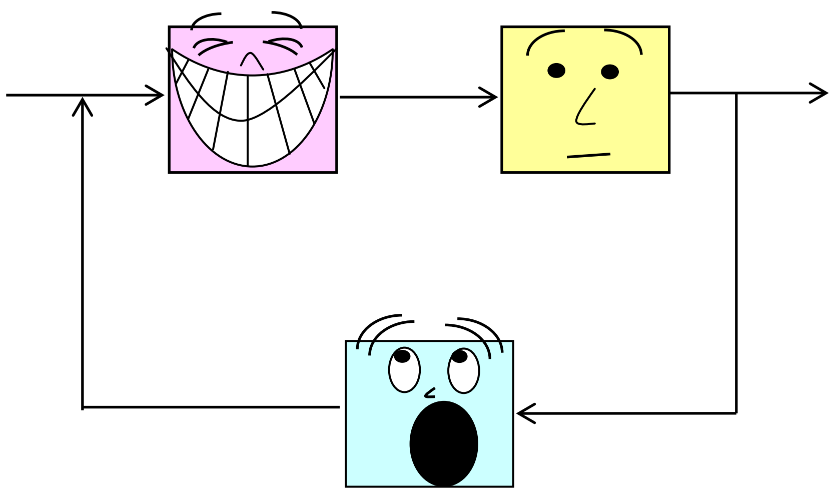
A Cartoon Tour of Control Theory

Part I- Classical Controls

S. M. Joshi



© 1990-2015 S. M. Joshi
All rights reserved

Blank page

A Cartoon Tour of Control Theory

Part I- Classical Controls

S. M. Joshi

© 1990 S. M. Joshi
All rights reserved.

Any redistribution, reproduction, or online posting of any of the contents in any form is prohibited except the following: 1) Contents may be printed on paper or downloaded to your personal local storage disk for your personal non-commercial use , and 2) Contents may be copied to individual third parties for their personal use, provided you acknowledge the website as the source. Commercial use or distribution of the contents is prohibited without our express written permission. The contents may not be transmitted or stored in any other website or other form of electronic retrieval system.

Additional Notes

Download Website (as of Nov 2, 2015): http://controlcartoons.com
(The author's "Out of Control" cartoons are also available at this website)

Author Contact Info: Send email to: sj.systemtheory@gmail.com

Blank page

Preface

This booklet is intended to be a light introduction to some basic ideas in controls. The objectives are to

- Promote student interest in control engineering and systems science
- Educate beginners and non-specialists
- Entertain specialists and geeks

Typically, ECE students who have taken a Junior level (3rd year) Signals & Systems course, and EE/ME/AE/ChE students taking a Senior level Controls course, should find it easily understandable. I hope to do additional parts (state-space methods, optimal control, nonlinear systems ….etc) in the future.

This material was originally prepared as a live presentation, therefore some narrative explanations are missing. I have been too lazy to add them, and I hope the reader can interpolate as necessary.

Finally, this is only a light introduction; if you are seriously interested in controls, please take a course or get a real textbook- there are many good ones.

Author Info

S. M. Joshi, received his bachelor's and master's degrees from India (Banaras and IIT-Kanpur), and his PhD from Rensselaer Polytechnic Institute (Troy, NY), all in Electrical Engineering. He is Fellow of the IEEE, AIAA, and ASME, and the author/coauthor of over 200 serious publications- including 3 books- in control theory and aerospace applications, his research area for many years. He also taught several controls courses at three universities and advised graduate dissertations. He was the originator and contributor of the "Out of Control" cartoons in the *IEEE Control Systems Magazine* (1985- 1994). He is a recipient of a number of prestigious technical awards from IEEE, ASME, AACC (American Automatic Control Council), and NASA.

Blank page

Disclaimer

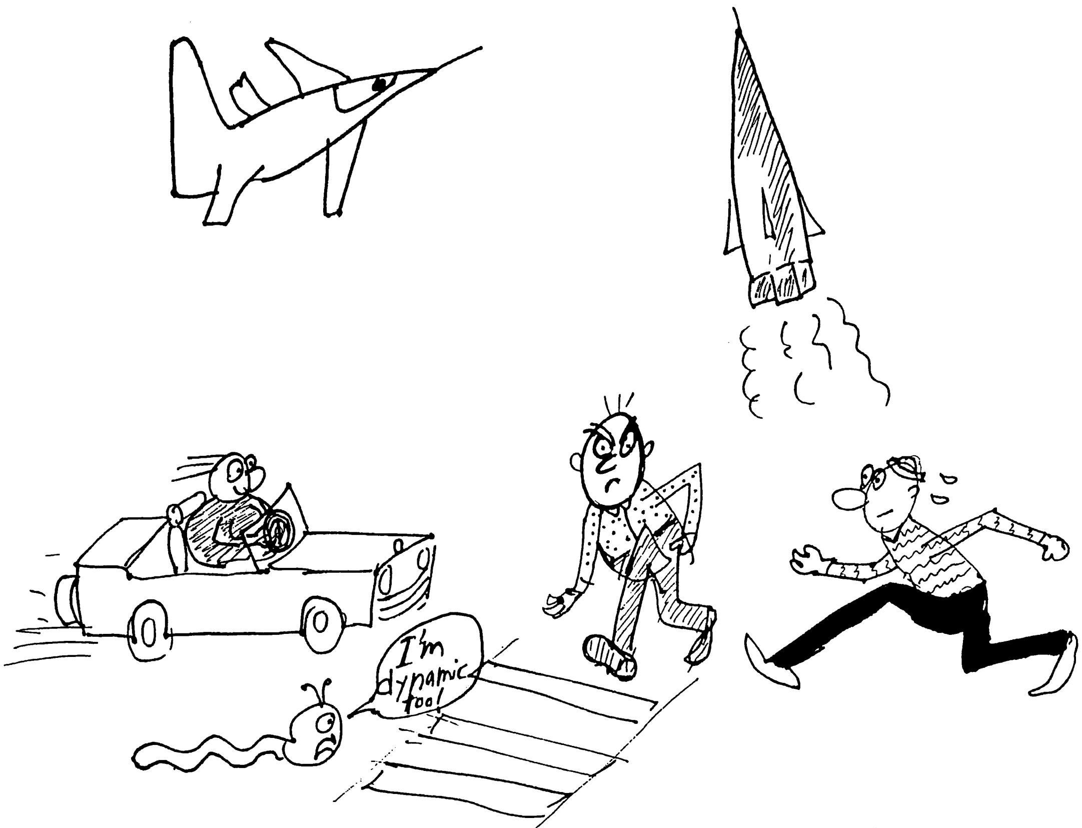
The contents, views, and opinions in this booklet are solely those of the author and not of any organization. Any similarities to real-life persons or situations (except for historical references) are purely coincidental.

Blank page

CONTENTS

DYNAMICS

Nothing† can move instantaneously!



† Even your mind cannot wander faster than the speed of light!

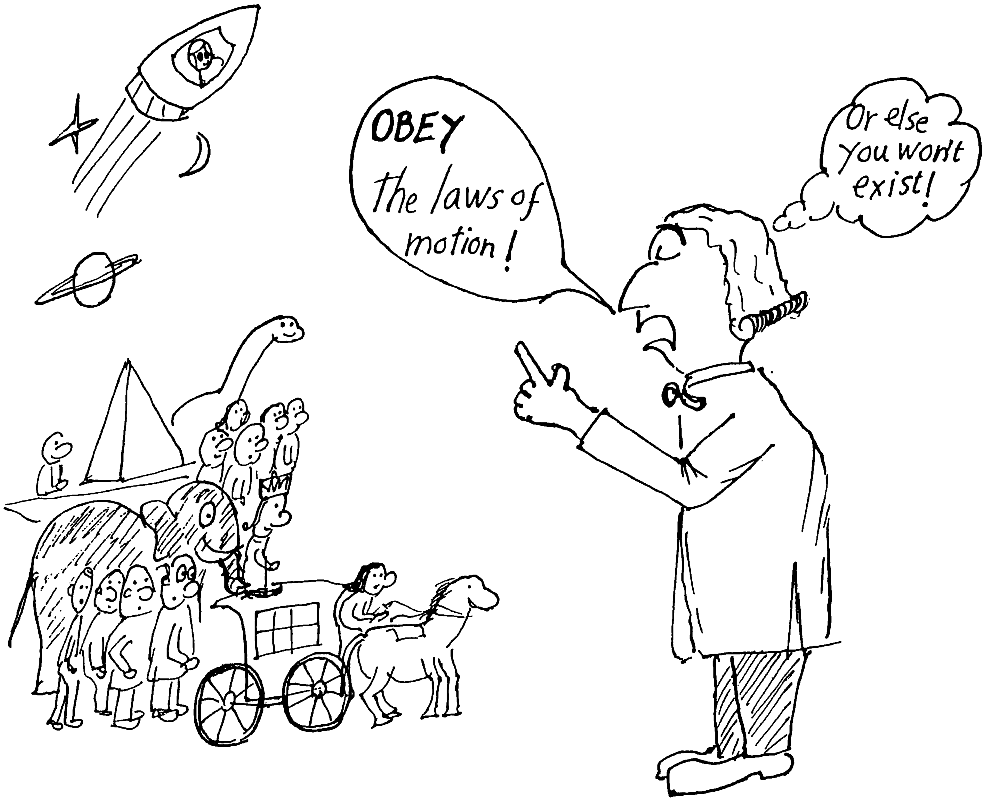
© 1990 S. M. Joshi

People didn't know for a long time how to predict dynamic behavior of things.

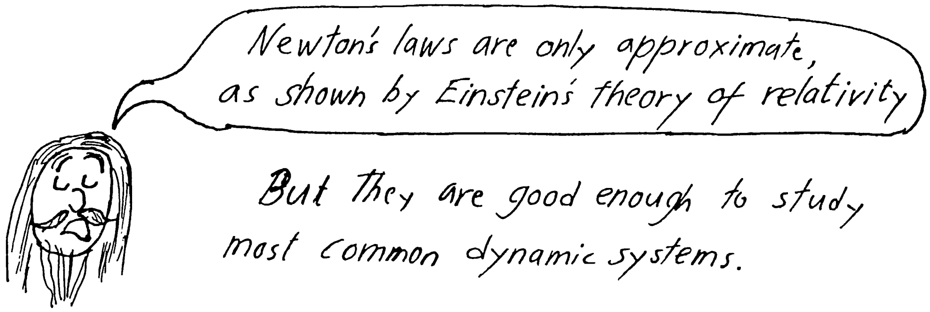
And then, one day†....

....SIR ISAAC NEWTON discovered the laws of motion AND Calculus.

(† not really! over a period of several years!) © 1990 S. M. Joshi

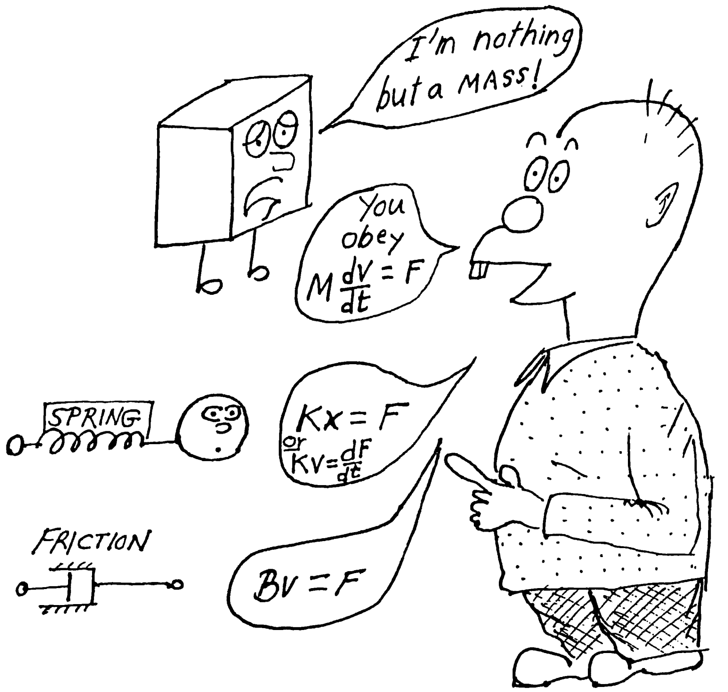


And so the science of DYNAMIC MATH-MODELING was born in the 17th century.



© 1990 S. M. Joshi

The art of analyzing physical systems using their math models came into being in the 17th and 18th centuries.



M = mass (lbm or kg)

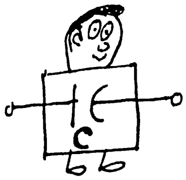
V = velocity (ft/sec or m/s)

F = Force (lbf or Newtons)

K = Spring constant; B = Friction parameter

© 1990 S. M. Joshi

ELECTRICAL SYSTEMS
obey similar laws.



