ORIE 6300 Mathematical Programming I November 22, 2016

Lecture 25

Lecturer: Damek Davis *Scribe: Johan Bjorck*

Throughout the lecture we assume that all functions are closed, convex, and proper. We first show a result about the subgradients of a sum of functions

Proposition 1 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be functions, where f is differentiable at $x \in \mathrm{dom}(g)$. Then $\partial(f+g)(x) = \nabla f(x) + \partial g(x)$.*

Proof: We first show "$\subseteq$". Let $v \in \partial(f+g)(x)$ then we have

$$\begin{aligned}
\forall y \in \mathbb{R}^n \qquad & f(y) + g(y) \geq f(x) + g(x) + \langle v, y-x\rangle \\
\implies & g(y) \geq g(x) + (f(x) - f(y)) + \langle v, y-x\rangle \\
& = g(x) + \Big(\langle -\nabla f(x), x-y\rangle - o(y-x) + \langle v, y-x\rangle\Big) \\
& = g(x) - o(y-x) + \langle v - \nabla f(x), y-x\rangle
\end{aligned}$$

If $\mathrm{dom}(g) = \{x\}$ then $g(y) = \infty$ for $y \neq x$, which would give $g(y) \geq g(x) + \langle v - \nabla f(x), y-x\rangle$. If on the other hand $y \in \mathrm{dom}(g)$ and $y \neq x$, then $\forall \epsilon > 0$ we define $y_\epsilon = \epsilon y + (1-\epsilon)x \in \mathrm{dom}(g)$. Thus

$$g(y_\epsilon) \leq \epsilon g(y) + (1-\epsilon) g(x) \implies g(y) \geq \frac{1}{\epsilon} g(y_\epsilon) - \frac{1-\epsilon}{\epsilon} g(x)$$

Hence we have

$$\begin{aligned}
g(y) &\geq \frac{1}{\epsilon} g(y_\epsilon) + \left(1 - \frac{1}{\epsilon}\right) g(x) \\
&\geq \frac{1}{\epsilon} g(x) + \left(1 - \frac{1}{\epsilon}\right) g(x) + \frac{1}{\epsilon}\langle v - \nabla f(x), y_\epsilon - x\rangle - \frac{1}{\epsilon} o(y_\epsilon - x) \\
&\geq g(x) + \langle v - \nabla f(x), y-x\rangle - \frac{\epsilon o(y-x)}{\epsilon} \\
&\overset{\epsilon \to 0}{\to} g(x) + \langle v - \nabla f(x), y-x\rangle
\end{aligned}$$

Thus $v - \nabla f(x) \in \partial g(x)$.
Let us now prove the other inclusion "$\supseteq$". Let $v \in \partial g(x)$, then

$$\begin{aligned}
(\forall y \in \mathbb{R}^n) \qquad & f(y) \geq f(x) + \langle \nabla f(x), y-x\rangle \\
& g(y) \geq g(x) + \langle v, y-x\rangle
\end{aligned}$$

Thus we clearly see that $(f+g)(y) \geq (f+g)(x) + \langle v + \nabla f(x), y-x\rangle$, and then $\nabla f(x) + v \in \partial(g+f)(x)$. □

With the basics of convex calculus we can now start developing algorithms. On Homework 10, we developed the projected subgradient method. In practice, such subgradient methods are much slower than smooth gradient methods. Thus, we now develop a general way to *smooth* convex functions, in order to apply fast methods from smooth optimization.

Definition 1 *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function, for every $\gamma > 0$ we define the Moreau envelope as*

$$e_{\gamma g}(x) = \inf_y \left\{ g(y) + \frac{1}{2\gamma}\|x - y\|^2 \right\}$$

The moreau envelope trades off minimization and proximity through the parameter γ. We now show the uniqueness of the point obtain through the moreau envelope

Proposition 2 *Let $x \in \mathbb{R}^n$, then there exists a unique point $\text{prox}_{\gamma,g}$ such that*

$$\text{prox}_{\gamma,g} := \text{argmin}_y \left\{ g(y) + \frac{1}{2\gamma}\|x - y\|^2 \right\}.$$

