Optimality Conditions for Constrained Optimization Problems

Robert M. Freund

February, 2004

1 Introduction

Recall that a constrained optimization problem is a problem of the form

$$\begin{array}{ll} \text{(P)} & \min_x f(x) \\ \text{s.t.} & g(x) \le 0 \\ & h(x) = 0 \\ & x \in X, \end{array}$$

where X is an open set and $g(x)) = (g_1(x), \ldots, g_m(x)) : \Re^n \to \Re^m$, $h(x) = (h_1(x), \ldots, h_l(x)) : \Re^n \to \Re^l$. Let S denote the *feasible region* of (P), i.e.,

$$S := \{x \in X : g(x) \le 0,\ h(x) = 0\}.$$

Then the problem (P) can be written as

$$\min_{x \in S} f(x) \ .$$

Recall that $\bar{x}$ is a local minimum of (P) if there exists $\epsilon > 0$ such that $f(\bar{x}) \le f(y)$ for all $y \in S \cap B(\bar{x}, \epsilon)$. Local, global minima and maxima, strict and non-strict, are defined analogously.

We will often use the following "shorthand" notation:

$$\nabla g(x) = \begin{bmatrix} \nabla g_1(x)^t \\ \vdots \\ \nabla g_m(x)^t \end{bmatrix} \text{ and } \nabla h(x) = \begin{bmatrix} \nabla h_1(x)^t \\ \vdots \\ \nabla h_l(x)^t \end{bmatrix},$$

i.e., $\nabla g(x) \in \Re^{m \times n}$ and $\nabla h(x) \in \Re^{l \times n}$ are Jacobian matrices, whose $i^{\mathrm{t}h}$ row is the transpose of the corresponding gradient.

2 Necessary Optimality Conditions

2.1 Geometric Necessary Conditions

A set $C \subseteq \Re^n$ is a *cone* if for every $x \in C$, $\alpha x \in C$ for any $\alpha > 0$.

A set C is a *convex cone* if C is a cone and C is a convex set.

Suppose $\bar{x} \in S$. We have the following definitions:

- $F_0 := \{d : \nabla f(\bar{x})^t d < 0\}$ is the cone of "improving" directions of $f(x)$ at $\bar{x}$.
- $I = \{i : g_i(\bar{x}) = 0\}$ is the set of indices of the binding inequality constraints at $\bar{x}$.
- $G_0 = \{d : \nabla g_i(\bar{x})^t d < 0 \text{ for all } i \in I\}$ is the cone of "inward" pointing directions for the binding constrains at $\bar{x}$.
- $H_0 = \{d : \nabla h_i(\bar{x})^t d = 0 \text{ for all } i = 1, \ldots, l\}$ is the set of tangent directions for the equality constraints at $\bar{x}$.

Theorem 1 *Assume that $h(x)$ is a linear function, i.e., $h(x) = Ax - b$ for $A \in \Re^{l \times n}$. If $\bar{x}$ is a local minimum of (P), then $F_0 \cap G_0 \cap H_0 = \emptyset$.*

Proof: Note that $\nabla h_i(\bar{x}) = A_{i\cdot}$, i.e., $H_0 = \{d : Ad = 0\}$. Suppose $d \in F_0 \cap G_0 \cap H_0$. Then for all $\lambda > 0$ sufficiently small $g_i(\bar{x} + \lambda d) \le g_i(\bar{x}) = 0$ for all $i \in I$ (for $i \notin I$, since λ is small, $g_i(\bar{x} + \lambda d) < 0$), and $h(\bar{x} + \lambda d) = (A\bar{x} - b) + \lambda Ad = 0$. Therefore $\bar{x} + \lambda d \in S$ for all $\lambda > 0$ sufficiently small. On the other hand, for all sufficiently small $\lambda > 0$, $f(\bar{x} + \lambda d) < f(\bar{x})$. This contradicts the assumption that $\bar{x}$ is a local minimum of (P). ∎

The following is the extension of Theorem 1 to handle general nonlinear functions $h_i(x), i = 1, \ldots, l$.

Theorem 2 *If $\bar{x}$ is a local minimum of (P) and the gradient vectors $\nabla h_i(\bar{x})$, $i = 1, \ldots, l$ are linearly independent, then $F_0 \cap G_0 \cap H_0 = \emptyset$.* ∎

Note that Theorem 2 is essentially saying that if a point $\bar{x}$ is (locally) optimal, there is no direction d which is an *improving* direction (i.e., such that $f(\bar{x}+\lambda d) < f(\bar{x})$ for small $\lambda > 0$), and at the same time is also a *feasible direction* (i.e., such that $g_i(\bar{x}+\lambda d) \leq g_i(\bar{x}) = 0$ for $i \in I$ and $h(\bar{x}+\lambda d) \approx 0$), which makes sense intuitively. Observe, however, that the condition in Theorem 2 is somewhat weaker than the above intuitive explanation: indeed, we can have a direction d which is an improving direction but $\nabla f(\bar{x})^t d = 0$ and/or $\nabla g(\bar{x})^t d = 0$.

The proof of Theorem 2 is rather awkward and involved, and relies on the Implicit Function Theorem. We present this proof at the end of this note, in Section 6.

2.2 Separation of Convex Sets

We will shortly attempt to restate the geometric necessary local optimality conditions ($F_0 \cap G_0 \cap H_0 = \emptyset$) into a constructive and "computable" algebraic statement about the gradients of the objective function and the constraints functions. The vehicle that will make this happen involves the separation theory of convex sets.

- If $p \neq 0$ is a vector in $\Re^n$ and α is a scalar, $H := \{x \in \Re^n : p^t x = \alpha\}$ is a *hyperplane*, and $H^+ = \{x \in \Re^n : p^t x \geq \alpha\}$, $H^- = \{x \in \Re^n : p^t x \leq \alpha\}$ are *halfspaces*.

- Let S and T be two non-empty sets in $\Re^n$. A hyperplane $H = \{x : p^t x = \alpha\}$ is said to *separate* S and T if $p^t x \geq \alpha$ for all $x \in S$ and $p^t x \leq \alpha$ for all $x \in T$, i.e., if $S \subseteq H^+$ and $T \subseteq H^-$. If, in addition, $S \cup T \not\subset H$, then H is said to *properly* separate S and T.

- H is said to *strictly* separate S and T if $p^t x > \alpha$ for all $x \in S$ and $p^t x < \alpha$ for all $x \in T$.

- H is said to *strongly* separate S and T if for some $\epsilon > 0$, $p^t x \geq \alpha + \epsilon$ for all $x \in S$ and $p^t x \leq \alpha - \epsilon$ for all $x \in T$.

Theorem 3 *Let S be a nonempty closed convex set in $\Re^n$, and suppose that $y \notin S$. Then there exists $p \neq 0$ and α such that $H = \{x : p^t x = \alpha\}$ strongly separates S and $\{y\}$.*

To prove the theorem, we need the following result:

Theorem 4 *Let S be a nonempty closed convex set in $\Re^n$, and $y \notin S$. Then there exists a unique point $\bar{x} \in S$ with minimum distance from y. Furthermore, $\bar{x}$ is the minimizing point if and only if $(y-\bar{x})^t(x-\bar{x}) \leq 0$ for all $x \in S$.*

Proof: Let $\hat{x}$ be an arbitrary point in S, and let $\bar{S} = S \cap \{x : \|x-y\| \leq \|\hat{x}-y\|\}$. Then $\bar{S}$ is a compact set. Let $f(x) = \|x-y\|$. Then $f(x)$ attains its minimum over the set $\bar{S}$ at some point $\bar{x} \in \bar{S}$. Note that $\bar{x} \neq y$.

