LECTURE 19

LECTURE OUTLINE

- Return to descent methods
- Fixing the convergence problem of steepest descent
- ϵ-descent method
- Extended monotropic programming

IMPROVING STEEPEST DESCENT

- Consider minimization of a convex function $f : \Re^n \mapsto \Re$, over a closed convex set X.

- Return to iterative descent: Generate $\{x_k\}$ with

$$f(x_{k+1}) < f(x_k)$$

(unless x_k is optimal).

- If f is differentiable, the gradient/steepest descent method is

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

Has good convergence for α_k sufficiently small or optimally chosen.

- If f is nondifferentiable, the steepest descent method is

$$x_{k+1} = x_k - \alpha_k g_k$$

where g_k is the vector of minimum norm on $\partial f(x_k)$ ... but has convergence difficulties.

- We will discuss another method, called **ϵ-descent**:

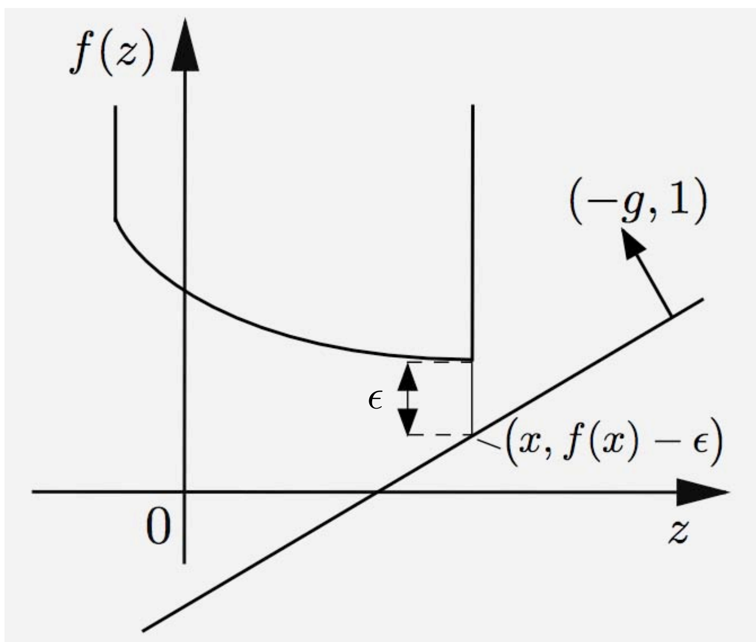
$$x_{k+1} = x_k - \alpha_k g_k$$

where g_k is the vector of minimum norm on $\partial_\epsilon f(x_k)$. **It fixes the convergence difficulties.**

REVIEW OF ϵ-SUBGRADIENTS

- For a proper convex $f : \Re^n \mapsto (-\infty, \infty]$ and $\epsilon > 0$, we say that a vector g is an ϵ-*subgradient* of f at a point $x \in \text{dom}(f)$ if

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



- The ϵ-*subdifferential* $\partial_\epsilon f(x)$ is the set of all ϵ-subgradients of f at x. By convention, $\partial_\epsilon f(x) = \emptyset$ for $x \notin \text{dom}(f)$.

- We have $\cap_{\epsilon \downarrow 0} \partial_\epsilon f(x) = \partial f(x)$ and

$$\partial_{\epsilon_1} f(x) \subset \partial_{\epsilon_2} f(x) \quad \text{if } 0 < \epsilon_1 < \epsilon_2$$

ϵ-SUBGRADIENTS AND CONJUGACY

- For any $x \in \mathrm{dom}(f)$, consider x-translation of f, i.e., the function f_x given by

$$f_x(d) = f(x+d) - f(x), \qquad \forall\, d \in \Re^n$$

and its conjugate

$$f_x^\star(g) = \sup_{d \in \Re^n} \big\{ d'g - f(x+d) + f(x) \big\} = f^\star(g) + f(x) - g'x$$

- We have

$$g \in \partial f(x) \qquad \text{iff} \qquad \sup_{d \in \Re^n} \big\{ d'g - f(x+d) + f(x) \big\} \leq 0,$$

