State Space Representations of Linear Physical Systems

Contents

Introduction

As systems become more complex, representing them with differential equations or transfer functions becomes cumbersome. This is even more true if the system has multiple inputs and outputs. This document introduces the state space method which largely alleviates this problem. The state space representation of a system replaces an n^{th} order differential equation with a single first order *matrix* differential equation. The state space representation of a system is given by two equations :

$$\dot{\mathbf{q}}(t) = \mathbf{Aq}(t) + \mathbf{Bu}(t)$$
$$\mathbf{y}(t) = \mathbf{Cq}(t) + \mathbf{Du}(t)$$

*Note: Bold face characters denote a vector or matrix.The variable **x** is more commonly used in textbooks and other references than is the variable **q** when state variables are discussed. The variable **q** will be used here since we will often use **x** to represent position.*

The first equation is called the state equation, the second equation is called the output equation. For an n^{th} order system (i.e., it can be represented by an n^{th} order differential equation) with *r* inputs and *m* outputs the size of each of the matrices is as follows:

- **q** is n*x*1 (n rows by 1 column); **q** is called the state vector, it is a function of time
- **A** is n*x*n; **A** is the state matrix, a constant
- **B** is n*x*r; **B** is the input matrix, a constant
- **u** is r*x*1; **u** is the input, a function of time
- **C** is m*x*n; **C** is the output matrix, a constant
- **D** is m*x*r; **D** is the direct transition (or feedthrough) matrix, a constant
- **y** is m*x*1; **y** is the output, a function of time

Note several features:

- The state equation has a single first order derivative of the state vector on the left, and the state vector, **q**(t), and the input **u**(t) on the right. There are no derivatives on the right hand side.
- The output equation has the output on the left, and the state vector, **q**(t), and the input **u**(t) on the right. There are no derivatives on the right hand side.

For systems with a single input and single output (i.e., most of the systems we will consider) these variables become (with *r*=1 and *m*=1):

$$\begin{aligned}\dot{\mathbf{q}}(t) &= \mathbf{A}\mathbf{q}(t) + \mathbf{B}u(t)\\ y(t) &= \mathbf{C}\mathbf{q}(t) + Du(t)\end{aligned}$$

where

- **q** is n*x*1 (n rows by 1 column)
- **A** is n*x*n
- **B** is n*x*1
- u is 1*x*1 (i.e., a scalar)
- **C** is 1*x*n
- D is 1*x*1 (i.e, a scalar)
- y is 1*x*1 (i.e, a scalar)

Advantages of this representation include:

- The notation is very compact. Even large systems can be represented by two simple equations.
- Because all systems are represented by the same notation, it is very easy to develop general techniques to solve these systems.
- Computers easily simulate first order equations.

A Simple Example

Consider an 4^{th} order system represented by a single 4^{th} order differential equation with input x and output z.

$$\ddddot{z} + a_1\dddot{z} + a_2\ddot{z} + a_3\dot{z} + a_4 z = b_0 x$$

We can define 4 new variables, q1 through q4.

$$
\begin{aligned}
q_1 &= z \\
q_2 &= \dot{q}_1 = \dot{z} \\
q_3 &= \dot{q}_2 = \ddot{z} \\
q_4 &= \dot{q}_3 = \dddot{z}, \quad \text{so} \\
&\ddddot{z} + a_1 q_4 + a_2 q_3 + a_3 q_1 + a_4 q_1 = b_0 x
\end{aligned}
$$

but

$$
\begin{aligned}
&\dot{q}_4 = \ddddot{z}, \quad \text{so} \\
&\ddddot{z} = \dot{q}_4 = -a_4 q_1 - a_3 q_2 - a_2 q_3 - a_1 q_4 + b_0 x
\end{aligned}
$$

