Consider the constrained problem

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

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

Recall:

Definition 1 (Local minimizer)**.** We say that $x^* \in \mathcal{X} \subseteq \mathrm{dom}(f)$ 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 $\nabla f(x^*) = 0$.

$x, y \in \mathbb{R}^d$.

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

$\Rightarrow p = 0$. That is, if $\mathcal{X} = \mathbb{R}^d$, $N_{\mathcal{X}}(x) = \{0\}$, $\forall x \in \mathbb{R}^d$.

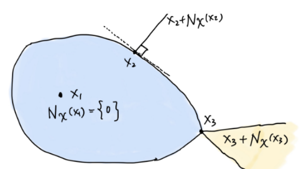
$-\nabla f(x^*) \in N_{\mathcal{X}}(x^*) \Rightarrow -\nabla f(x^*) = 0 \Rightarrow \nabla f(x^*) = 0$.

Theorem 1 (Thm 7.2 in Wright-Recht)**.** *Consider the problem* (P). (约束优化问题)

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:



if $x \in \mathrm{int}(\mathcal{X})$
then $N_{\mathcal{X}}(x) = \{0\}$.

pf: 1. Proof by contradiction, $-\nabla f(x^*) \notin N_{\mathcal{X}}(x^*)$.

$\exists y \in \mathcal{X}$, $\langle -\nabla f(x^*), y-x^* \rangle \geq \delta > 0$. $\Rightarrow$ $\langle \nabla f(x^*), y-x^* \rangle \leq -\delta$.
$\exists \delta \in \mathbb{R}$.

$\alpha \in [0,1)$

Taylor's thm, $\exists \nu \in [0,1)$

$$f(x^* + \alpha(y-x^*)) = f(x^*) + \alpha \langle \nabla f(x^* + \nu\alpha(y-x^*)), y-x^* \rangle$$

$x^* + \alpha(y-x^*) = (1-\alpha)x^* + \alpha y \in \mathcal{X}$.

Set $\alpha \to 0$. (sufficiently small) Since ∇f is continuous. $\langle \nabla f(x^* + \nu\alpha(y-x^*)), y-x^* \rangle \leq -\frac{\delta}{2}$.

$\Rightarrow f(x^* + \alpha(y-x^*)) \leq f(x^*) - \frac{\delta}{2} < f(x^*)$. Contradicts with x^* is local min.

2. $-\nabla f(x^*) \in N_{\mathcal{X}}(x^*)$ $\Rightarrow$ $\forall y \in \mathcal{X}$, $\langle -\nabla f(x^*), y-x^* \rangle \leq 0$.

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

$\Rightarrow x^*$ is a global min for (P).

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 $-\nabla f(x^*) \in N_{\mathcal{X}}(x^*)$.*

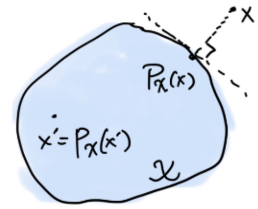
pf:

($\Leftarrow$) $\begin{cases} -\nabla f(x^*) \in N_{\mathcal{X}}(x^*) \\ f \text{ is strongly cvx.} \end{cases}$ This follows cvx case. $\Rightarrow x^*$ is a global min.
Moreover f is strongly cvx $\Rightarrow$ x^*, as a global min, is unique.

($\Rightarrow$). A global min x^* must be a local min $\Rightarrow$ $-\nabla f(x^*) \in N_{\mathcal{X}}(x^*)$, following Thm 1.

2 Euclidean (orthogonal) projection

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



$$P_{\mathcal{X}}(x) = \underset{y\in\mathcal{X}}{\operatorname{argmin}} \{\|y-x\|_2\}$$
$$= \underset{y\in\mathcal{X}}{\operatorname{argmin}} \left\{\frac{1}{2}\|y-x\|_2^2\right\}.$$

By Theorem 2:

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

$h(y) = \frac{1}{2}\|y-x\|^2.$

$y \in \mathcal{X}.$

- $P_{\mathcal{X}}(x)$, as a global min, satisfies 1-st order necessary optimality cond. (Thm I, part 1).