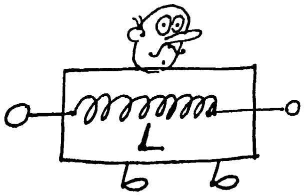
Capacitor

$$I = C\frac{dV}{dt}$$

V = Voltage (Volts)

C = Capacitance (Farads)

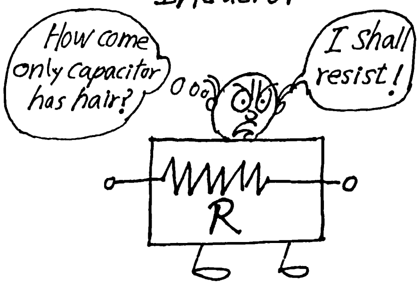
I = Current (Amperes)



Inductor

$$V = L\frac{dI}{dt}$$

L = Inductance (Henries)



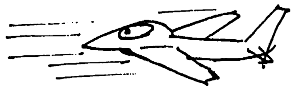
Resistor

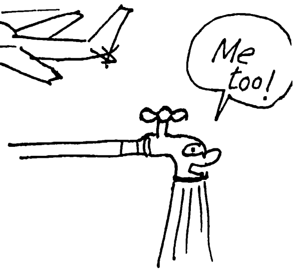
$$V = RI$$

R = Resistance (Ohms)

© 1990 S. M. Joshi

FLUID SYSTEMS





have similar laws!

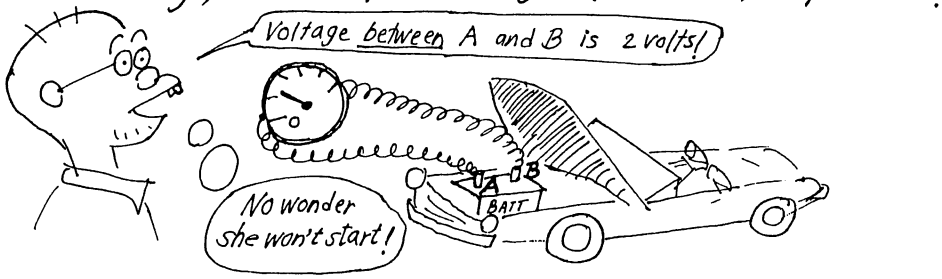
So do thermal systems,

and many others.

Equations of motion of a dynamic component usually 'connect' two types of variables.

1). "Across" variable - something that's defined relative to a reference point.

e.g., velocity, voltage, pressure, temperature.

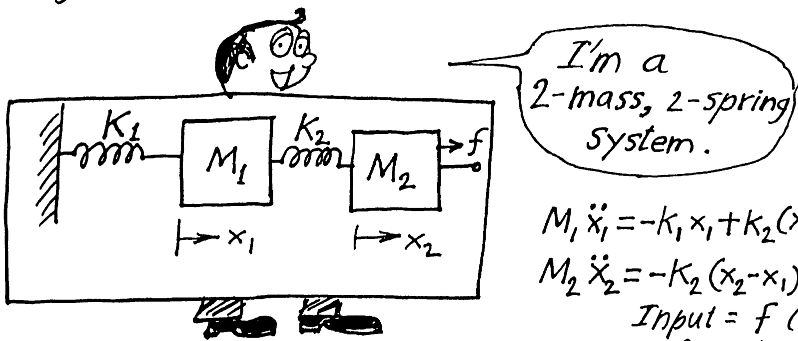


2). "Through" variable - the "other" variable:
e.g., current, flowrate, force, etc.

 © 1990 S. M. Joshi

A SYSTEM

is formed by connecting various components together in a certain manner.

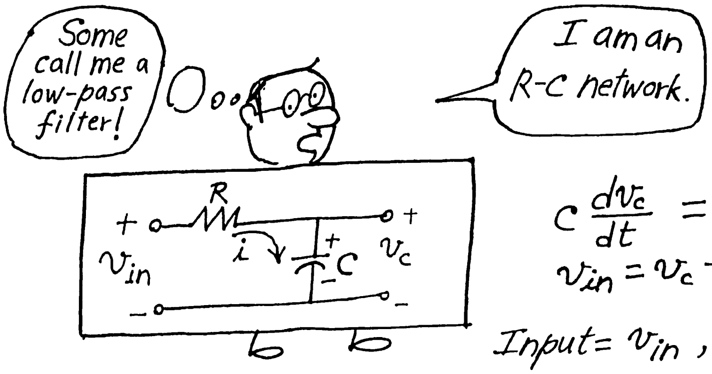


$$M_1 \ddot{x}_1 = -K_1 x_1 + K_2 (x_2 - x_1)$$
$$M_2 \ddot{x}_2 = -K_2 (x_2 - x_1) + f$$

Input = f (force)

Output = x_2

$$\left[\dot{x} \triangleq \frac{dx}{dt} ; \ \ddot{x} \triangleq \frac{d^2x}{dt^2} \right]$$



$$C \frac{dv_c}{dt} = i$$
$$v_{in} = v_c + Ri = v_c + RC \frac{dv_c}{dt}$$

Input = v_{in}, Output = v_c

A COMPLEX SYSTEM consists of many different components.

Each component obeys a differential or an algebraic equation.

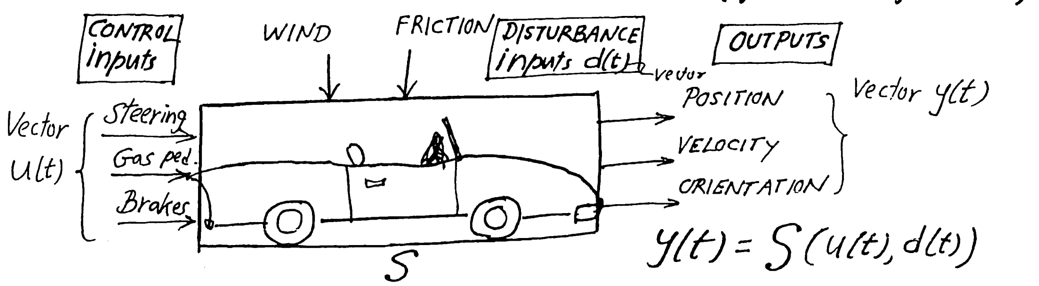
© 1990 S. M. Joshi

What do we get when we connect lots of different components?

We get a complex system represented by a bunch of coupled DIFFERENTIAL & ALGEBRAIC equations.

We will consider systems described by ORDINARY differential equations only.
(although partial diff. eqs. are neat - and are required for studying lots of systems).

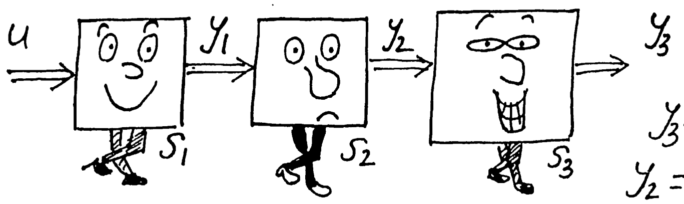
A SYSTEM is usually represented by a block (or a box) which has INPUTS & OUTPUTS (functions of time 't')



$$y(t) = S(u(t), d(t))$$

© 1990 S. M. Joshi

SYSTEMS can be connected with other systems to form a new system.

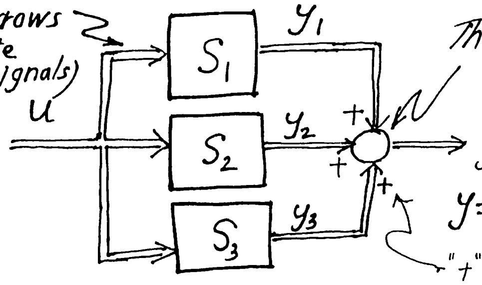


$y_3 = S_3(y_2)$
$y_2 = S_2(y_1)$
$y_1 = S_1(u)$

SERIES Connection

(Control folks use double arrows to denote VECTOR signals)

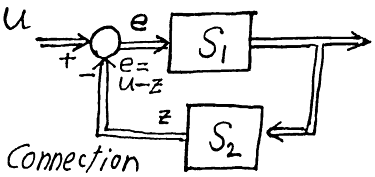
This is called a "Summing point"



$y = [S_1(u_1) + S_2(u) + S_3(u)]$

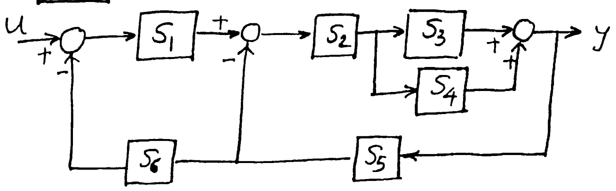
"+" or "−" indicates if the signal is added or subtracted.

PARALLEL Connection



$y = S_1[u - S_2(y)]$

FEEDBACK Connection



Genaral messy block diagram.

© 1990 S. M. Joshi

MATH MODELING of real physical systems involving different types of components and coordinate transformations is pretty complicated.

After Newton, the field of mechanics saw significant advancements in the 18th & 19th centuries.

JEAN LE ROND D'ALEMBERT (1717-1783) discovered his famous principle.

JOHN BERNOULLI (1667-1748)

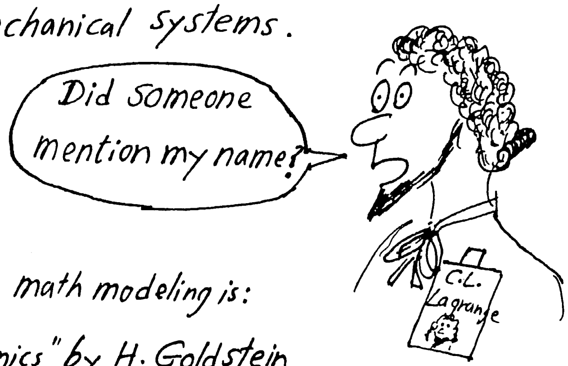
LEONHARD EULER (1707-1783)

COMTE LOUIS DE LAGRANGE (1736-1813)

laid the foundations of modern mechanics.

© 1990 S. M. Joshi

LAGRANGIAN FORMULATION became a standard method for math-modeling of complicated mechanical systems.

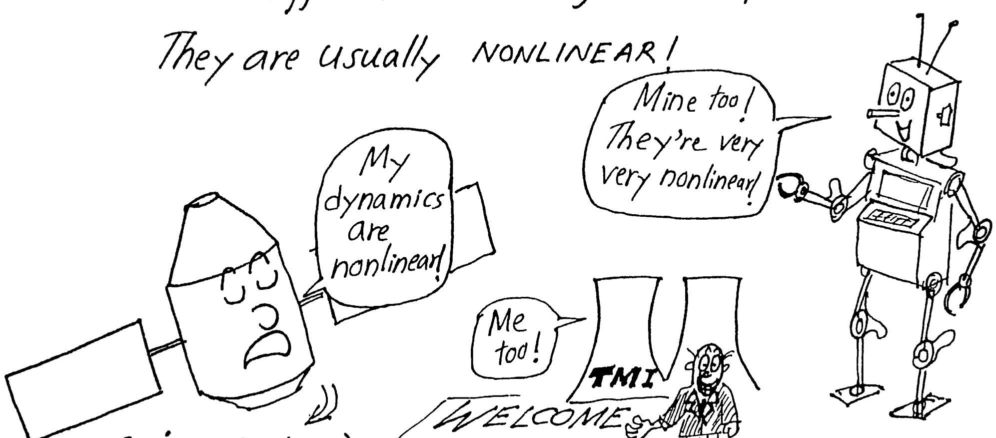


A standard text on math modeling is:

"Classical Mechanics" by H. Goldstein (Addison-Wesley, 1953).

What we finally get is a bunch of coupled differential and algebraic equations.

They are usually NONLINEAR!



$$[J]_{3\times3}\,\dot{\omega} + \omega \times (J\omega) = T$$

Moment of inertia matrix — angular velocity — applied torque

© 1990 S. M. Joshi

MATH MODELS

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} \qquad u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_m(t) \end{bmatrix} \qquad y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_\ell(t) \end{bmatrix}$$

$$\frac{dx(t)}{dt} \triangleq \dot{x}(t) = \overset{\text{FUNCTION}}{f}\Big(\underbrace{x(t)}_{\text{STATE Vector}}, \underbrace{u(t)}_{\text{INPUT Vector}}, t\Big)$$

$$y(t) = g(x(t), u(t), t)$$

EXAMPLE:

AIRPLANE DYNAMICS

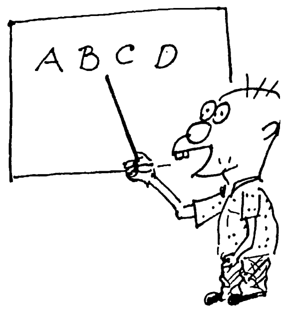
$$x = \begin{bmatrix} \omega \\ \theta \\ v \\ w \end{bmatrix} \begin{array}{l} \text{angular rate} \\ \text{orientation} \\ \text{velocity} \\ \text{position} \end{array} \qquad u = \begin{bmatrix} \delta_e \\ \delta_A \\ \delta_R \\ T \end{bmatrix} \begin{array}{l} \text{elevator} \\ \text{aileron} \\ \text{rudder} \\ \text{Thrust} \end{array}$$

(each is $\underline{3 \times 1}$)

If the flexibility of the airplane is significant, there will be additional state variables.

© 1990 S. M. Joshi

LINEARIZING EQS. OF MOTION
(Or "ABCD's of control!)



Nonlinear eqs:

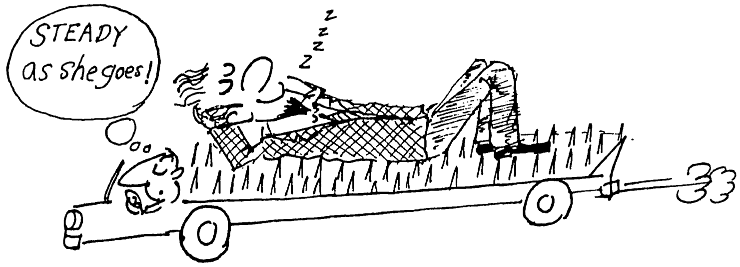
$$\dot{x}(t) = f(x(t), u(t))$$

"STATE" vector $(n \times 1)$ — "INPUT" vector $(m \times 1)$

"OUTPUT" vector $(\ell \times 1)$

$$y(t) = g(x(t), u(t))$$

We consider STEADY-STATE motion.
e.g., constant-speed, straight flight of an airplane.