We can now rewrite the 4th order differential equation as 4 first order equations

$$
\begin{aligned}
\dot{q}_1 &= q_2 = \dot{z} \\
\dot{q}_2 &= q_3 = \ddot{z} \\
\dot{q}_3 &= q_4 = \dddot{z} \\
\dot{q}_4 &= -a_4 q_1 - a_3 q_2 - a_2 q_3 - a_1 q_4 + b_0 x
\end{aligned}
$$

This is compactly written in state space format as

$$
\dot{\mathbf{q}} = \mathbf{A}\mathbf{q} + \mathbf{B}u \qquad
\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \\ \dot{q}_4 \end{bmatrix} =
\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -a_4 & -a_3 & -a_2 & -a_1 \end{bmatrix}
\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} +
\begin{bmatrix} 0 \\ 0 \\ 0 \\ b_0 \end{bmatrix} x
$$

$$
y = \mathbf{C}\mathbf{q} + Du \qquad
y = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}
\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix}
$$

with

$$
\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -a_4 & -a_3 & -a_2 & -a_1 \end{bmatrix}; \quad
\mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ b_0 \end{bmatrix}; \quad
\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}; \quad
D = 0; \quad \text{and the input} = u = x; \ \text{output} = z
$$

For this problem a state space representation was easy to find. In many cases (e.g., if there are derivatives on the right side of the differential equation) this problem can be much more difficult. Such cases are explained in the discussion of transformations between system representations.

The state space representation is not unique

Case 1: Alternate State Space Representation

Another important point is that the state space representation is not unique. As a simple example we could simply reorder the variables from the example above (the new state variables are labeled $\mathbf{q}_{new}$). This results in a new state space representation

$$
\begin{aligned}
q_{3,new} &= z \\
\dot{q}_{3,new} &= q_{2,new} = \dot{z} \\
\dot{q}_{2,new} &= q_{1,new} = \ddot{z} \\
\dot{q}_{1,new} &= q_{4,new} = \dddot{z}
\end{aligned}
$$

$$\dot{q}_{4,new} = -a_4 q_{3,new} - a_3 q_{2,new} - a_2 q_{1,new} - a_1 q_{4,new} + b_0 x$$

$$\dot{\mathbf{q}}_{new} = \mathbf{A}_{new}\mathbf{q}_{new} + \mathbf{B}_{new}u \qquad \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \\ \dot{q}_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -a_2 & -a_3 & -a_4 & -a_1 \end{bmatrix} \begin{bmatrix} q_{1,new} \\ q_{2,new} \\ q_{3,new} \\ q_{4,new} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ b_0 \end{bmatrix} u$$

$$y = \mathbf{C}_{new}\mathbf{q}_{new} + D_{new}u \qquad y = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} q_{1,new} \\ q_{2,new} \\ q_{3,new} \\ q_{4,new} \end{bmatrix}$$

$$\mathbf{A}_{new} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -a_2 & -a_3 & -a_4 & -a_1 \end{bmatrix}; \quad \mathbf{B}_{new} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ b_0 \end{bmatrix};$$

$$\mathbf{C}_{new} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}; \quad D_{new} = 0; \quad \text{and the input} = u = x; \text{ output} = z$$

Case 2: Alternate State Space Representations

In the previous case careful examination of the original and modified state space system reveals that they represent the same system. However we can make entirely new state variables by forming linear combination of the original state variables in which this equality is not obvious. Consider the state variable $\mathbf{q}_{new}$ defined as follows:

$$
\begin{aligned}
q_{1,new} &= z - \dot{z} \\
q_{2,new} &= z + \dot{z} \\
q_{3,new} &= \ddot{z} \\
q_{4,new} &= \dddot{z}
\end{aligned}
$$