To show uniqueness, suppose that there is some $x' \in S$ for which $\|y-\bar{x}\| = \|y-x'\|$. By convexity of S, $\frac{1}{2}(\bar{x}+x') \in S$. But by the triangle inequality, we have:

$$\left\| y - \frac{1}{2}(\bar{x}+x') \right\| \leq \frac{1}{2}\|y-\bar{x}\| + \frac{1}{2}\|y-x'\|.$$

If strict inequality holds, we have a contradiction. Therefore equality holds, and we must have $y - \bar{x} = \lambda(y-x')$ for some λ. Since $\|y-\bar{x}\| = \|y-x'\|$, $|\lambda| = 1$. If $\lambda = -1$, then $y = \frac{1}{2}(\bar{x}+x') \in S$, contradicting the assumption. Hence $\lambda = 1$, whereby $x' = \bar{x}$.

Finally we need to establish that $\bar{x}$ is the minimizing point if and only if $(y-\bar{x})^t(x-\bar{x}) \leq 0$ for all $x \in S$. To establish sufficiency, note that for any $x \in S$,

$$\|x-y\|^2 = \|(x-\bar{x})-(y-\bar{x})\|^2 = \|x-\bar{x}\|^2+\|y-\bar{x}\|^2-2(x-\bar{x})^t(y-\bar{x}) \geq \|\bar{x}-y\|^2.$$

Conversely, assume that $\bar{x}$ is the minimizing point. For any $x \in S$, $\lambda x +$

$(1-\lambda)\bar{x} \in S$ for any $\lambda \in [0,1]$. Also, $\|\lambda x + (1-\lambda)\bar{x} - y\| \geq \|\bar{x} - y\|$. Thus,

$$\begin{aligned} \|\bar{x} - y\|^2 &\leq \|\lambda x + (1-\lambda)\bar{x} - y\|^2 \\ &= \|\lambda(x - \bar{x}) + (\bar{x} - y)\|^2 \\ &= \lambda^2\|x - \bar{x}\|^2 + 2\lambda(x - \bar{x})^t(\bar{x} - y) + \|\bar{x} - y\|^2 \ , \end{aligned}$$

which when rearranged yields:

$$\lambda^2\|x - \bar{x}\|^2 \geq 2\lambda(y - \bar{x})^t(x - \bar{x}) \ .$$

This implies that $(y-\bar{x})^t(x-\bar{x}) \leq 0$ for any $x \in S$, since otherwise the above expression can be invalidated by choosing $\lambda > 0$ and sufficiently small. ∎

Proof of Theorem 3: Let $\bar{x} \in S$ be the point minimizing the distance from the point y to the set S. Note that $\bar{x} \neq y$. Let $p = y-\bar{x}$, $\alpha = \frac{1}{2}(y-\bar{x})^t(y+\bar{x})$, and $\epsilon = \frac{1}{2}\|y - \bar{x}\|^2$. Then for any $x \in S$, $(x - \bar{x})^t(y - \bar{x}) \leq 0$, and so

$$p^t x = (y-\bar{x})^t x \leq \bar{x}^t(y-\bar{x}) = \bar{x}^t(y-\bar{x}) + \frac{1}{2}\|y-\bar{x}\|^2 - \epsilon = \frac{1}{2}y^t y - \frac{1}{2}\bar{x}^t\bar{x} - \epsilon = \alpha - \epsilon.$$

Therefore $p^t x \leq \alpha - \epsilon$ for all $x \in S$. On the other hand, $p^t y = (y - \bar{x})^t y = \alpha + \epsilon$, establishing the result. ∎

Corollary 5 *If S is a closed convex set in $\Re^n$, then S is the intersection of all halfspaces that contain it.*

Theorem 6 *Let $S \in \Re^n$ and let C be the intersection of all halfspaces containing S. Then C is the smallest closed convex set containing S.*

Theorem 7 *Suppose S_1 and S_2 are disjoint nonempty closed convex sets and S_1 is bounded. Then S_1 and S_2 can be strongly separated by a hyperplane.*

Proof: Let $T = \{x \in \Re^n : x = y - z, \text{ where } y \in S_1, \ z \in S_2\}$. Then it is easy to show that T is a convex set. We also claim that T is a closed set.

To see this, let $\{x_i\}_{i=1}^{\infty} \subset T$, and suppose $\bar{x} = \lim_{i\to\infty} x_i$. Then $x_i = y_i - z_i$ for $\{y_i\}_{i=1}^{\infty} \subset S_1$ and $\{z_i\}_{i=1}^{\infty} \subset S_2$. By the Weierstrass Theorem, some subsequence of $\{y_i\}$ converges to a point $\bar{y} \in S_1$. Then $z_i = y_i - x_i \to \bar{y} - \bar{x}$ (over this subsequence), so that $\bar{z} = \bar{y} - \bar{x}$ is a limit point of $\{z_i\}$. Since S_2 is also closed, $\bar{z} \in S_2$, and then $\bar{x} = \bar{y} - \bar{x} \in T$, proving that T is a closed set.

By hypothesis, $S_1 \cap S_2 = \emptyset$, so $0 \notin T$. Since T is convex and closed, there exists a hyperplane $H = \{x : p^t x = \bar{\alpha}\}$ such that $p^t x > \bar{\alpha}$ for $x \in T$ and $p^t 0 < \bar{\alpha}$ (and hence $\bar{\alpha} > 0$).

Let $y \in S_1$ and $z \in S_2$. Then $x = y - z \in T$, and so $p^t(y - z) > \bar{\alpha} > 0$ for any $y \in S_1$ and $z \in S_2$.

Let $\alpha_1 = \inf\{p^t y : y \in S_1\}$ and $\alpha_2 = \sup\{p^t z : z \in S_2\}$ (note that $0 < \bar{\alpha} \le \alpha_1 - \alpha_2$); define $\alpha = \frac{1}{2}(\alpha_1 + \alpha_2)$ and $\epsilon = \frac{1}{2}\bar{\alpha} > 0$. Then for all $y \in S_1$ and $z \in S_2$ we have

$$p^t y \ge \alpha_1 = \frac{1}{2}(\alpha_1 + \alpha_2) + \frac{1}{2}(\alpha_1 - \alpha_2) \ge \alpha + \frac{1}{2}\bar{\alpha} = \alpha + \epsilon$$

and

$$p^t z \le \alpha_2 = \frac{1}{2}(\alpha_1 + \alpha_2) - \frac{1}{2}(\alpha_1 - \alpha_2) \le \alpha - \frac{1}{2}\bar{\alpha} = \alpha - \epsilon. \quad \blacksquare$$

Theorem 8 (Farkas' Lemma) *Given an $m \times n$ matrix A and an n-vector c, exactly one of the following two systems has a solution:*

(i) $Ax \le 0,\ c^t x > 0$

(ii) $A^t y = c,\ y \ge 0.$

Proof: First note that both systems cannot have a solution, since then we would have $0 < c^t x = y^t A x \le 0$.

