Preprints of Seventh Triennial World Congress of IFAC, Helsinki, June 1978, Vol. 2, pp. 1079-1084.

LOCAL CONVEX CONJUGACY AND FENCHEL DUALITY*

Dimitri P. Bertsekas

Department of Electrical Engineering and Coordinated Science Laboratory, University of Illinois, Urbana, Illinois 61801, USA

Abstract. In this paper we introduce a notion of a convex conjugate function of a nonlinear function defined on a manifold specified by nonlinear equality constraints. Under certain assumptions the conjugate is defined locally around a point and upon conjugation yields the original function. Local versions of the Fenchel duality theorem are also proved.

INTRODUCTION

The notion of a conjugate convex function permeates the subject of optimization under convexity assumptions and plays a fundamental role in the development of modern duality theory. Since at present the theory is limited to convex functions (see Rockafellar (1970) for an extensive treatment) it is interesting to attempt to provide notions of a convex conjugate of a nonconvex function which retain at least some of the important properties possessed by the one under convexity. One possibility is to consider real valued functions $f: R^n \to R$ which are convex only locally, i.e., over an open sphere $S(\bar{x};\delta)$ centered at a point $\bar{x}$ and having radius $\delta > 0$. Then one may define a convex conjugate $\phi: R^n \to (-\infty,+\infty]$ by means of

$$\phi(y) = \sup_{x \in S(\bar{x};\delta)} \{y'x - f(x)\}.$$

While, just as in the convex case, it is possible to recover the original function f over $S(\bar{x};\delta)$ by conjugation on ϕ, one has not really achieved much in the way of a novelty since ϕ is the ordinary conjugate convex function of the convex function $\bar{f}: R^n \to (-\infty,+\infty]$ defined by

$$\bar{f}(x) = \begin{cases} f(x) & \text{if } x \in S(\bar{x};\delta) \\ +\infty & \text{otherwise} \end{cases}$$

and its properties are obtained via the standard theory. A nontrivial extension is obtained by considering local conjugates of functions of the form

$$\bar{f}(x) = \begin{cases} f(x) & \text{if } x \in S(\bar{x};\delta),\ h(x) = 0 \\ +\infty & \text{otherwise} \end{cases}$$

where $f: R^n \to R$ is a real valued function and $h: R^n \to R^m$ is a perhaps nonlinear mapping from R^n to R^m.

In this paper we define under certain second order differentiability assumptions the conjugate of $\bar{f}$ above essentially via the relation

$$\phi(y) = \sup_{\substack{h(x)=0 \\ x \in S(\bar{x};\delta)}} \{y'x - f(x)\}.$$

The definition is local in nature, i.e., ϕ is defined over an open sphere centered at some point $\bar{y} \in R^n$ to be specified later. Under our assumptions ϕ is real-valued and convex locally around $\bar{y}$ and upon conjugation yields $f(x)$ for points $x \in S(\bar{x};\delta)$ which belong to the manifold $\{x | h(x) = 0\}$. We demonstrate various properties of the conjugate, and we show that local analogs of Fenchel's duality theorem hold. At this point it is difficult to assess the significance of local conjugate convex functions in applications related to optimization or other fields. They appear however naturally in the analysis of certain Augmented Lagrangian methods (Bertsekas, 1978), and this fact offers some hope that they will find application in other studies as well.

The analysis throughout the paper is conducted in n-dimensional Euclidean space denoted R^n and equipped with the usual norm denoted $|\cdot|$ (i.e., $|x| = (\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}}$ for $x = (x_1,\ldots,x_n)' \in R^n$). In our notation all vectors will be considered as column vectors. A prime denotes transposition. For $\varepsilon > 0$ and $x \in R^n$ we denote $S(x;\varepsilon)$ the open sphere centered at x with radius ε. For any function $h: R^n \to R$ we denote $\nabla h(x)$ and $\nabla^2 h(x)$ the gradient and Hessian matrix of h at a point x. For $h: R^n \to R^m$, $h = (h_1,\ldots,h_m)'$ we denote $\nabla h(x)$ the $n \times m$ matrix having as columns the gradients $\nabla h_1(x),\ldots,\nabla h_m(x)$. For any $x \in R^n$

*This work was carried out at the Coordinated Science Lab., University of Illinois, Urbana, Illinois and was supported by the National Science Foundation under Grant ENG 74-19332.

the notation $x \leq 0$ means that all coordinates of x are nonpositive.

2. LOCAL CONVEX CONJUGATE FUNCTIONS

It will be useful to introduce first the local conjugate of a twice continuously differentiable real valued function.

