Journal of Mathematical Analysis and Applications **253,** 290–296 (2001)
doi:10.1006/jmaa.2000.7115, available online at http://www.idealibrary.com on IDEAL®

A Necessary and Sufficient Condition for Minimizing a Convex Fréchet Differentiable Function on a Certain Hyperplane

Han-Lin Li

Institute of Information Management, National Chaio Tung University, Hsin-Chu, Taiwan

Yi-Hsin Liu and Valentin Matache

Department of Mathematics, University of Nebraska at Omaha, Omaha, Nebraska 68182-0243

and

Po-Lung Yu

School of Business, University of Kansas, Lawrence, Kansas 66045-2003; and Institute of Information Management, National Chaio Tung University, Hsin-Chu, Taiwan

Submitted by George Leitmann

Received May 30, 2000

A convex Fréchet differentiable function is minimized subject to a certain hyperplane at a point if the function is minimized in all directions which are defined by a finite set of vectors. The proposed approach is different from the Lagrange multiplier approach. At the end of this paper, a linear program is formulated to solve the case when the above given convex function is quadratic.

Key Words: convex function; Fréchet differentiable; quadratic programming.

1. INTRODUCTION

Let f be a real-valued convex Fréchet differentiable function on R^n. The function f is to be minimized subject to a constraint hyperplane

$$H_c = \left\{ x \in R^n : x = (x_1, x_2, \ldots, x_n),\ \sum_{j=1}^{n} x_j = c \right\},$$

$$\text{i.e., } \min_{x \in H_c} f(x). \tag{1}$$

0022-247X/01 $35.00

This problem occurs in many real life situations such as portfolio analysis, investment analysis, etc. The new necessary and sufficient condition proposed in this paper can be applied to improve the computation efficiency. This necessary and sufficient condition allows one to solve the above problem (1) by minimizing the convex objective function in every direction $v^j = (v_1^j, v_2^j, \ldots, v_n^j)^T \in R^n$, $j = 1, 2, \ldots, n$, where

$$v_k^j = \begin{cases} -1 & \text{if } k = j \\ \frac{1}{n-1} & \text{if } k \neq j. \end{cases} \tag{2}$$

We will show that if a convex Fréchet differentiable function is minimized at a point $x_0 \in H_c$ in every direction v^j, $j = 1, 2, \ldots, n$, as defined in (2), then the function attains its minimum over H_c at x_0. More precisely, we establish a condition for the solution of problem (1). The condition is stated and proved as Theorem 1, in Section 2 (Main Results). This allows one to obtain solutions of (1) by considering all directions v^j, $j = 1, 2, \ldots, n$, and the following necessary and sufficient condition: $f(x_0) \leq f(x_0 + tv^j)$, $j = 1, 2, \ldots, n$, $\forall t \in [0, \varepsilon]$, for some $\varepsilon > 0$, if and only if $f(x_0) \leq f(x)$, $\forall x \in H_c$.

The solution of (1) is usually obtained by applying the Lagrange multiplier condition and solving a system of equations to reach an optimal solution. The condition established in this paper allows one to approach the problem without using the Lagrange multiplier. This is an alternative method of solving this class of problems.

2. MAIN RESULTS

Let's consider the hyperplane $H_c = \{x \in R^n: \sum_{j=1}^n x_j = c\}$. Obviously H_0 is a linear subspace of R^n and $H_c = x + H_0$, $\forall x \in H_c$. The Fréchet differential of f at x_0, denoted by $d_{x_0} f$, can be calculated as $d_{x_0} f(x) = \langle \nabla f(x_0), x \rangle$, $x \in R^n$, where $\langle \cdot, \cdot \rangle$ is the usual inner product in R^n and $\nabla f(x_0) = ((\partial f/\partial x_1)(x_0), (\partial f/\partial x_2)(x_0), \ldots, (\partial f/\partial x_n)(x_0))$.

Let's denote $v_0 = (1, 1, \ldots, 1) \in R^n$. This vector is normalized as $q = v_0/\|v_0\| = v_0/\sqrt{n} = (1/\sqrt{n}, \ldots, 1/\sqrt{n})$.

