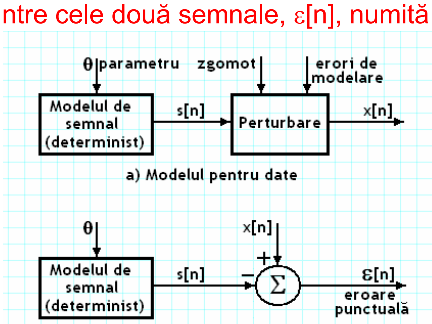
ESTIMAREA PRIN METODA CELOR MAI MICI PĂTRATE (LS)

METODA CELOR MAI MICI PĂTRATE

În capitolele precedente am căutat estimatori optimali sau asimptotic optimali, din categoria estimatorilor nedeplasați şi cu dispersie minimă (MVU). Pentru determinarea unor astfel de estimatori era necesară cunoaşterea densității de probabilitate a datelor sau măcar cunoaşterea momentelor de ordinul întâi şi doi ale acestora. Ce putem face dacă nu avem aceste informații ? În astfel de cazuri, vom aplica metoda celor mai mici pătrate (Least Squares), care conduce la un estimator care nu are, în cazul general, proprietăți de optimalitate asociate lui. Se presupune că semnalul util urmează un model cunoscut şi se minimizează suma pătratelor erorilor dintre date şi modelul de semnal. Ca urmare a modalității relativ simple de implementare, estimatorii bazați pe metoda LS sunt utilizați în mod frecvent.

1

Se presupune că semnalul util s[n] are un model cunoscut dar că depinde de un parametru necunoscut, θ, ce trebuie determinat. Semnalul util, s[n], aşa cum se arată în figură, este generat de un model (determinist), dependent de un parametru necunoscut θ. Semnalul util generat, s[n], este determinist, dar el este afectat de un zgomot de observare precum şi de erorile cauzate de diferența dintre model şi realitatea fizică. Acestea din urmă sunt erorile de model sau de modelare. În consecință datele sunt aleatoare şi doar ele sunt accesibile măsurării. Estimarea conform metodei celor mai mici pătrate, LS, va determina acea valoare a parametrului necunoscut θ care face ca semnalul să fie "cât mai apropiat" de datele măsurate. Gradul de apropiere a semnalului util s[n] de datele x[n], este măsurată, în metoda celor mai mici pătrate, prin energia diferenței dintre cele două semnale, ε[n], numită şi eroarea punctuală.



a) Modelul pentru date
b) Modelul pentru eroarea punctuală

2

Dependența de θ este realizată prin intermediul semnalului util, s[n]. Valoarea parametrului θ ce minimizează energia integrală a diferenței dintre cele două semnale, J(θ), este estimatorul LS. Se vede că nu am făcut nici o ipoteză privind statistica datelor. În consecință metoda LS este aplicabilă atunci când nu avem o caracterizare statistică a datelor x[n] sau în cazul în care estimatorul optimal nu se poate determina sau este prea complicat pentru a fi aplicat în practică.

$$J(\theta)=\sum_{n=0}^{N-1}\left(x[n]-s[n]\right)^2=\sum_{n=0}^{N-1}\varepsilon^2[n]$$

Pentru exemplificare, considerăm un semnal util de tip componentă continuă, A, căreia nu-i cunoaștem valoarea

$$s[n]=A$$

Pentru a determina estimatorul LS al componentei continue, A, anulăm derivata, după A, a funcției

$$J(A)=\sum_{n=0}^{N-1}\left(x[n]-A\right)^2$$

adică

$$\frac{dJ(A)}{dA}=-2\sum_{n=0}^{N-1}\left(x[n]-A\right)=-2\sum_{n=0}^{N-1}x[n]+2NA=0$$

Soluția ecuației este chiar estimatorul ML pentru componenta continuă, A

$$\hat{A}_{LS}=\frac{1}{N}\sum_{n=0}^{N-1}x[n]=\overline{x}$$

Estimatorul LS pentru componenta continuă este media eşantion, **indiferent de repartiția statistică a datelor** x[n]. Estimatorul nu este, în cazul general, optimal în sensul MVU. El minimizează doar eroarea LS. Dacă

$$x[n]=A+w[n]$$

în care zgomotul w[n] este alb, gaussian, de medie nulă, estimatorul LS este în acelaşi timp şi estimator MVU.

***

Un alt exemplu este cel al semnalului util sinusoidal căruia nu i se cunoaște frecvența digitală

$$s[n]=\cos 2\pi f_0 n;\quad n=0,1,\ldots,N-1$$

Determinarea estimatorului LS pentru frecvență impune minimizarea funcției

$$J(f_0)=\sum_{n=0}^{N-1}\left(x[n]-\cos 2\pi f_0 n\right)^2$$

Se anulează derivata funcției și se obține ecuația

$$\frac{dJ(f_0)}{dA}=2\sum_{n=0}^{N-1}\left(x[n]-\cos 2\pi f_0 n\right)\left(\sin 2\pi f_0 n\right)2\pi n=0$$

Fiind o ecuație neliniară, nu putem stabili o formulă pentru estimator. Cu datele **x** disponibile, se poate face o determinare numerică a estimatorului LS, așa cum am procedat pentru estimatorii ML

Dacă semnalul util este o funcție liniară de parametrul necunoscut, atunci funcția *J* este o funcție de gradul doi, al cărui minim se determină uşor. Se spune că un semnal ce depinde în mod liniar de parametrul necunoscut generează o problema liniară de cele mai mici pătrate. În celelalte cazuri se generează probleme neliniare de cele mai mici pătrate.

