4.1 Considerations on time variant linear systems

4.1.1 Properties of the Difference Polynomial Operator

- **Definition 1:** The time-shift operator \( q^{-i} \) maps \( \ell \) into \( \ell \) and is described by

\[
q^{-i} : \ell \rightarrow \ell \\
q^{-i} \{x(n)\} = x(n - i)
\]  

(22)

- **Property 2:** The time-shift operator is a linear operator, then

\[
q^{-i} \{c \, x(n) + d \, y(n)\} = c \, x(n - i) + d \, y(n - i)
\]  

(23)

- **Property 3:** A more general result for the time-shift operator results from its own definition and can be written as

\[
q^{-i} \{f(x(n), y(n))\} = f(x(n - i), y(n - i))
\]  

(24)

where \( f(\ldots) \) is any given function defined in the discrete-time domain.

- **Property 4:** The linear combination of time-shift operators is performed as follows

\[
(c \, q^{-i} + d \, q^{-j}) \{x(n)\} = c \, x(n - i) + d \, x(n - j)
\]  

(25)

- **Property 5:** The concatenation of time-shift operators is performed as follows

\[
q^{-j} \{q^{-i} \{x(n)\}\} = q^{-j} \{x(n - i)\} = x(n - i - j) = q^{-(i+j)} \{x(n)\}
\]  

(26)

- **Property 6:** The division of time shift operators follows the rule

\[
\left( \frac{q^{-i}}{q^{-j}} \right) \{x(n)\} = q^{-(i-j)} \{x(n)\} = x(n - i + j)
\]  

(27)
An important extension, the difference polynomial operator (DPO)

- Definition 7: The extrapolation of equation (25) for several terms results in

\[
\tilde{C}\{x(n)\} = (c_{n_c} + ... + c_1 q^{n_c-1} + c_0 q^{n_c})\{x(n)\} \\
= c_{n_c}x(n) + ... + c_1x(n + n_c - 1) + c_0x(n + n_c)
\]  \hspace{1cm} (28)

- With adaptive filters, is common the causal form of the DPO

\[
C(q) = q^{-n_c}\tilde{C}(q) = c_0 + c_1q^{-1} + ... + cn_cq^{-n_c}
\]  \hspace{1cm} (29)

- Then the DPO in the frequency domain \(Z\{q^{-i}\{x(n)\}\} = z^{-i}X(z)\), then

- Property 8: Extending ideas to transfer function

\[
y(n) = H(q)\{x(n)\} \leftrightarrow Y(z) = H(z)X(z)
\]  \hspace{1cm} (30)

- Property 10: a DPO (one with at least one nonzero coefficient) represents a bijective operator in the subspace of one-sided sequences \(x(n)\), such that \(x(n) = 0, \forall n < 0\).

- Property 11: The inverse DPO exists and it is defined by

\[
\left(\frac{1}{C(q)}\right)\{x(n)\} = C^{-1}(q)\{x(n)\}
\]  \hspace{1cm} (31)

in such a way that \(C^{-1}(q)\{C(q)\{x(n)\}\} = (C^{-1}(q)C(q))\{x(n)\}\).

- Property 12: The concatenation of direct and inverse DPOs is a commutative operation, i.e.,

\[
\left(\frac{1}{C(q)}\right)\left\{\frac{1}{D(q)}\{x(n)\}\right\} = \left(\frac{1}{D(q)}\right)\left\{\frac{1}{C(q)}\right\}
\]  \hspace{1cm} (32)
\[
C(q)\{D(q)\{x(n)\}\} = D(q)\{C(q)\{x(n)\}\}
\]  \hspace{1cm} (33)
\[
\left(\frac{C(q)}{D(q)}\right)\{x(n)\} = \left(\frac{1}{D(q)}\right)\{C(q)\{x(n)\}\}
\]  \hspace{1cm} (34)
4.1.2 The Time-varying Difference Polynomial Operator

- **Definition 13**: The TVDPO is defined as

\[
C(q, n)\{x(n)\} = (c_0(n) + c_1(n)q^{-1} + ... + c_n(n)q^{-n_e})\{x(n)\} \\
= c_0(n)x(n) + c_1(n)x(n-1) + ... + c_n(n)x(n-n_e)
\] (35)

- **Property 14**: The concatenation of a TVDPO with either a DPO or a TVDPO is not a commutative operation.

- **Example**: Consider the two first-order TVDPOs \(C(q, n) = 1 + c_1(n)q^{-1}\) and \(D(q, n) = 1 + d_1(n)q^{-1}\) with \(C(q, n) \neq D(q, n)\). Defining

\[
e_1(n) = C(q, n)\{D(q, n)\{x(n)\}\} \\
e_2(n) = D(q, n)\{C(q, n)\{x(n)\}\}
\] (36) (37)

it is easy to verify that

\[
e_1(n) - e_2(n) = [c_1(n)d_1(n-1) - c_1(n-1)d_1(n)]x(n - 2)
\] (38)

is generally different from zero, implying that \(e_1(n)\) and \(e_2(n)\) are two distinct sequences.

- **Property 15**:

\[
(C(q, n))^2\{x(n)\} = C(q, n)\{C(q, n)\{x(n)\}\} \\
= C^2(q, n)\{x(n)\}
\] (39)

\[
\left(\frac{C(q, n)}{D(q, n)}\right)\{x(n)\} = \left(\frac{1}{D(q, n)}\right)\{C(q, n)\{x(n)\}\} \\
\neq C(q, n)\left\{\frac{1}{D(q, n)}\{x(n)\}\right\}
\] (40)

- The difference between both sides of the inequality approximate zero if we assume the coefficients of the TVDPO essentially constant, i.e., the small step approximation.
4.1.3 Stability of time varying recursive filters

- Consider the state space description of a time-varying recursive filter

\[
\begin{bmatrix}
    x(n+1) \\
    \hat{y}(n)
\end{bmatrix} =
\begin{bmatrix}
    A(n) & b(n) \\
    c(n) & d(n)
\end{bmatrix}
\begin{bmatrix}
    x(n) \\
    u(n)
\end{bmatrix}
\]

- To study the small step approximation, consider also a time invariant system with the property that its fixed parameters agree with those of the previous equation at time \( n \), i.e.,

\[
\begin{bmatrix}
    A(k) & b(k) \\
    c(k) & d(k)
\end{bmatrix} =
\begin{bmatrix}
    A(n) & b(n) \\
    c(n) & d(n)
\end{bmatrix}
\]

for all \( k \leq n \).