By definition of $H_0 = \{x \in R^n: \sum_{j=1}^n x_j = 0\}$, one has $v_0 \perp H_0$, i.e., $\langle v_0, x \rangle = 0$, $\forall x \in H_0$ or $v_0, q \in H_0^\perp$, the orthogonal complement of H_0. Let U be a unitary operator (i.e., $U^T U = U U^T = I$, the identity) on R^n with the property $U((0, 0, \ldots, 0, 1)) = q = (1/\sqrt{n}, 1/\sqrt{n}, \ldots, 1/\sqrt{n})$.

Then $\langle U(x_1, \ldots, x_{n-1}, 0), \ U(0, \ldots, 0, 1)\rangle = \langle (x_1, \ldots, x_{n-1}, 0), U^T U(0, \ldots, 0, 1)\rangle = \langle (x_1, \ldots, x_{n-1}, 0), (0, \ldots, 0, 1)\rangle = 0$, which implies $U(x_1, x_2, \ldots, x_{n-1}, 0) \in H_0$, $\forall x_i \in R$, $i = 1, 2, \ldots, n-1$. Thus if we denote $S = \{(x_1, x_2, \ldots, x_{n-1}, 0)\colon x_i \in R, i = 1, \ldots, n-1\}$, U maps S onto H_0, i.e., $U(S) = H_0$. Let T_{x_0} be the translation mapping by x_0, i.e., $T_{x_0}(x) = x + x_0$, $\forall x \in R^n$. Define a function F on R^{n-1} by

$$F((x_1, x_2, \ldots, x_{n-1})) = f\big(T_{x_0}(U((x_1, x_2, \ldots, x_{n-1}, 0)))\big). \qquad (3)$$

Then $F(0) = f(x_0)$. Recall f is convex. Now T is affine and U is linear. Thus we have the following lemma.

LEMMA 1. *F is a convex function on R^{n-1}.*

Let τ be defined on R^{n-1} by $\tau(x_1, x_2, \ldots, x_{n-1}) = (x_1, x_2, \ldots, x_{n-1}, 0)$. Then $F = f \circ T_{x_0} \circ U \circ \tau$. Since U and τ are linear functions, U and τ are Fréchet differentiable at any point with the differential equal to U and τ, respectively. The map T_{x_0} is obviously differentiable at all points with differential equal to the identity map. Since $F(0) = f(x_0)$ and f is Fréchet differentiable at x_0, it follows by the chain rule that F is Fréchet differentiable at 0 with differential given by $[d_0 F] = [d_{x_0} F][I][U][\tau]$. Here we use $[\cdot]$ to denote the matrix of a linear transformation. This is stated as the following lemma.

LEMMA 2. *F is Fréchet differentiable at 0 with Fréchet differential given by $d_0 F(x_1, x_2, \ldots, x_{n-1}) = d_{x_0} f(U(x_1, x_2, \ldots, x_{n-1}, 0))$.*

Note that the set of all directions, $\{v^j\colon j = 1, 2, \ldots, n\}$ defined in (2), is a linearly dependent set of n vectors in H_0, and

(a) any proper nonempty subset of $\{v^j\colon j = 1, 2, \ldots, n\}$ is a linearly independent set of vectors,

(b) $v^1 + v^2 + \cdots + v^n = 0$.

The selection of the set $\{v^j\colon j = 1, 2, \ldots, n\}$ is not unique. In fact, one can select any subset of H_0 with properties (a) and (b) in order to prove the following.

LEMMA 3. *For every vector $v \in H_0$ there exist some nonnegative real numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that $v = \sum_{j=1}^n \alpha_j v^j$.*

Proof. It is easy to see that $v^1, v^2, \ldots, v^{n-1}$ are linearly independent and $v^1, v^2, \ldots, v^n$ belong to H_0, a $(n-1)$-dimensional subspace of R^n. Hence $v^1, v^2, \ldots, v^n$ span H_0. Thus, for each vector $v \in H_0$, there exist real numbers λ_j, $j = 1, 2, \ldots, n$ such that $v = \sum_{j=1}^n \lambda_j v^j$.

Let $M = \max\{|\lambda_j|: j = 1, 2, \ldots, n\}$. Set $\alpha_j = M + \lambda_j \geq 0$, $j = 1, 2, \ldots, n$. Observe that $\sum_{j=1}^n \alpha_j v^j = \sum_{j=1}^n (M + \lambda_j) v^j M \sum_{j=1}^n v^j + \sum_{j=1}^n \lambda_j v^j = 0 + v = v$. The lemma is proved.

