EE 381V: Large Scale Optimization **Fall 2012**

Lecture 18 — November 1

Lecturer: Caramanis & Sanghavi *Scribe: Christopher Hadlock, Bismark Singh*

18.1 Lecture Overview

We ultimately wish to develop algorithms which are tailored to: unconstrained optimization problems with non-smooth objective functions; constrained optimization problems; problems with special structures. In this lecture, we examine non-smooth functions. In particular, we cover the following: The Legendre-Fenchel Transform; Sub-differentials and Sub-gradients; Epigraphs.

While it might seem as if non-smooth functions are rarely encountered in practice, this is not the case. As a common example, Figure 14.1 illustrates the maximum of smooth functions as a non-smooth function.

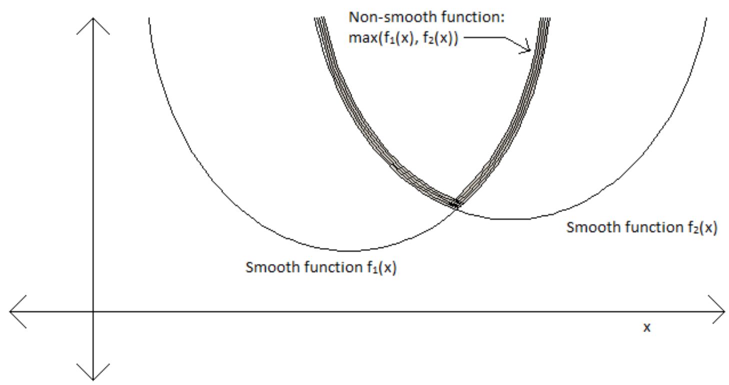

Figure 18.1. Example of a Non-Smooth Function – The Maximum of Two Smooth Functions

18.2 The Legendre-Fenchel Transform

Before introducing the Legendre-Fenchel Transform, some background concepts are in order.

18.2.1 Epigraphs and Semicontinuity

Recall that if C is a closed convex set, then we may represent C as the intersection of all halfspaces which contain it. That is,

$C = \cap H^+$, for all $H^+ \supseteq C$

In a similar fashion, we can represent a convex function as the supremum of all hyperplanes lying below f. Figure 14.2 provides a graphical illustration of this concept. Mathematically speaking, if f is a convex function, and h an arbitrary hyperplane, then we express this as follows.

$f(x) = \sup h(x)$, where h(x) is any hyperplane satisfying $h(x) \leq f(x)$

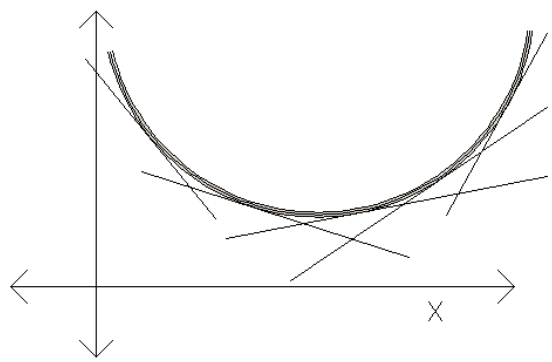

Figure 18.2. Representing a Convex Function as a Supremum of Hyperplanes Lying Below it

Recall that the epigraph of a function f is defined as follows.

$f : R^n \to R$ $epi(f) = \{(x, y) : f(x) \leq y\}$

Loosely speaking, the epigraph of a function f is the set of all points lying above f (See Figure 14.3)

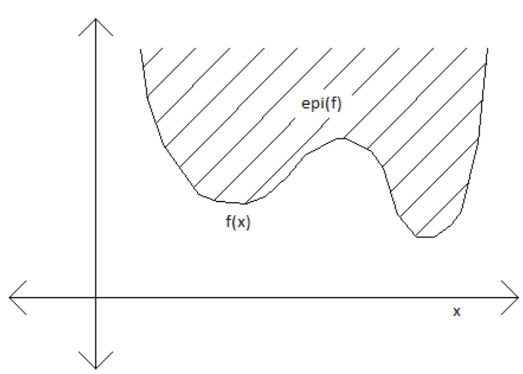

Figure 18.3. The Epigraph of Arbitrary Function f(x)

Theorem 18.1. *The epigraph of a function f is convex if and only if f is convex.*

Proof: Let f be convex and let (u,a),(v,b) $\in epif$. Then, $\forall \lambda \in [0, 1]$,

$$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda) f(v) \leq \lambda a + (1-\lambda) b \tag{18.1}$$

and hence $\lambda(u,a) + (1-\lambda a)(v,b) \in epif$ Conversely assume that epi f is convex. We verify the convexity of f on its domain. Let $u, v \in domf$ such that a $\geq f(u)$ and $b \geq f(v)$. Since $\lambda(u,a) + (1-\lambda)(v,b) \in epif \forall \lambda \in [0,1]$ we have that, such that $a \geq f(u)$ and $b \geq (v)$. Since $(u,a) + (v,b) \in epif$ for every $\lambda \in [0,1]$ it follows that $f(\lambda u + (1-\lambda)v) \leq \lambda a + (1-\lambda)b$. Now we can choose a and b as f(u), f(v) respectively and the proof is complete. □

Note that the epigraph of a function may not necessarily be a closed set, as shown in Figure 14.4. This motivates us to introduce the concept of lower semi continuity..

