbb{N}$, $a \in \{0, 1, \dots, s\}$, $p_A \in]0; 1[\setminus \{1/2\}$, $p_B := 1 - p_A$.
Betrachte den Zufallsprozess $(X_i)_{i \geq 0}$ mit $X_0 := a$ und

$$X_{i+1} = \begin{cases} X_i + 1 & \text{mit Wahrscheinlichkeit } p_A, \\ X_i - 1 & \text{mit Wahrscheinlichkeit } p_B. \end{cases}$$

Es bezeichne $T := \min\{X_t \mid X_t \in \{0, s\}\}$, $q := p_B/p_A$ und $q_A := \text{Prob}(X_T = 0)$.

Es gilt $q_A = \frac{q^a - q^s}{1 - q^s}$ und $E(T) = \frac{(1 - q_A) \cdot s - a}{p_A - p_B}$.

Zum Beweis des Gambler's-Ruin-Theorems

Plan Anwendung Optional-Stopping-Theorem (Theorem 10.13)

Beobachtung T ist Stoppzeit ✓

Betrachte $M_t := q^{X_t}$

Behauptung M_t ist Martingal in Bezug auf $(X_i)_{i \geq 0}$

$$\begin{aligned}
 & \mathbb{E}(M_{t+1} \mid X_1, X_2, \dots, X_t) \\
 = & \mathbb{E}(M_{t+1} \mid X_t) = \mathbb{E}(q^{X_{t+1}} \mid X_t) \\
 = & p_A \cdot q^{X_t+1} + p_B \cdot q^{X_t-1} \\
 = & q^{X_t} \cdot \left(p_A \cdot q + \frac{p_B}{q} \right) \\
 = & q^{X_t} \cdot \left(p_A \cdot \frac{p_B}{p_A} + p_B \cdot \frac{p_A}{p_B} \right) \\
 = & q^{X_t} \cdot (p_B + p_A) = q^{X_t} \checkmark
 \end{aligned}$$

Über die W'keit, ruiniert zu werden

Wir haben $q = p_B/p_A$, $T = \min\{X_t \mid X_t \in \{0, s\}\}$
 $M_t = q^{X_t}$ ist Martingal

Optional-Stopping-Theorem $E(M_T) = E(M_0) = E(q^{X_0}) = q^a$

Beobachtung $E(M_T) = q_A \cdot q^0 + q_B \cdot q^s$
 $= q_A + (1 - q_A) \cdot q^s = (1 - q^s) \cdot q_A + q^s$

also $q^a = (1 - q^s) \cdot q_A + q^s$

also $q_A = \frac{q^a - q^s}{1 - q^s}$ ✓

Beobachtung sagt **nichts** über $E(T)$

Zur erwarteten Dauer

Betrachte $N_t := X_t - t \cdot (p_A - p_B)$

Behauptung N_t ist Martingal in Bezug auf $(X_i)_{i \geq 0}$

$$\begin{aligned}
 & \mathbb{E}(N_{t+1} \mid X_1, X_2, \dots, X_t) \\
 = & \mathbb{E}(N_{t+1} \mid X_t) \\
 = & p_A \cdot (X_t + 1 - (t+1) \cdot (p_A - p_B)) \\
 & + p_B \cdot (X_t - 1 - (t+1) \cdot (p_A - p_B)) \\
 = & (p_A + p_B) \cdot X_t + p_A - p_B - (p_A + p_B) \cdot (t+1) \cdot (p_A - p_B) \\
 = & X_t + p_A - p_B - (t+1) \cdot (p_A - p_B) \\
 = & X_t - t \cdot (p_A - p_B) = N_t \checkmark
 \end{aligned}$$

Anwendung Optional-Stopping-Theorem

Wir haben $N_t = X_t - t \cdot (p_A - p_B)$ ist Martingal

Optional-Stopping-Theorem

$$\mathbb{E}(N_T) = \mathbb{E}(N_0) = \mathbb{E}(X_0 - 0 \cdot (p_A - p_B)) = a$$

Beobachtung

$$\mathbb{E}(N_T) = q_A \cdot (0 - \mathbb{E}(T) \cdot (p_A - p_B)) + q_B \cdot (s - \mathbb{E}(T) \cdot (p_A - p_B))$$

also

$$a = q_A \cdot (0 - \mathbb{E}(T) \cdot (p_A - p_B)) + q_B \cdot (s - \mathbb{E}(T) \cdot (p_A - p_B))$$