LEMMA 4. *If for some* $\varepsilon > 0$,

$$f(x_0) \leq f(x_0 + tv^j), \qquad j = 1, 2, \ldots, n, \ \forall t \in [0, \varepsilon] \tag{4}$$

then $(d_{x_0} f)(v) = 0$, $\forall v \in H_0$.

Proof. $f(x_0) \leq f(x_0 + tv^j)$, $j = 1, 2, \ldots, n$, $\forall t \in [0, \varepsilon]$ implies $(d_{x_0} f)(v^j) = \lim_{t \to 0^+} (f(x_0 + tv^j) - f(x_0)/t) \geq 0$, $j = 1, 2, \ldots, n$.
By Lemma 3 one deduces $(d_{x_0} f)(v) \geq 0$, $\forall v \in H_0$.
If $v \in H_0$ then $-v \in H_0$, hence $(d_{x_0} f)(-v) \geq 0$, i.e., $(d_{x_0} f)(v) \leq 0$, $\forall v \in H_0$. Thus, $(d_{x_0} f)(v) = 0$, $\forall v \in H_0$.

THEOREM 1. *Let f be a real valued convex function defined on R^n. If f is Fréchet differentiable at $x_0 \in R^n$ and satisfies condition* (4), *namely, for some* $\varepsilon > 0$, $f(x_0) \leq f(x_0 + tv^j)$, $j = 1, 2, \ldots, n$, $\forall t \in [0, \varepsilon]$, *then* $f(x_0) \leq f(x)$, $\forall x \in H_c$. *Conversely, if* $f(x_0) \leq f(x)$, $\forall x \in H_c$ *then f satisfies condition* (4).

Proof. For the sufficiency part, let F be as defined in (3). Since $T_{x_0}(U(\tau(R^{n-1})) = H_c$, it is sufficient to show $F(0) \leq F(x)$, $\forall x \in R^{n-1}$.
Observe that $U(\tau(R^{n-1})) = H_0$ and by Lemma 2 and Lemma 4 we have

$$d_0 F((x_1, x_2, \ldots, x_{n-1})) = d_{x_0} f(U(\tau((x_1, x_2, \ldots, x_{n-1})))) = 0,$$
$$\forall (x_1, x_2, \ldots, x_{n-1}) \in R^{n-1}.$$

By Lemma 1, F is convex. Hence according to Theorem 25.1 in Ref. [2, p. 242], we can deduce that $F(0) \leq F(x)$, $\forall x \in R^{n-1}$. The converse implication is obvious.

3. APPLICATIONS TO QUADRATIC PROGRAMMING

Let Q be a positive semidefinite $n \times n$ matrix, B a $1 \times n$ matrix, and $f(x) = \frac{1}{2} x^T Q x + Bx$. Then (1) becomes a quadratic program

$$\min_{x \in H_c} f(x) = \frac{1}{2} x^T Q x + Bx. \tag{QP}$$

Applying Lemma 4 and Theorem 1, we obtain the following theorem.

THEOREM 2. *x^* is an optimal solution of (QP) if and only if x^* is a solution of the following system of linear equations.*

$$\begin{aligned} x^T Q v^j + B v^j = 0, \qquad j = 1, 2, \dots, n \\ \mathbf{1} x = c, \end{aligned} \tag{5}$$

where $\mathbf{1} = (1, 1, \dots, 1)$ and v^j is defined in (2).

Proof. $x^* \in H_c$ if and only if $\mathbf{1} x^* = c$.
x^* is an optimal solution of (QP) if and only if $(d_{x^*} f)(v^j) = 0$, $\forall j = 1, 2, \dots, n$

$$\begin{aligned} (d_{x^*} f)(v^j) &= \lim_{t \to 0^+} \frac{f(x^* + t v^j) - f(x^*)}{t} \\ &= \lim_{t \to 0^+} \frac{(1/2)(x^* + t v^j)^T Q (x^* + t v^j) + B(x^* + t v^j) - \left((1/2) x^{*T} Q x^* + B x^*\right)}{t} \\ &= x^{*T} Q v^j + B v^j \\ &= 0. \end{aligned}$$

