

Centro Universitário de Brasília – UniCEUB

**OTIMIZAÇÃO ALEATÓRIA GLOBAL:
O ALGORITMO SIMULATED ANNEALING**

Bolsista: Rodrigo Pereira Pasturczak
Orientador: Juan Alberto Rojas Cruz
Curso: Matemática

Brasília, 1^o semestre de 2006

SUMÁRIO

INTRODUÇÃO E METODOLOGIA	03
Capítulo 1 – CONVERGÊNCIA DE MATRIZES ESTOCÁSTICA	
1.1 Matriz Estocástica	05
1.2 Matriz Irreduzível e Matriz Aperiódica	08
1.3 Resultados Assintóticos sobre Produto de Matrizes Estocásticas	14
Capítulo 2 – O ALGORITMO SIMULATED ANNEALING	
2.1 Introdução	19
2.2 Dinâmica do Algoritmo	20
2.3 Convergência do Algoritmo	21
CONCLUSÕES E CONSIDERAÇÕES FINAIS	23
REFERÊNCIAS BIBLIOGRÁFICAS	24

INTRODUÇÃO

O problema de otimização global, isto é, encontrar o máximo ou mínimo global de uma função, aparece em muitas situações práticas na economia, estatística e ciências da engenharia. Apesar de sua clara importância e dos esforços de pesquisas realizados, a situação dos algoritmos determinísticos para resolver problemas de otimização global, tem sido insatisfatória. A situação é satisfatória para funções relativamente simples quando a função (derivada) é diferenciável e as raízes da derivada da função podem ser computadas analiticamente.

Os processos determinísticos que, em geral, geram seqüências monótonas de soluções, encontram vários tipos de problemas: a solução final é fortemente dependente do ponto de partida; tendem a ficar presos em ótimos locais; requerem propriedades diferenciáveis da função a minimizar e; dependem da estrutura de seu domínio.

Os processos de busca aleatória tentam evitar estes problemas por um processo de aleatorização, realizando uma busca global no domínio da função e/ou relaxando transitoriamente a optimalidade. Dentre os métodos de otimização aleatória, destacamos o algoritmo simulated annealing, nome devido a sua similitude com um processo físico chamado recozimento, usado para remover defeitos de metais e cristais.

Algoritmos simulated annealing têm sido cada vez mais estudados, sobretudo do ponto de vista matemático, motivado pela sua boa performance para encontrar mínimos globais, evidenciada por simulações computacionais. Ainda existem questões fundamentais a serem pesquisadas tais como a rapidez de convergência do algoritmo. Acreditamos que a análise do algoritmo simulated annealing é uma boa maneira de iniciação ao estudo dos métodos de otimização global não determinísticos.

A Heurística do método simulated annealing encontra-se nos trabalhos de Metropolis et al (1953), que propôs um método para computar a distribuição de equilíbrio de um conjunto de partículas expostas a altas temperaturas, usando uma simulação computacional. Os primeiros autores que conectaram os métodos de

Metropolis et al (1953) com problemas de otimização foram Kirkpatrick et al (1983), dando assim origem aos algoritmos tipo simulated annealing.

O primeiro estudo rigoroso deste algoritmo no contexto de cadeias de Markov não – homogêneas pode ser encontrado em Gibas (1986). Condições necessárias e suficientes para a convergência do algoritmo em problemas de otimização discreta, foram apresentadas por Hajek (1988). Anily e Federgruen (1987) estudam estes algoritmos para probabilidades de aceitação gerais e Cruz e Dorea (1998) apresentam condições simples para a convergência do algoritmo que generalizam os resultados de Anily e Federgruen .

Ainda em relação à bibliografia básica destacamos os trabalhos de Dekkers & Aarts (1991) e Bélisle (1992) que mostram a convergência destes algoritmos quando a probabilidade de aceitação é exponencial e a função objetiva f é uniformemente contínua. A abordagem de Bélisle (1992) é direta, sem fazer uso da teoria de cadeia de Markov, sendo este um dos poucos trabalhos deste tipo, merecendo por isto uma atenção especial. Em Dorea, Cruz e Baigorri (1999) podemos encontrar condições suficientes para a convergência quando a hipótese de continuidade uniforme não é assumida.

Em relação à bibliografia básica em cadeia de Markov citamos o texto de Isaacson and Madsen (1976) e a tese de doutorado de Cruz (1998). O texto de Isaacson and Madsen constitui um dos poucos sobre cadeias de Markov não-homogêneas. Neste se encontra um bom resumo dos resultados teóricos mais relevantes, contendo também muitas referências de artigos de pesquisa.

METODOLOGIA

O estudo começou com uma revisão de conceitos básicos sobre matrizes estocásticas. Foram assim revisados os conceitos de matriz estocástica, matriz irredutível, matriz aperiódica, entre outros. Na continuação, foram sumariados e adaptados resultados sobre a convergência de produtos de matrizes estocásticas. Foi notado que o algoritmo simulated annealing tem associado uma cadeia de markov não-homogênea, e a convergência do algoritmo foi estabelecida através da convergência da cadeia associada.

CAPÍTULO 1

CONVERGÊNCIA DE MATRIZES ESTOCÁSTICAS

Para um melhor entendimento deste trabalho, precisamos destacar algumas definições e conceitos iniciais, o que faremos a seguir.

