Multivariable Calculus Review

Multi-Variable Calculus

Norms:

A function $\nu : \mathbb{R}^n \to \mathbb{R}$ is a vector norm on $\mathbb{R}^n$ if

Multi-Variable Calculus

Norms:

A function $\nu : \mathbb{R}^n \to \mathbb{R}$ is a vector norm on $\mathbb{R}^n$ if

- $\nu(x) \geq 0 \ \forall \ x \in \mathbb{R}^n$ with equality iff $x = 0$.

Multi-Variable Calculus

Norms:

A function $\nu : \mathbb{R}^n \to \mathbb{R}$ is a vector norm on $\mathbb{R}^n$ if

- $\nu(x) \geq 0 \ \forall \ x \in \mathbb{R}^n$ with equality iff $x = 0$.
- $\nu(\alpha x) = |\alpha|\nu(x) \ \forall \ x \in \mathbb{R}^n \ \alpha \in \mathbb{R}$
- $\nu(x + y) \leq \nu(x) + \nu(y) \ \forall \ x, y \in \mathbb{R}^n$

Multi-Variable Calculus

Norms:

A function $\nu : \mathbb{R}^n \to \mathbb{R}$ is a vector norm on $\mathbb{R}^n$ if

- $\nu(x) \geq 0 \ \forall \ x \in \mathbb{R}^n$ with equality iff $x = 0$.
- $\nu(\alpha x) = |\alpha| \nu(x) \ \forall \ x \in \mathbb{R}^n \ \alpha \in \mathbb{R}$
- $\nu(x + y) \leq \nu(x) + \nu(y) \ \forall \ x, y \in \mathbb{R}^n$

We usually denote $\nu(x)$ by $\|x\|$. Norms are convex functions.

Multi-Variable Calculus

Norms:

A function $\nu : \mathbb{R}^n \to \mathbb{R}$ is a vector norm on $\mathbb{R}^n$ if

- $\nu(x) \geq 0 \ \forall \ x \in \mathbb{R}^n$ with equality iff $x = 0$.
- $\nu(\alpha x) = |\alpha| \nu(x) \ \forall \ x \in \mathbb{R}^n \ \alpha \in \mathbb{R}$
- $\nu(x + y) \leq \nu(x) + \nu(y) \ \forall \ x, y \in \mathbb{R}^n$

We usually denote $\nu(x)$ by $\|x\|$. Norms are convex functions.

l_p norms

$$\begin{array}{rcl} \|x\|_p & := & \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}, \quad 1 \leq p < \infty \\ \|x\|_\infty & = & \max_{i=1,\ldots,n} |x_i| \end{array}$$

Elementary Topological Properties of Sets

Let $D \subset \mathbb{R}^n$.

Elementary Topological Properties of Sets

Let $D \subset \mathbb{R}^n$.

- D is *open* if for every $x \in D$ there exists $\epsilon > 0$ such that $x + \epsilon\mathbb{B} \subset D$ where

$$x + \epsilon\mathbb{B} = \{x + \epsilon u : u \in \mathbb{B}\}$$

  and $\mathbb{B}$ is the unit ball of some given norm on $\mathbb{R}^n$.

Elementary Topological Properties of Sets

Let $D \subset \mathbb{R}^n$.

- D is *open* if for every $x \in D$ there exists $\epsilon > 0$ such that $x + \epsilon\mathbb{B} \subset D$ where

$$x + \epsilon\mathbb{B} = \{x + \epsilon u : u \in \mathbb{B}\}$$

  and $\mathbb{B}$ is the unit ball of some given norm on $\mathbb{R}^n$.

- D is *closed* if every point x satisfying

$$(x + \epsilon\mathbb{B}) \cap D \neq \emptyset$$

  for all $\epsilon > 0$, must be a point in D.

Elementary Topological Properties of Sets

Let $D \subset \mathbb{R}^n$.

- D is *open* if for every $x \in D$ there exists $\epsilon > 0$ such that $x + \epsilon\mathbb{B} \subset D$ where
$$x + \epsilon\mathbb{B} = \{x + \epsilon u : u \in \mathbb{B}\}$$
and $\mathbb{B}$ is the unit ball of some given norm on $\mathbb{R}^n$.

