Least square interpretation:

$$X_n = H_n\theta + W_n, \qquad W_n \text{ iid} \sim N(0,\sigma^2)$$

$$\log L(x,\theta) = \underbrace{-\log(2\pi\sigma^2)^{N/2}}_{\text{const.}} - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(X_n - H_n\theta)^2$$

To ~~minimize~~ maximize $\log L(x,\theta)$ is to ~~maximize~~ minimize

$$\chi^2(\theta) = \sum_{n=1}^{N}(X_n - H_n\theta)^2$$

$$\hat{\theta} = \arg\min_{\theta}\left\{\sum_{n=1}^{N}(X_n - H_n\theta)^2\right\}$$

$\therefore$ $\hat{\theta}$ is the Least square sol. of

In the correlated case:

$$X = H\theta + W, \quad Cov(W W^T) = C$$

Cij = E(Wi*Wj)

p.d.f.

$$f(x, \theta) = \frac{1}{(2\pi)^{N/2} |C|^{1/2}} \exp\left\{-\frac{1}{2}(x - H\theta)^T C^{-1} (x - H\theta)\right\}$$

MLE is to max:

$$\log L(x,\theta) = -\log\left\{(2\pi)^{\frac{N}{2}} |C|^{\frac{1}{2}}\right\} - \underbrace{\frac{1}{2}(x - H\theta)^T C^{-1} (x - H\theta)}_{\text{this is quardratic}}$$

let $\chi^2(\theta) = (x - H\theta)^T C^{-1} (x - H\theta)$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - H_i\theta)(C^{-1})_{ij}(x_j - H_j\theta)$$

$$\hat{\theta} = \arg\min_{\theta} \left\{ \chi^2(\theta) \right\}$$

Again $\hat{\theta}$ is an LS solution to

---

* The MLE assume Gaussian noise.
* LS does not assume anything on noise. In fact, it does not think in terms of noise. It thinks in terms of model erros, and min. total error sq. sums.

Note: $\chi^2 = -2 \log L(x, \theta)$

$$\frac{\partial}{\partial\theta}\chi^2 = -2\frac{\partial}{\partial\theta}\log L(x,\theta)$$

$$\frac{\partial^2}{\partial\theta^2}\chi^2 = -2\frac{\partial^2}{\partial\theta^2}\log L(x,\theta)$$

Since $E\left(-\frac{\partial^2}{\partial\theta^2}\log L(x,\theta)\right) = \underset{\uparrow \text{ info.}}{I(\theta)}$

and $\operatorname{Var}(\hat{\theta}) \approx I^{-1}(\theta)$ for larg samples.

$$\Rightarrow \operatorname{Var}(\hat{\theta}) \approx 2^{-1}\left\{\frac{\partial^2 \chi^2(\theta)}{\partial\theta^2}\right\}^{-1}\Bigg|_{\theta=\hat{\theta}}$$

For a given set of obs., if $H_n \hat{\theta}$ matches x_n perfectly, $x_n - H_n \hat{\theta} = 0$

and we have $\chi^2(\hat{\theta}) = 0$.

$\therefore$ $\chi^2(\hat{\theta})$ value is an indicator for the "goodness-of-fit".

Since $\hat{\theta}$ depends on noisy data, $\chi^2(\hat{\theta})$ is a random variable.

It was shown that $\chi^2(\hat{\theta})$ follows the chi-square p.d.f.

$$f(\chi^2(\hat{\theta}) = t ; n_d) = \frac{1}{2^{n_d/2}\,\Gamma(n_d/2)}\, t^{n_d/2 - 1} \cdot e^{-t/2}$$

where n_d is the # deg. of free-dom of the chi-sq pdf

$$n_d = \#\text{data points} - \#\text{fitted para.}$$

$$= N - p$$

$$E(t) = \int t \cdot f(t\,;n_d)\,dt = n_d$$

The p-value:

$$p = \int_{\chi^2(\hat{\theta})}^{\infty} f(t\,;n_d)\,dt$$

= prob. of obtaining a $\chi^2(\hat{\theta})$ as high as the one we got or higher, if $\hat{\theta} = \theta$

$\Rightarrow$ Small $\chi^2(\hat{\theta})$: $\rightarrow$ good fit.

Small p-value : $\rightarrow \hat{\theta} = \theta$ is likely wrong.

Curvature of $\chi^2(\hat{\theta})$ near its minimum

flat $\longrightarrow$ Var. of $\hat{\theta}$ large

In the case of Gaussian i.i.d. noise,

$$E(w_i w_j) = \begin{cases} \sigma^2 & i=j \\ 0 & i \neq j \end{cases}$$

$$C = E(w\, w^T) = \begin{bmatrix} \sigma^2 & & & 0 \\ & \sigma^2 & & \\ & & \ddots & \\ 0 & & & \sigma^2 \end{bmatrix}_{N \times N}$$

$$\begin{aligned}
\chi^2(\theta) &= (x - H\theta)^T C^{-1} (x - H\theta) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - H_i\theta) \underbrace{(C^{-1})_{ij}}_{= \begin{cases} \frac{1}{\sigma^2} & i=j \\ 0 & i\neq j \end{cases}} (x_j - H_j\theta) \\
&= \sum_{i=1}^{N} \underbrace{(x_i - H_i\theta)}_{w_i} \frac{1}{\sigma^2} (x_i - H_i\theta)
\end{aligned}$$

$$E(\chi^2(\theta)) = \frac{1}{\sigma^2} \sum_{i=1}^{N} E(w_i^2) = \frac{1}{\sigma^2} \sum_{i=1}^{N} \sigma^2 = N$$

when $p = 1$, θ is scalar, H is $N \times 1$ $= H_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

$$\hat{\theta} = (H^T H)^{-1} H^T x = \left(\sum_{n=1}^{N} H_n^2\right)^{-1} \sum_{n=1}^{N} H_n x_n$$

$$\chi^2(\hat{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^{N} \left(x_i - H_i \hat{\theta}\right)^2$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{N} \left\{ x_i - \frac{\sum_{n=1}^{N} H_i H_n x_n}{\sum_{n=1}^{N} H_n^2} \right\}^2$$

$$E(\chi^2(\hat{\theta})) = \frac{1}{\sigma^2} \sum_{i=1}^{N} E \left\{ x_i - \frac{\sum x_n}{N} \right\}^2$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{N} \left\{ \sigma^2 - 2\frac{\sigma^2}{N} + \frac{N\sigma^2}{N^2} \right\}$$

$$= N - 1$$

In general: $E\{\chi^2(\hat{\theta})\} = N - p = n_d$

$\Rightarrow$ Use $\dfrac{\chi^2(\hat{\theta})}{N-p}$ ← actual realized value.

as an indicator of "goodness of fit".