Drept exemplu considerăm semnalul util sinusoidal de amplitudine necunoscută

$$s[n] = A\cos 2\pi f_0 n; \quad n = 0,1,\ldots,N-1$$

Se construiește funcția J și se anulează derivata ei

$$J(A) = \sum_{n=0}^{N-1} \left(x[n] - A\cos 2\pi f_0 n\right)^2$$

$$\frac{dJ(A)}{dA} = -2\sum_{n=0}^{N-1} \left(x[n] - A\cos 2\pi f_0 n\right)\cos 2\pi f_0 n = 0$$

Ecuația se rezolvă ușor, deoarece este liniară în A

$$\hat{A}\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n = \sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n$$

Dacă frecvența nu este foarte aproape de 0 sau 0.5 se poate scrie că

$$\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n = \sum_{n=0}^{N-1} \frac{1+\cos 4\pi f_0 n}{2} = \frac{N}{2} + \frac{1}{2}\sum_{n=0}^{N-1} \cos 4\pi f_0 n \cong \frac{N}{2}$$

În final rezultă estimatorul LS pentru amplitudine

$$\hat{A}_{LS} = \frac{2}{N}\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n$$

5

Ne vom ocupa de problema LS liniară, pentru care dependența semnalului util de parametrul necunoscut aste liniară, de forma

$$s[n] = \theta h[n]; \quad n = 0,1,\ldots,N-1$$

în care h[n] este o funcție cunoscută iar parametrul necunoscut θ este o amplitudine. Fucția ce trebuie minimizată are expresia

$$J(\theta) = \sum_{n=0}^{N-1} \left(x[n] - \theta h[n]\right)^2$$

După derivare și egalarea cu zero a derivatei

$$\frac{dJ(\theta)}{d\theta} = -2\sum_{n=0}^{N-1} \left(x[n] - \theta h[n]\right)h[n] = 0$$

Se obține estimatorul LS în această problemă liniară, oricare ar fi forma funcției h[n]

$$\hat{\theta}_{LS} = \frac{\sum_{n=0}^{N-1} x[n]h[n]}{\sum_{n=0}^{N-1} h^2[n]}$$

6

Ne interesează uneori minimul erorii medii pătratice. Acest minim se obține substituind în funcția J estimatorul LS, care o minimizează

$$J_{\min} = J\left(\hat{\theta}_{LS}\right) = \sum_{n=0}^{N-1}\left(x[n]-\hat{\theta}h[n]\right)^2$$
$$= \sum_{n=0}^{N-1}\left(x[n]-\hat{\theta}h[n]\right)\left(x[n]-\hat{\theta}h[n]\right)$$
$$= \sum_{n=0}^{N-1}x[n]\left(x[n]-\hat{\theta}h[n]\right)-\hat{\theta}\sum_{n=0}^{N-1}h[n]\left(x[n]-\hat{\theta}h[n]\right)$$

Din relația de minimizare avem

$$\sum_{n=0}^{N-1}h[n]\left(x[n]-\hat{\theta}h[n]\right) = \sum_{n=0}^{N-1}x[n]h[n]-\hat{\theta}\sum_{n=0}^{N-1}h^2[n]=0$$

Substituim expresia estimatorului LS și concluzia de mai sus în expresia minimului functiei J și obținem

$$J_{\min} = \sum_{n=0}^{N-1}x^2[n]-\frac{\left(\sum_{n=0}^{N-1}x[n]h[n]\right)^2}{\sum_{n=0}^{N-1}h^2[n]}$$

7

Energiile semnalelor x[n] și h[n] sunt

$$\mathcal{E}_{\mathbf{x}} = \sum_{n=0}^{N-1}x^2[n] \ ; \quad \mathcal{E}_{\mathbf{h}} = \sum_{n=0}^{N-1}h^2[n]$$

Avem, pentru intercorelația acelorași semnale, expresia

$$r_{\mathbf{xh}} = \sum_{n=0}^{N-1}x[n]h[n]$$

Cu aceste notații energia minimă devine

$$J_{\min} = \mathcal{E}_{\mathbf{x}} - \frac{(r_{\mathbf{xh}})^2}{\mathcal{E}_{\mathbf{h}}} = \mathcal{E}_{\mathbf{x}}\left[1-\underbrace{\frac{(r_{\mathbf{xh}})^2}{\mathcal{E}_{\mathbf{x}}\mathcal{E}_{\mathbf{h}}}}_{\rho^2}\right] = \mathcal{E}_{\mathbf{x}}\left[1-\rho^2\right]; \ \rho\in[-1,\,1]$$

în care cu ρ s-a notat coeficientul de intercorelație dintre semnalele x[n] și h[n].

