Lecture 3: “Primal” optimality conditions

Michael Patriksson

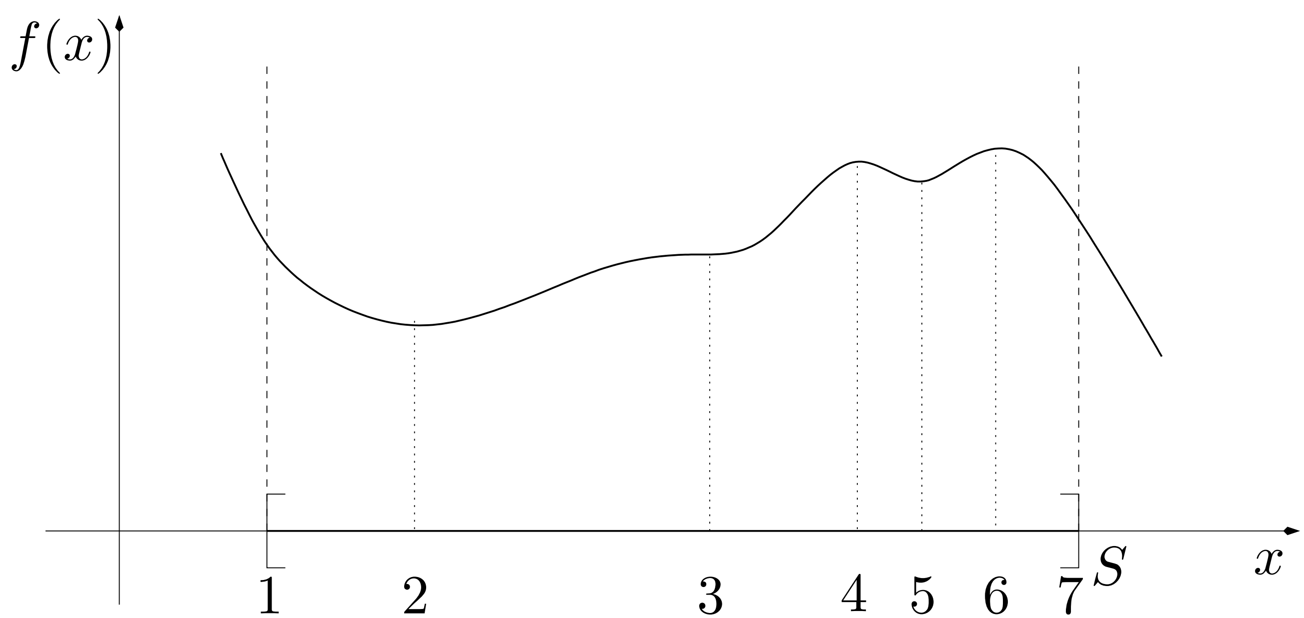
22 January 2003

Local and global optimality, $\mathbb{R}$ and $\mathbb{R}^n$

$$\text{minimize } f(\boldsymbol{x}), \tag{1a}$$

$$\text{subject to } \boldsymbol{x} \in S, \tag{1b}$$

$S \subseteq \mathbb{R}^n$ non-empty set, $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$ a given function.



Interesting points

(i) boundary points of S;

(ii) stationary points, that is, where $f'(x) = 0$;

(iii) discontinuities in f or f'.

Here:

(i) 1, 7;

(ii) 2, 3, 4, 5, 6;

(iii) none.

Global and local minimum

- $\boldsymbol{x}^* \in S$ is a *global minimum* of f over S if it attains the lowest value of f over S:

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \qquad \boldsymbol{x} \in S.$$

- $\boldsymbol{x}^* \in S$ is a *local minimum* of f over S if there exists a small enough ball intersected with S around $\boldsymbol{x}^*$ such that it is an optimal solution in that smaller set: with $B_\varepsilon(\boldsymbol{x}^*) := \{\, \boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y} - \boldsymbol{x}^*\| < \varepsilon \,\}$ being the Euclidean ball with radius ε centered at $\boldsymbol{x}^*$, we get

$$\exists \varepsilon > 0 \text{ such that } f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \qquad \boldsymbol{x} \in S \cap B_\varepsilon(\boldsymbol{x}^*).$$

- $\boldsymbol{x}^* \in S$ is a *strict local minimum* of f over S if $f(\boldsymbol{x}^*) < f(\boldsymbol{x})$ holds above for $\boldsymbol{x} \neq \boldsymbol{x}^*$.

