Statistical Digital Signal Processing

Week 14 Adaptive Filtering: IIR Adaptive Filter

Lecturer: ***Zelalem Hailu (Assistant Prof.)***

Addis Ababa Science and Technology University,
Addis Ababa, Ethiopia

Previous Topic (Week-13)

FIR Adaptive Filter

- Adaptive Filtering: Introduction
- FIR Adaptive Filter
- The Steepest Descent Adaptive Filter
- The LMS Algorithm
- Adaptive Linear Prediction

Lecture Learning Outcomes

1. Explain the fundamental concepts, structure, and applications of Infinite Impulse Response (IIR) adaptive filters and distinguish them from FIR adaptive filters
2. Analyze the operation of the IIR Steepest Descent Adaptive Filter, including gradient estimation and weight update equation
3. Apply the IIR LMS (Least Mean Squares) adaptive filtering algorithm to update filter coefficients
4. Evaluate and compare the performance, convergence characteristics, computational complexity, and stability issues of IIR Steepest Descent and IIR LMS adaptive filtering algorithms

Week 14: IIR Adaptive Filter

Outline

- IIR Adaptive Filtering: Introduction
- IIR Steepest Decent Adaptive Filter
- IIR LMS Adaptive Filter

IIR Adaptive Filtering: Introduction

- In previous session, we have seen methods of designing FIR (non-recursive) adaptive filters.
- Today, we will focus on the problem of designing IIR (recursive) adaptive filter which can produce minimum mean square estimation of a process from related measurement
- The IIR adaptive filter has the form [1]:

$$y(n) = \sum_{k=1}^{p} a_n(k) y(n-k) + \sum_{k=0}^{q} b_n(k) x(n-k) \qquad (1)$$

Where:

$a_n(k)$ and $b_n(k)$ are coefficient of the adaptive filter at time n

- Recursive adaptive filters have an advantage over non-recursive ones in providing better performance for a given filter order

IIR Adaptive Filtering: Introduction

- However, IIR (recursive) filters have issue of stability which may affect the convergence time and numerical sensitivity of the filter
- Despite of these challenges, there are many applications which require IIR adaptive filters
- Among those applications, echo cancelation is one good example
- Suppose that an echo is introduced by the channel which is used to transmit a signal $d(n)$ and the received signal $x(n)$ is given by:

$$x(n) = d(n) + \alpha d(n-N) \qquad (2)$$

Where:

$|\alpha| < 1$

N Is the delay associated with the echo

IIR Adaptive Filtering: Introduction

- Suppose both α and N are known, the ideal echo canceler for recovering $d(n)$ from $x(n)$ is IIR filter with the following system function:

$$H(z) = \frac{1}{1-\alpha z^{-N}} \qquad (3)$$

- However, α and N are generally unknown and possibly time varying,
- Therefore, the echo canceller is required to be adaptive IIR filter rather than the conventional IIR filter
- Even if FIR adaptive filters could be considered for tackling this problem, the order of the filter could be large to produce accurate estimate of $d(n)$ from $x(n)$
- Suppose that $H(z)$ can be expanded using geometric series as follows:

IIR Adaptive Filtering: Introduction

$$H(z) = \frac{1}{1-\alpha z^{-N}} = \sum_{k=0}^{\infty} \alpha^k z^{-Nk} \qquad (4)$$

- If we consider a large enough p, we will have

$$|\alpha|^p \ll 1 \qquad (5)$$

- Hence, we could have a finite order approximation for $H(z)$ as:

$$H(z) = \sum_{k=0}^{p} \alpha^k z^{-Nk} \qquad (6)$$

- Suppose, the adaptive non-recursive echo canceler has the form:

$$\hat{d}(n) = \sum_{k=0}^{Np} b_n(k) x(n-k) \qquad (7)$$

IIR Adaptive Filtering: Introduction

- However, if $\alpha \approx 1$, p will be forced to be large or if $N \gg 1$, the order of the adaptive filter, Np, which is required to produce sufficiently accurate estimate of $d(n)$ may become too large to be feasible solution
- Now let's focus on the derivation of necessary conditions required by the coefficient of IIR filter to minimize the mean square error:

$$\xi(n) = E\left|e(n)\right|^2 \qquad (8)$$

Where:

$$e(n) = d(n) - y(n) \qquad (9)$$

$d(n)$ is the desired process

$y(n)$ is the output of the adaptive filter

IIR Adaptive Filtering: Introduction

- Initially, let's assume the filter is shift invariant or non-adaptive with the following form:

$$y(n) = \sum_{l=1}^{p} a(l)y(n-l) + \sum_{l=0}^{q} b(l)x(n-l) \qquad (10)$$

- Denoting the filter coefficient by $\mathbf{\Theta}$ as:

$$\mathbf{\Theta} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \left[a(1), a(2), \ldots, a(p), b(0), b(1), \ldots, b(q) \right]^T \qquad (11)$$

- Similarly, denoting the aggregate data vector by $\mathbf{z}(n)$:

$$\mathbf{z}(n) = \begin{bmatrix} \mathbf{y}(n-1) \\ \mathbf{x}(n) \end{bmatrix}$$

$$= \left[y(n-1), y(n-2), \ldots, y(n-p), x(n), x(n-1), x(n-q) \right]^T \qquad (12)$$

IIR Adaptive Filtering: Introduction

- Now , the output of the filter at time n in terms of $\boldsymbol{\Theta}$ and $\mathbf{z}(n)$ as:

$$y(n) = \mathbf{a}^T \mathbf{y}(n-1) + \mathbf{b}^T \mathbf{x}(n)$$

$$= \Theta^T \mathbf{z}(n) \qquad (13)$$

- It is noted that $\boldsymbol{\Theta}$ should minimize the following mean square error:

$$\xi(n) = E\left\{|e(n)|^2\right\} = E\left\{|d(n) - y(n)|^2\right\} \qquad (14)$$

- Therefore, the derivative of $\xi(n)$ with respect to $a^*(k)$ and $b^*(k)$ should be equal to zero or the gradient vector equals to zero:

$$\nabla\xi(n) = \nabla E\left\{|e(n)|^2\right\} = E\left\{\nabla|e(n)|^2\right\}$$

$$= E\left\{e(n)\nabla e^*(n)\right\} \qquad (15)$$

IIR Adaptive Filtering: Introduction

- Hence, :

$$E\left\{ e(n)\nabla e^{*}(n) \right\} = 0 \qquad (16)$$

- Recalling that, $e(n) = d(n) - y(n)$, we will have

$$\nabla e^{*}(n) = -\nabla y^{*}(n) \qquad (17)$$

- Using eq(17, eq(16) can be further simplified as:

$$E\left\{ e(n)\nabla y^{*}(n) \right\} = 0 \qquad (18)$$

- Next we will evaluate $\nabla y^{*}(n)$ and we will have:

$$\frac{\partial y^{*}(n)}{\partial a^{*}(k)} = y^{*}(n-k) + \sum_{l=1}^{p} a^{*}(l)\frac{\partial y^{*}(n-l)}{\partial a^{*}(k)} \quad ; k = 1, 2, \ldots, p \qquad (19)$$

IIR Adaptive Filtering: Introduction

$$\frac{\partial y^*(n)}{\partial b^*(k)} = x^*(n-k) + \sum_{l=1}^{p} a^*(l) \frac{\partial y^*(n-l)}{\partial b^*(k)} \quad ; \; k = 0,1,2,\ldots,q \qquad (20)$$

- Writing eq(19 & 20) together in vector form, we will obtain the gradient of $y^*(n)$

$$\nabla y^*(n) = z^*(n) + \sum_{l=1}^{p} a^*(l) \nabla y^*(n-l) \qquad (21)$$

- Now, the necessary conditions required by $a(k)$ and $b(k)$ for minimizing the mean square error can be obtained from eq(18 & 21) as:

$$E\left\{ e(n) \left[y(n-k) + \sum_{l=1}^{p} a(l) \frac{\partial y(n-l)}{\partial a(k)} \right]^* \right\} = 0 \; ; \; k = 1,2,\ldots,p \qquad (22)$$