Notice that this is normalized w.r.t σ^2

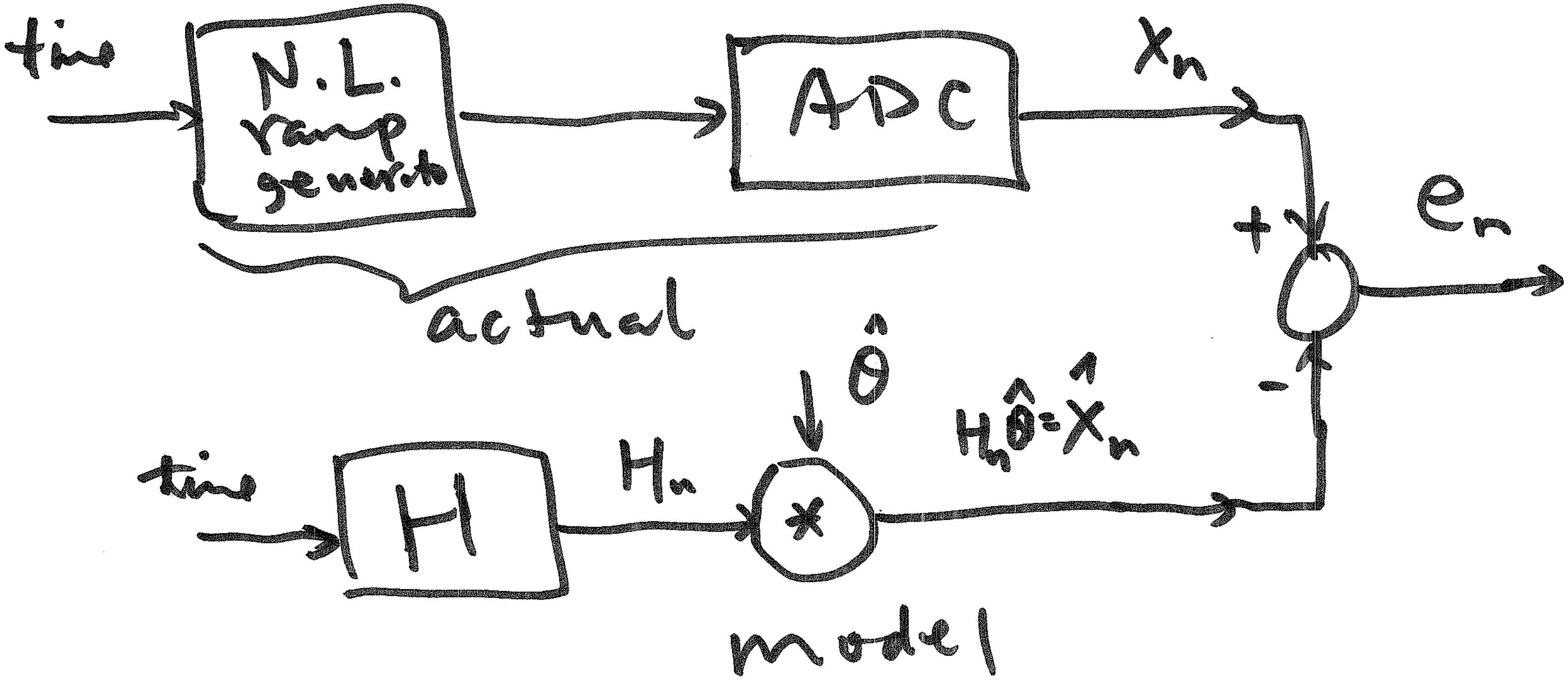
since $\chi^2(\hat{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^{N} (X_i - H_i \hat{\theta})^2$

Let $e^2_{tot}(\hat{\theta}) = \sum_{i=1}^{N} (X_i - H_i \hat{\theta})^2$

$\Rightarrow E\{e^2_{tot}(\hat{\theta})\} = (N-1)\sigma^2$

$$\Rightarrow \quad \frac{e^2_{tot}(\hat{\theta})}{N-p} = (\#) * \sigma^2$$

In the ramp N.L. measurement case:



As $P\uparrow$, $e^2_{tot}(\hat{\theta})$ will $\downarrow$

but $\dfrac{e^2_{tot}}{N-P}$ may $\downarrow$ and then $\uparrow$

$\Rightarrow$ help in determining # basis functions to use.

A-nother view at $\chi^2(\hat{\theta})$, or e^2_{tot}

When w_i i.i.d. : $\quad \hat{\theta} = (H^T H)^{-1} H^T x$

$$x = H\theta + ? + w$$

$$e^2_{tot} = (x - H\hat{\theta})^T (x - H\hat{\theta})$$

$$= x^T x - 2x^T H\hat{\theta} + \hat{\theta} H^T H \hat{\theta}$$

$$= x^T x - 2x^T H\hat{\theta} + x^T H \{ [\underset{\substack{\uparrow \\ \text{sym.}}}{(H^T H)^{-1}}]^T H^T H \hat{\theta}$$

$$= x^T x - 2x^T H\hat{\theta} + x^T H \underbrace{(H^T H)^{-1} (H^T H)} \hat{\theta}$$

