6.252 NONLINEAR PROGRAMMING

LECTURE 10

ALTERNATIVES TO GRADIENT PROJECTION

LECTURE OUTLINE

- Three Alternatives/Remedies for Gradient Projection
  - Two-Metric Projection Methods
  - Manifold Suboptimization Methods
  - Affine Scaling Methods

Scaled GP method with scaling matrix $H^k > 0$:

$$x^{k+1} = x^k + \alpha^k(\overline{x}^k - x^k),$$

$$\overline{x}^k = \arg\min_{x \in X} \left\{ \nabla f(x^k)'(x - x^k) + \frac{1}{2s^k}(x - x^k)' H^k (x - x^k) \right\}.$$

- The QP direction subproblem is complicated by:
  - Difficult inequality (e.g., nonorthant) constraints
  - Nondiagonal H^k, needed for Newton scaling

THREE WAYS TO DEAL W/ THE DIFFICULTY

- Two-metric projection methods:

$$x^{k+1} = \left[x^k - \alpha^k D^k \nabla f(x^k)\right]^+$$

  - Use Newton-like scaling but use a standard projection
  - Suitable for bounds, simplexes, Cartesian products of simple sets, etc
- Manifold suboptimization methods:
  - Use (scaled) gradient projection on the manifold of active inequality constraints
  - Each QP subproblem is equality-constrained
  - Need strategies to cope with changing active manifold (add-drop constraints)
- Affine Scaling Methods
  - Go through the interior of the feasible set
  - Each QP subproblem is equality-constrained, AND we don't have to deal with changing active manifold

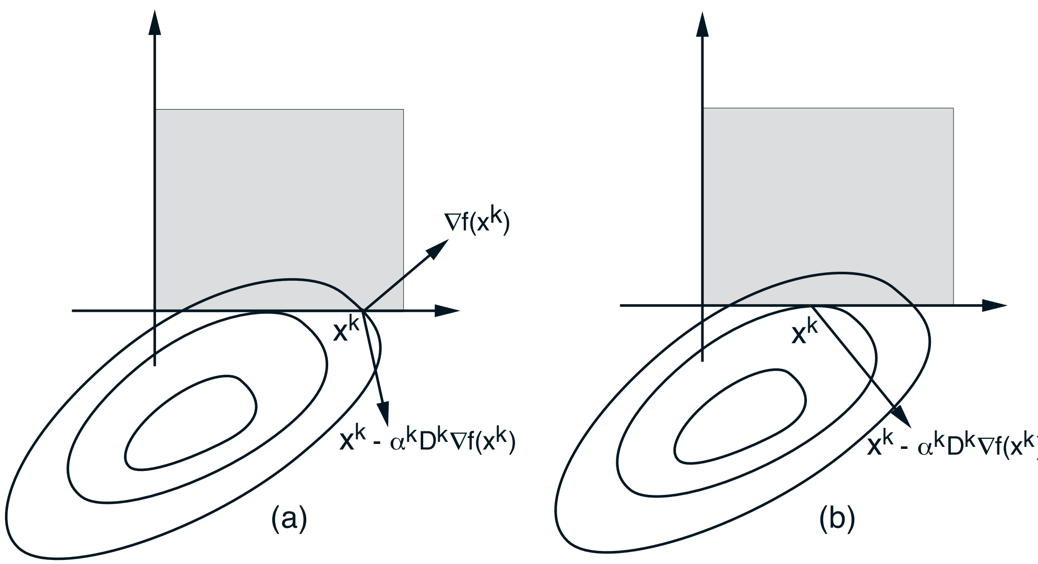
TWO-METRIC PROJECTION METHODS

- In their simplest form, apply to constraint: $x \geq 0$, but generalize to bound and other constraints