Steady-state: $\dot{\bar{x}} = 0 = f(\bar{x}, \bar{u})$

Let $x(t) = \bar{x} + \delta x(t)$; $u(t) = \bar{u} + \delta u(t)$

$y(t) = \bar{y} + \delta y(t)$

Then, TO FIRST-ORDER:

$$\delta\dot{x} = \left.\frac{\partial f}{\partial x}\right|_{\bar{x},\bar{u}} \delta x + \left.\frac{\partial f}{\partial u}\right|_{\bar{x},\bar{u}} \delta u$$

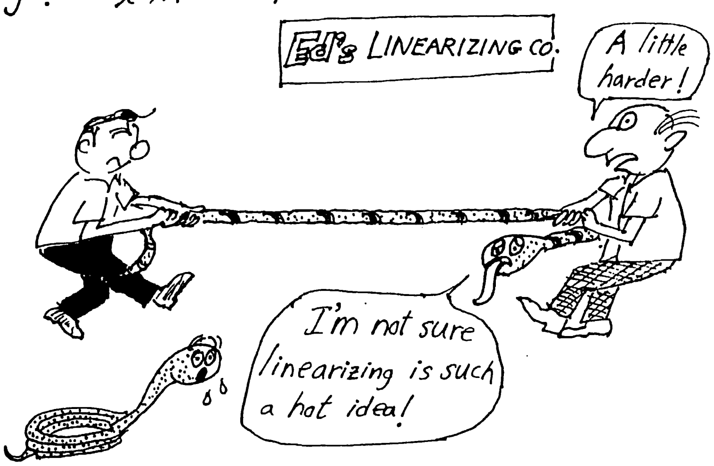
$$\delta y = \left.\frac{\partial g}{\partial x}\right|_{\bar{x},\bar{u}} \delta x + \left.\frac{\partial g}{\partial u}\right|_{\bar{x},\bar{u}} \delta u$$

© 1990 S. M. Joshi

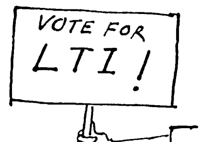
For simplicity, let's denote "δx" by "x", etc.
Linearized eqs. of motion are of the form:

$$\dot{x}(t) = A\,x(t) + B\,u(t)$$
$$y(t) = C\,x(t) + D\,u(t)$$

x : $n \times 1$ state vector ; u : $m \times 1$ input vector
y : $\ell \times 1$ output vector



When "A", "B", "C", "D" are fixed (i.e., not functions of time), we get a LINEAR, TIME-INVARIANT (LTI) system.
(The most popular kind of system).



© 1990 S. M. Joshi

STATE SPACE MODEL

$$\dot{x} = Ax + Bu$$

State vector → $\dot{x}$; A ← System matrix; B ← input matrix

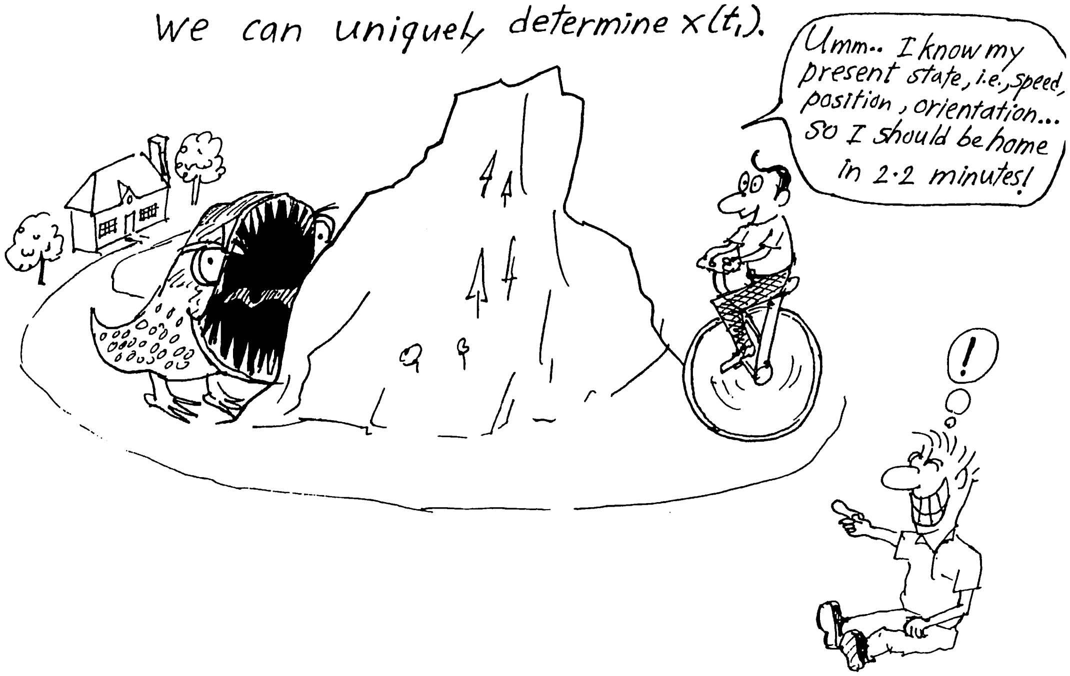
$$y = Cx + Du$$

Output vector → y; C ← Output matrix; D ← Direct transmission matrix

Given the <u>state</u> $x(t_0)$ at time t_0,

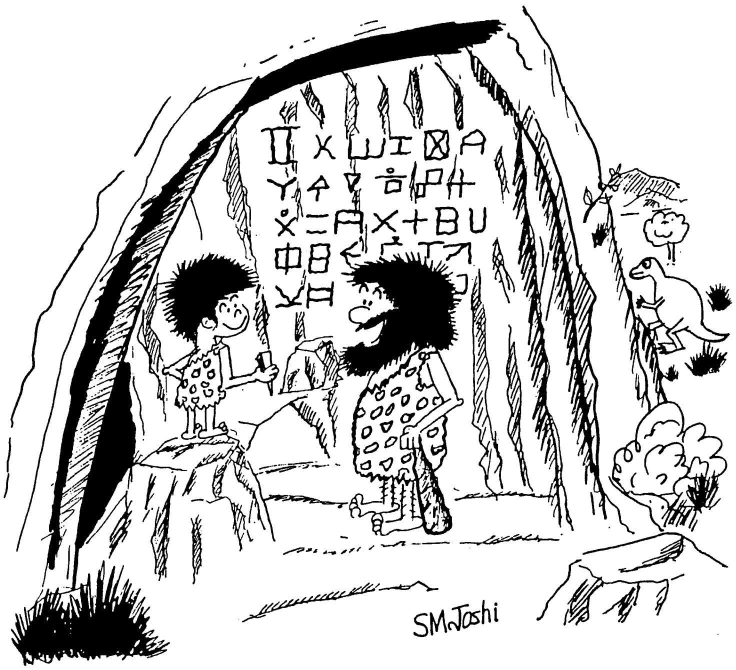
and <u>input</u> $u(t)$ for $t_0 \leq t \leq t_1$,

we can uniquely determine $x(t_1)$.



© 1990 S. M. Joshi

ANCIENT ORIGINS OF STATE EQUATION



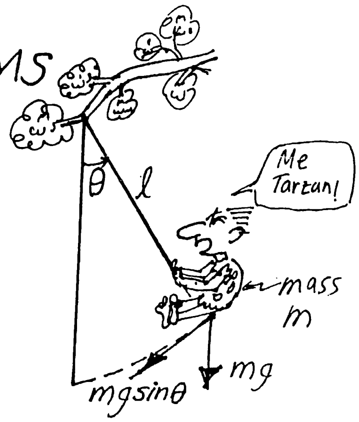
CSM-16
12/86

"Nice artwork, kiddo! I've got a gut feeling that a great many are going to make a living off that third line someday!"

© 1990 S. M. Joshi

LTI SYSTEMS

EXAMPLE: Simple pendulum



$$m\ell\ddot{\theta}(t) + mg\sin\theta(t) = 0$$

Let $\theta_1 = \theta$, $\theta_2 = \dot{\theta}$. Then

$$\dot{x} = \begin{bmatrix}\dot{\theta}_1\\ \dot{\theta}_2\end{bmatrix} = \begin{bmatrix}\theta_2\\ -\frac{g}{\ell}\sin\theta_1\end{bmatrix} = f(x,u)$$

(but $u=0$)

$$y = \theta_1 = [1 \quad 0]\, x$$

Linearize about $\theta = \dot{\theta} = 0$

(θ small $\Rightarrow$ $\sin\theta \simeq \theta$)

$$\Rightarrow \quad \dot{x} = \underbrace{\begin{bmatrix}0 & 1\\ -\frac{g}{\ell} & 0\end{bmatrix}}_{\text{"A"}} x + \underbrace{\begin{bmatrix}0\\0\end{bmatrix}}_{\text{"B"}} u$$

$$y = \overbrace{[1 \quad 0]}^{\text{"C"}}\, x + \overbrace{[0]}^{\text{"D"}}\, u$$

Suppose the guy has a jetpack attached, which produces force $u(t)$

Then

$$B = \begin{bmatrix}0\\ -\frac{1}{m}\end{bmatrix}$$

© 1990 S. M. Joshi

Basically an LTI System may consist of a number of LTI differential equations and algebraic equations. For example, consider

$$\frac{d^3y}{dt^3} + 2\frac{d^2y}{dt^2} - 5\frac{dy}{dt} + 3y = 7u$$

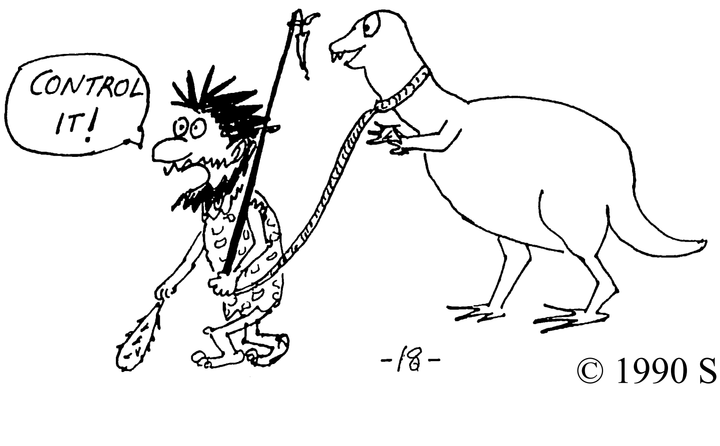
Define: $\quad y = x_1 \;; \quad \dot{x}_1 = x_2, \; \dot{x}_2 = x_3$

Then we have:
$$\dot{x} \triangleq \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -3 & 5 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix} u$$

$$y = [1 \;\; 0 \;\; 0]\, x$$

So we have converted one, third-order differential equation into 3, first-order coupled eqs.

Herein we will study single-input, single-output (SISO) systems, and in particular, how to make them behave the way WE want (i.e., CONTROL them).

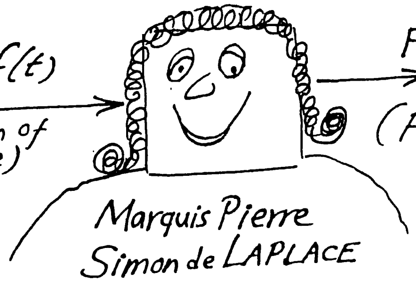


© 1990 S. M. Joshi

LAPLACE TRANSFORM

is a fundamental tool in control systems analysis and design.

$f(t)$ (function of time) →



→ $F(s)$ TRANSFORM (Function of Complex variable "s")

Marquis Pierre Simon de LAPLACE (1749 - 1827)

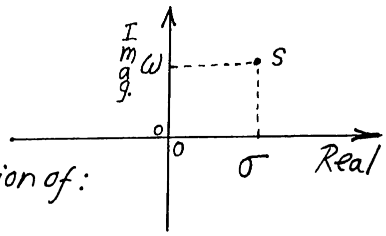
SOME BACKGROUND:

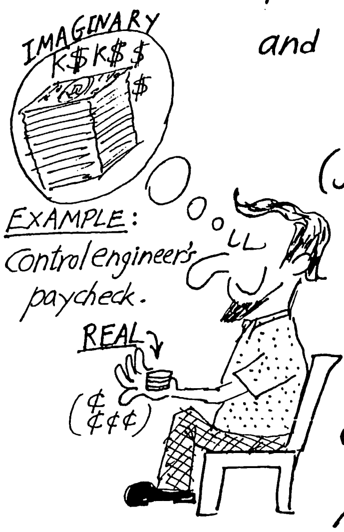
Complex variable "s" has a real part σ and an imaginary part ω.

$$s = \sigma + j\omega$$

$(j = \sqrt{-1})$

j is the solution of: $x^2+1=0$





EXAMPLE: Control engineer's paycheck.

($\sqrt{-1}$ does not exist - that's why its multiple is called the "imaginary" part).

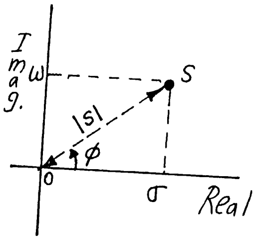
Q. Why do we use something that doesn't even exist?