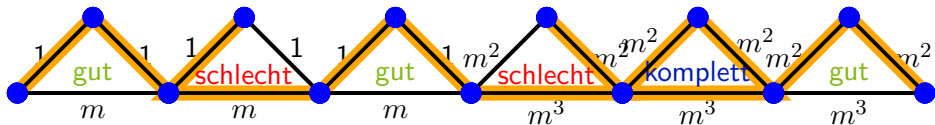
$$\text{also } \mathbb{E}(T) (p_A - p_B) \cdot (q_A + q_B) = q_B \cdot s - a$$

$$\text{also } \mathbb{E}(T) (p_A - p_B) = (1 - q_A)s - a$$

$$\text{also } \mathbb{E}(T) = \frac{(1 - q_A)s - a}{p_A - p_B}$$



Ein Beispielgraph



allgemein

- n „leichte“ Dreiecke
- n „schwere“ Dreiecke
- $4n + 1$ Knoten
- $m = 6n$ Kanten

wieder für $x \in \{0, 1\}^m$

$b(x)$ schlechte, $g(x)$ gute, $ct(x)$ komplette Dreiecke

Intuitionen für G

- anfangs $\Theta(m)$ Dreiecke schlecht
- schlechte Dreiecke müssen komplett werden
- schlecht \rightsquigarrow komplett erhöht f_2 um 1 in leichten und m^2 in schweren Dreiecken
- schlecht \rightsquigarrow gut sollte nicht zu unwahrscheinlich sein
- es sollte unwahrscheinlich sein, gute Dreiecke zu verlieren
- gut \rightsquigarrow komplett erhöht f_2 um m in leichten und m^3 in schweren Dreiecken
- gut \rightsquigarrow komplett sollte nicht zu wahrscheinlich sein

Fazit Unsere Wünsche

in schweren Dreiecken $f_2 \rightsquigarrow f_2 + m^3$ unwahrscheinlich

$f_2 \rightsquigarrow f_2 + m^2$ wahrscheinlich

in leichten Dreiecken $f_2 \rightsquigarrow f_2 + m$ unwahrscheinlich

$f_2 \rightsquigarrow f_2 + 1$ wahrscheinlich

mit nur einer festen Temperatur wohl nicht möglich

MA auf G

Wir erwarten

niedrige Temperatur \rightsquigarrow **schlechte schwere** Δ bleiben **schlecht**

hohe Temperatur \rightsquigarrow **gute leichte** Δ bleiben **nicht gut**

Theorem 13.20

Es gibt eine Konstante $c > 0$, so dass MA auf G mit Wahrscheinlichkeit $1 - e^{-\Omega(m)}$ mindestens e^{cm} Schritte zur Optimierung braucht.

Beweis.

anfangs alle Dreiecke komplett

für jedes **komplette** Δ Prob (mache Δ schlecht) = $\frac{2}{3}$

also Prob (\exists schlechtes schweres Δ) = $1 - e^{-\Omega(m)}$

Beweis von Theorem 13.20

1. Fall: $T < m$ (niedrige Temperatur)

Prob (füge m^2 -Kante im schlechten schweren Δ hinzu)

$$= \frac{1}{m} \cdot e^{-m^2/T} \leq \frac{e^{-m}}{m} = e^{-\Omega(m)}$$

2. Fall: $T \geq m$ (hohe Temperatur)

Betrachte nur leichte Dreiecke.

Betrachte $X_t =$ Anzahl guter leichter Δ nach t -ten Schritt.

klar $X_0 = 0$ gesucht $T = \min\{t \mid X_t = n\}$

Beobachtung X_{t+1} hängt nur von $b(x)$, $g(x)$ und $ct(x)$ ab
alle Werte bezogen auf gute Δ

Beweis von Theorem 13.20 (Fortsetzung)

Transitionswahrscheinlichkeiten

| | komplett | gut | schlecht |
|----------|------------------------------|----------------------------------|----------------------------------|
| komplett | $1 - \frac{3}{m}$ | $\frac{1}{m}$ | $\frac{2}{m}$ |
| gut | $\frac{1}{m} \cdot e^{-m/T}$ | $1 - \frac{1}{m} \cdot e^{-m/T}$ | 0 |
| schlecht | $\frac{1}{m} \cdot e^{-1/T}$ | 0 | $1 - \frac{1}{m} \cdot e^{-1/T}$ |

