

Regularização de Tikhonov:

Transformação de um problema mal-posto em um outro bem-posto através da introdução de informação a priori:

Na prática, um problema inverso é mal posto quando existem infinitas soluções que explicam as observações geofísicas dentro de uma precisão experimental. Heuristicamente, podemos dizer que um problema mal posto reflete um desbalanceamento entre a quantidade de informações contida nos dados observados e a quantidade de informações demandada pelo intérprete na solução do problema inverso.

Matematicamente, um problema é mal posto (Hadamard, 1962) se a sua solução não obedece a pelo menos uma das três condições:

- EXISTÊNCIA
- UNICIDADE
- ESTABILIDADE

O problema da Existência , Unicidade e Estabilidade:

Seja p uma função desconhecida pertencente ao sub-espço F do espaço de P de todos os possíveis parâmetros e seja uma função conhecida y pertencente ao sub-espço métrico AF de um espaço Y de todas as possíveis observações (Figura 1). Considere A um operador linear que mapeia p em y . Portanto, nosso problema reduz-se a resolver a equação linear dada por

$$y = A p \quad (1)$$

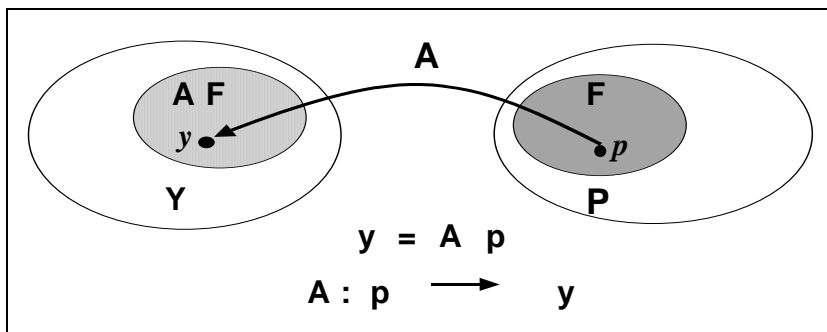


Figura 1

Note que um elemento $y \in AF$ representa um elemento produzido por uma transformação AP, cujo operador A envolve relações de causa e efeito obtidas da resolução de um sistema de equações diferenciais advindas da física-matemática (equação de Poisson, equação de Laplace, equações de Maxwell, equação da onda). No entanto, em problemas práticos não conhecemos y com total precisão, em geral, conhecemos uma medida aproximada y^0 que contem componentes aleatórias que não são explicadas por estas equações. Assim a medida y^0 pertence ao espaço Y porém pode não pertencer ao sub-espço AF. Neste caso **NÃO EXISTE SOLUÇÃO PARA O SISTEMA $y^0 = Ap$** , uma vez que o operador A não é definido (Figura 2)

Portanto, estamos violando a primeira condição imposta por Hadarmard para que um problema seja bem posto: **inexistência da solução**.

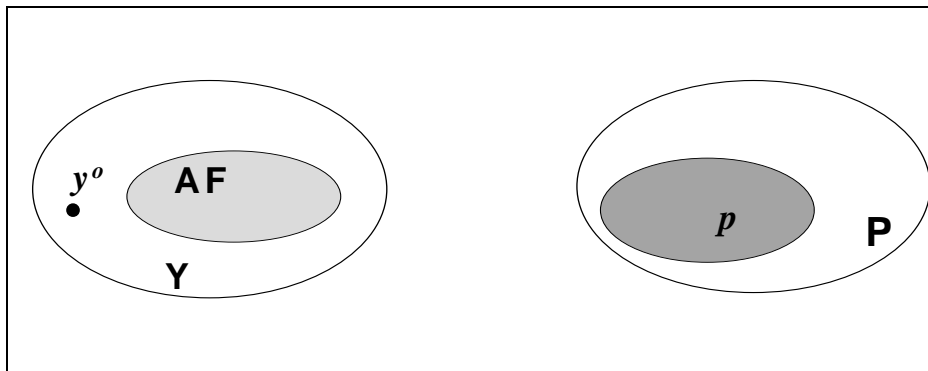


Figura 2

A inexistência de solução levou ao conceito de **Quasisolução**. Para tanto considere uma projeção de y^0 no sub-espço AF (Figura 3). Esta projeção y^c , será tal que a distância entre y^c e y^0 será mínima dentre todos os elementos $y \in AF$, em outras palavras, $d(y^c, y^0) \leq \delta$ em que d é uma métrica definida no espaço Y e δ um escalar positivo que tende a zero a medida que y^0 tende a y^c .

Agora o operador A é definido garantindo a existência da solução p^* denominadas de **Quasisolução** (Figura 3). Logo temos que o operador A mapeia p^* em y^c resultando na equação: $y^c = A p^*$.

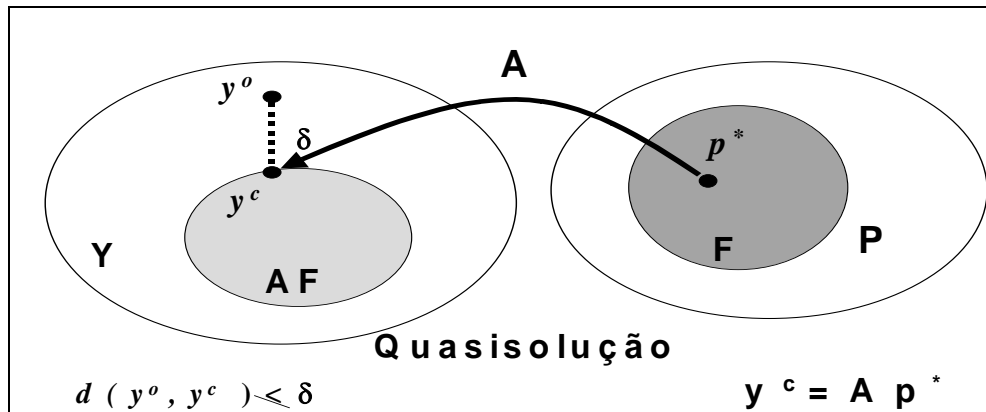


Figura 3

Considere que a distância entre y^c e y^o é a distância Euclidiana, então buscamos uma estimativa de p que minimize a distância Euclidiana entre y^c e y^o . A solução é a estimativa via MQ (Figura 4).

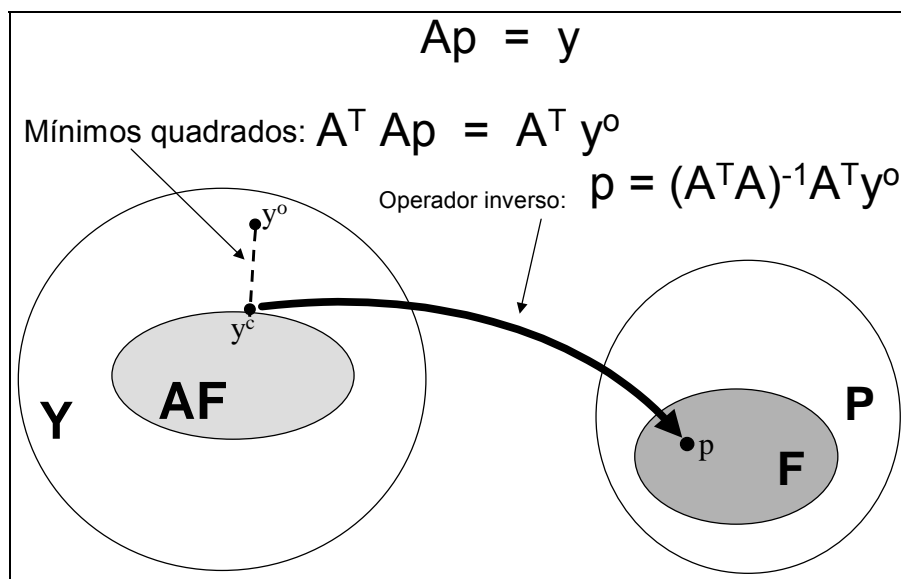


Figura 4

No entanto se o sub-espaco AF for não convexo (Figura 5) existirão infinitas projeções de y^o em AF uma vez que há infinitas métricas $d(y^c, y^o) \leq \delta$. Se há infinitas projeções y^c há portanto infinitas quasissoluções p^* . Neste caso a existência da solução foi garantida mas a unicidade da solução não (Figura 5).

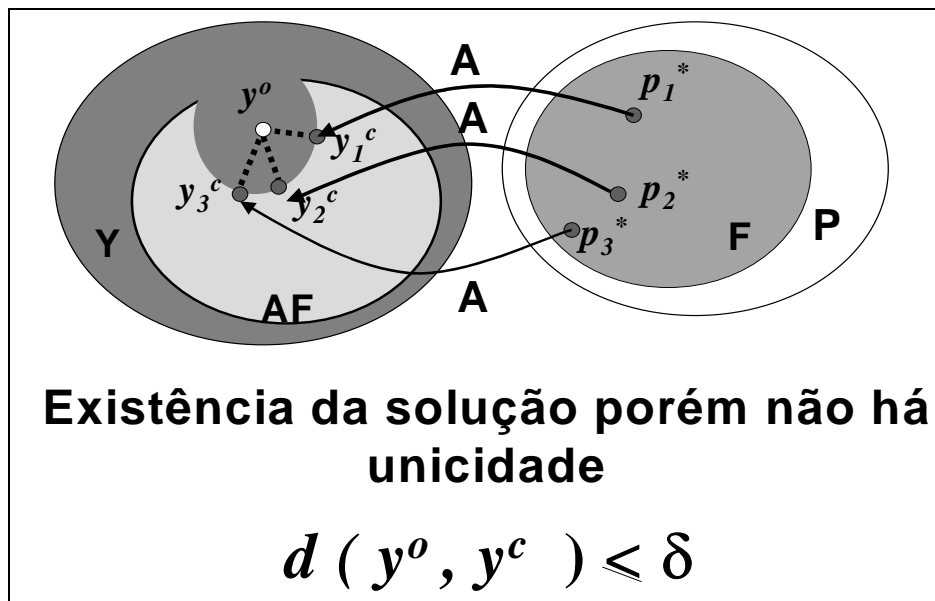


Figura 5

Desta forma temos a primeira condição de UNICIDADE DA SOLUÇÃO que garante que a situação mostrada na figura acima não ocorre.

