| | |
|---|---|
| **ORIE 6300 Mathematical Programming I** | November 15, 2016 |
| Lecture 22 | |
| *Lecturer: Damek Davis* | *Scribe: Lijun Ding* |

# 1 Last time

1. A point $\bar{x}$ is stationary for the problem $\inf_{x\in C} f(x)$ if it satisfies

$$-\nabla f(\bar{x}) \in N_C(\bar{x}) \quad (OPT).$$

2. The following theorem follows from the relation between the projection and the normal cone.

   **Theorem 1** *$\bar{x}$ satisfies OPT if and only if*

$$(\forall r > 0) \qquad \bar{x} = P_C(\bar{x} - \gamma \nabla f(\bar{x})).$$

3. The projected gradient can be summerized as follows. Let $L$ be the Lipschitz constant of the gradient $\nabla f$.

   *Input:* $x^0 \in C$ , $0 < \gamma < \frac{2}{L}$
   *For* $k = 0, 1, \ldots,$ *do*
   $x^{k+1} = P_C(x^k - \gamma \nabla f(x^k))$.

4. The following theorem asserts that OPT is a sufficient and necessary condtion under differentiability of $f$ and the convexity of $f$ and $C$.

   **Theorem 2** *Let $f$ be differentiable and convex, and $C \in \mathbb{R}^n$ be a closed convex set. Then*

$$-\nabla f(x) \in N_C(x) \iff x \in \arg\min_{y\in C} f(y).$$

5. Using the above theorem, we were able to develop necessary and sufficient conditions for optimality of $\inf_{x\mathbb{R}^n} \{(f+g)(x)\}$.

   **Theorem 3** *Suppose, $f$ and $g$ are convex, $f$ differentiable and $g$ continuous, then*

$$\bar{x} \in \arg\min\{f(x) + g(x)\} \iff 0 \in \nabla f(\bar{x}) + \partial g(\bar{x}).$$

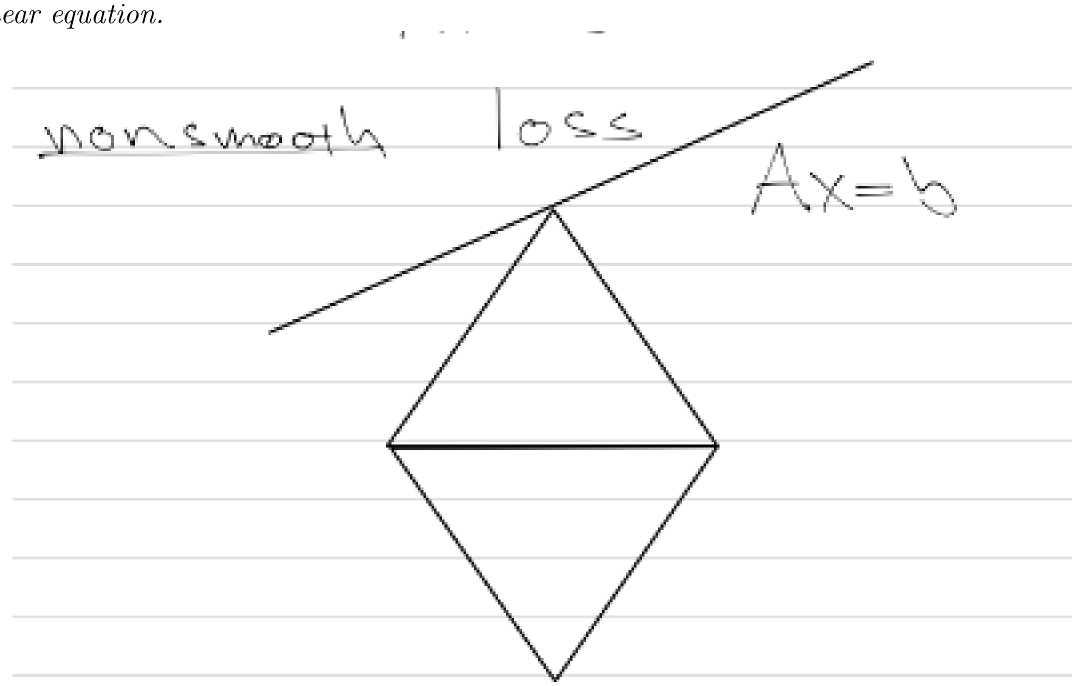
## 2 Nonsmooth Convex functions

Nonsmoothness is essential to accurately expressive modeling in applied science.

