Analysis of Discrete-Time Systems

Overview

- Stability
- Sensitivity and Robustness
- Controllability, Reachability, Observability, and Detectabiliy

Stability

Definitions

- We define stability first with respect to changes in the initial conditions.
- Consider

$$\boldsymbol{x}[k+1] = \boldsymbol{f}(\boldsymbol{x}[k], k).$$

- Let $\boldsymbol{x}^0[k]$ and $\boldsymbol{x}[k]$ be solutions when the initial conditions are $\boldsymbol{x}^0[k_0]$ and $\boldsymbol{x}[k_0]$, respectively.

Definition – Stability: The solution $\boldsymbol{x}^0[k]$ is stable if for a given ε, there exists a $\delta(\varepsilon, k_0) > 0$ such that all solutions with $||\boldsymbol{x}[k_0] - \boldsymbol{x}^0[k_0]|| < \delta$ are such that $||\boldsymbol{x}[k] - \boldsymbol{x}_0[k]|| < \varepsilon$ for all $k > k_0$.

Definition – Asymptotic Stability: The solution $\boldsymbol{x}^0[k]$ is asymptotically stable if it is stable and if δ can be chosen such that $||\boldsymbol{x}[k_0] - \boldsymbol{x}^0[k_0]|| < \delta$ implies that $||\boldsymbol{x}[k] - \boldsymbol{x}_0[k]|| \to 0$ when $k \to \infty$.

- Stability, in general, is a local concept.
- System is (asymptotically) stable if the trajectories do not change much if the initial condition is changed by a small amount.

Stability of Linear Discrete Time Systems

System

$$\boldsymbol{x}^0[k+1] = \boldsymbol{\Phi}\boldsymbol{x}^0[k] \qquad \boldsymbol{x}^0[0] = \boldsymbol{a}^0$$

System with perturbed initial value

$$\boldsymbol{x}[k+1] = \boldsymbol{\Phi}\boldsymbol{x}[k] \qquad \boldsymbol{x}[0] = \boldsymbol{a}$$

The difference $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{x}^0$ satisfies the equation

$$\tilde{\boldsymbol{x}}[k+1] = \boldsymbol{\Phi}\tilde{\boldsymbol{x}}[k] \qquad \tilde{\boldsymbol{x}}[0] = \boldsymbol{a} - \boldsymbol{a}^0$$

- $\rightsquigarrow$ If the solution $\boldsymbol{x}^0[k]$ is stable every other solution is also stable
- For linear, time-invariant systems, stability is a property of the system and not of a special trajectory!

Solution for the last system:

$$\tilde{\boldsymbol{x}}[k] = \boldsymbol{\Phi}^k\tilde{\boldsymbol{x}}[0]$$

- If it is possible to diagonalize $\boldsymbol{\Phi}$ then the solution is a combination of λ_i^k terms, where $\lambda_i^k, i = 1, \ldots, n$ are the eigenvalues of $\boldsymbol{\Phi}$.
- If it is not possible to diagonalize $\boldsymbol{\Phi}$ then the solution is a linear combination of the terms $p_i(k)\lambda_i^k$ where $p_i(k)$ are polynomials in k of the order one less the multiplicity of the corresponding eigenvalue.
- To get asymptotic stability, all solution must go to zero as k increases to infinity. $\rightsquigarrow$ Eigenvalues must have to property

$$|\lambda_i| < 1 \qquad i = 1, \ldots, n$$

Theorem – Asymptotic Stability of Linear Systems: A discrete-time linear time-invariant system is asymptotically stable if and only if all eigenvalues of $\boldsymbol{\Phi}$ are strictly inside the unit disk.

Input-Output Stability

Definition – Bounded-Input Bounded-Output Stability: A linear time-invariant system is defined bounded-input bounded-output (BIBO) stable if a bounded input gives a bounded output for every initial value.

Theorem – Relation between Stability Concepts: Asymptotic stability implies stability and BIBO stability.

- Stability does not imply BIBO stability, and vice versa!

