LECTURE 23

LECTURE OUTLINE

- Overview of Dual Methods
- Nondifferentiable Optimization

****

- Consider the primal problem

minimize $f(x)$
subject to $x \in X, \qquad g_j(x) \leq 0, \qquad j = 1, \ldots, r,$

assuming $-\infty < f^* < \infty$.

- Dual problem: Maximize

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

subject to $\mu \geq 0$.

PROS AND CONS FOR SOLVING THE DUAL

- The dual is concave.

- The dual may have smaller dimension and/or simpler constraints.

- If there is no duality gap and the dual is solved exactly for a geometric multiplier μ^*, all optimal primal solutions can be obtained by minimizing the Lagrangian $L(x, \mu^*)$ over $x \in X$.

- Even if there is a duality gap, $q(\mu)$ is a lower bound to the optimal primal value for every $\mu \geq 0$.

- Evaluating $q(\mu)$ requires minimization of $L(x, \mu)$ over $x \in X$.

- The dual function is often nondifferentiable.

- Even if we find an optimal dual solution μ^*, it may be difficult to obtain a primal optimal solution.

STRUCTURE

- Separability: Classical duality structure (Lagrangian relaxation).
- Partitioning: The problem

$$\begin{aligned}&\text{minimize } \ F(x) + G(y)\\ &\text{subject to } \ Ax + By = c, \quad x \in X, \ \ y \in Y\end{aligned}$$

can be written as

$$\begin{aligned}&\text{minimize } \ F(x) + \inf_{By = c - Ax,\, y \in Y} G(y)\\ &\text{subject to } \ x \in X.\end{aligned}$$

With no duality gap, this problem is written as

$$\begin{aligned}&\text{minimize } \ F(x) + Q(Ax)\\ &\text{subject to } \ x \in X,\end{aligned}$$

where

$$Q(Ax) = \max_{\lambda} q(\lambda, Ax)$$

$$q(\lambda, Ax) = \inf_{y \in Y} \big\{ G(y) + \lambda'(Ax + By - c) \big\}$$

DUAL DERIVATIVES

- Let

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

Then for all $\overline{\mu} \in \Re^r$,

$$\begin{aligned} q(\overline{\mu}) &= \inf_{x\in X}\big\{f(x)+\overline{\mu}' g(x)\big\} \\ &\leq f(x_\mu)+\overline{\mu}' g(x_\mu) \\ &= f(x_\mu)+\mu' g(x_\mu)+(\overline{\mu}-\mu)' g(x_\mu) \\ &= q(\mu)+(\overline{\mu}-\mu)' g(x_\mu). \end{aligned}$$

- Thus $g(x_\mu)$ *is a subgradient of* q *at* μ.

- **Proposition:** Let X be compact, and let f and g be continuous over X. Assume also that for every μ, $L(x,\mu)$ is minimized over $x \in X$ at a unique point x_μ. Then, q is everywhere continuously differentiable and

$$\nabla q(\mu) = g(x_\mu), \qquad \forall\, \mu \in \Re^r.$$

NONDIFFERENTIABLE DUAL

- If there exists a duality gap, the dual function is nondifferentiable at every dual optimal solution.
- Important nondifferentiable case: When q is polyhedral, that is,

$$q(\mu) = \min_{i \in I} \{ a_i' \mu + b_i \},$$

where I is a finite index set, and $a_i \in \Re^r$ and b_i are given (arises when X is a discrete set, as in integer programming).

- **Proposition:** Let q be polyhedral as above, and let I_μ be the set of indices attaining the minimum

$$I_\mu = \{ i \in I \mid a_i' \mu + b_i = q(\mu) \}.$$

The set of all subgradients of q at μ is

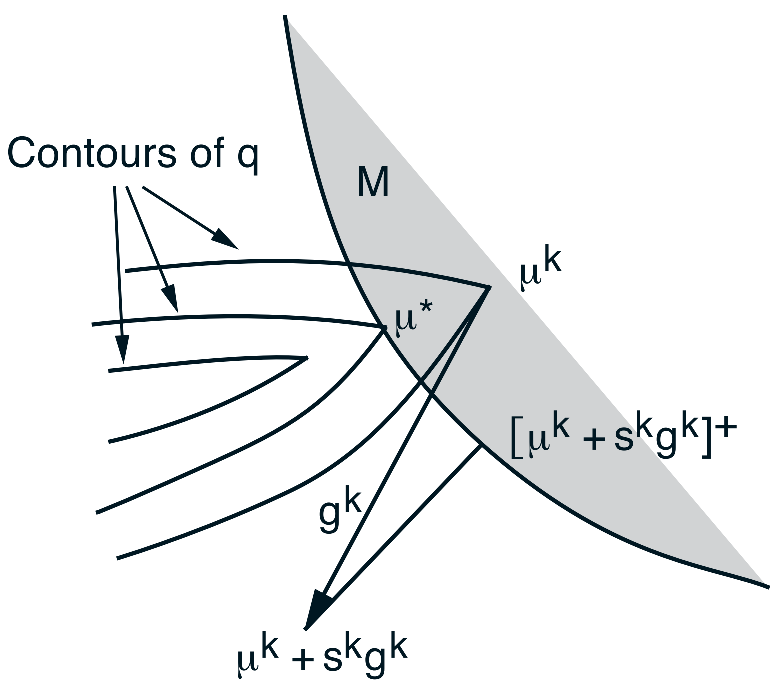
$$\partial q(\mu) = \left\{ g \;\middle|\; g = \sum_{i \in I_\mu} \xi_i a_i,\ \xi_i \geq 0,\ \sum_{i \in I_\mu} \xi_i = 1 \right\}.$$

