Mathematical Techniques for Computer Science Applications

CSCI-GA.1180-001

Parijat Dube

Fall 2025

Lecture 2: Matrices

Required Reading: Chapter 3 of Textbook

Lecture material from
Linear Algebra and Probability for Computer Science Applications , CRC Press, 2012.

Matrices

An mXn matrix is a rectangular array of numbers with m rows and n columns.

$$P = \begin{bmatrix} -5.2 & 3.1 & 2.9 \\ 1.8 & 4.4 & 0 \end{bmatrix}$$

P is a 2x3 matrix. $P[1,3] = 2.9$ $\quad P[2,:] = \begin{bmatrix} 1.8 & 4.4 & 0 \end{bmatrix}$

$$P[:,3] = \begin{bmatrix} 2.9 \\ 0 \end{bmatrix}$$

Operations on matrices

- Add two matrices of the same size, component by component

$$\begin{bmatrix} 3 & 1 & 2 \\ 0 & -1 & 4 \end{bmatrix} + \begin{bmatrix} 7 & -2 & 3 \\ 5 & 6 & 2 \end{bmatrix} = \begin{bmatrix} 3+7 & 1-2 & 2+3 \\ 0+5 & -1+6 & 4+2 \end{bmatrix} = \begin{bmatrix} 10 & -1 & 5 \\ 5 & 5 & 6 \end{bmatrix}$$

- Multiply a matrix by a scalar

$$3 \cdot \begin{bmatrix} 3 & 1 & 2 \\ 0 & -1 & 4 \end{bmatrix} = \begin{bmatrix} 3 \cdot 3 & 3 \cdot 1 & 3 \cdot 2 \\ 3 \cdot 0 & 3 \cdot -1 & 3 \cdot 4 \end{bmatrix} = \begin{bmatrix} 9 & 3 & 6 \\ 0 & -3 & 12 \end{bmatrix}$$

Transpose a matrix

Switch rows with columns and vice versa. Reflect across main diagonal for square matrices.

$$\text{If } A = \begin{bmatrix} 4 & 2 & -1 \\ 0 & 6 & 5 \end{bmatrix} \text{ then } A^T = \begin{bmatrix} 4 & 0 \\ 2 & 6 \\ -1 & 5 \end{bmatrix}$$

What is $(A^T)^T$?

How about $((A^T)^T)^T$?

For odd number of successive transpose operations you get the original matrix (True/False)

Algebraic properties

$$M + N = N + M.$$
$$M + (N + P) = (M + N) + P$$
$$a(M + N) = aM + aN$$
$$(a + b)M = aM + bM.$$
$$(ab)M = a(bM).$$
$$(M^T)^T = M.$$
$$M^T + N^T = (M + N)^T.$$
$$aM^T = (aM)^T.$$

Sum of transpose

- Same as Transpose of sum

$$M = \begin{bmatrix} m11 & m12 & m13 \\ m21 & m22 & m23 \\ m31 & m32 & m33 \end{bmatrix} \quad N = \begin{bmatrix} n11 & n12 & n13 \\ n21 & n22 & n23 \\ n31 & n32 & n33 \end{bmatrix} \quad M + N = \begin{bmatrix} m11 + n11 & m12 + n12 & m13 + n13 \\ m21 + n21 & m22 + n22 & m23 + n23 \\ m31 + n31 & m32 + n32 & m33 + n33 \end{bmatrix}$$

$$M^T = \begin{bmatrix} m11 & m21 & m31 \\ m12 & m22 & m32 \\ m13 & m23 & m33 \end{bmatrix} \quad N^T = \begin{bmatrix} n11 & n21 & n31 \\ n12 & n22 & n32 \\ n13 & n23 & n33 \end{bmatrix} \quad M^T + N^T = \begin{bmatrix} m11 + n11 & m21 + n21 & m31 + n31 \\ m12 + n12 & m22 + n22 & m32 + n32 \\ m13 + n13 & m23 + n23 & m33 + n33 \end{bmatrix}$$

$$(M + N)^T = \begin{bmatrix} m11 + n11 & m21 + n21 & m31 + n31 \\ m12 + n12 & m22 + n22 & m32 + n32 \\ m13 + n13 & m23 + n23 & m33 + n33 \end{bmatrix} \qquad (M + N)^T = M^T + N^T$$

How can you write $M + N$ in terms of M^T and N^T ?

Multiply matrix times vector: Dot product interpretation

Let P be an $m \times n$ matrix and let $\vec{v}$ be an n-dimensional vector. The product $P \cdot \vec{v}$ is an m-dimensional vector whose elements are dot products of each row of P with $\vec{v}$.

$$P \cdot \vec{v} = < P[1,:] \blacksquare \vec{v}, \ldots P[m,:] \blacksquare \vec{v} >$$

The vectors here are written as columns.

$$\begin{bmatrix} 3 & 1 & 2 \\ 0 & -1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 13 \\ 17 \end{bmatrix}$$

Application of matrix times vector

Grocery basket at multiple stores

- There are *n* items and *m* stores.
- *C* is an *m*X*n* matrix. *C[i,j]* is the cost of item *j* at store *i*.
- $\overrightarrow{v}$ is a shopping basket. $\overrightarrow{v}$[j] is the amount of item *j* in the basket.
- $C[i,:] \blacksquare \vec{v}$ is the total cost of the basket at store *i.*
- $C \cdot \vec{v}$ is an m-dimensional vector.
- $(C \cdot \vec{v})[i]$ = $C[i,:] \blacksquare \vec{v}$ is the total cost of the basket at store *i*.

