

## Least Squares Methods

### • Overdetermined linear equations

- $\mathbf{y} = \mathbf{Ax}$  where  $\mathbf{A} \in \mathbf{R}^{m \times n}$  and  $m > n$
- More equations than unknowns
- Cannot solve for  $\mathbf{x}$  in most cases.

### • Least squares solution of overdetermined linear equations

- Residual or error is  $\mathbf{r} = \mathbf{Ax} - \mathbf{y}$ .
- Find  $\mathbf{x}_{ls} = \arg \min_{\mathbf{x} \in \mathbf{R}^n} \|\mathbf{r}\|^2$ .
- $\mathbf{Ax}_{ls} \in R(\mathbf{A})$  and  $\mathbf{Ax}_{ls}$  is closest to  $\mathbf{y}$ .
- $\mathbf{Ax}_{ls}$  is projection of  $\mathbf{y}$  on  $R(\mathbf{A})$ .
- Assume  $\mathbf{A}$  is full rank. From

$$\frac{d\|\mathbf{r}\|^2}{d\mathbf{x}} = \frac{d}{d\mathbf{x}} (\mathbf{x}^T \mathbf{A}^T \mathbf{Ax} - 2\mathbf{y}^T \mathbf{Ax} + \mathbf{y}^T \mathbf{y}) = 2\mathbf{x}^T \mathbf{A}^T \mathbf{A} - 2\mathbf{y}^T \mathbf{A} = 0,$$

$$\mathbf{x}_{ls} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} = \mathbf{A}^+ \mathbf{y}$$

- The optimal residual is  $\mathbf{r}^* = \mathbf{Ax}_{ls} - \mathbf{y} = (\mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T - \mathbf{I})\mathbf{y}$ .
- Pseudo-inverse  $\mathbf{A}^+$  is a left inverse of  $\mathbf{A}$ ;  $\mathbf{A}^+ \mathbf{A} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A} = \mathbf{I}$
- Projection of  $\mathbf{y}$  on  $R(\mathbf{A})$  is  $\mathbf{Ax}_{ls} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} = \mathbf{P}_A \mathbf{y}$  where projection matrix is

$$\mathbf{P}_A = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T.$$

- Orthogonality principle:  $\mathbf{r} \perp R(\mathbf{A})$ , i.e.,  $\forall \mathbf{Az} \in R(\mathbf{A})$ ,

$$\langle \mathbf{r}, \mathbf{Az} \rangle = \mathbf{y}^T (\mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T - \mathbf{I})^T \mathbf{Az} = 0$$

### • Least squares estimation

- Model:  $\mathbf{y} = \mathbf{Ax} + \mathbf{v}$ 
  - (1)  $\mathbf{x}$  is what we want to estimate.
  - (2)  $\mathbf{y}$  is sensor measurements.
  - (3)  $\mathbf{v}$  is unknown noise or measurement error

- Least squares estimate,  $\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$ 
  - (1) If  $\mathbf{v} = \mathbf{0}$ ,  $\hat{\mathbf{x}} = \mathbf{x}$ .  $\Rightarrow$  unbiased
  - (2) Linear estimator, i.e.,  $\hat{\mathbf{x}} = \mathbf{B}\mathbf{y}$  for some  $\mathbf{B}$
  - (3)  $\hat{\mathbf{x}}$  is the best linear unbiased estimator (BLUE).

### • Least squares data fitting

#### - Preparation

- (1) Functions  $f_1, \dots, f_n : S \rightarrow \mathbf{R}$ ,  $S \subseteq \mathbf{R}^n$ : basis functions or regressors
- (2)  $\{(\mathbf{s}_i, g_i)\}_{i=1}^m$  with  $\mathbf{s}_i \in S$  and  $m > n$ : data or measurements

- Problem: find a linear combination of functions that fits data, i.e.,

$$\sum_{j=1}^n x_j f_j(\mathbf{s}_i) \approx g_i \quad \text{for } i = 1, 2, \dots, m$$

- Least squares solution:  $\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{g}\|^2 = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{g}$