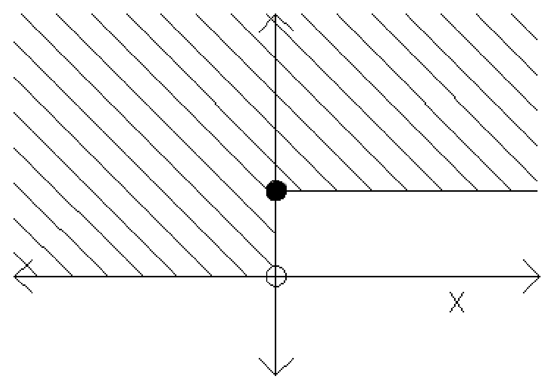

Figure 18.4. Example of a Function Whose Epigraph is Open

$f : R^n \to$ R$\cup\infty$ is lower semicontinuous at a point x for every sequence of points x_i converging to x one has, if $\liminf_{i\to\infty} f(x_i) \geq f(x)$ The above definition of lower semicontinuity is mathematically equivalent to declaring that the epigraph of f is closed.

Theorem 18.2. *A function f :$R^n \to R \cup \infty$ is lower semicontinuous iff its epigraph is closed*

Proof: Lsc $\to$ closed epigraph. Let t,x be the limit of sequence $\{t_i, x_i\} \subset$ Epi(f), then we have $t_i \geq f(x_i)$. Thus the following holds, $t = \lim_{i\to\infty} t_i \geq \lim_{i\to\infty} f(x_i) \geq \lim_{i\to\infty} f(x)$

Closed Epigraph $\to$ Lsc. Suppose that $f(x) > \gamma > \lim_{i\to\infty} f(x_i)$ for some non semicontinuous γ where x_i converges to x. Then $\exists$ a subsequence $\{x_i\}_\kappa$ such that $f(x_i) \leq \gamma \forall i \in \kappa$. Since the epigraph is closed then x must belong to this set giving f(x) $\leq \gamma$, which is a contradiction. □

It can be seen that the above example does not satisfy this property.

Fact: F is lower semicontinuous if and only if f is pointwise limit of monotone sequence of continuous functions $f(x)_n \to f(x); f(x)_n \leq f(x)_{n+1}$.

But the above is not so useful to us. What is useful is that the following are equivalent:

1. f is lsc
2. epi f is closed
3. sublevel sets of f are all closed

We conclude that epi f is closed convex when f is lsc and convex.

Lemma 18.3. *Let f be convex. If $f(x) > -\infty$ for some x, then $f(x) > -\infty$ everywhere.*

Proof: This follows trivially from the definition of convexity. If any of the terms on the RHS of Eq 14.1 are $-\infty$ then all terms must be $-\infty$. □

Further note that since epi f is convex and ∃ atleast one point where f(x) is not infinite , so ∃ atleast one affine function h s.t. $h \leq f$ (this follows from the Separation theorem - if there is a convex set and a point outside it then a non vertical hyperplane separates it). The 'non vertical' part is important here. Even in the most extreme cases we can express all points in the exterior of F as separable by non vertical affine functions(hyperplane). This is demonstrated via the following figure.

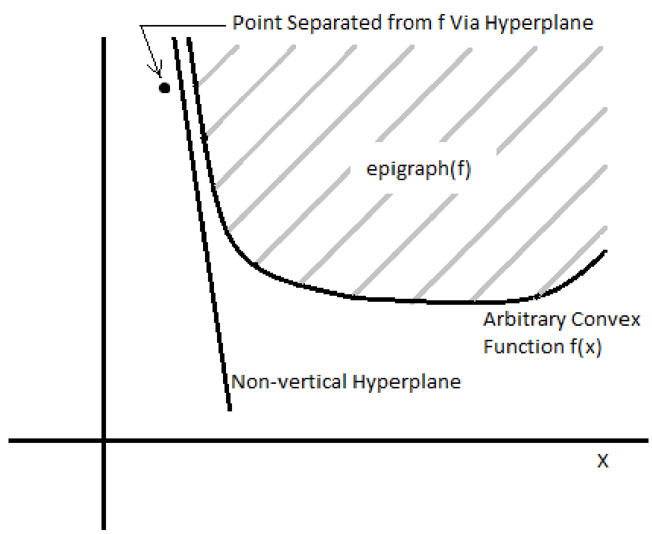

Figure 18.5. Convex Epigraph/Point Separation Via Non-Vertical Hyperplane

Observe that,
1. The domain of f has a relative interior, i.e. it is non empty
2. epi f $= \cap H^-$ where $H^- \supseteq$ epi f and H is non negative.