$-\nabla h(P_{\mathcal{X}}(x)) \in N_{\mathcal{X}}(P_{\mathcal{X}}(x)) \qquad \nabla h(y) = y - x.$

$\Rightarrow -(P_{\mathcal{X}}(x) - x) \in N_{\mathcal{X}}(P_{\mathcal{X}}(x))$

Apply this to verify a $P_{\mathcal{X}}(x)$ is correct.

$\Leftrightarrow \forall y \in \mathcal{X},\ -\langle P_{\mathcal{X}}(x) - x,\ y - P_{\mathcal{X}}(x)\rangle \leq 0. \quad \Leftrightarrow \quad \forall y \in \mathcal{X},\ \langle P_{\mathcal{X}}(x) - x,\ y - P_{\mathcal{X}}(x)\rangle \geq 0.$

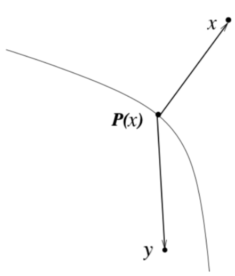
- Converse also true: If some $\bar{x}$ satisfies, for a fixed x,

$\forall y \in \mathcal{X},\ \langle \bar{x} - x,\ y - \bar{x}\rangle \geq 0$, then $\bar{x} = P_{\mathcal{X}}(x)$. $\rightarrow$ Apply this to find $P_{\mathcal{X}}(x)$.

$\downarrow$

$\bar{x}$ satisfies the same property as $P_{\mathcal{X}}(x)$, as a global min of $h(y)$. Since $h(y)$ is strongly cvx, global sol over $\mathcal{X}$ is unique. Hence $\bar{x} = P_{\mathcal{X}}(x)$.

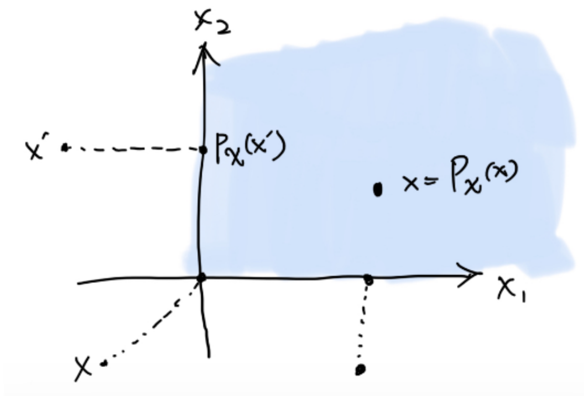
Characterization: Equation (2), which fully characterizes $P_{\mathcal{X}}(x)$, is also known as the *minimum principle.* Illustration:



Examples of $\mathcal{X}$, where associated projection is easy to compute:

2.1.1 Non-negative orthant

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



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

Verify: for $y \in \mathcal{X}$, $\langle P_\mathcal{X}(x) - x,\ y - P_\mathcal{X}(x)\rangle = \langle \max\{x, \vec{0}\} - x,\ y - \max\{x, \vec{0}\}\rangle$

$$= \sum_{i: x_i \geq 0} 0 \cdot (y_i - x_i) + \sum_{i: x_i < 0} (-x_i)(y_i - 0)$$

$x_i \geq 0$. $\quad 0$.

