Introduction to Optimization, and Optimality Conditions for Unconstrained Problems

Robert M. Freund

February, 2004

1 Preliminaries

1.1 Types of optimization problems

Unconstrained Optimization Problem:

$$
\begin{array}{lll}
(\mathrm{P}) & \min_x & f(x) \\
& \text{s.t.} & x \in X,
\end{array}
$$

where $x = (x_1, \ldots, x_n) \in \Re^n$, $f(x) : \Re^n \to \Re$, and X is an open set (usually $X = \Re^n$).

We say that x is a *feasible solution* of (P) if $x \in X$.

Constrained Optimization Problem:

$$
\begin{array}{llll}
(\mathrm{P}) & \min_x & f(x) & \\
& \text{s.t.} & g_i(x) \le 0 & i = 1, \ldots, m \\
& & h_i(x) = 0 & i = 1, \ldots, l \\
& & x \in X, &
\end{array}
$$

where $g_1(x), \ldots, g_m(x), h_1(x), \ldots, h_l(x) : \Re^n \to \Re$.

Let $g(x) = (g_1(x), \ldots, g_m(x)) : \Re^n \to \Re^m$, $h(x) = (h_1(x), \ldots, h_l(x))$:

$\Re^n \to \Re^l$. Then (P) can be written as

$$
\begin{aligned}
\text{(P)} \quad \min_x \quad & f(x) \\
\text{s.t.} \quad & g(x) \le 0 \\
& h(x) = 0 \\
& x \in X.
\end{aligned}
\tag{1}
$$

We say that x is a *feasible solution* of (P) if $g(x) \le 0, h(x) = 0$, and $x \in X$.

1.2 Local, Global, and Strict Optima

The *ball* centered at $\bar{x}$ with radius ϵ is the set:

$$B(\bar{x}, \epsilon) := \{x \mid \|x - \bar{x}\| \le \epsilon\}.$$

Consider the following optimization problem over the set $\mathcal{F}$:

$$
\begin{aligned}
P: \quad \min_x \text{ or } \max_x \quad & f(x) \\
\text{s.t.} \quad & x \in \mathcal{F}
\end{aligned}
$$

We have the following definitions of local/global, strict/non-strict minima/maxima.

Definition 1.1 *$x \in \mathcal{F}$ is a* local minimum *of P if there exists $\epsilon > 0$ such that $f(x) \le f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$.*

Definition 1.2 *$x \in \mathcal{F}$ is a* global minimum *of P if $f(x) \le f(y)$ for all $y \in \mathcal{F}$.*

Definition 1.3 *$x \in \mathcal{F}$ is a* strict local minimum *of P if there exists $\epsilon > 0$ such that $f(x) < f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$, $y \neq x$.*

Definition 1.4 *$x \in \mathcal{F}$ is a* strict global minimum *of P if $f(x) < f(y)$ for all $y \in \mathcal{F}$, $y \neq x$.*

Definition 1.5 *$x \in \mathcal{F}$ is a* local maximum *of P if there exists $\epsilon > 0$ such that $f(x) \geq f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$.*

Definition 1.6 *$x \in \mathcal{F}$ is a* global maximum *of P if $f(x) \geq f(y)$ for all $y \in \mathcal{F}$.*

Definition 1.7 *$x \in \mathcal{F}$ is a* strict local maximum *of P if there exists $\epsilon > 0$ such that $f(x) > f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$, $y \neq x$.*

Definition 1.8 *$x \in \mathcal{F}$ is a* strict global maximum *of P if $f(x) > f(y)$ for all $y \in \mathcal{F}$, $y \neq x$.*

1.3 Gradients and Hessians

Let $f(x) : X \to \Re$, where $X \subset \Re^n$ is open. $f(x)$ is *differentiable* at $\bar{x} \in X$ if there exists a vector $\nabla f(\bar{x})$ (the *gradient* of $f(x)$ at $\bar{x}$) such that for each $x \in X$

