1st-order optimality conditions

Andersen Ang
ECS, Uni. Southampton, UK
andersen.ang@soton.ac.uk
Homepage angms.science

Version: April 11, 2023
First draft: June 6, 2021

Content

Normal cone

$$\mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) := \begin{cases} \Big\{ \boldsymbol{v} \in \mathbb{R}^n \;\big|\; \big\langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{s} \big\rangle \;\leq\; 0 \;\forall \boldsymbol{x} \in \mathcal{S} \Big\} & \boldsymbol{s} \in \mathrm{bd}\mathcal{S} \\ \{\boldsymbol{0}\} & \boldsymbol{s} \in \mathrm{int}\mathcal{S} \\ \varnothing & \boldsymbol{s} \notin \mathcal{S} \end{cases}$$

Given convex differentiable $f : \mathbb{R}^n \to \mathbb{R}$ and a convex set $\mathcal{X} \subset \mathbb{R}^n$,

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x} \in \mathcal{X}}{\operatorname{argmin}}\; f(\boldsymbol{x}) \quad \Longleftrightarrow \quad -\nabla f(\boldsymbol{x}^*) \in \mathcal{N}_{\mathcal{X}}(\boldsymbol{x}^*)$$

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmin}}\; f(\boldsymbol{x}) \quad \Longleftrightarrow \quad -\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$$

Setup

- Given $f : \mathbb{R}^n \to \mathbb{R}^n$ is convex with $\mathrm{dom} f = \mathbb{R}^n$ is a convex set
- Unconstrained problem
$$(\mathcal{U}) \ : \ \underset{\boldsymbol{x}\in\mathbb{R}^n}{\operatorname{argmin}} \ f(\boldsymbol{x})$$
- Constrained problem
$$(\mathcal{C}) \ : \ \underset{\boldsymbol{x}\in\mathcal{X}}{\operatorname{argmin}} \ f(\boldsymbol{x})$$
where $\mathcal{X} \subset \mathbb{R}^n$ is a convex set.
- FoC: **F**irst-order **o**ptimality **Condition**
  - For $(\mathcal{U}) \ : \ \underset{\boldsymbol{x}\in\mathbb{R}^n}{\operatorname{argmin}} \ f(\boldsymbol{x})$
$$\text{if } \boldsymbol{x}^* \text{ is a minimizer of } f, \text{ then } -\nabla f(\boldsymbol{x}^*) = \mathbf{0} \qquad (*)$$
  - FoC is also called Fermat's rule by French optimizers.
  - What about $(\mathcal{C}) \ : \ \underset{\boldsymbol{x}\in\mathcal{X}}{\operatorname{argmin}} \ f(\boldsymbol{x})$? We have a constraint $\boldsymbol{x} \in \mathcal{X}$, so $(*)$ does not hold.
- This PDF: answer such question using the notation of normal cone.

Remark: we only assume f is convex but not strictly convex so it is possible $\mathcal{U}, \mathcal{C}$ have multiple solution.

FoC for $(\mathcal{C})$: $\underset{\boldsymbol{x}\in\mathcal{X}}{\operatorname{argmin}}\ f(\boldsymbol{x})$

- If $\boldsymbol{x}^*$ is solution of $\mathcal{C}$, then

$$-\nabla f(\boldsymbol{x}^*) = \mathbf{0} \quad \text{and} \quad \boldsymbol{x}^* \in \mathcal{C} \qquad \text{(WRONG)}$$

$$-\nabla f(\boldsymbol{x}^*) \ \in \ \mathcal{N}_{\mathcal{X}}(\boldsymbol{x}^*)$$

where $\mathcal{N}_{\mathcal{X}}$ is the **normal cone** of $\mathcal{X}$ at $\boldsymbol{x}^*$

