The University of Texas at Austin
Department of Electrical and Computer Engineering

**ECE 381V: Large-Scale Optimization II — Spring 2022**

Lecture 11

Caramanis & Mokhtari Wednesday, February 23, 2022

---

**Goal:** In this lecture, we first briefly mention some tips for computing subgradients and subdifferentials. Then, we talk about talk lower and upper bounds for solving convex nonsmooth optimization for both convex and strongly convex settings.

# 1 Subgradient Computation

**Definition 1.** *A vector* $\mathbf{g}$ *is called a subgradient of the function* $f$ *at point* $\mathbf{x}_0 \in dom(f)$*, if for any* $\mathbf{y} \in dom(f)$ *we have*

$$f(\mathbf{y}) \geq f(\mathbf{x}_0) + \mathbf{g}^\top(\mathbf{y} - \mathbf{x}_0).$$

Some useful tips for computing subdifferentials and subgradients

1. if $f$ is differentiable at $\hat{\mathbf{x}}$ we have $\partial f(\hat{\mathbf{x}}) = \{\nabla f(\hat{\mathbf{x}})\}$
2. if $f(\mathbf{x}) = \alpha_1 f_1(\mathbf{x}) + \alpha_2 f_2(\mathbf{x})$ then $\partial f(\hat{\mathbf{x}}) = \alpha_1 \partial f_1(\mathbf{x}) + \alpha_2 \partial f_2(\mathbf{x})$ (addition of sets)
3. if $f(\mathbf{x}) = h(\mathbf{A}\mathbf{x} + \mathbf{b})$ then $\partial f(\hat{\mathbf{x}}) = \mathbf{A}^\top \partial h(\mathbf{A}\mathbf{x} + \mathbf{b})$

**Example.** Consider the point-wise maximum function of $m$ functions defined as

$$f(\mathbf{x}) = \max_{i=1,\ldots,m} \{f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$$

Consider the active constraint $\mathcal{I}(\hat{\mathbf{x}}) = \{i \mid f_i(\hat{\mathbf{x}}) = f(\hat{\mathbf{x}})\}$ at $\hat{\mathbf{x}}$. Then, we have

$$\partial f(\mathbf{x}) = \text{convex hull of } \partial f_i(\mathbf{x}) \text{ where } i \in \mathcal{I}(\mathbf{x})$$

If the functions $f_i$ are differentiable, then

$$\partial f(\mathbf{x}) = \text{convex hull of } \nabla f_i(\mathbf{x}) \text{ where } i \in \mathcal{I}(\mathbf{x})$$

**Example.** Consider the following function

$$f(\mathbf{x}) = \|\mathbf{x}\|_1 = \max_{\mathbf{a} \in \{-1,1\}^n} \mathbf{a}^\top \mathbf{x}$$

then we have

$$\partial f(\mathbf{x}) = J_1 \times \cdots \times J_n \qquad J_i = \text{sign}(x(i)) = \begin{cases} 1 & \text{if } x(i) > 0 \\ -1 & \text{if } x(i) < 0 \\ \text{[-1,1]} & \text{if } x(i) = 0 \end{cases}$$

## 2 Lower Bound for the Convex Setting

We first formally define the class of functions, query/oracle, and the required approximation error.

**Model:** Unconstrained minimization where the objective function $f$ is convex on $\mathbb{R}^n$ and is Lipschitz on a bounded set, i.e., $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$, where $f$ is $G$-Lipschitz. More precisely, we assume that $\|\mathbf{x}_0 - \mathbf{x}^*\| \leq R$ and $f$ is $G$-Lipschitz over this bounded convex set.

**Oracle:** First-order Black Box: at each point $\hat{\mathbf{x}}$ we can compute $f(\hat{\mathbf{x}})$ and $\mathbf{g}(\hat{\mathbf{x}}) \in \partial f(\hat{\mathbf{x}})$.

**Approximation Solution:** Find $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that $f(\hat{\mathbf{x}}) - f^* \leq \epsilon$

**Methods:** Generate a sequence $\{\mathbf{x}_k\}$ according to $\mathbf{x}_k \in \mathbf{x}_0 + \text{Span}\{\mathbf{g}(\mathbf{x}_0), \ldots, \mathbf{g}(\mathbf{x}_{k-1})\}$

As described above, we assume that the initial point has a bounded distance from an optimal solution, i.e., $\|\mathbf{x}_0 - \mathbf{x}^*\| \leq R$, and function $f$ is $G$-Lipschitz over this set. Hence, we denote our class of problems as $\mathcal{P}(\mathbf{x}_0, R, G)$. In the following theorem, we establish a lower bound for this class of problems.