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^t (x - \bar{x}) + \|x - \bar{x}\| \alpha(\bar{x}, x - \bar{x}),$$

and $\lim_{y \to 0} \alpha(\bar{x}, y) = 0$. $f(x)$ is *differentiable on X* if $f(x)$ is differentiable for all $\bar{x} \in X$. The gradient vector is the vector of partial derivatives:

$$\nabla f(\bar{x}) = \left(\frac{\partial f(\bar{x})}{\partial x_1}, \ldots, \frac{\partial f(\bar{x})}{\partial x_n} \right)^t.$$

Example 1 *Let $f(x) = 3(x_1)^2(x_2)^3 + (x_2)^2(x_3)^3$. Then*

$$\nabla f(x) = \left(6(x_1)(x_2)^3, 9(x_1)^2(x_2)^2 + 2(x_2)(x_3)^3, 3(x_2)^2(x_3)^2 \right)^T .$$

The *directional derivative* of $f(x)$ at $\bar{x}$ in the direction d is:

$$\lim_{\lambda \to 0} \frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda} = \nabla f(\bar{x})^t d$$

The function $f(x)$ 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 $f(x)$ at $\bar{x}$) such that for each $x \in X$

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^t (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^t H(\bar{x})(x - \bar{x}) + \|x - \bar{x}\|^2 \alpha(\bar{x}, x - \bar{x}),$$

and $\lim_{y \to 0} \alpha(\bar{x}, y) = 0$. $f(x)$ is *twice differentiable on* X if $f(x)$ is twice differentiable for all $\bar{x} \in X$. The Hessian is the matrix of second partial derivatives:

$$H(\bar{x})_{ij} = \frac{\partial^2 f(\bar{x})}{\partial x_i \partial x_j}.$$

Example 2 *Continuing Example 1, we have*

$$H(x) = \begin{pmatrix} 6(x_2)^3 & 18(x_1)(x_2)^2 & 0 \\ 18(x_1)(x_2)^2 & 18(x_1)^2(x_2) + 2(x_3)^3 & 6(x_2)(x_3)^2 \\ 0 & 6(x_2)(x_3)^2 & 6(x_2)^2(x_3) \end{pmatrix}.$$

1.4 Positive Semidefinite and Positive Definite Matrices

An $n \times n$ matrix M is called

- *positive definite* if $x^t M x > 0$ for all $x \in \Re^n, \ x \neq 0$
- *positive semidefinite* if $x^t M x \geq 0$ for all $x \in \Re^n$
- *negative definite* if $x^t M x < 0$ for all $x \in \Re^n, \ x \neq 0$
- *negative semidefinite* if $x^t M x \leq 0$ for all $x \in \Re^n, \ x \neq 0$

- *indefinite* if there exists $x, y \in \Re^n$ for which $x^t M x > 0$ and $y^t M y < 0$

We say that M is SPD if M is symmetric and positive definite. Similarly, we say that M is SPSD if M is symmetric and positive semi-definite.

Example 3

$$M = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

is positive definite.

Example 4

$$M = \begin{pmatrix} 8 & -1 \\ -1 & 1 \end{pmatrix}$$

is positive definite. To see this, note that for $x \neq 0$,

$$x^T M x = 8x_1^2 - 2x_1 x_2 + x_2^2 = 7x_1^2 + (x_1 - x_2)^2 > 0 \ .$$

1.5 Existence of Optimal Solutions

Most of the topics of this course are concerned with

- existence of optimal solutions,
- characterization of optimal solutions, and
- algorithms for computing optimal solutions.

To illustrate the questions arising in the first topic, consider the following optimization problems:

- $$\begin{array}{lll} \text{(P)} & \min_x & \dfrac{1+x}{2x} \\ & \text{s.t.} & x \geq 1 \ . \end{array}$$

Here there is no optimal solution because the feasible region is unbounded

- $$\begin{array}{lll} (\text{P}) & \min_x & \dfrac{1}{x} \\ & \text{s.t.} & 1 \le x < 2 \ . \end{array}$$

  Here there is no optimal solution because the feasible region is not closed.

- $$\begin{array}{lll} (\text{P}) & \min_x & f(x) \\ & \text{s.t.} & 1 \le x \le 2 \ , \end{array}$$

  where
  $$f(x) = \begin{cases} 1/x, & x < 2 \\ 1, & x = 2 \ . \end{cases}$$

  Here there is no optimal solution because the function $f(\cdot)$ is not sufficiently smooth.

