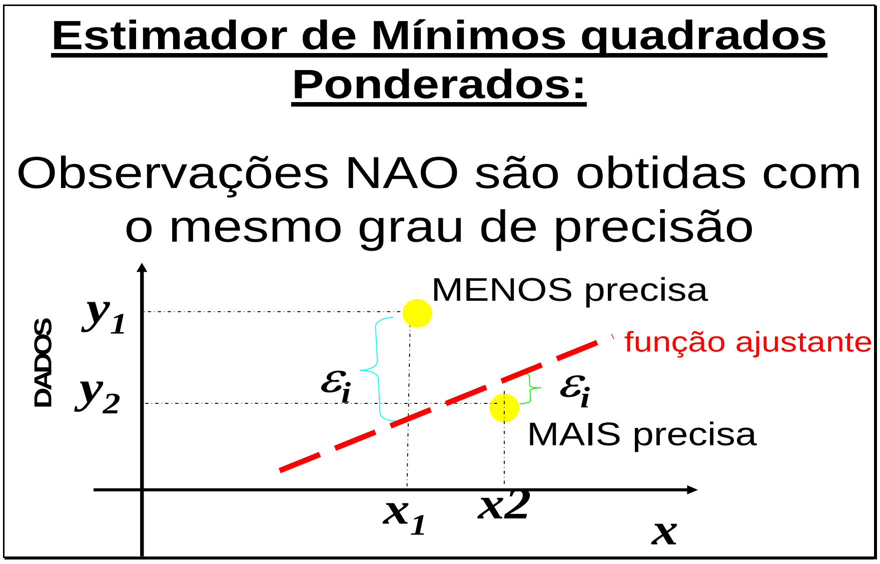
INVERSÃO LINEAR:

ESTIMADOR DE MÍNIMOS QUADRADOS PONDERADOS

Muitas vezes as observações geofísicas não são obtidas com o mesmo grau de precisão, sendo umas observações mais precisas que outras. Neste caso não é razoável exigir que a função ajustante produza a mesma proximidade a todas as observações. É mais razoável exigir um ajuste "melhor" nas observações geofísicas mais precisas e um ajuste "pior" nas observações geofísicas menos precisas (Figura 1).



Matematicamente, ao invés de minimizarmos a soma dos quadrados dos resíduos, i.e.,

$$\min_{\bar{\mathbf{p}}\in F}\{Q\}\equiv\min_{\bar{\mathbf{p}}\in F}\{\|\boldsymbol{\varepsilon}\|_2^2\}\equiv\min_{\bar{\mathbf{p}}\in F}\left\|\bar{\mathbf{y}}^{\mathbf{o}}-\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}}\right\|_2^2\equiv\min_{\bar{\mathbf{p}}\in F}\left(\bar{\mathbf{y}}^{\mathbf{o}}-\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}}\right)^{\mathrm{T}}\left(\bar{\mathbf{y}}^{\mathbf{o}}-\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}}\right)$$

minimizaremos a soma dos quadrados ponderados dos resíduos

$$\min_{\bar{\mathbf{p}}\in F}\{Q\}\equiv\min_{\bar{\mathbf{p}}\in F}\{\sum_{i=1}^{N}\varepsilon_i^2 w_i\}\equiv\min_{\bar{\mathbf{p}}\in F}\sum_{i=1}^{N}\left[y^o{}_i-y^c{}_i\right]^2 w_i$$

Em notação matricial temos:

$$\min_{\bar{\mathbf{p}} \in F} \{Q\} \equiv \min_{\bar{\mathbf{p}} \in F} \left\{ \bar{\boldsymbol{\varepsilon}}^T \,\overline{\overline{\mathbf{W}}}\, \bar{\boldsymbol{\varepsilon}} \right\} \equiv \min_{\bar{\mathbf{p}} \in F} \left\{ \left\| \overline{\overline{\mathbf{W}}}^{1/2} \bar{\boldsymbol{\varepsilon}} \right\|_2^2 \right\}$$

em que $\overline{\overline{\mathbf{W}}} \in R^{N \times N}$ é uma matriz diagonal de pesos, de modo que o i-ésimo elemento do resíduo (ε_i) será ponderado pelo i-ésimo elemento da diagonal de $\overline{\overline{\mathbf{W}}}$

$$Q = \varepsilon_1^2 w_1 + \varepsilon_2^2 w_2 + \varepsilon_3^2 w_3 + \ldots + \varepsilon_N^2 w_N$$

O método dos Mínimos Quadrados Ponderados (MQP) impõe que cada elemento do vetor das observações calculadas ou ajustadas ($\bar{\mathbf{y}}^{\mathbf{c}}$) deve estar o próximo ao correspondente elemento do vetor observado ($\bar{\mathbf{y}}^{\mathbf{o}}$), sendo que o grau desta proximidade é controlado pela escolha dos pesos w_i .