- If the parameters vary slowly, i.e.,

\[
\left\| \begin{bmatrix}
    A(k) & b(k) \\
    c(k) & d(k)
\end{bmatrix} - \begin{bmatrix}
    A(n) & b(n) \\
    c(n) & d(n)
\end{bmatrix} \right\| \leq \epsilon
\]

for all \( k \leq n \) and with \( \epsilon \) small. An when \( \epsilon \to 0 \) the two systems must coincide (in the limit).
• Consider now being approximating a rational system with instantaneous error given, with this approximation, by

\[ e(n) = \sum_{k=0}^{\infty} (h_k - \hat{h}_k)u(n - k) \]

\[ = (h_0 - d)u(n) + \sum_{k=1}^{\infty} (h_k - cA^{k-1}b)u(n - k) \]

where \( \begin{bmatrix} A(k) & b(k) \\ c(k) & d(k) \end{bmatrix} = \begin{bmatrix} A & b \\ c & d \end{bmatrix} \) for all \( k \) was used.

• Since the approximating system has time varying coefficients

\[ e(n) = (h_0 - d(n))u(n) + \sum_{k=1}^{\infty} (h_k - c(n)\Phi(n, n - k + 1)b(n - k))u(n - k) \]

where \( \Phi(n, n - k + 1) = \begin{cases} I & k = 1; \\ A(n - 1)A(n - 1)...A(n - k + 1) & k > 1. \end{cases} \)

• Clearly both errors are similar if the time-varying parameter change is sufficiently slow.

• If \( a_k(n) = 0 \), an FIR filter approximation, then the error is

\[ e(n) = \sum_{k=0}^{N} (h_0 - b_k(n))u(n - k) + \sum_{k=N+1}^{\infty} h_ku(n - k) \]

• To quantify the similarity of both systems we verify BIBO stability.

• Two remarks:
  
  – Stability of time-varying IIR filters is a generic necessary condition for a possible parameter updating algorithm to converge.
  
  – Stability properties of time-varying IIR filters can vary with the specific realization chosen.
• The concept of **exponential stability** and Liapunov methods are helpful to verify BIBO stability.

• If
  
  - The elements of the state space description remain bounded for all time \( n \);
  
  - \( \mathbf{x}(n+1) = \mathbf{A}(n)\mathbf{x}(n) \) (the homogeneous part) remains exponentially stable.

  the time-varying system is BIBO stable.

• By the second condition, for any bounded initial condition \( \|\mathbf{x}(n)\| < \infty \), then \( \|\mathbf{x}(m)\| \leq \beta \alpha^{m-n} \|\mathbf{x}(n)\| \), for all \( m \geq n \), where \( \beta > 0 \) and \( 0 \leq \alpha < 1 \).

• Then using \( \Phi(m,n) \) we obtain

\[
\mathbf{x}(m) = \Phi(m,n)\mathbf{x}(n) + \sum_{k=n}^{m-1} \Phi(m-1,k)\mathbf{b}(k)u(k)
\]

for all \( m > n \), from where it is not hard to see that \( \mathbf{x}(m) \) remains bounded, and then \( \dot{y}(n) \), which implies BIBO stability.
• A constructive way to shown exponential stability: Liapunov method, can lead to BIBO stability without appeal to the small step approximation or slow parameter variation approximation.

• To quantify the slow parameter approximation we consider

$$\|A(n + 1) - A(n)\| \leq \epsilon$$

or $x(n + N) = A(n + N - 1)\ldots A(n + 1)A(n)x(n)$ is approximated by $\bar{x}(n + N) = A^N\bar{x}(n)$.

• But this is stable if $A(n)$ has all its eigenvalues inside the unit circle. In particular, this is true if exist $P > 0$ for $P_2 = P - (A^N)^T PA^N > 0$.

• Then for the fixed case, this leads to

$$\bar{x}^T(n)P\bar{x}(n) - \bar{x}^T(n + N)P\bar{x}(n + N) = \bar{x}^T(n)P_2\bar{x}(n)$$

$$\|\bar{x}(n)\|_P^2 - \|\bar{x}(n + N)\|_P^2 = \|\bar{x}(n)\|_{P_2}^2 \geq c_1\|\bar{x}(n)\|_P^2$$

where $c_1\|v\|_P \leq \|v\|_{P_2} \leq c_2\|v\|_P$ was used. Then

$$\|\bar{x}(n + N)\|_P^2 \leq (1 - c_1)\|\bar{x}(n)\|_P^2$$

• The small step approximation leads to

$$\|\bar{x}(n + N) - x(n + N)\|_P^2 \leq \delta$$

where $\delta$ is a constant forced to be small if $\epsilon$ is small. This then must lead to

$$\|x(n + N)\|_P < \|x(n)\|_P$$
• A constructive way to verify the exponential stability of time-varying systems is using the Liapunov equation.

• Consider \{A(.)\} related to our time varying system, that satisfy \( \mathbf{P} - \mathbf{A}^T(n)\mathbf{P}\mathbf{A}(n) = \mathbf{C}(n)\mathbf{C}^T(n) > 0 \).

• Define \( \mathbf{y}(n) = \mathbf{C}^T(n)\mathbf{x}(n) \).

• Following similar ideas than previously, and considering a time window of \( N \) time instants

\[
\begin{align*}
\|\mathbf{x}(n)\|_\mathbf{P}^2 - \|\mathbf{x}(n+1)\|_\mathbf{P}^2 &= \|\mathbf{y}(n)\|_I^2 \\
\|\mathbf{x}(n+1)\|_\mathbf{P}^2 - \|\mathbf{x}(n+2)\|_\mathbf{P}^2 &= \|\mathbf{y}(n+1)\|_I^2 \\
&\vdots \\
\|\mathbf{x}(n+N-1)\|_\mathbf{P}^2 - \|\mathbf{x}(n+N)\|_\mathbf{P}^2 &= \|\mathbf{y}(n+N-1)\|_I^2
\end{align*}
\]

Summing

\[
\|\mathbf{x}(n)\|_\mathbf{P}^2 - \|\mathbf{x}(n+N)\|_\mathbf{P}^2 = \sum_{k=n}^{N+M-1} \|\mathbf{y}(k)\|_I^2
\]

