Notes:

Linear System Theory

ESE 502

Slide 1

1 Linear System Theory

1.1 Overview

System:

- Vague Term
- Inputs and Outputs
- Behavior in time
- Aspects:
  - Physical system (electronic, mechanical, economic, biological, etc.)
  - Mathematical Model (usually differential or difference equations)
  - Analysis (simulation and analytical)
  - Design (simulation, analytical, and experience)

Slide 2

Notes:

Notes:

Mathematical System Descriptions

1. Internal
2. External

Analysis: From internal to external description (usually unique)

Design: From external to internal description (usually not unique)

Slide 3

Notes:

System Properties

Linearity: If B denotes the action of a "black box", $u(t)$ and $v(t)$ are inputs, and a is a constant, then

$$B(u+v) = B(u)+B(v) \quad (1)$$

$$B(au) = aB(u) \quad (2)$$

Time-Invariance: Continuous and Discrete:

Continuous: If $v(t)$ is the output for an input $u(t)$, i.e. $v(t) = B(u(t))$, then, for all t_0,

$$v(t+t_0) = B(u(t+t_0))$$

Discrete: If v_n is the output for an input u_n, i.e., $v_n = B(u_n)$, then, for all n_0,

$$v_{n+n_0} = B(u_{n+n_0})$$

Slide 4

Notes:

System Properties continued:

Causality: Future inputs can not affect present and past outputs:

If $u_1(t)$ and $u_2(t)$ are two inputs, and $v_1(t)$ and $v_2(t)$ are the corresponding outputs, then, for every t_0:

- If $u_1(t) = u_2(t)$ for all $t < t_0$, then $v_1(t) = v_2(t)$ for all $t < t_0$.
- If linear: if $u(t) = 0$ for all $t < t_0$, then $v(t) = 0$ for all $t < t_0$.
- If linear and time-invariant: if $u(t) = 0$ for $t < 0$, then $v(t) = 0$ for $t < 0$.

"Lumpedness": finite # of variables

SISO or MIMO: Single-Input, Single-Output, or Multi-Input, Multi-Output

No system is perfectly linear or time-invariant, but there are enough systems which are approximately linear and time-invariant, or which can be modelled as linear or time-invariant, that these are very useful concepts.

Slide 5

Notes:

State:

- Fundamental concept of systems
- A set of variables internal to the system, whose value at any time, t_0 , together with the inputs for time $\geq t_0$, determines the outputs for time $\geq t_0$.
- A set of initial conditions
- Usually written as a column vector
- Encapsulates total effect of all past inputs on the system
- Past inputs can affect future outputs only through the state
- Lumpedness $\Longleftrightarrow$ State is a finite set of variables

Slide 6

Notes:

Examples:

L-C circuit: figure 2.2, p.7.

- State: capacitor voltages and inductor currents (or C charges and L fluxes)
- Finite-state (description at low frequencies)
- Causal
- Time-invariant
- Linear

Unit Delay: $y(t) = u(t-1)$ (continuous-time)

- *Infinite*-state
- Causal
- Time-invariant
- Linear

Slide 7

Examples (continued):

Unit Advance: $y(t) = u(t+1)$ (continuous-time)

- "Infinite-state"
- *Not* Causal
- Time-invariant
- Linear

Unit Delay: $y(n) = u(n-1)$ (discrete-time)

- *Finite*-state
- Causal
- Time-invariant
- Linear

Slide 8

Notes:

Notes:

Linear System Responses

Impulse response: limit of rectangular responses.

If

$$h_a(t, t_0) = \text{response to } r_a(t - t_0)$$

where

$$r_a(t) = \begin{cases} 1/a & \text{for } 0 < t < a \\ 0 & \text{otherwise} \end{cases}$$

Then define impulse response by

$$h(t, t_0) = \lim_{a \to 0} h_a(t, t_0)$$

Slide 9

Notes:

Kernel formula for input-output description: If input is $u(t)$, then

$$\begin{aligned} u(t) &= \lim_{\Delta t \to 0} \Delta t \sum_{n=-\infty}^{\infty} u(n\Delta t) r_{\Delta t}(t - n\Delta t) \\ &\approx \Delta t \sum_{n=-\infty}^{\infty} u(n\Delta t) r_{\Delta t}(t - n\Delta t) \end{aligned}$$

Output, $y(t)$, then is (use linearity!)

$$y(t) \approx \sum_{n=-\infty}^{\infty} u(n\Delta t) h_{\Delta t}(t, n\Delta t)\Delta t$$

and so

$$\begin{aligned} y(t) &= \lim_{\Delta t \to 0} \sum_{n=-\infty}^{\infty} u(n\Delta t) h_{\Delta t}(t, n\Delta t)\Delta t \\ &= \int_{-\infty}^{\infty} u(\tau) h(t, \tau) d\tau \end{aligned}$$

Slide 10

Notes:

Special Cases:

Time-invariant: $h(t,\tau) = h(t-\tau)$: then output is given by

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} u(\tau)h(t-\tau)d\tau \\ &= \int_{-\infty}^{\infty} u(t-\tau)h(\tau)d\tau \end{aligned}$$

— convolution.

Causal: In terms of impulse response

General: $h(t,\tau) = 0$ for $t < \tau$

Time-Invariant: $h(t) = 0$ for $t < 0$.

MIMO: With p inputs and q outputs, get a $q \times p$ impulse response matrix.

Slide 11

Notes:

Convolution:

The *convolution* of two functions $f(t)$ and $g(t)$ is defined by

$$\begin{aligned} f(t) * g(t) &= \int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau \\ &= \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \end{aligned}$$

If $f(t) = 0$ for all $t < 0$,

$$f(t) * g(t) = \int_{0}^{\infty} f(\tau)g(t-\tau)d\tau$$

If also $g(t) = 0$ for all $t < 0$,

$$\begin{aligned} f(t) * g(t) &= \int_{0}^{t} f(t-\tau)g(\tau)d\tau \\ &= \int_{0}^{t} f(\tau)g(t-\tau)d\tau \end{aligned}$$

Slide 12

Notes:

Laplace Transform:

For a function $f(t)$ with $f(t) = 0$ for all $t < 0$, the *Laplace Transform* of $f(t)$ is defined by:

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

Properties:

Linear:

$$\begin{aligned} \mathcal{L}\{f(t) + g(t)\} &= F(s) + G(s) \\ \mathcal{L}\{af(t)\} &= aF(s) \end{aligned}$$

for all *constants* a.

