Multivariable Calculus

Jeremy Irvin and Daniel Spokoyny

Derivative

- Let $E \subseteq \mathbb{R}$ be open. A function $f : E \to \mathbb{R}^m$ is differentiable at $x \in E$ if

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

exists. We call this value *f'(x)*.

- Intuitively, a function is differentiable if it is locally approximated by a line.

Derivative

- If f is differentiable, we can rewrite this as

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

- Or equivalently,

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

Derivative

- Now let $E \subseteq \mathbb{R}^n$ be open, and $f : E \to \mathbb{R}^m$. Then f is differentiable at $x \in E$ if $\exists f'(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$\lim_{h \to 0} \frac{||f(x+h) - f(x) - f'(x)h||_{\mathbb{R}^m}}{||h||_{\mathbb{R}^n}} = 0$$

- Intuitively, f is differentiable if it is locally approximated by a linear function.

Partial Derivative

$E \subseteq \mathbb{R}^n$

- Let $f : E \to \mathbb{R}$. The jth partial derivative of f at $x \in E$ is

$$\lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t} = D_j f(x) = \frac{\partial f}{\partial x_j}$$

provided this limit exists.

Total and Partial Derivative

- Now let $E \subseteq \mathbb{R}$ and $f : E \to \mathbb{R}^m$. We can write

$$f(x) = (f_1(x), \ldots, f_m(x))$$

- And if $E \subseteq \mathbb{R}^n$ and $f : E \to \mathbb{R}^m$, writing the component of *x* out explicitly,

$$f(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))$$

Total and Partial Derivative

- So we can define

$$\frac{\partial f_i}{\partial x_j} = D_j f_i(x) = \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t}$$

- Because *f'(x)* is linear, it can be represented as a matrix, call it $A \in \mathbb{R}^{m \times n}$. In fact,

$$A_{ij} = \frac{\partial f_i}{\partial x_j} = D_j f_i(x)$$

Total and Partial Derivative

- If *f* is real-valued ($f : E \to \mathbb{R}$), then the matrix representation of *f'(x)* is called the <u>gradient</u> of *f* at *x*, denoted

$$\nabla f(x_1, \ldots, x_n) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

- Think of the gradient as a direction in which the parameters move so that the function *f* increases the fastest.

Convexity

- Let $E \subseteq \mathbb{R}$. A function $f : E \to \mathbb{R}$ is <u>convex</u> if for all $x_1, x_2 \in E, t \in [0, 1]$,

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

- This means that the line between any two points on the graph of *f* lies above the graph of *f* (see blackboard).

- **Convex functions have a single global optimum.**

Lagrange Multipliers

Jeremy Irvin and Daniel Spokoyny

Lagrange Multipliers

- Suppose $g : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, and let *M* be the set of points $x \in \mathbb{R}^n$ such that $g(x) = 0$ and $\nabla g(x) \neq 0$. If the differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ attains its maximum or minimum on *M* at the point $a \in M$, then

$$\nabla f(a) = \lambda \nabla g(a)$$

λ is called the "Lagrange multiplier".

Lagrange Multiplier Example

- Find the rectangular box of volume 1000 which has the least total surface area A.
- Let $A = f(x, y, z) = 2xy + 2xz + 2yz$ and $g(x, y, z) = xyz - 1000$.
- We want to minimize f on the set of points which satisfy $g(x, y, z) = 0$.
- Sounds like Lagrange Multipliers!

Lagrange Multiplier Example

- $\nabla f = (2y + 2z, 2x + 2z, 2x + 2y)$
- $\nabla g = (yz, xz, xy)$
- We want to solve

$$\begin{aligned} 2y + 2z &= \lambda yz \\ 2x + 2z &= \lambda xz \\ 2x + 2y &= \lambda xy \\ xyz &= 1000 \end{aligned}$$

- It is easily seen that the unique solution to this set of equations is $x=y=z=10$.

Generalized Lagrange Multipliers

- Informally, given some constraints $g_1(x) = 0, \ldots, g_m(x) = 0$, and denoting *M* the set of points which satisfy them, if (under some conditions on these constraints) the differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ attains a local maximum or minimum on *M* at $a \in M$, then

$$\nabla f(a) = \lambda_1 \nabla g_1(a) + \cdots + \lambda_m \nabla g_m(a)$$

- So to find points which optimize *f* given some constraints, simply solve the set of equations above.

Matrix Calculus

Matrix Gradient

- Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ (input matrix, output real value). The gradient of f with respect to some input $A \in \mathbb{R}^{m \times n}$ is the matrix of partial derivatives:

$$\nabla_A f(A) \in \mathbb{R}^{m \times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \cdots & \frac{\partial f(A)}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \cdots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

- More compactly,

$$(\nabla_A f(A))_{ij} = \frac{\partial f(A)}{\partial A_{ij}}$$

Matrix Derivative Properties

$$
\begin{aligned}
\nabla_A \operatorname{tr} AB &= B^T \\
\nabla_{A^T} f(A) &= (\nabla_A f(A))^T \\
\nabla_A \operatorname{tr} ABA^T C &= CAB + C^T AB^T \\
\nabla_A |A| &= |A| (A^{-1})^T.
\end{aligned}
$$

Hessian

- Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$. The <u>Hessian</u> matrix with respect to *x* is the *n x n* matrix of partial derivatives:

$$\nabla_x^2 f(x) \in \mathbb{R}^{n\times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

- More compactly,

$$(\nabla_x^2 f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

Gradients/ Hessians of Quadratic/ Linear Functions

- $\nabla_x b^T x = b$
- $\nabla_x x^T A x = 2Ax$ (if A symmetric)
- $\nabla_x^2 x^T A x = 2A$ (if A symmetric)

What Just Happened?

- Multivariable Derivative
- Convexity
- Lagrange Multipliers
- Matrix Calculus

Done with math!! (kinda)