

## 4

# Técnicas de Processamento de Sinais e Otimização

Os perfis de poço e os traços sísmicos são sinais discretos no espaço e no tempo respectivamente. Os perfis de poço, vistos como sinais, são processados para atenuar suas perturbações, retirar informações estratigráficas e produzir as medições previstas. Por outro lado, os traços sísmicos são processados para obter informações das posições dos estratos de maior contraste e que são atravessadas pelos poços.

Uma parte importante do processamento consiste em identificar padrões semelhantes nos sinais. Isso, pode ser realizado com uma técnica de busca e otimização conhecida como Algoritmos Genéticos. Assim, nas seções seguintes são explicadas essa e outras técnicas de processamento de sinais, utilizadas para desenvolver o método de previsão proposto.

### 4.1. Amostragem

As variações das grandezas físicas, normalmente, são contínuas e os cálculos complexos para a análise dessas grandezas seriam inviáveis, embora existam as ferramentas de cálculo contínuo (processadores analógicos). Mas através da amostragem pode-se recolher uma série de valores para realizar cálculos aproximados. A amostragem é realizada com diversas técnicas determinísticas ou probabilísticas, dependendo do problema. No processamento de sinais, normalmente são utilizadas as determinísticas.

Os sinais analisados são funções contínuas e dependentes de uma variável independente como o tempo ( $t$ ) ou como a distância ( $z$ ). Na prática, esses sinais são amostrados, mas isso quebra sua continuidade e impossibilita seu processamento. Para manter a validade dos algoritmos de processamento, pode-se formular uma função que inclui as amostras num somatório de funções deslocadas e multiplicadas pelo valor da amostra correspondente nessa posição. Uma função ou relação simples de amostragem utiliza funções do tipo impulso-unitário, como a da eq. (4.1):

$$x_a(t) = \sum_{k=0}^{\infty} x(k \Delta t) \delta(t - k \Delta t), \quad (4.1)$$

em que,  $x(k \Delta t) = x(k)$  representa a amostra  $k$  de uma função ou uma medição de uma grandeza física,  $\Delta t$  é o tempo de amostragem ou diferença de tempo entre amostras;  $x_a(t)$  representa a função contínua construída com as amostras, normalmente, com  $k$  inteiro e finito.

Nesse ponto aparecem algumas questões como, será possível recuperar o sinal original  $x(k)$  a partir de  $x_a(t)$ ? Ou, até que ponto é possível uma recuperação aproximada? Essas questões podem ser respondidas através do teorema de *Nyquits* (Yarlagadda, 2009), no qual o tempo de amostragem  $\Delta t$  e a máxima frequência ( $B$ ) do sinal  $x(k)$  são relacionados para determinar o limite para uma recuperação aceitável. Normalmente, a recuperação de um sinal  $x(k)$  de banda limitada é realizada com um filtro passa baixa, com frequência de corte equivalente à frequência  $B$ . Assim, as condições para determinar o tempo de amostragem  $\Delta t$  são:

- Se  $\Delta t \leq 1 / (2 B)$ , então é possível a recuperação;
- Se  $\Delta t \ll 1 / (2 B)$ , então o sinal resultante contém muitas amostras e, talvez, mais do que as necessárias para o processamento eficiente;
- Se  $\Delta t > 1 / (2 B)$ , então o sinal resultante apresenta uma distorção conhecida como faseamento (*aliasing*).

## 4.2. Convolução

A convolução é uma operação matemática adaptada e aplicada em diferentes áreas como o processamento de sinais. Esta operação pode ser utilizada, por exemplo, para simular a resposta de um sistema físico ou abstrato como resposta a um determinado estímulo, ou simplesmente para extrair algum tipo de informação de um conjunto de medições.

A convolução é uma operação sobre duas funções reais ( $\mathbb{R}$ ) para produzir uma terceira função através do cálculo da área da sobreposição de uma das funções deslocada sobre a outra, de acordo com a eq. (4.2). A dependência das funções pode ser multidimensional, mas neste trabalho as operações são realizadas sobre funções que dependem do tempo ( $t$ ) ou da profundidade ( $z$ ). A operação de convolução permite analisar sistemas lineares e invariantes ao longo da variável independente, utilizando a eq. (4.2):

$$y(t) = \int_0^t h(t-\tau)x(\tau)d\tau = (h * x)(t), \quad (4.2)$$

em que  $x(t)$  e  $y(t)$  são sinais e  $h$  é o modelo do sistema. A eq. (4.2) e suas propriedades, mostradas na Tabela A.1 no Anexo, são válidas tanto para sinais periódicos quanto para não periódicos (Yarlagadda, 2009).

Para modelos de sistemas discretos ou digitais utiliza-se a convolução discreta definida como uma operação de valores discretos para produzir valores discretos. Assim, pode-se utilizar uma função de amostragem  $x_a$ , como da eq. (4.1), para definir um sinal de entrada para a eq. (4.2) e obter a eq. (4.3), da forma seguinte:

$$y_a(t) = \int_0^t h(t-\tau) \sum_{k=0}^{\infty} x(k \Delta t) \delta(\tau - k \Delta t) d\tau, \quad (4.3)$$

em que  $\Delta t$  é o tempo de amostragem e  $y_a$  é a função amostrada de  $y$ . Rearranjando a integral, dentro do somatório, tem-se a eq. (4.4):

$$y_a(t) = \sum_{k=0}^{\infty} x(k \Delta t) \int_0^t h(t-\tau) \delta(\tau - k \Delta t) d\tau. \quad (4.4)$$

O resultado da integral devido ao impulso unitário será  $\delta(t-k\Delta t)$  uma amostra dependente de  $k$ , no intervalo  $0 \leq k \leq n-1$ , em que  $n$  é o número de amostras. Logo,  $t$  também é discreto e redefinido como  $t_j = j \Delta t$ , em que  $j$  pode tomar valores no mesmo intervalo de  $k$ . O resultado da operação de convolução para uma amostra é calculado com a eq. (4.5):

$$y_a(t_j) = \sum_{k=0}^{n-1} x(k \Delta t) h(t_j - k \Delta t). \quad (4.5)$$

A eq. (4.5) pode ser escrita também de maneira sintética conforme a eq. (4.6):

$$y_j = \sum_{k=0}^{n-1} x_k h_{j-k}. \quad (4.6)$$

A partir de agora, a eq. (4.6) é utilizada de maneira implícita nas operações de filtragem e simulação dos sinais sísmicos.