Suppose the system (ii) has no solution. Let $S = \{x : x = A^t y \text{ for some } y \ge 0\}$. Then $c \notin S$. S is easily seen to be a convex set. Also, S is a closed set. (For an exact proof of this, see Appendix B.3 of *Nonlinear Programming* by Dimitri Bertsekas, Athena Scientific, 1999.) Therefore there exist p and α such that $c^t p > \alpha$ and $p^t(A^t y) = (Ap)^t y \le \alpha$ for all $y \ge 0$.

If $(Ap)_i > 0$ for some i, one could set y_i sufficiently large so that $(Ap)^t y > \alpha$, a contradiction. Thus $Ap \leq 0$. Taking $y = 0$, we also have that $\alpha \geq 0$, and so $c^t p > 0$, and p is a solution of (i). ∎

Lemma 9 (Key Lemma) *Given matrices A, B, and H of appropriate dimensions, exactly one of the two following systems has a solution:*

(i) $\bar{A}x < 0,\ Bx \leq 0,\ Hx = 0$

(ii) $\bar{A}^t u + B^t v + H^t w = 0,\ u \geq 0,\ v \geq 0,\ e^t u = 1.$

Proof: It is easy to show that both (i) and (ii) cannot have a solution. Suppose (i) does not have a solution. Then the system

$$\begin{aligned} \bar{A}x + e\theta &\leq 0, \quad \theta > 0 \\ Bx &\leq 0 \\ Hx &\leq 0 \\ -Hx &\leq 0 \end{aligned}$$

has no solution. This system can be re-written in the form

$$\begin{bmatrix} \bar{A} & e \\ B & 0 \\ H & 0 \\ -H & 0 \end{bmatrix} \cdot \begin{pmatrix} x \\ \theta \end{pmatrix} \leq 0,\ (0, \ldots, 0, 1) \cdot \begin{pmatrix} x \\ \theta \end{pmatrix} > 0.$$

From Farkas' Lemma, there exists a vector $(u; v; w^1; w^2) \geq 0$ such that

$$\begin{bmatrix} \bar{A} & e \\ B & 0 \\ H & 0 \\ -H & 0 \end{bmatrix}^t \cdot \begin{pmatrix} u \\ v \\ w^1 \\ w^2 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

This can be rewritten as

$$\bar{A}^t u + B^t v + H^t(w^1 - w^2) = 0, \qquad e^t u = 1\ .$$

Letting $w = w^1 - w^2$ completes the proof of the lemma. ∎

2.3 Algebraic Necessary Conditions

Theorem 10 (Fritz John Necessary Conditions) *Let $\bar{x}$ be a feasible solution of (P). If $\bar{x}$ is a local minimum of (P), then there exists (u_0, u, v) such that*

$$u_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{x}) = 0,$$

$$u_0, u \geq 0, \ (u_0, u, v) \neq 0,$$

$$u_i g_i(\bar{x}) = 0, \ i = 1, \ldots, m.$$

(Note that the first equation can be rewritten as

$$u_0 \nabla f(\bar{x}) + \nabla g(\bar{x})^t u + \nabla h(\bar{x})^t v = 0 \text{ .)}$$

Proof: If the vectors $\nabla h_i(\bar{x})$ are linearly dependent, then there exists $v \neq 0$ such that $\nabla h(\bar{x})^t v = 0$. Setting $(u_0, u) = 0$ establishes the result.

Suppose now that the vectors $\nabla h_i(\bar{x})$ are linearly independent. Then we can apply Theorem 2 and assert that $F_0 \cap G_0 \cap H_0 = \emptyset$. Assume for simplicity that $I = \{1, \ldots, p\}$. Let

$$A = \begin{bmatrix} \nabla f(\bar{x})^t \\ \nabla g_1(\bar{x})^t \\ \vdots \\ \nabla g_p(\bar{x})^t \end{bmatrix}, \ H = \begin{bmatrix} \nabla h_1(\bar{x})^t \\ \vdots \\ \nabla h_l(\bar{x})^t \end{bmatrix}.$$

Then there is no d that satisfies $Ad < 0$, $Hd = 0$. From the Key Lemma there exists $(u_0, u_1, \ldots, u_p)$ and $(v_1, \ldots, v_l)$ such that

$$u_0 \nabla f(\bar{x}) + \sum_{i=1}^{p} u_i \nabla g_i(\bar{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\bar{x}) = 0,$$

with $u_0+u_1+\cdots+u_p=1$ and $(u_0,u_1,\ldots,u_p)\geq 0$. Define $u_{p+1},\ldots,u_m=0$. Then $(u_0,u)\geq 0$, $(u_0,u)\neq 0$, and for any i, either $g_i(\bar{x})=0$, or $u_i=0$. Finally,

$$u_0\nabla f(\bar{x})+\nabla g(\bar{x})^t u+\nabla h(\bar{x})^t v=0. \quad \blacksquare$$

Theorem 11 (Karush-Kuhn-Tucker (KKT) Necessary Conditions) *Let $\bar{x}$ be a feasible solution of (P) and let $I=\{i : g_i(\bar{x})=0\}$. Further, suppose that $\nabla h_i(\bar{x})$ for $i=1,\ldots,l$ and $\nabla g_i(\bar{x})$ for $i\in I$ are linearly independent. If $\bar{x}$ is a local minimum, there exists (u,v) such that*

$$\nabla f(\bar{x})+\nabla g(\bar{x})^t u+\nabla h(\bar{x})^t v=0,$$

$$u\geq 0,$$

$$u_i g_i(\bar{x})=0\ ,\ i=1,\ldots,m.$$

Proof: $\bar{x}$ must satisfy the Fritz John conditions. If $u_0>0$, we can redefine $u\leftarrow u/u_0$ and $v\leftarrow v/u_0$. If $u_0=0$, then

$$\sum_{i\in I} u_i\nabla g_i(\bar{x})+\sum_{i=1}^{l} v_i\nabla h_i(\bar{x})=0\ ,$$

and so the above gradients are linearly dependent. This contradicts the assumptions of the theorem. $\blacksquare$

Example 1 *Consider the problem:*

$$\begin{array}{llll}
\min & 6(x_1-10)^2 & +4(x_2-12.5)^2 & \\
s.t. & x_1^2 & +(x_2-5)^2 & \leq 50 \\
 & x_1^2 & +3x_2^2 & \leq 200 \\
 & (x_1-6)^2 & +x_2^2 & \leq 37
\end{array}$$

In this problem, we have:

$$f(x) = 6(x_1 - 10)^2 + 4(x_2 - 12.5)^2$$

$$g_1(x) = x_1^2 + (x_2 - 5)^2 - 50$$

$$g_2(x) = x_1^2 + 3x_2^2 - 200$$

$$g_3(x) = (x_1 - 6)^2 + x_2^2 - 37$$

We also have:

$$\nabla f(x) = \begin{pmatrix} 12(x_1 - 10) \\ \\ 8(x_2 - 12.5) \end{pmatrix}$$

$$\nabla g_1(x) = \begin{pmatrix} 2x_1 \\ \\ 2(x_2 - 5) \end{pmatrix}$$

$$\nabla g_2(x) = \begin{pmatrix} 2x_1 \\ \\ 6x_2 \end{pmatrix}$$

$$\nabla g_3(x) = \begin{pmatrix} 2(x_1 - 6) \\ \\ 2x_2 \end{pmatrix}$$

Let us determine whether or not the point $\bar{x} = (\bar{x}_1, \bar{x}_2) = (7, 6)$ *is a candidate to be an optimal solution to this problem.*