Pentru exemplificare ne referim la semnalul componentă continuă, pentru care

$$s[n] = A\cdot 1; \quad h[n]=1; \quad n=0,1,\ldots,N-1$$

Înlocuind în relația energiei minime obținem

$$J_{\min} = \sum_{n=0}^{N-1}x^2[n]-\frac{1}{N}\left(\sum_{n=0}^{N-1}x[n]\right)^2 = \sum_{n=0}^{N-1}x^2[n]-N\bar{x}^2$$

8

Reamintim inegalitatea Schwartz-Cauchy-Buniakovski

$$\left(\sum_{n=0}^{N-1} x[n]h[n]\right)^2 \le \sum_{n=0}^{N-1} x^2[n]\sum_{n=0}^{N-1} h^2[n]$$

în care egalul are loc dacă şi numai dacă x[n]=ah[n]. Se deduce imediat că

$$0 \le \rho^2 = \frac{(r_{\mathbf{xh}})^2}{\mathcal{E}_{\mathbf{x}}\mathcal{E}_{\mathbf{h}}} \le 1$$

În lipsa zgomotului coeficientul de corelație e 1 iar valoarea minimă a funcției J este zero, ceeace înseamnă că amplitudinea se determină “exact”. Avem

$$0 \le J_{\min} \le \sum_{n=0}^{N-1} x^2[n] = \mathcal{E}_{\mathbf{x}}$$

Metoda LS poate fi extinsă şi pentru cazul parametrului vector, pentru care

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$$

Estimatorul LS vector se determină minimizând funcția

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1}\left(x[n]-s[n]\right)^2 = (\mathbf{x}-\mathbf{s})^T(\mathbf{x}-\mathbf{s}) = (\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})$$

care poate fi pusă sub forma

$$J(\boldsymbol{\theta}) = \mathbf{x}^T\mathbf{x} - \mathbf{x}^T\mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{x} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} = \mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

Reamintim regulile de derivare pe care le-am mai utilizat

$$\frac{\partial}{\partial\boldsymbol{\theta}}(\mathbf{b}\boldsymbol{\theta}) = \mathbf{b};\quad \frac{\partial}{\partial\boldsymbol{\theta}}\left(\boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}\right) = 2\mathbf{A}^T\boldsymbol{\theta},\ \text{dacă } \mathbf{A}^T = \mathbf{A}$$

Aplicând regulile de3 derivare, se determină derivata funcției J sub forma

$$\frac{\partial J(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

După anularea derivatei se obțin p ecuații, care au forma matriceală

$$\mathbf{H}^T\mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}^T\mathbf{x}$$

Soluția este estimatorul vector LS, formal identic cu estimatorul BLU

$$\hat{\boldsymbol{\theta}}_{LS} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}$$

Se poate determina, cu estimatorul LS stabilit, minimul funcției J

$$\begin{aligned} J_{\min} &= J\left(\hat{\boldsymbol{\theta}}_{LS}\right) = \left(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}\right)^T\left(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\theta}}\right) \\ &= \left[\mathbf{x}-\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}\right]^T\left[\mathbf{x}-\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}\right] \\ &= \mathbf{x}^T\left[\mathbf{I}_u-\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right]\left[\mathbf{I}_u-\mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right]\mathbf{x} \end{aligned}$$

Deoarece paranteza dreaptă este, aşa cum am arătat deja, o matrice idempotentă, acest minim al funcției se poate pune sub forma

$$J_{\min} = \mathbf{x}^T \left[\mathbf{I}_u - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T \right]\mathbf{x}$$

sau, în final

$$J_{\min} = \mathbf{x}^T \left[\mathbf{x} - \mathbf{H}\underbrace{\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}}_{\hat{\boldsymbol{\theta}}} \right]$$

$$= \mathbf{x}^T\left[\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}\right]$$

Uneori forma LS liniară include o matrice de ponderare **W** de dimensiuni NxN

$$J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T\, \mathbf{W}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

Un caz particular al matricei de ponderare, des utilizat în practică este forma diagonală

$$\mathbf{W} = \begin{bmatrix} w_0 & 0 & \cdots & 0 \\ 0 & w_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{N-1} \end{bmatrix}$$

11

Pentru exemplificare, vom considera cazul în care matricea de ponderare este diagonală. Dacă semnalul util este s[n]=A, funcția J devine

$$J(A) = \sum_{n=0}^{N-1} w_n \left(x[n] - A\right)^2$$

O relație de această formă ne permite să ponderăm în mod diferit contribuția eşantioanelor de semnal în expresia funcției J. Fie că datele sunt de forma

$$x[n] = A + w[n];\quad n = 0,1,\ldots,N-1;\quad w[n] \sim \mathcal{N}\left(0,\ \sigma_n^2\right)$$

eşantioanele de zgomot normal w[n] sunt necorelate dar au dispersii diferite. Luând ca ponderi inversul dispersiilor, adică

$$w_n = 1/\sigma_n^2$$

funcția J şi derivata acesteia iau formele

$$J(A) = \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}\left(x[n] - A\right)^2;\qquad \frac{dJ(A)}{dA} = -2\sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}\left(x[n] - A\right)$$

Anulăm derivata şi obținem estimatorul LS pentru parametrul necunoscut, A

$$\hat{A}_{LS} = \sum_{n=0}^{N-1}\frac{x[n]}{\sigma_n^2} \Big/ \sum_{n=0}^{N-1}\frac{1}{\sigma_n^2}$$

12

Dacă matricea **W** este pozitiv definită, ea poate fi pusă sub forma unui produs

$$\mathbf{W} = \mathbf{D}^T\mathbf{D}$$

în care matricea **D**, de dimensiuni NxN, este inversabilă

Funcția J se poate pune sub forma

$$J(\boldsymbol{\theta}) = (\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{D}^T\mathbf{D}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}) = \left[\mathbf{D}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})\right]^T(\mathbf{D}\mathbf{x}-\mathbf{D}\mathbf{H}\boldsymbol{\theta})$$

$$= (\mathbf{D}\mathbf{x}-\mathbf{D}\mathbf{H}\boldsymbol{\theta})^T(\mathbf{D}\mathbf{x}-\mathbf{D}\mathbf{H}\boldsymbol{\theta}) = (\mathbf{x}'-\mathbf{H}'\boldsymbol{\theta})^T(\mathbf{x}'-\mathbf{H}'\boldsymbol{\theta})$$

