273a: Optimization
Nonlinear optimization with inequality constraints

Instructor: Wotao Yin
Department of Mathematics, UCLA

Fall 2015

material taken from the textbook Chong-Zak, 4th Ed.

Overview

- we discuss how to recognize a solution to the problem

$$\begin{aligned} \text{minimize } & f(\boldsymbol{x}) \\ \text{subject to } & \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0} \\ & \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}. \end{aligned}$$

  where $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\boldsymbol{h} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\boldsymbol{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p$, and

$$\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} h_1(\boldsymbol{x}) \\ \vdots \\ h_m(\boldsymbol{x}) \end{bmatrix} \qquad \boldsymbol{g}(\boldsymbol{x}) = \begin{bmatrix} g_1(\boldsymbol{x}) \\ \vdots \\ g_p(\boldsymbol{x}) \end{bmatrix}$$

- assume **continuous differentiability:** $f, \boldsymbol{h}, \boldsymbol{g} \in C^1$
- **goals of this lecture**
  1. 1st and 2nd order optimality conditions
  2. solutions to certain NLPs with in/equality constraints

- $g_j(\boldsymbol{x}) > 0$ called *inactive*; $g_j(\boldsymbol{x}) = 0$ called *active*;
- set of active inequality constraints:

$$J(\boldsymbol{x}) := \{j : g_j(\boldsymbol{x}) = 0\}$$

- a feasible point $\boldsymbol{x}^* \in \mathbb{R}^n$ is **regular** if

$$\nabla h_1(\boldsymbol{x}^*), \ldots, \nabla h_m(\boldsymbol{x}^*), \text{ and } g_j(\boldsymbol{x}^*),\ j \in J(\boldsymbol{x})$$

  are linearly independent

Karush-Kuhn-Tucker (KKT) conditions

Theorem

Let $f, \boldsymbol{h}, \boldsymbol{g} \in C^1$ and $\boldsymbol{x}^*$ be a regular local minimizer for

$$\text{minimize}\, f(\boldsymbol{x}) \qquad \text{subject to } \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},\ \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0}.$$

Then, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^n$ and $\boldsymbol{\mu}^* \in \mathbb{R}^p$ such that

1. $\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{\lambda}^* + D\boldsymbol{g}(\boldsymbol{x}^*)^T \boldsymbol{\mu}^* = \boldsymbol{0}$;
2. $\boldsymbol{\mu}^* \geq \boldsymbol{0}$;
3. $\boldsymbol{g}(\boldsymbol{x}^*)^T \boldsymbol{\mu}^* = 0$.

The KKT conditions also conclude

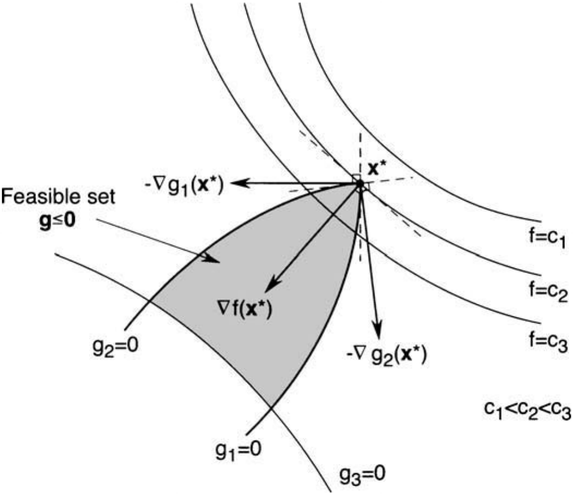
4. $\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$;
5. $\boldsymbol{g}(\boldsymbol{x}^*) \leq \boldsymbol{0}$.