We first check for feasibility:

$$g_1(\bar{x}) = 0 \leq 0$$

$$g_2(\bar{x}) = -43 < 0$$

$$g_3(\bar{x}) = 0 \leq 0$$

To check for optimality, we compute all gradients at $\bar{x}$*:*

$$\nabla f(x) = \begin{pmatrix} -36 \\ \\ -52 \end{pmatrix}$$

$$\nabla g_1(x) = \begin{pmatrix} 14 \\ \\ 2 \end{pmatrix}$$

$$\nabla g_2(x) = \begin{pmatrix} 14 \\ \\ 36 \end{pmatrix}$$

$$\nabla g_3(x) = \begin{pmatrix} 2 \\ \\ 12 \end{pmatrix}$$

We next check to see if the gradients "line up", by trying to solve for $u_1 \geq 0, u_2 = 0, u_3 \geq 0$ *in the following system:*

$$\begin{pmatrix} -36 \\ \\ -52 \end{pmatrix} + \begin{pmatrix} 14 \\ \\ 2 \end{pmatrix} u_1 + \begin{pmatrix} 14 \\ \\ 36 \end{pmatrix} u_2 + \begin{pmatrix} 2 \\ \\ 12 \end{pmatrix} u_3 = \begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}$$

Notice that $\bar{u} = (\bar{u}_1, \bar{u}_2, \bar{u}_3) = (2, 0, 4)$ *solves this system, and that* $\bar{u} \geq 0$ *and* $\bar{u}_2 = 0$*. Therefore* $\bar{x}$ *is a candidate to be an optimal solution of this problem.*

Example 2 *Consider the problem (P):*

$$
\begin{array}{llll}
(P): & \max_x & x^T Qx & \\
s.t. & & \|x\| & \leq 1
\end{array}
$$

where Q is symmetric. This is equivalent to:

$$
\begin{array}{llll}
(P): & \min_x & -x^T Qx & \\
s.t. & & x^T x & \leq 1 \ .
\end{array}
$$

The KKT conditions are:

$$
\begin{array}{rcl}
-2Qx + 2ux & = & 0 \\
x^T x & \leq & 1 \\
u & \geq & 0 \\
u(1 - x^T x) & = & 0 \ .
\end{array}
$$

One solution to the KKT system is $x = 0, u = 0$, with objective function value $x^T Qx = 0$. Are there any better solutions to the KKT system?

If $x \neq 0$ is a solution of the KKT system together with some value u , then x is an eigenvector of Q with nonnegative eigenvalue u. Also, $x^T Qx = ux^T x = u$, and so the objective value of this solution is u. Therefore the solution of (P) with the largest objective function value is $x = 0$ if the largest eigenvalue of Q is nonpositive. If the largest eigenvalue of Q is positive, then the optimal objective value of (P) is the largest eigenvalue, and the optimal solution is any eigenvector x corresponding to this eigenvalue, normalized so that $\|x\| = 1$.

Example 3 *Consider the problem:*

$$
\begin{array}{llll}
\min & (x_1-12)^2 & +(x_2+6)^2 & \\
s.t. & x_1^2+3x_1 & +x_2^2-4.5x_2 & \leq 6.5 \\
 & (x_1-9)^2 & +x_2^2 & \leq 64 \\
 & 8x_1 & +4x_2 & =20
\end{array}
$$

In this problem, we have:

$$f(x) = (x_1-12)^2 + (x_2+6)^2$$

$$g_1(x) = x_1^2 + 3x_1 + x_2^2 - 4.5x_2 - 6.5$$

$$g_2(x) = (x_1-9)^2 + x_2^2 - 64$$

$$h_1(x) = 8x_1 + 4x_2 - 20$$

Let us determine whether or not the point $\bar{x} = (\bar{x}_1, \bar{x}_2) = (2,1)$ *is a candidate to be an optimal solution to this problem.*

We first check for feasibility:

$$g_1(\bar{x}) = 0 \leq 0$$

$$g_2(\bar{x}) = -14 < 0$$

$$h_1(\bar{x}) = 0$$

To check for optimality, we compute all gradients at $\bar{x}$*:*

$$\nabla f(x) = \begin{pmatrix} -20 \\ \\ 14 \end{pmatrix}$$

$$\nabla g_1(x) = \begin{pmatrix} 7 \\ \\ -2.5 \end{pmatrix}$$

$$\nabla g_2(x) = \begin{pmatrix} -14 \\ \\ 2 \end{pmatrix}$$

$$\nabla h_1(x) = \begin{pmatrix} 8 \\ \\ 4 \end{pmatrix}$$

We next check to see if the gradients "line up", by trying to solve for $u_1 \geq 0, u_2 = 0, v_1$ *in the following system:*

$$\begin{pmatrix} -20 \\ \\ 14 \end{pmatrix} + \begin{pmatrix} 7 \\ \\ -2.5 \end{pmatrix} u_1 + \begin{pmatrix} -14 \\ \\ 2 \end{pmatrix} u_2 + \begin{pmatrix} 8 \\ \\ 4 \end{pmatrix} v_1 = \begin{pmatrix} 0 \\ \\ 0 \end{pmatrix}$$

Notice that $(\bar{u}, \bar{v}) = (\bar{u}_1, \bar{u}_2, \bar{v}_1) = (4, 0, -1)$ *solves this system and that* $\bar{u} \geq 0$ *and* $\bar{u}_2 = 0$. *Therefore* $\bar{x}$ *is a candidate to be an optimal solution of this problem.*

3 Generalizations of Convexity

Suppose X is a convex set in $\Re^n$. The function $f(x) : X \to \Re$ is a *quasiconvex* function if for all $x, y \in X$ and for all $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}.$$

$f(x)$ is *quasiconcave* if for all $x, y \in X$ and for all $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \geq \min\{f(x), f(y)\}.$$

If $f(x) : X \to \Re$, then the *level sets* of $f(x)$ are the sets

$$S_\alpha = \{x \in X : f(x) \leq \alpha\}$$

for each $\alpha \in \Re$.

Proposition 12 *If $f(x)$ is convex, then $f(x)$ is quasiconvex.*

Proof: If $f(x)$ is convex, for $\lambda \in [0,1]$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \leq \max\{f(x), f(y)\}. \blacksquare$$

Theorem 13 *A function $f(x)$ is quasiconvex on X if and only if S_α is a convex set for every $\alpha \in \Re$.*

Proof: Suppose that $f(x)$ is quasiconvex. For any given value of α, suppose that $x, y \in S_\alpha$.

Let $z = \lambda x + (1-\lambda)y$ for some $\lambda \in [0,1]$. Then $f(z) \leq \max\{f(x), f(y)\} \leq \alpha$, so $z \in S_\alpha$, which shows that S_α is a convex set.