A. The concept makes analysis consistent & easy.

CARL FRIEDRICH GAUSS (1777 - 1855) showed that all roots of a polynomial can be expressed as complex numbers. (Mathematicians use "i" to denote $\sqrt{-1}$; electrical engineers use "j" to avoid confusing with "current".)

© 1990 S. M. Joshi

A complex number can also be expressed in terms of its MAGNITUDE and PHASE ANGLE.



$$s = \sigma + j\omega$$
$$= |s| \angle s$$

where

magnitude of $s = \sqrt{\sigma^2 + \omega^2}$

Phase of $s = \angle s = \phi = \tan^{-1}\left(\frac{\omega}{\sigma}\right)$

A FUNCTION of a complex variable maps every complex number into another complex number.

e.g., $G(s) = \dfrac{s+1}{(s+2)(s+3)}$

The values of "s" for which $G(s) = 0$ are called "ZERO"s of G.

The values of "s" for which $G(s)$ becomes infinite are called "POLES" of G.

For the above G, zero is at : $s = -1$

poles are at: $s = -2$ and $s = -3$

If $G(s) = \dfrac{s-10}{2(s^2 + as + b)}$, zero is at $s = 10$,

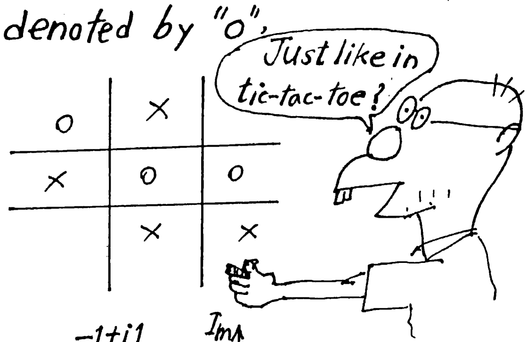
poles at: $s = \dfrac{-a \pm \sqrt{a^2 - 4b}}{2}$

(Can be real or complex).

© 1990 S. M. Joshi

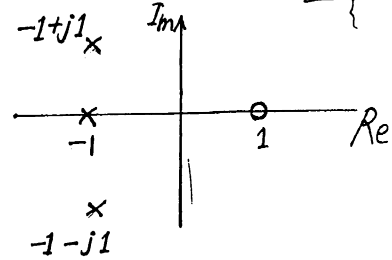
Zeros are denoted by "o",

and poles by "x".

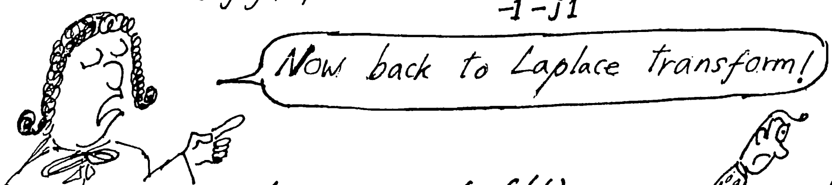


e.g.,

$$\mathcal{L}(s) = \frac{s-1}{(s+1)(s^2+2s+2)} \Rightarrow$$

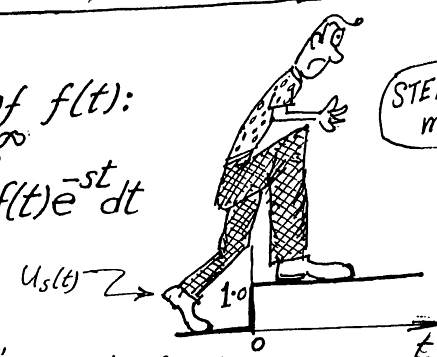


(The nice thing about complex poles or zeros is that they always occur in conjugate pairs.)



<u>Def.</u>:- Laplace transform of $f(t)$:

$$\mathcal{L}[f(t)] \triangleq F(s) = \int_0^\infty f(t) e^{-st} dt$$



For example, if

$f(t) = U_s(t)$ (i.e., "unit step function)

$$\Rightarrow \mathcal{L}[f(t)] = \int_0^\infty 1 \cdot e^{-st} dt = \frac{1}{s}$$

Also $\mathcal{L}[e^{at}] = \int_0^\infty e^{-(s-a)t} dt = \frac{1}{s-a}$

© 1990 S. M. Joshi

Similarly, $\mathcal{L}[t] = \frac{1}{s^2}$

(**Note**: all our functions are ZERO for negative time ($t<0$).

i.e., $\mathcal{L}[k_1 f_1(t) + k_2 f_2(t)] = k_1 F_1(s) + k_2 F_2(s)$

Also

$$\mathcal{L}\left[\frac{d^n f(t)}{dt^n}\right] = s^n F(s) - s^{n-1} f(0) - s^{n-2} f'(0) - \cdots - f^{(n-1)}(0)$$

("f^{ℓ}" denotes $d^{\ell} f/dt^{\ell}$).

Q. Why do control engineers just LOVE Laplace transform?

A. Because...

It helps them analyze LTI systems easily.

<u>Example</u>: R-C Network on p.7:

State-space model

$$\begin{cases} \frac{dv_c}{dt} = \left[-\frac{1}{RC}\right] v_c + \left[\frac{1}{RC}\right] v_{in} & (1) \\ (\dot{x} = [A]x + [B]u \\ y = v_c = [1] v_c + [0] v_{in} \\ (y = Cx + Du) \end{cases}$$

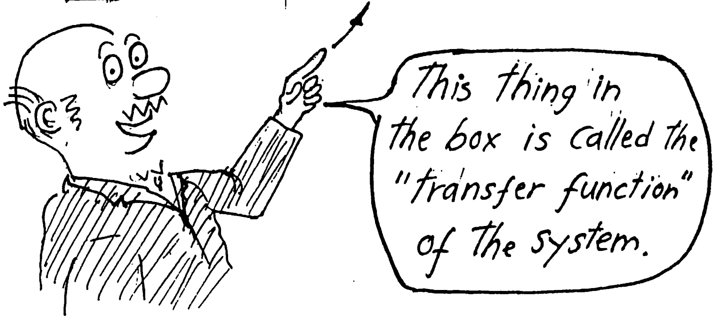
<u>Problem</u>: Suppose $v_c(0) = 3$ volts, and v_{in} = unit step. Find $v_c(t)$ for $t > 0$.

© 1990 S. M. Joshi

Take Laplace transform of both sides of (1):

$$sV_c(s) - v_c(0) = -\frac{1}{RC} V_c(s) + \frac{1}{RC} V_{in}(s)$$

$$\Rightarrow \boxed{Y(s) = V_c(s) = \boxed{\left(\frac{1/RC}{s + 1/RC}\right)} V_{in}(s) + \frac{1}{s + 1/RC} v_c(0)}$$



Let's denote it by $G(s)$.

Then, with ZERO initial conditions $(v_c(0) = 0)$

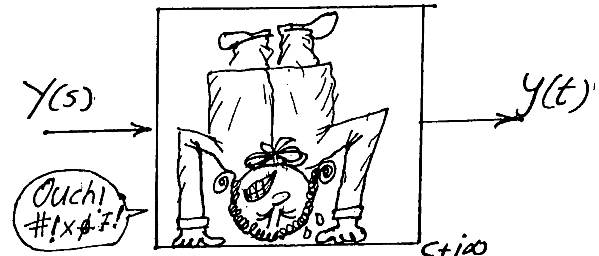
$$\boxed{Y(s) = G(s)\, U(s)}$$

So, given ANY input function $u(t)$, if we know the system's TRANSFER FUNCTION $G(s)$, we can determine the output's Laplace transform $Y(s)$.

We can then obtain $y(t)$ by "INVERSE-LAPLACE-TRANSFORMING" $Y(s)$:

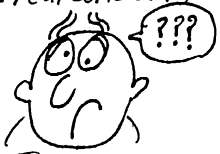
© 1990 S. M. Joshi

INVERSE LAPLACE TRANSFORM recovers the time function $y(t)$ from its transform $Y(s)$.

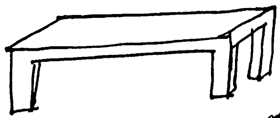


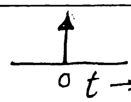
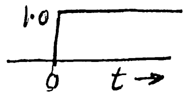
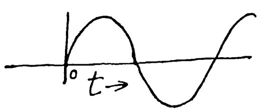
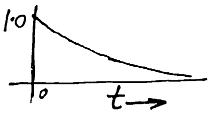
$$y(t) = \mathcal{L}^{-1}[Y(s)] = \int_{c-j\infty}^{c+j\infty} F(s)e^{st}\,ds \quad (c = \text{real constant})$$

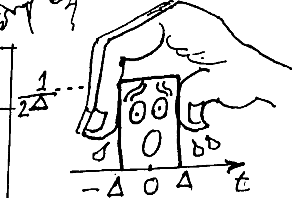
This formula is kinda hard to use!



It is easier to use a Laplace transform-table, instead. ⇒



Function	Picture	Transform
Impulse ($\delta(t)$)		1
Unit step ($U_s(t)$)		$\frac{1}{s}$
$\sin\omega t$		$\frac{\omega^2}{s^2+\omega^2}$
e^{-at}		$\frac{1}{s+a}$



Take a PULSE with UNIT area and SQUEEZE it to get the unit impulse, $\delta(t)$

It has infinite magnitude, but the area under it is 1·0

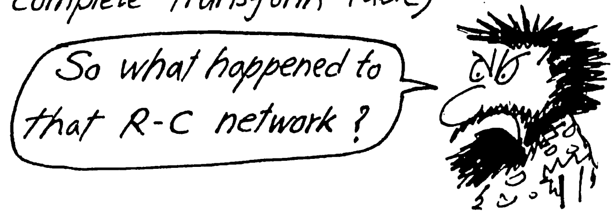
© 1990 S. M. Joshi

Other useful properties of Laplace Transform:

$$\mathcal{L}[e^{at} f(t)] = F(s-a)$$

$$\mathcal{L}\left[\int_0^t f(\sigma)\,d\sigma\right] = \frac{F(s)}{s}$$

(See any controls textbook (e.g., "Automatic Control Systems" by B.C. Kuo, Prentice-Hall, 1982) for a complete transform table)



Since $v_{in}(t) = U_s(t)$, $V_{in}(s) = \frac{1}{s}$

$$\therefore \quad Y(s) = \frac{1/RC}{s + 1/RC} \cdot \frac{1}{s} + \frac{1}{s + 1/RC}\,\overset{3\text{ volts}}{v_c(0)}$$

Suppose $R = \frac{1}{2}$ Ohm $\quad C = 1$ Farad. Then

$$Y(s) = \left(\frac{2}{s+2}\right)\left(\frac{1}{s}\right) + \frac{3}{s+2}$$

Resolve into partial fractions

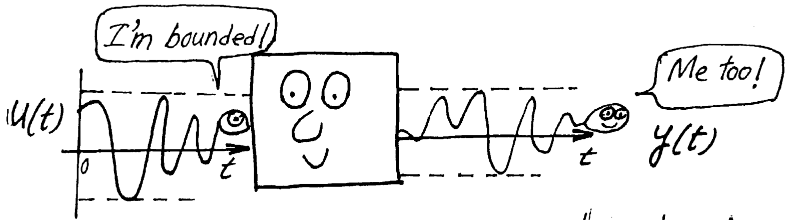
$$= \frac{-1}{s+2} + \frac{1}{s} + \frac{3}{s+2} = \frac{1}{s} + \frac{2}{s+2}$$

From the table, $\mathcal{L}^{-1}\left[\frac{1}{s+2}\right] = e^{-2t}$, and $\mathcal{L}^{-1}\left[\frac{1}{s}\right] = U_s(t) = 1$

$$\therefore \quad \boxed{y(t) = 1 + 2e^{-2t}}$$

© 1990 S. M. Joshi

WHAT IS STABILITY?



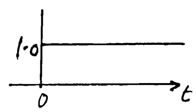
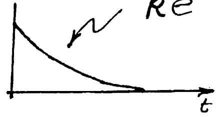
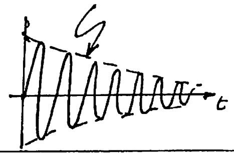
$$|u(t)| \le u_{MAX} \text{ (finite)} \Rightarrow |y(t)| \le y_{MAX} \text{ (finite)}$$

A system is said to be STABLE if EVERY bounded input produces a bounded output.

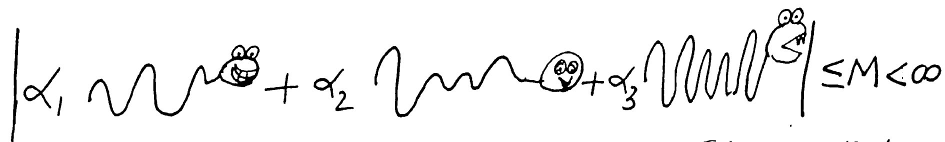
That's "bounded-input, bounded-output" or "BIBO"-stability.

(There are other kinds of stability, but let's not discuss them here).

What are bounded signals?

		ke^{-at}	$e^{-\rho\omega_0 t}\sin\omega_0 t$
$u(t) \rightarrow$	$u_s(t)$ 		
$U(s) \rightarrow$	$\frac{1}{s}$	$\frac{k}{s+a}$	$\frac{\omega_0^2}{s^2+2\rho\omega_0 s+(1+\rho^2)\omega_0^2}$

All linear combinations of bounded signals are bounded.

