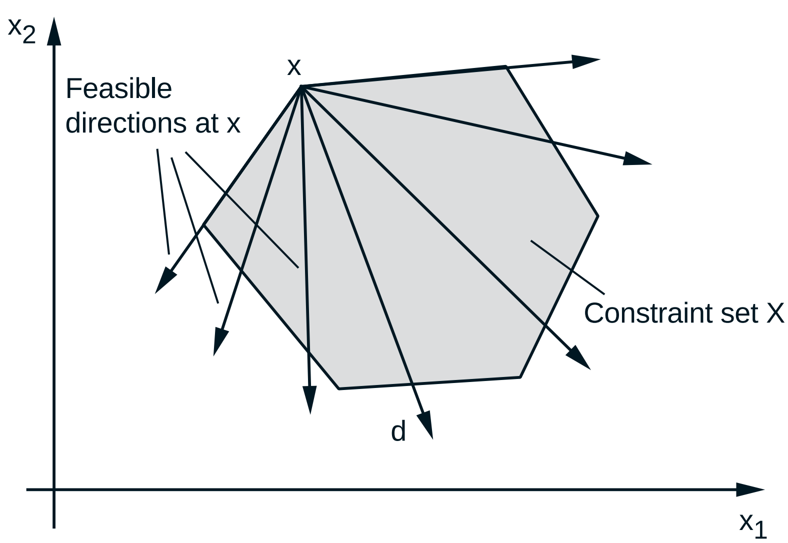
6.252 NONLINEAR PROGRAMMING

LECTURE 9: FEASIBLE DIRECTION METHODS

LECTURE OUTLINE

- Conditional Gradient Method
- Gradient Projection Methods

A *feasible direction* at an $x \in X$ is a vector $d \neq 0$ such that $x + \alpha d$ is feasible for all suff. small $\alpha > 0$



- Note: the set of feasible directions at x is the set of all $\alpha(z - x)$ where $z \in X$, $z \neq x$, and $\alpha > 0$

FEASIBLE DIRECTION METHODS

- A *feasible direction method*:

$$x^{k+1} = x^k + \alpha^k d^k,$$

where d^k: feasible *descent* direction [$\nabla f(x^k)' d^k < 0$], and $\alpha^k > 0$ and such that $x^{k+1} \in X$.

- Alternative definition:

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

where $\alpha^k \in (0, 1]$ and if x^k is nonstationary,

$$\overline{x}^k \in X, \qquad \nabla f(x^k)'(\overline{x}^k - x^k) < 0.$$

- Stepsize rules: Limited minimization, Constant $\alpha^k = 1$, Armijo: $\alpha^k = \beta^{m_k} s$, where m_k is the first nonnegative m for which

$$f(x^k) - f\big(x^k + \beta^m (\overline{x}^k - x^k)\big) \geq -\sigma \beta^m \nabla f(x^k)'(\overline{x}^k - x^k)$$

CONVERGENCE ANALYSIS

- Similar to the one for (unconstrained) gradient methods.

- The direction sequence $\{d^k\}$ is *gradient related* to $\{x^k\}$ if the following property can be shown:
For any subsequence $\{x^k\}_{k\in\mathcal{K}}$ that converges to a nonstationary point, the corresponding subsequence $\{d^k\}_{k\in\mathcal{K}}$ is bounded and satisfies

$$\limsup_{k\to\infty,\, k\in\mathcal{K}} \nabla f(x^k)'d^k < 0.$$

Proposition (Stationarity of Limit Points) Let $\{x^k\}$ be a sequence generated by the feasible direction method $x^{k+1} = x^k + \alpha^k d^k$. Assume that:

- $\{d^k\}$ is gradient related
- α^k is chosen by the limited minimization rule or the Armijo rule.

Then every limit point of $\{x^k\}$ is a stationary point.

- Proof: Nearly identical to the unconstrained case.

