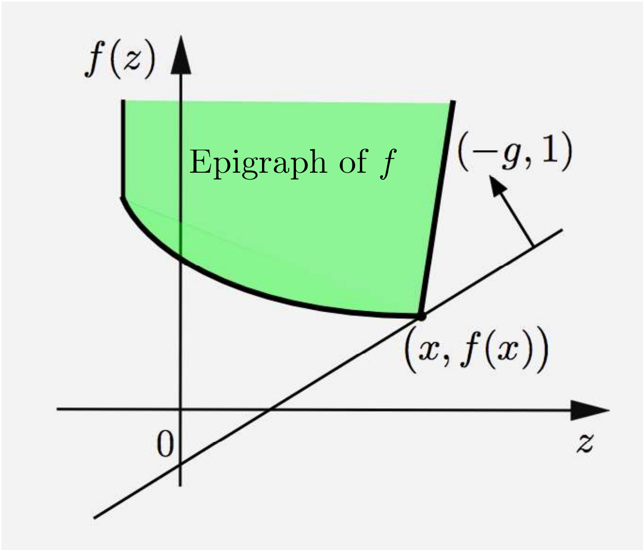
LECTURE 14

LECTURE OUTLINE

- Subgradients of convex functions
- Sugradients of real-valued convex functions
- Properties of subgradients
- Computation of subgradients
- Reading:
  - Section 5.4 of Convex Optimization Theory (focus on extended real-valued convex functions)
  - Section 2.1 of Convex Optimization Algorithms (focus on real-valued convex functions)

SUBGRADIENTS



- Let $f : \Re^n \mapsto (-\infty, \infty]$ be a convex function. A vector $g \in \Re^n$ is a subgradient of f at a point $x \in \text{dom}(f)$ if

$$f(z) \geq f(x) + (z - x)'g, \qquad \forall\, z \in \Re^n$$

- Support Hyperplane Interpretation: g is a subgradient if and only if

$$f(z) - z'g \geq f(x) - x'g, \qquad \forall\, z \in \Re^n$$

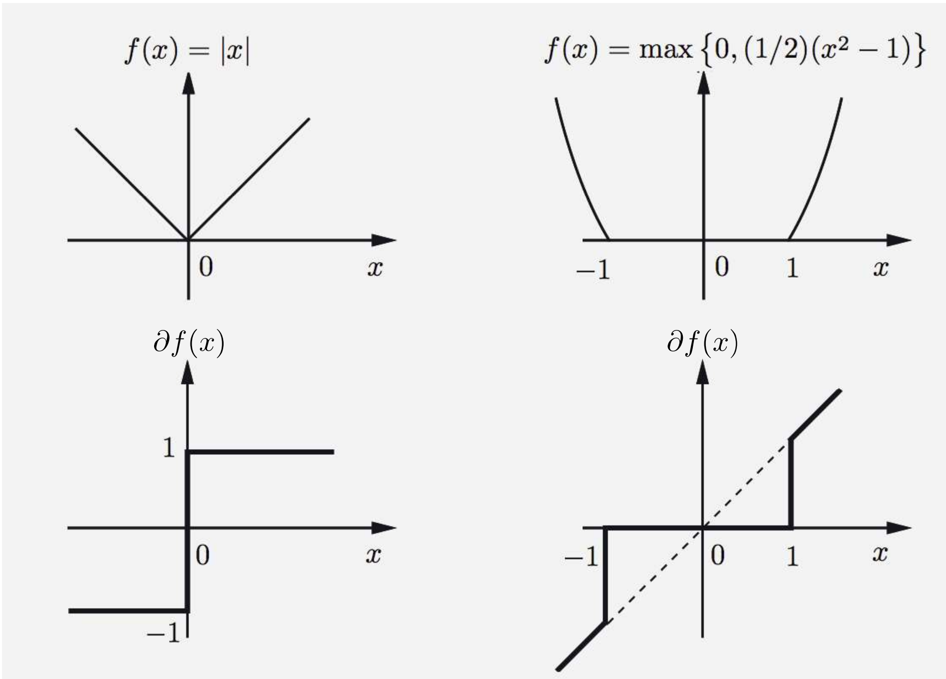
so g is a subgradient at x if and only if the hyperplane in $\Re^{n+1}$ that has normal $(-g, 1)$ and passes through $\big(x, f(x)\big)$ supports the epigraph of f.

- The set of all subgradients at x is the subdifferential of f at x, denoted $\partial f(x)$.

- x^* minimizes f if and only if $0 \in \partial f(x^*)$.

EXAMPLES OF SUBDIFFERENTIALS

- Some examples:



- If f is differentiable, then $\partial f(x) = \{\nabla f(x)\}$.

