Math 346 Lecture #6
6.6 Taylor's Theorem

Throughout let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces over the same field $\mathbb{F}$, and U an open set in X.

6.5.1 Higher-Order Derivatives

Higher-order derivatives are defined inductively as explained below after some technical results.

Definition 6.6.1. For $\mathscr{B}^1(X,Y) = \mathscr{B}(X,Y)$, and $k \in \mathbb{N}$ with $k \geq 2$, the Banach spaces $\mathscr{B}^k(X,Y)$ are defined inductively by

$$\mathscr{B}^k(X,Y) = \mathscr{B}(X, \mathscr{B}^{k-1}(X,Y)).$$

Note. That $\mathscr{B}^k(X,Y)$ is a Banach space follows by Theorem 5.7.1, i.e., X is Banach space and $\mathscr{B}(X,Y)$ is a Banach space, so $\mathscr{B}(X,\mathscr{B}(X,Y))$ is a Banach space in the induced norm, etc.

Note. An element of $\mathscr{B}^2(X,Y)$ is a linear transformation $L : X \to \mathscr{B}(X,Y)$ whose induced norm

$$\|L\|_{X,\mathscr{B}(X,Y)} = \sup\left\{ \frac{\|L(\mathrm{h}_1)\|_{X,Y}}{\|\mathrm{h}_1\|_X} : \mathrm{h}_1 \in X \setminus \{0\} \right\}$$

is finite, where for each $\mathrm{h}_1 \in X$ the linear transformation $L(\mathrm{h}_1) : X \to Y$ is bounded, i.e.,

$$\|L(\mathrm{h}_1)\|_{X,Y} = \sup\left\{ \frac{\|L(\mathrm{h}_1)\mathrm{h}_2\|_Y}{\|\mathrm{h}_2\|_X} : \mathrm{h}_2 \in X \setminus \{0\} \right\} < \infty.$$

We combine $\|L(\mathrm{h}_1)\|_{X,Y} \leq \|L\|_{X,\mathscr{B}(X,Y)}\|\mathrm{h}_1\|_X$ and $\|L(\mathrm{h}_1)\mathrm{h}_2\|_Y \leq \|L(\mathrm{h}_1)\|_{X,Y}\|\mathrm{h}_2\|_X$ to get

$$\|L(\mathrm{h}_1)\mathrm{h}_2\|_Y \leq \|L\|_{X,\mathscr{B}(X,Y)}\|\mathrm{h}_1\|_X\|\mathrm{h}_2\|_X,$$

or when both $\|\mathrm{h}_1\|_X$ and $\|\mathrm{h}_2\|_X$ nonzero, that

$$\|L\|_{X,\mathscr{B}(X,Y)} \geq \frac{\|L(\mathrm{h}_1)\mathrm{h}_2\|_Y}{\|\mathrm{h}_1\|_X\|\mathrm{h}_2\|_X}.$$

The upper bound $\|L\|_{X,\mathscr{B}(X,Y)}$ on the ratios is the supremum because for $\epsilon > 0$ there exists a nonzero $\mathrm{h}_1 \in X$ such that

$$\frac{\|L(\mathrm{h}_1)\|_{X,Y}}{\|h_1\|_X} > \|L\|_{X,\mathscr{B}(X,Y)} - \frac{\epsilon}{2}$$

and there exists a nonzero $\mathrm{h}_2 \in X$ such that

$$\frac{\|L(\mathrm{h}_1)\mathrm{h}_2\|_Y}{\|\mathrm{h}_2\|_X} > \|L(\mathrm{h}_1)\|_{X,Y} - \frac{\epsilon\|\mathrm{h}_1\|_X}{2},$$

so that

$$\frac{\|L(\mathrm{h}_1)\mathrm{h}_2\|_Y}{\|\mathrm{h}_1\|_X\|\mathrm{h}_2\|_X} > \frac{\|L(\mathrm{h}_1)\|_{X,Y}}{\|\mathrm{h}_1\|_X} - \frac{\epsilon}{2} > \|L\|_{X,\mathscr{B}(X,Y)} - \epsilon.$$

Thus

$$\|L\|_{X,\mathscr{B}(X,Y)} = \sup\left\{\frac{\|L(\mathrm{h}_1)\mathrm{h}_2\|_Y}{\|\mathrm{h}_1\|_X\|\mathrm{h}_2\|_X} : \mathrm{h}_1, \mathrm{h}_2 \in X\setminus\{0\}\right\}.$$

Note. Another interpretation of the elements L of $\mathscr{B}^2(X,Y)$ is as continuous multilinear transformations $L : X \times X \to Y$.

