LAIS, Lecture #17

"Recursive Least Squares Adaptive Filtering"

Reference: "Adaptive Filters", by Ali Sayed

$\underbrace{u(0), u(1), u(2), \dots, u(i)}_{\text{reference signal}}$

$\underbrace{d(0), d(1), d(2), \dots, d(i)}_{\text{desired signal}}$

goal: estimate $d(i)$ from $u(i)$ using a linear filter

$$\hat{d}(j) = w(1)u(j) + \dots + w(n)u(j-n+1)$$

$$\min_{w \in \mathbb{R}^n} \sum_{j=0}^{i} \left(d(j) - u_j w\right)^2 \quad (*)$$

$$u_j = [u(j) \; u(j-1) \; \dots \; u(j-n+1)] \in \mathbb{R}^{1\times n}$$

$$y_i = \begin{bmatrix} d(0) \\ \vdots \\ d(i) \end{bmatrix} \in \mathbb{R}^{i+1}, \quad \mathcal{U}_i = \begin{bmatrix} u_0 \\ \vdots \\ u_i \end{bmatrix}$$

$$\Leftrightarrow \min_{w \in \mathbb{R}^n} \|y_i - \mathcal{U}_i w\|_2^2$$

solution of minimal l_2-norm

$$w_i = \mathcal{U}_i^{\dagger} y_i$$

when $\mathcal{U}_i$ has full-column rank

$w_i = (\mathcal{U}_i^T \mathcal{U}_i)^{-1} \mathcal{U}_i^T y_i$ is the unique solution to $(*)$

ways of dealing with insufficiently exciting signals (so that $\mathcal{U}_i$ does not have full column rank)

Tikonov regularization

$$\min_{w\in\mathbb{R}^n} \|y_i-\mathcal{U}_i w\|_2^2+\epsilon\|w\|_2^2$$

$\Longleftrightarrow$

$$\min_{w\in\mathbb{R}^n} \left\| \begin{bmatrix} y_i \\ 0 \end{bmatrix} - \begin{bmatrix} \mathcal{U}_i \\ \sqrt{\epsilon} I_n \end{bmatrix} w \right\|_2^2$$

$$w_i=(\mathcal{U}_i^T\mathcal{U}_i+\epsilon I_n)^{-1}\mathcal{U}_i^T y_i$$

complexity $\mathcal{O}(n^3+n^2 i)$

dithering

intentionally add a small amount of white noise to u

adaptation: we need to "forget" past values of the signal

$0<\lambda<1$: forgetting factor

$$\min_{w\in\mathbb{R}^n} \sum_{j=0}^{i} \lambda^{i-j}(d(j)-u_j w)^2+\epsilon\lambda^{i+1}\|w\|_2^2$$

necessary for complexity reduction

"trade off between complexity and stability"

$\Longleftrightarrow$

$$\min_{w\in\mathbb{R}^n} \|\Lambda_i^{1/2}(y_i-\mathcal{U}_i w)\|_2^2+\|\sqrt{\epsilon}\,\lambda^{\frac{i+1}{2}}w\|_2^2$$

$$\Lambda_i=\mathrm{diag}(\lambda^i,\lambda^{i-1},\dots,\lambda,1)$$

$\Longleftrightarrow$

$$\min_{w\in\mathbb{R}^n} \left\| \begin{bmatrix} \Lambda_i^{1/2} y_i \\ 0 \end{bmatrix} - \begin{bmatrix} \Lambda_i^{1/2}\mathcal{U}_i \\ \sqrt{\epsilon}\,\lambda^{\frac{i+1}{2}} I_n \end{bmatrix} w \right\|_2^2$$

$$W_i = \Big(\underset{(n\times i)}{\mathcal{U}_i^T}\,\Lambda_i\,\underset{(i\times n)}{\mathcal{U}_i} + \epsilon\lambda^{i+1} I_n\Big)^{-1} \mathcal{U}_i^T \Lambda_i y_i$$

complexity: $\mathcal{O}(n^3 + n^2 i)$

"a new sample comes in..."

$$W_{i+1} = \left(\mathcal{U}_{i+1}^T \Lambda_{i+1} \mathcal{U}_{i+1} + \epsilon\lambda^{i+2} I_n\right)^{-1} \mathcal{U}_{i+1}^T \Lambda_{i+1} y_{i+1} \quad (**)$$

As time goes by y_i and $\mathcal{U}_i$ grow indefinetely so that we have a memory problem

The filter given by $(**)$ is impractical

Idea of RLS: compute W_{i+1} from $W_i, u(i+1), d(i+1)$ reccursively in $\mathcal{O}(n^2)$.

