Lecture 19: Optimization Programming

Hao Helen Zhang

Outline

1. Unconstrained Optimization
2. Constrained Optimization
   - Equality-constrained Optimization
   - Inequality-constrained Optimization
   - Mixture-constrained Optimization
3. Special Mathematical Programming
   - Quadratic Programming (QP), Linear Programming (LP)

Unconstrained Optimization

Consider the minimization problem

$$\min_{\mathbf{x}} \quad f(\mathbf{x}),$$

where $\mathbf{x} \in R^d$, and f is continuously differentiable.

- Necessary condition for local minimums
  - If $\mathbf{x}^*$ is a local minimum of $f(\mathbf{x})$, then

    $$\nabla f(\mathbf{x}^*) = 0.$$

  - A point satisfying the necessary condition is called a *stationary point*, which can be a local minimum, local maximum, or saddle point.
- Sufficient conditions for $\mathbf{x}^*$ be a local minimum are:

$$\begin{aligned} \nabla f(\mathbf{x}^*) &= 0, \\ \nabla^2 f(\mathbf{x}^*) & \text{ is positive definite.} \end{aligned}$$

Constrained Optimization

Consider the minimization problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \geq 0, \quad i = 1, \cdots, m. \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \cdots, k. \end{aligned}$$

- Assume $\mathbf{x} \in R^d$.
- $f : R^d \longrightarrow R$ is called the *objective function*
- $g_i : R^d \longrightarrow R, \ i = 1, \cdots, m$ are *inequality constraints*
- $h_j : R^d \longrightarrow R, \ j = 1, \cdots, k$ are *equality constraints.*
- Assume f, g_i's, h_j's are all continuously differentiable.

Single Equality Constraint

Consider the minimization problem

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) \tag{1}$$

$$\text{subject to} \quad h(\mathbf{x}) = 0. \tag{2}$$

Introduce the Lagrange function:

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda h(\mathbf{x})$$

Lagrange extreme value theory:

- If $\mathbf{x}^*$ is a local minimum under the constraint, then there exist some λ^* such that

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{x}^*, \lambda^*) &= 0, \\ \nabla_{\lambda} L(\mathbf{x}^*, \lambda^*) &= 0. \end{aligned}$$

Interpretations

- Above conditions imply that $(\mathbf{x}^*, \lambda^*)$ is a stationary point of L.
- The above equations are equivalent to

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \lambda^* \nabla_{\mathbf{x}} h(\mathbf{x}^*) = 0,$$
$$h(\mathbf{x}^*) = 0.$$

- Note the first condition gives d equations.

Multiple Equality Constraints

Consider the minimization problem

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) \tag{3}$$

$$\text{subject to} \quad h_j(\mathbf{x}) = 0, \quad j = 1, \cdots, k. \tag{4}$$

Introduce multiple Lagrange multipliers: $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_k)$

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{j=1}^{k} \lambda_j h_j(\mathbf{x})$$

Lagrange extreme theory:

- If $\mathbf{x}^*$ is a local minimum under the constraint, then there exist some $\boldsymbol{\lambda}^* = (\lambda_1^*, \cdots, \lambda_k^*)$ such that

$$\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0, \quad \nabla_{\boldsymbol{\lambda}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0.$$

In other words, $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a stationary point of L.

Example: Entropy Maximization Problem

Consider the minimization problem

$$\min_{p_1,\cdots,p_n} \sum_{i=1}^n p_i \log(p_i), \quad \text{subject to} \quad \sum_{i=1}^n p_i = 1.$$

The Lagrange function is

$$L(\mathbf{p}, \lambda) = \sum_{i=1}^n p_i \log(p_i) + \lambda(\sum_{i=1}^n p_i - 1).$$

$$\frac{\partial L}{\partial p_i} = \log(p_i) + 1 + \lambda = 0, \quad \log(p_i) = -1 - \lambda, \quad i = 1, \cdots, n.$$

Together with $\sum_{i=1}^n p_i = 1$, we have

$$p_1^* = \cdots = p_n^* = \frac{1}{n}.$$

This problem is known as "maximizing information entropy".

