Lecture 2
Linear functions and examples

- linear equations and functions
- engineering examples
- interpretations
Linear equations

consider system of linear equations

\[
\begin{align*}
y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
\vdots \\
y_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n
\end{align*}
\]

can be written in matrix form as \( y = Ax \), where

\[
\begin{align*}
y &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \\
A &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \\
x &= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
\end{align*}
\]
Linear functions

A function \( f : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is linear if

- \( f(x + y) = f(x) + f(y), \forall x, y \in \mathbb{R}^n \)
- \( f(\alpha x) = \alpha f(x), \forall x \in \mathbb{R}^n \forall \alpha \in \mathbb{R} \)

i.e., superposition holds
Matrix multiplication function

• consider function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ given by $f(x) = Ax$, where $A \in \mathbb{R}^{m \times n}$

• matrix multiplication function $f$ is linear

• converse is true: any linear function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ can be written as $f(x) = Ax$ for some $A \in \mathbb{R}^{m \times n}$

• representation via matrix multiplication is unique: for any linear function $f$ there is only one matrix $A$ for which $f(x) = Ax$ for all $x$

• $y = Ax$ is a concrete representation of a generic linear function
Interpretations of $y = Ax$

- $y$ is measurement or observation; $x$ is unknown to be determined
- $x$ is ‘input’ or ‘action’; $y$ is ‘output’ or ‘result’
- $y = Ax$ defines a function or transformation that maps $x \in \mathbb{R}^n$ into $y \in \mathbb{R}^m$
Interpretation of $a_{ij}$

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

$a_{ij}$ is gain factor from $j$th input ($x_j$) to $i$th output ($y_i$)

thus, e.g.,

- $i$th row of $A$ concerns $i$th output
- $j$th column of $A$ concerns $j$th input
- $a_{27} = 0$ means 2nd output ($y_2$) doesn’t depend on 7th input ($x_7$)
- $|a_{31}| \gg |a_{3j}|$ for $j \neq 1$ means $y_3$ depends mainly on $x_1$
• $|a_{52}| \gg |a_{i2}|$ for $i \neq 5$ means $x_2$ affects mainly $y_5$

• $A$ is lower triangular, i.e., $a_{ij} = 0$ for $i < j$, means $y_i$ only depends on $x_1, \ldots, x_i$

• $A$ is diagonal, i.e., $a_{ij} = 0$ for $i \neq j$, means $i$th output depends only on $i$th input

more generally, **sparsity pattern** of $A$, i.e., list of zero/nonzero entries of $A$, shows which $x_j$ affect which $y_i$
Linear elastic structure

- \( x_j \) is external force applied at some node, in some fixed direction
- \( y_i \) is (small) deflection of some node, in some fixed direction

\[
\begin{align*}
&x_1 \rightarrow \\
&x_2 \rightarrow \\
&x_3 \rightarrow \\
&x_4 \rightarrow \\
\end{align*}
\]

(provided \( x, y \) are small) we have \( y \approx Ax \)

- \( A \) is called the compliance matrix
- \( a_{ij} \) gives deflection \( i \) per unit force at \( j \) (in m/N)
Total force/torque on rigid body

- $x_j$ is external force/torque applied at some point/direction/axis
- $y \in \mathbb{R}^6$ is resulting total force & torque on body
  ($y_1, y_2, y_3$ are $x$-, $y$-, $z$- components of total force,
  $y_4, y_5, y_6$ are $x$-, $y$-, $z$- components of total torque)
- we have $y = Ax$
- $A$ depends on geometry
  (of applied forces and torques with respect to center of gravity CG)
- $j$th column gives resulting force & torque for unit force/torque $j$
Linear static circuit

interconnection of resistors, linear dependent (controlled) sources, and independent sources

- $x_j$ is value of independent source $j$
- $y_i$ is some circuit variable (voltage, current)
- we have $y = Ax$
- if $x_j$ are currents and $y_i$ are voltages, $A$ is called the impedance or resistance matrix
Final position/velocity of mass due to applied forces