In this case the new state space variables are given by (*the details of how these matrices are determined are not important for this discussion. They are given here if you are interested*):

$$\begin{aligned}\dot{\mathbf{q}}_{new} &= \mathbf{A}_{new}\mathbf{q}_{new} + \mathbf{B}_{new}u \\ y &= \mathbf{C}_{new}\mathbf{q}_{new} + D_{new}u\end{aligned}$$

$$\mathbf{A}_{new} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & -1 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \frac{a_3 - a_4}{2} & -\frac{a_3 + a_4}{2} & -a_2 & -a_1 \end{bmatrix}$$

$$\mathbf{B}_{new} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ b_0 \end{bmatrix}$$

$$\mathbf{C}_{new} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

$$D_{new} = 0$$

This new state space system is quite different from the original one, and it is not at all obvious that they represent the same system. (*It can be shown that the systems are identical by transforming the state space representation to a transfer function. Techniques for doing so are discussed elsewhere.*)

Key Concept: Defining a State Space Representation

A n^{th} order linear physical system can be represented using a state space approach as a single first order matrix differential equation:

$$\begin{aligned}\dot{\mathbf{q}}(t) &= \mathbf{Aq}(t) + \mathbf{Bu}(t) \\ \mathbf{y}(t) &= \mathbf{Cq}(t) + \mathbf{Du}(t)\end{aligned}$$

The first equation is called the state equation and it has a first order derivative of the state variable(s) on the left, and the state variable(s) and input(s), multiplied by matrices, on the right. The second equation is called the output equation and it has the output on the left and the state variable(s) and input(s), multiplied by matrices, on the right. No other terms are allowed in the equation. In these equations:

- **q** is n*x*1 (n rows by 1 column); **q** is called the state vector, it is a function of time
- **A** is n*x*n; **A** is the state matrix, a constant
- **B** is n*x*r; **B** is the input matrix, a constant
- **u** is r*x*1; **u** is the input, a function of time
- **C** is m*x*n; **C** is the output matrix, a constant
- **D** is m*x*r; **D** is the direct transition (or feedthrough) matrix, a constant
- **y** is m*x*1; **y** is the output, a function of time

For a single input, single output system (the case that interests us the most):

$$\begin{aligned}\dot{\mathbf{q}}(t) &= \mathbf{Aq}(t) + \mathbf{B}u(t) \\ y(t) &= \mathbf{Cq}(t) + Du(t)\end{aligned}$$

- **q** is n*x*1 (n rows by 1 column)
- **A** is n*x*n
- **B** is n*x*1

- u is 1x1 (i.e., a scalar)
- **C** is 1x*n*
- D is 1x1 (i.e, a scalar)
- y is 1x1 (i.e, a scalar)

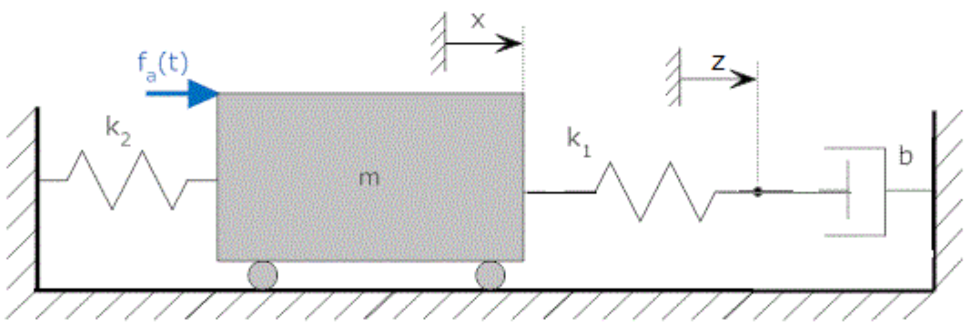
The state space representation is not unique; many (actually an infinite number) of state space systems can be used to represent any linear physical system.

