

O Problema de Mínimos Quadrados

6 de setembro de 2019

1 Relações básicas

1.1 Canal de medidas linear

Suponha um vetor de grandezas x de dimensão n é desconhecido. Ele é medido (ou observado) pela i -ésima vez através de um canal de medidas imperfeito $H_i \in \mathbb{R}^{m \times n}$, de modo que a i -ésima medida y_i é um vetor de dimensão m tal que

$$y_i \simeq H_i x, \text{ para } i = 1, \dots, I$$

I é o número total de medidas. Note que esse problema é formulado para um canal de medidas lineares, de outra forma, a relação adequada seria representada por

$$y_i \simeq h_i(x), \text{ para } i = 1, \dots, I.$$

O canal pode ser invariante nas distintas medidas de forma que ele seria representado por uma matriz H fixa e $y_i \simeq Hx$ para toda medida i . Ao empilharmos os resultados de I medidas, adotamos a notação vetorial,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_I \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_I \end{bmatrix} \cdot x = \mathbf{H}x$$

Considere agora o conjunto de equações lineares sobredeterminadas,

$$\mathbf{y} \simeq Hx$$

com matriz $H \in \mathbb{R}^{N \times n}$ e $N \geq n$, $\mathbf{y}$ é um vetor de dimensão $N \times 1$ e x é um vetor desconhecido de dimensão $n \times 1$. A notação escolhida é tal que $N = I \times m$. O vetor $\mathbf{y}$ não pertence ao $\mathcal{R}(H)$ (range da matriz $\mathbf{H}$), de forma que

$$\mathbf{y} = \mathbf{H}x + v$$

em que v é o vetor resíduo, e não há escolha para x que faça $v \equiv 0$. A solução de mínimos quadrados (MQ), $\hat{x}$ é o vetor que minimiza o tamanho do vetor v , isto é,

$$\|\mathbf{y} - H\hat{x}\|^2 \leq \|\mathbf{y} - Hx\|^2 \quad (\text{dispensando os negritos}),$$

para todo $x \in \mathbb{R}^{n \times 1}$ em que $\|\cdot\|^2$ é a norma euclidiana. Com essa escolha,

$$\|v\| \text{ é o mínimo valor.}$$

Lemma 1 (Mínimos Quadrados). *O vetor $\hat{x}$ é um mínimo da função custo $J(x) = \|y - Hx\|^2$, se e somente se este satisfaz a equação normal,*

$$H^\top H \hat{x} = H^\top y$$

e o custo mínimo pode ser escrito como

$$J(\hat{x}) = \|y - H\hat{x}\|^2 = \|y\|^2 - y^\top H \hat{x} = \|y\|^2 - \|H\hat{x}\|^2$$

Se $H^\top H \succ 0$, então

$$\hat{x} = x + (H^\top H)^{-1} H^\top (y - Hx)$$

($A \succeq 0$ e $A \succ 0$ indicam respectivamente que a matriz A é semidefinida positiva ou definida positiva.)

Prova

$$\frac{\partial}{\partial x} \|y - Hx\|^2 = \frac{\partial}{\partial x} (y - Hx)^\top (y - Hx) = \frac{\partial}{\partial x} (y^\top y - y^\top Hx - x^\top H^\top y + x^\top H^\top Hx) \Big|_{x=\hat{x}} = 0$$

então $\hat{x}^\top H^\top H - y^\top H = 0$ deve ser satisfeita, além disso notamos que a matrix Hessiana,

$$\frac{\partial^2}{\partial x^\top \partial x} \|y - Hx\|^2 = H^\top H \succeq 0$$

é uma matrix semidefinida positiva.

Caso $H^\top H \succ 0$, isto é H tem posto completo n de colunas, a matriz $H^\top H$ é não singular e neste caso $\hat{x}$ é uma solução única dada por,

$$\hat{x} = (H^\top H)^{-1} H^\top y$$

1.1.1 Mínimos Quadrados Ponderado

Em muitas aplicações utiliza-se o critério de mínimos quadrados ponderado,

$$J(x) = \|y - Hx\|_W^2 = (y - Hx)^\top W (y - Hx)$$

Em que $W \succ 0$ é uma matriz Hermitiana definida positiva.