Proof: Clearly $\nabla f(x) \in \partial f(x)$. Conversely, if $g \in \partial f(x)$, then for all $\alpha \in \Re$ and $d \in \Re^n$,

$$\alpha g' d \le f(x + \alpha d) - f(x) = \alpha \nabla f(x)' d + o(|\alpha|).$$

Let $d = \nabla f(x) - g$ to obtain

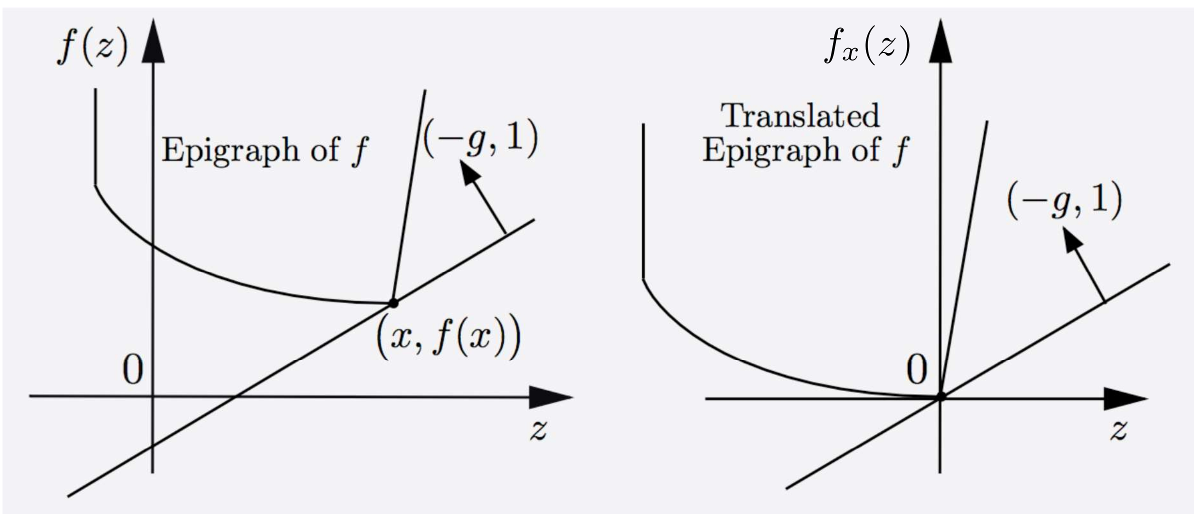
$$\|\nabla f(x) - g\|^2 \le -o(|\alpha|)/\alpha, \qquad \forall\ \alpha < 0$$

Take $\alpha \uparrow 0$ to obtain $g = \nabla f(x)$.

EXISTENCE OF SUBGRADIENTS

- Let $f : \Re^n \mapsto (-\infty, \infty]$ be proper convex.
- Consider MC/MC with

$$M = \text{epi}(f_x), \qquad f_x(z) = f(x+z) - f(x)$$



- By 2nd MC/MC Duality Theorem, $\partial f(x)$ is nonempty if $x \in \text{ri}\big(\text{dom}(f)\big)$.
- If f is real-valued, $\partial f(x)$ is nonempty for all x
- For $x \notin \text{ri}\big(\text{dom}(f)\big)$, $\partial f(x)$ may be empty.

SUBGRADIENTS OF REAL-VALUED FUNCTIONS

- Let $f : \Re^n \mapsto \Re$ be a real-valued convex function, and let $X \subset \Re^n$ be compact.

  (a) The set $\cup_{x \in X} \partial f(x)$ is bounded.

  (b) f is Lipschitz over X, i.e., for all $x, z \in X$,

$$\big|f(x) - f(z)\big| \leq L \, \|x - z\|, \quad L = \sup_{g \in \cup_{x \in X} \partial f(x)} \|g\|.$$

Proof: (a) Assume the contrary, so there exist $\{x_k\} \subset X$, and unbounded $\{g_k\}$ with

$$g_k \in \partial f(x_k), \qquad 0 < \|g_k\| < \|g_{k+1}\|, \quad k = 0, 1, \ldots.$$

Let $d_k = g_k / \|g_k\|$. Since $g_k \in \partial f(x_k)$, we have

$$f(x_k + d_k) - f(x_k) \geq g_k' d_k = \|g_k\|$$

Since $\{x_k\}$ and $\{d_k\}$ are bounded, we assume they converge to some vectors. By continuity of f, the left-hand side is bounded, contradicting the unboundedness of $\{g_k\}$.