Conversely, suppose S_α is a convex set for every α. Let x and y be given, and let $\alpha = \max\{f(x), f(y)\}$, and hence $x, y \in S_\alpha$. Then for any $\lambda \in [0,1]$, $f(\lambda x + (1-\lambda)y) \leq \alpha = \max\{f(x),\ f(y)\}$, and so $f(x)$ is a quasiconvex function. $\blacksquare$

Corollary 14 *If $f(x)$ is a convex function, its level sets are convex sets.* $\blacksquare$

Suppose X is a convex set in $\Re^n$. The differentiable function $f(x) : X \to \Re$ is a *pseudoconvex* function if for every $x, y \in X$ the following holds:

$$\nabla f(x)^t (y - x) \geq 0 \ \Rightarrow \ f(y) \geq f(x)\ .$$

Theorem 15

(i) A differentiable convex function is pseudoconvex.

(ii) A pseudoconvex function is quasiconvex.

Proof: To prove the first claim, we use the gradient inequality: if $f(x)$ is convex and differentiable, then $f(y) \geq f(x) + \nabla f(x)^t(y - x)$. Hence, if $\nabla f(x)^t(y - x) \geq 0$, then $f(y) \geq f(x)$, and so $f(x)$ is pseudoconvex.

To show the second claim, suppose $f(x)$ is pseudoconvex. Let x, y and $\lambda \in [0, 1]$ be given, and let $z = \lambda x + (1 - \lambda)y$. If $\lambda = 0$ or $\lambda = 1$, then $f(z) \leq \max\{f(x), f(y)\}$ trivially; therefore, assume $0 < \lambda < 1$. Let $d = y - x$.

If $\nabla f(z)^t d \geq 0$, then

$$\nabla f(z)^t(y - z) = \nabla f(z)^t(\lambda(y - x)) = \lambda \nabla f(z)^t d \geq 0,$$

so $f(z) \leq f(y) \leq \max\{f(x), f(y)\}$.

On the other hand, if $\nabla f(z)^t d \leq 0$, then

$$\nabla f(z)^t(x - z) = \nabla f(z)^t(-(1 - \lambda)(y - x)) = -(1 - \lambda)\nabla f(z)^t d \geq 0,$$

so $f(z) \leq f(x) \leq \max\{f(x), f(y)\}$. Thus $f(x)$ is quasiconvex. ∎

Incidentally, we define a differentiable function $f(x) : X \to \Re$ to be *pseudoconcave* if for every $x, y \in X$ the following holds:

$$\nabla f(x)^t(y - x) \leq 0 \;\Rightarrow\; f(y) \leq f(x)\ .$$

4 Sufficient Conditions for Optimality

Theorem 16 (KKT Sufficient Conditions) *Let $\bar{x}$ be a feasible solution of (P), and suppose $\bar{x}$ together with multipliers (u, v) satisfies*

$$\nabla f(\bar{x}) + \nabla g(\bar{x})^t u + \nabla h(\bar{x})^t v = 0,$$

$$u \geq 0,$$

$$u_i g_i(\bar{x}) = 0, \;\; i = 1, \ldots, m.$$

If $f(x)$ is a pseudoconvex function, $g_i(x), i = 1, \ldots, m$ are quasiconvex functions, and $h_i(x), i = 1, \ldots, l$ are linear functions, then $\bar{x}$ is a global optimal solution of (P).

Proof: Because each $g_i(x)$ is quasiconvex, the level sets

$$C_i := \{x \in X : g_i(x) \leq 0\}, \;\; i = 1, \ldots, m$$

are convex sets. Also, because each $h_i(x)$ is linear, the sets

$$D_i = \{x \in X : h_i(x) = 0\}, \;\; i = 1, \ldots, l$$

are convex sets. Thus, since the intersection of convex sets is also a convex set, the feasible region

$$S = \{x \in X : g(x) \leq 0, \;\; h(x) = 0\}$$

is a convex set.

Let $I = \{i \mid g_i(\bar{x}) = 0\}$ denote the index of active constraints at $\bar{x}$. Let $x \in S$ be any point different from $\bar{x}$. Then $\lambda x + (1 - \lambda)\bar{x}$ is feasible for all $\lambda \in (0, 1)$. Thus for $i \in I$ we have

$$g_i(\lambda x + (1 - \lambda)\bar{x}) = g_i(\bar{x} + \lambda(x - \bar{x})) \leq 0 = g_i(\bar{x})$$

for any $\lambda \in (0, 1)$, and since the value of $g_i(\cdot)$ does not increase by moving in the direction $x - \bar{x}$, we must have $\nabla g_i(\bar{x})^t (x - \bar{x}) \leq 0$ for all $i \in I$.

Similarly, $\nabla h_i(\bar{x} + \lambda(x - \bar{x})) = 0$, and so $\nabla h_i(\bar{x})^t (x - \bar{x}) = 0$ for all $i = 1, \ldots, l$.

Thus, from the KKT conditions,

$$\nabla f(\bar{x})^t (x - \bar{x}) = -\left(\nabla g(\bar{x})^t u + \nabla h(\bar{x})^t v\right)^t (x - \bar{x}) \geq 0,$$

and by pseudoconvexity, $f(x) \geq f(\bar{x})$ for any feasible x. ∎

The program

$$
\begin{array}{ll}
\text{(P)} & \min_x f(x) \\
\text{s.t.} & g(x) \leq 0 \\
& h(x) = 0 \\
& x \in X
\end{array}
$$

is called a *convex program* if $f(x)$, $g_i(x)$, $i = 1, \ldots, m$ are convex functions, $h_i(x)$, $i = 1 \ldots, l$ are linear functions, and X is an open convex set.

Corollary 17 *The KKT conditions are sufficient for optimality of a convex program.*

Example 4 *Continuing Example 1, note that $f(x), g_1(x), g_2(x)$, and $g_3(x)$ are all convex functions. Therefore the problem is a convex optimization problem, and the KKT conditions are necessary and sufficient. Therefore $\bar{x} = (7, 6)$ is the global minimum.*

Example 5 *Continuing Example 3, note that $f(x), g_1(x), g_2(x)$ are all convex functions and that $h_1(x)$ is a linear function. Therefore the problem is a convex optimization problem, and the KKT conditions are necessary and sufficient. Therefore $\bar{x} = (2, 1)$ is the global minimum.*

5 Constraint Qualifications

Recall that the statement of the KKT necessary conditions established herein has the form "if $\bar{x}$ is a local minimum of (P) and (*some requirement for the constraints*) then the KKT conditions must hold at $\bar{x}$." This additional requirement for the constraints that enables us to proceed with the proof of the KKT conditions is called a *constraint qualification*.

In (Theorem 11) we established the following constraint qualification:

Linear Independence Constraint Qualification: The gradients $\nabla g_i(\bar{x}),\ i \in I,\ \nabla h_i(\bar{x}),\ i = 1, \ldots, l$ are linearly independent.

