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

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

LECTURE 7

Caramanis & Mokhtari Wednesday, February 9, 2022

Goal: In this lecture, we talk about smooth constrained optimization over convex simple sets.

1 Convex Constrained Optimization

Consider a general constrained convex optimization problem where the constraint set is a simple convex set with no functional constraints:

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

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

2 Properties of convex and strongly convex constrained problems

We next provide optimality conditions for convex constrained optimization problems. These are the familiar first-order conditions of optimality for convex optimization. While simple, the condition is extremely useful both algorithmically and for analysis.

Theorem 1. *Let $f \in \mathcal{F}^1(\mathcal{Q})$ and the set $\mathcal{Q}$ be closed and convex. A point $\mathbf{x}^*$ is optimal for* (1) *if and only if $\mathbf{x}^* \in \mathcal{Q}$ and*

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{y} - \mathbf{x}^*) \geq 0, \qquad \text{for all } \mathbf{y} \in \mathcal{Q}.$$

What does it mean?
- If you move from $\mathbf{x}^*$ towards any feasible $\mathbf{y}$, you will increase f locally
- It means that $-\nabla f(\mathbf{x}^*)$ defines a supporting hyperplane to the feasible set at $\mathbf{x}^*$

Proof. (i) Suppose $\mathbf{x}^*$ satisfies the condition

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{y} - \mathbf{x}^*) \geq 0, \qquad \text{for all } \mathbf{y} \in \mathcal{Q}.$$

Further, by convexity of function f we have

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{y} - \mathbf{x}^*) \qquad \text{for all } \mathbf{y} \in \mathcal{Q}$$

Combining these two results we can write that

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) \qquad \text{for all } \mathbf{y} \in \mathcal{Q}$$

and therefore $\mathbf{x}^*$ is an optimal point.

(ii) Now suppose $\mathbf{x}^*$ is an optimal solution. We prove the optimality condition with contradiction. Suppose there exists a feasible $\mathbf{y}$ such that

$$\nabla f(\mathbf{x}^*)^\top(\mathbf{y} - \mathbf{x}^*) < 0$$

Now consider the feasible point $\mathbf{x}^* + \alpha(\mathbf{y} - \mathbf{x}^*)$ for $\alpha \in [0, 1]$ and define the function $g : \mathbb{R} \to \mathbb{R}$ as $g(\alpha) = f(\mathbf{x}^* + \alpha(\mathbf{y} - \mathbf{x}^*))$. Note that the first derivative of g is given by

$$g'(\alpha) = \nabla f(\mathbf{x}^* + \alpha(\mathbf{y} - \mathbf{x}^*))^\top(\mathbf{y} - \mathbf{x}^*)$$

Hence, we have

$$g'(0) = \nabla f(\mathbf{x}^*)^\top(\mathbf{y} - \mathbf{x}^*) < 0.$$

If $g(0) = f(\mathbf{x}^*)$ and $g'(0) < 0$ then for small positive α we have

$$f(\mathbf{x}^* + \alpha(\mathbf{y} - \mathbf{x}^*)) = g(\alpha) < g(0) = f(\mathbf{x}^*).$$

But this contradicts the optimality of $\mathbf{x}^*$.

□

Indeed, for convex problems the optimal solution may not be unique. In some cases it may not even exists. For instance, consider the problem

$$\begin{aligned} &\min \mathbf{1}^\top \mathbf{x} \\ &\text{s.t } \mathbf{x} \le \mathbf{0}, \end{aligned}$$

Indeed, the objective function is convex and differentiable and its constraint set is also convex and closed. Alas, the problem is unbounded below and hence it does not have a well-defined optimal solution.

However, when we add the strong convexity assumption as it provides a quadratic lower bound for the objective function we can show that the optimal solution exists. In fact, we show that the optimal solution in that case is unique.

