Practical Optimization Algorithms and Applications

Chapter IX: Theory of Constrained Optimization

Lingfeng NIU

Research Center on Fictitious Economy & Data Science,
University of Chinese Academy of Sciences

niulf@ucas.ac.cn

Outline

Outline

General Formulation for Constrained Optimization

A general formulation for constrained optimization is

$$\min_{x \in \Re^n} \quad f(x) \tag{1a}$$

$$\text{s.t.} \quad c_i(x) = 0, \quad i \in \mathcal{E}, \tag{1b}$$

$$c_i(x) \geq 0, \quad i \in \mathcal{I}. \tag{1c}$$

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

Define the *feasible set* Ω to be the set of points x that satisfy the constraints:

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

so that (1) can be written more compactly as

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

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.

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.

$$\|x\|_1 = |x_1| + |x_2| \leq 1,$$

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.

$$\|x\|_1 = |x_1| + |x_2| \leq 1,$$
$$x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1.$$

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.

$$\|x\|_1 = |x_1| + |x_2| \leq 1,$$
$$x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1.$$

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

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.

$$\|x\|_1 = |x_1| + |x_2| \leq 1,$$
$$x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1.$$

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

$$f(x) = \max(x^2, x),$$

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.

$$\|x\|_1 = |x_1| + |x_2| \leq 1,$$
$$x_1 + x_2 \leq 1, x_1 - x_2 \leq 1, -x_1 + x_2 \leq 1, -x_1 - x_2 \leq 1.$$

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

$$f(x) = \max(x^2, x),$$
$$\min t \text{ s.t. } t \geq x, t \geq x^2.$$

Introduction

Definition

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

We now derive mathematical characterization of the solutions of (1) or (3).

Definition

The active set $\mathcal{A}(x)$ at any feasible 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.

Outline

First-Order Optimality Conditions - A Geometric Viewpoint

Definition

The vector d is said to be a tangent to Ω at a point x if there are 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 $T_\Omega(x^*)$.

Theorem

If x^* is a local solution of (1), then we have

$$\nabla f(x^*)^T d \geq 0, \qquad \forall d \in T_\Omega(x^*). \tag{4}$$

The reverse of this result is not necessary true, i.e. 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

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

First-Order Optimality Conditions - A Geometric Viewpoint

Definition

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

$$N_\Omega(x) = \{v | c^T w \leq 0 \text{ for all } w \in T_\Omega(x)\},$$

where $T_\Omega(x)$ is the tangent cone. Each vector $v \in N_\Omega(x)$ is said to be a normal vector.

Geometrically, each normal vector v makes an angle of at least $\pi/2$ with every tangent vector.

Theorem

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

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

Linearized Feasible Direction

The tangent cone and norm cone only rely on the geometry of Ω. We want to obtained a first-order condition which could be expressed algebraically.

Definition

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

$$\mathcal{F}(x) = \{d | d^T \nabla c_i(x) = 0, \forall i \in \mathcal{E}; d^T \nabla c_i(x) \geq 0, \forall i \in \mathcal{A}(x) \cap \mathcal{I}\}.$$

The linearized feasible direction set depends on the algebraic specification of the constraint functions c_i, $i \in \mathcal{E} \cup \mathcal{I}$. Consider the following two cases:

- $\Omega = \{x | c_2(x) = 0\}$, where $c_2(x) = x_1^2 + x_2^2 - 2 = 0$;
- $\Omega = \{x | c_1(x) = 0\}$, where $c_1(x) = (x_1^2 + x_2^2 - 2)^2 = 0$.

Relating Tangent Cone and Linearized Feasible Direction

Theorem

Let x be a feasible point, we always have that $T_\Omega(x) \subset \mathcal{F}(x)$.

- Constraints qualifications are conditions under which the linearized feasible set $\mathcal{F}(x)$ is similar to the tangent cone $T_\Omega(x)$.
- 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 $T_\Omega(x)$.
- In fact, most constraint qualifications ensure that $\mathcal{F}(x)$ and $T_\Omega(x)$ are identical.

Constraint Qualifications

When one of the following constraint qualifications is satisfied at x, we have that $\mathcal{F}(x) = T_\Omega(x)$.

- LICQ: Given the point x and the active set $\mathcal{A}(x)$, 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.
- MFCQ: the Mangasarian-Fromovitz constraint qualification (MFCQ) holds if there exists a vector $w \in \Re^n$ such that

$$\begin{aligned} \nabla c_i(x)^T w > 0, &\quad \forall i \in \mathcal{A}(x) \cap \mathcal{I}, \\ \nabla c_i(x)^T w = 0, &\quad \forall i \in \mathcal{E}, \end{aligned}$$

  and the set of equality constraints gradients $\{\nabla c_i(x), i \in \mathcal{E}\}$ is linearly independent.
- Suppose that at some $x \in \Omega$, all active constraints $c_i(\cdot)$, $i \in \mathcal{A}(x)$, are linear functions.

