6
Metodologia de inversão

Nesta tese, a transformação de velocidades em pressão de poros é encarada como um problema de inversão. Pela natureza do problema, essa transformação apresenta caráter não único (existem muitas soluções possíveis), o que acarreta em grandes incertezas nos valores invertidos. Devido à necessidade de uma boa análise de incertezas, e devido ao fato de se tratar de um problema inverso não linear, adota-se nesta tese uma metodologia de inversão baseada em simulações de Monte Carlo. (Para um debate sobre a análise de incertezas em problemas inversos, ver Mosegaard e Sambridge (2002)).

6.1.
Inversão Bayesiana

A teoria do problema inverso é a teoria matemática que descreve como se pode obter informações de parâmetros não observáveis de um sistema físico a partir de observações realizadas sobre esse sistema, de relações teóricas entre os parâmetros não observáveis e as observações, e de informações *a priori* sobre os parâmetros não observáveis.

Sendo $d = \{d_1, d_2, ..., d_n\}$ e $m = \{m_1, m_2, ..., m_n\}$ respectivamente os conjuntos de observações realizadas e os parâmetros não observáveis de um sistema físico, representa-se as relações teóricas que correlacionam essas grandezas por:

$$d = g(m) + \varepsilon , \tag{79}$$

onde o operador $g(.)$ representa a formulação do problema direto (modelo proposto para representar o fenômeno físico) e ε representa as incertezas associadas às observações realizadas e à formulação do modelo proposto.

Considere agora que toda a informação *a priori* (antes de se considerar as observações) sobre o conjunto de parâmetros não observáveis m seja probabilisticamente resumida através de uma função densidade de probabilidade $p(m)$. E considere também que essa informação possa ser aumentada observando-se d, que é relacionada com m através de uma função densidade de probabilidade condicional $p(d|m)$. O aumento de informação sobre m após a realização de uma observação d é bastante intuitivo e é quantificado nesta tese através do teorema de Bayes[8]:

$$p(m|d) = \frac{p(m)p(d|m)}{p(d)}, \tag{80}$$

onde $p(m|d)$ representa a informação sobre m após a observação d (dita informação *a posteriori*). Nota-se que o termo $p(d)$ na equação não depende de m, e é tido como um termo normalizador, que garante que a integral da função densidade de probabilidade $p(m|d)$ no espaço **M** (espaço relativo às grandeza não observáveis m) seja igual a uma unidade.

Para um valor fixo de d, define-se a função $L(m;d) = p(d|m)$, chamada de função de verossimilhança, que fornece a plausibilidade de valores de m, dada as observações realizadas d.

Observa-se que existem alguns tipos de função de verossimilhança, e que estas variam de acordo com as características das incertezas associadas às observações realizadas e também àquelas associadas à modelagem do problema direto (para diferentes tipos de função, ver Tarantola (2004)). No caso em que as incertezas associadas à modelagem do problema direto são desconsideradas, e que as incertezas associadas às observações são representadas por distribuições Gaussianas centradas nos valores observados d, tem-se:

---

[8] É interessante notar que, a partir da conjunção de estados de informação sobre um sistema físico qualquer, Tarantola (2004) demonstra que a solução geral de um problema inverso é dada por uma expressão equivalente ao teorema de Bayes.

$$L(m;d) = k' \exp\left[-\frac{1}{2}\left(g(m)-d\right)^t Cov^{-1}\left(g(m)-d\right)\right], \tag{81}$$

onde Cov é a matriz de covariância, que descreve as incertezas associadas às observações (incertezas que podem ser independentes ou não) e k' é uma constante de normalização.

Seguindo essas definições, reescreve-se o teorema de Bayes através de sua forma não normalizada:

$$p(m|d) \propto p(m)L(m;d) \; . \tag{82}$$

Note que ao omitir o termo $p(d)$, a igualdade na equação (80) foi substituída por uma proporcionalidade. O lado direito da equação (82), ($p(m)L(m;d)$), representa $p(m|d)$ de uma forma não normalizada.