PRIMEIRA CONDIÇÃO DE UNICIDADE DA SOLUÇÃO:

Para a garantia da unicidade da solução temos que garantir a unicidade da projeção y^o no sub-espaço AF. Para tanto AF tem que ser CONVEXO. Como o operador A é linear se AF é convexo então F é convexo também (Figura 6).

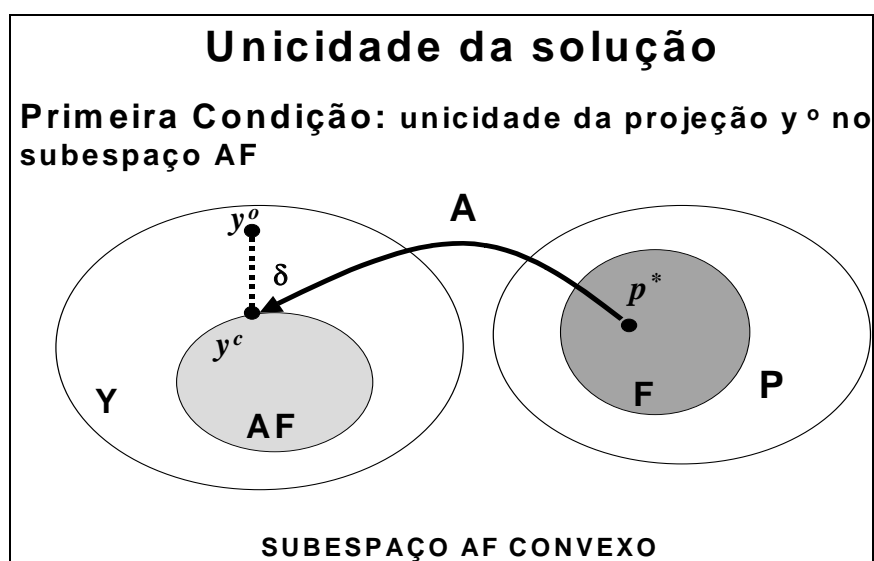


Figura 6

SEGUNDA CONDIÇÃO DE UNICIDADE DA SOLUÇÃO:

A segunda condição para a garantia da unicidade da solução é a garantia da unicidade do operador linear A que deve ser único um por um. Logo para o operador A que mapeia F em AF há um operador inverso A^{-1} que mapeia AF em F (Figura 7). Esta condição implica $p^* = 0$ é a única solução da equação $A p^* = 0$.

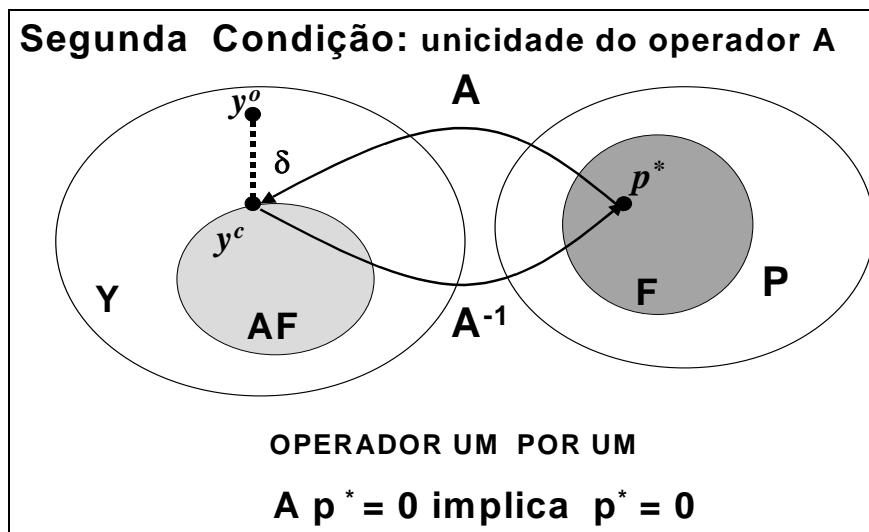


Figura 7

TERCEIRA CONDIÇÃO DE UNICIDADE DA SOLUÇÃO:

A terceira condição para a garantia da unicidade da solução é a garantia da continuidade do operador linear A . A continuidade do operador A leva ao importante conceito de estabilidade da solução.

Para entendermos o conceito de estabilidade vamos considerar que o **subespaço métrico AF é convexo** (logo garantimos a unicidade da projeção de y^o no subespaço AF) e que **o operador linear A é único e um-por-um** (logo garantimos a unicidade do operador linear A que mapeia p^* em y^c , portanto $p^* = 0$ é a única solução da equação $A p^* = 0$). Vamos considerar agora uma pequena perturbação δ , de modo que ao invés de conhecermos y^c conhecemos y^c_δ pertencente ao subespaço métrico AF , tal que a métrica $d(y^c, y^c_\delta) \leq \delta$. Logo o operador A mapeia p^* em y^c e o mesmo operador A mapeia p^*_δ em y^c_δ e que a métrica entre $d(p^*, p^*_\delta) \leq \varepsilon$.

Uma **solução é INSTÁVEL** se a métrica entre y^c e y^c_δ tender a zero **NÃO** implica que a métrica entre p^* e p^*_δ tenda a zero (Figura 8). Em outras palavras, uma solução é instável quando δ tender a zero **NÃO** implica que ε tenda a zero.

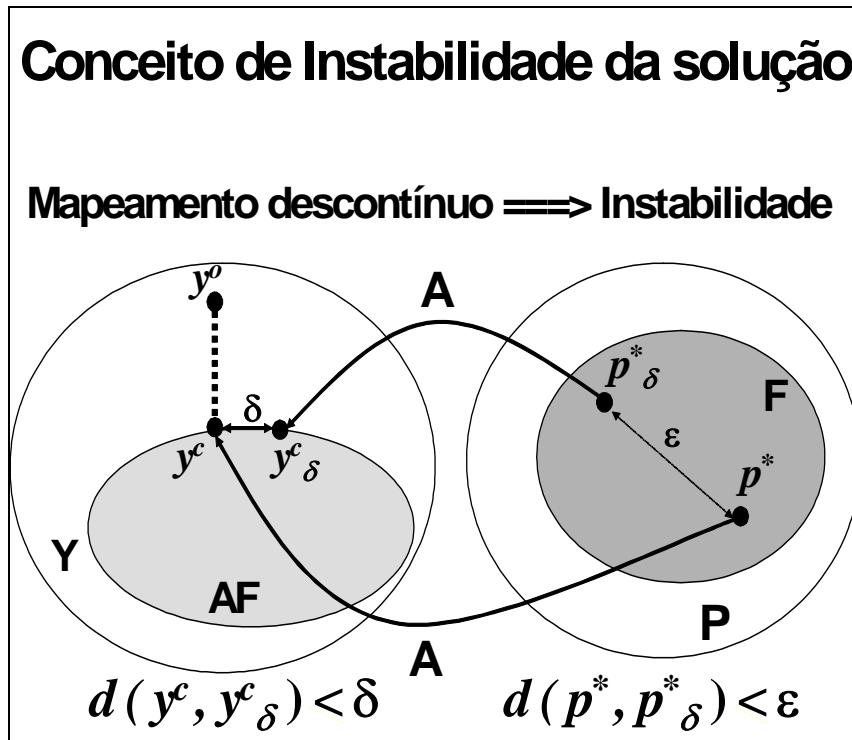


Figura 8

Neste caso de instabilidade da solução, o **operador linear A é único, um-por-um porém descontínuo** (Figura 8).

Heuristicamente, falando uma solução é instável se uma pequena perturbação no espaço da observações geofísicas leva a uma grande perturbação no espaço dos parâmetros (soluções).

Por outro lado, se **operador linear A é único, um-por-um e contínuo a solução é ESTÁVEL**. Neste caso, no mapeamento contínuo se a métrica entre y^c e y^c_δ tender a zero, ou seja se δ tender a zero, implica que a métrica entre p^* e p^*_δ tenda a zero, ou seja implica que ε tenda a zero (Figura 9). Neste caso, em que o mapeamento é contínuo p^*_δ tende a p^* . Uma solução é estável se uma pequena perturbação no

espaço das observações (AF) leva a uma pequena perturbação no espaço dos parâmetros (espaço das soluções F).

Temos então um lema: Dado um subespaço COMPACTO e CONVEXO F, se o mapeamento $F \rightarrow AF$ é CONTÍNUO e UM-POR-UM então o mapeamento inverso $AF \rightarrow F$ é também CONTÍNUO.