- unit mass, zero position/velocity at $t = 0$, subject to force $f(t)$ for $0 \leq t \leq n$
- $f(t) = x_j$ for $j - 1 \leq t < j$, $j = 1, \ldots, n$
  ($x$ is the sequence of applied forces, constant in each interval)
- $y_1, y_2$ are final position and velocity (i.e., at $t = n$)
- we have $y = Ax$
- $a_{1j}$ gives influence of applied force during $j - 1 \leq t < j$ on final position
- $a_{2j}$ gives influence of applied force during $j - 1 \leq t < j$ on final velocity
Gravimeter prospecting

\[ x_j = \rho_j - \rho_{\text{avg}} \] is (excess) mass density of earth in voxel \( j \);

\[ y_i \] is measured gravity anomaly at location \( i \), i.e., some component (typically vertical) of \( g_i - g_{\text{avg}} \)

\[ y = Ax \]
• $A$ comes from physics and geometry

• $j$th column of $A$ shows sensor readings caused by unit density anomaly at voxel $j$

• $i$th row of $A$ shows sensitivity pattern of sensor $i$
Thermal system

- $x_j$ is power of $j$th heating element or heat source
- $y_i$ is change in steady-state temperature at location $i$
- thermal transport via conduction
- $y = Ax$

Linear functions and examples
- $a_{ij}$ gives influence of heater $j$ at location $i$ (in °C/W)

- $j$th column of $A$ gives pattern of steady-state temperature rise due to 1W at heater $j$

- $i$th row shows how heaters affect location $i$
Illumination with multiple lamps

- $n$ lamps illuminating $m$ (small, flat) patches, no shadows
- $x_j$ is power of $j$th lamp; $y_i$ is illumination level of patch $i$
- $y = Ax$, where $a_{ij} = r_{ij}^{-2} \max\{\cos \theta_{ij}, 0\}$
  
  ($\cos \theta_{ij} < 0$ means patch $i$ is shaded from lamp $j$)
- $j$th column of $A$ shows illumination pattern from lamp $j$
Signal and interference power in wireless system

- $n$ transmitter/receiver pairs
- transmitter $j$ transmits to receiver $j$ (and, inadvertently, to the other receivers)
- $p_j$ is power of $j$th transmitter
- $s_i$ is received signal power of $i$th receiver
- $z_i$ is received interference power of $i$th receiver
- $G_{ij}$ is path gain from transmitter $j$ to receiver $i$
- we have $s = Ap$, $z = Bp$, where

$$a_{ij} = \begin{cases} G_{ii} & i = j \\ 0 & i \neq j \end{cases} \quad b_{ij} = \begin{cases} 0 & i = j \\ G_{ij} & i \neq j \end{cases}$$

- $A$ is diagonal; $B$ has zero diagonal (ideally, $A$ is ‘large’, $B$ is ‘small’)

Linear functions and examples
Cost of production

Production inputs (materials, parts, labor, . . .) are combined to make a number of products

- $x_j$ is price per unit of production input $j$

- $a_{ij}$ is units of production input $j$ required to manufacture one unit of product $i$

- $y_i$ is production cost per unit of product $i$

- we have $y = Ax$

- $i$th row of $A$ is *bill of materials* for unit of product $i$
production inputs needed

• $q_i$ is quantity of product $i$ to be produced

• $r_j$ is total quantity of production input $j$ needed

• we have $r = A^T q$

total production cost is

$$r^T x = (A^T q)^T x = q^T Ax$$
Network traffic and flows

- $n$ flows with rates $f_1, \ldots, f_n$ pass from their source nodes to their destination nodes over fixed routes in a network.

- $t_i$, traffic on link $i$, is sum of rates of flows passing through it.

- Flow routes given by flow-link incidence matrix $A_{ij} = \begin{cases} 1 & \text{flow } j \text{ goes over link } i \\ 0 & \text{otherwise} \end{cases}$

- Traffic and flow rates related by $t = Af$. 

Linear functions and examples 2–20
link delays and flow latency

• let $d_1, \ldots, d_m$ be link delays, and $l_1, \ldots, l_n$ be latency (total travel time) of flows

• $l = A^T d$

• $f^T l = f^T A^T d = (Af)^T d = t^T d$, total # of packets in network
Linearization

• if \( f : \mathbb{R}^n \to \mathbb{R}^m \) is differentiable at \( x_0 \in \mathbb{R}^n \), then

\[
x \text{ near } x_0 \implies f(x) \text{ very near } f(x_0) + Df(x_0)(x - x_0)
\]

where

\[
Df(x_0)_{ij} = \frac{\partial f_i}{\partial x_j} \bigg|_{x_0}
\]

is derivative (Jacobian) matrix

• with \( y = f(x) \), \( y_0 = f(x_0) \), define input deviation \( \delta x := x - x_0 \), output deviation \( \delta y := y - y_0 \)