$$= x^T x - 2x^T H\hat{\theta} + x^T H \hat{\theta}$$

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

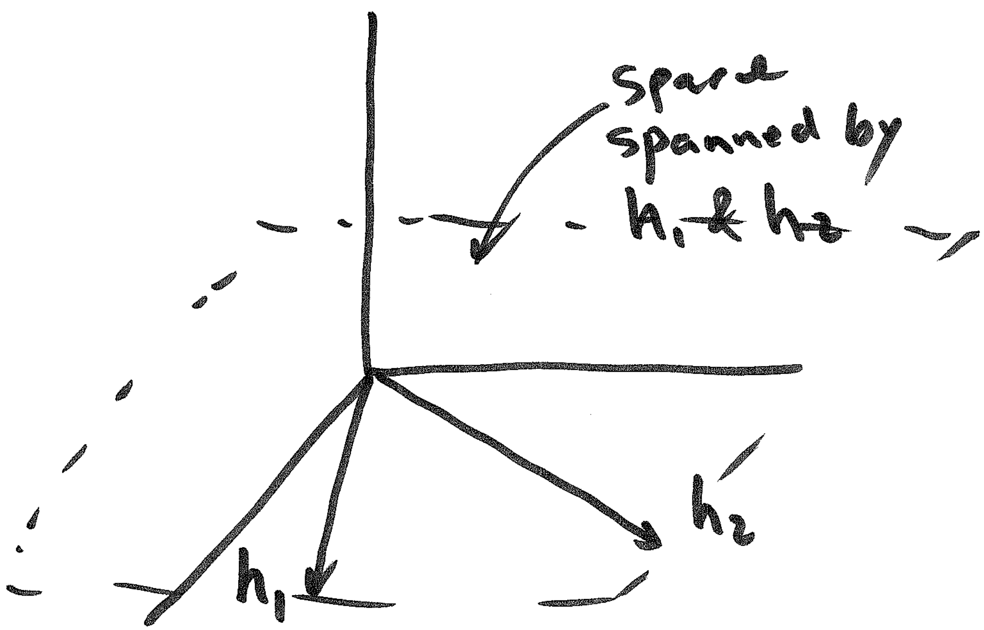
$$= x^T (x - H\hat{\theta}) = \langle \text{data}, \text{res. error} \rangle$$

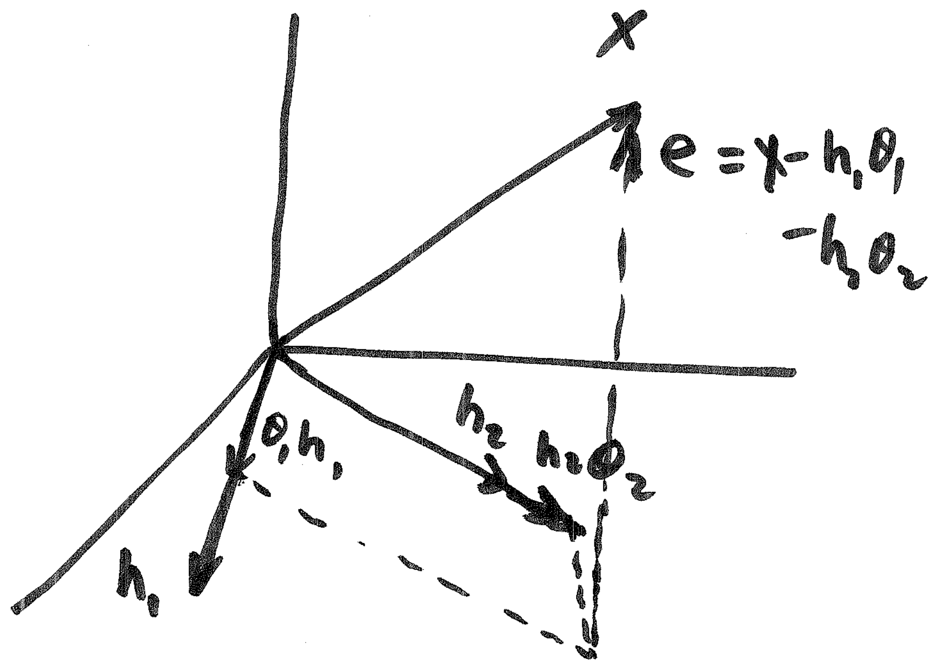
For correlated w : $\quad C = E(ww^T)$

$$\hat{\theta} = (H^T C^{-1} H)^{-1} H^T C^{-1} x$$

$$\chi^2(\hat{\theta}) = x^T C^{-1} (x - H\hat{\theta})$$

$$= \langle x, \; x - H\hat{\theta} \rangle_{C^{-1}}$$





When $h_1, h_2, \cdots h_p$ are orthogonal,

$$X = H\theta + ? + W$$

$$= h_1\theta_1 + h_2\theta_2 + \cdots + h_p\theta_p + ? + W$$

$$h_i^T X = \underbrace{h_i^T h_1}_{=0} \theta_1 + \cdots + h_i^T h_i \theta_i + \cdots + \cdots$$

$$= \underbrace{h_i^T h_i}_{\substack{=\text{const} \\ \overset{\text{scale}}{\equiv} 1}} \theta_i + \underbrace{h_i ? + h_i^T W}_{\text{ignore}}$$

$$\Rightarrow \hat{\theta}_i = h_i^T X \Rightarrow \theta = H^T X$$

Correlated W Case :

let $\langle x, y \rangle_{C^{-1}} = x^T C^{-1} y$, weighted inner prod.

If $\langle h_i, h_j \rangle_{C^{-1}} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

$$\langle h_i , x \rangle_{C^{-1}} = \quad \cdots \qquad \cdots$$

$$= \langle h_i , h_i \rangle_{C^{-1}} \theta_i + \cancel{\langle h_i , \ell + w \rangle_{C^{-1}}} \to 0$$

$$\hat{\theta}_i = \langle h_i , x \rangle$$

$$\hat{\theta} = H^T C^{-1} X$$

The orthogonality makes it really convenient to recursively add additional $\begin{cases} h_{p+1} & \text{basis function} \\ \theta_{p+1} & \text{parameter} \end{cases}$

$$H_{N\times p} \rightarrow [\underset{N\times p}{H} \ \underset{N\times 1}{h_{p+1}}]$$

$$\hat{\theta}_{p+1} = \langle h_{p+1}, x \rangle_{C^{-1}}$$

$$\begin{aligned}
e^2_{tot}(\hat{\theta})_{p+1} &= x^T C^{-1}(x - [H \ h_{p+1}]\hat{\theta}) \\
&= x^T C^{-1}(x - H\hat{\theta}_{1:p} - h_{p+1}\hat{\theta}_{p+1}) \\
&= x^T C^{-1}(x - H\hat{\theta}_{1:p}) - x^T C^{-1} h_{p+1}\hat{\theta}_{p+1} \\
&= e^2_{tot}(\hat{\theta})_p - x^T C^{-1} h_{p+1}\hat{\theta}_{p+1} \\
&= e^2_{tot}(\hat{\theta})_p - (x - H\theta_{1:p})^T C^{-1} h_{p+1}\hat{\theta}_{p+1}
\end{aligned}$$

start: $\quad p=0, \quad \hat{\theta}=[\,], \quad H=[\,], \quad e_{tot}^2=\langle x, x\rangle$

