1. Review of Multi-variable Calculus

Throughout this course we will be working with the vector space $\mathbb{R}^n$. For this reason we begin with a brief review of its metric space properties

Definition 1.1 (Vector Norm). *A function $\nu : \mathbb{R}^n \to \mathbb{R}$ is a vector norm on $\mathbb{R}^n$ if*

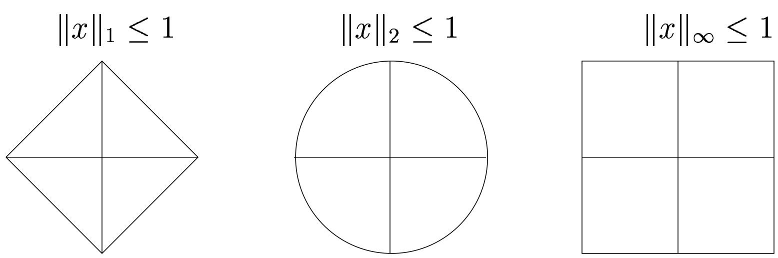
i. $\nu(x) \geq 0 \ \forall \ x \in \mathbb{R}^n$ *with equality iff* $x = 0$.
ii. $\nu(\alpha x) = |\alpha| \nu(x) \ \forall \ x \in \mathbb{R}^n \ \alpha \in \mathbb{R}$
iii. $\nu(x + y) \leq \nu(x) + \nu(y) \ \forall \ x, y \in \mathbb{R}^n$

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

Example: l_p norms

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

– $p = 1, 2, \infty$ are the most important cases.



– The unit ball of a norm is a convex set. We denote the unit ball by $\mathbb{B}$. The unit balls for the $p = 1, 2, \infty$ norms are denoted by $\mathbb{B}_1$, $\mathbb{B}_2$, and $\mathbb{B}_\infty$, respectively.

1.1. **Equivalence of Norms.**

$$\alpha(p, q)\|x\|_q \leq \|x\|_p \leq \beta(p, q)\|x\|_q$$

$\alpha(p, q)$

$p \backslash q$	1	2	3
1	1	1	1
2	$n^{-\frac{1}{2}}$	1	1
3	n^{-1}	$n^{-\frac{1}{2}}$	1

$\beta(p, q)$

$p \backslash q$	1	2	3
1	1	$n^{\frac{1}{2}}$	n
2	1	1	$n^{\frac{1}{2}}$
3	1	1	1

1.2. **Continuity and the Weierstrass Theorem.**

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

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

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

– A subset $D \subset \mathbb{R}^n$ is said to be <u>open</u> 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 $\mathbb{R}^n$.

– A subset $D \subset \mathbb{R}^n$ is said to be closed if every point x satisfying

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

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

– A subset $D \subset \mathbb{R}^n$ is said to be bounded if there exists $m > 0$ such that

$$\|x\| \leq m \text{ for all } x \in D.$$

– A subset $D \subset \mathbb{R}^n$ is said to be compact, if it is closed and bounded.

– A point $\overline{x}$ is said to be a cluster point of the set $D \subset \mathbb{R}^n$ if

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

for every $\epsilon > 0$.

Theorem 1.1 (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 in D.*

Theorem 1.2 (Bolzano-Weierstrass Theorem)**.** *A subset of $\mathbb{R}^{n \times n}$ is compact if and only if it is both closed and bounded.*

Theorem 1.3 (Weierstrass Extreme Value Theorem)**.** *Every continuous function on a compact set attains its extreme values on that set.*

1.3. **Dual Norms.** Let $\|\cdot\|$ be a given norm on $\mathbb{R}^n$ with associated closed unit ball $\mathbb{B}$. For each $x \in \mathbb{R}^n$ define

$$\|x\|_0 := \max\{x^T y : \|y\| \leq 1\}.$$

Since the transformation $y \mapsto x^T y$ is continuous (in fact, linear) and $\mathbb{B}$ is compact, Weierstrass's Theorem says that the maximum in the definition of $\|x\|_0$ is attained. Thus, in particular, the function $x \to \|x\|_0$ is well defined and finite-valued. Indeed, the mapping defines a norm on $\mathbb{R}^n$. This norm is said to be the norm dual to the norm $\|\cdot\|$. Thus, every norm has a norm dual to it.

We now show that the mapping $x \mapsto \|x\|_0$ is a norm.

