IIR Gradient Adaptive Filters

- IIR:fewer coefficients than FIR
- IIR:may be a function looked for
- Important class in modeling or identifying systems

$$H(z) = \frac{\sum_{i=0}^{N} a_i z^{-i}}{1 - \sum_{i=1}^{N} b_i z^{-i}} = \frac{A(z)}{B(z)}$$

$$\hat{Y}(z) = H(z)X(z)$$

Error and Coefficient Update

$$e(n+1) = y(n+1) - [A^T(n)\; B^T(n)]\begin{bmatrix} X(n+1) \\ \hat{Y}(n) \end{bmatrix}$$

$$\begin{bmatrix} A(n+1) \\ B(n+1) \end{bmatrix} = \begin{bmatrix} A(n) \\ B(n) \end{bmatrix} - \delta \underline{\frac{\partial}{\partial c_i} \frac{1}{2} e^2(n+1)}$$

gradient

c_i : a_i or b_i

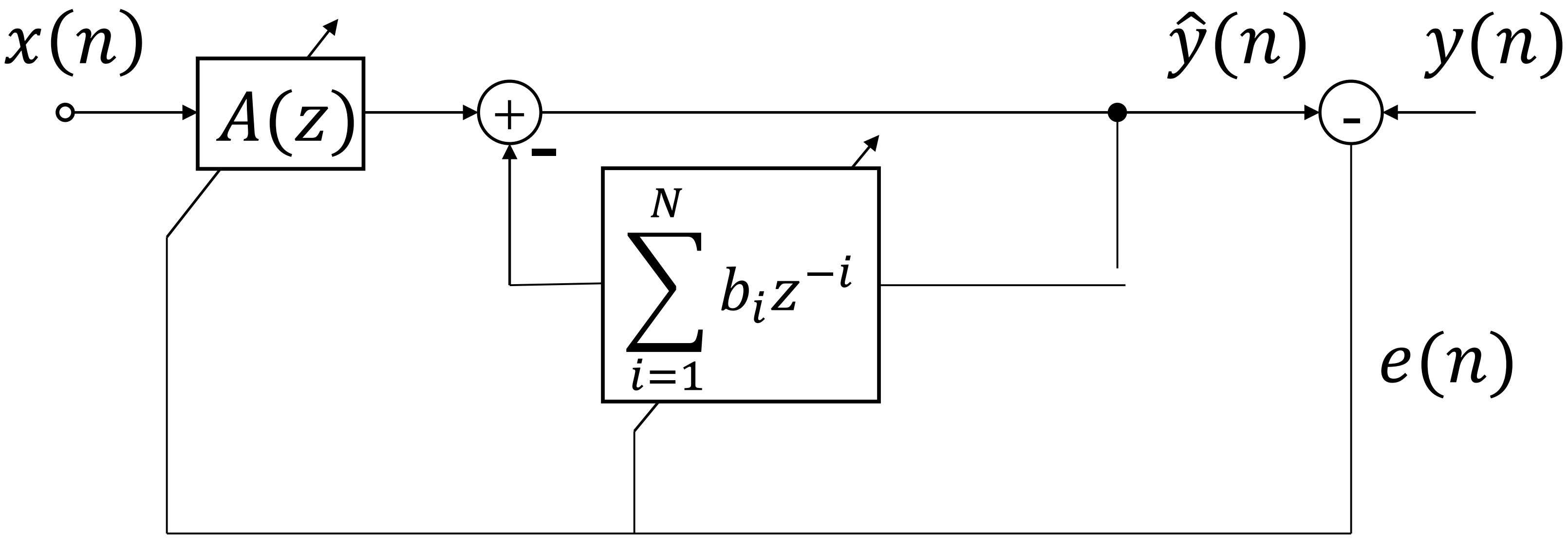
Gradient

$$\hat{y}(n) = \frac{1}{2\pi j}\oint z^{n-1} H(z)X(z)\,dz$$

$$\begin{cases} \dfrac{\partial \hat{y}(n)}{\partial a_i} = \dfrac{1}{2\pi j}\oint z^{n-1}\, z^{-i}\, \dfrac{X(z)}{B(z)}\,dz \\ \dfrac{\partial \hat{y}(n)}{\partial b_i} = \dfrac{1}{2\pi j}\oint z^{n-1}\, z^{-i}\, \dfrac{1}{B(z)} H(z)X(z)dz \end{cases}$$

