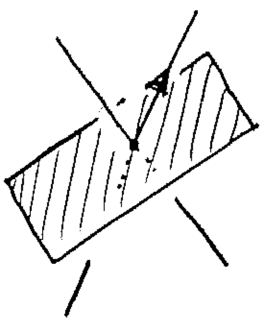
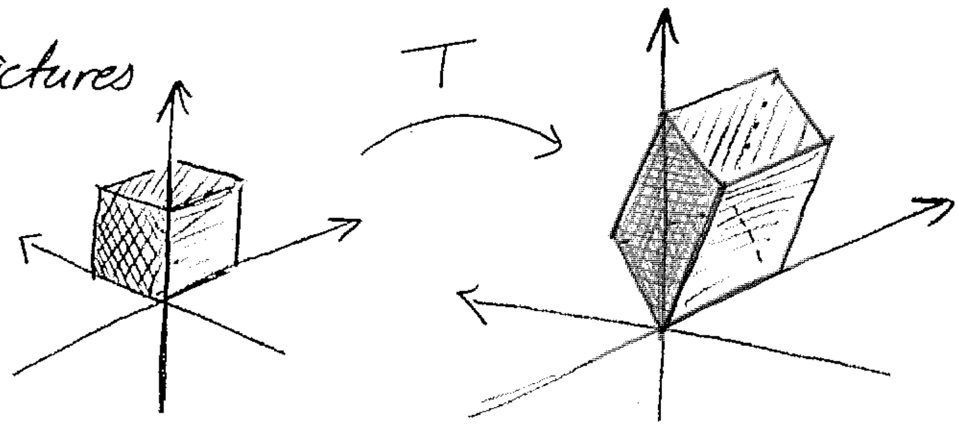
Lecture 40: Linear approximation, diagonalization, and the 2^{nd} derivative test.

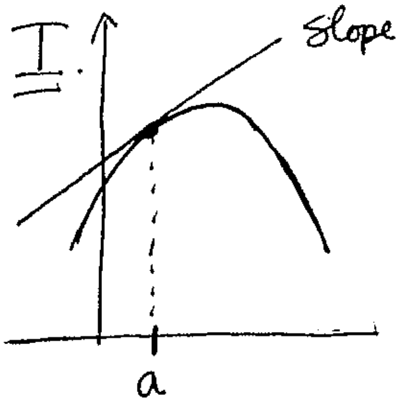
Linear algebra in pictures



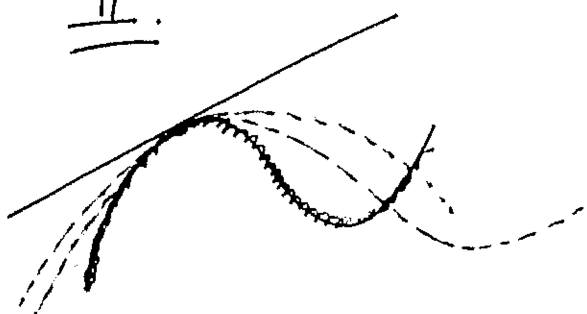


(Differential) Calculus in pictures.

I.



II.

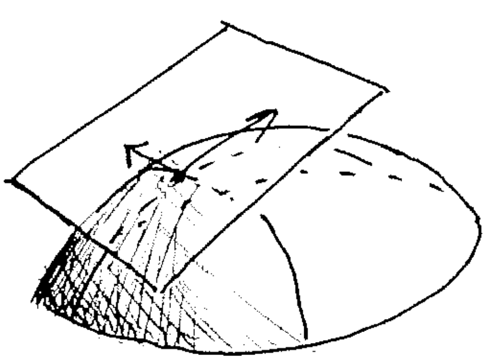


$f: \mathbb{R} \to \mathbb{R}$

$$f(a+h) \approx f(a) + f'(a)h$$

$$f(a+h) \approx f(a) + f'(a)h + \frac{f''(a)}{2}h^2 + \frac{f^{(3)}(a)}{3!}h^3 + \dots$$

III. $f: \mathbb{R}^2 \longrightarrow \mathbb{R}$



$$\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}$$

What about $f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$? Thinking of f as $(f_1, \dots, f_m)$ with $f_i : \mathbb{R}^n \to \mathbb{R}$ have $\dfrac{\partial f_i}{\partial x_j} : \mathbb{R}^n \to \mathbb{R}$ but is there more?

<u>Def</u>: A function $f: \mathbb{R}^n \to \mathbb{R}^m$ is <u>differentiable</u> at $a \in \mathbb{R}^n$ when there exists a linear transformation $D_a f : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(a+h) = f(a) + D_a f(h) + E(h)$$

where $\displaystyle\lim_{h \to 0} \frac{E(h)}{\|h\|} = 0$.

[<u>Moral</u>: f is differentiable at a if it can be well-approximated by a linear transformation (+ const) near a.]

called the derivative of f at a.