• Note that

\[
\begin{bmatrix}
\mathbf{y}(n) \\
\mathbf{y}(n+1) \\
\vdots \\
\mathbf{y}(n+N-1)
\end{bmatrix}
= \begin{bmatrix}
\mathbf{C}^T(n) \\
\mathbf{C}^T(n+1)\mathbf{A}(n) \\
\vdots \\
\mathbf{C}^T(n+N-1)\Phi(n+N-1, n)
\end{bmatrix}
\begin{bmatrix}
\mathbf{x}(n) \\
\mathbf{O}(n+N-1, n)\mathbf{x}(n)
\end{bmatrix}
\]
• Also note that

Definition: The pair \((A(.), C^T(.))\) over the time window \(n, n+1, \ldots, m\) is uniformly observable if there exists an integer \(N\) and positive constants \(c_1\) and \(c_2\) such that

\[
0 < c_1 I \leq O^T(n + N - 1, n)O(n + N - 1, n) \leq c_2 I < \infty
\]

• Then, if this property is verified

\[
\|x(n)\|^2_P - \|x(n + N)\|^2_P \geq \|x(n)\|^2_P O^T(n + N - 1, n)O(n + N - 1, n)x(n) \geq c_1\|x(n)\|^2_P \geq c_1c_3\|x(n)\|^2_P
\]

or

\[
\|x(n + N)\|^2_P \geq (1 - c_1c_3)\|x(n)\|^2_P
\]

Note that \(0 \leq (1 - c_1c_3) < 1\) since \(\|x(n + N)\|^2_P\) is positive.

• Increasing \(n\) by \(N\) steps successively we obtain

\[
\|x(n + kN)\|^2_P \geq (1 - c_1c_3)^k\|x(n)\|^2_P
\]

but the "worst" case decay is to maintain constant \(\|x(n)\|^2_P\) by \(N\) iterations.

• Then, this leads to the following equation

\[
\|x(m)\|^2_P \geq \beta^2 (1 - c_1c_3)^{(m-n)/N}\|x(n)\|^2_P
\]

for all \(m > n\), for some constant \(\beta\). This is nothing more than the exponential equation related to stability by recognizing that \(\alpha = (1 - c_1c_3)^{1/2N}.\)
• To summarize

Theorem: The homogeneous time varying system \( \dot{x}(n+1) = A(n)x(n) \) is exponentially stable provided there exists a symmetric positive definite matrix \( P \) fulfilling a Liapunov equation

\[
P - A^T(n)PA(n) = C(n)C^T(n)
\]

such that the resultant sequence \( \{C(.)\} \) gives \( (A(.), C^T(.)) \) as a uniformly observable pair.
4.2 The Ordinary Difference equation

4.2.1 ODE association to IIR Adaptive Algorithms

The ODE method can be used in adaptive algorithms of the following general form

\[
\begin{align*}
\theta(n + 1) &= \theta(n) + \alpha \hat{R}^{-1}(n + 1) \Psi(n) e(n) \\
\hat{R}(n + 1) &= \hat{R}(n) + \alpha [\Psi(n) \Psi^H(n) - \hat{R}(n)]
\end{align*}
\]  

(41)

where \( \theta(n) \) is the parameter vector, \( \Psi(n) \) is the regressor, and \( e(n) \) is the prediction error.

The average behavior of \( \theta(n) \) and \( \hat{R}(n) \) in the previous algorithm can be studied by the solution of the following associated ODE

\[
\begin{align*}
\frac{\partial \theta(t)}{\partial t} &= \mathcal{G}^{-1}(t) E[\Psi(n) e(n)] \\
\frac{\partial \mathcal{G}(t)}{\partial t} &= E[\Psi(n) \Psi^T(n)] - \mathcal{G}(t)
\end{align*}
\]  

(42)
In order to justify this association, there exist two behaviors that define the convergence analysis of interest:

- constant $\alpha$ (weak convergence or convergence in distribution).
- decrescent $\alpha$ (convergence with probability one).

The conditions for ODE association for both cases are summarized as follows.

- decrescent $\alpha$.
  - a) $\alpha(n) \to 0$ for $n \to \infty$, i.e., the convergence is possible independ-ently of the stochastic environment.
  - b) $\sum_{n=1}^{\infty} \alpha(n) = \infty$, i.e., the estimate is reached through an arbitrary number of iterations.

A typical example is $\alpha(n) = 1/n$. Note that $t = \sum_{k=1}^{n} \alpha(k)$.

1. Smoothness
2. Regularity
3. Boundness
4. Liapunov Function.

- $\alpha$ constant.

In this case $t = n\alpha$, and $t = n\alpha \to \infty$ for $\alpha \to 0$

1. 1 to 3.: Analogous to the previous case.
2. Stationarity.

Some properties of the analysis by the ODE associated

- $\vartheta(t)$ converge to the stable stationary points of equation (42).
- The ODE trajectories determine an average asymptotic path for the estimate $\theta(n + 1)$.
- A drawback of the analysis by the ODE method is that the information related to convergence speed is lost.
4.2.2 Heuristics to ODE approximation

Defining $\nabla(n)$ as the regressor vector and $p(n)$ as the parameter vector, the stochastic gradient version of the updating equation for can be written as follows

$$p(n+1) = p(n) + \mu e(n, \{p(n)\}) \nabla(n, \{p(n)\})$$

For $N$ successive time instants ($N$ sufficiently large)

$$p(n+1) - p(n) = \mu e(n, \{p(n)\}) \nabla(n, \{p(n)\})$$

$$p(n+2) - p(n+1) = \mu e(n+1, \{p(n+1)\}) \nabla(n+1, \{p(n+1)\})$$

$$\vdots$$

$$p(n+N) - p(n+N-1) = \mu e(n+N-1, \{p(n+N-1)\}) \nabla(n+N-1, \{p(n+N-1)\})$$

or

$$p(n + N) - p(n + N - 1) = \mu \sum_{k=n}^{N-1} e(k, \{p(k)\}) \nabla(k, \{p(k)\})$$

If $\mu$ is sufficiently small (i.e., the small step approximation) such that

$$p(n) \cong p(n+1) \ldots \cong p(n+N-1)$$

then

$$e(k, \{p(k)\}) \cong e(k/p(n))$$

$$\nabla(k, \{p(k)\}) \cong \nabla(k/p(n))$$

for $k \geq n$, i.e., each output of these filters is considered an stationary process.