so $\partial f(x)$ **is the 0-level set of** $f_x^\star$**:**

$$\partial f(x) = \big\{ g \mid f_x^\star(g) \leq 0 \big\}.$$

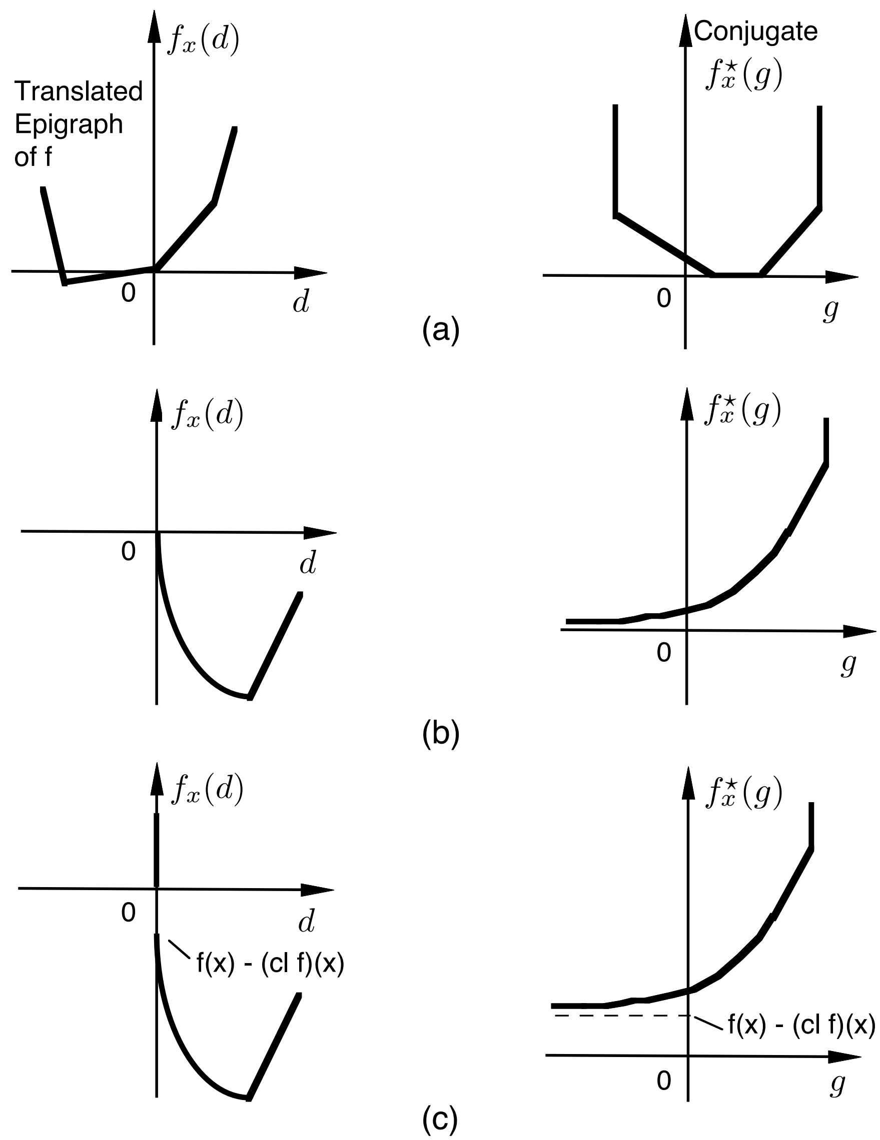
Similarly, $\partial_\epsilon f(x)$ **is the** ϵ**-level set of** $f_x^\star$**:**

$$\partial_\epsilon f(x) = \big\{ g \mid f_x^\star(g) \leq \epsilon \big\}$$

ϵ-SUBDIFFERENTIALS AS LEVEL SETS

- We have

$$\partial_\epsilon f(x) = \big\{g \mid f^\star(g)+f(x)-g'x \le \epsilon\big\} = \big\{g \mid f_x^\star(g) \le \epsilon\big\}$$



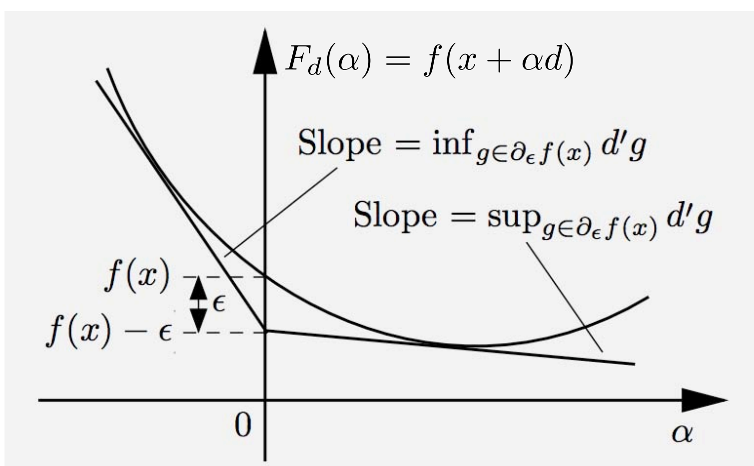
- If f is closed

$$\sup_{g\in\Re^n}\big\{-f_x^\star(g)\big\} = f_x^{\star\star}(0) = f_x(0) = 0$$

so $\partial_\epsilon f(x) \neq \emptyset$ for every $x \in \text{dom}(f)$ and $\epsilon > 0$.