Abschätzungen

$$\text{Prob}(X_{t+1} = a + 1 \mid X_t = a) \leq \frac{n - a}{m}$$

$$\text{Prob}(X_{t+1} = a - 1 \mid X_t = a) \geq \frac{a}{m} \cdot e^{-m/T} \geq \frac{a}{3m}$$

Beobachtung \tilde{X}_{t+1} nur noch von \tilde{X}_t abhängig (Markov-Kette)

Beweis von Theorem 13.20 (Fortsetzung (2))

$$\text{Prob}(X_{t+1} = a + 1 \mid X_t = a) \leq \frac{n - a}{m}$$

$$\text{Prob}(X_{t+1} = a - 1 \mid X_t = a) \geq \frac{a}{3m}$$

$$\text{Prob}(X_{t+1} = X_t - 1 \mid X_{t+1} \neq X_t) \geq \frac{a/(3m)}{a/(3m) + (n-a)/m} = \frac{a}{3n-2a}$$

Betrachte Situation erst ab $(10/11)n$ guten Dreiecken.

Brich Betrachtung ab bei n oder $(9/11)n$ guten Dreiecken.

Wie groß ist die W'keit, n gute Dreiecke zu erreichen?

$$\text{Prob}(X_{t+1} = X_t - 1 \mid X_{t+1} \neq X_t) \geq \frac{(9/11)n}{3n-2 \cdot (9/11)n} = \frac{3}{5}$$

nur ändernde Schritte: Start bei $(10/11)n$,

Änderung um ± 1 bis $(9/11)n$ oder n

W'keit Zunahme $2/5$, W'keit Abnahme $3/5$

Anwendung Gambler's Ruin Theorem

$$q_a = \frac{(p_B/p_A)^s - (p_B/p_A)^a}{(p_B/p_A)^s - 1}$$

mit $p_A = 3/5$, $a = (1/11)n$, $s = (2/11)n$

$$\begin{aligned} \text{Prob (erreiche } n \text{ gute } \Delta) &= q_a = \frac{(2/3)^{(2/11)n} - (2/3)^{(1/11)n}}{(2/3)^{(2/11)n} - 1} \\ &= \left(\frac{2}{3}\right)^{n/11} \cdot \frac{1 - (2/3)^{n/11}}{1 - (2/3)^{(2/11)n}} < \left(\frac{2}{3}\right)^{n/11} = e^{-\Omega(m)} \end{aligned}$$

also erwartete Anzahl solcher Anläufe $e^{\Omega(m)}$



MA also bei jeder Temperatur **hoffnungslos**.

Warum sollte es mit SA besser gehen?

Idee Beginne mit hoher Temperatur und löse schwere Dreiecke, senke Temperatur, darum bleiben schwere Dreiecke erhalten, bei niedriger Temperatur löse leichte Dreiecke.

Simulated Annealing auf G

Theorem 13.21

Sei p ein Polynom in m . Für jede ausreichend große Konstante c findet Simulated Annealing mit Annealingschedule $T(t) := m^3 \cdot \left(1 - \frac{1}{cm}\right)^{t-1}$ einen MST auf G in höchstens $3cm \ln m$ Schritten mit Wahrscheinlichkeit mindestens $1 - 1/p(m)$.

Beweis.

Wir unterteilen den Lauf in vier **Phasen**.

Phase 0: so lange $T(t) > m^{5/2}$

Phase 1: so lange $T(t) \in [m^2; m^{5/2}]$

Phase 2: so lange $T(t) \in]m^{1/2}; m^2[$

Phase 3: so lange $T(t) \in [1; m^{1/2}]$

Wie viele Schritte sind das insgesamt?

$$T(t) = m^3 \cdot \left(1 - \frac{1}{cm}\right)^{t-1} < 1$$

$$\left(1 - \frac{1}{cm}\right)^{cm \cdot \frac{t-1}{cm}} < \frac{1}{m^3} e^{-(t-1)/cm} < e^{-3 \ln m}$$

also $t > 1 + 3cm \ln m$ reicht aus

Beweis von Theorem 13.21

Was in **Phase 0** passiert, ist uns egal.