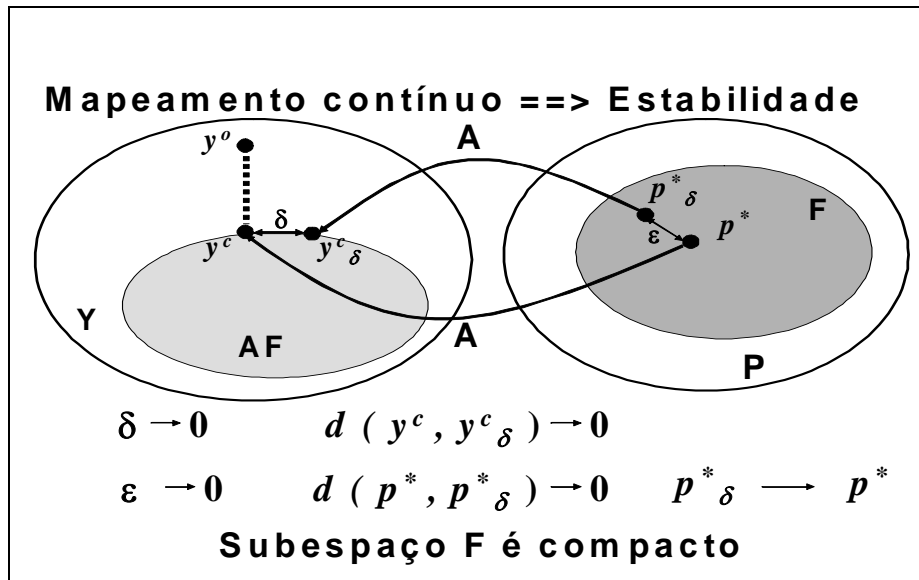


Figura 9

Conclusão: Se o operador linear A é contínuo e um-por-um e se o subespaço F que contem a solução é compacto então temos um problema bem posto.

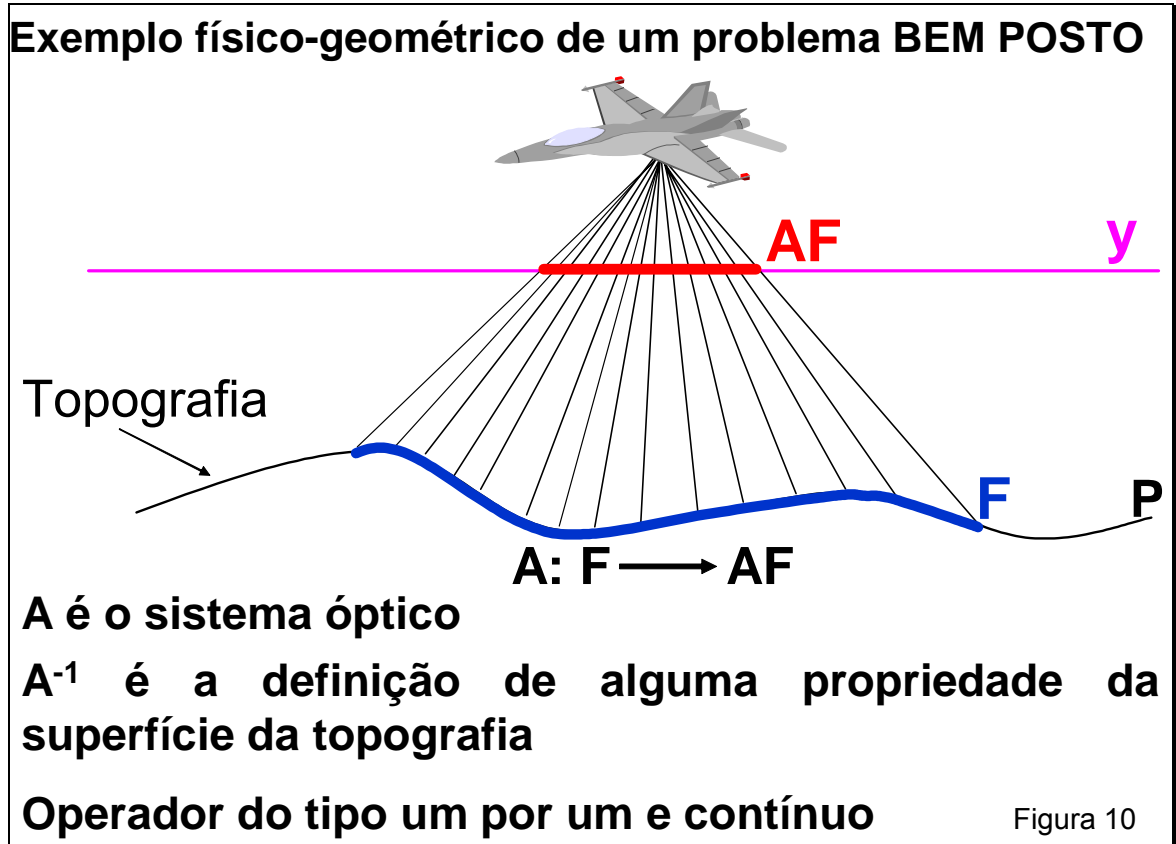
Heuristicamente, um conjunto compacto é um conjunto em que dado qualquer elemento podemos encontrar outro elemento tão próximo do primeiro quanto se queira. Geometricamente, um conjunto compacto é aquele que não há “buracos” ou “descontinuidades”.

Exemplo físico-geométrico de um problema bem posto:

Vamos ilustrar um problema bem posto com um exemplo físico-geométrico usando uma fotografia aérea, em que o centro do sistema óptico está em um avião que sobrevoa uma região (Figura 10). Então temos que:

- P (espaço de todos os possíveis parâmetros) é a topografia da região;

- Y (espaço de todas as possíveis observações) é o registro fotográfico;
- F (sub-espaço do espaço P que contém a solução) é o trecho da topografia visível quando o centro óptico está na atual posição (localização do avião);
- AF (sub-espaço do espaço Y que contém as observações) é a imagem de F ;
- A (operador que mapeia F em AF) é o sistema ótico;
- A^{-1} (operador inverso $AF \rightarrow F$) é a definição de alguma propriedade da topografia (por exemplo cor do solo);

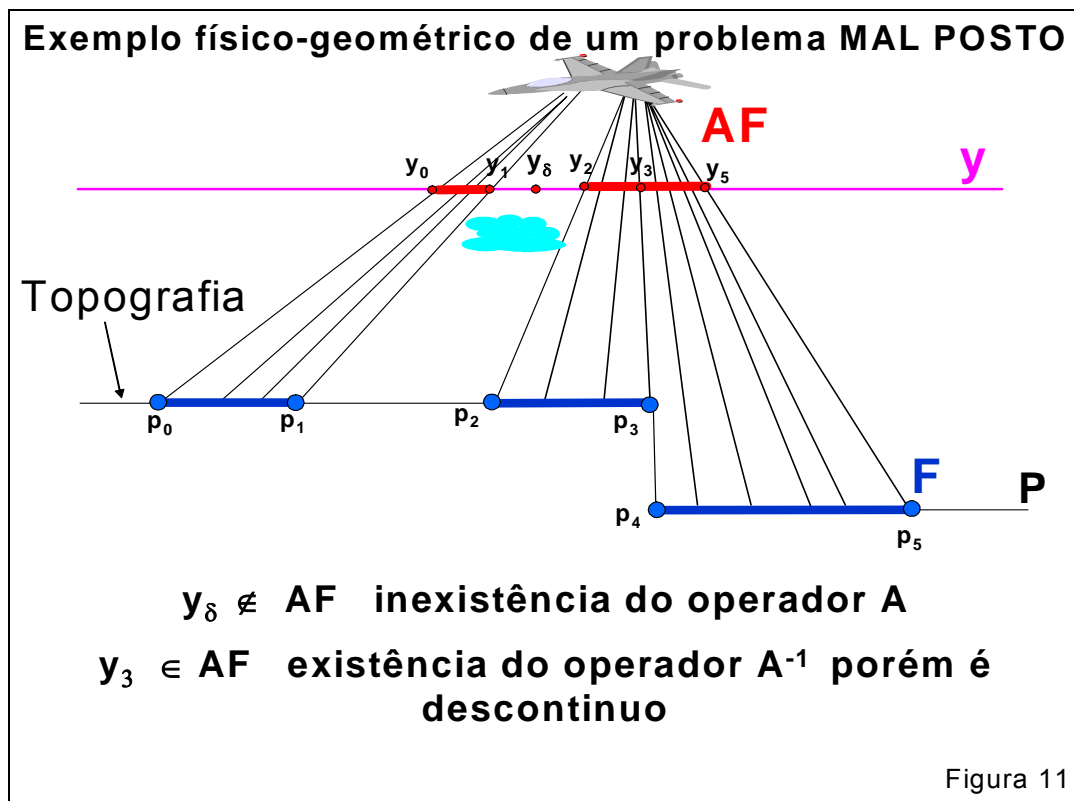


Note que o operador A que faz o mapeamento $F \rightarrow AF$ é UM-POR-UM isto porque a cada elemento em F corresponde apenas um elemento em AF . Note também que o operador A que faz o mapeamento $F \rightarrow AF$ é contínuo porque se dois elementos estão próximos em F suas correspondentes imagens em AF também estão próximas. Além disso note que o subespaço F é um espaço compacto, ou seja, não há buracos ou descontinuidades.

Concluimos que se o operador A é contínuo e um-por-um e se o subespaço F que contem a solução é compacto então temos um problema bem posto.

Exemplo físico-geométrico de um problema mal posto:

Vamos ilustrar um problema mal posto usando o mesmo exemplo físico-geométrico da fotografia aérea apresentado acima. No entanto, vamos supor que o avião esteja sobrevoando uma região com muitas nuvens (Figura 11). Especificamente, suponhamos agora que haja uma nuvem entre os pontos p_1 e p_2 então, neste trecho, a imagem fotográfica da topografia não foi registrada. Assim as operações A e A^{-1} não foram definidas em y_δ isto porque $y_\delta \notin AF$.



Em outras palavras, houve neste caso a violação da condição de existência da solução (primeira condição de Hadamard). Além disso, observe que uma pequena perturbação em torno de y_3 leva a grandes perturbações no espaço dos parâmetros. Especificamente, uma pequena perturbação para direita de y_3 leva através de A^{-1} a pontos na topografia a direita de p_4 , ou seja, a elementos (pontos) afastados de p_3 . Em outras palavras, neste caso o mapeamento inverso $AF \rightarrow F$ é descontínuo (A^{-1} é descontínua em y_3).