(b) If $g \in \partial f(x)$, then for all $x, z \in X$,

$$f(x) - f(z) \leq g'(x - z) \leq \|g\| \cdot \|x - z\| \leq L \, \|x - z\|$$

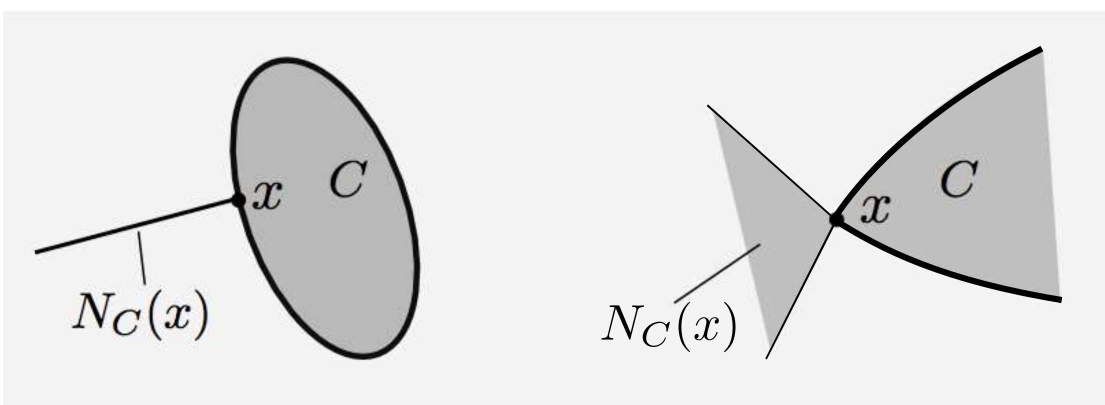
EXAMPLE: SUBDIFFERENTIAL OF INDICATOR

- Let C be a convex set, and δ_C be its indicator function.
- For $x \in C$, we have $g \in \partial\delta_C(x)$ iff

$$\delta_C(x) + g'(z - x) \leq \delta_C(z), \qquad \forall\ z \in C,$$

or equivalently $g'(z - x) \leq 0$ for all $z \in C$. Thus $\partial\delta_C(x)$ is the normal cone of C at x:

$$N_C(x) = \big\{g \mid g'(z - x) \leq 0,\ \forall\ z \in C\big\}.$$



CALCULUS OF SUBDIFFERENTIALS

- Chain Rule: Let $f : \Re^m \mapsto (-\infty, \infty]$ be convex, and A be a matrix. Consider $F(x) = f(Ax)$ and assume that F is proper. If

$$\text{Range}(A) \cap \text{ri}(\text{dom}(f)) \neq \emptyset,$$

then

$$\partial F(x) = A' \partial f(Ax), \qquad \forall\, x \in \Re^n.$$

- Subdifferential of a Sum: Let $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, be proper convex functions, and let

$$F = f_1 + \cdots + f_m.$$

Assume that $\cap_{1=1}^m \text{ri}\big(\text{dom}(f_i)\big) \neq \emptyset$. Then

$$\partial F(x) = \partial f_1(x) + \cdots + \partial f_m(x), \qquad \forall\, x \in \Re^n.$$

- Relative interior condition is needed as simple examples show.

- The relative interior conditions are automatically satisfied if the functions are real-valued.

- The relative interior conditions are unnecessary if the functions are polyhedral.

CONSTRAINED OPTIMALITY CONDITION

- Let $f : \Re^n \mapsto \Re$ and $X \subset \Re^n$ be convex. Then, a vector x^* minimizes f over X iff there exists $g \in \partial f(x^*)$ such that $-g$ belongs to the normal cone $N_X(x^*)$, i.e.,

