LECTURE 19

LECTURE OUTLINE

• Return to descent methods
• Fixing the convergence problem of steepest descent
• $\epsilon$-descent method
• Extended monotropic programming

All figures are courtesy of Athena Scientific, and are used with permission.
IMPROVING STEEPEST DESCENT

• Consider minimization of a convex function $f : \mathbb{R}^n \mapsto \mathbb{R}$, 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 $\alpha_k$ 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 $\varepsilon$-descent:

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

where $g_k$ is the vector of minimum norm on $\partial \varepsilon f(x_k)$. It fixes the convergence difficulties.
REVIEW OF $\epsilon$-SUBGRADIENTS

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

$$f(z) \geq f(x) + (z - x)'g - \epsilon, \quad \forall \ z \in \mathbb{R}^n$$

- The $\epsilon$-subdifferential $\partial_\epsilon f(x)$ is the set of all $\epsilon$-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

$$\partial_{\epsilon_1} f(x) \subset \partial_{\epsilon_2} f(x) \quad \text{if} \ 0 < \epsilon_1 < \epsilon_2$$
\( \varepsilon \)-SUBGRADIENTS AND CONJUGACY

- For any \( x \in \text{dom}(f) \), consider \( x \)-translation of \( f \), i.e., the function \( f_x \) given by
  \[
f_x(d) = f(x + d) - f(x), \quad \forall \ d \in \mathbb{R}^n
  \]
  and its conjugate
  \[
f_x^*(g) = \sup_{d \in \mathbb{R}^n} \{d'g - f(x + d) + f(x)\} = f^*(g) + f(x) - g'x
  \]

- We have
  \[
g \in \partial f(x) \quad \text{iff} \quad \sup_{d \in \mathbb{R}^n} \{d'g - f(x + d) + f(x)\} \leq 0,
  \]
  so \( \partial f(x) \) is the 0-level set of \( f_x^* \):
  \[
  \partial f(x) = \{g \mid f_x^*(g) \leq 0\}.
  \]

Similarly, \( \partial_\varepsilon f(x) \) is the \( \varepsilon \)-level set of \( f_x^* \):
  \[
  \partial_\varepsilon f(x) = \{g \mid f_x^*(g) \leq \varepsilon\}.
  \]
\( \varepsilon \)-SUBDIFFERENTIALS AS LEVEL SETS

- We have

\[
\partial_\varepsilon f(x) = \{ g \mid f^*(g) + f(x) - g'x \leq \varepsilon \} = \{ g \mid f^*_x(g) \leq \varepsilon \}
\]

- If \( f \) is closed

\[
\sup_{g \in \mathbb{R}^n} \{-f^*_x(g)\} = f^{**}(0) = f_x(0) = 0
\]

so \( \partial_\varepsilon f(x) \neq \emptyset \) for every \( x \in \text{dom}(f) \) and \( \varepsilon > 0 \).
PROPERTIES OF $\varepsilon$-SUBDIFFERENTIALS

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

Therefore,

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

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

- For \( f \): closed proper convex, by definition, \( 0 \in \partial f(x) \) if and only if
  \[
  f(x) \leq \inf_{z \in \mathbb{R}^n} f(z) + \epsilon
  \]

- For \( f \): closed proper convex and \( d \in \mathbb{R}^n \),
  \[
  \sup_{g \in \partial f(x)} d^t g = \inf_{\alpha > 0} f(x + \alpha d) - f(x) + \epsilon
  \]
  so
  \[
  \inf_{\alpha > 0} f(x + \alpha d) < f(x) - \epsilon \quad \text{iff} \quad \sup_{g \in \partial f(x)} d^t g < 0
  \]

- If \( 0 \notin \partial f(x) \), we have \( \sup_{g \in \partial f(x)} d^t g < 0 \) for
  \[
  g = \arg \min_{g \in \partial f(x)} \|g\|
  \]
  (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 $\alpha_k$ is a positive stepsize.

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

- If $g_k \neq 0$, choose $\alpha_k$ that reduces the cost function by at least $\epsilon$, 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 $\partial_\varepsilon f(x_k)$ is approximated by a set $A(x)$ such that

$$\partial_\varepsilon f(x_k) \subset A(x_k) \subset \partial_{\gamma_\varepsilon} f(x_k),$$

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

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

- Example of outer approximation for sum case

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

Take

$$A(x) = \text{cl}(\partial_\varepsilon f_1(x) + \cdots + \partial_\varepsilon f_m(x)),$$

based on the fact

$$\partial_\varepsilon f(x) \subset \text{cl}(\partial_\varepsilon f_1(x) + \cdots + \partial_\varepsilon f_m(x)) \subset \partial_{m_\varepsilon} f(x)$$

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

- Let
  - \( x = (x_1, \ldots, x_m) \) with \( x_i \in \mathbb{R}^{n_i} \)
  - \( f_i : \mathbb{R}^{n_i} \to (-\infty, \infty] \) is closed proper convex
  - \( S \) is a subspace of \( \mathbb{R}^{n_1+\cdots+n_m} \)

- Extended monotropic programming problem:

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

- **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

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{m} f_i(z_i) \\
\text{subject to} & \quad z_i = x_i, \quad i = 1, \ldots, m, \quad x \in S
\end{align*}
\]

• Assigning a dual vector \( \lambda_i \in \mathbb{R}^{n_i} \) to the constraint \( z_i = x_i \), the dual function is

\[
q(\lambda) = \inf_{x \in S} \lambda' x + \sum_{i=1}^{m} \inf_{z_i \in \mathbb{R}^{n_i}} \left\{ f_i(z_i) - \lambda'_i z_i \right\}
\]

\[
= \begin{cases} 
\sum_{i=1}^{m} q_i(\lambda_i) & \text{if } \lambda \in S^\perp, \\
-\infty & \text{otherwise,}
\end{cases}
\]

where \( q_i(\lambda_i) = \inf_{z_i \in \mathbb{R}} \left\{ f_i(z_i) - \lambda'_i z_i \right\} = -f_i^*(\lambda_i) \).

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

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{m} f_i^*(\lambda_i) \\
\text{subject to} & \quad \lambda \in S^\perp
\end{align*}
\]
OPTIMALITY CONDITIONS

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

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

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

$$x^* \in S, \lambda^* \in S^\perp, \quad (x^*_i, \lambda^*_i) \in \Gamma_i, \quad \forall \ i$$

where

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

• 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

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

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 \( \epsilon \)-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 \( \text{nonlinear} \) constraints. (Hard to prove ...)