Stability Tests

- Numerical computation of the eigenvalues of $\boldsymbol{\Phi}$ or the roots of the characteristic equation $\det(z\boldsymbol{I} - \boldsymbol{\Phi}) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n = 0$ (Scilab commands `spec` and `roots`)

Algebraic or graphical methods (help to understand how parameters of the system or controller will influence the stability):

- Direct algebraic computation of the eigenvalues
- **Methods based on the properties of characteristic polynomials**
- Root locus method (used to determine closed-loop stability for known open-loop system)
- **The Nyquist criterion** (used to determine closed-loop stability for known open-loop system)
- Lyapunov's method

Jury's Stability Criterion

- Characteristic polynomial

$$\mathsf{A}(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n = 0 \qquad (1)$$

See blackboard:

Theorem – Jury's Stability Test: If $a_0 > 0$, then Eq. (1) has all roots inside the unit disk if and only if all $a_0^k, k = 0, 1, \ldots, n-1$ are positive. If no a_0^k is zero, then the number of negative a_0^k is equal to the number of roots outside the unit disk.

Remark: If all a_0^k are positive for $k > 0$, then the condition a_0^0 can be shown to be equivalent to the conditions

$$\begin{aligned} \mathsf{A}(1) &> 0 \\ (-1)^n \mathsf{A}(-1) &> 0 \end{aligned}$$

These are necessary conditions for stability and hence can be used before forming the table above.

Example for Jury's Stability Criterion: See blackboard ...

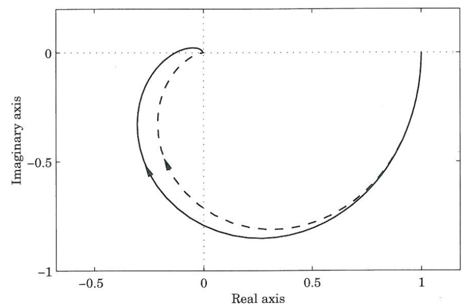
Nyquist and Bode Diagrams for Discrete-Time Systems

- Continuous-time system $G(s)$: The *Nyquist curve* or *frequency response* of the system is the map $G(j\omega)$ for $\omega \in [0, \infty)$.
- This curve is drawn in polar coordinates (Nyquist diagram) or as amplitude and phase curves as a function of frequency (Bode diagram)
- $G(j\omega)$ interpretation: Stationary amplitude and phase of system output when a sinusoidal input signal with frequency ω is applied.
- Discrete-time system with pulse-transfer function $H(z)$: The Nyquist curve or frequency curves is given by the map $H(\mathrm{e}^{j\omega\Delta})$ for $\omega\Delta \in [0, \pi]$ (up to the Nyquist frequency).
- $H(\mathrm{e}^{j\omega\Delta})$ is periodic with period $2\pi/\Delta$. Higher harmonics are created.

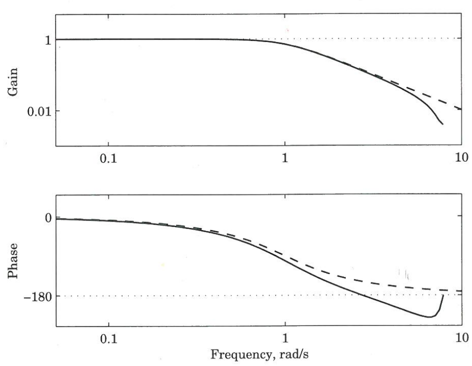
Example:

$$G(s) = \frac{1}{s^2 + 1.4s + 1} \to \text{ZOH \& } \Delta = 0.4\text{s} \ \to H(z) = \frac{0.066z + 0.055}{z^2 - 1.450z + 0.571}$$

Nyquist plot: $G(j\omega)$ (dashed), $H(\mathrm{e}^{j\omega\Delta})$ (solid)

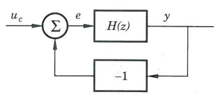


Bode plot: $G(j\omega)$ (dashed), $H(\mathrm{e}^{j\omega\Delta})$ (solid)



Nyquist Criterion

- Well-known stability test for continuous-time systems.
- To determine the stability of the closed-loop system when the open-loop system is given.
- Can be reformulated to handle discrete-time systems.
- Consider the discrete-time system:

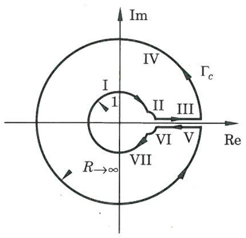


$$H_{cl}(z) = \frac{Y(z)}{U_c(z)} = \frac{H(z)}{1 + H(z)}$$

with the characteristic polynomial

$$1 + H(z) = 0$$

Nyquist contour Γ_c that encircles the area outside the unit disc (instability area):