Ilustra-se a seguir o processo de aumento de informação obtido através da aplicação do teorema de Bayes (Figura 18). Considere um simples problema onde o sistema físico é constituído por duas grandezas não observáveis $m = \{m_1, m_2\}$, e que seja possível a realização de uma observação de uma outra grandeza relacionada com m, denotada por $d = \{d_1\}$. Considere que as informações *a priori* sobre m sejam quantificadas pela distribuição de probabilidades *a priori* (representada no lado esquerdo da Figura 18). E considere também a plausibilidade dos valores de m, dada a observação d, que é avaliada pela função de verossimilhança (representação ao centro da Figura 18). Através do teorema de Bayes, é possível unir essas duas informações, obtendo a distribuição de probabilidades *a posteriori* de m (representada no lado direito da Figura 18). Ao se comparar as informações *a priori* e *a posteriori* sobre m, compreende-se o mecanismo de aumento de informação dado pela inferência bayesiana.

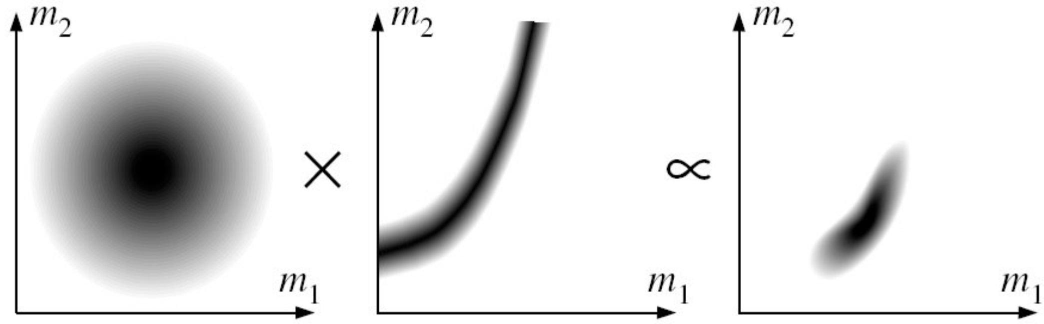

Figura 18. Representação do aumento de informação através do teorema de Bayes (Malinverno e Torres-Verdin, 2000).

6.2. Solução através de Simulações de Monte Carlo

Nesta tese, a solução do problema inverso, $p(m|d)$, é obtida através de uma técnica chamada de algoritmo de Metropolis-Hastings, que permite amostrar aleatoriamente a distribuição de probabilidades representada por $p(m|d)$ (Tarantola, 2004). Antes de se abordar o problema propriamente dito, é didática a apresentação desta técnica fora do contexto do problema inverso.

6.2.1. Algoritmo de Metropolis-Hastings (M-H)

O algoritmo de Metropolis-Hastings é um método de simulação estocástica que permite amostrar aleatoriamente qualquer distribuição de probabilidades, aqui representada pela função $f(x)$. Essencialmente, o algoritmo é composto por uma cadeia de Markov definida no mesmo suporte de $f(x)$ (ou seja, no espaço amostral X) e por um critério de aceitação ou rejeição das transições propostas pela cadeia de Markov (Gilks et al., 1998).

Por cadeia de Markov (de primeira ordem), entende-se o processo estocástico que simula uma seqüência de amostras aleatórias em um espaço, de modo que um estágio futuro dependa somente do estágio atual da cadeia, e não dos estágios percorridos pela cadeia no passado. O comportamento da seqüência aleatória é determinado por uma distribuição de probabilidades condicionada, também chamada de distribuição de probabilidades de transição, que determina a probabilidade de ocorrência de valores em X, dado o estágio atual da cadeia. A notação utilizada para descrever essa distribuição é $q(x_{i+1}|x_i)$, que representa a probabilidade de ocorrer a transição do estagio x_i para o estagio x_{i+1}.