IIR Adaptive Filtering: Introduction

$$E\left\{ e(n)\left[x(n-k)+\sum_{l=1}^{p} a(l)\frac{\partial y(n-l)}{\partial b(k)} \right]^{*} \right\} = 0 \qquad (23)$$

$$;\ k = 0,1,2,\ldots,q$$

- Eq(22 & 23) are non-linear equations and difficult to solved to obtain the optimal filter coefficients
- In addition, this non-linearity may impose the uncertainty that the solution may corresponds to the local minima rather than the global minima
- In general, these non linear equations can not be solved directly

IIR Adaptive Filtering: Introduction

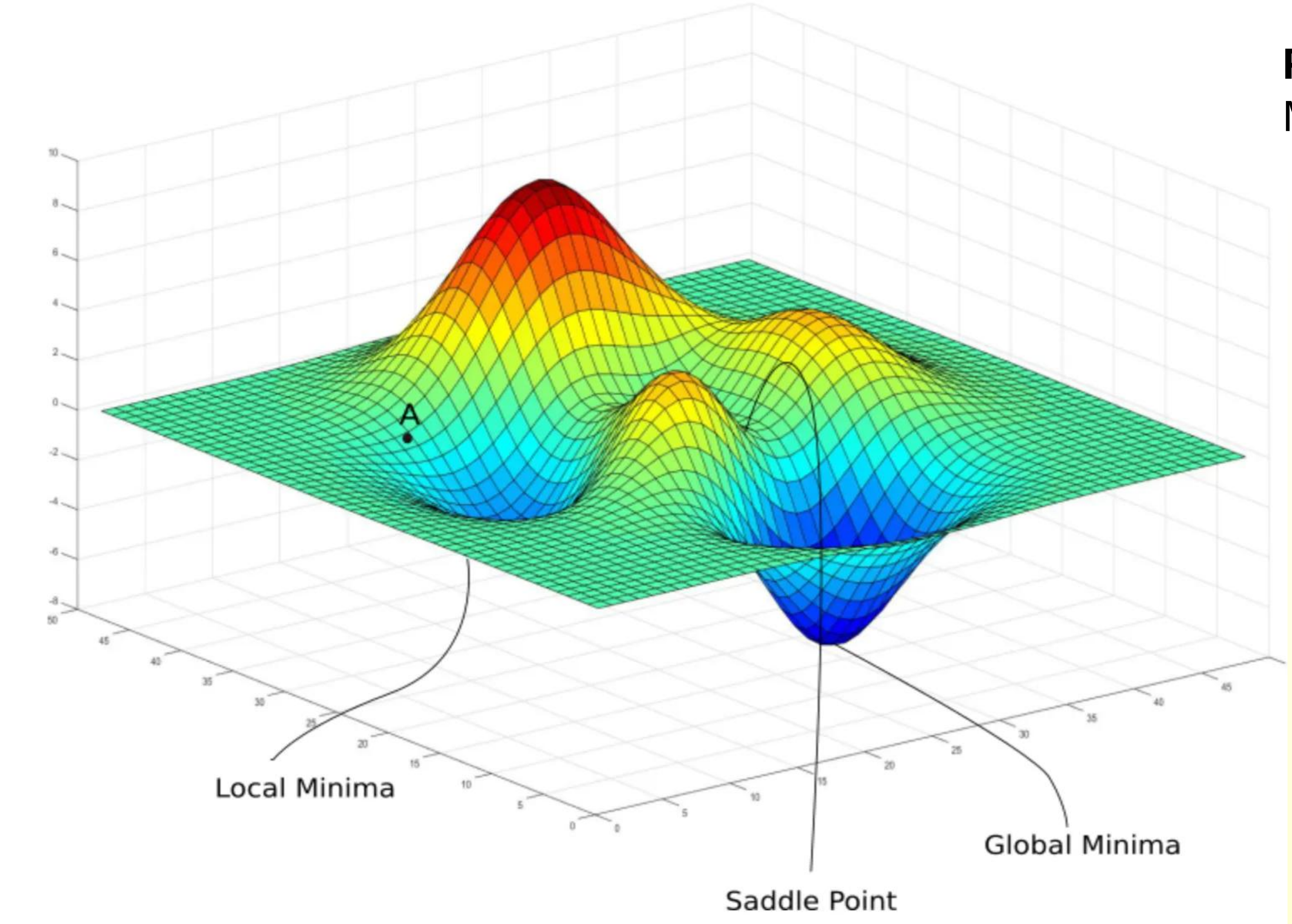


Figure 1: Local and Global Minima

Source: "The journey of Gradient Descent -From Local to Global", Medium. https://miro.medium.com/v2/resize:fit:4800/format:webp/1*ZC9qItK9wI0F6BwSVYMQGg.png

