

## 4 Fundamentação Teórica

### 4.1 Geoestatística

A Geoestatística é um ramo da Estatística Aplicada e da Matemática que nasceu da necessidade de se ter uma ferramenta, mais poderosa que a Estatística Clássica, que permitisse a interpretação e a manipulação mais adequadas de dados provenientes de medições efetuadas na Terra. Segundo [43], dados dessa natureza apresentam uma correlação espacial, que acontece em maior ou menor grau, dependendo da distância que os separa, ou seja, a similaridade entre duas medições será tão maior quanto menor for a distância entre os pontos de coleta. Esse comportamento é típico de fenômenos ditos regionais e são representados pelas variáveis regionalizadas que são o objeto de estudo da Geoestatística. Essa variabilidade leva em consideração tanto a heterogeneidade quanto a direcionalidade em conjuntos de dados.

A utilidade da Geoestatística fica evidente com a ilustração da Figura 4.1. Apesar da imagem à esquerda ter uma aparência quase aleatória, ainda assim é possível observar um alinhamento preferencial de noroeste para sudeste. No caso da imagem à direita, há um alto grau de continuidade espacial e de anisotropia que mantêm o alinhamento de noroeste para sudeste. Visualmente as imagens são bastante diferentes, contudo seus valores de média e de variância são exatamente os mesmos, o que expõe a limitação da Estatística Clássica ao representar esses dados.

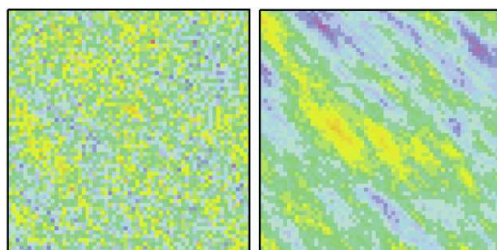


Figura 4.1: Distribuições de pontos diferentes com os mesmos valores de média e de variância (*Fonte: Chambers, 2000*)

As origens da Geoestatística remontam ao início dos anos 50, quando D. G. Krige e H. S. Sichel desenvolveram um método empírico para estimar reservas minerais ([44]). Porém, esse método só recebeu um tratamento formal no início dos anos 60, quando G. Matheron elaborou a sua *Teoria das Variáveis Regionalizadas* ([45]). Na área de petróleo, a Geoestatística só foi introduzida em meados dos anos 70, época em que se espalhou também por várias outras áreas devido, principalmente, ao advento dos computadores de alta velocidade.

O processo de caracterização de um reservatório envolve a construção de uma série de mapas que são modelos numéricos da distribuição espacial de propriedades do reservatório como, por exemplo, porosidade e permeabilidade. A grande dificuldade por trás da construção desses mapas é a escassez de dados reais medidos diretamente no campo, pois, é a partir desses poucos dados que se faz a interpolação de todos os pontos não amostrados do reservatório. A qualidade dessa caracterização é fundamental para se obter previsões mais precisas a respeito do comportamento do reservatório, por isso, os mapas que compõem o modelo de um reservatório devem representar as suas heterogeneidades com a maior fidelidade possível. Para atingir esse objetivo, a Geoestatística tem sido uma ferramenta amplamente aceita para a construção desses mapas, especialmente quando incorpora dados secundários como, por exemplo, dados provenientes de análise sísmica em 3D e, mais recentemente, em 4D ([46]).

#### 4.1.1

##### Variograma e Krigagem

O interesse por trás da análise de continuidade espacial é quantificar a variabilidade de um conjunto de amostras de um atributo que estão separadas por uma determinada distância  $h$  e dispostas em uma determinada direção. A forma mais comum de analisar a continuidade espacial de um conjunto de dados é através do cálculo do variograma. O variograma é uma medida de dissimilaridade entre amostras espacialmente distribuídas e, por isso, tende a aumentar à medida que aumenta a distância entre as amostras. Esse aumento ocorre até que um limite, denominado patamar, seja atingido. A partir daí a correlação entre as amostras praticamente deixa de existir. Segundo [47], o variograma mede o grau de dependência espacial entre um conjunto de amostras e, para a sua construção, é calculada a soma dos quadrados das diferenças das amostras, dada uma distância de separação e uma direção, conforme ilustra a Equação 4-1.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (x_{i+h} - x_i)^2 \quad (4-1)$$

Onde:

$h$  é a distância de separação dos pontos;

$\gamma(h)$  é o variograma dos pontos separados pela distância  $h$ ;

$N(h)$  é o número de pontos separados pela distância  $h$ ;

$x_{i+h}$  é o valor da amostra distante  $h$  unidades da amostra  $i$ ;

$x_i$  é o valor da amostra  $i$ .

Visto que tanto a direção quanto a distância de separação entre as amostras influenciam o cálculo do variograma, uma análise criteriosa da continuidade espacial implica em calcular o variograma para várias distâncias e direções diferentes. Com isso, é possível identificar:

- o tamanho da zona de influência em torno de uma amostra, pois toda amostra cuja zona de influência é maior ou igual à distância entre a amostra e um ponto a ser estimado fornece informações sobre aquele ponto;
- a anisotropia, caso os variogramas apresentem comportamentos diferentes para direções de linhas de busca diferentes;
- e a continuidade, de acordo com a forma do variograma.

Para estimar valores não amostrados de uma variável regionalizada é necessário conhecer a função variograma para todos os azimutes e distâncias possíveis. Isso exige que o variograma calculado experimentalmente seja ajustado a um modelo teórico para que a variância do erro seja conhecida em qualquer localização. Há diversos modelos teóricos possíveis, mas os mais comuns são os modelos esférico, exponencial, linear (sem patamar) e gaussiano. Segundo [47], a escolha do modelo teórico mais apropriado, geralmente, é feita depois de várias tentativas e depende da experiência do profissional. A validação cruzada é uma ferramenta que pode ser utilizada para avaliar a escolha adotada, apesar de ela não provar que o modelo escolhido é o mais correto e sim que ele não é totalmente incorreto.

Depois de estabelecido o modelo teórico do variograma, a estimação dos valores não amostrados é realizada através da *krigagem*. Em [47], a *krigagem* é definida como uma técnica determinística de interpolação que estima valores de variáveis distribuídas no espaço a partir de valores adjacentes que são

considerados interdependentes pelo variograma. A *krigagem* usa as informações do variograma para encontrar os pesos ótimos que são associados às amostras utilizadas para estimar um ponto. A estimativa de um ponto por *krigagem* simples é dada pela Equação 4-2.

$$x_0 = m_x + \sum_{i=1}^n w_i \cdot (x_i - m_x) \quad (4-2)$$

Onde:

$i = 0$  é o índice do ponto a ser estimado;

$x_0$  é o valor estimado do ponto;

$m_x$  é a média constante da variável  $x$ ;

$w_i$  é o peso da *krigagem* da amostra  $i$ ;

$x_i$  é o valor da amostra  $i$ ;

$n$  é o número de pontos.