- Like unconstr. gradient methods except for $[\cdot]^+$

$$x^{k+1} = \left[x^k - \alpha^k D^k \nabla f(x^k)\right]^+, \qquad D^k > 0$$

- Major difficulty: Descent is not guaranteed for D^k: arbitrary



- Remedy: Use D^k that is diagonal w/ respect to indices that “are active and want to stay active”

$$I^+(x^k) = \left\{ i \;\middle|\; x_i^k = 0,\ \partial f(x^k)/\partial x_i > 0 \right\}$$

PROPERTIES OF 2-METRIC PROJECTION

- Suppose D^k is diagonal with respect to $I^+(x^k)$, i.e., $d_{ij}^k = 0$ for $i, j \in I^+(x^k)$ with $i \neq j$, and let

$$x^k(a) = \left[x^k - \alpha D^k \nabla f(x^k)\right]^+$$

  - If x^k is stationary, $x^k = x^k(\alpha)$ for all $\alpha > 0$.
  - Otherwise $f\big(x(\alpha)\big) < f(x^k)$ for all sufficiently small $\alpha > 0$ (can use Armijo rule).

- Because $I^+(x)$ is discontinuous w/ respect to x, to guarantee convergence we need to include in $I^+(x)$ constraints that are "ϵ-active" [those w/ $x_i^k \in [0, \epsilon]$ and $\partial f(x^k)/\partial x_i > 0$].

- The constraints in $I^+(x^*)$ eventually become active and don't matter.

- Method reduces to unconstrained Newton-like method on the manifold of active constraints at x^*.

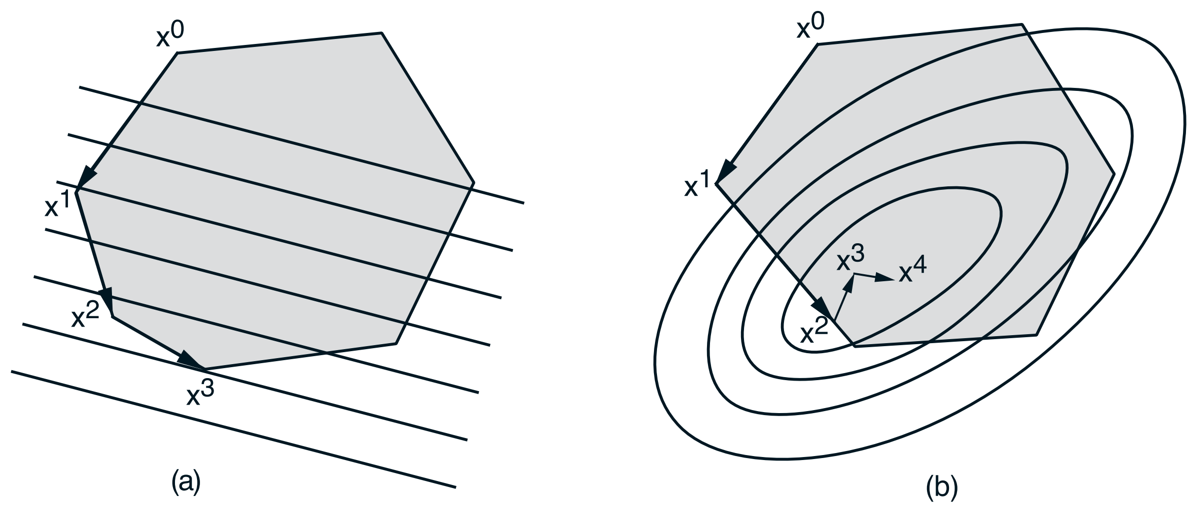
- Thus, superlinear convergence is possible w/ simple projections.

MANIFOLD SUBOPTIMIZATION METHODS

- Feasible direction methods for

$$\min f(x) \qquad \text{subject to } a_j' x \le b_j, \quad j = 1, \ldots, r$$

- Gradient is projected on a linear manifold of active constraints rather than on the entire constraint set (linearly constrained QP).



- Searches through sequence of manifolds, each differing by at most one constraint from the next.

- Potentially many iterations to identify the active manifold; then method reduces to (scaled) steepest descent on the active manifold.