Dacă datele **x** se prelucrează prin multiplicare cu matricea **D** se obțin

$$\mathbf{x}' = \mathbf{D}\mathbf{x} = \mathbf{D}\mathbf{H}\boldsymbol{\theta} + \mathbf{D}\mathbf{w}$$

sau

$$\mathbf{x}' = \mathbf{H}'\boldsymbol{\theta} + \mathbf{w}'$$

Adaptând relația de estimare vectorială LS anterior stabilită avem

$$\hat{\boldsymbol{\theta}}_{LS} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x} \;\rightarrow\; \hat{\boldsymbol{\theta}}_{LS} = \left(\mathbf{H}'^T\mathbf{H}'\right)^{-1}\mathbf{H}'^T\mathbf{x}'$$

$$= \left(\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{x}$$

$$= \left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\mathbf{x}$$

Se determină acum minimul funcției J

$$J_{\min} = \mathbf{x}'^T\left[\mathbf{I}_u - \mathbf{H}'\left(\mathbf{H}'^T\mathbf{H}'\right)^{-1}\mathbf{H}'^T\right]\mathbf{x}'$$

$$= \mathbf{x}^T\mathbf{D}^T\left[\mathbf{I}_u - \mathbf{D}\mathbf{H}\left(\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{D}^T\right]\mathbf{D}\mathbf{x}$$

$$= \mathbf{x}^T\left[\mathbf{D}^T\mathbf{D} - \mathbf{D}^T\mathbf{D}\mathbf{H}\left(\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{D}^T\mathbf{D}\right]\mathbf{x}$$

$$= \mathbf{x}^T\left[\mathbf{W} - \mathbf{W}\mathbf{H}\left(\mathbf{H}^T\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{W}\right]\mathbf{x}$$

O interpretare geometrică a estimării LS

Vom reexamina abordarea prin metoda LS liniară dând relațiilor interpretări geometrice. Modelul liniar corespunde semnalului **s=Hθ.** Dacă notăm coloanele matricei **H** cu **h**i, coloanele fiind vectori, semnalul util poate fi pus sub forma:

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_{N-1} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{N-1} \end{bmatrix} = \sum_{i=1}^{p} \theta_i \mathbf{h}_i$$

Semnalul util este deci o combinație liniară a vectorilor de semnal, **h**i.

15

Vom da ca exemplu analiza Fourier

Vom presupune că în semnalul util, s[n], este prezentă o singură armonică, adică

$$s[n] = a\cos 2\pi f_0 n + b\sin 2\pi f_0 n; \quad n = 0,1,\ldots,N-1$$

Parametrii necunoscuți, a şi b, sunt grupați în vectorul

$$\boldsymbol{\theta} = \begin{bmatrix} a & b \end{bmatrix}^T$$

Cele N relații de legătură între eşantioanele semnalului util şi necunoscutele a şi b, se pot pune într-o formă matriceală

$$\begin{bmatrix} s[0] \\ s[1] \\ \vdots \\ s[N-1] \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \cos 2\pi f_0 & \sin 2\pi f_0 \\ \vdots & \vdots \\ \underbrace{\cos 2\pi f_0 (N-1)}_{\mathbf{h}_1} & \underbrace{\sin 2\pi f_0 (N-1)}_{\mathbf{h}_2} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

16

Cei doi vectori coloană din structura matricei **H** sunt

$$\mathbf{h}_1 = \begin{bmatrix} 1 & \cos 2\pi f_0 & \cdots & \cos 2\pi f_0 (N-1) \end{bmatrix}^T$$

$$\mathbf{h}_2 = \begin{bmatrix} 0 & \sin 2\pi f_0 & \cdots & \sin 2\pi f_0 (N-1) \end{bmatrix}^T$$

Vectorul de semnal util, **s**, se poate exprima şi printr-o combinație liniară a celor doi vectori coloană, prin

$$\mathbf{s} = a\mathbf{h}_1 + b\mathbf{h}_2$$

Pentru un vector oarecare, ξ, având N componente

$$\boldsymbol{\xi} = \begin{bmatrix} \xi_1 & \xi_2 & \cdots & \xi_N \end{bmatrix}^T$$

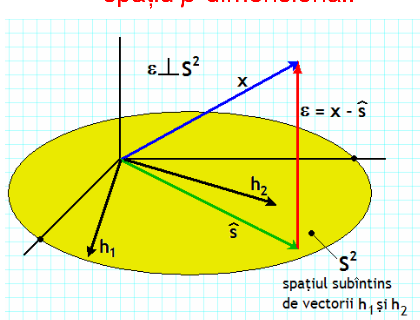
pătratul normei şi lungimea sa euclidiană se calculează cu relațiile

$$\|\boldsymbol{\xi}\|^2 = \boldsymbol{\xi}^T \boldsymbol{\xi} = \sum_{i=1}^{N} \xi_i^2; \qquad \|\boldsymbol{\xi}\| = \sqrt{\boldsymbol{\xi}^T \boldsymbol{\xi}_i} = \sqrt{\sum_{i=1}^{N} \xi_i^2}$$

Eroarea pătratică, J, ca energie a diferenței dintre datele **x** şi modelul de semnal util, **s=Hθ** se poate pune sub forma

$$J(\boldsymbol{\theta}) = \|\mathbf{x} - \mathbf{H}\boldsymbol{\theta}\|^2 = \left\| \mathbf{x} - \sum_{i=1}^{N} \theta_i \mathbf{h}_i \right\|^2$$