Fundamental Theorem of global optimality

Consider the problem (1), *where S is a convex set and f is convex on S. Then, every local minimum of f over S is also a global minimum.*

Proof. Suppose that $\boldsymbol{x}^*$ is a local minimum but not a global one, while $\bar{\boldsymbol{x}}$ is a global minimum. Then, $f(\bar{\boldsymbol{x}}) < f(\boldsymbol{x}^*)$. Let $\lambda \in (0, 1)$. By the convexity of S and f, $\lambda\bar{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^* \in S$, and $f(\lambda\bar{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^*) \leq \lambda f(\bar{\boldsymbol{x}}) + (1-\lambda)f(\boldsymbol{x}^*) < f(\boldsymbol{x}^*)$. Choosing $\lambda > 0$ small enough then leads to a contradiction to the local optimality of $\boldsymbol{x}^*$. □

Intuitive image: If $\boldsymbol{x}^*$ is a local minimum, then f cannot go down-hill from $\boldsymbol{x}^*$ in any direction, but if $\bar{\boldsymbol{x}}$ has a lower value, then f has to go down-hill sooner or later. This cannot be the shape of any convex function.

Weak coercivity

$S \subseteq \mathbb{R}^n$ non-empty and closed, $f : S \mapsto \mathbb{R}$.

- f is *weakly coercive* with respect to the set S if S is bounded or the sequence $\{f(\boldsymbol{x}_k)\}$ tends to infinity whenever the sequence $\{\boldsymbol{x}_k\} \subset S$ tends to infinity in norm.
  In other words, f is weakly coercive if either S is bounded or
  $$\lim_{\substack{\|\boldsymbol{x}_k\| \to \infty \\ \boldsymbol{x}_k \in S}} f(\boldsymbol{x}_k) = \infty$$
  holds.
- The weak coercivity of $f : S \mapsto \mathbb{R}$ is equivalent to the property that f has bounded level sets.

Existence of optimal solutions, I: Weierstrass

Weierstrass' Theorem *Let $S \subseteq \mathbb{R}^n$ be a non-empty and closed set, and $f : S \mapsto \mathbb{R}$ be a continuous function on S. If f is weakly coercive with respect to S, then there exists a non-empty, closed and bounded (thus compact) set of optimal solutions to the problem* (1).

Convex polynomials *Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex polynomial; S can be described by inequality constraints $g_i(\boldsymbol{x}) \leq 0$, $i = 1, \dots, m$, where each function g_i is convex and polyhedral. The problem* (1) *then is convex. Moreover, it has a non-empty (as well as closed and convex) set of optimal solutions if and only if f is lower bounded on S.*

Existence of solutions, II: The Frank–Wolfe Theorem with specialization to LP

- Suppose

$$S = \{\, \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}; \quad \boldsymbol{E}\boldsymbol{x} = \boldsymbol{d} \,\},$$

$\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{E} \in \mathbb{R}^{\ell \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, and $\boldsymbol{d} \in \mathbb{R}^\ell$.

- The *recession cone* to S is the following set, defining the set of directions that are feasible at every point in S:

$$\mathrm{rec}_S := \{\, \boldsymbol{p} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{p} \le \boldsymbol{0}^m; \quad \boldsymbol{E}\boldsymbol{p} = \boldsymbol{0}^\ell \,\}.$$

We also suppose that $f(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{x}$, $\boldsymbol{x} \in \mathbb{R}^n$, where $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ is a symmetric and positive semi-definite matrix and $\boldsymbol{q} \in \mathbb{R}^n$.