gradient is calculated by applying $x(n)$ and $\hat{y}(n)$ to $\frac{1}{B(z)}$

Parallel IIR Gradient Adaptive Filter



- analysis not simple
- stability problem

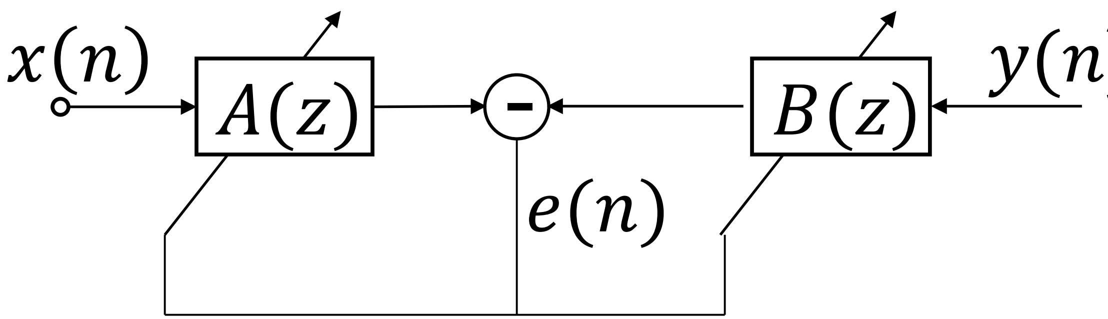
Series-Parallel IIR Gradient Adaptive Filter

After convergence

$$\hat{y}(n) \cong y(n)$$

then

$$e(n+1) \cong y(n+1) - [A^T(n)\, B^T(n)] \begin{bmatrix} X(n+1) \\ Y(n) \end{bmatrix}$$



No stability problem

Strength and Weakness of Gradient Filters

- Strength
  - Ease of design
  - Simplicity of realization
  - Flexibility
  - Robustness against signal characteristic evolution and computation errors
- Weakness
  - Dependence on signal statistics
    - best for flat spectrum
    - low speed or large residual errors for colored signal

Linear Prediction Filter

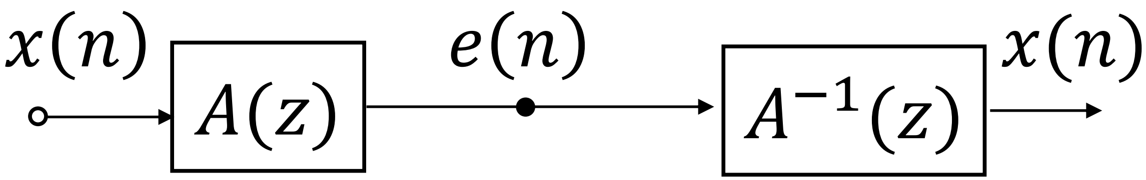
$$e(n) = x(n) - \sum_{i=1}^{\infty} a_i x(n-i)$$

- coefficients are calculated to minimize the variance of $e(n)$.
- The minimization leads to

$$E[e(n)x(n-i)] = 0, \qquad i \geq 1$$
$$E[e(n)e(n-i)] = 0, \qquad i \geq 1$$

- $e(n)$:white noise prediction error or innovation

Linear Prediction Filter and Inverse



whitening filter

model or innovation filter

$E_a = E[e^2(n)]$: prediction error variance

$S(e^{j\omega})$: input signal power spectrum density

$$E_a = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|A(e^{j\omega})\right|^2 S(e^{j\omega})d\omega$$

$$= \frac{1}{j2\pi}\oint A(z)A(z^{-1})\,S(z)\frac{dz}{z}$$

$A(z)$ is Minimum Phase

- if z_0, a zero of $A(z)$, is outside the unit circle, $|z_0| > 1$, consider $A'(z)$ given by

$$A'(z) = A(z)\frac{\left(z - \overline{z_0}^{-1}\right)\left(z - z_0^{-1}\right)}{(z - z_0)(z - \overline{z_0})}$$

$$E_a' = \frac{1}{|z_0|^2} E_a < E_a$$

which contradicts the assumption.