CONDITIONAL GRADIENT METHOD

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

$$\overline{x}^k = \arg\min_{x \in X} \nabla f(x^k)'(x - x^k).$$

- Assume that X is compact, so $\overline{x}^k$ is guaranteed to exist by Weierstrass.

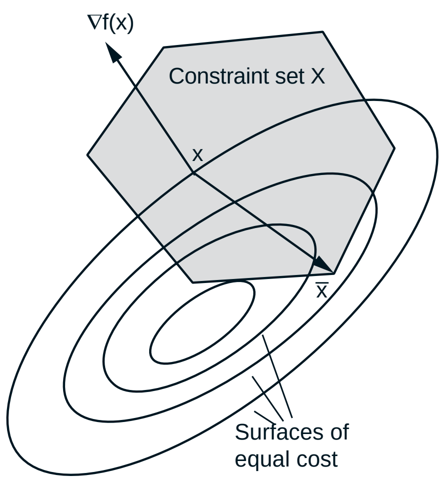
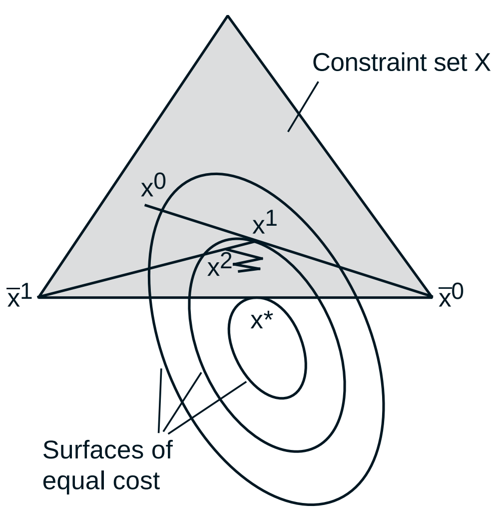


Illustration of the direction of the conditional gradient method.



Operation of the method. Slow (sublinear) convergence.

CONVERGENCE OF CONDITIONAL GRADIENT

- Show that the direction sequence of the conditional gradient method is gradient related, so the generic convergence result applies.

- Suppose that $\{x^k\}_{k\in K}$ converges to a nonstationary point $\tilde{x}$. We must prove that

$$\left\{\|\overline{x}^k - x^k\|\right\}_{k\in K} : \text{bounded}, \qquad \limsup_{k\to\infty,\, k\in K} \nabla f(x^k)'(\overline{x}^k - x^k) < 0.$$

- 1st relation: Holds because $\overline{x}^k \in X$, $x^k \in X$, and X is assumed compact.

- 2nd relation: Note that by definition of $\overline{x}^k$,

$$\nabla f(x^k)'(\overline{x}^k - x^k) \leq \nabla f(x^k)'(x - x^k), \qquad \forall\, x \in X$$

Taking limit as $k \to \infty,\ k \in K$, and min of the RHS over $x \in X$, and using the nonstationarity of $\tilde{x}$,

$$\limsup_{k\to\infty,\, k\in K} \nabla f(x^k)'(\overline{x}^k - x^k) \leq \min_{x\in X} \nabla f(\tilde{x})'(x - \tilde{x}) < 0,$$

thereby proving the 2nd relation.

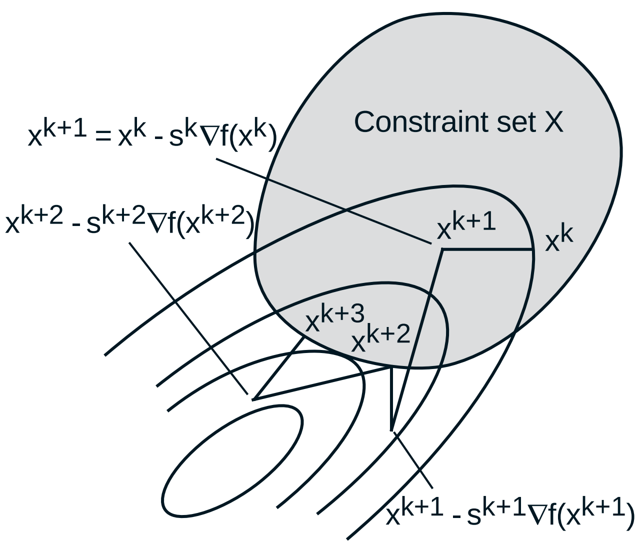
GRADIENT PROJECTION METHODS

• Gradient projection methods determine the feasible direction by using a quadratic cost subproblem. Simplest variant:

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

$$\overline{x}^k = \left[x^k - s^k \nabla f(x^k)\right]^+$$

where, $[\cdot]^+$ denotes projection on the set X, $\alpha^k \in (0, 1]$ is a stepsize, and s^k is a positive scalar.



Gradient projection iterations for the case

$$\alpha^k \equiv 1, \quad x^{k+1} \equiv \overline{x}^k$$

If $\alpha^k < 1$, x^{k+1} is in the line segment connecting x^k and $\overline{x}^k$.

• Stepsize rules for α^k (assuming $s^k \equiv s$): Limited minimization, Armijo along the feasible direction, constant stepsize. Also, Armijo along the projection arc ($\alpha^k \equiv 1$, s^k: variable).

CONVERGENCE

- If α^k is chosen by the limited minimization rule or by the Armijo rule along the feasible direction, every limit point of $\{x^k\}$ is stationary.

- Proof: Show that the direction sequence $\{\overline{x}^k - x^k\}$ is gradient related. Assume $\{x^k\}_{k\in K}$ converges to a nonstationary $\tilde{x}$. Must prove

$$\left\{\|\overline{x}^k - x^k\|\right\}_{k\in K} : \text{bounded}, \qquad \limsup_{k\to\infty,\, k\in K} \nabla f(x^k)'(\overline{x}^k - x^k) < 0.$$

1st relation holds because $\left\{\|\overline{x}^k - x^k\|\right\}_{k\in K}$ converges to $\|[\tilde{x} - s\nabla f(\tilde{x})]^+ - \tilde{x}\|$. By optimality condition for projections, $\big(x^k - s\nabla f(x^k) - \overline{x}^k\big)'(x - \overline{x}^k) \leq 0$ for all $x \in X$. Applying this relation with $x = x^k$, and taking limit,

$$\limsup_{k\to\infty,\, k\in K} \nabla f(x^k)'(\overline{x}^k - x^k) \leq -\frac{1}{s}\left\|\tilde{x} - \left[\tilde{x} - s\nabla f(\tilde{x})\right]^+\right\|^2 < 0$$

- Similar conclusion for constant stepsize $\alpha^k = 1$, $s^k = s$ (under a Lipschitz condition on ∇f).

- Similar conclusion for Armijo rule along the projection arc.

CONVERGENCE RATE – VARIANTS

- Assume $f(x) = \frac{1}{2}x'Qx - b'x$, with $Q > 0$, and a constant stepsize ($a^k \equiv 1$, $s^k \equiv s$). Using the nonexpansiveness of projection

$$\begin{aligned}\left\|x^{k+1} - x^*\right\| &= \left\|\left[x^k - s\nabla f(x^k)\right]^+ - \left[x^* - s\nabla f(x^*)\right]^+\right\| \\ &\le \left\|\left(x^k - s\nabla f(x^k)\right) - \left(x^* - s\nabla f(x^*)\right)\right\| \\ &= \left\|(I - sQ)(x^k - x^*)\right\| \\ &\le \max\left\{|1 - sm|, |1 - sM|\right\}\left\|x^k - x^*\right\|\end{aligned}$$

where m, M: min and max eigenvalues of Q.

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

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

and $H^k > 0$. Since the minimum value above is negative when x^k is nonstationary, $\nabla f(x^k)'(\overline{x}^k - x^k) < 0$. Newton's method for $H^k = \nabla^2 f(x^k)$.

- Variants: Projecting on an expanded constraint set, projecting on a restricted constraint set, combinations with unconstrained methods, etc.