Concluimos que neste exemplo não há garantia da existência do operador A . Além disso, quando o operador A existe é do tipo um-por-um porém é descontínuo e o subespaço F , que contem a solução, não é compacto (com buracos e descontinuidades). Então este segundo exemplo caracteriza um problema mal posto.

O método da Regularização de Tikhonov:

Suponhamos que o operador A na equação

$$A p = y \quad (2)$$

é tal que o operador inverso A^{-1} não seja contínuo no subespaço AF e o subespaço F de possíveis soluções não seja compacto, logo não há estabilidade da solução.

Para solucionarmos o problema da instabilidade da solução, há uma abordagem desenvolvida por Tikhonov (1963) conhecida como método de Regularização de Tikhonov que estabelece condições matemáticas para obtermos soluções estáveis de modo que uma pequena perturbação nos dados leve a pequenas perturbações nas soluções estimadas. Esta abordagem baseia-se no conceito de operador regularizador.

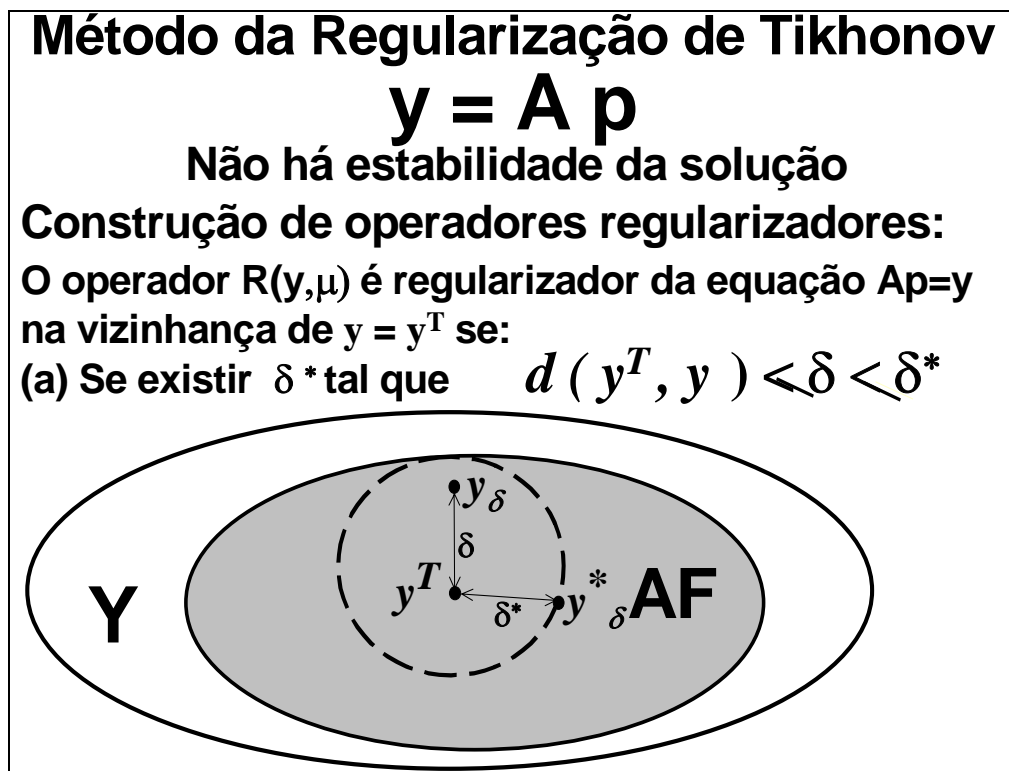
O conceito de operador regularizador:

Suponhamos que o operador A na equação (2) é tal que o operador inverso A^{-1} não seja contínuo no subespaço AF e o subespaço F de possíveis soluções não seja compacto. Se o membro do lado direito da equação $A \mathbf{p} = \mathbf{y}$ é um elemento $\mathbf{y}_\delta \in AF$ que difere do exato membro do lado direito \mathbf{y}^T de um valor δ , então se a métrica, definida no subespaço AF , $d(\mathbf{y}^T, \mathbf{y}_\delta) \leq \delta$, é óbvio que a solução aproximada \mathbf{p}_δ^* da equação (2) não pode ser definida como a solução exata desta equação com membro do lado direito aproximado ($\mathbf{y} = \mathbf{y}_\delta$), ou seja, conforme a equação: $\mathbf{p}_\delta^* = A^{-1} \mathbf{y}_\delta$. Isto porque se A^{-1} não é um operador contínuo no subespaço AF , então não há garantia que a solução obtida a partir de observações imprecisas \mathbf{y}_δ , ou seja a solução $\mathbf{p}_\delta^* = A^{-1} \mathbf{y}_\delta$, tenda a solução exata \mathbf{p}^{*T} quando $\delta \rightarrow 0$.

Para obtermos uma solução aproximada estável via método da regularização de Tikhonov teremos primeiro que encontrar um operador regularizador. Para tanto faremos algumas considerações preliminares. Vimos que o parâmetro numérico δ caracteriza o erro no lado direito da equação (2). Naturalmente os operadores regularizadores que iremos definir para buscar uma solução aproximada estável dependerá de um parâmetro chamado de parâmetro de regularização que é função de δ . Portanto, definiremos \mathbf{p}_δ^* com o auxílio de um operador regularizador que depende de um parâmetro tendo um valor escolhido de acordo com o erro δ no dado inicial \mathbf{y}_δ . Especificamente, este operador será escolhido de modo que se $\delta \rightarrow 0$, ou seja, se \mathbf{y}_δ aproximar-se do valor exato \mathbf{y}^T , então a solução aproximada \mathbf{p}_δ^* deve aproximar-se da solução exata \mathbf{p}^{*T} que estamos procurando para a equação $A \mathbf{p} = \mathbf{y}^T$. Sendo que a solução exata, o elemento $\mathbf{p}^{*T} \in F$, e as observações exatas, $\mathbf{y}^T \in AF$, são conectados pela equação $A \mathbf{p}^{*T} = \mathbf{y}^T$.

Como já abordamos, para solucionarmos o problema da instabilidade da solução, há uma abordagem desenvolvida por Tikhonov (1963) conhecida como método de Regularização de Tikhonov que estabelece condições matemáticas para obtermos soluções estáveis. Este método baseia-se na construção de um operador regularizador para a equação (2) na vizinhança de $\mathbf{y} = \mathbf{y}^T$. Um operador $R(\mathbf{y}, \mu)$ é chamado de regularizador para a equação $\mathbf{y} = \mathbf{A}\mathbf{p}$ na vizinhança de $\mathbf{y} = \mathbf{y}^T$ se duas condições forem satisfeitas:

- (1) Se na vizinhança de \mathbf{y} existir um número positivo δ^* , tal que o operador $R(\mathbf{y}, \mu)$ é definido para todo $\mu > 0$ e todo $\mathbf{y} \in \text{AF}$. Então existe um valor $\delta^* > 0$ tal que $d(\mathbf{y}, \mathbf{y}^T) \leq \delta \leq \delta^*$ e dentro desta região define-se um operador regularizador $R(\mathbf{y}, \mu)$.



(2) Se existe y_δ que satisfaz $d(y^T, y_\delta) \leq \delta \leq \delta^*$ em que o operador regularizador $R(y, \mu)$ está definido dentro desta região. Então existe uma função $\mu = \mu(\delta, y_\delta)$ de δ e dos dados y_δ tal que, para todo $\varepsilon > 0$, existe um número $\delta(\varepsilon) \leq \delta^*$ tal que a inclusão $y_\delta \in AF$ e da inequação

$$d(y^T, y_\delta) \leq \delta(\varepsilon)$$

implique

$$d(p^{*T}, p_{\delta}^*) \leq \varepsilon$$

onde

$$p_{\delta}^* = p_{\mu(\delta)}^* = R(y_\delta, \mu(\delta)).$$

(b) Se existir y_δ que satisfaz $d(y, y_\delta) \leq \delta \leq \delta^*$

Então posso definir um operador $R(y_\delta, \mu(\delta))$, em que $\mu(\delta) > 0$, tal que

$$d(p^*, p_{\mu(\delta)}^*) \leq \varepsilon, \quad \varepsilon > 0$$

implique $d(y, y_\delta) \leq \delta(\varepsilon) \leq \delta^*(\varepsilon)$

em que $\mu(\delta)$ será escolhido de modo que se :

$$\delta(\varepsilon) \rightarrow 0 \quad d(y, y_\delta) \rightarrow 0$$

$$\varepsilon \rightarrow 0 \quad d(p^*, p_{\mu(\delta)}^*) \rightarrow 0 \quad p_{\mu(\delta)}^* \rightarrow p^*$$

Se $d(y_\delta, y^T) \leq \delta \leq \delta^*$ podemos com o auxílio de um operador de regularização $R(y_\delta, \mu)$, escolher o parâmetro numérico $\mu = \mu(\delta, y_\delta)$ conforme o erro δ dos dados iniciais y_δ de modo que, se $\delta \rightarrow 0$, ou seja, se y_δ aproximar-se do valor exato y^T , então a **solução regularizada** da $p_\delta^* = p_{\mu(\delta)}^* = R(y_\delta, \mu(\delta))$ aproxima-se da solução exata p^{*T} , isto é, $d(p^{*T}, p_{\mu(\delta)}^*) \rightarrow 0$. Portanto a solução regularizada p_δ^* é uma solução estável.

Esta solução $p_\delta^* = p_{\mu(\delta)}^* = R(y_\delta, \mu(\delta))$, obtida com o auxílio do operador de regularização $R(y_\delta, \mu(\delta))$, é chamada uma **solução regularizada** da equação (2) e o parâmetro numérico μ é chamado de **parâmetro de regularização**.