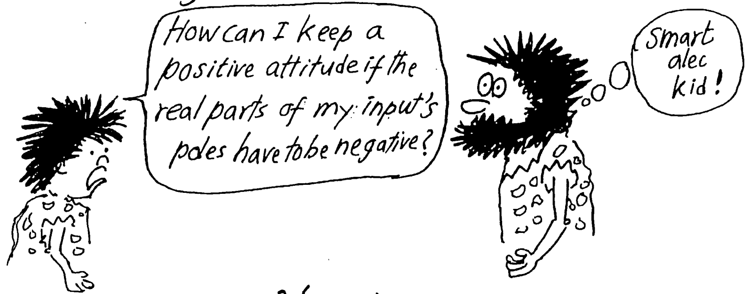


($\alpha_1, \alpha_2, \alpha_3$ finite)

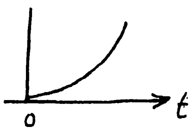
© 1990 S. M. Joshi

Notice the poles of all those* U(s) have NEGATIVE or ZERO real parts. (* The ones on previous page)

Unbounded signals have poles with POSITIVE real parts.



e.g., $f(t) = e^{2t}$



$\Rightarrow F(s) = \frac{1}{s-2}$ (Pole at $s = +2$)

$$f(t) = e^{t} \sin(2t) \Rightarrow F(s) = \frac{\omega_0^2}{s^2 - 2s + 5}$$

$\Rightarrow$ poles at $s = 1 + j4$ and $s = 1 - j4$

(i.e., real parts of poles are POSITIVE.)

So when you input a bounded signal $u(t)$, the poles of $U(s)$ all have negative or zero real parts.

Therefore,
$$Y(s) = G(s)\,U(s) = \underbrace{\left[\frac{(s-z_1)(s-z_2)\cdots}{(s-p_1)(s-p_2)\cdots}\right]}_{G(s)} \underbrace{\left[\frac{(s-zu_1)(s-zu_2)\cdots}{(s-pu_1)(s-pu_2)\cdots}\right]}_{U(s)}$$

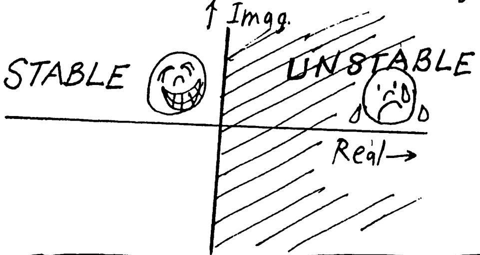
Resolve into partial fractions:

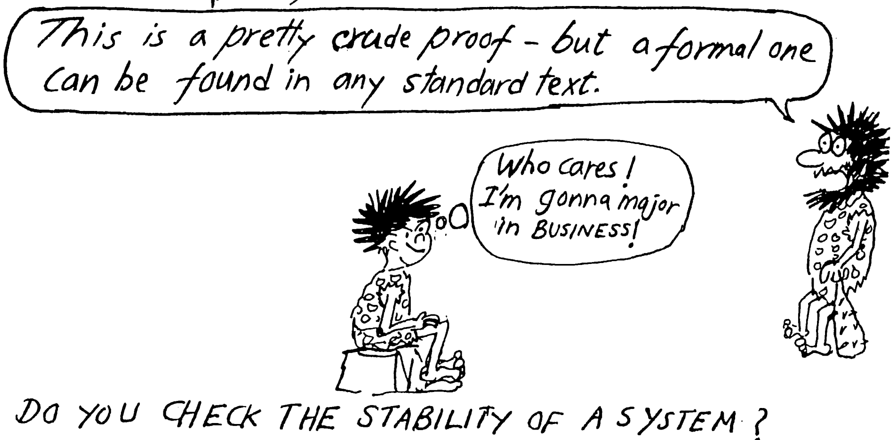
$$Y(s) = \frac{K_{p_1}}{s-p_1} + \cdots + \frac{K_{p_\ell}}{s-p_\ell} + \frac{K_{u_1}}{s-pu_1} + \cdots + \frac{K_{u_m}}{s-pu_m}$$

© 1990 S. M. Joshi

If $u(t)$ is bounded, real parts of p_{u_i} are all ≤ 0.
So, if real parts of p_i are < 0, then $y(t)$ will be bounded. Therefore,

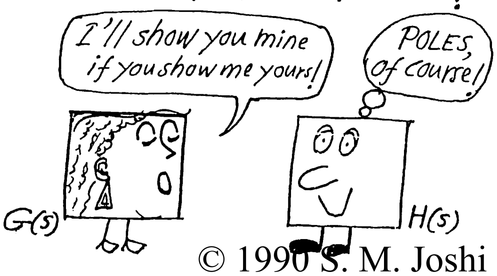
A system is STABLE if and only if all its poles have negative real parts, i.e., all its poles are in the left-half of the complex plane.



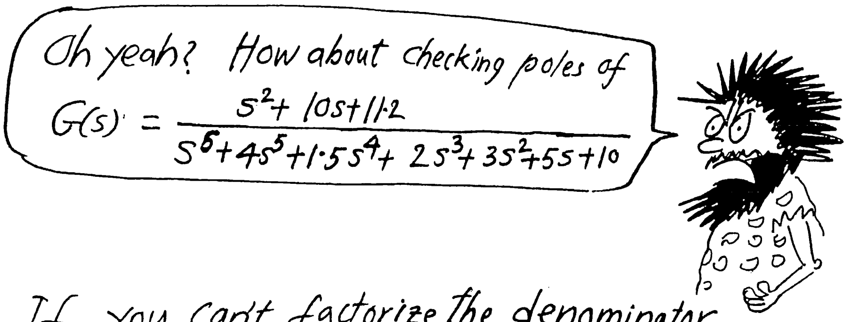


Q. HOW DO YOU CHECK THE STABILITY OF A SYSTEM?

A. Look at its poles.



© 1990 S. M. Joshi

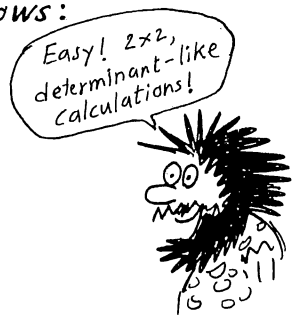


If you can't factorize the denominator, do not despair - the ROUTH-HURWITZ criterion is here!

Different versions of this were developed by E. J. Routh and A. Hurwitz during late 19th century to check the roots of a polynomial.

Suppose $p(s) = 2s^3 + 3s^2 + s + 5$

Form the "Routh array" as follows:



No. of sign changes				
	s^3	2	1	0
0	s^2	3	5	0
1	s^1	$\frac{3\times 1 - 5\times 2}{3} = -\frac{7}{3}$	$\frac{3\times 0 - 0\times 2}{3} = 0$	0
1	s^0	$\frac{(-\frac{7}{3})\times 5 - 0\times 3}{(-\frac{7}{3})} = 5$	$\frac{(-\frac{7}{3})\times 0 - 0\times 3}{(-\frac{7}{3})} = 0$	

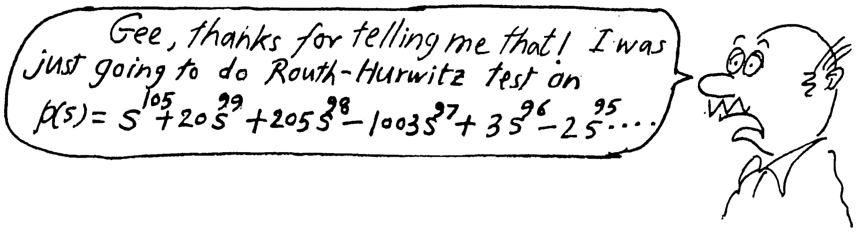
(2 sign changes)

No. of roots of $p(s)$ in the right-half-plane = No. of sign changes in the 1st column.

Note: There are some tricks for handling "0"s in the first column. Look for them in a standard text.

 © 1990 S. M. Joshi

If one or more coefficients of $p(s)$ are ≤ 0, then there is no need to do the Routh-Hurwitz test - $p(s)$ WILL have at least one root in the right-half-plane.



A word about **CAUSALITY** (not "Ca<u>su</u>alty"!)

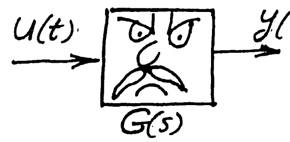
A system is called "causal" if it responds ONLY <u>AFTER</u> receiving an input - never before!

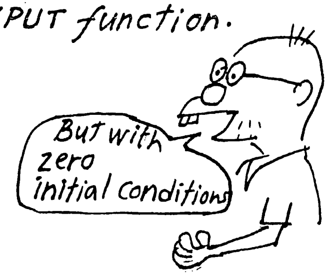
All physically realizable systems are causal.

 © 1990 S. M. Joshi

FREQUENCY RESPONSE

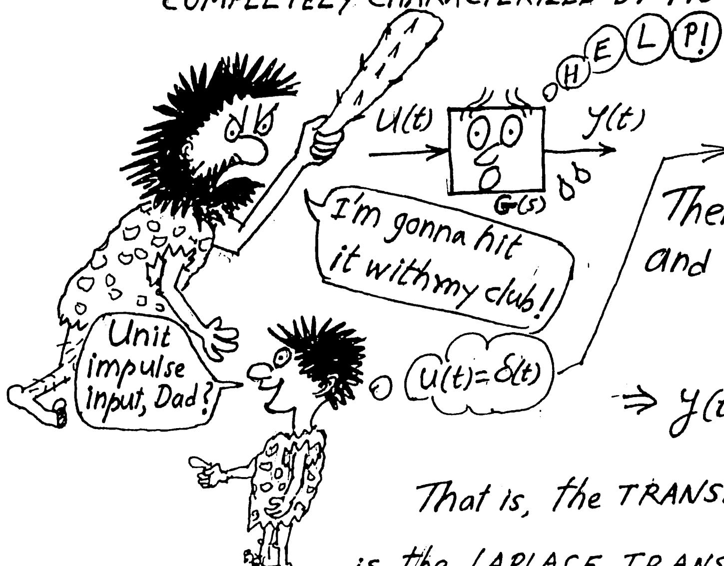
The TRANSFER FUNCTION lets us calculate the RESPONSE for any given INPUT function.





$$Y(s) = G(s)\,U(s)$$

A SYSTEM'S INPUT/OUTPUT BEHAVIOR IS COMPLETELY CHARACTERIZED BY ITS TRANSFER FUNCTION.



Suppose we input a unit impulse.

$$u(t) = \delta(t)$$

Then $U(s) = 1$

and

$$Y(s) = G(s)\cdot 1 = G(s)$$

$$\Rightarrow y(t) = \mathcal{L}^{-1}[G(s)] \triangleq g(t)$$

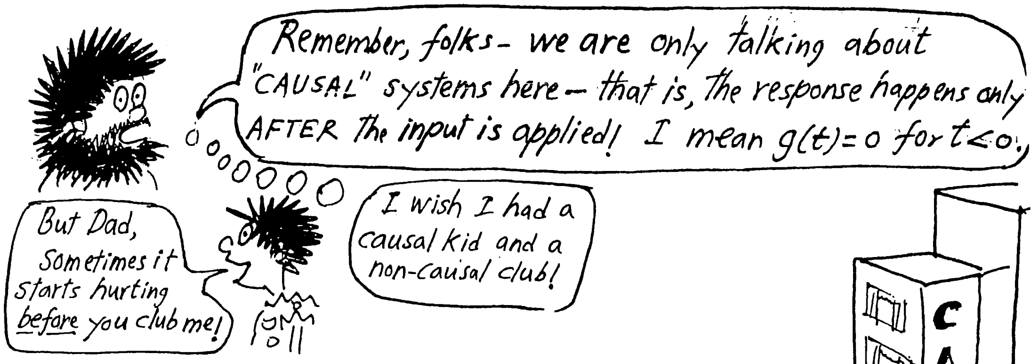
That is, the TRANSFER FUNCTION is the LAPLACE TRANSFORM OF THE SYSTEM'S "IMPULSE RESPONSE"!

<u>EXAMPLES</u>: $G(s) = \frac{3}{s+5} \Rightarrow g(t) = 3e^{-5t}$

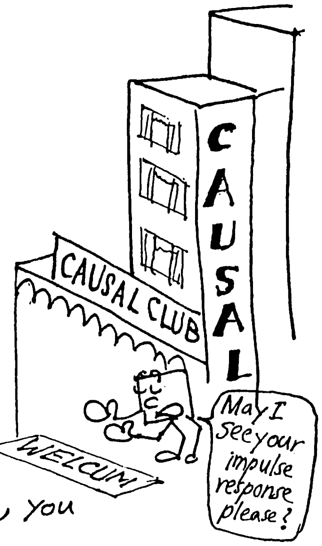
$$G(s) = \frac{\omega_0^2}{s^2+\omega_0^2} \Rightarrow g(t) = \sin(\omega_0 t)$$

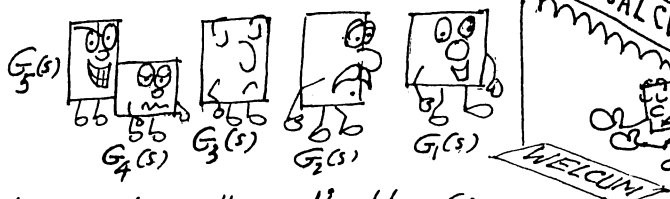
$$G(s) = \frac{1}{s^2} \Rightarrow g(t) = t$$

© 1990 S. M. Joshi



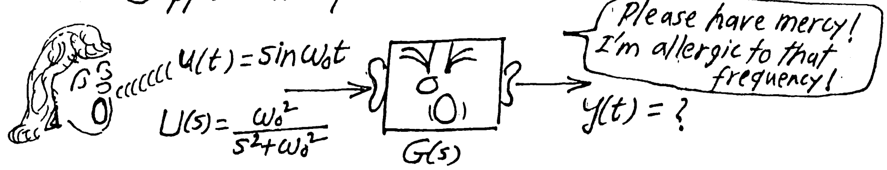
Speaking of clubs, all physical systems belong to the CAUSAL CLUB.