IIR Adaptive Filtering: Introduction

- Alternatively way of solving eq(22 & 23) is, by using the steepest decent to search the solutions iteratively
- Denoting the solution at time n by $\mathbf{\Theta}_n$:

$$\mathbf{\Theta}_n = \begin{bmatrix} \mathbf{a}_n \\ \mathbf{b}_n \end{bmatrix} = \; a_n(1), a_n(2), \ldots, a_n(p), b_n(0), b_n(1), \ldots, b_n(q) \;^T \qquad (24)$$

- The steepest decent weight update equation becomes:

$$\mathbf{\Theta}_{n+1} = \mathbf{\Theta}_n - \mu \nabla \xi(n) \qquad (25)$$

Where:

μ is the step size

$\nabla \xi(n)$ is the gradient vector

IIR LMS Adaptive Filter

- The challenge related to the update equation given in eq(25) is, the gradient vector involves expectation:

$$\nabla \xi(n) = E\{e(n)\nabla e^{*}(n)\} = -E\{e(n)\nabla y^{*}(n)\} \qquad (26)$$

- Therefore, to implement the steepest decent algorithm, the ensemble average given in eq(26) is required to be known
- However, If we adopt the LMS adaptive filter approach by replacing the expected values by instantaneous values, we will obtain:

$$\hat{\nabla} \xi(n) = -e(n)\nabla y^{*}(n) \qquad (27)$$

- Then the coefficient update equation in eq(25) becomes

$$\mathbf{\Theta}_{n+1} = \mathbf{\Theta}_{n} + \mu e(n)\nabla y^{*}(n) \qquad (28)$$

IIR LMS Adaptive Filter

- Eq(28) can also be expressed in terms of the filter coefficients, $a_n(k)$ and $b_n(k)$ as :

$$a_{n+1}(k) = a_n(k) + \mu e(n) \frac{\partial y^*(n)}{\partial a_n^*(k)} \qquad (29)$$

$$b_{n+1}(k) = b_n(k) + \mu e(n) \frac{\partial y^*(n)}{\partial b_n^*(k)} \qquad (30)$$

- The partial derivatives in eq(29 & 30) can be evaluated using eq(21) as:

$$\frac{\partial y^*(n)}{\partial a_n^*(k)} = y^*(n-k) + \sum_{l=1}^{p} a_n^*(l) \frac{\partial y^*(n-l)}{\partial a_n^*(k)} \quad ; \; k = 1, 2, \ldots, p \qquad (31)$$

$$\frac{\partial y^*(n)}{\partial b_n^*(k)} = x^*(n-k) + \sum_{l=1}^{p} a_n^*(l) \frac{\partial y^*(n-l)}{\partial b_n^*(k)} \quad ; \; k = 0, 1, 2, \ldots, q \qquad (32)$$

IIR LMS Adaptive Filter

- Eq(29, 30, 31, and 32) along with the filter output equation becomes:

$$y(n) = \mathbf{a}_n^T \mathbf{y}(n-1) + \mathbf{b}_n^T \mathbf{x}(n) = \mathbf{\Theta}_n^T \mathbf{z}(n) \qquad (33)$$

- Eq(33) forms the ***IIR LMS algorithm***
- For sufficiently small step size, the IIR LMS algorithm generally converges to the filter coefficients that minimize the mean square error
- However, Compared to the FIR LMS algorithm, the convergence of IIR LMS algorithm is much slower
- Further more, due to the presence of local minima in the mean square error surface, the convergence may not be to the global minima
- Lets look further ways of simplifying the IIR LMS Adaptive filter