A transformation $L : X \times X \to Y$ is multilinear if for fixed h_2, the map $\mathrm{h}_1 \to L(\mathrm{h}_1,\mathrm{h}_2)$ from X to Y is linear, and for each fixed h_1, the map $\mathrm{h}_2 \to L(\mathrm{h}_1,\mathrm{h}_2)$ from X to Y is linear.

Previously for $L \in \mathscr{B}^2(X,Y) = \mathscr{B}(X,\mathscr{B}(X,Y))$ we wrote $L(\mathrm{h}_1)\mathrm{h}_2$, but since this L is linear in h_1 and linear in h_2, it is multilinear, and we write $L(\mathrm{h}_1,\mathrm{h}_2)$ instead.

A multilinear transformation $L : X \times X \to Y$ is continuous if its norm

$$\|L\| = \sup\left\{\frac{\|L(\mathrm{h}_1,\mathrm{h}_2)\|_Y}{\|\mathrm{h}_1\|_X\|\mathrm{h}_2\|_X} : \mathrm{h}_1,\mathrm{h}_2 \in X\setminus\{0\}\right\}$$

is finite. [This is precisely the norm on L we got previous.]

Note. For Banach spaces $X_1,\dots,X_n$ and Y the Banach space $\mathscr{B}(X_1,\dots,X_n;Y)$ consists of multilinear transformations $L : X_1\times\cdots\times X_n \to Y$ whose norms

$$\|L\| = \sup\left\{\frac{\|L(\mathrm{h}_1,\dots,\mathrm{h}_n)\|_Y}{\|\mathrm{h}_1\|_{X_1}\cdots\|\mathrm{h}_n\|_{X_n}} : \mathrm{h}_i \in X_i\setminus\{0\}\right\}$$

are finite. The norm $\|L\|$ has the property that

$$\|L(\mathrm{h}_1,\dots,\mathrm{h}_n)\|_Y \le \|L\|\ \|\mathrm{h}_1\|_{X_1}\cdots\|\mathrm{h}_n\|_{X_n} \text{ for all } \mathrm{h}_i \in X_i.$$

If $X_1 = \cdots = X_n$ we write $\mathscr{B}^k(X,Y)$ instead of $\mathscr{B}(X,\dots,X;Y)$.

Definition 6.6.2. Let $f : U \to Y$ be differentiable on U.

We say f is twice differentiable on U if $Df : U \to \mathscr{B}(X,Y)$ is differentiable on U, and write $D^2 f = D(Df)$ for the second derivative.

For each $\mathrm{x} \in U$, the second derivative $D^2 f(\mathrm{x})$, if it exists, belongs to $\mathscr{B}^2(X,Y)$, so that $D^2 f(\mathrm{x})$ is a continuous multilinear transformation that acts on a pair of vectors $(\mathrm{h}_1,\mathrm{h}_2) \in X \times X$ to produce a vector in Y.

Proceeding inductively for $k \ge 2$, if the map $D^{k-1}f : U \to \mathscr{B}^{k-1}(X,Y)$ is differentiable on U, then we say that f is k-times differentiable on U and denote the k^{th} derivative by $D^k f = D(D^{k-1}f)$.

For each $\mathrm{x} \in U$, the k^{th} derivative $D^k f(\mathrm{x})$, if it exists, belongs to $\mathscr{B}^k(X,Y)$, so that $D^k f(\mathrm{x})$ is a continuous multilinear transformation that acts on k vectors $(\mathrm{h}_1,\dots,\mathrm{h}_k) \in X\times\cdots\times X$ to produce a vector in Y.

If the k^{th} derivative $D^k f$ is continuous on U, then we say that f is k-times continuously differentiable on U.

We denote the set of k-times continuously differentiable functions on U by $C^k(U,Y)$. This is a vector space of functions.

A function $f : U \to Y$ is called smooth if $f \in C^k(U, Y)$ for all $k \in \mathbb{N}$.

