Chapter 10: Theory of Constrained Optimization

Outline

1. Introduction
2. Constraint Qualifications
3. First-Order Optimality Conditions
4. Second-Order Optimality Conditions
5. Duality
6. References

Outline

1. Introduction
2. Constraint Qualifications
3. First-Order Optimality Conditions
4. Second-Order Optimality Conditions
5. Duality
6. References

A General Formulation for Constrained Optimization

- Constrained optimization:

$$\min_{x \in \Re^n} \quad f(x) \tag{1.1a}$$

$$s.\ t. \quad c_i(x) = 0, \quad i \in \mathcal{E}, \tag{1.1b}$$

$$c_i(x) \geq 0, \quad i \in \mathcal{I}. \tag{1.1c}$$

  where f and the functions c_i are all smooth, real-valued functions on a subset of $\Re^n$, and $\mathcal{I}$ and $\mathcal{E}$ are two finite sets of indices. As before, we call f the *objective function*, while c_i, $i \in \mathcal{E}$ are the *equality constraints* and c_i, $i \in \mathcal{I}$ are the *inequality constraints*.

- We define the *feasible set* Ω to be the set of points x that satisfy the constraints; that is,

$$\Omega = \{x | c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\}, \tag{1.2}$$

  so that we can rewrite (1.1) more compactly as

$$\min_{x \in \Omega} f(x). \tag{1.3}$$

Local and Global Solutions

- A vector x^* is *a local solution* of the problem (1.1) if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that

$$f(x^*) \leq f(x) \text{ for } x \in \mathcal{N} \cap \Omega$$

- A vector x^* is *a strict local solution* of the problem (1.1) if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that

$$f(x^*) < f(x) \text{ for } x \in \mathcal{N} \cap \Omega.$$

- A vector x^* is an *isolated local solution* of the problem (1.1) if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that x^* is the only local solution in $\mathcal{N} \cap \Omega$.

Smoothness

Smoothness ensures that the objective function and the constraints all behave in a reasonably predictable way and therefore allows algorithms to make good choices for search directions.

- The nonsmooth boundaries can often be described by a collection of smooth constraint functions.

$$
\begin{aligned}
& \|x\|_1 = |x_1| + |x_2| \leq 1, \\
\Leftrightarrow \quad & x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1.
\end{aligned}
$$

- Nonsmooth, unconstrained optimization problems can sometimes be reformulated as smooth constrained problems.

$$
\begin{aligned}
& \min \quad f(x) = \max(x^2, x), \\
\Leftrightarrow \quad & \min \quad t \quad \text{s.t. } t \geq x, t \geq x^2.
\end{aligned}
$$

Definition - Active Set

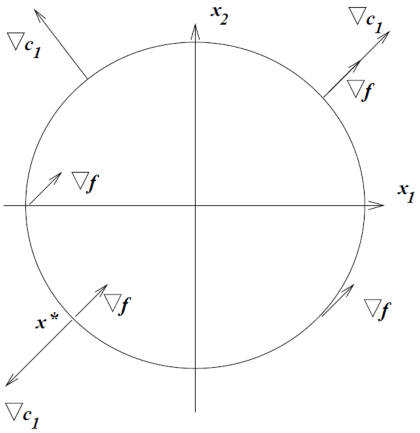
The active set $\mathcal{A}(x)$ at any feasible point x consists of the equality constraint indices from $\mathcal{E}$ together with the indices of the inequality constraints i for which $c_i(x) = 0$; that is,

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

At a feasible point x, the inequality constraint $i \in \mathcal{I}$ is said to be *active* if $c_i(x) = 0$ and *inactive* if the strict inequality $c_i(x) > 0$ is satisfied.

Example - A Single Equality Constraint

$$\min\ x_1 + x_2 \quad \text{s.t. } x_1^2 + x_2^2 - 2 = 0.$$



Example - A Single Equality Constraint

- To retain feasibility with respect to $c_1(x) = 0$, we require any small (but nonzero) step d to satisfy that $c_1(x + d) = 0$; that is,

$$0 = c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d = \nabla c_1(x)^T d$$

- If we want d to produce a decrease in f, we would have so that

$$0 > f(x + d) - f(x) \approx \nabla f(x)^T d.$$

Example - A Single Equality Constraint

- Suppose x^* is a local minimizer. The only way that a d satisfying

$$\nabla c_1(x^*)^T d = 0 \text{ and } \nabla f(x^*)^T d < 0$$

does *not* exist is if $\nabla f(x^*)$ and $\nabla c_1(x^*)$ are parallel, that is, if the condition

$$\nabla f(x^*) = \lambda_1 \nabla c_1(x^*) \tag{1.4}$$

holds at x^*, for some scalar λ_1^*.