- The *recession cone* to f is the recession cone to the level set of f for any value of b (same for every b!). In the special case of the convex quadratic function f,

$$\mathrm{rec}_f = \{\, \boldsymbol{p} \in \mathbb{R}^n \mid \boldsymbol{Q}\boldsymbol{p} = \boldsymbol{0}^n; \quad \boldsymbol{q}^{\mathrm{T}}\boldsymbol{p} \leq 0 \,\}.$$

(Set of directions that nowhere are ascent directions.)

- *Three equivalent statements:*

(a) *The problem* (1) *has a non-empty (as well as a closed and convex) set of optimal solutions.*

(b) f *is lower bounded on* S*.*

(c) *For every vector* $\boldsymbol{p}$ *in the intersection of the recession cone* rec_S *to* S *and the null space* $N(\boldsymbol{Q})$ *of the matrix* $\boldsymbol{Q}$*, it holds that* $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{p} \geq 0$*. In other words,*

$$\boldsymbol{p} \in \mathrm{rec}_S \cap N(\boldsymbol{Q}) \qquad \Longrightarrow \qquad \boldsymbol{q}^{\mathrm{T}}\boldsymbol{p} \geq 0.$$

- Improvement over Weierstrass' Theorem: If in (1), f is convex on S where the latter is non-empty, closed and convex, then the problem has a non-empty, convex and compact set of optimal solutions iff $\mathrm{rec}_S \cap \mathrm{rec}_f = \{\mathbf{0}^n\}$.
- Interesting implication for LP: *Suppose f is linear. Three equivalent statements:*
  (a) *The problem* (1) *has a non-empty (polyhedral) set of optimal solutions.*
  (b) *f is lower bounded on S.*
  (c) *For every vector $\boldsymbol{p}$ in the recession cone rec_S to S, it holds that $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{p} \geq 0$. In other words,*

$$\boldsymbol{p} \in \mathrm{rec}_S \qquad \Longrightarrow \qquad \boldsymbol{q}^{\mathrm{T}}\boldsymbol{p} \geq 0.$$

- Corresponds (of course) exactly to the LP results to follow.
- Lower bounded not enough in general; cf. $f(x) = 1/x$ on $x \geq 1$.

Optimality over $\mathbb{R}^n$, $f \in C^1$

- $\boldsymbol{x}^*$ *is a local minimum of* f *on* $\mathbb{R}^n$ $\quad\Longrightarrow\quad \nabla f(\boldsymbol{x}^*) = \mathbf{0}^n$.
- Proof by Taylor expansion, contradiction.
- Direction $\Longleftarrow$ not true: $f(x) = x^3$, $x = 0$.
- Let $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{\pm\infty\}$ be given. Let $\boldsymbol{x} \in \mathbb{R}^n$ be a vector such that $f(\boldsymbol{x})$ is finite. Let $\boldsymbol{p} \in \mathbb{R}^n$. We say that the vector $\boldsymbol{p} \in \mathbb{R}^n$ is a *descent direction* with respect to f at $\boldsymbol{x}$ if

$$\exists \delta > 0 \text{ such that } f(\boldsymbol{x} + \alpha \boldsymbol{p}) < f(\boldsymbol{x}) \text{ for every } \alpha \in (0, \delta].$$

- Sufficient condition: *Suppose that* $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$ *is in* C^1 *around a point* $\boldsymbol{x}$ *for which* $f(\boldsymbol{x}) < +\infty$, *and that* $\boldsymbol{p} \in \mathbb{R}^n$. *If* $\nabla f(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{p} < 0$ *then the vector* $\boldsymbol{p}$ *defines a direction of descent with respect to* f *at* $\boldsymbol{x}$.