We denote the vector space of smooth functions from U to Y by $C^\infty(U, Y)$.

Example (slight variation of 6.3.3). For U open in $\mathbb{R}^n$, suppose $f : U \to \mathbb{R}$ is differentiable on U, i.e., $Df(\mathrm{x}) \in \mathscr{B}(\mathbb{R}^n, \mathbb{R})$ exists at each $\mathrm{x} \in U$.

The Banach space $\mathscr{B}(\mathbb{R}^n, \mathbb{R})$ is the dual space of $\mathbb{R}^n$, which by the Riesz Representation Theorem is isomorphic to $\mathbb{R}^n$, i.e., for each $L \in \mathscr{B}(\mathbb{R}^n, \mathbb{R})$ there exists a unique vector $\mathrm{u} \in \mathbb{R}^n$ such that $L(\mathrm{v}) = \langle \mathrm{u}, \mathrm{v} \rangle = \mathrm{u}^{\mathrm{T}}\mathrm{v}$.

By writing the vector u as the row vector u^{T}, we represent $Df(\mathrm{x})$ as a row vector, which by Theorem 6.2.11, in the standard basis of $\mathbb{R}^n$, is

$$Df(\mathrm{x}) = \begin{bmatrix} D_1 f(\mathrm{x}) & \cdots & D_n f(\mathrm{x}) \end{bmatrix}$$

where $D_i f(\mathrm{x}) = Df(\mathrm{x})\mathrm{e}_i \in \mathbb{R}$, $i = 1, \ldots, n$, are the partial derivatives.

Now suppose that f is twice differentiable on U, i.e., $D^2 f(\mathrm{x}) \in \mathscr{B}^2(\mathbb{R}^n, \mathbb{R})$ exists at each $\mathrm{x} \in U$.

Since $\mathscr{B}^2(\mathbb{R}^n, \mathbb{R}) = \mathscr{B}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n, \mathbb{R})) = \mathscr{B}(\mathbb{R}^n, (\mathbb{R}^n)^*)$, the directional derivative of $Df(\mathrm{x})$ in the direction $\mathrm{u} \in \mathbb{R}^n$, i.e., $D^2 f(\mathrm{x})(\mathrm{u}) \in (\mathbb{R}^n)^*$, is a row vector.

We can still apply Theorem 6.2.11, but in transposed form, i.e., $D^2 f(\mathrm{x})(\mathrm{u}) = \mathrm{u}^{\mathrm{T}} H^{\mathrm{T}}$ where H is the "Hessian" of f at x,

$$\begin{aligned} H &= D \begin{bmatrix} D_1 f(\mathrm{x}) & \cdots & D_n f(\mathrm{x}) \end{bmatrix}^{\mathrm{T}} \\ &= \begin{bmatrix} D_1 D_1 f(\mathrm{x}) & \cdots & D_n D_1 f(\mathrm{x}) \\ \vdots & \ddots & \vdots \\ D_1 D_n f(\mathrm{x}) & \cdots & D_n D_n f(\mathrm{x}) \end{bmatrix}, \end{aligned}$$

so that

$$D^2 f(\mathrm{x})(\mathrm{u})(\mathrm{v}) = \mathrm{u}^{\mathrm{T}} H^{\mathrm{T}} \mathrm{v} \in \mathbb{R}$$

for all $\mathrm{u}, \mathrm{v} \in \mathbb{R}^n$.

From $D^2 f(\mathrm{x})(\mathrm{u})(\mathrm{v}) = \mathrm{u}^{\mathrm{T}} H^{\mathrm{T}} \mathrm{v} = \mathrm{v}^{\mathrm{T}} H \mathrm{u}$ (the transpose of a scalar is itself) we see that $D^2 f(\mathrm{x})$ does indeed act multilinearly on a pair of vectors $(\mathrm{u}, \mathrm{v}) \in \mathbb{R}^n \times \mathbb{R}^n$ to produce a scalar, i.e., $D^2 f(\mathrm{x})(\mathrm{u}, \mathrm{v}) = \mathrm{u}^{\mathrm{T}} H^{\mathrm{T}} \mathrm{v} \in \mathbb{R}$.

Definition 6.6.4. Let $(X_i, \|\cdot\|_{X_i})$, $i = 1, 2 \ldots, n$, be a finite collection of Banach spaces.