### 4.3. Filtragem

A filtragem é a operação de separação da informação desejada de um conjunto de dados. Os filtros são componentes utilizados para realizar a filtragem e podem ser agrupados, principalmente, em duas categorias: determinísticos e probabilísticos. Nesta seção são descritos três tipos de filtros para aplicar às medições dos perfis de poço.

A filtragem, na teoria de processamento de sinais contínuos, é a operação de separação de um sinal  $R(\omega)$ , na frequência  $\omega$ , em outros dois. Basicamente, um sinal com variações de alta frequência  $Y_H(\omega)$  e outro com variações de baixa frequência  $Y_B(\omega)$ , como são mostrados na Figura 4.1. Por outro lado, para realizar a filtragem das medições de poço, faz-se uma troca de variável, do tempo ( $t$ ) para a profundidade ( $z$ ). Assim, a frequência toma a forma de  $\omega = 2\pi/z$  e as expressões das saídas  $y_H(z)$  e  $y_B(z)$  são resultados das integrais de convolução.

Normalmente, para alguns tipos de medições de poço, as variações bruscas (de alta frequência) representam perturbações não desejadas e podem ser retiradas com os filtros.

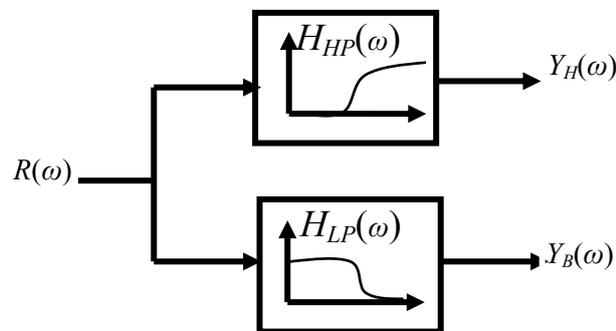


Figura 4.1 - Filtragem do sinal  $R(\omega)$  na frequência  $\omega$ .  $H_{HP}$ , bloco passa altas,  $H_{LP}$  bloco passa baixas.  $Y_H(\omega)$  sinal de altas frequências e  $Y_B(\omega)$  sinal de baixas frequências.

Na Figura 4.1 é mostrada a filtragem de um sinal contínuo, mas os sinais dos perfis de poço são discretos (amostras) e, por este motivo, os filtros utilizados são adaptações discretas de acordo com o número de amostras.

### 4.3.1. Filtro Binomial

O filtro binomial é usado para suavizar sinais discretos através do ajuste dos pesos  $\{w_{ik}\}$  de acordo com a distância entre os valores do sinal e o valor estimado. Esses pesos são calculados a partir de uma expansão binomial de expoente  $q$ . O filtro binomial é inteligível e fácil de calcular (Takezawa, 2006). Na eq. (4.7) é mostrado o filtro discreto como um somatório dos elementos da expansão binomial, centrada na posição  $i$ . Assim:

$$y_i = \sum_{k=i-q/2}^{i+q/2} w_{ik} r_k, \quad (4.7)$$

em que  $r_k$  é a medição  $k$  do perfil de poço e  $w_{ik}$  é o peso da medição  $r_k$  para calcular o valor suavizado  $y_i$ . A eq. (4.8) define os pesos como:

$$w_{ik} = \begin{cases} \frac{q!}{(q-o)!o!} \frac{1}{2^q}, & o = q/2 + i - k, \text{ Se } 0 \leq q/2 + i - k \leq q, \\ 0 & \text{outros,} \end{cases} \quad (4.8)$$

em que  $q$  é um número par positivo e expoente do binômio. Os pesos são associados aos fatores do triângulo de Pascal.

Do ponto de vista dos filtros digitais, o filtro binomial forma uma resposta rápida e compacta ao impulso finito (FIR), onde os pesos formam uma aproximação rápida da distribuição (ou filtro) de Gauss. A sua projeção pode ser feita como uma aproximação da variância desta distribuição, com  $\sigma \sim 0,5 q^{0,5}$ . A relação entre o mínimo comprimento de onda  $\lambda$  do sinal preservado e a variância deve ser  $\sigma = \lambda / 6$ .

A teoria de processamento de sinais considera outros critérios de frequência para a projeção do filtro. A transformada de *Fourier* de tempo discreto (DTFT), ou a transformada discreta de *Fourier* (DFT), do filtro binomial de expoente  $q = 2$  corresponde a um único período da função cosseno mais uma constante. Em outras palavras, corresponde a um filtro passa-baixa sem ondulações na banda de parada (Haddad, 1971). A eq. (4.9) é a função generalizante do filtro na frequência:

$$W_q(\omega_n) = \left( \frac{1 + \cos(\omega_n)}{2} \right)^{q/2}, \quad (4.9)$$

em que  $\omega_n$  é a frequência normalizada no intervalo  $[0, \pi]$  e é usado para projetar o filtro (Strang & Nguyen, 1996). A frequência de corte é o ponto de início para atenuar as altas frequências e é um parâmetro requerido para a projeção do filtro. Por outro lado, a frequência normalizada ( $\omega_n$ ) é proporcional à frequência do sinal ( $\omega_r$ ) e ao intervalo de amostragem  $\Delta z$ . Assim,  $\omega_n = \omega_r \Delta z$ , ou  $f_n = \Delta z f_r$ , de acordo com a condição de amostragem. Para atenuar as altas frequências do sinal, o filtro deve reduzir a energia dos componentes de alta frequência para menos de  $-3$  dB (ponto de meia potência). Em seguida,  $W_q(\omega_n)$  deve ser igualada a  $2^{-0.5}$  para obter a frequência de corte. Na prática, a potência  $q$  e a frequência de corte podem ser encontradas através de um procedimento iterativo e com um valor aproximado de  $W_q(\omega_n) = 2^{-1}$ .