Em resumo o método de Regularização de Tikhonov estabelece duas condições. A condição (1) estabelece que existe uma região em torno do elemento $y = y^T$ dentro da qual definimos um operador $R(y, \mu)$ chamado de operador de regularização. Este operador depende continuamente da variável μ (**parâmetro de regularização**). A condição (2) estabelece que se o elemento y_δ está dentro desta região definida pela condição (1), então podemos definir um operador de regularização $R(y_\delta, \mu(\delta))$ que depende continuamente da variável $\mu(\delta)$ tal que, se $\delta \rightarrow 0$, ou seja, se a distância $d(y_\delta, y^T) \rightarrow 0$, isto implica que $\varepsilon \rightarrow 0$, isto é, $d(p^{*T}, p_{\mu(\delta)}^*) \rightarrow 0$ e por consequência que $p_{\mu(\delta)}^* \rightarrow p^{*T}$.

O método da Regularização de Tikhonov:

O método de regularização de Tikhonov é um método de obtenção de soluções aproximadas da equação (2) que são estáveis sob pequenas perturbações no membro do lado direito da equação (dados geofísicos). Este método consiste:

- a) na construção de operadores de regularização ; e
- b) na determinação do parâmetro de regularização $\mu(\delta)$ que é função do tamanho do erro contido nos dados y_δ (no membro do lado direito da equação)

Então o nosso primeiro passo, consiste em construir operadores de regularização.

Método de construção de operadores de regularização:

O método para a construção de operadores de regularização para a equação (2) baseia-se no princípio variacional. Nós presumiremos que a equação $\mathbf{A}\mathbf{p} = \mathbf{y}^T$, em que $\mathbf{y}^T \in AF$ são os dados exatos, tem apenas uma ÚNICA solução \mathbf{p}^{*T} .

Seja $H(\mathbf{p})$ um funcional com as seguintes propriedades:

- 1) contínuo;
- 2) não negativo,
- 3) definido no subespaço compacto F_1 do espaço métrico F ;
- 4) em que a solução exata \mathbf{p}^{*T} pertence ao domínio de definição de $H(\mathbf{p})$; e
- 5) para todo número positivo C , o conjunto de elementos \mathbf{p} de F_1 para o qual

$$H(\mathbf{p}) \leq C \text{ é um subconjunto de } F_1 .$$

A este funcional $H(\mathbf{p})$ apresentando tais propriedades chamaremos de **funcional estabilizante** ou simplesmente o **estabilizador**.

O método de construção de operador regularizador reduz-se a um problema de extremo condicional de minimizar o funcional quasimonotônico $H(p)$ sujeito a explicar os dados observados (geofísicos) dentro do erro δ contido nos dados y_δ observados (erro experimental)

$$\begin{cases} \inf_{p \in F} H(p) \\ \text{sujeito a: } \|A p - y_\delta\| \leq \delta, \end{cases} \quad (3)$$

este problema é de difícil solução. Operacionalmente, torna-se mais fácil solucionarmos o seguinte problema de extremo condicional:

$$\begin{cases} \inf_{p \in F} H(p) \\ \text{sujeito a: } \|A p - y_\delta\| = \delta. \end{cases} \quad (4)$$

Este problema de extremo condicional (equação 4) é resolvido via método dos Multiplicadores de Lagrange. Assim temos o seguinte problema de extremo incondicional :

$$\inf_{p \in F} \left\{ \|A p - y_\delta\|^2 + \mu(\delta)H(p) \right\}, \quad (5)$$

em que o parâmetro de regularização $\mu(\delta)$ é determinado da condição

$$d(A p_{\mu(\delta)}^*, y_\delta) = \delta \quad (6)$$

sendo $p_{\mu(\delta)}^*$ o elemento em F no qual o funcional $M^\mu(p, y_\delta)$

$$M^\mu(p, y_\delta) = \|A p - y_\delta\|^2 + \mu(\delta)H(p) \quad (7)$$

atinge o seu maior limite inferior (inf).

O elemento $\mathbf{p}_{\mu(\delta)}^*$ pode ser considerado como o resultado da aplicação ao membro do lado direito ($\mathbf{y}=\mathbf{y}_\delta$) da equação (2) algum operador regularizador R_1 que depende do parâmetro de regularização $\mu(\delta)$:

$$\mathbf{p}_{\mu(\delta)}^* = R_1(\mathbf{y}_\delta, \mu(\delta)), \quad (8)$$

onde o parâmetro numérico $\mu=\mu(\delta)$ depende do erro δ .

Chamamos o funcional $H(\mathbf{p})$ de **funcional estabilizante** ou simplesmente o **estabilizador** e chamamos o funcional $M^\mu(\mathbf{p}, \mathbf{y}_\delta)$ de **funcional suavizante**.

Observe a generalidade da condição (5) para obtenção de soluções estáveis:

- \mathbf{p} e \mathbf{y} são elementos definidos em espaços métricos;
- \mathbf{A} é um operador linear ou não linear que realiza um mapeamento contínuo e um-
por-um.
- O funcional $H(\mathbf{p})$ sob as condições de ser um funcional estabilizante, isto é, contínuo, não negativo, definido em um subespaço compacto do espaço de parâmetros ao qual pertence a solução teórica, é um funcional bastante genérico no sentido de permitir incorporar uma classe bastante ampla de informações a priori sobre os parâmetros \mathbf{p} na solução de um problema geofísico.

Note também a generalidade do problema (7) que estabelece. Suponha que o conjunto F de possíveis soluções da equação (2) é um espaço métrico e que $H(\mathbf{p})$ é um funcional estabilizante definido no conjunto $F_1 \subset F$. Então temos o seguinte teorema:

Seja \mathbf{A} um operador contínuo que mapeia F em AF . Para todo elemento \mathbf{y} de AF e todo parâmetro positivo μ , existe um elemento $\mathbf{p}_\mu^* \in F_1$ para o qual o funcional

$$M^\mu(\mathbf{p}, \mathbf{y}) = d(\mathbf{A} \mathbf{p}, \mathbf{y}) + \mu H(\mathbf{p})$$

atinge o seu maior limite inferior:

$$\inf_{\mathbf{p} \in \mathbf{F}_1} \mathbf{M}^\mu(\mathbf{p}, \mathbf{y}) = \mathbf{M}^\mu(\mathbf{p}_\mu^*, \mathbf{y}).$$

Veja que \mathbf{p} e \mathbf{y}_δ são, simplesmente, elementos definidos em espaços métricos.

Estes elementos podem ser, por exemplo, funções. Considerando os elementos \mathbf{p} e \mathbf{y}_δ como os vetores $\bar{\mathbf{p}}$ e $\bar{\mathbf{y}}^0$ que contém, respectivamente, um número finito de parâmetros descrevendo o modelo interpretativo e as observações em pontos discretos, então ao invés de um operador contínuo \mathbf{A} , teremos a matriz de sensibilidade $\bar{\bar{\mathbf{A}}}(\bar{\mathbf{p}})$, que no caso linear não depende de $\bar{\mathbf{p}}$ e, finalmente, ao invés do maior limite inferior (inf) do funcional $\mathbf{H}(\mathbf{p})$ teremos o mínimo de uma função escalar $\mathbf{H}(\bar{\mathbf{p}})$. Tomando então na equação (4) a norma $\|\cdot\|$ e a função $\mathbf{H}(\bar{\mathbf{p}})$, como normas Euclidianas ponderadas:

$$\begin{aligned} \mathbf{H}(\bar{\mathbf{p}}) &= (\bar{\mathbf{p}} - \bar{\mathbf{p}}^0)^T \bar{\bar{\mathbf{W}}}_p (\bar{\mathbf{p}} - \bar{\mathbf{p}}^0), \\ \text{e} \\ \|\bar{\mathbf{y}}^0 - \bar{\bar{\mathbf{A}}} \bar{\mathbf{p}}\|^2 &= (\bar{\mathbf{y}}^0 - \bar{\bar{\mathbf{A}}} \bar{\mathbf{p}})^T \bar{\bar{\mathbf{W}}}_y (\bar{\mathbf{y}}^0 - \bar{\bar{\mathbf{A}}} \bar{\mathbf{p}}) \end{aligned} \quad (9)$$

então a condição (5) torna-se:

$$\text{MIN}\{\mathbf{M}(\bar{\mathbf{p}})\} = \text{MIN}\left\{\frac{1}{\mu(\delta)} (\bar{\mathbf{y}}^0 - \bar{\bar{\mathbf{A}}} \bar{\mathbf{p}})^T \bar{\bar{\mathbf{W}}}_y (\bar{\mathbf{y}}^0 - \bar{\bar{\mathbf{A}}} \bar{\mathbf{p}}) + (\bar{\mathbf{p}} - \bar{\mathbf{p}}^0)^T \bar{\bar{\mathbf{W}}}_p (\bar{\mathbf{p}} - \bar{\mathbf{p}}^0)\right\}, \quad (10)$$

em que $\bar{\mathbf{p}}^0$ é um vetor de referência e $\bar{\bar{\mathbf{W}}}_p$ e $\bar{\bar{\mathbf{W}}}_y$ são matrizes positivo-definidas nos espaços de parâmetros e observações, respectivamente. Note que a função escalar $\mathbf{H}(\bar{\mathbf{p}})$ na equação (9) introduz informações a priori no espaço dos parâmetros e depende somente do vetor $\bar{\mathbf{p}}^0$ e da matriz $\bar{\bar{\mathbf{W}}}_p$. Se a matriz $\bar{\bar{\mathbf{W}}}_p$ for diagonal então a função $\mathbf{H}(\bar{\mathbf{p}})$ (na equação (9)) pode ser escrita como:

$$H(\bar{\mathbf{p}}) = \sum_{i=1}^M (\mathbf{p}_i - \mathbf{p}^o_i)^2 \mathbf{W}_{p_i},$$

em que \mathbf{W}_{p_i} , \mathbf{p}_i e \mathbf{p}^o_i são o i -ésimo valor diagonal e os i -ésimos elementos de $\bar{\bar{\mathbf{W}}}_p$, $\bar{\mathbf{p}}$ e $\bar{\mathbf{p}}^o$ respectivamente.

