Continuous Optimization

Daniel Mao

Contents

Chapter 1

Unconstrained Optimization

Contents

1.1 Trust-Region Methods

1.1.1 The Dogleg Method

Chapter 2

Constrained Optimization

Contents

Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p$. Consider the following optimization problem

$$\begin{array}{lrl} \text{(P)} & \inf & f(x) \\ & \text{subject to:} & g(x) \leq \mathbb{0} \in \mathbb{R}^m, \\ & & h(x) = \mathbb{0} \in \mathbb{R}^p, \\ & & x \in \mathbb{R}^n. \end{array}$$

Let $\Omega \subseteq \mathbb{R}^n$ denote the feasible region of the above optimization problem.

2.1 Definitions

DEFINITION 2.1 (Local Minimizer)**.** ...

DEFINITION 2.2 (Active Set)**.** Let $x \in \Omega$. We define the **active set** at x, denoted by $\mathcal{A}(x)$, to be a subset of $\{1, ..., m\}$ given by

$$\mathcal{A}(x) := \{i \in \{1, ..., m\} : g_i(x) = 0\} = J.$$

We say that the inequality constraint $g_i(x) \leq 0$ is **active** if and only if $g_i(x) = 0$; and say that it is **inactive** if and only if $g_i(x) < 0$.

2.2 Constraint Qualifications

DEFINITION 2.3 (Tangent Vector). Let $d \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$. We say that d is **tangent** to Ω at x if and only if $\exists (z_i)_{i\in\mathbb{Z}_{++}} \subseteq \Omega$, $\lim_{i\in\mathbb{Z}_{++}} z_i = x$, and $\exists (t_i)_{i\in\mathbb{Z}_{++}} \subseteq \mathbb{R}_{++}$, $\lim_{i\in\mathbb{Z}_{++}} t_i = 0$, and $\lim_{i\in\mathbb{Z}_{++}} \frac{z_i - x}{t_i} = d$.

DEFINITION 2.4 (Tangent Cone). Let $x \in \mathbb{R}^n$. We define the **tangent cone** to Ω at x, denoted by $T_\Omega(x)$, to be the set of all tangent vectors to Ω at x.

DEFINITION 2.5 (Linearized Feasible Directions). Let $\bar{x} \in \Omega$. We define the set of **linearized feasible directions** at $\bar{x}$, denoted by $\mathcal{F}(\bar{x})$, to be a set given by

$$\mathcal{F}(\bar{x}) := \left\{ d \in \mathbb{R}^n : g'_J(\bar{x})d \le \mathbb{0} \in \mathbb{R}^J \text{ and } h'(\bar{x})d = \mathbb{0} \in \mathbb{R}^p \right\}.$$

DEFINITION 2.6 (Linear Independence Constraint Qualification). Let $x \in \mathbb{R}^n$. We say that **linear independence constraint qualification (LICQ)** holds at x if and only if $[\nabla g_J(x) | \nabla h(x)]$ has linearly independent columns

PROPOSITION 2.7. Let $x \in \Omega$. Then $T_\Omega(x) \subseteq \mathcal{F}(x)$. Moreover, if the LICQ condition holds at x, then $T_\Omega(x) = \mathcal{F}(x)$.

2.3 First-Order Optimality Conditions

In this section, we assume that $f, g, h \in \mathcal{C}^1$.

DEFINITION 2.8 (Lagrangain Function)**.** We define the **Lagrangian function** $\mathcal{L} : \mathbb{R}^n \oplus \mathbb{R}^m \oplus \mathbb{R}^p \to \mathbb{R}$ by

$$\mathcal{L}(x, \lambda, \mu) := f(x) + \lambda^\top g(x) + \mu^\top h(x).$$

DEFINITION 2.9 (Complementary Slackness)**.** Let $(x, \lambda, \mu) \in \mathbb{R}^n \oplus \mathbb{R}^m \oplus \mathbb{R}^p$. We say that **complementary slackness** holds for (x, λ, μ) if and only if

$$\lambda^\top g(x) + \mu^\top h(x) = 0.$$

DEFINITION 2.10 (KKT Triple)**.** Let $(x, \lambda, \mu) \in \mathbb{R}^n \oplus \mathbb{R}^m \oplus \mathbb{R}^p$. We say that (x, λ, μ) is a **KKT triple** for (P) if and only if it satisfies all of the following conditions

- primal feasibility:

$$\begin{cases} \nabla_\lambda \mathcal{L}(x, \lambda, \mu) = g(x) \leq \mathbb{0} \in \mathbb{R}^m; \\ \nabla_\mu \mathcal{L}(x, \lambda, \mu) = h(x) = \mathbb{0} \in \mathbb{R}^p; \end{cases}$$

- dual feasibility:

$$\begin{cases} \nabla_x \mathcal{L}(x, \lambda, \mu) = \nabla f(x) + \nabla g(x)\lambda + \nabla h(x)\mu = \mathbb{0} \in \mathbb{R}^n; \\ \lambda \geq \mathbb{0} \in \mathbb{R}^m; \end{cases}$$

- complementary slackness:

$$\lambda^\top \nabla_\lambda \mathcal{L}(x, \lambda, \mu) = 0.$$

DEFINITION 2.11 (Strict Complementary Slackness)**.** Let $(x, \lambda, \mu) \in \mathbb{R}^n \oplus \mathbb{R}^m \oplus \mathbb{R}^p$ be a KKT triple for (P). We say that **strict complementary slackness** holds for (x, λ, μ) if and only if $\forall i \in \{1, ..., m\}$, exactly one of $\lambda_i = 0$ and $g_i(x) = 0$ holds.