#### 4.3.2. Filtro de *Kalman*

Os métodos anteriores ao de *Kalman* (Kalman, 1960), como os de *Kolmogorov* e de *Wiener*, utilizam a caracterização estatística para separar um sinal do ruído. É por esse motivo que o nome de “filtro” foi estendido para esses tipos de método, entre eles o filtro de *Kalman*. Com o passar dos anos, suas aplicações foram além da simples separação de sinais. Atualmente, o filtro de *Kalman* é utilizado em muitas áreas de pesquisa, mas suas aplicações estão concentradas na estimação e na análise do desempenho de estimadores.

O filtro de *Kalman* é um estimador recursivo (Kalman, 1960; Trees, 1971), isto significa que só os valores do estado prévio de uma variável e a medida atual são necessários para calcular a estimativa do valor do estado atual. Em contraste com as técnicas de estimativa por blocos, a história das observações e / ou das estimativas não é necessária. As equações do filtro de *Kalman* foram desenvolvidas para estimar os estados de um sistema dinâmico a partir de medições indiretas com ruído. Essas equações são diferenciais e, quando envolvem processos aleatórios, são denominadas equações diferenciais estocásticas.

Este filtro tem variações e adaptações que dependem do sistema analisado (Grewal & Andrews, 2008). Por exemplo, o modelo discreto deste filtro, mostrado na Figura 4.2, pode ser aplicado para separar o ruído das amostras  $r_k$  de um perfil de poço. Esses perfis, como o de raio gama (GR), são definidos como sequências aleatórias e, às vezes, não estacionárias.

Na Figura 4.2 são mostrados os blocos de um sistema estocástico discreto, o modelo de medição e o filtro de *Kalman* para estimar os estados  $y_k$  do sistema.

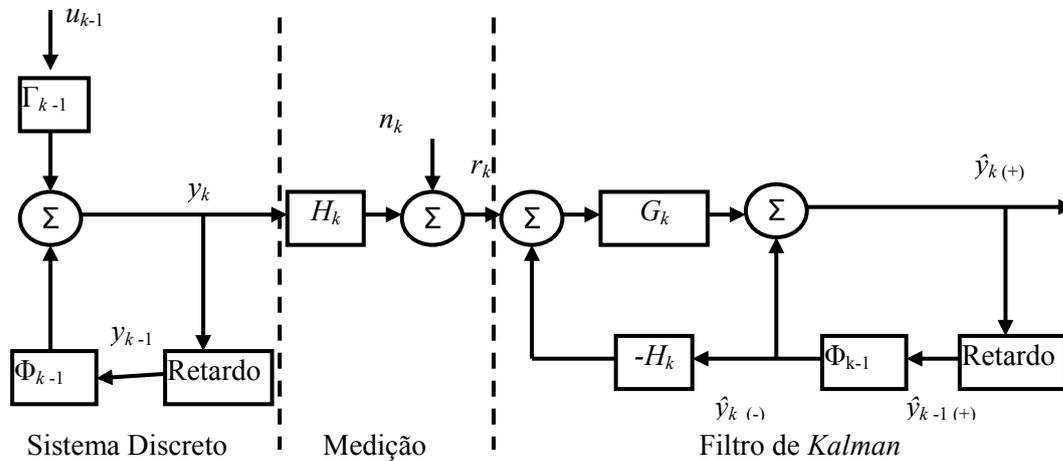


Figura 4.2 - Diagrama de blocos do sistema, do modelo de medição e do filtro discreto de *Kalman*.

Quando há pouca informação sobre o sistema medido, condições simplificadoras são aplicadas às variáveis  $u_{k-1}$ , e  $n_k$ . Uma dessas condições é assumir que essas variáveis são sequências aleatórias gaussianas com médias iguais a zero. Assim, os modelos das suas equações também podem ser simplificados, como na eq. (4.10) e na eq. (4.11).

$$y_k = \Phi y_{k-1} + \Gamma u_{k-1} \quad (4.10)$$

$$r_k = H y_k + n_k, \quad (4.11)$$

em que  $H$  é o modelo de observação,  $\Phi$  é o modelo da transição de estado e  $\Gamma$  o modelo de entrada de controle. Os modelos podem ser valores únicos ou matrizes, quando as medidas  $r_k$  são agrupadas em lotes. As variáveis aleatórias discretas, que representam o vetor de estados  $y_k$ , o vetor de ruído  $n_k$  e o vetor de controle  $u_{k-1}$ , têm distribuições de probabilidade normal  $N(m_y, K_y)$ ,  $N(0, K_n)$  e  $N(0, Q)$ , respectivamente. Os parâmetros  $K_y$ ,  $K_n$  e  $Q$  são as matrizes de covariância das variáveis aleatórias discretas e  $m_y$  é a média do vetor de estados  $y_k$ .

A estimativa iterativa de  $y_k$  pode ser feita desde a eq. (4.12) até a eq. (4.15):

$$\hat{y}_k = \Phi \hat{y}_{k-1} + G_k (r_k - H\Phi \hat{y}_{k-1}), \quad (4.12)$$

$$G_k = P_k H^T K_n^{-1} \quad (4.13)$$

$$M_k = \Phi P_{k-1} \Phi^T + \Gamma Q \Gamma^T, \quad (4.14)$$

$$P_k = (H^T K_n^{-1} H + M_k^{-1})^{-1} \quad (4.15)$$

em que  $\hat{y}_k$  é o estado estimado;  $G_k$  é a matriz de ganho do filtro;  $M_k$  é a matriz de atualização;  $P_k$  é a matriz de covariância (Trees, 1971). Quando existe apenas uma única variável de estado, estimada a cada unidade, estas matrizes se transformam em valores unitários.

Calcular todos os valores deste sistema de equações, claramente, implica conhecer profundamente o sistema medido. Se algumas variáveis são não Gaussianas, então é conveniente aplicar um procedimento de adaptação do filtro (Grewal & Andrews, 2008), utilizando a distribuição de probabilidades das amostras  $r_k$ . Porém, para os poços com informação incompleta dos seus parâmetros, é mais útil um filtro que retire a informação dos mesmos sinais de maneira automática.