Fix an open set $U \subset X_1 \times X_2 \times \cdots \times X_n$, and an ordered list of k integers $i_1, i_2, \cdots, i_k$ where $i_j \in \{1, \ldots, k\}$ (not necessarily distinct).

The k^{th}-order partial derivative of $f \in C^k(U, Y)$ corresponding to $i_1, \ldots, i_k$ is the function $D_{i_1} D_{i_2} \cdots D_{i_k} f \in C(U, \mathscr{B}(X_1, X_2, \ldots, X_k; Y))$.

[Recall from Definition 6.3.12, that the i^{th} partial derivative of a function $\alpha : X_1 \times \cdots \times X_n \to Y$ is the derivative of the function $\beta : X_i \to Y$ defined by $\beta(z) = \alpha(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n)$, i.e., the function obtained from α by fixing all of its inputs except the i^{th} variable.]

When $X_i = \mathbb{F}$ for all $i = 1, 2, \ldots, n$ and $Y = \mathbb{F}$, we often write $D_{i_1} D_{i_2} \cdots D_{i_k} f$ as the more familiar partial derivative

$$\frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}}.$$

The Hessian of $f : U \to \mathbb{R}$, U open in $\mathbb{R}^n$, is nothing more than the matrix of second-order partial derivatives:

$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}.$$

You might remember that this square matrix of second-order partial derivatives is usually symmetric. This is true when the second derivative is continuous.

Proposition 6.6.5. If $f \in C^2(U, Y)$ with Y finite dimensional, then for all $\mathrm{x} \in U$ and for all $(\mathrm{u}, \mathrm{v}) \in X \times X$, there holds

$$D^2 f(\mathrm{x})(\mathrm{u}, \mathrm{v}) = D^2 f(\mathrm{x})(\mathrm{v}, \mathrm{u}).$$

When U is an open subset of $X = X_1 \times X_2 \times \cdots \times X_n$ for Banach spaces X_i, and $f \in C^2(U, Y)$ for finite dimensional Y, then for all $\mathrm{x} \in U$ and for all $i, j \in \{1, 2, \ldots, n\}$, there holds

$$D_i D_j f(\mathrm{x}) = D_j D_i f(\mathrm{x}).$$

When U is a open subset of $X = \mathbb{F}^n = \mathbb{F} \times \mathbb{F} \times \cdots \times \mathbb{F}$, $Y = \mathbb{F}^m$, and $f = (f_1, f_2, \ldots, f_m) \in C^2(U, Y)$, then for all $\mathrm{x} \in U$ and for all $i, j \in \{1, 2, \ldots, n\}$ and all $k \in \{1, 2, \ldots, m\}$ there holds

$$\frac{\partial^2 f_k}{\partial x_i \partial x_j} = \frac{\partial^2 f_k}{\partial x_j \partial x_i}.$$

Proof. The hypothesis of finite dimensionality of Y implies that we can assume WLOG that $Y = \mathbb{F}^m$ and that $f = (f_1, \ldots, f_m)$.

With $\mathbb{C}^m$ and $\mathbb{R}^{2m}$ isomorphic as Banach spaces (with the standard norms) we can assume WLOG that $\mathbb{F} = \mathbb{R}$.

Then as $f_k : U \to \mathbb{R}$ it suffices to show the result for $Y = \mathbb{R}$.

For a fixed $\mathrm{x} \in U$ and $\mathrm{u}, \mathrm{v} \in X$ there exist $t, s > 0$ by the openness of U such that $\mathrm{x} + \xi \mathrm{u} + \eta \mathrm{v} \in U$ for all $\xi, \eta \in [0, \max\{s, t\}]$.

