Utrecht, 14 juni 2012

Modellen en Simulatie
Simulated Annealing

Gerard Sleijpen

Universiteit Utrecht
Department of Mathematics

http://www.staff.science.uu.nl/∼sleij101/

In deze les een toepassing van Markov ketens:

$$\mathbf{p}_{n+1} = \mathbf{P}\mathbf{p}_n$$

met $\mathbf{p}_n$ kans vectoren $\quad (\mathbf{1}^\top \mathbf{p}_n = 1$ en $\mathbf{p}_n \geq \mathbf{0})$
en $\mathbf{P} = (p_{i,j})$ en $N \times N$ kans matrix
$(\mathbf{1}^\top \mathbf{P} = \mathbf{1}^\top$, en $p_{i,j} \geq 0$ alle $i, j)$

Stelling. 1 is een eigenwaarde van $\mathbf{P}$
met een kans eigenvector $\mathbf{q}$: $\mathbf{P}\,\mathbf{q} = \mathbf{q}$ en $\mathbf{q}$ kansvector.

Als (de graaf van) $\mathbf{P}$ irreducibel en a-periodiek is,
dan is 1 dominant en $\mathbf{p}_n \to \mathbf{q}$.

Program

- Optimaliseren
- Het type probleem
- Het handelsreizigersprobleem.
- NP-compleet
- Simulated Annealing
- De kans op succes
- Hoe T te kiezen?
- Fysische interpretatie

Program

- Optimaliseren
- Het type probleem
- Het handelsreizigersprobleem.
- NP-compleet
- Simulated Annealing
- De kans op succes
- Hoe T te kiezen?
- Fysische interpretatie

Optimaliseren

$\mathcal{V} \subset \mathbb{R}^n$ en $\mathbf{f} : \mathcal{V} \to \mathbb{R}$.

Probleem. Vind $\mathbf{x}_0 \in \mathcal{V}$ zo dat $\mathbf{f}(\mathbf{x}_0) \leq \mathbf{f}(\mathbf{x})$ voor alle $\mathbf{x} \in \mathcal{V}$.

Optimaliseren

$\mathcal{V} \subset \mathbb{R}^n$ en $\mathbf{f} : \mathcal{V} \to \mathbb{R}$.

Probleem. Vind $\mathbf{x}_0 \in \mathcal{V}$ zo dat $\mathbf{f}(\mathbf{x}_0) \leq \mathbf{f}(\mathbf{x})$ voor alle $\mathbf{x} \in \mathcal{V}$.

Voorbeeld. $f : [a, b] \to \mathbb{R}$ continu differentieerbaar.

Dan $\quad x \in \{a, b\} \quad$ of $\quad \frac{\mathrm{d}f}{\mathrm{d}x}(x_0) = 0$

Oplosmethode.

1) Gebruik Newton om de nulpunten x_i van f' te bepalen

2) Vind minimum: vergelijk $f(x_i)$, $f(a)$, $f(b)$.

Optimaliseren

$\mathcal{V} \subset \mathbb{R}^n$ en $\mathbf{f} : \mathcal{V} \to \mathbb{R}$.

Probleem. Vind $\mathbf{x}_0 \in \mathcal{V}$ zo dat $\mathbf{f}(\mathbf{x}_0) \le \mathbf{f}(\mathbf{x})$ voor alle $\mathbf{x} \in \mathcal{V}$.

Voorbeeld. $n > 1$.

1) Vind nulpunten van

$$\mathbf{F} \equiv \mathbf{D}\,\mathbf{f} = \mathrm{grad}(\mathbf{f}) = \left(\frac{\partial \mathbf{f}}{\partial x_1}, \ldots \frac{\partial \mathbf{f}}{\partial x_n}\right)^{\mathsf{T}} : \mathcal{V} \to \mathbb{R}^n$$

Gebruik Newton: $\mathbf{x}_{n+1} = \mathbf{x}_n - (\mathbf{DF}(\mathbf{x}_n))^{-1}\mathbf{F}(\mathbf{x}_n)$

met
$$\mathbf{DF} = \begin{bmatrix} \frac{\partial \mathbf{F}_1}{\partial x_1} & \cdots & \frac{\partial \mathbf{F}_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial \mathbf{F}_n}{\partial x_1} & \cdots & \frac{\partial \mathbf{F}_n}{\partial x_n} \end{bmatrix}$$

2) Vergelijk $\mathbf{f}$ waarden in de gevonden $\mathbf{x}_i$ met die in de randpunten.

Newton

Voordelen. Snelle convergentie

Problemen.

1) Vindt geen randextrema
(in hogere dimensies is de rand groot)

2) Kan "blijven hangen" in lokale extrema

3) Werkt alleen als **f** glad is
(minimaal twee maal continu differentieerbaar).

4) De stappen zijn rekenintensief voor hogere dimensies

Vaak is er geen rand:

Voorbeeld. *n negatief geladen deeltjes op een bolopper-vlak. Verdeel de deeltjes zodat de energie minimaal is.*

Newton

Voordelen. Snelle convergentie

Problemen.

1) Vindt geen randextrema
   (in hogere dimensies is de rand groot)
2) Kan “blijven hangen” in lokale extrema
3) Werkt alleen als **f** glad is
   (minimaal twee maal continu differentieerbaar).
4) De stappen zijn rekenintensief voor hogere dimensies

Als het probleem structuur heeft, buit die structuur uit.

Voorbeeld. Lineair programmeringsprobleem

(Extrema op de rand (in hoekpunten),
op te lossen met lineaire algebra technieken (vegen))

Newton

Voordelen. Snelle convergentie

Problemen.