### 4.3.3. Ondeletas ou *Wavelets*

O processamento com ondeletas é uma técnica interessante que pode ser aplicada para muitos propósitos diferentes, inclusive filtrar ruído. As ondeletas podem analisar sinais não periódicos e com feições afiadas melhor do que outras técnicas. Deste modo, esta técnica pode contribuir para limpar os sinais dos perfis de poço como os de raios gama (Misiti et al., 2007).

Uma ondeleta é uma forma de onda de duração efetivamente limitada com uma média zero e com tendência assimétrica. A análise por ondeleta é a divisão de um sinal em versões deslocadas e amplificadas da ondeleta original (ou mãe). Assim, em uma dimensão, pode-se distinguir a ondeleta  $\psi$  a partir da função associada  $\varphi$ , chamada de função de escala. As propriedades básicas de  $\psi$  e  $\varphi$  são:

- $\int \psi(x) dx = 0$ , em que  $\psi$  é utilizada para definir os detalhes.
- $\int \varphi(x) dx = 1$ , em que  $\varphi$  é utilizada para definir as aproximações.

Na teoria de processamento de sinais é utilizado um banco de filtros para decompor o sinal original em outros de frequências baixas e altas. Este procedimento é conhecido como análise e pode ser utilizado para codificar o sinal e melhorar a relação sinal-ruído (Strang & Nguyen, 1996). O procedimento inverso é conhecido como síntese ou reconstrução do sinal. Isto é semelhante na transformada de ondeletas, onde é utilizado um banco de filtros que depende da ondeleta original e produz coeficientes de aproximação ( $C_A$ ) e detalhe ( $C_D$ ), calculados com a eq. (4.16):

$$C(J,l) = C(a,b) = \int y(z) \frac{1}{\sqrt{a}} \Psi\left(\frac{z-b}{a}\right) dz, \quad (4.16)$$

em que  $a = 2^J$ ,  $b = l 2^J$ ,  $(J, l) \in \mathbf{Z}^2$ ;  $C$  é um coeficiente  $\in \mathbf{R}$ ,  $z$  representa a profundidade ou distância;  $\Psi$  é a ondeleta original. Como não é possível processar o sinal original contínuo  $r(z)$ , utiliza-se uma relação contínua equivalente conhecida como sinal de amostragem  $y(z) = \sum r_k \delta(z - \Delta z k)$ . A soma ou reconstrução dos sinais de aproximação ( $A$ ) e detalhe ( $D$ ) produz o sinal  $y(z)$ . Quando eles são analisados em frequência,  $C_A$  representa o coeficiente de baixa frequência e  $C_D$  representa o coeficiente de alta frequência. Todo valor de  $J$  é um nível de decomposição do sinal com seus coeficientes  $C_A$  e  $C_D$ , em faixas de frequências cada vez mais específicas.

A síntese utiliza os coeficientes para reconstruir o sinal  $\hat{y}(z)$  com a eq. (4.17):

$$\hat{y}(z) = \sum_{J \in \mathbf{Z}} \sum_{l \in \mathbf{Z}} C(J,l) \Psi_{J,l}(z), \quad (4.17)$$

em que  $\Psi_{J,l}(z) = 2^{-J/2} \Psi(2^{-J} z - l)$ ;  $\hat{y}(z)$  é o sinal discreto reconstruído e deve ser quase idêntico a  $y(z)$ . No filtro de ruído com ondeletas é aplicado um limiar (normalmente suave) aos coeficientes de detalhe ( $C_D$ ) ou de uma faixa de alta frequência específica, depois é reconstruído o sinal filtrado (Misiti et al., 2007).

Se o sinal, perturbado só pelo ruído branco, é analisado, as amplitudes dos detalhes dos níveis crescentes diminuem e as variações dos detalhes diminuem gradualmente. Em outras palavras, os sinais de aproximação e detalhes não têm mais ruído branco. Desse modo, o ruído colorido é inserido através dos filtros (Misiti et al., 2007). De um ponto de vista estatístico, o filtro é um modelo de regressão e o método pode ser considerado como uma estimativa não paramétrica do sinal  $y(z)$  com bases ortogonais. O método é eficiente para  $y(z)$  com distribuições dispersas dos coeficientes (poucos diferentes de zero).

Para projetar um filtro passa baixa, aplicado a um perfil de poço, é necessário achar uma ondeleta adequada e um nível  $J$  ótimo (ou família). Há algumas famílias de ondeletas, mostradas na Tabela A.2 do Anexo, que poderiam ser utilizadas de acordo com a forma da onda. Por exemplo, a família de *Daubechies* tem algumas ondeletas importantes usadas em muitas aplicações. Estas estão rotuladas como 'db' seguidas de um número que indica a ordem. Quando a ordem aumenta, a ondeleta mostra mais oscilações e características suaves (veja a Figura 4.3); Por isso, as ondeletas *db2* até a *db5* são mais apropriados para conseguir formas suavizadas. Note-se que a ondeleta *db1* é a ondeleta *Haar*.

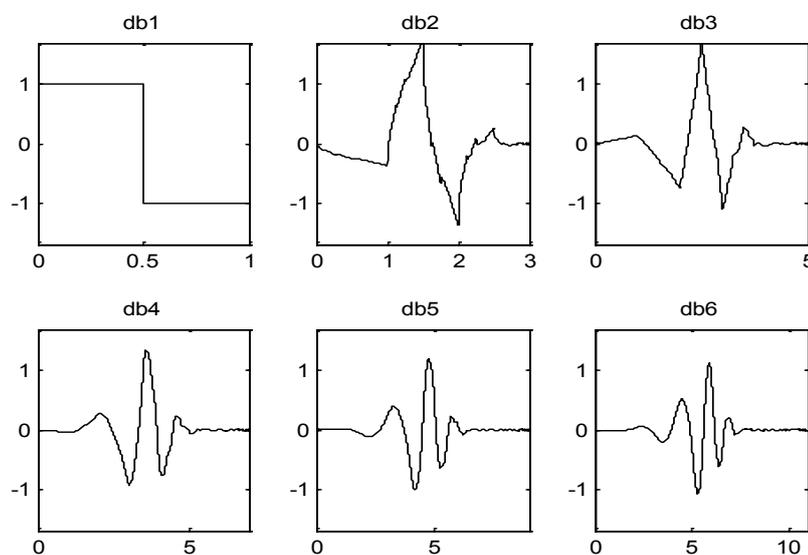


Figura 4.3 - Forma de Onda de seis ondeletas da família Daubechies.