Se todos os elementos que compõe o vetor $\bar{\mathbf{y}}^{\mathbf{c}}$ (anomalia ajustada) devem estar igualmente próximos aos elementos que compõe $\bar{\mathbf{y}}^{\mathbf{o}}$ (dados observados) então, neste caso, a matriz de peso será a matriz identidade $\overline{\overline{\mathbf{W}}} = \overline{\overline{\mathbf{I}}}$ e o método se reduz aos Mínimos Quadrados (MQ) já estudado anteriormente.

$$\min_{\bar{\mathbf{p}} \in F} \quad Q = \bar{\boldsymbol{\varepsilon}}^T \,\overline{\overline{\mathbf{W}}}\, \bar{\boldsymbol{\varepsilon}}$$

$$\min_{\bar{\mathbf{p}} \in F} \quad Q = \left(\bar{\mathbf{y}}^{\mathbf{o}} - \bar{\mathbf{y}}^{c} \right)^{\mathbf{T}} \overline{\overline{\mathbf{W}}} \left(\bar{\mathbf{y}}^{\mathbf{o}} - \bar{\mathbf{y}}^{c} \right)$$

como

$$\bar{\mathbf{y}}^{c} = \overline{\overline{\mathbf{A}}}\, \bar{\mathbf{p}}$$

então temos

$$\min_{\bar{\mathbf{p}} \in F} \; Q = \left(\bar{\mathbf{y}}^{\mathbf{o}} - \overline{\overline{\mathbf{A}}}\, \bar{\mathbf{p}} \right)^{\mathbf{T}} \overline{\overline{\mathbf{W}}} \left(\bar{\mathbf{y}}^{\mathbf{o}} - \overline{\overline{\mathbf{A}}}\, \bar{\mathbf{p}} \right)$$

A condição necessária para que Q tenha mínimo é

$$\overline{\nabla}_{\overline{\mathbf{p}}}\{Q\} = \overline{\mathbf{0}}$$

$$\overline{\nabla}_{\overline{\mathbf{p}}}\{Q\} = \overline{\nabla}_{\overline{\mathbf{p}}}\left\{\left(\overline{\mathbf{y}}^{\mathbf{o}} - \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}}\right)^{\mathrm{T}} \overline{\overline{\mathbf{W}}}\left(\overline{\mathbf{y}}^{\mathbf{o}} - \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}}\right)\right\} = \overline{\mathbf{0}}$$

Para o cálculo do vetor gradiente acima, veja na revisão de álgebra linear o item gradiente de uma forma quadrática do tipo $Q = \overline{\mathbf{x}}^{T}\,\overline{\overline{\mathbf{A}}}\;\overline{\mathbf{x}}$.

$$\overline{\nabla}_{\overline{\mathbf{p}}}\{Q\} = 2\,\overline{\nabla}_{\overline{\mathbf{p}}}\left\{\left(\overline{\mathbf{y}}^{\mathbf{o}} - \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}}\right)^{\mathrm{T}}\right\} \overline{\overline{\mathbf{W}}}\left(\overline{\mathbf{y}}^{\mathbf{o}} - \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}}\right) = \overline{\mathbf{0}}$$

$$\overline{\nabla}_{\overline{\mathbf{p}}}\{Q\} = 2\left(\overline{\nabla}_{\overline{\mathbf{p}}}(\overline{\mathbf{y}}^{\mathbf{o}\,T}) - \overline{\nabla}_{\overline{\mathbf{p}}}(\overline{\mathbf{p}}^{T}\,\overline{\overline{\mathbf{A}}}^{T})\right)\overline{\overline{\mathbf{W}}}\left(\overline{\mathbf{y}}^{\mathbf{o}} - \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}}\right) = \overline{\mathbf{0}}$$