This gives us that epi f= ∩ h where h is an affine function and h(x) ≤ f(x) ∀ x.

Thus to study the funcion f or to study the set of affine functions associated with it is equivalent.

Now we have the necessary background to introduce the Legendre-Fenchel Transform.

18.2.2 Introducing the Legendre-Fenchel Transform

Recall that in the previous section we stated we can represent a convex function f as the supremum over all hyperplanes lying below f. That is:

$f(x) = \sup h(x)$, where h(x) is any hyperplane satisfying $h(x) \le f(x)$

For any such hyperplane h(x) satisfying the above equation, we may paramaterize the h(x) as follows:

$h(x) =< s, x > -\beta$, where s and β are scalars

Now, let us define $F^* = \{(s, \beta) :< s, x > -\beta \le f(x) \forall x \in R^n\}$

That is, F^* represents the set of all hyperplanes lying below the function f. Now, we claim that F^* is the epigraph of some convex function f^*. Knowing f gives us F^* and hence f^*. In particular, note the following,

$(s, \beta) \in F^* \Leftrightarrow< s, x > -\beta \le f(x) \forall x \Leftrightarrow \sup\{< s, x > -f(x)\} \le \beta \forall x.$

Thus, $\sup < s, x > -f(x) \le \beta$

or $f^*(s) \le \beta$ where s is the quantity inside the sup of the previous expression.

Now, $f^*(s) \le \beta \Rightarrow F^* = epi(f^*)$

Theorem 18.4. *Both $f^* and F^*$ are closed and convex*

Proof: f^* is the sup over linear functions, and since the max of convex functions is again a convex function(refer to PS 2 question 4) , f^* is again convex. Also since f^* is the intersection of closed sets, it is in turn closed.

For F^* we note that it is the epi f^*, and the result follows from Theorem 14.1 □

We define,

$$f : R^n \to R : f^*(x^*) = \sup\{< x^*, x > -f(x)\} \tag{18.2}$$

as the Legendre-Fenchel transform (a.k.a. conjugate), where the supremum is taken over all x. This x^* must not be confused with our (previous) notation of x^* being the optimal of a function.

Theorem 18.5. *If the function f is closed and convex, then $f^{**} = f$*

Proof: We provide a proof for the one dimensional case. For n dimensions,it can be easily extended. Maximizing the RHS of Equation 14.2 wrt x we obtain ,$x = f^{'-1}(x^*)$. Substituting back we get, $f^*(x^*) = (x^*)f^{'-1}(x^*) - (x^*)$.

Now note that $f^{**} = \sup(kx - f^*(k))$, where the sup is taken over all k. Maximizing the RHS of this expression analogously to the previous, we obtain the desired result. Note that our proof rests on the assumption that f admits a supporting line at x and that f^* is differentiable at k. In general we have that f^{**} is the largest convex function satisfying $f^{**}(x) \le f(x)$ which gives with the definition of the convex hull of a function. Figure 14.5 below provides a graphical illustration.

ALITER : The Legendre transformation of f_* is

$$\sup_{d \in R^n} (x^T)d - f_*(d) = \sup_{d \in R^n, a \geq f^*(d)} (d^T x - a) \tag{18.3}$$

This is exactly the sup of all supporting hyperplanes of f. $\because (a \geq f^*(d) \iff (d^T x - a))$. the upper bound of all the affine hyperplanes of f is the closure of f, and hence if f is proper , the result follows.

