Linear Algebra & Calc

YUE WU

What is a Matrix?

Product

Inverse

Invertibility

Positive Semi-definite Matrices

Inverse

https://www.mathsisfun.com/algebra/matrix-inverse.html

Invertibility

In linear algebra, an *n*-by-*n* square matrix **A** is called **invertible** (also **nonsingular** or **nondegenerate**) if there exists an *n*-by-*n* square matrix **B** such that

$$AB = BA = I_n$$

where $\mathbf{I}_n$ denotes the *n*-by-*n* identity matrix and the multiplication used is ordinary matrix multiplication. If this is the case, then the matrix **B** is uniquely determined by **A** and is called the ***inverse*** of **A**, denoted by $\mathbf{A}^{-1}$.

Eigenvalues and Eigen Vectors

In linear algebra, an **eigenvector** (/ˈaɪɡənˌvɛktər/) or **characteristic vector** of a linear transformation is a nonzero vector that changes at most by a scalar factor when that linear transformation is applied to it.

Now consider the linear transformation of n-dimensional vectors defined by an n by n matrix A,

$$Av = w$$

If it occurs that v and w are scalar multiples, that is if

$$Av = w = \lambda v$$

then *v* is an **eigenvector** of the linear transformation *A* and the scale factor λ is the **eigenvalue** corresponding to that eigenvector. Equation (1) is the **eigenvalue equation** for the matrix *A*.

Eigenvalues and Eigen Vectors

$$Av = w = \lambda v$$

Can be stated equivalently as

$$(A - \lambda I)v = 0$$

where *I* is the *n* by *n* identity matrix and 0 is the zero vector.

SVD

https://en.wikipedia.org/wiki/Eigendecomposition_of_a_matrix

https://en.wikipedia.org/wiki/Singular_value_decomposition

Positive Definite.

A linear transformation $T: \mathbb{R}^d \to \mathbb{R}^d$ is called positive semi-definite if

$\forall v \in \mathbb{R}^d, \ (Tv) \cdot v \geq 0$

Note: A symmetric matrix is positive semi-definite iff. all eigen values≥ 0

What is Calc?

We say f is differentiable at a if $\lim_{h \to 0} \frac{f(a+h)-f(a)}{h}$ exists ($\lim_{x \to a} \frac{f(x)-f(a)}{x-a}$ exists)

Notation: $f'(a) = \lim_{x \to a} \frac{f(x)-f(a)}{x-a}$

Derivative = 0

$f'(x) = 0$ when f attains local min/max

Pf for min. Assume not. WLOG $f'(a) > 0$

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} > 0$$

$$\exists \delta > 0: |h| < \delta \Rightarrow \left| \frac{f(a+h) - f(a)}{h} - f'(a) \right| < f'(a)$$

$$\frac{f(a+h) - f(a)}{h} > 0 \Rightarrow$$

If $h < 0$: $f(a+h) < f(a)$

Contradiction

$$f''(a) = \lim_{h \to 0} \frac{f'(a+h) - f\prime(a)}{h} > 0$$

$$\exists \delta: |h| < \delta \Rightarrow \frac{f'(a+h) - f'(a)}{h} > 0$$

$$h > 0 \Rightarrow f'(a+h) > f'(a) = 0$$

$$h < 0 \Rightarrow f'(a+h) < f'(a) = 0$$

Suppose $f'(a)$ is not local min, $\exists h': f'(a+h') < f(a)$

WLOG we consider $h' > 0$, use MVT

$$\exists k \in (0, h'): f'(a+h) = \frac{f(a+h') + f(a)}{h\prime}$$

Extension to MultiDim

We say $f: \mathbb{R}^d \rightarrow \mathbb{R}$ is differentiable at a, if $\exists a$ a linear transformation $T: \mathbb{R}^d \rightarrow \mathbb{R}$

$$\lim_{\mathrm{h} \to 0} \frac{|f(a+h) - f(a) - Th|}{|h|} = 0$$

T is called the derivative of f.