Slide 13

Notes:

Laplace Transform Properties continued:

Shifting Theorem: For $t_0 \geq 0$

$$\mathcal{L}\{f(t-t_0)\} = e^{-st_0}F(s)$$

Proof:

$$\begin{aligned}
\mathcal{L}\{f(t-t_0)\} &= \int_0^\infty e^{-st} f(t-t_0)\, dt \\
&= \int_{-t_0}^\infty e^{-s(t'+t_0)} f(t')\, dt' \\
&= \int_0^\infty e^{-s(t'+t_0)} f(t')\, dt' \\
&= e^{-st_0} \int_0^\infty e^{-st'} f(t')\, dt' \\
&= e^{-st_0} \mathcal{L}\{f(t)\}
\end{aligned}$$

using the fact that $f(t) = 0$ for $t < 0$, and the substitution $t' = t - t_0$

Slide 14

Notes:

Laplace Transform Properties continued:

Convolution

$$\mathcal{L}\{f(t) * g(t)\} = F(s)G(s)$$

Proof:

$$\begin{aligned}
\mathcal{L}\{f(t) * g(t)\} &= \int_0^\infty e^{-st} f(t) * g(t)\ dt \\
&= \int_0^\infty e^{-st} \int_0^t f(\tau)g(t-\tau)\ d\tau\ dt \\
&= \int_0^\infty \int_\tau^\infty e^{-st} f(\tau)g(t-\tau)\ dt\ d\tau \\
&= \int_0^\infty f(\tau) \int_\tau^\infty e^{-st} g(t-\tau)\ dt\ d\tau \\
&= \int_0^\infty f(\tau) \int_0^\infty e^{-s(t'+\tau)} g(t')\ dt'\ d\tau \\
&= \int_0^\infty f(\tau)e^{-s\tau} \int_0^\infty e^{-st'} g(t')\ dt'\ d\tau \\
&= F(s)G(s)
\end{aligned}$$

where the interchange of integrals uses the fact that the integration is over the set $\{(t,\tau) | t \geq 0, \tau \geq 0, \tau \leq t\}$.

Slide 15

Notes:

Laplace Transform Properties continued:

Derivative:

$$\mathcal{L}\{f'(t)\} = sF(s) - f(0)$$

Proof:

$$\begin{aligned}
\mathcal{L}\{f'(t)\} &= \int_0^\infty e^{-st} f'(t)\, dt \\
&= \int_0^\infty e^{-st}\, df(t) \\
&= e^{-st} f(t)\Big|_0^\infty - \int_0^\infty f(t)\, de^{-st} \\
&= -f(0) + s\int_0^\infty f(t) e^{-st}\, dt \\
&= sF(s) - f(0)
\end{aligned}$$

(integration by parts).

Notes:

Laplace Transform Properties continued:

Integral:

$$\mathcal{L}\{\int_{0^-}^{t} f(\tau)d\tau\} = \frac{1}{s}F(s)$$

Exponential:

$$\mathcal{L}\{e^{at}U(t)\} = \frac{1}{s-a}$$

Impulse:

$$\mathcal{L}\{\delta(t)\} = 1$$

Unit Step:

$$\mathcal{L}\{U(t)\} = \frac{1}{s}$$

Slide 17

Notes:

More Laplace Transform Properties:

Exponential Multiplication: For any $f(t)$:

$$\begin{aligned}\mathcal{L}\{e^{at}f(t)\} &= \int_0^\infty f(t)e^{-(s-a)t}\,dt \\ &= F(s-a)\end{aligned}$$

Trigonometric Functions:

$$\begin{aligned}\mathcal{L}\{\sin(\omega t)\} &= \mathcal{L}\{(e^{j\omega t}-e^{-j\omega t})/(2j)\} \\ &= \left(1/(s-j\omega)-1/(s+j\omega)\right)/(2j) \\ &= \omega/(s^2+\omega^2)\end{aligned}$$

Similarly:

$$\mathcal{L}\{\cos(\omega t)\} = s/(s^2+\omega^2)$$

Exponentials and Sinusoids: From the above:

$$\mathcal{L}\{e^{at}\cos(\omega t)\} = (s-a)/((s-a)^2+\omega^2)$$

and

$$\mathcal{L}\{e^{at}\sin(\omega t)\} = \omega/((s-a)^2+\omega^2)$$

Slide 18

Notes:

Miscellaneous and Terminology:

1. Examples 2.2, 2.3, 2.4 and 2.5.
2. Laplace transform of impulse response matrix $h(t)$ is the *transfer function matrix* $H(s)$.
3. Lumped system $\Longrightarrow$ rational transfer function.
4. (a) $H(s)$ proper $\Longleftrightarrow$ $H(\infty)$ is finite.
   (b) Strictly proper: $H(\infty) = 0$
   (c) Biproper: proper and $H(\infty) \neq 0$
5. (a) Pole of $H(s)$: a complex number s_p such that $H(s_p) = \infty$
   (b) Zero of $H(s)$: a complex number s_z such that $H(s_z) = 0$
   (c) $H(s)$ can be factored into a product of first and second-order factors with real coefficients, or of first-order factors with complex coefficients. Numerator factors are of the form $(s - s_z)$; Denominator factors are of the form $(s - s_p)$

Slide 19

Notes:

2 State Variable (State Space) Equations

2.1 Continuous-Time

1. State variables form an n-dimensional column vector $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots x_n(t))^T$
2. State equations are:

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

   If there are p inputs and q outputs, then

   A is $n \times n$;

   B is $n \times p$;

   C is $q \times n$;

   D is $q \times p$.

Slide 20

Notes:

Example; Nonlinear state equations:

Pendulum of mass m, length l, external horizontal force $u(t)$ on mass; θ as position variable.

Dynamical Equation:

$$ml\ddot{\theta} = -mg\sin\theta - u\cos\theta$$

State variables: $x_1 = \theta$, $x_2 = \dot{\theta}$;

State equations (nonlinear, time-invariant):

$$\begin{aligned} \dot{x_1} &= x_2 \\ \dot{x_2} &= -\frac{g}{l}\sin x_1 - \frac{1}{ml}(\cos x_1)u \end{aligned}$$

— "input affine"

Read examples 2.6 – 2.10.