Indeed, if the reference and the input signals are stationary then $e(.)$ and $\nabla(.)$ are stationary if $p(.)$ is in a well defined domain.
Also, a natural approximation (ergodic assumption)

\[
\sum_{k=n}^{n+N-1} e(k/p(n)) \nabla(k/p(n)) \cong NE[e(n/p(n)) \nabla(n/p(n))]
\]

or

\[
\frac{p(n + N) - p(n)}{N} \cong E[e(n/p(n)) \nabla(n/p(n))]
\]

Introducing now a change of variables, specifically: a continuous time \( t \), such that \( t = n \) and \( \Delta t = \mu N \), we obtain

\[
\frac{p(t + \Delta t) - p(t)}{\Delta t} \cong E[e(n/p(n)) \nabla(n/p(n))] \overset{\Delta}{=} f(p)
\]

And finally, for \( \mu \ll 1 \) or \( \Delta t \to 0 \),

\[
\frac{\partial p(t)}{\partial t} = f(p)
\]
4.2.3 Stability analysis

Let \( p_* \) a convergence point of the ODE, then

- \( p_* \) is an **stationary point**, if \( f(p_*) = 0 \) such that \( \frac{\partial p(t)}{\partial t} = 0 \), or

\[
\{p_*\} = \{p : f(p) = 0\}
\]

- \( p_* \) is an **stable stationary point** (attractor), when

\[
\forall p(0), \text{ such that } |p_* - p(0)| < \epsilon, \text{ then } p(t) \xrightarrow{t \to \infty} p_*
\]

- if \( p_* \) not satisfies the previous condition is is an **unstable stationary point**.

**Liapunov direct method**

- Exist \( L(p) > 0 \) a scalar function, with local minimum in \( p_* \), i.e.

\[
L(p(t)) > L(p_*) \geq 0
\]

\[
\forall p(t) \neq p_*, \text{ such that } |p(t) - p_*| \leq \epsilon.
\]

- \( L(p(t)) \) decrescent in all trajectories of \( p(t) \), i.e.

\[
\frac{\partial L(p(t))}{\partial t} = \frac{\partial L(p(t))}{\partial p(t)} \frac{\partial p(t)}{\partial t} = \frac{\partial L(p(t))}{\partial p(t)} f(p(t)) \leq 0
\]

\[
\forall p(t) \neq p_*, \text{ such that } |p(t) - p_*| \leq \epsilon.
\]

then \( p_* \) is a local stationary point of the ODE.

**Liapunov indirect method**

Linearization around a stationary point, i.e.

\[
f(p_* + \Delta p) = f(p_*) + \left[ \frac{\partial f(p)}{\partial p} \right]_{p = p_*} \Delta p + O(\Delta p) \Delta p
\]

then \( p_* \) is a local stationary point if \( \frac{\partial p(u)}{\partial t} = \left[ \frac{\partial f(p)}{\partial p} \right]_{p = p_*} \) has all their eigenvalues with negative real part.
4.3 Some linear systems concepts

- Consider the state variable equation

\[
\begin{bmatrix}
  x(n+1) \\
  \dot{y}(n)
\end{bmatrix}
= \begin{bmatrix}
  A & b \\
  c & d
\end{bmatrix}
\begin{bmatrix}
  x(n) \\
  u(n)
\end{bmatrix}
\]

then \( \hat{H}(z) = d + c(zI - A)^{-1}b = \sum_{k=0}^{\infty} \hat{h}_k z^{-k} \) with \( \hat{h}_k = \begin{cases} 
  d, & k=0; \\
  cA^{k-1}b, & k=1,2,\ldots
\end{cases} \).

- The controllability and observability grammians fulfill

\[
K = AK^T A + bb^T \\
W = AW^T A + c^T c
\]

where \( K \) (respectively \( W \)) is definite positive if and only if \((A, b)\) (resp. \((A, c)\)) is completely controllable (resp. completely observable), or \( \hat{H}(z) \) is minimal.

This result is closely related to the ”infinite horizon” controllability (resp. observability) matrix

\[
\mathcal{C} = \begin{bmatrix}
  b & Ab & \cdots & A^{N-1}b & \cdots & A^{N+k}b & \cdots
\end{bmatrix} \\
\mathcal{O} = \begin{bmatrix}
  c & cA & \cdots & cA^{N-1} & \cdots & cA^{N+k} & \cdots
\end{bmatrix}^T
\]

since \( K = \mathcal{C} \mathcal{C}^T \) (resp. \( W = \mathcal{O}^T \mathcal{O} \)).
Also, driving with a unit-variance white noise sequence the state variable filter, it is possible to obtain

$$E\{x(n)x^T(n)\} = A E\{x(n)x^T(n)\} A^T + bb^T$$

in such a way that $K$ represent the state covariance matrix.

- Consider the following double infinite **Hankel form matrix**

$$\Gamma_H = \begin{bmatrix} \hat{h}_1 & \hat{h}_2 & \hat{h}_3 & \cdots \\ \hat{h}_2 & \hat{h}_3 & \hat{h}_4 & \cdots \\ \hat{h}_3 & \hat{h}_4 & \hat{h}_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

then $$\Gamma_H = \begin{bmatrix} c \\ cA \end{bmatrix} \begin{bmatrix} b & A \end{bmatrix} \begin{bmatrix} A & b & \cdots \end{bmatrix} = OC$$ has rank $N$ if the realization is minimal.