- D is *closed* if every point x satisfying
$$(x + \epsilon\mathbb{B}) \cap D \neq \emptyset$$
for all $\epsilon > 0$, must be a point in D.

- D is *bounded* if there exists $\beta > 0$ such that
$$\|x\| \leq \beta \text{ for all } x \in D.$$

Elementary Topological Properties of Sets: Compactness

Let $D \subset \mathbb{R}^n$.

Elementary Topological Properties of Sets: Compactness

Let $D \subset \mathbb{R}^n$.

- $\overline{x} \in D$ is said to be a *cluster point* of D if

$$(\overline{x} + \epsilon \mathbb{B}) \cap D \neq \emptyset$$

  for every $\epsilon > 0$.

Elementary Topological Properties of Sets: Compactness

Let $D \subset \mathbb{R}^n$.

- $\overline{x} \in D$ is said to be a *cluster point* of D if
$$(\overline{x} + \epsilon\mathbb{B}) \cap D \neq \emptyset$$
for every $\epsilon > 0$.

- D is *compact* if it is closed and bounded.

Elementary Topological Properties of Sets: Compactness

Let $D \subset \mathbb{R}^n$.

- $\overline{x} \in D$ is said to be a *cluster point* of D if

$$(\overline{x} + \epsilon\mathbb{B}) \cap D \neq \emptyset$$

  for every $\epsilon > 0$.

- D is *compact* if it is closed and bounded.

Theorem: [Weierstrass Compactness Theorem]
A set $D \subset \mathbb{R}^n$ is compact if and only if every infinite subset of D has a cluster point and all such cluster points lie in D.

Continuity and The Weierstrass Extreme Value Theorem

The mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at the point $\overline{x}$ if

$$\lim_{\|x-\overline{x}\|\to 0} \|F(x) - F(\overline{x})\| = 0.$$

Continuity and The Weierstrass Extreme Value Theorem

The mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at the point $\overline{x}$ if

$$\lim_{\|x-\overline{x}\|\to 0} \|F(x) - F(\overline{x})\| = 0.$$

F is continuous on a set $D \subset \mathbb{R}^n$ if F is continuous at every point of D.

Continuity and The Weierstrass Extreme Value Theorem

The mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at the point $\overline{x}$ if

$$\lim_{\|x-\overline{x}\|\to 0} \|F(x) - F(\overline{x})\| = 0.$$

F is continuous on a set $D \subset \mathbb{R}^n$ if F is continuous at every point of D.

Theorem: [Weierstrass Extreme Value Theorem]
Every continuous function on a compact set attains its extreme values on that set.

Operator Norms

$$A \in \mathbb{R}^{m \times n} \qquad \|A\|_{(p,q)} = \max\{\|Ax\|_p : \|x\|_q \leq 1\}$$

Operator Norms

$A \in \mathbb{R}^{m \times n}$ $\qquad \|A\|_{(p,q)} = \max\{\|Ax\|_p : \|x\|_q \leq 1\}$

Examples:

Operator Norms

$A \in \mathbb{R}^{m\times n}$ $\qquad \|A\|_{(p,q)} = \max\{\|Ax\|_p : \|x\|_q \leq 1\}$

Examples: $\|A\|_2 \quad = \quad \|A\|_{(2,2)} = \max\{\|Ax\|_2 : \|x\|_2 \leq 1\}$

Operator Norms

$A \in \mathbb{R}^{m\times n}$ $\qquad \|A\|_{(p,q)} = \max\{\|Ax\|_p : \|x\|_q \le 1\}$

Examples:

$$\begin{aligned}\|A\|_2 &= \|A\|_{(2,2)} = \max\{\|Ax\|_2 : \|x\|_2 \le 1\} \\ \|A\|_\infty &= \|A\|_{(\infty,\infty)} = \max\{\|Ax\|_\infty : \|x\|_\infty \le 1\} \\ &= \max_{1\le i\le m} \sum_{j=1}^n |a_{ij}| \qquad \text{(max row sum)}\end{aligned}$$

