Chapter 0

Mathematical Preliminaries

0.1 Norms

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 0.1.1 (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

\[
\|x\|_p := \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \quad 1 \leq p < \infty
\]

\[
\|x\|_\infty = \max_{i=1,...,n} |x_i|
\]

- \( p = 1, 2, \infty \) are most important cases

\( \|x\|_1 = 1 \quad \|x\|_2 = 1 \quad \|x\|_\infty = 1 \)

- The unit ball of a norm is a convex set.
0.1.1 Equivalence of Norms

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

\[
\begin{array}{c|ccc}
\alpha(p, q) & 1 & 2 & 3 \\
\hline
1 & 1 & 1 & 1 \\
2 & n^{-\frac{1}{2}} & 1 & 1 \\
3 & n^{-1} & n^{-\frac{1}{2}} & 1 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
\beta(p, q) & 1 & 2 & 3 \\
\hline
1 & 1 & n^{-\frac{1}{2}} & n \\
2 & 1 & 1 & n^{-\frac{1}{2}} \\
3 & 1 & 1 & 1 \\
\end{array}
\]

0.2 Open, Closed, and Compact Sets

- A subset \( D \subset \mathbb{R}^n \) is said to be **open** if for every \( x \in D \) there exists \( \epsilon > 0 \) such that \( x + \epsilon B \subset D \) where

  \[ x + \epsilon B = \{ x + \epsilon u : u \in B \} \]

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

- A point \( x \) is said to be a cluster point (or accumulation point) of the set \( D \subset \mathbb{R}^n \) if

  \[ (x + \epsilon B) \cap D \neq \emptyset \]

  for every \( \epsilon > 0 \).

- A subset \( D \subset \mathbb{R}^n \) is said to be closed if it contains all of its cluster points.

- 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.

**Fact:** [Bolzano–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 are in \( D \).
0.3 Continuity and the Existence of Extrema

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

$$
\lim_{\|x - \bar{x}\| \to 0} \|F(x) - F(\bar{x})\| = 0,
$$

or equivalently, for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$
\|F(x) - F(\bar{x})\| < \epsilon
$$

whenever $\|x - \bar{x}\| < \delta$. The function $F$ is said to be continuous on a set $D \subset \mathbb{R}^n$ if $F$ is continuous at every point of $D$.

\textbf{Weierstrass Extreme Value Theorem} \hspace{0.5cm} Every continuous function on a compact set attains its extreme values on that set.

0.4 Dual Norms

Let $\| \cdot \|$ be a given norm on $\mathbb{R}^n$ with associated closed unit ball $B$. 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 $B$ is compact, Weierstrass’s Theorem says that the maximum in the definition of $\|x\|_0$ is attained. Thus, in particular, the function $x \mapsto \|x\|_0$ 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

$$
\|\alpha x\|_0 = \max\{x^T(\alpha y) : \|y\| \leq 1\}, (z = \alpha y) = \max\{x^T z : 1 \geq \|\frac{z}{z} \| = \|\frac{x}{\|x\|z}\| \} \cdot (w = \frac{z}{z}) = \max\{x^T(\alpha |w| : 1 \geq \|w\|\} = \|\alpha\| \|x\|_0.
$$
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.

\[
\begin{align*}
(c) \quad \|x + z\|_0 &= \max\{x^T y + z^T y : \|y\| \leq 1\} \\
&= \max\left\{x^T y_1 + z^T y_2 : \|y_1\| \leq 1, \|y_2\| \leq 1, y_1 = y_2 \right\} \\
&= \sup_{y_1, y_2} \{x^T y_1 + z^T y_2 : \|y_1\| \leq 1, \|y_2\| \leq 1\} \\
&\leq \max_{\|y_1\| \leq 1, \|y_2\| \leq 1} \{x^T y_1 + z^T y_2\} \\
&= \|x\|_0 + \|z\|_0
\end{align*}
\]

**FACTS:**

(i) \( x^T y \leq \|x\| \|y\|_0 \) (apply definition)

