Lecture 14: Constrained Optimization over Closed Convex Sets

Yudong Chen

Consider the constrained problem

$$\min_{x \in \mathcal{X}} f(x), \tag{P}$$

where f is continuously differentiable and $\mathcal{X} \subseteq \text{dom}(f) \subseteq \mathbb{R}^d$ is a *closed*, *convex* and nonempty set.

Recall:

Definition 1 (Local minimizer)**.** We say that $x^* \in \mathcal{X} \subseteq \text{dom}(f)$ is a *local minimizer/solution* of (P) if there exists a neighborhood $\mathcal{N}_{x^*}$ of x^* such that we have $f(x) \geq f(x^*), \forall x \in \mathcal{N}_{x^*} \cap \mathcal{X}$.

For constrained problem, if x^* is a (local) minimizer of (P), it is not necessary that $\nabla f(x^*) = 0$. Example: $f(x) = x, \mathcal{X} = [2,3], x^* = 2, \nabla f(x^*) = 1 \neq 0$.

1 Optimality condition

A cone is a set that satisfies the following property: if z is in the set, then for any $t > 0$, tz is also in the set.

The optimality condition for constrained optimization would involve a special cone.

Definition 2 (Normal cone)**.** Let $\mathcal{X}$ be a closed convex set. At any point $x \in \mathcal{X}$, the normal cone $N_{\mathcal{X}}(x)$ is defined by

$$N_{\mathcal{X}}(x) = \left\{ p \in \mathbb{R}^d : \langle p, y - x \rangle \leq 0, \forall y \in \mathcal{X} \right\}.$$

Note that by definition,

$$-\nabla f(x) \in N_{\mathcal{X}}(x) \iff \langle -\nabla f(x), y - x \rangle \leq 0, \forall y \in \mathcal{X}. \tag{1}$$

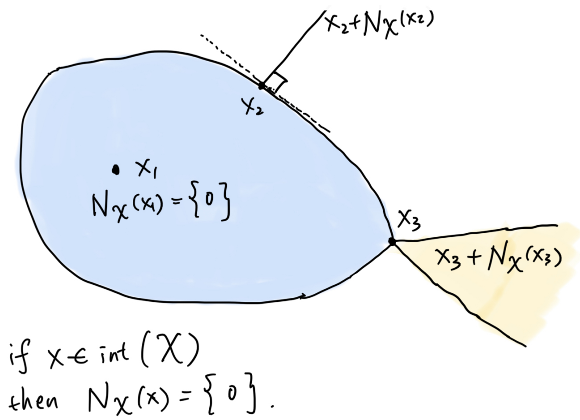
If $\mathcal{X} = \mathbb{R}^d$, then (1) reduces to $\nabla f(x^*) = 0$.

Theorem 1 (Thm 7.2 in Wright-Recht)**.** *Consider the problem (P).*

1. *(1st-order necessary condition) If $x^* \in \mathcal{X}$ is a local solution to (P), then $-\nabla f(x^*) \in N_{\mathcal{X}}(x^*)$.*
2. *(1st-order sufficient condition) If f is convex, then $-\nabla f(x^*) \in N_{\mathcal{X}}(x^*)$ implies that x^* is a global solution to (P).*

Any point x that satisfies (1) is called a *stationary point* for the constrained problem (P).

Illustration of normal cones:



Proof. **Part 1:** Want to show: x^* is a local solution $\Longrightarrow -\nabla f(x^*) \in N_{\mathcal{X}}(x^*)$.

Proof by contradiction. Suppose $-\nabla f(x^*) \notin N_{\mathcal{X}}(x^*)$. By definition of $N_{\mathcal{X}}(x^*)$, there exists $y \in \mathcal{X}$ such that

$$\begin{aligned} &\langle -\nabla f(x^*), y - x^* \rangle \geq \delta > 0 \\ \Longleftrightarrow &\langle \nabla f(x^*), y - x^* \rangle \leq -\delta. \end{aligned}$$

For each $\alpha > 0$, by Taylor's Theorem we have

$$f\Big(\underbrace{x^* + \alpha(y - x^*)}_{=(1-\alpha)x^* + \alpha y \in \mathcal{X}}\Big) = f(x^*) + \alpha \left\langle \nabla f(x^* + \gamma\alpha(y - x^*)), y - x^* \right\rangle$$

for some $\gamma \in (0, 1)$. Because ∇f is continuous, for all $\alpha > 0$ sufficiently small:

$$\langle \nabla f(x^* + \gamma\alpha(y - x^*)), y - x^* \rangle \leq -\frac{\delta}{2}.$$

It follows that

$$f\Big(x^* + \alpha(y - x^*)\Big) \leq f(x^*) - \frac{\alpha\delta}{2} < f(x^*),$$

which means x^* cannot be a local solution, a contradiction.