Theorem 1 (Weierstrass' Theorem for sequences) *Let $\{x_k\}$, $k \to \infty$ be an infinite sequence of points in the compact (i.e., closed and bounded) set F. Then some infinite subsequence of points x_{k_j} converges to a point contained in F.*

Theorem 2 (Weierstrass' Theorem for functions) *Let $f(x)$ be a continuous real-valued function on the compact nonempty set $F \subset \Re^n$. Then F contains a point that minimizes (maximizes) $f(x)$ on the set F.*

Proof: Since the set F is bounded, $f(x)$ is bounded below on F. Since $F \neq \emptyset$, there exists $v = \inf_{x \in F} f(x)$. By definition, for any $\epsilon > 0$, the set $F_\epsilon = \{x \in F : v \le f(x) \le v + \epsilon\}$ is non-empty. Let $\epsilon_k \to 0$ as $k \to \infty$, and let $x_k \in F_{\epsilon_k}$. Since F is bounded, there exists a subsequence of $\{x_k\}$ converging to some $\bar{x} \in F$. By continuity of $f(x)$, we have $f(\bar{x}) = \lim_{k\to\infty} f(x_k)$, and, since $v \le f(x_k) \le v + \epsilon_k$, it follows that $f(\bar{x}) = \lim_{k\to\infty} f(x_k) = v$. ∎

The assumptions of Weierstrass' Theorem can be somewhat relaxed. For example, for a minimization problem, we can assume that

- the set $\{x \in F : f(x) \leq f(x')\}$ is compact for some $x' \in F$, and
- $f(x)$ is *lower semi-continuous*, i.e., for any constant c, the set $\{x \in F : f(x) \leq c\}$ is closed.

The proof is similar to the proof of the Weierstrass' Theorem.

2 Optimality Conditions for Unconstrained Problems

$$\text{(P)} \quad \min \quad f(x)$$

$$\text{s.t.} \quad x \in X,$$

where $x = (x_1, \ldots, x_n) \in \Re^n$, $f(x) : \Re^n \to \Re$, and X is an open set (usually $X = \Re^n$).

Definition 2.1 *The direction $\bar{d}$ is called a* descent direction *of $f(x)$ at $x = \bar{x}$ if*

$$f(\bar{x} + \epsilon \bar{d}) < f(\bar{x}) \text{ for all } \epsilon > 0 \text{ and sufficiently small .}$$

A *necessary condition* for local optimality is a statement of the form: "if $\bar{x}$ is a local minimum of (P), then $\bar{x}$ must satisfy ..." Such a condition helps us identify all candidates for local optima.

Theorem 3 *Suppose that $f(x)$ is differentiable at $\bar{x}$. If there is a vector d such that $\nabla f(\bar{x})^t d < 0$, then for all $\lambda > 0$ and sufficiently small, $f(\bar{x} + \lambda d) < f(\bar{x})$, and hence d is a descent direction of $f(x)$ at $\bar{x}$.*

Proof: We have:

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x})^t d + \lambda \|d\| \alpha(\bar{x}, \lambda d),$$

where $\alpha(\bar{x}, \lambda d) \to 0$ as $\lambda \to 0$. Rearranging,

$$\frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda} = \nabla f(\bar{x})^t d + \|d\| \alpha(\bar{x}, \lambda d).$$

Since $\nabla f(\bar{x})^t d < 0$ and $\alpha(\bar{x}, \lambda d) \to 0$ as $\lambda \to 0$, $f(\bar{x} + \lambda d) - f(\bar{x}) < 0$ for all $\lambda > 0$ sufficiently small. ∎

Corollary 4 *Suppose $f(x)$ is differentiable at $\bar{x}$. If $\bar{x}$ is a local minimum, then $\nabla f(\bar{x}) = 0$.*

Proof: If it were true that $\nabla f(\bar{x}) \neq 0$, then $d = -\nabla f(\bar{x})$ would be a descent direction, whereby $\bar{x}$ would not be a local minimum. ∎

The above corollary is a *first order necessary optimality condition* for an unconstrained minimization problem. The following theorem is a *second order necessary optimality condition*