#### 4.4. Filtragem Multitaxa

A variação do número de amostras originais de uma série de medições é, geralmente, conhecida como reamostragem, e pode ser realizada através de diferentes técnicas. Uma delas é a filtragem multitaxa (Diniz et al., 2010) aplicada nos sistemas de processamento digital de sinais que utilizam mais de uma taxa de amostragem ( $1/\Delta t$ ).

Existem dois componentes básicos para obter diferentes números de amostras: o expansor e o dizimador. O primeiro para incrementar o número de amostras e o segundo para reduzir o número de amostras.

O expansor, também conhecido como interpolador, incrementa o número de amostras de uma sequência discreta  $r_k$ , com um componente que inclui

amostras (*up-sampler*) entre as amostras originais  $r_k$  e um filtro passa baixa  $H$  para a interpolação dos valores de saída, como é mostrado na Figura 4.4,

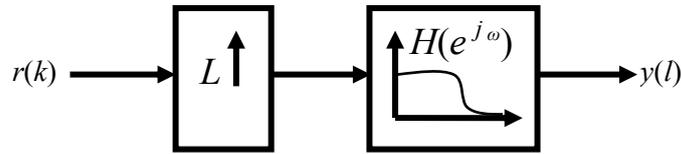


Figura 4.4 - Diagrama de blocos do expansor.  $H$  representa o filtro discreto passa baixa na frequência  $\omega$  e  $y(l)$  a saída a uma taxa de amostragem maior.

Na eq. (4.18) é mostrada a função discreta do componente que inclui  $L-1$  amostras nulas (zeros) entre as amostras originais, antes do filtro discreto passa baixa ou interpolador  $H$ .

$$r(l) = \begin{cases} r(k = l/L), & l, L \in \mathbb{Z}, \quad l = 0, \pm L, \pm 2L, \pm 3L, \dots \\ 0, & \text{outros} \end{cases}, \quad (4.18)$$

em que  $l$  é o índice do sinal reamostrado  $r(l)$ ,  $k$  é o índice do sinal original  $r(k)$  e  $L$  é o fator de expansão.

O dizimador reduz o número de amostras através da combinação do componente de eliminação de amostras (*down-sampler*) e um filtro passa baixa  $H$  (ver Figura 4.5). Esse filtro serve para eliminar a distorção por faseamento (filtro *anti-aliasing*), produzida pela sobreposição das réplicas do espectro original, depois da eliminação das amostras.

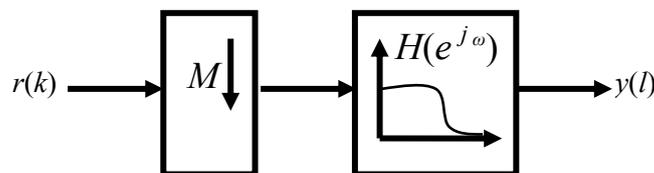


Figura 4.5 - Diagrama de blocos do dizimador.  $H$  representa o filtro discreto passa baixa na frequência  $\omega$ , e  $y(l)$  a saída a uma taxa de amostragem menor.

De forma semelhante, o componente de eliminação (*down-sampler*), de  $M-1$  amostras de cada  $M$  amostras, realiza essa operação através da eq. (4.19):

$$r(l) = \{r(k = lM), \quad l, k, M \in \mathbb{Z}. \quad (4.19)$$

em que  $l$  é o índice do sinal reamostrado  $r(l)$ ,  $k$  é o índice do sinal original  $r(k)$  e  $M$  é o fator de compressão. O filtro  $H$  para eliminação da distorção pode ser aplicado antes ou depois do componente de eliminação de amostras.

A combinação dos dois componentes, o expansor e o dizimador em cascata como na Figura 4.6, pode ser utilizada para a reamostragem de taxa fracionária ( $M/L$ ).

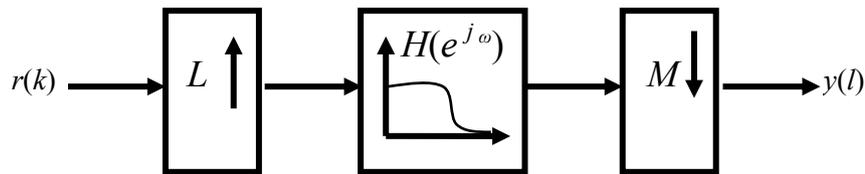


Figura 4.6 - Diagrama de blocos para a reamostragem a taxa fracionária ( $M/L$ ). Combinação dos componentes expansor e dizimador.

#### 4.5. O Spectrum e o Cepstrum de Potência

O *spectrum* de potência é uma técnica de análise espectral de sinais. O *spectrum* de potência de um sinal é representado como uma função real, conforme a eq. (4.20):

$$G_{xx}(\omega) = |F\{g(t)\}|^2, \quad (4.20)$$

em que  $F\{\}$  representa a transformada de *Fourier* de uma função  $g$ , do tempo  $t$  para a frequência  $\omega$ . Na eq. (4.20) é calculado o módulo ao quadrado da transformada de *Fourier*. Assim, quando  $G_{xx}$  é desenhada, não só a frequência fundamental e os harmônicos de um sinal são visíveis, mas também as frequências das perturbações ou o ruído. Para traçar o *spectrum* de uma sequência de amostras é utilizada a transformada discreta de *Fourier* (DFT), com um algoritmo conhecido como transformada rápida de *Fourier* (FFT).

Para obter informação mais clara do sinal, foi desenvolvida uma técnica como o *cepstrum* de potência (nome produzido ao inverter as primeiras palavras de *spec-trum*). O *cepstrum* de potência é uma técnica muito útil de análise espectral, (Norton & Karczub, 2003), representada com a função real mostrada na eq. (4.21):

$$C_{p_{xx}}(\tau) = F^{-1}\{\log_{10}(G_{xx}(\omega))\}, \quad (4.21)$$

em que  $F^{-1}\{\}$  representa a transformada inversa de *Fourier* do logaritmo do *spectrum* de potência  $G_{xx}$ . A transformação se realiza da frequência para o tempo  $\tau$  conhecido como “*quefrença*”. Com o *cepstrum* de potência é possível separar os efeitos da fonte e da função de transferência ou meio de transmissão sólido ou líquido. O processo de separação também é conhecido como *deconvolução*.