Definition 1: Given a function $f: R^n \to R$ we say that f is locally convex at a point $\bar{x} \in R^n$ if f is twice continuously differentiable in a neighborhood of $\bar{x}$ and the Hessian matrix $\nabla^2 f(\bar{x})$ is positive definite.

For a given function f which is locally convex at $\bar{x}$ we have that $\nabla^2 f(x)$ is positive definite within a neighborhood of $\bar{x}$. Let

$$\bar{y} = \nabla f(\bar{x})$$

and consider the problem

$$\text{maximize } y'x - f(x).$$

By using the inverse function theorem in the equation $y = \nabla f(x)$ and the positive definiteness of $\nabla^2 f(\bar{x})$ it follows that for every y in a sphere $S(\bar{y};\varepsilon)$ this problem has a unique local maximum closest to $\bar{x}$ and denoted $x(y)$. Define the local convex conjugate function of f at $\bar{x}$ by

$$\phi(y) = y'x(y) - f[x(y)] \qquad \forall y \in S(\bar{y};\varepsilon).$$

We have the following proposition which can be proved easily using the inverse function theorem.

Proposition 1: 1) The function ϕ is convex and twice continuously differentiable on $S(\bar{y};\varepsilon)$. Furthermore for all $y \in S(\bar{y};\varepsilon)$

$$\nabla\phi(y) = x(y)$$
$$\nabla^2\phi(y) = \{\nabla^2 f[x(y)]\}^{-1}.$$

2) Define

$$\tilde{f}(x) = \sup_{y \in S(\bar{y};\varepsilon)} \{x'y - \phi(y)\}.$$

Then there exists a $\delta > 0$ such that

$$\tilde{f}(x) = f(x) \qquad \forall x \in S(\bar{x};\delta).$$

Consider now the following definition:

Definition A.1: Given a function $f: R^n \to R$ and a mapping $h: R^n \to R^m$, $h = (h_1,\ldots,h_m)'$ we say that f is h-locally convex at a point $\bar{x} \in R^n$ if $h(\bar{x}) = 0$, f and h are twice continuously differentiable in a neighborhood of $\bar{x}$, the gradients $\nabla h_1(\bar{x}),\ldots,\nabla h_m(\bar{x})$ are linearly independent and there exists a vector $\bar{\lambda} = (\bar{\lambda}_1,\ldots,\bar{\lambda}_m)' \in R^m$ such that

$$z'[\nabla^2 f(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i \nabla^2 h_i(\bar{x})]z > 0, \quad \forall z \neq 0, \quad \nabla h(\bar{x})'z = 0.$$

The reader may at this point wonder whether the requirements in the definition above are so restrictive as to render the notion introduced essentially useless. It is thus worth pointing out that if $\bar{x}$ is a local minimum of the problem $\min\{f(x) \mid h(x) = 0\}$ and satisfies together with a unique Lagrange multiplier vector $\bar{\lambda}$ the standard second order sufficiency conditions for optimality (see below) then f is h-locally convex at $\bar{x}$. This fact demonstrates that h-local ~~conjugacy~~ convexity is a relevant notion at least for optimization studies (see Bertsekas, 1978 for an application).

Let f be an h-locally convex function at $\bar{x}$ and let $\bar{\lambda}$ be some vector satisfying

$$z'[\nabla^2 f(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i \nabla^2 h_i(\bar{x})]z > 0, \quad \forall z \neq 0, \quad \nabla h(\bar{x})z = 0. \tag{1}$$

Let

$$\bar{y} = \nabla f(\bar{x}) + \nabla h(\bar{x})\bar{\lambda}. \tag{2}$$

Consider for a fixed y the problem

$$\text{maximize } y'x - f(x) \quad \text{subject to } h(x) = 0. \tag{3}$$

The standard second order sufficiency conditions for this problem (see Luenberger, 1973 p. 226) are

$$y - \nabla f(x) - \nabla h(x)\lambda = 0, \quad h(x) = 0 \tag{4a}$$

$$z'[\nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x)]z > 0, \quad \forall z \neq 0, \quad \nabla h(x)'z = 0. \tag{4b}$$

The vectors $\bar{y}$, $\bar{x}$, $\bar{\lambda}$ satisfy conditions (4). From the implicit function theorem it follows that there exist scalars $\varepsilon > 0$ and $\sigma > 0$ such that for every $y \in S(\bar{y};\varepsilon)$ problem (3) has a unique local maximum within a sphere $S(\bar{x};\sigma)$ denoted $x(y)$ and a unique associated Lagrange multiplier vector within $S(\bar{x};\sigma)$ denoted $\lambda(y)$ and satisfying for all $y \in S(\bar{y};\varepsilon)$

$$y - \nabla f[x(y)] - \nabla h[x(y)]\lambda(y) = 0, \quad h[x(y)] = 0 \tag{5a}$$

$$z'[\nabla^2 f[x(y)] + \sum_{i=1}^m \lambda_i(y) \nabla^2 h_i[x(y)]]z > 0, \quad \forall z \neq 0, \quad \nabla h[x(y)]'z = 0. \tag{5b}$$