Define $g : [0, t] \to \mathbb{R}$ by

$$g_\xi(\mathrm{x}) = f(\mathrm{x} + \xi \mathrm{u}) - f(\mathrm{x})$$

and $S_{\eta,t}(\mathrm{x}) : [0, s] \to \mathbb{R}$ by

$$\begin{aligned} S_{\eta,t}(\mathrm{x}) &= g_t(\mathrm{x} + \eta v) - g_t(\mathrm{x}) \\ &= f(\mathrm{x} + t\mathrm{u} + \eta \mathrm{v}) - f(\mathrm{x} + \eta \mathrm{v}) - f(\mathrm{x} + t\mathrm{u}) + f(\mathrm{x}). \end{aligned}$$

Recognize that $S_{0,t}(\mathrm{x}) = 0$ and that with x and t fixed,

$$DS_{\eta,t}(\mathrm{x}) = Dg_t(\mathrm{x} + \eta\mathrm{v})\mathrm{v}.$$

The function $S_{\eta,t}$ is continuous on $[0, s]$ and differentiable on $(0, s)$, so by the Mean Value Theorem there exists $\sigma_{s,t} \in (0, s)$ such that

$$S_{s,t}(\mathrm{x}) = S_{s,t}(\mathrm{x}) - S_{0,t}(\mathrm{x}) = Dg_t(\mathrm{x} + \sigma_{s,t}\mathrm{v})(\mathrm{v})(s - 0) = Dg_t(\mathrm{x} + \sigma_{s,t}\mathrm{v})(s\mathrm{v}).$$

Since $g_t(\mathrm{x} + \eta\mathrm{v}) = f(\mathrm{x} + t\mathrm{u} + \eta\mathrm{v}) - f(\mathrm{x} + \eta\mathrm{v})$ we have

$$Dg_t(\mathrm{x} + \sigma_{s,t}\mathrm{v})(s\mathrm{v}) = Df(\mathrm{x} + t\mathrm{u} + \sigma_{s,t}\mathrm{v})(s\mathrm{v}) - Df(\mathrm{x} + \sigma_{s,t}\mathrm{v})(s\mathrm{v}).$$

The function

$$\xi \to Df(\mathrm{x} + \xi\mathrm{u} + \sigma_{s,t}\mathrm{v})(s\mathrm{v}) - Df(\mathrm{x} + \sigma_{s,t}\mathrm{v})(s\mathrm{v})$$

is zero when $\xi = 0$, is continuous on $[0, t]$ and differentiable on $(0, t)$, so by the Mean Value Theorem there exists $\tau_{s,t} \in (0, t)$ such that

$$Df(\mathrm{x} + t\mathrm{u} + \sigma_{s,t}\mathrm{v})(s\mathrm{v}) - Df(\mathrm{x} + \sigma_{s,t}\mathrm{v})(s\mathrm{v}) = D^2f(\mathrm{x} + \tau_{s,t}\mathrm{u} + \sigma_{s,t}\mathrm{v})(s\mathrm{v})(t\mathrm{u}).$$

Thus

$$S_{s,t}(\mathrm{x}) = D^2f(\mathrm{x} + \tau_{s,t}\mathrm{u} + \sigma_{s,t}\mathrm{v})(s\mathrm{v}, t\mathrm{u}).$$

Switching the roles of $t\mathrm{u}$ and $s\mathrm{v}$ in the above argument gives the existence of $\tau'_{s,t}$ and $\sigma'_{s,t}$ such that

$$S_{s,t}(\mathrm{x}) = D^2f(\mathrm{x} + \sigma'_{s,t}\mathrm{v} + \tau'_{s,t}\mathrm{u})(t\mathrm{u}, s\mathrm{v}).$$

[Needed that $\mathrm{x} + \xi\mathrm{u} + \eta\mathrm{v} \in U$ for all $\xi, \eta \in [0, \max\{s, t\}]$ here.] Equating the two expressions for the same quantity $S_{s,t}(\mathrm{x})$ gives

$$D^2f(\mathrm{x} + \tau_{s,t}\mathrm{u} + \sigma_{s,t}\mathrm{v})(s\mathrm{v}, t\mathrm{u}) = D^2f(\mathrm{x} + \sigma'_{s,t}\mathrm{v} + \tau'_{s,t}\mathrm{u})(t\mathrm{u}, s\mathrm{v}).$$

Since $D^2f(\mathrm{x})$ is multilinear, we can pull out the scalars s and t from the inputs $s\mathrm{v}$ and $t\mathrm{u}$ from both sides; they cancel, giving

$$D^2f(\mathrm{x} + \tau_{s,t}\mathrm{u} + \sigma_{s,t}\mathrm{v})(\mathrm{v}, \mathrm{u}) = D^2f(\mathrm{x} + \sigma'_{s,t}\mathrm{v} + \tau'_{s,t}\mathrm{u})(\mathrm{u}, \mathrm{v}).$$

As the scalars $s, t \to 0$ the quantities $\sigma_{s,t}, \sigma'_{s,t}, \tau_{s,t}, \tau'_{s,t}$ all go to zero.