A system is physically realizable (i.e., you can actually build it) only if it is CAUSAL.

Suppose we put in a sine wave input signal.



$$Y(s) = G(s) \cdot \frac{\omega_0^2}{s^2+\omega_0^2}.$$ It can be shown that,

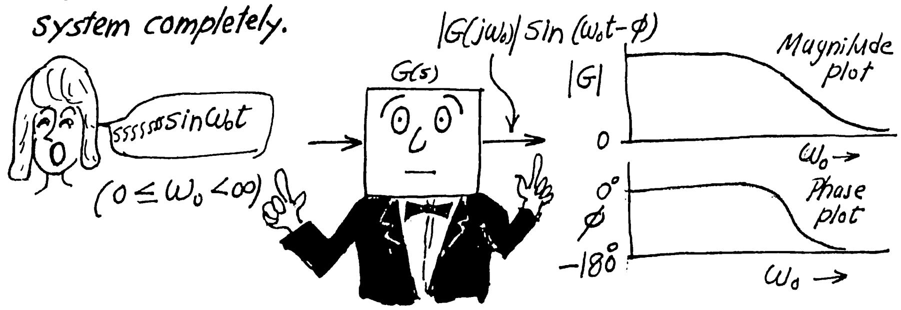
for large t, $$y(t) = |G(j\omega_0)| \sin(\omega_0 t + \phi)$$

where $|G(j\omega_0)|$ = magnitude of $G(j\omega_0) = \sqrt{(\text{Real part})^2 + (\text{Imag. pt.})^2}$

and ϕ = phase of $G(j\omega_0) = \tan^{-1}\left[\frac{\text{Imag. part of } G(j\omega_0)}{\text{Real part of } G(j\omega_0)}\right]$

© 1990 S. M. Joshi

Starting at zero frequency, if we plot the output's magnitude and phase, that too characterizes the system completely.



The plots of $|G(j\omega)|$ and ϕ, versus ω, are called the "frequency response" of the system.

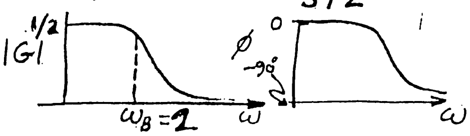
The frequency at which the magnitude starts to "roll off" (below 70.7% of the max. value) is called "bandwidth".

The part of input signal with frequencies $< \omega_B$ "passes" through the system.

The part of input signal with frequencies $> \omega_B$ is ATTENUATED or blocked.

Example: $G(s) = \frac{1}{s+2} \Rightarrow |G(j\omega)| = \left|\frac{1}{j\omega+2}\right| = \frac{1}{\sqrt{\omega^2+4}}$

$\phi(\omega) = \tan^{-1}(\omega/2)$



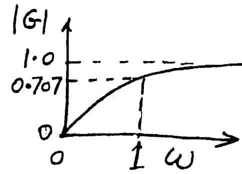
© 1990 S. M. Joshi

The idea of bandwidth is pretty common in signals and communication theory. A "low-pass filter" allows low frequencies and blocks high frequencies.

$$\frac{1}{s+1}\ ,\ \frac{1}{(s+1)(s+5)}\ ,\ \frac{s+3}{s^2+as+b}$$

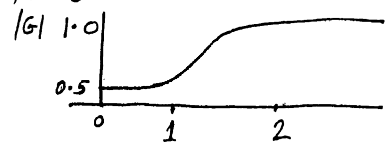
are all low-pass filters. As $s\,(=j\omega)\to\infty$, these functions $\to 0$.

How about $G(s)=\dfrac{s}{s+1}$?



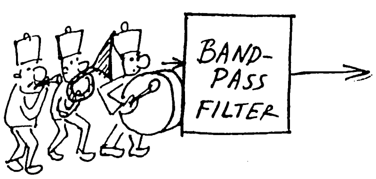
This is a HIGH-PASS filter.

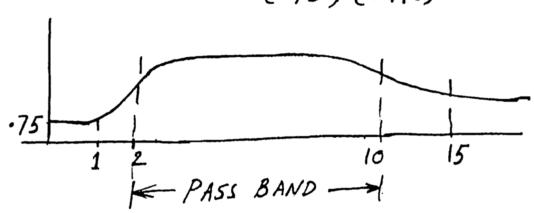
So is $G=\dfrac{s+1}{s+2}$.



and $G=\dfrac{(s+1)(s+2)}{(s+10)(s+20)}$.

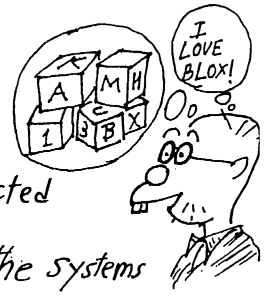
And then there are BAND-PASS filters: $G=\dfrac{(s+1)(s+15)}{(s+2)(s+10)}$





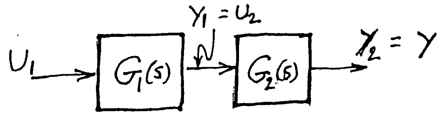
© 1990 S. M. Joshi

BLOCK DIAGRAMS



Two or more systems can be connected together to form a new system. If the systems are linear & time-invariant (LTI), each can be represented by its transfer function. The resulting block diagram can be manipulated fairly easily for analyzing the composite system. Three basic types of connections are:

1. SERIES Connection



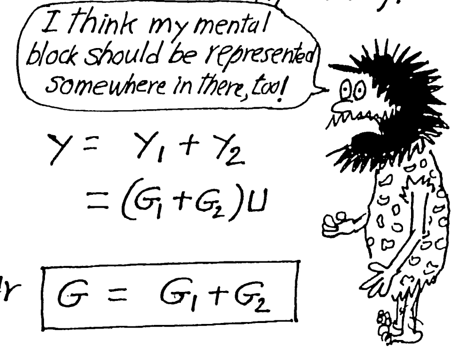
$$Y(s) = G_2(s)\,Y_1(s)$$

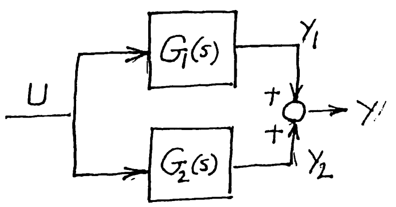
$$\therefore \quad Y(s) = \left[G_2(s)\,G_1(s)\right]U(s)$$

i.e., Equivalent transfer function is: $\boxed{G(s) = G_2(s)G_1(s)}$

(Note that if G_1, G_2 are MATRICES instead of scalars, $G_2G_1 \neq G_1G_2$; i.e, the order would be important).

2. PARALLEL Connection



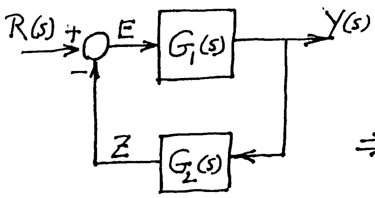


$$Y = Y_1 + Y_2$$
$$= (G_1 + G_2)U$$

or $\boxed{G = G_1 + G_2}$

© 1990 S. M. Joshi

3. FEEDBACK Connection

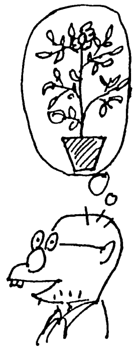


$$Y = G_1 E \quad ; \quad Z = G_2 Y$$

$$E = R - Z = R - G_2 Y$$

$$\Rightarrow \quad Y = G_1 (R - G_2 Y)$$

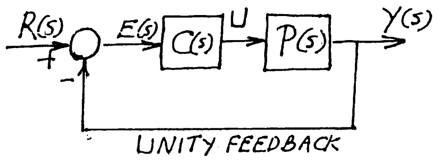
$$\Rightarrow \quad Y = \frac{G_1}{1 + G_1 G_2} R$$



$$\therefore \quad \boxed{G = \frac{G_1}{1 + G_1 G_2}}$$

(NOTE: Control folks like to refer to the system to be controlled as the "PLANT").

Another variation of this:



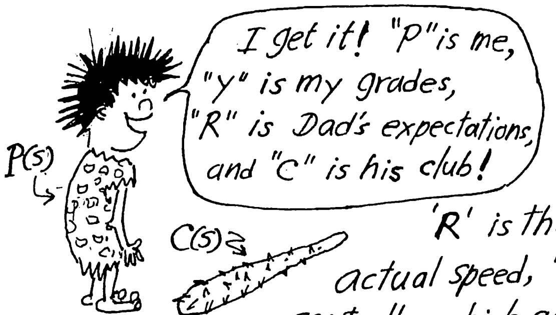
$P(s)$: "Plant"

$C(s)$: Controller

$R(s)$: Reference input (Desired behavior)

$Y(s)$: output (actual behavior)

$E(s)$: Error signal



In a speed control system, 'R' is the speed setpoint, 'y' is the actual speed, 'P' is the car, and 'C' is the controller which generates the required throttle (gas pedal) input 'U'. In this case,

$$G = \frac{PC}{1 + PC} \quad , \; (\text{where } Y(s) = G(s) R(s))$$

The equivalent transfer function G for feedback connection is called the "Closed-loop" transfer function.

© 1990 S. M. Joshi

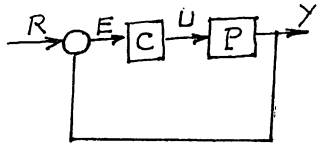
Feedback is important!

To maintain the correct setpoint we must continuously SENSE the actual output and compare it with the desired output. The controller (or compensator) then generates the corrective action based on the ERROR. This is called "closed-loop" or "feedback" control.

The other way to do it is "open-loop" control, where we pre-calculate the input u(t) [as a function of time] which will generate the desired output "y(t)". However, a small error in the math model of the plant, or a small unforeseen disturbance can make things go haywire!

© 1990 S. M. Joshi

For the closed-loop system:



$$G = \frac{PC}{1+PC}$$

is the "closed-loop" transfer function.

$(1+PC)$ is called the "return difference".

It can be seen that $E = (1+PC)^{-1}R$

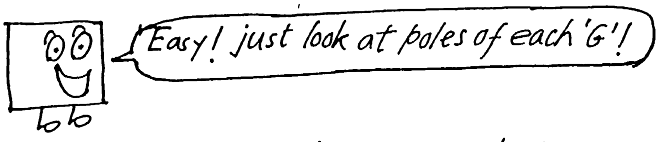
$(1+PC)^{-1}$ is denoted as "S", and is called the "sensitivity", or the "inverse return difference. Also,

$$S(s) + G(s) = 1$$

So much for jargon!

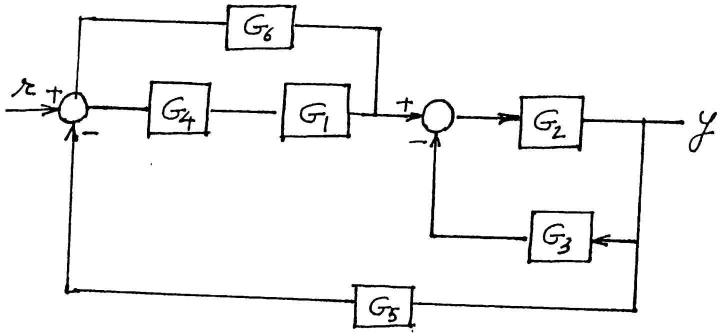
Anyways, we should look at the "closed-loop" behavior rather than the "open-loop" one.

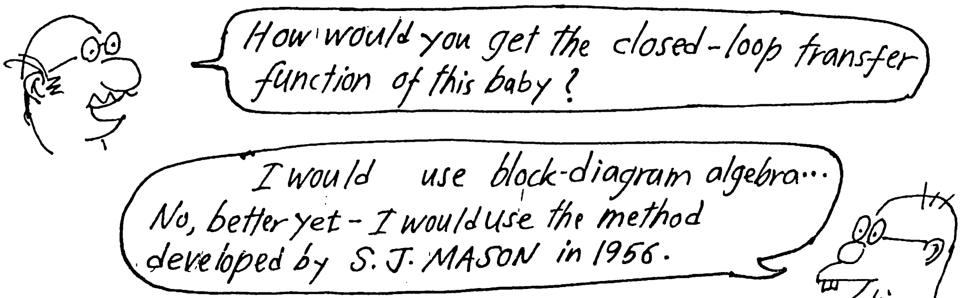
First of all, is the feedback system STABLE? (The <u>series</u>- and <u>parallel</u>- connected systems are stable if individual G_1 and G_2 are stable).



But feedback systems require more analysis. General systems consisting of many blocks are even more complicated.

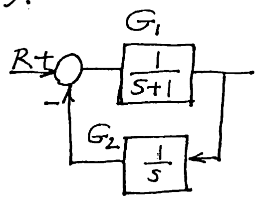
© 1990 S. M. Joshi





(This may be found in standard controls texts under the title "Signal flow graphs").