- Consequently, the prediction filter $A(z)$ is minimum phase.

Prediction Error Power

$$E_a = \frac{1}{j2\pi}\oint A(z)A(z^{-1})\,S(z)\frac{dz}{z}$$

$$2\pi j \ln E_a = \underbrace{\oint \ln A(z)\frac{dz}{z}}_{0} + \underbrace{\oint \ln A(z^{-1})\frac{dz}{z}}_{0} + \oint \ln S(z)\frac{dz}{z}$$

because

$A(z)$ is minimum phase --- $\ln A(z)$ is analytic for $|z| > 1$

integration contour can be a circle whose radius is arbitrarily large, and $\lim_{z\to\infty} A(z) = a_0 = 1$

$$E_a = \exp\left\{\frac{1}{j2\pi}\int_{-\pi}^{\pi} \ln S\left(e^{j\omega}\right)d\omega\right\}$$

Kolmogoroff-Szego formula

Linear Prediction Coefficients

$$e(n) = x(n) - \sum_{i=1}^{N} a_i x(n-i)$$

$$\frac{\partial}{\partial a_j} E[e^2(n)] = r(j) - \sum_{i=1}^{N} a_i r(j-i) = 0, \qquad 1 \le j \le N$$

which can be completed by the power relation

$$E_{aN} = E[e^2(n)] = r(0) - \sum_{i=1}^{N} a_i r(i)$$

In concise form $R_{N+1} \begin{bmatrix} 1 \\ -A_N \end{bmatrix} = \begin{bmatrix} E_{aN} \\ 0 \end{bmatrix}$

where

$$R_{N+1} = \left[\begin{array}{c|ccc} r(0) & r(1) & \cdots & r(N) \\ \hline r(1) & & & \\ \vdots & & R_N & \\ r(n) & & & \end{array}\right], R_N = E[X(n)X^T(n)]$$

First-Order FIR Predictor

$H(z) = 1 - az^{-1}$

can be applied to a constant signal in white noise with power σ_b^2

$x(n) = 1 + b(n)$

The prediction error power

$$
\begin{aligned}
E[e^2(n)] &= |H(1)|^2 + \sigma_b^2(1 + a^2) \\
&= (1 - a)^2 + \sigma_b^2(1 + a^2)
\end{aligned}
$$

$$\frac{\partial E[e^2(n)]}{\partial a} = 0 \quad \rightarrow \quad a = \frac{1}{1 + \sigma_b^2}$$

First-Order FIR Predictor (cont'd)