The assumed continuity of D^2f on U implies as $s, t \to 0$ that $D^2f(\mathrm{x})(\mathrm{v}, \mathrm{u}) = D^2f(\mathrm{x})(\mathrm{u}, \mathrm{v})$.

In the case that $X = X_1 \times \cdots \times X_n$, we take $\mathrm{u}_i \in X_i$ and $\mathrm{v}_j \in X_j$ and form the vectors

$$\mathrm{u} = (0, \ldots, \mathrm{u}_i, \ldots, 0),\ \ \mathrm{v} = (0, \ldots, \mathrm{v}_j, \ldots, 0)$$

where u_i is in the i^{th} slot and v_j is in the j^{th} slot, to get

$$D^2f(\mathrm{x})(\mathrm{u}, \mathrm{v}) = D^2f(\mathrm{x})(\mathrm{v}, \mathrm{u})$$

which implies that $D_iD_jf(\mathrm{x}) = D_jD_if(\mathrm{x})$.

Finally, the $D_iD_jf(\mathrm{x}) = D_jD_if(\mathrm{x})$ implies the equality

$$\frac{\partial^2 f_k}{\partial x_i \partial x_j} = \frac{\partial^2 f_k}{\partial x_j \partial x_i}$$

for all $k = 1, \ldots, m$. $\square$

Remark 6.6.6. For $U \subset \mathbb{R}^n$ and $f \in C^2(U, \mathbb{R})$, Proposition 6.6.5 guarantees that the Hessian of f is a symmetric matrix.

6.6.2 Higher-Order Directional Derivatives

For U open in $\mathbb{F}^n$ and $f \in C^k(U, \mathbb{F}^m)$, the derivative $D^k f(\mathrm{x})$ is an element of $\mathscr{B}^k(\mathbb{F}^n, \mathbb{F}^m)$ for each $\mathrm{x} \in U$.

This means that $D^k f(\mathrm{x})$ acts on k vectors from $\mathbb{F}^n$ giving a vector in $\mathbb{F}^m$.

When all of the inputs of $D^k f(\mathrm{x})$ are the same vector v, we call the output

$$D^k_{\mathrm{v}} f(\mathrm{x}) = D^k f(\mathrm{x})(\mathrm{v}, \ldots, \mathrm{v})$$

the k^{th} directional derivative of f at x in the direction v.

We express the directional derivatives of f in terms of standard coordinates for $\mathbb{F}^n$ and $\mathbb{F}^m$.

For $k = 1$ and

$$\mathrm{v} = \sum_{i=1}^{n} v_i \mathrm{e}_i \in \mathbb{F}^n,$$

the vector

$$Df(\mathrm{x})\mathrm{v} = \begin{bmatrix} D_1 f(\mathrm{x}) & \cdots & D_n f(\mathrm{x}) \end{bmatrix} \mathrm{v} = \sum_{j=1}^{n} D_j f(\mathrm{x}) v_j \in \mathbb{F}^m$$

is the first-order directional derivative $D_{\mathrm{v}} f(\mathrm{x})$ of f at x in the direction v.

The second-order directional derivative of f at x in the direction v is

$$D^2_{\mathrm{v}} f(\mathrm{x}) = D_{\mathrm{v}} \sum_{j=1}^{n} D_j f(\mathrm{x}) v_j = \sum_{i=1}^{n} \sum_{j=1}^{n} D_i D_j f(\mathrm{x}) v_i v_j = \mathrm{v}^{\mathrm{T}} H(\mathrm{x}) \mathrm{v},$$

where $H(\mathrm{x})$ is the Hessian of f at x.

Iterating k times gives the k^{th}-order directional derivative of f at x in the direction v:

$$D^k_{\mathrm{v}} f(\mathrm{x}) = \sum_{i_1, \ldots, i_k = 1}^{n} D_{i_1} \cdots D_{i_k} f(\mathrm{x}) v_{i_1} \cdots v_{i_k}.$$

Often we write $D^k f(\mathrm{x}) \mathrm{v}^{(k)}$ for $D^k_{\mathrm{v}} f(\mathrm{x})$ where by $\mathrm{v}^{(k)}$ we mean the k-component Cartesian product $(\mathrm{v}, \ldots, \mathrm{v})$.