Part 2: Want to show:

$$\underbrace{f \text{ is convex}}_{\text{(i)}} \text{ and } \underbrace{-\nabla f(x^*) \in N_{\mathcal{X}}(x^*)}_{\text{(ii)}} \implies x^* \text{ is a global solution}$$

From (i): $\forall x, y \in \mathbb{R}^d$: $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$. In particular, for $x = x^*$:

$$\forall y \in \mathcal{X}: \quad f(y) \geq f(x^*) + \langle \nabla f(x^*), y - x^* \rangle.$$

From (ii):

$$\forall y \in \mathcal{X}: \quad \langle -\nabla f(x^*), y - x^* \rangle \leq 0 \Longleftrightarrow \langle \nabla f(x^*), y - x^* \rangle \geq 0.$$

(i)+(ii) gives $f(y) \geq f(x^*), \forall y \in \mathcal{X}$. □

For strongly convex f, the minimizer is unique.

Theorem 2 (Thm 7.3 in Wright-Recht)**.** *Consider (P) and assume, in addition, that f is strongly convex. Then (P) has a unique global minimizer. Moreover, x^* is the global minimizer if and only if $-\nabla f(x^*) \in N_{\mathcal{X}}(x^*)$.*

Proof. Recall that Strong convexity means there exists $m > 0$ such that

$$\forall x, y : f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{m}{2} \|y - x\|_2^2 .$$

Existence of global solution: Fix an arbitrary $x \in \mathcal{X}$. Consider any y such that $f(y) \leq f(x)$. We have

$$\begin{aligned} \|y - x\|_2^2 &\leq \frac{2}{m} \Big(\underbrace{f(y) - f(x)}_{\leq 0} - \langle \nabla f(x), y - x \rangle \Big) \\ &\leq \frac{2}{m} \|\nabla f(x)\|_2 \|y - x\|_2 . \qquad \text{Cauchy-Schwarz} \end{aligned}$$

Hence

$$\|y - x\|_2 \leq \frac{2}{m} \|\nabla f(x)\|_2 < \infty.$$

Thus, the set $\{y \in \mathcal{X} \mid f(y) \leq f(x)\}$ is closed and bounded $\implies$ compact $\implies$ a global minimizer x^* exists by Weierstrass theorem.

"only if" part: follows from Theorem 1.

"if part" and uniqueness. Apply strong convexity to $x = x^*$:

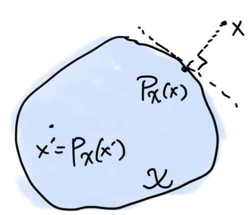
$$\begin{aligned} \forall y \in \mathcal{X} : f(y) &\geq f(x^*) + \underbrace{\langle \nabla f(x^*), y - x^* \rangle}_{\geq 0} + \frac{m}{2} \|y - x^*\|_2^2 \\ &\geq f(x^*) + \frac{m}{2} \|y - x^*\|_2^2 , \end{aligned}$$

where $\langle \nabla f(x^*), y - x^* \rangle \geq 0$ because $-\nabla f(x^*) \in N_{\mathcal{X}}(x^*)$. Therefore, $f(y) \geq f(x^*)$, and equality holds if and only if $y = x^*$. □

2 Euclidean (orthogonal) projection

The Euclidean projection of x onto the (closed and convex) set $\mathcal{X}$ is defined as

$$\begin{aligned} P_{\mathcal{X}}(x) &= \operatorname*{argmin}_{y \in \mathcal{X}} \{ \|y - x\|_2 \} \\ &= \operatorname*{argmin}_{y \in \mathcal{X}} \left\{ \frac{1}{2} \|y - x\|_2^2 \right\} . \end{aligned}$$



By Theorem 2:

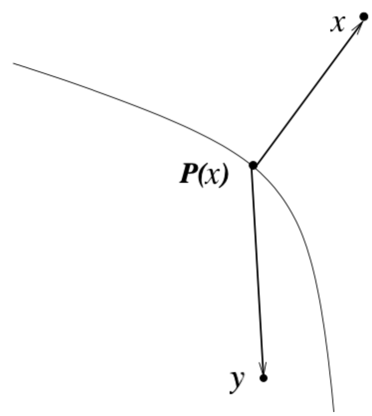
- $P_{\mathcal{X}}(x)$ exists and is unique, since we are minimizing a strongly convex function over a closed convex set.