Slide 21

Notes:

Transfer Functions of State-Variable Systems

Take Laplace transform of state equations:

$$\begin{aligned} s\mathbf{X}(s) - \mathbf{x}(\mathbf{0}) &= A\mathbf{X}(s) + B\mathbf{U}(s) \\ \mathbf{Y}(s) &= C\mathbf{X}(s) + D\mathbf{U}(s) \end{aligned}$$

Solve for $\mathbf{X}(s)$:

$$(sI - A)\mathbf{X}(s) = B\mathbf{U}(s) + \mathbf{x}(\mathbf{0})$$

and so

$$\mathbf{X}(s) = (sI - A)^{-1}B\mathbf{U}(s) + (sI - A)^{-1}\mathbf{x}(\mathbf{0})$$

First term: zero-state state response;

Second term: zero-input state response.

Use second equation:

$$\mathbf{Y}(s) = \left(C(sI - A)^{-1}B + D\right)\mathbf{U}(s) + C(sI - A)^{-1}\mathbf{x}(\mathbf{0})$$

First term: zero-state (output) response;

Second term: zero-input (output) response.

Slide 22

Notes:

Transfer Function of State Variable Equations (continued):

For transfer function matrix, **x**(0)=**0**: then

$$\mathbf{Y}(s) = H(s)\mathbf{U}(s)$$

where

$$H(s) = C(sI - A)^{-1}B + D$$

Notes:

- Op-amp implementation: need only integrators, adders, and constant gains; all “easily” done with op-amps.
- Linearization of a nonlinear, time-invariant system:
  - About a point: linear, time-invariant system
  - About a trajectory: linear, *time-varying* system

Slide 23

Notes:

Linearization: Example: Pendulum as previously:

State variables: $x_1 = \theta$, $x_2 = \dot{\theta}$;

State equations (nonlinear, time-invariant):

$$\begin{aligned} \dot{x_1} &= x_2 \\ \dot{x_2} &= -\frac{g}{l}\sin x_1 - \frac{1}{ml}(\cos x_1)u \end{aligned}$$

Linearize about equilibrium point $\mathbf{x}(t) = 0$, $u(t) = 0$:

$$\begin{aligned} \dot{x_1} &= x_2 \\ \dot{x_2} &= -\frac{g}{l}x_1 - \frac{1}{ml}u \end{aligned}$$

– linear, time-invariant.

Linearize about known natural trajectory $\mathbf{x}(t)$, $u(t) = 0$:

$$\begin{aligned} \delta\dot{x}_1 &= \delta x_2 \\ \delta\dot{x}_2 &= -\frac{g}{l}\cos(x_1(t))\delta x_1 - \frac{1}{ml}\cos(x_1(t))\delta u \end{aligned}$$

– linear, time-varying.

Slide 24

Notes:

Circuits

1. RLC Networks, p.26; example 2.11
2. RLC procedure:
   (a) Normal tree: branches in the order v_{src}, C, R, L, i_{src}
   (b) State variables: v_c in tree and i_L in links
   (c) Apply KVL to fundamental loops of state variable links, and KCL to fundamental cutsets of state variable branches.

Read example 2.13: tunnel diode $\Longrightarrow$ negative resistance

Slide 25

Notes:

2.2 Discrete-time systems.

Basic element: unit delay

Discrete convolution:

$$(p_n) = (f_n) * (g_n)$$

where

$$p_n = \sum_{k=-\infty}^{\infty} f_k g_{n-k}$$

For "causal" sequences (f_n) and g_n, (i.e., with $f_n = g_n = 0$ for all $n < 0$)

$$p_n = \sum_{k=0}^{n} f_k g_{n-k}$$

(One-Sided) Z-Transform: If $f_n = 0$ for $n < 0$

$$\mathcal{Z}\{f_n\} = F(z) = \sum_{k=0}^{\infty} f_n z^{-n}$$

Slide 26

Notes:

$\mathcal{Z}$-Transform Properties.

1. Non-causal shift fomula:

$$\mathcal{Z}\{x(n+1)\} = zX(z) - zx(0)$$

2. Convolution:

$$\mathcal{Z}\{f * g\} = F(z)G(z)$$

3. State equations

$$\begin{aligned} \mathbf{x}(n+1) &= A\mathbf{x}(n) + B\mathbf{u}(n) \\ \mathbf{y}(n) &= C\mathbf{x}(n) + D\mathbf{u}(n) \end{aligned}$$

Transfer function for state equations: take $\mathcal{Z}$-transform

$$\begin{aligned} z\mathbf{X}(z) - z\mathbf{x}(\mathbf{0}) &= A\mathbf{X}(z) + B\mathbf{U}(z) \\ \mathbf{Y}(z) &= C\mathbf{X}(z) + D\mathbf{U}(z) \end{aligned}$$

and so

$$\mathbf{Y}(z) = H(z)\mathbf{U}(z)$$

where

$$H(z) = C(zI - A)^{-1}B + D$$

Slide 27

Notes:

3 Linear Algebra (Chapter 3):

3.1 Fundamentals:

Vector space: *vectors* and *scalars* (here always real or complex) which satisfy usual properties:

1. Vectors can be added and subtracted;
2. Scalars can be added,subtracted, multiplied, and divided (except by 0);
3. Vectors can be multiplied by scalars to give another vector: distributive, etc.

Examples:

1. $\mathbf{R}^n$: column vectors of real numbers;
2. $\mathbf{C}^n$: column vectors of complex numbers;
3. l_2: sequences (x_n) with $\sum_{n=-\infty}^{\infty} x_n^2$ finite.
4. L_2: functions $f(t)$ with $\int_{-\infty}^{\infty} f^2(t)dt$ finite
5. Many others . . .

Slide 28

Notes:

Basis:

1. A set of vectors $\{\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_m\}$ is *linearly independent* if the only scalars $\alpha_1, \alpha_2, \ldots \alpha_m$ which satisfy the equation

$$\alpha_1 \mathbf{q}_1 + \alpha_2 \mathbf{q}_2 + \ldots \alpha_m \mathbf{q}_m = 0$$

are $\alpha_1 = \alpha_2 = \ldots \alpha_m = 0$.

2. A set of vectors $\{\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_m\}$ *spans* a vector space if every vector $\mathbf{q}$ in the space can be written in the form

$$\mathbf{q} = \alpha_1 \mathbf{q}_1 + \alpha_2 \mathbf{q}_2 + \ldots \alpha_m \mathbf{q}_m$$

for some scalars $\alpha_1, \alpha_2, \ldots \alpha_m$.

3. A set of vectors $\{\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_n\}$ is *a basis* of a vector space if it is linearly independent and spans the space.