Select h_{p+1} to be orthog. to H

$$\hat{\theta}_{p+1}=\langle h_{p+1}, x\rangle_{C^{-1}}$$

$$\hat{\theta}=\begin{bmatrix}\hat{\theta}\\ \hat{\theta}_{p+1}\end{bmatrix}, \quad H=[H \;\; h_{p+1}]$$

$$e_{tot}^2 = e_{tot}^2 - \langle x, h_{p+1}\hat{\theta}_{p+1}\rangle_{C^{-1}}$$

If $|\hat{\theta}_{p+1}| <$ crit.
or change in $e_{tot}^2 <$ crit $\Big\}$ done.

else go to

$p = p+1$

General order-recursive L.S.

Given X, $N \times 1$ obs.

Suppose with $H = [h_1 \ h_2 \cdots h_p]$, we have

$$\underset{\substack{\uparrow \\ p\times 1}}{\hat{\theta}} = (H^T H)^{-1} H X$$

$$e^2_{tot}(\hat{\theta}) = (X - H\hat{\theta})^T (X - H\hat{\theta})$$

To add anothe basis function h_{p+1} & para θ_{p+1},

$$X = H\hat{\theta} + h_{p+1}\theta_{p+1} + \text{new residual}$$

Let P be a projection ~~into~~ orthog. space $\perp H$,

$$\Rightarrow \quad P = I - H(H^T H)^{-1} H^T$$

[To verify : if y is spanned by H,

i.e. $\underset{N\times 1}{y} = \underset{N\times p}{H} \cdot \underset{p\times 1}{c}$

then $Py = IHC - H\underbrace{(H^TH)^{-1}H^TH}_{I}C$

$= 0$

no component in orthog. space]

$$\Rightarrow \quad Px = \cancel{PH\hat{\theta}}^{\to 0} + Ph_{p+1}\theta_{p+1} + \underbrace{P \cdot \text{resid}}_{\text{ignore}}$$

$$Px = Ph_{p+1}\hat{\theta}_{p+1}$$

$$h_{p+1}^T Px = \underbrace{h_{p+1}^T P h_{p+1}}\,\hat{\theta}_{p+1}$$

size of comp. of h_{p+1} in orth. space

$$\hat{\theta}_{p+1} = \frac{h_{p+1}^T P x}{h_{p+1}^T P h_{p+1}}$$

plug back:

$$x = H\hat{\theta} + h_{p+1}\hat{\theta}_{p+1} + \cdots$$

$$x - \frac{h_{p+1}h_{p+1}^T P x}{h_{p+1}^T P h_{p+1}} = H\hat{\theta} + \cdots -$$

$$\hat{\theta}_{new} = (H^TH)^{-1}H^T(\quad)$$

$$= (H^TH)^{-1}H^Tx - \frac{(H^TH)^{-1}H^T \overbrace{h_{p+1}\, h_{p+1}^T \overbrace{Px}^{x \text{ comp. in } H^{\perp}}}^{\text{proj on } h_{p+1} \text{ axis}}}{h_{p+1}^T P h_{p+1}}$$

$$= \hat{\theta}_{old} - (H^TH)^{-1}H^T \{x \text{ comp.} \perp H \ \& \ \parallel h_{p+1}\}$$

$$\Rightarrow \begin{cases} H = [H \quad h_{p+1}] \\ \hat{\theta} = \begin{bmatrix} \hat{\theta}_{new} \\ \hat{\theta}_{p+1} \end{bmatrix} \\ e^2_{tot}(\hat{\theta}) = e^2_{tot\ prev} - \dfrac{(h_{p+1}^T P x)^2}{h_{p+1}^T P h_{p+1}} \end{cases}$$

Note: 1). If $h_{p+1} \perp H$, ~~no~~ $\hat{\theta}_{old} = \hat{\theta}_{new}$

2) Px is x comp ~~in~~ $\perp H$

3) If h_{p+1} comp in $\perp H$, i.e. Ph_{p+1}, is small, $h_{p+1}^T P h_{p+1}$ small $\Rightarrow$ alg. blow up.

So, pick h_{p+1} to be $\perp$ to H

4) e_{tot}^2 is reducing, can stop when it is reducing slowly.

$$\frac{e_{tot}^2}{N-P}$$ may bottom & go up again as $p \uparrow$

5) No need to compute $(H^T H)^{-1}$ with matrix inversion.

Let $D_p = (H^T H)^{-1}$, when $p = 1$

$D_p = 1/h_1^T h_1$.

$$D_{p+1} = \begin{bmatrix} D_p + \dfrac{D_p H^T h_{p+1} h_{p+1}^T H D}{h_{p+1}^T P h_{p+1}} & -\dfrac{D_p H h_{p+1}}{\cdot\cdot} \\ (\text{same})^T & \dfrac{1}{\cdot\cdot} \end{bmatrix}$$

$$= \begin{bmatrix} p \times p & p \times 1 \\ 1 \times p & 1 \times 1 \end{bmatrix}$$

Proof of 5) above:

$$\theta_{new} = \begin{pmatrix} \theta \\ \theta_{p+1} \end{pmatrix}, \quad H_{new} = [H$$ [redacted] $$h_{p+1}]$$

$$\hat{\theta}_{new} = (H_{new}^T H_{new})^T H_{new}^T x$$

$$= \begin{bmatrix} \downarrow \\ {}_{(p+1)\times(p+1)} \end{bmatrix} \begin{bmatrix} H^T x \\ h_{p+1}^T x \end{bmatrix}$$

$$\overset{\text{match}}{=} \begin{bmatrix} (H^TH)^{-1}H^Tx - (H^TH)^{-1}H^T \dfrac{h_{p+1} h_{p+1}^T P x}{h_{p+1}^T P h_{p+1}} \\[2ex] \dfrac{h_{p+1}^T P x}{h_{p+1}^T P h_{p+1}} \end{bmatrix}$$