como $\overline{\nabla}_{\overline{\mathbf{p}}}(\overline{\mathbf{y}}^{\mathbf{o}\,T}) = \overline{\mathbf{0}}$, então temos

$$\overline{\nabla}_{\overline{\mathbf{p}}}\{Q\} = -\,2\left(\overline{\nabla}_{\overline{\mathbf{p}}}(\overline{\mathbf{p}}^{T}\,\overline{\overline{\mathbf{A}}}^{T})\right)\overline{\overline{\mathbf{W}}}\left(\overline{\mathbf{y}}^{\mathbf{o}} - \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}}\right) = \overline{\mathbf{0}}$$

como $\overline{\nabla}_{\overline{\mathbf{p}}}(\overline{\mathbf{p}}^{T}) = \overline{\overline{\mathbf{I}}}$, então temos

$$\overline{\nabla}_{\overline{\mathbf{p}}}\{Q\} = -\,2\,\overline{\overline{\mathbf{A}}}^{T}\,\overline{\overline{\mathbf{W}}}\left(\overline{\mathbf{y}}^{\mathbf{o}} - \overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}}\right) = \overline{\mathbf{0}}$$

$$-\,2\,\overline{\overline{\mathbf{A}}}^{T}\,\overline{\overline{\mathbf{W}}}\,\overline{\mathbf{y}}^{\mathbf{o}} + 2\,\overline{\overline{\mathbf{A}}}^{T}\,\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} = \overline{\mathbf{0}}$$

$$\overline{\overline{\mathbf{A}}}^{T}\,\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} = \overline{\overline{\mathbf{A}}}^{T}\,\overline{\overline{\mathbf{W}}}\,\overline{\mathbf{y}}^{\mathbf{o}}$$

Resultando no estimador dos Mínimos Quadrados Ponderados (MQP)

$$\hat{\overline{\mathbf{p}}} = \left(\overline{\overline{\mathbf{A}}}^{T}\,\overline{\overline{\mathbf{W}}}\;\overline{\overline{\mathbf{A}}}\right)^{-1}\overline{\overline{\mathbf{A}}}^{T}\,\overline{\overline{\mathbf{W}}}\,\overline{\mathbf{y}}^{\mathbf{o}}$$

Análise da Unicidade da solução estimada via MQP:

A solução $\hat{\mathbf{p}}$ só será única se a matriz a ser invertida ($\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}$) for NÃO SINGULAR. Um meio de avaliarmos se a matriz é não singular é através do seu determinante. Se o determinante a matriz $\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}$ é diferente de zero ($\mathrm{Det}(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}) \neq 0$) esta matriz é não singular. Um segundo meio de avaliarmos se uma matriz é não singular é através do seu posto[1]. Como a matriz de sensibilidade $\overline{\overline{\mathbf{A}}}$ é N x M, o produto $\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}$ é M x M então, a matriz $\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}$ deve ter posto M (número de parâmetros) para ser uma matriz não singular ($\mathrm{r}(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}) = \mathrm{M}$). Tal como no estimador MQ, para usarmos o estimador de mínimos quadrados ponderados é necessário, embora não suficiente que N $\geq$ M, isto é que o número de observações (N) seja pelo menos igual ao número de parâmetros (M). Para garantirmos a unicidade da solução estimada via estimador MQP é necessário e suficiente que N $\geq$ M e o posto da matriz $\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}$ seja igual a M ($\mathrm{r}(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}) = \mathrm{M}$).

Caso contrário, se ($\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}$) for singular então, $\hat{\mathbf{p}} \equiv \overline{\mathbf{0}}$ não é a única solução da equação normal $\overline{\overline{\mathbf{A}}}^T\,\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\,\overline{\mathbf{p}} = \overline{\overline{\mathbf{A}}}^T\,\overline{\overline{\mathbf{W}}}\,\overline{\mathbf{y}}^{\mathbf{o}}$ e a condição mencionada anteriormente para a unicidade da solução não é satisfeita. Neste caso em que não há unicidade da solução a matriz $\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}$ é singular, então o seu determinante é igual a zero ($\mathrm{Det}(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}) = 0$) e o seu posto é menor que o número de parâmetros a serem estimados ($\mathrm{r}(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}) < \mathrm{M}$).