4. If $\{\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_n\}$ is a basis of a vector space, then every vector $\mathbf{q}$ in the space can be written in the form

$$\mathbf{q} = \alpha_1 \mathbf{q}_1 + \alpha_2 \mathbf{q}_2 + \ldots \alpha_m \mathbf{q}_m$$

for a unique set of scalars $\{\alpha_1, \alpha_2, \ldots \alpha_m\}$.

Slide 29

Notes:

Dimension and Notation:

Fundamental fact: every basis of a given vector space has the same number of elements; this number is called the *dimension* of the space.

Notation: if Q is the matrix formed from the column vectors $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$, then the equation

$$\mathbf{x} = \alpha_1 \mathbf{q}_1 + \ldots + \alpha_n \mathbf{q}_n$$

can be written as

$$\mathbf{x} = Q\mathbf{a}$$

where

$$\mathbf{a} = [\alpha_1, \ldots, \alpha_n]'$$

Basis example: *standard basis* for R^n : $\mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_n$ where

$$\mathbf{i}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} ; \mathbf{i}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} ; \cdots \mathbf{i}_n = \begin{bmatrix} 0 \\ 0 \\ \ldots \\ 1 \end{bmatrix}$$

Slide 30

Notes:

Norms:

Properties:

1. $||\mathbf{x}|| \geq 0; ||\mathbf{x}|| = 0 \Longrightarrow \mathbf{x} = 0$
2. $||\alpha \mathbf{x}|| = |\alpha| \; ||\mathbf{x}||$
3. $||\mathbf{x}_1 + \mathbf{x}_2|| \leq ||\mathbf{x}_1|| + ||\mathbf{x}_2||$

Norm Examples: $1, 2, \infty, p$ norms

1. $||\mathbf{q}||_1 = |q_1| + |q_2| \ldots + |q_n|$
2. $||\mathbf{q}||_\infty = \max\{|q_1|, |q_2|, \ldots |q_n|\}$
3. $||\mathbf{q}||_2 = \sqrt{|q_1|^2 + |q_2|^2 \ldots + |q_n|^2}$
4. $||\mathbf{q}||_p = (|q_1|^p + |q_2|^p \ldots + |q_n|^p)^{1/p}$ for $p \geq 1$

Notes:

$||\mathbf{q}||_2$ is the usual Euclidean norm;

The subscript is usually omitted when only one norm is being used.

Slide 31

Notes:

Inner Product:

A scalar-valued product of two vectors: $<\mathbf{x},\mathbf{y}>$ with the properties

1. $<\mathbf{x},\mathbf{x}> > 0$ unless $\mathbf{x}=0$
2. $<\mathbf{x},\mathbf{y}>=<\mathbf{y},\mathbf{x}>$ (real);
   $<\mathbf{x},\mathbf{y}>=<\mathbf{y},\mathbf{x}>^*$ (complex)
3. $<\mathbf{x}+\mathbf{y},\mathbf{z}>=<\mathbf{x},\mathbf{z}>+<\mathbf{y},\mathbf{z}>$
4. $<\alpha\mathbf{x},\mathbf{z}>=\alpha<\mathbf{x},\mathbf{z}>$

Can be proved that $||\mathbf{x}||=\sqrt{<\mathbf{x},\mathbf{x}>}$ defines a norm.

Inner Product Examples:

1. $\mathbf{R}^n$; $<\mathbf{x},\mathbf{y}>=\mathbf{y}^T\mathbf{x}$
2. $\mathbf{C}^n$; $<\mathbf{x},\mathbf{y}>=\mathbf{y}^*\mathbf{x}$
3. $L_2(-\pi:\pi)$; $<f(x),g(t)>=\int_{-\pi}^{\pi} f(t)g^*(t)\,dt$
4. l_2; $<(x_n),(y_n)>=\sum_{n=-\infty}^{\infty} x_n y_n^*$
5. $L_2(-\infty:\infty)$; $<f(x),g(t)>=\int_{-\infty}^{\infty} f(t)g^*(t)\,dt$

Slide 32

Notes:

Norm and Inner Product Terminology:

1. Normalized: $\mathbf{x}$ normalized iff $||\mathbf{x}|| = 1$
2. Orthogonal: $\mathbf{x}$ and $\mathbf{y}$ orthogonal iff $< \mathbf{x}, \mathbf{y} >= 0$
3. Orthonormal: $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_m\}$ orthonormal iff $||\mathbf{x}_i|| = 1$ for all i, and $< \mathbf{x}_i, \mathbf{x}_j >= 0$ for all $i \neq j$.
4. Projection of one vector on another: projection of $\mathbf{x}$ on $\mathbf{y}$ is the vector $\frac{<\mathbf{x},\mathbf{y}>}{||\mathbf{y}||}\mathbf{y}$
5. If $A = [\mathbf{a}_1, \ldots, \mathbf{a}_m]$ with $m \leq n$, and the $\mathbf{a}_i$ are orthonormal, then $A'A = I_m$, but not necessarily $AA' = I_n$

Notes:

Orthonormalization (Gram-Schmidt):

Given a set of vectors $\{\mathbf{e_1}, \mathbf{e_2}, \ldots \mathbf{e_m}\}$, a set of orthonormal vectors with the same span can be found as follows:

$$\mathbf{u}_1 = \mathbf{e}_1; \qquad \mathbf{q}_1 = \mathbf{u}_1/||\mathbf{u}_1||$$

$$\mathbf{u}_2 = \mathbf{e}_2 - (\mathbf{q}_1'\mathbf{e}_2)\mathbf{q}_1; \qquad \mathbf{q}_2 = \mathbf{u}_2/||\mathbf{u}_2||$$

$$\vdots$$

$$\mathbf{u}_m = \mathbf{e}_m - \sum_{k=1}^{m-1} (\mathbf{q}_k'\mathbf{e}_m)\mathbf{q}_k; \quad \mathbf{q}_m = \mathbf{u}_m/||\mathbf{u}_m||$$

— not necessarily optimal numerically.