- Indentation at $z=1$: to exclude the integrators in the open-loop system $\rightsquigarrow$ infinitesimal semicircle with decresing arguments from $\pi/2$ to $-\pi/2$ is mapped into the $H(z)$-plane as an infinitely large circle from $-n\pi/2$ to $n\pi/2$ (n - number of integrators in the open-loop system)
- The map of the unit circle is $H(\mathrm{e}^{j\omega\Delta})$ for $\omega\Delta \in (0, 2\pi)$

- Stability is determined by analysing how Γ_c is mapped by $H(z)$.

  The number of encirclements N in the positive direction around (-1,0) by the map Γ_c is equal to

  $$N = Z - P$$

  where Z and P are the numbers of zeros and poles, respectively, of $1 + H(z)$ outside the unit disk.
- Zeros (Z) of $1 + H$ are unstable closed-loop poles
- Poles (P) of $1 + H$ are unstable open-loop poles
- If the open-loop system is stable ($P = 0$) then $N = Z$. $\rightsquigarrow$ The stability of the system is then ensured if the map of Γ_c does not encircle the point (-1,0).
- If $H(z) \to 0$ when $z \to \infty$, the parallel lines III and V do not influence the stability test, and it is sufficient to find the map of the unit circle and the small semicircle at $z = 1$.

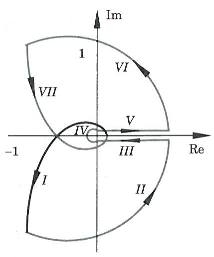
Simplified Nyquist Criterium: If the open-loop system and its inverse are stable then the stability of the closed-loop system is ensured if the point (-1,0) in the $H(z)$-plane is to the left of the map of $H(\mathrm{e}^{j\omega\Delta})$ for $\omega\Delta = 0$ to π – that is, to the left of the Nyquist curve.

Example:

- Consider the system

  $$H(z) = \frac{0.25K}{(z-1)(z-0.5)}$$

  with $\Delta = 1$.
- The black line on the right graphic shows $H(\mathrm{e}^{j\omega\Delta})$ for $\omega\Delta = 0$ to π, the Nyquist curve.



Relative Stability

- Amplitude and phase margins can be defined for discrete-time systems analogously to continuous-time systems.

Definition – Amplitude Margin: Let the open-loop system have the pulse-transfer function $H(z)$ and let ω_0 the smallest frequency such that

$$\arg H(\mathrm{e}^{j\omega_0\Delta}) = -\pi.$$

The *amplitude* or *gain margin* is then defined as

$$A_{\text{marg}} = \frac{1}{H(\mathrm{e}^{j\omega_0\Delta})}$$

- The amplitude margin is how much the gain can be increased before the closed-loop system becomes unstable.

Definition – Phase Margin: Let the open-loop system have the pulse-transfer function $H(z)$ and further let the crossover frequency ω_c be the smallest frequency such that

$$|H(\mathrm{e}^{j\omega_c\Delta})| = 1.$$

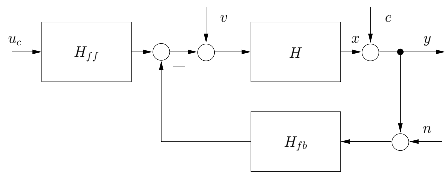
The *phase margin* is then defined as

$$\phi_{\text{marg}} = \pi + \arg H(\mathrm{e}^{j\omega_c\Delta})$$

- The phase margin is how much extra phase lag is allowed before the closed-loop system becomes unstable.