IIR LMS Adaptive Filter

- It is noted that the derivatives of $y^*(n-l)$ in eq(31 & 32) are taken with respect to the current values of $a_n^*(k)$ and $b_n^*(k)$
- If the step size is small enough, the coefficients will adapt slowly and we will have:

$$\frac{\partial y^*(n-l)}{\partial a_n^*(k)} \approx \frac{\partial y^*(n-l)}{\partial a_{n-1}^*(k)} \qquad (34)$$

$$\frac{\partial y^*(n-l)}{\partial b_n^*(k)} \approx \frac{\partial y^*(n-l)}{\partial b_{n-1}^*(k)} \qquad (35)$$

- Based on eq(34 & 35), the derivatives in eq(31 & 32) can be simplifies as follows

IIR LMS Adaptive Filter

$$\frac{\partial y^*(n)}{\partial a_n^*(k)} \approx y^*(n-k) + \sum_{l=1}^{p} a_n^*(l) \frac{\partial y^*(n-l)}{\partial a_{n-1}^*(k)} \qquad (36)$$

$$\frac{\partial y^*(n)}{\partial b_n^*(k)} \approx x^*(n-k) + \sum_{l=1}^{p} a_n^*(l) \frac{\partial y^*(n-l)}{\partial b_{n-1}^*(k)} \qquad (37)$$

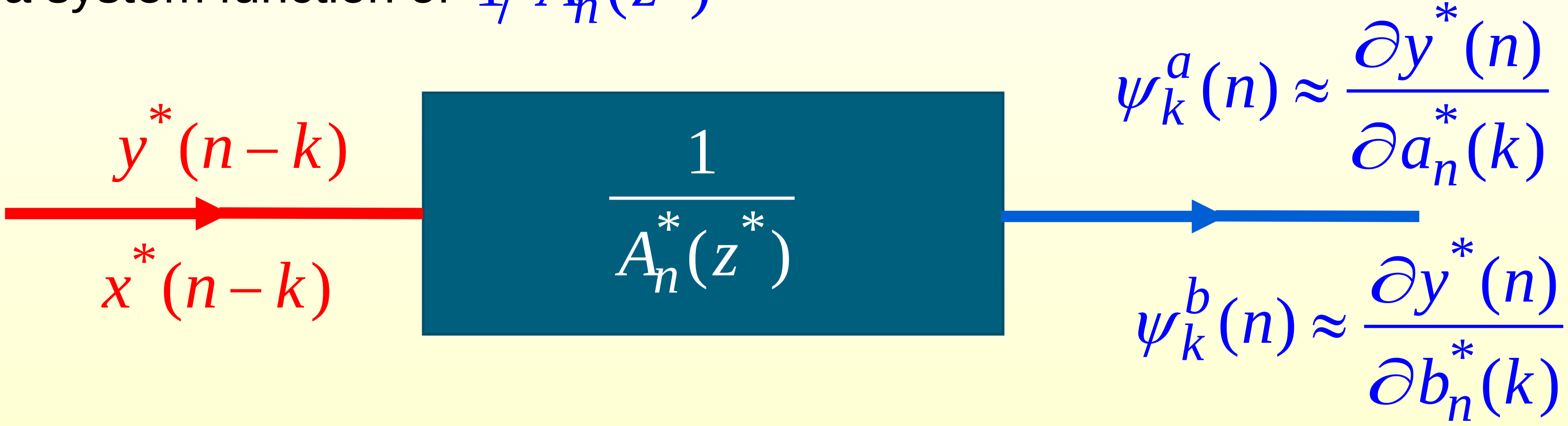
- Note that the above expressions for the partial derivatives are now recursive since the partial derivatives in the sum correspond to delayed versions of the derivatives on the left
- Let $\psi_k^a(n)$ and $\psi_k^b(n)$ are the approximation of the partial derivatives that are generated by eq(36 & 37) respectively

IIR LMS Adaptive Filter

$$\psi_k^a(n) = y^*(n-k) + \sum_{l=1}^{p} a_n^*(l)\psi_k^a(n-1) \qquad (38)$$

$$\psi_k^b(n) = x^*(n-k) + \sum_{l=1}^{p} a_n^*(l)\psi_k^b(n-1) \qquad (39)$$

- These approximations given in eq(38 & 39) can be generated by filtering $x^*(n-k)$ and $y^*(n-k)$ with the shift varying recursive filter having a system function of $1/A_n^*(z^*)$



IIR LMS Adaptive Filter

- The following figure, Figure (2), represents the simplified block diagram of IIR LMS adaptive filter