Slide 34

Notes:

Cauchy-Schwartz Inequality:

For any vectors $\mathbf{x}$ and $\mathbf{y}$ in an inner-product space:

$$|<\mathbf{x},\mathbf{y}>| \leq ||\mathbf{x}||\ ||\mathbf{y}||$$

Proof:

$$\begin{aligned} 0 &\leq\ <\mathbf{x}-\lambda\mathbf{y},\mathbf{x}-\lambda\mathbf{y}> \\ &=\ <\mathbf{x},\mathbf{x}> -\lambda <\mathbf{y},\mathbf{x}> -\lambda^* <\mathbf{x},\mathbf{y}> \\ &\quad +|\lambda|^2 <\mathbf{y},\mathbf{y}> \end{aligned}$$

Now pick $\lambda = <\mathbf{x},\mathbf{y}>/||\mathbf{y}||^2$: then

$$\begin{aligned} 0 &\leq ||\mathbf{x}||^2 - 2\frac{|<\mathbf{x},\mathbf{y}>|^2}{||\mathbf{y}||^2} + \frac{|<\mathbf{x},\mathbf{y}>|^2}{||\mathbf{y}||^4}||\mathbf{y}||^2 \\ &= ||\mathbf{x}||^2 - \frac{|<\mathbf{x},\mathbf{y}>|^2}{||\mathbf{y}||^2} \end{aligned}$$

and so

$$|<\mathbf{x},\mathbf{y}>|^2 \leq ||\mathbf{x}||^2\ ||\mathbf{y}||^2$$

Slide 35

Notes:

3.2 Linear Equations:

Assume an equation

$$A\mathbf{x} = \mathbf{y}$$

where A is $m \times n$, $\mathbf{x}$ is $n \times 1$, and $\mathbf{y}$ is $m \times 1$.

Then:

1. Range(A)=all possible linear combinations of columns of A
2. $\rho(A) = rank(A) = \dim(range(A))$; note that this causes numerical difficulties.
3. $\mathbf{x}$ is a *null* vector of A: $A\mathbf{x} = \mathbf{0}$;
4. Nullspace(A)=set of all null vectors of A
5. $\nu(A) = nullity(A) = \dim(nullspace(A))$
6. Fundamental result: $\rho(A) + \nu(A) =$ # columns of A.

Slide 36

Notes:

Solutions:

Theorem 3.1:

1. There exists a solution of $A\mathbf{x} = \mathbf{y}$ if, and only if, $\mathbf{y}$ is in $range(A)$
2. If A is an $m \times n$ matrix, there exists a solution of $A\mathbf{x} = \mathbf{y}$ for every $\mathbf{y}$ if, and only if, $\rho(A) = m$.

Theorem 3.2:

If A is an $m \times n$ matrix, and if $\nu(A) = 0$ (i.e., $\rho(A) = n$), any solution is unique. Otherwise, all solutions are given by:

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n$$

where $\mathbf{x}_n$ is any vector in the nullspace, and $\mathbf{x}_p$ is any one solution.

Slide 37

Determinants:

For a square matrix $A = (a_{ij})$:

1. $\det(A) = \Sigma a_{ij} c_{ij}$, where c_{ij} is the cofactor of a_{ij}
2. $A^{-1} = Adj(A)/\det(A)$, where $Adj(A) = (c_{ij})'$
3. Determinant properties:
   (a) $\det(AB) = \det(A)\det(B)$;
   (b) $\det(A) \neq 0$ if, and only if, A^{-1} exists (i.e., A is nonsingular).

Slide 38

Notes:

Notes:

3.3 Change of Basis:

If $A = (a_{ij})$ is an $n \times n$ matrix , and $\mathbf{x}$ is a vector, with

$$\mathbf{x} = x_1\mathbf{i_1} + ... + x_n\mathbf{i_n}$$

where $\{\mathbf{i_1}, ...\mathbf{i_n}\}$ is the standard basis, then

$$A\mathbf{x} = y_1\mathbf{i_1} + ... + y_n\mathbf{i_n}$$

where $y_i = a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n$, or

$$y_i = \Sigma_{j=1}^{n} a_{ij}x_j$$

Similarly, if $\{\mathbf{q_1}, \mathbf{q_2}, ...\mathbf{q_n}\}$ is any other basis, and $\mathbf{x}$ is expressed as

$$\mathbf{x} = \overline{x}_1\mathbf{q_1} + ... + \overline{x}_n\mathbf{q_n}$$

and $\mathbf{y}$ is given by

$$\mathbf{y} = A\mathbf{x} = \overline{y}_1\mathbf{q_1} + ... + \overline{y}_n\mathbf{q_n}$$

then

$$\overline{y}_i = \Sigma_{j=1}^{n} \overline{a}_{ij}\overline{x}_j$$

The matrix $\overline{A} = (\overline{a}_{ij})$ is the representation of A with respect to the basis $\{\mathbf{q_1}, \mathbf{q_2}, ...\mathbf{q_n}\}$

Slide 39

Notes:

Change of Basis (Continued):

As usual, let $Q = [\mathbf{q}_1, \ldots, \mathbf{q}_n]$, where Q is nonsingular, since the $\{\mathbf{q_i}\}$ form a basis. Then, $\mathbf{x} = Q\overline{\mathbf{x}}$, and $\mathbf{y} = Q\overline{\mathbf{y}}$.

Substitute in the equation $A\mathbf{x} = \mathbf{y}$ to get:

$$AQ\overline{\mathbf{x}} = Q\overline{\mathbf{y}}$$

or

$$Q^{-1}AQ\overline{\mathbf{x}} = \overline{\mathbf{y}}$$

and so

$$Q^{-1}AQ = \overline{A}$$

This can also be written as

$$A[\mathbf{q}_1, \ldots, \mathbf{q}_n] = [\mathbf{q}_1, \ldots, \mathbf{q}_n]\overline{A}$$

The last equation implies that the $i-th$ column of $\overline{A}$ is the representation of $A\mathbf{q_i}$ with respect to the basis $\{\mathbf{q_1}, \ldots \mathbf{q_n}\}$; this is often the easiest way to find $\overline{A}$

Slide 40

Notes:

Example:

If

$$A = \begin{bmatrix} 3 & 2 & -1 \\ -2 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix}$$

and

$$\mathbf{b} = [0, 0, 1]'$$

then the vectors

$$\{\mathbf{b}, A\mathbf{b}, A^2\mathbf{b}\} = \{\mathbf{q_1}, \mathbf{q_2}, \mathbf{q_3}\} = \begin{bmatrix} 0 & -1 & -4 \\ 0 & 0 & 2 \\ 1 & 1 & -3 \end{bmatrix}$$

form a basis.