**Theorem 1.** *For any class $\mathcal{P}(\mathbf{x}_0, R, G)$ and any $1 \leq k \leq n-1$, there exists a function $f \in \mathcal{P}(\mathbf{x}_0, R, G)$, such that*

$$\min_{i=1,\ldots,k} f(\mathbf{x}_i) - f^* \;\geq\; \frac{GR}{2(2+\sqrt{k+1})}.$$

*for any optimization scheme which generates iterates according to $\mathbf{x}_k \in \mathbf{x}_0 + Span\{\mathbf{g}(\mathbf{x}_0), \ldots, \mathbf{g}(\mathbf{x}_{k-1})\}$.*

Before, proving this result let us define the function that we need for proving this result. Consider the following function

$$f_k(\mathbf{x}) = \gamma \max_{i=1,\ldots,k} x(i) + \frac{\mu}{2}\|\mathbf{x}\|^2$$

Note that as we will see $\gamma$ and $\mu$ will depend on $k$. Now, let us first compute the subdifferential of this function.

$$\partial f_k(\hat{\mathbf{x}}) = \mu\hat{\mathbf{x}} + \gamma \,\text{Conv}\{\mathbf{e}_i \mid i \in \mathcal{I}(\hat{\mathbf{x}})\},$$

where $\mathcal{I}(\hat{\mathbf{x}}) = \{j \mid 1 \leq j \leq k,\ x(j) = \max_{i=1,\ldots,k} x(i)\}$.
Hence, for any $\mathbf{x}, \mathbf{y} \in B(\mathbf{x}^*, \rho)$ we have

$$f_k(\mathbf{x}) - f_k(\mathbf{y}) \leq \mathbf{g}_k(\mathbf{y})^\top(\mathbf{y}-\mathbf{x}) \leq \|\mathbf{g}_k(\mathbf{y})\|\|\mathbf{y}-\mathbf{x}\| \leq (\mu\|\mathbf{y}\|+\gamma)\|\mathbf{y}-\mathbf{x}\| \leq (\mu\|\mathbf{x}^*\|+\mu\rho+\gamma)\|\mathbf{y}-\mathbf{x}\|$$

Hence, the function is $(\mu\|\mathbf{x}^*\| + \mu\rho + \gamma)$-Lipschitz. Further, we can show that $\mathbf{x}_k^*$ is given by

$$x_k^*(i) = -\frac{\gamma}{\mu k} \qquad 1 \leq i \leq k$$
$$x_k^*(i) = 0 \qquad k+1 \leq i \leq n$$

which implies that

$$\|\mathbf{x}_k^*\| = \frac{\gamma}{\mu\sqrt{k}}, \qquad f_k^* = -\frac{\gamma^2}{2\mu k}$$

and hence the function is Lipschitz with constant

$$G_k = \mu\|\mathbf{x}^*\| + \mu\rho + \gamma = \mu\rho + \gamma\frac{\sqrt{k}+1}{\sqrt{k}}$$

Now we need to assume something about the subgradient that the oracle returns. In this case, we assume

$$\mathbf{g}(\hat{\mathbf{x}}) = \mu\hat{\mathbf{x}} + \gamma\mathbf{e}_{i^*}$$

where $i^* = \min\{j \mid x(j) = \max_{i=1,\ldots,k+1} x(i)\}$. This means that every time that we query a subgradient, the coordinate selected is the first maximal component of the vector $\mathbf{x}$. This is a resisting oracle and it consists in providing us with the worst possible subgradient at each test point.
This way for $\mathbf{x}_k$, at most the first $k$ components are nonzero, if we start from $\mathbf{x}_0 = \mathbf{0}$. Hence, for $i \leq k-1$

$$f_k(\mathbf{x}_i) \geq \gamma \max_{i=1,\ldots,k} x(i) \geq 0$$

Now since we start form $\mathbf{x}_0 = \mathbf{0}$, if we define $f(\mathbf{x}) = f_{k+1}(\mathbf{x})$ and set

$$\gamma = \frac{\sqrt{k+1}G}{2+\sqrt{k+1}}, \qquad \mu = \frac{G}{(2+\sqrt{k+1})R}$$