1.1 Matriz Estocástica

Matriz Estocástica: É uma matriz quadrada com entradas maiores ou iguais a zero (o) na qual a soma de cada linha é igual a 1. Suas linhas são uma probabilidade. Assim $P = (P_{ij})$ é matriz estocástica se:

- i) $0 \leq P_{ij} \leq 1$
- ii) $\sum_{j \geq 1} P_{ij} = 1, \forall i.$

Exemplos de matrizes estocásticas:

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \quad P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}$$

As matrizes estocásticas estão associadas com processos estocásticos chamados cadeias de Markov.

Formalmente, uma cadeia de Markov é uma seqüência $X_1, X_2, \dots, X_n, \dots$ de variáveis aleatórias tomando valores num conjunto S satisfazendo a propriedade:

$$P(X_{n+1} = J / X_n = i_n, X_{n-1} = i_{n-1}, X_0 = i_0) =$$

$$P(X_{n+1} = J / X_n = i_n), \forall J, i_1, i_2, \dots, i_n \in S.$$

A propriedade acima é chamada propriedade Markoviana e pode ser verbalizada como:

“Para se fazer a previsão do futuro de um processo, basta conhecermos o seu presente, não importando o seu passado”

Dada uma cadeia de Markov, tomando valores em S podemos associar uma seqüência de matrizes estocásticas $P_1, P_2, \dots, P_n, \dots$ definidas como a seguir:

$$P_n = (P_n(i, j))$$

onde

$$P_n(i, j) = P(X_{n+1} = J / X_n = i)$$

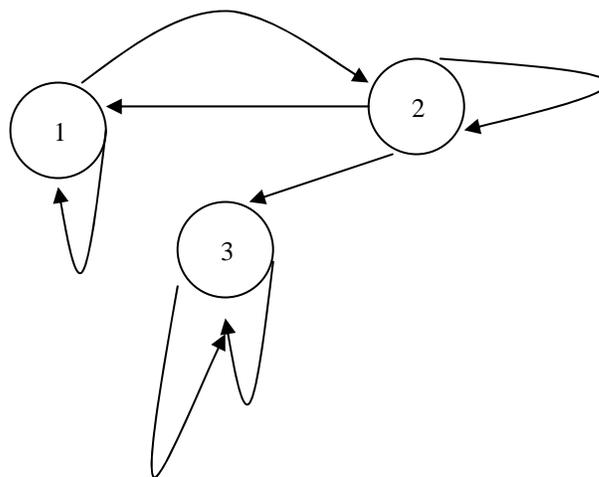
Se todas as matrizes P_n são iguais a cadeia é chamada cadeia homogênea, caso contrário, chama-se cadeia não homogênea.

Se $P(X_{n+1} = J / X_n = i) = P(X_{m+1} = J / X_m = i) \forall n, m$ anotamos simplesmente P_{ij} e falamos em cadeia homogênea.

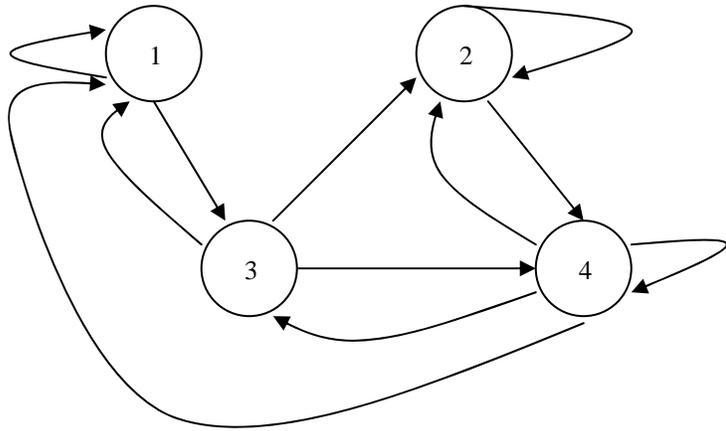
Logo, se X_1, X_2, \dots é uma cadeia homogênea, teremos somente uma matriz estocástica associada. No caso não homogênea, teremos uma seqüência de matrizes estocásticas associadas com a cadeia, que serão chamadas de matrizes de transição da cadeia.

Vejamos exemplos de matrizes estocásticas e sua representação por digrafos

$$P_1 = \begin{bmatrix} \frac{1}{2} & \frac{2}{4} & 0 \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \\ 0 & 0 & \frac{1}{1} \end{bmatrix}$$



$$-P_2 = \begin{bmatrix} X & 0 & X & 0 \\ 0 & X & 0 & X \\ X & X & 0 & X \\ X & X & X & X \end{bmatrix}$$



Propriedade 1.1: Se P e Q são matrizes estocásticas $n \times n$, então PQ também é matriz estocástica.

Corolário 1.2: Se P é matriz estocástica, então P^n também é matriz estocástica para qualquer número natural n .

1.2 Matriz Irredutível e Matriz Aperiódica

Os conceitos matriz irredutível e matriz aperiódica apresentados a seguir serão necessários na obtenção de resultados assintóticos sobre produto de matrizes estocásticas.

Matriz Irredutível: Uma matriz estocástica $P = (P_{i,j})$ é chamada irredutível se para qualquer i, j existe um número natural $n = n(i,j)$ tal que $P_{i,j}^n > 0$, isto é, a entrada i, j na matriz P^n é maior que zero.

Vejamos um exemplo:

Seja $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

Note que $P^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ assim,

$p_{(1,2)}^1 = 1 > 0$; $p_{(2,1)}^1 = 1 > 0$; $p_{(1,1)}^2 = 1 > 0$; $p_{(2,2)}^2 = 1 > 0$, logo a matriz P é irredutível.