Sensitivity and Complementary Sensitivity Function

Consider the closed-loop system with the feedforward filter H_{ff} and the feedback controller H_{fb}:



Pulse-transfer operator from the inputs to y:

$$y = \frac{H_{ff}H}{1+\mathcal{L}}u_c + \frac{H}{1+\mathcal{L}}v + \frac{1}{1+\mathcal{L}}e - \frac{\mathcal{L}}{1+\mathcal{L}}n$$

Open-loop transfer function:

$$\mathcal{L} = H_{fb}H$$

Closed-loop transfer function:

$$H_{cl} = \frac{H_{ff}H}{1+\mathcal{L}}$$

Sensitivity of H_{cl} with respect to variations in H is given by

$$\frac{\mathrm{d}H_{cl}}{\mathrm{d}H} = \frac{H_{ff}}{(1+\mathcal{L})^2}$$

The relative sensitivity of H_{cl} with respect to H thus can be written as

$$\frac{\mathrm{d}H_{cl}}{H_{cl}} = \frac{1}{1+\mathcal{L}}\frac{\mathrm{d}H}{H} = \mathcal{S}\frac{\mathrm{d}H}{H}$$

Sensitivity function $\mathcal{S}$:

$$\mathcal{S} = \frac{1}{1+\mathcal{L}}$$

- Pulse-transfer function from e to y, should be small at low frequencies

Complementary sensitivity function $\mathcal{T}$:

$$\mathcal{T} = 1 - \mathcal{S} = \frac{\mathcal{L}}{1+\mathcal{L}}$$

- negative Pulse-transfer function from n to y (or e to x), should be small at higher frequencies
- Remark: Due to $\mathcal{T} + \mathcal{S} = 1$, we can not make $\mathcal{T}$ and $\mathcal{S}$ for all frequencies and must make a trade-off as outlined above!

Robustness

- The distance r from the Nyquist curve $\mathcal{L}(e^{-j\omega\Delta})$ to the critical point (-1,0) is a good measure for the robustness and is given by $r = |1+\mathcal{L}|$.
- Note that the complex number $1+\mathcal{L}(e^{-j\omega\Delta})$ can be represented as the vector from the point -1 to the point $L(e^{-j\omega\Delta})$ on the Nyquist curve.
- Reciprocal of the smallest distance r_{min} from the Nyquist curve $\mathcal{L}(e^{-j\omega\Delta})$ to the critical point:

$$\frac{1}{r_{min}} = \max\left(\frac{1}{|1+\mathcal{L}(e^{-j\omega\Delta})|}\right) = \max|\mathcal{S}(e^{-j\omega\Delta})|. \qquad (2)$$

- Guaranteed bounds on the margins:

$$\begin{aligned} A_{\text{marg}} &\geq 1/(1-r_{min}), && (3)\\ \phi_{\text{marg}} &\geq 2\arcsin(r_{min}/2). && (4) \end{aligned}$$

- By requiring $\max|S(e^{-j\omega\Delta})| < 2$, the system will have at least the robustness margins $A_{\text{marg}} \geq 2$ and $\phi_{\text{marg}} \geq 29°$.

Controllability, Reachability

- Question: Is it possible to steer a system from a given initial state to any other state?
- Consider the system

$$\begin{aligned} \boldsymbol{x}[k+1] &= \boldsymbol{\Phi}\boldsymbol{x}[k] + \boldsymbol{\Gamma}\boldsymbol{u}[k] && (5)\\ y[k] &= \boldsymbol{C}\boldsymbol{x}[k] && (6) \end{aligned}$$

- Assume that the initial state $\boldsymbol{x}[0]$ is given. At state the sample instant n, where n is the order of the system, is given by

$$\begin{aligned} \boldsymbol{x}[n] &= \boldsymbol{\Phi}^n\boldsymbol{x}[0] + \boldsymbol{\Phi}^{n-1}\boldsymbol{\Gamma}\boldsymbol{u}[0] + \cdots + \boldsymbol{\Gamma}\boldsymbol{u}[n-1]\\ &= \boldsymbol{\Phi}^n\boldsymbol{x}[0] + \boldsymbol{W}_c\boldsymbol{U} \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{W}_c &= \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Phi}\boldsymbol{\Gamma} & \cdots & \boldsymbol{\Phi}^{n-1}\boldsymbol{\Gamma} \end{pmatrix}\\ \boldsymbol{U} &= \begin{pmatrix} \boldsymbol{u}^T[n-1] & \cdots & \boldsymbol{u}^T[0] \end{pmatrix} \end{aligned}$$

- The matrix $\boldsymbol{W}_c$ is referred to as the *controllability matrix*.
- If $\boldsymbol{W}_c$ has rank, then it is possible to find n equations from which the control signal can be found such that the initial state is transferred to the desired state $\boldsymbol{x}[n]$.
- The solution is not unique if there is more than one input!