Is a vector a 1 column matrix or a 1 row matrix?

Different authors do it different ways. There are very minor stylistic pros and cons.

Matlab accommodates both, to the extent possible. When it is forced to choose, it uses a 1 row matrix.

If you take a vector to be a 1 row matrix, then (as we shall see) you have to write the product of a vector with a matrix as $\vec{v} \cdot M$.

It makes no actual difference, but you end up with things in backwards order.

Multiply matrix times vector: Linear combination interpretation

Alternatively, $M \cdot \vec{v}$ is the weighted sum of the columns of M, weighted by the components of $\vec{v}$.

Linear combination of the columns of M

$$M \cdot \vec{v} = \vec{v}[1] \cdot M[:,1] + \ldots + \vec{v}[n] \cdot M[:,n]$$

$$\begin{bmatrix} 3 & 1 & 2 \\ 0 & -1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} = \underline{2} \cdot \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \underline{-1} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \underline{4} \cdot \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 13 \\ 17 \end{bmatrix}$$

Vector-matrix multiplication in terms of linear combinations

Vector-matrix multiplication is different from matrix-vector multiplication:

Let M be an $R \times C$ matrix.
Linear-Combinations Definition of matrix-vector multiplication: If $\mathbf{v}$ is a C-vector then

$$M * \mathbf{v} = \sum_{c \in C} \mathbf{v}[c] \text{ (column } c \text{ of } M)$$

Linear-Combinations Definition of vector-matrix multiplication: If $\mathbf{w}$ is an R-vector then

$$\mathbf{w} * M = \sum_{r \in R} \mathbf{w}[r] \text{ (row } r \text{ of } M)$$

$$[3,4] \quad * \quad \begin{bmatrix} 1 & 2 & 3 \\ 10 & 20 & 30 \end{bmatrix} \quad = \quad 3\,[1,2,3] \quad + \quad 4\,[10,20,30]$$

Transition matrix (to be discussed further in chap. 10)

Suppose that, on Jan. 1, 2019, the populations of cities A, B, C are 400K, 200K, 100K

In 2019:

Of the population of A: 70% stays in A; 20% moves to B; 10% moves to C.
Of the population of B: 25% moves to A; 65% stays in B; 10% moves to C.

Of the population of C: 5% moves to A; 5% moves to B; 90% moves to C.

Then one can calculate the population of the cities in 2020 as a matrix product

$$\begin{bmatrix} 0.7 & 0.25 & 0.05 \\ 0.2 & 0.65 & 0.05 \\ 0.1 & 0.1 & 0.9 \end{bmatrix} \cdot \begin{bmatrix} 400,000 \\ 200,000 \\ 100,000 \end{bmatrix} = \begin{bmatrix} 335,000 \\ 215,000 \\ 150,000 \end{bmatrix}$$

Signal processing: Vector is a time-series

- Compute running sum.
  Given <3,5,8,1,-2,4> compute <3,8,16,17,15,19>

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 5 \\ 8 \\ 1 \\ -2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \\ 16 \\ 17 \\ 15 \\ 19 \end{bmatrix}$$

Signal processing: Vector is a time-series

- Compute successive differences
  Given <3,8,9,17,15,19> compute <3,5,1,8,-2,4>

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 8 \\ 9 \\ 17 \\ 15 \\ 19 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 1 \\ 8 \\ -2 \\ 4 \end{bmatrix}$$

Signal processing: Vector is a time-series

- Time shifting a signal.
  Given <3,5,1,8,-2,4> compute <0,0,3,5,1,8,-2,4>

$$
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\cdot
\begin{bmatrix}
3 \\ 5 \\ 1 \\ 8 \\ -2 \\ 4
\end{bmatrix}
=
\begin{bmatrix}
0 \\ 0 \\ 3 \\ 5 \\ 1 \\ 8 \\ -2 \\ -4
\end{bmatrix}
$$

Smoothing a signal

- Observation = True signal + Noise $\vec{d} = \vec{s} + \vec{n}$
- To remove noise, estimate the true value of the signal at each time point by the weighted average of the signal and its immediate neighbors

$$\vec{d} = \langle 0.0941, 0.4514, 0.6371, 0.9001, 0.8884, 0.9844, 1.0431, 0.8984, 0.7319, 0.3911, -0.0929 \rangle$$