$$P = I - H(H^TH)^{-1}H^T, \quad Px = x - H(H^TH)^{-1}H^Tx$$

$$
= \left[\begin{array}{c:c}
(H^TH)^{-1}H^Tx + \dfrac{(H^TH)^{-1}H^Th_{p+1}h_{p+1}^TH(H^TH)^{-1}H^Tx}{h_{p+1}^T P h_{p+1}} & -\dfrac{(H^TH)^{-1}H^Th_{p+1}h_{p+1}^Tx}{h_{p+1}^T P h_{p+1}} \\[3ex]
\dfrac{-h_{p+1}^TH(H^TH)^{-1}H^Tx}{h_{p+1}^T P h_{p+1}} \quad + & \dfrac{h_{p+1}^Tx}{h_{p+1}^T P h_{p+1}}
\end{array}\right]
$$

$$
= \underbrace{\left[\begin{array}{c|c}
(H^TH)^{-1} + \dfrac{(H^TH)^{-1}H^Th_{p+1}h_{p+1}^TH(H^TH)^{-1}}{h_{p+1}^T P h_{p+1}} & \dfrac{-(H^TH)^{-1}H^Th_{p+1}}{h_{p+1}^T P h_{p+1}} \\[3ex]
\hdashline
-\dfrac{h_{p+1}^TH(H^TH)^{-1}}{h_{p+1}^T P h_{p+1}} & \dfrac{1}{h_{p+1}^T P h_{p+1}}
\end{array}\right]}_{D_{p+1}}
\begin{bmatrix} H^Tx \\[2ex] h_{p+1}^Tx \end{bmatrix}
$$

$$
D_{p+1} = \left[\begin{array}{c:c}
D_p + \dfrac{D_pH^Th_{p+1}h_{p+1}^THD_p}{h_{p+1}^T P h_{p+1}} & \dfrac{-D_pH^Th_{p+1}}{h_{p+1}^T P h_{p+1}} \\[3ex]
-\dfrac{h_{p+1}^THD_p}{h_{p+1}^T P h_{p+1}} & \dfrac{1}{h_{p+1}^T P h_{p+1}}
\end{array}\right]
$$

Note: $h_{p+1}^T P h_{p+1}$

$$= h_{p+1}^T h_{p+1} - \left(h_{p+1}^T H\right)\left(H^T H\right)^{-1} H^T h_{p+1}$$

$$= \langle h_{p+1}, h_{p+1} \rangle - \langle h_{p+1}\ H \rangle D_p (\quad)^T$$

size of new basis function

components of new b.f. in the spanned space of old b.f.

Start with 1 basis function (could start 0)

$$h_1(t) = 1$$

$$v(t_n) = 1 \cdot \theta_1 + \cdots \quad , \quad n = 1, 2, \ldots$$

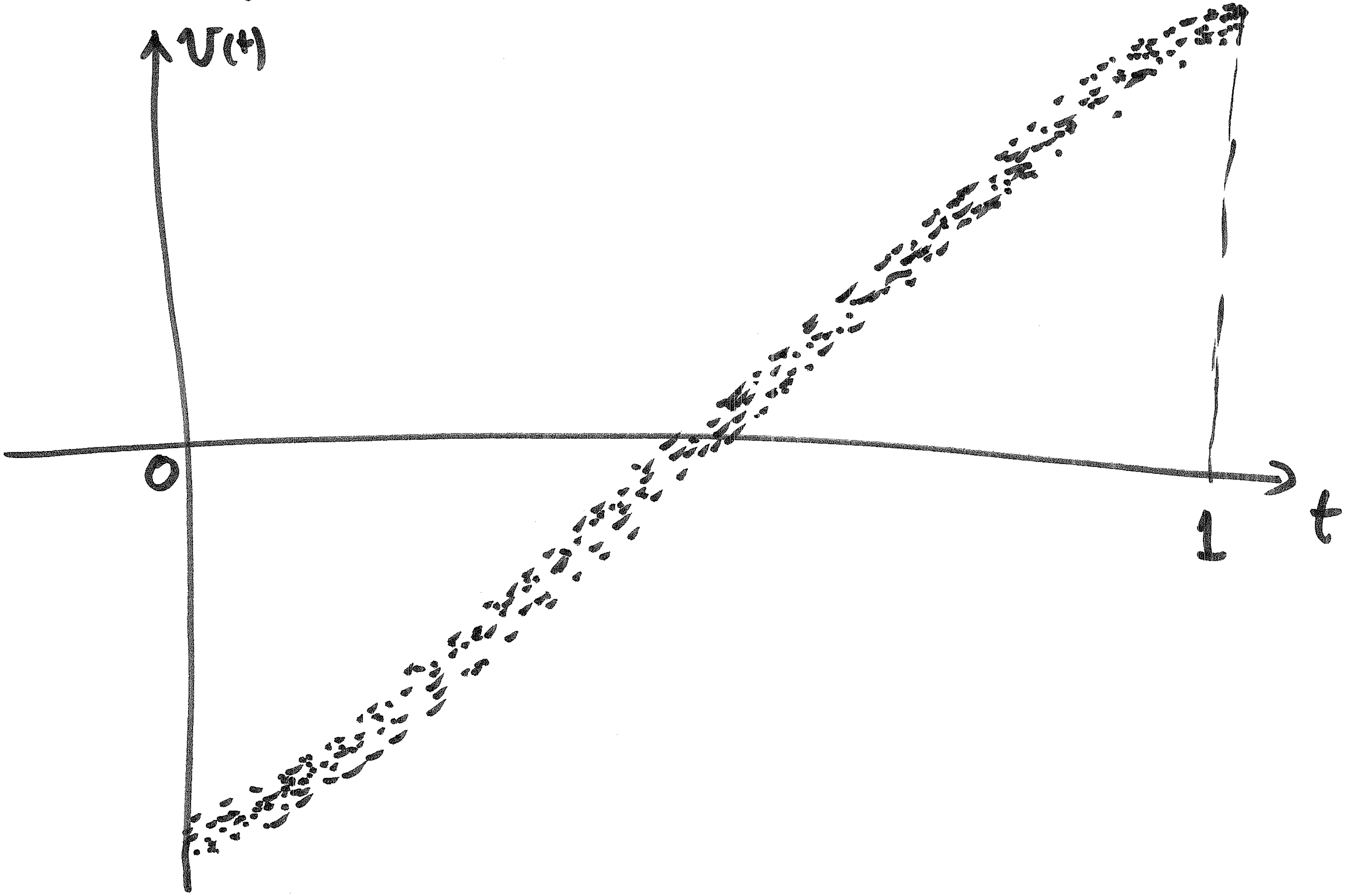
$$h_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{N \times 1} , \quad H = [h_1] , \quad x = \begin{bmatrix} V(t_1) \\ V(t_2) \\ \vdots \\ \cdot \end{bmatrix}_{N \times 1}$$

$$\hat{\theta}_1 = (H^T H)^{-1} H^T x = (h_1^T h_1)^{-1} h_1^T x = \frac{1}{N} \sum_{n=1}^{N} V(t_n)$$

$$= \text{average of } x_n \left\{ \equiv V(t_n) \right\}$$

$$D_1 = (H^T H)^{-1} = (h_1^T h_1)^{-1} = \frac{1}{N}$$

Example: ramp fitting:



2^{nd} step: $h_2 = $ 2 choices $\begin{cases} \text{simple}, & h_2 = t \\ \text{orth.} & h_2 = t - 0.5 \text{ or } 2t-1 \end{cases}$

$h_2(t) = t$ not orth to H

$h_2 = [t_1 \; t_2 \dots \; t_N]^T$

so need D_1, $\Rightarrow$ $P = I - H D_1 H^T$

$$= I - h_1 D_1 h_1^T$$

$$= \underset{N\times 1}{I} - \frac{1}{N} \underset{N\times N}{\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix}}$$

$$P h_2 = h_2 - \begin{bmatrix} \text{avg } h_2 \\ \text{avg } h_2 \\ \vdots \end{bmatrix} = h_2 - 0.5 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = h_2 - 0.5 h_1$$

$$h_2^T P h_2 = h_2^T h_2 - 0.5 \, h_2^T h_1 = \frac{N}{3} - \frac{N}{4} = \frac{N}{12}$$

$$\hat{\theta}_2 = \dots$$

$\Rightarrow$

$$\hat{\theta}_1 = \hat{\theta}_{1\,old} - \dots$$

HW: complete the example.

For the non-orth. choices of h_i, just finish step 2 by finding: New H, $\hat{\theta}_2$, updated $\hat{\theta}_1$, D_2, P, e^2_{tot} after

1st & 2nd step

For orthogonal, finish step 2 with $h_2(t)=2t-1$

& step 3 with $h_3(t)=(2t-1)^2$

Notice that in this case, No D, P, needed.

Compute $\hat{\theta}_2$ in step 2, $\hat{\theta}_3$ in step 3.

Show $\left\{ \begin{array}{l} H \\ e^2_{tot} \end{array} \right.$ in step 2 & 3

Constrained Least square :

$$X = H\theta + ? + w$$

but req θ to satisfy some conditions

$$A\theta = b$$

Use Lagrange multiplier :

$$J = e^2_{tot} + \lambda^T \cdot (\text{constraints})$$

$$= (X - H\theta)^T (X - H\theta) + \lambda^T (A\theta - b)$$

Solve :

$$\begin{cases} \dfrac{\partial J}{\partial \theta} = 0 \Rightarrow -2H^T(X - H\theta) + A^T\lambda = 0 \\ \dfrac{\partial J}{\partial \lambda} = 0 \Rightarrow A\theta - b = 0 \end{cases}$$

$$2H^T x + A^T \lambda = 2H^T H \theta$$

$$\hat{\theta} = (H^T H)^{-1} H^T x - (H^T H)^{-1} A^T \frac{\lambda}{2}$$

$$A (H^T H)^{-1} H^T x - A (H^T H)^{-1} A^T \frac{\lambda}{2} - b = 0$$

$$\frac{\lambda}{2} = \left(A (H^T H)^{-1} A^T \right)^{-1} \left(A (H^T H)^{-1} H^T x - b \right)$$

$$\hat{\theta} = (\hat{H} H)^{-1} H^T x - (H^T H)^{-1} A^T \left(A (H^T H)^{-1} A^T \right)^{-1} \left(\quad \right)$$

Nonlinear fitting problems:

Frequently, there is a "natural" way for fitting the data,

- driven by physics behind
- required by customer / boss

$$\underbrace{X_n}_{\text{data recorded}} = \underbrace{S_n(\theta)}_{\text{modeled signal}} + \underbrace{r}_{\text{residual det. error}} + \underbrace{w_n}_{\text{true noise}}$$

Need to identify θ from $\begin{cases} X_n \text{ data} \\ \text{assumption on } W_n \end{cases}$

For example:

$$x_n = A\cos(2\pi f t_n + \phi) + w_n$$

$$\theta = \begin{bmatrix} A \\ f \\ \phi \end{bmatrix} \leftarrow$$ appear nonlinearly in signal model

Another example:

Op amp output in step resp:

$$V_o(t_n) = \underset{V_{of}}{V_o(\infty)} + (V_{oi} - V_{of})\, e^{-\frac{t_n}{\tau}} \cos(2\pi f_d t_n + \phi)$$

$$\theta = [V_{oi},\ V_{of},\ \tau,\ f_d,\ \phi]^T$$

3 approaches to nonlinear L.S.

a) Remodel it, call a nonlinear exp. of old para as new para that appear linearly in model

e.g.: $$X_n = A\cos(2\pi f t_n + \phi) + w_n$$

$$= (A\cos\phi)\cos(2\pi f t_n)$$

$$- (A\sin\phi)(\cos 2\pi f t_n) + w_n$$

originally only one para A appear lin.

Now let $\begin{cases} a = A\cos\phi \\ b = A\sin\phi \end{cases}$, 2 para. apear linearly

In geneal, if θ appear nonlinearly as $g(\theta)$ in eq.

Call $\alpha = g(\theta)$ as new para.