Residual prediction error $(1-a)^2$

Amplified noise power $\sigma_b^2(1+a^2)$

The former is much smaller

Forward and Backward Prediction

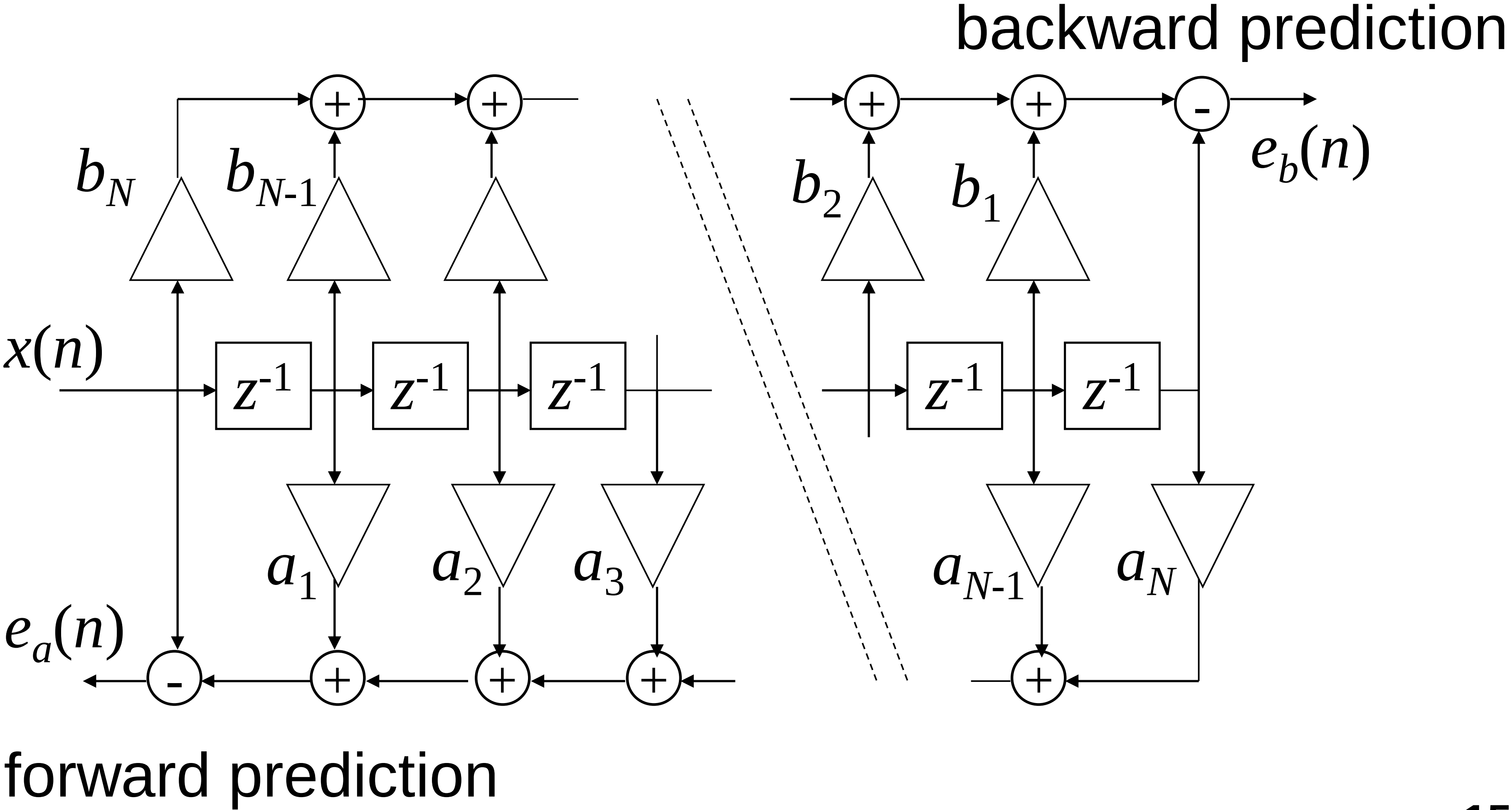
For finite order, the oldest sample is discarded every time a new sample is acquired.
the loss of the oldest sample is characterized by backward linear prediction.

- Forward linear prediction $e_a(n) = x(n) - \sum_{i=1}^{N} a_i x(n-i)$

  or $e_a(n) = x(n) - A_N^T X(n-1)$

- Backward linear prediction $e_b(n) = x(n-N) - B_N^T X(n)$

Forward and Backward Linear Prediction Error Filter



Backward Prediction

Minimization of prediction error power leads to

$$R_{N+1}\begin{bmatrix}-B_N\\1\end{bmatrix}=\begin{bmatrix}0\\E_{bN}\end{bmatrix}$$

