Lecture 13: CS395T Numerical Optimization for Graphics and AI — Theory of Constrained Optimization

Qixing Huang
The University of Texas at Austin
huangqx@cs.utexas.edu

Disclaimer

This note is adapted from

- Section 12 of *Numerical Optimization* by Jorge Nocedal and Stephen J. Wright. Springer series in operations research and financial engineering. Springer, New York, NY, 2. ed. edition, (2006)

1 Introduction

The second part of this class is about minimizing functions subject to constraints on the variables. A general formulation for these problems is

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & c_i(\boldsymbol{x}) = 0, \quad i \in \mathcal{E}, \\ & c_i(\boldsymbol{x}) \geq 0, \quad i \in \mathcal{I}, \end{aligned} \tag{1}$$

where f and the functions c_i are all smooth, real-valued functions on a subset of $\mathbb{R}^n$, and $\mathcal{I}$ and $\mathcal{E}$ are two finite sets of indices.

If we define the feasible set Ω to be the set of points $\boldsymbol{x}$ that satisfies the constraints, that is,

$$\Omega = \{\boldsymbol{x} | c_i(\boldsymbol{x}) = 0, i \in \mathcal{E}; c_i(\boldsymbol{x}) \geq 0, i \in \mathcal{I}\}, \tag{2}$$

then we can always rewrite (1) more compactly as

$$\min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x}) \tag{3}$$

In this lecture, we will go through some optimally conditions. They are generalized from their counterparts in the constrained case. They are summarized below.

2 First-Order Optimality Conditions

To define the first-order optimality conditions, we first consider the notion of active sets:

Definition 2.1. *The active set* $\mathcal{A}(\boldsymbol{x})$ *at any feasible* $\boldsymbol{x}$ *consists of the equality constraint indices from* $\mathcal{E}$ *together with the indices of the inequality constraints* i *for which* $c_i(\boldsymbol{x}) = 0$*; that is,*

$$\mathcal{A}(\boldsymbol{x}) = \mathcal{E} \cup \{i \in \mathcal{I} | c_i(\boldsymbol{x}) = 0\}.$$

We will also need the characterization of a local solution:

Definition 2.2. *A vector* $\boldsymbol{x}^\star$ *is a local solution of the problem (3) if* $\boldsymbol{x}^\star \in \Omega$ *and there is a neighborhood* $\mathcal{N}$ *of* $\boldsymbol{x}^\star$ *such that* $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^\star)$ *for* $\boldsymbol{x} \in \mathcal{N} \cap \Omega$*.*

To define the optimality conditions, we will also need the so-called LICQ condition.

Definition 2.3. *Given the point* $\boldsymbol{x}$ *and the active set* $\mathcal{A}(\boldsymbol{x})$ *defined in definition 2.1, we say that the linear independence constraint qualification (LICQ) holds if the set of active constraint gradients* $\{\nabla c_i(\boldsymbol{x}), i \in \mathcal{A}(\boldsymbol{x})\}$ *is linearly independent.*

Now we are ready to introduce first-order necessary conditions, which often known as the Karush-Kuhn-Tucker conditions, or KKT conditions for short.

Theorem 2.1. *Consider the Lagrangian given by*

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) - \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i c_i(\boldsymbol{x}).$$

Suppose $\boldsymbol{x}^\star$ *is a local solution of (1), that the functions* f *and* c_i *in (1) are continuously differentiable, and that the LICQ holds at* $\boldsymbol{x}^\star$*. Then there is a Lagrangian multipler vector* $\lambda^\star$*, with components* $\lambda_i^\star, i \in \mathcal{E} \cup \mathcal{I}$*, such that the following conditions are satisfied at* $(\boldsymbol{x}^\star, \lambda^\star)$

$$\begin{aligned} \nabla_{\boldsymbol{x}} L(\boldsymbol{x}^\star, \lambda^\star) &= 0, && (4) \\ c_i(\boldsymbol{x}^\star) &= 0, \quad \text{for all } i \in \mathcal{E}, && (5) \\ c_i(\boldsymbol{x}^\star) &\geq 0, \quad \text{for all } i \in \mathcal{I}, && (6) \\ \lambda_i^\star &\geq 0, \quad \text{for all } i \in \mathcal{I}, && (7) \\ \lambda_i^\star c_i(\boldsymbol{x}^\star) &= 0, \quad \text{for all } i \in \mathcal{I} \cup \mathcal{E}. && (8) \end{aligned}$$

3 Second-Order Optimality Conditions

To derive second-order optimality conditions, we begin with defining the feasible direction set, which we define as follows.