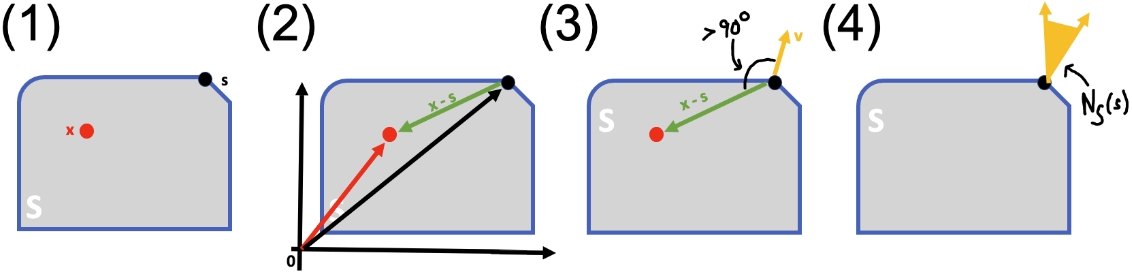
- **Definition** The **normal cone** of a set $\mathcal{S} \subset \mathbb{R}^n$ at the point $\boldsymbol{s}$ is defined as

$$\mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) \ := \begin{cases} \left\{\boldsymbol{v} \in \mathbb{R}^n \mid \langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{s}\rangle \ \leq \ 0 \ \forall \boldsymbol{x} \in \mathcal{S}\right\} & \boldsymbol{s} \in \mathcal{S} \\ \varnothing & \boldsymbol{s} \notin \mathcal{S} \end{cases}$$

You are given two mathematical objects

- a set $\mathcal{S}$ — $\mathcal{S}$ can be convex or nonconvex
- a point $\boldsymbol{s}$ — $\boldsymbol{s}$ can be inside $\mathcal{S}$ or outside $\mathcal{S}$

Normal cone $\mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) := \begin{cases} \left\{ \boldsymbol{v} \in \mathbb{R}^n \mid \langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{s} \rangle \leq 0 \ \forall \boldsymbol{x} \in \mathcal{S} \right\} & \boldsymbol{s} \in \mathcal{S} \\ \varnothing & \boldsymbol{s} \notin \mathcal{S} \end{cases}$



Explanation

1. Given a set $\mathcal{S}$, a point $\boldsymbol{x} \in \mathcal{S}$ and a point $\boldsymbol{s} \in \operatorname{bd}\mathcal{S}$
2. The vector $\boldsymbol{x} - \boldsymbol{s}$
3. A vector $\boldsymbol{v}$ that $\langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{s} \rangle < 0$, the angle between is larger than 90°
4. The normal cone at $\boldsymbol{s}$ = the set of all possible $\boldsymbol{v}$

Properties of normal cone

- It is a cone: $\boldsymbol{v} \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) \implies \lambda\boldsymbol{v} \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s})$ for all $\lambda \geq 0$.
  - **Proof** Fix $\boldsymbol{v} \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s})$, by definition of normal cone

$$
\begin{aligned}
\boldsymbol{v} \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) \quad &\iff \quad \big\langle \boldsymbol{v}, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \\
&\iff \quad \lambda\big\langle \boldsymbol{v}, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \\
&\iff \quad \big\langle \lambda\boldsymbol{v}, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \\
&\iff \quad \lambda\boldsymbol{v} \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s})
\end{aligned}
$$

- It is a convex set: $\boldsymbol{v}_1 \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s})$ and $\boldsymbol{v}_2 \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) \implies \lambda\boldsymbol{v}_1 + (1-\lambda)\boldsymbol{v}_2 \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s})$ for all $\lambda \in (0,1)$.
  - **Proof** Fix $\boldsymbol{v}_1 \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s})$ and $\boldsymbol{v}_2 \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s}$, by definition of normal cone