$$J_{N+1}R_{N+1}\begin{bmatrix}-B_N\\1\end{bmatrix}=\begin{bmatrix}E_{bN}\\0\end{bmatrix}$$

$$R_{N+1}\begin{bmatrix}1\\-J_N B_N\end{bmatrix}=\begin{bmatrix}E_{bN}\\0\end{bmatrix}$$

$$J=\begin{bmatrix}0&\cdots&1\\\vdots&\iddots&\vdots\\1&\cdots&0\end{bmatrix}$$

co-identity matrix

Hence

$$A_N=J_N B_N,\qquad E_{aN}=E_{bN}=E_N$$

Forward and Backward Prediction

For the stationary signals, the two are equal, and the coefficients are the same. Difference appears in the transition phases.

- Forward linear prediction error filter is minimum phase.
- Backward filter is maximum phase.

Order Iterative Relations

Let $r_N^a = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(N) \end{bmatrix}$, $r_N^b = J_N r_N^a$

$$\left[\begin{array}{c|c} R_N & r_N^b \\ \hline \left(r_N^b\right)^T & r(0) \end{array}\right] \left[\begin{array}{c} 1 \\ -A_{N-1} \\ \hline 0 \end{array}\right] = \left[\begin{array}{c} E_{N-1} \\ 0 \\ \hline K_N \end{array}\right]$$

where $K_N = r(N) - \sum_{i=1}^{N-1} a_{i,N-1}\, r(N-i)$

For backward linear prediction,

$$\left[\begin{array}{c|c} r(0) & \left(r_N^a\right)^T \\ \hline r_N^a & R_N \end{array}\right] \left[\begin{array}{c} 0 \\ \hline -B_{N-1} \\ 1 \end{array}\right] = \left[\begin{array}{c} K_N \\ \hline 0 \\ E_{N-1} \end{array}\right]$$

Order Iterative Relations (cont'd)

Multiplying both side by $k_N = K_N / E_{N-1}$

$$R_{N+1}\begin{bmatrix} 0 \\ \begin{bmatrix} -B_{N-1} \\ 1 \end{bmatrix} k_N \end{bmatrix} = \begin{bmatrix} k_N^2 E_{N-1} \\ 0 \\ K_N \end{bmatrix}$$

Subtracting this from the equation in the previous page leads to order N linear prediction equation, which implies

$$A_N = \begin{bmatrix} A_{N-1} \\ 0 \end{bmatrix} - k_N \begin{bmatrix} B_{N-1} \\ -1 \end{bmatrix},$$ the last row is $a_{NN} = k_N$

and $E_N = E_{N-1}(1 - k_N^2)$

This recursive solution of prediction matrix equation is called Levinson-Durbin algorithm.

Lattice Linear Prediction Filter

The coefficient k_n establish direct relations between forward and backward prediction errors for consecutive orders.

from $\begin{cases} e_{aN}(n) = x(n) - A_N^T X(n-1) \\ A_N = \begin{bmatrix} A_{N-1} \\ 0 \end{bmatrix} - k_N \begin{bmatrix} B_{N-1} \\ -1 \end{bmatrix} \end{cases}$