Phase 1: $T(t) \in [m^2; m^{5/2}]$

Betrachte **nur** schwere Dreiecke.

Prob (gutes $\Delta \rightsquigarrow$ komplettes Δ) $\leq e^{-m^3/T(t)} \leq e^{-m^{1/2}}$

also mit ausreichend großer W'keit bleiben alle guten Δ gut

$E(\# \text{Kante } e \text{ flippt}) \geq \frac{(c/4)m \ln m}{m} = \frac{c \ln m}{4}$

also jede Kante mindestens $c'' \ln m$ Flips in Phase 1 (**Chernoff**)

Prob (schlechtes $\Delta \rightsquigarrow$ gutes Δ in 2 nächsten Flips des Δ)

$\geq \frac{1}{3} \cdot e^{-m^2/T(t)} \cdot \frac{1}{3} \geq \frac{1}{9e}$

Prob (passiert allen schweren Dreiecken in Phase 1) = $1 - n^{-k}$
wenn c groß genug

Was in **Phase 2** passiert, ist uns **fast** egal.

wie in Phase 1: gute schwere Δ bleiben erhalten

Zur letzten Phase

Phase 3: $T(t) \in [1; m^{1/2}]$

Betrachte **nur** leichte Dreiecke.

Prob (gutes $\Delta \rightsquigarrow$ komplettes Δ) $\leq e^{-m/T(t)} \leq e^{-m^{1/2}}$

also mit ausreichend großer W'keit bleiben alle guten Δ gut

$E(\#Kante\ e\ \text{flippt}) \geq \frac{(c/4)m \ln m}{m} = \frac{c \ln m}{4}$

also jede Kante mindestens $c' \ln m$ Flips in Phase 1 (Chernoff)

Prob (schlechtes $\Delta \rightsquigarrow$ gutes Δ in 2 nächsten Flips des Δ)

$\geq \frac{1}{3} \cdot e^{-1/T(t)} \cdot \frac{1}{3} \geq \frac{1}{9e}$

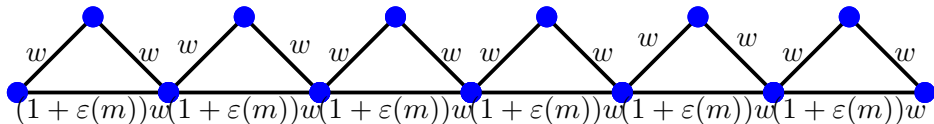
Prob (passiert allen leichten Dreiecken in Phase 1) $= 1 - n^{-k}$

wenn c groß genug



Verallgemeinerung auf mehr Graphen — negativ

nur ein Beispielgraph ist ein bisschen sehr speziell. . .
eine Klasse von Beispielgraphen



Theorem 13.22

Sei $\varepsilon > 0$ konstant. Für $\varepsilon(m) \geq \varepsilon$ findet MA mit $T := \varepsilon w / (3 \ln m)$ einen MST im Erwartungswert in polynomieller Zeit.

Beweis.

$$\text{Prob (akzeptiere neue } w\text{-Kante)} = e^{-w/T} = m^{-3/\varepsilon}$$

$$\text{Prob (akzeptiere neue } (1 + \varepsilon(m))w\text{-Kante)} \leq e^{-(1+\varepsilon)w/T} = m^{-3-3/\varepsilon}$$

Beweis von Theorem 13.22

$$\text{Prob}(\text{akzeptiere neue } w\text{-Kante}) = m^{-3/\varepsilon}$$

$$\text{Prob}(\text{akzeptiere neue } (1 + \varepsilon(m))w\text{-Kante}) \leq m^{-3-3/\varepsilon}$$

Betrachte $m^{2+(3/\varepsilon)}$ Schritte.

