1 Calculus and analysis review

Almost all books on nonlinear programming have an appendix reviewing the relevant notions. Most of these should be familiar to you from a course in analysis. Most of material in this course is based in some form on these concepts, therefore, to succeed in this course you should be not just familiar, but comfortable working with these concepts.

A few words on symbols and wording in mathematical statements

Three symbols used frequently in these notes are

- \( \forall \) — read as “for every,” or “for any,” or “for all.”
  - E.g., the expression “for any \( x \) in \( \mathbb{R}^n \) such that \( x \) is nonnegative...” is abbreviated as “\( \forall x \in \mathbb{R}^n : x \geq 0... \)”

- \( \exists \) — read as “there exists a...” or, less formally, “one can find a...”
  - E.g., the expression “for any \( x \in S \), there exists (or one can find) a value \( \delta > 0 \) such that \( \| x \| \leq \delta \)” is abbreviated as “\( \exists \delta > 0 \text{ such that } \| x \| \leq \delta \).”
  - Note that the above is meant to indicate that the value of \( \delta \) can depend on the specific \( x \); however, if the statement read “one can find (or there exists) a value \( \delta > 0 \) such that \( \| x \| \leq \delta \text{ for any } x \in S \)” (abbreviated as “\( \exists \delta > 0 : \| x \| \leq \delta \text{ } \forall x \in S \)”), that would mean that the same value of \( \delta \) has to “work” for every \( x \in S \).

- \( \leftrightarrow \) — read as “if and only if.”

Vectors and Norms

- \( \mathbb{R}^n \): set of \( n \)-dimensional real vectors \( (x_1, \ldots, x_n)^T \) (“\( x^T \)” — transpose)

- Definition: norm \( \| \cdot \| \) on \( \mathbb{R}^n \): a mapping of \( \mathbb{R}^n \) into \( \mathbb{R} \) such that:
  1. \( \| x \| \geq 0 \forall x \in \mathbb{R}^n; \| x \| = 0 \iff x = 0. \)
  2. \( \| cx \| = |c| \cdot \| x \| \forall c \in \mathbb{R}, x \in \mathbb{R}^n. \)
  3. \( \| x + y \| \leq \| x \| + \| y \| \forall x, y \in \mathbb{R}^n. \)

- Euclidean norm: \( \| \cdot \|_2: \| x \|_2 = \sqrt{x^T x} = \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2}. \)
  - Schwartz inequality: \( |x^T y| \leq \| x \|_2 \cdot \| y \|_2 \); equality holds if and only if \( x = \alpha y \).

- Other norm examples: \( \| x \|_1 = \sum_{i=1}^{n} |x_i|; \| x \|_\infty = \max_{i=1,\ldots,n} |x_i|, \) etc.

- All norms in \( \mathbb{R}^n \) are equivalent, i.e., for any \( \| \cdot \|_1 \) and \( \| \cdot \|_2 \) \( \exists \alpha_1, \alpha_2 > 0 \) s.t. \( \alpha_1 \| x \|_1 \leq \| x \|_2 \leq \alpha_2 \| x \|_1 \forall x \in \mathbb{R}^n. \)

- \( \epsilon \)-Neighborhood: \( N_\epsilon(x) = B(x, \epsilon) = \{ y : \| y - x \| \leq \epsilon \} \) (sometimes — strict inequality).

Sequences and Limits.

Sequences in \( \mathbb{R} \)

- Notation: a sequence: \( \{ x_k : k = 1, 2, \ldots \} \subset \mathbb{R}, \{ x_k \} \) for short.
• Definition: \( \{x_k\} \subset \mathbb{R} \) converges to \( x \in \mathbb{R} \) (\( x_k \to x \), \( \lim_{k \to \infty} x_k = x \)) if

\[
\forall \varepsilon > 0 \ \exists K : |x_k - x| \leq \varepsilon \ \text{ (equiv. } x_k \in B(x, \varepsilon) \ \forall k \geq K).
\]

\( x_k \to -\infty \) if \( \forall \varepsilon > 0 \ \exists K : x_k \leq x - \varepsilon \) \( \forall k \geq K \).

\( x_k \to +\infty \) if \( \forall \varepsilon > 0 \ \exists K : x_k \geq x + \varepsilon \) \( \forall k \geq K \).

• Definition: \( \{x_k\} \) is **bounded above** (below): \( \exists A : x_k \leq A \ (x_k \geq A) \ \forall k \).