To find $\overline{A}$ w.r.t. this basis, use the representation of $A\mathbf{q_j}$ in terms of the basis $\{\mathbf{q_1}, \mathbf{q_2}, ...\mathbf{q_n}\}$

Slide 41

Notes:

Example (continued):

Then:

$$A\mathbf{q_1} = A\mathbf{b} = \mathbf{q_2}$$

and

$$A\mathbf{q_2} = A^2\mathbf{b} = \mathbf{q_3}$$

Also, the characteristic equation of A (to be done) is:

$$A^3 - 5A^2 + 15A - 17I = 0$$

and so

$$A^3\mathbf{b} = 17\mathbf{b} - 15A\mathbf{b} + 5A^2\mathbf{b}$$

Therefore $\overline{A} = \begin{bmatrix} 0 & 0 & 17 \\ 1 & 0 & -15 \\ 0 & 1 & 5 \end{bmatrix}$ (companion form).

Slide 42

Notes:

3.4 Diagonal and Jordan Form:

Definitions:

Eigenvalue: λ is an eigenvalue of A if there is a vector $\mathbf{x} \neq \mathbf{0}$ such that

$$A\mathbf{x} = \lambda\mathbf{x}$$

or

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

The vector $\mathbf{x}$ is called an *eigenvector* for the λ.

Characteristic Polynomial of A is

$$\Delta(\lambda) = \det(\lambda I - A)$$

— a monic polynomial of order n in λ, with n roots, counting multiplicity.

Roots of $\Delta(\lambda)$: λ_0 is an eigenvalue of $A \Leftrightarrow \Delta(\lambda_0) = 0$

Then, if $\lambda_1, \lambda_2, \ldots \lambda_k$ are the eigenvalues, with multiplicities $n_1, n_2, \ldots n_k$, the characteristic polynomial is given by:

$$\Delta(\lambda) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \ldots (\lambda - \lambda_k)^{n_k}$$

Slide 43

Notes:

Companion form:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\alpha_4 & -\alpha_3 & -\alpha_2 & -\alpha_1 \end{bmatrix}$$

— characteristic polynomial is

$$\Delta(\lambda) = \lambda^4 + \alpha_1\lambda^3 + \alpha_2\lambda^2 + \alpha_3\lambda^1 + \alpha_4$$

Jordan Block:

$$\begin{bmatrix} \lambda_0 & 1 & 0 & 0 \\ 0 & \lambda_0 & 1 & 0 \\ 0 & 0 & \lambda_0 & 1 \\ 0 & 0 & 0 & \lambda_0 \end{bmatrix}$$

— characteristic polynomial is

$$\Delta(\lambda) = (\lambda - \lambda_0)^4$$

Slide 44

Notes:

Diagonalization:

Simplest case: $\Delta(\lambda)$ has distinct roots.

Then the eigenvalues are all distinct; it follows that the eigenvectors are linearly independent.

To see this, assume we have eigenvalues $\lambda_1, \ldots \lambda_n$ with corresponding eigenvectors $\mathbf{q}_1, \ldots \mathbf{q}_n$, and suppose $\alpha_1 \mathbf{q}_1 + \ldots + \alpha_n \mathbf{q}_n = 0$.

Then pick any k , and apply the operator

$$\Pi_{j=1, j \neq k}^{n} (A - \lambda_j I)$$

to the vector $\alpha_1 \mathbf{q}_1 + \ldots + \alpha_n \mathbf{q}_n$ to obtain

$$\Pi_{j=1, j \neq k}^{n} (\lambda_k - \lambda_j) \alpha_k \mathbf{q}_k = 0$$

Since k is arbitrary, linear independence follows.

The representation of A for the basis $\mathbf{q}_1, \ldots \mathbf{q}_n$, is then diagonal: that is, $D = Q^{-1} A Q$.

Because the roots may be complex, must allow complex vectors.

Slide 45

Notes:

Diagonalization: Non-distinct eigenvalues:

Let λ_0 be an eigenvalue of multiplicity k_0 .

Assume that the nullity of $(A - \lambda_0 I)$ is $p_0 \leq k_0$.

If $p_0 = k_0$, then can pick *any* k_0 linearly independent vectors in the nullspace and get diagonal form again for this eigenvalue.

If $p_0 \neq k_0$, need the concept of *generalized eigenvector*:

A *generalized eigenvector* $\mathbf{q}_k$ of grade k satisfies

$$(A - \lambda_0 I)^k \mathbf{q}_k = 0$$

and

$$(A - \lambda_0 I)^{k-1} \mathbf{q}_k \neq 0$$

Slide 46

Notes:

Generalized Eigenvectors (continued):

Given a generalized eigenvector $\mathbf{q}$ of grade k, can get a chain of generalized eigenvectors

$$\begin{aligned}
\mathbf{q}_k &= \mathbf{q} \\
\mathbf{q}_{k-1} &= (A-\lambda_0 I)\mathbf{q}_k = (A-\lambda_0 I)\mathbf{q} \\
\mathbf{q}_{k-2} &= (A-\lambda_0 I)\mathbf{q}_{k-1} = (A-\lambda_0 I)^2\mathbf{q} \\
\vdots &\;\vdots\; \vdots \\
\mathbf{q}_1 &= (A-\lambda_0 I)\mathbf{q}_2 = (A-\lambda_0 I)^{k-1}\mathbf{q}
\end{aligned}$$

and these are linearly independent (multiply by $(A-\lambda_0 I)^j$, for $k-1 \geq j \geq 1$).

Note that $\mathbf{q}_1$ is an ordinary eigenvector ($A\mathbf{q}_1 = \lambda_0\mathbf{q}_1$), and that these equations can be solved by first finding a generalized eigenvector, and evaluating from the top, or by first finding an ordinary eigenvector, and solving from the bottom.

For each $j > 1$,

$$A\mathbf{q}_j = \lambda_0\mathbf{q}_j + \mathbf{q}_{j-1}$$

Slide 47

Notes:

Jordan Canonical Form:

With respect to these vectors, the block therefore has the representation

$$J_k = \begin{bmatrix} \lambda_0 & 1 & 0 & \dots & 0 \\ 0 & \lambda_0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & \lambda_0 & 1 \\ 0 & 0 & \dots & 0 & \lambda_0 \end{bmatrix}$$

Jordan canonical form: block diagonal matrix with these blocks.

Note: For any Jordan block, with *zero* eigenvalue: $J_k^k = 0$ (nilpotent) and so

$$(J_k - \lambda_0 I_k)^k = 0$$

for any Jordan block with eigenvalue λ_0

Example: Problem 3.13(4), p. 81.