$$\begin{aligned} \text{Prob}(\text{akzeptiere irgendwann schwere Kante}) &\leq m^{-3-3/\varepsilon} \cdot m^{2+(3/\varepsilon)} \\ &= \frac{1}{m} \end{aligned}$$

mit W'keit sehr dicht an 1 für jedes Δ

$\Omega(m^{1+(3/\varepsilon)})$ Kantenflips

$$\text{Prob}(\text{schlechtes } \Delta \rightsquigarrow \text{gutes } \Delta \text{ dann}) = \Omega(m^{-3/\varepsilon})$$

$$\text{Prob}(\text{komplettes } \Delta \rightsquigarrow \text{gutes } \Delta \text{ dann}) = \Omega(m^{-3/\varepsilon})$$

also mit W'keit sehr dicht an 1 wird jedes Dreieck gut

zusätzlich bleiben alle guten Δ gut mit W'keit $\geq 1 - 1/m$

im Misserfolgsfall „nicht schlimmer als neu gestartet“



Verallgemeinerung auf mehr Graphen — negativ

Theorem 13.23

Für $\varepsilon(m) = o(1)$ finden MA mit jeder Temperatur und SA mit jedem Annealingschedule für jede Konstante k und jedes Polynom p in $O(p(m))$ Schritten einen MST^k mit Wahrscheinlichkeit $o(m^{-k})$.

Beweis.

wie zuvor mit genügend großer W'keit $\Theta(m)$ viele schlechte Δ
zunächst nur MA betrachten

Prob (akzeptiere w -Kante irgendwann im Lauf) = $O(p(m) \cdot e^{-w/T})$

nur für $T \geq w/(\gamma \cdot \ln m)$ ($\gamma > 0$ konstant) ist das groß genug

$p^*(T) := \text{Prob}(\text{akzeptiere } w\text{-Kante})$

$p_\varepsilon^*(T) := \text{Prob}(\text{akzeptiere } (1 + \varepsilon(m))w\text{-Kante})$

zu zeigen Unterschied zwischen $p^*(T)$ und $p_\varepsilon^*(T)$ zu klein

Beweis von Theorem 13.23

$p^*(T) = \text{Prob}(\text{akzeptiere } w\text{-Kante})$

$p_\varepsilon^*(T) = \text{Prob}(\text{akzeptiere } (1 + \varepsilon(m))w\text{-Kante})$

$$\begin{aligned} \frac{p^*(T)}{p_\varepsilon^*(T)} &= e^{-w/T} \cdot e^{(1+\varepsilon(m))w/T} \\ &= e^{\varepsilon(m)w/T} \leq e^{\varepsilon(m)\gamma \ln m} \\ &= m^{\varepsilon(m)\gamma} < m^\delta \quad \forall \delta > 0 \end{aligned}$$

Zufallsprozess beschreibbar durch $b(x)$ und $ct(x)$

Potenzialfunktion $\Phi(x) = 2b(x) + ct(x)$

anfangs $\Phi(x) = n$

MST $\Phi(x) = 0$

Beweis von Theorem 13.23 (Fortsetzung)

Potenzialfunktion $\Phi(x) = 2b(x) + ct(x)$ sinkt von n auf 0

$$\begin{aligned} & \text{Prob}(\Phi(x_{t+1}) = a + 1 \mid \Phi(x_t) = a) \\ = & \underbrace{\frac{2ct(x_t)}{m}}_{\text{flippe leichte Kante in kompletten } \Delta \text{ raus}} + \underbrace{\frac{n - b(x_t) - ct(x_t)}{m} \cdot p_\varepsilon^*(T)}_{\text{flippe schwere Kante in gutem } \Delta \text{ rein}} \end{aligned}$$

$$\begin{aligned} & \text{Prob}(\Phi(x_{t+1}) = a - 1 \mid \Phi(x_t) = a) \\ = & \underbrace{\frac{ct(x_t)}{m}}_{\text{flippe schwere Kante in kompletten } \Delta \text{ raus}} + \underbrace{\frac{b(x_t)}{m} \cdot p^*(T)}_{\text{flippe leichte Kante in schlechten } \Delta \text{ rein}} \end{aligned}$$

$$\begin{aligned} & \text{Prob}(\Phi(x_{t+1}) = a \mid \Phi(x_t) = a) = \\ & 1 - \text{Prob}(\Phi(x_{t+1}) = a + 1 \mid \Phi(x_t) = a) - \\ & \text{Prob}(\Phi(x_{t+1}) = a - 1 \mid \Phi(x_t) = a) \end{aligned}$$

Beweis von Theorem 13.23 (Fortsetzung (2))