Abordarea LS liniară impune minimizarea pătratului unei distanțe de la vectorul de date **x** la un vector de semnal util **s=Hθ**, o combinație liniară a coloanelor matricei **H**. Vectorul **x** este un vector în spațiul real *N*-dimensional. Vectorii de semnal util posibili sunt *p*-dimensionali, p<N. Ei se află în spațiul real, *p*-dimensional, subîntins (generat) de vectorii **{$\mathbf{h}_1$,$\mathbf{h}_2$,…,$\mathbf{h}_p$ }**. Am admis că matricea are rangul *p* şi deci $\mathbf{h}_i$ sunt vectori liniar independenti, ei generând un spațiu *p*-dimensional.



În figură se ilustrează cazul N=3 şi p=2. Este evident că vectorul estimator LS al semnalului util, care are distanța minimă - în sens euclidian – la vectorul **x** este proiecția ortogonală a lui **x** pe subspațiul bidimensional, subîntins de vectorii $\mathbf{h}_1$ şi $\mathbf{h}_2$.

Vectorul eroare

$$\boldsymbol{\varepsilon} = \mathbf{x} - \hat{\mathbf{s}}$$

este, în consecință, ortogonal pe subspațiul bidimensional

$$S^2 \subset \mathbb{R}^3$$

şi deci pe orice vector din subspațiul bidimensional. Reamintim că doi vectori, ξ şi η, sunt ortogonali, dacă produsul lor scalar este nul

$$\xi^T \boldsymbol{\eta} = 0 \iff \xi \perp \boldsymbol{\eta}$$

Pentru cazul din figură

$$\mathbf{x} - \hat{\mathbf{s}} \perp \mathbf{S}^2 \iff \mathbf{x} - \hat{\mathbf{s}} \perp \mathbf{h}_1 \text{ şi, în acelaşi timp } \mathbf{x} - \hat{\mathbf{s}} \perp \mathbf{h}_2$$

Relația de mai sus este echivalentă cu relațiile

$$\left(\mathbf{x} - \hat{\mathbf{s}}\right)^T \mathbf{h}_1 = 0; \quad \left(\mathbf{x} - \hat{\mathbf{s}}\right)^T \mathbf{h}_2 = 0$$

Cele două relații de mai sus pot fi puse sub o formă matriceală

$$\left(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}\right)^T \left[\mathbf{h}_1 \quad \mathbf{h}_2\right] = \left[0 \quad 0\right]$$

Cu notațiile matriceale uzuale obținem

$$\left(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}\right)^T \mathbf{H} = \mathbf{0}^T$$

19

Efectuăm produsul şi obținem ecuația

$$\mathbf{H}^T \mathbf{x} - \mathbf{H}^T \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{0}$$

a cărei soluție este estimatorul LS pentru parametrul vector

$$\hat{\boldsymbol{\theta}}_{LS} = \left(\mathbf{H}^T \mathbf{H}\right)^{-1} \mathbf{H}^T \mathbf{x}$$

Vectorul de eroare, ε, se poate pune sub forma

$$\boldsymbol{\varepsilon} = \mathbf{x} - \hat{\mathbf{s}} = \mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}$$

cu care relația de ortogonalitate devine

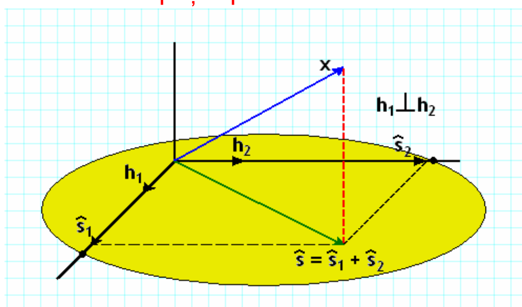
$$\boldsymbol{\varepsilon}^T \mathbf{H} = \mathbf{0}^T \iff \boldsymbol{\varepsilon} \perp \mathbf{H} \iff \boldsymbol{\varepsilon} \perp \mathbf{h}_i; \quad i = 1, 2, \ldots, p$$

Energia minimă a diferenței dintre date şi modelul de semnal se poate calcula cu relația

$$\begin{aligned}
J_{\min} &= \left(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}\right)^T \left(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}\right) = \left\|\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}\right\|^2 = \left\|\boldsymbol{\varepsilon}\right\|^2 \\
&= \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^T \left(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}\right) = \boldsymbol{\varepsilon}^T \mathbf{x} - \boldsymbol{\varepsilon}^T \mathbf{H}\hat{\boldsymbol{\theta}} = \boldsymbol{\varepsilon}^T \mathbf{x} = \left(\mathbf{x}^T - \hat{\boldsymbol{\theta}}^T \mathbf{H}^T\right) \mathbf{x} \\
&= \left[\mathbf{x}^T - \mathbf{x}^T \mathbf{H}\left(\mathbf{H}^T \mathbf{H}\right)^{-1} \mathbf{H}^T\right] \mathbf{x} = \mathbf{x}^T \left[\mathbf{I}_u - \mathbf{H}\left(\mathbf{H}^T \mathbf{H}\right)^{-1} \mathbf{H}^T\right] \mathbf{x}
\end{aligned}$$

20

În concluzie, abordarea LS poate fi interpretată ca fiind o problemă de aproximare a unui vector de date **x,** dintr-un spațiu N dimensional, printr-un vector **s** ce este o combinație liniară a p vectori **h**$_i$ care fac parte dintr-un subspațiu, cu dimensiunea p, unde p<N. Problema se rezolvă alegând pentru estimator proiecția ortogonală a vectorului de date din spațiul N dimensional pe subspațiul p dimensional.