Por outro lado o *cepstrum* de potência também é definido com a eq. (4.22):

$$C_{p_{xx}}(\tau) = \left| F^{-1} \{ \log_{10}(G_{xx}(\omega)) \} \right|^2. \quad (4.22)$$

Ambas, a eq. (4.21) e a eq. (4.22), são equivalentes e a única diferença está nas suas amplitudes. Neste trabalho é utilizada a eq. (4.22) porque realça os picos dos harmônicos.

O *cepstrum* de potência é aplicado para encontrar qualquer estrutura periódica no *spectrum* de potência como padrões harmônicos ou medição de propriedades das superfícies de reflexão de vibrações mecânicas (Norton & Karczub, 2003). O *cepstrum* de potência é uma técnica complementar à análise espectral e permite identificar características que não são observáveis na frequência. Mas esta técnica tende a suprimir informação do conteúdo do espectro do sinal. Por este motivo, se aconselha a análise do *cepstrum* conjuntamente com o *spectrum*.

#### 4.6. Algoritmos Genéticos

Os Algoritmos Genéticos constituem uma técnica de busca paralela de soluções ótimas, inspirada no princípio da seleção natural advindo da teoria evolucionária de Charles Darwin, associado à noção de reprodução genética (Goldberg, 1989). O código genético constitui a identidade de um indivíduo, diferenciando-o de todos os outros, e é organizado em cromossomos. De acordo com a teoria de Darwin, o princípio de seleção privilegia os indivíduos mais aptos com maior longevidade com uma conseqüente maior probabilidade de reprodução. Com mais descendentes, estes indivíduos têm mais chance de perpetuarem seus códigos genéticos nas próximas gerações. Estes são os princípios que servem de base para a construção de algoritmos computacionais que buscam uma melhor solução para um determinado problema de otimização, evoluindo populações de soluções codificadas através de cromossomos artificiais.

Uma das grandes vantagens dos Algoritmos Genéticos é a forma como integra informações do problema para encontrar uma boa solução (indivíduo). Se as equações exatas do problema não são conhecidas, então podem ser utilizadas informações empíricas para encontrar aproximações à melhor solução. A própria avaliação de uma solução serve como bússola, indicando a direção em que devem estar localizadas as melhores soluções.

A seguir, os componentes de um algoritmo genético são apresentados em detalhes.

**Representação:** neste estágio define-se a estrutura ou forma do cromossomo, com os respectivos genes que o compõem, de maneira que este seja capaz de descrever todo o espaço de busca relevante ao problema. Além disso, o espaço de busca deve ser finito. Os principais tipos de representação são: binária, números reais, inteiros, baseada em ordem e estruturas de árvore (Back et al., 2000).

**Codificação e Decodificação:** a solução de um problema pode ser representada por um conjunto de parâmetros (genes), organizados em sequência (cromossomo). Este processo é denominado como codificação. Cada gene é representado por um símbolo de um sistema alfabético. Originalmente, utilizou-se o alfabeto binário (0, 1), porém, novos modelos de Algoritmos Genéticos codificam as soluções com outros alfabetos, como, por exemplo, com números reais (Michalewicz, 1996).

A decodificação do cromossomo consiste, basicamente, na construção da solução real do problema a partir do cromossomo. O processo de decodificação constrói a solução para que esta seja avaliada pelo problema.

**Avaliação:** a avaliação permite ao algoritmo genético determinar sua proximidade à solução ótima do problema. Ela é feita através de uma função que melhor representa o problema e tem por objetivo oferecer uma medida de aptidão de cada indivíduo na população corrente, que irá dirigir o processo de busca. Dado um cromossomo, a avaliação consiste em se associar um valor numérico de "adaptação", o qual se supõe proporcional à sua "utilidade" ou "habilidade" de solucionar o problema em questão.

**Reprodução:** se refere ao processo de selecionar e copiar um determinado cromossomo para a população seguinte de acordo com sua aptidão. Isto significa que os cromossomos mais aptos têm maior probabilidade de contribuir para a formação de um ou mais indivíduos da população seguinte. Existem basicamente os seguintes métodos: troca de toda população, troca de toda população com elitismo, onde o cromossomo mais apto da população

corrente é copiado para população seguinte, e troca parcial da população (*steady state*), onde alguns dos melhores indivíduos, da população corrente, são copiados para população seguinte (Goldberg, 1989; Koza, 1992). Algumas vezes, a troca parcial da população também é denominada de elitismo.

**Seleção:** é um método essencial de Algoritmos Genéticos e consiste em escolher cromossomos de acordo com a sua função de avaliação, onde os indivíduos melhor adaptados são mais susceptíveis de ser escolhidos. O método de seleção é responsável pela velocidade da evolução e geralmente citado como responsável nos casos que a convergência prematura retardar o sucesso do algoritmo evolucionário (Banzhaf et al.,1998). A literatura identifica cinco principais mecanismos de seleção: proporcional, por torneios, com truncamento, por normalização linear e por normalização exponencial. (Blickle, 1996).

**Operadores Genéticos:** os operadores mais conhecidos nos Algoritmos Genéticos são Cruzamento e Mutação (Michalewicz, 1996).

O Cruzamento é um operador baseado na troca de partes dos cromossomos (pais), formando-se duas novas soluções (filhos), para manter o tamanho da população. Este processo pode ser observado no exemplo da Figura 4.7, em que a solução está codificada com alfabeto binário.