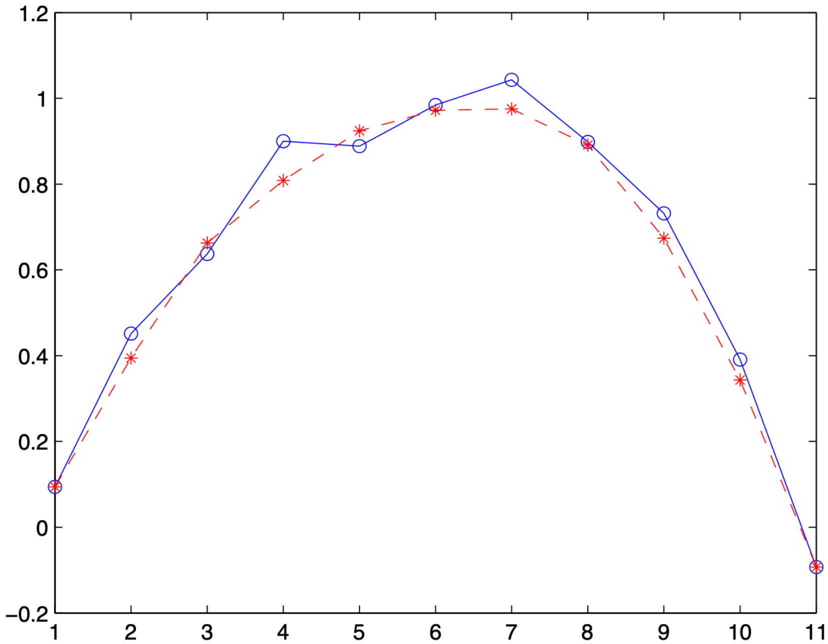
$$\begin{array}{ccccc} & & \vec{d} & & \vec{q} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} & \cdot & \begin{bmatrix} 0.0941 \\ 0.4514 \\ 0.6371 \\ 0.9001 \\ 0.8884 \\ 0.9844 \\ 1.0431 \\ 0.8984 \\ 0.7319 \\ 0.3911 \\ -0.0929 \end{bmatrix} & = & \begin{bmatrix} 0.0941 \\ 0.3942 \\ 0.6629 \\ 0.8085 \\ 0.9243 \\ 0.9720 \\ 0.9753 \\ 0.8912 \\ 0.6738 \\ 0.3434 \\ -0.0929 \end{bmatrix} \end{array}$$

Smoothing matrix

Observation vector (over 11 time periods)

Estimate of true signal (over 11 time periods)

Smoothing a signal (contd.)



The noisy points are the circles on the solid blue line.
The smoothed points are the asterisks on the dashed red line.

Figure 3.1: Smoothing

$$\vec{s} = \langle 0, 0.36, 0.64, 0.84, 0.96, 1, 0.96, 0.84, 0.64, 0.36, 0 \rangle$$

Correlation between $\vec{d}$ and $\vec{s}$ = 0.9841

Correlation between $\vec{q}$ and $\vec{s}$ = 0.9904

Smoothing improves the prediction

Time series operations

These may seem like somewhat indirect and remarkably inefficient ways of computing running sums, successive differences, and time delay, and in themselves, they are, but:

- The inefficiencies of successive differences and time delay can be largely mitigated by using a sparse matrix representation. (Running sum, not so much.)
- Much more importantly, as we will see, representing all these as matrices makes it possible to combine them with one another and with other matrix operations.

Dot product as matrix multiplication

The dot product is a row (1Xn) matrix times a column (nX1) matrix.

For that reason, in the literature, the dot product of $\vec{u}$ and $\vec{v}$ is often written as $\vec{u}^T \cdot \vec{v}$ or as $\vec{u} \cdot \vec{v}^T$, depending on whether vectors are taken to be column or row matrices.

$\vec{u}^T \cdot \vec{v}$: $\vec{u}$ is a column matrix and $\vec{v}$ is a column matrix

$\vec{u} \cdot \vec{v}^T$: $\vec{u}$ is a row matrix and $\vec{v}$ is a row matrix

What happens when you multiply a column matrix (mx1) with a row matrix (1xn) ? What is the output ?

Solving systems of linear equations

Given a matrix M and a vector $\vec{c}$, find a vector $\vec{x}$ such that $M\vec{x} = \vec{c}$.

E.g. find a vector $\vec{x}$ such that

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 1 & 2 \\ 1 & 2 & -1 \end{bmatrix} \cdot \vec{x} = \begin{bmatrix} 9 \\ 4 \\ 1 \end{bmatrix} \qquad \text{Answer: } \vec{x} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

Simultaneous linear equations

Writing out $\vec{x}$ as a tuple of three variables <a,b,c>, this becomes

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 1 & 2 \\ 1 & 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 9 \\ 4 \\ 1 \end{bmatrix}$$

or, in other words, three linear equations in three unknowns.

$$\begin{aligned} a + 2b + 3c &= 9 \\ -a + b + 2c &= 4 \\ a + 2b - c &= 1 \end{aligned}$$

Solving a matrix-vector equation: 2×2 special case

Simple formula to solve

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} * [x, y]^{\mathsf{T}} = [p, q]^{\mathsf{T}}$$

if $ad \neq bc$:

$$x = \frac{dp-cq}{ad-bc} \text{ and } y = \frac{aq-bp}{ad-bc}$$

For example, to solve

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} * [x, y]^{\mathsf{T}} = [-1, 1]^{\mathsf{T}}$$

we set

$$x = \frac{4 \cdot -1 - 2 \cdot 1}{1 \cdot 4 - 2 \cdot 3} = \frac{-6}{-2} = 3$$

and

$$y = \frac{1 \cdot 1 - 3 \cdot -1}{1 \cdot 4 - 2 \cdot 3} = \frac{4}{-2} = -2$$

Later we study algorithms for more general cases.

Applications of linear equations

Inverting any of the applications of matrix multiplication discussed earlier:

- Given: A cost matrix C[i,j] of the cost of item i at store j and the cost of a basket at the different stores, find the contents of the basket.
- Given: A transition matrix and the population of the cities after the transition, find their populations before the transition.