Optimality over $\mathbb{R}^n$, $f \in C^2$

- $\boldsymbol{x}^*$*is a local minimum of* f *on* $\mathbb{R}^n$ $\Longrightarrow$
$$\begin{cases} \nabla f(\boldsymbol{x}^*) = \mathbf{0}^n; \\ \nabla^2 f(\boldsymbol{x}^*) \text{ is positive semi-definite.} \end{cases}$$
- [*Note:* $n = 1$: $x^* \in \mathbb{R}$ is a local minimum $\Longrightarrow f'(x^*) = 0$ and $f''(x^*) \geq 0$.]
- $$\left.\begin{array}{r} \nabla f(\boldsymbol{x}^*) = \mathbf{0}^n \\ \nabla^2 f(\boldsymbol{x}^*) \text{ is positive definite} \end{array}\right\} \Longrightarrow$$
$\boldsymbol{x}^*$*is a strict local minimum of* f *on* $\mathbb{R}^n$.
- [*Note:* $n = 1$: $f'(x^*) = 0$ and $f''(x^*) > 0 \Longrightarrow x^* \in \mathbb{R}$ is a strict local minimum.]

Optimality over $\mathbb{R}^n$, f convex in C^1

- *Let $f \in C^1$, and f be convex. Then,*

$$\boldsymbol{x}^* \textit{ is a global minimum of } f \textit{ on } \mathbb{R}^n \quad \Longleftrightarrow \quad \nabla f(\boldsymbol{x}^*) = \mathbf{0}^n.$$

- *Proof.* $\Longrightarrow$ Global min means local min means stationary.
$\Longleftarrow$ Convexity of f yields that for every $\boldsymbol{y} \in \mathbb{R}^n$,

$$\begin{aligned} f(\boldsymbol{y}) &\geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}^*) \\ &= f(\boldsymbol{x}^*), \end{aligned}$$

the equality from the property that $\nabla f(\boldsymbol{x}^*) = \mathbf{0}^n$, by assumption.

Active constraints

Consider

$$\text{minimize } f(\boldsymbol{x}),$$
$$\text{subject to } \boldsymbol{x} \in S.$$

$S \subseteq \mathbb{R}^n$ non-empty, closed, convex, $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$ in C^1 on S.

- Feasible directions at $\boldsymbol{x}^*$ depend on active constraints.
- Let $\boldsymbol{x} \in S$, where $S \subseteq \mathbb{R}^n$, and that $\boldsymbol{p} \in \mathbb{R}^n$. Then, $\boldsymbol{p}$ defines a *feasible direction* at $\boldsymbol{x}$ if

$$\exists \delta > 0 \text{ such that } \boldsymbol{x} + \alpha \boldsymbol{p} \in S \text{ for all } \alpha \in [0, \delta].$$

- Suppose

$$S = \{\, \boldsymbol{x} \in \mathbb{R}^n \mid g_i(\boldsymbol{x}) \leq 0, \quad i \in \mathcal{I}; \quad g_i(\boldsymbol{x}) = 0, \quad i \in \mathcal{E} \,\}.$$

  Suppose $\boldsymbol{x} \in S$. The set of *active constraints* is the union of all the equality constraints and the set of inequality constraints that are satisfied with equality, that is, the set $\mathcal{E} \cup \mathcal{I}(\boldsymbol{x})$, where $\mathcal{I}(\boldsymbol{x}) := \{\, i \in \mathcal{I} \mid g_i(\boldsymbol{x}) = 0 \,\}$.

- Linear constraints: $g_i(\boldsymbol{x}) := \boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{x} - d_i$ $(i \in \mathcal{E})$, $g_i(\boldsymbol{x}) := \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} - b_i$ $(i \in \mathcal{I})$.

- Matrix notation: $S = \{\, \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{E}\boldsymbol{x} = \boldsymbol{d}; \quad \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \,\}$.

- Feasible directions at $\boldsymbol{x} \in S$:

$$\{\, \boldsymbol{p} \in \mathbb{R}^n \mid \boldsymbol{E}\boldsymbol{p} = \boldsymbol{0}^{\ell}; \quad \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{p} \leq 0, \ i \in \mathcal{I}(\boldsymbol{x}) \,\}.$$

- For nonlinear constraints: more technical! Later!