- (1)  $\mathbf{x} = [x_1 \cdots x_n]^T$
- (2)  $\mathbf{g} = [g_1 \cdots g_m]^T$
- (3)  $\mathbf{A} = [a_{ij}] = [f_j(\mathbf{s}_i)] \in \mathbf{R}^{m \times n}$

- Fitting function:  $f_s(\mathbf{s}) = x_1 f_1(\mathbf{s}) + \cdots + x_n f_n(\mathbf{s})$

- Choosing the value of  $n$ : plot  $\|\mathbf{r}\| = \|\mathbf{A}\mathbf{x} - \mathbf{g}\|$  as a function of  $n$

- Least squares polynomial fitting for  $S \subseteq \mathbf{R}$

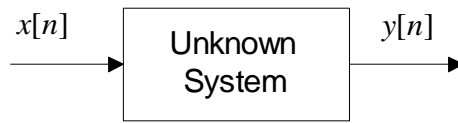
- (1)  $f_j(s) = s^{j-1}$
- (2)  $a_{ij} = s_i^{j-1}$  (Vandermonde matrix)

#### - Applications

- (1) Interpolation, extrapolation, smoothing of data
- (2) Simple and approximate model of data

### • Least squares system identification

- Problem: find a model for unknown system from input-output data



- Assume a model: for example,  $MA(n)$

$$y[n] = \sum_{i=0}^n w_i x[n-i]$$

- Collect input-output data with  $N > (n + 1)$

$$(1) \mathbf{A}\mathbf{w} = \begin{bmatrix} x[n] & x[n-1] & \cdots & x[0] \\ x[n+1] & x[n] & \cdots & x[1] \\ \vdots & \vdots & \ddots & \vdots \\ x[n+N-1] & x[n+N-2] & \cdots & x[N-1] \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$$

$$(2) \mathbf{y} = \begin{bmatrix} y[n] \\ y[n+1] \\ \vdots \\ y[n+N-1] \end{bmatrix}$$

- Least squares solution:  $\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\mathbf{A}\mathbf{w} - \mathbf{y}\|^2 = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$

- Choosing the order of the model ( $n$ )

(1) Larger  $n$

- (a) Small error for a particular set of data (smaller bias)
- (b) Try to fit even noise (overfit or overmodeling)
- (c) Poor generalization and poor predictive ability (larger variance)

(2) Cross-validation: use different set of data and monitor the change of error on this data set

• **Multi-objective least squares**

- Two (competing) objectives

$$(1) J_1 = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2$$

$$(2) J_2 = \|\mathbf{B}\mathbf{x} - \mathbf{z}\|^2$$

- Weighted sum objective

$$(1) J = J_1 + \mu J_2 = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \mu \|\mathbf{B}\mathbf{x} - \mathbf{z}\|^2 \text{ with } \mu \geq 0$$

$$(2) \quad \|\mathbf{Ax} - \mathbf{y}\|^2 + \mu \|\mathbf{Bx} - \mathbf{z}\|^2 = \left\| \begin{bmatrix} \mathbf{A} \\ \sqrt{\mu}\mathbf{B} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{y} \\ \sqrt{\mu}\mathbf{z} \end{bmatrix} \right\|^2 = \|\tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{y}}\|^2$$

- Least squares solution:  $\hat{\mathbf{x}} = (\tilde{\mathbf{A}}^T \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^T \tilde{\mathbf{y}} = (\mathbf{A}^T \mathbf{A} + \mu \mathbf{B}^T \mathbf{B})^{-1} (\mathbf{A}^T \mathbf{y} + \mu \mathbf{B}^T \mathbf{z})$

### • Regularized least squares

- With  $\mathbf{B} = \mathbf{I}$  and  $\mathbf{z} = \mathbf{0}$ ,  $J = J_1 + \mu J_2 = \|\mathbf{Ax} - \mathbf{y}\|^2 + \mu \|\mathbf{x}\|^2$

- (1) There is a penalty for large  $\mathbf{x}$ .
- (2) It stabilizes the algorithm.
- (3) It works for any  $\mathbf{A}$ .
- (4) Tikhonov regularization

- Least squares solution:  $\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A} + \mu \mathbf{I})^{-1} \mathbf{A}^T \mathbf{y}$

- (1) Useful when  $\mathbf{A}^T \mathbf{A}$  is ill-conditioned.
- (2) Useful when we know  $\mathbf{x}$  is small.
- (3) Useful when model is accurate only for small  $\mathbf{x}$ .