Lemma 2 (Mínimos Quadrados Ponderado). *A solução de mínimos quadrados ponderado das equações inconsistentes $y \simeq Hx$ são tal que*

$$\|y - H\hat{x}\|_W^2 \leq \|y - Hx\|_W^2$$

para $x \in \mathbb{R}^n$. Ela é dada pela solução da equação normal,

$$H^\top W H \hat{x} = H^\top W y$$

e o custo mínimo pode ser escrito como

$$J(\hat{x}) = \|y - H\hat{x}\|_W^2 = \|y\|_W^2 - y^\top W H \hat{x} = \|y\|_W^2 - \|\hat{x}\|_{H^\top W H}^2$$

Se $H^\top W H \succ 0$, então

$$\hat{x} = x + (H^\top W H)^{-1} H^\top W (y - Hx)$$

Note que até aqui não se tratou o problema de estimar o parâmetro x com estrutura probabilística, bastou utilizar aqui a idéia de resíduos para as medições.

2 Resíduo como ruído estocástico

Entenda agora o resíduo de uma forma estatística. A equação inconsistente e sobre determinada passa a ser a relação

$$y = Hx + \epsilon$$

em que x é um vetor constante (determinístico) desconhecido e ϵ é um ruído ou perturbação aleatória, com media conhecida $E[\epsilon] = 0$ e $\text{Cov}(\epsilon) = E[\epsilon\epsilon^\top] = R_\epsilon$. Nesse caso o estimador

$$\hat{x} = (H^\top W H)^{-1} H^\top W y$$

é variável aleatória. Sua média é

$$E[\hat{x}] = E[(H^\top W H)^{-1} H^\top W y] = (H^\top W H)^{-1} H^\top W E[y] = (H^\top W H)^{-1} H^\top W H x = x$$

portanto o estimador é não polarizado. Sua matrix de covariância é calculada como,

$$\text{Cov}(\hat{x} - x) = (H^\top W H)^{-1} H^\top W R_\epsilon W H (H^\top W H)^{-1} \quad (1)$$

Theorem 2.1 (Teorema de Gauss-Markov). *Considere o modelo de medidas $y = Hx + \xi$ em que ξ é um vetor de media nula em $\text{Cov}(\xi) = I$, x é determinístico e H tem posto completo de colunas. O estimador definido como $\hat{x} = (H^\top H)^{-1} H^\top y$ é o estimador de mínimos quadrados ótimo linear não polarizado de x .*

Quando a matriz de covariância é não unitária, $\text{Cov}(\epsilon) = R_\epsilon$ pode-se escalar a equação $y = Hx + \epsilon$ pela inversa de qualquer matriz L que satisfaça $R_\epsilon = LL^\top$. Com isso podemos escrever

$$L^{-1}y = L^{-1}Hx + L^{-1}\epsilon = L^{-1}Hx + \xi, \\ \text{com } \text{Cov}(\xi) = \text{Cov}(L^{-1}\epsilon) = I.$$

E de acordo como teorema de Gauss-Markov, o estimador de mínimos quadrados ótimo linear não polarizado de x dado as medidas transformadas $L^{-1}y$, é dado por

$$\hat{x} = (H^T L^{-T} L^{-1} H)^{-1} H^T L^{-T} L^{-1} y = (H^T R_\epsilon^{-1} H)^{-1} H^T R_\epsilon^{-1} y$$

o que implica que $W = R_\epsilon^{-1}$ é a escolha que conduz ao estimador ótimo. Nesse caso a covariância do estimador é obtida de (1), e é dada por

$$\text{Cov}(\hat{x} - x) = (H^T R_\epsilon^{-1} H)^{-1}$$

e é a expressão da covariância mínima, no sentido semidefinido positivo.

Note que esse resultado estende a noção que o estimador MQ faz sentido para v.a.'s gaussianas. Quando tomamos $W = R_\epsilon^{-1}$ para o caso gaussiano, isso faz que a solução do MQ coincida com a solução de máxima verossimilhança (MV).

Aqui não podemos fazer afirmações sobre a MV e MQ se as v.a.'s aleatórias envolvidas não forem gaussianas, porém o teorema de Gauss-Markov garante que a escolha $W = R_\epsilon^{-1} = \text{Cov}(\epsilon)^{-1}$ é a melhor escolha para o MQ ponderado no caso estocástico.