Vejam os outros exemplos:

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/3 & 2/3 \\ 1 & 0 & 0 \end{pmatrix}$$

Note que $P^2 = \begin{pmatrix} 0 & 1/3 & 2/3 \\ 2/3 & 1/9 & 2/9 \\ 0 & 1 & 0 \end{pmatrix}$

e

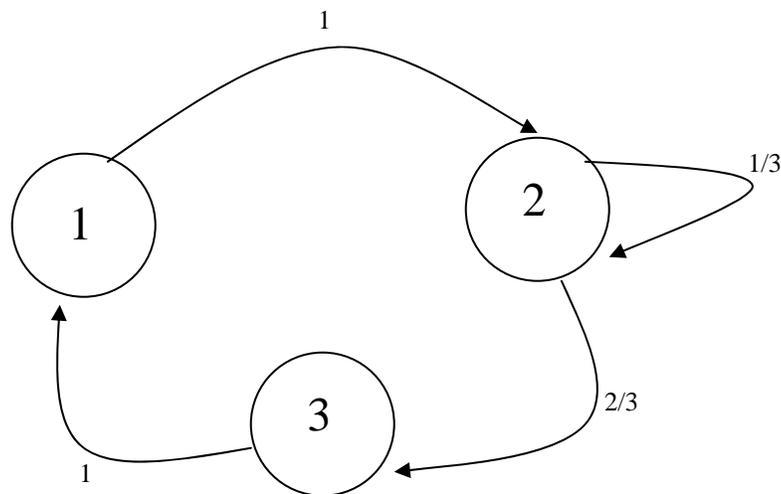
$$P^3 = \begin{pmatrix} 2/3 & 1/9 & 2/9 \\ 2/9 & 19/27 & 2/27 \\ 0 & 1/3 & 2/3 \end{pmatrix}$$

Assim para todo par (i,j) existe um natural n tal que $P_{i,j}^n > 0$, logo P é matriz irredutível.

A verificação da irredutibilidade de uma matriz pode ser analisada de maneira gráfica como veremos a seguir.

Consideremos novamente a matriz $P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/3 & 2/3 \\ 1 & 0 & 0 \end{pmatrix}$ a qual tem associada o seguinte

digrafo:



Note que para qualquer par de estado (i,j) existe um caminho seguindo a orientação do digrafo, por exemplo se começamos do estado 1 temos um caminho de comprimento 2 para chegarmos ao estado 3, a saber o caminho que vai de 1 a 2 e de 2 a 3.

Observe também que $P_{(3,3)}^1 = 0$; $P_{(3,3)}^2 = 0$ porém

$P_{(3,3)}^3 = 2/3 > 0$, o que mostra que a verificação da irredutibilidade de uma matriz pode envolver o calculo de varias potências desta.

Surge o seguinte questionamento: Considere a situação no qual o calculo das n primeiras potências $P^1, P^2, P^3, \dots, P^n$ de uma matriz $P_{(n \times n)}$ obtemos zero numa entrada fixa (i_0, j_0) , então podemos concluir que $P_{(i_0, j_0)}^k = 0$ para qualquer número natural k ?

A seguinte proposição responde o questionamento acima.

Proposição 1.3: Seja P uma matriz estocástica de ordem $n \times n$ tal que

$$P_{(i_0, j_0)} = 0 \quad ; \quad P_{(i_0, j_0)}^2 = 0 \quad ; \quad P_{(i_0, j_0)}^3 = 0 \quad \dots \quad P_{(i_0, j_0)}^n = 0$$

então $P_{(i_0, j_0)}^m = 0$ para todo número natural m .

A demonstração do resultado acima utiliza o seguinte lema:

Lema 1.4: Seja P uma matriz estocástica $n \times n$ tal que $P_{(i_0, j_0)}^k > 0$. Então existem índices $i_1, i_2, \dots, i_k, j_0$ tais que:

$$P_{(i_0, i_1)} \cdot P_{(i_1, i_2)} \cdot \dots \cdot P_{(i_k, j_0)} > 0. \text{ O recíproco também é verdadeiro.}$$

A seqüência de índices $i_1, i_2, \dots, i_k, j_0$ é chamada de **caminho** de comprimento k unindo i_0 com j_0 . Este caminho e denotado por $i_1 \rightarrow i_2 \rightarrow i_3 \rightarrow \dots \rightarrow i_k \rightarrow j_0$.

Demonstração :

O resultado segue-se diretamente da definição de multiplicação de matrizes. Consideremos o caso $k = 3$, isto é, suponhamos

$$P_{(i_0, j_0)}^3 > 0 \quad \text{logo}$$

$$\sum_{k=1}^n P_{(i_0,k)} \cdot P_{(k,j_0)}^2 > 0$$

portanto existe um i_1 tal que

$$P_{(i_0,i_1)} \cdot P_{(i_1,j_0)}^2 > 0 \quad \text{logo} \quad P_{(i_0,i_1)} > 0 \quad \text{e} \quad P_{(i_1,j_0)}^2 > 0$$

De maneira similar, $P_{(i_1,j_0)}^2 > 0$ implica que existe i_2 tal que

$P_{(i_1,i_2)} > 0$ e $P_{(i_2,j_0)} > 0$ assim temos $P_{(i_0,i_1)} \cdot P_{(i_1,i_2)} \cdot P_{(i_2,j_0)} > 0$, o que finaliza a prova.

O conceito de **ciclo** será utilizado na demonstração da proposição 1.3.