Notes:

Functions of a square matrix A:

Power of A: $A^n = \underbrace{AA\ldots A}_{n \text{ times}}$

Polynomial in A: If

$$p(x) = a_n x^n + a_{n-1}x^{n-1} \cdots + a_1 x + a_0$$

then $p(A)$ is defined by

$$p(A) = a_n A^n + a_{n-1}A^{n-1} \cdots + a_1 A + a_0 I$$

Similarity: $p(QAQ^{-1}) = Qp(A)Q^{-1}$

Slide 49

Notes:

Functions of a square matrix A (continued):

Block Diagonal: If A is block diagonal:

$$A = \begin{bmatrix} A_1 & 0 & 0 & \dots & 0 \\ 0 & A_2 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & A_{r-1} & 0 \\ 0 & 0 & \dots & 0 & A_r \end{bmatrix}$$

Then

$$p(A) = \begin{bmatrix} p(A_1) & 0 & 0 & \dots & 0 \\ 0 & p(A_2) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & p(A_{r-1}) & 0 \\ 0 & 0 & \dots & 0 & p(A_r) \end{bmatrix}$$

Slide 50

Notes:

A^k **for Jordan Block:**

If

$$J_k = \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & \dots & 0 & \lambda \end{bmatrix}$$

then

$$J_{k+1}^r = \begin{bmatrix} \lambda^r & r\lambda^{r-1} & \frac{r(r-1)}{2!}\lambda^{r-2} & \dots & \frac{1}{k!}\frac{d^k(\lambda^r)}{d\lambda^k} \\ 0 & \lambda^r & r\lambda^{r-1} & \dots & \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & \lambda^r & r\lambda^{r-1} \\ 0 & 0 & \dots & 0 & \lambda^r \end{bmatrix}$$

Slide 51

Notes:

A^k for Jordan Block (continued):

Therefore, $p(J_{k+1})$ for any polynomial of a Jordan Block is given by:

$$p(J_{k+1}) = \begin{bmatrix} p(\lambda) & p'(\lambda) & \frac{p''(\lambda)}{2!} & \dots & \frac{p^{(k)}(\lambda)}{k!} \\ 0 & p(\lambda) & p'(\lambda) & \dots & \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \dots & p(\lambda) & p'(\lambda) \\ 0 & 0 & \dots & 0 & p(\lambda) \end{bmatrix}$$

— example 3.10, p.66

Notes:

Minimal Polynomial:

For any eigenvalue λ_i, the *index* of $\lambda_i = m_i$ is the largest order of all Jordan blocks with eigenvalue λ_i.

The *multiplicity* of $\lambda_i = n_i$ is the highest power of $(\lambda - \lambda_i)$ in the characteristic polynomial

$$\Delta(\lambda) = \det(\lambda I - A)$$

Therefore $m_i \leq n_i$.

Define the *minimal* polynomial of A to be the product of the terms $(\lambda - \lambda_j)$ to power of index, i.e

$$\psi(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \ldots (\lambda - \lambda_k)^{m_k}$$

Apply this polynomial to each block of the Jordan Canonical Form, and the entire matrix becomes zero:

$$\psi(A) = 0$$

The **Cayley-Hamilton Theorem** follows immediately:

$$\Delta(A) = 0$$

Consequence: for any polynomial $f(x)$, $f(A)$ can be expressed as a polynomial of degree $n-1$ in A.

Slide 53

Notes:

Matrix Functions continued:

How is this polynomial calculated?

In principle:

$$\begin{aligned} A^n &= -\alpha_0 I - \alpha_1 A - \alpha_2 A^2 \ldots - \alpha_{n-1}A^{n-1} \\ &= p_1(A) \\ A^{n+1} &= -\alpha_0 A - \alpha_1 A^2 - \alpha_2 A^3 \ldots - \alpha_{n-1}A^n \\ &= -\alpha_0 A \ldots - \alpha_{n-2}A^{n-1} - \alpha_{n-1}p_1(A) \\ &= \ldots \end{aligned}$$

More realistic: division with remainder gives:

$$f(\lambda) = q(\lambda)\Delta(\lambda) + h(\lambda)$$

where $h(\lambda)$ is the remainder, with order $< n$.

Therefore, for any eigenvalue λ_i

$$f(\lambda_i) = h(\lambda_i)$$

More generally, if n_i is the multiplicity of λ_i

$$f^{(l)}(\lambda_i) = h^{(l)}(\lambda_i) \qquad \text{for } 1 \leq l \leq n_i - 1$$

Slide 54

Notes:

Matrix Functions continued:

If these equations hold for all eigenvalues, we say "$f = h$ on the spectrum of A", and, by the Cayley-Hamilton theorem,

$$f(A) = h(A)$$

Note: Also works with $\bar{n}$ (the degree of the *minimal* polynomial) in place of n, but $\bar{n}$ is not normally known.

It is often more convenient to use these conditions directly: assume a polynomial of degree $n-1$ with unknown coefficients:

$$h(\lambda) = \beta_0 + \beta_1\lambda + \beta_2\lambda^2 + \cdots + \beta_{n-1}\lambda^{n-1}$$

and use the n conditions above to solve for the β_l.

Example 3.10: If A is a Jordan block with eigenvalue λ_0, it is more convenient to assume the form

$$h(\lambda) = \beta_0 + \beta_1(\lambda - \lambda_0) + \cdots + \beta_{n-1}(\lambda - \lambda_0)^{n-1}$$

and the formula for $f(J_k)$ follows.

Note: Formula for $f(J_k)$ shows that derivatives are necessary.

Slide 55

Notes:

Transcendental Matrix Functions:

Can define transcendental functions of A by means of (infinite) power series.

Simpler: *define* a transcendental function $f(A)$ of A to be a polynomial $h(A)$ of order $n-1$ in A with $f = h$ on the spectrum of A.

Most important transcendental function: e^{At}.

Example: Problem 3.22 (3.13(4), p. 81).

Properties of matrix exponentials:

1. Differentiation:
$$\frac{d}{dt}e^{At} = Ae^{At} = e^{At}A$$
2. $$e^{(A+B)t} \neq e^{At}e^{Bt}$$
unless $AB = BA$
3. $$\mathcal{L}\{e^{At}\} = (sI - A)^{-1}$$

Slide 56

Notes:

Lyapunov equation:

If A is $n \times n$, B is $m \times m$, and M and C are $n \times m$, then the equation

$$AM + MB = C$$

with A, B, and C known, and M unknown, is called a Lyapunov equation: nm equations in nm unknowns.