We will now establish two other useful constraint qualifications. Before doing so we have the following important definition:

Definition 5.1 *A point x is called a* Slater point *if x satisfies $g(x) < 0$ and $h(x) = 0$, that is, x is feasible and satisfies all inequalities strictly.*

Theorem 18 (Slater condition) *Suppose that $g_i(x),\ i = 1, \ldots, m$ are pseudoconvex, $h_i(x), i = 1, \ldots, l$ are linear, and $\nabla h_i(x),\ i = 1, \ldots, l$ are linearly independent, and (P) has a Slater point. Then the KKT conditions are necessary to characterize an optimal solution.*

Proof: Let $\bar{x}$ be a local minimum. The Fritz-John conditions are necessary for this problem, whereby there must exist $(u_0, u, v) \neq 0$ such that $(u_0, u) \geq 0$ and

$$u_0 \nabla f(\bar{x}) + \nabla g(\bar{x})^t u + \nabla h(\bar{x})^t v = 0,\ u_i g_i(\bar{x}) = 0.$$

If $u_0 > 0$, dividing through by u_0 demonstrates KKT conditions. Now suppose $u_0 = 0$. Let x^0 be Slater point, and define $d := x^0 - \bar{x}$. Then for each $i \in I$, $0 = g_i(\bar{x}) > g_i(x^0)$, and by the pseudo-convexity of $g_i(\cdot)$ we have $\nabla g_i(\bar{x})^t d < 0$. Also, since $h_i(x), i = 1, \ldots, l$ are linear, $d^t \nabla h(\bar{x}) = 0$. Thus,

$$0 = 0^t d = (\nabla g(\bar{x})^t u + \nabla h(\bar{x})^t v)^t d < 0,$$

unless $u_i = 0$ for all $i \in I$. But if this is true, then we would have $v \neq 0$ and $\nabla h(\bar{x})^t v = 0$, violating the linear independence assumption. This is a contradiction, and so $u_0 > 0$. ∎

Theorem 19 (Linear constraints) *If all constraints are linear, the KKT conditions are necessary to characterize an optimal solution.*

Proof: Our problem is

$$
\begin{array}{ll}
(\mathrm{P}) & \min_x f(x) \\
\text{s.t.} & Ax \le b \\
& Mx = g\ .
\end{array}
$$

Suppose $\bar{x}$ is a local optimum. Without loss of generality, we can partition the constraints $Ax \le b$ into groups $A_I x \le b_I$ and $A_{\bar{I}} x \le b_{\bar{I}}$ such that $A_I \bar{x} = b_I$ and $A_{\bar{I}} \bar{x} < b_{\bar{I}}$. Then at $\bar{x}$, the set $\{d : A_I d \le 0,\ Md = 0\ \}$ is precisely the set of feasible directions. Thus, in particular, for every d as above, $\nabla f(\bar{x})^t d \ge 0$, for otherwise d would be a feasible descent direction at $\bar{x}$, violating its local optimality. Therefore, the linear system

$$\nabla f(\bar{x})^t d < 0, A_I d \le 0, Md = 0$$

has no solution. From the Key Lemma, there exists (u, v, w) satisfying $u = 1$, $v \ge 0$, and $\nabla f(\bar{x})u + A_I^T v + M^T w = 0$ which are precisely the KKT conditions. ∎

5.1 Second-Order Optimality Conditions

To describe the second order conditions for optimality, we will define the following function, known as the *Lagrangian function*, or simply the Lagrangian:

$$L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{i=1}^{l} v_i h_i(x) = f(x) + u^t g(x) + v^t h(x).$$

Using the Lagrangian, we can, for example, re-write the gradient conditions of the KKT necessary conditions as follows:

$$\nabla_x L(\bar{x}, u, v) = 0, \tag{1}$$

since $\nabla_x L(x, u, v) = \nabla f(x) + \nabla g(x)^t u + \nabla h(x)^t v$.

Also, note that $\nabla^2_{xx} L(x, u, v) = \nabla^2 f(x) + \sum_{i=1}^{m} u_i \nabla^2 g_i(x) + \sum_{i=1}^{l} v_i \nabla^2 h_i(x)$. Here we use the standard notation: $\nabla^2 q(x)$ denotes the Hessian of the function $q(x)$, and $\nabla^2_{xx}L(x,u,v)$ denotes the submatrix of the Hessian of $L(x,u,v)$ corresponding to the partial derivatives with respect to the x variables only.

Theorem 20 (KKT second order necessary conditions) *Suppose $\bar{x}$ is a local minimum of (P), and $\nabla g_i(\bar{x}),\ i \in I$ and $\nabla h_i(\bar{x}),\ i = 1,\ldots,l$ are linearly independent. Then $\bar{x}$ must satisfy the KKT conditions. Furthermore, every d that satisfies:*

$$\begin{aligned} \nabla g_i(\bar{x})^t d &\le 0, \quad i \in I, \\ \nabla h_i(\bar{x})^t d &= 0, \quad i = 1\ldots,l \end{aligned}$$

must also satisfy

$$d^t \nabla_{xx} L(\bar{x},u,v) d \ge 0 \ .$$

∎

Theorem 21 (KKT second order sufficient conditions) *Suppose the point $\bar{x} \in S$ together with multipliers (u,v) satisfies the KKT conditions. Let $I^+ = \{i \in I : u_i > 0\}$ and $I^0 = \{i \in I : u_i = 0\}$. Additionally, suppose that every $d \neq 0$ that satisfies*

$$\begin{aligned} \nabla g_i(\bar{x})^t d &= 0, \quad i \in I^+, \\ \nabla g_i(\bar{x})^t d &\le 0, \quad i \in I^0, \\ \nabla h_i(\bar{x})^t d &= 0, \quad i = 1\ldots,l \end{aligned}$$

also satisfies

$$d^t \nabla^2_{xx} L(\bar{x},u,v) d > 0 \ .$$

Then $\bar{x}$ is a strict local minimum of (P).

∎

6 A Proof of Theorem 2

The proof of Theorem 2 relies on the Implicit Function Theorem. To motivate the Implicit Function Theorem, consider a system of linear functions:

$$h(x) := Ax - b$$

and suppose that we are interested in solving

$$h(x) = Ax - b = 0 \; .$$

Let us assume that $A \in \Re^{l \times n}$ has full row rank (i.e., its rows are linearly independent). Then we can partition columns of A and elements of x as follows: $A = [B \mid N]$, $x = (y; z)$, so that $B \in \Re^{l \times l}$ is non-singular, and $h(x) = By + Nz - b$.

Let $s(z) = B^{-1}b - B^{-1}Nz$. Then for any z, $h(s(z), z) = Bs(z) + Nz - b = 0$, i.e., $x = (s(z), z)$ solves $h(x) = 0$. This idea of "invertability" of a system of equations is generalized (although only locally) by the following version of the Implicit Function Theorem, where we will preserve the notation used above:

Theorem 22 (Implicit Function Theorem) *Let* $h(x) : \Re^n \to \Re^l$ *and* $\bar{x} = (\bar{y}_1, \ldots, \bar{y}_l, \bar{z}_1, \ldots, \bar{z}_{n-l}) = (\bar{y}, \bar{z})$ *satisfy:*

1. $h(\bar{x}) = 0$
2. $h(x)$ *is continuously differentiable in a neighborhood of* $\bar{x}$
3. *The* $l \times l$ *Jacobian matrix*

$$\begin{bmatrix} \frac{\partial h_1(\bar{x})}{\partial y_1} & \cdots & \frac{\partial h_1(\bar{x})}{\partial y_l} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_l(\bar{x})}{\partial y_1} & \cdots & \frac{\partial h_l(\bar{x})}{\partial y_l} \end{bmatrix}$$

is nonsingular.