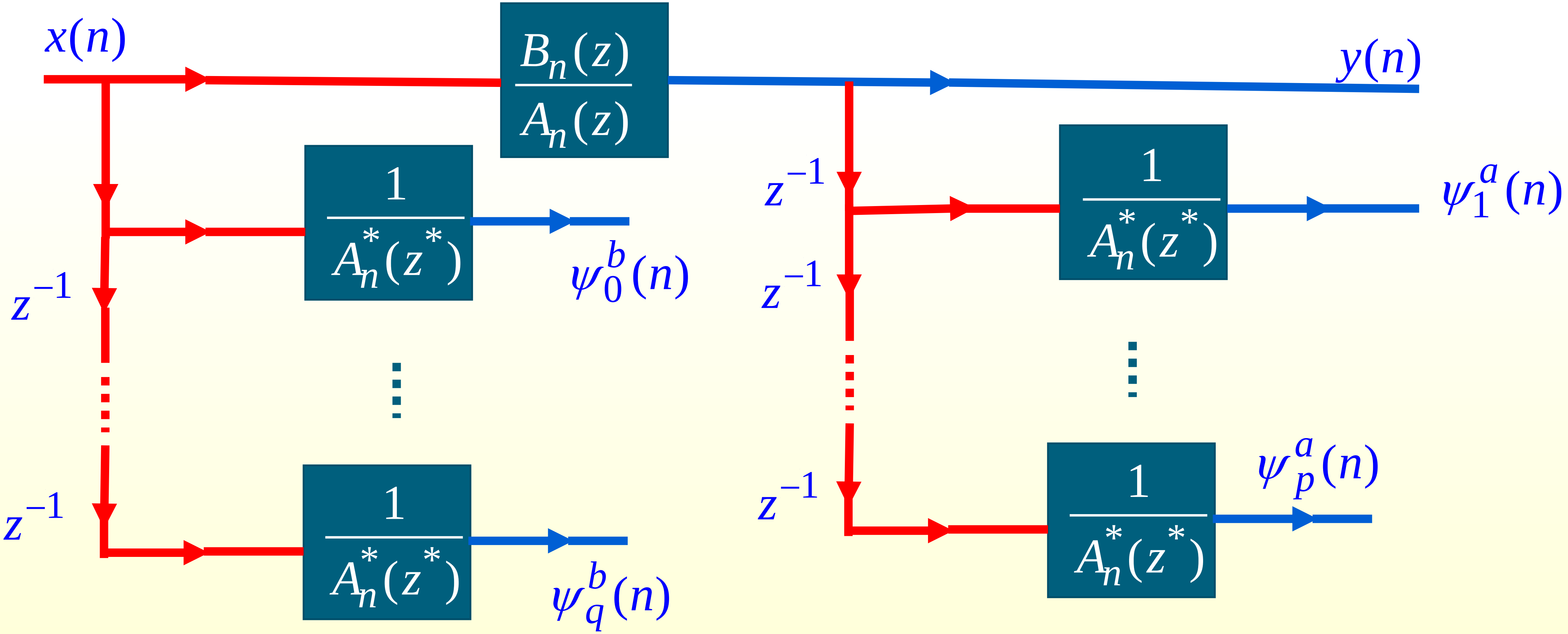


Figure 2: Simplified block diagram of IIR LMS adaptive filter
((The input of the filter must be conjugated before filtering))

IIR LMS Adaptive Filter

- It is noted that there are p+q+1 recursive filters operating in parallel to produce those approximations, $\psi_k^a(n)$ and $\psi_k^b(n)$, for the gradient vector
- The filter outputs are used to adaptively update the filter coefficients
- Although the approximation in Eq. (34 & 35) simplifies the algorithm, implementing p+q+1 parallel shift-varying filters still requires high computational complexity and considerable memory storage
- However, by assuming a sufficiently small step size μ, the IIR LMS adaptive filter can be further simplified
- the gradient estimate $\psi_k^a(n)$ is obtained by delaying $y^*(n)$ and filtering it through the shift-varying filter $1/A_n^*(z^*)$ as shown in the net figure, Figure (3)
- Similarly, $\psi_k^b(n)$ can be computed by filtering delayed versions of $x^*(n)$

