Jim Lambers
MAT 419/519
Summer Session 2011-12
Lecture 2 Notes

These notes correspond to Section 1.2 in the text.

Functions of Several Variables

We now generalize the results from the previous section, pertaining to optimization of functions of one variable, to functions of several variables. However, we first need some notation and definitions.

Definition An *n-vector* in $\mathbb{R}^n$ is an ordered n-tuple $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of real numbers x_i, called the *components* of $\mathbf{x}$.

Vectors belong to *vector spaces*, which support two essential operations. We define addition of two vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$ by

$$\mathbf{x} + \mathbf{y} = (x_1 + y1, x_2 + y_2, \ldots, x_n + y_n),$$

and multiplication of $\mathbf{x}$ and a real number λ by

$$\lambda\mathbf{x} = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n).$$

Multiplication of numbers needs to be generalized to a sort of multiplication operation involving two vectors.

Definition If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ are vectors in $\mathbb{R}^n$, their *dot product* or *inner product* $\mathbf{x} \cdot \mathbf{y}$ is defined by

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{k=1}^{n} x_k y_k.$$

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* if $\mathbf{x} \cdot \mathbf{y} = 0$.

We also need to generalize the notion of absolute value, or a number's magnitude, to the magnitude of a vector.

Definition The *norm* or *length* $\|\mathbf{x}\|$ of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$ is defined by

$$\|\mathbf{x}\| = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2} = (\mathbf{x} \cdot \mathbf{x})^{1/2}.$$

The norm is a real-valued function on $\mathbb{R}^n$ with the following properties:

1. $\|\mathbf{x}\| \geq 0$ for all vectors $\mathbf{x} \in \mathbb{R}^n$.
2. $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
3. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for all vectors $\mathbf{x} \in \mathbb{R}^n$ and all real numbers α.
4. $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, the *Triangle Inequality*
5. $|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$, the *Cauchy-Schwarz Inequality*

Using the norm, the dot product can also be defined as

$$\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$.

Just as the distance between two numbers x and y is given by $|x-y|$, the distance between two points in n-dimensional space can be defined similarly.

Definition If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the *distance* $d(\mathbf{x}, \mathbf{y})$ between $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{1/2}.$$

The *ball* $B(\mathbf{x}, r)$ *centered at* $\mathbf{x}$ *of radius* r is the set of all vectors $\mathbf{y} \in \mathbb{R}^n$ such that $d(\mathbf{x}, \mathbf{y}) < r$.

A point $\mathbf{x}$ in a set $D \subseteq \mathbb{R}^n$ is an *interior point* of D if there exists an $r > 0$ such that $B(\mathbf{x}, r) \subseteq D$. The *interior* D^0 is the set of all interior points of D. A set $G \subseteq \mathbb{R}^n$ is *open* if $G^0 = G$. A set $F \subseteq \mathbb{R}^n$ is *closed* if its complement $G = F^c$ in $\mathbb{R}^n$ is open.

Now, we are prepared to define minimizers and maximizers of functions of n variables.

Definition Suppose that $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$. A point $\mathbf{x}^* \in D$ is a

1. *global minimizer* of $f(\mathbf{x})$ on D if $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in D$;
2. *strict global minimizer* of $f(\mathbf{x})$ on D if $f(\mathbf{x}^*) < f(\mathbf{x})$ for all $\mathbf{x} \in D$;
3. *local minimizer* of $f(\mathbf{x}$ if there exists a $\delta > 0$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ whenever $\mathbf{x} \in B(\mathbf{x}^*, \delta)$;
4. *strict local minimizer* of $f(\mathbf{x})$ if there exists a $\delta > 0$ such that $f(\mathbf{x}^*) < f(\mathbf{x})$ whenever $\mathbf{x} \in B(\mathbf{x}^*, \delta)$ and $\mathbf{x} \neq \mathbf{x}^*$;
5. *critical point* of $f(\mathbf{x})$ if the first partial derivatives of $f(\mathbf{x})$ exist at $\mathbf{x}^*$ and

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0, \quad i = 1, 2, \ldots, n.$$

Using this definition of a critical point, we can now characterize the location of maximizers and minimizers, as in Fermat's theorem in the single-variable case.