Notation:

T alone is called the derivative of f at a denoted by Df_a (Δf)

Note: If f differentiable, derivative is unique

Directional Derivatives:

Def: The directional derivative of f at a in a direction v is defined to be

$$D_v f(a) = \frac{d}{dt}\big(f(a+tv)\big)\Big|_{t=0}$$

If $v = e_i$ (i^{th} basis vector). $D_{e_i} f(a)$ is called the i^{th} partial derivative of f at a. Notation: $\partial_i f(a)$ or $\frac{\partial f(a)}{\partial x_i}$

Note: $\partial_1 f(a) = \lim_{h\to 0} \frac{f(a_1+h, a_2, a_3, \dots) - f(a)}{h}$

Prop: If f is differentiable at a then all directional derivatives (at a) exist (Need not be continuous).

Moreover $D_v f(a) = Df_a(v)$

Attaching the Proof for Completeness

Prop: If f is differentiable at a then all directional derivatives (at a) exist (Need not be continuous).

Moreover $D_v f(a) = Df_a(v)$

Pf: WTS $\lim_{t\to 0} \frac{f(a+tv)-f(a)}{t} = (Df_a)v$

Know $\lim_{h\to 0} \frac{|f(a+h)-f(a)-(Df_a)h|}{|h|}$, so we plug in $h = tv$

$$\lim_{t\to 0} \frac{|f(a+tv)-f(a)-t(Df_a)v|}{|t||v|} = \frac{1}{|v|} \lim_{t\to 0} \frac{|f(a+tv)-f(a)-t(Df_a)v|}{|t|} = \frac{1}{|v|} \lim_{t\to 0} \left| \frac{f(a+tv)-f(a)}{t} - Df_a(v) \right| = 0$$

Jacobian

If f: $\mathbb{R}^d \to \mathbb{R}$ is differentiable at a

Df_a: $\mathbb{R}^d \to \mathbb{R}$ is a linear transformation

$$Df_a = (Df_a(e_1), \dots, Df_a(e_d))$$

y Differentiable at $a \Rightarrow Df_a = (\partial_1 f(a), \dots, \partial_d f(a))$

We call the matrix of partials the Jacobian.

$$Df_a = \begin{bmatrix} \partial_1 f_1(a) & \partial_2 f_1(a) & \dots & \partial_n f_1(a) \\ \vdots & \ddots & & \vdots \\ \partial_1 f_n(a) & \partial_2 f_n(a) & \dots & \partial_n f_n(a) \end{bmatrix}$$

Chain Rule

$g: \mathbb{R}^d \to \mathbb{R}^m$ Differentiable at $a \in \mathbb{R}^d$

$f: \mathbb{R}^m \to \mathbb{R}^n$ Differentiable at $g(a) \in \mathbb{R}^m$

Theorem: $f \circ g$ **differentiable at** a **and** $D(f \circ g)_a = (Df)_{g(a)} Dg_a$

Non-math version: f some function of y, y some function of x

$$\frac{\partial f}{\partial x_i} = \partial_i (f \circ g) = \sum_{j=1}^{m} \frac{\partial f}{\partial g_j} \frac{\partial g}{\partial x_i}$$

Higher Order Partials

$f: \mathbb{R}^d \to \mathbb{R}$

$\partial_i f: \mathbb{R}^d \to \mathbb{R}$

$\partial_j(\partial_i f)$ is the second order derivative of f

Taylor’s Theorem

Theorem (Talor's): Suppose f is a $C^n(\mathbb{R}^d)$ function, then for any $a, h \in \mathbb{R}^d$, $\exists R_n : \mathbb{R}^d \to \mathbb{R}$ s.t.

$$f(a+h) = f(a) + \sum_i \partial_i f(a) h_i + \frac{1}{2} \sum_{i,j} \partial_i \partial_j f(a) h_i h_j + R_n(h)$$