Theorem 5 *Suppose that $f(x)$ is twice continuously differentiable at $\bar{x} \in X$. If $\bar{x}$ is a local minimum, then $\nabla f(\bar{x}) = 0$ and $H(\bar{x})$ is positive semidefinite.*

Proof: From the first order necessary condition, $\nabla f(\bar{x}) = 0$. Suppose $H(\bar{x})$ is not positive semi-definite. Then there exists d such that $d^t H(\bar{x}) d < 0$. We have:

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x})^t d + \frac{1}{2}\lambda^2 d^t H(\bar{x}) d + \lambda^2 \|d\|^2 \alpha(\bar{x}, \lambda d)$$

$$= f(\bar{x}) + \frac{1}{2}\lambda^2 d^t H(\bar{x}) d + \lambda^2 \|d\|^2 \alpha(\bar{x}, \lambda d),$$

where $\alpha(\bar{x}, \lambda d) \to 0$ as $\lambda \to 0$. Rearranging,

$$\frac{f(\bar{x} + \lambda d) - f(\bar{x})}{\lambda^2} = \frac{1}{2} d^t H(\bar{x}) d + \|d\|^2 \alpha(\bar{x}, \lambda d).$$

Since $d^t H(\bar{x}) d < 0$ and $\alpha(\bar{x}, \lambda d) \to 0$ as $\lambda \to 0$, $f(\bar{x} + \lambda d) - f(\bar{x}) < 0$ for all $\lambda > 0$ sufficiently small, yielding the desired contradiction. ∎

Example 5 *Let*

$$f(x) = \frac{1}{2}x_1^2 + x_1 x_2 + 2x_2^2 - 4x_1 - 4x_2 - x_2^3 \ .$$

Then

$$\nabla f(x) = \left(x_1 + x_2 - 4, x_1 + 4x_2 - 4 - 3x_2^2\right)^T \ ,$$

and

$$H(x) = \begin{pmatrix} 1 & 1 \\ 1 & 4 - 6x_2 \end{pmatrix} \ .$$

$\nabla f(x) = 0$ *has exactly two solutions:* $\bar{x} = (4, 0)$ *and* $\tilde{x} = (3, 1)$. *But*

$$H(\tilde{x}) = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$$

is indefinite, therefore, the only possible candidate for a local minimum is $\bar{x} = (4, 0)$.

A *sufficient condition* for local optimality is a statement of the form: "if $\bar{x}$ satisfies ..., then $\bar{x}$ is a local minimum of (P)." Such a condition allows us to automatically declare that $\bar{x}$ is indeed a local minimum.

Theorem 6 *Suppose that $f(x)$ is twice differentiable at $\bar{x}$. If $\nabla f(\bar{x}) = 0$ and $H(\bar{x})$ is positive definite, then $\bar{x}$ is a (strict) local minimum.*

Proof:

$$f(x) = f(\bar{x}) + \frac{1}{2}(x - \bar{x})^t H(\bar{x})(x - \bar{x}) + \|x - \bar{x}\|^2 \alpha(\bar{x}, x - \bar{x}).$$

Suppose that $\bar{x}$ is not a strict local minimum. Then there exists a sequence $x_k \to \bar{x}$ such that $x_k \neq \bar{x}$ and $f(x_k) \leq f(\bar{x})$ for all k. Define $d_k = \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|}$. Then

$$f(x_k) = f(\bar{x}) + \|x_k - \bar{x}\|^2 \left(\frac{1}{2} d_k^t H(\bar{x}) d_k + \alpha(\bar{x}, x_k - \bar{x}) \right) ,$$

and so

$$\frac{1}{2} d_k^t H(\bar{x}) d_k + \alpha(\bar{x}, x_k - \bar{x}) = \frac{f(x_k) - f(\bar{x})}{\|x_k - \bar{x}\|^2} \leq 0 .$$

Since $\|d_k\| = 1$ for any k, there exists a subsequence of $\{d_k\}$ converging to some point d such that $\|d\| = 1$. Assume without loss of generality that $d_k \to d$. Then

$$0 \geq \lim_{k \to \infty} d_k^t H(\bar{x}) d_k + \alpha(\bar{x}, x_k - \bar{x}) = \frac{1}{2} d^t H(\bar{x}) d,$$

which is a contradiction of the positive definiteness of $H(\bar{x})$. ∎

Note:

- If $\nabla f(\bar{x}) = 0$ and $H(\bar{x})$ is negative definite, then $\bar{x}$ is a local maximum.
- If $\nabla f(\bar{x}) = 0$ and $H(\bar{x})$ is positive *semi*definite, we cannot be sure if $\bar{x}$ is a local minimum.