$\square$

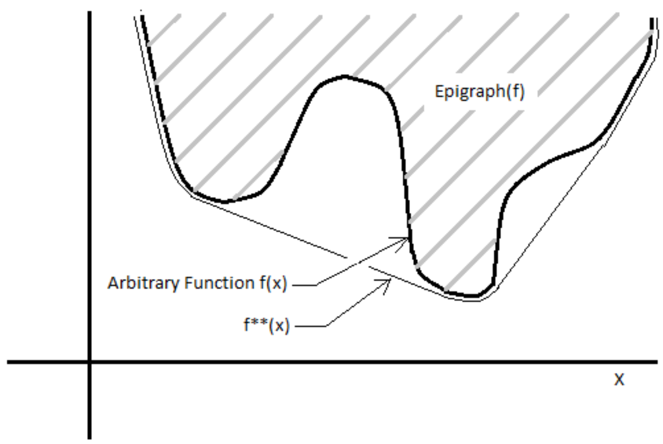

Figure 18.6. Graphical Illustration of f(x) and f**(x)

Now, we claim that $< x, x^* > \leq f(x) + f^*(x^*)$. However, for which points x^* do we achieve equality? We again look at the smooth case and develop an analogy. Note that if the function f is continuous, then the vector normal to the tangent plane at the point (x, f(x)) is simply $(\nabla f(x), -1)$. We look at non vertical hyperplanes in R^n.

$H = [x :< s, x >= r]$

And for n+1 dimensions we write explicitly,

$$H = [x, x_{n+1} :< s, s_{n+1}, x, x_{n+1} >= r] \tag{18.4}$$

so for the non vertical case we should have $s_{n+1} \neq 0$. We can scale the above equation to get, $< (s, -1), (x, x_{n+1}) >= r$ However, if the function f is not smooth at some point x, then we will have a set of tangent planes at (x, f(x)) lying below the f. This motivates the definition of a subdifferential.

Definition: The set of x^* such that $< x^*, x >= f(x) + f^*(x^*)$ is called the subdifferential of f at x: $\partial f(x)$

$y \in \partial f(x) \Leftrightarrow x \in \partial f^*(y) \Leftrightarrow < x, y >= f(x) + f^*(y)$.

We can describe the concept of the subdifferential in an alternate manner as follows. Recall that if the function f is continuous and convex, then $f(y) \geq f(x) + < \nabla f(x), y - x >$

$\forall y$. By an alternative definition, the subdifferential $\partial f(x)$ is the set of vectors x^* such that: $f(y) \geq f(x) + < x^*, y - x > \forall y$

Returning to the setting of unconstrained optimization problems, recall that when f is continuous, then $\nabla f(x) = 0$ at the minimum. However, for the general case in which f is not necessarily smooth everywhere, we have $0 \in \partial f(x)$ at the minimum.

Similarly for the case of constrained minization, we had for smooth functions $0 \in \nabla f(x) + N_\chi(x)$. Now analogously we have,$0 \in \partial f(x) + N_\chi(x)$.

Recall that in the case of unconstrained optimization problems involving a smooth objective function, all descent algorithms used the gradient in some way. We want to develop an algorithm in an analogous was again but remember that -the subdifferential being a set- the algorithm must work for all elements of the set, and not a specific element alone. We relied on two propoerties in the smooth case:

1. $-\nabla f(x)$ is a descent direction
2. $\nabla f(x) \to 0$ as $x \to x^*$.

However, in the case where the objective function is not smooth everywhere, it is possible that neither of these hold. Infact, we can achieve a worse (greater) objective value with these holding.

As an example, consider the function: $f(x) = |x_1| + 2 * |x_2|$. What is the subdifferential at the point (1,0)? We observe: $\partial|x| = sign(x)$ if $x \neq 0$, and $\partial|x| = z$, where $-1 \leq z \leq 1, if x = 0$. Thus, for f as defined in our example, we have $(1,1) \in \partial f(1,0)$, and $x_+ = x - \varepsilon * \nabla f(x) \Rightarrow (1,0) - \varepsilon * (1,1) = (1-\varepsilon, \varepsilon)$. Now, $f(1-\varepsilon, \varepsilon) = 1 - \varepsilon + 2 * |-\varepsilon| = 1 + \varepsilon > 1$. Thus, we see how this example illustrates how taking a step in the direction of the negative gradient(Property 1 of above) can lead to a worse objective value. Property (2) is easily seen to not hold always as there are elements in the set of the subgradient which may not go to 0 as x approaches the optimal.

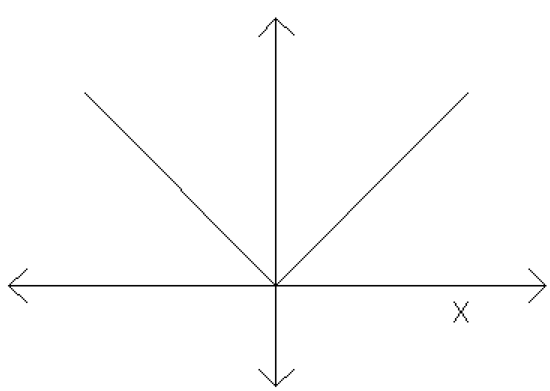

Figure 18.7. Plot of $f(x_1, 0$ Versus x)