- By introducing the *Lagrangian function*

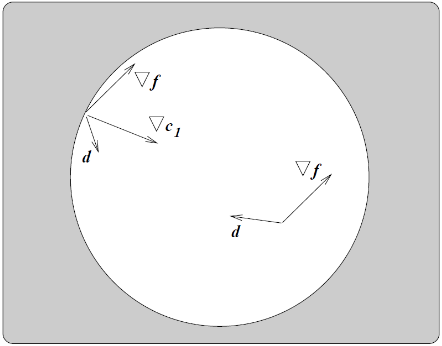
$$\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x),$$

we can state the condition (1.4) equivalently as that $\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0$, here λ_1^* is called a *Lagrange multiplier* for the constraint $c_1(x^*) = 0$.

- This observation suggests that we can search for solutions of the equality constrained problem by searching for stationary points of the Lagrangian function.

Example - A Single Inequality Constraint

$$\min x_1 + x_2 \text{ s.t. } 2 - x_1^2 - x_2^2 \geq 0.$$



Example - A Single Inequality Constraint

- As before, we conjecture that a given feasible point x is not optimal if we can find a small steps that both retains feasibility and decreases the objective function f to first-order.
- The step d improves the objective function, to first order, if

$$\nabla f(x)^T d < 0.$$

- Meanwhile, d retains feasibility if

$$0 \leq c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d,$$

so, to first order, feasibility is retained if

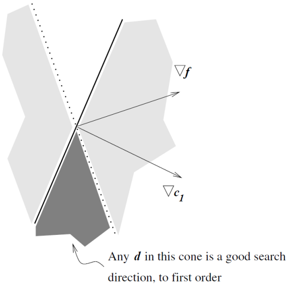
$$c_1(x) + \nabla c_1(x)^T d \geq 0.$$

Example - A Single Inequality Constraint

Suppose that there exists no d such that $c_1(x) + \nabla c_1(x)^T d \geq 0$ and $\nabla f(x)^T d < 0$.

- Case I: x lies strictly inside the circle, that is $c(x) > 0$, so we must have
$$\nabla f(x) = 0;$$
- Case II: x lies on the boundary of the circle, that is $c_1(x) = 0$, so if there exists no d such that $\nabla c_1(x)^T d \geq 0$ and $\nabla f(x)^T d < 0$ then we must have
$$\nabla f(x) = \lambda_1 \nabla c_1(x), \text{ for some } \lambda_1 \geq 0.$$

Note that the sign of the multiplier is significant here:



Example - A Single Inequality Constraint

- The optimality conditions for both cases I and II can be summarized neatly with reference to the Lagrangian function $\mathcal{L}$.
- Suppose that x^* is a local minimizer. When no first-order feasible descent direction exists at some point x^*, we have that

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0, \text{ for some } \lambda_1^* \geq 0.$$

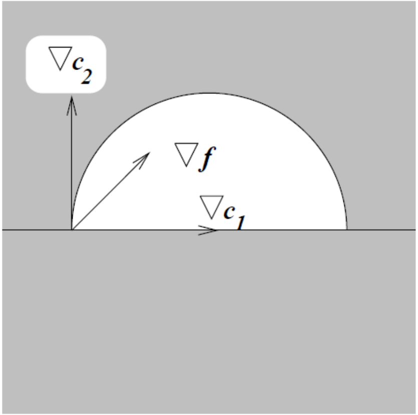
- We also require that

$$\lambda_1^* c_1(x^*) = 0.$$

  This condition is known as a *complementarity condition*; it implies that the Lagrange multiplier λ_1 can be strictly positive *only when the corresponding constraint* c_1 *is active*.

Example - Two Inequality Constraints

$$\min x_1 + x_2 \text{ s.t. } 2 - x_1^2 - x_2^2 \geq 0, x_2 \geq 0.$$



The gradients of the active constraints and objective at the solution $(-\sqrt{2}, 0)^T$.