Eigenvalues: η is an eigenvalue iff

$$AM + MB = \eta M$$

The eigenvalues are given by

$$\eta_k = \lambda_i + \mu_j$$

— nm eigenvalues for $1 \leq n$, and $1 \leq m$, where λ_i is a (right) eigenvalue of A

$$A\mathbf{x} = \lambda_i \mathbf{x}$$

and μ_j is a left eigenvalue of B:

$$\mathbf{x}B = \mu_j \mathbf{x}$$

E.g., let $\mathbf{u}$ be a right eigenvector of A, $\mathbf{v}'$ a left eigenvector of B, and $M = \mathbf{u}\mathbf{v}'$

Slide 57

Notes:

Miscellaneous Formulae (sec. 3.8)

1. $\rho(AB) \leq \min(\rho(A), \rho(B))$
2. if C and D are invertible, $\rho(AC) = \rho(DA) = \rho(A)$
3. If A is $m \times n$ and B is $n \times m$, then

$$\det(I_m + AB) = \det(I_n + BA)$$

For the last property, define

$$N = \begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} I_m & 0 \\ -B & I_n \end{pmatrix}$$

$$P = \begin{pmatrix} I_m & -A \\ B & I_n \end{pmatrix}$$

Then

$$\det(P) = \det(NP) = \det \begin{pmatrix} I_m + AB & 0 \\ B & I_n \end{pmatrix}$$

$$\det(P) = \det(QP) = \det \begin{pmatrix} I_m & -A \\ 0 & I_n + BA \end{pmatrix}$$

Notes:

3.5 Quadratic Forms (Sec.3.9):

A Quadratic Form is a product of the form $\mathbf{x}'M\mathbf{x}$.

Since $\mathbf{x}'S\mathbf{x} = 0$ for any skew-symmetric ($S' = -S$) matrix, only the symmetric part of M is significant, so assume M is symmetric.

Since eigenvalues can be complex, initially allow $\mathbf{x}$ to be complex, and look at $\mathbf{x}^\star M\mathbf{x}$.

$\mathbf{x}^\star M\mathbf{x}$ is real: $(\mathbf{x}^\star M\mathbf{x})^\star = \mathbf{x}^\star M\mathbf{x}$.

Theorem: The eigenvalues of a symmetric matrix M are real:

Proof: Let λ be a (possibly complex) eigenvalue of M. Then

$$M\mathbf{x} = \lambda\mathbf{x} \Longrightarrow \mathbf{x}^\star M\mathbf{x} = \lambda\mathbf{x}^\star\mathbf{x}$$

and so λ is real.

So all eigenvalues of a symmetric matrix are real, and so we need consider only real eigenvalues and real eigenvectors.

Slide 59

Notes:

Quadratic Forms (continued):

Theorem: If M is symmetric, then its range and nullspace are orthogonal.

Proof: Suppose $\mathbf{y} = M\mathbf{z}$ and $M\mathbf{x} = \mathbf{0}$. Then

$$\begin{aligned} <\mathbf{x}, \mathbf{y}> &= \mathbf{x}'M\mathbf{z} \\ &= \mathbf{z}'M'\mathbf{x} \\ &= \mathbf{z}'M\mathbf{x} \\ &= 0 \end{aligned}$$

Theorem: If M is symmetric, then M is diagonalizable.

Proof: Suppose there is a generalized eigenvector. Then there is a vector $\mathbf{x}$ and a real eigenvalue λ such that $(M - \lambda I)^2\mathbf{x} = 0$, but $\mathbf{y} = (M - \lambda I)\mathbf{x} \neq 0$.

So $\mathbf{y} \neq \mathbf{0}$ is both in the range and nullspace of $N = (M - \lambda I)$, a contradiction.

So there is a Q such that $M = QDQ^{-1}$.

Slide 60

Notes:

Quadratic Forms (continued):

Theorem: For a symmetric matrix, eigenvectors of different eigenvalues are orthogonal.

Proof: If $M\mathbf{x}_1 = \lambda_1 \mathbf{x}_1$ and $M\mathbf{x}_2 = \lambda_2 \mathbf{x}_2$ with $\lambda_1 \neq \lambda_2$, then

$$\mathbf{x}_1' M \mathbf{x}_2 = \mathbf{x}_1' \lambda_2 \mathbf{x}_2 = \lambda_2 \mathbf{x}_1' \mathbf{x}_2$$

but also

$$\mathbf{x}_1' M \mathbf{x}_2 = \mathbf{x}_2' M' \mathbf{x}_1 = \mathbf{x}_2' M \mathbf{x}_1 = \lambda_1 \mathbf{x}_2' \mathbf{x}_1 = \lambda_1 \mathbf{x}_1' \mathbf{x}_2$$

Therefore $(\lambda_1 - \lambda_2)\mathbf{x}_1' \mathbf{x}_2 = 0$, and since $\lambda_1 - \lambda_2 \neq 0$, it follows that $\mathbf{x}_1' \mathbf{x}_2 = 0$.

Consequence: A symmetric matrix M has an orthonormal basis of eigenvectors, and so the diagonalizing matrix Q such that $M = QDQ^{-1}$ can be taken to have orthonormal columns.

Definition: A matrix Q is called *orthogonal* if the columns of Q are orthonormal, or equivalently $QQ' = Q'Q = I$, or $Q^{-1} = Q'$

Result: if M is symmetric, $M = QDQ'$ with D diagonal, and Q orthogonal.

Slide 61

Notes:

Quadratic Forms (continued):

Positive Definite: $\mathbf{x}'M\mathbf{x} > 0$ unless $\mathbf{x} = 0$, or all eigenvalues of M are > 0.

Positive Semidefinite: $\mathbf{x}'M\mathbf{x} \geq 0$ for all $\mathbf{x}$, or all eigenvalues of M are ≥ 0.

Singular Values: If H is an $m \times n$ matrix, the *singular values* of H are defined to be the square roots of eigenvalues of $M = H'H$.

Since $\mathbf{x}'H'H\mathbf{x} = ||H\mathbf{x}||^2 \geq 0$, the singular values are all real and nonnegative.

Singular Value Decomposition: H can be decomposed into the form

$$H = RSQ'$$

where $R'R = RR' = I_m$, $Q'Q = QQ' = I_n$, and S is $m \times n$ with the singular values of H on the diagonal.

Slide 62