Developing a state space model from a system diagram (Mechanical Translating)

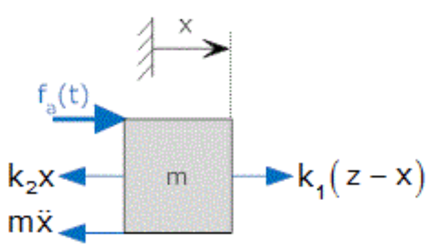
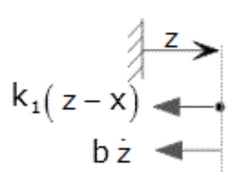
Another, powerful, way to develop a state space model is directly from the free body diagrams. If you choose as your state variables those quantities that determine the energy in the system, a state space system is often easy to derive. For example, in a mechanical system you would choose extension of springs (potential energy, ½kx²) and the velocity of masses (kinetic energy, ½mv²); for electrical systems choose voltage across capacitors, ½Ce² (e=voltage)) and current through inductors (½Li²). This is best illustrated by several examples, two rotating and one electrical.

Example: Direct Derivation of State Space Model (Mechanical Translating)

Derive a state space model for the system shown. The input is f_a and the output is z.



We can write free body equations for the system at x and at y.

Freebody Diagram	**Equation**
	$m\cdot\ddot{x}+k_1\cdot x+k_2x-k_1\cdot z=f_a$
	$b\cdot\dot{z}+k_1z-k_1\cdot x=0$

There are three energy storage elements, so we expect three state equations. The energy storage elements are the spring, k_2, the mass, m, and the spring, k_1. Therefore we choose as our state variables x (the energy in spring k_2 is $\frac{1}{2}k_2x^2$), the velocity at x (the energy in the mass m is ½mv², where v is the first derivative of x), and y (the energy in spring k_1 is $½k_1(z-x)^2$, so we could pick z-x as a state variable, but we'll just use z (since x is already a state variable; recall that the choice of state variables is not unique). Our state variables become:

$$
\begin{aligned}
q_1 &= x \\
q_2 &= \dot{x} \\
q_3 &= z
\end{aligned}
$$

Now we want equations for their derivatives. The equations of motion from the free body diagrams yield

$$
\begin{aligned}
\dot{q}_1 &= \dot{x} = q_2 \\
\dot{q}_2 &= \ddot{x} = \frac{1}{m}\left(f_a - k_1 x - k_2 x + k_1 z\right) \\
&= \frac{1}{m}\left(f_a - k_1 q_1 - k_2 q_1 + k_1 q_3\right) \\
\dot{q}_3 &= \dot{z} = \frac{k_1}{b}\left(x - z\right) = \frac{k_1}{b}\left(q_1 - q_3\right)
\end{aligned}
$$

or

$$
\dot{\mathbf{q}} = \mathbf{A}\mathbf{q} + \mathbf{B}u \qquad \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{k_1 + k_2}{m} & 0 & \frac{k_1}{m} \\ \frac{k_1}{b} & 0 & -\frac{k_1}{b} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 \\ \frac{1}{m} \\ 0 \end{bmatrix}
$$

$$
y = \mathbf{C}\mathbf{q} + Du \qquad \mathbf{C} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \qquad D = 0
$$

with the input $u=f_a$, and the output y=z.

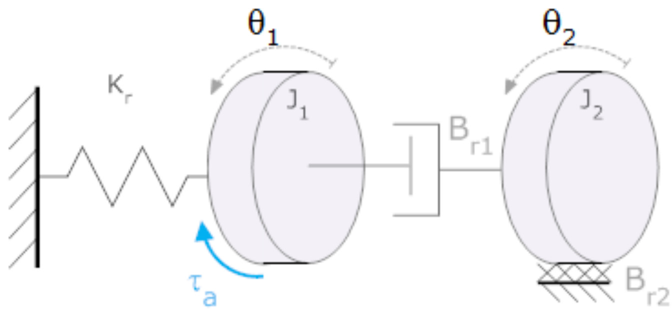
This technique does not always easily yield a set of state equations (read about some examples here). In some cases it is easier to develop a transfer function model and convert this to a state space model. Transfer functions are discussed elsewhere.