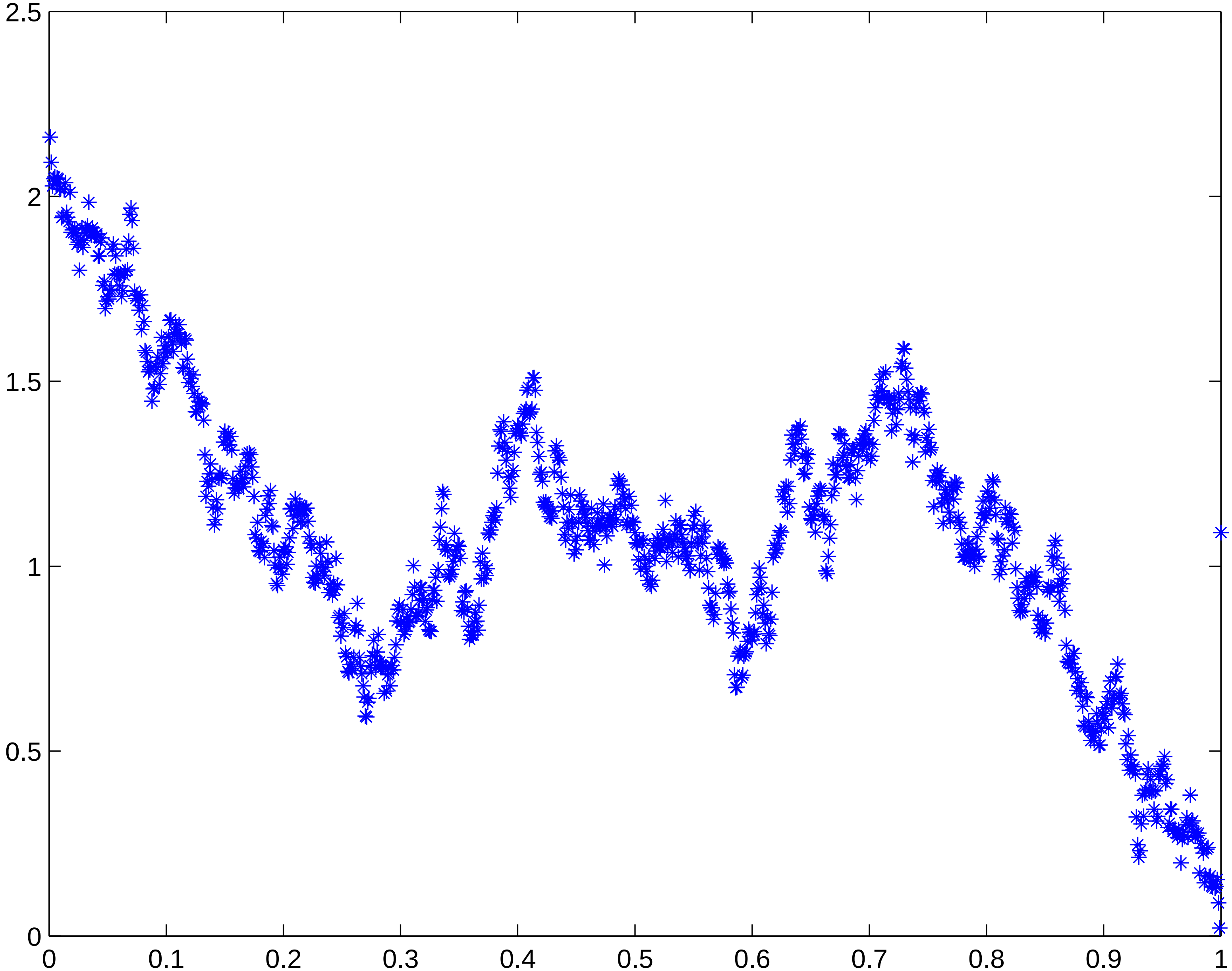
1) Vindt geen randextrema
   (in hogere dimensies is de rand groot)
2) Kan "blijven hangen" in lokale extrema
3) Werkt alleen als $\mathbf{f}$ glad is
   (minimaal twee maal continu differentieerbaar).
4) De stappen zijn rekenintensief voor hogere dimensies

Als het probleem geen (duidelijke) structuur heeft. . .

Program

- Optimaliseren
- Het type probleem
- Het handelsreizigersprobleem.
- NP-compleet
- Simulated Annealing
- De kans op succes
- Hoe T te kiezen?
- Fysische interpretatie

Voorbeeld.



Voorbeeld. $\mathbf{f} : \{t_k = k/N \mid k = 0, 1, \ldots, N\} \to \mathbb{R}$, $N = 1000$.

Probleem heeft eindig veel **f**-waarden:

brute-force methode werkt

```
m = 0, k = 0

while m ≤ N
    if f(m) < f(k),
        k ← m
    end if
    m = m + 1
end while
```

Program

- Optimaliseren
- Het type probleem
- Het handelsreizigersprobleem.
- NP-compleet
- Simulated Annealing
- De kans op succes
- Hoe T te kiezen?
- Fysische interpretatie

Handelsreizigersprobleem

Gegevens. n steden $S_1, S_2, \ldots, S_n$. De afstand L_{ij} (over de weg) tussen ieder tweetal steden S_i en S_j is bekend.

Traveling salesman problem. Vind een rondreis met de kleinste totaal afstand die langs **alle** steden gaat.

Vaak wordt **symmetrie** aangenomen: $L_{ij} = L_{ji}$.

Een andere populaire aanname is:

$$L_{ij} + L_{jk} \geq L_{ik} \text{ (\textbf{metrisch})}.$$

Wij gebruiken deze aannames hier niet.

Handelsreizigersprobleem

Gegevens. n steden $S_1, S_2, \ldots, S_n$. De afstand L_{ij} (over de weg) tussen ieder tweetal steden S_i en S_j is bekend.

Traveling salesman problem. Vind een rondreis met de kleinste totaal afstand die langs **<u>alle</u>** steden gaat.

Als $$S_1 \to S_2 \to S_3 \to \ldots \to S_n \to S_1$$

de rondreis beschrijft, dan is de totaal afstand voor deze reis:

$$f = L_{1,2} + L_{2,3} + \ldots + L_{n-1,n} + L_{n,1}.$$

Handelsreizigersprobleem

Gegevens. n steden $S_1, S_2, \ldots, S_n$. De afstand L_{ij} (over de weg) tussen ieder tweetal steden S_i en S_j is bekend.

Traveling salesman problem. Vind een rondreis met de kleinste totaal afstand die langs **<u>alle</u>** steden gaat.

Als $\pi : \{2, \ldots, n\} \to \{2, \ldots, n\}$ een permutatie is en

$$S_1 \to S_{\pi(2)} \to S_{\pi(3)} \to \ldots \to S_{\pi(n)} \to S_1$$

beschrijft de rondreis, dan is de totaal afstand voor deze reis:

$$\mathbf{f}(\pi) \equiv L_{1,\pi(2)} + L_{\pi(2),\pi(3)} + \ldots + L_{\pi(n-1),\pi(n)} + L_{\pi(n),1}.$$

Handelsreizigersprobleem

Gegevens. n steden $S_1, S_2, \ldots, S_n$. De afstand L_{ij} (over de weg) tussen ieder tweetal steden S_i en S_j is bekend.

Traveling salesman problem. Vind een rondreis met de kleinste totaal afstand die langs **alle** steden gaat.

Als $\pi : \{2, \ldots, n\} \to \{2, \ldots, n\}$ een permutatie is en

$$S_1 \to S_{\pi(2)} \to S_{\pi(3)} \to \ldots \to S_{\pi(n)} \to S_1$$

beschrijft de rondreis, dan is de totaal afstand voor deze reis:

$$\mathbf{f}(\pi) \equiv L_{1,\pi(2)} + L_{\pi(2),\pi(3)} + \ldots + L_{\pi(n-1),\pi(n)} + L_{\pi(n),1}.$$

Permutatie: Neem $n-1$ speelkaarten. Zet op iedere kaart de naam van een van de $n-1$ steden $S_2, \ldots, S_n$; op iedere kaart een andere naam. Schudt de kaarten en je heb een permutatie.

