Constrained optimization

$$\min_{x \in \mathbb{R}^n} f(x) \quad s.t \quad x \in S \qquad \text{— (1)}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable
- S is convex set
- If $S = \mathbb{R}^n$: x^* is stationary $\Longleftrightarrow$ $\nabla f(x^*) = 0$.

Defn: (Stationary point) x^* is called a stationary point of (1) if $\nabla f(x^*)^T (x - x^*) \geq 0$ for any $x \in C$.

There is no feasible descent direction of f at x^*.

Stationary point in $\mathbb{R}^n$.

If $C = \mathbb{R}^n$, the stationary points are x^* satisfying

$$\nabla f(x^*)^T(x - x^*) \geq 0$$

for all $x \in \mathbb{R}^n$.

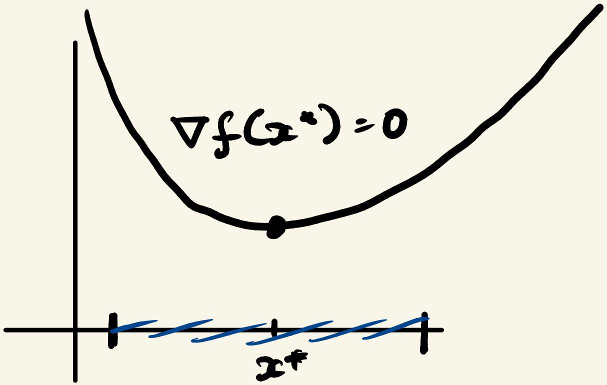
pick $x = x^* - \nabla f(x^*)$ $\Rightarrow -\|\nabla f(x^*)\| \geq 0$

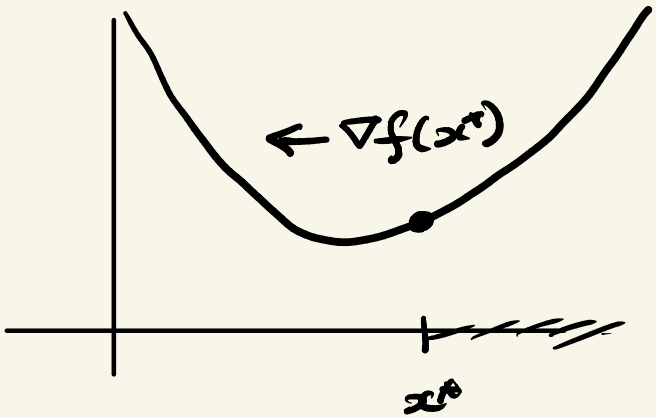
$\Rightarrow \nabla f(x^*) = 0$.

Reduces to stationary condition for $S = \mathbb{R}^n$.

Necessary Condition

Thm: Let f be a continuously differentiable function over a convex set S and let x^* be a local minimum of (1). Then x^* is a stationary point:





Example

$$\min_{x \in \mathbb{R}^n} f(x) \quad s.t. \quad x_i \geq 0 \quad i = 1, \dots, n$$

Characterize the stationary points: Assume x^* is stationary.

$$\Longleftrightarrow \quad \nabla f(x^*)^T (x - x^*) \geq 0 \quad \forall\, x_i \geq 0$$

$$\Longleftrightarrow \quad \nabla f(x^*)^T x - \nabla f(x^*)^T x^* \geq 0 \quad \forall\, x_i \geq 0$$

$$\Longleftrightarrow \quad \nabla f(x^*) \geq 0 \quad \text{and} \quad -\nabla f(x^*)^T x^* \geq 0$$

(Why? because $a^T x + b \geq 0$ for all $x \geq 0 \Longleftrightarrow a \geq 0, b \geq 0$)

Note that $x^* \geq 0$ and $\nabla f(x^*) \geq 0 \Rightarrow x_i^* \dfrac{\partial f}{\partial x_i}(x^*) = 0$