Vamos supor que a matriz $\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}$ é não singular, portanto garantimos a unicidade da solução. Já falamos anteriormente, que a unicidade da solução é uma condição necessária, mas não é suficiente para afirmarmos que estamos resolvendo um problema bem-posto. Falta analisar a estabilidade da solução estimada que estudaremos aqui através de uma análise estatística.

[1] posto de uma matriz é o número de vetores colunas (ou vetores linhas) que formam um conjunto LI de vetores

Análise da Estabilidade da solução MQP via análise estatística:

Como já abordamos anteriormente, o estudo da estabilidade da solução pode ser realizado através da variância dos parâmetros, uma vez que o estimador $\hat{\mathbf{p}}$ é uma variável aleatória. Vamos analisar se o estimador $\hat{\bar{\mathbf{p}}}$ apresenta variância mínima. Lembramos que, à luz da estatística um bom estimador $\hat{\bar{\mathbf{p}}}$ deve ser não tendencioso e de variância mínima. Então vamos calcular a esperança e variância do estimador de mínimos quadrados $\hat{\mathbf{p}}$. Para tanto precisamos estabelecer alguma premissas estatísticas sobre o ruído que contamina os dados (veja Anexo 1 do tópico 3) e também sobre as demais variáveis do modelo matemático.

(1) Esperança do estimador MQP ($\hat{\mathbf{p}}$) :

Considerando as premissas estatísticas estabelecidas na Figura 2 e as propriedades da esperança, temos que

$$\mathrm{E}\left\{ \hat{\bar{\mathbf{p}}} \right\} = \mathrm{E}\left\{ \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}} \bar{\bar{\mathbf{A}}} \right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}} \bar{\mathbf{y}}^{\mathbf{o}} \right\}$$

Premissas estatísticas sobre o ruído que contamina os dados geofísicos e outras variáveis para calcularmos:
a Esperança do estimador MQP
(Mínimos Quadrados Ponderados)

Esperança :

Premissas estatísticas 1 1 — — — — 1 1

Erro aditivo

Erro com média nula

variáveis independentes sem erros

parâmetros não são aleatórios

Figura 2

Usando a premissa 1 que os erros são aditivos ($\bar{\mathbf{y}}^o = \bar{\mathbf{y}}^c + \bar{\boldsymbol{\varepsilon}}$) e usando a informação que em um problema linear a componente determinística (vetor dos dados ajustados ou calculados) é $\bar{\mathbf{y}}^C = \bar{\bar{\mathbf{A}}}\bar{\mathbf{p}}$ temos que

$$E\left\{\hat{\bar{\mathbf{p}}}\right\} = E\left\{\left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,(\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} + \bar{\boldsymbol{\varepsilon}})\right\}$$

$$E\left\{\hat{\bar{\mathbf{p}}}\right\} = E\left\{\left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \overline{\overline{\mathbf{W}\,\mathbf{A}}}\,\bar{\mathbf{p}} + \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\boldsymbol{\varepsilon}}\right\}$$

como $\left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \overline{\overline{\mathbf{W}\,\mathbf{A}}} = \bar{\bar{\mathbf{I}}}$

$$E\left\{\hat{\bar{\mathbf{p}}}\right\} = E\left\{\bar{\mathbf{p}}\right\} + E\left\{\left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\boldsymbol{\varepsilon}}\right\}$$

Usando a premissa 7 que as variáveis independentes não são v.a. temos,

$$E\left\{\hat{\bar{\mathbf{p}}}\right\} = E\left\{\bar{\mathbf{p}}\right\} + \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\; E\left\{\bar{\boldsymbol{\varepsilon}}\right\}$$

Pela premissa 2 os erros tem média nula o que implica $E[\bar{\boldsymbol{\varepsilon}}] = \bar{0}$ então

$$E\left\{\hat{\bar{\mathbf{p}}}\right\} = E\left\{\bar{\mathbf{p}}\right\}$$

Usando a premissa 7 que os parâmetros não são aleatórios temos

$$E\left\{\hat{\bar{\mathbf{p}}}\right\} = \bar{\mathbf{p}}$$

Vemos então que o estimador MQP ($\hat{\bar{\mathbf{p}}}$) é NÃO TENDENCIOSO. De fato, o estimador MQP não introduz nenhum tipo de informação a priori sobre os parâmetros a serem estimados. A informação introduzida é no espaço das observações.