Handelsreizigersprobleem

Gegevens. n steden $S_1, S_2, \ldots, S_n$. De afstand L_{ij} (over de weg) tussen ieder tweetal steden S_i en S_j is bekend.

Traveling salesman problem. Vind een rondreis met de kleinste totaal afstand die langs **alle** steden gaat.

Als $\pi : \{2, \ldots, n\} \to \{2, \ldots, n\}$ een permutatie is en

$$S_1 \to S_{\pi(2)} \to S_{\pi(3)} \to \ldots \to S_{\pi(n)} \to S_1$$

beschrijft de rondreis, dan is de totaal afstand voor deze reis:

$$\mathbf{f}(\pi) \equiv L_{1,\pi(2)} + L_{\pi(2),\pi(3)} + \ldots + L_{\pi(n-1),\pi(n)} + L_{\pi(n),1}.$$

Permutatie: bijectie.

Handelsreizigersprobleem

Gegevens. n steden $S_1, S_2, \ldots, S_n$. De afstand L_{ij} (over de weg) tussen ieder tweetal steden S_i en S_j is bekend.

Traveling salesman problem. Vind een rondreis met de kleinste totaal afstand die langs **alle** steden gaat.

Als $\pi : \{2, \ldots, n\} \to \{2, \ldots, n\}$ een permutatie is en

$$S_1 \to S_{\pi(2)} \to S_{\pi(3)} \to \ldots \to S_{\pi(n)} \to S_1$$

beschrijft de rondreis, dan is de totaal afstand voor deze reis:

$$\mathbf{f}(\pi) \equiv L_{1,\pi(2)} + L_{\pi(2),\pi(3)} + \ldots + L_{\pi(n-1),\pi(n)} + L_{\pi(n),1}.$$

Karakteristieken: alle afstanden L_{ij} zijn bekend. Gegeven een rondreis (permutatie π) dan is $\mathbf{f}(\pi)$ **zeer** snel uit te rekenen.

Handelsreizigersprobleem

Gegevens. n steden $S_1, S_2, \ldots, S_n$. De afstand L_{ij} (over de weg) tussen ieder tweetal steden S_i en S_j is bekend.

Traveling salesman problem. Vind een rondreis met de kleinste totaal afstand die langs **alle** steden gaat.

Als $\pi : \{2, \ldots, n\} \to \{2, \ldots, n\}$ een permutatie is en

$$S_1 \to S_{\pi(2)} \to S_{\pi(3)} \to \ldots \to S_{\pi(n)} \to S_1$$

beschrijft de rondreis, dan is de totaal afstand voor deze reis:

$$\mathbf{f}(\pi) \equiv L_{1,\pi(2)} + L_{\pi(2),\pi(3)} + \ldots + L_{\pi(n-1),\pi(n)} + L_{\pi(n),1}.$$

Brute-force?

Moeilijkheid: het aantal permutaties is heel groot.

[wikipedia]
An optimal TSP tour through Germanys 15 largest cities. It is the shortest among 43 589 145 600 possible tours visiting each city exactly once.

Brute-force

Met $n = 31$ steden is het aantal permutaties

$$30! \approx 2.6\,10^{30}$$

Stel dat je 10^9 (1 Terra) mogelijke rondreizen per seconde kunt checken, dan duurt de brute-force methode

$$\approx 2.6\,10^{21}\,\text{sec} \approx 8\,10^{13}\,\text{jaar.}$$

($1\,\text{jaar} \approx 3\,10^7\,\text{sec.}$).

Brute-force

Met $n = 31$ steden is het aantal permutaties

$$30! \approx 2.6\,10^{30}$$

Stel dat je 10^9 (1 Terra) mogelijke rondreizen per seconde kunt checken, dan duurt de brute-force methode

$$\approx 2.6\,10^{21}\,\text{sec} \approx 8\,10^{13}\,\text{jaar}.$$

($1\,\text{jaar} \approx 3\,10^7\,\text{sec.}$).

[Wikipedia] In the theory of computational complexity, the decision version of TSP belongs to the class of NP-complete problems. Thus, it is assumed that there is no efficient algorithm for solving TSP problems. In other words, it is likely that the worst case running time for any algorithm for TSP increases exponentially with the number of cities, so even some instances with only hundreds of cities will take many CPU years to solve exactly.

Program

- Optimaliseren
- Het type probleem
- Het handelsreizigersprobleem.
- NP-compleet
- Simulated Annealing
- De kans op succes
- Hoe T te kiezen?
- Fysische interpretatie

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als “gemakkelijk” na te gaan is dat een oplossing correct is.

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als “gemakkelijk” na te gaan is dat een oplossing correct is.

Voorbeeld [het deelverzameling probleem]. Zij $\mathcal{F}$ een eindige collecte gehele getallen (vb $\mathcal{F} = \{-7, -3, -2, 5, 8\}$). Bepaal, indien mogelijk, een deelverzameling $\mathcal{G}$ van $\mathcal{F}$ waarvan alle getallen optellen tot 0 (als $\mathcal{G} = \{-3, -2, 5\}$)?

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als “gemakkelijk” na te gaan is dat een oplossing correct is.

Voorbeeld [het deelverzameling probleem]. Zij $\mathcal{F}$ een eindige collecte gehele getallen (vb $\mathcal{F} = \{-7, -3, -2, 5, 8\}$). Bepaal, indien mogelijk, een deelverzameling $\mathcal{G}$ van $\mathcal{F}$ waarvan alle getallen optellen tot 0 (als $\mathcal{G} = \{-3, -2, 5\}$)?

“Gemakkelijk” Benodigde rekentijd (of aantal rekenoperaties) is $\sim n^k$, waarbij n het aantal input parameters is en k een positief geheel getal. (**polynomiale tijd**)

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als "gemakkelijk" na te gaan is dat een oplossing correct is.

Voorbeeld [het deelverzameling probleem]. Zij $\mathcal{F}$ een eindige collecte gehele getallen (vb $\mathcal{F} = \{-7, -3, -2, 5, 8\}$). Bepaal, indien mogelijk, een deelverzameling $\mathcal{G}$ van $\mathcal{F}$ waarvan alle getallen optellen tot 0 (als $\mathcal{G} = \{-3, -2, 5\}$)?

"Gemakkelijk" Benodigde rekentijd (of aantal rekenoperaties) is $\sim n^k$, waarbij n het aantal input parameters is en k een positief geheel getal. (**polynomiale tijd**)

Bestaat er voor ieder NP probleem een oplosmethode die in polynomiale tijd 'n oplossing vindt? P=NP?

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als “gemakkelijk” na te gaan is dat een oplossing correct is.

