Lecture 6: Optimization over a Convex Set

- Optimality conditions
- Projection theorem
- Feasible direction methods
- Conditional gradient method
- Gradient projection methods

Optimality Conditions

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in X, \end{array}$$

where f is continuously differentiable, X is convex.

- At a local minimum $\boldsymbol{x}^*$, the gradient $\nabla f(\boldsymbol{x}^*)$ makes an angle less than or equal to 90 degrees with all feasible variations $\boldsymbol{x} - \boldsymbol{x}^*$, $\boldsymbol{x} \in X$.

  a) If $\boldsymbol{x}^*$ is a local minimum of f over X, then

  $$\nabla f(\boldsymbol{x}^*)'(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \forall\, \boldsymbol{x} \in X.$$

  b) If f is convex over X, then this condition is also sufficient for $\boldsymbol{x}^*$ to minimize f over X.

- The optimality condition fails when X is not convex. For example, $\boldsymbol{x}^*$ is a local min but we have $f(\boldsymbol{x}^*)'(\boldsymbol{x} - \boldsymbol{x}^*) < 0$ for some feasible vector $x \in X$.

Proof

a) Suppose that $\nabla f(\boldsymbol{x}^*)'(\boldsymbol{x} - \boldsymbol{x}^*) < 0$ for some $\boldsymbol{x} \in X$. By the Mean Value Theorem, for every $\epsilon > 0$ there exists an $s \in [0, 1]$ such that

$$f\left(\boldsymbol{x}^* + \epsilon(\boldsymbol{x} - \boldsymbol{x}^*)\right) = f(\boldsymbol{x}^*) + \epsilon \nabla f\left(\boldsymbol{x}^* + s\epsilon(\boldsymbol{x} - \boldsymbol{x}^*)\right)'(\boldsymbol{x} - \boldsymbol{x}^*).$$

Since ∇f is continuous, for sufficiently small $\epsilon > 0$,

$$\nabla f\left(\boldsymbol{x}^* + s\epsilon(\boldsymbol{x} - \boldsymbol{x}^*)\right)'(\boldsymbol{x} - \boldsymbol{x}^*) < 0,$$

so that $f\left(\boldsymbol{x}^* + \epsilon(\boldsymbol{x} - \boldsymbol{x}^*)\right) < f(\boldsymbol{x}^*)$. The vector $\boldsymbol{x}^* + \epsilon(\boldsymbol{x} - \boldsymbol{x}^*)$ is feasible for all $\epsilon \in [0, 1]$ because X is convex, contradicting the local optimality of $\boldsymbol{x}^*$.

b) Using the convexity of f

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)'(\boldsymbol{x} - \boldsymbol{x}^*)$$

for every $\boldsymbol{x} \in X$. If the condition $\nabla f(\boldsymbol{x}^*)'(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0$ holds for all $\boldsymbol{x} \in X$, we obtain $f(x) \geq f(\boldsymbol{x}^*)$, so $\boldsymbol{x}^*$ minimizes f over X.

Optimization Subject to Bounds

- Let $X = \{\boldsymbol{x} \mid \boldsymbol{x} \geq 0\}$. Then the necessary condition for $\boldsymbol{x}^* = (x_1^*, ..., x_n^*)'$ to be a local min is

$$\sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x}^*)}{\partial x_i}(x_i - x_i^*) \geq 0, \quad \forall\, x_i \geq 0, i = 1, ..., n.$$

- Fix i. Let $x_j = x_j^*$ for $j \neq i$ and $x_i = x_i^* + 1$:

$$\frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} \geq 0, \quad \forall\, i.$$

- If $x_i^* > 0$, let also $x_j = x_j^*$ for $j \neq i$ and $x_i = \frac{1}{2}x_i^*$. Then $\frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} \leq 0$, so

$$\frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} = 0, \quad \text{if } x_i^* > 0.$$

Optimization over a Simplex

Simplex: $X = \left\{ \boldsymbol{x} \mid \boldsymbol{x} \geq \mathbf{0}, \sum_{i=1}^{n} x_i = r \right\},$ where $r > 0$ is a given scalar.