Dacă vectorii **h**$_1$ şi **h**$_2$ sunt nu numai ortogonali ci şi ortonormali, ceeace presupune şi

$$\|\mathbf{h}_1\| = \|\mathbf{h}_2\| = 1$$

suma proiecțiilor vectorului estimator ML pe cei doi vectori ortonormali din subspațiul cu dimensiunea p=2, este

$$\hat{\mathbf{s}}_{ML} = \hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2 = \underbrace{\left(\mathbf{h}_1^T\mathbf{x}\right)}_{proiecție}\mathbf{h}_1 + \left(\mathbf{h}_2^T\mathbf{x}\right)\mathbf{h}_2$$

Folosind notații matriceale, relația anterioară devine

$$\hat{\mathbf{s}} = \begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2\end{bmatrix}\begin{bmatrix}\mathbf{h}_1^T\mathbf{x} \\ \mathbf{h}_2^T\mathbf{x}\end{bmatrix} = \begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2\end{bmatrix}\begin{bmatrix}\mathbf{h}_1^T \\ \mathbf{h}_2^T\end{bmatrix}\mathbf{x} = \mathbf{H}\mathbf{H}^T\mathbf{x}$$

Dar

$$\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}}$$

şi deci

$$\hat{\boldsymbol{\theta}} = \mathbf{H}^T\mathbf{x}$$

relație valabilă pentru o dimensiune p oarecare, dar p<N

Deoarece vectorii **h** sunt ortonormali, rezultă că **H** este o matrice unitară, pentru care matricea inversă este transpusa

$$\mathbf{H}^T\mathbf{H} = \mathbf{I}_u$$

Continuarea exemplului privind calculul coeficienților Fourier

Vom presupune că N este un număr par şi că frecvența digitală a fundamentalei este 1/N. Conform cu relațiile stabilite în Cap. 3 avem

$$\mathbf{h}_1^T\mathbf{h}_2 = \sum_{n=0}^{N-1} \cos 2\pi \frac{k}{N} n \sin 2\pi \frac{k}{N} n = 0$$

$$\mathbf{h}_1^T\mathbf{h}_1 = \|\mathbf{h}_1\|^2 = \sum_{n=0}^{N-1} \cos^2 2\pi \frac{k}{N} n = \frac{N}{2}$$

$$\mathbf{h}_2^T\mathbf{h}_2 = \|\mathbf{h}_2\|^2 = \frac{N}{2}$$

Calculând produsul

$$\mathbf{H}^T\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1^T\mathbf{h}_1 & \mathbf{h}_1^T\mathbf{h}_2 \\ \mathbf{h}_2^T\mathbf{h}_1 & \mathbf{h}_2^T\mathbf{h}_2 \end{bmatrix} = \begin{bmatrix} \frac{N}{2} & 0 \\ 0 & \frac{N}{2} \end{bmatrix} = \frac{N}{2}\mathbf{I}_u$$

deducem că vectorii **h** sunt ortogonali, dar nu ortonormali.

23

se pot determina acum estimatorii LS pentru amplitudinile a şi b

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x} = \frac{2}{N}\mathbf{H}^T\mathbf{x} = \frac{2}{N}\begin{bmatrix} \mathbf{h}_1^T\mathbf{x} \\ \mathbf{h}_2^T\mathbf{x} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{2}{N}\sum_{n=0}^{N-1} x[n]\cos 2\pi \frac{k}{N} n \\ \frac{2}{N}\sum_{n=0}^{N-1} x[n]\sin 2\pi \frac{k}{N} n \end{bmatrix}$$

În cazul general vectorii coloană **h** extraşi din matricea **H** nu sunt ortogonali şi estimatorul LS are expresia

$$\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x} = \mathbf{P}\mathbf{x}$$

în care matricea **P** de dimensiuni NxN defită cu relația

$$\mathbf{P} = \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T$$

este simetrică şi idempotentă

$$\mathbf{P}^T = \mathbf{P}; \qquad \mathbf{P}^2 = \mathbf{P}$$

24

Proiecția vectorului **x** pe subspațiul p dimensional este

$$\mathbf{Px}$$

Matricea **P** nu este inversabilă, ceeace înseamnă că din relația de estimare LS

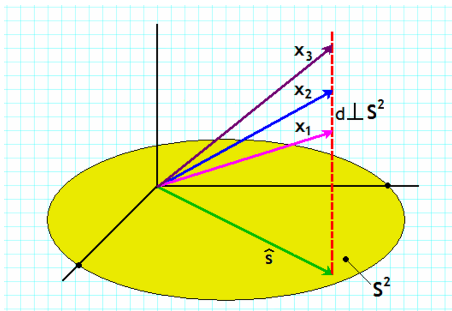
$$\hat{\mathbf{s}} = \mathbf{Px}$$

nu putem reconstrui vectorul de date. Aşa cum rezultă din figură, există mai mulți vectori **x** care au aceeaşi proiecție