- $\boldsymbol{\lambda}^*$ is called the Lagrange multipliers
- $\boldsymbol{\mu}^*$ is called the KKT multipliers
- Since $\mu_j^* \geq 0$ and $\boldsymbol{g}_j(\boldsymbol{x}^*) \leq 0$ for each j, the condition

$$\boldsymbol{g}(\boldsymbol{x}^*)^T \boldsymbol{\mu}^* = \mu_1^* g_1(\boldsymbol{x}^*) + \cdots + \mu_p^* g_p(\boldsymbol{x}^*) = 0$$

  implies
  - if $g_j(\boldsymbol{x}^*) < 0$ (inactive), then $\mu_j^* = 0$;
  - if $\mu_j^* > 0$, the $g_j(\boldsymbol{x}^*) = 0$ (active);
  - possible: $\mu_j^* = 0$ and $g_j(\boldsymbol{x}^*) = 0$;
  - impossible: $\mu_j^* > 0$ and $g_j(\boldsymbol{x}^*) < 0$.

Illustration



Second-order conditions

- let $\boldsymbol{F}, \boldsymbol{H}_i, \boldsymbol{G}_j$ be the **Hessians** of f, $\boldsymbol{h}$, and $\boldsymbol{g}$
- Lagrangian: $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\lambda} + \boldsymbol{g}(\boldsymbol{x})^T \boldsymbol{\mu}$
- Hessian of $\mathcal{L}$: $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{F}(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \boldsymbol{H}_i(\boldsymbol{x}) + \sum_{j=1}^{p} \mu_j \boldsymbol{G}_j(\boldsymbol{x})$
- Tangent space, defined by all of the active constraints:

$$T(\boldsymbol{x}^*) := \{\boldsymbol{y} \in \mathbb{R}^n : D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0},\ \nabla g_j(\boldsymbol{x}^*)^T \boldsymbol{y} = 0\ \forall j \in J(\boldsymbol{x}^*)\}$$

Theorem (2nd order necessary conditions)

Let $\boldsymbol{x}^*$ be a regular local minimizer for

$$\text{minimize}\, f(\boldsymbol{x}) \qquad \text{subject to } \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0},\ \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0}.$$

Then, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^n$ and $\boldsymbol{\mu}^* \in \mathbb{R}^p$ such that

1. $\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{\lambda}^* + D\boldsymbol{g}(\boldsymbol{x}^*)^T \boldsymbol{\mu}^* = \boldsymbol{0}$, $\boldsymbol{\mu}^* \geq 0$, *and* $\boldsymbol{g}(\boldsymbol{x}^*)^T \boldsymbol{\mu}^* = 0$;
2. *for all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, we have $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \boldsymbol{y} \geq 0$.*

- set of **strongly active inequality constraints**:

$$\tilde{J}(\boldsymbol{x}, \boldsymbol{\mu}) := \{j : g_j(\boldsymbol{x}) = 0, \mu_j > 0\} \subseteq J(\boldsymbol{x})$$

- corresponding tangent space

$$\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) := \{\boldsymbol{y} \in \mathbb{R}^n : D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0},\ \nabla g_j(\boldsymbol{x}^*)^T \boldsymbol{y} = 0\ \forall j \in \tilde{J}(\boldsymbol{x}^*, \mu^*)\}$$

Theorem (2nd-order sufficient conditions)

Suppose that $f, \boldsymbol{h}, \boldsymbol{g} \in C^2$ *and there exists feasible* $\boldsymbol{x}^* \in \mathbb{R}^n$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, *and* $\boldsymbol{\mu}^* \in \mathbb{R}^p$ *such that*

1. $\nabla f(\boldsymbol{x}^*) + D\boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{\lambda}^* + D\boldsymbol{g}(\boldsymbol{x}^*)^T \boldsymbol{\mu}^* = \boldsymbol{0}$, $\boldsymbol{\mu}^* \geq 0$, *and* $\boldsymbol{g}(\boldsymbol{x}^*)^T \boldsymbol{\mu}^* = 0$;
2. *for all* $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}^*, \mu^*), \boldsymbol{y} \neq \boldsymbol{0}$, *we have* $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{y} > 0$.

