CS-565 Computer Vision

Nazar Khan

Department of Computer Science
University of the Punjab

2. Background Mathematics

Notation

- Scalars are denoted by lower-case letters like s, a, b.
- Vectors are denoted by lower-case bold letters like $\mathbf{x}, \mathbf{y}, \mathbf{v}$.
- Matrices are denoted by upper-case bold letters like $\mathbf{M}, \mathbf{D}, \mathbf{A}$.
- Any vector $\mathbf{x} \in \mathbb{R}^d$ is by default a column vector.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

- The corresponding row vector is obtained as $\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \dots & x_d \end{bmatrix}$.

Inner Product

For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$

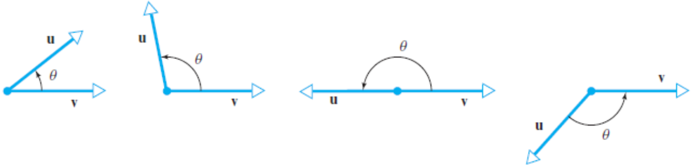
- *Inner product* is a scalar value.

$$\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_dy_d = \|\mathbf{x}\|\|\mathbf{y}\| \cos\theta$$

  where θ is the angle between vectors $\mathbf{x}$ and $\mathbf{y}$.
- Also called *dot product* or *scalar product*. Other representations:

$$\mathbf{x} \cdot \mathbf{y}, (\mathbf{x}, \mathbf{y}) \text{ and } < \mathbf{x}, \mathbf{y} >$$

- Represents similarity of vectors.
  - If $\mathbf{x}^T\mathbf{y} = 0$, then $\mathbf{x}$ and $\mathbf{y}$ are orthogonal vectors (in 2D, this means they are perpendicular).



Euclidean Norm

- *Euclidean norm* of vector

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{x_1x_1 + x_2x_2 + \cdots + x_dx_d}$$

represents the magnitude of the vector.

- *Euclidean distance* between points $\mathbf{x}$ and $\mathbf{y}$ can be computed as

$$\begin{aligned}\|\mathbf{x}-\mathbf{y}\| &= \sqrt{(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y})} \\ &= \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + \cdots + (x_d-y_d)^2}\end{aligned}$$

- *Unit vector* has norm 1. Also called *normalised vector*.
- If $\|\mathbf{x}\| = 1$ and $\|\mathbf{y}\| = 1$, and $\mathbf{x}^T\mathbf{y} = 0$, then $\mathbf{x}$ and $\mathbf{y}$ are *orthonormal vectors*.

Outer Product

For vectors $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{z} \in \mathbb{R}^k$

- *Outer-product* $\mathbf{x}\mathbf{z}^T$ is a $d \times k$ matrix.

$$\mathbf{x}\mathbf{z}^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \begin{bmatrix} z_1 & z_2 & \dots & z_k \end{bmatrix} = \begin{bmatrix} x_1 z_1 & x_1 z_2 & \dots & x_1 z_k \\ x_2 z_1 & x_2 z_2 & \dots & x_2 z_k \\ \vdots & \vdots & \vdots & \vdots \\ x_d z_1 & x_d z_2 & \dots & x_d z_k \end{bmatrix}$$

Matrix and Vector Calculus

For vector $\mathbf{x} \in \mathbb{R}^d$, scalar function $f(\mathbf{x})$ and vector function $\mathbf{g}(\mathbf{x}) \in \mathbb{R}^k$

- The gradient operator $\frac{d}{d\mathbf{x}}$ is also written as $\nabla_\mathbf{x}$ or simply ∇ when the differentiation variable is implied.
- $\nabla_\mathbf{x} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_d} \end{bmatrix}$ so that $\nabla_\mathbf{x}(f(\mathbf{x})) = \frac{d}{d\mathbf{x}}(f(\mathbf{x})) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_d} \end{bmatrix}$
- $\nabla_\mathbf{x}(\mathbf{g}(\mathbf{x})) = \frac{d}{d\mathbf{x}}(\mathbf{g}(\mathbf{x})) = \begin{bmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \frac{\partial g_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial g_k(\mathbf{x})}{\partial x_1} \\ \frac{\partial g_1(\mathbf{x})}{\partial x_2} & \frac{\partial g_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_k(\mathbf{x})}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1(\mathbf{x})}{\partial x_d} & \frac{\partial g_2(\mathbf{x})}{\partial x_d} & \cdots & \frac{\partial g_k(\mathbf{x})}{\partial x_d} \end{bmatrix}$

Matrix and Vector Calculus

For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and matrices $\mathbf{M} \in \mathbb{R}^{k\times d}$ and $\mathbf{A} \in \mathbb{R}^{d\times d}$

- $\nabla_{\mathbf{x}}(\mathbf{y}^T\mathbf{x}) = \nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{y}) = \mathbf{y}$
- $\nabla_{\mathbf{x}}(\mathbf{M}\mathbf{x}) = \mathbf{M}^T$
- $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}$
- For symmetric $\mathbf{A}$, $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}$

Matrices as linear operators

- In a matrix transformation $\mathbf{Mx}$, components of $\mathbf{x}$ are acted upon in a linear fashion.