3 Mínimos Quadrados Regularizados

O ponto de partida do critério conhecido como Mínimos Quadrados Regularizado (MQR) é o critério MQ modificado,

$$\min_x (\|x - x_0\|_G^2 + \|y - Hx\|_W^2) \quad (2)$$

para um canal de medida m -dimensional definido por $H : \mathbb{R}^{n \times m}$, em que $y \in \mathbb{R}^m$ é o vetor de medidas na relação padrão,

$$y = Hx + \epsilon,$$

com $E\{\epsilon\} = 0$ e variância conhecida. $W \succeq 0$, como visto para o MQ estocástico, o teorema de Gauss-Markov indica a escolha $W = \text{Cov}(\epsilon)^{-1}$.

A introdução do par $x_0 \in \mathbb{R}^n$ e $G \in \mathbb{R}^{n \times n} \succeq 0$ é associada às informações *a priori*, a respeito da media e da covariância, respectivamente. Assim o critério em (2) pode ser visto como uma *versão bayesiana do MQ*, em que se consideram apenas os dois primeiros parâmetros estatísticos—media e variância—sendo análogo às noções fracas que se aplicam por exemplo, à noção de estacionariedade em outro cenário.

Lemma 3 (Mínimos Quadrados Regularizado). *Dado par $x_0 \in \mathbb{R}^n$ e $G \in \mathbb{H}^{n \times n} \succeq 0$, a solução de mínimos quadrado regularizado das equações inconsistentes $y \simeq Hx$ são tal que*

$$\|\hat{x} - x_0\|_G^2 + \|y - H\hat{x}\|_W^2 \leq \|x - x_0\|_G^2 + \|y - Hx\|_W^2$$

para todo $x \in \mathbb{R}^{n \times 1}$. Elas são dadas por soluções da equação normal

$$G + H^T W H \hat{x} = Gx_0 + H^T W y$$

e o custo mínimo pode ser escrito como

$$J(\hat{x}) = \|y\|_W^2 + \|x_0\|_G^2 - (y^\top W H - x_0^\top G) \hat{x} = \|y\|_W^2 - \|\hat{x}\|_{G+H^\top W H}^2$$

Se $G \succ 0$ e/ou $H^\top W H \succ 0$, então

$$\hat{x} = x_0 + (G + H^\top W H)^{-1} H^\top W (y - H x_0) \quad (3)$$

($\mathbb{H}^{n \times n}$ representa as matrizes hermitianas (matrizes simétricas no caso real) de dimensão n).

Como em geral $x_0 \neq x$, note que a diferença $y - H x_0$ é composta do resíduo/ruido ϵ , possivelmente acrescido de um valor determinístico diferente de zero.

Atualizando a solução: MQ recursivo Conforme se aumenta o número de observações I , o número total de dados $N = m \times I$ cresce. Seja porque esses dados são tomados sequencialmente ou seja porque desejamos diminuir a necessidade de armazenamento, é sempre conveniente um método recursivo para o MQ, que trate as medidas conforme elas se tornam disponíveis. O MQR serve de estrutura para essa ideia, o que será visto a seguir.

Suponha que no passo $i - 1$ foi resolvido o problema MQ para

$$\mathbf{y}_{i-1} \simeq \mathbf{H}_{i-1} x$$

com

$$\mathbf{H}_{i-1} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_{i-1} \end{bmatrix}, \text{ e } \mathbf{y}_{i-1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{i-1} \end{bmatrix}$$

Agora suponha que no passo i , temos uma nova medida y_i obtida pelo canal H_i , tal que

$$\begin{bmatrix} \mathbf{y}_{i-1} \\ y_i \end{bmatrix} \simeq \begin{bmatrix} \mathbf{H}_{i-1} \\ H_i \end{bmatrix} x$$

Resolvendo novamente o MQ ponderado, considere $\mathbf{W}_i = \text{Diag}[W_1 \ W_2 \ \cdots \ W_i]$, caso cada conjunto de medidas i tenha um peso distinto, ou $\mathbf{W}_i = \text{Diag}[W \ W \ \cdots \ W]$ formada com todos elementos diagonais idênticos, caso todos os conjuntos de medidas forem tomados com o mesmo peso. Em qualquer situação, $\mathbf{W}_i$ será formada por blocos diagonais de matrizes Hessianas semidefinidas positivas, pois supomos que as medidas não sejam correlacionadas entre si para diferentes observações i .