Theorem 2. *Let $f \in \mathcal{F}^1_\mu(\mathcal{Q})$ and the set $\mathcal{Q}$ be closed and convex. Then there exists a unique solution $\mathbf{x}^*$ of problem* (1).

Proof. Let $\mathbf{x}_0 \in \mathcal{Q}$. Consider the following optimization problem

$$\begin{aligned} &\min f(\mathbf{x}) \\ &\text{s.t } \mathbf{x} \in \mathcal{Q}', \end{aligned} \tag{2}$$

where $\mathcal{Q}' := \{\mathbf{x} \in \mathcal{Q} \mid f(\mathbf{x}) \le f(\mathbf{x}_0)\}$. It can be easily verified that these two problems are equivalent: the optimal objective function value of both are the same and the optimal solution of any of them (if exists) is also an optimal solution of the other one.

The major advantage of this problem is that the second set $\mathcal{Q}'$ is a bounded set. We show this by proving that any $\mathbf{x} \in \mathcal{Q}'$ has a bounded distance from a the point $\mathbf{x}_0$. To do so, note that for any $\mathbf{x} \in \mathcal{Q}'$ we have

$$f(\mathbf{x}_0) \ge f(\mathbf{x}) \ge f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top(\mathbf{x} - \mathbf{x}_0) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{x}_0\|^2$$

and hence,

$$\nabla f(\mathbf{x}_0)^\top(\mathbf{x} - \mathbf{x}_0) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{x}_0\|^2 \le 0$$

leading to

$$\frac{\mu}{2}\|\mathbf{x}-\mathbf{x}_0\|^2 \leq -\nabla f(\mathbf{x}_0)^\top(\mathbf{x}-\mathbf{x}_0) \leq \|\nabla f(\mathbf{x}_0)\|\|\mathbf{x}-\mathbf{x}_0\|$$

which implies that for any $\mathbf{x} \in \mathcal{Q}'$ we have

$$\|\mathbf{x}-\mathbf{x}_0\|^2 \leq \frac{2}{\mu}\|\nabla f(\mathbf{x}_0)\|$$

and hence the set is bounded. Using this observation we obtain that (2) has a solution, as it is a convex minimization over a bounded convex set. Hence, there exists a solution for problem (1).

Now we need to show that if $\mathbf{x}^*$ is a solution of (1), it has to be unique. Note that using strong convexity, for any $\mathbf{x} \in \mathcal{Q}$ we have

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top(\mathbf{x}-\mathbf{x}^*) + \frac{\mu}{2}\|\mathbf{x}-\mathbf{x}^*\|^2$$

and as $\mathbf{x}^*$ is an optimal solution we know that for any $\mathbf{x} \in \mathcal{Q}$ we have $\nabla f(\mathbf{x}^*)^\top(\mathbf{x}-\mathbf{x}^*)$. Hence, for any $\mathbf{x} \in \mathcal{Q}$ we have

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \frac{\mu}{2}\|\mathbf{x}-\mathbf{x}^*\|^2$$

Now if there exists another point $\hat{\mathbf{x}}^* \in \mathcal{Q}$ who is also an optimal solution of (1), then we have $f(\mathbf{x}^*) = f(\hat{\mathbf{x}}^*)$ which implies that

$$0 \geq \frac{\mu}{2}\|\hat{\mathbf{x}}^*-\mathbf{x}^*\|^2$$

and therefore $\hat{\mathbf{x}}^* = \mathbf{x}^*$. Hence, the optimal solution of the problem is unique when the objective function is strongly convex. □

3 Euclidean Projection

Definition 1. *Let $\mathcal{Q}$ be a closed set and $\mathbf{x}_0 \in \mathbb{R}^n$. Then, we define the Euclidean projection of the point $\mathbf{x}_0$ onto the set $\mathcal{Q}$ as*

$$\pi_{\mathcal{Q}}(\mathbf{x}_0) := \operatorname*{argmin}_{\mathbf{x}\in\mathcal{Q}} \|\mathbf{x}-\mathbf{x}_0\|.$$

This definition implies that $\pi_{\mathcal{Q}}(\mathbf{x}_0)$ is the closest point in the set $\mathcal{Q}$ to the point $\mathbf{x}_0$. Indeed, if $\mathbf{x}_0 \in \mathcal{Q}$, we have $\pi_{\mathcal{Q}}(\mathbf{x}_0) = \mathbf{x}_0$ (considering the closeness of the set).