**Example 1** *The goal of compressed sensing is to solve the following problem:*

$$\min \|x\|_1$$
$$\textit{subject to: } Ax = b$$

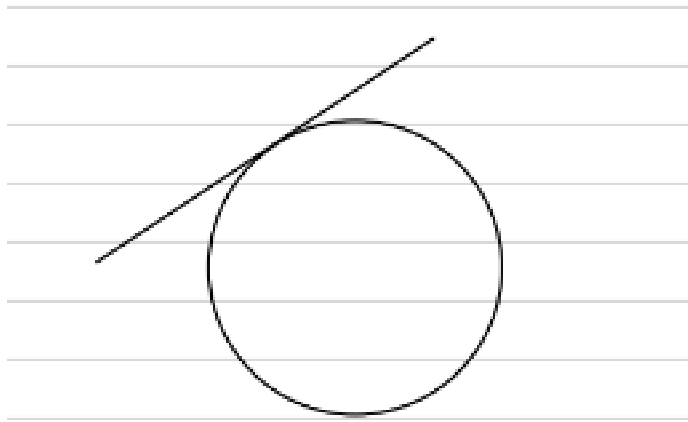
*The loss function is nonsmooth. As the following figure indicates, random affine spaces are are highly likely to hit the corners of the $\ell_1$ ball. These corners are precisely sparse solutions to the linear equation.*




*In contrast, solutions to the $\ell_2$ minimization problem*

$$\min \|x\|_2$$
$$\textit{subject to:} Ax = b,$$

*are not likely to be sparse. Indeed, random affine spaces are equally likely to touch the $\ell_2$ ball at any point on the boundary.*



The following two question come up when we want to solve nonsmooth optimization problems.

- How do we differentiate nonsmooth problems? (Today)
- How do we perform "projected gradient" descent on nonsmooth problems?

First a piece of notation:

**Notation 1** *Throughout the rest of the course, we will denote* $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$.

In nonsmooth optimization, it is convenient to consider extended real valued functions, i.e., functions which possibly take on the value $\infty$. Such functions are able to implicitly enforce constraints such as $x \in C$ by taking on infinite values outside of $C$. In that spirit, we introduce the notation

$$\mathrm{dom}(g) := \{x \in \mathbb{R}^n \mid g(x) < \infty\}.$$

Thus, we allow the function $g \equiv \infty$, but since $\mathrm{dom}(g) = \emptyset$, it doesn't seem like a proper function. This motivates the terminology.

**Definition 1** *A function* $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is called* *proper* *if*

$$\exists x \in \mathbb{R}^n, s.t.\ g(x) < \infty$$

Thus, proper functions are exactly the functions with nonempty domain.

**We will only deal with convex sets throughout the rest of the course.** The definition of convexity for extended real-valued functions is exactly the same as the standard definition.

**Definition 2 (Convexity)** *A function* $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is called convex if*

$$(\forall \alpha \in [0,1]), (\forall x, y \in \mathbb{R}^n) \qquad g((1-\alpha)x + \alpha y) \leq (1-\alpha)g(x) + \alpha g(y).$$

Functions which take on the value $\infty$ cannot be continuous. Thus, we need to introduce a new notion of continuity tailored to $\infty$-valued functions.

**Definition 3** *A function $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called lower-semicontinuous, or closed, if*

$$\mathrm{epi}(g) := \{(x,t) \in \mathbb{R}^{n+1} | g(x) \leq t\}$$

*is a closed set.*

We arrive at the above definition of lower-semicontinuity because it is analogous to the following property of continuous functions:

**Exercise 1** *The function $g : \mathbb{R}^n \to \mathbb{R}$ is continuous if, and only if, $g$ is everywhere finite and $\mathrm{gph}(g) = \{(x, g(x)) \mid x \in \mathbb{R}^n\}$ is a closed set.*

One useful property of continuity is that continuous functions commute with limits:

$$\lim_{k\to\infty} x_k = x \implies \lim_{k\to\infty} g(x_k) = g(x)$$

Lower-semicontinuous functions do not commute with limits in general, but they do satisfy another useful property:

**Exercise 2** *Let $g : \mathbb{R} \to \overline{\mathbb{R}}$, let $x \in \mathbb{R}^n$, and suppose that $\{x_k\}_{k\in\mathbb{N}}$ is a sequence with the property that $\lim_{k\to\infty} x_k = x$. Prove that*

$$\liminf g(x_k) \geq g(x).$$

The above result shows that $\mathrm{epi}(g)$ can have sudden downward drops, such as from $\infty$ to 0.

We give some examples of lower-semicontinuous functions below. You should convince yourself about the assertions in the examples.