Tomando a notação $P_i = (\mathbf{H}_i^\top \mathbf{W}_i \mathbf{H}_i)^{-1}$, então $P_i \in \mathbb{R}^{n \times n}$, e obtemos a relação,

$$\begin{aligned} P_i^{-1} &= \mathbf{H}_i^\top \mathbf{W}_i \mathbf{H}_i = \begin{pmatrix} \mathbf{H}_{i-1}^\top & H_i^\top \end{pmatrix} \begin{pmatrix} \mathbf{W}_{i-1} & 0 \\ 0 & W_i \end{pmatrix} \begin{pmatrix} \mathbf{H}_{i-1} \\ H_i \end{pmatrix} \\ &= \mathbf{H}_{i-1}^\top \mathbf{W}_{i-1} \mathbf{H}_{i-1} + H_i^\top W_i H_i = P_{i-1}^{-1} + H_i^\top W_i H_i \quad (4) \end{aligned}$$

Interpretação: a informação (no sentido de Fisher) é equivalente ao inverso das covariâncias, e ela é atualizada tomando-se a informação até o estágio $i - 1$ adicionada à nova informação sobre x , obtida de y_i . Isso porque o parâmetro x é fixo e as observações são tomadas de forma *independentes entre si*.

Teremos então,

$$\begin{aligned}\hat{x}_i &= (\mathbf{H}_i^\top \mathbf{W}_i \mathbf{H}_i)^{-1} \mathbf{H}_i^\top \mathbf{W}_i \mathbf{y}_i = P_i \mathbf{H}_i^\top \mathbf{W}_i \mathbf{y}_i = P_i (\mathbf{H}_{i-1}^\top \mathbf{W}_{i-1} \mathbf{y}_{i-1} + H_i^\top W_i y_i) \\ &= P_i (\mathbf{H}_{i-1}^\top \mathbf{W}_{i-1} \mathbf{y}_{i-1} + H_i^\top W_i (y_i - H_i \hat{x}_{i-1}) + H_i^\top W_i H_i \hat{x}_{i-1}) \\ &= P_i (\mathbf{H}_{i-1}^\top \mathbf{W}_{i-1} \mathbf{H}_{i-1} \hat{x}_{i-1} + H_i^\top W_i H_i \hat{x}_{i-1} + H_i^\top W_i (y_i - H_i \hat{x}_{i-1}))\end{aligned}\quad (5)$$

pois, $\mathbf{H}_{i-1}^\top \mathbf{W}_{i-1} \mathbf{y}_{i-1} = \mathbf{H}_{i-1}^\top \mathbf{W}_{i-1} \mathbf{H}_{i-1} \hat{x}_{i-1}$ (eq. normal). Como também,

$$\mathbf{H}_{i-1}^\top \mathbf{W}_{i-1} \mathbf{H}_{i-1} \hat{x}_{i-1} + H_i^\top W_i H_i \hat{x}_{i-1} = \mathbf{H}_i^\top \mathbf{W}_i \mathbf{H}_i \hat{x}_{i-1} = P_i^{-1} \hat{x}_{i-1}$$

é verdadeiro, tem-se de (5) que,

$$\begin{aligned}\hat{x}_i &= \hat{x}_{i-1} + P_i H_i^\top W_i (y_i - H_i \hat{x}_{i-1}) \\ &= \hat{x}_{i-1} + (P_{i-1}^{-1} + H_i^\top W_i H_i)^{-1} H_i^\top W_i (y_i - H_i \hat{x}_{i-1})\end{aligned}\quad (6)$$

O que a eq. (6) permite concluir é que:

1. Dado um par $(\hat{x}_0, P_0)$ teremos da fórmula recursiva em (6), a forma do estimador recursivo para o MQ;
2. Por inspeção de (6) e da solução do MQR em (3) vemos que a solução recursiva é o ótimo do critério MQR,

$$\min_x J(x, y_i) = \min_x \|x - \hat{x}_{i-1}\|_{G_{i-1}}^2 + \|y_i - H_i x\|_{W_i}^2$$

em que $G_{i-1} = P_{i-1}^{-1}$.

