1 Definitions

Functions

We consider functions $f : D \to \mathbb{R}$ where typically $D = \mathbb{R}^n$. In some instances, and later for constrained optimization problems, D will most often be a simply connected closed subset of $\mathbb{R}^n$.

Definition 1.1. A function f is said to be bounded above (below) if $f(x) \leq M$ $(f(x) \geq m)$ for all $x \in D$ and some $M \in \mathbb{R}$ $(m \in \mathbb{R})$. If f is both bounded above and bounded below, then we say that f is bounded.

Definition 1.2. The neighborhood of $y \in \mathbb{R}^n$ of radius r is the set $\mathcal{N}(y, r) = \{x \in \mathbb{R}^n \mid \|x - y\| < r\}$, for some choice of norm $\|\cdot\|$.

Definition 1.3. Function f is said to be continuous at $y \in D$ if for every $\epsilon > 0$, there exists $r > 0$ such that $|f(x) - f(y)| < \epsilon$ for all $x \in \mathcal{N}(y, r) \cap D$.

Derivatives

Definition 1.4. The gradient of f, if it exists, is the $n \times 1$ vector $\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n$.

Definition 1.5. The directional derivative of f in direction $p \in \mathbb{R}^n$, if it exists, is $D_p f(x) = \lim\limits_{a \searrow 0} \dfrac{f(x + ap) - f(x)}{a}$. If $\nabla f(x)$ exists, then $D_p f(x) = \nabla f^{\mathsf{T}} p$.

Definition 1.6. The Hessian of f, if it exists, is the $n \times n$ symmetric matrix

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Definition 1.7. Function f is said to be continuously differentiable if ∇f exists and is continuous on D. Furthermore, f is said to be twice continuously differentiable if $\nabla^2 f$ exists and is continuous on D.

Definition 1.8. We say that function f is sufficiently smooth (or simply smooth) if any required derivatives of f exist.

Extremal Points

Definition 1.9. A vector $y \in D$ is a global minimizer (maximizer) of f if $f(y) \leq f(x)$ $(f(y) \geq f(x))$ for all $x \in D$.

Definition 1.10. A vector $y \in D$ is a local minimizer (maximizer) of f if, for some $r > 0$, $f(y) \leq f(x)$ $(f(y) \geq f(x))$ for all $x \in D \cap \mathcal{N}(y, r)$.

Definition 1.11. A vector $y \in D$ is a strict global minimizer (maximizer) of f if $f(y) < f(x)$ $(f(y) > f(x))$ for all $x \in D \backslash \{y\}$.

Definition 1.12. A vector $y \in D$ is a strict local minimizer (maximizer) of f if, for some $r > 0$, $f(y) < f(x)$ $(f(y) > f(x))$ for all $x \in D \cap \mathcal{N}(y, r) \backslash \{y\}$.

Definition 1.13. An extremal point (minimizer or maximizer) y of f is said to be isolated if it is the unique such extremal point in some neighborhood of y.

Approximating Functions

Theorem 1.14. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and $x, p \in \mathbb{R}^n$. Then*

1. $f(x + p) = f(x) + \int_0^1 \nabla f(x + tp)^\mathsf{T} p \, dt$ *for some* $t \in (0, 1)$, *and*
2. $f(x + p) = f(x) + \nabla f(x + tp)^\mathsf{T} p$ *for some* $t \in (0, 1)$.

Proof. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and $x, p \in \mathbb{R}^n$. Consider the function $g(t) = f(x + tp)$ and $y(t) = x + tp$. We have

$$\frac{dg}{dt} = \sum_{i=1}^{n} \frac{\partial f}{\partial y_i} \frac{\partial y_i}{\partial t}, \quad \text{so} \quad \frac{dg}{dt}(a) = \nabla f(y(a))^\mathsf{T} p = \nabla f(x + ap)^\mathsf{T} p.$$

By the fundamental theorem of calculus, $g(1) = g(0) + \int_0^1 g'(t) \, dt$. Then, by substitution

$$f(x + p) = f(x) + \int_0^1 \nabla f(x + tp)^\mathsf{T} p \, dt.$$

Finally, by the mean value theorem, $g(1) = g(0) + g'(\xi)$ for some $\xi \in (0, 1)$. Then, by substitution,

$$f(x + p) = f(x) + \nabla f(x + \xi p)^\mathsf{T} p \text{ for some } \xi \in (0, 1).$$

□

Theorem 1.15. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and $x, p \in \mathbb{R}^n$. Then*

1. $\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)p \, dt$ *for some $t \in (0,1)$, and*
2. $f(x+p) = f(x) + \nabla f(x)^\mathsf{T} p + \frac{1}{2} p^\mathsf{T} \nabla^2 f(x+tp)p$ *for some $t \in (0,1)$.*

Proof. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and $x, p \in \mathbb{R}^n$. Consider the function $g(t) = \nabla f(x+tp)$ and $y(t) = x + tp$. We have

$$\frac{dg}{dt} = \sum_{i=1}^{n} \left(\frac{\partial}{\partial y_i} \nabla f(y) \right) \left(\frac{\partial y_i}{\partial t} \right) = \sum_{i=1}^{n} \left[\frac{\partial}{\partial y_i} \begin{pmatrix} \frac{\partial f}{\partial y_1} \\ \vdots \\ \frac{\partial f}{\partial y_n} \end{pmatrix} \right] \left(\frac{\partial y_i}{\partial t} \right) = \sum_{i=1}^{n} \begin{pmatrix} \frac{\partial^2 f}{\partial y_1 \partial y_i} \\ \vdots \\ \frac{\partial^2 f}{\partial y_n \partial y_i} \end{pmatrix} p_i = \nabla^2 f(y)p.$$

So, $\frac{dg}{dt}(a) = \nabla^2 f(y(a))p$. By the fundamental theorem of calculus, $g(1) = g(0) + \int_0^1 g'(t) \, dt$, and by substitution,

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)p \, dt.$$

(The proof of the second claim is somewhat involved, making careful use of FTC and MVT.)