Voorbeeld [het deelverzameling probleem]. Zij $\mathcal{F}$ een eindige collecte gehele getallen (vb $\mathcal{F} = \{-7, -3, -2, 5, 8\}$). Bepaal, indien mogelijk, een deelverzameling $\mathcal{G}$ van $\mathcal{F}$ waarvan alle getallen optellen tot 0 (als $\mathcal{G} = \{-3, -2, 5\}$)?

“Gemakkelijk” Benodigde rekentijd (of aantal rekenoperaties) is $\sim n^k$, waarbij n het aantal input parameters is en k een positief geheel getal. (**polynomiale tijd**)

Bestaat er voor ieder NP probleem een oplosmethode die in polynomiale tijd ’n oplossing vindt? P=NP?

Antwoord levert 1MEuro op!

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als "gemakkelijk" na te gaan is dat een oplossing correct is.

Voorbeeld [het deelverzameling probleem]. Zij $\mathcal{F}$ een eindige collecte gehele getallen (vb $\mathcal{F} = \{-7, -3, -2, 5, 8\}$). Bepaal, indien mogelijk, een deelverzameling $\mathcal{G}$ van $\mathcal{F}$ waarvan alle getallen optellen tot 0 (als $\mathcal{G} = \{-3, -2, 5\}$)?

"Gemakkelijk" Benodigde rekentijd (of aantal rekenoperaties) is $\sim n^k$, waarbij n het aantal input parameters is en k een positief geheel getal. (**polynomiale tijd**)

Een NP-probleem is **NP-compleet** als er geen proces **bekend** is dat een oplossing vindt in polynomiale tijd.

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als "gemakkelijk" na te gaan is dat een oplossing correct is.

Een NP-probleem is **NP-compleet** als er geen proces **bekend** is dat een oplossing vindt in polynomiale tijd.

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als "gemakkelijk" na te gaan is dat een oplossing correct is.

Een NP-probleem is **NP-compleet** als er geen proces **bekend** is dat een oplossing vindt in polynomiale tijd.

Moeten we blij zijn met polynomiale tijd?

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als "gemakkelijk" na te gaan is dat een oplossing correct is.

Een NP-probleem is **NP-compleet** als er geen proces **bekend** is dat een oplossing vindt in polynomiale tijd.

Voorbeelden van NP-complete problemen.

- Handelsreizigersprobleem
- Deelverzamelingprobleem
- Het vinden van een Hamilton pad (of cyclus) in een graaf (indien mogelijk).

  dwz: vind een pad dat **alle** punten in de graaf **precies één keer** aan doet.

NP-compleet

Een probleem is van complexiteitsklasse **Nondeterministic Polynomial time** als “gemakkelijk” na te gaan is dat een oplossing correct is.

Een NP-probleem is **NP-compleet** als er geen proces **bekend** is dat een oplossing vindt in polynomiale tijd.

Voorbeelden van NP-complete problemen.

- Handelsreizigersprobleem
- Deelverzamelingprobleem
- Het vinden van een Hamilton pad (of cyclus) in een graaf (indien mogelijk).
- Bottleneck TSP

Handelsreizigersprobleem

Brute-force werkt niet.

Er zijn diverse ‘intelligentere’ exacte oplosmethoden bekend die de oplossing vinden in $\sim n^2\,2^n$

Handelsreizigersprobleem

Brute-force werkt niet.

Er zijn diverse 'intelligentere' exacte oplosmethoden bekend die de oplossing vinden in $\sim n^2\,2^n$ en zelfs in $\sim 1.9999^n$

Handelsreizigersprobleem

Brute-force werkt niet.

Er zijn diverse ‘intelligentere’ exacte oplosmethoden bekend die de oplossing vinden in $\sim n^2\, 2^n$ en zelfs in $\sim 1.9999^n$

Exacte oplossing bekend (gebruik makend van lineair programmeer-achtige technieken en vereenvoudigingstechnieken die probleem afhankelijk zijn) voor
15.112 Duitse steden (2001):
$\approx$ 4 jaar op een 2.8GHz computer
24.978 Zweedse steden (2004)
33.810 circuit board (2005)
85.900 circuit board (2006):
$\approx$ 136 jaar op een 2.8GHz computer

Handelsreizigersprobleem

Brute-force werkt niet.

Er zijn diverse 'intelligentere' exacte oplosmethoden bekend die de oplossing vinden in $\sim n^2\,2^n$ en zelfs in $\sim 1.9999^n$

Exacte oplossing bekend (gebruik makend van lineair programmeer-achtige technieken en vereenvoudigingstechnieken die probleem afhankelijk zijn) voor
15.112 Duitse steden (2001):
$\approx$ 4 jaar op een 2.8GHz computer
24.978 Zweedse steden (2004)
33.810 circuit board (2005)
85.900 circuit board (2006):
$\approx$ 136 jaar op een 2.8GHz computer

Blijft rekenintensief

Handelsreizigersprobleem

Brute-force werkt niet.

Er zijn diverse 'intelligentere' exacte oplosmethoden bekend die de oplossing vinden in $\sim n^2\,2^n$ en zelfs in $\sim 1.9999^n$

Alternatief. Tevreden zijn als π_0 'vlug' te berekenen is met π_0 bijvoorbeeld zo dat

$$\mathbf{f}(\pi_0) \leq 1.2 \min_\pi \mathbf{f}(\pi)$$

Handelsreizigersprobleem

Brute-force werkt niet.

Er zijn diverse 'intelligentere' exacte oplosmethoden bekend die de oplossing vinden in $\sim n^2\,2^n$ en zelfs in $\sim 1.9999^n$

Alternatief. Tevreden zijn als π_0 'vlug' te berekenen is met π_0 bijvoorbeeld zo dat

$$\mathbf{f}(\pi_0) \leq 1.2 \min_\pi \mathbf{f}(\pi)$$

Alternatief. Tevreden zijn als π_0 'vlug' te berekenen is met π_0 zo dat met grote waarschijnlijkheid

$$\mathbf{f}(\pi_0) \approx \min_\pi \mathbf{f}(\pi)$$

Simulated annealing

Program

- Optimaliseren
- Het type probleem
- Het handelsreizigersprobleem.
- NP-compleet
- Simulated Annealing
- De kans op succes
- Hoe T te kiezen?
- Fysische interpretatie