Um caminho $i_0 \rightarrow i_1 \rightarrow i_2 \rightarrow \dots \rightarrow j_0$ é chamado um ciclo se $i_0 = j_0$.

Demonstração da Proposição 1.3:

Mostremos que $P_{(i_0,j_0)}^{n+1} = 0$. Suponhamos por absurdo que $P_{(i_0,j_0)}^{n+1} > 0$, segue-se do lema 1.4 a existência de um caminho $i_0 \rightarrow i_1 \rightarrow i_2 \rightarrow i_3 \rightarrow \dots \rightarrow i_{n+1} \rightarrow j_0$ de i_0 para j_0 , assim $P_{(i_0,i_1)} \cdot P_{(i_1,i_2)} \cdot \dots \cdot P_{(i_n,i_{n+1})} > 0$, como $i_0, i_1, \dots, i_{n+1} \in \{1, 2, \dots, n\}$ concluímos que os índices $i_0, i_1, \dots, i_{n+1}, j_0$ não podem ser todos distintos, logo deve existir um ciclo $i_r, i_{r+1}, \dots, i_{r+s} = i_r$ que faz parte do caminho $(i_0, i_1, \dots, i_r, \dots, i_{r+s}, i_{r+s+1}, \dots, j_0)$ unindo i_0, j_0 . Retirando o ciclo $i_r, i_{r+1}, \dots, i_{r+s}$ obtemos o caminho $(i_0, i_1, \dots, i_{r-1}, i_{r+s+1}, \dots, j_0)$ unindo i_0 com j_0 de comprimento menor que $n+1$, utilizando novamente o lema 1.4, obtemos uma contradição com a hipótese $P_{(i_0,j_0)}^i = 0$ se $i \in \{1, 2, \dots, n\}$.

Matriz Aperiódica: É uma matriz estocástica P de ordem $n \times n$ tal que:

$\text{mdc} \{ n \in \mathbb{N}^* / P_{i,i}^n > 0 \} = 1$ para todo índice $i \in \{1, 2, \dots, n\}$, onde mdc denota o máximo comum divisor.

Os exemplos a seguir, esclareceram o conceito de matriz aperiódica.

a) Seja
$$P = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}$$

verifiquemos para $i = 1$ temos:

$$P_{(1,1)}^1 = 1 > 0, \quad P_{(1,1)}^2 = 1 > 0, \quad \dots, \quad P_{(1,1)}^n = 1 > 0,$$

Assim, $\text{mdc} \{ n \in \mathbb{N}^* / P_{i,i}^n > 0 \} = \text{mdc} \{ 1,2,3,\dots \} = 1$. De maneira similar mostra-se que $\text{mdc} \{ n \in \mathbb{N}^* / P_{i,i}^n > 0 \} = 1$ para os outros valores de i .

b) Seja
$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

para $i = 1$ temos:

$$P_{(1,1)}^1 = 0, \quad P_{(1,1)}^2 = 0, \quad P_{(1,1)}^3 = \frac{1}{2}, \quad P_{(1,1)}^4 = \frac{1}{2}, \quad P_{(1,1)}^5 = 0$$

...

logo, $\text{mdc} \{ n \in \mathbb{N}^* / P_{i,i}^n > 0 \} = \text{mdc} \{ 3,4,\dots \} = \text{mdc} \{ 3,4 \} = 1$.

Para determinar se uma matriz é aperiódica, temos que calcular o mdc associado a todos os estados. Faremos então a verificação para os estados 2, 3 e 4 da matriz P .

para $i = 2$

$$\text{mdc} \{ n \in \mathbb{N}^* / P_{2,2}^n > 0 \} = \text{mdc} \{ 3,4 \} = 1$$

para $i = 3$

$$\text{mdc} \{ n \in \mathbb{N}^* / P_{3,3}^n > 0 \} = \text{mdc} \{ 3,4 \} = 1$$

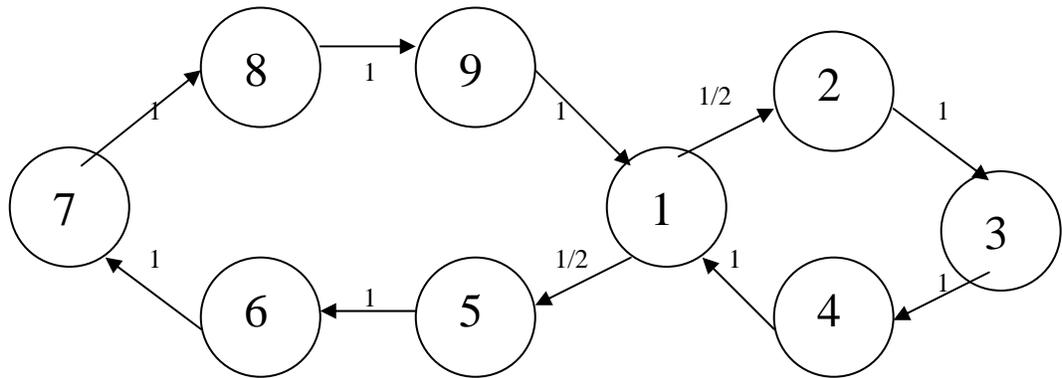
para $i = 4$

$$\text{mdc} \{ n \in \mathbb{N}^* / P_{4,4}^n > 0 \} = \text{mdc} \{ 4,11 \} = 1$$

Assim, $\text{mdc} \{ n \in \mathbb{N}^* / P_{i,i}^n > 0 \} = 1$ para todo $i \in \{ 1,2,3,4 \}$.