### • Underdetermined linear equations

- $\mathbf{y} = \mathbf{Ax}$  where  $\mathbf{A} \in \mathbf{R}^{m \times n}$  with  $m < n$
- More unknowns than equations
- $\mathbf{x}$  is under specified and there are many solutions.
- Assuming  $\text{rank}(\mathbf{A}) = m$ , for each  $\mathbf{y} \in \mathbf{R}^m$

- (1) Set of solutions is  $\{\mathbf{x} : \mathbf{Ax} = \mathbf{y}\} = \{\mathbf{x} + \mathbf{z} : \mathbf{Ax} = \mathbf{y}, \mathbf{z} \in N(\mathbf{A})\}$
- (2) Solution has  $\dim N(\mathbf{A}) = n - m$  degree of freedom
- (3) What is the best solution?

### • Minimum norm solution

- $\mathbf{y} = \mathbf{Ax}$  where  $\mathbf{A} \in \mathbf{R}^{m \times n}$  with  $m < n$  and  $\text{rank}(\mathbf{A}) = m$
- Minimum norm solution:  $\tilde{\mathbf{x}} = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{y}$
- For any  $\mathbf{x}$  such that  $\mathbf{Ax} = \mathbf{y}$ ,

- (1)  $\|\mathbf{x} - \tilde{\mathbf{x}}, \tilde{\mathbf{x}}\| = (\mathbf{x} - \tilde{\mathbf{x}})^T \tilde{\mathbf{x}} = 0$ , i.e.,  $(\mathbf{x} - \tilde{\mathbf{x}}) \perp \tilde{\mathbf{x}}$

$$(2) \|\mathbf{x}\|^2 \geq \|\tilde{\mathbf{x}}\|^2$$

- Orthogonality condition:  $\tilde{\mathbf{x}} \perp N(\mathbf{A})$

-  $\tilde{\mathbf{x}}$  is projection of  $\mathbf{0}$  on the solution set  $\{\mathbf{x} : \mathbf{Ax} = \mathbf{y}\} = \{\mathbf{x} + \mathbf{z} : \mathbf{Ax} = \mathbf{y}, \mathbf{z} \in N(\mathbf{A})\}$

-  $\mathbf{A}^+ = \mathbf{A}^T (\mathbf{AA}^T)^{-1}$  is pseudo-inverse of  $\mathbf{A}$

- Derivation

$$(1) \text{ Formulation: } \begin{array}{l} \min_{\mathbf{x}} \mathbf{x}^T \mathbf{x} \\ \text{subject to } \mathbf{Ax} = \mathbf{y} \end{array}$$

$$(2) \text{ Lagrange multiplier, } L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{x}^T \mathbf{x} + \boldsymbol{\lambda}^T (\mathbf{Ax} - \mathbf{y})$$

(3) Optimality conditions

$$(a) \frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{x}^T + \boldsymbol{\lambda}^T \mathbf{A} = \mathbf{0} \Rightarrow \mathbf{x} = -\frac{\mathbf{A}^T \boldsymbol{\lambda}}{2}$$

$$(b) \frac{\partial L}{\partial \boldsymbol{\lambda}} = (\mathbf{Ax} - \mathbf{y})^T = \mathbf{0} \Rightarrow \boldsymbol{\lambda} = -2(\mathbf{AA}^T)^{-1} \mathbf{y}$$

$$(4) \tilde{\mathbf{x}} = \mathbf{A}^T (\mathbf{AA}^T)^{-1} \mathbf{y}$$

- Comparing regularized least squares solution,

$$(\mathbf{A}^T \mathbf{A} + \mu \mathbf{I})^{-1} \mathbf{A}^T \xrightarrow{\mu \rightarrow 0} \mathbf{A}^T (\mathbf{AA}^T)^{-1}$$