the we have

$$f^* = f_k^* = -\frac{\gamma^2}{2\mu(k+1)} = -\frac{GR}{2(2+\sqrt{k+1})}, \qquad \|\mathbf{x}_0 - \mathbf{x}^*\| = \|\mathbf{x}_0 - \mathbf{x}_{k+1}^*\| = \frac{\gamma}{\mu\sqrt{k+1}} = R,$$

and the function $f$ is $G$-Lipschitz, since $G = \mu R + \gamma\frac{\sqrt{k+1}+1}{\sqrt{k+1}}$. Further, we know that for $i \leq k$

$$f(\mathbf{x}_i) = f_{k+1}(\mathbf{x}_k) \geq 0$$

Hence,

$$\min_{i=1,\ldots,k} f(\mathbf{x}_i) - f^* \geq -f^* = \frac{GR}{2(2+\sqrt{k+1})}.$$

# 3 Upper Bound for the Convex Setting: Subgradient Method

Consider the following general convex (nonsmooth) constrained problem:

$$\begin{aligned} &\min f(\mathbf{x}) \\ &\text{s.t } \mathbf{x} \in \mathcal{Q}, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex on $\mathbb{R}^n$ and the set $\mathcal{Q}$ is convex.

The update of subgradient method is formally given by

$$\mathbf{x}_{k+1} = \pi_{\mathcal{Q}}\left(\mathbf{x}_k - \eta_k\mathbf{g}_k\right)$$

where $\mathbf{g}_k \in \partial f(\mathbf{x}_k)$. We assume that $\mathbf{x}_0 \in \mathcal{Q}$ is given.

**Theorem 2.** *Let $f$ be a convex function that is $G$-Lipschitz continuous on the set $\|\mathbf{x}-\mathbf{x}^*\| \leq R$, and suppose $\|\mathbf{x}_0-\mathbf{x}^*\| \leq R$. Then, we have*

$$\sum_{i=0}^{k} \eta_i (f(\mathbf{x}_i) - f(\mathbf{x}^*)) \leq \frac{R^2 + G^2 \sum_{i=0}^{k} \eta_k^2}{2}$$

*Proof.* Note that we can show that

$$\begin{aligned}
\|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 &= \|\pi_{\mathcal{Q}}(\mathbf{x}_k - \eta_k \mathbf{g}_k) - \pi_{\mathcal{Q}}(\mathbf{x}^*)\|^2 \\
&\leq \|\mathbf{x}_k - \eta_k \mathbf{g}_k - \mathbf{x}^*\|^2 \\
&= \|\mathbf{x}_k - \mathbf{x}^*\|^2 - 2\eta_k \mathbf{g}_k^\top (\mathbf{x}_k - \mathbf{x}^*) + \eta_k^2 \|\mathbf{g}_k\|^2 \\
&\leq \|\mathbf{x}_k - \mathbf{x}^*\|^2 - 2\eta_k (f(\mathbf{x}_k) - f(\mathbf{x}^*)) + \eta_k^2 \|\mathbf{g}_k\|^2 \\
&\leq \|\mathbf{x}_k - \mathbf{x}^*\|^2 - 2\eta_k (f(\mathbf{x}_k) - f(\mathbf{x}^*)) + \eta_k^2 G^2
\end{aligned}$$

By regrouping the terms we obtain

$$2\eta_k (f(\mathbf{x}_k) - f(\mathbf{x}^*)) \leq \|\mathbf{x}_k - \mathbf{x}^*\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 + \eta_k^2 G^2$$

Now if we sum both sides for all steps we obtain

$$\sum_{i=0}^{k} 2\eta_i (f(\mathbf{x}_i) - f(\mathbf{x}^*)) \leq \|\mathbf{x}_0 - \mathbf{x}^*\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2 + \sum_{i=0}^{k} \eta_k^2 G^2 \leq \|\mathbf{x}_0 - \mathbf{x}^*\|^2 + G^2 \sum_{i=0}^{k} \eta_k^2$$

which implies the claim.