(ii) \( \|x\|_\infty = \|x\|_\square 

(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-Schwarz Inequality:

\[
|x^T y| \leq \|x\|_2 \|y\|_2
\]

### 0.5 Operators

#### 0.5.1 Operator Norms

\( A \in \mathbb{R}^{m \times n} \)

\[
\|A\|_{(a,b)} = \max\{\|Ax\|_{(a)} : \|x\|_{(b)} \leq 1\}
\]
0.5. OPERATORS

**Example:**
\[
\|A\|_2 = \max\{\|Ax\|_2 : \|x\|_2 \leq 1\}
\]
\[
\|A\|_\infty = \max\{\|Ax\|_\infty : \|x\|_\infty \leq 1\}
= \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|, \text{ max row sum}
\]
\[
\|A\|_1 = \max\{\|Ax\|_1 : \|x\|_1 \leq 1\}
= \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|, \text{ max column sum}
\]

**Fact:**
\[
\|Ax\|_{(a)} \leq \|A\|_{(a,b)} \|x\|_{(b)}.
\]

(a) \(\|A\| \geq 0\) with equality \(\iff \|Ax\| = 0 \forall x \text{ or } A \equiv 0\).

(b) \(\|\alpha A\| = \max\{\|\alpha Ax\| : \|x\| \leq 1\}
= \max\{\|\alpha\| \|Ax\| : \|\alpha\| \leq 1\} = |\alpha| \|A\|
\]

(c) \(\|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\|
\]

0.5.2 Spectral Radius

\(A \in \mathbb{R}^{n \times n}\)

\[
\rho(A) := \max\{|\lambda| : \lambda \in \Sigma(A)\}
\]

\(\Sigma(A) = \{\lambda \in \mathbb{C} : Ax = \lambda x \text{ for some } x \neq 0\}.
\]

\(\rho(A) \sim \text{spectral radius of } A\)

\(\Sigma(A) \sim \text{spectrum of } A\)

**Fact:**

(i) \(\|A\|_2 = (\rho(A^T A))^{\frac{1}{2}}\)

(ii) \(\rho(A) < 1 \iff \lim_{k \to \infty} A^k = 0\)

(iii) \(\rho(A) < 1 \Rightarrow (I - A)^{-1} = \sum_{i=0}^{\infty} A^i\) (Neumann Lemma)
0.5.3 Condition number

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

**Fact:** [Error estimates in the solution of linear equations] If \( Ax_1 = b \) and \( Ax_2 = b + e \), then

\[ \frac{\|x_1 - x_2\|}{\|x_1\|} \leq \kappa(A) \frac{\|e\|}{\|b\|} \]

**Proof:**

\[ \|b\| = \|Ax_1\| \leq \|A\| \|x_1\| \Rightarrow \frac{1}{\|x_1\|} \leq \frac{\|A\|}{\|b\|}, \] so

\[ \frac{\|x_1 - x_2\|}{\|x_1\|} \leq \frac{\|A\|}{\|b\|} \|A^{-1}(A(x_1 - x_2))\| \leq \|A\| \|A^{-1}\| \frac{1}{\|b\|} \|Ax_1 - Ax_2\| \]

\[ \blacksquare \]

0.5.4 The Frobenius Norm

There is one further norm for matrices that is very useful. It is called the Frobenius norm. 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 mapping from \( \mathbb{R}^{m \times n} \) to \( \mathbb{R}^{mn} \) defined in this way is denoted by \( \text{vec} \cdot \). The Frobenius norm of a matrix \( A \in \mathbb{R}^{m \times n} \) is then the 2-norm of \( \text{vec}(A) \). It can be verified that

\[ \|A\|_F = \text{tr}(A^T A). \]

0.6 Review of Differentiation

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”. We denote \( J \) by \( J = F'(x) \) and write
\[
F(y) = F(x) + F'(x)(y - x) + o(\|y - x\|),
\]
where the “little-o” notation signifies
\[
\lim_{t \to 0} \frac{o(t)}{t} = 0.
\]