Operator Norms

$$A \in \mathbb{R}^{m\times n} \qquad \|A\|_{(p,q)} = \max\{\|Ax\|_p : \|x\|_q \le 1\}$$

Examples:

$$\|A\|_2 = \|A\|_{(2,2)} = \max\{\|Ax\|_2 : \|x\|_2 \le 1\}$$

$$\begin{aligned} \|A\|_\infty &= \|A\|_{(\infty,\infty)} = \max\{\|Ax\|_\infty : \|x\|_\infty \le 1\} \\ &= \max_{1\le i\le m} \textstyle\sum_{j=1}^n |a_{ij}| \qquad \text{(max row sum)} \end{aligned}$$

$$\begin{aligned} \|A\|_1 &= \|A\|_{(1,1)} = \max\{\|Ax\|_1 : \|x\|_1 \le 1\} \\ &= \max_{1\le j\le n} \textstyle\sum_{i=1}^m |a_{ij}| \quad \text{(max column sum)} \end{aligned}$$

Operator Norms

Properties:

Operator Norms

Properties:

- $\|Ax\|_p \leq \|A\|_{(p,q)} \|x\|_q$.

Operator Norms

Properties:

- $\|Ax\|_p \leq \|A\|_{(p,q)}\|x\|_q$.
- $\|A\| \geq 0$ with equality $\Leftrightarrow$ $\|Ax\| = 0\ \forall\ x$ or $A \equiv 0$.

Operator Norms

Properties:

- $\|Ax\|_p \leq \|A\|_{(p,q)} \|x\|_q$.
- $\|A\| \geq 0$ with equality $\Leftrightarrow$ $\|Ax\| = 0 \; \forall \; x$ or $A \equiv 0$.
- $$\begin{aligned} \|\alpha A\| &= \max\{\|\alpha A x\| : \|x\| \leq 1\} \\ &= \max\{|\alpha| \, \|Ax\| : \|\alpha\| \leq 1\} = |\alpha| \, \|A\| \end{aligned}$$

Operator Norms

Properties:

- $\|Ax\|_p \leq \|A\|_{(p,q)} \|x\|_q$.
- $\|A\| \geq 0$ with equality $\Leftrightarrow$ $\|Ax\| = 0 \ \forall \ x$ or $A \equiv 0$.
- $$\begin{aligned} \|\alpha A\| &= \max\{\|\alpha Ax\| : \|x\| \leq 1\} \\ &= \max\{|\alpha| \, \|Ax\| : \|\alpha\| \leq 1\} = |\alpha| \, \|A\| \end{aligned}$$
- $$\begin{aligned} \|A + B\| &= \max\{\|Ax + Bx\| : \|x\| \leq 1\} \leq \max\{\|Ax\| + \|Bx\| \ A \leq 1\} \\ &= \max\{\|Ax_1\| + \|Bx_2\| : x_1 = x_2, \|x_1\| \leq 1, \|x_2\| \leq 1\} \\ &\leq \max\{\|Ax_1\| + \|Bx_2\| : \|x_1\| \leq 1, \|x_2\| \leq 1\} \\ &= \|A\| + \|B\| \end{aligned}$$

Frobenius Norm and Inner Product

$$A \in \mathbb{R}^{m\times n} \qquad \|A\|_F = \left[\mathrm{trace}(A^TA)\right]^{1/2}$$

Frobenius Norm and Inner Product

$$A \in \mathbb{R}^{m\times n} \qquad \|A\|_F \;=\; \left[\mathrm{trace}(A^T A)\right]^{1/2}$$

Identify $\mathbb{R}^{m\times n}$ with $\mathbb{R}^{(mn)}$ by simply stacking the columns of a matrix one on top of the other to create a very long vector in $\mathbb{R}^{(mn)}$.

Frobenius Norm and Inner Product

$$A \in \mathbb{R}^{m \times n} \qquad \|A\|_F = \left[\mathrm{trace}(A^T A)\right]^{1/2}$$

Identify $\mathbb{R}^{m \times n}$ with $\mathbb{R}^{(mn)}$ by simply stacking the columns of a matrix one on top of the other to create a very long vector in $\mathbb{R}^{(mn)}$.