Now a natural question is the solution to the above problem unique? Meaning, is projection unique? The answer is yes, if the set $\mathcal{Q}$ is convex!

Theorem 3. *If the set $\mathcal{Q}$ is closed and convex, then $\pi_{\mathcal{Q}}(\mathbf{x}_0)$ is always unique.*

Proof. The projection problem is a strongly convex and smooth problem over a convex compact set, and hence its solution is unique. □

3.1 properties of projection

Lemma 1. *For any two points $\mathbf{x}_1$ and $\mathbf{x}_2 \in \mathbb{R}^n$ we have*

$$\|\pi_{\mathcal{Q}}(\mathbf{x}_1) - \pi_{\mathcal{Q}}(\mathbf{x}_2)\| \leq \|\mathbf{x}_1-\mathbf{x}_2\|$$

Lemma 2. *For any two points where* $\mathbf{x} \in \mathcal{Q}$ *and* $\mathbf{y} \in \mathbb{R}^n$ *we have*

$$\|\pi_{\mathcal{Q}}(\mathbf{y}) - \mathbf{x}\|^2 + \|\pi_{\mathcal{Q}}(\mathbf{y}) - \mathbf{y}\|^2 \leq \|\mathbf{x} - \mathbf{y}\|^2$$

In the following theorem, we show that the set of optimal solutions of problem (1) can be characterized as the set of points whose projection after a gradient step would be themselves. I.e., if we are at the optimal solution by following a projected gradient update we will return to the same point.

Theorem 4. *Let* $\mathbf{x}^*$ *be an optimal solution of* (1) *where the loss function is convex and differentiable. Then, for any* $\gamma > 0$*, we have*

$$\pi_{\mathcal{Q}}\left(\mathbf{x}^* - \frac{1}{\gamma}\nabla f(\mathbf{x}^*)\right) = \mathbf{x}^*$$

Proof. Let's defined $\hat{\mathbf{x}}^*$ as

$$\hat{\mathbf{x}}^* := \pi_{\mathcal{Q}}\left(\mathbf{x}^* - \frac{1}{\gamma}\nabla f(\mathbf{x}^*)\right) := \underset{\mathbf{x}\in\mathcal{Q}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{x}^* + \frac{1}{\gamma}\nabla f(\mathbf{x}^*)\|^2.$$

Hence, $\hat{\mathbf{x}}^*$ satisfies the condition:

$$\left(\hat{\mathbf{x}}^* - \mathbf{x}^* + \frac{1}{\gamma}\nabla f(\mathbf{x}^*)\right)^\top (\mathbf{x} - \hat{\mathbf{x}}^*) \geq 0$$

for any $\mathbf{x} \in \mathcal{Q}$. Now if we have $\hat{\mathbf{x}}^* \neq \mathbf{x}^*$, then by setting $\mathbf{x} = \mathbf{x}^*$ we obtain

$$\left(\frac{1}{\gamma}\nabla f(\mathbf{x}^*)\right)^\top (\mathbf{x}^* - \hat{\mathbf{x}}^*) \geq \|\hat{\mathbf{x}}^* - \mathbf{x}^*\|^2$$

which implies that

$$\nabla f(\mathbf{x}^*)^\top (\hat{\mathbf{x}}^* - \mathbf{x}^*) \leq -\gamma\|\hat{\mathbf{x}}^* - \mathbf{x}^*\|^2 < 0$$

which is a contradiction. Hence, $\hat{\mathbf{x}}^* = \mathbf{x}^*$ and the proof is complete. □