Convex Optimization

CPSC 406 – Computational Optimization

Convex Optimization

- optimality for convex problems
- normal cone
- Lagrange multipliers for linearly constrained problems

Optimality for convex problems

$$\min_{x} \{f(x) \mid x \in C\}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex differentiable
- $C \subseteq \mathbb{R}^n$ is convex
- x^* is optimal if all feasible directions are non-increasing in f
- if $C = \mathbb{R}^n$ the problem is **unconstrained**

$$x^* \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x) \iff 0 \leq f'(x^*, d) = \nabla f(x^*)^T d \quad \text{for all} \quad x^* + d \in \mathbb{R}^n$$

- implies $\nabla f(x^*) = 0$

Optimality – constrained

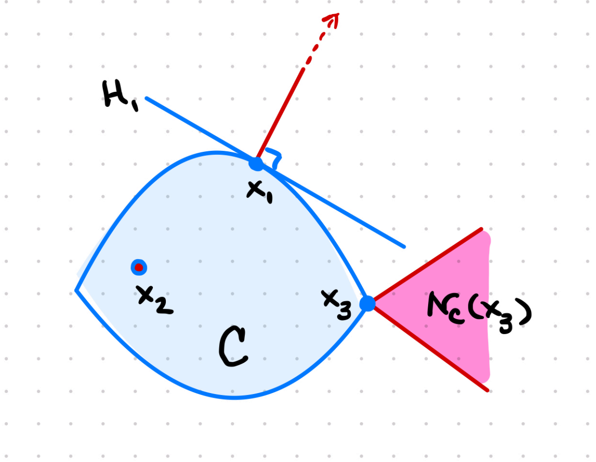
$$x^* \in \operatorname*{argmin}_{x \in C} f(x) \iff 0 \le f'(x^*, d) = \nabla f(x^*)^T d \quad \forall x^* + d \in C$$

- does not imply $\nabla f(x^*) = 0$

Normal cone

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

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



- $\mathcal{N}_C(x_1)$ is the normal to supporting hyperplane $H_1 = \{z \in \mathbb{R}^n \mid g^T z \leq g^T x_1\}$
- $\mathcal{N}_C(x_2) = \{0\}$ because x_2 is an interior point
- $\mathcal{N}_C(x_3)$ is the cone of normals at the vertex x_3

Question

What is the normal cone to the set

$$\mathbb{B}^n = \{x \in \mathbb{R}^2 \mid \|x\|_2 \leq 1\}$$

at the point $x = (1/\sqrt{2}, 1/\sqrt{2})$?

a. $\{0\}$
b. $\{(\lambda, \lambda) \mid \lambda \geq 0\}$
c. $\left\{\lambda(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \mid \lambda \leq 0\right\}$
d. $\left\{\mu(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}) \mid \mu \in \mathbb{R}\right\}$

Example

$$\min_{x_1, x_2 \geq 0} \tfrac{1}{2}(x_1 - 1)^2 + \tfrac{1}{2}(x_2 + 1)^2$$

Solution and gradient:

$$x^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \nabla f(x^*) = \begin{bmatrix} x_1^* - 1 \\ x_2^* + 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Normal cone at x^*:

$$\mathcal{N}_{\mathbb{R}_+^2}(x^*) = \left\{ \lambda \begin{bmatrix} 0 \\ -1 \end{bmatrix} : \lambda \geq 0 \right\}$$

Optimality:

$$-\nabla f(x^*) \in \mathcal{N}_{\mathbb{R}_+^2}(x^*)$$

Necessary and sufficient optimality

a point $x^* \in \mathrm{argmin}_{x \in C}\, f(x)$ if and only if

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

Use the definition of the normal code to deduce the equivalent condition

$$-\nabla f(x^*) \in \mathcal{N}_C(x^*)$$

Interior point

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

$$x + \epsilon d \in C \quad \forall d \in \mathbb{R}^n \text{ and } \epsilon > 0 \text{ small}$$

- if $g \in \mathcal{N}_C(x)$ and $x \in \mathrm{int}\, C$ then for every direction d,

$$\begin{aligned} 0 \le g^T(z - x) &= \epsilon g^T d && \text{for all} \quad z = x + \epsilon d \in C \\ 0 \le g^T(z - x) &= -\epsilon g^T d && \text{for all} \quad z = x - \epsilon d \in C \end{aligned}$$

- together, these imply $g = 0$, and thus

$$x \in \mathrm{int}\, C \implies \mathcal{N}_C(x) = \{0\}$$

[aside; the opposite implication is also true, but requires the supporting hyperplane theorem.]

- unconstrained optimality:

$$x^* \in \underset{x \in \mathbb{R}^n}{\mathrm{argmin}}\, f(x) \quad \iff \quad -\nabla f(x^*) \in \mathcal{N}_C(x) = \{0\} \quad \iff \quad \nabla f(x^*) = 0$$

Normal cone to an affine set

$$C = \{x \in \mathbb{R}^n \mid Ax = b\}, \quad A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m$$

For any $x \in C$, define the translated set

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

Then,

$$\begin{aligned}
\mathcal{N}_C(x) &= \{g \mid g^T(z - x) \leq 0 \quad \forall z \in C\} \\
&= \{g \mid g^T d \leq 0 \quad \forall d \in C_x\} \\
&= \{g \mid g^T d \leq 0 \quad \forall d \in \mathbf{null}(A)\} \\
&= \{g \mid g^T d = 0 \quad \forall d \in \mathbf{null}(A)\} \\
&= \mathbf{range}(A^T)
\end{aligned}$$

Linearly constrained optimization

$$\min_{x\in\mathbb{R}^n} \{f(x) \mid Ax = b\}$$

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

$$-\nabla f(x) \in \mathcal{N}_C(x^*) = \mathbf{range}(A^T)$$

or, equivalently,

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

- the vector $y = (y_1, \ldots, y_m)$ contains the **Lagrange multipliers** for each constraint $a_i^T x = b_i$