Assim a informação a priori introduzida ao minimizar $H(\bar{\mathbf{p}})$ é a de buscar uma solução aproximada $\tilde{\mathbf{p}}$ que esteja o mais próximo possível de $\bar{\mathbf{p}}^o$, sendo que o grau de proximidade desejada é dado pela matriz $\bar{\bar{\mathbf{W}}}_p$. Valores altos na diagonal de $\bar{\bar{\mathbf{W}}}_p$ impõe uma grande proximidade de $\tilde{\mathbf{p}}$ a $\bar{\mathbf{p}}^o$ e vice-versa. A solução geral da equação (10) é obtida tomando o gradiente da função $M(\bar{\mathbf{p}})$ e igualando o resultado ao vetor nulo:

$$\begin{aligned} \nabla_p \{M(\bar{\mathbf{p}})\} &= \nabla_p \left\{ (\bar{\mathbf{y}}^o - \bar{\mathbf{A}} \bar{\mathbf{p}})^T \bar{\bar{\mathbf{W}}}_y (\bar{\mathbf{y}}^o - \bar{\mathbf{A}} \bar{\mathbf{p}}) \right\} + \mu(\delta) \nabla_p \left\{ (\bar{\mathbf{p}} - \bar{\mathbf{p}}^o)^T \bar{\bar{\mathbf{W}}}_p (\bar{\mathbf{p}} - \bar{\mathbf{p}}^o) \right\} \\ \nabla_p \{M(\bar{\mathbf{p}})\} &= \nabla_p \left\{ \bar{\mathbf{p}}^T \bar{\mathbf{A}}^T \bar{\bar{\mathbf{W}}}_y \bar{\mathbf{A}} \bar{\mathbf{p}} - 2 \bar{\mathbf{p}}^T \bar{\mathbf{A}}^T \bar{\bar{\mathbf{W}}}_y \bar{\mathbf{y}}^o + \bar{\mathbf{y}}^{oT} \bar{\bar{\mathbf{W}}}_y \bar{\mathbf{y}}^o \right\} + \\ &\quad \mu(\delta) \nabla_p \left\{ \bar{\mathbf{p}}^T \bar{\bar{\mathbf{W}}}_p \bar{\mathbf{p}} - 2 \bar{\mathbf{p}}^T \bar{\bar{\mathbf{W}}}_p \bar{\mathbf{p}}^o + \bar{\mathbf{p}}^{oT} \bar{\bar{\mathbf{W}}}_p \bar{\mathbf{p}}^o \right\} \end{aligned}$$

Calculando os gradientes e igualando o resultado ao vetor nulo temos:

$$2 \bar{\mathbf{A}}^T \bar{\bar{\mathbf{W}}}_y \bar{\mathbf{A}} \tilde{\mathbf{p}} - 2 \bar{\mathbf{A}}^T \bar{\bar{\mathbf{W}}}_y \bar{\mathbf{y}}^o + 2 \mu(\delta) \bar{\bar{\mathbf{W}}}_p \tilde{\mathbf{p}} + 2 \mu(\delta) \bar{\bar{\mathbf{W}}}_p \bar{\mathbf{p}}^o = \mathbf{0} \quad (12)$$

$$\bar{\mathbf{A}}^T \bar{\bar{\mathbf{W}}}_y \bar{\mathbf{A}} \tilde{\mathbf{p}} + \mu(\delta) \bar{\bar{\mathbf{W}}}_p \tilde{\mathbf{p}} - \mu(\delta) \bar{\bar{\mathbf{W}}}_p \bar{\mathbf{p}}^o = \bar{\mathbf{A}}^T \bar{\bar{\mathbf{W}}}_y \bar{\mathbf{y}}^o \quad (13)$$

Vamos somar nos dois lados da equação o termo $-\bar{\mathbf{A}}^T \bar{\bar{\mathbf{W}}}_y \bar{\mathbf{A}} \bar{\mathbf{p}}^o$ então teremos:

$$\bar{\mathbf{A}}^T \bar{\mathbf{W}}_y \bar{\mathbf{A}} (\tilde{\mathbf{p}} - \bar{\mathbf{p}}^0) + \mu(\delta) \bar{\mathbf{W}}_p (\tilde{\mathbf{p}} - \bar{\mathbf{p}}^0) = \bar{\mathbf{A}}^T \bar{\mathbf{W}}_y (\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}}^0) \quad (14)$$

Resultando na equação normal:

$$\left(\bar{\mathbf{A}}^T \bar{\mathbf{W}}_y \bar{\mathbf{A}} + \mu(\delta) \bar{\mathbf{W}}_p \right) (\tilde{\mathbf{p}} - \bar{\mathbf{p}}^0) = \bar{\mathbf{A}}^T \bar{\mathbf{W}}_y (\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}}^0) \quad (15)$$

Portanto obtemos o **estimador genérico**:

$$\tilde{\mathbf{p}} = \bar{\mathbf{p}}^0 + \left(\bar{\mathbf{A}}^T \bar{\mathbf{W}}_y \bar{\mathbf{A}} + \mu(\delta) \bar{\mathbf{W}}_p \right)^{-1} \bar{\mathbf{A}}^T \bar{\mathbf{W}}_y (\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}}^0) \quad (16)$$

Tomando a identidade matricial:

$$(\bar{\mathbf{C}} \bar{\mathbf{D}} + \bar{\mathbf{I}}_m) \bar{\mathbf{C}} = \bar{\mathbf{C}} (\bar{\mathbf{D}} \bar{\mathbf{C}} + \bar{\mathbf{I}}_n) \quad (17)$$

sendo $\bar{\mathbf{C}}$ uma matriz com dimensão M x N e $\bar{\mathbf{D}}$ uma matriz com dimensão N x M.

Pré e pós-multiplicando ambos os lados da identidade (17) por $(\bar{\mathbf{C}} \bar{\mathbf{D}} + \bar{\mathbf{I}}_m)^{-1}$ e

$(\bar{\mathbf{D}} \bar{\mathbf{C}} + \bar{\mathbf{I}}_n)^{-1}$, respectivamente, obtemos:

$$\begin{aligned} (\bar{\mathbf{C}} \bar{\mathbf{D}} + \bar{\mathbf{I}}_m)^{-1} (\bar{\mathbf{C}} \bar{\mathbf{D}} + \bar{\mathbf{I}}_m) \bar{\mathbf{C}} (\bar{\mathbf{D}} \bar{\mathbf{C}} + \bar{\mathbf{I}}_n)^{-1} &= (\bar{\mathbf{C}} \bar{\mathbf{D}} + \bar{\mathbf{I}}_m)^{-1} \bar{\mathbf{C}} (\bar{\mathbf{D}} \bar{\mathbf{C}} + \bar{\mathbf{I}}_n) (\bar{\mathbf{D}} \bar{\mathbf{C}} + \bar{\mathbf{I}}_n)^{-1} \\ \bar{\mathbf{C}} (\bar{\mathbf{D}} \bar{\mathbf{C}} + \bar{\mathbf{I}}_n)^{-1} &= (\bar{\mathbf{C}} \bar{\mathbf{D}} + \bar{\mathbf{I}}_m)^{-1} \bar{\mathbf{C}} \end{aligned} \quad (18)$$

Podemos reescrever a equação (16):

$$\tilde{\mathbf{p}} = \bar{\mathbf{p}}^0 + \bar{\mathbf{B}} (\bar{\mathbf{y}}^0 - \bar{\mathbf{A}} \bar{\mathbf{p}}^0) \quad (19)$$

em que

$$\bar{\mathbf{B}} = \left(\bar{\mathbf{A}}^T \bar{\mathbf{W}}_y \bar{\mathbf{A}} + \mu(\delta) \bar{\mathbf{W}}_p \right)^{-1} \bar{\mathbf{A}}^T \bar{\mathbf{W}}_y$$

vamos colocar em evidência o termo $\mu(\delta) \bar{\mathbf{W}}_p$ dentro do termo a ser invertido:

$$\bar{\mathbf{B}} = \left[\left(\bar{\mathbf{A}}^T \bar{\mathbf{W}}_y \bar{\mathbf{A}} \left(\frac{1}{\mu(\delta)} \bar{\mathbf{W}}_p^{-1} \right) + \bar{\mathbf{I}}_m \right) \mu(\delta) \bar{\mathbf{W}}_p \right]^{-1} \bar{\mathbf{A}}^T \bar{\mathbf{W}}_y$$

$$\bar{\bar{\mathbf{B}}} = \left(\mu(\delta) \bar{\bar{\mathbf{W}}}_p \right)^{-1} \left[\left(\bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}_y \bar{\bar{\mathbf{A}}} \left(\frac{1}{\mu(\delta)} \bar{\bar{\mathbf{W}}}_p^{-1} \right) + \bar{\bar{\mathbf{I}}}_m \right) \right]^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}_y \quad (20)$$

Vamos chamar:

$$\bar{\bar{\mathbf{C}}} = \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}_y \quad \text{e}$$

$$\bar{\bar{\mathbf{D}}} = \bar{\bar{\mathbf{A}}} \left(\frac{1}{\mu(\delta)} \bar{\bar{\mathbf{W}}}_p^{-1} \right).$$