Example - Two Inequality Constraints

At the point $(-\sqrt{2}, 0)^T$, both constraints are active. By repeating the arguments for the previous examples, we would expect a direction d of first-order feasible descent to satisfy

$$\nabla c_i(x)^T d \geq 0, \quad i \in \mathcal{I} = \{1, 2\}, \quad \nabla f(x)^T d < 0.$$

However, from the figure there is no such direction at $(-\sqrt{2}, 0)^T$. In this case, the Lagrangian function is defined as

$$\mathcal{L}(x, \lambda) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x).$$

It is easy to check that at $x^* = (-\sqrt{2}, 0)^T$,

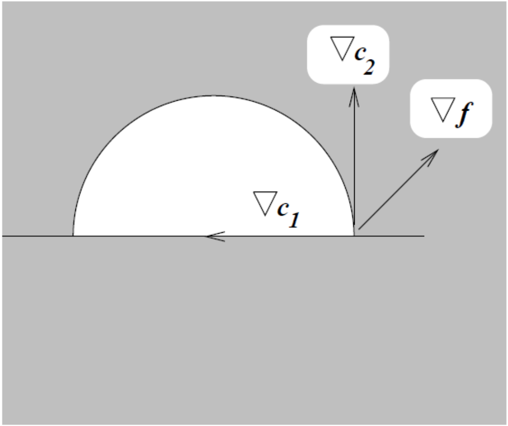
$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \text{ for } \lambda_1^* = 1/(2\sqrt{2}), \lambda_2^* = 1.$$

and

$$\lambda_1^* c_1(x^*) = 0, \quad \lambda_2^* c_2(x^*) = 0.$$

Example - Two Inequality Constraints

$$\min x_1 + x_2 \text{ s.t. } 2 - x_1^2 - x_2^2 \geq 0, x_2 \geq 0.$$



The gradients of the active constraints and objective at a nonoptimal solution.

Outline

1. Introduction
2. Constraint Qualifications
3. First-Order Optimality Conditions
4. Second-Order Optimality Conditions
5. Duality
6. References

Motivation

- Previously, we determined whether or not it was possible to take a feasible descent step away from a given feasible point x by examining the first derivatives of f and the constraint functions c_i.
- We used the first-order Taylor series expansion of these functions about x to form an approximate problem in which both objective and constraints are linear.
- This approach makes sense, however, only when the linearized approximation captures the essential geometric features of the feasible set near the point x in question.

Motivation

- If, near x, the linearization is fundamentally different from the feasible set, then we cannot expect the linear approximation to yield useful information about the original problem. For example, the equality constraint is

$$x^2 = 0.$$

  At the point $x = 0$, the linearization corresponds to the entire space, while the feasible set is a single point.

- Hence, we need to make assumptions about the nature of the constraints c_i that are active at x to ensure that the linearized approximation is similar to the feasible set, near x.

- Constraint qualifications are assumptions that ensure similarity of the constraint set Ω and its linearized approximation, in a neighborhood of x.

Definition: Linearized Feasible Directions

- Given a feasible points x and the active constraint set $\mathcal{A}$, *the set of linearized feasible directions* $\mathcal{F}(x)$ is

$$\mathcal{F}(x) = \left\{ d \mid \begin{array}{ll} d^T \nabla c_i(x) = 0, & \forall i \in \mathcal{E}, \\ d^T \nabla c_i(x) \geq 0, & \forall i \in \mathcal{A}(x) \cap \mathcal{I} \end{array} \right\}.$$

- The linearized feasible direction set depends on the definition of the constraint functions c_i, $i \in \mathcal{E} \cup \mathcal{I}$.

Tangent Cone

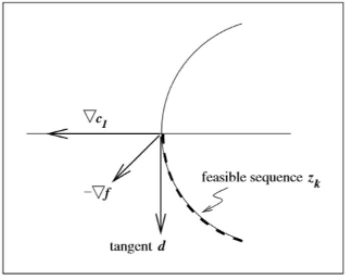
- Given a feasible point x, we call $\{z_k\}$ a *feasible sequence approaching x* if $z_k \in \Omega$ for all k sufficiently large and $z_k \to x$.
- A vector d is called *tangent* to Ω at a point x, if it is a limiting direction of a feasible sequence. Namely, there exists a feasible sequence $\{z_k\}$ approaching x and a sequence of positive scalars $\{t_k\}$ with $t_k \to 0$ such that

$$\lim_{k\to\infty} \frac{z_k - x}{t_k} = d.$$

- The set of all tangents to Ω at x^* is called the *tangent cone* and is denoted by $T_\Omega(x^*)$.
- The definition of tangent cone does not rely on the algebraic specification of the set Ω, only on its geometry.

Illustrations

$$\min\ x_1 + x_2 \quad \text{s.t. } x_1^2 + x_2^2 - 2 = 0.$$



- $x = (-\sqrt{2}, 0)^T$,
  - $z_k = (-\sqrt{2 - 1/k^2}, -1/k)^T,\ t_k = \|z_k - x\| \Rightarrow d = (0, -1)^T$.
  - $z_k = (-\sqrt{2 - 1/k^2}, 1/k)^T,\ t_k = \|z_k - x\| \Rightarrow d = (0, 1)^T$.
- $T_\Omega(x) = \mathcal{F}(x) = \{(0, d_2)^T | d_2 \in \mathbb{R}\}$.