Interpolating a polynomial

Given n points in the plane <x_1,y_1> … <x_n,y_n> where the x_i are all different

Find an n-1 degree polynomial

$$y = t_{n-1}x^{n-1} + t_{n-2}x^{n-2} + \cdots + t_1 x^1 + t_0$$

that fits all the points.

Interpolating a polynomial

For instance, the points are <-3,1>, <-1,0>,<0,5>,<2,0>,<4,1>

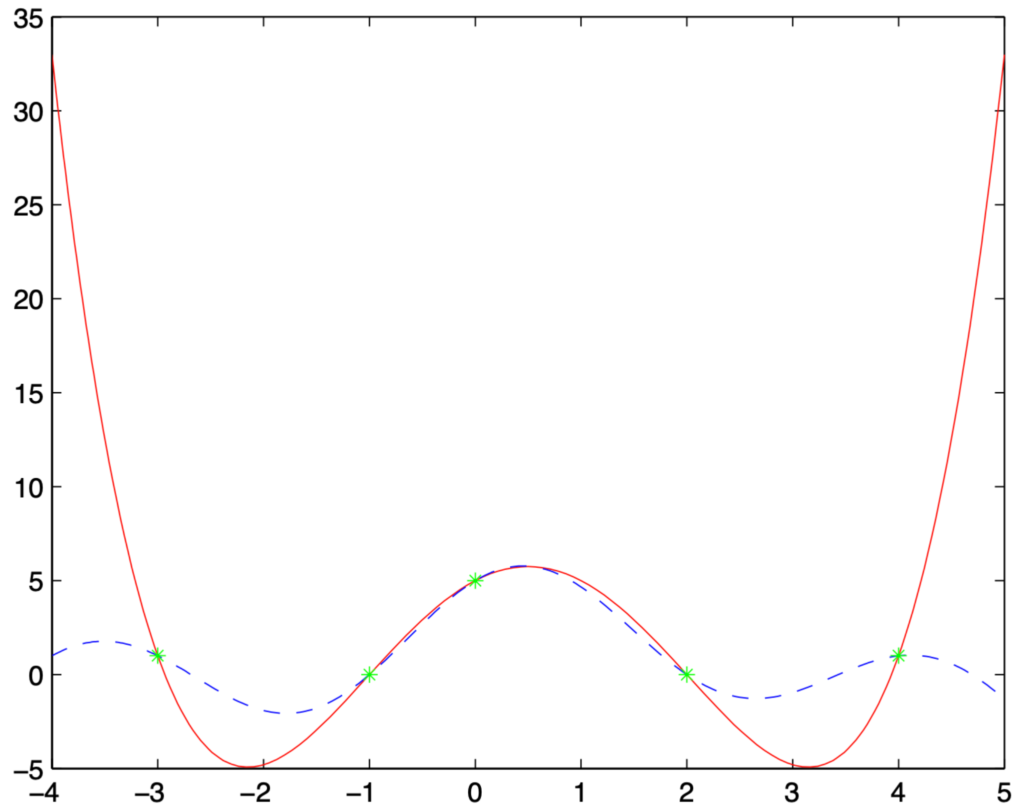
$$
\begin{aligned}
t_4(-3)^4 + t_3(-3)^3 + t_2(-3)^2 + t_1(3) + t_0 &= 1 \\
t_4(-1)^4 + t_3(-1)^3 + t_2(-1)^2 + t_1(1) + t_0 &= 0 \\
t_4(0)^4 + t_3(0)^3 + t_2(0)^2 + t_1(0) + t_0 &= 5 \\
t_4(2)^4 + t_3(2)^3 + t_2(2)^2 + t_1(2) + t_0 &= 0 \\
t_4(4)^4 + t_3(4)^3 + t_2(4)^2 + t_1(4) + t_0 &= 1
\end{aligned}
\qquad
\begin{bmatrix} 81 & -27 & 9 & -3 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 16 & 8 & 4 & 2 & 1 \\ 256 & 64 & 16 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} t_4 \\ t_3 \\ t_2 \\ t_1 \\ t_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 5 \\ 0 \\ 1 \end{bmatrix}
$$

The solution is $\langle 13/60, -26/60, -163/60, 176/60, 5 \rangle$.

Nothing very much particular about power functions

You can use sums of any functions. E.g. interpolate the points with a function $y = t_4 + t_3 \sin\frac{\pi x}{2} + t_2 \sin\frac{\pi x}{4} + t_1 \cos\frac{\pi x}{2} + t_0 \cos\frac{\pi x}{6}$

$$\begin{bmatrix} 1 & 1 & -1/\sqrt{2} & 0 & 0 \\ 1 & -1 & -1/\sqrt{2} & 0 & \sqrt{3}/2 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 & 0.5 \\ 1 & 0 & 0 & 1 & -0.5 \end{bmatrix} \cdot \begin{bmatrix} t_4 \\ t_3 \\ t_2 \\ t_1 \\ t_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 5 \\ 0 \\ 1 \end{bmatrix}$$



The polynomial interpolation is the solid red line.
The sinusoidal interpolation is the dashed blue line.

Figure 3.4: Interpolation

Algebraic properties of matrix times vector

$$(M + N)\vec{v} = M\vec{v} + N\vec{v}$$
$$M(\vec{u} + \vec{v}) = M\vec{u} + M\vec{v}$$
$$(aM)\vec{v} = a(M\vec{v}) = M(a\vec{v})$$

Linear transformation

Definition: Let $f(\vec{v})$ be a function that maps an *n*-dimensional vector $\vec{v}$ to an *m*-dimensional vector. f is a *linear transformation* if it satisfies the following properties:

For all c and $\vec{v}$, $f(c \cdot \vec{v}) = c \cdot f(\vec{v})$.