**Example 2** *Let $C \subset \mathbb{R}^n$ closed. Then the indicator function*

$$\iota_C(x) = \begin{cases} \infty & x \notin C; \\ 0 & x \in C. \end{cases}$$

*is closed. If $C$ is conex, then $\iota_C$ is convex.*

**Example 3** *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be closed and convex and $A \in \mathbb{R}^{n\times m}$. Then $g(Ax)$ is convex and closed.*

**Example 4** *The sum of closed convex functions is convex.*

**Example 5** *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$, be a function such that for all $y \in \mathbb{R}^n$, the function $f(\cdot, y)$ is closed and convex. Then $g(x) := \sup_y f(x, y)$ is closed and convex.*

**Example 6** *The following is a list of convex functions that often appear in real optimization problems: $\exp(x), \log(1+\exp(x)), \frac{1}{2}\|Ax-b\|^2, \|x\|_p$ for $p \in [1,\infty], -\sqrt{x}, -\log(x),$, $\max\{0, 1-x\}$ (hinge).*

**Example 7** *Example of a function which is convex, but not closed:*

$$g(x,y) = \begin{cases} 0 & x^2+y^2 < 1 \\ \phi(x,y) & x^2+y^2 = 1, \end{cases}$$

*where $\phi$ is an arbitrary nonnegative function. The function $g$ may not be always closed.*

Note that $\mathrm{dom}(g)$ may be open even if $\mathrm{epi}(g)$ is closed. Indeed, consider the function defined by $g(x) = \frac{1}{x}$ if $x > 0$ and $g(x) = \infty$ otherwise.

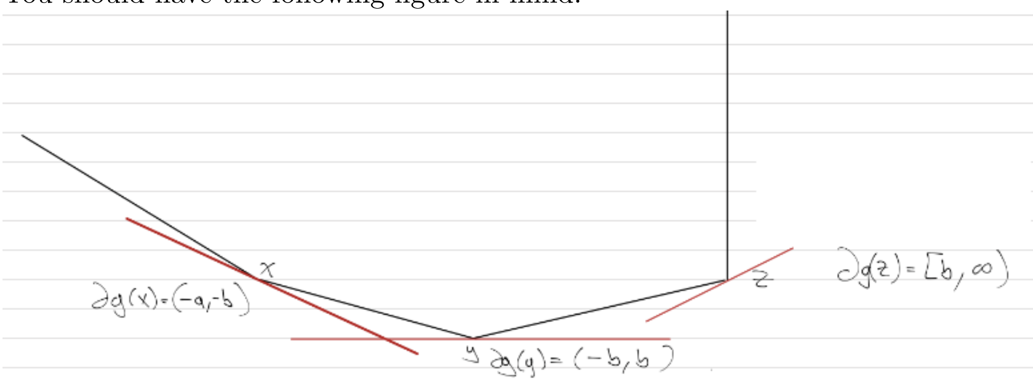
## 3 Differentiating nonsmooth $\infty$-valued functions

**Definition 4** *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. The set*

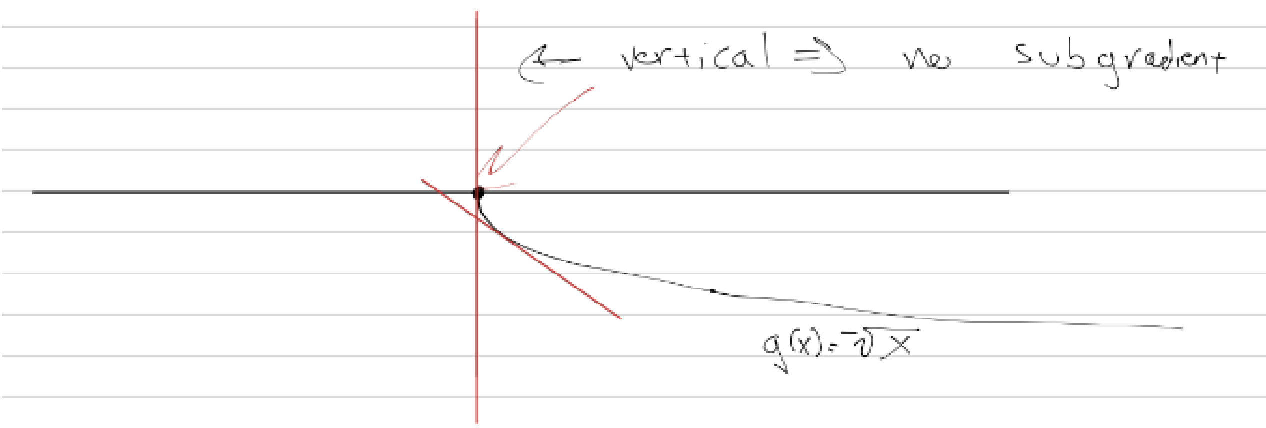
$$\partial g(x) := \{v \in \mathbb{R}^n | (\forall y)\ g(y) \geq g(x) + \langle v, y - x\rangle\}$$