• Definition: \( \{x_k\} \) is **bounded**: \( \{|x_k|\} \) is bounded; equivalently, \( \{x_k\} \) bounded above and below.

• Definition: \( \{x_k\} \) is **nonincreasing** (nondecreasing): \( x_{k+1} \leq x_k \ (x_{k+1} \geq x_k) \ \forall k \);

**monotone**: nondecreasing or nonincreasing.

• Proposition: Every monotone sequence in \( \mathbb{R} \) has a limit (possibly infinite). If it is also bounded, the limit is finite.

**Sequences in** \( \mathbb{R}^n \)

• Definition: \( \{x_k\} \subset \mathbb{R}^n \) converges to \( x \in \mathbb{R}^n \) (is bounded) if \( \{x^i_k\} \) (the sequence of \( i \)th coordinates of \( x_k \)’s) converges to the \( x^i \) (is bounded) \( \forall i \).

• Propositions:
  - \( x_k \to x \iff \|x_k - x\| \to 0 \)
  - \( \{x_k\} \) is bounded \( \iff \{\|x_k\|\} \) is bounded

• Note: \( \|x_n\| \to \|x\| \) does not imply that \( x_n \to x \)!! (Unless \( x = 0 \)).

**Limit Points**

• Definition: \( x \) is a limit point of \( \{x_k\} \) if there exists an infinite subsequence of \( \{x_k\} \) that converges to \( x \).

• Definition: \( x \) is a limit point of \( A \subset \mathbb{R}^n \) if there exists an infinite sequence \( \{x_k\} \subset A \) that converges to \( x \).

• To see the difference between limits and limit points, consider the sequence

\[
\{(1,0), (0,1), (-1,0), (0,-1), (1,0), (0,1), (-1,0), (0,-1), \ldots\}
\]

• Proposition: let \( \{x_k\} \subset \mathbb{R}^n \)
  - If \( \{x_k\} \) is bounded, then \( \{x_k\} \) converges if and only if it has a unique limit point
  - If \( \{x_k\} \) is bounded, it has at least one limit point

**Infimum and Supremum**

• Let \( A \subset \mathbb{R} \).

  **Supremum** of \( A \) (sup \( A \)): smallest \( y \in \mathbb{R} \) that satisfies \( x \leq y \ \forall x \in A \).

  **Infimum** of \( A \) (inf \( A \)): largest \( y \in \mathbb{R} \) that satisfies \( x \geq y \ \forall x \in A \).

• Not the same as **maximum** and **minimum**, which are the largest and smallest elements of the set \( A \)! Consider, for example, \( A = (0,1) \).
Closed and Open Sets

- Definition: a set $A \subseteq \mathbb{R}^n$ is closed if it contains all its limit points. In other words, for any sequence $\{x_k\} \subset A$ that has a limit $x$, $x \in A$.
- Definition: a set $A \subseteq \mathbb{R}^n$ is open if its complement, $\mathbb{R}^n \setminus A$, is closed
- Definition: a point $x \in A$ is interior if there is a neighborhood of $x$ contained in $A$
- Proposition
  1. A set is open $\iff$ All of its elements are interior points.
  2. Union of finitely many closed sets is closed.
  3. Intersection of closed sets is closed.
  4. Union of open sets is open.
  5. Intersection of finitely many open sets is open.
  6. Every subspace of $\mathbb{R}^n$ is closed.
- Examples: neighborhoods of $x$:
  $\{ y : \|y - x\| \leq \epsilon \} \quad \text{— closed}$
  $\{ y : \|y - x\| < \epsilon \} \quad \text{— open}$
- Some sets are neither: $(0,1]$

Functions and Continuity

- $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}$ - a function.
- Definition: $f$ is continuous at $\bar{x}$ if
  $$\forall \epsilon > 0 \ \exists \delta > 0 : \ x \in A, \ ||x - \bar{x}|| < \delta \Rightarrow |f(x) - f(\bar{x})| < \epsilon.$$  
- Proposition: $f$ is continuous at $\bar{x} \iff$ for any $\{x_n\} \subset A : \ x_n \to \bar{x}$ we have $f(x_n) \to f(\bar{x})$. (In other words, $\lim f(x_n) = f(\lim x_n)$.)
- Proposition:
  - Sums, products and inverses of continuous functions are continuous (in the last case, provided the function is never zero).
  - Composition of two continuous functions is continuous.
  - Any vector norm is a continuous function.