Segundo [46], a *krigagem* simples parte do princípio de que a média das amostras ( $m_x$ ) é constante e este valor de média deve ser inicialmente fixado. Isso significa que as médias locais, ou seja, as médias das amostras em torno dos pontos a serem estimados, são consideradas relativamente constantes e de valor muito semelhante à média de todas as amostras.

Os pesos  $w_i$  ótimos associados às amostras envolvidas na estimativa de um ponto são aqueles que minimizam a variância do erro de estimação. Isso corresponde a minimizar uma função de erro quadrático parametrizada em termos da covariância espacial da variável  $x$ . Na prática, a obtenção desses pesos ótimos corresponde à solução de um sistema de equações semelhante ao apresentado na Equação 4-3. Segundo [46], os elementos da diagonal principal da matriz são todos iguais à variância da variável regionalizada  $x$ . O vetor à direita do sinal de igualdade é obtido a partir do cálculo das covariâncias espaciais entre as amostras disponíveis e o ponto a ser estimado. A matriz de *krigagem* captura a informação sobre a correlação das amostras enquanto o vetor de *krigagem* apresenta a informação sobre a correlação espacial entre a localização do ponto a ser estimado e as localizações das amostras.

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} C_{01} \\ C_{02} \\ \vdots \\ C_{0n} \end{bmatrix} \quad (4-3)$$

Onde:

$i = 0$  é o índice do ponto a ser estimado;

$C_{ij}$  é a covariância espacial estimada entre os pontos  $i$  e  $j$ ;

$w_i$  é o peso associado ao ponto  $i$ .

É importante observar que o sistema de *krigagem* depende apenas da configuração das amostras em relação à localização do ponto a ser estimado e não propriamente dos valores das amostras. Uma vez calculados, esses pesos são usados para estimar o valor de um ponto, através da Equação 4-2, em uma situação semelhante à apresentada na Figura 4.2. A Figura 4.3 ilustra um exemplo de um mapa completo construído por meio de *krigagem*, onde os pontos em destaque representam as amostras.

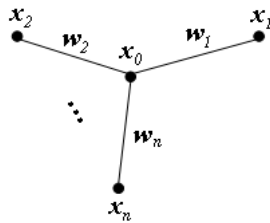


Figura 4.2: Estimação do ponto  $x_0$  por *krigagem*

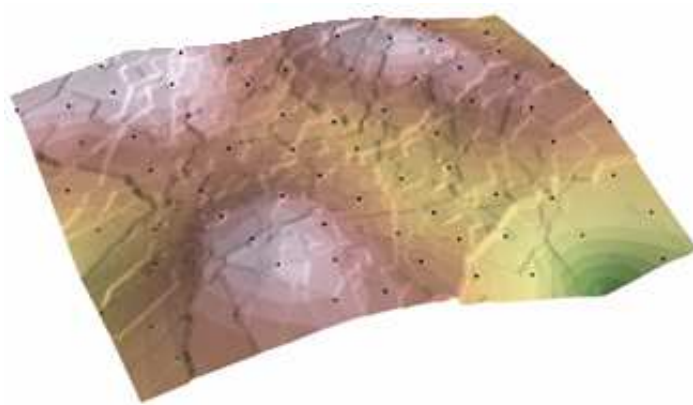


Figura 4.3: Mapa construído através de *krigagem*

De forma semelhante à *krigagem* simples, a interpolação de pontos pode ser feita também por *krigagem* ordinária. No caso da *krigagem* ordinária, a estimativa de um ponto também é dada pela Equação 4-2. Contudo, o cálculo da média  $m_x$  considera que as médias locais não estão necessariamente próximas da média de todas as amostras. Por isso, nesse caso, a cada nova

estimativa o valor de  $m_x$  deve ser calculado a partir dos valores das amostras que estão na vizinhança do ponto a ser estimado.

Os métodos de *krigagem* simples e ordinária, porém, não levam em consideração as relações existentes entre a propriedade em estudo e as outras propriedades. Para considerar essas relações pode-se usar a *cokrigagem*, uma extensão multivariada da *krigagem* cujo objetivo é estimar o valor de uma propriedade em um local, combinando as medidas da mesma propriedade em outras localizações, com medidas de outra variável que apresente uma correlação cruzada com a primeira. Além dessas, há outras formas de *krigagem* cujos detalhes podem ser encontrados em [48] e [49].

#### 4.1.2

##### Simulação Sequencial

Segundo [50], as propriedades de interesse para a caracterização de um reservatório como, por exemplo, porosidade e permeabilidade, são o resultado de processos físicos e químicos altamente complexos. Esses processos estabelecem um padrão espacial para essas propriedades e é importante entender suas escalas e seus aspectos direcionais, a fim de criar modelos realistas que colaborem para aumentar a eficiência na produção de hidrocarbonetos. Além da variabilidade espacial, que aumenta a complexidade dessas propriedades, a escassez de dados reais também contribui para dificultar ainda mais a tarefa de estimar os valores dessas propriedades em outros pontos do reservatório. Devido a isso, é inevitável a consideração de existência de incerteza na estimação desses pontos.

Como visto na Seção 4.1.1, a *krigagem* é um método determinístico de interpolação que retorna apenas uma solução que é a melhor estimativa, ou seja, aquela que minimiza a variância do erro de estimação. Portanto, o mapa construído com a *krigagem* é uma estimativa suavizada. Porém, obter a melhor estimativa pode não ser o objetivo do especialista, uma vez que este pode estar interessado em saber mais a respeito da variabilidade espacial da variável regionalizada. Para estudar a variabilidade espacial utiliza-se a simulação sequencial que, segundo [47], permite a geração de uma infinidade de realizações diferentes de mapas onde cada um tem aproximadamente o mesmo variograma e a mesma variância dos dados originais. Teoricamente a média de uma grande quantidade de mapas gerados por simulação fornece resultados mais realistas e, portanto, mais confiáveis para prever o comportamento do reservatório. Além disso, a variação identificada entre as realizações representa uma medida de incerteza associada às estimativas.

A simulação sequencial é um processo iterativo que consiste em, inicial-

mente, estabelecer um caminho aleatório de visitação dos pontos não amostrados de um mapa. Em seguida, a cada iteração, visita-se um ponto do caminho, estima-se a distribuição probabilística condicional local e sorteia-se um valor aleatório para o ponto a partir da distribuição probabilística estimada. Para estimar a distribuição de probabilidade em torno de um ponto leva-se em consideração tanto os dados reais disponíveis no mapa, quanto os pontos estimados em iterações anteriores. Vale lembrar que, assim como acontece na interpolação determinística, é possível também incorporar dados secundários ao processo de simulação. Para gerar as diversas realizações de um mapa, basta repetir o processo de simulação com um caminho aleatório diferente. O algoritmo de simulação sequencial está ilustrado na Figura 4.4.