□

**Corollary 1.** *If we run the algorithm for $N$ steps and set the stepsize az $\eta_k = \frac{R}{G\sqrt{N+1}}$, then we have*

$$f\left(\frac{1}{N+1}\sum_{i=0}^{N} \mathbf{x}_i\right) - f(\mathbf{x}^*) \leq \frac{RG}{\sqrt{N+1}}$$

The proof simply follows from the fact that for convex functions we have

$$f\left(\frac{1}{N+1}\sum_{i=0}^{N} \mathbf{x}_i\right) \leq \frac{1}{N+1}\sum_{i=0}^{N} f(\mathbf{x}_i)$$

**Corollary 2.** *If set the stepsize az $\eta_k = \frac{R}{G\sqrt{k+1}}$, then we have*

$$\min_{i=0,\ldots,k} f(\mathbf{x}_i) - f(\mathbf{x}^*) \leq \mathcal{O}\left(\frac{RG\ln(k+1)}{\sqrt{k+1}}\right)$$

The proof simply follows from the fact that

$$\min_{i=0,\ldots,k} f(\mathbf{x}_i) - f(\mathbf{x}^*) \leq \frac{R^2 + G^2 \sum_{i=0}^{k} \eta_k^2}{2\sum_{i=0}^{k} \eta_i} = RG\left(\frac{1 + \sum_{i=0}^{k} \frac{1}{i+1}}{2\sum_{i=0}^{k} \frac{1}{\sqrt{i+1}}}\right) \leq \frac{RG(1+\ln(k+1))}{4\sqrt{k+1}}$$

# 4 Lower Bound for the Strongly Convex Setting

We first formally define the class of functions, query/oracle, and the required approximation error.

**Model:** Unconstrained minimization where the objective function $f$ is strongly convex on $\mathbb{R}^n$ with $\mu$ and is Lipschitz on a bounded set, i.e., $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$, where $f$ is $G$-Lipschitz. More precisely, we assume that $\|\mathbf{x}_0 - \mathbf{x}^*\| \leq R$ and $f$ is $G$-Lipschitz over this bounded convex set.

**Oracle:** First-order Black Box: at each point $\hat{\mathbf{x}}$ we can compute $f(\hat{\mathbf{x}})$ and $\mathbf{g}(\hat{\mathbf{x}}) \in \partial f(\hat{\mathbf{x}})$.

**Approximation Solution:** Find $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that $f(\hat{\mathbf{x}}) - f^* \leq \epsilon$

**Methods:** Generate a sequence $\{\mathbf{x}_k\}$ according to $\mathbf{x}_k \in \mathbf{x}_0 + \text{Span}\{\mathbf{g}(\mathbf{x}_0), \ldots, \mathbf{g}(\mathbf{x}_{k-1})\}$

As described above, we assume that the initial point has a bounded distance from an optimal solution, i.e., $\|\mathbf{x}_0 - \mathbf{x}^*\| \leq R$, and function $f$ is $G$-Lipschitz over this set. Hence, we denote our class of problems as $\mathcal{P}(\mathbf{x}_0, R, G, \mu)$. In the following theorem, we establish a lower bound for this class of problems.

**Theorem 3.** *For any class $\mathcal{P}(\mathbf{x}_0, R, G, \mu)$ and any $1 \leq k \leq n-1$, there exists a function $f \in \mathcal{P}(\mathbf{x}_0, R, G, \mu)$, such that*

$$\min_{i=1,\ldots,k} f(\mathbf{x}_i) - f^* \;\geq\; \frac{G^2}{2\mu(2+\sqrt{k+1})^2}.$$

*for any optimization scheme which generates iterates according to $\mathbf{x}_k \in \mathbf{x}_0 + Span\{\mathbf{g}(\mathbf{x}_0), \ldots, \mathbf{g}(\mathbf{x}_{k-1})\}$.*

Recall the function

$$f(\mathbf{x}) = f_{k+1}(\mathbf{x}) = \gamma \max_{i=1,\ldots,k+1} x(i) + \frac{\mu}{2}\|\mathbf{x}\|^2$$

with

$$\gamma = \frac{G\sqrt{k+1}}{2+\sqrt{k+1}}$$

Note that we have

$$\|\mathbf{x}_0 - \mathbf{x}^*\| = \|\mathbf{x}^*_{k+1}\| = \frac{\gamma}{\mu\sqrt{k+1}} = \frac{G}{\mu(2+\sqrt{k+1})}, \qquad f^* = f^*_{k+1} = -\frac{\gamma^2}{2\mu(k+1)} = -\frac{G^2}{2\mu(2+\sqrt{k+1})^2}$$

It can be easily verified that the function is Lipschitz with constant:

$$\mu\|\mathbf{x}^*\| + \mu\rho + \gamma = 2\mu\|\mathbf{x}^*_{k+1}\| + \gamma = G$$