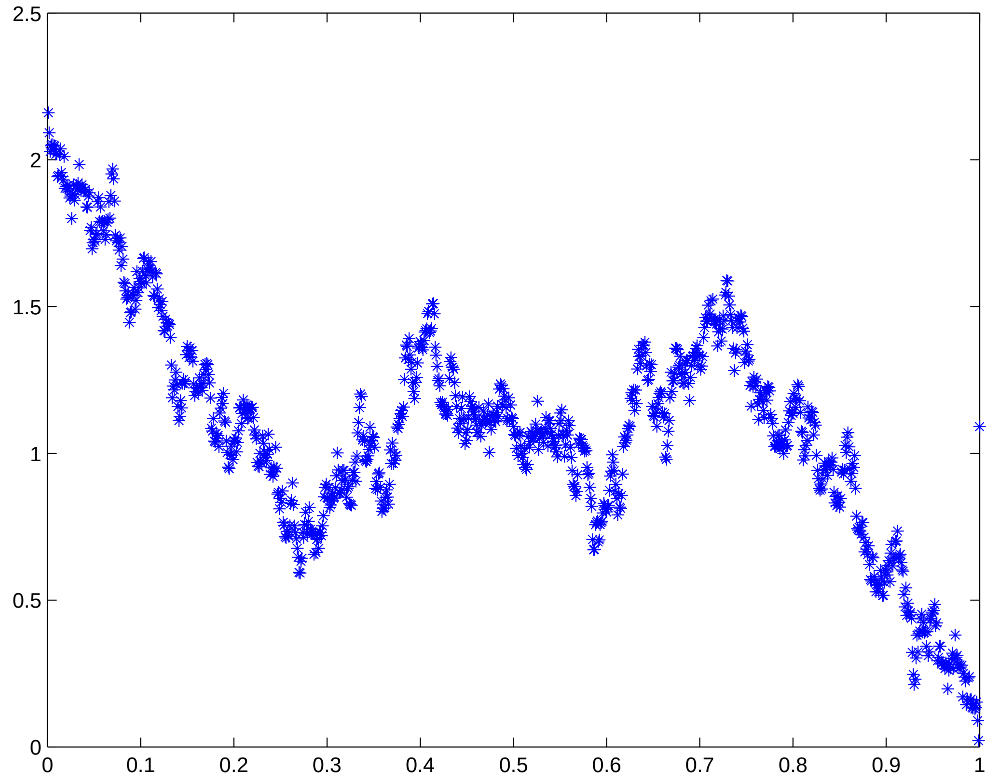
Simulated annealing

$\mathbf{f} : \mathcal{V} \to \mathbb{R}$. Vind $\mathbf{x}_0 \in \mathcal{V}$ zodat $\mathbf{f}(\mathbf{x}_0) \leq \mathbf{f}(\mathbf{x})$ alle $\mathbf{x} \in \mathcal{V}$.

Vb. 1) $\mathcal{V} = \{\pi \text{ permutatie van } \mathbf{N} \equiv \{1, 2, \ldots, n\}\}$,

$$\mathbf{f}(\pi) = L_{1,\pi(2)} + \sum_{j=2}^{n-1} L_{\pi(j),\pi(j+1)} + L_{\pi(n),1}$$

2) $\mathcal{V} = \{t_k = \frac{k-1}{N-1} \mid k = 1, \ldots, N\}$, $\quad \mathbf{f}(t_k) = f_k$.



Simulated annealing

$\mathbf{f} : \mathcal{V} \to \mathbb{R}$. Vind $\mathbf{x}_0 \in \mathcal{V}$ zodat $\mathbf{f}(\mathbf{x}_0) \leq \mathbf{f}(\mathbf{x})$ alle $\mathbf{x} \in \mathcal{V}$.

Vb. 1) $\mathcal{V} = \{\pi$ permutatie van $\mathbf{N} \equiv \{1, 2, \ldots, n\}\}$,

$$\mathbf{f}(\pi) = L_{1,\pi(2)} + \sum_{j=2}^{n-1} L_{\pi(j),\pi(j+1)} + L_{\pi(n),1}$$

2) $\mathcal{V} = \{t_k = \frac{k-1}{N-1} \mid k = 1, \ldots, N\}, \qquad \mathbf{f}(t_k) = f_k.$

Voor $\mathbf{x} \in \mathcal{V}$ zij $\mathcal{B}(\mathbf{x}) \subset \mathcal{V}$ een set van '**buur**' punten van $\mathbf{x}$.

Als $\mathbf{x}$ een buur is van $\mathbf{y}$ dan is $\mathbf{y}$ een buur van $\mathbf{x}$.

vb 1) π_1 is een buur van π als
er $j_1, j_2 \in \mathbf{N}$ zijn $j_1 \neq j_2$ zo dat
$\pi_1(j_1) = \pi(j_2)$, $\pi_1(j_2) = \pi(j_1)$
en $\pi_1(i) = \pi(i)$ voor de andere $i \in \mathbf{N}$.

vb 2) t_k is een buur van t_m als, voor $M = 10$,
$|n - j| \leq M$ voor 'n k met $m = j \bmod N$

Simulated annealing

$\mathbf{f} : \mathcal{V} \to \mathbb{R}$. Vind $\mathbf{x}_0 \in \mathcal{V}$ zodat $\mathbf{f}(\mathbf{x}_0) \leq \mathbf{f}(\mathbf{x})$ alle $\mathbf{x} \in \mathcal{V}$.

Vb. 1) $\mathcal{V} = \{\pi$ permutatie van $\mathbf{N} \equiv \{1, 2, \ldots, n\}\}$,

$$\mathbf{f}(\pi) = L_{1,\pi(2)} + \sum_{j=2}^{n-1} L_{\pi(j),\pi(j+1)} + L_{\pi(n),1}$$

2) $\mathcal{V} = \{t_k = \frac{k-1}{N-1} \mid k = 1, \ldots, N\}$, $\quad \mathbf{f}(t_k) = f_k$.

Voor $\mathbf{x} \in \mathcal{V}$ zij $\mathcal{B}(\mathbf{x}) \subset \mathcal{V}$ een set van '**buur**' punten van $\mathbf{x}$.

Kies $\mathbf{x}$ en $T > 0$
Herhaal
1) Kies random $\mathbf{y} \in \mathcal{B}(\mathbf{x})$
2) Als $\mathbf{f}(\mathbf{y}) < \mathbf{f}(\mathbf{x})$ dan $\mathbf{x} \leftarrow \mathbf{y}$
anders $\mathbf{x} \leftarrow \mathbf{y}$ met kans $e^{(\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y}))/T}$

Simulated annealing

$\mathbf{f} : \mathcal{V} \to \mathbb{R}$. Vind $\mathbf{x}_0 \in \mathcal{V}$ zodat $\mathbf{f}(\mathbf{x}_0) \leq \mathbf{f}(\mathbf{x})$ alle $\mathbf{x} \in \mathcal{V}$.

Vb. 1) $\mathcal{V} = \{\pi$ permutatie van $\mathbf{N} \equiv \{1, 2, \ldots, n\}\}$,

$$\mathbf{f}(\pi) = L_{1,\pi(2)} + \sum_{j=2}^{n-1} L_{\pi(j),\pi(j+1)} + L_{\pi(n),1}$$

2) $\mathcal{V} = \{t_k = \frac{k-1}{N-1} \mid k = 1, \ldots, N\}$, $\quad \mathbf{f}(t_k) = f_k$.