Wenn wir zeigen können

$$\frac{\text{Prob}(\Phi(x_{t+1}) = a + 1 \mid \Phi(x_t) = a)}{\text{Prob}(\Phi(x_{t+1}) = a + 1 \mid \Phi(x_t) = a) + \text{Prob}(\Phi(x_{t+1}) = a - 1 \mid \Phi(x_t) = a)} \geq \frac{3}{5}$$

Behauptung folgt aus **Gambler's Ruin Problem** wie vorhin

$$\begin{aligned} & \frac{(2ct(x_t))/m + ((n-b(x_t)-ct(x_t))/m) \cdot p_\varepsilon^*(T)}{(2ct(x_t))/m + ((n-b(x_t)-ct(x_t))/m) \cdot p_\varepsilon^*(T) + (ct(x_t))/m + ((b(x_t))/m) \cdot p^*(T)} \geq \frac{3}{5} \\ & \Leftrightarrow \\ & 2ct(x_t) + (n - b(x_t) - ct(x_t)) \cdot p_\varepsilon^*(T) \\ & \geq \frac{9}{5}ct(x_t) + \frac{3}{5}(n - b(x_t) - ct(x_t)) \cdot p_\varepsilon^*(T) + \frac{3}{5}b(x_t) \cdot p^*(T) \\ & \Leftrightarrow \\ & \frac{1}{5}ct(x_t) + \frac{2}{5}(n - b(x_t) - ct(x_t)) \cdot p_\varepsilon^*(T) \geq \frac{3}{5}b(x_t) \cdot p^*(T) \end{aligned}$$

Beweis von Theorem 13.23 (Fortsetzung (3))

zu zeigen $\frac{1}{5}ct(x_t) + \frac{2}{5}(n - b(x_t) - ct(x_t)) \cdot p_\varepsilon^*(T) \geq \frac{3}{5}b(x_t) \cdot p^*(T)$

1. Fall $ct(x_t) \geq 3b(x_t)$ ✓

2. Fall $ct(x_t) < 3b(x_t)$

$$n - b(x_t) - ct(x_t) > n - 4b(x_t)$$

Uns genügt letzte Phase mit $b(x_t) \leq n^{1/2}$.

also $n - b(x_t) - ct(x_t) > n - 4n^{1/2}$

es genügt $2(n - 4n^{1/2}) \cdot p_\varepsilon^*(T) \geq 3b(x_t) \cdot p^*(T)$

$$\Leftrightarrow \frac{p^*(T)}{p_\varepsilon^*(T)} \leq \frac{2}{3}(n^{1/2} - 4)$$

wir haben $\frac{p^*(T)}{p_\varepsilon^*(T)} < m^\delta \quad \forall \delta > 0$

weil $n = \Theta(m)$ ✓

Beweis von Theorem 13.23 (Fortsetzung (4))

Was ist mit Simulated Annealing?

mit W 'keit sehr dicht an 1 \exists schlechtes Dreieck

also $T(t) \geq w/(\gamma \ln m)$ erforderlich

Beweis für MA funktioniert für alle solche Temperaturen

also auch SA nicht erfolgreich mit W 'keit sehr dicht an 1 für jede polynomielle Schrittzahl □

eher entmutigend für einfache Graphklasse nicht erfolgreich

Kann man Graphklassen finden, für die SA in Polynomialzeit mit W 'keit dicht bei 1 erfolgreich ist?

Verallgemeinerung

Man kann zeigen $(1 + \varepsilon)$ -Trennung von Kantengewichten reicht auch aus.

hier ohne Beweis

Theorem

Sei $\varepsilon > 0$ konstant, sei $G = (V, E, w)$ ein zusammenhängender, ungerichteter, gewichteter Graph mit $w: V \rightarrow \{1, \dots, 2^{|E|}\}$ und $w(e) > w(e') \Rightarrow w(e) \geq (1 + \varepsilon)w(e')$ für alle $e \in E$.

Simulated Annealing findet mit passendem Annealingschedule einen MST für G in polynomieller Zeit mit einer Wahrscheinlichkeit polynomiell dicht an 1.

Beweisidee bei hinreichend langsamen Abkühlen werden nach fallendem Gewicht gute Kanten hinzugefügt und nicht wieder entfernt, zu schwere Kanten kommen später nicht mehr dazu

bei Interesse Ingo Wegener: Simulated annealing beats Metropolis in Combinatorial Optimization. ICALP 2005.