Apresentamos a seguir uma matriz que não é aperiódica.

Seja P uma matriz estocástica de ordem 9×9 com a seguinte representação de digrafos:



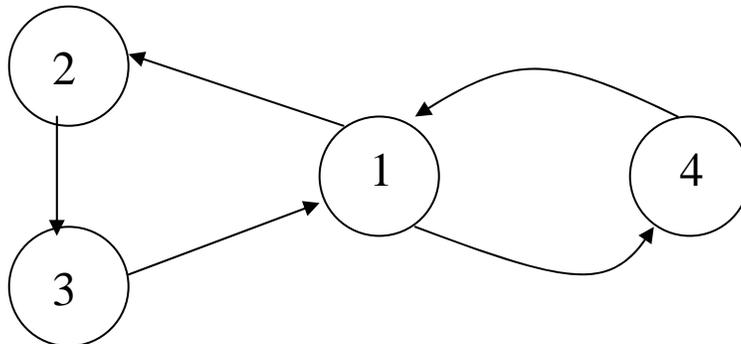
$$P_{(1,1)}^6 = \frac{1}{6} > 0 \quad ; \quad P_{(1,1)}^4 = \frac{1}{6} > 0$$

$$\text{mdc} \{ n \in \mathbb{N}^* / P_{1,1}^n > 0 \} = \text{mdc} \{ 6, 4, 8, 10, 12, \dots \} = 2.$$

logo a matriz P não é aperiódica.

O exemplo acima mostra uma matriz com “ciclos”, a qual não é aperiódica; uma conjectura ingênua seria pensar que uma matriz com ciclos não é aperiódica. O exemplo a seguir esclarece essa conjectura.

Consideremos a matriz P com a representação por digrafos



Note que

$$P_{(1,1)}^3 > 0 \quad \text{e} \quad P_{(1,1)}^2 > 0$$

para $i = 1$

$$\text{mdc} \{ n \in \mathbb{N}^* / P_{1,1}^n > 0 \} = \text{mdc} \{ 2,3,5, \dots \} = 1$$

para $i = 2$

$$\text{mdc} \{ n \in \mathbb{N}^* / P_{2,2}^n > 0 \} = \text{mdc} \{ 3,5, \dots \} = 1 .$$

para $i = 3$

$$\text{mdc} \{ n \in \mathbb{N}^* / P_{3,3}^n > 0 \} = \text{mdc} \{ 3,5, \dots \} = 1$$

para $i = 4$

$$\text{mdc} \{ n \in \mathbb{N}^* / P_{4,4}^n > 0 \} = \text{mdc} \{ 2,5, \dots \} = 1$$

Assim, $\text{mdc} \{ n \in \mathbb{N}^* / P_{i,i}^n > 0 \} = 1$ para todo $i \in \{ 1,2,3,4 \}$. Logo a matriz P é aperiódica.

O exemplo acima é significativo, pois é comum verbalizar matriz aperiódica dizendo que ela não possui ciclos. Essa verbalização não é correta.

1.3 Resultados Assintóticos sobre Produto de Matrizes Estocásticas

Os resultados a seguir serão fundamentais no estudo assintótico do algoritmo Simulated Annealing. Fixemos algumas definições. Uma matriz estocástica P é chamada **constante** se todas as linhas são iguais. Uma matriz estocástica P que converge para uma matriz constante é chamada **ergódica**.

Teorema 1.5: Se P é matriz estocástica $n \times n$ irredutível e aperiódica, então:

$$i) \quad \lim_{k \rightarrow \infty} P_{(i,j)}^k = \pi_j \quad \text{para todo } i, j \in \{ 1,2, \dots, n \}$$

$$\text{ii)} \quad \pi_j \geq 0 \quad \text{e} \quad \sum_{j=1}^n \pi_j = 1$$

O próximo resultado mostra como pode ser encontrado o limite de uma matriz ergódica.

Teorema 1.6: Seja P é matriz ergódica tal que $\lim_{k \rightarrow \infty} P^k = Q$, então Q é a única matriz constante satisfazendo a condição $Q = QP$.

O teorema 1.5 pode ser reformulado num contexto mais geral, o que facilita sua aplicabilidade;

Teorema 1.7: Considere uma matriz estocástica P de ordem n com a decomposição $P = \begin{pmatrix} P_1 & 0 \\ Q & T \end{pmatrix}$ onde P_1 é matriz $n_1 \times n_1$ e T é matriz $n_2 \times n_2$. Suponha também:

i) P_1 é irredutível e aperiódica.

ii) $\sum_{k=1}^{\infty} P_{(i,i)}^k < \infty$ para todo $i \in C_2$, onde $C_2 = \{ n_1+1, n_1+2, \dots, n_1+n_2 \}$.

então a matriz P é ergódica, mais ainda:

$$\lim_{k \rightarrow \infty} P_{(i,j)}^k = \begin{cases} \pi_j & \text{se } j \in \{ 1,2,\dots,n_1 \} \\ 0 & \text{se } j \in C_2 \end{cases}$$

A condição ii) aparentemente artificial possui uma interessante interpretação em termos de probabilidade de retorno ao estado i . De fato, $\sum_{k=1}^{\infty} P_{(i,i)}^k = \infty$ se e somente se a probabilidade de retorno ao estado i é um.