Voor $\mathbf{x} \in \mathcal{V}$ zij $\mathcal{B}(\mathbf{x}) \subset \mathcal{V}$ een set van '**buur**' punten van $\mathbf{x}$.

Kies $\mathbf{x}$ en $T > 0$
Herhaal
  1) Kies random $\mathbf{y} \in \mathcal{B}(\mathbf{x})$
  2) Als $\mathbf{f}(\mathbf{y}) < \mathbf{f}(\mathbf{x})$ dan $\mathbf{x} \leftarrow \mathbf{y}$
    anders $\mathbf{x} \leftarrow \mathbf{y}$ met kans $e^{(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y}))/T}$

Wat is de kans dat we $\quad \mathbf{x}_0 \equiv \mathit{argmin}_{\mathbf{x}}\, \mathbf{f}(\mathbf{x}) \quad$ *vinden?*

Kies $\mathbf{x}$ en $T > 0$
Herhaal
1) Kies random $\mathbf{y} \in \mathcal{B}(\mathbf{x})$
2) Als $\mathbf{f}(\mathbf{y}) < \mathbf{f}(\mathbf{x})$ dan $\mathbf{x} \leftarrow \mathbf{y}$
anders $\mathbf{x} \leftarrow \mathbf{y}$ met kans $e^{(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y}))/T}$

Kies $\mathbf{x}$ en $T > 0$
Herhaal
    1) Kies random $\mathbf{y} \in \mathcal{B}(\mathbf{x})$
    2) Als $\mathbf{f}(\mathbf{y}) < \mathbf{f}(\mathbf{x})$ dan $\mathbf{x} \leftarrow \mathbf{y}$
    anders $\mathbf{x} \leftarrow \mathbf{y}$ met kans $e^{(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y}))/T}$

Model. Zij $p_{\mathbf{x},\mathbf{y}}$ de kans dat $\mathbf{x}$ vervangen wordt door $\mathbf{y}$.

Zij b het aantal buren in $\mathcal{B}(\mathbf{x})$
(neem aan dit aantal is hetzelfde voor iedere $\mathbf{x}$.)

vb 1) $b = \frac{1}{2}(n-1)(n-2)$.
vb 2) $b = 2\,M$.

Kies $\mathbf{x}$ en $T > 0$
Herhaal
1) Kies random $\mathbf{y} \in \mathcal{B}(\mathbf{x})$
2) Als $\mathbf{f}(\mathbf{y}) < \mathbf{f}(\mathbf{x})$ dan $\mathbf{x} \leftarrow \mathbf{y}$
anders $\mathbf{x} \leftarrow \mathbf{y}$ met kans $e^{(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y}))/T}$

Model. Zij $p_{\mathbf{x},\mathbf{y}}$ de kans dat $\mathbf{x}$ vervangen wordt door $\mathbf{y}$.

Zij b het aantal buren in $\mathcal{B}(\mathbf{x})$
(neem aan dit aantal is hetzelfde voor iedere $\mathbf{x}$.)

Dan, als $\mathbf{y} \in \mathcal{B}(\mathbf{x})$

$$p_{\mathbf{x},\mathbf{y}} = \begin{cases} \frac{1}{b} & \text{als} \quad \mathbf{f}(\mathbf{y}) < \mathbf{f}(\mathbf{x}) \\ \frac{1}{b}\, e^{(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y}))/T} & \text{als} \quad \mathbf{f}(\mathbf{y}) \geq \mathbf{f}(\mathbf{x}) \end{cases}$$

als $\mathbf{y} \notin \mathcal{B}(\mathbf{x})$, $\mathbf{y} \neq \mathbf{x}$,

$$p_{\mathbf{x},\mathbf{y}} = 0$$

$p_{\mathbf{x},\mathbf{x}} = 1 - \sum p_{\mathbf{x},\mathbf{y}}$ waarbij gesommeerd over alle $\mathbf{y} \neq \mathbf{x}$.

vb 2) $\mathcal{V} = \{t_k = \frac{k-1}{N-1} \mid k = 1, \ldots, N\}$ met $N = 6$
en $\mathcal{B}(t_k) = \{t_{k-1}, t_{k+1}\}$ (de index mod N).

$\mathbf{p}_n(k)$ is de kans dat we in de n-de slag t_k vinden.

$$\mathbf{P} = \begin{bmatrix} * & \frac{1}{2}e^{(f_2-f_1)/T} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & * & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2}e^{(f_2-f_3)/T} & 0 & \frac{1}{2}e^{(f_4-f_3)/T} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & * & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2}e^{(f_4-f_5)/T} & * & \frac{1}{2} \\ \frac{1}{2}e^{(f_1-f_6)/T} & 0 & 0 & 0 & \frac{1}{2}e^{(f_5-f_6)/T} & 0 \end{bmatrix}$$

$$\mathbf{p}_n = \mathbf{P}^n \mathbf{p}_0.$$

vb 2) $\mathcal{V} = \{t_k = \frac{k-1}{N-1} \mid k = 1, \ldots, N\}$ met $N = 6$
en $\mathcal{B}(t_k) = \{t_{k-1}, t_{k+1}\}$ (de index mod N).

$\mathbf{p}_n(k)$ is de kans dat we in de n-de slag t_k vinden.

$$\mathbf{P} = \begin{bmatrix} * & \frac{1}{2}e^{(f_2-f_1)/T} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & * & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2}e^{(f_2-f_3)/T} & 0 & \frac{1}{2}e^{(f_4-f_3)/T} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & * & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2}e^{(f_4-f_5)/T} & * & \frac{1}{2} \\ \frac{1}{2}e^{(f_1-f_6)/T} & 0 & 0 & 0 & \frac{1}{2}e^{(f_5-f_6)/T} & 0 \end{bmatrix}$$

$$\mathbf{p}_n = \mathbf{P}^n \mathbf{p}_0.$$

$\mathbf{P}$ is irreducibel en a-periodiek. Dus: 1 is de dominante eigenwaarde en is $\mathbf{q}$ de kans eigenvector bij 1, dan $\mathbf{p}_n \to \mathbf{q}$. Op den duur is (ongeacht de start) de kans dat we t_k vinden gelijk aan $\mathbf{q}(k)$.

Program

- Optimaliseren
- Het type probleem
- Het handelsreizigersprobleem.
- NP-compleet
- Simulated Annealing
- De kans op succes
- Hoe T te kiezen?
- Fysische interpretatie

De dominante kans eigenvector

Stelling. Zij **P** een $N \times N$ irreducibele, a-periodieke kans-matrix. Zij **q** een kansvector van dimensie N. Als

$$(*) \qquad p_{i,j}\, q_j = p_{j,i}\, q_i \qquad (i, j = 1, \ldots, N),$$

dan is **q** de dominante kans eigenvector van **P**.