2) Matriz de covariância do estimador MQP ($\hat{\bar{\mathbf{p}}}$):

Considerando as premissas estatísticas estabelecidas na Figura 3, as propriedades da variância e partindo-se da equação da matriz de covariância de um vetor (Anexo 1) temos que

$$\mathbf{COV}\left\{\hat{\bar{\mathbf{p}}}\right\} = E\left\{\left[\hat{\bar{\mathbf{p}}} - E(\hat{\bar{\mathbf{p}}})\right]\left[\hat{\bar{\mathbf{p}}} - E(\hat{\bar{\mathbf{p}}})\right]^T\right\}$$

Premissas estatísticas sobre o ruído que contamina os dados geofísicos e outras variáveis para calcularmos:
a Variância do estimador MQP
(Mínimos Quadrados Ponderados)

Variância :
Premissas estatísticas 1 1 — 1 — — 1 1

Erro aditivo
Erro com média nula
Erros não correlacionáveis
variáveis independentes sem erros
parâmetros não são aleatórios

Figura 3

Usando o resultado anterior e $E\left\{\hat{\bar{\mathbf{p}}}\right\} = \bar{\mathbf{p}}$

$$\text{cov}\left\{\hat{\bar{\mathbf{p}}}\right\} = E\left\{\left[\hat{\bar{\mathbf{p}}} - \bar{\mathbf{p}}\right]\left[\hat{\bar{\mathbf{p}}} - \bar{\mathbf{p}}\right]^T\right\}$$

Usando a premissa 1 que os erros são aditivos ($\bar{\mathbf{y}}^O = \bar{\mathbf{y}}^C + \bar{\boldsymbol{\varepsilon}}$) e a informação que o problema é linear i.e., $\bar{\mathbf{y}}^C = \bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}}$ então temos que

$$\hat{\bar{\mathbf{p}}} = \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\mathbf{y}}^{\mathbf{o}}$$

$$\hat{\bar{\mathbf{p}}} = \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,(\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} + \bar{\boldsymbol{\varepsilon}})$$

$$\hat{\bar{\mathbf{p}}} = \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\,\bar{\mathbf{p}} + \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\boldsymbol{\varepsilon}}$$

como $\left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}} = \bar{\bar{\mathbf{I}}}$

$$\hat{\bar{\mathbf{p}}} = \bar{\mathbf{p}} + \left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}\,\bar{\boldsymbol{\varepsilon}}$$

substituindo o estimador acima na equação da covariância temos que

$$\operatorname{cov}\left\{\hat{\overline{\mathbf{p}}}\right\} = E\left\{\left[\overline{\mathbf{p}} + \left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\overline{\boldsymbol{\varepsilon}} - \overline{\mathbf{p}}\right]\left[\overline{\mathbf{p}} + \left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\overline{\boldsymbol{\varepsilon}} - \overline{\mathbf{p}}\right]^T\right\}$$

$$\mathrm{COV}\left\{\hat{\overline{\mathbf{p}}}\right\} = E\left\{\left[\left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\overline{\boldsymbol{\varepsilon}}\right]\left[\left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\overline{\boldsymbol{\varepsilon}}\right]^T\right\}$$

Usando propriedades da transposta temos

$$\mathrm{COV}\left\{\hat{\overline{\mathbf{p}}}\right\} = E\left\{\left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\overline{\boldsymbol{\varepsilon}}\,\overline{\boldsymbol{\varepsilon}}^T\,\overline{\overline{\mathbf{W}}}^T\,\overline{\overline{\mathbf{A}}}\left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}\right\}$$

Usando a premissa 7 que as variáveis independentes não são v.a. temos,

$$\mathrm{COV}\left\{\hat{\overline{\mathbf{p}}}\right\} = \left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}E\left\{\overline{\boldsymbol{\varepsilon}}\,\overline{\boldsymbol{\varepsilon}}^T\right\}\overline{\overline{\mathbf{W}}}^T\,\overline{\overline{\mathbf{A}}}\left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}$$

Temos que calcular a esperança de $E\left\{\overline{\boldsymbol{\varepsilon}}\overline{\boldsymbol{\varepsilon}}^T\right\}$ que nada mais é que a esperança de uma forma quadrática (veja <u>Apêndice - Matriz de Covariância dos Erros)</u> que é expressa como

$$E\left\{\overline{\boldsymbol{\varepsilon}}\,\overline{\boldsymbol{\varepsilon}}^T\right\} = \operatorname{cov}(\overline{\boldsymbol{\varepsilon}}) + E(\overline{\boldsymbol{\varepsilon}})E(\overline{\boldsymbol{\varepsilon}}^T)$$