$x_i < 0$. $\quad -x_i$

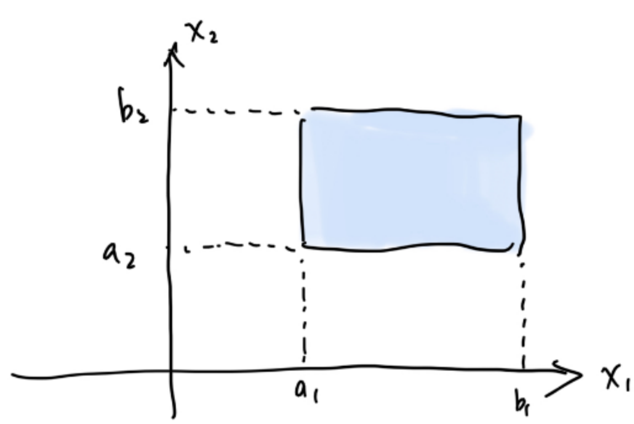
$$= -\sum_{i: x_i < 0} x_i y_i \geq 0$$

$\forall i,\ y_i \geq 0$.

$\Rightarrow \max\{x, \vec{0}\} = P_\mathcal{X}(x)$.

2.1.2 Hyper-rectangle

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



$$\boxed{[P_\mathcal{X}(x)]_i = \min(\max(x_i, a_i), b_i)}$$

Verify 类似. 分3类.

$$\begin{cases} x_i \leq a_i \Rightarrow [P_\mathcal{X}(x)]_i = a_i \\ a_i < x_i < b_i \Rightarrow [P_\mathcal{X}(x)]_i = x_i \\ x_i \geq b_i \Rightarrow [P_\mathcal{X}(x)]_i = b_i \end{cases}$$

$$\langle P_{\mathcal{X}}(x) - x, y - P_{\mathcal{X}}(x)\rangle$$

$$= \sum_{i: x_i \le a_i} \underbrace{(a_i - x_i)}_{\ge 0}\underbrace{(y_i - a_i)}_{\ge 0} + \sum_{i:\, a_i < x_i < b_i} 0 \cdot (y_i - x_i) + \sum_{i: x_i \ge b_i} \underbrace{(b_i - x_i)}_{\le 0}\underbrace{(y_i - b_i)}_{\le 0}.$$

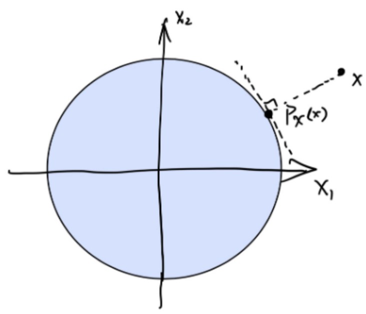
$$\ge 0. \qquad (\forall i,\ a_i \le y_i \le b_i)$$

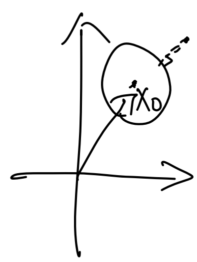
2.1.3 Euclidean ball (ℓ_2-ball)

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





Verify:

$x \in \mathcal{X}$. ✓.

$x \notin \mathcal{X}$. for $y \in \mathcal{X}$, $\langle P_{\mathcal{X}}(x) - x, y - P_{\mathcal{X}}(x)\rangle = \langle \frac{x}{\|x\|_2} - x, y - \frac{x}{\|x\|_2}\rangle$

$$= \frac{1}{\|x\|_2}\langle x, y\rangle - \frac{\langle x, x\rangle}{\|x\|_2^2} - \langle x, y\rangle + \frac{1}{\|x\|_2}\|x\|_2^2$$

$$= \underbrace{\left(\frac{1}{\|x\|_2} - 1\right)}_{<0}\langle x, y\rangle + \|x\|_2 \ge \left(\frac{1}{\|x\|_2} - 1\right)\|x\|_2\|y\|_2 + \|x\|_2 = \|y\|_2 + \underbrace{(1 - \|y\|_2)\|x\|_2}_{\ge 0} \ge 0.$$

Ex 1:

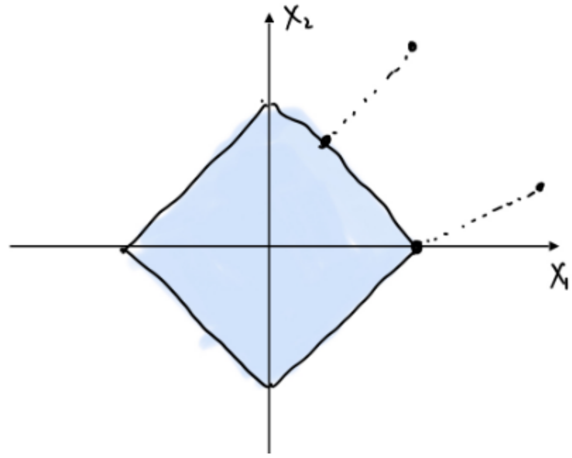
$$\mathcal{X} = \{x \in \mathbb{R}^d : \|x\|_2 \le R\}. \Rightarrow P_{\mathcal{X}}(x) = \begin{cases} x & x \in \mathcal{X} \\ R\frac{x}{\|x\|_2} & x \notin \mathcal{X}. \end{cases}$$

Not centered:

$$\mathcal{X} = \{x \in \mathbb{R}^d : \|x - x_0\|_2 \le R\}. \Rightarrow P_{\mathcal{X}}(x) = \begin{cases} x & x \in \mathcal{X}. \\ x_0 + R\frac{x - x_0}{\|x - x_0\|_2} & x \notin \mathcal{X}. \end{cases}$$

2.1.4 ℓ_1 ball

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



2.1.5 Probability simplex

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

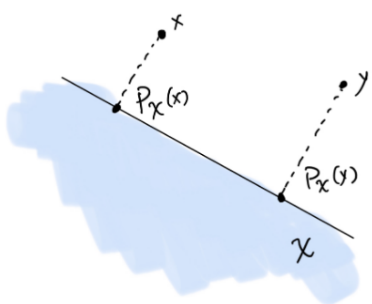
Above, we're discussing the property of $P_{\mathcal{X}}(x)$ when x is fixed, $\mathcal{X}$ varies.
Now we consider $P_{\mathcal{X}}(\cdot)$, as a function of x, $\mathcal{X}$ fixed.

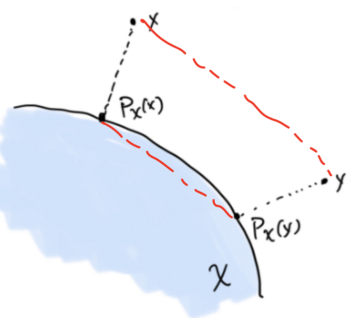
2.2 $P_{\mathcal{X}}$ is nonexpansive (非扩张映射)

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.,*

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

Illustrations:





Pf: We show that $\|x-y\|_2^2 \ge \|P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y)\|_2^2$.

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

$$= \|\underbrace{x - P_{\mathcal{X}}(x) - (y - P_{\mathcal{X}}(y))}_{\ge 0}\|^2 + \|P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y)\|^2 + 2\underbrace{\langle P_{\mathcal{X}}(x) - x, \underbrace{P_{\mathcal{X}}(y)}_{\in \mathcal{X}} - P_{\mathcal{X}}(x)\rangle}_{\ge 0, \text{ by property of } P_{\mathcal{X}}(x)}$$

$$+ 2\langle P_{\mathcal{X}}(y) - y, \underbrace{P_{\mathcal{X}}(x)}_{\in \mathcal{X}} - P_{\mathcal{X}}(y)\rangle$$

≥ 0, by property of $P_{\mathcal{X}}(y)$.

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

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

暂时不被缩到0.

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

In particular, if $\boxed{y \in \mathcal{X}}$, 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 $\boxed{x \notin \mathcal{X}}$.

$\Downarrow$

$\|x - P_{\mathcal{X}}(x)\|^2 > 0$.

投影后与原始点距离严格减小.