Suponha que $f(x)$ represente a distribuição da qual se quer obter amostras aleatórias. Suponha também que se tenha uma função geradora de amostras candidatas, cujo comportamento segue $q(x_{i+1}|x_i)$. O algoritmo de Metropolis-Hastings é então dado por:

1. Considere a cadeia iniciada em $x_{i=1}$.
2. Repetir:

- Gerar um candidato x_{i+1} usando $q(x_{i+1}|x_i)$;
- Avaliar a probabilidade de aceitação do movimento proposto através de:

$$\tau = min\left\{1, \frac{f(x_{i+1})q(x_i|x_{i+1})}{f(x_i)q(x_{i+1}|x_i)}\right\}; \tag{83}$$

- Gerar u aleatoriamente dentro da distribuição uniforme U (0,1);
- Se $\tau \geq u$, aceitar x_{i+1} como amostra de $f(x)$, e fazer $x_i = x_{i+1}$;
- Senão, rejeitar a proposta de movimento e descartar x_{i+1}.

Em outras palavras, em cada estágio da cadeia, um novo candidato é proposto de acordo com a função geradora. Se este candidato é aceito, entende-se que a cadeia anda (ocorre a transição), senão, a cadeia fica parada e o candidato é descartado. Pode-se demonstrar que a coleção de amostras aceitas tem uma distribuição de freqüências que é (assintoticamente) proporcional à $f(x)$ (Chib e Greenberg, 1995).

Nota-se que a função de interesse $f(x)$ entra na probabilidade de aceitação (equação (83)) como uma razão $\left(f(x_{i+1})/f(x_i)\right)$. Sendo assim, o algoritmo M-H pode também ser utilizado quando $f(x)$é conhecida somente sobre uma constante de normalização (não é preciso conhecer $f(x)$ na forma normalizada). Observe que esta característica torna o algoritmo muito útil em um contexto de inferência Bayesiana, pois permite o uso do teorema de Bayes em sua forma não normalizada (equação (82)).

6.2.1.1.
Caso de cadeias independentes

Considere agora um esquema especial do algoritmo M-H, que é caracterizado pelo uso de cadeias independentes. Ou seja, a proposta de transição $(x_i \rightarrow x_{i+1})$ é feita independentemente da posição anterior da cadeia, e portanto $q(x_{i+1}|x_i)$ é dada por uma função não condicionada, denominada $h(x)$ (da qual se sabe gerar amostras aleatórias, é claro). Neste caso, o algoritmo de M-H é utilizado com a seguinte probabilidade de aceitação (τ):

$$\tau = min\left\{1, \frac{f(x_{i+1})h(x_i)}{f(x_i)h(x_{i+1})}\right\} \quad . \tag{84}$$

Em outras palavras, amostras candidatas são propostas a partir de uma distribuição geradora (representada por $h(x)$), e através da regra de aceitação do algoritmo, decide-se se estas são amostras aleatórias da distribuição alvo (representada por $f(x)$). Observa-se que a eficiência deste algoritmo aumenta (maior taxa de aceitação de transições propostas) ao se escolher uma $h(x)$ que seja o mais semelhante possível à $f(x)$.

6.2.2.
O algoritmo M-H no problema inverso

Baseando-se no fato do algoritmo M-H ser capaz de amostrar aleatoriamente qualquer distribuição de probabilidades, este vem sendo utilizado para obter a solução de problemas inversos. Exemplos de aplicação do algoritmo M-H em conjunto com o teorema de Bayes em problemas inversos podem ser encontrados em Malinverno e Torres-Verdín (2000), Mosegaard e Tarantola (1995), Pedersen e Knudsen (1990) e Koren et al. (1991).

No caso de uma abordagem bayesiana de inversão, a distribuição alvo é dada por $p(m|d)$, conforme definição dada na equação (80). Para obter amostras aleatórias de $p(m|d)$, uma apropriada cadeia de Markov é definida no espaço $\mathbf{M}$ (espaço das grandezas não observáveis). Esta cadeia propõe amostras aleatórias de m, que são ou não aceitas pelo algoritmo. A coleção de amostras aceitas tem uma distribuição de freqüências que é (assintoticamente) proporcional à $p(m|d)$.