Definition – Controllability: The system (5) is controllable if it is possible to find a control sequence such that **the origin** can be reached from any initial state in finite time.

Definition – Reachability: The system (5) is reachable if it is possible to find a control sequence such that **an arbitrary state** can be reached from any initial state in finite time.

- Controllability does not imply reachability: If $\boldsymbol{\Phi}^n\boldsymbol{x}[0] = \boldsymbol{0}$, then the origin will be reached with zero input, but the system is not necessarily reachable. See example ...

Theorem – Reachability: The system (5) is reachable if and only if the matrix $\boldsymbol{W}_c$ has rand n.

The reachable states belong to the linear subspace spanned by the columns of $\boldsymbol{W}_c$. Example ...

- Reachability is independent of the coordinates:

$$\begin{aligned} \tilde{\boldsymbol{W}}_c &= \begin{pmatrix} \tilde{\boldsymbol{\Gamma}} & \tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{\Gamma}} & \cdots & \tilde{\boldsymbol{\Phi}}^{n-1}\tilde{\boldsymbol{\Gamma}} \end{pmatrix}\\ &= \begin{pmatrix} \boldsymbol{T}\boldsymbol{\Gamma} & \boldsymbol{T}\boldsymbol{\Phi}\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{\Gamma} & \cdots & \boldsymbol{T}\boldsymbol{\Phi}^{-n-1}\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{\Gamma} \end{pmatrix}\\ &= \boldsymbol{T}\boldsymbol{W}_c \end{aligned}$$

- If $\boldsymbol{W}_c$ has rank n, than $\tilde{\boldsymbol{W}}_c$ has also rank n for a non-singular transformation matrix $\boldsymbol{T}$.

- Assume a SISO system, a non-singular $\boldsymbol{W}_c$ and

$$\det(\lambda \boldsymbol{I} - \boldsymbol{\Phi}) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n = 0 \tag{7}$$

Then there exists a transformation on *controllable canonical form*

$$\begin{aligned}
\boldsymbol{z}[k+1] &= \begin{pmatrix} -a_1 & -a_2 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \boldsymbol{z}[k] + \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} u[k] \\
y[k] &= \begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix} \boldsymbol{z}[k]
\end{aligned}$$

  - Good for computation of I/O model and for the design of a state-feedback-control law.

State Trajectory Following

- Question: Does reachability also imply that is possible to follow a given trajectory in the state space?
- In order to drive a system from $\boldsymbol{x}[k]$ to a desired $\boldsymbol{x}[k+1]$ the matrix $\boldsymbol{\Gamma}$ must have rank n.
- For a reachable SISO system it is, in general possible, to reach desired states only at each n-th sample instant, provided that the desired points are known n steps ahead.

Output Trajectory Following

- Assume that a desired trajectory $u_c[k]$ is given, the control signal should satisfy

$$y[k] = \frac{\mathsf{B}(q)}{\mathsf{A}(q)} u[k] = u_c[k]$$

or

$$u[k] = \frac{\mathsf{A}(q)}{\mathsf{B}(q)} u_c[k]$$

- For a time-delay of d steps the generation of $u[k]$ is only causal if the trajectory is known d steps ahead.
- The signal $u[k]$ is bounded if $u_c[k]$ is bounded and if the system has a stable inverse.

Observability and Detectability

- To solve the problem of finding the state of a system from observations of the output, the concept of unobservable states is introduced at first:

Definition - Unobservable States: The state $\boldsymbol{x}^0 \neq \boldsymbol{0}$ is unobservable if there exists a finite $k_1 \geq n-1$ such that $y[k] = 0$ for $0 \leq k \leq k_1$ when $\boldsymbol{x}(0) = \boldsymbol{x}^o$ and $u[k] = 0$ for $0 \leq k \leq k_1$.

