LECTURE 21

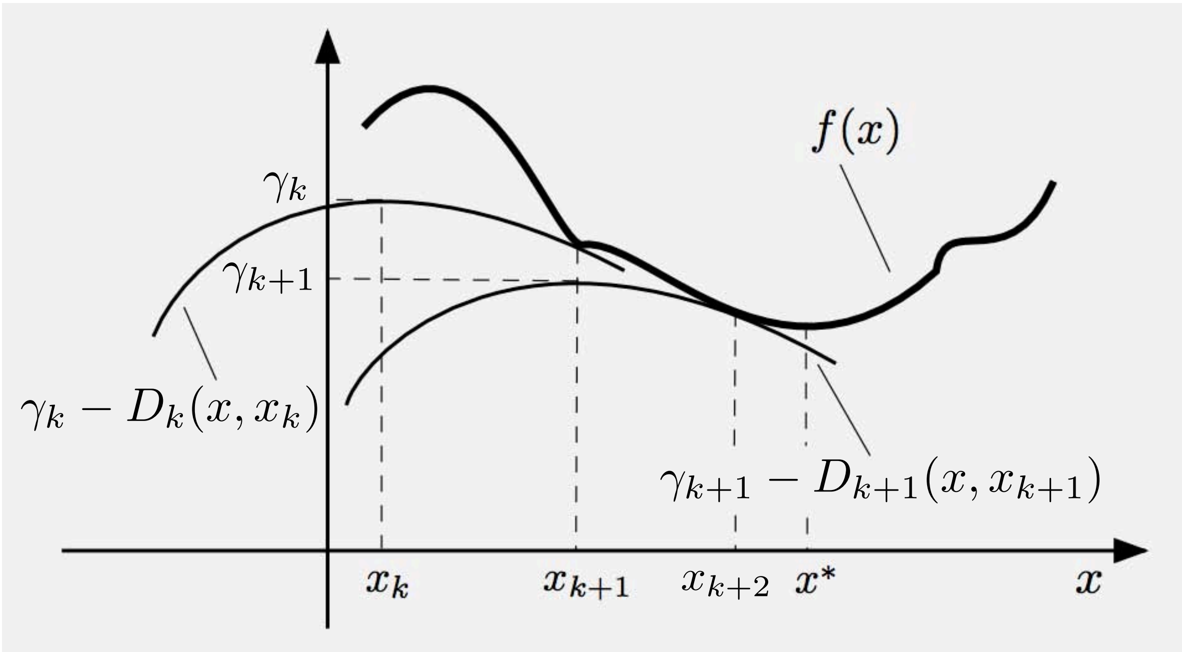
LECTURE OUTLINE

- Generalized forms of the proximal point algorithm
- Interior point methods
- Constrained optimization case - Barrier method
- Conic programming cases

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

GENERALIZED PROXIMAL ALGORITHM

- Replace quadratic regularization by more general proximal term.
- Minimize possibly nonconvex $f :\mapsto (-\infty, \infty]$.



- Introduce a general regularization term $D_k : \Re^{2n} \mapsto (-\infty, \infty]$:

$$x_{k+1} \in \arg \min_{x \in \Re^n} \left\{ f(x) + D_k(x, x_k) \right\}$$

- Assume attainment of min (but this is not automatically guaranteed)
- Complex/unreliable behavior when f is nonconvex

SOME GUARANTEES ON GOOD BEHAVIOR

- Assume

$$D_k(x, x_k) \geq D_k(x_k, x_k), \qquad \forall\, x \in \Re^n,\, k \qquad (1)$$

Then we have a cost improvement property:

$$\begin{aligned} f(x_{k+1}) &\leq f(x_{k+1}) + D_k(x_{k+1}, x_k) - D_k(x_k, x_k) \\ &\leq f(x_k) + D_k(x_k, x_k) - D_k(x_k, x_k) \\ &= f(x_k) \end{aligned}$$

- Assume algorithm stops only when x_k in optimal solution set X^*, i.e.,

$$x_k \in \arg\min_{x \in \Re^n} \big\{ f(x) + D_k(x, x_k) \big\} \quad \Rightarrow \quad x_k \in X^*$$