Agora vamos usar a identidade (18) na equação (20)

$$\bar{\bar{\mathbf{B}}} = \left(\mu(\delta) \bar{\bar{\mathbf{W}}}_p \right)^{-1} \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}_y \left[\left(\bar{\bar{\mathbf{A}}} \left(\frac{1}{\mu(\delta)} \bar{\bar{\mathbf{W}}}_p^{-1} \right) \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}_y + \bar{\bar{\mathbf{I}}}_n \right) \right]^{-1}$$

Colocando o termo $\mu(\delta)^{-1} \bar{\bar{\mathbf{W}}}_y$ para dentro do termo a ser invertido temos:

$$\begin{aligned} \bar{\bar{\mathbf{B}}} &= \bar{\bar{\mathbf{W}}}_p^{-1} \bar{\bar{\mathbf{A}}}^T \left[\left(\bar{\bar{\mathbf{A}}} \left(\frac{1}{\mu(\delta)} \bar{\bar{\mathbf{W}}}_p^{-1} \right) \bar{\bar{\mathbf{A}}}^T \bar{\bar{\mathbf{W}}}_y + \bar{\bar{\mathbf{I}}}_n \right) \mu(\delta) \bar{\bar{\mathbf{W}}}_y^{-1} \right]^{-1} \\ \bar{\bar{\mathbf{B}}} &= \bar{\bar{\mathbf{W}}}_p^{-1} \bar{\bar{\mathbf{A}}}^T \left[\left(\bar{\bar{\mathbf{A}}} \bar{\bar{\mathbf{W}}}_p^{-1} \bar{\bar{\mathbf{A}}}^T + \mu(\delta) \bar{\bar{\mathbf{W}}}_y^{-1} \right) \right]^{-1} \end{aligned} \quad (21)$$

Substituindo a equação 21 em 19 temos:

$$\tilde{\bar{\mathbf{p}}} = \bar{\mathbf{p}}^0 + \bar{\bar{\mathbf{W}}}_p^{-1} \bar{\bar{\mathbf{A}}}^T \left[\bar{\bar{\mathbf{A}}} \bar{\bar{\mathbf{W}}}_p^{-1} \bar{\bar{\mathbf{A}}}^T + \mu(\delta) \bar{\bar{\mathbf{W}}}_y^{-1} \right]^{-1} (\bar{\mathbf{y}}^0 - \bar{\bar{\mathbf{A}}} \bar{\mathbf{p}}^0) \quad (22)$$

As equações (16) e (22) só são identidades entre si se as matrizes $\bar{\bar{\mathbf{W}}}_y$ e $\bar{\bar{\mathbf{W}}}_p$ e suas inversas ($\bar{\bar{\mathbf{W}}}_y^{-1}$ e $\bar{\bar{\mathbf{W}}}_p^{-1}$) não forem singulares. Note que se $\bar{\bar{\mathbf{W}}}_y$ e/ou $\bar{\bar{\mathbf{W}}}_p$ forem singulares, então só a equação (16) pode ser usada. Por outro lado, se $\bar{\bar{\mathbf{W}}}_y^{-1}$ e/ou $\bar{\bar{\mathbf{W}}}_p^{-1}$ forem singulares, então só a equação (22) pode ser usada. Se

nem uma das quatro matrizes acima for singular, então os estimadores genéricos (16) e (22) são idênticos.

Portanto as duas equações que chamaremos de **estimador genérico** são:

$$\tilde{\mathbf{p}} = \mathbf{p}^0 + \left(\overline{\mathbf{A}}^T \overline{\mathbf{W}}_y \overline{\mathbf{A}} + \mu(\delta) \overline{\mathbf{W}}_p \right)^{-1} \overline{\mathbf{A}}^T \overline{\mathbf{W}}_y \left(\mathbf{y}^0 - \overline{\mathbf{A}} \mathbf{p}^0 \right)$$

e

$$\tilde{\mathbf{p}} = \mathbf{p}^0 + \overline{\mathbf{W}}_p^{-1} \overline{\mathbf{A}}^T \left[\overline{\mathbf{A}} \overline{\mathbf{W}}_p^{-1} \overline{\mathbf{A}}^T + \mu(\delta) \overline{\mathbf{W}}_y^{-1} \right]^{-1} \left(\mathbf{y}^0 - \overline{\mathbf{A}} \mathbf{p}^0 \right)$$

Seleção do parâmetro de regularização (selecionando uma boa solução)

Para a maioria dos problemas lineares em geofísica resolvidos via estimador de mínimos quadrados resultam em infinitas soluções que ajustam os dados. Se considerarmos que os dados contem ruído não nenhuma explicação para ajustarmos perfeitamente tal ruído, portanto torna-se evidente que pode haver muitas soluções que podem adequadamente ajustar os dados no sentido que $\|\overline{\mathbf{A}} \overline{\mathbf{p}} - \overline{\mathbf{y}}^o\|_2$ seja bastante pequeno. Na regularização de Tikhonov

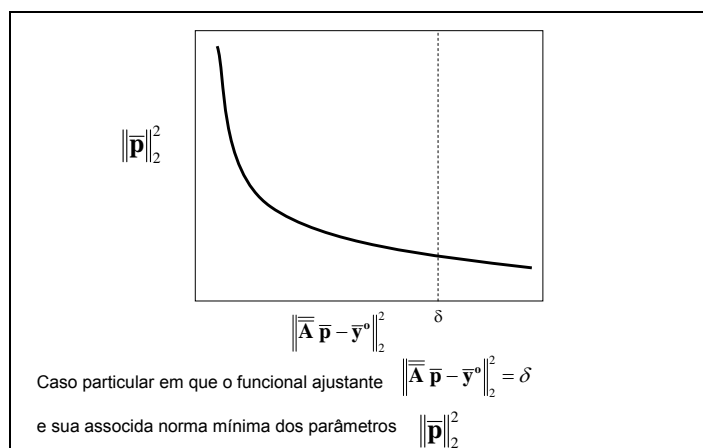
regularization, consideramos todas as soluções com $\|\overline{\mathbf{A}} \overline{\mathbf{p}} - \overline{\mathbf{y}}^o\|_2 \leq \delta$ e selecionamos dentre essas soluções aquela que minimiza a norma de $\overline{\mathbf{p}}$, i.e.,

$$\begin{aligned} \text{Minimizar} \quad & \|\overline{\mathbf{p}}\|_2 = \overline{\mathbf{p}}^T \overline{\mathbf{p}} \\ \text{sujeito a :} \quad & \|\overline{\mathbf{A}} \overline{\mathbf{p}} - \overline{\mathbf{y}}^o\|_2 \leq \delta. \quad (1) \end{aligned}$$

Por que selecionamos a solução de mínima norma dos parâmetros dentre aquelas soluções que ajustam adequadamente os dados ? Uma explicação intuitiva é que toda a característica diferente de zero que aparecer na solução regularizada aumentará a norma de $\overline{\mathbf{p}}$. Tais características aparecem na solução porque são necessárias para ajustar os dados.

Note que como δ aumenta, o conjunto de possíveis soluções também aumenta e o valor mínimo de $\|\overline{\mathbf{p}}\|_2$ diminui. Nós podemos assim construir uma curva de valores mínimos de

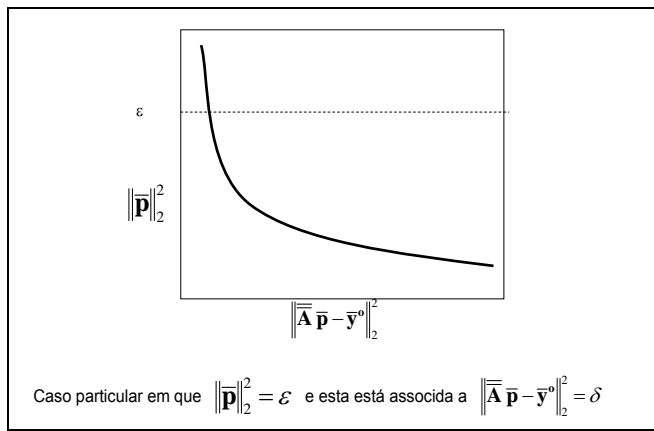
$\|\overline{\mathbf{p}}\|_2$ versus $\|\overline{\mathbf{A}} \overline{\mathbf{p}} - \overline{\mathbf{y}}^o\|_2$, i.e., uma curva relacionada ao problema condicionado acima:



Nós também podemos construir uma outra curva considerando o problema

$$\begin{aligned} &\text{Minimizar} && \left\| \overline{\mathbf{A}} \overline{\mathbf{p}} - \overline{\mathbf{y}}^o \right\|_2 \\ &\text{sujeito} && \text{a : } \left\| \overline{\mathbf{p}} \right\|_2 \leq \varepsilon. \end{aligned} \quad (2)$$

Considerando que ε diminui, o conjunto de possíveis soluções também começa diminuir, por outro lado o mínimo de $\left\| \overline{\mathbf{A}} \overline{\mathbf{p}} - \overline{\mathbf{y}}^o \right\|_2$ aumenta.



Uma terceira opção considerara o problema dos minimos quadrados amortecido (Damped Least-Square) i.e.,

$$\min \left\{ \left\| \overline{\mathbf{A}} \overline{\mathbf{p}} - \overline{\mathbf{y}}^o \right\|_2 + \mu \left\| \overline{\mathbf{p}} \right\|_2 \right\} \quad (3)$$

Este problema (3) surge quando aplicamos o método dos multiplicadores de Lagrange ao problema (1) em que μ é o parâmetro de regularização. Pode ser mostrado que uma escolha apropriada de μ , δ e ε os três problemas acima (1), (2) e (3) resultam na mesma solução.