Furthermore the vectors $\nabla h_i[x(y)]$, $i=1,\ldots,m$ are linearly independent for all $y \in S(\bar{y};\varepsilon)$. We define the h-local convex conjugate of f at $(\bar{x},\bar{\lambda})$ by

$$\phi(y) = y'x(y) - f[x(y)], \quad \forall y \in S(\bar{y};\varepsilon). \tag{6}$$

Let us now investigate some of the properties of h-local conjugates. Consider the set S of all $y \in R^n$ around which a conjugate can be defined locally

$$S = \{y \mid y = \nabla f(x) + \nabla h(x)\lambda \text{ for some } (x,\lambda) \text{ with } h(x) = 0,\ z'[\nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 h_i(x)]z > 0, \forall z \neq 0,\ \nabla h(x)'z = 0, \text{ and } \nabla h_1(x),\ldots,\nabla h_m(x) \text{ linearly independent}\}. \tag{7}$$

We first observe that S is an open set. This can be seen from the fact that if $\bar{y} \in S$ and an h-local convex conjugate at $(\bar{x},\bar{\lambda})$ is defined within $S(\bar{y};\varepsilon)$, then for every $y \in S(\bar{y};\varepsilon)$ the vectors $x(y)$, $\lambda(y)$ satisfy (5) and hence $S(\bar{y};\varepsilon) \subset S$. Now to each vector $y \in S$ there may correspond more than one pair (x,λ) satisfying (4)[†] and the value of the conjugate ϕ at y will depend on the corresponding pair (x,λ). Thus a perhaps more appropriate notion for ϕ and indeed an alternative (and equivalent) definition would be to set

$$\phi(y;x,\lambda) = y'x-f(x)$$

for all $y \in S$, and (x,λ) satisfying (4). This completely specifies the local conjugate for all points where it can be defined. In our definition (6) the dependence on (x,λ) is suppressed since ϕ is defined only locally within $S(\bar{y},\varepsilon)$ rather than over the whole set S. There is one case, however, where the value of the conjugate does not depend on (x,λ), namely when to each $y \in S$ there corresponds a unique pair $[x(y),\lambda(y)]$ satisfying (5). Under these circumstances we have

$$\phi(y) = y'x(y)-f[x(y)] \qquad \forall y \in S,$$

and specification of y determines $x(y)$ and hence also $\phi(y)$. Some examples may be helpful in clarifying this situation.

Example 1: Let

$$f(x) = \alpha'x, \qquad h(x) = \tfrac{1}{2}(|x|^2-1)$$

where α is a given vector in R^n. Let $\bar{x}$ be any point satisfying $|\bar{x}| = 1$. Then (1) holds for all $\bar{\lambda} \in R$ with $\bar{\lambda} > 0$. Hence the h-local conjugate of f can be defined within a sphere centered at each point $\bar{y}$ of the form

$$\bar{y} = \alpha + \bar{\lambda}\bar{x}, \quad \text{with } \bar{\lambda} > 0, \quad |\bar{x}| = 1. \tag{8}$$

Thus a local conjugate can be defined for every $\bar{y} \neq \alpha$ and to each $\bar{y} \neq \alpha$ there corresponds a unique pair $(\bar{x},\bar{\lambda})$ satisfying (8). We have by straightforward calculation

$$\phi(y) = |y-\alpha| \qquad \forall y \neq \alpha. \tag{9}$$

One may attempt to define a conjugate convex function of ϕ by

$$\phi^*(x) = \sup_{y \neq \alpha}\{y'x-\phi(y)\}$$

and straightforward calculation using (9) yields

$$\phi^*(x) = \begin{cases} \alpha'x & \text{if } |x| \leq 1 \\ +\infty & \text{otherwise.} \end{cases}$$

Thus the original function f is recovered on the manifold $\{x \mid h(x) = 0\}$, i.e.,

$$\phi^*(x) = f(x), \qquad \text{if } h(x) = 0,$$

although we may have $\phi^*(x) \neq f(x)$ if $h(x) \neq 0$.