This mapping takes a matrix in $\mathbb{R}^{m \times n}$ to a vector in $\mathbb{R}^{(mn)}$ by stacking columns. It is called vec (or sometimes cvec).

Frobenius Norm and Inner Product

$A \in \mathbb{R}^{m\times n} \qquad \|A\|_F \;=\; \left[\text{trace}(A^TA)\right]^{1/2}$

Identify $\mathbb{R}^{m\times n}$ with $\mathbb{R}^{(mn)}$ by simply stacking the columns of a matrix one on top of the other to create a very long vector in $\mathbb{R}^{(mn)}$.

This mapping takes a matrix in $\mathbb{R}^{m\times n}$ to a vector in $\mathbb{R}^{(mn)}$ by stacking columns. It is called vec (or sometimes cvec).

Using vec we can define an inner product on $\mathbb{R}^{m\times n}$ (called the Frobenius inner prodiuct) by writting

$$\langle A, B\rangle_F = \mathrm{vec}\,(A)^T\mathrm{vec}\,(B) = \text{trace}(A^TB)\ .$$

Frobenius Norm and Inner Product

$A \in \mathbb{R}^{m\times n} \qquad \|A\|_F \;=\; \left[\mathrm{trace}(A^TA)\right]^{1/2}$

Identify $\mathbb{R}^{m\times n}$ with $\mathbb{R}^{(mn)}$ by simply stacking the columns of a matrix one on top of the other to create a very long vector in $\mathbb{R}^{(mn)}$.

This mapping takes a matrix in $\mathbb{R}^{m\times n}$ to a vector in $\mathbb{R}^{(mn)}$ by stacking columns. It is called vec (or sometimes cvec).

Using vec we can define an inner product on $\mathbb{R}^{m\times n}$ (called the Frobenius inner prodiuct) by writting

$$\langle A, B\rangle_F = \mathrm{vec}\,(A)^T\mathrm{vec}\,(B) = \mathrm{trace}(A^TB)\ .$$

Note $\|A\|_F^2 = \langle A, A\rangle_F$.

Differentiation

Let $F : \mathbb{R}^n \to \mathbb{R}^m$, and let F_i denote the ith component functioon of F:

$$F(x) = \begin{bmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_m(x) \end{bmatrix} ,$$

where each F_i is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Differentiation

Let $F : \mathbb{R}^n \to \mathbb{R}^m$, and let F_i denote the ith component functioon of F:

$$F(x) = \begin{bmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_m(x) \end{bmatrix},$$

where each F_i is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$.
If the limit

$$\lim_{t \downarrow 0} \frac{F(x+td) - F(x)}{t} =: F'(x; d)$$

exists, it is called the directional derivative of F at x in the direction h.

Differentiation

Let $F : \mathbb{R}^n \to \mathbb{R}^m$, and let F_i denote the ith component functioon of F:

$$F(x) = \begin{bmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_m(x) \end{bmatrix},$$

where each F_i is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$.

If the limit

$$\lim_{t \downarrow 0} \frac{F(x+td) - F(x)}{t} =: F'(x; d)$$

exists, it is called the directional derivative of F at x in the direction h.

If this limit exists for all $d \in \mathbb{R}^n$ and is linear in the d argument,

$$F'(x; \alpha d_1 + \beta d_2) = \alpha F'(x; d_1) + \beta F'(x; d_2),$$

then F is said to be differentiable at x, and denote the associated linear operator by $F'(x)$.