For all $\vec{u}$ and $\vec{v}$, $f(\vec{u} + \vec{v}) = f(\vec{u}) + f(\vec{v})$

Linear Transformation and Matrix

- **Theorem:** If f is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, then there exists a unique mXn matrix F such that, for all $\vec{v}$, $f(\vec{v}) = F \cdot \vec{v}$

  Proof: Exactly the same as in the one-dimensional case, except that this time you have m different stores, each of which is using a different linear function to compute the cost of a basket. You find the *m*X*n* cost matrix *C*, where *C[i,j]* is the cost of item *j* at store *i*, by bringing unit baskets of each item time to each cash register.

- For details of the proof see the textbook

Matrix multiplication is a linear transformation

Matrices and linear transformations

Given any $m \times n$ matrix B, we can define a function $\mathbf{g} : \mathbf{R}^n \to \mathbf{R}^m$ (note the order of m and n switched) by $\mathbf{g}(\mathbf{x}) = B\mathbf{x}$, where $\mathbf{x}$ is an n-dimensional vector. As another example, if

$$C = \begin{bmatrix} 5 & -3 \\ 1 & 0 \\ -7 & 4 \\ 0 & -2 \end{bmatrix},$$

then the function $\mathbf{h}(\mathbf{y}) = C\mathbf{y}$, where $y = (y_1, y_2)$, is $\mathbf{h}(\mathbf{y}) = (5y_1 - 3y_2, y_1, -7y_1 + 4y_2, -2y_2)$.

- We can associate a function with any matrix. How ?
- What about the other way round. With any function can you always associate a matrix ? When can you and when you cannot ?

https://mathinsight.org/matrices_linear_transformations

Check for linear transformation

- Given some function $\mathbf{g} : \mathbf{R}^n \to \mathbf{R}^m$

The function $\mathbf{g}(\vec{v})$ is a linear transformation if each term of each component of $\mathbf{g}(\vec{v})$ is a scalar times one of the components of $\vec{v}$.

$$\vec{u} = \begin{bmatrix} a \\ b \end{bmatrix} \vec{v} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$g(\vec{v}) = \begin{bmatrix} 2x + y \\ y \\ x - 3y \end{bmatrix} \qquad cg(\vec{v}) = c\begin{bmatrix} 2x + y \\ y \\ x - 3y \end{bmatrix} = \begin{bmatrix} 2cx + cy \\ cy \\ cx - 3cy \end{bmatrix} = g(c\vec{v})$$

$$g(\vec{u} + \vec{v}) = \begin{bmatrix} 2(a + x) + (b + y) \\ (b + y) \\ (a + x) - 3(b + y) \end{bmatrix} = \begin{bmatrix} 2a + b \\ b \\ a - 3b \end{bmatrix} + \begin{bmatrix} 2x + y \\ y \\ x - 3y \end{bmatrix} = g(\vec{u}) + g(\vec{v})$$

Which functions are linear transformations ?
What is the corresponding matrix?

$$\mathbf{f}(x, y) = (2x + y, y/2)$$

$$\mathbf{g}(x, y, z) = (z, 0, 1.2x)$$

$$\mathbf{f}(x, y) = (x^2, y, x)$$

$$\mathbf{g}(x, y, z) = (y, xyz)$$

$$\mathbf{h}(x, y, z) = (x + 1, y, z)$$

Examples of functions which are not linear transformations

- max(v)
- sort (v)
- product(v)
- Why are they not linear transformations ? Verify if they satisfy the two properties (multiplication by a scalar and operation on sum of vectors; scalar can be any real number)

Find matrix associated with a linear transformation

- $g(\vec{v}) = \begin{bmatrix} 2x + y \\ y \\ x - 3y \end{bmatrix}$

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}.$$

- We want to find A such that $g(\vec{v}) = A \cdot \vec{v}$
- When $\vec{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $A \cdot \vec{v}$ will give the first column of A
- When $\vec{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $A \cdot \vec{v}$ will give the second column of A

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

Operations on linear transformation

If f and g are linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$, then the function $h(\vec{v}) = f(\vec{v}) + g(\vec{v})$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Let F and G be the matrices corresponding to f and g. Then $F + G$ is the matrix corresponding to h.

If c is a constant, then the function $h(\vec{v}) = cf(\vec{v})$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. The matrix corresponding to h is cF.

Identity matrix

The n-dimensional identity function is the function $f(\vec{v}) = \vec{v}$.

The corresponding matrix is the $n \times n$ identity matrix, I_n with 1's on the main diagonal and 0's everywhere else.

$$I_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Composition of linear transformation

Let f be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and let g be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^p$. Let $h = g \circ f$ be the *composition* of f and g, defined by $h(\vec{v}) = g\big(f(\vec{v})\big)$. Then $h(\vec{v})$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^p$.