Example 6 *Continuing Example 5, we compute*

$$H(\bar{x}) = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}$$

is positive definite. To see this, note that for any $d = (d_1, d_2)$*, we have*

$$d^T H(\bar{x}) d = d_1^2 + 2d_1 d_2 + 4d_2^2 = (d_1 + d_2)^2 + 3d_2^2 > 0 \text{ for all } d \neq 0 \ .$$

Therefore, $\bar{x}$ *satisfies the sufficient conditions to be a local minimum, and so* $\bar{x}$ *is a local minimum.*

Example 7 *Let*

$$f(x) = x_1^3 + x_2^2 \ .$$

Then

$$\nabla f(x) = \left(3x_1^2, 2x_2\right)^T \ ,$$

and

$$H(x) = \begin{pmatrix} 6x_1 & 0 \\ 0 & 2 \end{pmatrix} \ .$$

At $\bar{x} = (0, 0)$*, we have* $\nabla f(\bar{x}) = 0$ *and*

$$H(\bar{x}) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

is positive semi-definite, but $\bar{x}$ *is not a local minimum, since* $f(-\epsilon, 0) = -\epsilon^3 < 0 = f(0, 0) = f(\bar{x})$ *for all* $\epsilon > 0$.

Example 8 *Let*

$$f(x) = x_1^4 + x_2^2 \ .$$

Then

$$\nabla f(x) = \left(4x_1^3, 2x_2\right)^T \ ,$$

and

$$H(x) = \begin{pmatrix} 12x_1^2 & 0 \\ 0 & 2 \end{pmatrix} \ .$$

At $\bar{x} = (0, 0)$*, we have* $\nabla f(\bar{x}) = 0$ *and*

$$H(\bar{x}) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

is positive semi-definite. Furthermore, $\bar{x}$ *is a local minimum, since for all* x *we have* $f(x) \geq 0 = f(0, 0) = f(\bar{x})$.

2.1 Convexity and Minimization

- Let $x, y \in \Re^n$. Points of the form $\lambda x + (1-\lambda)y$ for $\lambda \in [0,1]$ are called *convex combinations* of x and y .
- A set $S \subset \Re^n$ is called a *convex* set if for all $x, y \in S$ and for all $\lambda \in [0,1]$ it holds that $\lambda x + (1-\lambda)y \in S$.
- A function $f(x) : S \to \Re$, where S is a nonempty convex set, is a *convex function* if

  $$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y)$$

  for all $x, y \in S$ and for all $\lambda \in [0,1]$.
- A function $f(x)$ as above is called a *strictly convex* function if the inequality above is strict for all $x \ne y$ and $\lambda \in (0,1)$.
- A function $f(x) : S \to \Re$, where S is a nonempty convex set, is a *concave function* if

  $$f(\lambda x + (1-\lambda)y) \ge \lambda f(x) + (1-\lambda)f(y)$$

  for all $x, y \in S$ and for all $\lambda \in [0,1]$.
- A function $f(x)$ as above is called a *strictly concave* function if the inequality above is strict for all $x \ne y$ and $\lambda \in (0,1)$.

Consider the problem:

$$\begin{array}{lll} \text{(CP)} & \min_x & f(x) \\ & \text{s.t.} & x \in S\ . \end{array}$$

Theorem 7 *Suppose S is a convex set, $f(x) : S \to \Re$ is a convex function, and $\bar{x}$ is a local minimum of (CP). Then $\bar{x}$ is a global minimum of $f(x)$ over S.*

Proof: Suppose $\bar{x}$ is not a global minimum, i.e., there exists $y \in S$ for which $f(y) < f(\bar{x})$. Let $y(\lambda) := \lambda\bar{x} + (1-\lambda)y$, which is a convex combination of $\bar{x}$ and y for $\lambda \in [0,1]$ (and therefore, $y(\lambda) \in S$ for $\lambda \in [0,1]$). Note that $y(\lambda) \to \bar{x}$ as $\lambda \to 0$.

