Optimality for Convex Optimization

March 2, 2022

OPTIMALITY FOR CONVEX OPTIMIZATION (Smooth Objective)

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad x \in C$$

- $f : \mathbb{R}^n \to \mathbb{R}$ convex differentiable function
- $C \subseteq \mathbb{R}^n$ convex set

Unconstrained ($C = \mathbb{R}^n$)

$$x^* \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \ f(x) \iff f'(x^*; x - x^*) = \underbrace{\nabla f(x^*)^T (x - x^*) \geq 0 \quad \forall x \in \mathbb{R}^n}_{\text{all directions away from } x^* \text{ are nondecreasing}}$$

$$\iff \nabla f(x^*) = 0$$

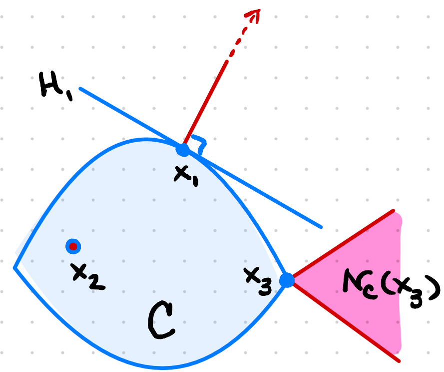
Constrained ($C \subset \mathbb{R}^n$)

$$x^* \in \underset{x \in C}{\operatorname{argmin}} \ f(x) \iff f'(x^*; x - x^*) = \underbrace{\nabla f(x^*)^T (x - x^*) \geq 0 \quad \forall x \in C}_{\text{all feasible directions are nondecreasing}}$$

NORMAL CONE

The normal cone to $C \subseteq \mathbb{R}^n$ at the point $x \in C$ is

$$N_C(x) := \{ g \in \mathbb{R}^n \mid g^T(z-x) \leq 0 \quad \forall z \in C \}$$



$N_C(x_1) = \{$normal to supporting hyperplane $H_1 = \{z \in \mathbb{R}^n \mid g^T z \leq g^T x_3\}\}$

$N_C(x_2) = \{0\}$ because $x_2 \in \operatorname{int} C$, ie, "strictly feasible"

$N_C(x_3) = \{$ cone of normals at "vertex" $\}$

Optimality

a point $x^* \in \underset{x \in C}{\operatorname{argmin}} f(x)$ if and only if

$$\nabla f(x^*)^T (z - x) \geq 0 \quad \forall z \in C$$

use the normal cone definition to deduce the equivalent optimality condition:

$$-\nabla f(x^*) \in N_C(x^*)$$

interior of C

a point x is in the interior of C (ie, $x \in \text{int}\, C$) if all directions are feasible:

$$x + \varepsilon d \in C \quad \text{for all } d \in \mathbb{R}^n \text{ and } \varepsilon > 0 \text{ small enough}$$

If $g \in N_C(x)$ and $x \in \text{int}\, C$, then either

① $g^T(z-x) > 0$ for $z := x + \varepsilon d \in C$

or

② $g^T(z-x) < 0$ for $z := x - \varepsilon d \in C$

① and ② together imply that $g = 0$. Thus,

$$x \in \text{int}\, C \implies N_C(x) = \{0\}.$$

[also $\Leftarrow$ via supporting hyperplace theorem (extra credit!)]

Unconstrained Optimality

$$x^* \in \underset{x \in \mathbb{R}^n}{\text{argmin}}\ f(x) \iff -\nabla f(x^*) \in N_C(x) = \{0\} \iff \nabla f(x^*) = 0.$$

example: normal cone to affine set

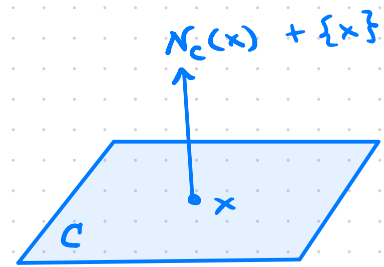
$$C = \{ x \in \mathbb{R}^n \mid Ax = b \} \qquad A \;\; m \times n \;\; \text{matrix}$$

define the "shifted" set

$$C_x = \{ z - x \mid z \in C \} = \text{null}(A)$$

(why?)

then



$$\begin{aligned} N_C(x) &= \{ g \mid g^T(z-x) \le 0 \;\; \forall z \in C \} \\ &= \{ g \mid g^T d \le 0 \;\; \forall d \in C_x \} \\ &= \{ g \mid g^T d \le 0 \;\; \forall d \in \text{null}(A) \} \\ &= \{ g \mid g^T d = 0 \;\; \forall d \in \text{null}(A) \} \\ &= \text{range}(A^T) \end{aligned}$$

because $d \in \text{null}(A) \iff -d \in \text{null}(A)$

Optimality for linearly-constrained optimization

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subj to} \quad Ax = b$$

a point $x^* \in C = \{x \mid Ax = b\}$ is optimal if and only if

$$-\nabla f(x^*) \in N_C(x^*)$$

$$\Updownarrow \quad \text{by previous slide}$$

$$-\nabla f(x^*) \in \operatorname{range}(A^T) = \{A^T y \mid y \in \mathbb{R}^m\}$$

$$\Updownarrow$$

$$\nabla f(x^*) = A^T y \quad \text{for some } y \in \mathbb{R}^m$$

vector of "Lagrange multipliers"