Differentiation

Real-valued functions: Let $f : X \to \mathbb{R}$, where $X \subset \mathbb{R}^n$ is open.
- Definition: $f$ is differentiable at $\bar{x} \in X$ if there exists a vector $\nabla f(\bar{x})$ (the gradient of $f$ at $\bar{x}$) and a function $\alpha_{\bar{x}}(y) : X \to \mathbb{R}$ satisfying $\lim_{y \to 0} \alpha_{\bar{x}}(y) = 0$, such that for each $x \in X$
  $$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + ||x - \bar{x}|| \alpha_{\bar{x}}(x - \bar{x}).$$
  $f$ is differentiable on $X$ if $f$ is differentiable $\forall \bar{x} \in X$. The gradient vector is a vector of partial
derivatives:
\[ \nabla f(\bar{x}) = \left( \frac{\partial f(\bar{x})}{\partial x_1}, \ldots, \frac{\partial f(\bar{x})}{\partial x_n} \right)^T. \]

The directional derivative of \( f \) at \( \bar{x} \) in the direction \( d \) is
\[
\lim_{\lambda \to 0} \frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda} = \nabla f(\bar{x})^T d
\]

- Definition: the function \( f \) is twice differentiable at \( \bar{x} \in X \) if there exists a vector \( \nabla f(\bar{x}) \) and an \( n \times n \) symmetric matrix \( H(\bar{x}) \) (the Hessian of \( f \) at \( \bar{x} \)) such that for each \( x \in X \)
\[
f(x) = f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}) + \frac{1}{2}(x - \bar{x})^TH(\bar{x})(x - \bar{x}) + \|x - \bar{x}\|^2\alpha_3(x - \bar{x}),
\]
and \( \lim_{y \to 0} \alpha_3(y) = 0 \). \( f \) is twice differentiable on \( X \) if \( f \) is twice differentiable \( \forall \bar{x} \in X \). The Hessian, which we often denote by \( H(x) \) for short, is a matrix of second partial derivatives:
\[
[H(x)]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j},
\]
and for functions with continuous second derivatives, it will always be symmetric:
\[
\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}
\]

- Example:
\[
f(x) = 3x_1^2x_2^3 + x_2^2x_3^2
\]
\[
\nabla f(x) = \begin{pmatrix} 6x_1x_2^3 \\ 9x_1^2x_2^2 + 2x_2x_3^3 \\ 3x_2^2x_3^2 \end{pmatrix}
\]
\[
H(x) = \begin{bmatrix} 6x_2^3 & 18x_1x_2^2 & 0 \\ 18x_1x_2^2 & 18x_1^2x_2 + 2x_3^3 & 6x_2x_3^2 \\ 0 & 6x_2x_3^2 & 6x_2^2x_3 \end{bmatrix}
\]

- See additional handout to verify your understanding and derive the gradient and Hessian of linear and quadratic functions.

Vector-valued functions: Let \( f : X \to \mathbb{R}^m \), where \( X \subset \mathbb{R}^n \) is open.

- \[
f(x) = f(x_1, \ldots, x_n) = \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ f_2(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{pmatrix},
\]
where each of the functions \( f_i \) is a real-valued function.

- Definition: the Jacobian of \( f \) at point \( \bar{x} \) is the matrix whose \( j \)th row is the gradient of \( f_j \) at \( \bar{x} \), transposed. More specifically, the Jacobian of \( f \) at \( \bar{x} \) is defined as \( \nabla f(\bar{x})^T \), where \( \nabla f(\bar{x}) \) is the matrix with entries:
\[
[\nabla f(\bar{x})]_{ij} = \frac{\partial f_j(\bar{x})}{\partial x_i}.
\]
Notice that the \( j \)th column of \( \nabla f(\bar{x}) \) is the gradient of \( f_j \) at \( \bar{x} \) (what happens when \( m = 1 \))
Example:

\[ f(x) = \begin{pmatrix} \sin x_1 + \cos x_2 \\ e^{3x_1+x_2^2} \\ 4x_1^3 + 7x_1x_2^2 \end{pmatrix}. \]

Then

\[ \nabla f(x)^T = \begin{pmatrix} \cos x_1 & -\sin x_2 \\ 3e^{3x_1+x_2^2} & 2x_2e^{3x_1+x_2^2} \\ 12x_1^2 + 7x_2^2 & 14x_1x_2 \end{pmatrix}. \]

Other well-known results from calculus and analysis will be introduced throughout the course as needed.