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} m_{11}x_1 + m_{12}x_2 \\ m_{21}x_1 + m_{22}x_2 \end{bmatrix}$$

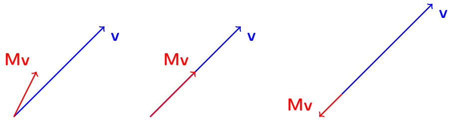
- *Every* matrix multiplication represents a linear transformation.
- *Every* linear transformation can be represented as a matrix multiplication.

Eigenvectors

- When a square matrix $\mathbf{M}$ is multiplied with a vector $\mathbf{v}$, the vector is linearly transformed.
  - Rotation/Shearing/Scaling
  - Scaling does not change the direction of the vector.
- If vector $\mathbf{Mv}$ is only a scaled version of $\mathbf{v}$, then $\mathbf{v}$ is called an *eigenvector of* $\mathbf{M}$.
- That is, if $\mathbf{v}$ is an eigenvector of $\mathbf{M}$ then

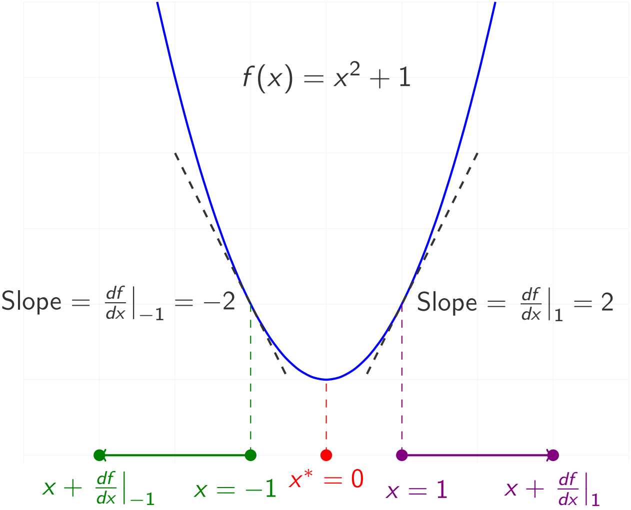
$$\mathbf{Mv} = \lambda \mathbf{v}$$

where scaling factor λ is also called the *eigenvalue of* $\mathbf{M}$ *corresponding to eigenvector* $\mathbf{v}$.



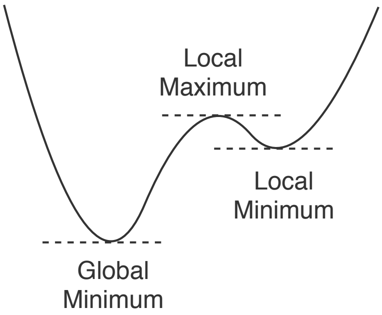
Not an eigenvector Eigenvector Eigenvector

Minimization



What is the slope/derivative/gradient at the minimizer $x^* = 0$?

Minimization
Local vs. Global Minima



- *Stationary point*: where derivative is 0.
- A stationary point can be a minimum or a maximum.
- A minimum can be local or global. Same for maximum.

Constrained Optimization

- For optimizing a function $f(\mathbf{x})$, the gradient of f must vanish at the optimizer $\mathbf{x}^*$.

$$\nabla f|_{\mathbf{x}^*} = \mathbf{0}$$

- For optimizing a function $f(\mathbf{x})$ *subject to some constraint* $g(\mathbf{x}) = 0$, the gradient of the so-called Lagrange function

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$$

must vanish at the optimizer $\mathbf{x}^*$. That is,

$$\nabla L(\mathbf{x}, \lambda) = \nabla f|_{\mathbf{x}^*} + \lambda \nabla g|_{\mathbf{x}^*} = \mathbf{0}$$

where λ is the Lagrange (or undetermined) multiplier.

Constrained Optimization

- Quite often, we will need to maximize $\mathbf{x}^T\mathbf{M}\mathbf{x}$ with respect to $\mathbf{x}$ where $\mathbf{M}$ is a symmetric, positive-definite[1] matrix.
  - Trivial solution: $\mathbf{x} = \mathbf{inf}$
- To prevent trivial solution, we must constrain the norm of $\mathbf{x}$. For example, $\mathbf{x}^T\mathbf{x} = 1$.
- This gives us a constrained optimization problem.

$$\textit{Maximize } f(\mathbf{x}) = \mathbf{x}^T\mathbf{M}\mathbf{x} \textit{ subject to the constraint } g(\mathbf{x}) = \mathbf{x}^T\mathbf{x} - 1 = 0.$$

- Lagrangian becomes $L(\mathbf{x}, \lambda) = \mathbf{x}^T\mathbf{M}\mathbf{x} + \lambda(1 - \mathbf{x}^T\mathbf{x})$
- Use $\nabla_{\mathbf{x}} L|_{\mathbf{x}^*} = \mathbf{0}$ and $\nabla_{\lambda} L|_{\lambda^*} = 0$ to solve for optimal $\mathbf{x}^*$.