Where $\lim_{h \to 0} \frac{R_n(h)}{|h|^n} = 0$

Introduce our Guests

Gradient $\nabla f = (Df)^T = \left(\partial_1 f(a), \dots, \partial_d f(a)\right)^{\mathrm{T}}$

Hessian $Hf_a = \begin{bmatrix} \partial_1\partial_1 f & \dots & \partial_1\partial_d f \\ \vdots & \ddots & \vdots \\ \partial_d\partial_1 f & \dots & \partial_d\partial_d f \end{bmatrix}$

Local Min/Max in $\mathbb{R}^d$

Let $f: \mathbb{R}^d \to \mathbb{R}$

$U \subseteq \mathbb{R}^d$ compact. We say f attain a local min at $a \in U$ if $\exists \epsilon > 0, \forall x \in B(a, \epsilon), f(x) \geq f(a)$

suppose f differentiable and attains a local min at a ,then $\nabla f(a) = 0$

Pf. If f has a local min at a $+>$ $\forall v \in \mathbb{R}^d, v \neq 0$

consider $g(t) = f(a + tv)$ has a local min at $t = 0$

$g'(t) = \sum_{i=1}^{d} \partial_i f(a + tv) \, \frac{d}{dt}(a_i + tv_i) = v \times \nabla f(a + tv)$

Hessian

If f is $C^2(U)$ and f allows a local min at a then $\nabla f(a) = 0$ and Hf_a is positive semi-definite

Pf.

$g''(0) \geq 0$

$g''(t) = \frac{d}{dt}\left(\sum_i v_i \partial_i f(a + tv)\right) - \sum_{i,j} v_i v_j \partial_i \partial_j f(a + tv)$

$g''(0) \geq 0$ iff $\sum_{i,j} v_i v_j \partial_i \partial_j f(a) \geq 0\,,\ \forall v$

$\Rightarrow\ Hf_a$ positive semi-definite

Thank You

SURPRISE COMING AFTER THIS SLIDE

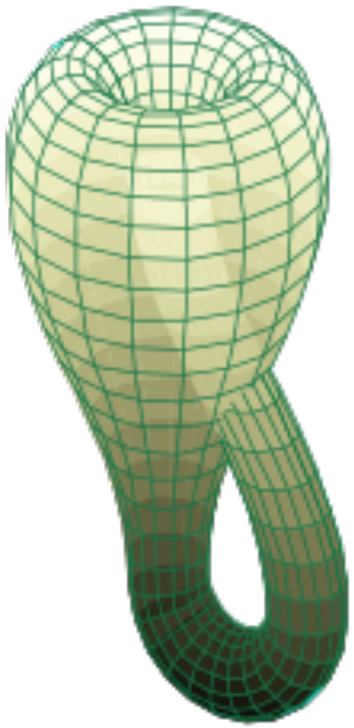
Manifolds

Def: $M \subseteq \mathbb{R}^d$ is called an m-dim manifold if $\forall x \in M, \exists U \subseteq \mathbb{R}^d$

s.t.

1. $x \in U$

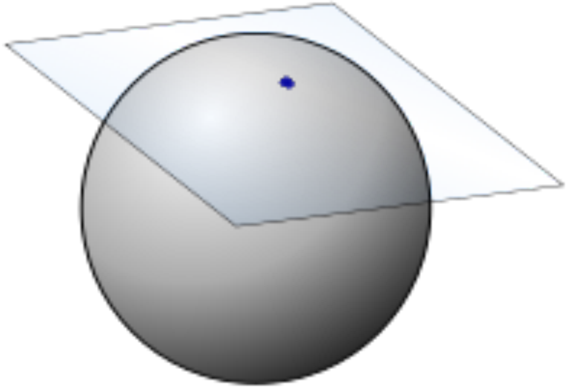
2. $\exists \varphi: U \to B(0,1) \subseteq \mathbb{R}^d$
   - Such that φ is a coordinate change transformation (C^1, bijective, $D\varphi$ invertible) & $\varphi(\mathrm{M} \cap U) = B(-,1) \cap \{x \in \mathbb{R}^d | x_2 = x_3 = \cdots = x_m = 0\}$