Developing state space model from system diagram (Mechanical Rotating)

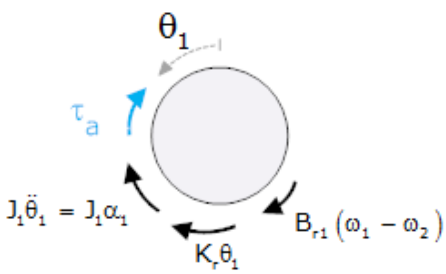
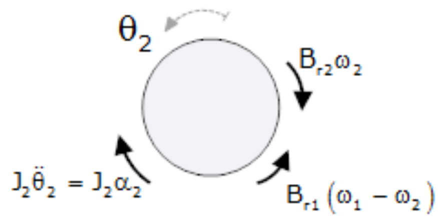
The energy variables for rotating systems are potential energy stored in springs ($½K_r\theta^2$) and kinetic energy stored in inertial elements ($½J\omega^2$).

Example: Direct Derivation of State Space Model (Mechanical Rotating)

Derive a state space model for the system shown. The input is τ_a and the output is θ_1.



We can write free body equations for the system at θ_1 and θ_2

Freebody Diagram	Equation
	$J_1\ddot{\theta}_1 + B_{r1}\dot{\theta}_1 + K_r\theta_1 - B_{r1}\dot{\theta}_2 = -\tau_a$
	$J_2\ddot{\theta}_2 + (B_{r2} + B_{r1})\dot{\theta}_2 - B_{r1}\dot{\theta}_1 = 0$

There are three energy storage elements, so we expect three state equations. Energy is stored as potential energy in the spring ($\frac{1}{2}K_r\theta_1^2$) and kinetic energy in the two flywheels ($\frac{1}{2}J_1\alpha_1^2$, $\frac{1}{2}J_2\alpha_2^2$). Our state variable equations become:

$$\begin{aligned} q_1 &= \theta_1 \\ q_2 &= \dot{\theta}_1 \\ q_3 &= \dot{\theta}_2 \end{aligned}$$

Now we want equations for their derivatives. The equations of motion from the free body diagrams yield

$$\begin{aligned} \dot{q}_1 &= \dot{\theta}_1 = q_2 \\ \dot{q}_2 &= \ddot{\theta}_1 = \frac{1}{J_1}\left(-\tau_a - B_{r1}\dot{\theta}_1 - K_r\theta_1 + B_{r1}\dot{\theta}_2\right) \\ &= -\frac{1}{J_1}\tau_a - \frac{B_{r1}}{J_1}q_2 - \frac{K_r}{J_1}q_1 + \frac{B_{r1}}{J_1}q_3 \\ \dot{q}_3 &= \ddot{\theta}_2 = \frac{1}{J_2}\left(-(B_{r2} + B_{r1})\dot{\theta}_2 + B_{r1}\dot{\theta}_1\right) \\ &= -\frac{(B_{r2} + B_{r1})}{J_2}q_3 + \frac{B_{r1}}{J_2}q_2 \end{aligned}$$

or

$$\dot{q} = \mathbf{A}\mathbf{q} + \mathbf{B}u \quad \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{K_r}{J_1} & -\frac{B_{r1}}{J_1} & \frac{B_{r1}}{J_1} \\ 0 & \frac{B_{r1}}{J_2}q_2 & -\frac{(B_{r2}+B_{r1})}{J_2} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 0 \\ -\frac{1}{J_1} \\ 0 \end{bmatrix}$$

$$y = \mathbf{C}\mathbf{q} + Du \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \quad \mathbf{D} = 0$$

with the input $u=\tau_a$, and the output $y=\theta_1$.