[1] $\mathbf{x}^T\mathbf{M}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$

Taylor Series Approximation

- If values of a function $f(a)$ and its derivatives $f'(a), f''(a), \ldots$ are known at a value a, then we can approximate $f(x)$ for x close to a via the *Taylor series expansion*

$$f(x) \approx f(a) + (x-a)^1 \frac{f'(a)}{1!} + (x-a)^2 \frac{f''(a)}{2!} + (x-a)^3 \frac{f'''(a)}{3!} + O((x-a)^4)$$

- For example, for x around $a = 0$
  - $\sin(x) \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$
  - $e^x \approx 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots$
- It is often convenient to use the first-order Taylor expansion

$$f(x) \approx f(a) + (x-a)f'(a)$$

Taylor Series Approximation

Not very useful for x not close to a

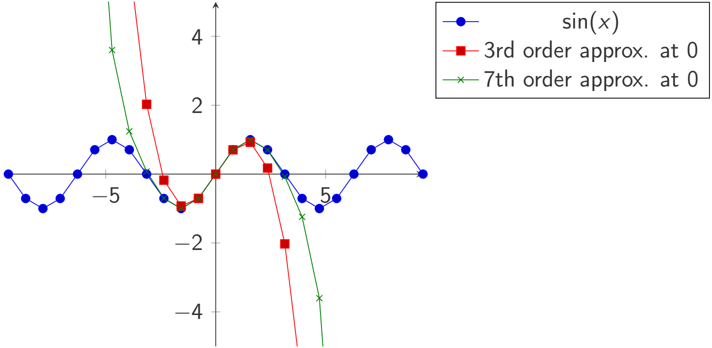
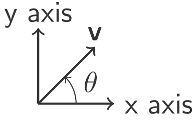


Figure: The sin() function (blue) is closely approximated around 0 by its Taylor polynomials. The 7th order approximation (green) is good for a full period centered at 0. However, it becomes poor for $|x - 0| > \pi$.

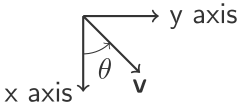
Image Coordinates

Cartesian axis



+ve x-axis from left to right.
+ve y-axis goes upwards.

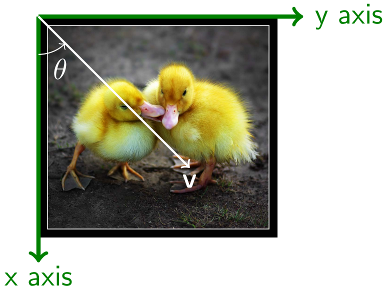
Image axis

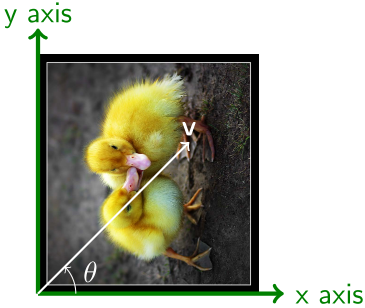


+ve x-axis goes downwards.
+ve y-axis from left to right.

For both coordinate systems, angles are always measured in counter-clockwise direction from positive x-axis.

Image Coordinates





Rotate by 90° anticlockwise

By rotating the axis, the mathematics on the image axes will remain the same as for the Cartesian axes. For example, a line in the image can still be represented via $y = mx + c$ and slope $m = \tan\theta$.

Take-home Quiz 1

1. **(4 marks)** For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and matrices $\mathbf{M} \in \mathbb{R}^{k \times d}$ and $\mathbf{A} \in \mathbb{R}^{d \times d}$, prove the following derivatives.
   1.1 $\nabla_{\mathbf{x}}(\mathbf{y}^T\mathbf{x}) = \nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{y}) = \mathbf{y}$
   1.2 $\nabla_{\mathbf{x}}(\mathbf{M}\mathbf{x}) = \mathbf{M}^T$
   1.3 $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}$
   1.4 $\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}$ when $\mathbf{A}$ is symmetric
2. **(6 marks)** For a symmetric, positive-definite matrix $\mathbf{A}$, show that the non-trivial maximizer of $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is the eigenvector of $\mathbf{A}$ corresponding to the largest eigenvalue.