- Then strict cost improvement for $x_k \notin X^*$
- Guaranteed if f is convex and
  - (a) $D_k(\cdot, x_k)$ satisfies (1), and is convex and differentiable at x_k
  - (b) We have

$$\mathrm{ri}\big(\mathrm{dom}(f)\big) \cap \mathrm{ri}\big(\mathrm{dom}(D_k(\cdot, x_k))\big) \neq \emptyset$$

EXAMPLE METHODS

- Bregman distance function

$$D_k(x, y) = \frac{1}{c_k}\big(\phi(x) - \phi(y) - \nabla\phi(y)'(x - y)\big),$$

where $\phi : \Re^n \mapsto (-\infty, \infty]$ is a convex function, differentiable within an open set containing $\mathrm{dom}(f)$, and c_k is a positive penalty parameter.

- Majorization-Minimization algorithm:

$$D_k(x, y) = M_k(x, y) - M_k(y, y),$$

where M satisfies

$$M_k(y, y) = f(y), \qquad \forall\ y \in \Re^n,\ k = 0, 1,$$

$$M_k(x, x_k) \geq f(x_k), \qquad \forall\ x \in \Re^n,\ k = 0, 1, \ldots$$

- Example for case $f(x) = R(x) + \|Ax - b\|^2$, where R is a convex regularization function

$$M(x, y) = R(x) + \|Ax - b\|^2 - \|Ax - Ay\|^2 + \|x - y\|^2$$

- Expectation-Maximization (EM) algorithm (special context in inference, f nonconvex)

INTERIOR POINT METHODS

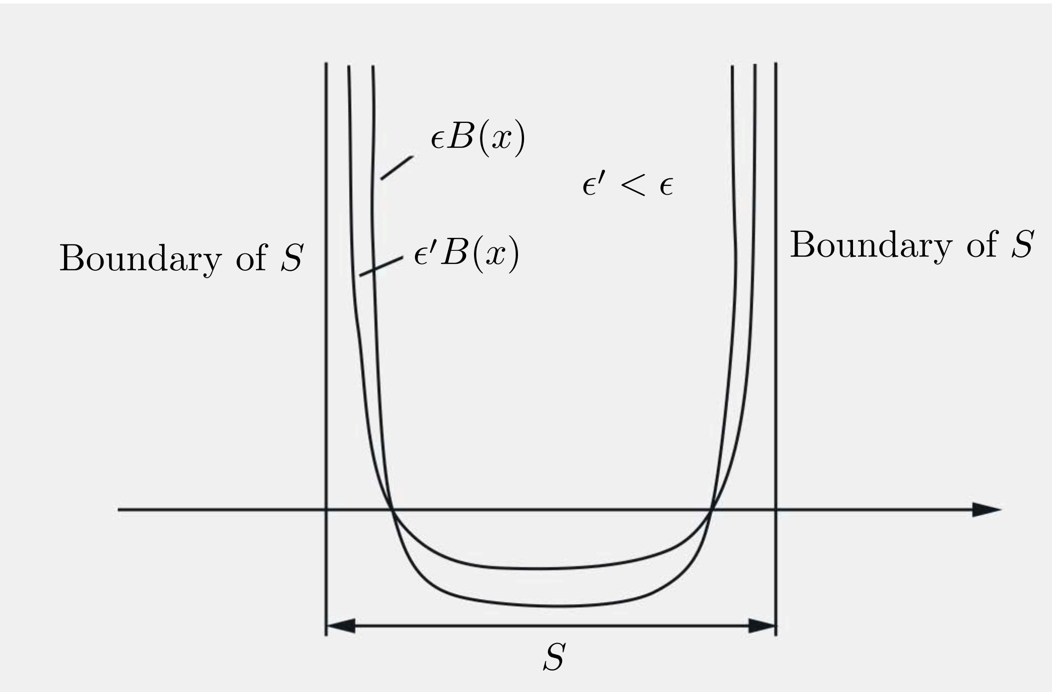
- Consider $\min f(x)$ s. t. $g_j(x) \leq 0,\ j = 1, \ldots, r$
- A **barrier function**, that is continuous and goes to ∞ as any one of the constraints $g_j(x)$ approaches 0 from negative values; e.g.,

$$B(x) = -\sum_{j=1}^{r} \ln\bigl\{-g_j(x)\bigr\}, \qquad B(x) = -\sum_{j=1}^{r} \frac{1}{g_j(x)}.$$