3. Equação de atualização de P_i é,

$$P_i = (P_{i-1}^{-1} + H_i^\top W_i H_i)^{-1}.$$

Para completar o resultado para o MQ recursivo, é possível evitar a inversão direta da matriz P_i a cada passo indica no cálculo acima. O lema (ou a identidade) de inversão matricial garante que, para todas matrizes A, B, C e D de dimensões compatíveis, com A e C inversíveis, vale a identidade,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

Aplicada às necessidades aqui, temos

$$P_i = (P_{i-1}^{-1} + H_i^\top W_i H_i)^{-1} = P_{i-1} - P_{i-1} H_i^\top (W_i^{-1} + H_i P_{i-1} H_i^\top)^{-1} H_i P_{i-1} \quad (7)$$

1. Lembre-se que de acordo com o teorema de Gauss-Markov, para o caso estocástico devemos tomar $W_i^{-1} = \text{Cov}(\epsilon)$, ou $W_i^{-1} = \text{Cov}(\epsilon_i)$ se as diferentes medidas forem sujeitas à ruídos com características distintas.
2. As equações finais do MQ recursivo são

$$\begin{cases} \hat{x}_i = \hat{x}_{i-1} + P_i H_i^\top W_i (y_i - H_i \hat{x}_{i-1}), & (\text{eq. (6)}) \\ P_i = P_{i-1} - P_{i-1} H_i^\top (W_i^{-1} + H_i P_{i-1} H_i^\top)^{-1} H_i P_{i-1}, & (\text{eq. (7)}) \end{cases}$$

com alguma condição inicial $(\hat{x}_0, P_0)$ dada.

Atualizando a solução: MQ recursivo com fator de esquecimento Suponha que as medidas venham a se tornar menos significativas, por exemplo, no caso em que o parâmetro x não é fixo, mas varia lentamente. Com isso as medidas mais recentes têm maior significado e podemos adotar o MQ recursivo com a matriz de pesos no instante i , $W_i = \text{Diag}[W_1 \ W_2 \ \cdots \ W_i]$ de forma ad-hoc a atender essa situação. Dois casos simples podem ser utilizados,

1. *Janelamento Simples.* Neste caso tomamos $W_i = \text{Diag}[0 \ \cdots \ 0 \ W_{i-N} \ \cdots \ W_i]$ tal que $W_j = 0$ se $j < i - N$ e $W_j = W$ para $i - N \leq j \leq i$ a cada medida i . Significa que o tamanho da janela é N e as medidas anteriores ao índice $i - N$ serão descartadas no estágio i .

Para operar o algoritmo recursivo, temos que adotar uma estimativa “inicial” $\hat{x}_{i-N-1}$ com uma matrix P_{i-N-1} correspondente, e recalcular os N passos iterativamente a cada nova medida y_i . Isso não será diferente de resolver o MQR em lotes de tamanho N , tomando-se em (2),

$$x_0 = \hat{x}_{i-N-1}, \ G = P_{i-N-1}^{-1} \text{ e } W = \text{Diag}[P_{i-N} \ \cdots \ P_i].$$

2. *Fator de esquecimento geométrico.* Aqui toma-se para $0 < \lambda < 1$,

$$W_i = \text{Diag}[\lambda^{i-1}W \ \lambda^{i-2}W \ \cdots \ \lambda W \ W],$$

isto é, cada elemento de peso é dado por $W_j = \lambda^{i-j}W, j \leq i$. Da eq. (4) deduz-se que

$$P_i^{-1} = \lambda P_{i-1}^{-1} + H_i^\top W H_i$$

o que leva a equação recursiva (mostre!)

$$\begin{cases} \hat{x}_i = \hat{x}_{i-1} + P_i H_i^\top W (y_i - H_i \hat{x}_{i-1}), & (\text{eq. (6)}) \\ P_i = \frac{1}{\lambda} \left(P_{i-1} - P_{i-1} H_i^\top (\lambda W^{-1} + H_i P_{i-1} H_i^\top)^{-1} H_i P_{i-1} \right) \end{cases}$$

É fácil estender essas expressões para o caso variante com W_j e λ_j distintos para cada estágio j .

4 Mínimos Quadrados Não-Linear

4.1 Regressão não-linear: Família de curvas conhecidas

Exemplo: Identificação do atuador não-linear O levitador magnético com dois discos de ímãs permanentes e duas bobinas que se encontra disponível no laboratório é mostrado esquematicamente na Fig. 1.

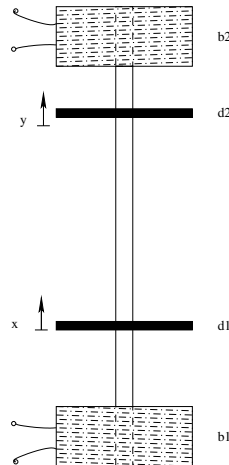


Figura 1: Levitador Magnético; emissor laser e sensor laser fixados em cada bobina.