Generalized Lagrange Methods

Consider the minimization problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \geq 0, \quad i = 1, \cdots, m. \end{aligned}$$

Introduce multiple non-negative Lagrange multipliers:

$$\mu_1 \geq 0, \cdots, \mu_m \geq 0.$$

$$L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) - \sum_{i=1}^{m} \mu_i g_i(\mathbf{x}).$$

- μ_i's are also called *dual variables* or *slack variables*. Correspondingly, we call $\mathbf{x}$ as *primal variables*.
- We would like to minimize L with respect to $\mathbf{x}$ and maximize L with respect to $\boldsymbol{\mu}$.

KKT Necessary Optimality Conditions

If $\mathbf{x}^*$ is a local minimum under the above inequality constraints, then there exist some $\boldsymbol{\mu}^* = (\mu_1^*, \cdots, \mu_m^*)$ such that

- Stationarity: $\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*) = 0$, i.e.,

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) - \sum_{i=1}^{m} \mu_k^* \nabla_{\mathbf{x}} g_i(\mathbf{x}^*) = 0.$$

- Primal feasibility: $g_i(\mathbf{x}^*) \geq 0$ for $i = 1, \cdots, m$.
- Dual feasibility: $\mu_i \geq 0$ for $i = 1, \cdots, m$.
- Complementary slackness:

$$\mu_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \cdots, m.$$

KKT = Karush-Kuhn-Tucker

Sufficient Optimality Conditions

The above necessary conditions are also sufficient for optimality, if

- f, g_i's are continuously differentiable
- f and g_i's are convex functions.

Under this case,

- There exists a solution $\mathbf{x}^*$ satisfying the KKT condition,
- $\mathbf{x}^*$ is actually a global optimum.

Examples: SVM problems

Alternative Constraints

Consider the minimization problem

$$\begin{array}{ll} \min\limits_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m. \end{array}$$

Introduce multiple non-negative Lagrange multipliers:

$$\mu_1 \geq 0, \cdots, \mu_m \geq 0.$$

$$L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \mu_i g_i(\mathbf{x}).$$

- μ_i's are also called *dual variables* or *slack variables*. Correspondingly, we call $\mathbf{x}$ as *primal variables*.
- We would like to minimize L with respect to $\mathbf{x}$ and maximize L with respect to $\boldsymbol{\mu}$.

KKT Necessary Optimality Conditions

If $\mathbf{x}^*$ is a local minimum under the above inequality constraints, then there exist some $\boldsymbol{\mu}^* = (\mu_1^*, \cdots, \mu_m^*)$ such that

- Stationarity: $\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*) = 0$, i.e.,

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^{m} \mu_k^* \nabla_{\mathbf{x}} g_i(\mathbf{x}^*) = 0.$$

- Primal feasibility: $g_i(\mathbf{x}^*) \leq 0$ for $i = 1, \cdots, m$.
- Dual feasibility: $\mu_i \geq 0$ for $i = 1, \cdots, m$.
- Complementary slackness:

$$\mu_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \cdots, m.$$

Generalized Lagrange Methods

Consider the minimization problem

$$\begin{array}{ll} \min_{\mathbf{x}} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \geq 0, \quad i = 1, \cdots, m. \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \cdots, k. \end{array}$$

Introduce two sets of Lagrange multipliers:

$$\mu_1 \geq 0, \cdots, \mu_m \geq 0, \lambda_1, \cdots, \lambda_k.$$

Define the Lagrange function:

$$L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{i=1}^{m} \mu_i g_i(\mathbf{x}) + \sum_{j=1}^{k} \lambda_j h_j(\mathbf{x})$$

- μ_i's and λ_j's are *dual variables* or *slack variables*.