(a) It is easily seen that $\|x\|_0 = 0$ if and only if $x = 0$. If $x \neq 0$, then

$$\|x\|_0 = \max\{x^T y : \|y\| \leq 1\} \geq x^T \left(\frac{x}{\|x\|}\right) = \frac{\|x\|_2}{\|x\|} > 0.$$

(b) From (a), $\|0 \cdot x\|_0 = 0 = 0 \cdot \|x\|_0$. Next suppose $\alpha \in \mathbb{R}$ with $\alpha \neq 0$. Then

$$\begin{aligned} \|\alpha x\|_0 &= \max\{x^T(\alpha y) : \|y\| \leq 1\}, (z = \alpha y) \\ &= \max\left\{x^T z : 1 \leq \left\|\tfrac{z}{\alpha}\right\| = \tfrac{1}{|\alpha|}\|z\| = \left\|\tfrac{z}{|\alpha|}\right\|\right\}, \left(w = \tfrac{z}{|\alpha|}\right) \\ &= \max\{x^T(|\alpha| z) : 1 \geq \|w\|\} \\ &= |\alpha|\, \|x\|_0. \end{aligned}$$

In order to establish the triangle inequality, we make use of the following elementary, but very useful, fact.

FACT: If $f : \mathbb{R}^n \to \mathbb{R}$ and $C \subset D \subset \mathbb{R}^n$, then

$$\sup_{x \in C} f(x) \leq \sup_{x \in D} f(x).$$

That is, the supremum over a larger set must be larger. Similarly, the infimum over a larger set must be smaller.

(c)
$$\begin{aligned} \|x + z\|_0 &= \max\{x^T y + z^T y : \|y\| \leq 1\} \\ &= \max\left\{x^T y_1 + z^T y_2 : \begin{array}{l} \|y_1\| \leq 1 \\ \|y_2\| \leq 1 \end{array}, y_1 = y_2\right\} \end{aligned}$$

(max over a larger set)

$$\begin{aligned} &= \leq \max\{x^T y_1 + z^T y_2 : \|y_1\| \leq 1, \|y_2\| \leq 1\} \\ &= \|x\|_0 + \|z\|_0 \end{aligned}$$

FACTS:

(i) $x^T y \leq \|x\|\, \|y\|_0$ (apply definition)
(ii) $\|x\|_\infty = \|x\|$
(iii) $(\|x\|_p)_0 = \|x\|_q$ where $\frac{1}{p} + \frac{1}{q} = 1$, $1 \leq p \leq \infty$
(iv) Hölder's Inequality: $|x^T y| \leq \|x\|_p \|y\|_q$

$$\frac{1}{p} + \frac{1}{q} = 1$$

(v) Cauchy-Schwartz Inequality:

$$|x^T y| \leq \|x\|_2 \|y\|_2$$

1.4. **Operator Norms.** $A \in \mathbb{R}^{m \times n}$

$$\|A\|_{(p,q)} = \max\{\|Ax\|_p : \|x\|_q \leq 1\}$$

EXAMPLE:
$$\begin{aligned} \|A\|_2 &= \|A\|_{(2,2)} = \max\{\|Ax\|_2 : \|x\|_2 \leq 1\} \\ \|A\|_\infty &= \|A\|_{(\infty,\infty)} = \max\{\|Ax\|_\infty : \|x\|_\infty \leq 1\} \\ &= \max_{1 \leq i \leq m} \textstyle\sum_{j=1}^n |a_{ij}|, \text{max row form} \\ \|A\|_1 &= \|A\|_{(1,1)} = \max\{\|Ax\|_1 : \|x\|_1 \leq 1\} \\ &= \max_{1 \leq j \leq n} \textstyle\sum_{i=1}^m |a_{ij}|, \text{max column sum} \end{aligned}$$

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

(a) $\|A\| \geq 0$ with equality $\Leftrightarrow \|Ax\| = 0\ \forall\ x$ or $A \equiv 0$.

$$\begin{array}{rrcl}
\text{(b)} & \|\alpha A\| & = & \max\{\|\alpha Ax\| : \|x\| \le 1\} \\
 & & = & \max\{|\alpha|\,\|Ax\| : \|\alpha\| \le 1\} = |\alpha|\,\|A\| \\
\text{(c)} & \|A+B\| & = & \max\{\|Ax+Bx\| : \|x\| \le 1\} \le \max\{\|Ax\| + \|Bx\|\ A \le 1\} \\
 & & = & \max\{\|Ax_1\| + \|Bx_2\| : x_1 = x_2, \|x_1\| \le 1, \|x_2\| \le 1\} \\
 & & \le & \max\{\|Ax_1\| + \|Bx_2\| : \|x_1\| \le 1, \|x_2\| \le 1\} \\
 & & = & \|A\| + \|B\|
\end{array}$$