NONDIFFERENTIABLE OPTIMIZATION

- Consider maximization of $q(\mu)$ over $M = \{\mu \geq 0 \mid q(\mu) > -\infty\}$
- Subgradient method:

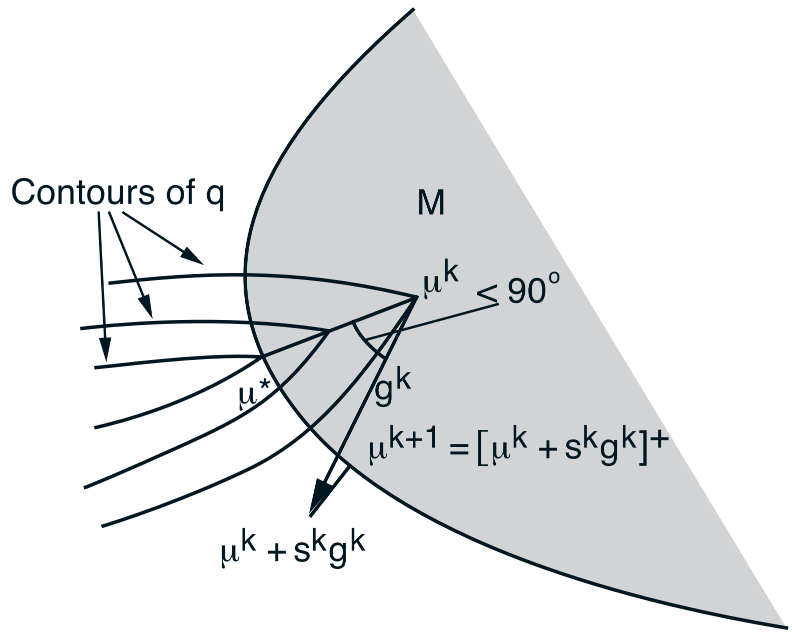
$$\mu^{k+1} = \left[\mu^k + s^k g^k\right]^+,$$

where g^k is the subgradient $g(x_{\mu^k})$, $[\cdot]^+$ denotes projection on the closed convex set M, and s^k is a positive scalar stepsize.



KEY SUBGRADIENT METHOD PROPERTY

- For a small stepsize it reduces the Euclidean distance to the optimum.



- **Proposition:** For any dual optimal solution μ^*, we have

$$\|\mu^{k+1} - \mu^*\| < \|\mu^k - \mu^*\|,$$

for all stepsizes s^k such that

$$0 < s^k < \frac{2\big(q(\mu^*) - q(\mu^k)\big)}{\|g^k\|^2}.$$

STEPSIZE RULES

- Constant stepsize: $s^k \equiv s$ for some $s > 0$.
- If $\|g^k\| \leq C$ for some constant C and all k,

$$\|\mu^{k+1}-\mu^*\|^2 \leq \|\mu^k-\mu^*\|^2-2s\big(q(\mu^*)-q(\mu^k)\big)+s^2C^2,$$

so the distance to the optimum decreases if

$$0 < s < \frac{2\big(q(\mu^*)-q(\mu^k)\big)}{C^2}$$

or equivalently, if μ^k belongs to the level set

$$\left\{\mu \;\middle|\; q(\mu) < q(\mu^*) - \frac{sC^2}{2}\right\}.$$

- With a little further analysis, it can be shown that the method, at least asymptotically, reaches this level set, i.e.

$$\limsup_{k\to\infty} q(\mu^k) \geq q(\mu^*) - \frac{sC^2}{2}.$$

OTHER STEPSIZE RULES

- Diminishing stepsize: $s^k \to 0$ with some restrictions.
- Dynamic stepsize rule:

$$s^k = \frac{\alpha^k \big(q^k - q(\mu^k)\big)}{\|g^k\|^2},$$

where $q^k \approx q^*$ and $0 < \alpha^k < 2$.

- Some possibilities:
  - q^k is the best known upper bound to q^*: start with $\alpha^0 = 1$ and decrease α^k by a certain factor every few iterations.
  - $\alpha^k = 1$ for all k and

$$q^k = \big(1 + \beta(k)\big)\hat{q}^k,$$

where $\hat{q}^k = \max_{0 \le i \le k} q(\mu^i)$, and $\beta(k) > 0$ is adjusted depending on algorithmic progress of the algorithm.