Pela premissa 2 os erros tem média nula o que implica $E[\overline{\boldsymbol{\varepsilon}}] = \overline{0}$ então

$$E\left\{\overline{\boldsymbol{\varepsilon}}\,\overline{\boldsymbol{\varepsilon}}^T\right\} = \operatorname{cov}(\overline{\boldsymbol{\varepsilon}})$$

Substituindo o resultado acima na expressão obtemos a expressão da matriz de covariância do estimador MQP, i.e,

$$\mathrm{COV}\left\{\hat{\overline{\mathbf{p}}}\right\} = \left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\operatorname{cov}\left\{\overline{\boldsymbol{\varepsilon}}\right\}\overline{\overline{\mathbf{W}}}^T\,\overline{\overline{\mathbf{A}}}\left(\overline{\overline{\mathbf{A}}}^T\overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}$$

Usando a premissa 4 que os erros não são correlacionáveis ($\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$, $\forall\ i \neq j$), isto significa dizer que a matriz de covariância dos erros

$\left[\mathrm{cov}(\bar{\boldsymbol{\varepsilon}})\right]$ é uma matriz diagonal cujo i-ésimo elemento da diagonal é a variância σ_i^2 do i-ésimo elemento do vetor dos resíduos ($\bar{\boldsymbol{\varepsilon}}$) então temos

$$\mathrm{COV}\{\bar{\varepsilon}\} = \begin{bmatrix} \sigma_1^2 & & & \\ & {\sigma_2}^2 & & \\ & & \ddots & \\ & & & \sigma^2{}_N \end{bmatrix}_{N \times N}$$

2.1) Matriz de covariância do estimador MQP ($\hat{\mathbf{p}}$) para um caso particular:

Vamos considerar o caso particular em que o i-ésimo elemento da matriz de peso $\overline{\overline{\mathbf{W}}}$ é definido como:

$$w_i = \frac{1}{\sigma_i^2}$$

Em outras palavras, a matriz de peso $\overline{\overline{\mathbf{W}}}$ foi definida como a inversa da matriz de covariância dos erros $\left[\mathrm{cov}(\bar{\boldsymbol{\varepsilon}})\right]$. Então podemos dizer que

$$\overline{\overline{\mathbf{W}}} = \left(\mathrm{COV}\{\bar{\varepsilon}\}\right)^{-1}$$

Veja que podemos dizer que

$$\mathrm{COV}\{\bar{\varepsilon}\} = \overline{\overline{\mathbf{W}}}^{\,-1}$$

Portanto, neste caso particular a matriz de covariância dos parâmetros do estimador $\hat{\mathbf{p}}$ dos MQP

$$\mathrm{COV}\left\{\hat{\mathbf{p}}\right\} = \left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1} \overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{W}}}\,\mathrm{cov}\left\{\bar{\boldsymbol{\varepsilon}}\right\} \overline{\overline{\mathbf{W}}}^T\,\overline{\overline{\mathbf{A}}}\left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}$$

será simplificada:

$$\mathrm{COV}\left\{\hat{\mathbf{p}}\right\} = \left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1} \overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{W}}}\left(\overline{\overline{\mathbf{W}}}^{-1}\right) \overline{\overline{\mathbf{W}}}^T\,\overline{\overline{\mathbf{A}}}\left(\overline{\overline{\mathbf{A}}}^T \overline{\overline{\mathbf{W}}}\,\overline{\overline{\mathbf{A}}}\right)^{-1}$$

como $\bar{\bar{\mathbf{W}}}\left(\bar{\bar{\mathbf{W}}}\right)^{-1} = \bar{\bar{\mathbf{I}}}$

$$\mathrm{COV}\left\{\hat{\bar{\mathbf{p}}}\right\} = \left(\bar{\bar{\mathbf{A}}}^T\bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1}\bar{\bar{\mathbf{A}}}^T\,\bar{\bar{\mathbf{W}}}^T\,\bar{\bar{\mathbf{A}}}\left(\bar{\bar{\mathbf{A}}}^T\bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1}$$

como $\left(\bar{\bar{\mathbf{A}}}^T\,\bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1}\bar{\bar{\mathbf{A}}}^T\,\bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}} = \bar{\bar{\mathbf{I}}}$ temos

$$\mathrm{COV}\left\{\hat{\bar{\mathbf{p}}}\right\} = \left(\bar{\bar{\mathbf{A}}}^T\bar{\bar{\mathbf{W}}}\,\bar{\bar{\mathbf{A}}}\right)^{-1}$$