Then there exists $\epsilon > 0$ along with functions $s(z) = (s_1(z), \ldots, s_l(z))$ such that for all $z \in B(\bar{z}, \epsilon)$, $h(s(z), z) = 0$ and $s_k(z)$ are continuously differentiable. Moreover, for all $i = 1, \ldots, m$ and $j = 1, \ldots, n-l$ we have:

$$\sum_{k=1}^{l} \frac{\partial h_i(y,z)}{\partial y_k} \cdot \frac{\partial s_k(z)}{\partial z_j} + \frac{\partial h_i(y,z)}{\partial z_j} = 0 \ .$$

∎

Proof of Theorem 2: Let $A = \nabla h(\bar{x}) \in \Re^{l \times n}$. Then A has full row rank, and its columns (along with corresponding elements of $\bar{x}$) can be re-arranged so that $A = [B \mid N]$ and $\bar{x} = (\bar{y}; \bar{z})$, where B is non-singular. Let z lie in a small neighborhood of $\bar{z}$. Then, from the Implicit Function Theorem, there exists $s(z)$ such that $h(s(z), z) = 0$.

Suppose that $d \in F_0 \cap G_0 \cap H_0$, and let us write $d = (q; p)$. Then $0 = Ad = Bq + Np$, whereby $q = -B^{-1}Np$. Let $z(\theta) = \bar{z} + \theta p$, $y(\theta) = s(z(\theta)) = s(\bar{z} + \theta p)$, and $x(\theta) = (y(\theta), z(\theta))$. We will derive a contradiction by showing that d is an improving feasible direction, i.e., for small $\theta > 0$, $x(\theta)$ is feasible and $f(x(\theta)) < f(\bar{x})$.

To show feasibility of $x(\theta)$, note that for $\theta > 0$ sufficiently small, it follows from the Implicit Function Theorem that:

$$h(x(\theta)) = h(s(z(\theta)), z(\theta)) = 0 \ .$$

Furthermore, for $i = 1, \ldots, l$ we have:

$$0 = \frac{\partial h_i(x(\theta))}{\partial \theta} = \sum_{k=1}^{l} \frac{\partial h_i(s(z(\theta)), z(\theta))}{\partial y_k} \cdot \frac{\partial s_k(z(\theta))}{\partial \theta} + \sum_{k=1}^{n-l} \frac{\partial h_i(s(z(\theta)), z(\theta))}{\partial z_k} \cdot \frac{\partial z_k(\theta)}{\partial \theta} .$$

Let $r_k = \frac{\partial s_k(z(\theta))}{\partial \theta}$, and recall that $\frac{\partial z_k(\theta)}{\partial \theta} = p_k$. The above equation system can then be re-written as $0 = Br + Np$, or $r = -B^{-1}Np = q$. Therefore, $\frac{\partial x_k(\theta)}{\partial \theta} = d_k$ for $k = 1, \ldots, n$.

For $i \in I$,

$$\begin{aligned} g_i(x(\theta)) &= g_i(\bar{x}) + \theta \left. \frac{\partial g_i(x(\theta))}{\partial \theta} \right|_{\theta=0} + |\theta|\alpha_i(\theta) \\ &= \theta \sum_{k=1}^{n} \frac{\partial g_i(x(\theta))}{x_k} \cdot \left. \frac{\partial x_k(\theta)}{\partial \theta} \right|_{\theta=0} \\ &= \theta \nabla g_i(\bar{x})^t d + |\theta|\alpha_i(\theta), \end{aligned}$$

where $\alpha_i(\theta) \to 0$ as $\theta \to 0$. Hence $g_i(x(\theta)) < 0$ for all $i = 1, \ldots, m$ for $\theta > 0$ sufficiently small, and therefore, $x(\theta)$ is feasible for any $\theta > 0$ sufficiently small.

On the other hand,

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

for $\theta > 0$ sufficiently small, which contradicts the local optimality of $\bar{x}$. Therefore no such d can exist, and the theorem is proved. ∎

7 Constrained Optimization Exercises

1. Suppose that $f(x)$ and $g_i(x)$, $i = 1, \ldots, m$ are convex real-valued functions over $\Re^n$, and that $X \subset \Re^n$ is a closed and bounded convex set. Let $I = \{(s, z) \in \Re^{m+1} : \text{ there exists an } x \in X \text{ for which } g(x) \leq s, \; f(x) \leq z\}$. Prove that I is a closed convex set.

2. Suppose that $f(x)$ and $g_i(x)$, $i = 1, \ldots, m$ are convex real-valued functions over $\Re^n$, and that $X \subset \Re^n$ is a closed and bounded convex set. Consider the perturbation function:

$$\begin{array}{lrcl} z^*(y) = \text{minimum}_x & f(x) & & \\ \text{s.t.} & g_i(x) & \leq & y_i, \quad i = 1, \ldots, m \\ & x \in X \; . & & \end{array}$$

   - Prove that $z^*(\cdot)$ is a convex function.

- Show that $y_1 \le y_2$ implies that $z^*(y_1) \ge z^*(y_2)$.

3. Consider the program

$$\begin{array}{lll} (\mathrm{P}): & z^* = \mathrm{minimum}_x & \|c - x\| \\ & & \\ & \mathrm{s.t.} & \|x\| = \alpha \ , \end{array}$$

where α is a given nonnegative scalar. What are the necessary optimality conditions for this problem? Use these conditions to show that $z^* = |\|c\| - \alpha|$. What is the optimal solution x^*?

4. Let S_1 and S_2 be convex sets in $\Re^n$. Recall the definition of *strong* separation of convex sets in the notes, and show that there exists a hyperplane that strongly separates S_1 and S_2 if and only if

$$\inf\{\|x_1 - x_2\| \mid x_1 \in S_1, \ x_2 \in S_2\} \ > \ 0 \ .$$

5. Consider $S = \{x \in \Re^2 \mid x_1^2 + x_2^2 \le 1\}$. Represent S as the intersection of a collection of half-spaces. Find the half-spaces explicitly.

6. Let C be a nonempty set in $\Re^n$. Show that C is a convex cone if and only if $x_1, x_2 \in C$ implies that $\lambda_1 x_1 + \lambda_2 x_2 \in C$ whenever $\lambda_1, \lambda_2 \ge 0$ and $\lambda_1 + \lambda_2 \ne 0$.

7. Let S be a nonempty convex set in $\Re^n$ and let $f(\cdot) : S \to \Re$. Show that $f(\cdot)$ is a convex function on S if and only if for any integer $k \ge 2$ the following holds true:

$$x^1, \ldots, x^k \in S \ \Rightarrow \ f\left(\sum_{j=1}^k \lambda_j x^j\right) \le \sum_{j=1}^k \lambda_j f(x^j)$$

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

8. Let $f_1(\cdot), \ldots, f_k(\cdot) : \Re^n \to \Re$ be convex functions, and consider the function $f(\cdot)$ defined by:

$$f(x) := \max\{f_1(x), \ldots, f_k(x)\} \ .$$

Prove that $f(\cdot)$ is a convex function. State and prove a similar result for concave functions.

9. Let $f_1(\cdot), \ldots, f_k(\cdot) : \Re^n \to \Re$ be convex functions, and consider the function $f(\cdot)$ defined by:

$$f(x) := \alpha_1 f_1(x) + \cdots + \alpha_k f_k(x) \ ,$$

where $\alpha_1, \ldots, \alpha_k > 0$. Prove that $f(\cdot)$ is a convex function. State and prove a similar result for concave functions.