Métodos de escolha de μ , os quais se baseiam na tarefa de se encontrar $\mu(\varepsilon)$, tal que

$\left\| \overline{\mathbf{p}} \right\|_2 = \varepsilon$, são denominados métodos “a priori” [problema (2)]. A premissa deste tipo de método, é a de que se tem uma boa informação prévia da solução a ser estimada. No entanto na maioria das vezes este dado não é disponível ou então não é suficientemente confiável ou preciso. Em geral, uma alternativa mais viável consiste em se empregar métodos, onde a escolha do parâmetro de regularização, μ , seja orientada pela busca de um $\mu(\delta)$, o qual resulte em $\left\| \overline{\mathbf{A}} \overline{\mathbf{p}} - \overline{\mathbf{y}}^o \right\|_2 = \delta$. Este é o chamado Princípio da Discrepância de Morozov (1984). Parte-se do

pressuposto de que o erro (ou *discrepância*) δ , para $\mu(\delta)$, é consequência de ruído nos dados observados, e este é corretamente estimado. Sob certas premissas estatísticas para a contaminação por ruído dos dados experimentais, δ é o erro médio quadrático (EMQ) da sequência de realizações da variável aleatória que contamina as observações. Logo, dada uma sequência $\sigma_i, i = 1, \dots, N$ o erro médio quadrático é dado por

$$\delta = \sqrt{\frac{\sum_{i=1}^N \sigma_i^2}{N}}$$

Sob certas premissas estatísticas, o EMQ é uma estimativa do desvio padrão da variável aleatória que contamina as observações.

Logo, tendo uma estimativa para o desvio padrão do ruído a busca da solução utilizando o critério de Morozov pode ser feita utilizando o Algoritmo

Início

passo 1: $k = 0$

passo 2: $\mu_k = 0$

passo 3: $\sigma =$ desvio padrão do ruído dos dados (valor presumido pelo interprete)

passo 4: compute $\tilde{\mathbf{p}} = \min \left\{ \left\| \overline{\mathbf{A}} \mathbf{p} - \mathbf{y}^o \right\|_2 + \mu_k \left\| \mathbf{p} \right\|_2 \right\}$

passo 5: **Enquanto** $\left\| \overline{\mathbf{A}} \mathbf{p} - \mathbf{y}^o \right\|_2 > \sqrt{\frac{\sum_{i=1}^N \sigma_i^2}{N}}$ **faça**

passo 6: $k = k + 1$

passo 7: $\mu_k = \mu_{k-1} \cdot 10^{1/10}$

passo 7: compute $\tilde{\mathbf{p}} = \min \left\{ \left\| \overline{\mathbf{A}} \mathbf{p} - \mathbf{y}^o \right\|_2 + \mu_k \left\| \mathbf{p} \right\|_2 \right\}$

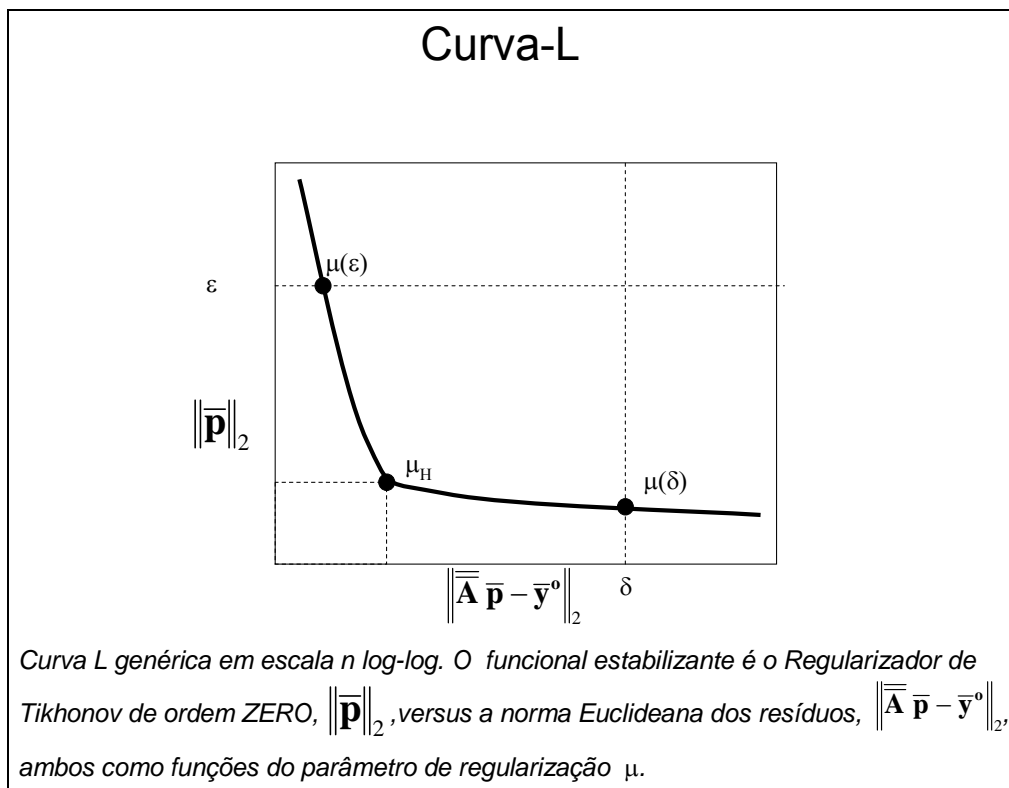
Fim Enquanto

Fim

Nós iremos nos concentrar na solução do problema (3) Mínimos quadrados amortecido.

Curva L

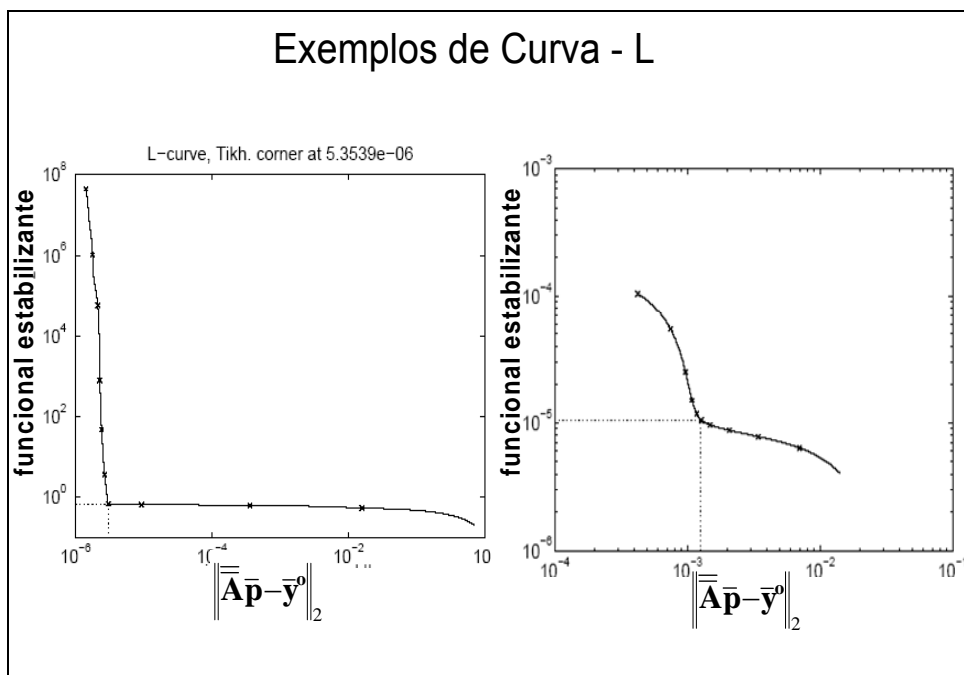
Vamos agora construir um gráfico na escala log-log da função $\|\bar{\mathbf{p}}\|_2$ versus $\|\bar{\mathbf{A}} \bar{\mathbf{p}} - \bar{\mathbf{y}}^0\|_2$ em função do parâmetro de regularização $\mu(\delta)$. Esta curva é chamada convenientemente de **curva-L**. Esta forma L ocorre porque $\|\bar{\mathbf{p}}\|_2$ é uma função estritamente decrescente com μ e $\|\bar{\mathbf{A}} \bar{\mathbf{p}} - \bar{\mathbf{y}}^0\|_2$ é uma função estritamente crescente com μ .



Note que em intervalos da curva-L mais próximos ao eixo vertical, uma variação pequena de μ , resulta numa acentuada variação de $\|\bar{\mathbf{p}}\|_2$. O mesmo se observa com relação a $\|\bar{\mathbf{A}} \bar{\mathbf{p}} - \bar{\mathbf{y}}^0\|_2$, para intervalos da curva-L mais próximos do eixo horizontal. Portanto, o passo logarítmico ameniza o efeito destas variações bruscas nestes intervalos críticos, tornando a busca de μ mais eficiente.

No gráfico, aparece um novo valor de parâmetro de regularização μ_H que corresponde um terceiro método de busca de μ , localizado no ponto da curva onde a sua curvatura é máxima. A escolha recai neste ponto, porque é onde se dá a relação mais equilibrada de suavidade(estabilidade) \times ajuste, e onde a diferença no efeito da variação de μ é reduzido. Denominado de método de Hansen (1992), ou método da curva-L, não requer nem conhecimento prévio de ε nem qualquer informação sobre o erro (ou *discrepância*) δ .

A nitidez do "canto" varia de um problema para outro, mas é frequentemente bem definido. Por esta razão, a curva é chamada uma curva L. Portanto além do princípio da discrepância (critério acima descrito) para escolher o valor de μ temos o critério da L-curva em que o valor de μ selecionado, μ_H , é aquele cuja a solução localiza-se o mais próximo ao canto da L-curva. Veja alguns exemplos reais de curva L.



V. A. Morozov. Methods for Solving Incorrectly Posed Problems. Springer– Verlag, New York, 1984.

Per Christian Hansen. Analysis of discrete ill-posed problems by means of the L–curve. SIAM Review, 34(4):561–580, December 1992.