From the convexity of $f(x)$,

$$f(y(\lambda)) = f(\lambda\bar{x}+(1-\lambda)y) \leq \lambda f(\bar{x})+(1-\lambda)f(y) < \lambda f(\bar{x})+(1-\lambda)f(\bar{x}) = f(\bar{x})$$

for all $\lambda \in (0,1)$. Therefore, $f(y(\lambda)) < f(\bar{x})$ for all $\lambda \in (0,1)$, and so $\bar{x}$ is not a local minimum, resulting in a contradiction. ∎

Note:

- If $f(x)$ is strictly convex, a local minimum is the *unique* global minimum.
- If $f(x)$ is concave, a local maximum is a global maximum.
- If $f(x)$ is strictly concave, a local maximum is the unique global maximum.

Theorem 8 *Suppose S is a non-empty open convex set, and $f(x) : S \to \Re$ is differentiable. Then $f(x)$ is a convex function if and only if $f(x)$ satisfies the following* gradient inequality*:*

$$f(y) \geq f(x) + \nabla f(x)^t (y-x) \quad \text{for all } x, y \in S.$$

Proof: Suppose $f(x)$ is convex. Then for any $\lambda \in [0,1]$,

$$f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda)f(x)$$

which implies that

$$\frac{f(x+\lambda(y-x)) - f(x)}{\lambda} \leq f(y) - f(x) \ .$$

Letting $\lambda \to 0$, we obtain: $\nabla f(x)^t(y-x) \le f(y) - f(x)$, establishing the "only if" part.

Now, suppose that the gradient inequality holds for all $x, y \in S$. Let w and z be any two points in S. Let $\lambda \in [0,1]$, and set $x = \lambda w + (1-\lambda)z$. Then

$$f(w) \ge f(x) + \nabla f(x)^t(w-x) \text{ and } f(z) \ge f(x) + \nabla f(x)^t(z-x).$$

Taking a convex combination of the above inequalities, we obtain

$$\begin{aligned} \lambda f(w) + (1-\lambda)f(z) &\ge f(x) + \nabla f(x)^t(\lambda(w-x) + (1-\lambda)(z-x)) \\ &= f(x) + \nabla f(x)^t 0 \\ &= f(\lambda w + (1-\lambda)z) \ , \end{aligned}$$

which shows that $f(x)$ is convex. ∎

Theorem 9 *Suppose S is a non-empty open convex set, and $f(x) : S \to \Re$ is twice differentiable. Let $H(x)$ denote the Hessian of $f(x)$. Then $f(x)$ is convex if and only if $H(x)$ is positive semidefinite for all $x \in S$.*

Proof: Suppose $f(x)$ is convex. Let $\bar{x} \in S$ and d be any direction. Then for $\lambda > 0$ sufficiently small, $\bar{x} + \lambda d \in S$. We have:

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \nabla f(\bar{x})^t(\lambda d) + \frac{1}{2}(\lambda d)^t H(\bar{x})(\lambda d) + \|\lambda d\|^2 \alpha(\bar{x}, \lambda d),$$

where $\alpha(\bar{x}, y) \to 0$ as $y \to 0$. Using the gradient inequality, we obtain

$$\lambda^2 \left(\frac{1}{2} d^t H(\bar{x}) d + \|d\|^2 \alpha(\bar{x}, \lambda d) \right) \ge 0.$$

Dividing by $\lambda^2 > 0$ and letting $\lambda \to 0$, we obtain $d^t H(\bar{x}) d \ge 0$, proving the "only if" part.

Conversely, suppose that $H(z)$ is positive semidefinite for all $z \in S$. Let $x, y \in S$ be arbitrary. Invoking the second-order version of Taylor's theorem, we have:

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

for some z which is a convex combination of x and y (and hence $z \in S$). Since $H(z)$ is positive semidefinite, this means that

$$f(y) \geq f(x) + \nabla f(x)^t (y - x) \ .$$

Therefore the gradient inequality holds, and hence $f(x)$ is convex. ∎

Returning to the optimization problem (P), knowing that the function $f(x)$ is convex allows us to establish a *global* optimality condition that is both necessary and sufficient:

Theorem 10 *Suppose $f(x) : X \to \Re$ is convex and differentiable on X. Then $\bar{x} \in X$ is a global minimum if and only if $\nabla f(\bar{x}) = 0$.*

Proof: The necessity of the condition $\nabla f(\bar{x}) = 0$ was established in Corollary 4 regardless of the convexity of the function $f(x)$.