- Well-suited for a small number of constraints, and for quadratic programming.

OPERATION OF MANIFOLD METHODS

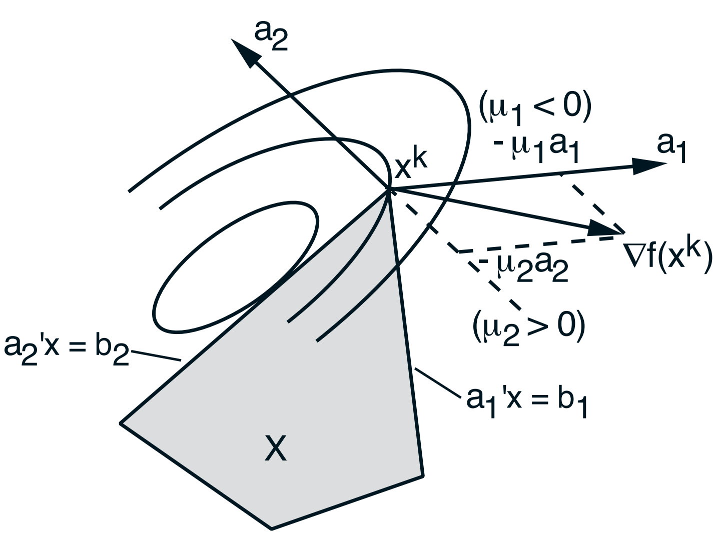
- Let $A(x) = \{j \mid a_j' x = b_j\}$ be the active index set at x. Given x^k, we find

$$d^k = \arg \min_{a_j' d = 0,\, j \in A(x^k)} \nabla f(x^k)' d + \tfrac{1}{2} d' H^k d$$

- If $d^k \neq 0$, then d^k is a feasible descent direction. Perform feasible descent on the current manifold.

- If $d^k = 0$, either (1) x^k is stationary or (2) we enlarge the current manifold (drop an active constraint). For this, use the scalars μ_j such that

$$\nabla f(x^k) + \sum_{j \in A(x^k)} \mu_j a_j = 0$$



If $\mu_j \geq 0$ for all j, x^k is stationary, since for all feasible x, $\nabla f(x^k)'(x - x^k)$ is equal to

$$-\sum_{j \in A(x^k)} \mu_j a_j'(x - x^k) \geq 0$$

Else, drop a constraint j with $\mu_j < 0$.

AFFINE SCALING METHODS FOR LP

- Focus on the LP $\min_{Ax=b,\, x\ge 0} c'x$, and the scaled gradient projection $x^{k+1} = x^k + \alpha^k(\overline{x}^k - x^k)$, with

$$\overline{x}^k = \arg \min_{Ax=b,\, x\ge 0} c'(x - x^k) + \frac{1}{2s^k}(x - x^k)'H^k(x - x^k)$$

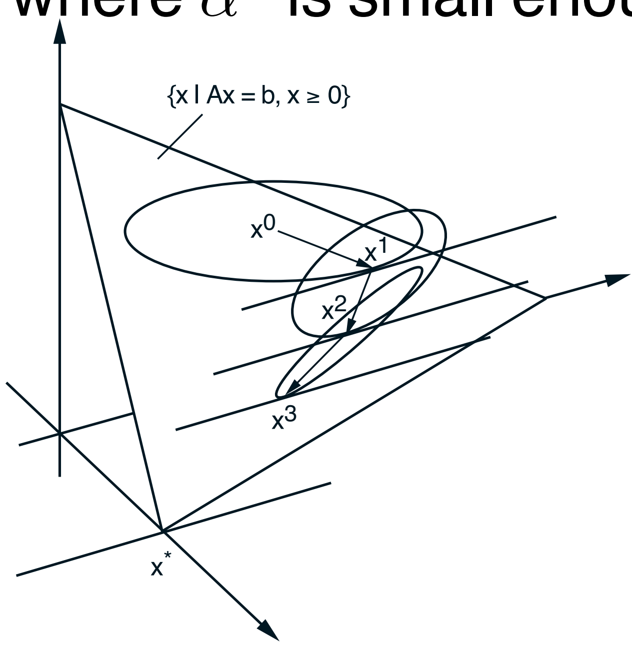
- If $x^k > 0$ then $\overline{x}^k > 0$ for s^k small enough, so $\overline{x}^k = x^k - s^k(H^k)^{-1}(c - A'\lambda^k)$ with