Illustrations

Suppose that the feasible set is defined instead by the formula

$$\Omega = \{x \mid (x_1^2 + x_2^2 - 2)^2 = 0\}.$$

Note that Ω is the same, but its algebraic specification has changed. Still at $x = (-\sqrt{2}, 0)^T$, the vector d belongs to the linearized feasible set if

$$0 = \nabla c_1(x)^T d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}.$$

So $T_\Omega(x) \neq \mathcal{F}(x) = \mathbb{R}^2$.

Constraint Qualifications

- Constraints qualifications are conditions under which the linearized feasible set $\mathcal{F}(x)$ is similar to the tangent cone $T_\Omega(x)$.
- In fact, most constraint qualifications ensure that these two sets are identical.
- These conditions ensure that the $\mathcal{F}(x)$, which is constructed by linearizing the algebraic description of the set Ω at x, captures the essential geometric features of the set Ω in the vicinity if x, as represented by $T_\Omega(x)$.

linear independence constraint qualification

- LICQ: Given the point x and the active set $\mathcal{A}(x)$, we say that the linear independence constraint qualification (LICQ) holds if

  the set of active constraint gradients $\{\nabla c_i(x), i \in \mathcal{A}(x)\}$ is linearly independent.

- In general, if LICQ holds, none of the active constraint gradients can be zero.

Outline

1. Introduction
2. Constraint Qualifications
3. First-Order Optimality Conditions
4. Second-Order Optimality Conditions
5. Duality
6. References

First-Order Necessary Conditions

- Define the Lagrangian function for the general problem (1.1):

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

- Suppose that x^* is a local solution of (1.1) and that the LICQ holds at x^*. Then there is a Lagrange multiplier vector λ^*, with components λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at (x^*, λ^*)

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, && && \text{(3.1a)}\\
c_i(x^*) &= 0, && \forall i \in \mathcal{E}, && \text{(3.1b)}\\
c_i(x^*) &\geq 0, && \forall i \in \mathcal{I}, && \text{(3.1c)}\\
\lambda_i^* &\geq 0, && \forall i \in \mathcal{I}, && \text{(3.1d)}\\
\lambda_i^* c_i(x^*) &= 0, && \forall i \in \mathcal{E} \cup \mathcal{I}. && \text{(3.1e)}
\end{aligned}$$

- These conditions are often known as the *Karush-Kuhn-Tucker* conditions, or KKT conditions for short.
- Ex: write the KKT conditions for

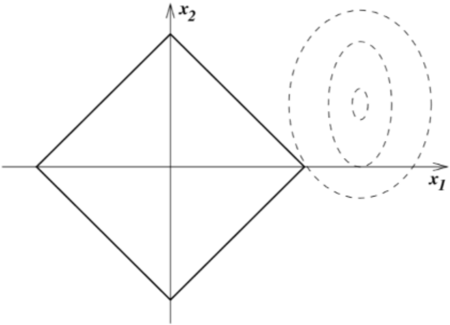
$$\min \frac{1}{2} x^T G x + x^T c \quad s.t. \quad Ax = b,$$

Strict Complementarity

- Given a local solution x^* of (1.1) and a vector λ^* satisfying (3.1), we say that the strict complementarity condition holds if exactly one of λ_i^* and $c_i(x^*)$ is zero for each index $i \in \mathcal{I}$. In other words, we have that $\lambda_i^* > 0$ for each $i \in \mathcal{I} \cap \mathcal{A}(x^*)$.
- Satisfaction of the strict complementary property usually makes it easier for algorithms to determine the active set $\mathcal{A}(x^*)$ and converge rapidly to the solution x^*.
- For a given problem (1.1) and solution point x^*, there may be many vectors λ^* for which the conditions (3.1) are satisfied. When the LICQ holds, however, the optimal λ^* is unique.

An Example

$$\min_x (x_1 - 1.5)^2 + (x_2 - 0.5)^4, \quad s.t. \quad \begin{bmatrix} 1 - x_1 - x_2 \\ 1 - x_1 + x_2 \\ 1 + x_1 - x_2 \\ 1 + x_1 + x_2 \end{bmatrix} \geq 0.$$



An Example

The solution is $x^* = (1,0)^T$. At x^*, the first two constraints are active, while the last two are inactive. So

$$\lambda_3^* = \lambda_4^* = 0.$$

Since

$$\nabla f(x^*) = \begin{bmatrix} -1 \\ -0.5 \end{bmatrix}, \quad \nabla c_1(x^*) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \nabla c_2(x^*) = \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

we obtain that $\lambda_1^* = 0.75, \lambda_2^* = 0.25$.