[†] If $M_{\bar{y}}$ is the set of pairs (x,λ) corresponding to $\bar{y} \in S$ as in (7), then for any $(\bar{x},\bar{\lambda}) \in M_{\bar{y}}$ there exists a sphere $S[(\bar{x},\bar{\lambda});\bar{\gamma}], \bar{\gamma} > 0$, such that $S[(\bar{x},\bar{\lambda});\bar{\gamma}] \cap M_{\bar{y}} = \{(\bar{x},\bar{\lambda})\}$. This follows from the implicit function theorem and implies that $M_{\bar{y}}$ is a countable set. (Pick a vector $\bar{r} \in R^{n+m}$ with rational coordinates in each $S[(\bar{x},\bar{\lambda});\bar{\gamma}]$ and establish a one-one correspondence of $M_{\bar{y}}$ with a countable subset of R^{n+m}).

In Example 1 the functions f and h are convex and it can be verified that the conjugate ϕ equals within its domain of definition the function

$$\bar{\phi}(y) = \sup_{h(x) \leq 0}\{y'x-f(x)\}.$$

It is thus of interest to consider an example where h is not convex.

Example 2: For $n = 2$ let

$$f(x) = 0, \qquad h(x) = h(x_1,x_2) = \tfrac{1}{3}\,x_1^3 - x_2.$$

Let $\bar{x} = (\bar{x}_1,\bar{x}_2)'$, be any vector satisfying $\frac{1}{3}\bar{x}_1^3 = \bar{x}_2$. Then, for $z = (z_1,z_2)'$, (1) is written as

$$2\bar{\lambda}\,\bar{x}_1 z_1^2 > 0, \quad \forall z \neq 0 \quad \bar{x}_1^2 z_1 - z_2 = 0.$$

The above relation is satisfied for all $(\bar{x},\bar{\lambda})$ with $\bar{\lambda}\bar{x}_1 > 0$. The h-local conjugate of f can be defined locally at each point $\bar{y}$ of the form

$$\bar{y} = \begin{bmatrix} \bar{\lambda}\bar{x}_1^2 \\ -\bar{\lambda} \end{bmatrix} \quad \text{with } \bar{\lambda}\bar{x}_1 > 0, \; \tfrac{1}{3}\bar{x}_1^3 = \bar{x}_2.$$

Hence the domain of definition of ϕ is the set

$$S = \{(y_1,y_2) \mid y_1 > 0,\ y_2 < 0\} \cup \{(y_1,y_2) \mid y_1 < 0,\ y_2 > 0\}.$$

Notice that this set is nonconvex and disconnected. Straightforward calculation yields

$$\phi(y) = \begin{cases} \dfrac{2}{3}\dfrac{(y_1)^{3/2}}{(-y_2)^{1/2}} & \text{if } y_1 > 0,\ y_2 < 0 \\[2ex] \dfrac{2}{3}\dfrac{(-y_1)^{3/2}}{(y_2)^{1/2}} & \text{if } y_1 < 0,\ y_2 > 0. \end{cases}$$

It is possible to verify that the functions ϕ_1^* and ϕ_2^* defined by

$$\phi_1^*(x) = \sup_{\substack{y_1>0 \\ y_2<0}}\{x_1'y_1 + x_2'y_2 - \phi(y)\}$$

$$\phi_2^*(x) = \sup_{\substack{y_1<0 \\ y_2>0}}\{x_1'y_1 + x_2'y_2 - \phi(y)\}$$

satisfy

$$\phi_1^*(x) = f(x) = 0 \qquad \text{if } x_1 > 0,\ \tfrac{1}{3}x_1^3 = x_2$$

$$\phi_2^*(x) = f(x) = 0 \qquad \text{if } x_1 < 0,\ \tfrac{1}{3}x_1^3 = x_2.$$

In the preceding examples we have that for each $y \in S$ there is unique (x,λ) satisfying (4). As a result the value of ϕ at each $y \in S$ does not depend on (x,λ). In the following example the situation is different.

Example 3: For $n = 2$ let

$$f(x) = f(x_1,x_2) = x_1 + x_2, \quad h(x) = \cos x_1 - x_2 \ \text{ for } |x_2| \leq 1.$$

Let $\bar{x} = (\bar{x}_1, \bar{x}_2)'$ be any vector satisfying $\cos \bar{x}_1 = \bar{x}_2$. Then (1) is written as

$$-\bar{\lambda} \cos \bar{x}_1 z_1^2 > 0, \quad \forall z \neq 0, \; z_1 \sin \bar{x}_1 + z_2 = 0.$$

The above relation is satisfied for $\bar{\lambda} \cos \bar{x}_1 < 0$ and a conjugate can be defined locally at each point $\bar{y}$ of the form

$$\bar{y} = \begin{bmatrix} 1-\bar{\lambda} \sin \bar{x}_1 \\ 1-\bar{\lambda} \end{bmatrix} \text{ with } \bar{\lambda} \cos \bar{x}_1 < 0, \; \cos \bar{x}_1 = \bar{x}_2. \tag{10}$$