Theorem Suppose that $f(\mathbf{x})$ is a real-valued function for which all first partial derivatives of $f(\mathbf{x})$ exist on a subset D of $\mathbb{R}^n$. If $\mathbf{x}^*$ is an interior point of D that is a local minimizer of $f(\mathbf{x})$, then $\mathbf{x}^*$ is a critical point of $f(\mathbf{x})$.

This theorem can be proved by reduction to the single-variable case, in which all variables except one are fixed.

We need to generalize Taylor's Formula to the multi-variable case. Given a function $f : \mathbb{R}^n \to \mathbb{R}$ whose first and second partial derivatives are continuous on an open set containing the line segment

$$[\mathbf{x}^*, \mathbf{x}] = \{\mathbf{w} \in \mathbb{R}^n \,|\, \mathbf{w} = \mathbf{x}^* + t(\mathbf{x} - \mathbf{x}^*), 0 \leq t \leq 1\}$$

joining $\mathbf{x}$ and $\mathbf{x}^*$. By defining the function

$$\varphi(t) = f(\mathbf{x}^* + t(\mathbf{x} - \mathbf{x}^*))$$

and applying Taylor's Formula in conjunction with the multi-variable Chain Rule, we obtain the following result.

Theorem Suppose that $\mathbf{x}^*$, $\mathbf{x} \in \mathbb{R}^n$ and that $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ with continuous first and second partial derivatives on some open set containing the line segment $[\mathbf{x}^*, \mathbf{x}]$. Then there exists a $\mathbf{z} \in [\mathbf{x}^*, \mathbf{x}]$ such that

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*) \cdot Hf(\mathbf{z})(\mathbf{x} - \mathbf{x}^*)$$

where

$$\nabla f(\mathbf{x}) = \left[\begin{array}{cccc} \frac{\partial f}{\partial x_1}(\mathbf{x}) & \frac{\partial f}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{array} \right]$$

is the *gradient* of f, and

$$Hf(\mathbf{x}) = \left[\begin{array}{cccc} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{array} \right]$$

is the *Hessian* of f.

Now we can characterize local or global maximizers or minimizers based on the second partial derivatives, in the same way as in the single-variable case.

Theorem Suppose that $\mathbf{x}^*$ is a critical point of $f(\mathbf{x})$ with continuous first and second partial derivatives on $\mathbb{R}^n$. Then:

1. $\mathbf{x}^*$ is a global minimizer of $f(\mathbf{x})$ if $(\mathbf{x}-\mathbf{x}^*)\cdot Hf(\mathbf{z})(\mathbf{x}-\mathbf{x}^*) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and all $\mathbf{z} \in [\mathbf{x}^*, \mathbf{x}]$;

2. $\mathbf{x}^*$ is a strict global minimizer of $f(\mathbf{x})$ if $(\mathbf{x}-\mathbf{x}^*)\cdot Hf(\mathbf{z})(\mathbf{x}-\mathbf{x}^*) > 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and all $\mathbf{z} \in [\mathbf{x}^*, \mathbf{x}]$;

3. $\mathbf{x}^*$ is a global maximizer of $f(\mathbf{x})$ if $(\mathbf{x}-\mathbf{x}^*)\cdot Hf(\mathbf{z})(\mathbf{x}-\mathbf{x}^*) \leq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and all $\mathbf{z} \in [\mathbf{x}^*, \mathbf{x}]$;

4. $\mathbf{x}^*$ is a strict global maximizer of $f(\mathbf{x})$ if $(\mathbf{x}-\mathbf{x}^*)\cdot Hf(\mathbf{z})(\mathbf{x}-\mathbf{x}^*) < 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and all $\mathbf{z} \in [\mathbf{x}^*, \mathbf{x}]$.