Proof:

$$h(c \cdot \vec{v}) = g\big(f(c \cdot \vec{v})\big) = g\big(c \cdot f(\vec{v})\big) = c \cdot g\big(f(\vec{v})\big) = c \cdot h(\vec{v})$$

$$h(\vec{u} + \vec{v}) = g\big(f(\vec{u} + \vec{v})\big) = g\big(f(\vec{u}) + f(\vec{v})\big) = g\big(f(\vec{u})\big) + g\big(f(\vec{v})\big) = h(\vec{u}) + h(\vec{v})$$

Composition of linear transformations

Let f be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and let g be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^p$. Let $h = g \circ f$, a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^p$.

Let F be the mXn matrix corresponding to f, and let G be the pXm matrix corresponding to g . Then there is a pXn matrix H corresponding to h. H is said to be the product of G times F: $H = G \cdot F$

The i, j component of H is the dot product of the *i*th row of G with the *j*th column of F.

$$H[i,j] = G[i,:] \blacksquare F[:,j] = \sum_{k=1}^{m} G[i,k] \cdot F[k,j]$$

Product of two matrices

If G is a $p \times m$ matrix and F is an $m \times n$ matrix, then the product $H = G \cdot F$ is the $p \times n$ matrix defined by $H[i,j] = \sum_{k=1}^{m} G[i,k] \cdot F[k,j]$

Two-handed method: left hand moves left-to-right across a row of G, right hand moves top-to-bottom down a column of F, multiply, keep a running tally.

$$\begin{bmatrix} 1 & 3 & -1 & 4 \\ 2 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 3 \\ 0 & 1 & 2 \\ -1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 8 & 16 \\ 1 & 4 & 6 \end{bmatrix}$$

Product of two matrices

For any vector $\vec{v}$, $(G \cdot F) \cdot \vec{v} = G \cdot (F \cdot \vec{v})$. Example:

$$\begin{bmatrix} 1 & 3 & -1 & 4 \\ 2 & 0 & 1 & 1 \end{bmatrix} \cdot \left(\begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 3 \\ 0 & 1 & 2 \\ -1 & 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \right) = \begin{bmatrix} 1 & 3 & -1 & 4 \\ 2 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 7 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 43 \\ 17 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 8 & 16 \\ 1 & 4 & 6 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 43 \\ 17 \end{bmatrix}$$

Matrix Multiplication: when it is allowed ?

- When we allow to multiply the two matrices A and B ?

  $A \cdot B$
- A and B are both square matrices; A is mxm and B is nxn
  - Condition: m = n   output shape: mxm
- A is square and B is rectangular; A is mxm and B is nxp
  - Condition: m=n   output shape: mxp
- A is rectangular and B is square; A is mxn and B is pxp
  - Condition: n=p   output shape: mxp
- Both A and B are rectangular; A is mxn and B is pxq
  - Condition: n=p   output shape: mxq

Important!
Matrix multiplication does not commute.

Suppose that G is a $p \times m$ matrix and F is an $m \times n$ matrix, so $G \cdot F$ is a $p \times n$ matrix.

If $p \neq n$ then $F \cdot G$ is not defined.

If $p = n \neq m$ then $F \cdot G$ is an $m \times m$ matrix while $G \cdot F$ is an $n \times n$ matrix

Even if $p = n = m$, in most cases $F \cdot G \neq G \cdot F$

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 4 & 2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 7 \\ 2 & 1 \end{bmatrix}$$

Matrix Multiplication

We need to define matrix $G \cdot F$ to correspond to $g \circ f$.

$$(g \circ f)(\vec{v}) = g\big(f(v)\big) = G \cdot (F \cdot \vec{v})$$

so that $(G \cdot F) \cdot \vec{v} = G \cdot (F \cdot \vec{v})$

Consider the special case $\vec{v} = \vec{e}^{\,j}$.

$F \cdot \vec{e}^{\,j}$ is just the *j*th column of F

So $G \cdot \left(F \cdot \vec{e}^{\,j}\right)$ is all the dot products of rows of G with the *j*th column of F

But $(G \cdot F) \cdot \vec{e}^{\,j}$ is the *j*th column of $G \cdot F$.

So $(G \cdot F)[i,j] = G[i,:] \blacksquare F[:,j]$

Matrix Multiplication

$$\left(\begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 4 & 2 \end{bmatrix} \cdot \begin{bmatrix} 4 & 0 & 1 & 2 \\ 3 & 5 & -1 & 4 \end{bmatrix}\right) \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 4 & 2 \end{bmatrix} \cdot \left(\begin{bmatrix} 4 & 0 & 1 & 2 \\ 3 & 5 & -1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}\right) =$$

$$\begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 4 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \langle 1, 3 \rangle \bullet \langle 1, -1 \rangle \\ \langle 2, -1 \rangle \bullet \langle 1, -1 \rangle \\ \langle 4, 2 \rangle \bullet \langle 1, -1 \rangle \end{bmatrix}$$