- Furthermore, $P_{\mathcal{X}}(x)$ satisfies the first-order optimality condition

$$\begin{aligned} \forall y \in \mathcal{X}: \quad & \langle P_{\mathcal{X}}(x) - x, y - P_{\mathcal{X}}(x) \rangle \geq 0 \\ & \Updownarrow \\ & -\left(P_{\mathcal{X}}(x) - x\right) \in N_{\mathcal{X}}\left(P_{\mathcal{X}}(x)\right). \end{aligned} \tag{2}$$

- The converse is also true: if some $\bar{x}$ satisfies $\langle \bar{x} - x, y - \bar{x} \rangle \geq 0, \forall y \in \mathcal{X}$, then we must have $\bar{x} = P_{\mathcal{X}}(x)$.

Equation (2), which fully characterizes $P_{\mathcal{X}}(x)$, is also known as the *minimum principle.* Illustration:

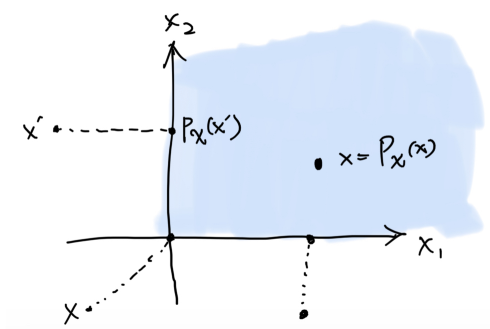


2.1 Examples

Some examples of $\mathcal{X}$ for which the associated projection is easy to compute.

2.1.1 Non-negative orthant

$\mathcal{X} = \left\{x \in \mathbb{R}^d \mid x \geq 0 \text{ element-wise}\right\}.$



Claim 1. $P_{\mathcal{X}}(x) = \max\left\{x, \vec{0}\right\}$, where the max is elementwise.

Proof. Check (2):

$$\begin{aligned} \forall y \in \mathcal{X}: & \langle P_{\mathcal{X}}(x) - x, y - P_{\mathcal{X}}(x) \rangle \\ &= \sum_{i=1}^{d} \left(\max\{x_i, 0\} - x_i\right)\left(y_i - \max\{x_i, 0\}\right) \\ &\geq 0, \end{aligned}$$

where the last inequality holds because

$$\max\{x_i, 0\} - x_i \begin{cases} = 0 & \text{if } x_i \geq 0 \\ = -x_i > 0 & \text{if } x_i < 0 \end{cases}$$

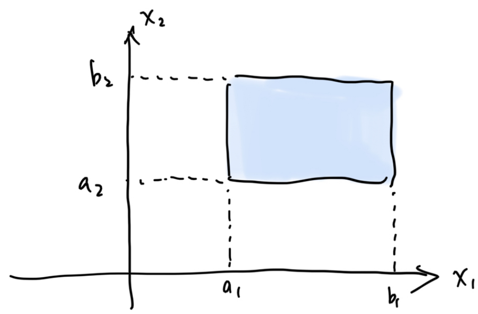
and

$$y_i - \max\{x_i, 0\} \begin{cases} = y_i - x_i & \text{if } x_i \geq 0 \\ = y_i \geq 0 & \text{if } x_i < 0 \end{cases}$$

□

2.1.2 Hyper-rectangle

$\mathcal{X} = \{x \in \mathbb{R}^d \mid \forall i \in \{1, \dots, d\} : x_i \in [a_i, b_i]\}$, where $a_i < b_i$. See HW4.

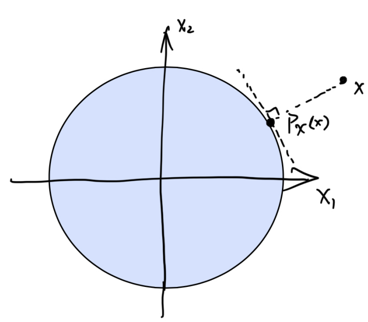


2.1.3 Euclidean ball

$\mathcal{X} = \{x \in \mathbb{R}^d \mid \|x\|_2 \leq 1\}$. Then

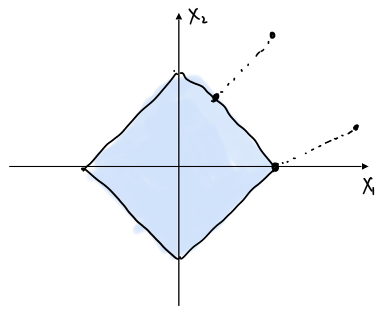
$$P_{\mathcal{X}}(x) = \begin{cases} x, & \text{if } x \in \mathcal{X} \\ \frac{x}{\|x\|_2} & \text{if } x \notin \mathcal{X} \end{cases}$$

Exercise 1. What if the ball was of radius $R > 0$? What if the ball was not centered at zero?