The Lagrangian Function

Theorem

Suppose that the LICQ assumption holds at x. Then the normal cone $N_\Omega(x)$ could be simply defined as

$$N_\Omega(x) = -\{\sum_{i\in\mathcal{A}(x)} \lambda_i \nabla c_i(x), \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x) \cup \mathcal{I}\}.$$

Definition

The Lagrangian function for the general problem (1) is

$$\mathcal{L}(x,\lambda) = f(x) - \sum_{i\in\mathcal{E}\cap\mathcal{I}} \lambda_i c_i(x). \tag{5}$$

Here λ_i is called the Lagrange multiplier for constraint $c_i(x)$.

First-Order Necessary Conditions

Theorem

Suppose that x^* is a local solution of (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^*)

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, & & \text{(6a)}\\
c_i(x^*) &= 0, \quad \forall i \in \mathcal{E}, & & \text{(6b)}\\
c_i(x^*) &\geq 0, \quad \forall i \in \mathcal{I}, & & \text{(6c)}\\
\lambda_i^* &\geq 0, \quad \forall i \in \mathcal{I}, & & \text{(6d)}\\
\lambda_i^* c_i(x^*) &= 0, \quad \forall i \in \mathcal{E} \cup \mathcal{I}. & & \text{(6e)}
\end{aligned}$$

Conditions (6) are often known as the *Karush-Kuhn-Tucker* conditions, or KKT conditions for short.

First-Order Necessary Conditions

Theorem

Suppose that x^* is a local solution of (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^*)

$$
\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, & & \quad (6a)\\
c_i(x^*) &= 0, & \forall i \in \mathcal{E}, & \quad (6b)\\
c_i(x^*) &\geq 0, & \forall i \in \mathcal{I}, & \quad (6c)\\
\lambda_i^* &\geq 0, & \forall i \in \mathcal{I}, & \quad (6d)\\
\lambda_i^* c_i(x^*) &= 0, & \forall i \in \mathcal{E} \cup \mathcal{I}. & \quad (6e)
\end{aligned}
$$

Conditions (6) are often known as the *Karush-Kuhn-Tucker* conditions, or KKT conditions for short.

For a given problem (1) and solution point x^*, there may be many vectors λ^* for which the conditions (6) are satisfied. When the LICQ holds, the optimal λ^* is unique.

The Complementary Conditions

The conditions (6e) are complementary conditions, they imply that either constraint i is active or $\lambda_i^* = 0$, or possibly both. Since the Lagrange multiplier corresponding to inactive inequality constraints are zero, we can omit the terms for indices $i \notin \mathcal{A}(x^*)$ from (6a) and rewrite it as

$$0 = \nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*).$$

Definition

Given a local solution x^* of (1) and a vector λ^* satisfying (6), 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 $i \in \mathcal{I} \cap \mathcal{A}(x^*)$.

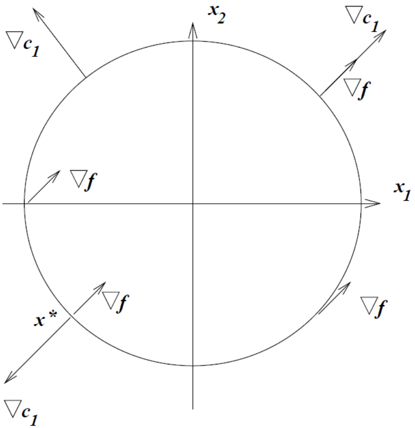
Usually, satisfaction of the strict complementary property makes it easier for algorithms to determine the active set $\mathcal{A}(x^*)$ and converge rapidly to the solution x^*.