1.4.1. *Spectral Radius.* $A \in \mathbb{R}^{n\times n}$

$$\rho(A) := \max\{|\lambda| : z \in \Sigma(A)\}$$

$\Sigma(A) = \{\lambda \in \mathbb{C} : Ax = \lambda x \text{ for some } x \neq 0)$.
$\rho(A) \sim$ spectral radius of A
$\Sigma(A) \sim$ spectrum of A

FACT:

(i) $\|A\|_2 = (\rho(A^TA))^{\frac{1}{2}}$
(ii) $\rho(A) < 1 \Leftrightarrow \lim\limits_{k\to\infty} A^k = 0$
(iii) $\rho(A) < 1 \Rightarrow (I-A)^{-1} = \sum\limits_{i=0}^{\infty} A^i$ (Neumann Lemma)

1.4.2. *Condition number.* $A \in \mathbb{R}^{n\times n}$

$$\kappa(A) = \begin{cases} \|A\|\,\|A^{-1}\| & \text{if } A^{-1} \text{ exists} \\ \infty & \text{otherwise} \end{cases}$$

FACT: [Error estimates in the solution of linear equations] If $Ax_1 = b$ and $Ax_2 = b + e$, then

$$\frac{\|x_1 - x_2\|}{\|x_1\|} \le \kappa(A)\frac{\|e\|}{\|b\|}$$

Proof. $\|b\| = \|Ax_1\| \le \|A\|\,\|x_1\| \Rightarrow \frac{1}{\|x_1\|} \le \frac{\|A\|}{\|b\|}$, so

$$\frac{\|x_1 - x_2\|}{\|x_1\|} \le \frac{\|A\|}{\|b\|}\|A^{-1}(A(x_1 - x_2)\| \le \|A\|\,\|A^{-1}\|\frac{1}{\|b\|}\|Ax_1 - Ax_2\|$$

$\square$

1.5. **The Frobenius Norm.** There is one further norm for matrices, called the Frobenius norm, that is very useful. Observe that we can 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 $\mathbb{R}^{(mn)}$. The function that takes a matrix in $\mathbb{R}^{m\times n}$ to a vector in $\mathbb{R}^{(mn)}$ by stacking columns is called vec (or sometimes cvec).

EXAMPLE:

$$\mathrm{vec}\left(\begin{bmatrix} 1 & 2 & -3 \\ 0 & -1 & 4 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 2 \\ -1 \\ -3 \\ 4 \end{bmatrix}$$

Using vec we can define an inner product on $\mathbb{R}^{m\times n}$ by writting

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

This is called the *Frobenius* inner product on $\mathbb{R}^{m\times n}$. It is easy to show that

$$\langle A, B\rangle_F = \mathrm{tr}\left(A^T B\right)\ .$$

This inner product gives rise to the Frobenius norm by the formula

$$\|A\|_F = \sqrt{\langle A, A\rangle_F} = \|\mathrm{vec}\,(A)\|_2.$$

1.6. **Review of Differentiation.** Let $F : \mathbb{R}^n \to \mathbb{R}^m$. In this course we 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 $\mathbb{R}^m$.

1) Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and let $x, d \in \mathbb{R}^n$. 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 Gâteaux differentiable at x.

2) Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and let $x \in \mathbb{R}^n$. If there exists $J \in \mathbb{R}^{m\times n}$ such that

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

then F is said to be Fréchet differentiable at x and J is said to be its "Fréchet derivative". It can be shown that this definition is independent of the choice of norm. We denote J by $J = F'(x)$ or $J = \nabla F(x)$.

3) In the case where $f : \mathbb{R}^n \to \mathbb{R}$, the notation differs a bit from that given above. In this case we write $\nabla f(x) = f'(x)^T$, and we call $\nabla f(x)$ the gradient of f at x.