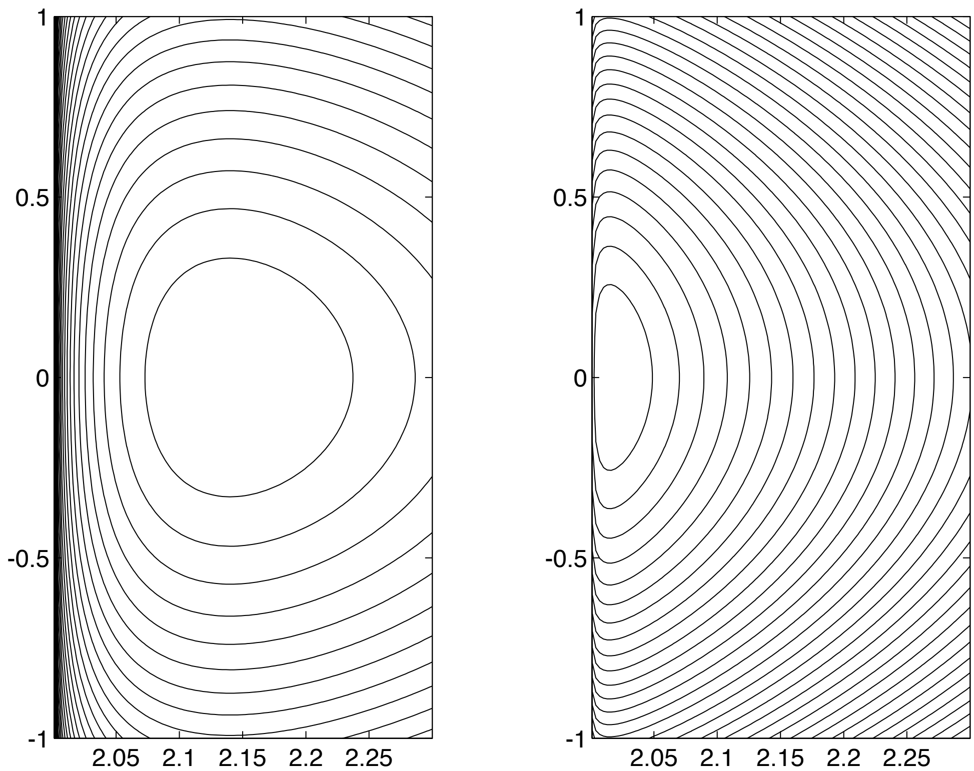
- **Barrier method:** Let

$$x_k = \arg\min_{x \in S} \bigl\{ f(x) + \epsilon_k B(x) \bigr\}, \qquad k = 0, 1, \ldots,$$

where $S = \{x \mid g_j(x) < 0,\ j = 1, \ldots, r\}$ and the parameter sequence $\{\epsilon_k\}$ satisfies $0 < \epsilon_{k+1} < \epsilon_k$ for all k and $\epsilon_k \to 0$.



BARRIER METHOD - EXAMPLE



$$\text{minimize} \quad f(x) = \tfrac{1}{2}\big((x^1)^2 + (x^2)^2\big)$$
$$\text{subject to} \quad 2 \leq x^1,$$

with optimal solution $x^* = (2, 0)$.

- Logarithmic barrier: $B(x) = -\ln(x^1 - 2)$
- We have $x_k = \big(1 + \sqrt{1 + \epsilon_k}\,, 0\big)$ from

$$x_k \in \arg\min_{x^1 > 2} \Big\{ \tfrac{1}{2}\big((x^1)^2 + (x^2)^2\big) - \epsilon_k \ln(x^1 - 2) \Big\}$$

- As ϵ_k is decreased, the unconstrained minimum x_k approaches the constrained minimum $x^* = (2, 0)$.