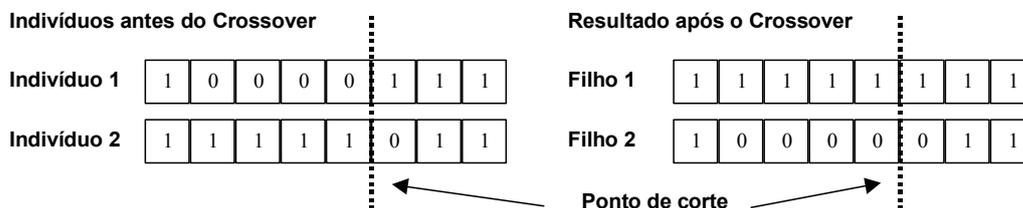


Figura 4.7 – Operação de Cruzamento de um ponto para indivíduos binários

O ponto onde ocorre o corte para a realização do cruzamento é escolhido aleatoriamente e o operador é conhecido como cruzamento de um ponto. O operador com cortes em mais de um ponto é denominado cruzamento multiponto (Goldberg, 1989; Michalewicz, 1996).

Para realizar o cruzamento, primeiro é necessária a escolha, por sorteio, dos cromossomos “pais”. Em seguida, o cruzamento acontece se um parâmetro entre 0 e 1, denominado taxa de cruzamento, é maior do que a probabilidade sorteada para cada par de pais. Deste modo, se a taxa de cruzamento é menor do que essa probabilidade, então os cromossomos “pais” são repassados sem modificação para a geração seguinte, criando “filhos” idênticos a eles.

Para a representação baseada em números reais existem vários tipos de operadores de cruzamento como, por exemplo, o heurístico (Mathworks, 2007). Esse operador gera os filhos conforme a eq. (4.23):

$$filho = Indivíduo2 + R * (Indivíduo1 - Indivíduo2), \quad (4.23)$$

em que os dois indivíduos (pais) são combinados linearmente aplicando uma taxa  $R$  para produzir um filho na direção do melhor  $Indivíduo1$  (normalmente  $R > 1$ ).

A ideia do operador de cruzamento é tirar vantagem do material genético presente na população. Em outras palavras, transmitir à geração seguinte a informação contida nos genes que aproximam os indivíduos ao indivíduo ótimo.

Outro operador genético é a Mutação, definida como a troca aleatória do valor contido nos genes de um cromossomo por outro valor válido do alfabeto. No caso de alfabeto binário troca-se de 0 para 1 e vice-versa. Da mesma forma que no cruzamento, utiliza-se uma taxa de mutação que, para cada bit da sequência de caracteres, sorteia-se se ocorrerá ou não a mutação; caso o número sorteado entre 0 e 1 (probabilidade) seja menor ou igual à taxa de mutação, o bit será trocado por outro valor válido pertencente ao alfabeto (ver Figura 4.8).

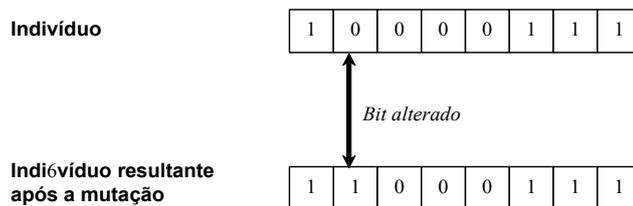


Figura 4.8 – Operação de Mutação de um indivíduo binário

Esta forma de mutação também é conhecida como mutação uniforme porque todos os genes têm a mesma probabilidade de mutação. Quando a mutação é realizada em cromossomos com genes reais, a troca se realiza por um número aleatório no intervalo do parâmetro representado pelo gene (Mathworks, 2007).

A mutação garante a diversidade das características dos indivíduos da população e permite que sejam introduzidas informações que não estiveram presentes em nenhum dos indivíduos. Além disso, proporciona uma busca

aleatória (exploração) no Algoritmo Genético, oferecendo oportunidade para que mais pontos do espaço de busca sejam avaliados.

**Parâmetros da Evolução:** Os parâmetros que mais influenciam no desempenho do algoritmo genético, (Pacheco, 2006), são:

- Tamanho da População
- Taxa de Cruzamento
- Taxa de Mutação
- Taxa de *Steady State* (ou número de indivíduos para o elitismo)
- Número de gerações

**Crítérios de Parada:** existe um conjunto de critérios de parada para interromper ou deter os ciclos da evolução. Entre os principais estão: cumprir o número máximo de gerações; a diferença mínima entre duas avaliações; a diferença mínima entre os genes do cromossomo e o tempo de execução total.

Na Figura 4.9 é mostrado o ciclo evolutivo de uma geração. Os ciclos devem se repetir até cumprir pelo menos um dos critérios de parada.

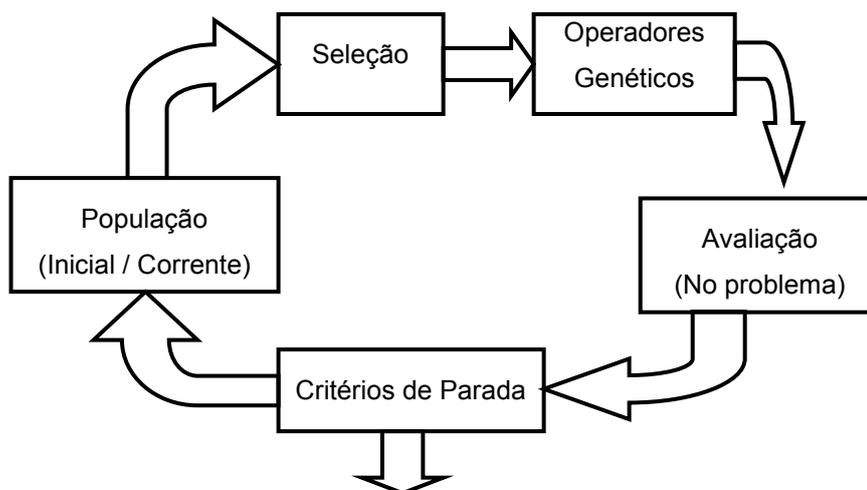


Figura 4.9 - Diagrama de blocos das etapas do ciclo evolutivo do algoritmo genético.