IIR LMS Adaptive Filter

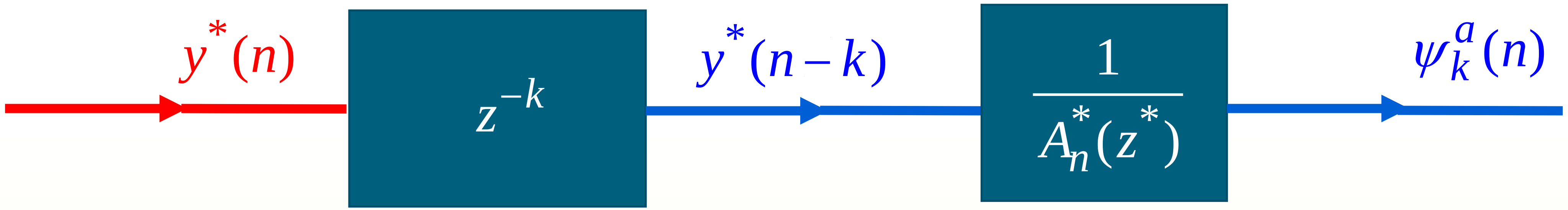


Figure 3: The gradient estimate $\psi_k^a(n)$ obtained by delaying $y^*(n)$ first

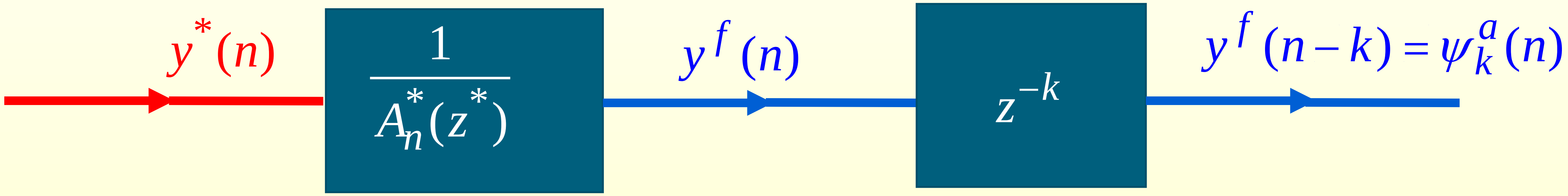


Figure 4: The gradient estimate $\psi_k^a(n)$ obtained by delaying $y^f(n)$

IIR LMS Adaptive Filter

- If the step size (μ) is sufficiently small, the coefficients an $a_n(k)$ vary only slightly over intervals of length p
- Therefore, the filter $1/A_n^*(z^*)$ can be treated as shift-invariant, allowing the delay and filtering operations to be interchanged, as illustrated in Figure(4)
- As a result, $\psi_k^a(n)$ and $\psi_k^b(n)$ may be estimated first by simply generating the filtered signals as follows:

$$y^f(n) = \psi_0^a(n) \qquad (40)$$

$$x^f(n) = \psi_0^b(n) \qquad (41)$$

- Then, by delaying the signal given in equation (40 & 41) we will obtain $\psi_k^b(n)$ and $\psi_k^a(n)$ as shown in the next figure, Figure(5)
- This method is called ***Filtered Signal Approach***