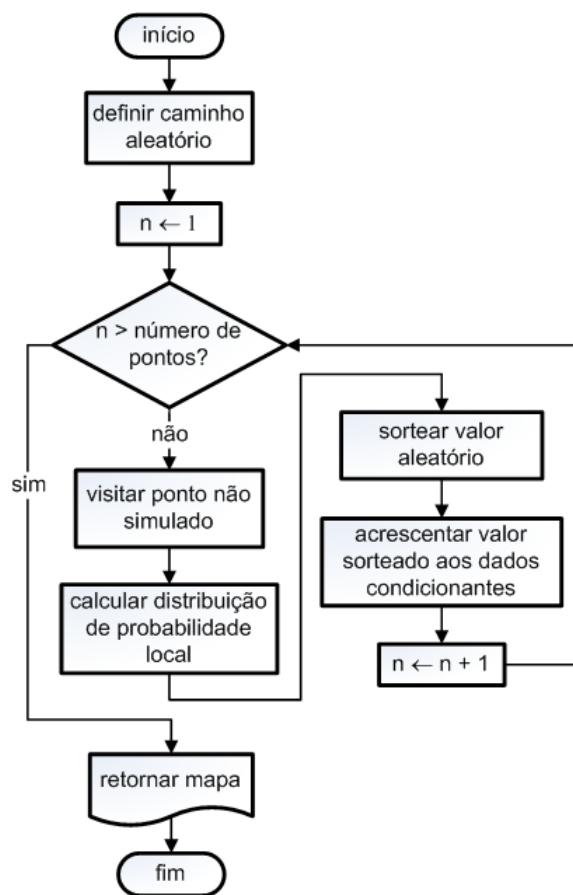


Figura 4.4: Fluxograma da simulação sequencial

Há pelo menos dois tipos de simulação sequencial, a simulação sequencial indicadora (*SIS*) e a simulação sequencial gaussiana (*SGS*). A *SIS* é utilizada na simulação de propriedades categóricas como, por exemplo, as litofácies arenito e folhelho. Já a *SGS* é utilizada na simulação de propriedades contínuas como, por exemplo, as propriedades petrofísicas permeabilidade e porosidade.

A diferença entre esses dois tipos de simulação está justamente na forma como a distribuição probabilística condicional local é estimada.

Na *SIS*, a distribuição é calculada por meio da *krigagem* indicadora que, para o caso mais comum, que envolve apenas duas categorias, é determinada segundo as Equações 4-4 e 4-5.

$$P_{c_1}(i) = \pi_{c_1} + \sum_{j=1}^{i-1} w_j \cdot (x_j - \pi_{c_1}) \quad (4-4)$$

$$P_{c_2}(i) = 1 - P_{c_1}(i) \quad (4-5)$$

Onde:

$i$  é o índice do ponto a ser estimado;

$j$  é o índice de um ponto amostrado ou estimado anteriormente;

$P_{c_1}$  é a probabilidade de ocorrência da categoria  $c_1$ ;

$P_{c_2}$  é a probabilidade de ocorrência da categoria  $c_2$ ;

$\pi_{c_1}$  é o valor esperado da categoria  $c_1$ ;

$w_j$  é o peso da *krigagem* do ponto  $j$  ;

$x_j$  é o valor do ponto  $j$ .

A *krigagem* indicadora aproxima a distribuição condicional. Logo, a probabilidade de ocorrência da categoria  $c_1$  pode ser reescrita conforme a Equação 4-6.

$$P_{c_1}(i) = P(x_i = 1 | x_1, \dots, x_{i-1}) \quad (4-6)$$

Na *SGS* a *krigagem* é utilizada para calcular os parâmetros de média e de variância da distribuição gaussiana que é utilizada para o sorteio aleatório dos pontos. Os valores de média e de variância são determinados segundo as Equações 4-7 e 4-8.

$$x_i = m_x + \sum_{j=1}^{i-1} w_j \cdot (x_j - m_x) \quad (4-7)$$



$$\sigma_i^2 = \sigma_x^2 - \sum_{j=1}^{i-1} w_j \cdot C_{ij} \quad (4-8)$$

Onde:

- $i$  é o índice do ponto a ser estimado;
- $j$  é o índice de um ponto amostrado ou estimado anteriormente;
- $x_i$  é a média da distribuição normal local no ponto  $i$ ;
- $m_x$  é a média constante da variável  $x$ ;
- $w_j$  é o peso da *krigagem* do ponto  $j$ ;
- $\sigma_i$  é a variância da distribuição normal local no ponto  $i$ ;
- $\sigma_x$  é a variância constante da variável  $x$ ;
- $C_{ij}$  é a covariância espacial estimada entre os pontos  $i$  e  $j$ .

### 4.1.3 Geoestatística de Múltiplos Pontos

Segundo [51], os métodos geoestatísticos convencionais, baseados no cálculo do variograma, não são os mais adequados para a construção de modelos onde os objetos ou estruturas a serem representados são curvilíneos. Exemplos bastante comuns são os canais sinuosos de permeabilidade em um reservatório de petróleo. Para a criação de modelos dessa natureza, que sejam realmente representativos, é necessário lançar mão de métodos estatísticos que considerem múltiplos pontos de uma só vez. Enquanto o variograma se limita a calcular a variabilidade espacial entre pares de pontos, a Geoestatística de Múltiplos Pontos, como o próprio nome indica, calcula essa variabilidade entre vários pontos distribuídos no espaço. Com isso, torna-se possível modelar com maior fidelidade as não linearidades próprias de determinadas propriedades de um reservatório.

Porém, durante a caracterização de um reservatório, os dados reais disponíveis, ou seja, os dados vindos de medições realizadas no campo, raramente são suficientes para permitir a inferência das estatísticas de múltiplos pontos. Para contornar essa limitação foi proposto em [52] um algoritmo de simulação sequencial por indicadores que infere as estatísticas de múltiplos pontos a partir de imagens de treinamento e, em seguida, utiliza essas estatísticas para a geração do modelo geoestatístico numérico.

Uma imagem de treinamento é uma representação puramente conceitual dos padrões esperados das heterogeneidades geológicas presentes no modelo

do reservatório. Na construção de um modelo geoestatístico, mais de uma imagem de treinamento pode ser usada, tanto para representar as diferentes escalas de variabilidade, quanto os estilos das heterogeneidades. Segundo [53], fotografias interpretadas, dados de sensoriamento remoto, ou ainda esboços desenhados à mão podem ser usados para construir imagens de treinamento. A inferência das estatísticas a partir da imagem de treinamento elimina a necessidade de calcular o variograma para identificar a variabilidade espacial dos dados e também a necessidade da *krigagem* para derivar as probabilidades condicionais. A Figura 4.5 apresenta dois exemplos possíveis de imagens de treinamento. A primeira representa uma propriedade categórica que pode assumir três valores distintos: branco, preto ou cinza. A segunda representa uma propriedade contínua que pode assumir qualquer valor entre branco e preto, segundo uma escala de cinza.