For the simple feedback system:



$$G = \frac{G_1}{1+G_1G_2} = \frac{\frac{1}{s+1}}{1+\frac{1}{s(s+1)}}$$

$$= \frac{s}{s^2+s+1}$$

The POLES of G are at: $-\frac{1}{2} \pm j\frac{\sqrt{3}}{2}$
i.e., they are inside the open left-half plane
("Open" excludes the imaginary axis) $\Rightarrow$ STABLE.

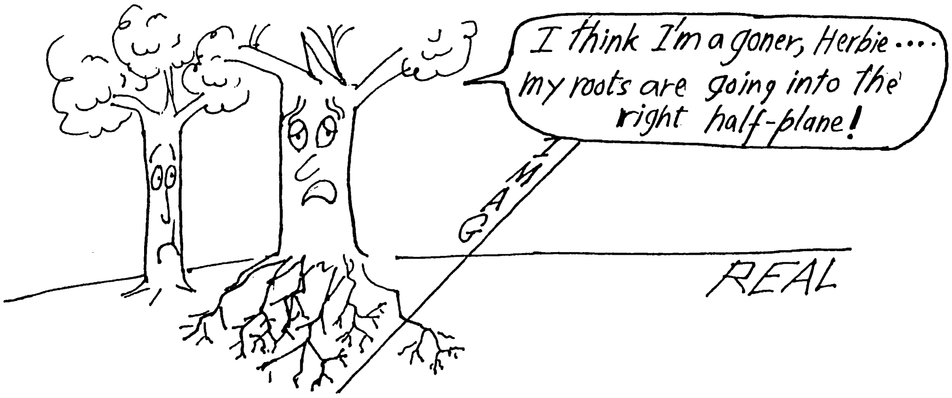
© 1990 S. M. Joshi

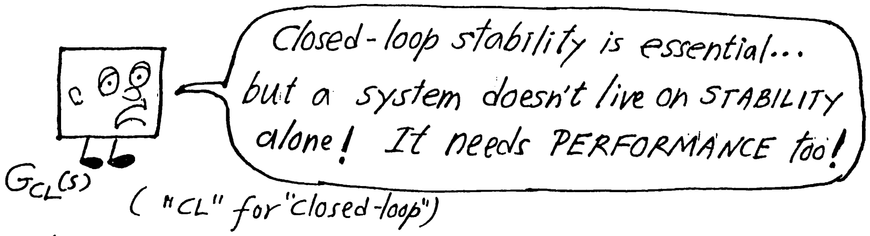
The same method can be used for more complicated systems. i.e.,

1) Find the closed-loop transfer function

(Use Mason's formula if necessary).

2) Check if all the roots of the DENOMINATOR polynomial are inside the left half of the complex plane.



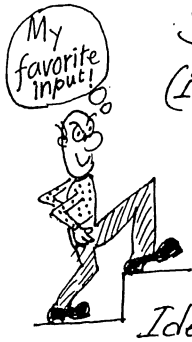


("CL" for "closed-loop")

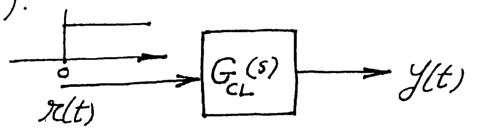
A good control system responds FAST and WITHOUT TOO MANY WIGGLES.

Pole locations indicate the nature of the response.

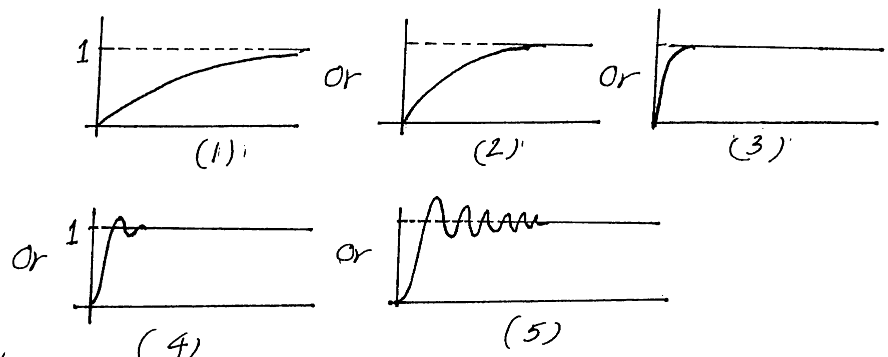
© 1990 S. M. Joshi

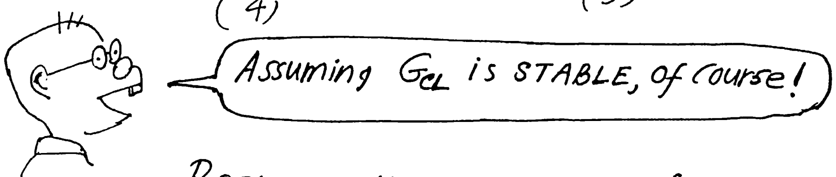


Suppose we put in a UNIT STEP command; (i.e., increase the reference input $r(t)$ suddenly): from $\underline{0}$ to $\underline{1}$).



Ideally the response $y(t)$ should also be the unit step. But the actual response may look like:





Responses (1), (2) and (3) are from an "Overdamped" system. (i.e., the closed-loop poles are REAL, so we get something like this:

$$y(t) = 1 - \alpha_1 e^{-p_1 t} - \alpha_2 e^{-p_2 t} \ldots$$

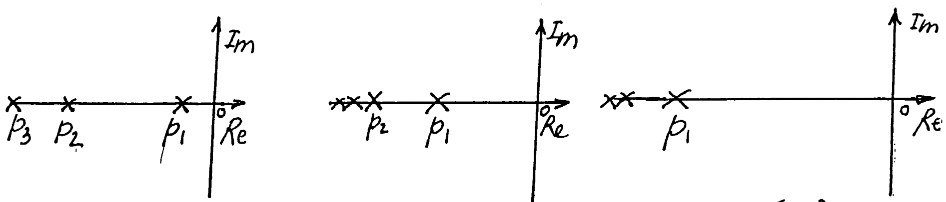
If some closed-loop poles are COMPLEX, we have responses like (4) and (5).

© 1990 S. M. Joshi

The *more negative* the real part of a pole, the *faster* is the decay of its contribution to $y(t)$.

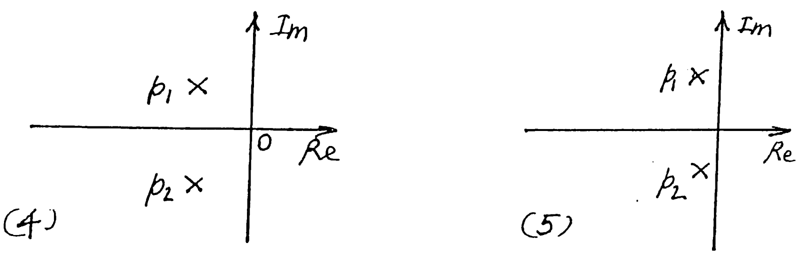
Thus, the poles which are CLOSEST to the imaginary axis dominate the transient response.

The pole locations for cases (1), (2) & (3) may look like this:



(1) slow (2) fast (3) faster

The pole locations corresponding to (4) & (5) may look like this:



(4) (5)

Suppose the denominator of $G_{CL}(s)$ is:

$$s^2 + 2\rho\omega s + \omega^2 \quad (\text{where } \rho < 1)$$

© 1990 S. M. Joshi

i.e., the poles are at: $s = -\rho\omega \pm j\sqrt{1-\rho^2}\,\omega$

$\Rightarrow \quad |p_1| = |p_2| = \omega$

Then $\quad \rho = -\dfrac{Re[p_1]}{|p_1|}$

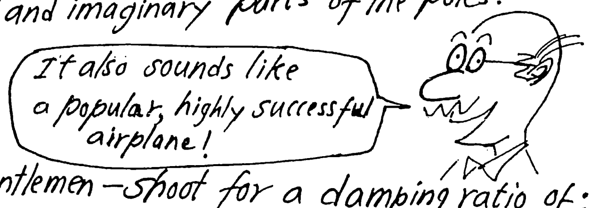
"ρ" is called the DAMPING RATIO (because it determines how fast the wiggles in the response dampout).

Small ρ $\Rightarrow$ p_1, p_2 are close to imaginary axis
(e.g., $\rho = 0.1$) $\Rightarrow$ more oscillatory response (like (5)).

Medium ρ $\Rightarrow$ less oscillatory response (like (4)).
(e.g., $\rho = \frac{1}{\sqrt{2}} = 0.707$)

Large ρ $\Rightarrow$ non-oscillatory response.
(e.g., $\rho \geq 1$)

Controls folks like to have "$\rho = 0.707$". This is the ideal they shoot for. It gives a fast response with very small overshoot and wiggles. It corresponds to equal real and imaginary parts of the poles.



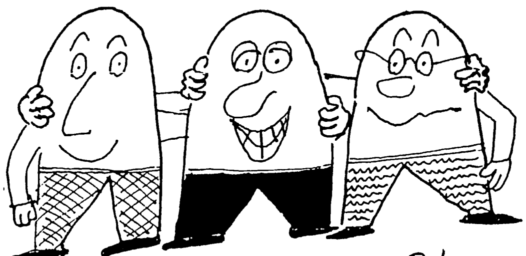
So ladies and gentlemen—shoot for a damping ratio of:
0.707.

© 1990 S. M. Joshi

There is a direct relationship between

TIME-RESPONSE, FREQUENCY-RESPONSE, and POLE-LOCATION.

of the system.



Time response　　Frequency response　　Pole locations

HIGH BANDWIDTH Corresponds to:

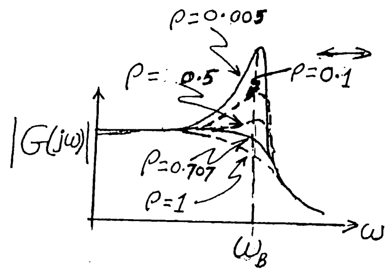
$\leftrightarrow$ poles farther left of the imag. axis

$\leftrightarrow$ fast time-response

SMALL DAMPING RATIO $\Rightarrow$

$\leftrightarrow$ Oscillatory time-response

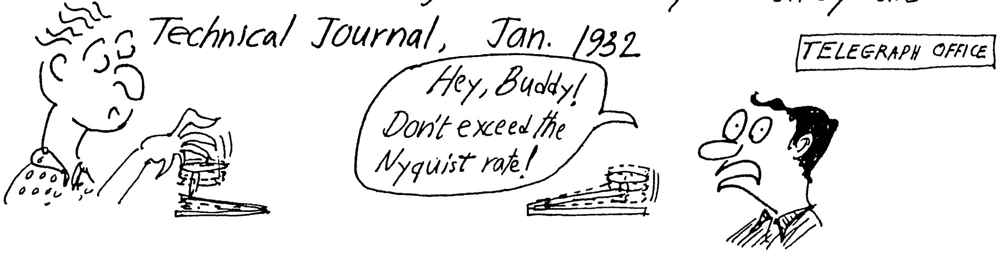
$\leftrightarrow$ Freq. response with higher peaks.



© 1990 S. M. Joshi

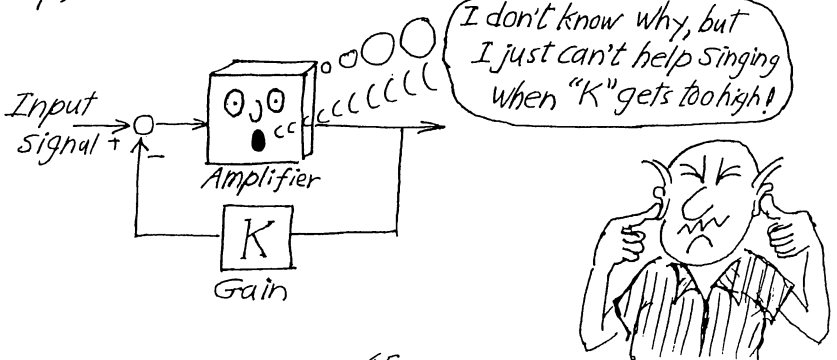
NYQUIST STABILITY CRITERION

H. Nyquist: "Regeneration Theory". Bell Systems Technical Journal, Jan. 1932

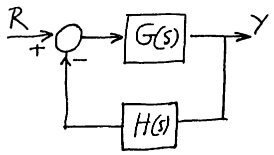


HARRY NYQUIST (1889–1976), a Bell Labs researcher, was one of the greatest contributors to communication and control theory. In 1928 he discovered the theoretical upper limit on the transmission rate for telegraph pulse transmission (NYQUIST rate).

In 1932, while studying why some feedback amplifiers "sang" when the feedback gain was turned up, he discovered the NYQUIST CRITERION for stability.