10. Consider the following problem:

$$\begin{array}{lccc} \text{minimum}_x & (x_1-4)^2 + (x_2-6)^2 & & \\ \text{s.t.} & -x_1^2 + x_2 & \geq & 0 \\ & x_2 & \leq & 4 \\ & x \in \Re^2 \ . & & \end{array}$$

Write a necessary condition for optimality and verify that it is satisfied by the point $(2, 4)$. Is this the optimal point? Why or why not?

11. Consider the problem to minimize $f(x)$ subject to $x \in S$ where S is a convex set in $\Re^n$ and $f(\cdot)$ is a differentiable convex function on S. Prove that $\bar{x}$ is an optimal solution of this problem if and only if $\nabla f(\bar{x})^t (x - \bar{x}) \geq 0$ for every $x \in S$.

12. Consider the following problem:

$$\begin{array}{lccc} \text{maximize}_x & 3x_1 - x_2 + x_3^2 & & \\ \text{s.t.} & x_1 + x_2 + x_3 & \leq & 0 \\ & -x_1 + 2x_2 + x_3^2 & = & 0 \\ & x \in \Re^3 \ . & & \end{array}$$

- Write down the KKT optimality conditions.
- Argue why this problem is unbounded.

13. Consider the following problem:

$$
\begin{array}{lccc}
\text{minimize}_x & \left(x_1 - \frac{9}{4}\right)^2 + (x_2 - 2)^2 & & \\
\text{s.t.} & x_2 - x_1^2 & \geq & 0 \\
 & x_1 + x_2 & \leq & 6 \\
 & x_1 \geq 0 & & \\
 & x_2 \geq 0 & & \\
 & x \in \Re^2 \ . & &
\end{array}
$$

- Write down the KKT optimality conditions and verify that these conditions are satisfied at the point $\bar{x} = \left(\frac{3}{2}, \frac{9}{4}\right)$.
- Present a graphical interpretation of the KKT conditions at $\bar{x}$.
- Show that $\bar{x}$ is the optimal solution of the problem.

14. Let $f(\cdot) : \Re^n \to \Re$, $g_i(\cdot) : \Re^n \to \Re$, $i = 1, \ldots, m$, be convex functions. Consider the problem to minimize $f(x)$ subject to $g_i(x) \leq 0$ for $i = 1, \ldots, m$, and suppose that the optimal objective value of this problem is z^* and is attained at some feasible point x^*. Let M be a proper subset of $\{1, \ldots, m\}$ and suppose that $\hat{x}$ solves the problem to minimize $f(x)$ subject to $g_i(x) \leq 0$ for $i \in M$. Let $V := \{i \mid g_i(\hat{x}) > 0\}$. If $z^* > f(\hat{x})$, show that $g_i(x^*) = 0$ for some $i \in V$. (This shows that if an unconstrained minimum of $f(\cdot)$ is infeasible and has an objective value that is less than z^*, then any constrained minimum lies on the boundary of the feasible region.)

15. Consider the following problem, where $c \neq 0$ is a vector in $\Re^n$:

$$
\begin{array}{lccc}
\text{minimize}_d & c^T d & & \\
\text{s.t.} & d^t d & \leq & 1 \\
 & d \in \Re^n \ . & &
\end{array}
$$

- Show that $\bar{d} := -\frac{c}{\|c\|_2}$ is a KKT point of this problem. Furthermore, show that $\bar{d}$ is indeed the unique optimal solution.

- How is this result related to the definition of the direction of steepest descent in the steepest descent algorithm?

16. Consider the following problem, where b and $a_j, c_j, j = 1, \ldots, n$ are positive constants:

$$\begin{array}{rrcl} \text{minimize}_x & \sum_{j=1}^{n} \frac{c_j}{x_j} & & \\ \text{s.t.} & \sum_{j=1}^{n} a_j x_j & = & b \\ & x_j & \geq & 0, \; j = 1, \ldots, n \\ & x \in \Re^n \; . & & \end{array}$$

Write down the KKT optimality conditions, and solve for the point $\bar{x}$ that solves this problem.

17. Let $c \in \Re^n$, $b \in \Re^m$, $A \in \Re^{m \times n}$, and $H \in \Re^{n \times n}$. Consider the following two problems:

$$\begin{array}{rrrcl} P_1 : & \text{minimize}_x & c^t x + \frac{1}{2} x^T H x & & \\ & \text{s.t.} & Ax & \leq & b \\ & & x \in \Re^n & & \end{array}$$

and

$$\begin{array}{rrrcl} P_2 : & \text{minimize}_u & h^t u + \frac{1}{2} u^T G u & & \\ & \text{s.t.} & u & \geq & 0 \\ & & u \in \Re^m \; , & & \end{array}$$

where $G := AH^{-1}A^T$ and $h := AH^{-1}c + b$. Investigate the relationship between the KKT conditions of these two problems.

18. Consider the problem to minimize $f(x)$ subject to $Ax \leq b$. Suppose that $\bar{x}$ is a feasible solution such that $A_\beta \bar{x} = b_\beta$ and $A_\eta \bar{x} < b_\eta$ where

β, η are a partition of the rows of A. Assuming that A_β has full rank, the matrix P that projects any vector onto the nullspace of A_β is given by:

$$P = I - A_\beta^T [A_\beta A_\beta^T]^{-1} A_\beta \ .$$

- Let $\bar{d} = -P\nabla f(\bar{x})$. Show that if $\bar{d} \neq 0$ then $\bar{d}$ is an improving direction, that is, $\bar{x} + \lambda \bar{d}$ is feasible and $f(\bar{x} + \lambda \bar{d}) < f(\bar{x})$ for all $\lambda > 0$ and sufficiently small.
- Suppose that $\bar{d} = 0$ and that $u := -A_\beta^T [A_\beta A_\beta^T]^{-1} A_\beta \nabla f(\bar{x}) \geq 0$. Show that $\bar{x}$ is a KKT point.
- Show that $\bar{d}$ is a positive multiple of the optimal solution of the following problem:

$$\begin{array}{lccc} \text{minimize}_d & \nabla f(\bar{x})^T d & & \\ \text{s.t.} & A_\beta d & 0 & 0 \\ & d^T d & \leq & 1 \\ & d \in \Re^n \ . & & \end{array}$$

- Suppose that $A = -I$ and $b = 0$, that is, the constraints are of the form "$x \geq 0$". Develop a simple way to construct $\bar{d}$ in this case.

19. Consider the problem to minimize $f(x)$ subject to $x \in X$ and $g_i(x) \leq 0, i = 1, \ldots, m$. Let $\bar{x}$ be a feasible point, and let $I := \{i \mid g_i(\bar{x}) = 0\}$. Suppose that X is an open set and $g_i(x), i = 1, \ldots, m$ are continuous functions, and let $J := \{i \mid g_i(\cdot) \text{ is pseudoconcave}\}$. Furthermore, suppose that

$$\left\{ d \mid \nabla g_i(\bar{x})^t d \leq 0 \text{ for } i \in J,\ \nabla g_i(\bar{x})^t d < 0 \text{ for } i \in I \setminus J \right\}$$

is nonempty. Show that this condition is sufficient to validate the KKT conditions at $\bar{x}$. (This is called the "Arrow-Hurwicz-Uzawa constraint qualification.")