Necessary optimality conditions, I: VIP

- *Suppose* $S \subseteq \mathbb{R}^n$, $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$ *is in* C^1 *around* $\boldsymbol{x} \in S$ *for which* $f(\boldsymbol{x}) < +\infty$.
  (a) *If* $\boldsymbol{x}^* \in S$ *is a local minimum of* f *on* S *then* $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{p} \geq 0$ *holds for every feasible direction* $\boldsymbol{p}$ *at* $\boldsymbol{x}^*$.
  (b) *Suppose that* S *is convex and that* f *is in* C^1 *on* S. *If* $\boldsymbol{x}^* \in S$ *is a local minimum of* f *on* S *then*

$$\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \qquad \boldsymbol{x} \in S. \tag{2}$$

- *Proof.* (a) Taylor expansion of f around $\boldsymbol{x}^*$ and proof by contradiction. Similar to unconstrained case.
  (b) If S is convex then every feasible direction $\boldsymbol{p}$ is a positive scalar times $\boldsymbol{x} - \boldsymbol{x}^*$ for *some* vector $\boldsymbol{x} \in S$. The expression (2) follows from the statement in (a). □

Convex case

- We refer to (2) as a *variational inequality*.
- *Suppose* $S \subseteq \mathbb{R}^n$ *is non-empty and convex. Let* $f \in C^1$ *on* S*, convex. Then,*

$$\boldsymbol{x}^* \textit{ is a global minimum of } f \textit{ on } S \quad \iff \quad (2) \textit{ holds.}$$

- *Proof.* [$\Longrightarrow$] A global min is a local min. Follows then from the above result.

  [$\Longleftarrow$] The convexity of f yields for every $\boldsymbol{y} \in S$ that

$$\begin{aligned} f(\boldsymbol{y}) &\geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}^*) \\ &\geq f(\boldsymbol{x}^*). \end{aligned}$$

  Second inequality from (2). □

- Compare with the case $S = \mathbb{R}^n$!

Separation Theorem revisited (proof)

- *Suppose that* $C \subseteq \mathbb{R}^n$ *is closed and convex, and that the point* $\boldsymbol{y}$ *does not lie in* C*. Then there exist a real* α *and an* $n \times 1$ *vector* $\boldsymbol{\pi} \neq \mathbf{0}^n$ *such that* $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{y} > \alpha$ *and* $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{x} \leq \alpha$ *for all* $\boldsymbol{x} \in C$*.*
- *Proof.* Assume C is nonempty, and define the function $f : \mathbb{R}^n \mapsto \mathbb{R}$ by $f(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2/2$. By Weierstrass' Theorem there exists a minimizer $\boldsymbol{x}^*$, which by the above conditions satisfies $(\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0$ for all $\boldsymbol{x} \in C$ (since $-\nabla f(\boldsymbol{x}^*) = \boldsymbol{y} - \boldsymbol{x}^*$). Setting $\boldsymbol{\pi} = \boldsymbol{y} - \boldsymbol{x}^*$ and $\alpha = (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{x}^*$ gives the result. □
- $\boldsymbol{y} \notin C$: $0 < \boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{y} - \alpha = (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{y} - (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{x}^* = \|\boldsymbol{y} - \boldsymbol{x}^*\|^2$.
- $\boldsymbol{x} \in C$: $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{x} \leq \alpha \iff (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{x} \leq (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{x}^* \iff (\boldsymbol{y} - \boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0$.
- The hyperplane is a tangent to C, the normal is $\boldsymbol{y} - \boldsymbol{x}^*$.

Necessary optimality conditions, II: Projection

- $\boldsymbol{x}^*$ *is stationary iff*

$$\boldsymbol{x}^* = \mathrm{Proj}_S[\boldsymbol{x}^* - \nabla f(\boldsymbol{x}^*)].$$

- In other words, $\boldsymbol{x}^*$ is stationary if and only if a step in the direction of the steepest descent direction followed by a Euclidean projection onto S means that we have not moved at all. (If not, then we obtain a descent direction towards that projected point—basis for the *projection method* in Chapter 12.)
- *Proof.* Write the projection problem as ($\boldsymbol{z} = \boldsymbol{x}^* - \nabla f(\boldsymbol{x}^*)$)

$$\underset{\boldsymbol{x} \in S}{\text{minimize}}\ h(\boldsymbol{x}) := \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|^2.$$