we have $e_{aN}(n) = e_{aN-1}(n) - k_N [B_{N-1}^T \quad 1] X(n-1)$

backward prediction error

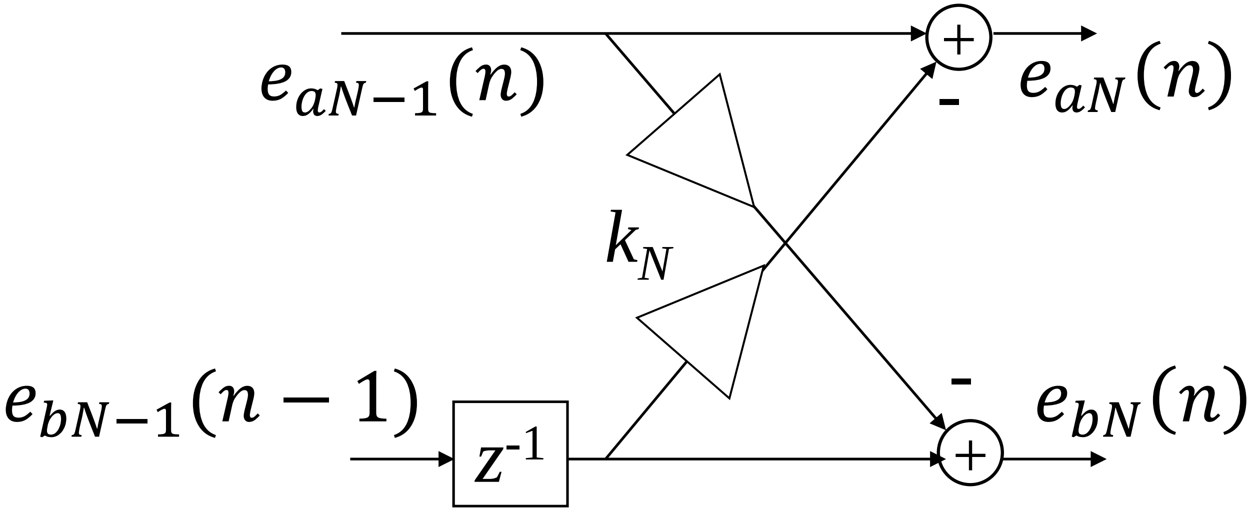
$e_{bN}(n) = x(n-N) - B_N^T X(n)$

for order $N-1$

$e_{bN-1}(n) = [-B_{N-1}^T \quad 1] X(n)$

Lattice Linear Prediction Filter

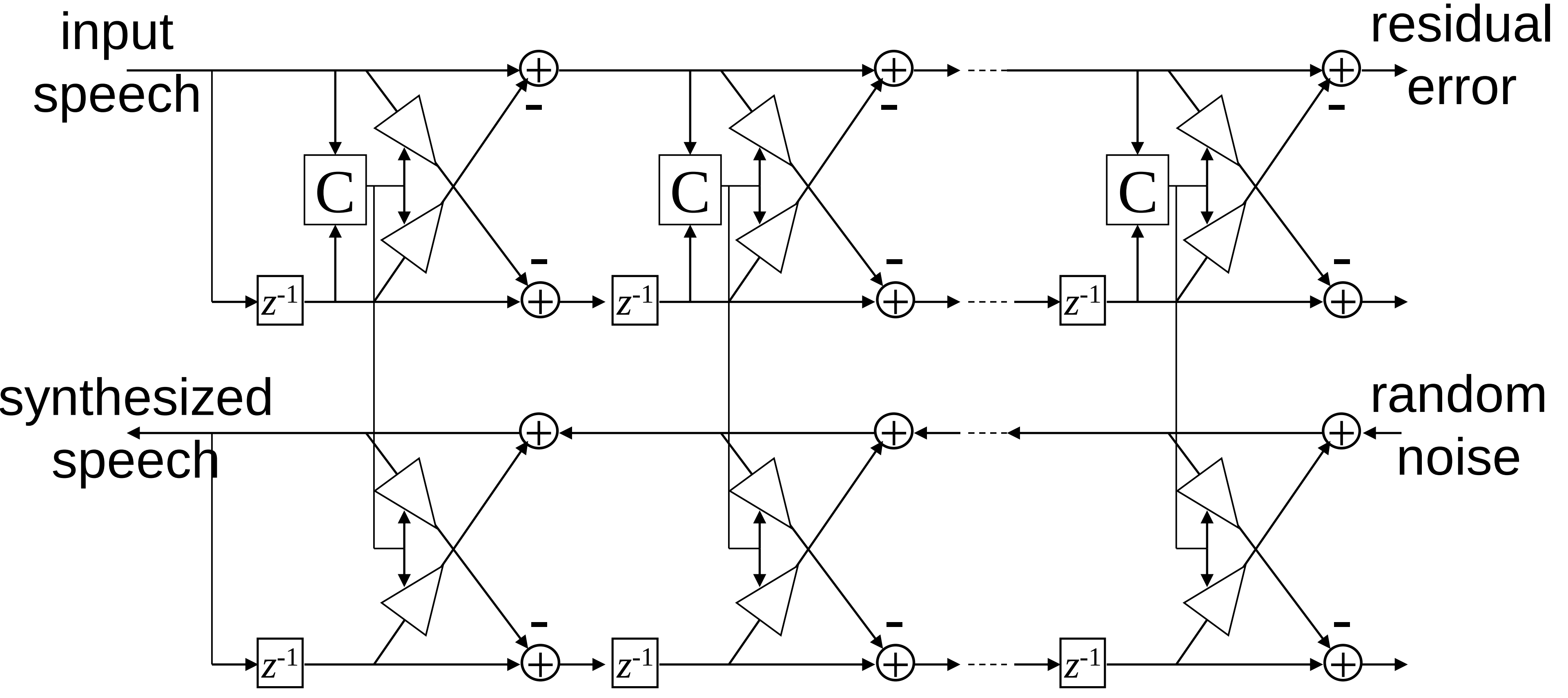
$$\begin{cases} e_{aN}(n) = e_{aN-1}(n) - k_N e_{bN-1}(n-1) \\ e_{bN}(n) = e_{bN-1}(n) - k_N e_{aN-1}(n) \end{cases}$$