$$
\begin{aligned}
\boldsymbol{v}_1 \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s}),\ \boldsymbol{v}_2 \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) \quad &\iff \quad \big\langle \boldsymbol{v}_1, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \ \text{ and } \ \big\langle \boldsymbol{v}_2, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \\
&\iff \quad \lambda\big\langle \boldsymbol{v}_1, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \ \text{ and } \ (1-\lambda)\big\langle \boldsymbol{v}_2, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \\
&\iff \quad \big\langle \lambda\boldsymbol{v}_1, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \ \text{ and } \ \big\langle (1-\lambda)\boldsymbol{v}_2, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \\
&\iff \quad \big\langle \lambda\boldsymbol{v}_1, \boldsymbol{x}-\boldsymbol{s} \big\rangle + \big\langle (1-\lambda)\boldsymbol{v}_2, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \\
&\iff \quad \big\langle \lambda\boldsymbol{v}_1 + (1-\lambda)\boldsymbol{v}_2, \boldsymbol{x}-\boldsymbol{s} \big\rangle \leq 0 \\
&\iff \quad \lambda\boldsymbol{v}_1 + (1-\lambda)\boldsymbol{v}_2 \in \boldsymbol{N}_{\mathcal{S}}(\boldsymbol{s})
\end{aligned}
$$

The "full" formula $\mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) \coloneqq \begin{cases} \Big\{\boldsymbol{v} \in \mathbb{R}^n \mid \langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{s} \rangle \le 0 \; \forall \boldsymbol{x} \in \mathcal{S} \Big\} & \boldsymbol{s} \in \mathrm{bd}\mathcal{S} \\ \{\mathbf{0}\} & \boldsymbol{s} \in \mathrm{int}\mathcal{S} \\ \varnothing & \boldsymbol{s} \notin \mathcal{S} \end{cases}$

- If $\boldsymbol{s}$ inside $\mathcal{S}$, there are two sub-cases
  - $\boldsymbol{s} \in \mathrm{int}\mathcal{S}$: the point $\boldsymbol{s}$ is inside the interior of $\mathcal{S}$
  - $\boldsymbol{s} \in \mathrm{bd}\mathcal{S}$: the point $\boldsymbol{s}$ is on the boundary of $\mathcal{S}$

- Proving $\mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) \overset{\boldsymbol{s} \in \mathrm{int}\mathcal{S}}{=} \{\mathbf{0}\}$
  - $\boldsymbol{s} \in \mathrm{int}\mathcal{S} \implies \exists \delta > 0 \,\text{s.t.}\, \mathbb{B}(\boldsymbol{s}; \delta) \subset \mathcal{S}$. interior implies existence of internal enclosing ball
  - Let $\boldsymbol{v} \in \mathcal{N}_{\mathcal{S}}(\boldsymbol{s})$. $\boldsymbol{v}$ is the normal cone of $\boldsymbol{s}$
    Then $\langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{s} \rangle \le 0 \; \forall \boldsymbol{x} \in \mathcal{S}$. by definition of normal cone
  - Let $\boldsymbol{x} \in \mathcal{S}$ and $t > 0$ sufficiently small s.t. $\boldsymbol{s} + t\boldsymbol{v} \in \mathbb{B}(\boldsymbol{s}; \delta)$. move $\boldsymbol{s}$ slightly along $\boldsymbol{v}$ will still stay in $\mathbb{B}$
  - By [blue] and [red], $\boldsymbol{s} + t\boldsymbol{v}$ is inside $\mathcal{S}$ so we can make use of [green], which gives

$$\langle \boldsymbol{v}, \underbrace{\boldsymbol{s} + t\boldsymbol{v}}_{\boldsymbol{x} \text{ in [green]}} - \boldsymbol{s} \rangle \le 0 \; \forall \boldsymbol{x} \in \mathcal{S}$$

$$\iff t\langle \boldsymbol{v}, \boldsymbol{v} \rangle \le 0$$

As $t > 0$, so $\|\boldsymbol{v}\|^2 \le 0$ which implies $\boldsymbol{v} = \mathbf{0}$

