Zusammenfassend: $v^{(k+1)} = P^T v^{(k)}$.

bzw: $v^{(k)} = (P^T)^{k-1} v^{(1)}$ $(*)$

Beachte $P \geq 0$, $Pe = e$

$$\overbrace{\|P^T\|_1}^{\text{Spaltensummennorm!}} = \max_i \sum_j |P_{ij}| = 1 \quad \overbrace{\Rightarrow}^{\text{Der Spektralradius ist } \leq \text{ jeder induzierten Norm}} \rho(P^T) \leq 1.$$

Da P den Eigenwert 1 hat hat auch P^T den Eigenwert 1

$\Rightarrow \rho(P^T) = 1$.

Jede Rundtour ist von jeder anderen aus "erreichbar" (durch obige Wechsel zu benachbarten Touren).

$\Rightarrow P$ ist irreduzibel (4. Beh. S. 43)

und primitiv (da z.B. $P_{11} > 0$).

Aus dem Satz von Perron-Frobenius folgt

P^T hat (bis auf Vielfache) nur einen Eigenvektor zu dem einzigen Eigenwert 1 vom Betrag 1, d.h.

$\exists_{=1} v^*$ mit $P^T v^* = v^*$ und für allgemein gewählte $v^{(0)} \geq 0$ konvergiert die Potenz-methode gegen einen Eigenvektor $v^* \geq 0$ zu $\lambda = 1$. (vgl. Numa 2 "Potenzmethode"). (Wegen $P^T \geq 0$ und $v^{(0)} \geq 0$ ist induktiv auch $v^{(k)} \geq 0 \ \forall k \geq 0$.)

Ferner: $e^T v^{(k)} = e^T P^T v^{(k-1)} = (Pe)^T v^{(k-1)} = e^T v^{(k-1)} = \dots = 1$.

Also $v^* > 0$ und $e^T v^* = 1$. Dabei hängt v^* von der Wahl der "Temperatur" T (S. 47) ab, $v^* = v^*(T)$.

Bem:

Eine große Wahl von T erhöht die WS. dass eine "lokale Minimalstelle" $x^{(j)}$ (d.h. eine Rundtour, zu der alle benachbarten Rundtouren eine größere Länge haben) durch eine andere Iterierte $x^{(l)}$ ersetzt wird. Für $T \to 0$ wird die WS. eine lokale Minimalstelle zu verlassen geringer.

Sei nun $D \in \mathbb{R}^{m \times m}$ die Diagonalmatrix mit $D_{ii} = e^{\frac{f(x^{(1)}) - f(x^{(i)})}{T}} \le 1$ und sei $v^* = v^*(T)$ wie oben.

Setze $\hat{v}^* := D^{-1} v^* \Rightarrow \hat{v}^*$ ist Eigenvektor von $D^{-1} P^T D$ zum Eigenwert 1. Dabei ist $Q := D P D^{-1}$ ähnlich zu P^T, d.h. auch Q hat nur den einfachen Eigenwert 1 vom Betrag 1.

$$Q_{ij} = \begin{cases} w_{ij} \exp((f(x^{(j)}) - f(x^{(i)}))/T) & \text{falls } i > j \\ 1 - \sum\limits_{l=1,\, l \ne i}^{m} P_{il} & \text{falls } i = j \\ w_{ij} & \text{falls } i < j \end{cases} \qquad \text{(s. S. 47)}$$

$$\Big(\text{wegen } \underbrace{D_{jj}^{-1} D_{ii}}_{\text{Faktor von } Q_{ji}} = e^{(f(x^{(j)}) - f(x^{(1)}))/T}\, e^{(f(x^{(1)}) - f(x^{(i)}))/T} = e^{(f(x^{(j)}) - f(x^{(i)}))/T}.\Big)$$

Für kleine $T > 0$ ist dies beschränkt und strebt gegen eine obere Dreiecksmatrix, sofern $f(x^{(1)}) < f(x^{(2)}) \le \dots < f(x^{(m)})$. (Bei Gleichheit treten Diagonalblöcke auf.)

Falls $f(x^{(1)}) < f(x^{(2)})$ (eindeutige kürzeste Rundtour)

folgt $P_{11} = Q_{11} = 1 - \sum_{j=2}^{m} P_{1j} = 1 - \sum_{j=2}^{m} w_{ij} \underbrace{\exp((f(x^{(1)}) - f(x^{(j)}))/T)}_{\xrightarrow{T\to 0} 0}$

und $Q_{1j} \xrightarrow{T\to 0} 0$ für $j \geq 2$.

Für festes $T > 0$ sind P und Q irreduzibel.

Sei $\hat{v}^* = \frac{D^{-1} v^*}{\|D^{-1} v^*\|_2}$ der Eigenvektor von Q^T zum Eigenwert 1.

Für $T \to 0$ konvergiert $Q^T = Q(T)$ gegen eine untere (Block)-Dreiecksmatrix mit $Q_{11} = 1$.

$\overset{\text{Übung}}{\Rightarrow}$ Q^T hat einen Eigenvektor $\tilde{v}$ zum Ew. 1 mit $\tilde{v}_1 \neq 0$.

Wenn[a] mit $T \to 0$ auch $\hat{v}_1^* > 0$ gilt, so folgt für

$$v_i^* = \|D^{-1} v^*\|_2 \underbrace{D_{ii}}_{\exp((f(x^1) - f(x^i))/T) \xrightarrow{T\to 0} 0 \text{ für } i \geq 2} \hat{v}_i^*$$

dass die Wahrscheinlichkeit für das Auftreten von x^1 in Schritt k für große k gegen 1 strebt.

Das Problem ist, dass k sehr groß sein muss, insbesondere wenn T klein ist. (Bei $T = 0$ geht die "Irreduzibilität" von P^T verloren, für große T wird die Wahrscheinlichkeit von x^1 sehr klein.)

[a] Siehe 6. Übung

Bewertung

⊕ Eine "begrenzte" Konvergenzaussage ist möglich.

○ Das Gegenbeispiel des 6. Übungsblatts ist konstruiert und nicht unbedingt repräsentativ.

(?) Kann eine symmetrische Definition von Nachbarschaften verhindern, dass $\lim_{T \to 0} \hat{v}_1^*(T) = 0$ gilt?

("symmetrisch" wenn ein Wechsel $x^i \to x^j$ möglich ist, dann auch $x^j \to x^i$.)

⊖ Selbst wenn $\hat{v}_1^*(0) > 0$ gilt, kann die Konvergenz sehr langsam sein und mehr als m Iterationen nötig sein um mit Wahrscheinlichkeit $\geq \frac{1}{2}$ die Tour x^1 einmal zu erreichen. (Ausgabe: beste gefundene Tour!)

⊕ Das Verfahren (und ähnliche Verfahren s. S. 53) sind auf viele andere Optimierungsprobleme insbesondere mit "diskreten Variablen" übertragbar und liefern bei Anwendungen häufig "zufriedenstellende Lösungen."

Effizient ist der Ansatz nur, wenn die meisten möglichen Lösungen nie erzeugt werden und daher die Wahrscheinlichkeit, gute oder optimale Lösungen zu erzeugen größer als "$\frac{1}{m}$" ist.