$$g'(x - x^*) \geq 0, \qquad \forall\, x \in X.$$

Proof: x^* minimizes

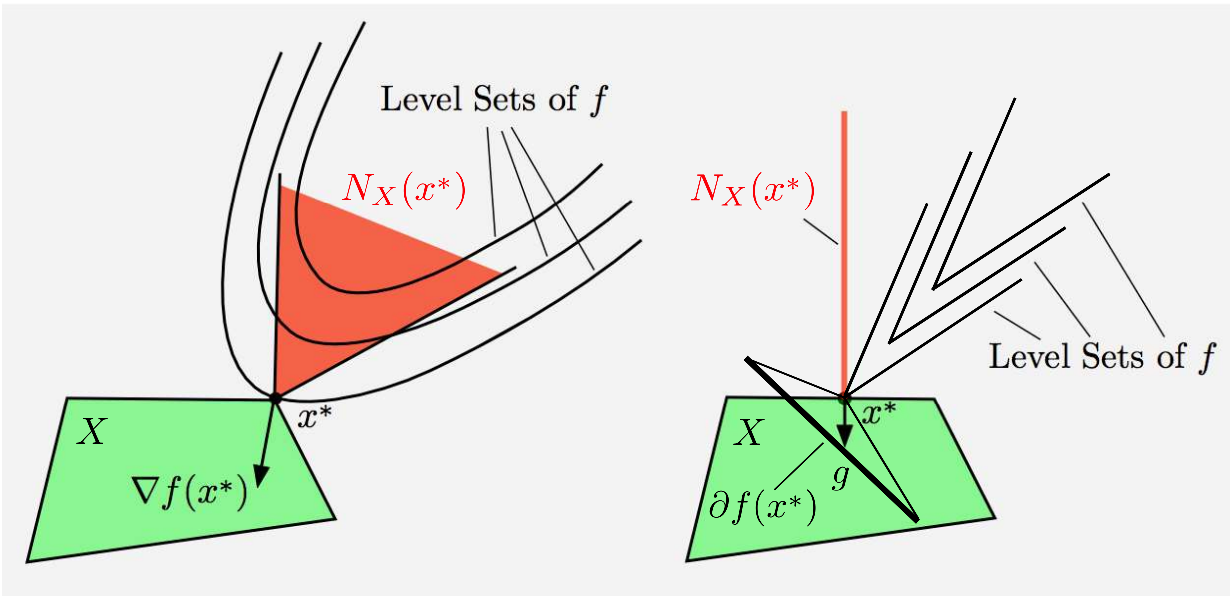
$$F(x) = f(x) + \delta_X(x)$$

if and only if $0 \in \partial F(x^*)$. Use the formula for subdifferential of sum to write

$$0 \in \partial F(x^*) = \partial f(x^*) + N_X(x^*)$$

Q.E.D.

ILLUSTRATION OF OPTIMALITY CONDITION



• In the figure on the left, f is differentiable and the optimality condition is

$$-\nabla f(x^*) \in N_X(x^*),$$

which is equivalent to

$$\nabla f(x^*)'(x - x^*) \geq 0, \qquad \forall\, x \in X.$$

• In the figure on the right, f is nondifferentiable, and the optimality condition is

$$-g \in N_X(x^*) \quad \text{for some } g \in \partial f(x^*).$$

DANSKIN'S THEOREM FOR MAX FUNCTIONS

- Let
$$f(x) = \max_{z \in Z} \phi(x, z),$$
where $x \in \Re^n$, $z \in \Re^m$, $\phi : \Re^n \times \Re^m \mapsto \Re$ is a function, Z is a compact subset of $\Re^m$, $\phi(\cdot, z)$ is convex and differentiable for each $z \in Z$, and $\nabla_x \phi(x, \cdot)$ is continuous on Z for each x. Then

$$\partial f(x) = \text{conv}\big\{\nabla_x \phi(x, z) \mid z \in Z(x)\big\}, \qquad x \in \Re^n,$$

where $Z(x)$ is the set of maximizing points

$$Z(x) = \left\{ \overline{z} \;\middle|\; \phi(x, \overline{z}) = \max_{z \in Z} \phi(x, z) \right\}$$

- Special case: $f(x) = \max\big\{\phi_1(x), \ldots, \phi_m(x)\big\}$ where ϕ_i are differentiable convex. Then

$$\partial f(x) = \text{conv}\big\{\nabla \phi_i(x) \mid i \in I(x)\big\},$$

where
$$I(x) = \big\{i \mid \phi_i(x) = f(x)\big\}$$

IMPORTANT ALGORITHMIC POINT

- Computing a single subgradient is often much easier than computing the entire subdifferential.
- Special case of dual functions: Consider

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad g(x) \leq 0, \end{aligned}$$

where $f : \Re^n \mapsto \Re$, $g : \Re^n \mapsto \Re^r$, $X \subset \Re^n$. Consider the dual problem $\max_{\mu \geq 0} q(\mu)$, where

$$q(\mu) = \inf_{x \in X} \big\{ f(x) + \mu' g(x) \big\}.$$

For a given $\mu \geq 0$, suppose that x_μ minimizes the Lagrangian over $x \in X$,

$$x_\mu \in \arg\min_{x \in X} \big\{ f(x) + \mu' g(x) \big\}.$$

Then $-g(x_\mu)$ is a subgradient of the negative of the dual function $-q$ at μ.

- Verification: For all $\nu \in \Re^r$,

$$q(\nu) = \inf_{x \in X} \big\{ f(x) + \nu' g(x) \big\} \leq f(x_\mu) + \nu' g(x_\mu)$$

$$= f(x_\mu) + \mu' g(x_\mu) + (\nu - \mu)' g(x_\mu) = q(\mu) + (\nu - \mu)' g(x_\mu)$$