Note also that, their eigenvalues (or singular values) verify

$$\sigma_k(\Gamma_H) = \sqrt{\lambda_k(\Gamma_H^T \Gamma_H)}$$

$$= \sqrt{\lambda_k(OC^TO)}$$

$$= \sqrt{\lambda_k(OCC^T)}$$

$$= \sqrt{\lambda_k(KW)}$$

**Theorem:** The Hankel form $\Gamma_H$ is of finite rank $N$ if and only if $H(z)$ is an $N$-th order rational function. Then, for $\sigma_1(\Gamma_H) \geq \sigma_2(\Gamma_H) \geq \sigma_3(\Gamma_H) \geq \cdots$, rank $(\Gamma_H) = N$ if and only if $\sigma_N > 0$ and $\sigma_{N+1} = 0$.

then if $K = W = diag(\sigma_1, ..., \sigma_N)$ the realization is **internally balanced**.
Let \( x(n) = \frac{1}{A(z)} u(n) \) an \( N \)-order autoregressive process (i.e., with \( u(n) \) white noise), then

\[
E \{ x(n)x^T(n) \} = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & a_1 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & a_N & 1
\end{bmatrix} [\cdot]^T - \begin{bmatrix}
0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & a_N & 0 \\
a_1 & a_2 & \cdots & a_N & 0
\end{bmatrix} [\cdot]^T
\]

- Orthonormal realizations:

Theorem: Let \( V(z) \) be an stable all-pass transfer function of McMillan degree \( N \), and let \((A, b, g, v_0)\) be a balanced realization of \( V(z) \). Denote \( F_k(z) = z(zI - A)^{-1}bV^k(z) \). Then \( \{e_i^TF_k(z)\}_{i=1,\ldots,N; k=0,\ldots,\infty} \) constitutes an orthonormal basis of the space \( \mathcal{H}_2 \) (stable and causal functions).

Corollary: for every proper stable transfer function \( \hat{H}(z) \in \mathcal{H}_2 \) exist unique \( d \) and \( \nu = \{\nu_k\}_{k=0,1,\ldots} \in l_2 \) such that \( \hat{H}(z) = \nu_0 + z^{-1} \sum_{k=0}^{\infty} \nu_k F_k(z) \). Then \( \nu_0, \nu_k \) are the \textbf{orthogonal expansion coefficients} of \( \hat{H}(z) \).

Particular cases

- Obviously, if \( V(z) = z^{-1} \), it corresponds to \( \nu_k = cA^kb \).
- If \( V(z) = \frac{1-az}{z-a} \), with some real-valued \( a, |a| < 1 \) with \( \alpha = 1-a^2 \), then \( F_k(z) = \alpha z \frac{(1-az)^k}{(z-a)^{k+1}} \), that corresponds to discrete-time \textbf{Laguerre functions}.
- An orthonormal extension with real-valued poles

\[
F_k(z) = \frac{\alpha_k z}{z - a_{k+1}} \prod_{i=1}^{k-1} \left( \frac{1 - a_i z}{z - a_i} \right)
\]

where \( a_k \) is the \( k \)-th pole and \( \alpha_k = 1-a_k^2 \) is a normalization constant.
- If $F_k(z)$ are the Szego polynomials $\overline{D}_k(z)$, that related to the direct for realization, $\overline{H}(z) = \frac{B(z)}{A(z)} = \sum_{k=0}^{N} \nu_k \frac{\overline{D}_k(z)}{\overline{D}_N(z)}$, then the normalized lattice realization can be obtained.

- By the state space description

$$
\begin{bmatrix}
  \mathbf{x}(n+1) \\
  w(n)
\end{bmatrix} = Q
\begin{bmatrix}
  \mathbf{x}(n) \\
  u(n)
\end{bmatrix} \\
\hat{y}(n) = [\nu_0 \ n \ u_1 \ ... \ \nu_N]
\begin{bmatrix}
  \mathbf{x}(n+1) \\
  w(n)
\end{bmatrix}
$$

where $Q = Q_1 Q_2 ... Q_N$ is orthogonal and

$$
Q_k = \begin{bmatrix}
  I_{k-1} & -\sin \theta_k & \cos \theta_k \\
  \cos \theta_k & \sin \theta_k & 0 \\
  0 & 0 & I_{N-k}
\end{bmatrix}
$$

Figure 25: Third order recursive lattice filter.
4.4 Rational approximation theory

4.4.1 Some definitions

We denote $f(z) \in L_2$ to mean

- $\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(e^{iw})|^2 \,dw < \infty$, that admits a Fourier series $f(e^{iw}) = \sum_{k=-\infty}^{\infty} f_k e^{jkw}$ such that $\sum_{k=-\infty}^{\infty} |f_k|^2 < \infty$.

- Also, if $F(z) = \sum_{k=-\infty}^{\infty} f_k z^{-k}$ and $G(z) = \sum_{k=-\infty}^{\infty} g_k z^{-k}$ for $|z| = 1$, then the inner product becomes

$$< F(z), G(z) > = \frac{1}{2\pi j} \int_{|z|=1} F(z^{-1}) G(z) \frac{dz}{z} = \sum_{k=-\infty}^{\infty} f_k g_k$$

We denote $f(z) \in \mathcal{H}_p$ if

- For $1 \leq p < \infty$, $\|f(e^{iw})\| = \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(e^{iw})|^p \,dw \right)^{1/p} < \infty$.
  
  For $p \to \infty$, $\|f(e^{iw})\|_{\infty} = \sup_{w} |f(e^{iw})| < \infty$.

- The Fourier series expansion is one-sided: $f(e^{iw}) = \sum_{k=0}^{\infty} f_k e^{jkw}$. This implies that $f(e^{iw})$ can be analytically continued to all points outside the unit circle by $f(z) = \sum_{k=0}^{\infty} f_k z^{-k}$, $|z| > 1$.

In particular, ”$f(z)$ stable and causal” and ”$f(z) \in \mathcal{H}_2$” may be used interchangeably.

Note: If $f(z) \in L_2$ but not fully in $\mathcal{H}_2$ we can write $f(z) = [f(z)]_- + [f(z)]_+$, the sum of the anti-causal and strict causal part. Also, for $f(z), g(z) \in L_2$, $< [f(z)]_-, [g(z)]_+ >= 0$. 

4.4.2 Decomposition of $\mathcal{H}_2$

The doubly infinite Hankel form $\Gamma_H$, the matrix representation of a rational system, defines two important $\mathcal{H}_2$ subspaces, called **shift-invariant subspaces**, left shift-invariant subspace (lisis) and right shift-invariant subspace (rsis), i.e., given

\[
\begin{bmatrix} g_1 & g_2 & g_3 & \ldots \end{bmatrix} = \begin{bmatrix} f_0 & f_1 & f_2 & \ldots \end{bmatrix} \begin{bmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & h_4 & \cdots \\ h_3 & h_4 & h_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}
\]

a right shift to $f(z)$ gives

\[
\begin{bmatrix} g_2 & g_3 & g_4 & \ldots \end{bmatrix} = \begin{bmatrix} 0 & f_0 & f_1 & \ldots \end{bmatrix} \begin{bmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & h_4 & \cdots \\ h_3 & h_4 & h_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}
\]

that results in a left shift to $g(z)$.