COROLLARY. *x^* is an optimal solution of (QP) if and only if x^* solves the following linear program (LP-Q) with the optimal objective function value* 0,

$$\begin{aligned} &\min \sum_{j=1}^{n+1} s_j \\ &\text{s.t.} \\ &x^T Q v^j + B v^j + s_j = 0, \qquad j = 1, 2, \dots, n \\ &\mathbf{1} x + s_{n+1} = c \\ &s_j \geq 0,\ j = 1, 2, \dots, n+1. \end{aligned} \tag{LP-Q}$$

In order to obtain the solution of (QP) via the Lagrange multiplier theorem one solves the system of equations

$$\begin{aligned} \left(x^T Q + B\right) + \lambda \mathbf{1} = 0 \\ \mathbf{1} x = c. \end{aligned} \tag{6}$$

To find an optimal solution of (QP) applying Theorem 1, one solves (5). Conditions (5) and (6) have the same number of equations and (6) contains a Lagrange multiplier. Thus, solving the system (6), usually, one has to determine more variables than solving (5).

4. AN EXAMPLE

Based on the above theorems, one can apply the linear program (LP-Q) to solve (1) when the objective is a quadratic function. It is easy to see that solving (1) by the method proposed in this paper can be more efficient than applying the traditional Lagrange multiplier theorem. The following example is solved by both methods to illustrate the differences between them.

EXAMPLE.

$$\min f(x,y,z) = \frac{1}{2}x^2 + y^2 + \frac{3}{2}z^2 + xz + yz + x + y + 2z$$

$$\text{s.t.}$$

$$x + y + z = 1.$$

Solution.

$$\begin{aligned} f(x,y,z) &= \frac{1}{2}x^2 + y^2 + \frac{3}{2}z^2 + xz + yz + x + y + 2z \\ &= \frac{1}{2}(x,y,z)\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} + (1,1,2)\begin{pmatrix} x \\ y \\ z \end{pmatrix} \end{aligned}$$

$$Q = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}, \qquad V = (v^1 \quad v^2 \quad v^3) = \begin{pmatrix} -1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -1 \end{pmatrix}.$$

Applying Theorem 2, we solve the system of equations:

$$\begin{cases} (x,y,z)QV + BV = 0 \\ x + y + z = 1, \end{cases}$$

i.e.,

$$\begin{cases} -x + 3y + 2z + 1 = 0 \\ 2x - 3y + 2z + 1 = 0 \\ -x \qquad\quad - 4z - 2 = 0 \\ x + y + z = 1. \end{cases} \tag{7}$$

By the Corollary, an alternative is to solve the linear program

$$
\begin{aligned}
&\min s_1 + s_2 + s_3 + s_4 \\
&\text{s.t.} \\
&-x + 3y + 2z + s_1 = -1 \\
&2x - 3y + 2z + s_2 = -1 \\
&-x \quad - 4z + s_3 = 2 \\
&x + y + z + s_4 = 1 \\
&s_j \geq 0, \qquad j = 1,2,3,4.
\end{aligned}
\tag{8}
$$

The optimal solution is $s_i = 0$, $i = 1,2,3,4$, $x = 1.2$, $y = 0.6$, $z = -0.8$ and the optimal objective function value of this linear program is 0. Thus $x = 1.2$, $y = 0.6$, and $z = -0.8$ minimizes f over the set $H_1 = \{(x, y, z)\colon x + y + z = 1\}$.

To apply the Lagrange multiplier theorem, one solves the equations

$$
\begin{cases}
(x, y, z)Q + B + \lambda(1,1,1) = 0 \\
x + y + z = 1,
\end{cases}
$$

i.e., one solves the system of equations

$$
\begin{cases}
x \quad + z \quad + \lambda + 1 = 0 \\
\quad 2y + z + \lambda + 1 = 0 \\
x + y + 3z + \lambda + 2 = 0 \\
x + y + z - 1 = 0.
\end{cases}
$$

Solving this system one obtains the same solutions as expected.

This example demonstrates the methods based on Theorem 2 and its Corollary. Also, the method developed in this paper is an alternative to the Lagrange multiplier theorem for (1).

REFERENCES

1. D. G. Luenberger, “Optimization by Vector Space Methods,” Wiley, New York, 1969.
2. R. T. Rockefeller, “Convex Analysis,” Princeton Univ. Press, Princeton, NJ, 1970.
3. A. Wouk, “A Course of Applied Functional Analysis,” Wiley, New York, 1979.
4. P. L. Yu, “Multiple-Criteria Decision Making: Concepts, Techniques, and Extensions,” Plenum, New York, 1985.