ORIE 6300 Mathematical Programming I	November 29, 2016
Lecture 26	
Lecturer: Damek Davis	*Scribe: Benjamin Grimmer*

1 Last Time

Recall that for $\gamma > 0$, the Moreau envelope is defined through the minimization problem $e_{\gamma g}(x) = \inf_y \left(g(y) + \frac{1}{2\gamma}\|x - y\|^2 \right)$. Last time we proved a number of properties showing that $e_{\gamma g}$ is well behaved. We proved the following proposition:

Proposition 1 *Let $x \in \mathbb{R}^n$. Then there exists a unique point denoted by $\operatorname{prox}_{\gamma g}(x)$ such that*

1. $\{\operatorname{prox}_{\gamma g}(x)\} = \operatorname{argmin}_y \left\{ g(y) + \frac{1}{2\gamma}\|x - y\|^2 \right\}$;
2. $\operatorname{Fix}(\operatorname{prox}_{\gamma g}) = \operatorname{argmin}_y \{g(y)\}$;
3. $\operatorname{prox}_{\gamma g}$ *is firmly non-expansive.*

Then a direct application of the KM convergence theorem gives the following result.

Corollary 2 *Suppose $\operatorname{argmin}_y\{g(y)\} \neq \emptyset$. Then the proximal point algorithm converges to a minimizer of g.*

2 Smoothness of Moreau Envelope

Theorem 3 *$e_{\gamma g}$ is C^1 and for all $x \in \mathbb{R}^n$, $\nabla e_{\gamma g}(x) = \frac{1}{\gamma}(x - \operatorname{prox}_{\gamma g}(x))$. Consequently, the Moreau envelope has a $1/\gamma$ Lipschitz continuous gradient.*

Proof: Let $x, y \in \mathbb{R}^n$. Then the Theorem will follow if we show the following:

$$\lim_{y \to x} \frac{e_{\gamma g}(y) - e_{\gamma g}(x) - \langle y - x, \frac{1}{\gamma}(x - \operatorname{prox}_{\gamma g}(x))\rangle}{\|y - x\|} = 0$$

As a shorthand, we will denote $p_x = \operatorname{prox}_{\gamma g}(x)$ and $p_y = \operatorname{prox}_{\gamma g}(y)$. From optimality conditions, we know that $(1/\gamma)(x - p_x) \in \partial g(p_x)$. Then we find the following lower bound.

$$
\begin{aligned}
e_{\gamma g}(y) - e_{\gamma g}(x) &= g(p_y) + \frac{1}{2\gamma}\|p_y - y\|^2 - g(p_x) - \frac{1}{2\gamma}\|p_x - x\|^2 \\
&\geq \frac{1}{2\gamma}\Big(2\langle p_y - p_x, x - p_x\rangle + \|y - p_y\|^2 - \|x - p_x\|^2\Big) \\
&= \frac{1}{2\gamma}\Big(2\langle (p_y - y) - (p_x - x), x - p_x\rangle + 2\langle y - x, x - p_x\rangle + \|y - p_y\|^2 - \|x - p_x\|^2\Big) \\
&= \frac{1}{2\gamma}\Big(\|(p_y - y) - (p_x - x)\|^2 + \|x - p_x\|^2 - \|y - p_y\|^2 \\
&\qquad\qquad + 2\langle y - x, x - p_x\rangle + \|y - p_y\|^2 - \|x - p_x\|^2\Big) \\
&= \frac{1}{2\gamma}\Big(\|(p_y - y) - (p_x - x)\|^2 + 2\langle y - x, x - p_x\rangle\Big) \\
&\geq \frac{1}{\gamma}\langle y - x, x - p_x\rangle
\end{aligned}
$$

We can apply the same argument swapping the roles of x and y to find a symmetric bound:

$$
e_{\gamma g}(y) - e_{\gamma g}(x) \leq \langle y - x, y - p_y\rangle
$$

Combining these two bounds, we get the following:

$$
\begin{aligned}
0 \leq e_{\gamma g}(y) - e_{\gamma g}(x) - \frac{1}{\gamma}\langle y - x, x - p_x\rangle &\leq \frac{1}{\gamma}\langle y - x, (y - p_y) - (x - p_x)\rangle \\
&= \frac{1}{\gamma}\langle y - x, (y - x) - (p_y - p_x)\rangle \\
&= \frac{1}{2\gamma}\Big(\|y - x\|^2 + \|(y - p_y) - (x - p_x)\|^2 - \|p_x - p_y\|^2\Big) \\
&\leq \frac{1}{\gamma}\|y - x\|^2
\end{aligned}
$$

Then the Theorem follows from dividing the above equation by $\|y - x\|$ and applying the squeeze theorem as $y \to x$. □

As an immediate consequence of this Theorem, we have the following corollary by recalling the optimality condition for $\text{prox}_{\gamma g}(\cdot)$ (i.e. $(1/\gamma)(x - \text{prox}_{\gamma g}(x)) \in \partial g(\text{prox}_{\gamma g}(x))$).