- Necessary condition for $\boldsymbol{x}^* = (x_1^*, \ldots, x_n^*)'$ to be a local min:

$$\sum_{i=1}^{n} \frac{\partial f(\boldsymbol{x}^*)}{\partial x_i}(x_i - x_i^*) \geq 0, \quad \forall\, x_i \geq 0 \text{ with } \sum_{i=1}^{n} x_i = r.$$

- Fix i with $x_i^* > 0$ and let j be any other index. Use $\boldsymbol{x}$ with $x_i = 0$, $x_j = x_j^* + x_i^*$, and $x_m = x_m^*$ for all $m \neq i, j$:

$$\left(\frac{\partial f(\boldsymbol{x}^*)}{\partial x_j} - \frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} \right) x_i^* \geq 0,$$

$$x_i^* > 0 \quad \Longrightarrow \quad \frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} \leq \frac{\partial f(\boldsymbol{x}^*)}{\partial x_j}, \quad \forall\, j.$$

Projection Over A Convex Set

- Let $\boldsymbol{z} \in \mathbb{R}^n$ and a closed convex set X be given. Problem:

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) = \|\boldsymbol{z} - \boldsymbol{x}\|^2 \\ \text{subject to} & \boldsymbol{x} \in X. \end{array}$$

  has a unique solution $\boldsymbol{x}^* = [\boldsymbol{z}]^+$ (the projection of z).

- Necessary and sufficient condition for $\boldsymbol{x}^*$ to be the projection: The angle between $\boldsymbol{z} - \boldsymbol{x}^*$ and $\boldsymbol{x} - \boldsymbol{x}^*$ should be greater or equal to 90 degrees for all $\boldsymbol{x} \in X$, or $(\boldsymbol{z} - \boldsymbol{x}^*)'(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0$

- If X is a subspace, $\boldsymbol{z} - \boldsymbol{x}^* \perp X$.

- The mapping $f : \mathbb{R}^n \mapsto X$ defined by $f(\boldsymbol{x}) = [\boldsymbol{x}]^+$ is continuous and non-expansive, that is,

$$\|[\boldsymbol{x}]^+ - [\boldsymbol{y}]^+\| \le \|\boldsymbol{x} - \boldsymbol{y}\|, \quad \forall\, \boldsymbol{x},\ \boldsymbol{y} \in \mathbb{R}^n.$$

  Why? [Add $\langle \boldsymbol{x} - [\boldsymbol{x}]^+, [\boldsymbol{y}]^+ - [\boldsymbol{x}]^+\rangle \le 0$ to $\langle \boldsymbol{y} - [\boldsymbol{y}]^+, [\boldsymbol{x}]^+ - [\boldsymbol{y}]^+\rangle \le 0$]

- **Exercise:** Assume X is convex. A vector $\boldsymbol{x}^* \in X$ is a stationary point of

$$\boxed{\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in X, \end{array}}$$

  iff $\boldsymbol{x}^*$ satisfies the following fixed point equation

$$\boldsymbol{x}^* = [\boldsymbol{x}^* - \alpha \nabla f(\boldsymbol{x}^*)]^+$$

  for any $\alpha > 0$.

Feasible Directions Method

- A feasible direction at an $\boldsymbol{x} \in X$ is a vector $\boldsymbol{d} \neq \mathbf{0}$ such that $\boldsymbol{x} + \alpha \boldsymbol{d}$ is feasible for all sufficiently small $\alpha > 0$

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

- A feasible direction method:

$$\boldsymbol{x}^{r+1} = \boldsymbol{x}^r + \alpha_r \boldsymbol{d}^r,$$

  where $\boldsymbol{d}^r$: feasible descent direction, i.e., $\nabla f(\boldsymbol{x}^r)' \boldsymbol{d}^r < 0$, and $\alpha_r > 0$ is such that $\boldsymbol{x}^{r+1} \in X$.