Proof: Because g is convex $\exists a \in \mathbb{R}^n, b \in \mathbb{R}$ s.t. $\forall y \in \mathbb{R}^n$ we have $g(y) \geq \langle a, y \rangle + b$ (Exercise!). Thus

$$(\forall y \in \mathbb{R}^n) \qquad g(y) + \frac{1}{2\gamma}\|x - y\|^2 \geq \langle a, y \rangle + b + \frac{1}{2\gamma}\|x - y\|^2 \geq \text{const}_{x,a,b}$$

Thus $e_{\gamma g} > -\infty$, note also that $e_{\gamma g} < \infty$ as $g \neq \infty$ so $\exists x_0 \in \mathbb{R}^n$ such that $\infty > g(x_0) + \|x - x_0\|^2/(2\gamma) > e_{\gamma,f}$.

For existence, $\exists \{x^i\}_{i \in N}$ such that $g(x^i) + \frac{1}{2\gamma}\|x^i - x\|^2 \to e_\gamma(x)$. We now claim that $\{x^i\}_{i \in N}$ is bounded, let us quickly prove this claim.

We have

$$\begin{aligned} g(x^i) + \frac{1}{2\gamma}\|x^i - x\|^2 &\geq \langle a, x^i \rangle + b + \frac{1}{2\gamma}\|x^i - x\|^2 \\ &= \langle a, x \rangle + b + \langle a, x^i - x \rangle + \frac{1}{2\gamma}\|x - x^i\|^2 \\ &= \langle a, x \rangle + b + \frac{1}{2\gamma}\|x^i - x + \gamma a\|^2 - \frac{\gamma}{2}\|a\|^2 \end{aligned}$$

Thus, $\|x^i - x + \gamma a\|^2 \leq \max_{j \in \mathbb{N}}\{g(x^j) + \frac{1}{2\gamma}\|x^j - x\|^2\} - \langle a, x \rangle + b + \frac{\gamma}{2}\|a\|^2 < \infty$, which is finite because $\{g(x^j) + \frac{1}{2\gamma}\|x^j - x\|^2\}_{j \in \mathbb{N}}$ is a convergent sequence. Thus $\{x^i\}$ is bounded and we have proved our claim.

Now, WLOG, assume that $x^i \to y \in \mathbb{R}^n$, then by lower semicontinuity of $g + \frac{1}{2\gamma}\|\cdot - x\|^2$ we have

$$e_{\gamma g}(x) = \lim_{i \to \infty} \left[g(x^i) + \frac{1}{2\gamma}\|x^i - x\|^2 \right] \geq g\left(\lim_{i \to \infty} x^i\right) + \frac{1}{2\gamma}\left\|\lim_{i \to \infty} x^i - x\right\|^2 = g(y) + \frac{1}{2\gamma}\|x - y\|^2.$$

Thus $e_{\gamma g} = g(y) + \frac{1}{2\gamma}\|x - y\|^2$. To show the uniqueness, note that if $g(y) + \frac{1}{2\gamma}\|x - y\|^2 = e_{\gamma g}(x)$ then

$$0 \in \partial g(y) + \frac{1}{\gamma}(y - x) \implies \frac{1}{\gamma}(x - y) \in \partial g(y).$$

Thus $\forall z \in \mathbb{R}^n$ we have

$$\begin{aligned} g(z) &\geq g(y) + \left\langle \frac{1}{\gamma}(x - y), z - y \right\rangle \\ &= g(y) + \frac{1}{2\gamma}\Big(\|x - y\|^2 + \|y - z\|^2 - \|x - z\|^2 \Big) \\ \implies g(z) + \frac{1}{2\gamma}\|z - x\|^2 &\geq g(y) + \frac{1}{2\gamma}\Big(\|x - y\|^2 + \|z - y\|^2 \Big). \end{aligned}$$

So $g(z) + \frac{1}{2\gamma}\|z - x\|^2 = e_{\gamma g}(x)$ is equivalent to $z = y$, which proves uniqueness. □

Note that we in the proof we showed that $e_{\gamma g}$ is finite and always sits beneath $g(x)$. As a corollary of the above results we learn that