IIR LMS Adaptive Filter

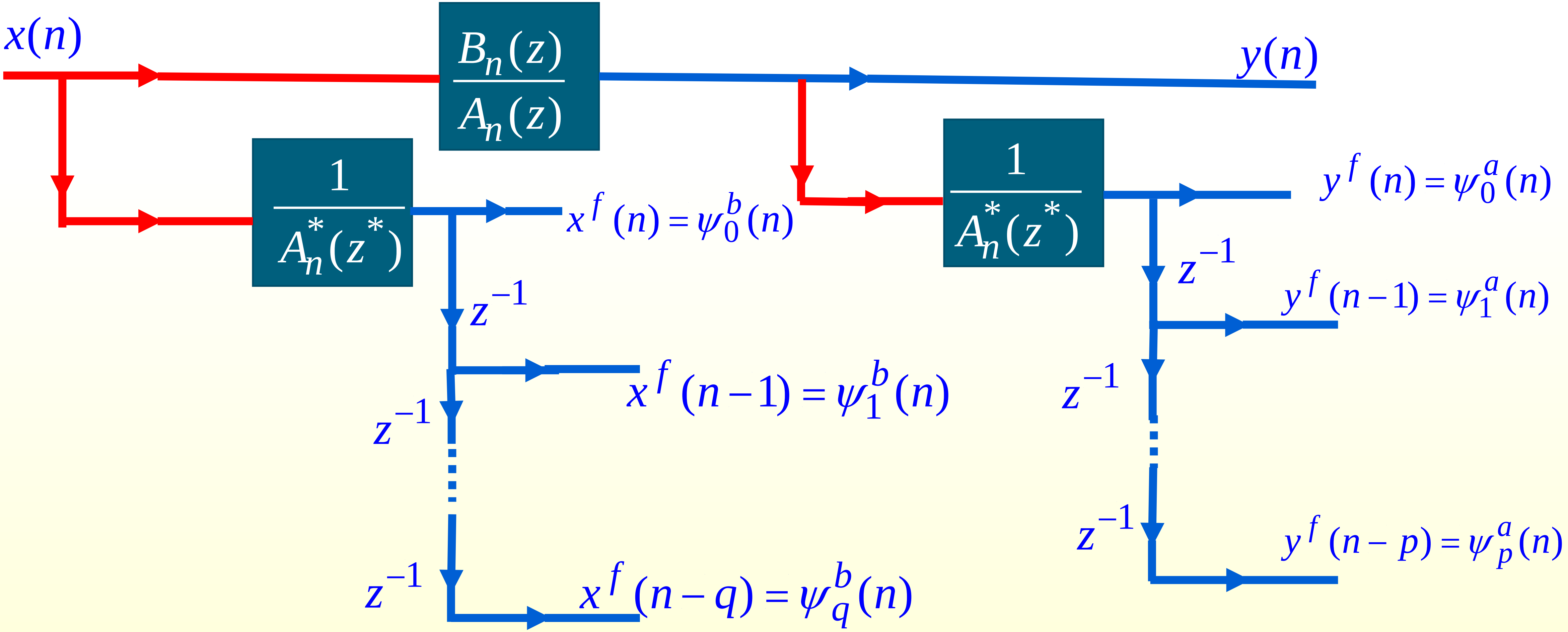


Figure 5: The IIR adaptive filter using filtered signal approach (The input of the filter must be conjugated before filtering)

Summary

- **Key Differences between IIR and FIR adaptive filters**
  - ✓ IIR adaptive filter is recursive type and has a feedback
  - ✓ FIR adaptive filter is non recursive filter and it is a Feed-forward type
- **IIR Adaptive Filter**
  - ✓ have an advantage over non-recursive ones in providing better performance for a given filter order
  - ✓ However, IIR (recursive) filters have issue of stability which may affect the convergence time and numerical sensitivity of the filter
  - ✓ Despite this challenge, there are many applications which require IIR adaptive filters such as **Echo Cancelation**
  - ✓ For application such as echo cancelation, Fir adaptive filters are impractical due to the requirement of large filter order

Summary

- **Algorithms for IIR Adaptive Filter**
  - ✓ Steepest Decent Algorithm
  - ✓ LMS Algorithm
- **Steepest Decent Algorithm**
  - ✓ The challenge related to steepest decent algorithm is , the gradient vector in the weight update equation involves expectation
  - ✓ Due to that the algorithm is computationally intensive and has issue of convergence
- **LMS Algorithm**
  - ✓ Replace the expected value in the weight update equation by instantaneous value
  - ✓ Offers less computational complexity and better convergence compared to the Steepest Decent Algorithm

Summary

- ✓ It is possible to simplify the LMS algorithm by considering small enough step size and by transforming the partial derivatives associate with the gradient vector from non recursive to recursive type
- ✓ However, the simplified IIR LMS adaptive filter involves p+q+1 recursive filters operating parallelly which increases the computation and storage requirement
- ✓ However, IIR LMS adaptive filter can be further simplified by using **Filtered Signal Approach**

References

[1] Monson H. Hayes, “Statistical Digital Signal Processing and Modeling’’, John Wiley and sons, Pp.534, 1996.

Contents Here

Thank You!