LAIS, Lecture #17

Lem [The matrix inversion lemma]

A, B, C, D matrices s.t. $A+BCD$ exists.

Suppose A, C, $C^{-1}+DA^{-1}B$ are invertible.

Then $A+BCD$ is invertible and

$$(A+BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1}+DA^{-1}B\right)^{-1}DA^{-1}. \blacksquare$$

$$\Phi_i \equiv \mathcal{H}_i^T \Lambda_i \mathcal{H}_i + \epsilon \lambda^{i+1} I_n$$

$$P_i = \Phi_i^{-1}, \quad W_i = P_i \mathcal{H}_i^T \Lambda_i y_i$$

$$y_{i+1} = \begin{bmatrix} y_i \\ d(i+1) \end{bmatrix}, \quad \mathcal{H}_{i+1} = \begin{bmatrix} \mathcal{H}_i \\ u_{i+1} \end{bmatrix}$$

$$W_{i+1} = P_{i+1} \mathcal{H}_{i+1}^T \Lambda_{i+1} y_{i+1}$$

$$\Phi_{i+1} = \mathcal{H}_{i+1}^T \Lambda_{i+1} \mathcal{H}_{i+1} + \epsilon \lambda^{i+2} I_n$$

$$= \begin{bmatrix} \mathcal{H}_i^T & u_{i+1}^T \end{bmatrix} \begin{bmatrix} \lambda^{i+1} & & \\ & \ddots & \\ & & 1 \end{bmatrix} \begin{bmatrix} \mathcal{H}_i \\ u_{i+1} \end{bmatrix} + \epsilon \lambda^{i+2} I_n$$

$$= \mathcal{H}_i^T \begin{bmatrix} \lambda^{i+1} & & \\ & \ddots & \\ & & \lambda \end{bmatrix} \mathcal{H}_i + u_{i+1}^T u_{i+1} + \epsilon \lambda^{i+2} I_n$$

$$= \lambda \Phi_i + u_{i+1}^T u_{i+1}$$

$$P_{i+1} = \Big(\underbrace{\lambda \Phi_i}_{A} + \underbrace{u_{i+1}^T}_{B} \underbrace{u_{i+1}}_{D}\Big)^{-1}$$

$C = 1$

$$C^{-1} + DA^{-1}B = 1 + \lambda^{-1} u_{i+1} P_i u_{i+1} > 0$$

From the lemma

$$P_{i+1} = \lambda^{-1}P_i - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1} \lambda^{-1} P_i}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}$$

complexity : $O(n^2)$

$$W_{i+1} = P_{i+1} U_{i+1}^T \Lambda_{i+1} y_{i+1} = \left[\lambda^{-1}P_i - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1} \lambda^{-1} P_i}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}\right] \underbrace{\left[U_i^T \ u_{i+1}^T\right] \begin{bmatrix} \lambda\Lambda_i & \\ & 1 \end{bmatrix} \begin{bmatrix} y_i \\ d(i+1) \end{bmatrix}}_{\left(\lambda U_i^T \Lambda_i y_i + d(i+1) u_{i+1}^T\right)}$$

$$= W_i - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1} W_i}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T} + \lambda^{-1}P_i u_{i+1}^T d(i+1) - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1} \lambda^{-1}P_i u_{i+1}^T d(i+1)}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}$$

$$= W_i + \frac{\lambda^{-1}P_i u_{i+1}^T}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}\left[-u_{i+1}W_i + d(i+1)\left(1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T\right) - \lambda^{-1}u_{i+1}P_i u_{i+1}^T d(i+1)\right]$$

$$W_{i+1} = W_i + \underbrace{\frac{\lambda^{-1} P_i u_{i+1}^T}{1 + \lambda^{-1} u_{i+1} P_i u_{i+1}^T}}_{\text{Kalman gain}} \underbrace{\left(d(i+1) - u_{i+1} W_i \right)}_{\text{error}}$$