First-Order Optimality Conditions: Proof

Step I : Relating the Tangent Cone and the First-Order Feasible Direction Set

Let x^* be a feasible point. The following two statements are true.

(i) $T_\Omega(x^*) \subset \mathcal{F}(x^*)$.

(ii) If the LICQ condition is satisfied at x^*, then $\mathcal{F}(x^*) = T_\Omega(x^*)$.

First-Order Optimality Conditions: Proof

Step II: A Fundamental Necessary Condition

If x^* is a local solution of (1.1), then we have

$$\nabla f(x^*)^T d \geq 0, \qquad \forall d \in T_\Omega(x^*). \tag{3.3}$$

The reverse of this result is not necessary true,. That is, we may have $\nabla f(x^*)^T d \geq 0$ for all $d \in T_\Omega(x^*)$, yet x^* is not a local minimizer. An example is the following problem

$$\min x_2 \quad \text{s.t. } x_2 \geq -x_1^2.$$

Consider the point $x = (0, 0)^T$.

First-Order Optimality Conditions: Proof

Step III: Farkas' Lemma

Let the cone K be defined as

$$K = \{By + Cw | y \geq 0\},$$

where B and C are matrices of dimension $n \times m$ and $n \times p$, respectively, and y and w are vector of appropriate dimensions. Given any vector $g \in \Re^n$, we have either that $g \in K$ or that there exists $d \in \Re^n$ satisfying

$$g^T d < 0, \qquad B^T d \geq 0, \qquad C^T d = 0,$$

but not both.

Farkas' Lemma

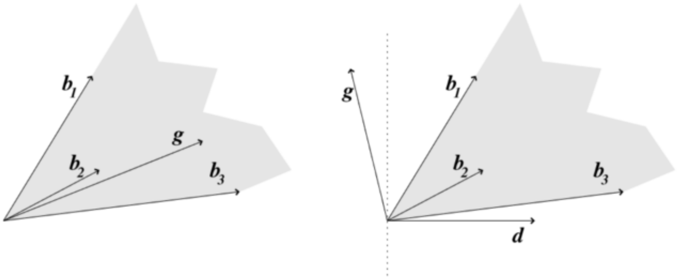


Figure 1: Farkas' Lemma: Either $g \in K$ (left) or there is a separating hyperplane (right). Here $B = [b_1, b_2, b_3]$, $C = 0$.

Farkas' Lemma

Let

$$g = \nabla f(x), \quad B = [\nabla c_i(x)]_{i \in \mathcal{A}(x) \cap \mathcal{I}}, \quad C = [\nabla c_i(x)]_{i \in \mathcal{E}}.$$

Farkas' lemma shows that either $g \in K$, namely, there exist $\lambda_i \geq 0, i \in \mathcal{A}(x) \cap \mathcal{I}$ such that

$$\nabla f(x) = \sum_{i \in \mathcal{A}(x) \cap \mathcal{I}} \lambda_i \nabla c_i(x) + \sum_{i \in \mathcal{E}} \lambda_i \nabla c_i(x),$$

or there exists d such that

$$\nabla f(x)^T d < 0, \quad \nabla c_i(x)^T d \geq 0, i \in \mathcal{A}(x) \cap \mathcal{I}, \quad \nabla c_i(x)^T d = 0, i \in \mathcal{E},$$

which means $\nabla f(x)^T d < 0$ and $d \in \mathcal{F}(x)$.

Proof Sketch of First-Order Optimality conditions

Suppose x^* is a local solution of (1.1). Obviously it satisfies the feasibility conditions

$$c_i(x^*) = 0, \quad i \in \mathcal{E}; \quad c_i(x^*) \geq 0, \quad i \in \mathcal{I}.$$

Suppose that LICQ holds at x^*. Then

$$\nabla f(x^*)^T d \geq 0, \quad \forall d \in \mathcal{F}(x^*).$$

According to Farkas' lemma, we have

$$\begin{aligned} \nabla f(x^*) &= \sum_{i \in \mathcal{A}(x^*) \cap \mathcal{I}} \lambda_i \nabla c_i(x^*) + \sum_{i \in \mathcal{E}} \lambda_i \nabla c_i(x^*), \\ \lambda_i^* &\geq 0, \quad i \in \mathcal{A}(x) \cap \mathcal{I}. \end{aligned}$$

We further define $\lambda_i^* = 0$, $i \in \mathcal{I} \setminus \mathcal{A}(x^*)$. Then x^*, λ^* satisfy the KKT conditions.