Lattice linear prediction filter section

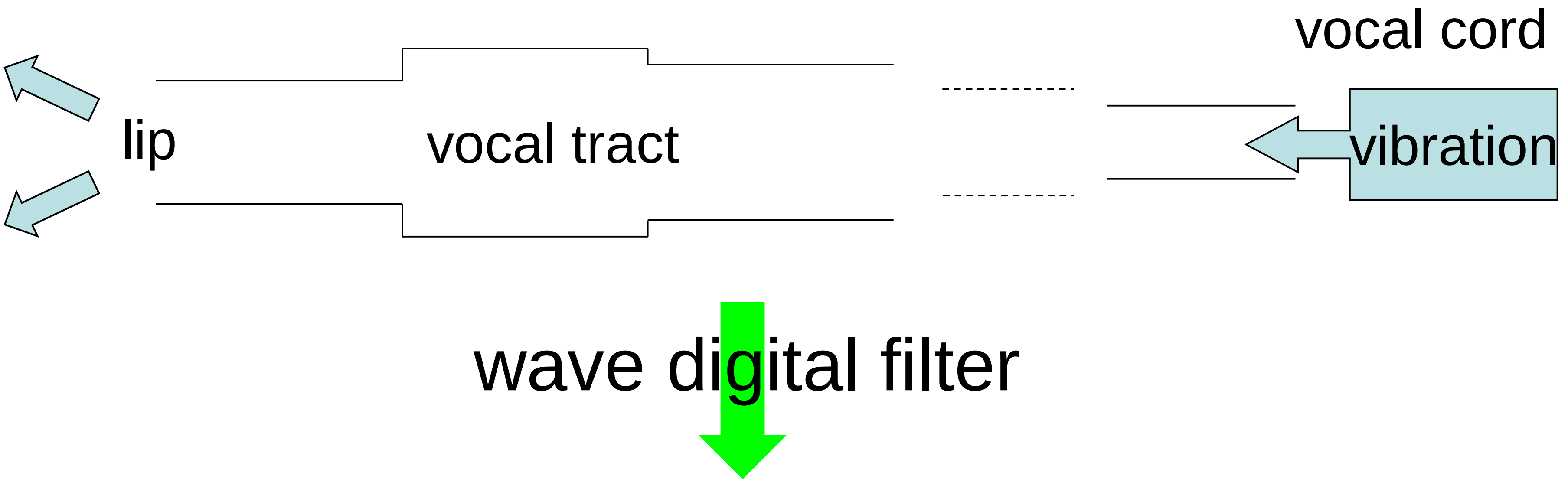
PARCOR

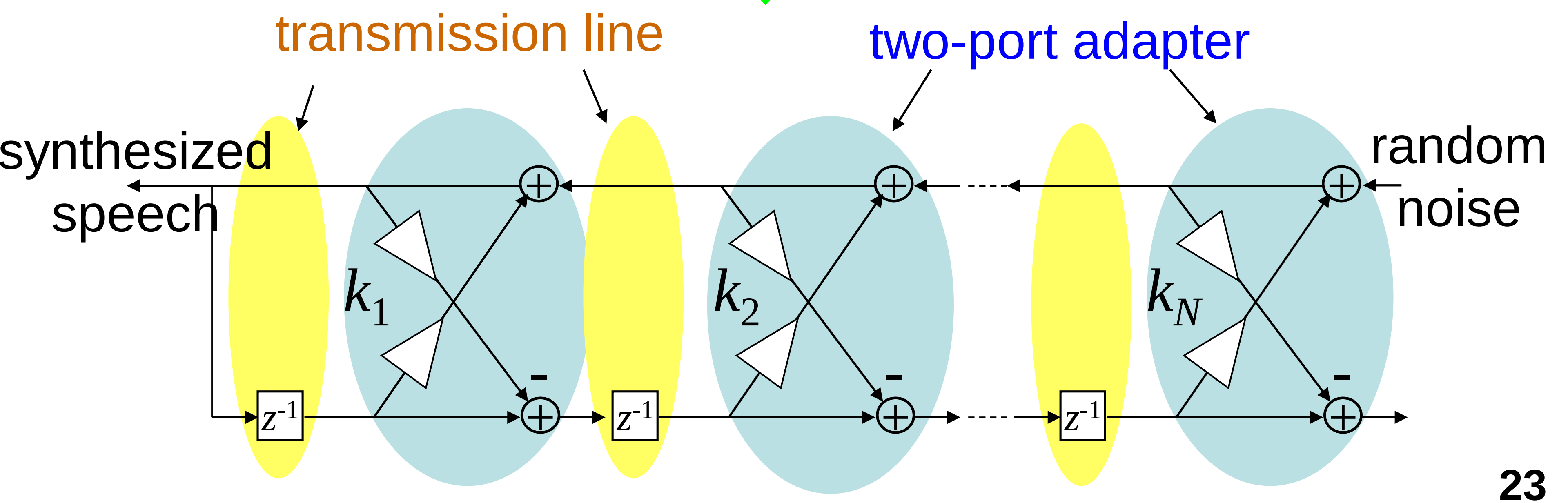
Partial autocorrelation



C: correlator

Speech Synthesis Filter





Exercise 9

1. Express the error outputs of parallel IIR gradient AF and series-parallel IIR gradient AF in *z*-domain, and compare them.
2. Calculate the first four terms of the autocorrelation function of the signal
   $x(n) = \sin\frac{\pi n}{4}$.
   Using the normal equations, calculate the coefficients of the predictor of order $N = 3$.
3. Consider the three systems $(0 < a, b < 1)$
   $y_n = x_n - (a+b)x_{n-1} + abx_{n-2}$,
   $y_n = abx_n - (a+b)x_{n-1} + x_{n-2}$,
   $y_n = ax_n - (a+b)x_{n-1} + bx_{n-2}$.
   What are the system functions for these systems? Which system is minimum phase, which maximum phase, and which mixed phase? Take x_n to be a zero mean stationary white noise, with $\langle x_n x_{n+m} \rangle = \delta_m$ and $\left\langle x_n x_{n+m_1} x_{n+m_2} \right\rangle = \delta_{m_1 m_2}$. Compare the autocorrelations of the outputs of those systems.
4. Read the following paper;
   J. Makhoul,."Linear prediction: A tutorial review", Proceedings of the IEEE, Vol.63, 4, pp. 561-580, Apr. 1975