From the lemma

$$\boxed{P_{i+1} = \lambda^{-1}P_i - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1} \lambda^{-1} P_i}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}}$$

RLS filter

complexity : $\mathcal{O}(n^2)$

$$W_i = \left(\underset{(n\times i)}{\mathcal{H}_i^T} \underset{(i\times n)}{\Delta_i \mathcal{H}_i} + \epsilon\lambda^{i+1} I_n\right)^{-1} \mathcal{H}_i^T \Delta_i y_i$$

complexity $\mathcal{O}(n^3 + n^2 i)$

"a new sample comes in..."

$$W_{i+1} = \left(\mathcal{H}_{i+1}^T \Delta_{i+1} \mathcal{H}_{i+1} + \epsilon\lambda^{i+2} I_n\right)^{-1} \mathcal{H}_{i+1}^T \Delta_{i+1} y_{i+1}$$

$$W_{i+1} = P_{i+1}\mathcal{H}_{i+1}^T \Delta_{i+1} y_{i+1} = \left[\lambda^{-1}P_i - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1}\lambda^{-1}P_i}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}\right] \underbrace{\left[\mathcal{H}_i^T \; u_{i+1}^T\right] \begin{bmatrix} \lambda\Delta_i & \\ & 1 \end{bmatrix} \begin{bmatrix} y_i \\ d(i+1) \end{bmatrix}}_{\left(\lambda \mathcal{H}_i^T \Delta_i y_i + d(i+1) u_{i+1}^T\right)}$$

$$= W_i - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1} W_i}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T} + \lambda^{-1}P_i u_{i+1}^T d(i+1) - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1}\lambda^{-1}P_i u_{i+1}^T d(i+1)}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}$$

$$= W_i + \frac{\lambda^{-1}P_i u_{i+1}^T}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}\left[-u_{i+1}W_i + d(i+1)\left(1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T\right) - \lambda^{-1}u_{i+1}P_i u_{i+1}^T d(i+1)\right]$$

$$\boxed{W_{i+1} = W_i + \underbrace{\frac{\lambda^{-1}P_i u_{i+1}^T}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}}_{\text{Kalman gain}} \underbrace{\left(d(i+1) - u_{i+1}W_i\right)}_{\text{error}}}$$

$$W_{i+1} = W_i + \underbrace{\frac{\lambda^{-1} P_i u_{i+1}^T}{1+\lambda^{-1} u_{i+1} P_i u_{i+1}^T}}_{\text{Kalman gain}} \underbrace{\left(d(i+1) - u_{i+1} W_i\right)}_{\text{error}}$$

RLS: superior convergence rate and steady-state error compared to $\mathcal{O}(n)$ adaptive filters like LMS

From the lemma

$$\boxed{P_{i+1} = \lambda^{-1}P_i - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1} \lambda^{-1} P_i}{1+\lambda^{-1} u_{i+1} P_i u_{i+1}^T}}$$

RLS filter

complexity : $O(n^2)$

robust numerical implementations

"square-root RLS"

$^{i+2} I_n$

$^{i+2} I_n$

$$W_{i+1} = P_{i+1} U_{i+1}^T \Lambda_{i+1} y_{i+1} = \left[\lambda^{-1}P_i - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1}\lambda^{-1}P_i}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}\right] \underbrace{\begin{bmatrix} U_i^T & u_{i+1}^T \end{bmatrix} \begin{bmatrix} \lambda \Lambda_i & \\ & 1 \end{bmatrix} \begin{bmatrix} y_i \\ d(i+1) \end{bmatrix}}_{\left(\lambda U_i^T \Lambda_i y_i + d(i+1) u_{i+1}^T\right)}$$

$$= W_i - \frac{\lambda^{-1} P_i u_{i+1}^T u_{i+1} W_i}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T} + \lambda^{-1} P_i u_{i+1}^T d(i+1) - \frac{\lambda^{-1}P_i u_{i+1}^T u_{i+1} \lambda^{-1} P_i u_{i+1}^T d(i+1)}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}$$

$$= W_i + \frac{\lambda^{-1}P_i u_{i+1}^T}{1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T}\left[-u_{i+1}W_i + d(i+1)\left(1+\lambda^{-1}u_{i+1}P_i u_{i+1}^T\right) - \lambda^{-1}u_{i+1}P_i u_{i+1}^T d(i+1)\right]$$