Proposition 6.6.5 and its application to higher-order derivatives shows that many of the terms in the k^{th}-order directional derivative are repeated.

Combining these repeated terms gives

$$D_{\mathrm{v}}^k f(\mathrm{x}) = \sum_{j_1+\cdots+j_n=k} \frac{k!}{j_1!\cdots j_n!} D_1^{j_1}\cdots D_n^{j_n} f(\mathrm{x}) v_1^{j_1}\cdots v_n^{j_n}$$

where the quantities $j_1,\dots,j_n$ are nonnegative integers summing to the integer k.

6.6.3 Taylor's Theorem

We first recall Taylor's Theorem and Lagrange's Remainder Theorem for functions $f : U \to \mathbb{R}$ for an open interval $U \subset \mathbb{R}$.

Theorem 6.6.8. For an open interval U of $\mathbb{R}$, if $f : U \to \mathbb{R}$ is $(k+1)$-times differentiable on U, then for all $a \in U$ and $h \in \mathbb{R}$ satisfying $a+h \in U$ there exists c between a and $a+h$ (h could be negative) such that

$$f(a+h) = f(a) + f'(a)h + \frac{f''(a)}{2}h^2 + \cdots + \frac{f^{(k)}(a)}{k!}h^k + \frac{f^{(k+1)}(c)}{(k+1)!}h^{k+1},$$

where $f^{(k)}$ is the k^{th} derivative of f.

We extend this to the Banach space setting where we make use of the higher-order directional derivatives $D^k f(\mathrm{x})\mathrm{v}^{(k)}$.

Theorem 6.6.9. If $f \in C^k(U,Y)$, then for all $\mathrm{x} \in U$ and $\mathrm{h} \in X$ such that the line segment $\ell(\mathrm{x},\mathrm{x}+\mathrm{h}) = \{(1-t)\mathrm{x} + t\mathrm{h} : t \in [0,1]\}$ lies in U, there holds

$$f(\mathrm{x}+\mathrm{h}) = f(\mathrm{x}) + Df(\mathrm{x})\mathrm{h} + \frac{D^2 f(\mathrm{x})\mathrm{h}^{(2)}}{2!} + \cdots + \frac{D^{k-1} f(\mathrm{x})\mathrm{h}^{(k-1)}}{(k-1)!} + R_k$$

where the remainder R_k is given by the integral form

$$R_k(\mathrm{x},\mathrm{h}) = \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} D^k f(\mathrm{x}+t\mathrm{h})\mathrm{h}^{(k)}\, dt.$$

The proof is by induction (see the text).

Remark 6.6.10. The $(k-1)^{\text{th}}$ Taylor polynomial approximation of $f \in C^k(U,Y)$ is given in Theorem 6.6.9 by ignoring the remainder term R_k.

If $f \in C^\infty(U,Y)$ and the remainder R_k goes to zero as $k \to \infty$, then the Taylor series for f converges to f, and we say that f is analytic.

Corollary 6.6.14. If $\|D^k f(\mathrm{x}+t\mathrm{h})\| < M$ for all $t \in [0,1]$, then

$$\|R_k\|_Y \le \frac{M}{k!}\|\mathrm{h}\|_X^k.$$

The proof of this is HW (Exercise 6.33. Hint: use a property of the norm for bounded multilinear transformations).

Example. Find the first-order Taylor polynomial approximation for

$$f(x,y) = \sin(x+y)$$

at the origin.

We compute several quantities:

$$\begin{aligned} f(0,0) &= \sin(0+0) = 0, \\ D_1 f(0,0) &= \cos(0+0) = 1, \\ D_2 f(0,0) &= \cos(0+0) = 1. \end{aligned}$$

Thus

$$\begin{aligned} f(h_1,h_2) &= f(0,0) + \begin{bmatrix} D_1 f(0,0) & D_2 f(0,0) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \\ &= h_1 + h_2. \end{aligned}$$

This is the tangent plane of the graph of $z = f(x,y) = \sin(x+y)$ at the point $(0,0)$.