O sistema é composto de duas bobinas que ao serem submetidas a corrente elétrica, criam campos magnéticos que vão interagir com os campos dos respectivos disco magnéticos permanentes. O sentido da corrente em cada bobina, bem como a polaridade de cada magneto são estabelecidos de forma a criar uma força mecânica resultante sobre cada disco magneto que seja:

- a) Repulsiva para o magneto inferior acionada pela bobina inferior,
- b) Atrativa para o magneto superior acionado pela bobina superior.

Utiliza-se um emissor de laser e um sensor correspondente, fixados na superfície de cada uma das bobinas, de forma a se fazer medidas de posição dos discos de forma independente e não intrusiva.

Exemplo Deseja-se através de algumas medidas obter o modelo da força de repulsão entre o disco e a bobina #1 quando se aplica uma corrente na bobina. Para isso faz-se algumas medidas, variando-se a corrente na bobina e mede-se a distância do disco, construindo a tabela seguinte. Sabe-se a massa do disco portanto o seu peso. No equilíbrio a força

Tabela 1: Medidas da interação entre a bobina #1 e disco #1.

posição y_1 (cm)	0,8	1,3	1,55	2,15	2,65	2,75	2,85	3,4	3,5
medidas u (mA)	300	400	500	600	700	800	900	1.000	1.100

magnética produzida iguala a força peso. A relação entre os campos magnéticos produz em forças que são descritas na forma geral por,

$$F(u, y) = \frac{a}{(y + b)^n} \cdot u$$

em que u é a corrente que passa pela bobina e y é a posição de equilíbrio do disco, e assim em equilíbrio, $F(u, y) = mg$. Pode haver variações devido à geometria mas em geral $n = 4$.

Exercício Deseja-se estimar através de medidas sucessivas a curva de calibração dos medidores de posição laser, isto é, determinar as funções $y_i = f_i(x_i)$, $i = 1, 2$ que estabeleça a relação entre a resposta do sensor x_i e o valor da posição y_i . Para o disco inferior foi obtido o conjunto de medidas mostrados na Fig. 2.

Obtenha com o Matlab a melhor curva de regressão no sentido de erro quadrado mínimo que represente a curva de calibração do medidor. Faça uso do comando '\ ' que calcula a solução.

Tabela 2: Medidas da distância do disco obtidas pelo sensor laser.

posição y_1 (cm)	0	0,5	1	2	3	4	5	6
medidas	11875	8709	7013	4536	2884	1852	1076	642

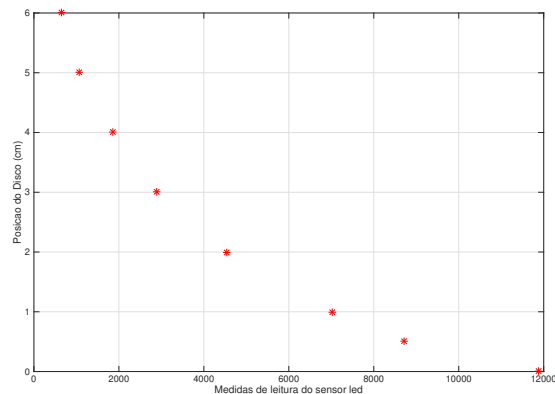


Figura 2: Medidas da distância do disco obtidas pelo sensor laser.

4.2 Mínimos Quadrados Iterado (ILS)

Suponha que se tenha uma única medida $y \in \mathbb{R}^m$ de um canal de medidas não linear,

$$y = h(x) + \epsilon$$

e a função $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é conhecida porém não linear e com $E\{\epsilon\} = 0$ e covariância R_ϵ . Suponha que temos uma estimativa inicial para x , $\hat{x}_0 = x_0$, com covariância Σ_0 . Inspirado no mínimos quadrados regularizados, MQR, podemos escrever o custo análogo,

$$J(y) = \min_x \left(\|x - x_0\|_{\Sigma_0^{-1}}^2 + \|y - h(x)\|_{R_\epsilon^{-1}}^2 \right) \quad (8)$$

porém esse problema não tem solução direta. Uma ideia que pode funcionar se houver convergência, ela irá garantir a obtenção de ao menos um mínimo local de J . Tomando a aproximação linear da função do canal de medidas h em torno da estimativa inicial, tem-se

