Linear Algebra and Multivariable Calc.

We are now going to weave together some ideas from calculus and linear algebra.

Definition. A linear map $V: \mathbb{R}^n \to \mathbb{R}$ is called a <u>linear functional on $\mathbb{R}^n$</u>.

Proposition. Every linear functional V on $\mathbb{R}^n$ can be written as $V\vec{\omega} = \langle \vec{\omega}, \vec{v} \rangle$ for some vector $\vec{v} \in \mathbb{R}^n$.

Proof. V is a $1 \times n$ matrix, $\vec{v} = V^T$ is an $n \times 1$ column vector.

and $\langle \vec{\omega}, \vec{v} \rangle = \vec{v}^T \vec{\omega} = (V^T)^T \vec{\omega} = V\vec{\omega}$. $\square$

We already know a linear functional from calculus class.

Definition. The <u>directional derivative</u>

$$(D_{\vec{v}} f)(\vec{x}_0) = \lim_{h \to 0} \frac{f(\vec{x}_0 + h\vec{v}) - f(\vec{x}_0)}{h}.$$

Notice that $D_{\vec{v}} f$ is a scalar.

Prop. $(D_{\vec{v}} f)(\vec{x})$ is linear in $\vec{v}$.

Proof. We learned in calculus that

$$(D_{\vec{v}} f)(\vec{x}_0) = \sum_j v_j \frac{\partial f}{\partial x_j}(x_0)$$

$$= \left\langle \vec{v}, \begin{bmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix} \right\rangle = \langle \vec{v}, \nabla f \rangle$$

where the column vector of partials of f is called the gradient vector. $\square$

Homework. Prove that

$$\max_{\{\vec{v} \mid \|\vec{v}\|=1\}} D_{\vec{v}} f = \|\nabla f\|$$

(the norm of the gradient is the ~~st~~ largest rate of ascent) and

$$\operatorname{argmax}_{\{\vec{v} \mid \|\vec{v}\|=1\}} D_{\vec{v}} f = \frac{\nabla f}{\|\nabla f\|}$$

(the direction of the gradient is the direction of steepest ascent), using

$$D_{\vec{v}} f = \langle \vec{v}, \nabla f \rangle = \|\vec{v}\| \, \|\nabla f\| \cos\theta.$$

---

a) $\max_{x \in S} f(x)$ is the max value of f over all x in the set S while $\operatorname{argmax}_{x \in S} f(x)$ is the x_0 so that $f(x_0) = \max_{x \in S} f(x)$

Taylor's Theorem (version 1).

If $f: \mathbb{R}^n \to \mathbb{R}$ is a C^2 function, then

$$f(\vec{x}_0 + \vec{v}) \approx f(\vec{x}_0) + D_{\vec{v}} f(\vec{x}_0)$$
$$= f(\vec{x}_0) + \langle \vec{v}, \nabla f(\vec{x}_0) \rangle$$

and the right hand side (which is linear in $\vec{v}$) is the best linear approximation to f near $\vec{x}_0$.

Example. If $n = 1$,

$$f(x_0 + v) \approx f(x_0) + f'(x_0)\, v$$

where the rhs is the tangent line to f at x_0.

We are now going to generalize this a little further!

Definition. A map $Q: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called <u>bi</u>linear if $Q(\vec{v}, \vec{w})$ is linear in $\vec{v}$ and linear in $\vec{w}$:

$$Q(\lambda \vec{v} + \mu \vec{u}, \vec{w}) = \lambda Q(\vec{v}, \vec{w}) + \mu Q(\vec{u}, \vec{w})$$

$$Q(\vec{v}, \lambda \vec{u} + \mu \vec{w}) = \lambda Q(\vec{v}, \vec{u}) + \mu Q(\vec{v}, \vec{w})$$

We say Q is <u>symmetric</u> if $Q(\vec{v}, \vec{w}) = Q(\vec{w}, \vec{v})$.

Finally, a symmetric, bilinear $Q: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called a <u>quadratic</u> <u>form</u>.

Just as every linear functional can be written in terms of a vector, every quadratic form can be written in terms of a matrix.

Proposition. Every quadratic form on $\mathbb{R}^n$ can be written as $Q(\vec{v},\vec{w}) = \langle \vec{v}, A\vec{w} \rangle$ for some $n \times n$ symmetric matrix A.

We will use the notations

$$Q_A(\vec{v},\vec{w}) = \langle \vec{v}, A\vec{w} \rangle = \langle \vec{v}, \vec{w} \rangle_A$$

to refer to the quadratic form.

Again, we've already met a quadratic form in calculus class!

Definition. If $f: \mathbb{R}^n \to \mathbb{R}$ and $\vec{v}, \vec{w} \in \mathbb{R}^n$ the mixed partial in directions $\vec{v}, \vec{w}$ is given by $D_{\vec{v}}(D_{\vec{w}} f)$.

Note that if $\vec{v} = \vec{e}_i$ and $\vec{w} = \vec{e}_j$, then

$$D_{\vec{e}_i}(D_{\vec{e}_j} f) = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f = \frac{\partial^2 f}{\partial x_i \partial x_j},$$

which motivates the name. Now you learned

Clairaut's Theorem. If $f: \mathbb{R}^n \to \mathbb{R}$ is C^2 at a point $\vec{x}_0$ then $\dfrac{\partial^2 f}{\partial x_i \partial x_j} = \dfrac{\partial^2 f}{\partial x_j \partial x_i}$ at $\vec{x}_0$.

or "mixed partials commute".

Homework: Show $D_{\vec{v}}(D_{\vec{w}} f)$ is linear in $\vec{v}$ and in $\vec{w}$.

We will prove in homework that $D_{\vec{v}}(D_{\vec{\omega}} f) = D_{\vec{\omega}}(D_{\vec{v}} f)$ by showing that

$$\begin{aligned} D_{\vec{v}}(D_{\vec{\omega}} f) &= \langle \vec{v}, A\vec{\omega} \rangle \\ &= \langle \vec{\omega}, A\vec{v} \rangle = D_{\vec{\omega}}(D_{\vec{v}} f) \end{aligned}$$

where A is the symmetric $n \times n$ matrix of <u>second</u> partials of f: $A_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$.

Definition. The matrix $A_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$ is called the <u>Hessian matrix</u> of f (and usually denoted Hf).

Taylor's Theorem (version 2)

If $f : \mathbb{R}^n \to \mathbb{R}$ is a C^3 function, then

$$f(\vec{x}_0 + \vec{v}) \approx f(x_0) + D_{\vec{v}} f(x_0) + \frac{1}{2}\left(D_{\vec{v}}(D_{\vec{v}} f)\right)(x_0)$$

$$= f(x_0) + \langle \vec{v}, \nabla f(\vec{x}_0) \rangle + \frac{1}{2} \langle \vec{v}, Hf(x_0)\, \vec{v} \rangle$$

the right hand side (which is quadratic in $\vec{v}$) is the best quadratic approximation to f near x_0.

Example. If $n = 1$,

$$f(x_0 + v) \approx f(x_0) + f'(x_0) v + \frac{1}{2} f''(x_0) v^2$$

We now want to understand the meaning of the matrix Hf in the same way we understood the vector ∇f.

Doing so will require more linear algebra!