Differentiation

One can show that if $F'(x)$ exists, then

$$\lim_{\|y-x\|\to 0} \frac{\|F(y) - (F(x) + F'(x)(y-x))\|}{\|y-x\|} = 0,$$

Differentiation

One can show that if $F'(x)$ exists, then

$$\lim_{\|y-x\|\to 0} \frac{\|F(y)-(F(x)+F'(x)(y-x))\|}{\|y-x\|} = 0,$$

or equivalently,

$$F(y) = F(x) + F'(x)(y-x) + o(\|y-x\|),$$

Differentiation

One can show that if $F'(x)$ exists, then

$$\lim_{\|y-x\|\to 0} \frac{\|F(y) - (F(x) + F'(x)(y-x))\|}{\|y-x\|} = 0,$$

or equivalently,

$$F(y) = F(x) + F'(x)(y-x) + o(\|y-x\|),$$

where $o(t)$ is a function satisfying

$$\lim_{t\to 0} \frac{o(t)}{t} = 0 \ .$$

Differentiation Facts

Differentiation Facts

(i) If $F'(x)$ exists, it is unique.

Differentiation Facts

(i) If $F'(x)$ exists, it is unique.

(ii) If $F'(x)$ exists, then $F'(x;d)$ exists for all d and $F'(x;d) = F'(x)d$.

Differentiation Facts

(i) If $F'(x)$ exists, it is unique.

(ii) If $F'(x)$ exists, then $F'(x; d)$ exists for all d and $F'(x; d) = F'(x)d$.

(iii) If $F'(x)$ exists, then F is continuous at x.

Differentiation Facts

(i) If $F'(x)$ exists, it is unique.

(ii) If $F'(x)$ exists, then $F'(x;d)$ exists for all d and $F'(x;d) = F'(x)d$.

(iii) If $F'(x)$ exists, then F is continuous at x.

(iv) (Matrix Representation) Suppose $F'(x)$ is continuous at $\bar{x}$, Then

$$F'(\bar{x}) = \nabla F(\bar{x}) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \dots & \frac{\partial F_1}{\partial x_n} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \dots & \frac{\partial F_2}{\partial x_n} \\ \vdots & & & \\ \frac{\partial F_n}{\partial x_1} & \dots & \dots & \frac{\partial F_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla F_1(\bar{x})^T \\ \nabla F_2(\bar{x})^T \\ \vdots \\ \nabla F_m(\bar{x})^T \end{bmatrix},$$

where each partial derivative is evaluated at $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)^T$.

Chain Rule

Let $F : A \subset \mathbb{R}^m \to \mathbb{R}^k$ be differentiable on the open set $\mathcal{A}$.

Chain Rule

Let $F : A \subset \mathbb{R}^m \to \mathbb{R}^k$ be differentiable on the open set $\mathcal{A}$.

Let $G : B \subset \mathbb{R}^k \to \mathbb{R}^n$ be differentiable on the open set $\mathcal{B}$.

Chain Rule

Let $F : A \subset \mathbb{R}^m \to \mathbb{R}^k$ be differentiable on the open set $\mathcal{A}$.

Let $G : B \subset \mathbb{R}^k \to \mathbb{R}^n$ be differentiable on the open set $\mathcal{B}$.

If $F(A) \subset \mathcal{B}$, then the composite function $G \circ F$ is differentiable on $\mathcal{A}$ and

$$(G \circ F)'(x) = G'(F(x)) \circ F'(x).$$

The Mean Value Theorem

The Mean Value Theorem

(a) If $f : \mathbb{R} \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}$ there exists z between x and y such that

$$f(y) = f(x) + f'(z)(y - x).$$

The Mean Value Theorem

(a) If $f : \mathbb{R} \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}$ there exists z between x and y such that

$$f(y) = f(x) + f'(z)(y - x).$$

(b) If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}$ there is a $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(z)^T (y - x).$$

The Mean Value Theorem

(a) If $f : \mathbb{R} \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}$ there exists z between x and y such that

$$f(y) = f(x) + f'(z)(y - x).$$

(b) If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}$ there is a $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(z)^T (y - x).$$

(c) If $F : \mathbb{R}^n \to \mathbb{R}^m$ continuously differentiable, then for every $x, y \in \mathbb{R}$

$$\|F(y) - F(x)\|_q \leq \left[\sup_{z \in [x,y]} \|F'(z)\|_{(p,q)} \right] \|x - y\|_p .$$

The Mean Value Theorem: Proof

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}^n$ there exists $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(z)(y - x).$$

The Mean Value Theorem: Proof

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}^n$ there exists $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(z)(y - x).$$

Proof: Set $\varphi(t) = f(x + t(y - x))$.