#### 4.7. Métricas

A previsão de medições é feita, principalmente, para perfis de poço de raio gama, assim como para os perfis sônicos, de indução e densidade. Isso complica um pouco a avaliação dos resultados porque há maior quantidade de características para serem consideradas. Além disso, cada tipo de medição está em torno de valores característicos de determinado tipo de rocha. Isso obriga a

utilização de mais de uma métrica para avaliar os resultados da previsão, porque algumas características não avaliadas por uma métrica podem ser avaliadas pelas outras. Dessa maneira é possível ter uma melhor perspectiva dos resultados da previsão.

Uma métrica simples e pontual é o erro absoluto médio percentual, MAPE (Chen & Yang, 2004) definido como:

$$MAPE = \frac{100}{k_{\max}} \sum_{k=1}^{k_{\max}} \left| \frac{\hat{r}_k - r_k}{r_k} \right|, \quad (4.24)$$

em que  $k_{\max}$  é o número de amostras do poço fictício,  $r_k$  é a amostra prevista (ou saída de um filtro) do perfil, e  $r_k$  é a amostra do perfil real ou padrão.

O MAPE é pontual porque calcula o erro em relação a uma amostra padrão. Logo, é tirada a média de todos os erros. Esta operação não é um problema para o conjunto de amostras com média fixa ou convergente (para um tipo rocha), calculada com a eq. (4.25):

$$m_r = \sum_{k=1}^{k_{\max}} \frac{\hat{r}_k}{k_{\max}}, \quad (4.25)$$

mas o perfil de raio gama (GR) tem média variável e dependente do tipo de material predominante nos estratos. Por exemplo, se o erro de previsão de medições de raio gama estiver num intervalo fixo, então os valores elevados de radiação gama, de materiais como o folhelho, produzirá um MAPE menor do que para materiais como o arenito. O cálculo do MAPE para o conjunto completo de amostras não é um indicador significativo desde o ponto de vista litológico.

Outra métrica afetada pela média é o desvio padrão, calculado como:

$$\sigma_r = \sqrt{k_{r,r}} \cong \sqrt{\frac{1}{k_{\max}} \sum_{k=1}^{k_{\max}} (\hat{r}_k - m_r)^2}, \quad (4.26)$$

em que  $k_{r,r}$  é a variância de um conjunto de medições de poço.

No histograma do perfil de raio gama é possível observar várias distribuições ou agrupamentos que indicam os diversos tipos de rocha, mas para simplificar o cálculo, com uma perda mínima de informação, agrupam-se em duas categorias principais: folhelho e arenito. O MAPE e as outras métricas são calculados para esses dois tipos de agrupamentos.

O coeficiente de correlação  $\rho_{r,y}$  da eq. (4.27), é uma métrica limitada a um intervalo e quantifica o relacionamento estatístico entre duas variáveis aleatórias

(Lira, 2004,Albuquerque et al., 2008), como as medições previstas e reais dos perfis de poço. O coeficiente de correlação entre dois sinais é calculado com a eq. (4.27):

$$\rho_{r,y} = \frac{k_{r,y}}{\sqrt{k_{r,r} k_{y,y}}} \quad 0 \leq |\rho_{r,y}| \leq 1, \quad (4.27)$$

em que  $k_{r,y}$  é a covariância de dois sinais  $r$  e  $y$ ;  $k_{r,r}$  e  $k_{y,y}$  são as variâncias de cada sinal. Quando o  $\rho_{r,y} = 0$ , não existe correlação linear entre as medições, mas quando  $\rho_{r,y} = 1$  existe correlação perfeita. Como  $\rho_{r,y}$  frequentemente toma valores intermediários, alguns autores definem intervalos para avaliá-lo qualitativamente. Por exemplo, para Callegari-Jacques (2003):

- Se  $0,0 < |\rho_{r,y}| < 0,3$ , então existe correlação linear fraca;
- Se  $0,3 < |\rho_{r,y}| < 0,6$ , então existe correlação linear moderada;
- Se  $0,6 < |\rho_{r,y}| < 0,9$ , então existe correlação linear forte;
- Se  $0,9 < |\rho_{r,y}| < 1,0$ , então existe correlação linear muito forte.

Um modelo multivariado de previsão, desde o ponto de vista estatístico, é visto como um modelo causal e, por isso, pode ser analisado através de procedimentos e métricas como os coeficientes regressão, coeficientes de correlação parcial, cargas fatoriais ou coeficientes canônicos (Kenny, 2004). Neste caso é utilizado o coeficiente de correlação parcial que permite conhecer o grau de correlação entre duas variáveis dado que a variação entre outras variáveis está controlada. O coeficiente de correlação parcial pode ter diversas ordens. Porém, apenas é necessário o coeficiente de primeira ordem, mostrada na eq. (4.28):

$$\rho_{r,y_z} = \frac{\rho_{r,y} - \rho_{r,z}\rho_{y,z}}{\sqrt{(1 - \rho_{r,z}^2)(1 - \rho_{y,z}^2)}}, \quad (4.28)$$

em que  $\rho_{r,y_z}$  é o coeficiente de correlação parcial entre as variáveis  $r$  e  $y$  controlando  $z$ ,  $\rho_{r,y}$  é o coeficiente de correlação entre as variáveis  $r$  e  $y$ ;  $\rho_{r,z}$  é o coeficiente de correlação entre as variáveis  $r$  e  $z$ ; e  $\rho_{y,z}$  é o coeficiente de correlação entre as variáveis  $y$  e  $z$ .

Durante os testes para a correlação lateral de poço se observou uma diferença considerável nos resultados quando foram incluídos perfis diferentes ao do raio gama. Esta observação produz uma pergunta: qual é a característica principal de um perfil para obter informação relevante para a previsão? Uma resposta intuitiva indicaria a maior variação (ou variabilidade) entre as medições

que caracterizam os materiais. Dessa maneira a estratigrafia e as propriedades geofísicas seriam realçadas.

Para comparar dois ou mais tipos diferentes de medições (ou variáveis aleatórias), de acordo com sua variabilidade, utiliza-se o coeficiente de variação de Pearson (Guimarães, 2008), que é dado pela eq. (4.29):

$$CV_r = \frac{\sigma_r}{m_r}. \quad (4.29)$$

Este coeficiente pode ser multiplicado por 100 para ser expresso em porcentagem. Observa-se que é uma medida relativa de variabilidade e é independente da natureza ou tipo de medição de poço.