Theorem

1. A 1 dim manifold in $\mathbb{R}^d$ is called a curve
2. An "orientable" 2 dim manifold is called a surface

Tangent Spaces



Tangent Plane

$f: \mathbb{R}^2 \rightarrow \mathbb{R}$

$S = \{x \in \mathbb{R}^3 | x_3 = f(x_1, x_2)\}$ (2d manifold, a surface)

$$x_3 = f(a_1, a_2) + Df_{(a_1,a_2)} \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix}$$

Tangent space = Tangent plane shifted to pass through the origin

Tangent Spaces

Def (Tangent Space):

Let $M \subseteq \mathbb{R}^d$ be a m-dim manifold

Let $a \in M$, the tangent space of M at a is defined as follows.

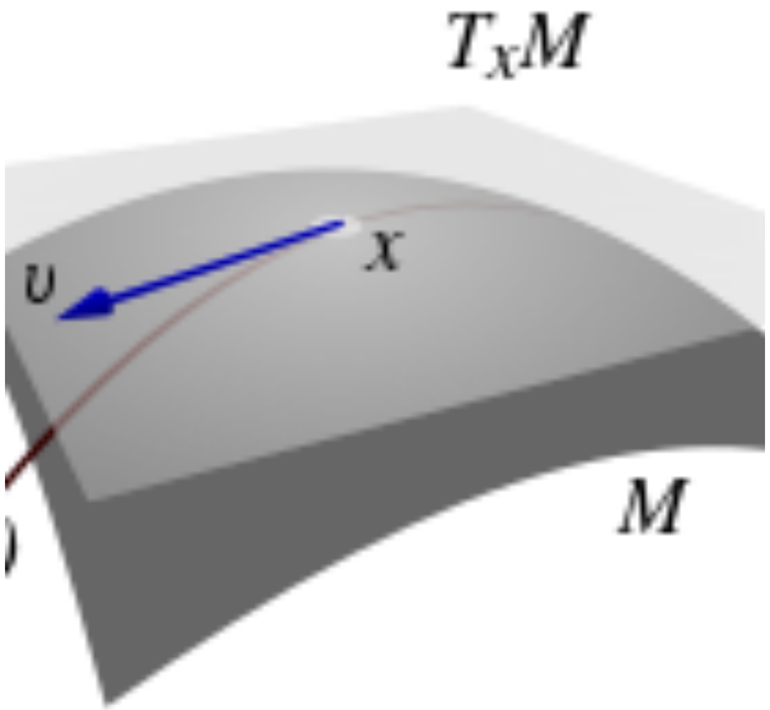
- $\exists U \ni a$ open, & $\phi: U \to B(0,1) \subseteq \mathbb{R}^d$ C^1 diffeomorphic (C^1, bijective, inverse C^1) s.t. $\phi(M \cap U) = B(0,1) \cap (\mathbb{R}^m \times \mathbf{0})$, $\mathbf{0} \in \mathbb{R}^{d-n}$
- Let $\psi = \phi^{-1}$ (WLOG $\psi(0) = a$)
- Define TM_a = Tangent space of M at a $= D\psi(\mathbb{R}^m \times \mathbf{0})$

Tangent space of $S = \{x \in \mathbb{R}^3 | x_3 = f(x_1, x_2)\}$ at the point $a = \left(a_1, a_2, f(a_1, a_2)\right)$ is defined to be

$$\left\{x \in \mathbb{R}^3 \middle| x_3 = \nabla f(a).\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right\} = \{x \in \mathbb{R}^3 | x_3 = \partial_1 f(a) x_1 + \partial_2 f(a) x_2\}$$

Note that the tangent space is a subspace of $\mathbb{R}^3$

Basis: $\begin{pmatrix} 1 \\ 0 \\ \partial_1 f(a) \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \partial_2 f(a) \end{pmatrix}$



Lagrange Multipliers (Constrained Optimization)

Say $f\colon \mathbb{R}^3 \to \mathbb{R}$, (want to minimize/maximize f), $g\colon \mathbb{R}^3 \to \mathbb{R}$ a constraint.