- Alternative definition:

$$\boldsymbol{x}^{r+1} = \boldsymbol{x}^r + \alpha_r (\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r),$$

where $\alpha_r \in (0, 1]$ and if $\boldsymbol{x}^r$ is nonstationary, then there exists an

$$\bar{\boldsymbol{x}}^r \in X, \quad \nabla f(\boldsymbol{x}^r)'(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r) < 0.$$

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

$$f(\boldsymbol{x}^r) - f(\boldsymbol{x}^r + \beta^m(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r) \geq -\sigma\beta^m \nabla f(\boldsymbol{x}^r)'(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r)$$

Convergence Analysis

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

- The direction sequence $\{\boldsymbol{d}^r\}$ is gradient related to $\{\boldsymbol{x}^r\}$ if the following property can be shown: For any subsequence $\{\boldsymbol{x}^r\}_{r\in K}$ that converges to a nonstationary point, the corresponding subsequence $\{\boldsymbol{d}^r\}_{r\in K}$ is bounded and satisfies
$$\limsup_{r\to\infty, r\in K} \nabla f(\boldsymbol{x}^r)'\boldsymbol{d}^r < 0.$$

- **Proposition (Stationarity of Limit Points)** Let $\{\boldsymbol{x}^r\}$ be a sequence generated by the feasible direction method $\boldsymbol{x}^{r+1} = \boldsymbol{x}^r + \alpha_r \boldsymbol{d}^r$. Assume that:
  - ⋆ $\{\boldsymbol{d}^r\}$ is gradient related
  - ⋆ α_r is chosen by the limited minimization rule or the Armijo rule.

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

- Proof is nearly identical to the unconstrained case.

Conditional Gradient Method

- Define $\boldsymbol{x}^{r+1} = \boldsymbol{x}^r + \alpha_r(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r)$, where

$$\bar{\boldsymbol{x}}^r = \arg\min_{\boldsymbol{x} \in X} \nabla f(\boldsymbol{x}^r)'(\boldsymbol{x} - \boldsymbol{x}^r).$$

- Assume that X is compact, so $\bar{\boldsymbol{x}}^r$ is guaranteed to exist.

- Slow (sublinear) convergence.

Convergence of the Conditional Gradient Method

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

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

$$\{\|\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r\|\}_{r\in K}: \text{ bounded}, \quad \limsup_{r\to\infty, r\in K} \nabla f(\boldsymbol{x}^r)'(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r) < 0.$$

- 1st relation: Holds because $\bar{\boldsymbol{x}}^r \in X$, $\boldsymbol{x}^r \in X$, and X is compact.

- 2nd relation: Note that by definition of $\bar{\boldsymbol{x}}^r$,

$$\nabla f(\boldsymbol{x}^r)'(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r) \le \nabla f(\boldsymbol{x}^r)'(\boldsymbol{x} - \boldsymbol{x}^r), \quad \text{for all } \boldsymbol{x} \in X$$

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

$$\limsup_{r\to\infty, r\in K} \nabla f(\boldsymbol{x}^r)'(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r) \le \min_{\boldsymbol{x}\in X} \nabla f(\tilde{x})'(\boldsymbol{x} - \tilde{\boldsymbol{x}}) < 0,$$

thereby proving the 2nd relation.

Gradient Projection Methods

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

$$\begin{aligned} \boldsymbol{x}^{r+1} &= \boldsymbol{x}^r + \alpha_r(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r) \\ \bar{\boldsymbol{x}}^r &= \mathrm{proj}_X\left[\boldsymbol{x}^r - s_r \nabla f(\boldsymbol{x}^r)\right] \end{aligned}$$

  where, $\mathrm{proj}_X[\cdot]$ denotes projection on the set X, $\alpha_r \in (0, 1]$ is a stepsize, and s_r is a positive scalar.