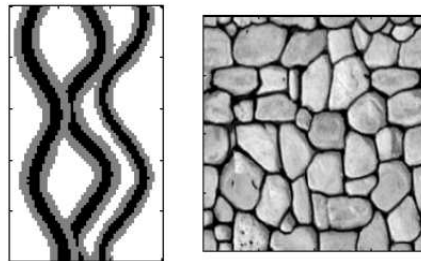


Figura 4.5: Imagens de treinamento para propriedades categórica e contínua  
(Fonte: Switzer, 2008)

Segundo [51], o algoritmo proposto originalmente em [52] apresenta um custo computacional extremamente elevado, pois, durante a simulação, para cada ponto não amostrado, é necessário percorrer toda a imagem de treinamento para calcular a probabilidade condicional associada ao ponto. A fim de melhorar sua eficiência, foi proposto em [51] o algoritmo de simulação sequencial *SNESIM*, que é uma extensão do algoritmo original em que a imagem de treinamento é percorrida apenas uma vez no processo de simulação de pontos. Em [54] são apresentados os detalhes de utilização do algoritmo *SNESIM* e, além disso, são feitas análises de sensibilidade que avaliam os impactos causados pela variação de uma série de parâmetros que são considerados pelo algoritmo quando uma simulação é realizada.

Apesar do aumento de eficiência em relação ao algoritmo original, o *SNESIM* está limitado à simulação apenas de propriedades categóricas como, por exemplo, mapas de fácies, não se aplicando, portanto, à simulação de mapas de propriedades como permeabilidade e porosidade, que são constituídos de valores contínuos. Para superar a limitação imposta pelo *SNESIM*, foi proposto

em [55] e [56] o algoritmo de simulação sequencial *FILTERSIM* que, além de simular tanto propriedades categóricas quanto contínuas, apresenta também um baixo consumo de memória a um custo razoável de CPU quando comparado ao *SNESIM*. O *FILTERSIM* também extrai as estatísticas de múltiplos pontos a partir de imagens de treinamento, porém de uma forma diferente de como são extraídas no algoritmo *SNESIM*. O algoritmo *FILTERSIM*, que é parte do modelo de solução proposto no Capítulo 5, é apresentado em detalhes a seguir.

### Algoritmo FILTERSIM

O algoritmo *FILTERSIM* é dividido em duas etapas, uma de classificação dos padrões presentes na imagem de treinamento e uma de simulação da propriedade em questão. Na etapa de classificação, inicialmente o algoritmo recebe uma imagem de treinamento sobre a qual a classificação deve ser realizada. Vale ressaltar que essa imagem de treinamento deve representar adequadamente as heterogeneidades da propriedade em estudo, seja essa uma propriedade categórica ou contínua (Figura 4.5). O passo seguinte, ainda dentro da etapa de classificação, consiste em atribuir um conjunto de pontuações a cada ponto da imagem de treinamento de acordo com o padrão ao redor daquele ponto. Para isso, seja  $X(i, j)$  o valor na posição  $(i, j)$  da imagem de treinamento. A pontuação  $S_f(i, j)$ , onde  $f \in \{1, \dots, 6\}$ , para o padrão na vizinhança de  $(i, j)$  é definida pela aplicação de um filtro  $f(u, v)$ , segundo a Equação 4-9.

$$S_f(i, j) = \sum_{v=-n}^n \sum_{u=-n}^n X(i+u, j+v) \cdot f(u, v) \quad (4-9)$$

Onde:

a dimensão da vizinhança local é  $(2n + 1) \times (2n + 1)$ .

A representação gráfica dos filtros a serem aplicados à imagem de treinamento são apresentados na Figura 4.6. Nesse exemplo o tamanho dos filtros é de  $15 \times 15$  pixels ( $n = 7$ ), mas este valor pode variar.

As definições matemáticas dos filtros  $f(u, v)$  são dadas pelas Equações 4-10 a 4-15:

- $f_1$ : filtro de média Norte - Sul.

$$f_1(u, v) = 1 - \frac{|v|}{n} \quad (4-10)$$

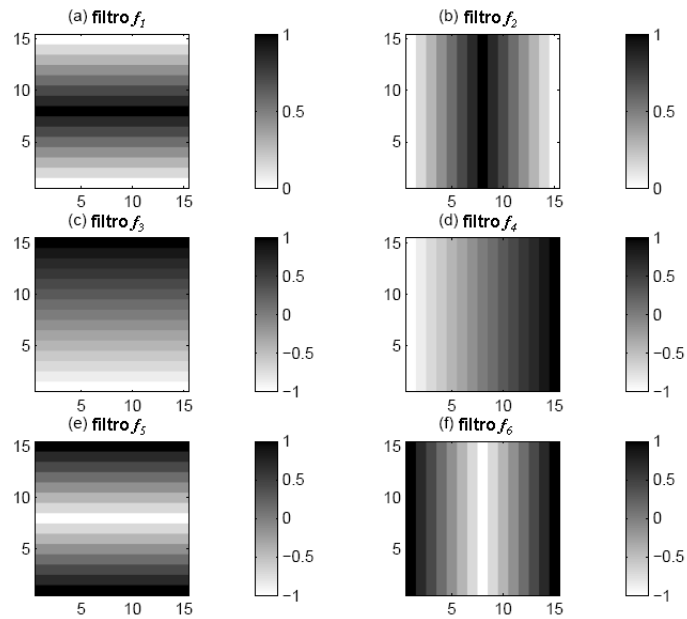


Figura 4.6: Filtros a serem aplicados à imagem de treinamento (*Adaptado de Switzer, 2008*)

Onde:

$$v = -n, \dots, n.$$

- $f_2$ : filtro de média Leste - Oeste, que é obtido mediante uma rotação de  $90^\circ$  em  $f_1$ .

$$f_2(u, v) = 1 - \frac{|u|}{n} \tag{4-11}$$

- $f_3$ : filtro de gradiente Norte - Sul.

$$f_3(u, v) = \frac{v}{n} \tag{4-12}$$

- $f_4$ : filtro de gradiente Leste - Oeste, que é obtido mediante uma rotação de  $90^\circ$  em  $f_3$ .

$$f_4(u, v) = \frac{u}{n} \tag{4-13}$$

- $f_5$ : filtro de curvatura Norte - Sul.

$$f_5(u, v) = \frac{2|v|}{n} - 1 \quad (4-14)$$

- $f_6$ : filtro de curvatura Leste - Oeste, que é obtido mediante uma rotação de  $90^\circ$  em  $f_5$ .

$$f_6(u, v) = \frac{2|u|}{n} - 1 \quad (4-15)$$

O cálculo da pontuação em um dado ponto da imagem é exemplificado na Figura 4.7, onde o filtro  $f_1$  é aplicado a um padrão específico da imagem de treinamento.