Necessary optimality conditions [$\nabla h(\boldsymbol{x}^*) = \boldsymbol{x}^* - \boldsymbol{z}$]

$$(\boldsymbol{x}^* - \boldsymbol{z})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}^*) \geq 0, \qquad \boldsymbol{y} \in S \quad \Longleftrightarrow \quad (2)! \qquad \square$$

Necessary optimality conditions, III: LP

- *x^* is stationary iff*

$$\underset{\boldsymbol{x} \in S}{\text{minimum}} \; \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) = 0.$$

- *Proof.* If $\boldsymbol{x}^*$ does not minimize $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*)$ over $\boldsymbol{x} \in S$ then the optimal value, $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\bar{\boldsymbol{x}} - \boldsymbol{x}^*)$, is negative, hence $\boldsymbol{p} := \bar{\boldsymbol{x}} - \boldsymbol{x}^*$ feasible descent direction. (Zero always attainable, by letting $\boldsymbol{x} = \boldsymbol{x}^*$.) □

- Method basis: given $\boldsymbol{x}_k \in S$, find out if we are stationary by minimizing $\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_k)$ over $\boldsymbol{x} \in S$. In some sense, we find the $\boldsymbol{x} \in S$ which "violates optimality the most." Perform a line search in the direction from $\boldsymbol{x}_k$ towards that point. Repeat until convergence.

- Names: *Frank–Wolfe, Simplicial decomposition.* Chapter 12.

Necessary optimality conditions, IV: Normal cone

- If we wish to project $\boldsymbol{z} \in \mathbb{R}^n$ onto S, then the resulting (unique) projection is the vector $\boldsymbol{x}$ for which the following holds:

$$[\boldsymbol{x} - \boldsymbol{z}]^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \geq 0, \qquad \boldsymbol{y} \in S$$

  that is

$$[\boldsymbol{z} - \boldsymbol{x}]^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \leq 0, \qquad \boldsymbol{y} \in S.$$

- Interpretation: the angle between the two vectors $\boldsymbol{z} - \boldsymbol{x}$ (the vector that points towards the point being projected) and the vector $\boldsymbol{y} - \boldsymbol{x}$ (the vector that points towards any vector $\boldsymbol{y} \in S$) is $\geq 90°$. So, the projection operation has the characterization

$$[\boldsymbol{z} - \mathrm{Proj}_S(\boldsymbol{z})]^{\mathrm{T}}(\boldsymbol{y} - \mathrm{Proj}_S(\boldsymbol{z})) \leq 0, \qquad \boldsymbol{y} \in S. \tag{3}$$

- Interesting with $\boldsymbol{z} = \boldsymbol{x}^* - \nabla f(\boldsymbol{x}^*)$:

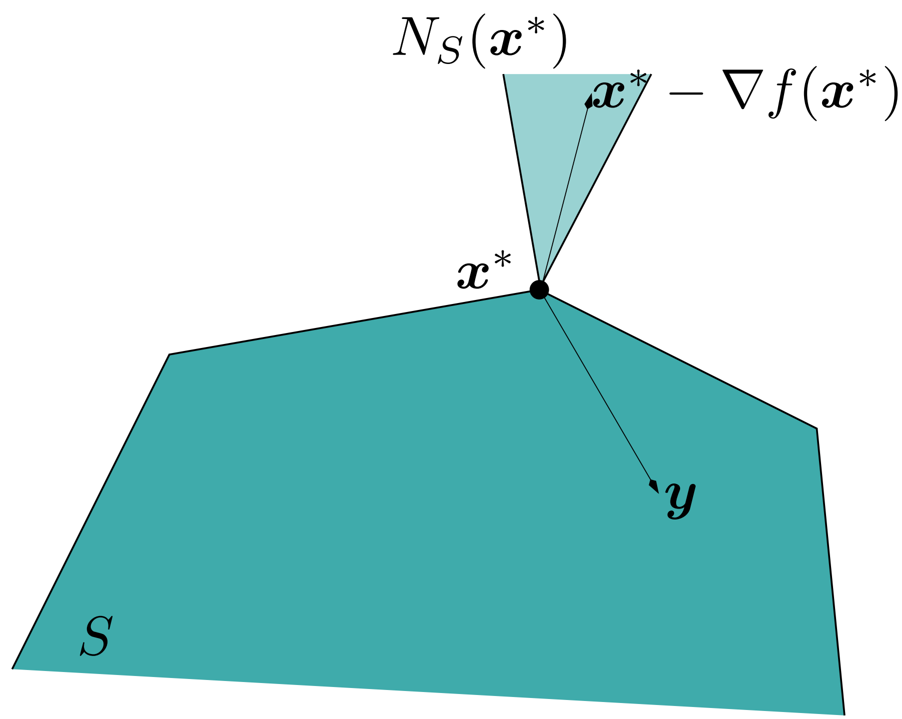


Figure 1: Normal cone characterization of a stationary point.

- Suppose $S \subseteq \mathbb{R}^n$ is closed and convex. Let $\boldsymbol{x} \in \mathbb{R}^n$. Then, the *normal cone* to S at $\boldsymbol{x}$ is the set

$$N_S(\boldsymbol{x}) := \begin{cases} \{\, \boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{v}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \leq 0, \qquad \boldsymbol{y} \in S \,\}, & \text{if } \boldsymbol{x} \in S, \\ \emptyset & \text{otherwise.} \end{cases}$$

- Characterization of stationary point at $\boldsymbol{x}^*$, number IV:

$$-\nabla f(\boldsymbol{x}^*) \in N_S(\boldsymbol{x}^*). \tag{4}$$

- Geometric interpretation: the angle between the negative gradient and any feasible direction is $\geq 90°$ ($\nexists$ feasible descent directions).

- S is a subspace $\Longrightarrow$ (4) states that $\nabla f(\boldsymbol{x}^*)$ is parallel to the normal of the subspace!

Characterization of globally optimal solutions, convex case

- In $\mathbb{R}^n$: $\nabla f(\boldsymbol{x}^*) = \mathbf{0}^n$.
- In $S \subset \mathbb{R}^n$, S convex? Let S^* denote the set of optimal solutions.
- *The value of* $\nabla f(\boldsymbol{x})$ *is constant on* S^**. Suppose* $\boldsymbol{x}^* \in S^*$*. Then,*

$$S^* = \{\, \boldsymbol{x} \in S \mid \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) = 0 \textit{ and } \nabla f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}^*) \,\}.$$

- *Proof.* Let $\boldsymbol{x}^* \in S^*$. The convexity of f gives

$$f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geq \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*), \qquad \boldsymbol{x} \in \mathbb{R}^n.$$

Let $\bar{\boldsymbol{x}} \in S^*$. It follows that $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\bar{\boldsymbol{x}} - \boldsymbol{x}^*) = 0$. Substitute $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{x}^*$ for $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\bar{\boldsymbol{x}}$ above and use $f(\boldsymbol{x}^*) = f(\bar{\boldsymbol{x}}) \Longrightarrow$

$$f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}}) \geq \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \bar{\boldsymbol{x}}), \qquad \boldsymbol{x} \in \mathbb{R}^n,$$

which is equivalent to the statement that $\nabla f(\bar{\boldsymbol{x}}) = \nabla f(\boldsymbol{x}^*)$. $\square$

Near-optimality (ε-optimality)

- f convex and in C^1 on S, S non-empty, convex. Wish to find means to terminate an algorithm when we are "near" a solution.
- The "Frank–Wolfe" subproblem is:

$$\bar{\boldsymbol{y}} \in \arg\operatorname*{minimum}_{\boldsymbol{y}\in S}\ z(\boldsymbol{y}) := \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}).$$

$$f(\boldsymbol{x}) + z(\bar{\boldsymbol{y}}) = f(\boldsymbol{x}) + \min_{\boldsymbol{y}\in S}\ z(\boldsymbol{y}) \le f(\boldsymbol{x}) + z(\boldsymbol{x}^*) \le f(\boldsymbol{x}^*) \le f(\boldsymbol{x}).$$