Thus a local conjugate can be defined at any point in the set

$$S = \{(y_1, y_2) \mid |y_1 - 1| < |y_2 - 1|\}.$$

However to each $\bar{y} \in S$ there correspond more than one pair $(\bar{x}, \bar{\lambda})$ satisfying (10). For example if $\bar{y} = (\frac{1}{2}, 2)$ then (10) is satisfied for

$$\bar{\lambda} = -1, \; \bar{x}_1 = 2k\pi - \frac{\pi}{6}, \; k = \text{integer}, \; \bar{x}_2 = \frac{\sqrt{3}}{2}.$$

We have for the local conjugate at $(2k\pi - \frac{\pi}{6}, \frac{\sqrt{3}}{2}, -1)$

$$\phi(\bar{y}) = \bar{y}'\bar{x} - f(\bar{x}) = \frac{1}{2}(2k\pi - \frac{\pi}{6}) + \sqrt{3} - (2k\pi - \frac{\pi}{6}) - \frac{\sqrt{3}}{2}$$

or

$$\phi(\frac{1}{2}, 2) = \frac{\sqrt{3}}{2} - k\pi + \frac{\pi}{12}$$

and thus the value of ϕ depends on the integer k, i.e., the point $(\bar{x}, \bar{\lambda})$ which is used in the local definition of ϕ.

The following proposition provides essential characterizations of the function ϕ. Part 2) in particular shows that by conjugation on ϕ one obtains the original function f for points near $\bar{x}$ which lie on the manifold $\{x \mid h(x) = 0\}$.

<u>Proposition 2</u>: Let f be an h-locally convex function at $\bar{x}$. Consider a vector $\bar{\lambda}$ satisfying (1) and let ϕ be the h-local convex conjugate of f at $(\bar{x}, \bar{\lambda})$ defined by (6). Then:

1) ϕ is convex and twice continuously differentiable in $S(\bar{y};\varepsilon)$ and for all $y \in S(\bar{y};\varepsilon)$ we have

$$\nabla \phi(y) = x(y) \tag{11}$$

$$\nabla^2 \phi(y) = [\tilde{L}(y)]^{-1} - [\tilde{L}(y)]^{-1} \nabla h[x(y)] \cdot \{\nabla h[x(y)]'[\tilde{L}(y)]^{-1} \nabla h[x(y)]\}^{-1} \cdot \nabla h[x(y)]'[\tilde{L}(y)]^{-1} \tag{12}$$

where $\tilde{L}(y)$ equals any symmetric invertible matrix such that

$$\tilde{L}(y)z = L(y)z, \quad \forall z, \quad \nabla h[x(y)]'z = 0 \tag{13}$$

$$L(y) = \nabla^2 f[x(y)] + \sum_{i=1}^{m} \lambda_i(y) \nabla^2 h_i[x(y)]. \tag{14}$$

(<u>Note</u>: We show that $\nabla^2 \phi(y)$ is uniquely defined in this manner. In particular one may take $\tilde{L}(y)$ equal to $L(y)$ whenever $L(y)$ is invertible.)

2) Let $\tilde{f} : R^n \to R$ be defined by

$$\tilde{f}(x) = \sup_{y \in S(\bar{y};\varepsilon)} \{x'y - \phi(y)\}. \tag{15}$$

Then there exists a $\delta > 0$ such that

$$\tilde{f}(x) = f(x), \quad \forall x \in S(\bar{x};\delta), \; h(x) = 0. \tag{16}$$

<u>Proof</u>: 1) We have for $y \in S(\bar{y};\varepsilon)$

$$y - \nabla f[x(y)] - \nabla h[x(y)]\lambda(y) = 0 \tag{17}$$

$$h[x(y)] = 0 \tag{18}$$

$$\nabla \phi(y) = x(y) + \nabla x(y)\{y - \nabla f[x(y)]\} \tag{19}$$

where the $n \times n$ matrix $\nabla x(y)$ has as columns the gradients of the coordinates of x with respect to y. From (17) we obtain

$$\nabla x(y)\{y - \nabla f[x(y)]\} = \nabla x(y) \nabla h[x(y)]\lambda(y) \tag{20}$$