Essa condição pode ser de difícil verificação, o que nos leva a desenvolver e demonstrar o lema a seguir:

Lema 1.8: No teorema 1.7 podemos substituir a hipótese ii) pela condição ii*)

$$\text{ii*)} \quad P_{(i,j)} = \begin{cases} > 0 & \text{se } i \in C_2, j \in C_1 \\ 0 & \text{se } i \in C_1, j \in C_2 \end{cases}$$

onde $c_1 = \{ 1, 2, \dots, n_1 \}$ e $c_2 = \{ n_1 + 1, \dots, n_1 + n_2 \}$

Resulta claro que a condição ii*) é de mais fácil verificação que ii).

Demonstração:

Como P é matriz estocástica $P_{ic_1} = \sum_{k \in c_1} P_{(i,k)} + \sum_{k \in c_2} P_{(i,k)} = 1$.

Se $i \in c_2$ então por hipóteses ii)* segue-se que

$$P_{ic_1} = \sum_{k \in c_1} P_{(i,k)} = \alpha_1, \text{ com } 0 < \alpha_1 \leq 1$$

Assim

$$P_{ic_1} = \sum_{k \in c_1} P_{(i,k)} = \alpha_1 \text{ para todo } i \in c_2$$

e

$$P_{ic_2} = \sum_{k \in c_2} P_{(i,k)} = 1 - \alpha_1, \quad 0 \leq 1 - \alpha_1 < 1 \text{ (*)}$$

Seja

$$P_{ic_2}^2 = \sum_{k=1}^n P_{(i,k)} P_{kc_2}, \text{ para todo } i \in c_2, \text{ logo}$$

$$\begin{aligned} P_{ic_2}^2 &= \sum_{k \in c_1} P_{(i,k)} P_{kc_2} + \sum_{k \in c_2} P_{(i,k)} P_{kc_2} \\ &= \sum_{k \in c_2} P_{(i,k)} P_{kc_2} \quad (\text{pois } P_{kc_2} = 0 \text{ se } k \in c_1) \\ &= (1 - \alpha_1) \sum_{k \in c_2} P_{(i,k)} \quad (\text{usando (*)}) \\ &= (1 - \alpha_1) P_{ic_2} \\ &= (1 - \alpha_1)^2 \quad (\text{usando (*)}) \end{aligned}$$

Logo

$$P_{ic_2}^2 = (1 - \alpha_1)^2 \quad \text{para todo } i \in c_2.$$

De maneira similar, mostra-se que:

$$P_{ic_2}^3 = \sum_{k=1}^n P_{(i,k)} \cdot P_{kc_2}^2 = (1 - \alpha_1)^3, \quad \text{para todo } i \in c_2.$$

Em geral temos

$$P_{ic_2}^{m+1} = \sum_{k=1}^n P_{(i,k)} P_{ic_2}^m = (1 - \alpha_1)^{m+1}, \quad \text{para todo } i \in c_2.$$

Como

$$P_{(i,i)}^{m+1} \leq P_{ic_2}^{m+1}$$

concluimos que

$$P_{(i,i)}^{m+1} \leq (1 - \alpha_1)^{m+1}.$$

Por último, a convergência da série geométrica $\sum_{m=1}^{\infty} (1 - \alpha_1)^m$ ($|1 - \alpha_1| < 1$), garante

a convergência da série $\sum_{m=1}^{\infty} P_{(i,i)}^m$, para todo $i \in c_2$, o que finaliza a demonstração.

O resultado a seguir é um dos resultados mais importantes sobre produto de matrizes estocásticas, o qual será fundamental para estabelecer a convergência do algoritmo Simulated Annealing.

Teorema 1.9: Suponha uma seqüência de matrizes estocásticas $P_{(1)}, P_{(2)}, P_{(3)}, \dots$ tais que o $\lim_{n \rightarrow \infty} P_{(n)} = P$, com P matriz ergódica, então:

$$\lim_{n \rightarrow \infty} P_{(1)} \cdot P_{(2)} \cdot \dots \cdot P_{(n)} = Q \quad \text{onde } Q = \lim_{n \rightarrow \infty} P^n$$

Apresentemos uma aplicação didática do teorema 4.

Exemplo:

Seja $(P_{(n)})$ a seqüência de matrizes definida por:

$$P_n = \begin{pmatrix} \frac{1}{4} - \frac{|\cos n|}{n} & \frac{3}{4} + \frac{|\sin n|}{n} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \quad \text{logo}$$

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} 1/4 & 3/4 \\ 1/3 & 2/3 \end{pmatrix} = P$$

segue-se do teorema 1.5 que P é matriz ergodica, assim existe o $\lim_{n \rightarrow \infty} P^n = Q$ onde Q é matriz constante e $Q = QP$, ou seja:

$$(x \ y) = (x \ y) \begin{pmatrix} 1/4 & 3/4 \\ 1/3 & 2/3 \end{pmatrix} \text{ e } \quad x + y = 1 \quad ; \quad x, y > 0$$

resolvendo:

$$\begin{cases} x = \frac{x}{4} + \frac{y}{4} \\ x + y = 1 \end{cases}$$

encontramos $x=1/4$ e $y=3/4$

do Teorema 1.8 concluímos que o produto de matrizes $P_{(1)} \cdot P_{(2)} \cdot \dots \cdot P_{(n)}$ converge para a matriz constante

$$\begin{pmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \end{pmatrix}$$

A demonstração dos teoremas 1.5,1.6,1.7 e 1.9 podem ser encontradas em **Isaacson, D and Madsen, Markov chains**, Wiley, new York, (1975).O Lema lema 1.8 é de nossa autoria.