KKT Necessary Optimality Conditions

If $\mathbf{x}^*$ is a local minimum under the above inequality constraints, then there exist some $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ such that

- Stationarity: $\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = 0$, i.e.,

$$\nabla_{\mathbf{x}} f(\mathbf{x}^*) - \sum_{i=1}^{m} \mu_i^* \nabla_{\mathbf{x}} g_i(\mathbf{x}^*) + \sum_{j=1}^{k} \lambda_j^* \nabla_{\mathbf{x}} h_j(\mathbf{x}^*) = 0.$$

- Primal feasibility:

$$\begin{aligned} g_i(\mathbf{x}^*) &\geq 0, \quad i = 1, \cdots, m; \\ h_j(\mathbf{x}^*) &= 0, \quad j = 1, \cdots, k. \end{aligned}$$

- Dual feasibility: $\mu_i \geq 0$ for $i = 1, \cdots, m$.
- Complementary slackness:

$$\mu_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \cdots, m.$$

Sufficient Optimality Conditions

The above necessary conditions are also sufficient for optimality, if

- f, g_i's are continuously differentiable.
- f and g_i's are convex functions,
- h_j's are linear constraints.

Under this case,

- There exists a solution $\mathbf{x}^*$ satisfying the KKT condition,
- $\mathbf{x}^*$ is actually a global optimum.

Quadratic Programming (QP)

- The objective function is quadratic in $\mathbf{x}$
- The constraints are linear in $\mathbf{x}$.

Consider the minimization problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x}, \\ \text{subject to} \quad & A\mathbf{x} \leq \mathbf{b}, \\ & E\mathbf{x} = \mathbf{d}. \end{aligned}$$

where Q is a symmetric matrix of $d \times d$, A is a matrix of $m \times d$, E is a matrix of $k \times d, \mathbf{c} \in R^d, \mathbf{b} \in R^m$, and $\mathbf{d} \in R^k$.
Example: SVM

Properties of QP

- If Q is a non-negative definite matrix, then $f(\mathbf{x})$ is convex.
- If Q is a non-negative definite, the above QP has a global minimizer if there exists at least one $\mathbf{x}$ satisfying the constraints and $f(\mathbf{x})$ is bounded below in the feasible region.
- If Q is a positive definite, the above QP has a unique global minimizer.
- If $Q = 0$, the problem becomes a linear programming (LP).

From optimization theory,

- If Q is non-negative definite, the necessary and sufficient condition for a point $\mathbf{x}$ to be a global minimizer is for it to satisfy the KKT condition.

Solving QP

- If there are only equality constraints, then the QP can be solved by a linear system.
- In general, a variety of methods for solving the QP are commonly used, including
    - Methods: ellipsoid, interior point, active set, conjugate gradient, etc
    - Software packages: Matlab, CPLEX, AMPL, GAMS, etc.
- For positive definite Q, the ellipsoid method solves the problem in polynomial time. In other words, the running time is upper bounded by a polynomial in the size of the input for the algorithm, i.e., $T = O(d^k)$ for some k.

Dual Problem for QP

Consider the minimization problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x}, \\ \text{subject to} \quad & A\mathbf{x} \leq \mathbf{b}. \end{aligned}$$

The Lagrange function.

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T (A\mathbf{x} - \mathbf{b}).$$

We need to minimize L with respect to $\mathbf{x}$ and maximize L with respect to $\boldsymbol{\lambda}$. The *dual* function is defined as a function of dual variables only by solving $\mathbf{x}$ first with any fixed $\boldsymbol{\lambda}$:

$$G(\boldsymbol{\lambda}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}).$$

Dual Problem Derivation for QP

With fixed $\boldsymbol{\lambda}$, we solve the equation $\nabla L_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}) = 0$ and obtain

$$\mathbf{x}^* = -Q^{-1}A^T\boldsymbol{\lambda}.$$

Plug this into L, we get the dual function

$$G(\boldsymbol{\lambda}) = -\frac{1}{2}\boldsymbol{\lambda}^T A Q^{-1} A^T - \mathbf{b}^T\boldsymbol{\lambda}.$$

The dual problem of the QP is

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \quad & G(\boldsymbol{\lambda}) = -\frac{1}{2}\boldsymbol{\lambda}^T A Q^{-1} A^T - \mathbf{b}^T\boldsymbol{\lambda}, \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq 0. \end{aligned}$$

- We first find $\boldsymbol{\lambda}^* = \arg\max_{\boldsymbol{\lambda}} G(\boldsymbol{\lambda})$, and then compute
$$\mathbf{x}^* = -Q^{-1}A^T\boldsymbol{\lambda}^*.$$
- The idea is similar to the profile method.