- Need $g(\theta)$ invertible, at least near the true value of θ, i.e. $g^{-1}(\alpha)$ exist
- i.e. $\frac{\partial g}{\partial \theta}$ is square nonsingular.
- $s(\theta) \stackrel{\text{now}}{=\!=} s(g^{-1}(\alpha)) = H\alpha$ linear.

$$\Rightarrow \quad \hat{\alpha} = (H^T H)^{-1} H^T x$$

$$\hat{\theta} = g^{-1}(\hat{\alpha})$$

In the second case, we have made some progress: we separated linear & nonlinear parts:

$$S(A, \phi, f) = A\cos(2\pi f t + \phi)$$

$$= \begin{bmatrix} \cos 2\pi f t & -\sin 2\pi f t \end{bmatrix} \begin{bmatrix} A\cos\phi \\ A\sin\phi \end{bmatrix}$$

$\cancel{\sin 2\pi}$

$$= \underbrace{H(f)}_{\text{nonlinear function}} \cdot \underbrace{\theta(A, \phi)}_{\text{linear para}}$$

In general: we may have

$$S(\theta) = S(\alpha, \beta) = H(\alpha) \cdot \beta$$

$$\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \begin{array}{l} (p-q) \times 1 \\ \leftarrow q \times 1 \end{array} \Bigg\} \; p \times 1$$

$$J(\alpha, \beta) = e^2_{tot} = (x - H(\alpha)\beta)^T (x - H(\alpha)\beta)$$

$$\frac{\partial J}{\partial \beta} = -2 H^T(\alpha)(x - H(\alpha)\beta) \overset{set}{=} 0$$

$$\Rightarrow \hat{\beta} = \left(H^T(\alpha) H(\alpha)\right)^{-1} H^T(\alpha)\, x$$

but α is unknown, $\therefore$ ↑ not an estimate

plug into J:

$$J(\alpha) = J(\alpha, \hat{\beta}) = x^T\left(I - H(\alpha)\left(H^T(\alpha) H(\alpha)\right)^{-1} H^T(\alpha)\right) x$$

This is truely nonlinear. When solved, plug $\hat{\alpha}$

b) nonlinear transform of data:

$$i_D(V_g) = i_0 \frac{W}{L} e^{\frac{V_g}{V_T}} + \dots$$

$$\Big(\log i_D(V_g)\Big) = \underbrace{\Big(\log\big(i_0 \tfrac{W}{L}\big)\Big)}_{\theta_1} \cdot \underset{h_1 = 1}{\uparrow} + \underset{h_2 = V_g}{(V_g)} \cdot \underset{\theta_2}{\left(\frac{1}{V_T}\right)} + \dots$$

here "V_g" is like "t", the ind. variable.

As in a), this may only turn some para into lin. model.

There may still be some nonlin. para.

3). Use Newton or Newton-Raphson to solve the nonlinear problem.

$$J(\theta) = (x - s(\theta))^T (x - s(\theta))$$

$$\text{or} \quad = J(s(\theta))$$

Let $g(\theta) = \frac{\partial J}{\partial \theta} = \frac{\partial s}{\partial \theta} \cdot \frac{\partial J}{\partial s}$

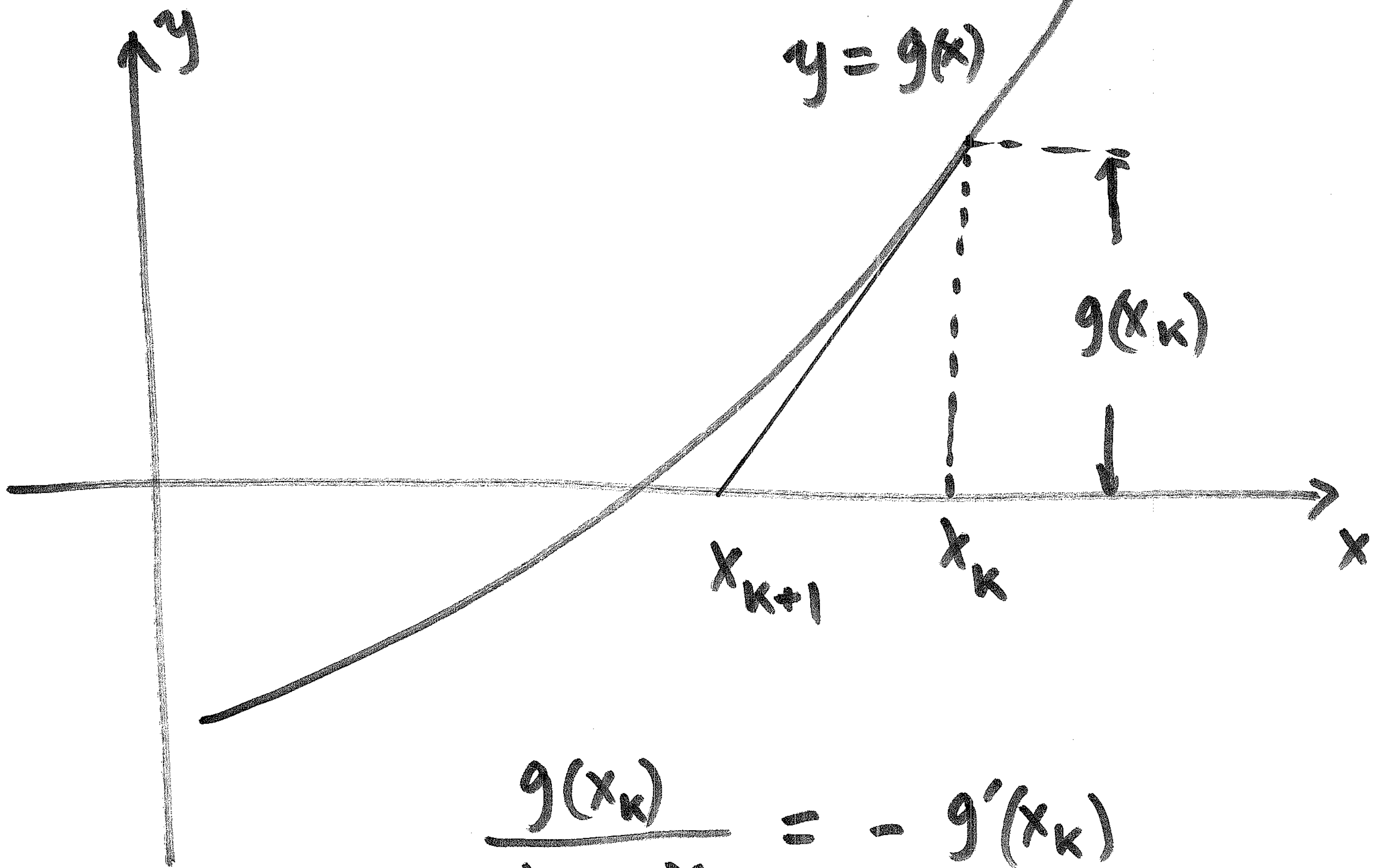
$$= -\frac{\partial s(\theta)^T}{\partial \theta} (x - s(\theta))$$