We further using a similar argument can show that

$$f(\mathbf{x}_k) - f^* \geq -f^* = \frac{\gamma^2}{2\mu(k+1)} = \frac{G^2}{2\mu(2+\sqrt{k+1})^2}.$$

# 5 Upper Bound for the Strongly Convex Setting: Subgradient Method

**Theorem 4.** *Let $f$ be a $\mu$-strongly convex function that is $G$-Lipschitz continuous on the set $\|\mathbf{x}-\mathbf{x}^*\| \leq R$, and suppose $\|\mathbf{x}_0-\mathbf{x}^*\| \leq R$. Then, after at most $N = \frac{G^2}{\mu\epsilon}\log\left(\frac{\|\mathbf{x}_0-\mathbf{x}^*\|}{\epsilon}\right)$ we have*

$$\min_{k=1,\dots,N} f(\mathbf{x}_i) - f^* \leq \epsilon$$

*Proof.* Suppose for all $k = 0,\dots,N$ we have $f(\mathbf{x}_k) - f(\mathbf{x}^*) > \epsilon$ Note that we can show that

$$\begin{aligned}
\|\mathbf{x}_{k+1}-\mathbf{x}^*\|^2 &= \|\pi_{\mathcal{Q}}\left(\mathbf{x}_k - \eta_k\mathbf{g}_k\right) - \pi_{\mathcal{Q}}\left(\mathbf{x}^*\right)\|^2 \\
&\leq \|\mathbf{x}_k - \eta_k\mathbf{g}_k - \mathbf{x}^*\|^2 \\
&= \|\mathbf{x}_k-\mathbf{x}^*\|^2 - 2\eta_k\mathbf{g}_k^\top(\mathbf{x}_k-\mathbf{x}^*) + \eta_k^2\|\mathbf{g}_k\|^2 \\
&\leq \|\mathbf{x}_k-\mathbf{x}^*\|^2 - 2\eta_k(f(\mathbf{x}_k)-f(\mathbf{x}^*) + \frac{\mu}{2}\|\mathbf{x}_k-\mathbf{x}^*\|^2) + \eta_k^2\|\mathbf{g}_k\|^2 \\
&\leq (1-\mu\eta_k)\|\mathbf{x}_k-\mathbf{x}^*\|^2 - 2\eta_k(f(\mathbf{x}_k)-f(\mathbf{x}^*) + \eta_k^2\|\mathbf{g}_k\|^2
\end{aligned}$$

If we simply set $\eta_k = \frac{2\epsilon}{G^2}$ or $\eta_k = \frac{2\epsilon}{\|\mathbf{g}_k\|^2}$ we have

$$\|\mathbf{x}_{k+1}-\mathbf{x}^*\|^2 \leq (1-\mu\eta_k)\|\mathbf{x}_k-\mathbf{x}^*\|^2$$

Therefore,

$$\|\mathbf{x}_N-\mathbf{x}^*\|^2 \leq \left(1-\frac{2\mu\epsilon}{G^2}\right)^N\|\mathbf{x}_0-\mathbf{x}^*\|^2$$

Hence, we have

$$f(\mathbf{x}_N)-f^* \leq G\left(1-\frac{2\mu\epsilon}{G^2}\right)^{N/2}\|\mathbf{x}_0-\mathbf{x}^*\| \leq G\exp\left\{-\frac{\mu\epsilon N}{G^2}\right\}\|\mathbf{x}_0-\mathbf{x}^*\|$$

Therefore,

$$G\exp\left\{-\frac{\mu\epsilon N}{G^2}\right\}\|\mathbf{x}_0-\mathbf{x}^*\| \geq \epsilon$$

otherwise we have a contradiction. Hence,

$$N \leq \frac{G^2}{\mu\epsilon}\log\left(\frac{G\|\mathbf{x}_0-\mathbf{x}^*\|}{\epsilon}\right)$$

□

Considering the fact that $\frac{\mu}{2}\|\mathbf{x}_0-\mathbf{x}^*\|^2 \leq f(\mathbf{x}_0)-f^* \leq G\|\mathbf{x}_0-\mathbf{x}^*\|$ we have $\|\mathbf{x}_0-\mathbf{x}^*\| \leq 2G/\mu$. Hence, to find an $\epsilon$ accurate solution we need at most $N = \frac{G^2}{\mu\epsilon}\log\left(\frac{2G^2}{\mu\epsilon}\right)$ iterations of the subgradient method.