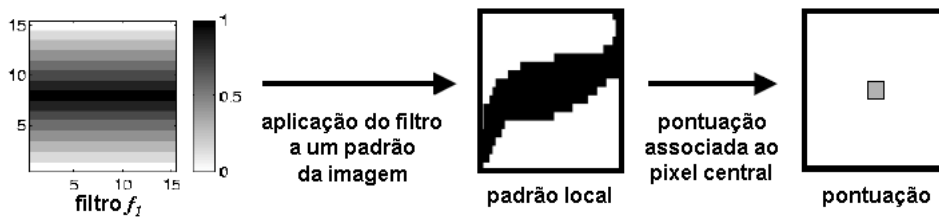


Figura 4.7: Aplicação do filtro à imagem de treinamento (*Adaptado de Switzer, 2008*)

Após a aplicação do conjunto de filtros a todos os pontos da imagem de treinamento, classificam-se todos os padrões identificados na imagem de acordo com os 6 valores de pontuação atribuídos ao seu ponto central. Para isso, calcula-se a distribuição de frequência dos valores de pontuação de cada filtro. Em seguida divide-se cada distribuição de frequência em  $n$  partes de forma que todas elas tenham a mesma frequência. Em [53] sugere-se a divisão em 5 partes, conforme ilustra a Figura 4.8. Assim, o espaço hexadimensional de pontuações é particionado em  $5^6$  células.

Uma vez estabelecidas as células, associa-se cada padrão de treinamento à célula que corresponde à combinação das 6 pontuações atribuídas ao padrão. Apesar da grande quantidade de células ( $5^6$ ), na prática muitas dessas células não terão padrões associados devido à inexistência de padrões que apresentem as combinações de pontuações correspondentes. As células que possuem padrões associados representam as classes e, para cada classe, calcula-se um protótipo que é dado pela média de todos os padrões presentes na classe. Na

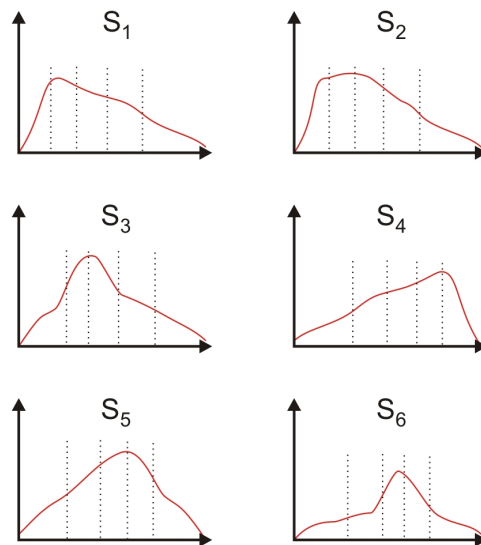


Figura 4.8: Particionamento das distribuições de frequência

Figura 4.9 é apresentado um exemplo de um conjunto de padrões pertencentes a uma mesma classe e o seu protótipo correspondente. Na Figura 4.10 é apresentado um fluxograma que resume os passos da etapa de classificação do algoritmo *FILTERSIM*.

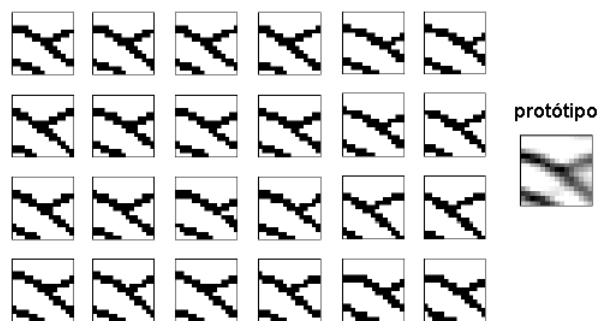


Figura 4.9: Padrões pertencentes a uma classe e seu protótipo correspondente (Adaptado de Switzer, 2008)

Uma vez finalizada a etapa de classificação dos padrões da imagem de treinamento, inicia-se a etapa de simulação dos valores da propriedade em estudo. A etapa de simulação segue os princípios da simulação sequencial e, por isso, inicialmente é preciso estabelecer um caminho aleatório para a visitação de todos os blocos do modelo do reservatório onde o valor da propriedade deve ser estimado.

Para determinar o valor da propriedade em um bloco é necessário, primeiramente, verificar se existem dados condicionantes na região em torno desse bloco. A região a ser verificada tem uma quantidade de blocos igual à quantidade de pontos dos padrões que foram inicialmente classificados,

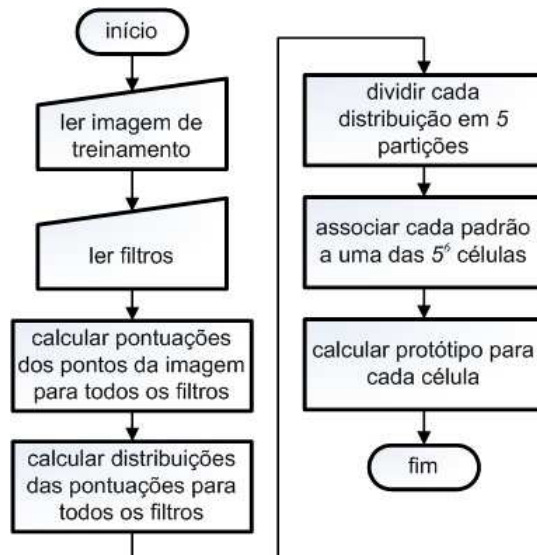


Figura 4.10: Fluxograma da etapa de classificação do algoritmo *FILTERSIM*

sendo o bloco a ser simulado o centro dessa região. Caso não existam dados condicionantes em torno do bloco a ser simulado, seleciona-se um padrão aleatório, pertencente a um protótipo aleatório, e substitui-se a região do modelo pelo padrão selecionado. Caso existam dados condicionantes em torno do bloco, seleciona-se o protótipo mais próximo ao padrão em torno do bloco segundo a função de distância dada pela Equação 4-16:

$$d(R, P) = \sum_{k=1}^3 \omega_k \frac{\sum_{i_k=1}^{n_k} |x_k(i_k) - y_k(i_k)|}{n_k} \quad (4-16)$$

Onde:

- $R$  é a região em torno do bloco a ser simulado;
- $P$  é o protótipo;
- $k$  é o tipo de dado condicionante;
- $\omega$  é o peso atribuído a cada tipo de dado;
- $i$  é a localização do ponto na imagem;
- $x$  é o valor da propriedade no padrão;
- $y$  é o valor da propriedade no protótipo;
- $n$  é o número de pontos.

Os dados condicionantes podem ser de 3 tipos:

- valores rígidos, provenientes de medições feitas no próprio reservatório;

- valores estimados em iterações anteriores durante o processo de simulação;
- valores provenientes de partes de padrões associados aos blocos simulados anteriormente.

Uma vez identificado o protótipo mais próximo do padrão em torno do bloco, seleciona-se aleatoriamente um padrão pertencente à classe representada pelo protótipo e substitui-se a região do modelo pelo padrão selecionado. Esse procedimento é repetido até que todos os blocos do reservatório sejam simulados.