$$\lambda^k = \big(A(H^k)^{-1}A'\big)^{-1}A(H^k)^{-1}c$$

Lumping s^k into α^k:

$$x^{k+1} = x^k - \alpha^k(H^k)^{-1}(c - A'\lambda^k),$$

where α^k is small enough to ensure that $x^{k+1} > 0$



Importance of using time-varying H^k (should bend $\overline{x}^k - x^k$ away from the boundary)

AFFINE SCALING

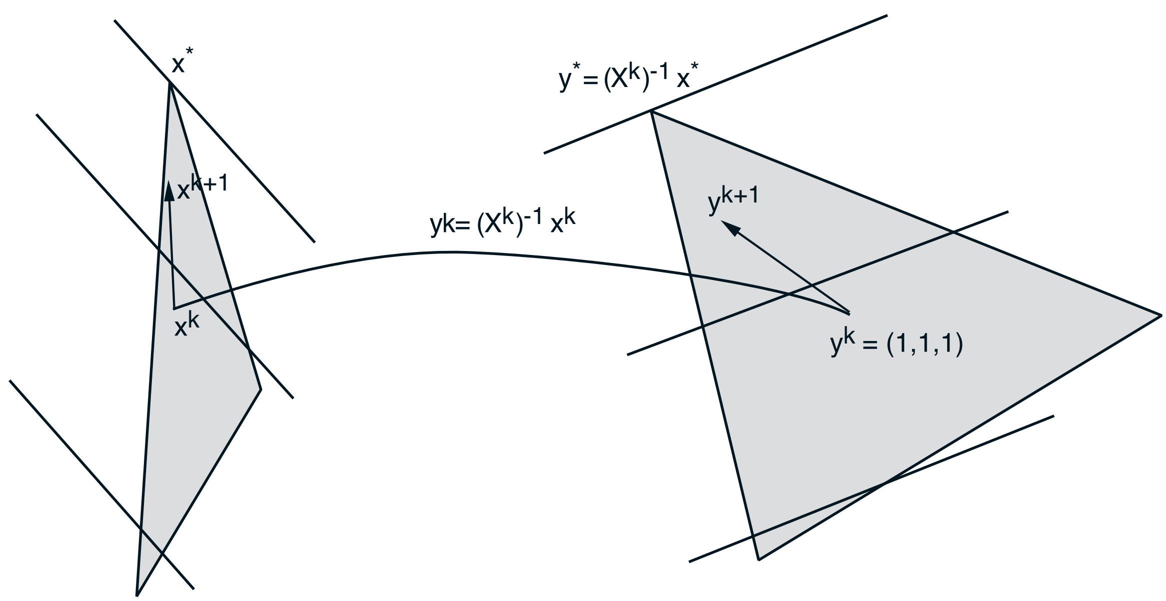
- Particularly interesting choice (affine scaling)

$$H^k = (X^k)^{-2},$$

where X^k is the diagonal matrix having the (positive) coordinates x_i^k along the diagonal:

$$x^{k+1} = x^k - \alpha^k (X^k)^2 (c - A'\lambda^k), \ \ \lambda^k = \left(A(X^k)^2 A'\right)^{-1} A(X^k)^2 c$$

- Corresponds to unscaled gradient projection iteration in the variables $y = (X^k)^{-1} x$. The vector x^k is mapped onto the unit vector $y^k = (1, \ldots, 1)$.



- Extensions, convergence, practical issues.