**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 \( \bar{x} \) and that the mapping \( x \mapsto F'(x) \) is continuous at \( \bar{x} \),
\[
\lim_{\|x-\bar{x}\| \to 0} \|F'(x) - F'(\bar{x})\| = 0,
\]
then \( \partial F_i / \partial x_j \) exist for each \( i = 1, \ldots, m \), \( j = 1, \ldots, n \) and with respect to the standard basis the linear operator \( F'(\bar{x}) \) has the representation
\[
\nabla F(\bar{x}) = \left[ \begin{array}{ccccc} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_n} \\ \vdots \\ \frac{\partial F_m}{\partial x_1} & \frac{\partial F_m}{\partial x_2} & \cdots & \frac{\partial F_m}{\partial x_n} \end{array} \right]^T = \left[ \frac{\partial F_i}{\partial x_j} \right]^T
\]
where each partial derivative is evaluated at \( \bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)^T \). This matrix is called the Jacobian matrix for \( F \) at \( \bar{x} \).

**Notation:** For \( f : \mathbb{R}^n \to \mathbb{R} \), \( f'(x) = \left[ \frac{\partial f_1}{\partial x_1}, \ldots, \frac{\partial f_n}{\partial x_n} \right] \) we write \( \nabla f(x) = f'(x)^T \).
(v) If \( F : \mathbb{R}^n \to \mathbb{R}^m \) has continuous partials \( \partial F_i / \partial x_i \) on an open set \( D \subset \mathbb{R}^n \), then \( F \) is differentiable on \( D \). Moreover, in the standard basis the matrix representation for \( F'(x) \) is the Jacobian of \( F \) at \( x \).

(vi) (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

\[
(G \circ F)'(x_0) = G'(F(x_0)) \circ F'(x_0).
\]

Remarks: Let \( F : \mathbb{R}^n \to \mathbb{R}^m \) be differentiable. If \( L(\mathbb{R}^n, \mathbb{R}^m) \) denotes the set of linear maps from \( \mathbb{R}^n \) to \( \mathbb{R}^m \), then

\[
F' : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m).
\]

(In a standard basis we usually identify \( L(\mathbb{R}^n, \mathbb{R}^m) \) with \( \mathbb{R}^{m \times n} \).) Therefore hierarchy for higher derivatives:

\[
F : \mathbb{R}^n \to \mathbb{R}^m \\
F' : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m) \approx \mathbb{R}^{m \times n} \\
F'' : \mathbb{R}^n \to L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) \approx \mathbb{R}^{m \times n \times n} \\
F''' : \mathbb{R}^n \to L(\mathbb{R}^n, L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)))) \approx \mathbb{R}^{m \times n \times n \times n} \\
\vdots
\]

(v) 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 \( x, y \in \mathbb{R} \)

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

Proof of (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 \( \varphi \) 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 \( z = x + \bar{t}(y - x) \).
0.6.1 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 invertible, 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 $(x, y) \in U$ and $F(x, y) = 0$.

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

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

0.6.2 Some facts about the Second Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$ so that $f' : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}) (\approx \mathbb{R}^{n \times 1} = \mathbb{R}^2)$ and $f'' : \mathbb{R}^n \to L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R})) (\approx \mathbb{R}^{n \times n \times 1} = \mathbb{R}^{n \times n}).$

(i) If $f''$ exists and is continuous at $x_0$, then in the standard basis

$$f''(x_0) \approx \nabla^2 f(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_i \partial x_j} \end{bmatrix}_{x=x_0}$$

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 $\nabla^2 f(x_2)$ is called the Hessian of $f$ at $x_0$. 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

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

We also obtain

$$\|f(y) - (f(x) + f'(x)(y - x))\| \leq \frac{1}{2} \|y - x\|^2 \sup_{z \in [x, y]} \|f''(z)\|.$$
0.6.3 Integration
Let \( f : \mathbb{R}^n \to \mathbb{R} \) be differentiable and set \( \varphi(t) = f(x + t(y - x)) \) so that \( \varphi : \mathbb{R} \to \mathbb{R} \). Then
\[
\begin{align*}
    f(y) - f(x) &= \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) \, dt \\
    &= \int_0^1 \nabla f(x + t(y - x))^T (y - x) \, dt
\end{align*}
\]
Similarly, if \( F : \mathbb{R}^n \to \mathbb{R}^m \), then
\[
\begin{align*}
    F(y) - F(x) &= \begin{bmatrix}
        \int_0^1 \nabla F_1(x + t(y - x))^T (y - x) \, dt \\
        \vdots \\
        \int_0^1 \nabla F_m(x + t(y - x))^T (y - x) \, dt
    \end{bmatrix} \\
    &= \int_0^1 F'(x + t(y - x))(y - x) \, dt
\end{align*}
\]