<u>Thm</u>: If f is differentiable at a, then $D_a f$ above is unique and

$$\left[D_a f\right]_{\mathrm{Std}(\mathbb{R}^n)}^{\mathrm{Std}(\mathbb{R}^m)} = \underbrace{\begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}}_{\text{Jacobian matrix}}$$

Thm: If all $\dfrac{\partial f_i}{\partial x_{j_1} \cdots \partial x_{j_k}}$ exist on all of $\mathbb{R}^n$, then f is differentiable at all $a \in \mathbb{R}^n$.

Ex: $f(x,y,z) = \left(x^2 + 3y + \sin(z),\ 3x + z^2,\ y + \cos(x+z^2) \right)$ is a differentiable function $\mathbb{R}^3 \to \mathbb{R}^3$.

Chain Rule: Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a and $g : \mathbb{R}^m \to \mathbb{R}^k$ is differentiable at $f(a)$. Then $g \circ f : \mathbb{R}^n \to \mathbb{R}^k$ is diff at a with

$$D_a(g \circ f) = \left(D_{f(a)} g \right) \circ \left(D_a f \right)$$

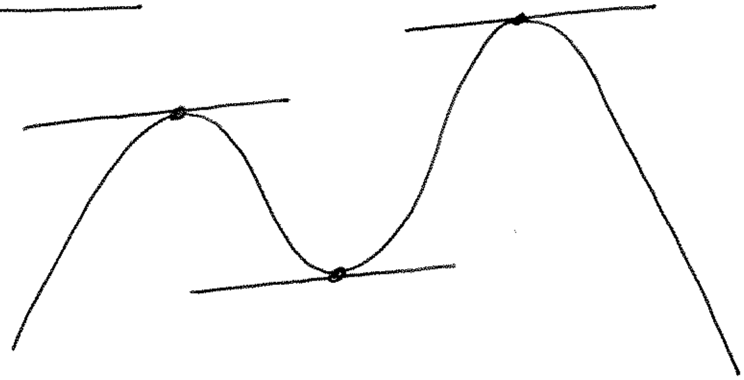
[**Compare**: Converting to Jacobian matrices, this gives the special cases of the chain rule you learned about in Calculus III.]

[**Other connections**: Change of variables formula in multivariable integration...]

2^{nd} derivative test:

1-variable

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



Critical points: $f'(a) = 0$

Test:

$f''(a) > 0$: local min.

$f''(a) < 0$: local max.

$f''(a) = 0$: inconclusive.

2-variables: $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$

Critical points: $\nabla f(a) = \underbrace{\left(\frac{\partial f}{\partial x}(a), \frac{\partial f}{\partial y}(a) \right)}_{= [D_a f]^{std}_{std}} = 0$

Test: Set $M = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(a) & \frac{\partial^2 f}{\partial x \partial y}(a) \\ \frac{\partial^2 f}{\partial y \partial x}(a) & \frac{\partial^2 f}{\partial y^2}(a) \end{pmatrix}$

Then $\det(M) > 0$ and $\frac{\partial^2 f}{\partial x^2}(a) > 0$: local min

and $\frac{\partial^2 f}{\partial x^2}(a) < 0$: local max

$\det(M) < 0$: saddle

$\det(M) = 0$ inconclusive.

<u>n-variables</u>: $f : \mathbb{R}^n \longrightarrow \mathbb{R}$

<u>Critical Points</u>: $\nabla f(a) = 0 \iff D_a f = 0$. For a crit pt a, define $M \in M_{n\times n}(\mathbb{R})$ by $M_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(a)$

<u>Key</u>: If f is reasonably smooth at a, then $M^t = M$. By Thm, M is diagonalizable via an orthonormal Q: $Q^t M Q = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$

<u>Test</u>: If all $\lambda_i > 0$, then a is a local min. If all $\lambda_i < 0$, then a is a local max. If some $\lambda_i > 0$ but other $\lambda_j < 0$ then a is a saddle. If some $\lambda_i = 0$, who knows?

<u>Exercise</u>: Check that the test in 2^d follows from this.

Proof idea: Suppose a is a critical point of f.

Then it turns out that right Taylor series is

$$f(a+h) = f(a) + \frac{1}{2}h^t M h + E(h)$$

where $\lim_{h\to 0} \frac{E(h)}{\|h\|^2} = 0$. Changing coordinates

via $h' = Q^t h \Rightarrow h = Qh'$ and so

$$\begin{aligned} f(a+h) &= f(a) + \frac{1}{2}(Qh')^t M (Qh') + E'(h') \\ &= f(a) + \frac{1}{2}(h')^t \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} h' + E'(h') \\ &= f(a) + \frac{1}{2}(\lambda_1 x_1^2 + \dots \lambda_n x_n^2) + \underbrace{E'(h')}_{\text{small}} \end{aligned}$$

if $h' = (x_1, \dots, x_n)$.

Thus if all $\lambda_i > 0$ we have that a is a local min and similarly for the other cases.

$\blacksquare$

Note: We're actually using here that Q is orthognal to understand how the mixed partials transform under this change of coordinates...