Properties of matrix multiplication

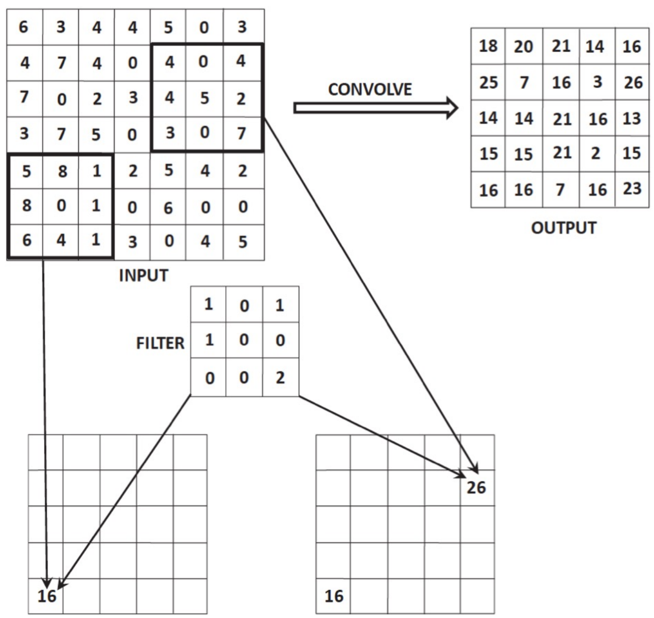
1. $A \cdot (C \cdot E) = (A \cdot C) \cdot E$. (Associativity)
2. $A \cdot (C + D) = (A \cdot C) + (A \cdot D)$. (Right distributivity).
3. $(A + B) \cdot C = (A \cdot C) + (B \cdot C)$. (Left distributivity).
4. $a \cdot (A \cdot B) = (a \cdot A) \cdot B = A \cdot (a \cdot B)$.
5. $A \cdot I_n = A$. (Right identity)
6. $I_n \cdot C = C$. (Left identity)
7. $(A \cdot B)^T = B^T \cdot A^T$.

Simple proof using linear transformations

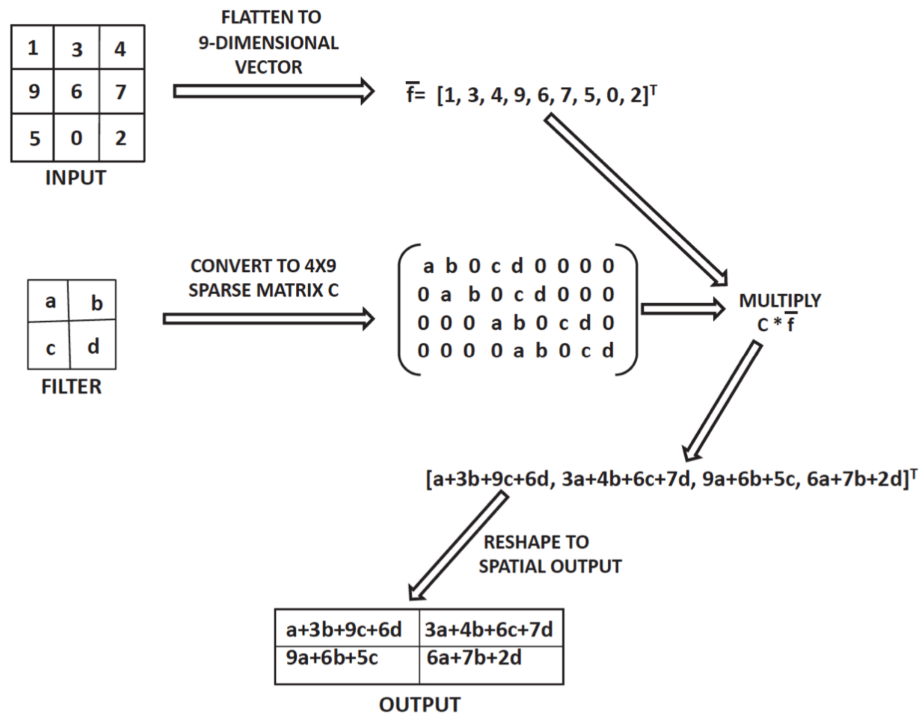
To prove that $A \cdot (C + D) = (A \cdot C) + (A \cdot D)$:

$$
\begin{aligned}
&(a \circ (c + d))(\vec{v}) = \text{(by definition of composition)} \\
&a((c + d)(\vec{v})) = \text{(by definition of } c + d\text{)} \\
&a(c(\vec{v}) + d(\vec{v})) = \text{(by linearity of } a\text{)} \\
&a(c(\vec{v})) + a(d(\vec{v})) = \text{(by definition of composition)} \\
&(a \circ c)(\vec{v}) + (a \circ d)(\vec{v}) = \text{(by definition of } (a \circ c) + (a \circ d)\text{)} \\
&((a \circ c) + (a \circ d))(\vec{v}).
\end{aligned}
$$

Convolution – Basic building block of deep learning models for image classification



Convolution as a Matrix Operation



Inverse Matrix

- $AB = I$ B is *right inverse* of A
- $CA = I$ C is *left inverse* of A
- For any square matrix, if its inverse exists, then left inverse is same as right inverse, denote as A^{-1}
- $AA^{-1} = A^{-1}A = I$
- If a square matrix A has a right inverse B and a left inverse C, then B and C must be the same matrix – Prove it !
- If inverse of matrix exists, matrix is called invertible/non-singular
- Singular matrix has no inverse

When does a matrix has no inverse ?

- If some non-zero combinations of column of a matrix gives 0 then the matrix has no inverse
- A=$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$ -- Does this matrix has an inverse ?
- Can you find a non-zero vector $\vec{v}$ such that $\mathrm{A}\vec{v} = 0$?