De dominante kans eigenvector

Stelling. Zij $\mathbf{P}$ een $N \times N$ irreducibele, a-periodieke kansmatrix. Zij $\mathbf{q}$ een kansvector van dimensie N. Als

$$(*) \qquad p_{i,j}\, q_j = p_{j,i}\, q_i \qquad (i, j = 1, \ldots, N),$$

dan is $\mathbf{q}$ de dominante kans eigenvector van $\mathbf{P}$.

Opmerking. Met $\mathbf{D} \equiv \mathrm{diag}(q_1, \ldots, q_N)$ geldt dat

$$(*) \quad \Leftrightarrow \quad \mathbf{PD} = \mathbf{DP}^\mathsf{T}:$$

$\mathbf{PD}$ is symmetrisch $\Leftrightarrow (*)$.

De dominante kans eigenvector

Stelling. Zij **P** een $N \times N$ irreducibele, a-periodieke kansmatrix. Zij **q** een kansvector van dimensie N. Als

$$(*) \qquad p_{i,j}\, q_j = p_{j,i}\, q_i \qquad (i,j = 1,\ldots,N),$$

dan is **q** de dominante kans eigenvector van **P**.

Opmerking. Met $\mathbf{D} \equiv \mathrm{diag}(q_1,\ldots,q_N)$ geldt dat

$$(*) \quad \Leftrightarrow \quad \mathbf{P}\mathbf{D} = \mathbf{D}\mathbf{P}^{\top}:$$

PD is symmetrisch $\Leftrightarrow (*)$.

Bewijs. We bewijzen dat **q** een eigenvector is bij eigenwaarde 1 als $(*)$ geldt. We gebruiken dat $\mathbf{1}^{\top}\mathbf{P} = \mathbf{1}^{\top}$ voor een kans matrix **P** en dat voor een irreducible, a-periodiek kansmatrix de dominante kans eigenvector uniek is.

$$\mathbf{P}\mathbf{q} = \mathbf{P}\mathbf{D}\mathbf{1} = \mathbf{D}\mathbf{P}^{\top}\mathbf{1} = \mathbf{D}(\mathbf{1}^{\top}\mathbf{P})^{\top} = \mathbf{D}\mathbf{1} = \mathbf{q}.$$

De dominante kans eigenvector

Stelling. $p_{i,j}\, q_j = p_{j,i}\, q_i \; \forall i, j \qquad \Rightarrow \qquad \mathbf{P}^n \mathbf{p}_0 \to \mathbf{q} \; (n \to \infty)$.

vb 2) $N = 6$, $M = 1$:

Met $\quad \widetilde{q}_k \equiv e^{-f_k/T} \quad$ is $\quad p_{ij}\, \widetilde{q}_j = p_{ji}\, \widetilde{q}_i$. Omdat $\widetilde{q}_k \geq 0$ is

$$q_k = \frac{\widetilde{q}_k}{\sigma} = \frac{e^{-f_k/T}}{\sigma} \qquad \text{met} \qquad \sigma = \widetilde{q}_1 + \ldots + \widetilde{q}_N.$$

De dominante kans eigenvector

Stelling. $p_{i,j}\, q_j = p_{j,i}\, q_i \;\forall i,j \quad \Rightarrow \quad \mathbf{P}^n \mathbf{p}_0 \to \mathbf{q}\ (n \to \infty)$.

vb 2)

Met $\quad \widetilde{q}_k \equiv e^{-f_k/T} \quad$ is $\quad p_{ij}\, \widetilde{q}_j = p_{ji}\, \widetilde{q}_i$. Omdat $\widetilde{q}_k \geq 0$ is

$$q_k = \frac{\widetilde{q}_k}{\sigma} = \frac{e^{-f_k/T}}{\sigma} \quad \text{met} \quad \sigma = \widetilde{q}_1 + \ldots + \widetilde{q}_N.$$

De dominante kans eigenvector

Stelling. $p_{i,j}\, q_j = p_{j,i}\, q_i \;\forall i, j \qquad \Rightarrow \qquad \mathbf{P}^n \mathbf{p}_0 \to \mathbf{q} \;(n \to \infty)$.

vb 2)

Met $\quad \tilde{q}_k \equiv e^{-f_k/T} \quad$ is $\quad p_{ij}\, \tilde{q}_j = p_{ji}\, \tilde{q}_i$. Omdat $\tilde{q}_k \geq 0$ is

$$q_k = \frac{\tilde{q}_k}{\sigma} = \frac{e^{-f_k/T}}{\sigma} \qquad \text{met} \qquad \sigma = \tilde{q}_1 + \ldots + \tilde{q}_N .$$

Stel $f_k < f_j$ alle $j \neq k$. Dan

$$S = e^{-f_k/T} \left(e^{(f_k - f_1)/T} + \ldots + 1 + \ldots + e^{(f_k - f_N)/T} \right):$$

$$q_k \to 1 \qquad \text{als} \qquad T \to 0.$$

De dominante kans eigenvector

Stelling. $p_{i,j}\, q_j = p_{j,i}\, q_i \;\forall i, j \qquad \Rightarrow \qquad \mathbf{P}^n \mathbf{p}_0 \to \mathbf{q} \;(n \to \infty)$.

vb 2)

Met $\quad \widetilde{q}_k \equiv e^{-f_k/T} \quad$ is $\quad p_{ij}\, \widetilde{q}_j = p_{ji}\, \widetilde{q}_i$. Omdat $\widetilde{q}_k \geq 0$ is

$$q_k = \frac{\widetilde{q}_k}{\sigma} = \frac{e^{-f_k/T}}{\sigma} \qquad \text{met} \qquad \sigma = \widetilde{q}_1 + \ldots + \widetilde{q}_N.$$

Stel $f_k < f_j$ alle $j \neq k$. Dan

$$S = e^{-f_k/T} \left(e^{(f_k - f_1)/T} + \ldots + 1 + \ldots + e^{(f_k - f_N)/T} \right):$$

$$q_k \to 1 \qquad \text{als} \qquad T \to 0.$$

Als het minimum wordt aangenomen voor twee k's dan is voor kleine T de kans om in van de minima terecht te komen $\approx \frac{1}{2}$.

De dominante kans eigenvector

Stelling. $p_{i,j}\, q_j = p_{j,i}\, q_i \ \forall i, j \qquad \Rightarrow \qquad \mathbf{P}^n \mathbf{p}_0 \to \mathbf{q} \ (n \to \infty).$

vb 1) Met $\widetilde{q}(\pi) \equiv e^{-\mathbf{f}(\pi)/T}$ is

$$q(\pi) = \frac{1}{\sigma}\, e^{-\mathbf{f}(\pi)/T} \qquad \text{met} \qquad \sigma \equiv \sum_{\pi} e^{-\mathbf{f}(\pi)/T}.$$

Bewering.

q_π is de kans dat het proces op den duur π oplevert.