A Geometric Viewpoint

- An alternative first-order optimality condition that depends only on the geometry of the feasible set Ω and not on its particular algebraic description in terms of the functions c_i, $i \in \mathcal{E} \cup \mathcal{I}$.
- In geometric terms, our problem (1.1) can be stated as

$$\min f(x) \text{ subject to } x \in \Omega \tag{3.5}$$

where Ω is the feasible set defined by the conditions (1.2).

- The *normal cone* to the set Ω at the point $x \in \Omega$ is defined as

$$N_\Omega(x) = \{v | v^T w \leq 0 \qquad \forall w \in T_\Omega(x)\}. \tag{3.6}$$

Theorem 1

Suppose that x^* is a local minimizer of f in Ω. Then

$$-\nabla f(x^*) \in N_\Omega(x^*).$$

Lagrange Multipliers and Sensitivity

- The value of each Lagrange multiplier λ_i tells us something about the sensitivity of the optimal objective value $f(x^*)$ to the presence of constraint c_i.
- Suppose that c_i is active and let us perturb the ith constraint, requiring $c_i(x) \geq -\epsilon\|\nabla c_i(x^*)\|$.
- Suppose that $x^*(\epsilon)$ still has the same set of active constraints and the Lagrange multipliers are not much affected.
- Then the family of solutions $x^*(\epsilon)$ satisfies

$$\frac{df(x^*(\epsilon))}{d\epsilon} = -\lambda_i^* \|\nabla c_i(x^*)\|.$$

- If $\lambda_i^* \|\nabla c_i(x^*)\|$ is large, then the optimal value is sensitive to the placement of the ith constraint, while if this quantity is small, the dependence is not too strong.
- If λ_i^* is exactly zero for some active constraint, small perturbations to c_i in some directions will hardly affect the optimal objective value at all; the change is zero, to first order.

Outline

1 Introduction

2 Constraint Qualifications

3 First-Order Optimality Conditions

4 Second-Order Optimality Conditions

5 Duality

6 References

Second-Order Conditions

- When the first-order conditions are satisfied, a move along any vector w from $\mathcal{F}(x^*)$ either increases the first-order approximation to the objective function (that is, $w^T\nabla f(x^*)$ 0), or else keeps this value the same (that is, $w^T\nabla f(x^*) = 0$).
- For the directions $w \in \mathcal{F}(x^*)$ for which $w^T\nabla f(x^*) = 0$, we cannot determine from first derivative information alone whether a move along this direction will increase or decrease the objective function f. Second-order conditions examine the second derivative terms in the Taylor series expansions of f and c_i, to see whether this extra information resolves the issue of increase or decrease in f.
- Essentially, the second-order conditions concern the curvature of the Lagrangian function in the "undecided" directions - the directions $w \in \mathcal{F}(x^*)$ for which $w^T\nabla f(x^*) = 0$.

Critical Cone

- Define the critical cone as follows:

$$\mathcal{C}(x^*, \lambda^*) = \{w \in \mathcal{F}(x^*) | \nabla c_i(x^*)^T w = 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0\}.$$

- Equivalently,

$$w \in \mathcal{C}(x^*, \lambda^*) \Leftrightarrow \begin{cases} \nabla c_i(x^*)^T w = 0, & \forall i \in \mathcal{E}, \\ \nabla c_i(x^*)^T w = 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0, \\ \nabla c_i(x^*)^T w \geq 0, & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* = 0, \end{cases}$$

- The critical cone $\mathcal{C}(x^*, \lambda^*)$ contains directions from $\mathcal{F}(x^*)$ for which it is not clear from first derivative information alone whether f will increase or decrease.
- It follows that

$$w \in \mathcal{C}(x^*, \lambda^*) \Rightarrow \lambda_i^* \nabla c_i(x^*)^T w = 0 \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}.$$

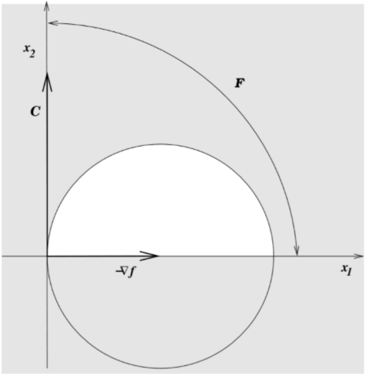
- From KKT condition, we have

$$w \in \mathcal{C}(x^*, \lambda^*) \Rightarrow w^T \nabla f(x^*) = 0.$$

An Example

Consider

$$\min x_1 \quad s.t. \quad x_2 \geq 0, 1 - (x_1 - 1)^2 - x_2^2 \geq 0.$$



The solution is $x^* = (0,0)^T$ and $\lambda^* = (0, 0.5)^T$. We have

$$\mathcal{F}(x^*) = \{d \mid d \geq 0\} \quad C(x^*, \lambda^*) = \{(0, w_2)^T \mid w_2 \geq 0\}.$$