Como estamos no caso particular que $\bar{\bar{\mathbf{W}}} = \left(\mathrm{COV}\{\bar{\varepsilon}\}\right)^{-1}$ temos

$$\mathrm{COV}\left\{\hat{\bar{\mathbf{p}}}\right\} = \left(\bar{\bar{\mathbf{A}}}^T\left[\mathrm{COV}\{\bar{\varepsilon}\}\right]^{-1}\bar{\bar{\mathbf{A}}}\right)^{-1}$$

Veja que se considerarmos que a variância dos erros contidos nas observações geofísicas (incerteza dos dados) é a mesma em todas as observações, i.e., $\sigma_i^2 = \sigma^2, \ \forall\ i = 1,\ldots, N$ então temos que

$$\mathrm{COV}\{\bar{\varepsilon}\} = \begin{bmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2 \end{bmatrix} = \sigma^2\bar{\bar{\mathbf{I}}}$$

Substituindo $\mathrm{COV}\{\bar{\varepsilon}\} = \sigma^2\bar{\bar{\mathbf{I}}}$ temos

$$\mathrm{COV}\left\{\hat{\bar{\mathbf{p}}}\right\} = \left(\bar{\bar{\mathbf{A}}}^T\left[\sigma^2\bar{\bar{\mathbf{I}}}\right]^{-1}\bar{\bar{\mathbf{A}}}\right)^{-1}$$

$$\mathrm{COV}\left\{\hat{\bar{\mathbf{p}}}\right\} = \left(\bar{\bar{\mathbf{A}}}^T\left(\frac{1}{\sigma^2}\bar{\bar{\mathbf{I}}}\right)\bar{\bar{\mathbf{A}}}\right)^{-1}$$

$$\mathrm{COV}\left\{\hat{\bar{\mathbf{p}}}\right\} = \sigma^2\left(\bar{\bar{\mathbf{A}}}^T\bar{\bar{\mathbf{A}}}\right)^{-1}$$

Veja que a matriz de covariância dos parâmetros do estimador $\hat{\mathbf{p}}$ dos MQP no caso particular em que $\mathrm{COV}\{\bar{\varepsilon}\} = \overline{\overline{\mathbf{W}}}^{-1} = \sigma^2 \overline{\overline{\mathbf{I}}}$ é exatamente igual a matriz de covariância dos parâmetros do estimador $\hat{\mathbf{p}}$ dos MQ, portanto não há melhoria significativa com a introdução da matriz peso da variância dos ruídos com relação a estabilidade da solução estimada. A pequena melhoria que poderá ocorre NÃO é devido a estabilização da solução, mas devido a um processo semelhante a escolha das observações com maior razão sinal/ruído.

<u>Papel dos Pesos nos Mínimos Quadrados Ponderados:</u>

No método dos Mínimos Quadrados Ponderados (MQP) a matriz diagonal $\overline{\overline{\mathbf{W}}} \in R^{N \times N}$ define a contribuição relativa de cada erro individual ao erro total estimado. Para observações mais precisas desejamos um ajuste melhor, ou seja, desejamos que os erros ε_i sejam menores. Como minimizaremos a soma dos quadrados dos erros ponderados, i.e., $\min\{\varepsilon_1^2 w_1 + \ldots + \varepsilon_N^2 w_N\}$, se atribuirmos para a i-ésima observação um peso w_i grande para minimizar $Q = \sum_{i=1}^{N} \varepsilon_i^2 w_i$ o erro ε_i deverá ser pequeno implicando um melhor ajuste. Por outro lado, se atribuirmos para a i-ésima observação um peso w_i pequeno para minimizar $Q = \sum_{i=1}^{N} \varepsilon_i^2 w_i$ o erro ε_i poderá ser grande implicando um pior ajuste.

Em resumo, desejamos que o erro ε_i das observações mais precisas e menos precisas tenham um peso grande e um peso pequeno, respectivamente, na quantificação do erro total.