So,

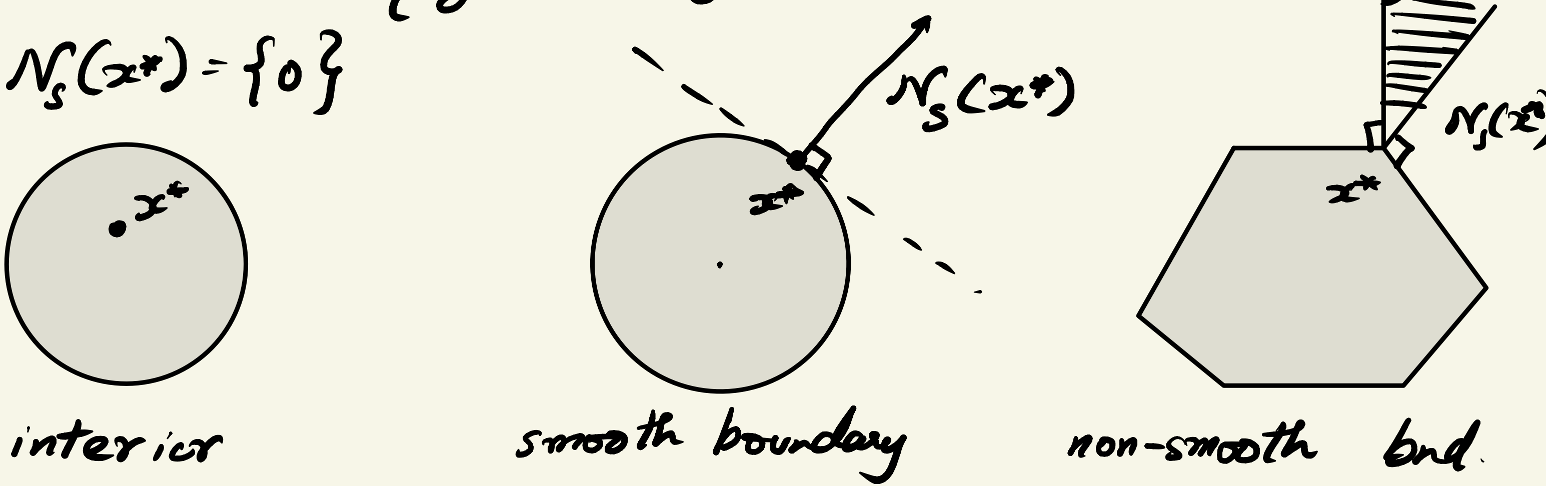
$$\boxed{\frac{\partial f}{\partial x_i}(x^*) \begin{cases} = 0, & x_i^* \geq 0 \\ \geq 0, & x_i^* = 0 \end{cases}}$$

Normal cone.

We can express necessary condition in terms of normal cone.

For $x \in S$, define the normal cone of S at x

$$N_S(x) = \{ g \in \mathbb{R}^n : g^T(z-x) \le 0 , \ \forall z \in S \}$$



Example

1. Let $S = \{z\}$.

$$N_S(x) = \begin{cases} \mathbb{R}^n & \text{if } x = z \\ \emptyset & \text{otherwise.} \end{cases}$$

2. Let $S = \{x \mid \|x\|_2 \le 1\}$

$$N_S(x) = \begin{cases} x & \text{if } \|x\|_2 = 1 \\ \{0\} & \text{if } \|x\|_2 < 1 \\ \emptyset & \text{otherwise} \end{cases}$$

Stationary point and normal cone

Thm: Let f be a continuously differentiable function over a convex set S and let x^* be a local minimum of (1). Then

$$-\nabla f(x^*) \in N_S(x^*).$$

Example: x is in interior of S.

We say that $x \in \mathbb{R}^n$ is in the interior of S ($x \in \text{int}(S)$) if

there exists $\varepsilon > 0$ such that for all $v \in \mathbb{R}^n$, $x + \varepsilon v \in S$.

let $S = \{x \mid \|x\|_2 = 1\}$. $\quad \text{int}(S) = \phi$.

let $S = \{x \mid \|x\|_2 \leq 1\}$ $\quad \text{int}(S) = \{x \mid \|x\|_2 < 1\}$.