A distância calculada segundo a Equação 4-16 é mais adequada para os casos em que o modelo a ser simulado tem apenas duas dimensões. Para modelos com três dimensões o custo computacional aumenta consideravelmente, principalmente em casos em que a imagem de treinamento é mais complexa. Essa limitação é contornada em [57], onde é proposta uma nova maneira de calcular essa distância, que reduz significativamente o seu custo computacional. Na Figura 4.11 é apresentado um fluxograma que resume os passos da etapa de simulação do algoritmo *FILTERSIM*.

## 4.2 Inteligência Computacional

Desde a invenção dos primeiros computadores eletrônicos digitais, a quantidade e a complexidade das tarefas atribuídas a essas máquinas vêm aumentando consideravelmente. O que antes era restrito à geração de tabelas de cálculo de trajetórias balísticas, como foi o caso do computador ENIAC durante a 2ª Guerra Mundial, evoluiu para a execução de tarefas bem mais ambiciosas. Técnicas computacionais essencialmente exatas e determinísticas passaram a dividir espaço com técnicas aproximadas e não determinísticas, que estão bem mais próximas da forma como o ser humano e a própria natureza se comportam. Técnicas desse tipo, que se inspiram em comportamentos observados na natureza, são comumente denominadas “inteligentes” e constituem um ramo da Ciência da Computação conhecido como Inteligência Computacional. Elas são empregadas, principalmente, na solução de problemas para os quais não existem procedimentos efetivos capazes de solucioná-los satisfatoriamente. Problemas assim, geralmente podem ser modelados como tarefas de aprendizagem, percepção, previsão, adaptação ou evolução, e estas são as principais características presentes nas técnicas de Inteligência Computacional. Atualmente a Inteligência Computacional compreende um grande conjunto de técnicas, mas as de maior destaque são as Redes Neurais Artificiais ([58], [59], [60]), a Lógica Nebulosa ([61], [62]) e os Algoritmos Genéticos ([63], [64]).



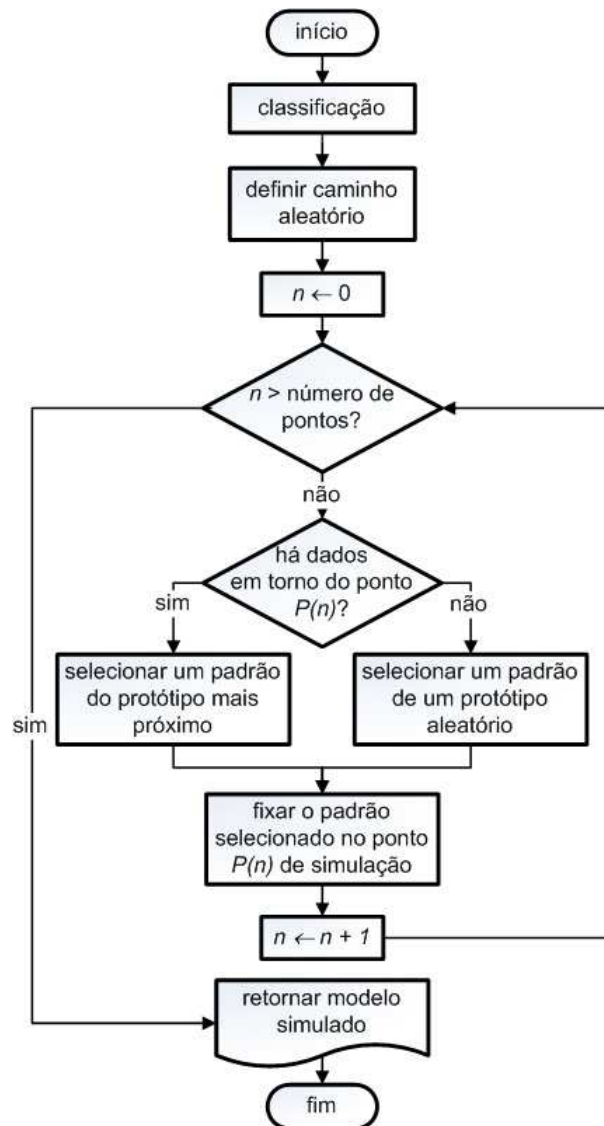


Figura 4.11: Fluxograma da etapa de simulação do algoritmo *FILTERSIM*

#### 4.2.1 Redes Neurais

Uma rede neural artificial, ou simplesmente rede neural, é um modelo matemático não-linear, inspirado na estrutura e operação do cérebro humano, usado para encontrar relacionamentos complexos entre conjuntos de dados de entrada e de saída. Esse modelo é usado principalmente na solução de problemas que envolvem reconhecimento de padrões e aproximação de funções. Semelhantemente ao sistema nervoso biológico, uma rede neural também é formada por unidades elementares, denominadas neurônios artificiais, que efetuam operações muito simples. Esses neurônios encontram-se interconectados de forma que cada neurônio recebe e envia informações para neurônios vizinhos. A interconexão e o paralelismo dos neurônios tornam a rede neural eficaz

na tarefa de encontrar uma função que seja capaz de aproximar dados não lineares, incompletos e ruidosos. Esse potencial de aproximação é a principal vantagem das Redes Neurais perante outros métodos de interpolação.

### 4.2.2

#### Lógica Nebulosa

A Lógica Nebulosa, por sua vez, é uma técnica que se inspira nas imprecisões inerentes ao discurso em linguagem natural, para a modelagem de sistemas de maneira mais realista e flexível. Quando se diz, por exemplo, que “*o dia está quente*”, o significado do termo “*quente*” não é precisamente estabelecido. Entretanto, essa imprecisão não compromete o entendimento do conceito que se quer expressar. Apesar de os seres humanos lidarem com essas imprecisões de uma maneira bastante natural, a matemática tradicional, especificamente a teoria clássica de conjuntos, não é suficientemente adequada para modelar essas imprecisões. Uma vez que um conjunto clássico apresenta uma fronteira muito bem definida, é necessário estabelecer um limiar, a partir do qual um valor passa a pertencer ao conjunto das temperaturas quentes e abaixo do qual deixa de pertencer.

O senso comum, no entanto, sugere que a transição entre as temperaturas quentes e não quentes não acontece de forma tão abrupta, e sim de uma forma mais suave. Assim, foram propostos em [65] os conjuntos nebulosos, que permitem que situações como a que foi exemplificada sejam adequadamente representadas. A partir dos conjuntos nebulosos foi criado todo um arcabouço lógico ([66], [67], [68]) que permitiu o surgimento da Lógica Nebulosa. Através dessa nova lógica é possível construir sistemas baseados em regras, onde essas regras manipulam termos imprecisos, que podem ser aplicados na solução de problemas complexos que envolvem, principalmente, controle e aproximação de funções.