• then we have \( \delta y \approx Df(x_0)\delta x \)

• when deviations are small, they are (approximately) related by a linear function
Navigation by range measurement

- \((x, y)\) unknown coordinates in plane
- \((p_i, q_i)\) known coordinates of beacons for \(i = 1, 2, 3, 4\)
- \(\rho_i\) measured (known) distance or range from beacon \(i\)
• $\rho \in \mathbb{R}^4$ is a nonlinear function of $(x, y) \in \mathbb{R}^2$:

$$\rho_i(x, y) = \sqrt{(x - p_i)^2 + (y - q_i)^2}$$

• linearize around $(x_0, y_0)$: $\delta \rho \approx A \begin{bmatrix} \delta x \\ \delta y \end{bmatrix}$, where

$$a_{i1} = \frac{(x_0 - p_i)}{\sqrt{(x_0 - p_i)^2 + (y_0 - q_i)^2}}, \quad a_{i2} = \frac{(y_0 - q_i)}{\sqrt{(x_0 - p_i)^2 + (y_0 - q_i)^2}}$$

• $i$th row of $A$ shows (approximate) change in $i$th range measurement for (small) shift in $(x, y)$ from $(x_0, y_0)$

• first column of $A$ shows sensitivity of range measurements to (small) change in $x$ from $x_0$

• obvious application: $(x_0, y_0)$ is last navigation fix; $(x, y)$ is current position, a short time later
Broad categories of applications

linear model or function $y = Ax$

some broad categories of applications:

- estimation or inversion
- control or design
- mapping or transformation

(this list is not exclusive; can have combinations . . . )
Estimation or inversion

\[ y = Ax \]

- \( y_i \) is \( i \)th measurement or sensor reading (which we know)
- \( x_j \) is \( j \)th parameter to be estimated or determined
- \( a_{ij} \) is sensitivity of \( i \)th sensor to \( j \)th parameter

Sample problems:

- find \( x \), given \( y \)
- find all \( x \)'s that result in \( y \) (\( i.e., \) all \( x \)'s consistent with measurements)
- if there is no \( x \) such that \( y = Ax \), find \( x \) s.t. \( y \approx Ax \) (\( i.e., \) if the sensor readings are inconsistent, find \( x \) which is almost consistent)
Control or design

\[ y = Ax \]

- \( x \) is vector of design parameters or inputs (which we can choose)
- \( y \) is vector of results, or outcomes
- \( A \) describes how input choices affect results

Sample problems:

- find \( x \) so that \( y = y_{\text{des}} \)
- find all \( x \)'s that result in \( y = y_{\text{des}} \) (i.e., find all designs that meet specifications)
- among \( x \)'s that satisfy \( y = y_{\text{des}} \), find a small one (i.e., find a small or efficient \( x \) that meets specifications)
Mapping or transformation

- \( x \) is mapped or transformed to \( y \) by linear function \( y = Ax \)

Sample problems:

- determine if there is an \( x \) that maps to a given \( y \)
- (if possible) find *an* \( x \) that maps to \( y \)
- find *all* \( x \)'s that map to a given \( y \)
- if there is only one \( x \) that maps to \( y \), find it (*i.e.*, decode or undo the mapping)
Matrix multiplication as mixture of columns

write \( A \in \mathbb{R}^{m \times n} \) in terms of its columns:

\[
A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}
\]

where \( a_j \in \mathbb{R}^m \)

then \( y = Ax \) can be written as

\[
y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n
\]

(\( x_j \)'s are scalars, \( a_j \)'s are \( m \)-vectors)

- \( y \) is a (linear) combination or mixture of the columns of \( A \)
- coefficients of \( x \) give coefficients of mixture
an important example: $x = e_j$, the $j$th unit vector