- As $\epsilon_k \to 0$, computing x_k becomes more difficult because of ill-conditioning (a Newton-like method is essential for solving the approximate problems).

CONVERGENCE

- Every limit point of a sequence $\{x_k\}$ generated by a barrier method is a minimum of the original constrained problem.

Proof: Let $\{\overline{x}\}$ be the limit of a subsequence $\{x_k\}_{k\in K}$. Since $x_k \in S$ and X is closed, $\overline{x}$ is feasible for the original problem.

If $\overline{x}$ is not a minimum, there exists a feasible x^* such that $f(x^*) < f(\overline{x})$ and therefore also an interior point $\tilde{x} \in S$ such that $f(\tilde{x}) < f(\overline{x})$. By the definition of x_k,

$$f(x_k) + \epsilon_k B(x_k) \le f(\tilde{x}) + \epsilon_k B(\tilde{x}), \qquad \forall\ k,$$

so by taking limit

$$f(\overline{x}) + \liminf_{k\to\infty,\, k\in K} \epsilon_k B(x_k) \le f(\tilde{x}) < f(\overline{x})$$

Hence $\liminf_{k\to\infty,\, k\in K} \epsilon_k B(x_k) < 0$.

If $\overline{x} \in S$, we have $\lim_{k\to\infty,\, k\in K} \epsilon_k B(x_k) = 0$, while if $\overline{x}$ lies on the boundary of S, we have by assumption $\lim_{k\to\infty,\, k\in K} B(x_k) = \infty$. Thus

$$\liminf_{k\to\infty} \epsilon_k B(x_k) \ge 0,$$

– a contradiction.

SECOND ORDER CONE PROGRAMMING

- Consider the SOCP

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \;\; A_i x - b_i \in C_i, \;\; i = 1, \ldots, m, \end{aligned}$$

where $x \in \Re^n$, c is a vector in $\Re^n$, and for $i = 1, \ldots, m$, A_i is an $n_i \times n$ matrix, b_i is a vector in $\Re^{n_i}$, and C_i is the second order cone of $\Re^{n_i}$.

- We approximate this problem with

$$\text{minimize} \quad c'x + \epsilon_k \sum_{i=1}^{m} B_i(A_i x - b_i)$$

$$\text{subject to} \;\; x \in \Re^n, \quad A_i x - b_i \in \text{int}(C_i), \;\; i = 1, \ldots, m,$$

where B_i is the logarithmic barrier function:

$$B_i(y) = -\ln\Big(y_{n_i}^2 - (y_1^2 + \cdots + y_{n_i - 1}^2)\Big), \quad y \in \text{int}(C_i),$$

and $\{\epsilon_k\}$ is a positive sequence with $\epsilon_k \to 0$.

- Essential to use Newton's method to solve the approximating problems.

- Interesting complexity analysis

SEMIDEFINITE PROGRAMMING

- Consider the dual SDP

$$\begin{aligned} &\text{maximize} && b'\lambda \\ &\text{subject to} && D - (\lambda_1 A_1 + \cdots + \lambda_m A_m) \in C, \end{aligned}$$

where $b \in \Re^m$, $D, A_1, \ldots, A_m$ are symmetric matrices, and C is the cone of positive semidefinite matrices.

- The logarithmic barrier method uses approximating problems of the form

$$\text{maximize} \quad b'\lambda + \epsilon_k \ln\Big(\det(D - \lambda_1 A_1 - \cdots - \lambda_m A_m)\Big)$$

over all $\lambda \in \Re^m$ such that $D - (\lambda_1 A_1 + \cdots + \lambda_m A_m)$ is positive definite.

- Here $\epsilon_k > 0$ and $\epsilon_k \to 0$.

- Furthermore, we should use a starting point such that $D - \lambda_1 A_1 - \cdots - \lambda_m A_m$ is positive definite, and Newton's method should ensure that the iterates keep $D - \lambda_1 A_1 - \cdots - \lambda_m A_m$ within the positive definite cone.

MIT OpenCourseWare
http://ocw.mit.edu

6.253 Convex Analysis and Optimization
Spring 2012

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.