CAPÍTULO 2

O ALGORITMO SIMULATED ANNEALING

2.1 INTRODUÇÃO

Dada uma função $f : S \rightarrow R$ considere o problema de se estimar o mínimo global de f , isto é, $\min \{ f(i) / i \in S \}$.

Uma abordagem aleatória bastante conhecida é o algoritmo Simulated Annealing, o qual gera uma seqüência $x_1, x_2, \dots, x_n, \dots$ de soluções as quais, eventualmente, identificam o mínimo global da função f . A principal virtude deste algoritmo é evitar os mínimos locais da função. Como veremos a seguir, a dinâmica do algoritmo é Markoviana, isto é, dados os valores x_1, x_2, \dots, x_n geradas pelo algoritmo, o próximo valor x_{n+1} é independente dos valores x_1, x_2, \dots, x_n . É justamente esta perda de memória que convida a utilizar a teoria das Cadeias de Markov no estudo assintótico do algoritmo.

Entre as principais questões associadas ao algoritmo destacamos:

- i) Existe independência assintótica da solução inicial x_i ?
- ii) A seqüência de valores $x_1, x_2, \dots, x_n, \dots$ gerada pelo algoritmo, converge para $S_m = \{ x \in S / x \text{ ponto de mínimo de } f \}$?

Formularemos com mais precisão os questionamentos i) e ii). Se x_1, x_2, \dots, x_n é a seqüência de variáveis aleatórias associadas com o algoritmo, então i) é equivalente a:

$$\lim_{n \rightarrow \infty} (P(x_n = i / x_1 = j_1) - (x_n = i / x_1 = j_2)) = 0$$

para todo i, j_1, j_2 em S . A condição ii) é equivalente a:

existe um vetor probabilidade $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ tal que:

$$\lim_{n \rightarrow \infty} P(x_n = i / x_1 = j) = \pi_i \text{ para todo } i \in S.$$

2.2 Dinâmica do Algoritmo

A formulação do algoritmo implica a escolha de uma matriz estocástica $Q = (Q_{(i,j)})$ e funções chamadas probabilidades de aceitação a_n . A matriz Q gera candidatos a solução do problema $\min \{ f(i) / i \in S \}$

Se por exemplo, considerarmos $Q = \begin{pmatrix} 1/n & 1/n & \dots & 1/n \\ \vdots & & & \\ \vdots & & & \\ 1/n & .. & .. & 1/n \end{pmatrix}$ então todos os pontos

de $S = \{ 1, 2, \dots, n \}$ possuem a mesma chance de ser escolhidos. Esta escolha de Q é razoável se não temos informações adicional sobre o conjunto $S_m = \{ i \in S / i \text{ ponto de mínimo} \}$.

A Dinâmica do Algoritmo é:

Etapa 1. Escolha-se uma solução aleatória $x_1 \in S$.

Etapa $K+1$. Se x_k é a solução da etapa k , gera-se um novo candidato Y_{K+1} com probabilidade

$$Q_{ij} = P(Y_{k+1} = j / X_k = i)$$

e toma-se

$$X_{k+1} = \begin{cases} Y_{k+1} & \text{com probabilidade } a_{k+1}(Y_{k+1} / X_k) \\ X_k & \text{com probabilidade } 1 - a_{k+1}(Y_{k+1} / X_k) \end{cases} .$$

Note que a seqüência de soluções X_1, X_2, \dots geradas pelo algoritmo é uma cadeia de Markov uma vez que X_{k+1} é determinado pelo conhecimento de X_k, Y_{k+1} , não sendo necessário o conhecimento dos valores anteriores X_1, X_2, \dots, X_{k-1} .

As matrizes estocásticas associadas $P_1, P_2, \dots, P_n, \dots$ são dadas por:

$$P_n(i,j) = \begin{cases} Q_{ij} \cdot a_n(j/i) & \text{se } i \neq j \\ 1 - \sum_{k \neq i} P_{ik}(n) & \text{se } i = j \end{cases} .$$

observe que as matrizes $P_1, P_2, \dots, P_n, \dots$ são distintas, pois as probabilidades de aceitação mudam a cada iteração do algoritmo.

2.3 Convergência do Algoritmo

Para estabelecer a convergência utilizaremos os resultados assintóticos sobre produto de matrizes estocásticas do capítulo 1.

Suponhamos as seguintes hipóteses sobre a matriz Q e sobre as probabilidades de aceitação:

i) $Q = (Q_{(i,j)})$ tal que $Q_{(i,j)} > 0$ para todo i, j em S

ii)
$$\lim_{k \rightarrow \infty} a_k(j/i) = \begin{cases} 1 & \text{se } f(j) \leq f(i) \\ 0 & \text{se } f(j) > f(i) \end{cases} .$$

a) Vejamos que a seqüência de matrizes $P_1, P_2, \dots, P_n, \dots$ associadas ao algoritmo, convergem para uma matriz ergódica.

Por hipóteses a seqüência na é convergente, o que implica a existência do limite:

$$\lim_{n \rightarrow \infty} P_n^{(i,j)} = \lim_{n \rightarrow \infty} \begin{cases} Q_{ij} a_n(j/i) & i \neq j \\ 1 - \sum_{k \neq i} P_n^{(i,k)} & i = j \end{cases} = P_{(i,j)}$$

Mostremos a ergodicidade da matriz $P = (P_{(i,j)})$ utilizando o lema 1.8.