- $\bar{\boldsymbol{x}}^r$ can be defined as

$$\bar{\boldsymbol{x}}^r = \arg\min_{\boldsymbol{x} \in X} \nabla f(\boldsymbol{x}^r)'(\boldsymbol{x} - \boldsymbol{x}^r) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^r\|^2$$

  so $(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r)$ is a descent direction. The proximal term $\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^r\|^2$ provides regularization. [No need for X to be compact.]

- Stepsize rules for α_r
  - ⋆ assuming $s_r \equiv s$: Limited minimization, Armijo along the feasible direction, constant stepsize.
  - ⋆ Also, assuming $\alpha_r \equiv 1$: Armijo along the projection arc (s_r: variable).

Convergence Analysis of GP Methods

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

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

$$\{\|\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r\|\}_{r\in K}: \text{ bounded}, \quad \limsup_{r\to\infty, r\in K} \nabla f(\boldsymbol{x}^r)'(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r) < 0.$$

- 1st relation holds because $\{\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r\}_{r\in K}$ converges to $\mathrm{proj}_X[\tilde{\boldsymbol{x}} - s\nabla f(\tilde{\boldsymbol{x}})] - \tilde{\boldsymbol{x}}$. By optimality condition for projections,

$$(\boldsymbol{x}^r - s\nabla f(\boldsymbol{x}^r) - \bar{\boldsymbol{x}}^r)'(\boldsymbol{x} - \bar{\boldsymbol{x}}^r) \le 0, \quad \text{for all } \boldsymbol{x} \in X.$$

Applying this relation with $\boldsymbol{x} = \boldsymbol{x}^r$, and taking limit,

$$\limsup_{r\to\infty, r\in K} \nabla f(\boldsymbol{x}^r)'(\bar{\boldsymbol{x}}^r - \boldsymbol{x}^r) \le -\frac{1}{s}\|\tilde{\boldsymbol{x}} - \text{proj}_X[\tilde{\boldsymbol{x}} - s\nabla f(\tilde{\boldsymbol{x}})]\|^2 < 0$$

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

- Similar conclusion for Armijo rule along the projection arc.

Convergence Rate Analysis

- Consider a strongly convex quadratic function $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x}$, with $\boldsymbol{A} \succ \boldsymbol{0}$.

- $\exists$ a unique solution $\boldsymbol{x}^* \in X$ satisfying $\boldsymbol{x}^* = \mathrm{proj}_X[\boldsymbol{x}^* - \alpha^r \nabla f(\boldsymbol{x}^*)]$, so

$$
\begin{aligned}
\|\boldsymbol{x}^{r+1} - \boldsymbol{x}^*\| &= \|\,\mathrm{proj}_X[\boldsymbol{x}^r - \alpha_r \nabla f(\boldsymbol{x}^r)] - \mathrm{proj}_X[\boldsymbol{x}^* - \alpha_r \nabla f(\boldsymbol{x}^*)]\| \\
&\leq \|(\boldsymbol{x}^r - \boldsymbol{x}^*) - \alpha_r(\nabla f(\boldsymbol{x}^r) - \nabla f(\boldsymbol{x}^*))\| \\
&= \|(\boldsymbol{I} - \alpha_r \boldsymbol{A})(\boldsymbol{x}^r - \boldsymbol{x}^*)\| \\
&\leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)\|\boldsymbol{x}^r - \boldsymbol{x}^*\| = \left(1 - \frac{2}{\kappa + 1}\right)\|\boldsymbol{x}^r - \boldsymbol{x}^*\|.
\end{aligned}
$$

- Convergence rate depends on $\kappa = \lambda_{\max}/\lambda_{\min}$, but *independent* of dimension.

- Requires $O(1)\kappa \ln(1/\epsilon)$ to find an ϵ-relative optimal solution.