while from (18) we have by differentiation

$$\nabla x(y) \nabla h[x(y)] = 0. \tag{21}$$

Combining (19)-(21) we obtain

$$\nabla \phi(y) = x(y)$$

and (11) is proved. From the above equality we also obtain

$$\nabla^2 \phi(y) = \nabla x(y). \tag{22}$$

Differentiating (17) with respect to y we obtain

$$I - \nabla \lambda(y) \nabla h[x(y)]' - \nabla x(y) L(y) = 0 \tag{23}$$

where I is the $n \times n$ identity matrix, $\nabla \lambda(y)$ is the $n \times m$ matrix having as columns the gradients $\nabla \lambda_i(y)$, and $L(y)$ is given by (14). We have

$$z'L(y)z > 0, \quad \forall z \neq 0, \; \nabla h[x(y)]'z = 0. \tag{24}$$

Let $\tilde{L}(y)$ be any symmetric invertible matrix such that

$$L(y)z = \tilde{L}(y)z, \quad \forall z, \quad \nabla h[x(y)]'z = 0. \tag{25}$$

For example we can take $\tilde{L}(y) = L(y)$ if $L(y)$ is invertible. Another choice is given by

$$\tilde{L}(y) = L(y) + r \nabla h[x(y)] \nabla h[x(y)]'. \tag{26}$$

This matrix, in view of (24), is positive definite for $r > 0$ sufficiently large by a well known result. From (21) and (25) we have

$$\nabla x(y) L(y) = \nabla x(y) \tilde{L}(y) \tag{27}$$

and we can write (23) as

$$I - \nabla \lambda(y) \nabla h[x(y)]' - \nabla x(y) \tilde{L}(y) = 0.$$

Postmultiplying this relation with $[\tilde{L}(y)]^{-1} \nabla h[x(y)]$ and using (21) we obtain

$$[\tilde{L}(y)]^{-1} \nabla h[x(y)] - \nabla \lambda(y) \nabla h[x(y)]'[\tilde{L}(y)]^{-1} \nabla h[x(y)] = 0$$

from which

$$\nabla \lambda(y) = [\tilde{L}(y)]^{-1} \nabla h[x(y)] \cdot \{\nabla h[x(y)]'[\tilde{L}(y)]^{-1} \nabla h[x(y)]\}^{-1}. \tag{28}$$

Combining (22), (23), (27), and (28) we obtain

$$\nabla^2 \phi(y) = [\tilde{L}(y)]^{-1} - [\tilde{L}(y)]^{-1} \nabla h[x(y)] \cdot \{\nabla h[x(y)]'[\tilde{L}(y)]^{-1} \nabla h[x(y)]\}^{-1} \cdot \nabla h[x(y)][\tilde{L}(y)]^{-1} \tag{29}$$

and the desired relation (12) is proved. In order to show that ϕ is convex in $S(\bar{y};\varepsilon)$ it is sufficient to show that $\nabla^2\phi(y)$ is positive semidefinite. But this follows from (28) since $L(y)$ can be taken to be a positive definite matrix (for example of the form (26) for r sufficiently large) and hence $\nabla^2\phi(y)$ is a projection matrix.

2) We first observe that we have for each $y \in S(\bar{y};\varepsilon)$

$$\nabla[x'y-\phi(y)] = x-\nabla\phi(y) = x-x(y).$$

Hence y attains the supremum of $x'y-\phi(y)$ if $x=x(y)$ and it follows that

$$\tilde{f}[x(y)] = x(y)'y-\phi(y) = x(y)'y-y'x(y) + f[x(y)] \tag{30}$$

and finally

$$\tilde{f}[x(y)] = f[x(y)] \qquad \forall y \in S(\bar{y};\varepsilon).$$

Hence in order to prove part 2) it will be sufficient to show that there exists a $\delta > 0$ such that for each $x \in S(\bar{x};\delta)$ with $h(x)=0$ there is a $y \in S(\bar{y};\varepsilon)$ such that $x = x(y)$. The set

$$\{(x,\lambda) \mid [\nabla f(x)+\nabla h(x)\lambda] \in S(\bar{y};\varepsilon)\}$$

is clearly open and contains $(\bar{x},\bar{\lambda})$. Take $\delta > 0$ such that (x,λ) belongs to this set for all $x \in S(\bar{x};\delta)$, $\lambda \in S(\bar{\lambda};\delta)$ and

$$z'[\nabla^2 f(x)+\sum_{i=1}^{m}\lambda_i\nabla^2 h_i(x)]z > 0, \qquad \forall z \neq 0 \quad \nabla h(x)'z = 0$$

and furthermore $\delta \leq \sigma$ where σ is the scalar introduced in the application of the implicit function theorem following (4). Let x be such that $x \in S(\bar{x};\delta)$, $h(x)=0$ and let λ be any vector in $S(\bar{\lambda};\varepsilon)$. Then if $y = \nabla f(x)+\nabla h(x)\lambda$ we clearly have $y \in S(\bar{y};\varepsilon)$ and (x,λ) form a local maximum-Lagrange multiplier pair for problem (3). Since $\delta \leq \sigma$, by the uniqueness part of the implicit function theorem we obtain $x = x(y)$, $\lambda = \lambda(y)$. Q.E.D.