Second-Order Necessary Conditions

Theorem 2

Suppose that x^* is a local solution of (1.1) and that the LICQ condition is satisfied. Let λ^* be a Lagrange multiplier vector such that the KKT conditions (3.1) are satisfied. Then

$$w^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) w \geq 0, \quad \forall w \in \mathcal{C}(x^*, \lambda^*). \tag{4.1}$$

Second-Order Sufficient Conditions

Theorem 3

Suppose that for some feasible point $x^* \in \Re^n$, there is a Lagrange multiplier vector λ^* such that the KKT conditions (3.1) are satisfied. Suppose also that

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w > 0, \quad \forall w \in \mathcal{C}(x^*, \lambda^*), w \neq 0. \tag{4.2}$$

Then x^* is a strict local solution for (1.1).

Second-Order Conditions and Projected Hessians

- When the multiplier λ^* that satisfies the KKT conditions (3.1) is unique (as happens, for example, when the LICQ condition holds) and strict complementarity holds. In this case, the definition of $\mathcal{C}(x^*, \lambda^*)$ reduces to

$$\mathcal{C}(x^*, \lambda^*) = \text{Null}[\nabla c_i(x^*)^T]_{i \in \mathcal{A}(x^*)} = \text{Null}A(x^*).$$

  Here $A(x^*)^T = [\nabla c_i(x^*)]_{i \in \mathcal{A}(x^*)}$.

- Define the matrix Z with full column rank whose columns span the space $\mathcal{C}(x^*, \lambda^*)$, that is

$$\mathcal{C}(x^*, \lambda^*) = \{Zu | u \in \mathbb{R}^{|\mathcal{A}(x^*)|}\}.$$

- The second-order necessary condition (4.1) can be restated as

$$Z^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z \text{ is positive semidefinite.}$$

  Similarly, the second-order sufficient condition (4.2) can be restated as

$$Z^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z \text{ is positive definite.}$$

- Z can be computed numerically, so that the positive (semi)definiteness conditions can actually be checked by forming these matrices and finding their eigenvalues.

Computation of Z

In the case above (in which the multiplier λ is unique and strictly complementary holds), we write the QR factorization of $A(x^*)^T$ as

$$A(x^*)^T = Q \begin{bmatrix} R \\ 0 \end{bmatrix} = [Q_1 \quad Q_2] \begin{bmatrix} R \\ 0 \end{bmatrix} = Q_1 R.$$

If R is nonsingular, we can set $Z = Q_2$. Otherwise, a slight enhancement of this procedure that makes use of column pivoting during the QR procedure can be used to identify Z.

Outline

1. Introduction
2. Constraint Qualifications
3. First-Order Optimality Conditions
4. Second-Order Optimality Conditions
5. Duality
6. References

Duality

- Duality theory is used to motivate and develop some important algorithms, including the augmented Lagrangian algorithms.
- In its full generality, duality theory ranges beyond nonlinear programming to provide important insight into the fields of convex nonsmooth optimization and even discrete optimization.
- Duality theory shows how we can construct an alternative problem from the functions and data that define the original optimization problem.
- In some cases, the dual problem is easier to solve computationally than the original problem. In other cases, the dual can be used to obtain easily a lower bound on the optimal value of the objective for the primal problem.

Primal and Dual Problem

- Consider

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \;\; c(x) \geq 0, \tag{5.1}$$

where, $c(x) \equiv (c_1(x), c_2(x), \cdots, c_m(x))^T$, f and $-c_i$ are all convex function.

- The dual problem to (5.1) is defined as follows:

$$\max_{\lambda\in\mathbb{R}^m} q(\lambda) \quad \text{s.t.} \;\; \lambda \geq 0, \tag{5.2}$$

where $q(\lambda) \equiv \inf_x \mathcal{L}(x, \lambda)$.