$$h(x) = h(x_0) + H_0(x - x_0) + O(\|x - x_0\|^2)$$

em que $H_0 = \frac{\partial}{\partial x} h(x)|_{x=x_0}$ é a matriz jacobiana de h calculada em x_0 . Assim, (13) é dado de forma aproximada por

$$\tilde{J}(z_0) = \min_x \left(\|x - x_0\|_{\Sigma_0^{-1}}^2 + \|z_0 - H_0 x\|_{R_\epsilon^{-1}}^2 \right) \quad (9)$$

com $z_0 = y - h(x_0)$. A solução é conhecida, dada pelo Lema 3,

$$\hat{x}_1 = x_0 + (G + H_0^\top W H_0)^{-1} H_0^\top W (z_0 - H_0 x_0) \quad (10)$$

com $W = R_\epsilon^{-1}$ e $G = \Sigma_0^{-1}$. Note que o resíduo pode ser melhor calculado como $y - h(x_0)$ e com isso a forma final do estimador é

$$\hat{x}_1 = x_0 + (G + H_0^\top W H_0)^{-1} H_0^\top W (y - h(x_0)) \quad (11)$$

Podemos propor um algoritmo recursivo, notando-se que a matriz de covariância do estimador x_1 seria dada por

$$P_1^{-1} = \Sigma_0^{-1} + H_0^\top R_\epsilon^{-1} H_0$$

o que leva ao algoritmo MQ iterado, ou ILS (Iterated Least Square),

Dados x_0 e Σ_0 e a medida y , toma-se $k = 0$,

$$\begin{cases} \hat{x}_{k+1} = \hat{x}_k + P_k H_k^\top R_\epsilon^{-1} (y - h(x_k)), & \text{com } H_k = \frac{\partial}{\partial x} h(x)|_{x=x_0}, \\ P_k = (\Sigma_0^{-1} + H_k^\top R_\epsilon^{-1} H_k)^{-1}, & k = k + 1, \end{cases}$$

MQ não-linear com medidas sucessivas simplificar o texto

Suponha uma sequência de medidas única medida $y \in \mathbb{R}^m$ de um canal de medidas não linear,

$$y = h(x) + \epsilon$$

e a função $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é conhecida porém não linear e com $E\{\epsilon\} = 0$ e covariância R_ϵ . Suponha que temos uma estimativa inicial para x , $\hat{x}_0 = x_0$, com covariância P_0 . Inspirado no mínimos quadrados regularizados, MQR, podemos escrever o custo análogo,

$$J(y) = \min_x \left(\|x - x_0\|_{\Sigma_0^{-1}}^2 + \|y - h(x)\|_{R_\epsilon^{-1}}^2 \right) \quad (12)$$

porém esse problema não tem solução direta. Uma ideia que funciona e se convergir garante ao menos um mínimo local de J é obtido se tomarmos a aproximação linear em torno da estimativa inicial para h ,

$$h(x) = h(x_0) + H_0(x - x_0) + O(\|x - x_0\|^2)$$

em que $H_0 = \frac{\partial}{\partial x} h(x)|_{x=x_0}$ é a matriz jacobiana de h calculada em x_0 . Assim, (13) é dado de forma aproximada por

$$\tilde{J}(z_0) = \min_x \left(\|x - x_0\|_{\Sigma_0^{-1}}^2 + \|z_0 - H_0 x\|_{R_\epsilon^{-1}}^2 \right) \quad (13)$$

com $z_0 = y - h(x_0)$. A solução é conhecida, dada pelo Lema 3,

$$\hat{x}_1 = x_0 + (G + H_0^\top W H_0)^{-1} H_0^\top W (z_0 - H_0 x_0) \quad (14)$$

com $W = R_\epsilon^{-1}$ e $G = P_0^{-1}$. Podemos adotar, análogo ao algoritmo recursivo que a matriz de covariância do estimador x_1 seria dada por

$$P_1^{-1} = P_0^{-1} + H_0^\top W H_0$$

o que leva ao algoritmo MQ não-linear recursivo,

Dados x_0 e P_0 e a medida y , toma-se $k = 0$,

$$\begin{cases} \hat{x}_{k+1} = \hat{x}_k + P_k H_k^\top R_\epsilon^{-1} (y - h(x_k)), & \text{com } H_k = \frac{\partial}{\partial x} h(x)|_{x=x_0}, \\ P_{k+1} = P_k - P_k H_k^\top (R_\epsilon + H_k P_k H_k^\top)^{-1} H_k P_k, & k = k + 1, \end{cases}$$