*is called the convex subdifferential operator of $g$. Elements $v \in \partial g(x)$ are called subgradients.*

You should have the following figure in mind:




Even for convex functions, subgradients do not necessarily exist. Indeed, consider $-\frac{1}{\sqrt{x}}$ at $x = 0$:




**Example 8** *Let $C$ be closed and convex. Then $\partial\iota_C(x) = N_C(x)$*

Thus, we have been studying subgradients since week 3! On Homework 3, problem 3, you showed that

$$\begin{bmatrix} v \\ -1 \end{bmatrix} \in N_{\mathrm{epi}(g)}((x, g(x))) \iff v \in \partial g(x).$$

In that exercise, we assumed that $g$ is convex and continuous, but that condition can be relaxed to merely assuming that $g$ is convex and closed. So subdifferential operators are deeply connected to normal cones of epigraphs.

When

$$\begin{bmatrix} v \\ 0 \end{bmatrix} \in N_{\mathrm{epi}(g)}((x, g(x)))$$

we get horizontal normals (or vertical hyperplanes); see the the plot of $-\sqrt{x}$ at the point $(0,0)$.

These horizontal normals are called horizon subgradients.

**Definition 5** *For closed, convex $g : \overline{\mathbb{R}}^n \to \bar{\mathbb{R}}$, the* <u>*horizon subdifferential*</u> *is the set-valued operator*

$$\partial^\infty g(x) = \left\{ v \in \mathbb{R}^n \middle| \begin{bmatrix} v \\ 0 \end{bmatrix} \in N_{\mathrm{epi}(g)}(x, g(x)) \right\}.$$

**Exercise 3** *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be closed and convex. Show that for all $x \in \mathbb{R}^n$, $\partial^\infty g(x) = N_{\mathrm{dom}(g)}(x)$. (Note that there is a slight technicality here. The domain of $g$ is not necessarily closed, but we can still define its normal cone in the usual way. In particular, for points $x \in \mathrm{int}(\mathrm{dom}(g))$, we still have $N_{\mathrm{dom}(g)}(x) = \{0\}$.)*

We can completely characterize the normal cone of the epigraph through the convex and horizon subdifferentials.

**Theorem 4** *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be closed and convex. Then $\forall x \in \mathrm{dom}(g)$,*

$$N_{\mathrm{epi}(g)}((x, g(x))) = \left\{ \lambda \begin{bmatrix} v \\ -1 \end{bmatrix} \middle| v \in \partial g(x), \lambda > 0 \right\} \cup \left\{ \begin{bmatrix} v \\ 0 \end{bmatrix} \middle| v \in \partial^\infty g(x) \right\}.$$

**Proof:** From the the definition of the horizon subdifferential and the equivalence

$$\begin{bmatrix} v \\ -1 \end{bmatrix} \in N_{\mathrm{epi}(g)}((x, g(x)) \iff v \in \partial g(x),$$

we see we only need to that $\nexists v \in \mathbb{R}^n$ s.t.

$$\begin{bmatrix} v \\ 1 \end{bmatrix} \in N_{\mathrm{epi}(g)}((x, g(x))).$$

Suppose for contradicition that such $v$ exists. Then

$$0 \geq \left\langle \begin{bmatrix} v \\ 1 \end{bmatrix}, \begin{bmatrix} x \\ g(x) + 1 \end{bmatrix} - \begin{bmatrix} x \\ g(x) \end{bmatrix} \right\rangle = g(x) + 1 - g(x) = 1$$

which is a contradiction. $\square$

**Corollary 5** *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be closed and convex. Then*

1. $(\forall x \in \mathrm{dom}(g)) \qquad \partial g(x) \cup (\partial g(x) \backslash \{0\}) \neq \emptyset$
2. $(\forall x \in \mathrm{int}(\mathrm{dom}(g))) \qquad \partial g(x) \neq \emptyset.$

**Proof:**

1. Since $(x, g(x)) \in \mathrm{boundary}(\mathrm{epi}(g))$,

   $$N_{\mathrm{epi}(g)}((x, g(x))) \neq \{0\}.$$

   Then use the previous theorem.
2. Since $\partial^\infty g(x) = N_{\mathrm{dom}(g)}(x) = \{0\}$ by previous exercise, part 1 implies $\partial g(x) \neq \emptyset$.

$\square$

The above corollary tells us that subgradients (in a broader sense) exist at any any point in $\mathrm{dom}(g)$ and they are either vertical (horizon) or non vertical.