A função $p(m|d)$, por sua vez, representa a solução completa do problema inverso, ou seja, representa toda a informação que se pode inferir sobre os parâmetros não-observáveis do sistema. Qualquer inferência sobre m, como estimativa de valores ótimos e quantificação de incertezas é realizada através da inspeção de $p(m|d)$.

Para ilustrar o procedimento de obtenção da solução do problema inverso através do algoritmo M-H, recorre-se à Figura 19, que se refere ao exemplo descrito na Figura 18. Os pontos brancos representam a amostragem aleatória da solução do problema inverso através do algoritmo M-H.

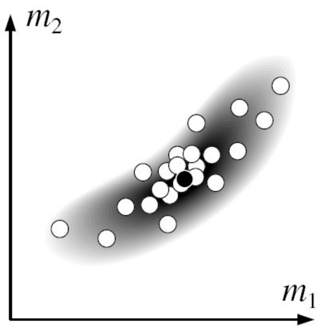

Figura 19. Amostragem da solução do problema inverso através de simulações de Monte Carlo (Figura original em Malinverno e Torres-Verdin, 2000).

Observa-se que a possibilidade da definição de uma cadeia de Markov é muito propícia para certos tipos de problema inverso, onde existem grandes proporções do espaço amostral $\mathbf{M}$ que não correspondem à solução do problema (ou seja, existe uma grande região compatível com as informações *a priori*, mas que não corresponde às observações realizadas). Neste caso, a definição de uma cadeia de Markov permite que, após a localização da "zona de interesse" (zona correspondente às soluções do problema), que esta se fixe nesta região, evitando uma amostragem desnecessária do espaço "vazio". Em problemas inversos onde a maior parte do espaço amostral é "vazia", este procedimento é essencial para garantir a convergência do algoritmo.

No caso do problema abordado nesta tese, não existe uma grande proporção de "regiões vazias" no espaço amostral **M**, o que permite o uso do esquema de cadeias independentes sem maiores comprometimentos com a eficiência do algoritmo. Observa-se que fazendo a distribuição geradora (representada por $h(x)$) ser a distribuição *a priori* (representada por $p(m)$), é possível reescrever o algoritmo M-H para um contexto de inferência bayesiana de uma forma muito conveniente. Eis o algoritmo utilizado nesta tese:

1. Iniciar o algoritmo gerando $m_{i=1}$ através de $p(m)$;
2. Repetir:

- Gerar um candidato m_{i+1} usando $p(m)$;
- Avaliar a probabilidade de aceitação do movimento proposto através de:

$$\tau = min\left\{1, \frac{L(m_{i+1};d)}{L(m_i;d)}\right\}; \tag{85}$$

- Gerar u aleatoriamente dentro da distribuição uniforme U (0,1);
- Se $\tau \geq u$, aceitar m_{i+1} como amostra de $p(m|d)$, e fazer $m_i = m_{i+1}$;
- Senão, rejeitar a proposta de movimento e descartar m_{i+1}.

Em outras palavras, gera-se uma amostra candidata de acordo com a distribuição *a priori*, e decide-se, através do algoritmo de M-H, se esta é aceita ou não como amostra aleatória da distribuição *a posteriori*. As amostras aceitas constituem amostras aleatórias da solução do problema inverso.

Observa-se que através deste algoritmo, não é preciso conhecer a função $p(m|d)$ de uma forma normalizada, e que também não é necessária a avaliação da constante de normalização k', referente à equação ((81)). Esse modo de uso o algoritmo de Metropolis-Hastings vem sendo chamado de *Metropolized Independence*

Sampler (MIS) na literatura. Para maiores detalhes, Pursiainem (2003) e Zhai e Yeary (2004).