## Least Squares Estimation

- **Linear model:**  $\mathbf{y} = \mathbf{x} + \mathbf{n}$  and  $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_p]\boldsymbol{\theta} = \sum_{i=1}^p \theta_i \mathbf{h}_i$  or  $x_i = \tilde{\mathbf{h}}_i^T \boldsymbol{\theta}$ 
  - $\mathbf{y} = [y_1 y_2 \dots y_N]^T \in \mathbf{R}^N$  : measurements, known
  - $\mathbf{x} = [x_1 x_2 \dots x_N]^T \in \mathbf{R}^N$  : model output, signal component, unknown
  - $\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1p} \\ h_{21} & h_{22} & \dots & h_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \dots & h_{Np} \end{bmatrix} \in \mathbf{R}^{N \times p}$  : model structure, assumed to be known
  - $\boldsymbol{\theta} = [\theta_1 \theta_2 \dots \theta_p]^T \in \mathbf{R}^p$  : model parameter, unknown
  - $\mathbf{n} = [n_1 n_2 \dots n_N]^T \in \mathbf{R}^N$  : measurement error, noise component, unknown or known statistics

- **Least squares solution**

- $\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n}$  and  $e^2 = \mathbf{n}^T \mathbf{n} = (\mathbf{y} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{H}\boldsymbol{\theta}) = \text{tr}[(\mathbf{y} - \mathbf{H}\boldsymbol{\theta})(\mathbf{y} - \mathbf{H}\boldsymbol{\theta})^T]$
- $\frac{\partial}{\partial \boldsymbol{\theta}} e^2 = 2\mathbf{H}^T (\mathbf{y} - \mathbf{H}\boldsymbol{\theta})$  and  $\frac{\partial^2}{\partial \boldsymbol{\theta}^2} e^2 = 2\mathbf{H}^T \mathbf{H} \geq 0 \Rightarrow \hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} = \mathbf{G}_H^{-1} \mathbf{H}^T \mathbf{y}$
- Normal equation:  $\mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}} = \mathbf{H}^T \mathbf{y}$  or  $\mathbf{G}_H \hat{\boldsymbol{\theta}} = \mathbf{H}^T \mathbf{y}$
- Gram matrix or Grammian,  $\mathbf{G}_H = \mathbf{H}^T \mathbf{H}$  is nonsingular iff  $\{\mathbf{h}_i\}_{i=1}^p$  are linearly independent.
- Let  $\langle \mathbf{H} \rangle = \text{span}\{\mathbf{h}_i\}_{i=1}^p$  and assume  $\langle \mathbf{H} \rangle \oplus \langle \mathbf{A} \rangle = \mathbf{R}^N$  and  $\langle \mathbf{H} \rangle \perp \langle \mathbf{A} \rangle$
- Projections
  - (a)  $\hat{\mathbf{x}} = \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} = \mathbf{P}_H \mathbf{y}$ ,  $\mathbf{P}_H = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$  : projection onto  $\langle \mathbf{H} \rangle$



• **Orthogonality**

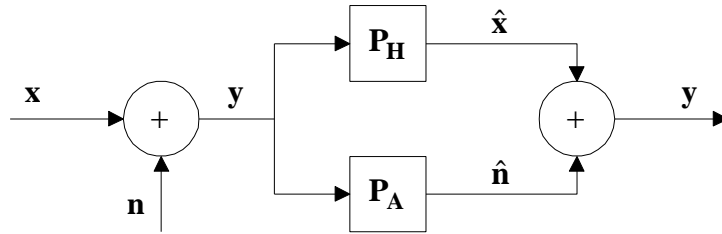
-  $\forall \mathbf{y} \in \mathbf{R}^N, \mathbf{y} = \mathbf{I}\mathbf{y} = \mathbf{P}_H\mathbf{y} + \mathbf{P}_A\mathbf{y} = \hat{\mathbf{x}} + \hat{\mathbf{n}} \Rightarrow \hat{\mathbf{n}}^T \hat{\mathbf{x}} = \mathbf{y}^T \mathbf{P}_A \mathbf{P}_H \mathbf{y} = 0$