THEOREM 2.12 (First-Order Necessary Conditions)**.** Let x^* be a local minimizer of (P). Suppose that the LICQ holds at x^*. Then $\exists \lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ such that (x^*, λ^*, μ^*) is a KKT triple for (P).

2.4 Second-Order Optimality Conditions

In this section, we assume that $f, g, h \in \mathbb{C}^2$.

DEFINITION 2.13 (Critical Cone)**.** Let $(x, \lambda, \mu) \in \mathbb{R}^n \oplus \mathbb{R}^m \oplus \mathbb{R}^p$ be a KKT triple for (P). We define the **critical cone**, denoted by $\mathcal{C}(x, \lambda, \mu)$, to be a subset of $\mathcal{F}(x)$ given by

$$\begin{aligned}\mathcal{C}(x, \lambda, \mu) &:= \left\{ w \in \mathcal{F}(x) : \forall i \in J : \lambda_i > 0, \nabla g_i(x)^\top w = 0 \right\} \\ &= \left\{ w \in \mathbb{R}^n : \begin{array}{l} h'(x)w = \mathbb{0} \in \mathbb{R}^p \\ \nabla g_i(x)^\top w = 0, \text{ if } i \in J \text{ and } \lambda_i > 0, \\ \nabla g_i(x)^\top w \leq 0, \text{ otherwise.} \end{array} \right\}.\end{aligned}$$

REMARK 2.14. If (x, λ, μ) is a KKT triple for (P) and $w \in \mathcal{C}(x, \lambda, \mu)$, then

$$-w^\top \nabla f(x) = w^\top \Big[\nabla g(x)\lambda + \nabla h(x)\mu \Big] = 0.$$

THEOREM 2.15 (Second-Order Necessary Conditions)**.** Let x^* be a local minimizer of (P). Suppose that the LICQ holds at x^*. Let $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ be such that (x^*, λ^*, μ^*) is a KKT triple for (P). Then

$$\forall w \in \mathcal{C}(x^*, \lambda^*, \mu^*), \quad w^\top \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) w \geq 0.$$

Proof. Since the LICQ holds at x^*, $T_\Omega(x^*) = \mathcal{F}(x^*)$. So $\exists (z_k)_{k \in \mathbb{Z}_{++}} \subseteq \Omega$, $\lim\limits_{k \in \mathbb{Z}_{++}} z_k = x^*$, and $\exists (t_k)_{k \in \mathbb{Z}_{++}} \subseteq \mathbb{R}_{++}$, $\lim\limits_{k \in \mathbb{Z}_{++}} t_k = 0$, and $\lim\limits_{k \in \mathbb{Z}_{++}} \dfrac{z_k - x^*}{t_k} = w$. So $z_k - x^* = t_k w + o(t_k)$. ... So

$$g_i(z_k) = t_k \nabla g_i(x^*)^\top w, \quad \forall i \in J.$$

So

$$\mathcal{L}(z_k, \lambda^*, \mu^*) = f(z_k) + \lambda^\top g(z_k) + \mu^\top h(z_k).$$

So

$$f(z_k) = f(x^*) + \frac{t_k^2}{2} w^\top \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w + o(t_k^2).$$

Since x^* is a local minimizer, we must have $f(z_k) \geq f(x^*)$ for sufficiently large k. So $w^\top \nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)w \geq 0$. $\square$

THEOREM 2.16 (Second-Order Sufficient Conditions)**.** Let $\bar{x} \in S$. Suppose that $\exists \bar{\lambda} \in \mathbb{R}^m$ and $\bar{\mu} \in \mathbb{R}^p$ such that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple. Suppose also that

$$w^\top \nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\mu})w > 0, \quad \forall w \in \mathcal{C}(\bar{x}, \bar{\lambda}) \setminus \{\mathbb{0}\}.$$

Then $\bar{x}$ is a strict local minimizer for (P).

2.5 Augmented Lagrangian

Chapter 3

Unconstrained Optimization

Contents

Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^m$. Consider the following optimization problem:

$$
\begin{array}{lll}
\text{(P)} & \inf\limits_{x} & f(x) \\
 & \text{subject to:} & x \in \mathbb{R}^n, \\
 & & g(x) \geq \mathbb{0}_m, \text{ and} \\
 & & h(x) = \mathbb{0}_p.
\end{array}
$$

3.1 Inequality-Constrained Quadratic Programming (IQP) Approach

At each iteration $k \in \mathbb{Z}_{++}$, we solve the following quadratic subproblem

$$
\begin{aligned}
\text{(Q)} \qquad & \inf_{p} && f(x^{(k)}) + [\nabla f(x^{(k)})]^\top p + \frac{1}{2} p^\top [\nabla^2_{xx} \mathcal{L}(x^{(k)}, \lambda^{(k)}, \mu^{(k)})] p \\
& \text{subject to:} && p \in \mathbb{R}^n, \\
& && [\nabla g(x_k)]^\top p + g(x_k) \geq \mathbb{0}_m, \text{ and} \\
& && [\nabla h(x_k)]^\top p + h(x_k) = \mathbb{0}_p.
\end{aligned}
$$

Algorithm 1: The IQP Algorithm

Input: Initial $(x^{(0)}, \lambda^{(0)}, \mu^{(0)}) \in \mathbb{R}^n \oplus \mathbb{R}^m \oplus \mathbb{R}^p$.

while *stopping criterion is not satisfied* **do**

  Solve the above subproblem to obtain $p^{(k)} \in \mathbb{R}^n$, $\lambda^{(k+1)} \in \mathbb{R}^m$, and $\mu^{(k+1)} \in \mathbb{R}^p$;

  Set $x^{(k+1)} := x^{(k)} + p^{(k)}$;