- Consider the dual problem of

$$\min_{(x_1,x_2)} 0.5(x_1^2 + x_2^2) \quad \text{s.t.} \;\; x_1 - 1 \geq 0.$$

Dual Problem

- In many problems, this infimum is $-\infty$ for some values of λ. We define the domain of q as the set of λ values for which q is finite, that is,

$$\mathcal{D} \equiv \{\lambda | q(\lambda) > -\infty\}.$$

- Note that calculation of the infimum in requires finding the global minimizer of the function $\mathcal{L}(x, \lambda)$ for the given λ which may be extremely difficult in practice.
- However, when f and $-c_i$ are convex functions and $\lambda \geq 0$, the function $\mathcal{L}(\cdot, \lambda)$ is also convex. In this situation, all local minimizers are global minimizers, so computation of $q(\lambda)$ becomes a more practical proposition.

Weak Duality

Theorem 4

The function q is concave and its domain $\mathcal{D}$ is convex.

Theorem 5

For any $\bar{x}$ feasible for (5.1) and any $\bar{\lambda} \geq 0$, we have $q(\bar{\lambda}) \leq f(\bar{x})$.

KKT conditions

The KKT conditions (3.1) specialized to (5.1) are as follows:

$$\nabla f(\bar{x}) - \nabla c(\bar{x})\bar{\lambda} = 0, \tag{5.3a}$$

$$c(\bar{x}) \geq 0, \tag{5.3b}$$

$$\bar{\lambda} \geq 0, \tag{5.3c}$$

$$\bar{\lambda}_i c_i(\bar{x}) = 0, \qquad i = 1, \cdots, m. \tag{5.3d}$$

where $\nabla c(x) = [\nabla c_1(x), \nabla c_2(x), \cdots, \nabla c_m(x)]$.

Duality

Theorem 6

Suppose that $\bar{x}$ is a solution of (5.1) and that f and $-c_i$, $i = 1, \cdots, m$ are convex functions on $\Re^n$ that are differentiable at x. Then any $\bar{\lambda}$ for which $(\bar{x}, \bar{\lambda})$ satisfies the KKT conditions (5.3) *is a solution of (5.2).*

Theorem 7

Suppose that f and $-c_i$, $i = 1, \cdots, m$ are convex and continuously differentiable on $\Re^n$. Suppose that $\bar{x}$ is a solution of (5.1) at which LICQ holds. Suppose that $\hat{\lambda}$ solves (5.2) and that the infimum in $\inf_x \mathcal{L}(x, \hat{\lambda})$ is attained at $\hat{x}$. Assume further that $\mathcal{L}(\cdot, \hat{\lambda})$ is a strictly convex function. Then $\bar{x} = \hat{x}$ (that is, $\hat{x}$ is the unique solution of (5.1)), and $f(\bar{x}) = \mathcal{L}(\hat{x}, \hat{\lambda})$

Wolfe Dual

$$\max_{x,\lambda} \quad \mathcal{L}(x,\lambda) \tag{5.4a}$$

$$s.t. \quad \nabla_x \mathcal{L}(x,\lambda) = 0, \lambda \geq 0. \tag{5.4b}$$

Theorem 8

Suppose that f and $-c_i$, $i = 1, \cdots, m$ are convex and continuously differentiable on $\Re^n$. Suppose that $(\bar{x}, \bar{\lambda})$ is a solution pair of (5.1) at which LICQ holds. Then $(\bar{x}, \bar{\lambda})$ solves the problem (5.4).

Examples

Linear Programming:

$$\min c^T x \text{ s.t. } Ax - b \geq 0.$$

Quadratic Programming:

$$\min \frac{1}{2} x^T G x + c^T x \text{ s.t. } Ax - b \geq 0.$$

Quadratic Programming

Consider

$$\min \frac{1}{2} x^T G x + c^T x \quad s.t. \quad Ax - b \geq 0,$$

where G is symmetric positive definite. Then the ducal objective for this problem is

$$\begin{aligned} q(\lambda) = \inf_x \mathcal{L}(x, \lambda) &= \inf_x \frac{1}{2} x^T G x + c^T x - \lambda^T (Ax - b) \\ &= -\frac{1}{2} (A^T \lambda - c)^T G^{-1} (A^T \lambda - c)^T + b^T \lambda. \end{aligned}$$

The Wolfe dual is

$$\begin{aligned} \max_{(\lambda, x)} \quad & \tfrac{1}{2} x^T G x + c^T x - \lambda^T (Ax - b) \\ s.t. \quad & Gx + c - A^T \lambda = 0, \quad \lambda \geq 0, \end{aligned}$$

which requires only positive semidefiniteness of G.

Outline

1. Introduction
2. Constraint Qualifications
3. First-Order Optimality Conditions
4. Second-Order Optimality Conditions
5. Duality
6. References

References

D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, second ed., 1999.

R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, New York, second ed., 1987.

Thanks for your attention!