Corollary 3 Fix $\left(\text{prox}_{\gamma g}\right) = \text{argmin}(g)$

Proof:

$$\text{prox}_{\gamma,g}(x) = x \iff x = \text{argmin}\left\{g(y) + \frac{1}{2\gamma}\|x-y\|^2\right\} \iff 0 \in \partial g(x) + \frac{1}{\gamma}(x-x) = \partial g(x).$$

□

As an example take C to be a closed convex set, then we have

$$e_{\gamma\iota_C}(x) = \text{argmin}_y \left\{\iota_C(y) + \frac{1}{2\gamma}\|y-x\|^2\right\} = \text{argmin}_{y\in C}\left\{\frac{1}{2\gamma}\|y-x\|^2\right\} = \frac{1}{2\gamma}\text{dist}_C^2(x)$$

The unique y such that $e_{\gamma\iota_C}(x) = \text{argmin}_{y\in C}\frac{1}{2}\|y-x\|^2$ is just $\text{prox}_{\gamma\iota_C}(x) = P_C(x)$.

We have previously used the firm-non-expansiveness to develop efficient algorithms, and it turns out that $\text{prox}_{\gamma g}$ is also firmly nonexpansive. To show firm nonexpansiveness, we first need the following result

Lemma 4 (Subgradient Monotonicity) *Let $x, y \in \text{dom}(g)$ and take $u \in \partial g(x)$, $v \in \partial g(y)$, then the following holds*

$$\langle x-y, u-v\rangle \geq 0$$

Proof: We have

$$g(y) \geq g(x) + \langle x-y, u\rangle$$
$$g(x) \geq g(y) + \langle y-x, v\rangle$$

Adding these two together yields the result . □

Theorem 5 *The mapping $\text{prox}_{\gamma g}$ is firmly nonexpansive.*

Proof: Let $x, y \in \mathbb{R}^n$, the we have

$$\text{prox}_{\gamma g}(x) = \text{argmin}_z \left\{g(z) + \frac{1}{2\gamma}\|x-z\|^2\right\}$$

so $0 \in \partial g(\text{prox}_{\gamma g}(x)) + \frac{1}{\gamma}(\text{prox}_{\gamma g}(x) - x)$, equivalently $\frac{1}{\gamma}(x - \text{prox}_{\gamma g}(x)) \in \partial g(\text{prox}_{\gamma g}(x))$. The same holds for y, and thus

$$\begin{aligned}
0 &\leq \left\langle \frac{1}{\gamma}(x-\text{prox}_{\gamma g}(x)) - \frac{1}{\gamma}(y - \text{prox}_{\gamma g}(y)), \text{prox}_{\gamma g}(x) - \text{prox}_{\gamma g}(y)\right\rangle \\
&= \frac{1}{\gamma}\langle x - y - (\text{prox}_{\gamma,g}(x) - \text{prox}_{\gamma,g}(y)), \text{prox}_{\gamma,g}(x) - \text{prox}_{\gamma,g}(y)\rangle \\
&= \frac{1}{2\gamma}\Big(- \|(x-\text{prox}_{\gamma,g}(x)) - (y - \text{prox}_{\gamma,g}(y))\|^2 - \|\text{prox}_{\gamma,g}(x) - \text{prox}_{\gamma,g}(y)\|^2 + \|x-y\|^2\Big)
\end{aligned}$$

Thus we have

$$\|\text{prox}_{\gamma,g}(x) - \text{prox}_{\gamma,g}(y)\|^2 + \|(x - \text{prox}_{\gamma,g}(x)) - (y - \text{prox}_{\gamma,g}(y))\|^2 \leq \|x-y\|^2.$$

□

Corollary 6 *Assume* $\text{argmin}\{g(x)\} \neq \emptyset$*, then the sequence $\{x_i\}_{i\in N}$ converges to a minimizer of g where $x_0 \in \mathbb{R}^n$ is arbitrary and*

$$x_{k+1} = \text{prox}_{\gamma,g}(x_k).$$

For proving this we simply apply the KM iteration to $2\text{prox}_{\gamma,g} - I$.