3. LOCAL VERSIONS OF FENCHEL'S DUALITY THEOREM

We now develop two results which may be viewed as analogs of the Fenchel Duality Theorem within our setting.

Consider the problem

$$\text{minimize } f_1(x)+f_2(x) \qquad \text{subject to } h(x)=0 \tag{31}$$

where $f_1 : R^n \to R$, $f_2 : R^n \to R$, $h : R^n \to R^m$, $h = (h_1,\ldots,h_m)'$. Let $\bar{x}$ be a local minimum of this problem and consider the following strengthened second order sufficiency assumptions.

A1) $f_1, f_2, h_1, \ldots, h_m$ are twice continuously differentiable in a neighborhood of $\bar{x}$, the gradients $\nabla h_1(\bar{x}),\ldots,\nabla h_m(\bar{x})$ are linearly independent and hence there exists a unique Lagrange multiplier vector $\bar{\lambda}$ such that

$$\nabla f_1(\bar{x})+\nabla f_2(\bar{x})+\nabla h(\bar{x})\bar{\lambda} = 0.$$

A2) The matrix $\nabla^2 f_2(\bar{x})$ is positive definite and there holds

$$z'[\nabla^2 f_1(\bar{x})+\sum_{i=1}^{m}\bar{\lambda}_i\nabla^2 h_i(\bar{x})]z > 0, \qquad \forall z \neq 0, \quad \nabla h(\bar{x})'z = 0.$$

We note that assumption A2) is stronger than the usual second order sufficiency assumption

$$z'[\nabla^2 f_1(\bar{x})+\nabla^2 f_2(\bar{x})+\sum_{i=1}^{m}\bar{\lambda}_i\nabla^2 h_i(\bar{x})]z > 0, \qquad \forall z \neq 0, \quad \nabla h(\bar{x})'z = 0.$$

Our stronger assumption is necessary for our results.

Now under A1) and A2) it follows that f_1 and f_2 are h-locally convex at $\bar{x}$ and locally convex at $\bar{x}$ respectively. The h-local conjugate of f_1, denoted ϕ_1, is defined in a sphere $S(\bar{y};\varepsilon_1)$ centered at

$$\bar{y} = \nabla f_1(\bar{x})+\nabla h(\bar{x})\bar{\lambda} = -\nabla f_2(\bar{x})$$

while the local conjugate of f_2, denoted ϕ_2, is defined in a sphere $S(-\bar{y};\varepsilon_2)$ centered at $-\bar{y}$. The function

$$\phi(y) = \phi_1(y)+\phi_2(-y)$$

is defined and is convex in the sphere $S(\bar{y};\varepsilon)$ where $\varepsilon = \min\{\varepsilon_1,\varepsilon_2\}$. We have $\nabla\phi_1(\bar{y}) = \nabla\phi_2(-\bar{y}) = \bar{x}$ and hence

$$\nabla\phi(\bar{y}) = \nabla\phi_1(\bar{y})-\nabla\phi_2(-\bar{y}) = 0.$$

We also have

$$\phi(\bar{y}) = \phi_1(\bar{y})+\phi_2(-\bar{y}) = \bar{y}'\bar{x}-f_1(\bar{x})-\bar{y}'\bar{x}-f_2(\bar{x}) = -[f_1(\bar{x})+f_2(\bar{x})].$$

It follows that $\bar{y}$ minimizes ϕ and $-\phi(\bar{y})$ equals the optimal value of problem (31). The above developments are summarized in the following proposition.

Proposition 3: Consider problem (31) under assumptions A1) and A2). Then there exist scalars $\delta > 0$ and $\varepsilon > 0$ such that

$$\min_{\substack{h(x)=0\\ x\in S(\bar{x};\delta)}} \{f_1(x)+f_2(x)\} = -\min_{y\in S(\bar{y};\varepsilon)} \{\phi_1(y)+\phi_2(-y)\}$$

where ϕ_1 is the h-local conjugate of f_1 at $(\bar{x},\bar{\lambda})$ and ϕ_2 is the local conjugate of f_2 at $\bar{x}$. The vector $\bar{x}$ attains the minimum on the left while the vector $\bar{y} = \nabla f_1(\bar{x})+\nabla h(\bar{x})\bar{\lambda} = -\nabla f_2(\bar{x})$ attains the minimum on the right.