The Mean Value Theorem: Proof

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}^n$ there exists $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(z)(y - x).$$

Proof: Set $\varphi(t) = f(x + t(y - x))$.
Chain rule $\Rightarrow \varphi'(t) = \nabla f(x + t(y - x))^T (y - x)$.

The Mean Value Theorem: Proof

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}^n$ there exists $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(z)(y - x).$$

Proof: Set $\varphi(t) = f(x + t(y - x))$.
Chain rule $\Rightarrow \varphi'(t) = \nabla f(x + t(y - x))^T (y - x)$.
In particular, φ is differentiable.

The Mean Value Theorem: Proof

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}^n$ there exists $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(z)(y - x).$$

Proof: Set $\varphi(t) = f(x + t(y - x))$.
Chain rule $\Rightarrow \varphi'(t) = \nabla f(x + t(y - x))^T (y - x)$.
In particular, φ is differentiable.
By the 1-D MVT there exists $\bar{t} \in (0, 1)$ such that

$$\varphi(1) = \varphi(0) + \varphi'(\bar{t})(1 - 0),$$

The Mean Value Theorem: Proof

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then for every $x, y \in \mathbb{R}^n$ there exists $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(z)(y - x).$$

Proof: Set $\varphi(t) = f(x + t(y - x))$.
Chain rule $\Rightarrow \varphi'(t) = \nabla f(x + t(y - x))^T (y - x)$.
In particular, φ is differentiable.
By the 1-D MVT there exists $\bar{t} \in (0, 1)$ such that

$$\varphi(1) = \varphi(0) + \varphi'(\bar{t})(1 - 0),$$

or equivalently,

$$f(y) = f(x) + \nabla f(z)^T (y - x)$$

where $z = x + \bar{t}(y - x)$.

The Second Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ so that $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$.
The second derivative of f is just the derivative of ∇f as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$:

$$\nabla[\nabla f(x)] = \nabla^2 f(x).$$

The Second Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ so that $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$.
The second derivative of f is just the derivative of ∇f as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$:

$$\nabla[\nabla f(x)] = \nabla^2 f(x).$$

This is an $n \times n$ matrix:

The Second Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ so that $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$.
The second derivative of f is just the derivative of ∇f as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$:

$$\nabla[\nabla f(x)] = \nabla^2 f(x).$$

This is an $n \times n$ matrix:
If ∇f exists and is continuous at x, then it is given as the matrix of second partials of f at x:

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right] .$$

The Second Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ so that $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$.
The second derivative of f is just the derivative of ∇f as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$:

$$\nabla[\nabla f(x)] = \nabla^2 f(x).$$

This is an $n \times n$ matrix:
If ∇f exists and is continuous at x, then it is given as the matrix of second partials of f at x:

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right] .$$

Moreover, $\frac{\partial f}{\partial x_i \partial x_j} = \frac{\partial f}{\partial x_j \partial x_i}$ for all $i, j = 1, \ldots, n$.

The Second Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ so that $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$.
The second derivative of f is just the derivative of ∇f as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$:

$$\nabla[\nabla f(x)] = \nabla^2 f(x).$$

This is an $n \times n$ matrix:
If ∇f exists and is continuous at x, then it is given as the matrix of second partials of f at x:

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right] .$$

Moreover, $\frac{\partial f}{\partial x_i \partial x_j} = \frac{\partial f}{\partial x_j \partial x_i}$ for all $i, j = 1, \ldots, n$. The matrix $\nabla^2 f(x)$ is called the Hessian of f at x. It is a symmetric matrix.

Second-Order Taylor Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable on an open set containing $[x, y]$, then there is a $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(z)(y - x).$$

shown that

$$|f(y) - (f(x) + f'(x)(y - x))| \leq \frac{1}{2}\|x - y\|_p^2 \sup_{z \in [x,y]} \|\nabla^2 f(z)\|_{(p,q)},$$

for any choice of p and q from $[1, \infty]$.