The goal is to solve iteratively

$$g(\theta) = 0$$

starting from some initial guess θ_0

Gauss-Newton method, Newton-Raphson



$$\frac{g(x_k)}{x_{k+1} - x_k} = -g'(x_k)$$

$$\Rightarrow \quad x_{k+1} = x_k - \frac{1}{g'(x_k)}\, g(x_k) \quad \leftarrow \text{Scalar}$$

~~matrix:~~ $\cancel{x_{k+1} = x_k}$

Vector case:
$$g(x_k) = -g'(x_k)\,(x_{k+1} - x_k)$$

$$x_{k+1} = x_k - \left[g'(x_k)\right]^{-1} g(x_k)$$

Apply to N.L. L.S.:
$$g(\theta) = \frac{\partial J}{\partial \theta} = \frac{\partial s^T}{\partial \theta} \cdot \frac{\partial J}{\partial s} = \frac{\partial s^T}{\partial \theta}(x - s(\theta))$$
$$= H(\theta)(x - s(\theta))$$

Start at initial guess, use

$$\theta_{k+1} = \theta_k - \left(\frac{\partial g(\theta)}{\partial \theta}\right]^{-1} g(\theta)\Big|_{\theta=\theta_k}$$

If linear: $\frac{\partial s^T}{\partial \theta} = H^T$, $\quad g(\theta) = -2H^T(x - H\theta)$

$$\frac{\partial g}{\partial \theta} = 2H^TH, \quad \theta_{k+1} = \theta_k + (H^TH)^{-1}H^T(x - H\theta_k)$$
$$= (H^TH)^{-1}H^Tx$$

Gauss - Newton update for N.L.

$$\theta_{k+1} = \theta_k + \left(H^T(\theta_k)\, H(\theta_k)\right)^{-1} H^T(\theta_k) \left(x - s(\theta_k)\right)$$

where $H(\theta_k) = \left. \dfrac{\partial s(\theta)}{\partial \theta} \right|_{\theta = \theta_k}$

Newton - Raphson update law:

$$\theta_{k+1} = \theta_k + \left[G(\theta_k)\right]^{-1} H^T(\theta_k) \left(x - s(\theta_k)\right)$$

$$G(\theta_k) = H^T(\theta_k)\, H(\theta_k) - \sum_{n=1}^{N} \left\{ G_n(\theta_k) \left(x_n - s(\theta_k)_n \right) \right\}$$

where $G_n(\theta_k) = \left. \dfrac{\partial^2 s}{\partial \theta^2} \right|_{\substack{\theta = \theta_k \\ \text{time} = n}}$

Ex: phase and freq measure

$$x_n = \underbrace{\cos(2\pi f t_n + \phi)}_{S(f,\phi)_n} + w_n$$

$$S(f,\phi)_n \, , \quad \theta = \begin{bmatrix} f \\ \phi \end{bmatrix}$$

$$\frac{\partial s}{\partial f} = -2\pi t_n \sin(2\pi f t_n + \phi) \leftarrow h_1(t_n)$$

$$\frac{\partial s}{\partial \phi} = -\sin(2\pi f t_n + \phi) \leftarrow h_2(t_n)$$

$$H(\theta) = \begin{bmatrix} h_1(t_1) & h_2(t_1) \\ h_1(t_2) & h_2(t_2) \\ h_1(t_3) & h_2(t_3) \\ \vdots & \vdots \end{bmatrix}$$

For $H(\theta_k)$: put in estimate of f & ϕ at k-th iteration: f_k, ϕ_k into

$$G_n(\theta) = \begin{bmatrix} \frac{\partial^2 s}{\partial f^2} & \frac{\partial^2 s}{\partial f \partial \phi} \\ \frac{\partial^2 s}{\partial f \partial \phi} & \frac{\partial^2 s}{\partial \phi^2} \end{bmatrix} = \begin{bmatrix} -4\pi^2 t_n^2 \cos(2\pi f t_n + \phi) & -2\pi t_n \cos \\ -2\pi t_n & -1 \end{bmatrix}$$

$$= \begin{bmatrix} -4\pi^2 t_n^2 & -2\pi t_n \\ -2\pi t_n & -1 \end{bmatrix} \cos(2\pi f t_n + \phi)$$

int.

If we take t_n's over $\left[-\frac{k}{f}, +\frac{k}{f}\right]$, i.e. k periods each side

and take $2M+1$ points, i.e. $t_{-M} =$, $t_{+M} =$

Then h_1 & h_2 will be orth. $\sum_n h_1(t_n)\, h_2(t_n) \approx 0$

$\sum h_2(t_n)\, h_2(t_n) = \text{avg} * \# = \frac{1}{2} * (2M+1) \approx M$

$\sum h_1(t_n)\, h_1(t_n) = 2 * 4\pi^2 \int_0^1 \frac{t^2}{2} * (2M+1) \approx \frac{8M^3\pi^2}{3}$

$$H^T(\theta_k)\,H^T(\theta_k) \approx \begin{bmatrix} \frac{8M^3\pi^2}{3} & 0 \\ 0 & M \end{bmatrix} \quad \forall k$$

$$\big[\quad \big]^{-1} = \begin{bmatrix} \frac{3}{8\pi^2 M^3} & 0 \\ 0 & \frac{1}{M} \end{bmatrix}$$

$$\begin{pmatrix} f \\ \phi \end{pmatrix}_{k+1} = \begin{pmatrix} f \\ \phi \end{pmatrix}_{k} + \big[\quad \big] \begin{bmatrix} h_1^T \\ h_2^T \end{bmatrix} (x - \cos..)$$

Due to orth. $h_1^T \begin{bmatrix} \cos.. \\ \vdots \end{bmatrix} = 0$, $h_2^T \cos = 0$

$$\therefore \begin{pmatrix} f \\ \phi \end{pmatrix}_{k+1} = \begin{pmatrix} f \\ \phi \end{pmatrix}_{k} + \big[\quad \big] \begin{bmatrix} h_1^T x \\ h_2^T x \end{bmatrix}$$

This is Gauss-Newton.