<u>Claim</u> For any $x \in \text{int}(S)$, S convex,

$$N_S(x) = \{0\}.$$

<u>proof</u>: For any g, pick $v = g$ or $v = -g$.

$$\varepsilon g^T v = g^T(z - x) > 0, \quad z = x + \varepsilon v \in S.$$

or $$-\varepsilon g^T v = g^T(z - x) > 0, \quad z = x - \varepsilon v \in S.$$

Example: Normal cone to affine set

$$N_S(x) = \{g \in \mathbb{R}^n \mid g^T(z-x) \leq 0, \forall z \in S\}$$

Affine set $S = \{x \mid Ax = b\}$.

- Trick: shift the set. Define $S' = \{\underbrace{z-x}_{\in \mathcal{N}(A)} \mid z \in S\}$.

  Then $N_S(x) = \{g \in \mathbb{R}^n \mid g^T u \leq 0, \; \forall u \in S'\}$

  $= \{g \in \mathbb{R}^n \mid g^T u \leq 0, \; \forall u \in \text{Null}(A)\}$

- If $u \in \mathcal{N}(A)$ and $g^T u \leq 0$, then $-u \in \mathcal{N}(A)$ and $-g^T u \geq 0$.
- Therefore, $N_S(x) = \{g \in \mathbb{R}^n \mid g^T u = 0, \forall \in \mathcal{N}(A)\} = \mathcal{R}(A^T)$
- Verify: pick any $g \in \mathcal{R}(A^T)$. Then

$$g^T(z-x) = v^T A(z-x) = v^T(b-b) = 0.$$

Example: Normal cone to affine halfspace.

Affine halfspace: $S = \{x \mid Ax \le b\} = \{x : a_i^T x \le b_i,\ i=1,\dots,m\}$

What is the normal cone of S?

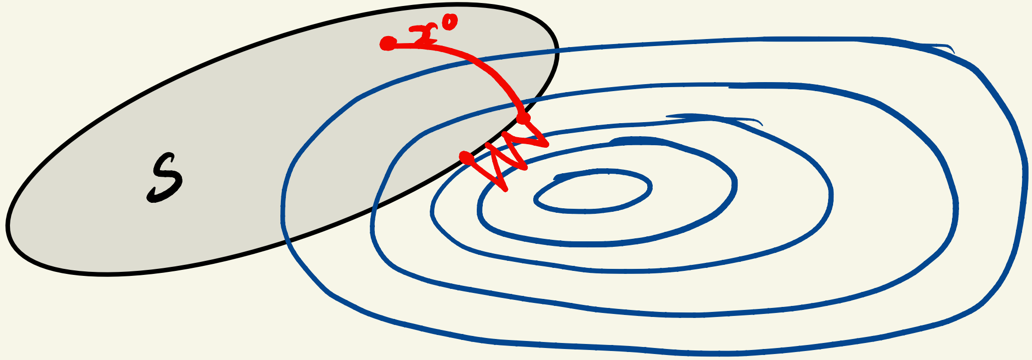
Projected gradient descent:

$$\min_{x} f(x) \quad \text{s.t.} \quad x \in S.$$

projected gradient descent:

$$x^{k+1} = \mathrm{proj}_S\left(x^{(k)} - \alpha^{(k)} \nabla f(x^{(k)})\right),$$

where $\mathrm{proj}_S(z) = \underset{x \in S}{\arg\min} \|x - z\|_2$.



Geometry of projection.

$$\mathrm{proj}_S(z) = \underset{x\in S}{\mathrm{argmin}}\ \frac{1}{2}\|x - z\|_2^2$$

- If $x \in S$ then $\mathrm{proj}_S(x) = x$.
- optimality cond. for projection : since $\nabla p(x) = x - z$

$$x = \mathrm{proj}_S(z) \iff (z-x)^T(y-x) \le 0 \quad \forall\, y \in S.$$