-  $\mathbf{n}^T \mathbf{n} = \mathbf{y}^T \mathbf{P}_A \mathbf{y} = \mathbf{y}^T (\mathbf{I}_N - \mathbf{P}_H) \mathbf{y} = \mathbf{y}^T \mathbf{y} - \mathbf{x}^T \mathbf{x} \Rightarrow \mathbf{y}^T \mathbf{y} = \mathbf{x}^T \mathbf{x} + \mathbf{n}^T \mathbf{n}$

-  $\mathbf{n} = \hat{\mathbf{n}} + (\hat{\mathbf{x}} - \mathbf{x})$ : orthogonal decomposition of  $\mathbf{n}$

(a) 
$$\begin{aligned} \mathbf{n}^T \mathbf{n} &= (\mathbf{y} - \mathbf{x} + \mathbf{x} - \mathbf{x})^T (\mathbf{y} - \mathbf{x} + \mathbf{x} - \mathbf{x}) = [\mathbf{n} - (\mathbf{x} - \mathbf{x})]^T [\mathbf{n} - (\mathbf{x} - \mathbf{x})] \\ &= \mathbf{n}^T \mathbf{n} + (\mathbf{x} - \mathbf{x})^T (\mathbf{x} - \mathbf{x}) \geq \mathbf{n}^T \mathbf{n} \end{aligned}$$

(b)  $\hat{\mathbf{n}} = \mathbf{P}_A \mathbf{y} = \mathbf{y} - \hat{\mathbf{x}}$ : minimum norm, i.e., least squares



• **Example: complex exponential modes analysis**

- Model:  $x(t) = \sum_{i=1}^p \theta_i e^{j\omega_i t} = \theta_1 e^{j\omega_1 t} + \theta_2 e^{j\omega_2 t} + \dots + \theta_p e^{j\omega_p t}$ , i.e., sum of  $p$  complex exponentials

- From  $y(t) = x(t) + n(t)$ , take  $N$  measurements at  $t = kT, k = 0, 1, 2, \dots, (N - 1)$ . WLOG, assume  $T = 1$ .

- 
$$\mathbf{y} = \begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ y(N-1) \end{bmatrix} = \begin{bmatrix} \theta_1 + \theta_2 + \dots + \theta_p \\ \theta_1 e^{j\omega_1} + \theta_2 e^{j\omega_2} + \dots + \theta_p e^{j\omega_p} \\ \theta_1 e^{j\omega_1 2} + \theta_2 e^{j\omega_2 2} + \dots + \theta_p e^{j\omega_p 2} \\ \vdots \\ \theta_1 e^{j(N-1)\omega_1} + \theta_2 e^{j(N-1)\omega_2} + \dots + \theta_p e^{j(N-1)\omega_p} \end{bmatrix} + \begin{bmatrix} n(0) \\ n(1) \\ n(2) \\ \vdots \\ n(N-1) \end{bmatrix}, \text{ or}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ e^{j\omega_1} & e^{j\omega_2} & e^{j\omega_3} & \dots & e^{j\omega_p} \\ e^{j2\omega_1} & e^{j2\omega_2} & e^{j2\omega_3} & \dots & e^{j2\omega_p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{j(N-1)\omega_1} & e^{j(N-1)\omega_2} & e^{j(N-1)\omega_3} & \dots & e^{j(N-1)\omega_p} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_p \end{bmatrix} + \begin{bmatrix} n(0) \\ n(1) \\ n(2) \\ \vdots \\ n(N-1) \end{bmatrix}, \text{ or}$$



$$\mathbf{y} = \mathbf{x} + \mathbf{n} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n}$$

- Least squares solution:  $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}$  and  $\hat{\mathbf{x}} = \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} = \mathbf{P}_H \mathbf{y}$