- Since $\Gamma_H$ is not necessarily square it can have a range space and a non empty null space. Since $\Gamma_H$ is symmetric its range and null space are orthogonal.

- If $g(z)$ is in the range space of $\Gamma_H$ so is its left shifted version. Also if $f(z)$ is in the null space of $\Gamma_H$ so must its right shifted version.
Revisiting orthogonal realizations based on shift invariant subspaces

Theorem:

- To every *rsis* is associated a unique all-pass function, \( V(z) \), which causally divides every element of the *rsis*. Thus each *rsis* may be written as \( V(z)\mathcal{H}_2 \) to denote the set of functions \( V(z)f(z) \) as \( f(z) \) varies over \( \mathcal{H}_2 \);
- Since every *lsis* is the orthogonal complement of a *rsis*, each *lsis* may be written \( \mathcal{H}_2 \oplus V(z)\mathcal{H}_2 \), to denote the orthogonal complement to \( V(z)\mathcal{H}_2 \);
- The dimension of the subspace \( \mathcal{H}_2 \oplus V(z)\mathcal{H}_2 \) is the McMillan degree of \( V(z) \).
- If degree \( V(z) = N \), then a set of linearly independent basis functions for the subspace \( \mathcal{H}_2 \oplus V(z)\mathcal{H}_2 \) may be taken as the \( N \) functions of the transfer vector \( C(z) = z(zI - A)^{-1}b \), where the \( N \times N \) matrix \( A \) and the \( N \times 1 \) vector \( b \) are such the pair \((A, b)\) is completely controllable and the eigenvalues of \( A \) coincide with the zeros of \( V(z) \).

In particular, suppose \((A, b, g, v_0)\) is an orthogonal realization, then

Lemma: Let \( C(z) \) the controllability transfer vector, and let \( V(z) = \frac{\det(I-zA)}{\det(zI-A)} \), then

- \( <z^{-k}V(z), z^{-l}V(z)> = \delta_{kl} \),
- \( <C(z), z^{-k}V(z)> = 0 \) for all \( k \geq 0 \).
- \( f(z) \in \mathcal{H}_2 \) satisfies \( <C(z), f(z)> = 0 \) if and only if \( f(z) \) is causally divisible by \( V(z) \), i.e.,

\[
f(z) = V(z)g(z) \quad \text{for some } g(z) \in \mathcal{H}_2
\]
4.4.3 Relation with Hankel form

Problem: Given $h_k$, find the parameters of a rational description (with finite unknown degree).

In terms of the usual transfer function operator, the matrix Hankel form

$$
\begin{bmatrix}
g_1 \\
g_2 \\
g_3 \\
\vdots
\end{bmatrix}
= 
\begin{bmatrix}
h_1 & h_2 & h_3 & \cdots \\
h_2 & h_3 & h_4 & \cdots \\
h_3 & h_4 & h_5 & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\begin{bmatrix}
f_0 \\
f_1 \\
f_2 \\
\vdots
\end{bmatrix}
$$

i.e., $g = \Gamma_H f$, with $g(z)$ strictly casual and $f(z)$ casual, can be rewritten as

$$
g(z) = [H(z)f(z^{-1})]_+
$$

where $[\cdot]_+$ is the strictly causal projection operator.

Let $\Gamma_H$ of finite rank $N$, then

Theorem: Let $\mathcal{N}(\Gamma_H) = \{f : \Gamma_H f = 0\}$ denote the set of vectors $f = [f_0, f_1, f_2, \ldots]^T$ lying in the null space of $\Gamma_H$, or equivalently

$$
\mathcal{N}(\Gamma_H) = \{f(z) : [H(z)f(z^{-1})]_+ = 0\}
$$

Then

- exist a causal all-pass $V(z)$, determined by $H(z)$, such that $f(z) = \sum_{k=0}^{\infty} f_k z^{-k} = V(z)R(z)$ for some $R(z) \in \mathcal{H}_2$.
- Since $H(z) = \frac{B(z)}{A(z)}$ then $V(z) = \frac{z^{-N}A(z^{-1})}{A(z)}$. 
Particular interesting cases can be obtained by chosen $R(z)$ as follows

1. $R(z) = A(z)$, then $f(z) = z^{-N}A(z^{-1})$ a finite length sequence (equation error methods).

   - Pade approximant: $\Gamma_H f = \begin{bmatrix} 0_N \\ \times \end{bmatrix}$.

   - Equation error: $\Gamma_H f = \begin{bmatrix} 0_N \\ 0_N \end{bmatrix}$.

2. $R(z) = 1$, then $f(z) = V(z)$ a unit norm function (output error methods).

   - Output error: Consider the minimization of $\|H(z) - \hat{H}(z)\|^2$ using orthogonal representation for $H(z) = \sum_{k=0}^{\infty} h_k F_k(z)$ ($\hat{h}_k = \langle H(z), F_k(z) \rangle$) and $\hat{H}(z) = \sum_{k=0}^{N} \nu_k F_k(z)$, then the optimal choice of $\nu_k = \hat{h}_k = \langle H(z), F_k(z) \rangle$ leads to the remaining error

   $$H(z) - \hat{H}(z) = \hat{h}_{N+1} F_{N+1}(z) + \hat{h}_{N+2} F_{N+2}(z) + \hat{h}_{N+3} F_{N+3}(z) + \ldots$$

   $$= V(z) \sum_{k=1}^{\infty} \hat{h}_{N+k} z^{-k}$$

   where $F_{N+k}(z) = z^{-k} V(z)$ was used.