0.6.4 More Facts about Continuity
Let \( F : \mathbb{R}^n \to \mathbb{R}^m \).

- We say that \( F \) is continuous on 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
\[
    \|F(y) - F(x)\| \leq \epsilon \quad \text{whenever} \quad \|y - x\| \leq \delta(x, \epsilon).
\]

- We say that \( F \) is uniformly continuous on \( D \subset \mathbb{R}^n \) if for every \( \epsilon > 0 \) there exists a \( \delta(\epsilon) > 0 \) such that
\[
    \|F(y) - F(x)\| \leq \epsilon \quad \text{whenever} \quad \|y - x\| \leq \delta(\epsilon).
\]

\textbf{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 on a set \( D \subset \mathbb{R}^n \) if there exists a constant \( K \geq 0 \) such that
\[
    \|F(x) - F(y)\| \leq K \|x - y\|
\]
for all \( x, y \in D \).

\textbf{Fact:} Lipschitz continuity implies uniform continuity.

\textbf{Proof:} \( \delta = \epsilon/K \). \( \blacksquare \)

\textbf{Examples:}
1. \( f(x) = x^{-1} \) 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:

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

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

---

**Quadratic Bound Lemma**

**Lemma 0.6.1** 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:**

\[
F(y) - F(x) - F'(x)(y - x) = \int_0^1 F'(x + t(y - x))(y - x)dt - F'(x)(y - x)
\]

\[
= \int_0^1 [F'(x + t(y - x)) - F'(x)](y - x)dt
\]

\[
\|F(y) - (F(x) + F'(x)(y - x))\| = \| \int_0^1 [F'(x + t(y - x)) - F'(x)](y - x)dt\|
\]

\[
\leq \int_0^1 \|(F'(x + t(y - x)) - F'(x))(y - x)\|dt
\]

\[
\leq \int_0^1 \|F'(x + t(y - x)) - F'(x)\| \|y - x\|dt
\]

\[
\leq \int_0^1 Kt\|y - x\|^2dt
\]

\[
= \frac{K}{2}\|y - x\|^2.
\]
**Extended Quadratic Bound Lemma**

**Lemma 0.6.2** Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable in an open convex set $D \subset \mathbb{R}^n$. If we assume that $F'$ is Lipschitz continuous in $D$ with Lipschitz constant $K > 0$, then for all $x, y, z \in D$ we have

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

**Proof:** Just as in the proof of the quadratic bound lemma

$$F(y) - F(x) - F'(z)(y - x) = \int_0^1 (F'(x + t(y - x)) - F'(z))(y - x) dt.$$ 

Therefore,

$$\|F(y) - F(x) - F'(z)(y - x)\| \leq \|y - x\| \int_0^t \|x + t(y - x) - z\| dt \leq \|y - x\| \int_0^t K \|t(y - z) + (1 - t)(x - z)\| dt \leq \|y - x\| K \int_0^t \|y - z\| + (1 - t)\|x - z\| dt \leq K \frac{\|y - z\| + \|x - z\|}{2} \|y - x\|.$$ 

\[\Box\]

**0.6.5 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 $\lambda_i q_i = H q_i$ for $i = 1, \ldots, n$ with $q_i^T q_j = \delta_{ij}$. Equivalently, there exists a unitary transformation $Q = \{q_1, \ldots, q_n\}$ such that

$$H = Q \Lambda Q^T$$

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

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 $\forall \lambda \in \sum \left(\frac{1}{2}(H + H^T)\right)$ with $\lambda \geq 0$. 