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots \quad e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

then $Ae_j = a_j$, the $j$th column of $A$

($e_j$ corresponds to a pure mixture, giving only column $j$)
Matrix multiplication as inner product with rows

write $A$ in terms of its rows:

$$A = \begin{bmatrix}
\tilde{a}_1^T \\
\tilde{a}_2^T \\
\vdots \\
\tilde{a}_m^T
\end{bmatrix}$$

where $\tilde{a}_i \in \mathbb{R}^n$

then $y = Ax$ can be written as

$$y = \begin{bmatrix}
\tilde{a}_1^T x \\
\tilde{a}_2^T x \\
\vdots \\
\tilde{a}_m^T x
\end{bmatrix}$$

thus $y_i = \langle \tilde{a}_i, x \rangle$, i.e., $y_i$ is inner product of $i$th row of $A$ with $x$
geometric interpretation:

\[ y_i = \tilde{a}_i^T x = \alpha \] is a hyperplane in \( \mathbb{R}^n \) (normal to \( \tilde{a}_i \))
Block diagram representation

$y = Ax$ can be represented by a *signal flow graph* or *block diagram*

e.g. for $m = n = 2$, we represent

$$
\begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix} =
\begin{bmatrix}
  a_{11} & a_{12} \\
  a_{21} & a_{22}
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  x_2
\end{bmatrix}
$$

as

- $a_{ij}$ is the gain along the path from $j$th input to $i$th output
- (by not drawing paths with zero gain) shows sparsity structure of $A$
  (e.g., diagonal, block upper triangular, arrow . . . )
example: block upper triangular, \( i.e., \)

\[
A = \begin{bmatrix}
A_{11} & A_{12} \\
0 & A_{22}
\end{bmatrix}
\]

where \( A_{11} \in \mathbb{R}^{m_1 \times n_1}, A_{12} \in \mathbb{R}^{m_1 \times n_2}, A_{21} \in \mathbb{R}^{m_2 \times n_1}, A_{22} \in \mathbb{R}^{m_2 \times n_2} \)

partition \( x \) and \( y \) conformably as

\[
x = \begin{bmatrix}
x_1 \\
x_2
\end{bmatrix}, \quad y = \begin{bmatrix}
y_1 \\
y_2
\end{bmatrix}
\]

\((x_1 \in \mathbb{R}^{n_1}, x_2 \in \mathbb{R}^{n_2}, y_1 \in \mathbb{R}^{m_1}, y_2 \in \mathbb{R}^{m_2}) \) so

\[
y_1 = A_{11}x_1 + A_{12}x_2, \quad y_2 = A_{22}x_2,
\]

\( i.e., y_2 \) doesn’t depend on \( x_1 \)
block diagram:

... no path from $x_1$ to $y_2$, so $y_2$ doesn’t depend on $x_1$
Matrix multiplication as composition

for $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$, $C = AB \in \mathbb{R}^{m \times p}$ where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

**composition interpretation:** $y = Cz$ represents composition of $y = Ax$ and $x = Bz$

(note that $B$ is on left in block diagram)
Column and row interpretations

can write product $C = AB$ as

$$C = \begin{bmatrix} c_1 & \cdots & c_p \end{bmatrix} = AB = \begin{bmatrix} Ab_1 & \cdots & Ab_p \end{bmatrix}$$

i.e., $i$th column of $C$ is $A$ acting on $i$th column of $B$

similarly we can write

$$C = \begin{bmatrix} \tilde{c}_1^T \\ \vdots \\ \tilde{c}_m^T \end{bmatrix} = AB = \begin{bmatrix} \tilde{a}_1^T B \\ \vdots \\ \tilde{a}_m^T B \end{bmatrix}$$

i.e., $i$th row of $C$ is $i$th row of $A$ acting (on left) on $B$
Inner product interpretation

inner product interpretation:

\[ c_{ij} = \tilde{a}_i^T b_j = \langle \tilde{a}_i, b_j \rangle \]

i.e., entries of \( C \) are inner products of rows of \( A \) and columns of \( B \)

- \( c_{ij} = 0 \) means \( i \)th row of \( A \) is orthogonal to \( j \)th column of \( B \)

- **Gram matrix** of vectors \( f_1, \ldots, f_n \) defined as \( G_{ij} = f_i^T f_j \)
  (gives inner product of each vector with the others)

- \( G = [f_1 \cdots f_n]^T [f_1 \cdots f_n] \)
Matrix multiplication interpretation via paths

- $a_{ik} b_{kj}$ is gain of path from input $j$ to output $i$ via $k$

- $c_{ij}$ is sum of gains over all paths from input $j$ to output $i$