De dominante kans eigenvector

Stelling. $p_{i,j}\, q_j = p_{j,i}\, q_i \ \forall i, j \qquad \Rightarrow \qquad \mathbf{P}^n \mathbf{p}_0 \to \mathbf{q} \ (n \to \infty).$

vb 1) Met $\tilde{q}(\pi) \equiv e^{-\mathbf{f}(\pi)/T}$ is

$$q(\pi) = \frac{1}{\sigma}\, e^{-\mathbf{f}(\pi)/T} \qquad \text{met} \qquad \sigma \equiv \sum_{\pi} e^{-\mathbf{f}(\pi)/T}.$$

Bewering.
q_π is de kans dat het proces op den duur π oplevert.

Als het minimum in precies een π_0 wordt aangenomen dan

$$q(\pi_0) \to 1 \qquad \text{als} \qquad T \to 0.$$

Bewering. Voor kleine T is de kans zeer groot dat het proces op den duur π_0 vindt.

De dominante kans eigenvector

Stelling. $p_{i,j}\,q_j = p_{j,i}\,q_i \;\forall i,j \qquad \Rightarrow \qquad \mathbf{P}^n\mathbf{p}_0 \to \mathbf{q}\ (n\to\infty)$.

Stelling. Met $q(\mathbf{x}) = \frac{1}{\sigma}\,e^{-\mathbf{f}(\mathbf{x})/T}$ en $\sigma \equiv \sum_{\mathbf{x}\in\mathcal{V}} e^{-\mathbf{f}(\mathbf{x})/T}$

is $q(\mathbf{x})$ de kans dat het proces op den duur $\mathbf{x}$ oplevert.

Opmerking. Om tot deze conclusie te komen hebben we de matrix $\mathbf{P}$ nodig (en de $p_{\mathbf{x},\mathbf{y}}$), maar $\mathbf{P}$ hebben we **niet** explicitiet nodig om het proces uit te voeren.

Als het minimum in precies in een $\mathbf{x}$ wordt aangenomen, zeg in $\mathbf{x} = \mathbf{x}_0$, dan $q(\mathbf{x}_0) \to 1$ als $T \to 0$.

Stelling. Voor kleine T is de kans zeer groot dat het proces op den duur $\mathbf{x}_0$ vindt.

Program

- Optimaliseren
- Het type probleem
- Het handelsreizigersprobleem.
- NP-compleet
- Simulated Annealing
- De kans op succes
- Hoe T te kiezen?
- Fysische interpretatie

Hoe kiezen we T?

$+$ T klein dan is de kans groot dat we op den duur in een minimaliserende $\mathbf{x}$ terecht komen.

$-$ T klein dan duurt ‘op den duur’ heel lang

Hoe kiezen we T?

+ T klein dan is de kans groot dat we op den duur in een minimaliserende **x** terecht komen.

− T klein dan duurt ‘op den duur’ heel lang

Praktische strategie.
Begin met grotere T, verklein T na een aantal slagen.

Met grotere T hebben we al ‘vlug’ een grotere kans dat we in de kleinere **f**-waarden terecht komen (maar er is ook nog een redelijke kans dat het proces andere punten oplevert). Die vormen een goed uitgangspunt voor de kleinere T, waarbij de kans dat het proces ‘echt’ verkeerde punten oplevert kleiner is.

Met welke strategie moeten we T verkleinen?
Getemperd (annealed) Trial and error.

Program

- Optimaliseren
- Het type probleem
- Het handelsreizigersprobleem.
- NP-compleet
- Simulated Annealing
- De kans op succes
- Hoe T te kiezen?
- Fysische interpretatie

Waarom met kans $e^{(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y}))/T}$ *werken als* $\mathbf{f}(\mathbf{x}) < \mathbf{f}(\mathbf{y})$*?*

Waarom met kans $e^{(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y}))/T}$ *werken als* $\mathbf{f}(\mathbf{x}) < \mathbf{f}(\mathbf{y})$*?*

1) Gemakkelijk in te zien dat $p_{\mathbf{x},\mathbf{y}}\, q(\mathbf{y}) = p_{\mathbf{y},\mathbf{x}}\, q(\mathbf{x})$

2) De distributie $e^{-\mathbf{f}(\mathbf{x})/T}$ bevoordeelt de $\mathbf{x}$ met kleinste **f**-waarde. (Hangt af van de waarde van T).

Waarom met kans $e^{(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y}))/T}$ *werken als* $\mathbf{f}(\mathbf{x}) < \mathbf{f}(\mathbf{y})$*?*

1) Gemakkelijk in te zien dat $p_{\mathbf{x},\mathbf{y}}\, q(\mathbf{y}) = p_{\mathbf{y},\mathbf{x}}\, q(\mathbf{x})$

2) De distributie $e^{-\mathbf{f}(\mathbf{x})/T}$ bevoordeelt de $\mathbf{x}$ met kleinste $\mathbf{f}$-waarde. (Hangt af van de waarde van T).

Hoe kom je op het idee?

Waarom met kans $e^{(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y}))/T}$ *werken als* $\mathbf{f}(\mathbf{x}) < \mathbf{f}(\mathbf{y})$*?*

1) Gemakkelijk in te zien dat $\quad p_{\mathbf{x},\mathbf{y}}\, q(\mathbf{y}) = p_{\mathbf{y},\mathbf{x}}\, q(\mathbf{x})$

2) De distributie $e^{-\mathbf{f}(\mathbf{x})/T}$ bevoordeelt de **x** met kleinste **f**-waarde. (Hangt af van de waarde van T).

Hoe kom je op het idee?

Statistische mechanica: de kans dat een veel deeltjes systeem (bv een stuk metaal) in toestand is met energie E bij temperatuur T (in graden Kelvin) is

$$e^{-\frac{E}{kT}} \quad \text{met } k \text{ de Boltzmann contante.}$$

k schaalt T. We kunnen een andere schaling kiezen.

Waarom met kans $e^{(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y}))/T}$ *werken als* $\mathbf{f}(\mathbf{x}) < \mathbf{f}(\mathbf{y})$*?*

1) Gemakkelijk in te zien dat $p_{\mathbf{x},\mathbf{y}}\, q(\mathbf{y}) = p_{\mathbf{y},\mathbf{x}}\, q(\mathbf{x})$

2) De distributie $e^{-\mathbf{f}(\mathbf{x})/T}$ bevoordeelt de $\mathbf{x}$ met kleinste **f**-waarde. (Hangt af van de waarde van T).

Hoe kom je op het idee?

Statistische mechanica: de kans dat een veel deeltjes systeem (bv een stuk metaal) in toestand is met energie E bij temperatuur T (in graden Kelvin) is

$e^{-\frac{E}{kT}}$ met k de Boltzmann contante.

Ervaringsfeit: door getemperd te koelen is de kans groter dat een systeem in een van de laagste energie toestanden komt.