Segue-se diretamente das hipóteses i) e ii) que:

$$P_{(i,j)} = 0 \text{ se } i \in S_m \text{ e } j \notin S_m, \text{ onde } S_m = \{ l \in S / f(l) = f_{\min} \} \text{ e}$$

$$f_{\min} = \min\{f(l) / l \in S\}$$

$$P_{(i,j)} = Q_{ij} > 0 \text{ se } j \in S_m \text{ e } j \neq i$$

$$P_{(i,i)} = 1 - \sum_{l \neq i} Q_{il} = Q_{ii} > 0 \text{ se } i \in S_m.$$

Temos assim as hipóteses do lema 1.8, o que mostra que a matriz $P = (P_{(i,j)})$ é ergódica e:

$$\lim_{n \rightarrow \infty} P_n^{(i,j)} = \pi_j > 0 \text{ para todo } j \in S_m \text{ e todo } i \in S$$

$$\lim_{n \rightarrow \infty} P_n^{(i,j)} = 0 \text{ para todo } j \notin S_m.$$

b) O Teorema 1.9 garante a convergência do produto de matrizes $P_{(1)}P_{(2)}\dots P_{(n)}$, mais ainda:

$$\lim_{n \rightarrow \infty} (P_{(1)}P_{(2)}\dots P_{(n)})_{(i,j)} = \begin{cases} \pi_j > 0 & \text{se } j \in S_m \\ 0 & \text{se } j \notin S_m \end{cases}$$

Ou seja, temos a convergência da seqüência x_1, x_2, \dots para S_m . Também segue-se a independência da primeira escolha $x_1 = i$.

observe que as matrizes $P_{(1)}, P_{(2)}, \dots$ são distintas, pois as probabilidades de aceitação mudam a cada iteração do algoritmo.

CONCLUSÕES E CONSIDERAÇÕES FINAIS

Neste trabalho foi analisada a matemática do algoritmo, isto é, a fundamentação teórica que garante a convergência e independência assintótica do algoritmo. A convergência do algoritmo foi estudada utilizando a teoria das Cadeias de Markov não – homogêneas.

Como resultado desta pesquisa destacamos,¹ Foi mostrada com detalhe a convergência e independência assintótica do algoritmo Simulated Annealing, ² Alguns resultados teóricos sobre produtos de matrizes estocásticas foram reformulados para facilitar sua utilização no estudo do algoritmo, entre os quais destacamos um critério para testar a irredutibilidade de uma matriz (Proposição 1.3), também foi reformulado um resultado que garante a ergodicidade de uma matriz estocástica (Lema 1.8)

A teoria de Cadeias de Markov não homogêneas mostrou ser um instrumento adequado ao estudo assintótico do algoritmo. Acreditamos que adaptações ou refinamentos desta teoria poderão contribuir na abordagem de questões ainda não fechadas, como a rapidez de convergência deste popular algoritmo.

Referências Bibliográficas

- Anily, S e Federgruen, A.**, *Ergodicity in parametric non-stationary Markov chains: Application to simulated annealing methods*, Operat. Res. (1985).
- Anily, S e Federgruen, A.**, *Ergodicity in parametric nonstationary Markov chains: na application to simulated annealing methods*, Operat. Res., vol. 35, No. 6 (1987), 867-874.
- Bélisle, C.I.P.**, *Convergence theorems for a classe of simulated annealing algorithms on R*, J. Appl. Prob. 29, (1992), 885-895.
- Connors, D. P. e Kumar, P. R.**, *Simulated annealing type Markov chains e their order balance equations*, SIAM J. Control e Optimization, Vol. 27, No.6, (1989), 1440-1461.
- Cruz, Juan Alberto Rojas; Dorea, Chang Chung Yu.** Simple Conditions for the Convergence of Simulated Annealing Type Algorithms. Journal of Applied Probability, v. 35, n. 4, p. 885-892, (1998).
- Cruz, Juan Alberto Rojas.** *Convergência de cadeias de Markov Não-Homogêneas: Ergodicidade Fraca e Forte*, Tese de Doutorado, departamento de Matemática, Universidade de Brasília, Brasil. (1998)

- Cruz, Juan Alberto Rojas; DOREA, Chang Chung; R, Baigorri A.** *Asymptotic Behavior of Markovian Algorithms with General State Space.* Stochastic Modelling And Applications, Índia, v. 2, n. 2, p. 1-12, (1999).
- Cruz, Juan Alberto Rojas; DOREA, Chang Chung.** *Apprximation Results for Non-Homogeneous MarkovChains and Some Applications.* SANKHYÃ, The Indian Journal of Statistics, Índia, (2004 - 2° semestre).
- Dorea, Chang Chung Yu.** *On the efficiency of a continuous version of the simulated annealing algorithm,* Stat. And Probability Letters, 31 (1997a).
- Dorea, Chang Chung Yu.** *Stationary distribution of Markov chains in R with applications to global random aptimization,* Bernoulli, vol. 3, No. 4, 415-427, (1997b).
- Dekkers, Anton e Aarts, Emile.** *Global Optimization and simulated annealing,* Mathematical Programming, vol. 50, 367-393, (1991).
- Gidas, B.,** *Nonstationary Markov chains and convergence f the annealing algorithm,* Journal of Statistical Physicis, Vol. 39 (1985), 73-131.
- Isaacson, D and Madsen, D.,** *Markov chains,* Wiley, new York, (1975).
- Kirkpatrick, S. Gelatt Jr., C.D. and Vecchi, M. P.** *Optimization by simulated annealing,* Science 220, 671-680. (1983).