Goal: maximize/minimize f on the manifold $M = \{g = c\}$ (usually written as $\{g = 0\}$ for simplicity)

https://en.wikipedia.org/wiki/Lagrange_multiplier

Lagrange Multipliers

Theorem. If f attains a constrained mim/max subject to the constraint $g = c$, then at all points a at which the constrained local min/max is attained, we have:

$\exists \lambda_1, \ldots, \lambda_n : \nabla f(a) = \sum_i \lambda_i \nabla g_i(a)$ ($m + n + n$ variables, $m + n$ equations)

$g(a) = c$ (n equations, $m + n$ variables)

($m + 2n$ variables and $m + 2n$ equations in total)

Proof

Lemma 1: If f attains a constrained max/min at $a \in M$, then $\nabla f \perp TM_a$ (i.e. $(\nabla f) \cdot v = 0$)

Lemma 2: Let $v \in \mathbb{R}^d$ be any vector, $v \perp TM_a, \Leftrightarrow \exists \lambda_1, \dots, \lambda_n: v = \sum_i \lambda_i \nabla g_i(a)$ $(v \in span(\nabla g_1, \dots, \nabla g_n))$

v is called a **normal vector**

Note that lemma 1+ lemma 2⇒ if φ attains a local min/max at a on M then 1. (Assume $g \in C^1, \forall x$: $rank(Dg_x) = n$)

$\nabla f - \sum \lambda_j \nabla g_j(a) = 0$, 2. $g(a) = c$

Pf. of **Lemma 1.**

Say f attains a constrained local min/max at $a \in M$. Use implicit function theorem and write M locally as the graph of some C^1 function.

Case 1. Assume that $M = \{(x, h(x)) | x \in U\}$, where

$U \subseteq \mathbb{R}^m$ open

$h: \mathbb{R}^m \to \mathbb{R}^n$ C^1,

In this case we know: $TM_a = Im\begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{Dh} \end{pmatrix} = span\{\begin{pmatrix} e_1 \\ \partial_1 h(b) \end{pmatrix}, \dots, \begin{pmatrix} e_m \\ \partial_m h(b) \end{pmatrix}\}$, where I the $m \times m$ identity, Dh $n \times m$ matrix

Let $a = h(b)$. Note that the function $x \to f(x)$ has a local max/min at $a \Leftrightarrow y \to f(y, h(y))$ has an unconstrained min/max at b. $\forall i \in \{1, \dots, M\}, \frac{\partial}{\partial y_i}(f(y, h(y)))\Big|_{y=b} = 0 \Leftrightarrow$ Let $F(y) = f(y, h(y)), DF_b = Df_a \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{Dh_b} \end{bmatrix} = 0$. $\Leftrightarrow \nabla f(a)$ orthogonal to each column of $\begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{Dh_b} \end{bmatrix}$, the basis of TM_a

Case 2. Just permute coordinate

*Pf of **Lemma 2***

Know $TM_a = \ker(Dg_a) \Leftrightarrow \forall v \in TM_a, (Dg_a)v = 0 \Leftrightarrow \nabla g_i(a) \cdot v = 0$

Let $V = \{v | v \perp TM_a\}$, $\dim(v) = d - \dim(TM_a)$, and ∇g_i are linearly independent in V, $V = span\{\nabla g_i\}$