Similarly one may develop a related result for the problem

$$\text{minimize } f_1(x)+f_2(x) \qquad \text{subject to } h_1(x)=0,\ h_2(x)=0 \tag{32}$$

where $f_1 : R^n \to R$, $f_2 : R^n \to R^{m_1}$, $h_1 = (h_{11},\ldots,h_{1m_1})' : R^n \to R^{m_1}$, $h_2 = (h_{21},\ldots,h_{2m_2})' : R^n \to R^{m_2}$. Let $\bar{x}$ be a local minimum for this problem and consider the following strengthened second order sufficiency assumptions:

B1) f_1, f_2, h_1, h_2 are twice continuously differentiable in a neighborhood of $\bar{x}$, the gradients $\nabla h_{11}(\bar{x}), \ldots, \nabla h_{1m_1}(\bar{x})$, $\nabla h_{21}(\bar{x}), \ldots, \nabla h_{2m_2}(\bar{x})$ are linearly independent and hence there exist unique Lagrange multiplier vectors $\bar{\lambda}_1$, $\bar{\lambda}_2$ such that

$$\nabla f_1(\bar{x}) + \nabla f_2(\bar{x}) + \nabla h_1(\bar{x})\bar{\lambda}_1 + \nabla h_2(\bar{x})\bar{\lambda}_2 = 0.$$

B2) There holds

$$z'[\nabla^2 f_1(\bar{x}) + \sum_{i=1}^{m_1} \bar{\lambda}_{1i} \nabla^2 h_{1i}(\bar{x})]z > 0, \quad \forall z \neq 0, \ \nabla h_1(\bar{x})'z = 0$$

$$z'[\nabla^2 f_2(\bar{x}) + \sum_{i=1}^{m_2} \bar{\lambda}_{2i} \nabla^2 h_{2i}(\bar{x})]z > 0, \quad \forall z \neq 0, \ \nabla h_2(\bar{x})'z = 0.$$

Under B1) and B2) it follows that f_1 and f_2 are h_1-locally convex at $\bar{x}$ and h_2-locally convex at $\bar{x}$ respectively. The h_1-local conjugate of f_1, denoted ϕ_1, is defined in a sphere $S(\bar{y};\epsilon_1)$ while the h_2-local conjugate of f_2, denoted ϕ_2, is defined in a sphere $S(-\bar{y};\epsilon_2)$ where

$$\bar{y} = \nabla f_1(\bar{x}) + \nabla h_1(\bar{x})\bar{\lambda} = -\nabla f_2(\bar{x}) - \nabla h_2(\bar{x})\bar{\lambda}.$$

Similarly as earlier we obtain the following proposition.

Proposition 4: Consider problem (32) under assumptions B1) and B2). Then there exist scalars $\delta > 0$ and $\epsilon > 0$ such that

$$\min_{\substack{h_1(x)=0 \\ h_2(x)=0 \\ x \in S(\bar{x};\delta)}} \{f_1(x) + f_2(x)\} = - \min_{y \in S(\bar{y};\epsilon)} \{\phi_1(y) + \phi_2(-y)\}$$

where ϕ_1 is the h_1-local conjugate of f_1 at $(\bar{x}, \bar{\lambda}_1)$ and ϕ_2 is the h_2-local conjugate of f_2 at $(\bar{x}, \bar{\lambda}_2)$. The vector $\bar{x}$ attains the minimum on the left while the vector $\bar{y} = \nabla f_1(\bar{x}) + \nabla h_1(\bar{x})\bar{\lambda}_1 = -\nabla f_2(\bar{x}) - \nabla h_2(\bar{x})\bar{\lambda}_2$ attains the minimum on the right.

We note that similar results may be obtained for problems with inequality constraints. Also one may obtain alternative statements of Propositions 3 and 4 which are more in line with the standard statement of Fenchel's duality theorem (see e.g. Rockafellar (1970) p. 327) by introducing notions of locally concave functions and local conjugate concave functions. Such results can be patterned along the lines of the developments of this paper and those of the classical theory under convexity and can be obtained in a routine manner.

REFERENCES

Bertsekas, D. P. (1978). Convexification Procedures and Decomposition Methods for Nonconvex Optimization Problems. J.O.T.A., (to appear).

Kort, B. W., and D. P. Bertsekas. (1976). Combined Primal-Dual and Penalty Methods for Convex Programming. SIAM J. on Control and Optimization, Vol. 14, pp. 268-294.

Luenberger, D. G. (1973). Introduction to Linear and Nonlinear Programming. Addison-Wesley, Reading, Mass.

Rockafellar, R. T. (1970). Convex Analysis. Princeton Univ. Press, Princeton, N.J.