   Using the expansions $H(z) = \sum_{k=0}^{\infty} h_k z^{-k}$ and $V(z) = \sum_{k=0}^{\infty} \nu_k z^{-k}$, we can express

   $$\hat{h}_{N+k} = \langle H(z), z^{-k} V(z) \rangle = h_k v_0 + h_{k+1} v_1 + h_{k+2} v_2 + \ldots$$

   $$\begin{bmatrix} \hat{h}_{N+1} \\ \hat{h}_{N+2} \\ \hat{h}_{N+3} \\ \vdots \end{bmatrix} = \Gamma_H \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \end{bmatrix}$$

   or $\sum_{k=1}^{\infty} \hat{h}_{N+k} z^{-k} = [H(z) V(z^{-1})]_+$. Then the output error is

   $$\|H(z) - \hat{H}(z)\|^2 = \|[H(z) V(z^{-1})]_+\|^2$$

   In the general case $\text{deg} H(z) < N$ the best we can do is to force $V(z)$ to lie in an approximate null space of $\Gamma_H$ in the sense that $\|\Gamma_H v\|$ is minimized.
4.4.4 Hankel norm rational approximation

Let $\Gamma_H$ be approximated by $\tilde{\Gamma}_H$, then the approximation problem has a closed form solution with

$$\min_{\text{rank } \tilde{\Gamma}_H \leq N} \|\Gamma_H - \tilde{\Gamma}_H\| = \sigma_{N+1}(\Gamma_H)$$

A physical interpretation of $\|\Gamma_H\| = \sigma_1$ with rank equal $N$, for a balanced realization.

- Considering $\|\Gamma_H\| = \max_{\|u\|=1} \|\Gamma_H u\|$, $u = [u(0)u(-1)u(-2)\cdots]^T$, then $y = [y(1) y(2) y(3) \cdots]^T = \Gamma_H u = OCu$.

- It is not hard to see that $Cu = x(1)$, where $x(1)$ is the state vector subject to an initial condition having been produced by a unit norm vector.

- Using the singular value decomposition of $\Gamma_H$

$$\Gamma_H = [\eta_1 \eta_2 \cdots \eta_N] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_N \end{bmatrix} \begin{bmatrix} \xi_1^T \\ \xi_2^T \\ \vdots \\ \xi_N^T \end{bmatrix}$$

then $y = [\eta_1 \eta_2 \cdots \eta_N] \begin{bmatrix} \sigma_1^{1/2} & & & \\ & \sigma_2^{1/2} & & \\ & & \ddots & \\ & & & \sigma_N^{1/2} \end{bmatrix} \begin{bmatrix} x_1(1) \\ x_2(1) \\ \vdots \\ x_N(1) \end{bmatrix}$

$= \sum_{k=1}^N \eta_k \sigma_k^{1/2} x_k(1)$ or, since $\{\eta_k\}$ are orthonormal $\|y\|^2 = \sum_{k=1}^N \sigma_k [x_k(1)]^2$

Another important property

$$\|H(z) - \hat{H}(z)\|^2 \leq \sigma_1(\Gamma_H - \tilde{\Gamma}_H) \leq \sup_{|z|=1} |H(z) - \hat{H}(z)|$$

an upper bound for the approximation in $\mathcal{H}_2$. 
Indeed, if we consider the Frobenius norm, i.e., that is defined for a matrix

\[
P = \begin{bmatrix}
  p_{1,1} & p_{1,2} & p_{1,3} & \cdots \\
  p_{2,1} & p_{2,2} & p_{2,3} & \cdots \\
  p_{3,1} & p_{3,2} & p_{3,3} & \cdots \\
  \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]

as \( \|P\|_F = \left( \sum_{k,l=1}^{\infty} p_{k,l}^2 \right)^{1/2} \).

Lemma: Let \( D = \text{diag}[d_0, d_1, d_2, \ldots] \), where \( d_k = d_{k-1}\sqrt{\frac{2k-1}{2k}}, d_0 = 1 \). Then for any Hankel form \( \Gamma_H \),

\[
\|D\Gamma_H D\|_F = \left( \sum_{k=1}^{\infty} h_k^2 \right)^{1/2} = \|[H(z)]_+\|^2
\]

That leads to a priori lower bound

\[
\min_{\deg H(z)=N} \|H(z) - \hat{H}(z)\|^2 = \min_{\text{rank } \Gamma_H = N} \|D(\Gamma_H - \hat{\Gamma}_H)D\|_F \\
\geq \sum_{k=N+1}^{\infty} \sigma_k^2(D\Gamma_H D)
\]
4.5 Stability theory concepts

4.5.1 Stability of quasi-invariant systems

Convergence in the mean of an IIR adaptive algorithm can be studied by a related difference equation.

If the average behavior of the algorithm can be written as

\[ E\{\hat{\theta}(n+1)\} = [R_1 + R_2(n)]E\{\hat{\theta}(n)\} + R_3(n) \quad (44) \]

where \( E\{\theta(n) - \theta_o\} = E\{\hat{\theta}(n)\} \) with \( \theta_o \) defining the ideal parameters, and \( R_1 \) is positive definite and \( R_2(n) \) has norm sufficiently small, then this system is called quasi-invariant.

Theorem: Let the quasi-invariant system defined by equation (44), i.e.,

- \( R_1 \) satisfy \( \|R_1^n\| < c\beta^n \) with \( c \) and \( \beta \) are constants such that \( c > 0 \) and \( 0 < \beta < 1 \),
- \( R_2(n) \) has a norm sufficiently low, i.e., \( \|R_2(n)\| \leq \kappa_2 \), for \( \kappa_2 \) a positive constant.
- \( R_3(n) \) has bounded norm, i.e., \( \|R_3(n)\| \leq \kappa_3 \), for \( \kappa_3 \) a positive constant.

Then if \( 0 < (\beta + c\kappa_2) < 1 \) the system of equation (44) is asymptotically stable.

Corollary: Note that if \( \|R_3(n)\| \) tends to zero for \( n \to \infty \), then the system of equation (44) converge asymptotically to the origin, i.e.,

\[ E\{\theta(n+1)\} \to \theta_o \quad (45) \]

for \( n \to \infty \).
4.5.2 Stability of a non linear feedback system

Consider $\mathcal{F}(z)$ a rational transfer function and the feedback law related to the figure of the form (Popov inequality)

$$\sum_{n=0}^{N} s(n)\epsilon(n) \leq \gamma^2$$

Theorem: The closed-loop system of the figure is asymptotically stable (i.e., $s(n)$ and $\epsilon(n)$ remain bounded and tend to zero) for all feedback laws as the specified, and for all initial conditions, if and only if $\mathcal{F}(z)$ is strictly positive real, i.e., a stable and causal function such that: $\text{Re} \mathcal{F}(e^{jw}) \geq c \geq 0$ for all $w$.

Properties of positive real functions

- If $\text{Re} \mathcal{F}(e^{jw}) \geq c > 0$ for all $w$, then $\text{Re} \mathcal{F}(z) \geq c > 0$ for all $|z| \geq 1$.