PROPERTIES OF ϵ-SUBDIFFERENTIALS

- Let f: closed proper convex, $x \in \text{dom}(f)$, $\epsilon > 0$.
- Then $\partial_\epsilon f(x)$ is nonempty and closed.
- $\partial_\epsilon f(x)$ is compact iff $f_x^\star$ has no nonzero directions of recession. True if f is real-valued or $x \in \text{int}\big(\text{dom}(f)\big)$ [support fn of $\text{dom}(f_x)$ is recession fn of $f_x^\star$].
- In one dimension: $g \in \partial_\epsilon f(x)$ i ff$f(x+\alpha d) \geq f(x) - \epsilon + \alpha d'g$ for all $d \in \Re^n$ and $\alpha > 0$.
- So $g \in \partial_\epsilon f(x)$ iff the line with slope $d'g$ that passes through $f(x) - \epsilon$ lies under $f(x+\alpha d)$.



- Therefore,

$$\sup_{g\in\partial_\epsilon f(x)} d'g = \inf_{\alpha>0} \frac{f(x+\alpha d) - f(x) + \epsilon}{\alpha}$$

This formula for the support function $\sigma_{\partial_\epsilon f(x)}(d)$ can be shown also in multiple dimensions.

ϵ-DESCENT PROPERTIES

- For f: closed proper convex, by definition, $0 \in \partial_\epsilon f(x)$ i ff

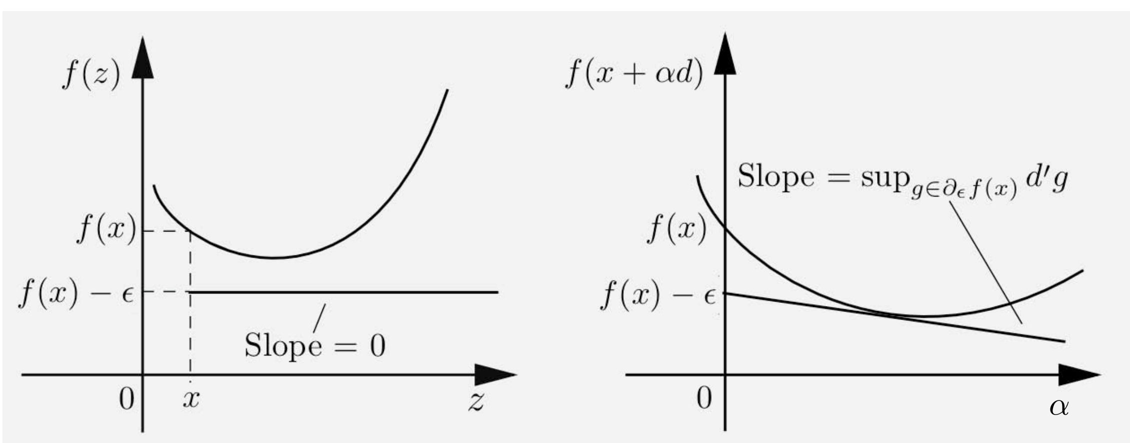
$$f(x) \leq \inf_{z \in \Re^n} f(z) + \epsilon$$

- For f: closed proper convex and $d \in \Re^n$,

$$\sup_{g \in \partial_\epsilon f(x)} d'g = \inf_{\alpha > 0} \frac{f(x + \alpha d) - f(x) + \epsilon}{\alpha}$$

so

$$\inf_{\alpha > 0} f(x + \alpha d) < f(x) - \epsilon \quad \text{iff} \quad \sup_{g \in \partial_\epsilon f(x)} d'g < 0$$



- If $0 \notin \partial_\epsilon f(x)$, we have $\sup_{g \in \partial_\epsilon f(x)} d'g < 0$ for

$$g = \arg \min_{g \in \partial_\epsilon f(x)} \|g\|,$$

(Projection Th.), so $\inf_{\alpha > 0} f(x - \alpha g) < f(x) - \epsilon$.

ϵ-DESCENT METHOD

- Method to minimize closed proper convex f:

$$x_{k+1} = x_k - \alpha_k g_k$$

where

$$-g_k = \arg \min_{g \in \partial_\epsilon f(x_k)} \|g\|,$$

and α_k is a positive stepsize.

- If $g_k = 0$, i.e., $0 \in \partial_\epsilon f(x_k)$, then x_k is an ϵ-optimal solution.

- If $g_k \neq 0$, choose α_k that reduces the cost function by at least ϵ, i.e.,

$$f(x_{k+1}) = f(x_k - \alpha_k g_k) \leq f(x_k) - \epsilon$$

- **Drawback:** Must know $\partial_\epsilon f(x_k)$.

- Motivation for a variant where $\partial_\epsilon f(x_k)$ is approximated by a set $A(x_k)$ that can be computed more easily than $\partial_\epsilon f(x_k)$.