Corollary 4 $\nabla e_{\gamma g}(x) = \frac{1}{\gamma}(x - \text{prox}_{\gamma g}(x)) \in \partial g(\text{prox}_{\gamma g}(x))$

Example. Let $C \subseteq \mathbb{R}^n$ be a closed convex set. Then the function $e_{\iota_C}(x) = \frac{1}{2}d_C^2(x)$ is smooth. Further, we know that $\nabla(\frac{1}{2}d_C^2(x)) = x - P_C(x) \in \partial\iota_C(P_C(x)) = N_C(P_C(x))$.

Question. Why is the Moreau envelope smooth?

We can define the addition of two sets C_1 and C_2 as $C_1 + C_2 = \{a_1 + a_2 | a_1 \in C_1 a_2 \in C_2\}$. Then we observe that adding a smooth set to one with corners will always produce a smooth set. In essence, the smoothness is inherited from either of the underlying sets. An instructive example is given by the sum of a triangle and a disk in the plane. Now to relate this smoothness to the Moreau envelope, we decompose the epigraph of the Moreau envelope into the sum of two simpler epigraphs.

Theorem 5 *(epigraph addition formula)* $\mathrm{epi}(e_{\gamma g}) = \mathrm{epi}(g) + \mathrm{epi}(\frac{1}{2\gamma}\|\cdot\|^2)$

Proof: First we prove the $\subseteq$ inclusion. Let $(x,t) \in \mathrm{epi}(e_{\gamma g})$. Then we know the following:

$$e_{\gamma g}(x) = g(\mathrm{prox}_{\gamma g}(x)) + \frac{1}{2\gamma}\|x - \mathrm{prox}_{\gamma g}(x)\|^2 \le t = t_1 + t_2$$

where $(\mathrm{prox}_{\gamma g}(x), t_1) \in \mathrm{epi}(g)$ and $(x - \mathrm{prox}_{\gamma g}(x), t_2) \in \mathrm{epi}\left(\frac{1}{2\gamma}\|\cdot\|^2\right)$.

Now we prove the $\supseteq$ inclusion. Let $(x_1, t_1) \in \mathrm{epi}(g)$ and $(x_2, t_2) \in \mathrm{epi}(\frac{1}{2\gamma}\|\cdot\|^2)$. Let $x := x_1 + x_2$. Then we know the following:

$$\begin{aligned}
t_1 + t_2 &\ge g(x_1) + \frac{1}{2\gamma}\|x_2\|^2 \\
&\ge \inf_{y+z=x}\left\{g(y) + \frac{1}{2\gamma}\|z\|^2\right\} \\
&= \inf_{y}\left\{g(y) + \frac{1}{2\gamma}\|y - x\|^2\right\} \\
&= e_{\gamma g}(x).
\end{aligned}$$

□

This decomposition of the epigraph of the Moreau envelope shows how its smoothness is inherited from the smoothness of $\|\cdot\|^2$.

3 How Well Does The Moreau Envelope Approximate $g(x)$?

Observe that setting γ near zero will cause the quadratic term to dominate the envelope. Thus we find that as $\gamma \to 0$, we will have $e_{\gamma g}(x) = g(x)$. Conversely, as $\gamma \to \infty$, we will have $e_{\gamma g}(x) = \inf g(y)$. This motivates the question. How good of an approximation of $g(x)$ is the Moreau envelope for a fixed γ. In the case of Lipschitz functions, we can give a guarantee of its quality.

Theorem 6 *(Constant approximation for Lipschitz functions) Let g be an L-Lipschitz function.*

$$0 \le g(x) - e_{\gamma g}(x) \le L^2\gamma$$

Proof: Notice that $v_p := (1/\gamma)(x - \mathrm{prox}_{\gamma g}(x)) \in \partial g(\mathrm{prox}_{\gamma g}(x))$. Let $v \in \partial g(x)$. By question 3 in Homework 10, we know that $\|v\|, \|v_p\| \le L$. Moreover, we know that the Moreau envelope is always a lower bound. Thus, we know the following:

$$\begin{aligned}
0 &\le g(x) - e_{\gamma g}(x) \\
&\le g(x) - g(\mathrm{prox}_{\gamma g}(x)) - \frac{1}{2\gamma}\|\mathrm{prox}_{\gamma g}(x)\|^2 \\
&\le \langle v, x - \mathrm{prox}_{\gamma g}(x)\rangle - \frac{1}{2\gamma}\|\mathrm{prox}_{\gamma g}(x)\|^2 \\
&\le \|v\|\|v_p\|\gamma \\
&\le L^2\gamma
\end{aligned}$$

□

Finally, note that if g is L-Lipschitz, then its Moreau envelope is also L-Lipschitz.