• **Example: polynomial curve fitting**

- Model:  $x(t) = \sum_{i=1}^p \theta_i t^{i-1} = \theta_1 + \theta_2 t + \theta_3 t^2 + \dots + \theta_p t^{p-1}$ , i.e., polynomial
- From  $y(t) = x(t) + n(t)$ , take  $N$  measurements at  $t = kT$ ,  $k = 1, 2, \dots, N$ . WLOG, assume  $T = 1$ .

$$\mathbf{y} = \begin{bmatrix} y(1) \\ y(2) \\ y(3) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} \theta_1 + \theta_2 + \dots + \theta_p \\ \theta_1 + \theta_2 2 + \dots + \theta_p 2^{p-1} \\ \theta_1 + \theta_2 3 + \dots + \theta_p 3^{p-1} \\ \vdots \\ \theta_1 + \theta_2 N + \dots + \theta_p N^{p-1} \end{bmatrix} + \begin{bmatrix} n(1) \\ n(2) \\ n(3) \\ \vdots \\ n(N) \end{bmatrix}, \text{ or}$$

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 2^2 & \dots & 2^{p-1} \\ 1 & 3 & 3^2 & \dots & 3^{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & N & N^2 & \dots & N^{p-1} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \vdots \\ \theta_p \end{bmatrix} + \begin{bmatrix} n(1) \\ n(2) \\ n(3) \\ \vdots \\ n(N) \end{bmatrix}, \text{ or } \mathbf{y} = \mathbf{x} + \mathbf{n} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n}$$

- Least squares solution:  $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}$  and  $\hat{\mathbf{x}} = \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} = \mathbf{P}_H \mathbf{y}$

• **Recursive least squares (RLS)**

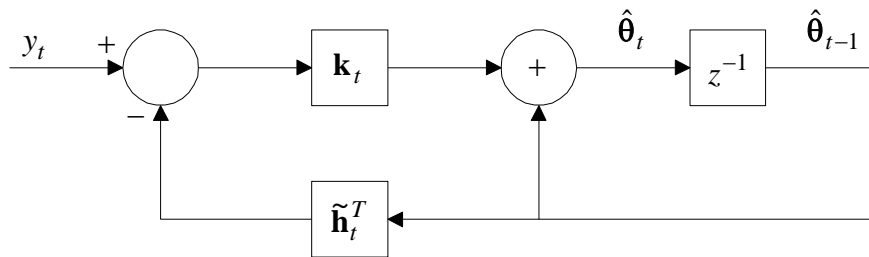
$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n} \Leftrightarrow \begin{bmatrix} \mathbf{y}_{t-1} \\ y_t \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{t-1} \\ \tilde{\mathbf{h}}_t^T \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{t-1} \\ n_t \end{bmatrix} \Leftrightarrow \begin{cases} \mathbf{y}_{t-1} = \mathbf{H}_{t-1} \boldsymbol{\theta} + \mathbf{n}_{t-1} \\ y_t = \tilde{\mathbf{h}}_t^T \boldsymbol{\theta} + n_t \end{cases}$$

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n} \Leftrightarrow y_t = \tilde{\mathbf{h}}_t^T \boldsymbol{\theta} + n_t \text{ for } t = 1, 2, 3, \dots$$

$$\text{Define } \mathbf{P}_t^{-1} = \mathbf{G}_{H_t} = \mathbf{H}_t^T \mathbf{H}_t = \mathbf{H}_{t-1}^T \mathbf{H}_{t-1} + \tilde{\mathbf{h}}_t \tilde{\mathbf{h}}_t^T = \mathbf{P}_{t-1}^{-1} + \tilde{\mathbf{h}}_t \tilde{\mathbf{h}}_t^T.$$

$$\mathbf{H}_t^T \mathbf{y}_t = \begin{bmatrix} \mathbf{H}_{t-1}^T \\ \tilde{\mathbf{h}}_t^T \end{bmatrix} \begin{bmatrix} \mathbf{y}_{t-1} \\ y_t \end{bmatrix} = \mathbf{H}_{t-1}^T \mathbf{y}_{t-1} + \tilde{\mathbf{h}}_t y_t$$