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 $F'(x)$ exists for all x near $\overline{x}$ and that the mapping $x \mapsto F'(x)$ is continuous at $\overline{x}$,

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

(it can again be shown that continuity is independent of the choice of norm, in). Then $\partial F_i/\partial x_j$ exist for each $i = 1, \ldots, m$, $j = 1, \ldots, n$ and $F'(\overline{x})$ has the representation

$$\nabla F(\overline{x}) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_n} \\ \vdots & & & \\ \frac{\partial F_n}{\partial x_1} & \cdots & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla F_1(\bar{x})^T \\ \nabla F_2(\bar{x})^T \\ \vdots \\ \nabla F_m(\bar{x})^T \end{bmatrix},$$

where each partial derivative is evaluated at $\overline{x} = (\overline{x}_1, \ldots, \overline{x}_n)^T$. This matrix is called the Jacobian matrix for F at $\overline{x}$. However, in the case where $m = 1$, recall from above that $\nabla f(x)$ is called the gradient and $\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right]^T$.

(v) (Chain Rule) Let $F : A \subset \mathbb{R}^m \to \mathbb{R}^k$ be differentiable on the open set A and let $G : B \subset \mathbb{R}^k \to \mathbb{R}^n$ be differentiable on the open set B. If $F(A) \subset B$, then the composite function $G \circ F$ is differentiable on A and

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

(vi) <u>The Mean Value Theorem</u>:

(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 $x, y \in \mathbb{R}$

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

Proof of the Mean Value Theorem: (b): Set $\varphi(t) = f(x + t(y - x))$. Then, by the chain rule, $\varphi'(t) = \nabla f(x+t(y-x))^T(y-x)$ so that φ is differentiable. Moreover, $\varphi : \mathbb{R} \to \mathbb{R}$. Thus, by (a), 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 $z = x + \bar{t}(y - x)$. ■

1.6.1. *The Implicit Function Theorem.* Let $F : \mathbb{R}^{n+m} \to \mathbb{R}^n$ be continuously differentiable on an open set $E \subset \mathbb{R}^{n+m}$. Further suppose that there is a point $(\bar{x}, \bar{y}) \in \mathbb{R}^{n+m}$ at which $F(\bar{x}, \bar{y}) = 0$. If $\nabla_x F(\bar{x}, \bar{y})$ is invertable, then there exist open sets $U \subset \mathbb{R}^{n+m}$ and $W \subset \mathbb{R}^m$, with $(\bar{x}, \bar{y}) \in U$ and $\bar{y} \in W$, having the following property:
To every $y \in W$ corresponds a unique $x \in \mathbb{R}^n$ such that

$$(x, y) \in U \qquad \text{and} \qquad F(x, y) = 0\,.$$

Moreover, if x is defined to be $G(y)$, then G is a continuously differentiable mapping of W into $\mathbb{R}^n$ satisfying

$$G(\bar{y}) = \bar{x}, \quad F(G(y), y) = 0 \;\forall\; y \in W, \quad \text{and} \quad G'(\bar{y}) = -(\nabla_x F(\bar{x}, \bar{y}))^{-1}\nabla_y F(\bar{x}, \bar{y})\;.$$

1.6.2. *Some facts about 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 $\mathbb{R}^n$. Hence, $\nabla[\nabla f(x)] = \nabla^2 f(x)$ is an $n \times n$ matrix (note that we can also denote $\nabla^2 f(x)$ by $f''(x)$).

(i) 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.

(ii) 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

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

It can be shown that

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

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

1.6.3. *Integration.* Let $f : \mathbb{R}^n \to \mathbb{R}^1$ be differentiable and set $\varphi(t) = f(x + t(y - x))$ so that $\varphi : \mathbb{R} \to \mathbb{R}$. Then

$$\begin{aligned} f(y) - f(x) &= \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt \\ &= \int_0^1 \nabla f(x_t(y-x))^T(y-x)dt \end{aligned}$$

Similarly, if $F : \mathbb{R}^n \to \mathbb{R}^m$, then

$$\begin{aligned} F(y) - F(x) &= \begin{bmatrix} \int_0^1 \nabla F_1(x + t(y-x))^T(y-x)dt \\ \vdots \\ \int_0^1 \nabla F_m(x + t(y-x))^T(y-x)dt \end{bmatrix} \\ &= \int_0^1 F'(x + t(y-x))(y-x)dt \end{aligned}$$

1.6.4. *More Facts about Continuity.* Let $F : \mathbb{R}^n \to \mathbb{R}^m$.