□

Special Functions and Sets

Definition 1.16. The <u>z-sublevel set</u> of f is $C_z = \{x \in D \mid f(x) \leq z\}$.

Definition 1.17. Set $C \subset \mathbb{R}^n$ is said to be <u>convex</u> if for every $x, y \in C$ and all $\lambda \in (0,1)$, $\lambda x + (1-\lambda)y \in C$.

Definition 1.18. Function f is said to be <u>convex</u> on $C \subseteq D$is convex and if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ for every $x, y \in C$ and all $\lambda \in (0,1)$.

Theorem 1.19. *Suppose f is continuously differentiable on convex set $C \subseteq D$.*

1. *f is convex over C if and only if $f(z) \geq f(x) + (z-x)^\mathsf{T} \nabla f(x), \forall x, z \in C$, and*
2. *f is strictly convex over C if and only if $f(z) > f(x) + (z-x)^\mathsf{T} \nabla f(x), \forall x, z \in C$, whenever $x \neq z$.*

Proof. Let f be a continuously differentiable function on convex set $C \subseteq D$.

($\Rightarrow$) Suppose f is convex. Then, $f(x + \lambda(z - x)) \leq (1 - \lambda) f(x) + \lambda f(z)$. or equivalently

$$f(z) \geq f(x) + \frac{f(x + \lambda(z - x)) - f(x)}{\lambda}.$$

Taking the limit $\lambda \to 0$, the rhs is a directional derivative:

$$f(z) \geq f(x) + (z - x)^\mathsf{T} \nabla f(x).$$

($\Leftarrow$) Suppose $f(z) \geq f(x) + (z - x)^\mathsf{T} \nabla f(x), \forall x, z \in C$. Let $z = \lambda x + (1 - \lambda) y$ for $0 \leq \lambda \leq 1$. We have, $f(x) \geq f(z) + (x - z)^\mathsf{T} \nabla f(z)$ and $f(y) \geq f(z) + (y - z)^\mathsf{T} \nabla f(z)$. Then,

$$\lambda f(x) + (1 - \lambda) f(y) \geq f(z) + \left[\lambda (x - z)^\mathsf{T} + (1 - \lambda)(y - z)^\mathsf{T}\right] f(z) = f(z).$$

That is, f is convex. The proof of part 2 is analogous. $\square$

Theorem 1.20. *Suppose f is twice continuously differentiable on convex set $C \subseteq D$.*

1. *f is convex over C if $\nabla^2 f(x)$ is psd for all $x \in C$, and*
2. *f is strictly convex over C if $\nabla^2 f(x)$ is pd for all $x \in C$, and*
3. *If C is open and f is convex over C, then $\nabla^2 f(x)$ is psd for all $x \in C$.*

Proof. Let f be twice continuously differentiable on convex set $C \subseteq D$.

1. Suppose $\nabla^2 f(x)$ is positive semi-definite for all $x \in C$. Then, for some z, a convex combination of $x, y \in C$,

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

   That is, f is convex.
2. The proof is analogous to that of part 1.
3. (not done)

$\square$

Definition 1.21. Function f is said to be <u>quasiconvex</u> if every sublevel set of f is convex.

Definition 1.22. Function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be <u>coercive</u> if, for every sequence $\{x_k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ satisfying $\lim_{k\to\infty} \|x_k\| = \infty$, we have $\lim_{k\to\infty} f(x_k) = \infty$. We use the simplified notation $\lim_{\|x\|\to\infty} f(x) = \infty$.

Linear Algebra

In this course, we will work almost exclusively with real-valued symmetric matrices. Definitions and results that follow assume real-valued symmetric matrices.

Definition 1.23. An $n \times n$ matrix A with elements a_{ij} is said to be symmetric if $a_{ij} = a_{ji}$ for all $1 \leq i, j \leq n$.

Definition 1.24. A matrix A whose entries satisfy $a_{ij} = 0$ whenever $i \neq j$ is said to be diagonal.

Definition 1.25. An $n \times n$ matrix A satisfying $AA^\top = A^\top A = I$ is said to be orthogonal.

Theorem 1.26. *Every real $n \times n$ symmetric matrix A is diagonalizable as $A = QDQ^\top$, where D is the $n \times n$ diagonal matrix whose entries are n (not necessarily unique) eigenvalues of A, and Q is an $n \times n$ orthogonal matrix whose columns are the corresponding eigenvectors.*

Definition 1.27. The rank of an $n \times n$ symmetric matrix A is the number of nonzero (possibly repeated) eigenvalues of A.

Definition 1.28. Symmetric matrix A is positive definite (pd) if every eigenvalue of A is strictly positive. Symmetric matrix A is positive semidefinite (psd) if every eigenvalue of A is nonnegative. Symmetric matrix A is indefinite if A has at least one positive and at least one negative eigenvalue.

Definition 1.29. Symmetric $n \times n$ matrix A is invertible if the rank of A equals n.