2.1.4 ℓ_1 ball

$\mathcal{X} = \{x \in \mathbb{R}^d \mid \|x\|_1 \leq 1\}$. Then $P_{\mathcal{X}}(x)$ can be computed with $O(d \log d)$ arithmetic operations (involves sorting).



2.1.5 Probability simplex

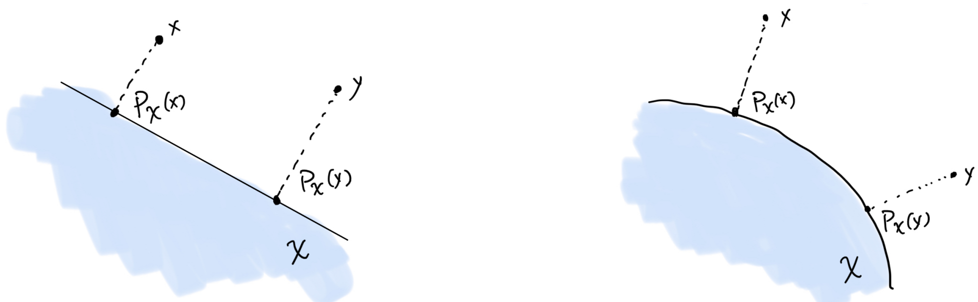
$\mathcal{X} = \left\{ x \in \mathbb{R}^d \mid x \geq 0, \sum_{i=1}^{d} x_i = 1 \right\}$. (A picture) Similar to ℓ_1 ball. Computable in $O(d \log d)$.

2.2 $P_{\mathcal{X}}$ is nonexpansive

Proposition 1 (Prop 7.7 in Wright-Recht)**.** *Let $\mathcal{X}$ be a closed, convex and nonempty set. Then $P_{\mathcal{X}}(\cdot)$ is a* non-expansive *operator, i.e.,*

$$\forall x, y \in \mathbb{R}^d : \quad \|P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y)\|_2 \leq \|x - y\|_2 .$$

Illustrations:



Proof. Equivalently, want to show that

$$\|x - y\|_2^2 \geq \|P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y)\|_2^2 .$$

We have

$$\begin{aligned} \|x - y\|_2^2 &= \|x - P_{\mathcal{X}}(x) - (y - P_{\mathcal{X}}(y)) + P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y)\|_2^2 \\ &= \underbrace{\|x - P_{\mathcal{X}}(x) - (y - P_{\mathcal{X}}(y))\|_2^2}_{\geq 0} + \|P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y)\|_2^2 \\ &\quad + 2 \underbrace{\langle x - P_{\mathcal{X}}(x), P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y) \rangle}_{\geq 0} + 2 \underbrace{\langle y - P_{\mathcal{X}}(y), P_{\mathcal{X}}(y) - P_{\mathcal{X}}(x) \rangle}_{\geq 0} \\ &\geq \|P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y)\|_2^2 , \end{aligned}$$

where we use the minimum principle (2) to lower bound the two inner products. □

Remark 1 (Firmly nonexpansive). The proof above shows that $P_{\mathcal{X}}(\cdot)$ actually satisfies a stronger property: it is *firmly nonexpansive,* in the sense that

$$\|P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y)\|_2^2 + \|x - P_{\mathcal{X}}(x) - (y - P_{\mathcal{X}}(y))\|_2^2 \leq \|x - y\|_2^2.$$

In particular, if $y \in \mathcal{X}$, then

$$\|P_{\mathcal{X}}(x) - y\|_2^2 + \|x - P_{\mathcal{X}}(x)\|_2^2 \leq \|x - y\|_2^2$$

and hence the strict inequality $\|P_{\mathcal{X}}(x) - y\|_2^2 < \|x - y\|_2^2$ holds whenever $x \notin \mathcal{X}$.