© 1990 S. M. Joshi

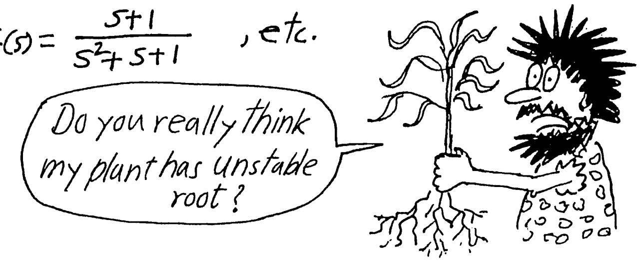


$$G_{CL} = \frac{G}{1+GH}$$

To check STABILITY, we need to check the roots of:

$\psi(s) = 1 + G(s)H(s)$. [G and H are 'rational functions', i.e., they are ratios of polynomials in 's';

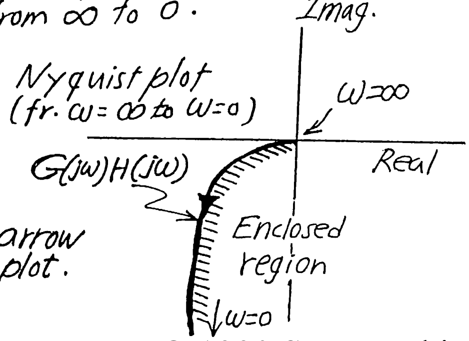
e.g., $G(s) = \frac{s+1}{s^2+s+1}$, etc.



Roots of $\psi(s)$ are those values of s at which $\psi(s) = 0$, or $G(s)H(s) = -1$.

A NYQUIST PLOT is the graph of the "loop gain" $G(s)H(s)$ when s varies along a certain closed contour in the complex plane. Of particular interest is the case when $s = j\omega$ and ω varies from ∞ to 0.

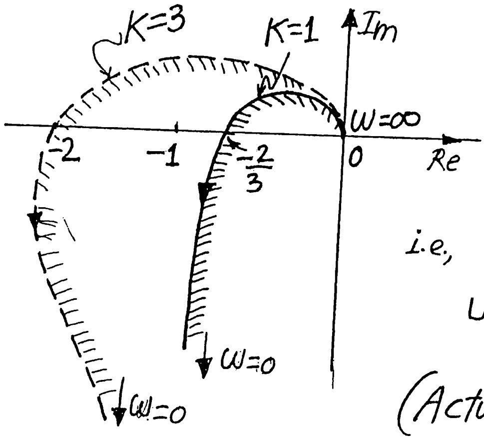
e.g., $\left.\begin{array}{l} G(s) = \frac{1}{s+1} \\ H(s) = \frac{1}{s} \end{array}\right\} \Rightarrow$



The region to the left of the arrow is said to be ENCLOSED by the plot.

© 1990 S. M. Joshi

NYQUIST CRITERION: Suppose G and H have no poles in the open right half-plane (i.e., they are stable except possibly for poles on the imaginary axis). Then the closed-loop system is stable if and only if the Nyquist path does not enclose the point "$-1+j0$".



Example: $G = \dfrac{4}{s(s+2)}$

$H = \dfrac{K}{s+1}$

i.e., STABLE for $K = 1$

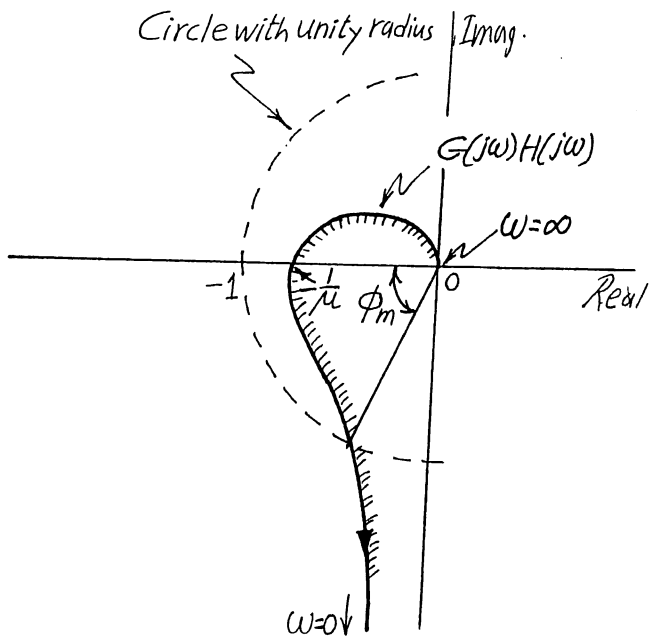
UNSTABLE for $K = 3$

(Actually it's stable for $0 < K < \frac{3}{2}$)

© 1990 S. M. Joshi

Nyquist criterion *also* addresses the question:

"*How stable* is it?" (i.e., degree of stability)



This system is stable.

We can make it unstable by *either*

a) Increasing the "gain" by "μ"
(So μ is called the "GAIN MARGIN")

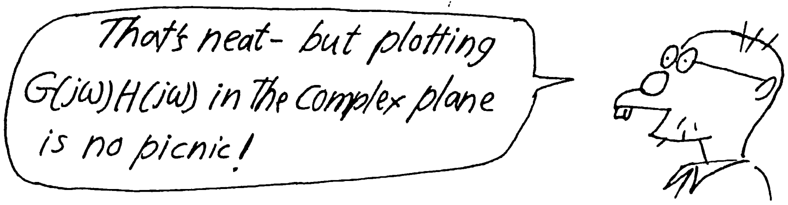
Or b) Reducing the phase angle by ϕ_m° (i.e., adding a "phase lag*" of ϕ_m° to GH)
(ϕ_m is called the "PHASE MARGIN").

*("lag" means negative phase).

© 1990 S. M. Joshi

The larger the gain- and phase-margins, the more 'ROBUST' the system is to errors in the system's math model.

Another measure of robustness is the smallest distance of the $G(j\omega)H(j\omega)$ plot from the point: "-1", i.e., $\min_{\omega}|1+GH|$.



However, if the numerator and denominator of GH are in FACTORED form, we can calculate $|G(j\omega)H(j\omega)|$ and $\angle G(j\omega)H(j\omega)$ rather easily, as shown by HENDRIK W. BODE (1905-1982), another Bell Labs scientist (and later a professor at Harvard).

H. W. Bode : "Network Analysis and Feedback Amplifier Design" (book). Van Nostrand, New York, 1945. (A classic book, it is still in print after 45 years!)

(Bode is generally pronounced as "Boh-dee", however, the original Dutch pronounciation is "Boh-dah").

© 1990 S. M. Joshi

BODE PLOTS

Suppose

$$G(s)H(s) = \frac{K(s+z_1)(s+z_1)}{(s+p_1)(s+p_2)} = \underbrace{\left(\frac{Kz_1z_2}{p_1p_2}\right)}_{K_1 \text{ (always real)}} \frac{(1+N_1s)(1+N_2s)}{(1+D_1s)(1+D_2s)}$$

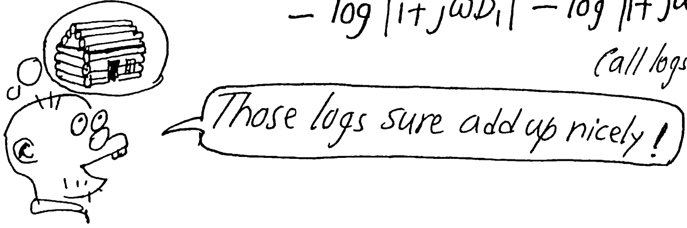
$N_1 = 1/z_1$, $D_1 = 1/p_1$, etc.

(real or complex).

Then, for $s = j\omega$ $\quad |GH| = |K_1| \cdot |1+N_1s| \cdot |1+N_2s| / |1+D_1s||1+D_2s|$

$$\Rightarrow \log|G(j\omega)H(j\omega)| = \log K_1 + \log|1+j\omega N_1| + \log|1+j\omega N_2| - \log|1+j\omega D_1| - \log|1+j\omega D_2|$$

(all logs are base 10)



and the phase angle $\angle GH = \angle 1+j\omega N_1 + \angle 1+j\omega N_2 - \angle 1+j\omega D_1 - \angle 1+j\omega D_2$

Consider a factor: $(1+Ns)$ where N is real.

Then $|1+j\omega N| = \sqrt{1+N^2\omega^2}$

LOW FREQ. APPROX.: $\omega \ll 1/N \Rightarrow |1+j\omega N| \simeq 1$

$\Rightarrow \log|1+j\omega N| \simeq 0$

HIGH-FREQ. APPROX: $\omega \gg 1/N \Rightarrow |1+j\omega N| \simeq \omega N$

$\Rightarrow \log|1+j\omega N| \simeq \log(\omega N)$

© 1990 S. M. Joshi

For the high frequency approximation,

$$\log|1+j\omega N| \simeq \log(\omega N)$$

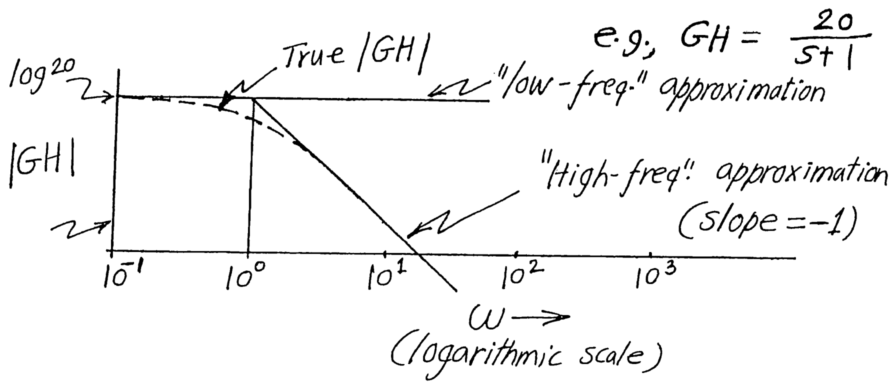
If "ω" is increased 10-fold,

$$\log(10\omega N) = \log(10) + \log(\omega N) = 1 + \log(\omega N)$$

i.e., "$\log(\omega N)$" increases by unity

i.e., on the logarithmic scale, the "slope" of the high-freq. approximation is: 1 unit (per decade in frequency ω). The high-freq. approximation intersects the "low freq. approximation" (i.e., the 0 line at: $\log(\omega N)=0$, or $\omega = 1/N$.

If "$(1+Ns)$" is in the denominator, the slope of the high-freq. approx. would be "-1".



© 1990 S. M. Joshi

Communications engineers like to express magnitude in "decibels" (denoted dB).

i.e., $|GH|_{dB} \triangleq 20 \log|GH|$

So the "slope" of "-1" per decade becomes "-20 dB per decade".

All this makes life easy – by converting tough complex MULTIPLICATIONS and DIVISIONS into tame ADDITIONS and SUBTRACTIONS!

© 1990 S. M. Joshi

For example, let $GH = \dfrac{100\,(s+1)(s+200)}{(s+10)(s+1000)}$

$= \dfrac{2\,(1+s)(1+s/200)}{(1+s/10)(1+s/1000)}$

1. Mark frequency on the X-axis of a "semilog" graph paper

2. At $\omega = 0$: $|GH| = 2$ gives the low-freq. approximation (horizontal line) at: $20 \log 2$ dB.

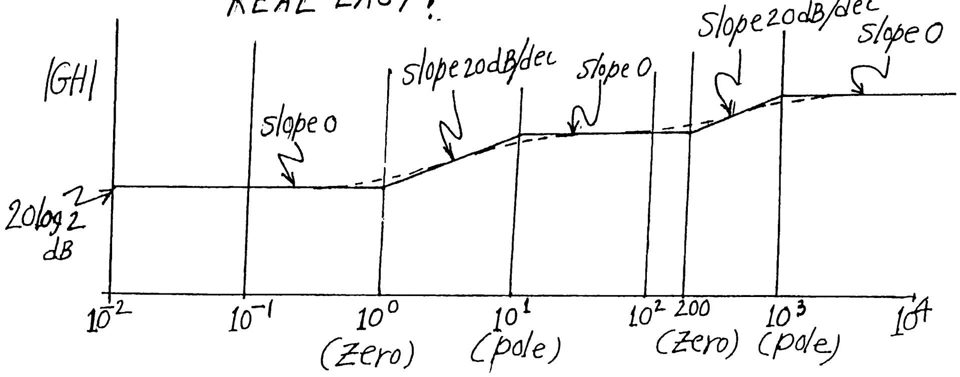
3. As ω increases, a ZERO is encountered at $\omega = 1$; so start a line sloping up at 20dB/decade

4. As ω increases further, a POLE is encountered at $\omega = 10$, so decrease the previous slope by 20dB/dec.

5. A ZERO comes next, at $\omega = 5$; increase the previous slope (zero dB/dec) by 20 dB/dec.

6. A POLE comes next, at $\omega = 10$; so decrease the previous slope by 20 dB/dec.

REAL EASY! (Finally, smooth it at corners (dashed line)).



© 1990 S. M. Joshi

The procedure is basically the same for complex poles/zeros, etc., with minor variations.

The PHASE ANGLE plot is very similar — just additions and subtractions! But it requires a bit more smoothing. No logarithms are required, so no "dB" jazz — just plain old "degrees" will do!

Bode plot can quickly give the info. necessary to draw a Nyquist plot. It also gives:

GAIN MARGIN: Negative of magnitude $|GH|_{dB}$ at the frequency (ω) where the phase becomes $-180°$; and

PHASE MARGIN: as $(\phi - 180°)$, where ϕ is the phase at the unity-gain (i.e., the zero dB) point.

© 1990 S. M. Joshi

BODE PLOTS are extensively used for SHAPING the frequency response (by designing compensator "$H(s)$") to give the required gain and phase margins (and of course the bandwidth).

More variations on this basic frequency-domain technique include: Gain-phase plots, Nichol's chart, M- circles, etc.

These "classical" methods have been successfully used for many years in many practical control system designs.

 © 1990 S. M. Joshi