- We say that F is continuous relative to a set $D \subset \mathbb{R}^n$ if for every $x \in D$ and $\epsilon > 0$ there exists a $\delta(x, \epsilon) > 0$ such that

$$\|F(y) - F(x)\| \le \epsilon \text{ whenever } \|y - x\| \le \delta(x, \epsilon) \ \text{ and } \ y \in D.$$

- We say that F is <u>uniformly</u> continuous relative to $D \subset \mathbb{R}^n$ if for every $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that

$$\|F(y) - F(x)\| \le \epsilon \text{ whenever } \|y - x\| \le \delta(\epsilon) \ \text{ and } \ y \in D.$$

FACT: If F is continuous on a compact set $D \subset \mathbb{R}^n$, then F is uniformly continuous on D.

- We say that F is Lipschitz continuous relative to a set $D \subset \mathbb{R}^n$ if there exists a constant $K \ge 0$ such that

$$\|F(x) - F(y)\| \le K\|x - y\|$$

for all $x, y \in D$.

FACT: Lipschitz continuity implies uniform continuity.

Proof. $\delta = \epsilon / K$. □

EXAMPLES:

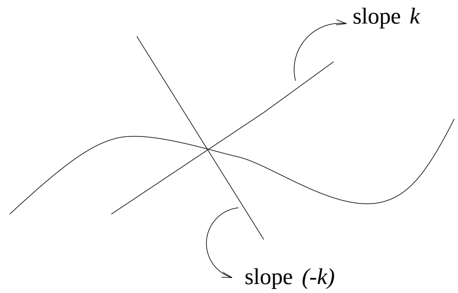
(1) $4(x) = x^{-1}$ is continuous on $(0, 1)$, but it is not uniformly continuous on $(0, 1)$.
(2) $f(x) = \sqrt{x}$ is uniformly continuous on $[0, 1]$, but it is not Lipschitz continuous on $[0, 1]$.

FACT: If F' exists and is continuous on a compact convex set $D \subset \mathbb{R}^m$, then F is Lipschitz continuous on D.

Proof. Mean value Theorem:

$$\|F(x) - F(y)\| \leq (\sup_{z\in[x,y]} \|F'(z)\|)\|x-y\|.$$

Lipschitz continuity is almost but not quite a differentiability hypothesis. The Lipschitz constant provides bounds on rate of change.



□

1.6.5. *Quadratic Bound Lemma.* Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be such that F' is Lipschitz continuous on the convex set $D \subset \mathbb{R}^n$. Then

$$\|F(y) - (F(x) + F'(x)(y-x))\| \leq \frac{K}{2}\|y-x\|^2$$

for all $x, y \in D$ where K is a Lipschitz constant for F' on D.

Proof.

$$\begin{aligned} F(y) - F(x) - F'(x)(y-x) &= \int_0^1 F'(x+t(y-x))(y-x)dt - F'(x)(y-x) \\ &= \int_0^1 [F'(x+t(y-x)) - F'(x)](y-x)dt \end{aligned}$$

$$\begin{aligned} \|F(y) - (F(x) + F'(x)(y-x))\| &= \| \int_0^1 [F'(x+t(y-x)) - F'(x)](y-x)dt\| \\ &\leq \int_0^1 \|(F'(x+t(y-x) - F'(x))(y-x)\|dt \\ &\leq \int_0^1 \|F'(x+t(y-x)) - F'(x)\| \, \|y-x\|dt \\ &\leq \int_0^1 Kt\|y-x\|^2 dt \\ &= \frac{K}{2}\|y-x\|^2. \end{aligned}$$

□

1.6.6. *Some Facts about Symmetric Matrices.* Let $H \in \mathbb{R}^{n\times n}$ be symmetric, i.e. $H^T = H$

(1) There exists an orthonormal basis of eigen-vectors for H, i.e. if $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the n eigenvalues of H (not necessarily distinct), then there exist vectors $q_1, \ldots, q_n$ such that $\lambda_i q_i = H q_i$ $i = 1, \ldots, n$ with $q_i^T q_j = \delta_{ij}$. Equivalently, there exists a unitary transformation $Q = \{q_1, \ldots, q_n\}$ such that

$$H = Q \wedge Q^T$$

where $\wedge = \text{diag}[\lambda_1, \ldots, \lambda_n]$.

(2) $H \in \mathbb{R}^{n\times n}$ is positive semi-definite, i.e.

$$x^T H x \geq 0 \text{ for all } x \in \mathbb{R}^n,$$

if and only if $\forall\ \lambda \in \Sigma\left(\frac{1}{2}(H + H^T)\right) \qquad \lambda \geq 0.$