FoC of constrained minimization

- For $(\mathcal{C}) \;:\; \underset{\boldsymbol{x}\in\mathcal{X}}{\operatorname{argmin}}\; f(\boldsymbol{x})$ with a convex differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and a convex set $\mathcal{X} \subset \mathbb{R}^n$,

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x}\in\mathcal{X}}{\operatorname{argmin}}\; f \quad \iff \quad -\nabla f(\boldsymbol{x}^*) \in \mathcal{N}_{\mathcal{X}}(\boldsymbol{x}^*)$$

- Remark: we assume f is convex but not strictly convex so $\mathcal{C}$ may have multiple solutions, thus the set $\operatorname{argmin} f$ is possibly not a singleton, so we write $\boldsymbol{x}^* \in \operatorname{argmin} f$ but not $\boldsymbol{x}^* = \operatorname{argmin} f$.
- **Proof by contradiction**
  - Assume $\boldsymbol{x}^* \in \operatorname{argmin} f$ and $-\nabla f(\boldsymbol{x}^*) \notin \mathcal{N}_{\mathcal{X}}(\boldsymbol{x}^*)$.
  - By definition of normal cone, if $-\nabla f(\boldsymbol{x}^*) \notin \mathcal{N}_{\mathcal{X}}(\boldsymbol{x}^*)$, it means there exists a direction $\boldsymbol{d} \in \mathcal{X}$ that is positively correlates with $-\nabla f(\boldsymbol{x}^*)$
  - Moving along $\boldsymbol{d}$ will decrease the objective function, therefore $\boldsymbol{x}^*$ cannot be a minimizer, contradiction.

Corollary: FoC of unconstrained mininization

- For $(\mathcal{U}) \; : \; \underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{argmin}} \; f(\boldsymbol{x})$ with a convex differentiable function $f : \mathbb{R}^n \to \mathbb{R}$

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

- Remark: we assume f is convex but not strictly convex so $\mathcal{C}$ may have multiple solutions, thus the set $\text{argmin } f$ is possibly not a singleton, so we write $\boldsymbol{x}^* \in \text{argmin } f$ but not $\boldsymbol{x}^* = \text{argmin } f$.
- **Proof**
  - $-\nabla f(\boldsymbol{x}^*) \in \mathcal{N}_{\mathbb{R}^n}(\boldsymbol{x}^*)$
  - By , $\boldsymbol{x}^* \in \mathbb{R}^n$ so $\boldsymbol{x}^*$ is an interior point of $\mathbb{R}^n$
  - By "normal cone at interior point is singleton of zero", so $\mathcal{N}_{\mathbb{R}^n}(\boldsymbol{x}^*) = \{\mathbf{0}\}$
  - We now have $-\nabla f(\boldsymbol{x}^*) \in \{\mathbf{0}\} \implies -\nabla f(\boldsymbol{x}^*) = \mathbf{0}$

Last page - summary

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable

1. $\boldsymbol{x}^* \in \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{argmin}} \; f(\boldsymbol{x}) \iff -\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ unconstrained minimization

2. $\boldsymbol{x}^* \in \underset{\boldsymbol{x} \in \mathcal{C}}{\operatorname{argmin}} \; f(\boldsymbol{x}) \iff -\nabla f(\boldsymbol{x}^*) \in \mathcal{N}_{\mathcal{C}}(\boldsymbol{x}^*)$ constrained minimization

3. Normal cone

$$\mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) := \begin{cases} \left\{ \boldsymbol{v} \in \mathbb{R}^n \mid \langle \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{s} \rangle \leq 0 \; \forall \boldsymbol{x} \in \mathcal{S} \right\} & \boldsymbol{s} \in \mathcal{S} \\ \varnothing & \boldsymbol{s} \notin \mathcal{S} \end{cases}$$

4. $\mathcal{N}_{\mathcal{S}}(\boldsymbol{s}) = \{\mathbf{0}\}$ if $\boldsymbol{s} \in \operatorname{int}\mathcal{S}$

5. $\{3 \implies 2\} + 4 \implies 1$