- $\therefore f^* \in [f(\boldsymbol{x}) + z(\bar{\boldsymbol{y}}), f(\boldsymbol{x})]$ (lower and upper bounds on f^*)
- Suppose $z(\bar{\boldsymbol{y}}) \ge -\varepsilon$, $\varepsilon > 0$ small.
- $f(\boldsymbol{x}^*) \ge f(\boldsymbol{x}) + z(\bar{\boldsymbol{y}}) \ge f(\boldsymbol{x}) - \varepsilon$, that is,

$$f(\boldsymbol{x}^*) \ge f(\boldsymbol{x}) - \varepsilon, \quad \text{or,} \quad f(\boldsymbol{x}) \le f^* + \varepsilon. \tag{5}$$

We call $\boldsymbol{x} \in S$ satisfying (5) an ε-*optimal solution*.

Contraction property of the projection operation

- Suppose $S \subseteq \mathbb{R}^n$ is closed and convex. Let $P : S \mapsto S$ denote a vector-valued operator from S to S. We say that P is *non-expansive* if, as a result of applying the mapping P, the distance between any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in S does not increase:

$$\|P(\boldsymbol{x}) - P(\boldsymbol{y})\| \leq \|\boldsymbol{x} - \boldsymbol{y}\|, \qquad \boldsymbol{x}, \boldsymbol{y} \in S.$$

- *For every* $\boldsymbol{x} \in \mathbb{R}^n$*, its projection* $\mathrm{Proj}_S(\boldsymbol{x})$ *is uniquely defined. The operator* $\mathrm{Proj}_S : \mathbb{R}^n \mapsto S$ *is non-expansive, and therefore in particular continuous.*
- *Proof.* Uniqueness follows since the objective function $\boldsymbol{x} \mapsto \|\boldsymbol{x} - \boldsymbol{z}\|^2$ is both weakly coercive and strictly convex on S for every $\boldsymbol{z} \in \mathbb{R}^n$ (e.g., Weierstrass + strict convexity).

Next, take $\boldsymbol{x}^1, \boldsymbol{x}^2 \in \mathbb{R}^n$. Then, by the characterization (3) of the Euclidean projection,

$$[\mathrm{Proj}_S(\boldsymbol{x}^2) - \mathrm{Proj}_S(\boldsymbol{x}^1)]^{\mathrm{T}}(\boldsymbol{x}^1 - \mathrm{Proj}_S(\boldsymbol{x}^2)) \leq 0,$$

$$[\mathrm{Proj}_S(\boldsymbol{x}^1) - \mathrm{Proj}_S(\boldsymbol{x}^2)]^{\mathrm{T}}(\boldsymbol{x}^2 - \mathrm{Proj}_S(\boldsymbol{x}^1)) \leq 0.$$

Summing the two inequalities yields

$$\begin{aligned} \|\mathrm{Proj}_S(\boldsymbol{x}^2) - \mathrm{Proj}_S(\boldsymbol{x}^1)\|^2 &\leq [\mathrm{Proj}_S(\boldsymbol{x}^2) - \mathrm{Proj}_S(\boldsymbol{x}^1)]^{\mathrm{T}}(\boldsymbol{x}^2 - \boldsymbol{x}^1) \\ &\leq \|\mathrm{Proj}_S(\boldsymbol{x}^2) - \mathrm{Proj}_S(\boldsymbol{x}^1)\| \cdot \|\boldsymbol{x}^2 - \boldsymbol{x}^1\|, \end{aligned}$$

that is, $\|\mathrm{Proj}_S(\boldsymbol{x}^2) - \mathrm{Proj}_S(\boldsymbol{x}^1)\| \leq \|\boldsymbol{x}^2 - \boldsymbol{x}^1\|$. Since this is true for every pair $(\boldsymbol{x}^1, \boldsymbol{x}^2) \in \mathbb{R}^n$, we have shown that the operator Proj_S is non-expansive on $\mathbb{R}^n$. In particular, non-expansive functions are continuous. (Why?)