This theorem can be proved using the multi-variable generalization of Taylor's Theorem, in conjunction with the continuity of the second partial derivatives.

Unfortunately, the sign of $(\mathbf{x}-\mathbf{x}^*)\cdot Hf(\mathbf{z})(\mathbf{x}-\mathbf{x}^*)$ is not so easily determined, in comparison to the single-variable counterpart $f''(z)(x-x^*)^2$. To that end, we turn to concepts from linear algebra.

Definition Let A be an $n \times n$ symmetric matrix. The *quadratic form associated with* A is a function $Q_A : \mathbb{R}^n \to \mathbb{R}$ defined by

$$Q_A(\mathbf{y}) = \mathbf{y} \cdot A\mathbf{y} = \sum_{i,j=1}^{n} a_{ij} y_i y_j, \quad \mathbf{y} \in \mathbb{R}^n.$$

Example Let $f(x, y, z) = x^2 - y^2 + 4z^2 - 2xy + 4yz$. Then we have

$$\nabla f(x, y, z) = (2x - 2y, -2y - 2x + 4z, 8z + 4y)$$

and

$$Hf(x, y, z) = \begin{bmatrix} 2 & -2 & 0 \\ -2 & -2 & 4 \\ 0 & 4 & 8 \end{bmatrix}.$$

It follows that

$$\begin{aligned} Q_{Hf}(x, y, z) &= (x, y, z) \cdot Hf(x, y, z)(x, y, z) \\ &= (x, y, z) \cdot (2x - 2y, -2x - 2y + 4z, 4y + 8z) \\ &= 2x^2 - 2y^2 + 8z^2 - 4xy + 8yz \\ &= 2f(x, y, z). \end{aligned}$$

□

The following terms will enable us to more easily describe the conditions for local or global minimizers or maximizers.

Definition Suppose that A is an $n \times n$ symmetric matrix and that $Q_A(\mathbf{y}) = \mathbf{y} \cdot A\mathbf{y}$ is the quadratic form associated with A. Then A and Q_A are called:

1. *positive semidefinite* if $Q_A(\mathbf{y}) \geq 0$ for all $\mathbf{y} \in \mathbb{R}^n$;
2. *positive definite* if $Q_A(\mathbf{y}) > 0$ for all $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{y} \neq \mathbf{0}$;
3. *negative semidefinite* if $Q_A(\mathbf{y}) \leq 0$ for all $\mathbf{y} \in \mathbb{R}^n$;
4. *negative definite* if $Q_A(\mathbf{y}) < 0$ for all $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{y} \neq \mathbf{0}$;
5. *indefinite* if $Q_A(\mathbf{y}) > 0$ for some $\mathbf{y} \in \mathbb{R}^n$ and $Q_A(\mathbf{y}) < 0$ for other $\mathbf{y} \in \mathbb{R}^n$.

With these terms, the preceding theorem can be restated more concisely as follows:

Theorem Suppose that $\mathbf{x}^*$ is a critical point of a function $f(\mathbf{x})$ with continuous first and second partial derivatives on $\mathbb{R}^n$ and that $Hf(\mathbf{x})$ is the Hessian of $f(\mathbf{x})$. Then $\mathbf{x}^*$ is a

1. global minimizer of $f(\mathbf{x})$ if $Hf(\mathbf{x})$ is positive semidefinite on $\mathbb{R}^n$;
2. strict global minimizer of $f(\mathbf{x})$ if $Hf(\mathbf{x})$ is positive definite on $\mathbb{R}^n$;
3. global maximizer of $f(\mathbf{x})$ if $Hf(\mathbf{x})$ is negative semidefinite on $\mathbb{R}^n$;
4. strict global maximizer of $f(\mathbf{x})$ if $Hf(\mathbf{x})$ is negative definite on $\mathbb{R}^n$.

It remains to determine when a given matrix is positive (or negative) definite (or semidefinite). This will be taken up in subsequent lectures.

Exercises

1. Chapter 1, Exercise 3
2. Chapter 1, Exercise 4
3. Chapter 1, Exercise 5