$A^{-1}\mathrm{A}\vec{v} = 0$ => $\vec{v}$ =0

- Let X=$\begin{bmatrix} x \\ y \end{bmatrix}$. Recall multiplication of a matrix by a vector

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} x + \begin{bmatrix} 2 \\ 6 \end{bmatrix} y = 0$$

Transpose of an inverse

- Since $AA^{-1} = I$ we can take transpose of both the sides to get
  $(AA^{-1})^T = (A^{-1})^T A^T = I$
- Thus, we get $(A^T)^{-1}$= $(A^{-1})^T$ i.e., *the transpose of an inverse is inverse of the transpose.*

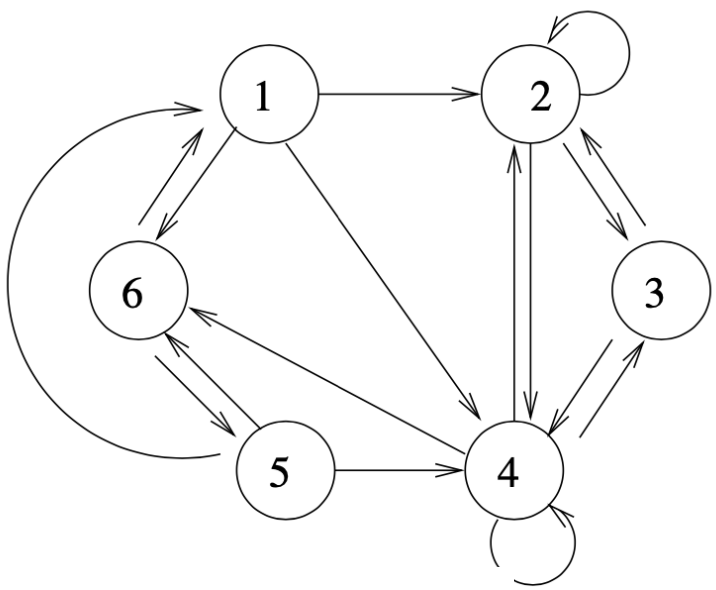
Matrix exponentiation

- For a square matrix $A \cdot A \cdot \ldots \cdot A$ (k times) is denoted as A^k

In the population transition example, let the transition matrix M be constant year after year.
Then if v is the population vector at the start:

- After 1 year the population vector will be $M \cdot \vec{v}$
- After 2 years the population vector will be $M \cdot (M \cdot \vec{v})$ or $M^2 \cdot \vec{v}$
- After k years the population vector will be $M^k \cdot \vec{v}$

Matrix Exponentiation Example



$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

If there is a two-step path exists from node U to node W going through node V then A[U,V]=1 and A[V,W]=1 or in other words

$$A[U,V] \cdot A[V,W] = 1$$

There can be multiple two-step paths between U and W. To find how many such two steps path exists in the graph between nodes U and W we need to sum over all the possible values for the intermediate node V, i.e., we need to find:

$$\sum_{V=1}^{N} A[U,V] \cdot A[V,W]$$

which is [U,W] element of matrix obtained by the product of A with itself, i.e.,

$$\sum_{V=1}^{N} A[U,V] \cdot A[V,W] = A^2[U,W]$$

In general, the number of k step paths between two nodes U and W is given by $A^k[U,W]$

Permutation Matrices

- Any matrix obtained by reordering of rows of identity matrix
- A square binary matrix that has exactly one entry of 1 in each row and each column and 0s elsewhere.

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

- $P^T P = I \Rightarrow P^T = P^{-1}$
- Is identity matrix a permutation matrix ?

Permutations matrices properties

- There are six 3x3 permutation matrices:

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P_{21} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P_{31} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$P_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P_{31}P_{21} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$P_{21}P_{31} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

- 4 of them are transpose of themselves:
  $I, P_{21}, P_{31}, P_{32}$
- $P_{31}P_{21}$ is transpose of $P_{21}P_{31}$
- For any permutation matrix P,
  $$P^T P = I$$
  This implies the inverse of a permutation matrix is its transpose
  $$P^T = P^{-1}$$
- Since transpose of a permutation matrix is also a permutation matrix, the inverse of a permutation matrix is also a permutation matrix
- There are 2 2x2 permutation matrices
  $$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$
- There are n! nxn permutation matrices

Inverse of transpose and transpose of inverse

- Transpose of inverse $(A^{-1})^T$
- Inverse of transpose $(A^T)^{-1}$
- How are they related?
- Left multiply $(A^{-1})^T$ with A^T

$$A^T(A^{-1})^T = (A^{-1}A)^T = \mathrm{I}$$

- Right multiply $(A^{-1})^T$ with A^T

$$(A^{-1})^T A^T = (AA^{-1})^T = \mathrm{I}$$

- $(A^{-1})^T$ is inverse of A^T

Symmetric matrices

- If A is a symmetric matrix then $A^T = A$ like

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 3 & 2 & 7 \\ 2 & 4 & 6 \\ 7 & 6 & 5 \end{bmatrix}$$

- The inverse of a symmetric matrix is also symmetric

  $(A^{-1})^T = (A^T)^{-1}$ and $A^T = A$

  => $(A^{-1})^T = A^{-1}$ hence A^{-1} is also symmetric

- Multiplying any matrix R by its transpose R^T gives a symmetric matrix
  - R can be a rectangular matrix also
  - Transpose of $R^T R$ is $R^T (R^T)^T = R^T R$, thus $R^T R$ is symmetric
- RR^T is also a symmetric matrix
- $R^T R \neq RR^T$ in general

Example

- Say $R = [1\ 2]$ then $R^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$
- $R^T R = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$
- $RR^T = [5]$