Dacă vectorul odată proiectat pe subspațiu se proiectează a doua oară, rezultatul este acelaşi, adică

$$\mathbf{P}(\mathbf{Px}) = \mathbf{P}^2\mathbf{x} = \mathbf{Px}$$

Idempotența este deci stabilită



25

Matricea de proiecție **P** este pozitiv semidefinită. Pentru a arăta acest fapt, construim forma pătratică

$$\mathbf{x}^T\mathbf{Px} = \mathbf{x}^T\mathbf{P}^2\mathbf{x} = \mathbf{x}^T\mathbf{PPx} = \mathbf{x}^T\mathbf{P}^T\mathbf{Px}$$

$$= (\mathbf{Px})^T(\mathbf{Px}) = \|\mathbf{Px}\|^2 \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^N$$

şi ea este nenegativă, oricare ar fi vectorul **x**

Vectorul de eroare ε este ortogonal pe subspațiul p dimensional şi are forma

$$\boldsymbol{\varepsilon} = \mathbf{x} - \hat{\mathbf{s}} = \mathbf{x} - \mathbf{Px} = (\mathbf{I}_u - \mathbf{P})\mathbf{x}$$

Vectorul eroare aparține subspațiului complementar spațiului de aproximare. Matricea de proiecție pe acest subspațiu este, şi ea, simetrică şi idempotentă

$$\mathbf{P}^{\perp} = \mathbf{I}_u - \mathbf{P}; \quad (\mathbf{P}^{\perp})^T = \mathbf{P}^{\perp}; \quad (\mathbf{P}^{\perp})^2 = \mathbf{P}^{\perp}$$

Se poate acum calcula minimul funcției J

$$J_{\min} = \mathbf{x}^T\left[\mathbf{x} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}\right] = \mathbf{x}^T(\mathbf{x} - \mathbf{Px})$$

$$= \mathbf{x}^T(\mathbf{I}_u - \mathbf{P})\mathbf{x} = \mathbf{x}^T\mathbf{P}^{\perp}\mathbf{x} = \mathbf{x}^T\mathbf{P}^{\perp}\mathbf{P}^{\perp}\mathbf{x}$$

$$= (\mathbf{P}^{\perp}\mathbf{x})^T(\mathbf{P}^{\perp}\mathbf{x}) = \|\mathbf{P}^{\perp}\mathbf{x}\|^2 = \|\boldsymbol{\varepsilon}\|^2$$

care este pătratul lungimii (norma) vectorului de eroare, ε

26

Stabilirea recursivă a modelului semnalului util, aplicând metoda LS

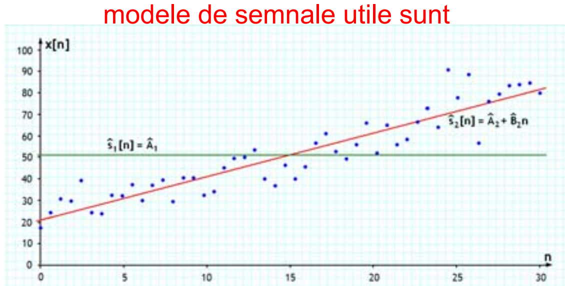
În unele situații nu suntem siguri în ceea ce priveşte modelul aplicabil semnalului util şi trebuie să adoptăm unul. Dacă ne referim la datele experimentale din figură, putem încerca, succesiv, două modele pentru semnalul util şi anume, mai întâi:

$$s_1[n] = A_1 \qquad n = 0,1,\ldots,N-1$$

și apoi

$$s_2[n] = A_2 + B_2 n \qquad n = 0,1,\ldots,N-1$$

şi să ne hotărâm apoi care dintre ele este mai potrivit. În mod evident, parametrii necunoscuți sunt A_1 şi apoi cuplul A_2, B_2 respectiv. Nu e greu să stabilim că cele două matrice de observare, **H**, corespunzătoare celor două modele de semnale utile sunt



27

$$\mathbf{H}_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad \mathbf{H}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix}$$

În cazul primului model, aplicarea estimării conduce la

$$\hat{A}_1 = \overline{x}$$

Pentru al doilea model efectuăm calculele pentru aplicarea estimatorului LS

$$\mathbf{H}^T\mathbf{H} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & N-1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_0^{N-1} 1 & \sum_0^{N-1} n \\ \sum_0^{N-1} n & \sum_0^{N-1} n^2 \end{bmatrix}$$

$$= \begin{bmatrix} N & \dfrac{N(N-1)}{2} \\ \dfrac{N(N-1)}{2} & \dfrac{N(N-1)(2N-1)}{6} \end{bmatrix}$$

28

Determinantul produsului are expresia

$$\left|\mathbf{H}^T\mathbf{H}\right| = \frac{N(N-1)(2N-1)}{6} - \frac{N^2(N-1)^2}{4}$$
$$= \frac{N^2(N^2-1)}{12}$$

Se poate acum inversa matricea produs

$$\left(\mathbf{H}^T\mathbf{H}\right)^{-1} = \frac{12}{N^2(N^2-1)}\begin{bmatrix} \frac{N(N-1)(2N-1)}{6} & -\frac{N(N-1)}{2} \\ -\frac{N(N-1)}{2} & N \end{bmatrix}$$
$$= \begin{bmatrix} \frac{2(2N-1)}{N(N+1)} & -\frac{6}{N(N+1)} \\ -\frac{6}{N(N+1)} & \frac{12}{N(N^2-1)} \end{bmatrix}$$