Definition 3.1. *Given a feasible point* $\boldsymbol{x}$ *and the active constraint set* $\mathcal{A}(\boldsymbol{x})$ *of Definition 2.1, the set of linearized feasible directions* $\mathcal{F}(\boldsymbol{x})$ *is*

$$\mathcal{F}(\boldsymbol{x}) = \left\{ \boldsymbol{d} \middle| \begin{array}{ll} \boldsymbol{d}^T \nabla c_i(\boldsymbol{x}) = 0, & \text{for all } i \in \mathcal{E}, \\ \boldsymbol{d}^T \nabla c_i(\boldsymbol{x}) \geq 0, & \text{for all } i \in \mathcal{A}(\boldsymbol{x}) \cap \mathcal{I} \end{array} \right\} \quad (9)$$

Definition 3.2. *Given* $\mathcal{F}(\boldsymbol{x}^\star)$ *from Definition 3.1 and some Lagrangian multipler vector* $\lambda^\star$ *satisfying the KKT conditions, we define the critical cone* $\mathcal{C}(\boldsymbol{x}^\star, \lambda^\star)$ *as follows:*

$$\mathcal{C}(\boldsymbol{x}^\star, \lambda^\star) = \{\boldsymbol{w} \in \mathcal{F}(\boldsymbol{x}^\star) | \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^\star) = 0, \text{all } i \in \mathcal{A}(\boldsymbol{x}^\star) \cap \mathcal{I} \text{ with } \lambda_i^\star > 0\} \quad (10)$$

Equivalently,

$$\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^\star, \lambda^\star) \leftrightarrow \begin{cases} \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^\star) = 0, & \text{for all } i \in \mathcal{E}, \\ \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^\star) = 0, & \text{for all } i \in \mathcal{A}(\boldsymbol{x}^\star) \cap \mathcal{I} \text{ with } \lambda_i^\star > 0, \\ \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^\star) \geq 0, & \text{for all } i \in \mathcal{A}(\boldsymbol{x}^\star) \cap \mathcal{I} \text{ with } \lambda_i^\star = 0. \end{cases} \tag{11}$$

The critical cone contains those directions $\boldsymbol{w}$ that would tend to "adhere" to the active inequality constraints even when we were to make small changes to the objective (those indices $i \in \mathcal{I}$ for which the Lagrange multiplier component $\lambda_i^\star$ is positive), as well as to the equality constraints. An important property of these directions is:

$$\boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^\star, \lambda^\star) \rightarrow \boldsymbol{w}^T \nabla f(\boldsymbol{x}^\star) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^\star \boldsymbol{w}^T \nabla c_i(\boldsymbol{x}^\star) = 0.$$

Theorem 3.1. *(**Second-Order Necessary Conditions**.) Suppose $\boldsymbol{x}^\star$ is a local solution of (1) and that LICQ condition is satisfied. Let $\lambda^\star$ be the Lagrangian multiplier vector for which the KKT conditions are satisfied. Then*

$$\boldsymbol{w}^T \nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}^\star, \lambda^\star) \boldsymbol{w} \geq 0, \quad \text{for all } \boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^\star, \lambda^\star).$$

The corresponding Second-Order Sufficient Conditions are given below

Theorem 3.2. *(**Second-Order Sufficient Conditions**.) Suppose that for some feasible point $\boldsymbol{x}^\star \in \mathbb{R}^n$ there is a Lagrangian multipler vector $\lambda^\star$ such that the KKT conditions are satisfied. Suppose also that*

$$\boldsymbol{w}^T \nabla^2_{\boldsymbol{xx}} L(\boldsymbol{x}^\star, \lambda^\star) \boldsymbol{w} > 0, \quad \text{for all } \boldsymbol{w} \in \mathcal{C}(\boldsymbol{x}^\star, \lambda^\star) \backslash \{\boldsymbol{0}\}.$$

Then $\boldsymbol{x}^\star$ is a strict local solution for (1).