Suppose $\nabla f(\bar{x}) = 0$. Then by the gradient inequality we have $f(y) \geq f(\bar{x}) + \nabla f(\bar{x})^t (y - \bar{x}) = f(\bar{x})$ for all $y \in X$, and so $\bar{x}$ is a global minimum. ∎

Example 9 *Continuing Example 5, recall that*

$$H(x) = \begin{pmatrix} 1 & 1 \\ 1 & 4 - 6x_2 \end{pmatrix} \ .$$

Suppose that the domain of of $f(\cdot)$ is $X = \{(x_1, x_2) \mid x_2 < 0\}$. Then $f(\cdot)$ is a convex function on this domain.

Example 10 *Let*

$$f(x) = -\ln(1 - x_1 - x_2) - \ln x_1 - \ln x_2 \ .$$

Then

$$\nabla f(x) = \begin{bmatrix} \frac{1}{1-x_1-x_2} - \frac{1}{x_1} \\ \\ \frac{1}{1-x_1-x_2} - \frac{1}{x_2} \end{bmatrix},$$

and

$$H(x) = \begin{bmatrix} \left(\frac{1}{1-x_1-x_2}\right)^2 + \left(\frac{1}{x_1}\right)^2 & \left(\frac{1}{1-x_1-x_2}\right)^2 \\ \\ \left(\frac{1}{1-x_1-x_2}\right)^2 & \left(\frac{1}{1-x_1-x_2}\right)^2 + \left(\frac{1}{x_2}\right)^2 \end{bmatrix}.$$

It is actually easy to prove that $f(x)$ is a strictly convex function, and hence that $H(x)$ is positive definite on its domain $X = \{(x_1, x_2) \mid x_1 > 0, x_2 > 0, x_1 + x_2 < 1\}$. At $\bar{x} = \left(\frac{1}{3}, \frac{1}{3}\right)$ we have $\nabla f(\bar{x}) = 0$, and so $\bar{x}$ is the unique global minimum of $f(x)$.

3 Exercises on Unconstrained Optimization

1. Find points satisfying necessary conditions for extrema (i.e., local minima or local maxima) of the function

$$f(x) = \frac{x_1 + x_2}{3 + x_1^2 + x_2^2 + x_1 x_2}.$$

   Try to establish the nature of these points by checking sufficient conditions.

2. Find minima of the function

$$f(x) = (x_2^2 - x_1)^2$$

   among all the points satisfying necessary conditions for an extremum.

3. Consider the problem to minimize $\|Ax - b\|^2$, where A is an $m \times n$ matrix and b is an m vector.

   a. Give a geometric interpretation of the problem.

   b. Write a necessary condition for optimality . Is this also a sufficient condition?

c. Is the optimal solution unique? Why or why not?

d. Can you give a closed form solution of the optimal solution? Specify any assumptions that you may need.

e. Solve the problem for A and b given below:

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 2 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

4. Let S be a nonempty set in $\Re^n$. Show that S is convex if and only if for each integer $k \geq 2$ the following holds true:

$$x^1, \ldots, x^k \in S \;\Rightarrow\; \sum_{j=1}^{k} \lambda_j x^j \in S$$

whenever $\lambda_1, \ldots, \lambda_k$ satisfy $\lambda_1, \ldots, \lambda_k \geq 0$ and $\sum_{j=1}^{k} \lambda_j = 1$.

5. Bertsekas, Exercise 1.1.1, page 16. (Note: x^* is called a *stationary point* of $f(\cdot)$ if $\nabla f(x^*) = 0$.)

6. Bertsekas, Exercise 1.1.2, page 16, parts (a), (b), (c), and (d). (Note: x^* is called a *stationary point* of $f(\cdot)$ if $\nabla f(x^*) = 0$.)