- If $\mathcal{F}(z)$ is strictly positive real (SPR), then it can have no zeros in $|z| \geq 1$, i.e., if SPR then minimum phase (the converse is not true).

- If $\mathcal{F}(z)$ is SPR, so is its inverse $1/\mathcal{F}(z)$.

- Suppose $\mathcal{F}(z)$ SPR, and let $\epsilon(n) = \mathcal{F}(z)s(n)$. Then for all non zero square summable $\{s(n)\}$: $\sum_{k=-\infty}^{\infty} s(k)\epsilon(k) > 0$. 
Proof of the Hyperstability theorem

Consider \( u(n) \) and \( y(n) \) such that

\[
\begin{align*}
u(n) &= s(n) + \epsilon(n) = [F(z) + 1]s(n) \\
y(n) &= s(n) - \epsilon(n) = [F(z) - 1]s(n)
\end{align*}
\]

that form \( y(n) = \frac{F(z)-1}{F(z)+1}u(n) = \mathcal{G}(z)u(n) \), also a rational function with realization \((A, b, c, d)\). If the bounded sequences \( u(n) \) and \( y(n) \) tend asymptotically to zero the same apply to \( s(n) \) and \( \epsilon(n) \).

From the properties of SPR functions, if \( F(z) \) is SPR, then \( |G(z)| \leq c < 1 \) for all \( |z| \geq 1 \). Based on an bounded initial condition of the state vector of the system realizing \( G(z) \), \( x(0) \), that can be written as

\[
x(0) = \begin{bmatrix} b & Ab & A^2b & \ldots \end{bmatrix} \begin{bmatrix} u(-1) \\
u(-2) \\
u(-3) \\
\vdots
\end{bmatrix}
\]

and that for \( N \) time instant we can write

\[
\begin{bmatrix}
y(0) \\
y(1) \\
\vdots \\
y(N)
\end{bmatrix} = \begin{bmatrix} c^T \\
c^TA \\
\vdots \\
c^TA^N
\end{bmatrix} \begin{bmatrix} x(0) \\
\vdots
\end{bmatrix} + \begin{bmatrix} d & 0 & \cdots & 0 \\
d & c^Tb & d & \cdots & \vdots \\
\vdots & \ddots & \ddots & \ddots & 0 \\
c^Tb & \cdots & c^TA^{N-1}b & \cdots & c^Tb & d
\end{bmatrix} \begin{bmatrix} u(0) \\
u(1) \\
\vdots \\
u(N)
\end{bmatrix}
\]

Then is not hard to shown that (extending the use of Parseval theorem to a bounded initial condition)

\[
\sum_{n=0}^{N} y^2(n) \leq c^2 \sum_{n=0}^{N} u^2(n) + f[x(0)] \tag{46}
\]

where \( f[x(0)] \) is a bounded function of the initial condition of the state vector of the system realizing \( G(z) \).
On the other hand, substituting \( s(n) \) and \( \epsilon(n) \) in the Popov inequality we obtain
\[
\sum_{n=0}^{N} u^2(n) \leq \sum_{n=0}^{N} y^2(n) + 4\gamma^2
\] (47)

By using this in (46)
\[
\sum_{n=0}^{N} y^2(n) \leq c^2 \sum_{n=0}^{N} y^2(n) + 4c^2\gamma^2 + f[x(0)]
\]
or
\[
\sum_{n=0}^{N} y^2(n) \leq \frac{4c^2\gamma^2 + f[x(0)]}{1 - c^2} < \infty
\]

this implies that \( y^2(n) \to 0 \) for \( n \to \infty \) and by (47) the same can be inferred for \( u^2(n) \).
Passive Impedance functions

Using \( p = \frac{z-1}{z+1} \) and \( \mathcal{F}(p) \) (a continuous time transfer function) an **impedance**. If \( S(p) \) is the Laplace transform of a casual \( s(t) \) electrical current, then \( e(t) \) is the resulting voltage.

If \( \mathcal{F}(p) \) is SPR,

\[
\int_0^\infty s(t)e(t)dt = \int_{-\infty}^{\infty} \mathcal{F}(j\Omega)S(j\Omega)S^*(j\Omega)d\Omega
= \int_{-\infty}^{\infty} \mathcal{F}(j\Omega)|S(j\Omega)|^2d\Omega > 0
\]

then the impedance is said to be **passive**.

Spectral factorization

Since for an stationary stochastic process with correlation \( \{r_k\} \), with \( r_k = r_{-k} \), \( S(z) = \sum_{k=-\infty}^{\infty} r_k z^{-k} \) is nonnegative along \( |z| = 1 \), then by chosen \( \mathcal{F}(z) = r_0/2 + r_1 z^{-1} + r_2 z^{-2} + \ldots \), is easy to see that \( \mathcal{F}(z^{-1}) + \mathcal{F}(z) = S(z) \). Or, \( \mathcal{F}(z) \) is SPR if and only if it is the (unilateral) \( z \)-transform of a correlation sequence \( \{r_k\} \).

Also, if \( S(z) \) has positive geometric mean, i.e.,

\[
\exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log[S(e^{j\omega})]\right) > 0
\]

then it admits a **spectral factorization**: \( S(z) = F(z)F(z^{-1}) \), for some stable and causal \( F(z) \). The stochastic process which furnishes the correlation \( r_k \) could be modelled as the output of \( F(z) \) driven by unit-variance white noise.

Positive real lemma: A rational function \( \mathcal{F}(z) = d + c(zI - A)^{-1}b \) is positive real if and only if there exists a symmetric, positive definite \( P \) for which the symmetric matrix

\[
\begin{bmatrix}
P - A^T PA & c - A^T Pb \\
ct - b^T PA & 2d - cPC
\end{bmatrix}
= \begin{bmatrix}
L^T \\
N^T
\end{bmatrix}
\begin{bmatrix}
L \\
N
\end{bmatrix}
\]

is positive definite.

Then

\[
F(z) = N + L(zI - A)^{-1}b
\]
5 MSOE minimization

MSOE minimization and related algorithms

- Stationary points (existence of local minima),
- ODE (convergence to local minima and instability).
- Direct-form realization of an adaptive IIR filter: implementation of the derivatives, simplifications.
- Lattice realization: simplifications.
- Other realizations