### 4.2.3

#### Algoritmos Genéticos

Os Algoritmos Genéticos, propostos em [69], pertencem ao grupo dos métodos evolutivos que são baseados em populações de potenciais soluções de um problema. Métodos dessa natureza se mostram interessantes na resolução de problemas complexos de otimização porque conseguem um equilíbrio entre a capacidade de exploração do espaço de soluções e também de aproveitamento das melhores soluções ao longo da evolução. Isso os torna menos suscetíveis ao aprisionamento em ótimos locais, porém o tempo de convergência costuma ser bem mais elevado que o tempo gasto por métodos de otimização baseados em

gradiente. O consumo de tempo está diretamente relacionado ao número de chamadas à função objetivo, que no caso de um algoritmo genético costuma ser bem maior que num método baseado em gradiente. Devido a essas características, os métodos evolutivos são mais adequados para a busca de uma solução aceitável, que não necessariamente seja uma solução ótima, para um problema de otimização.

A Computação Evolutiva é uma abordagem para a resolução de problemas inspirada, principalmente, no princípio da sobrevivência dos mais aptos, proposto pelo naturalista inglês Charles Darwin em sua *Teoria da Evolução das Espécies*. Segundo Darwin:

*“Quanto melhor um indivíduo se adaptar ao seu meio ambiente, maior será sua chance de sobreviver e gerar descendentes.”*

Os Algoritmos Genéticos são um dos principais modelos pertencentes à Computação Evolutiva e são tipicamente utilizados para resolver problemas da forma  $f : S \rightarrow \mathfrak{R}$ , onde  $S$  é um espaço de busca constituído por todas as possíveis soluções para um problema particular. Dependendo das peculiaridades do problema, as soluções podem ser representadas por vetores  $n$ -dimensionais de números binários, inteiros, reais, ou estruturas mais complexas. Para todas as soluções existentes no domínio de  $S$ , um número real é associado, medindo quão adequada é a solução para resolver o problema em questão. A tarefa principal de um algoritmo genético é buscar, de forma eficiente, em amostras do espaço de busca  $S$ , soluções que estejam de acordo com o objetivo do problema. É importante mencionar que essas soluções não precisam ser necessariamente ótimas, mas sim satisfatórias. Ao se lidar com espaços de busca muito grandes e complexos, o ótimo pode ser difícil de ser atingido e, nesse caso, pode-se apenas esperar achar uma solução satisfatória.

### **Etapas de um Algoritmo Genético Clássico**

Em um dado problema de otimização, uma representação adequada deve ser estabelecida para codificar possíveis soluções em estruturas de dados do tipo definido no conjunto  $S$ . Na implementação clássica, as soluções são codificadas por vetores de números binários. Logo, primeiramente é preciso encontrar uma forma adequada de representar as soluções do problema por meio de vetores binários. Após a escolha da representação, uma população de  $n$  potenciais soluções, também chamadas de indivíduos, é gerada aleatoriamente e, em seguida, os indivíduos são avaliados. A avaliação consiste em associar a cada indivíduo da população um número real, também chamado de aptidão, cuja finalidade é medir o quanto aquela solução é adequada para satisfazer às

especificações de um problema particular. O cálculo da avaliação se dá pela chamada da função objetivo, que recebe como entrada a solução representada pelo indivíduo.

Uma vez avaliada a população inicial do algoritmo genético, inicia-se o processo evolutivo em que os indivíduos são submetidos a quatro etapas básicas: seleção, cruzamento, mutação e novamente avaliação. Essas etapas se repetem por um determinado número de ciclos de evolução, também conhecidos por gerações, até que uma condição de parada seja satisfeita.

Na etapa de seleção os indivíduos passam por um procedimento que, baseado no princípio de Darwin, privilegia os mais aptos. Esse procedimento consiste em distribuir os indivíduos em uma roleta, de maneira que os mais aptos ocupem porções maiores dessa roleta. Assim, ao ser girada, os indivíduos que ocupam porções maiores terão mais chances de serem selecionados para compor a população da próxima geração. Isso significa que cada indivíduo da população tem uma probabilidade de seleção associada que é diretamente proporcional à sua aptidão.

Como o tamanho da população é mantido constante ao longo de todo o processo evolutivo, na transição de uma geração para a outra, a roleta deve ser girada um número de vezes igual ao tamanho da população. Devido às características da roleta, espera-se que na geração seguinte, a nova população seja formada pelos indivíduos mais aptos da geração anterior e que os menos aptos sejam eliminados.

A Figura 4.12 exhibe um exemplo de uma roleta com 5 indivíduos. Observa-se que o indivíduo identificado como 1 apresenta a maior aptidão, uma vez que ocupa a maior porção. O contrário acontece com o indivíduo identificado como 3. Portanto, a probabilidade de o indivíduo 1 compor a população de indivíduos da geração seguinte é bem maior que a do indivíduo 3. Assim, há uma probabilidade bastante alta de o indivíduo 3 ser eliminado na transição de uma geração para a outra.



Figura 4.12: Roleta de seleção dos indivíduos mais aptos

Após a seleção, alguns indivíduos da nova população sofrem a ação de operadores genéticos que são responsáveis pela modificação desses indivíduos, ou seja, geram novas soluções para o problema em questão. A cada operador genético é associada uma probabilidade de aplicação e, a partir dela, é possível selecionar o subconjunto de indivíduos que devem ser modificados. Os principais operadores genéticos são o cruzamento e a mutação. O cruzamento consiste em recombinar o material genético de dois indivíduos a fim de criar dois novos, ao passo que a mutação consiste em aplicar modificações aleatórias em uma ou mais características de um indivíduo para criar um novo. Considerando indivíduos codificados por vetores binários, as Figuras 4.13 e 4.14 ilustram, respectivamente, como funcionam os operadores de cruzamento e de mutação.

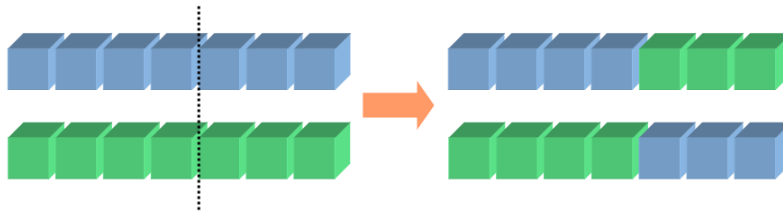


Figura 4.13: Operador clássico de cruzamento



Figura 4.14: Operador clássico de mutação

Enquanto o operador de mutação é responsável pela exploração do espaço de busca, o operador de cruzamento é responsável por tirar proveito de um determinado subespaço de busca. O equilíbrio adequado entre essas duas características é fundamental para a obtenção de bons resultados e, esse equilíbrio, é alcançado mediante a configuração criteriosa das taxas de aplicação dos operadores genéticos.