- LS solution is  $\hat{\boldsymbol{\theta}}_t = \mathbf{P}_t \mathbf{H}_t^T \mathbf{y}_t = (\mathbf{P}_{t-1}^{-1} + \tilde{\mathbf{h}}_t \tilde{\mathbf{h}}_t^T)^{-1} (\mathbf{H}_{t-1}^T \mathbf{y}_{t-1} + \tilde{\mathbf{h}}_t y_t)$ .
- From matrix inversion lemma,  $(\mathbf{P}_{t-1}^{-1} + \tilde{\mathbf{h}}_t \tilde{\mathbf{h}}_t^T)^{-1} = \mathbf{P}_{t-1} - \gamma_t \mathbf{P}_{t-1} \tilde{\mathbf{h}}_t \tilde{\mathbf{h}}_t^T \mathbf{P}_{t-1}$  and  $\gamma_t^{-1} = 1 + \tilde{\mathbf{h}}_t^T \mathbf{P}_{t-1} \tilde{\mathbf{h}}_t$  or  $\gamma_t \tilde{\mathbf{h}}_t^T \mathbf{P}_{t-1} \tilde{\mathbf{h}}_t = 1 - \gamma_t$
- $\hat{\boldsymbol{\theta}}_t = (\mathbf{P}_{t-1} - \gamma_t \mathbf{P}_{t-1} \tilde{\mathbf{h}}_t \tilde{\mathbf{h}}_t^T \mathbf{P}_{t-1}) (\mathbf{H}_{t-1}^T \mathbf{y}_{t-1} + \tilde{\mathbf{h}}_t y_t) = \hat{\boldsymbol{\theta}}_{t-1} + \gamma_t \mathbf{P}_{t-1} \tilde{\mathbf{h}}_t (y_t - \tilde{\mathbf{h}}_t^T \hat{\boldsymbol{\theta}}_{t-1})$
- Define  $\mathbf{k}_t = \gamma_t \mathbf{P}_{t-1} \tilde{\mathbf{h}}_t$



- Summary of RLS
  - (1) Initialization:  $\mathbf{P}_0 = \mathbf{I}_p$  and  $\hat{\boldsymbol{\theta}}_0 = \mathbf{0}_p$
  - (2)  $\gamma_t^{-1} = 1 + \tilde{\mathbf{h}}_t^T \mathbf{P}_{t-1} \tilde{\mathbf{h}}_t$
  - (3)  $\mathbf{k}_t = \gamma_t \mathbf{P}_{t-1} \tilde{\mathbf{h}}_t$
  - (4)  $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \mathbf{k}_t (y_t - \tilde{\mathbf{h}}_t^T \boldsymbol{\theta}_{t-1})$
  - (5)  $\mathbf{P}_t = \mathbf{P}_{t-1} - \gamma_t \mathbf{P}_{t-1} \tilde{\mathbf{h}}_t \tilde{\mathbf{h}}_t^T \mathbf{P}_{t-1}$

• **Weighted least squares**

- Problem:  $\min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{W} (\mathbf{y} - \mathbf{H}\boldsymbol{\theta})$  with nonsingular symmetric  $\mathbf{W} \in \mathbf{R}^{N \times N}$
- $\frac{\partial}{\partial \boldsymbol{\theta}} = 0 \Rightarrow \mathbf{H}^T \mathbf{W} (\mathbf{y} - \mathbf{H}\boldsymbol{\theta}) = \mathbf{0}$
- Solution:  $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{y}$
- Choice of  $\mathbf{W}$

(1) If  $\mathbf{n} : \mathcal{N}[\mathbf{0}, \mathbf{R}]$ ,  $\mathbf{W} = \mathbf{R}^{-1}$ .

(2)  $\mathbf{W} = \text{diag}(w_1, \dots, w_N)$  where  $w_i = \frac{1}{SNR_i}$ .