Then, $\boldsymbol{x}^*$ *is a strict local minimizer of* $\{f(\boldsymbol{x}) : \boldsymbol{h}(\boldsymbol{x}) = 0, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$.

Example 21.6

- consider

$$\begin{aligned} \text{minimize } f(\boldsymbol{x}) &= (x_1 - 1)^2 + x_2 - 2 \\ \text{subject to } \boldsymbol{h}(\boldsymbol{x}) &= x_2 - x_1 - 1 = 0 \\ \boldsymbol{g}(\boldsymbol{x}) &= x_1 + x_2 - 2 \leq 0. \end{aligned}$$

- for all $\boldsymbol{x} \in \mathbb{R}^2$, we have $\nabla h(\boldsymbol{x}) = [-1, 1]^T$ and $\nabla g(\boldsymbol{x}) = [1, 1]^T$
- list the KKT conditions
  1. $\nabla f(\boldsymbol{x}) + D\boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\lambda} + D\boldsymbol{g}(\boldsymbol{x})^T \boldsymbol{\mu} = [2x_1 - 2 - \lambda + \mu, 1 + \lambda + \mu]^T = 0$
  2. $\mu(x_1 + x_2 - 2) = 0$
  3. $\mu \geq 0$
  4. $x_2 - x_1 - 1 = 0$
  5. $x_1 + x_2 - 2 \leq 0$

- first try: $\mu > 0$, implies $x_1 + x_2 - 2 = 0$ and reduces the KKT conditions to

$$\begin{aligned} 2x_1 - 2 - \lambda + \mu &= 0 \\ 1 + \lambda + \mu &= 0, \\ x_2 - x_1 - 1 &= 0, \\ x_1 + x_2 - 2 &= 0, \end{aligned}$$

  which has the solution $[x_1, x_2, \lambda, \mu] = [\frac{1}{2}, \frac{3}{2}, -1, 0]$. but contradicts $\mu > 0$.

- next try: $\mu = 0$, reduces the KKT conditions to

$$\begin{aligned} 2x_1 - 2 - \lambda &= 0 \\ 1 + \lambda &= 0, \\ x_1 + x_2 - 2 &= 0, \end{aligned}$$

  which has the solution $[x_1^*, x_2^*, \lambda^*, \mu^*] = [\frac{1}{2}, \frac{3}{2}, -1, 0]$.

- check the 2nd-order sufficient conditions on $[x_1^*, x_2^*, \lambda^*, \mu^*] = [\frac{1}{2}, \frac{3}{2}, -1, 0]$.

$$\boldsymbol{L}(\boldsymbol{x}^*, \lambda^*, \mu^*) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} + (-1) \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + (0) \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

find the subspace

$$\tilde{T}(\boldsymbol{x}^*, \mu^*) = \{\boldsymbol{y} \in \mathbb{R}^2 : [-1, 1]\boldsymbol{y} = 0\} = \{[a, a]^T : a \in \mathbb{R}\}.$$

(since $\mu^* = 0$, g does not enter the $\tilde{T}$.) pick $\boldsymbol{y} \neq 0$ from $\tilde{T}(\boldsymbol{x}^*, \mu^*)$, then

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \lambda^*, \mu^*)\boldsymbol{y} = 2a^2 > 0.$$

Therefore, $\boldsymbol{x}^* = [x_1^*, x_2^*]^T = [\frac{1}{2}, \frac{3}{2}]^T$ is a strict local minimizer.

Summary

- active and inactive inequality constraints have different roles in NLP
- active constraints participate in deciding $\boldsymbol{x}$
- the KKT conditions equalize ∇f with a linear combination of all ∇h_i and active ∇g_j
- the coefficients of active ∇g_j is signed since $g_j(\boldsymbol{x}) \leq 0$ is sided
- like before, 1st-order conditions give candidates, 2nd-order conditions help determine extremal status