- Then use

$$g_k = \arg \min_{g \in A(x_k)} \|g\|,$$

[project on $A(x_k)$ rather than $\partial_\epsilon f(x_k)$].

ϵ-DESCENT - OUTER APPROXIMATION

- Here $\partial_\epsilon f(x_k)$ is approximated by a set $A(x)$ such that

$$\partial_\epsilon f(x_k) \subset A(x_k) \subset \partial_{\gamma\epsilon} f(x_k),$$

where γ is a scalar with $\gamma > 1$.

- Then the method terminates with a $\gamma\epsilon$-optimal solution, and effects at least ϵ-reduction on f otherwise.

- Example of outer approximation for sum case

$$f = f_1 + \cdots + f_m$$

Take

$$A(x) = \mathrm{cl}\big(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)\big),$$

based on the fact

$$\partial_\epsilon f(x) \subset \mathrm{cl}\big(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)\big) \subset \partial_{m\epsilon} f(x)$$

- Application to separable problems where each $\partial_\epsilon f_i(x)$ is a one-dimensional interval. Then to find an ϵ-descent direction, we must solve a quadratic programming/projection problem.

EXTENDED MONOTROPIC PROGRAMMING

- Let
  - $x = (x_1, \ldots, x_m)$ with $x_i \in \Re^{n_i}$
  - $f_i : \Re^{n_i} \mapsto (-\infty, \infty]$ is closed proper convex
  - S is a subspace of $\Re^{n_1+\cdots+n_m}$
- Extended monotropic programming problem:

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(x_i) \\ &\text{subject to} \ \ x \in S \end{aligned}$$

- **Monotropic programming** is the special case where each x_i is 1-dimensional.
- Models many important optimization problems (linear, quadratic, convex network, etc).
- Has a powerful symmetric duality theory.

DUALITY

- Convert to the equivalent form

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(z_i) \\ &\text{subject to} \;\; z_i = x_i, \quad i = 1, \ldots, m, \qquad x \in S \end{aligned}$$

- Assigning a dual vector $\lambda_i \in \Re^{n_i}$ to the constraint $z_i = x_i$, the dual function is

$$\begin{aligned} q(\lambda) &= \inf_{x \in S} \lambda' x + \sum_{i=1}^{m} \inf_{z_i \in \Re^{n_i}} \big\{ f_i(z_i) - \lambda_i' z_i \big\} \\ &= \begin{cases} \sum_{i=1}^{m} q_i(\lambda_i) & \text{if } \lambda \in S^{\perp}, \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}$$

where $q_i(\lambda_i) = \inf_{z_i \in \Re} \big\{ f_i(z_i) - \lambda_i' z_i \big\} = -f_i^{\star}(\lambda_i)$.

- The dual problem is the (symmetric) extended monotropic program

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i^{\star}(\lambda_i) \\ &\text{subject to} \;\; \lambda \in S^{\perp} \end{aligned}$$

OPTIMALITY CONDITIONS

- Assume that $-\infty < q^* = f^* < \infty$. Then (x^*, λ^*) are optimal primal and dual solution pair if and only if

$$x^* \in S, \lambda^* \in S^\perp, \qquad \lambda_i^* \in \partial f_i(x_i^*), \quad \forall\ i$$

- **Specialization to the monotropic case ($n_i = 1$ for all i):** The vectors x^* and λ^* are optimal primal and dual solution pair if and only if

$$x^* \in S, \lambda^* \in S^\perp, \qquad (x_i^*, \lambda_i^*) \in \Gamma_i, \ \ \forall\ i$$

where

$$\Gamma_i = \big\{ (x_i, \lambda_i) \mid x_i \in \mathrm{dom}(f_i),\ f_i^-(x_i) \leq \lambda_i \leq f_i^+(x_i) \big\}$$

- Interesting application of these conditions to electrical networks.

STRONG DUALITY THEOREM

• Assume that the extended monotropic programming problem is feasible, and that for all feasible solutions x, the set

$$S^\perp + \partial_\epsilon D_{1,\epsilon}(x) + \cdots + D_{m,\epsilon}(x)$$

is closed for all $\epsilon > 0$, where

$$D_{i,\epsilon}(x) = \big\{(0, \ldots, 0, \lambda_i, 0, \ldots, 0) \mid \lambda_i \in \partial_\epsilon f_i(x_i)\big\}$$

Then $q^* = f^*$.

• An unusual duality condition. It is satisfied if each set $\partial_\epsilon f_i(x)$ is either compact or polyhedral. Proof is also unusual - uses the ϵ-descent method!

• **Monotropic programming case:** If $n_i = 1$, $D_{i,\epsilon}(x)$ is an interval, so it is polyhedral, and $q^* = f^*$.

• There are some other cases of interest. See the text.

• The monotropic duality result extends to convex separable problems with *nonlinear* constraints. (Hard to prove ...)