29

Calculăm produsul

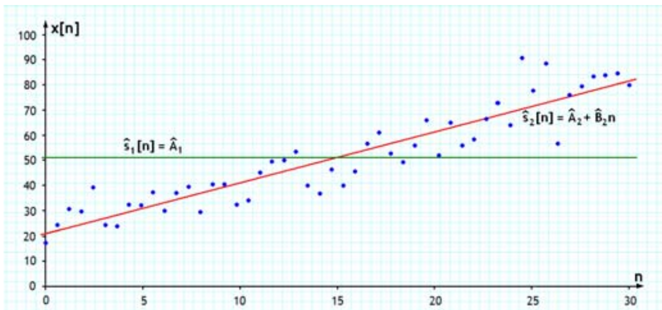
$$\mathbf{H}^T\mathbf{x} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & N-1 \end{bmatrix}\begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} = \begin{bmatrix} \sum_0^{N-1} x[n] \\ \sum_0^{N-1} nx[n] \end{bmatrix}$$

Înlocuind în relația de calcul a estimatorului LS și obținem

$$\begin{bmatrix} \hat{A}_2 \\ \hat{B}_2 \end{bmatrix}_{LS} = \begin{bmatrix} \frac{2(2N-1)}{N(N+1)} & -\frac{6}{N(N+1)} \\ -\frac{6}{N(N+1)} & \frac{12}{N(N^2-1)} \end{bmatrix}\begin{bmatrix} \sum_0^{N-1} x[n] \\ \sum_0^{N-1} nx[n] \end{bmatrix}$$
$$= \begin{bmatrix} \frac{2(2N-1)}{N(N+1)}\sum_0^{N-1} x[n] - \frac{6}{N(N+1)}\sum_0^{N-1} nx[n] \\ -\frac{6}{N(N+1)}\sum_0^{N-1} x[n] + \frac{12}{N(N^2-1)}\sum_0^{N-1} nx[n] \end{bmatrix}$$

Aceste formule de estimare au fost deduse și în Cap. 2

30



În figură sunt arătate datele (stabilite prin măsurare) şi cele două modele pentru semnalul util. Se observă că "potrivirea" dintre date şi model este mai bună pentru al doilea model, cel cu doi parametri. Pentru fiecare model se determină J_{min}. Pentru primul model avem J_{1min}. Pentru al doilea model se calculează $J_{2min} < J_{1min}$. Dacă se consideră un model parabolic cu necunoscutele A_3, B_3 și C_3, puteam obține J_{3min}, etc. Vom adăuga noi termeni modelului, până ce scăderea valorii J_{min} este nesemnificativă. În acest fel stabilim parametrii modelului pentru semnalul util. Trebuie spus că adăugând în model prea mulți termeni, acesta se va "potrivi" nu numai cu semnalul util ci şi cu zgomotul, ceea ce nu este de dorit.

31

Estimarea secvențială prin metoda celor mai mici pătrate

Îmbunătățirea modelului semnalului util prin adăugarea de noi termeni, ar fi bine să se facă în mod recursiv, utilizând rezultatele obținute pentru un model mai simplu. Vom aplica metoda LS în cazul unei matrice **H** de dimensiuni Nx(k+1), utilizând soluția obținută în cazul unei matrice **H** de dimensiuni Nxk.

Considerăm modelele succesive de semnal util

$$s_1[n] = A_1 \qquad -M \le n \le M$$

$$s_2[n] = A_2 + B_2 n \qquad -M \le n \le M$$

Pentru primul model matricea **H** se reduce la vectorul **1**, în timp ce pentru al doilea model de semnal

$$\mathbf{H}_2 = \begin{bmatrix} 1 & -M \\ 1 & -(M-1) \\ \vdots & \vdots \\ 1 & M \end{bmatrix}$$

Avem

$$\mathbf{H}_2^T \mathbf{H}_2 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ -M & -(M-1) & \cdots & M \end{bmatrix} \begin{bmatrix} 1 & -M \\ 1 & -(M-1) \\ \vdots & \vdots \\ 1 & M \end{bmatrix}$$

32

$$\mathbf{H}_2^T\mathbf{H}_2 = \begin{bmatrix} \sum_{-M}^{M} 1 & 0 \\ 0 & \sum_{-M}^{M} n^2 \end{bmatrix} = \begin{bmatrix} 2M+1 & 0 \\ 0 & \frac{M(M+1)(2M+1)}{6} \end{bmatrix}$$

Pentru cele două modele estimatorii LS sunt

$$\hat{A}_1 = \frac{1}{2M+1}\sum_{n=-M}^{M} x[n] = \bar{x} \qquad \hat{A}_2 = \bar{x}$$

$$\hat{B}_2 = \frac{\sum_{n=-M}^{M} n x[n]}{\sum_{n=-M}^{M} n^2}$$

Dacă vectorii **h** sunt ortogonali (sau ortogonalizați prin procedura Gram-Schmidt), se arată că sunt aplicabile relații de calcul recurent. Dacă $\mathbf{H}_k$ este matricea de observare de dimensiuni Nxk iar $\mathbf{H}_{k+1}$ este matricea de dimensiuni Nx(k+1), atunci

$$\hat{\boldsymbol{\theta}}_k = \left(\mathbf{H}_k^T\mathbf{H}_k\right)^{-1}\mathbf{H}_k^T\mathbf{x} = \mathbf{D}_k\mathbf{H}_k^T\mathbf{x}$$

$$J_{k\min} = \left(\mathbf{x} - \mathbf{H}_k\hat{\boldsymbol{\theta}}_k\right)^T\left(\mathbf{x} - \mathbf{H}_k\hat{\boldsymbol{\theta}}_k\right)$$

$$\mathbf{H}_{k+1} = \begin{bmatrix} \mathbf{H}_k & \mathbf{h}_{k+1} \end{bmatrix}$$