- The system (5) is observable if there is a finite k such that knowledge of the inputs $\boldsymbol{u}[0], \cdots, \boldsymbol{u}[k-1]$ and the outputs $\boldsymbol{y}[0], \cdots, \boldsymbol{y}[k-1]$ is sufficient to determine the initial state of the system.
- Effect of $\boldsymbol{u}[k]$ always can be determined. Without loss of generality we assume $\boldsymbol{u}[k] = \boldsymbol{0}$

$$\begin{aligned}
y[0] &= \boldsymbol{C}\boldsymbol{x}[0] \\
y[1] &= \boldsymbol{C}\boldsymbol{x}[1] = \boldsymbol{C}\boldsymbol{\Phi}\boldsymbol{x}(0) \\
&\vdots \\
y[n-1] &= \boldsymbol{C}\boldsymbol{\Phi}^{n-1}\boldsymbol{x}[0]
\end{aligned}$$

Using vector notations:

$$\begin{pmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{\Phi} \\ \vdots \\ \boldsymbol{C}\boldsymbol{\Phi}^{n-1} \end{pmatrix} \boldsymbol{x}[0] = \begin{pmatrix} y[0] \\ y[1] \\ \vdots \\ y[n-1] \end{pmatrix}$$

The initial state $\boldsymbol{x}[0]$ can be obtained if and only if the observability matrix

$$\boldsymbol{W}_o = \begin{pmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{\Phi} \\ \vdots \\ \boldsymbol{C}\boldsymbol{\Phi}^{n-1} \end{pmatrix}$$

has rank n.

Theorem - Observability The system (5) is observable if and only if $\boldsymbol{W}_0$ has rank n.

Theorem - Detectability A system is detectable if the only unobservable states are such that they decay to the origin. That is, the corresponding eigenvalues are stable.

Observable Canonical Form

Assume a SISO system, a non-singular matrix $\boldsymbol{W}_0$ and

$$\det(\lambda \boldsymbol{I} - \boldsymbol{\Phi}) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n = 0 \tag{8}$$

then there exists a transformation such that the transformed system is

$$\begin{aligned}
\boldsymbol{z}[k+1] &= \begin{pmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_{n-1} & 0 & 0 & \cdots & 1 \\ -a_n & 0 & 0 & \cdots & 0 \end{pmatrix} \boldsymbol{z}[k] + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{pmatrix} u[k] \\
y[k] &= \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \boldsymbol{z}[k]
\end{aligned}$$

which is called *observable canonical form.*

- This form has the advantage that it is easy to find the I/O model and to determine a suitable observer.
- The transformation is given by

$$T = \tilde{W}_o^{-1} W_o$$

where $\tilde{W}_o$ is the observable matrix of the system in *observable canonical form*.

Kalman's Decomposition

- The reachable and unobservable parts of a system are two linear subspaces of the state space.
- Kalman showed that it is possible to introduce coordinates such that a system can be partitioned in the following way:

$$\begin{aligned} x[k+1] &= \begin{pmatrix} \Phi_{11} & \Phi_{12} & 0 & 0 \\ 0 & \Phi_{22} & 0 & 0 \\ \Phi_{31} & \Phi_{32} & \Phi_{33} & \Phi_{34} \\ 0 & \Phi_{42} & 0 & \Phi_{44} \end{pmatrix} x[k] + \begin{pmatrix} \Gamma_1 \\ 0 \\ \Gamma_3 \\ 0 \end{pmatrix} u[k] \\ y[k] &= \begin{pmatrix} C_1 & C_2 & 0 & 0 \end{pmatrix} x[k] \end{aligned}$$

The state space is partitioned into four parts yielding for subsystems:

- S_{or} Observable and reachable
- $S_{o\bar{r}}$ Observable but not reachable
- $S_{\bar{o}r}$ Not observable but reachable
- $S_{\overline{or}}$ Neither observable nor reachable

The pulse-transfer operator for the observable and reachable subsystem is given by:

$$H(q) = C_1(qI - \Phi_{11})^{-1}\Gamma_1$$

See picture ...

Loss of Reachability and Observability Trough Sampling

- To get a reachable discrete-time system, it is necessary that the continuous-time system is also reachable.
- However it may happen that reachability is lost for some sampling periods.
- Conditions for unobservability are more restricted in the continuous-time case: output has to be zero over a time interval (for discrete-time system only at sampling instants).
- Continuous-time system may oscillate between sampling instants and remain zero at sampling instants (*hidden oscillations*).
- The sampled-data system thus can be unobservable even if the corresponding continuous-time system is observable.