Depois da aplicação dos operadores genéticos tem-se uma nova população e, para encerrar um ciclo de evolução, avalia-se os indivíduos dessa nova população. Conforme mencionado, esses ciclos (gerações) se repetem até que uma condição de parada seja satisfeita. Essa condição de parada pode variar de um problema para outro, mas geralmente é dada pelo número de gerações que, juntamente com o tamanho da população e com as taxas de aplicação dos operadores genéticos, compõem o conjunto de parâmetros que devem ser inicialmente informados ao algoritmo genético. Ao final da evolução espera-se obter soluções que, lembrando mais uma vez o princípio de Darwin, sejam

melhores que as soluções encontradas nas gerações anteriores. O fluxograma de um algoritmo genético clássico está ilustrado na Figura 4.15.

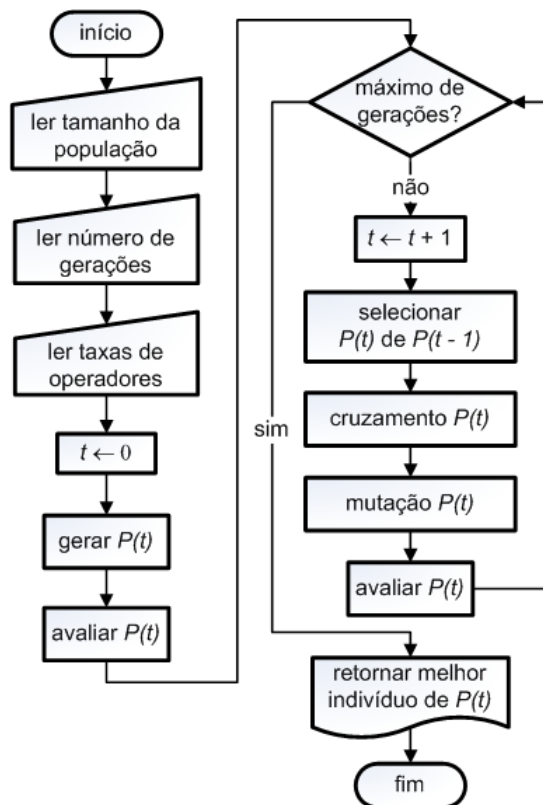


Figura 4.15: Fluxograma de um algoritmo genético clássico

### Aprimoramentos

Nos Algoritmos Genéticos, além da capacidade de lidar com problemas complexos de otimização, outra característica que merece ser destacada é a flexibilidade. A partir da versão clássica (Figura 4.15) é possível modificar a forma de representação, acrescentar novas instruções ou novos operadores e até mesmo associar novas técnicas, a fim de criar modelos híbridos. Essa flexibilidade é importante porque abre possibilidades para a criação de modelos de solução mais adequados ao problema que se quer resolver.

A representação por números binários, por exemplo, pode ser substituída por representações baseadas em números reais ou inteiros, baseada em ordem ou por alguma representação específica para o problema. As representações específicas podem ser tão complexas quanto as necessidades impostas pelo problema e, dentre outras particularidades, podem envolver duas ou mais formas de representação no mesmo cromossomo. É importante observar que, para cada tipo de representação, é necessário estabelecer operadores genéticos de cruzamento e de mutação que estejam alinhados àquela representação. Assim, cada

tipo de representação deve sempre ter os seus respectivos operadores associados. No caso das representações específicas, dependendo da complexidade, é necessário construir operadores particulares que estejam de acordo a forma de representação.

O procedimento de geração da população inicial pode ser alterado para que nem todos os indivíduos sejam gerados aleatoriamente. Entre esses indivíduos podem ser incluídas algumas sementes iniciais, que são indivíduos que representam soluções promissoras previamente conhecidas. Geralmente essas soluções surgem a partir da experiência adquirida por um especialista no problema em questão. Por isso, a inclusão de sementes iniciais é atrativa, uma vez que tende a conduzir o algoritmo genético a resultados finais melhores do que em situações nas quais a evolução parte de uma população inicial composta apenas por indivíduos aleatórios.

Outra modificação que pode impactar positivamente no desempenho do algoritmo genético é a garantia de preservação dos melhores indivíduos da população, na transição de uma geração para outra. Através a utilização de um mecanismo denominado *steady state*, tem-se a garantia de que um percentual dos melhores indivíduos de uma geração são automaticamente transportados para a geração seguinte, sem a necessidade de serem submetidos à seleção pela roleta. A vantagem da utilização do *steady state* é a garantia de que o desempenho da evolução não decresce durante as gerações. No algoritmo clássico, ainda que o melhor indivíduo da população tenha uma probabilidade de seleção maior que a dos indivíduos piores, não há garantia de que esse indivíduo estará presente na próxima população. Portanto, em determinado momento, essa solução pode ser perdida e isso acarreta uma queda no desempenho da evolução.

Além de ser aplicado entre uma geração e outra do algoritmo genético, o *steady state* também pode ser aplicado entre uma rodada e outra. Nesse contexto, uma rodada representa um experimento, ou seja, uma execução completa de um algoritmo genético. Da mesma forma como acontece entre as gerações dentro de uma rodada, o *steady state* entre rodadas garante que as melhores soluções sejam preservadas entre uma rodada e outra.

A realização de várias rodadas é interessante porque permite encontrar soluções melhores que aquelas encontradas com apenas uma rodada. Ao final de uma rodada, espera-se que o algoritmo genético tenha convergido para uma solução. Porém, a solução encontrada pode não ser a solução ótima. Assim, inicia-se uma nova rodada em que um percentual dos indivíduos é transferido da rodada anterior para a seguinte por *steady state*. O restante dos indivíduos da população inicial são gerados aleatoriamente. Assim, é possível aumentar

a diversidade da população sem a perda das melhores soluções encontradas até então. Essa diversidade dá chances ao algoritmo de encontrar uma solução ainda melhor que aquela encontrada na rodada anterior.

A utilização de taxas adaptativas, principalmente de cruzamento e de mutação, também é interessante porque permite que o equilíbrio entre o aproveitamento das melhores soluções e a exploração do espaço de busca seja ajustado ao longo da evolução. Com isso, é possível, por exemplo, aplicar taxas de cruzamento maiores no início da evolução e taxas menores no final. De maneira inversa, pode-se aplicar taxas de mutação menores no início e maiores no final da evolução. Esse ajuste no equilíbrio é importante para que, no início, a evolução tire maior proveito das soluções disponíveis e, à medida que se aproxima do final, comece a explorar mais o espaço de busca a fim de evitar uma convergência prematura.

Essas possibilidades de modificação tornam os Algoritmos Genéticos extremamente versáteis e, por isso, aplicáveis a problemas de otimização de diversas áreas do conhecimento. A qualidade dos resultados obtidos nessas aplicações está diretamente relacionada à qualidade da etapa inicial de modelagem do problema. Na modelagem, uma vez conhecidas as características do problema, são definidos os detalhes do algoritmo genético de forma a aproveitar ao máximo o seu potencial para a solução de problemas de otimização.