Developing State space model from system diagram (Electrical)

To develop a state space system for an electrical system, try choosing the voltage across capacitors, and current through inductors as state variables. Recall that

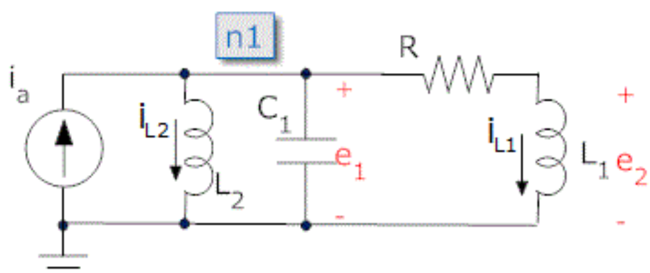
$$e_{inductor} = L\frac{di_{inductor}}{dt}$$

$$i_{capacitor} = C\frac{de_{capacitor}}{dt}$$

so if we can write equations for the voltage across an inductor, it becomes a state equation when we divide by the inductance (i.e., if we have an equation for $e_{inductor}$ and divide by L, it becomes an equation for $di_{inductor}/dt$ which is one of our state variables). Likewise if we can write an equation for the current through the capacitor and divide by the capacitance it becomes a state equation for $e_{capacitor}$. This is best illustrated by an example.

Example: Direct Derivation of State Space Model (Electrical)

Derive a state space model for the system shown. The input is i_a and the output is e_2.



There are three energy storage elements, so we expect three state equations. Try choosing i_1, i_2 and e_1 as state variables. Now we want equations for their derivatives. The voltage across the inductor L_2 is e_1 (which is one of our state variables)

$$L_2\frac{di_{L2}}{dt} = e_1$$

so our first state variable equation is

$$\frac{di_{L2}}{dt} = \frac{1}{L_2}e_1$$

If we sum currents into the node labeled n1 we get

$$i_a - i_{L2} - i_{C1} - i_{L1} = 0$$

This equation has our input (ia) and two state variable (iL2 and iL1) and the current through the capacitor. So from this we can get our second state equation

$$i_{C1} = C_1 \frac{de_1}{dt} = i_a - i_{L2} - i_{L1}$$
$$\frac{de_1}{dt} = \frac{1}{C_1}\left(i_a - i_{L2} - i_{L1}\right)$$

Our third, and final, state equation we get by writing an equation for the voltage across L_1 (which is e_2) in terms of our other state variables

$$e_2 = L_1 \frac{di_{L1}}{dt} = e_1 - Ri_{L1}$$
$$\frac{di_{L1}}{dt} = \frac{1}{L_1}\left(e_1 - Ri_{L1}\right)$$

We also need an output equation:

$$e_2 = e_1 - Ri_{L1}$$

So our state space representation becomes

$$\mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} i_{L2} \\ e_1 \\ i_{L1} \end{bmatrix}$$

$$\dot{\mathbf{q}} = \mathbf{Aq} + \mathbf{B}u \qquad \mathbf{A} = \begin{bmatrix} 0 & \frac{1}{L_2} & 0 \\ -\frac{1}{C_1} & 0 & -\frac{1}{C_1} \\ 0 & \frac{1}{L_1} & -\frac{R}{L_1} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 \\ \frac{1}{C_1} \\ 0 \end{bmatrix}$$

$$y = \mathbf{Cq} + Du \qquad \mathbf{C} = \begin{bmatrix} 0 & 1 & -R \end{bmatrix} \qquad D = 0$$

This technique does not always easily yield a set of state equations. In some cases it is easier to develop a transfer function model and convert this to a state space model. Transfer functions are discussed elsewhere.

Problems when developing a state space model from a system diagram

There are several cases when it is not so straightforward to develop a state space model from a system diagram. Some of these are discussed here.

Solution of State Space Problems

The state space representation of a system is a common and extremely powerful method of representing a system mathematically. This page only discusses how to develop the state space representation, the solution of state space problems are discussed elsewhere.

Transformations to other forms

Since state space is equivalent to the other representations, there must be a way to transform from one representation to another. These methods are discussed here.

References