Example - A Single Equality Constraint

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

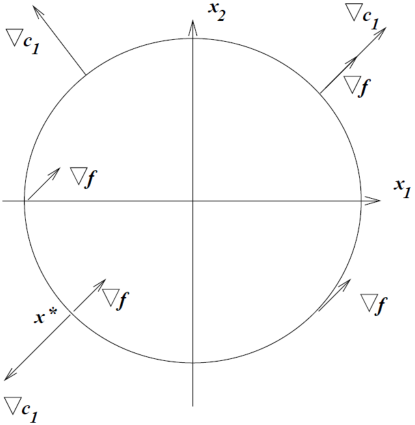
Example - A Single Equality Constraint

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



Example - A Single Equality Constraint

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



At the solution $x^* = (-1, -1)^T$, there is a scalar $\lambda_1^* = -1/2)$ such that

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*) \tag{7}$$

Example - A Single Equality Constraint

Consider a feasible x, we have $c_1(x) = 0$. To retain feasibility, we require $c_1(x + d) = 0$. By examining first-order Taylor series approximation, we obtain

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

Hence, the direction d retains feasibility with respect to c_1, to first order, when it satisfies

$$\nabla c_1(x)^T d = 0. \tag{8}$$

Example - A Single Equality Constraint

Consider a feasible x, we have $c_1(x) = 0$. To retain feasibility, we require $c_1(x + d) = 0$. By examining first-order Taylor series approximation, we obtain

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

Hence, the direction d retains feasibility with respect to c_1, to first order, when it satisfies

$$\nabla c_1(x)^T d = 0. \tag{8}$$

Similarly, a direction of improvement must produce a decrease in f, so that

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

or, to first order,

$$\nabla f(x)^T d < 0. \tag{9}$$

Example - A Single Equality Constraint

Consider a feasible x, we have $c_1(x) = 0$. To retain feasibility, we require $c_1(x + d) = 0$. By examining first-order Taylor series approximation, we obtain

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

Hence, the direction d retains feasibility with respect to c_1, to first order, when it satisfies

$$\nabla c_1(x)^T d = 0. \tag{8}$$

Similarly, a direction of improvement must produce a decrease in f, so that

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

or, to first order,

$$\nabla f(x)^T d < 0. \tag{9}$$

If there exists a direction d that satisfies both (8) and (9), we conclude that improvement on our current point x is possible. It follows that a necessary condition for optimality is that there exist no direction d satisfying both (8) and (9).

Example - A Single Equality Constraint

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

$$\nabla f(x) = \lambda_1 \nabla c_1(x) \tag{10}$$

holds at x, for some scalar λ_1.

Example - A Single Equality Constraint

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

$$\nabla f(x) = \lambda_1 \nabla c_1(x) \tag{10}$$

holds at x, for some scalar λ_1.

If $\triangledown f(x)$ and $\triangledown c_1(x)$ are not parallel, the direction defined by

$$d = -(I - \frac{\triangledown c_1(x) \triangledown c_1(x)^T}{\|\triangledown c_1(x)\|^2}) \triangledown f(x)$$

satisfies both conditions (8) and (9).

Example - A Single Equality Constraint

By introducing the *Lagrangian function*

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

and noting that $\nabla_x \mathcal{L}(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x)$ we can state the condition (10) equivalently as follows: At the solution x^*, there is a scalar λ_1^* such that

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

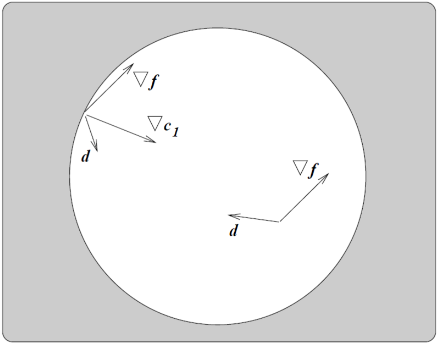
This observation suggests that we can search for solutions of the constrained problem by searching for stationary points of the Lagrangian function. Though the condition (10) appears to be *necessary* for an optimal solution of the equality-constrained problem, it is clearly not sufficient. (Consider the point $x = (1, 1)^T$ in the above example.)

Example - A Single Inequality Constraint

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

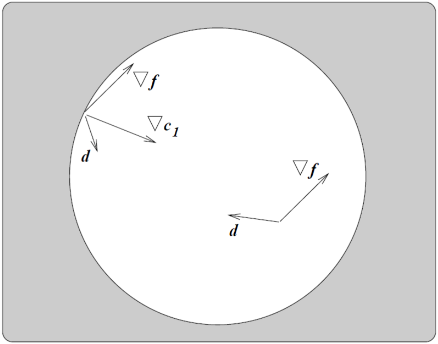
Example - A Single Inequality Constraint

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



Example - A Single Inequality Constraint

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



The solution is still $x^* = (-1, -1)^T$ and the condition (10) holds for the value $\lambda_1^* = -1/2$.

Example - A Single Inequality Constraint

We conjecture a given feasible point x is not optimal if we can find a step d that both retains feasibility and decreases the objective function f to first order.

The direction d improves the objective function, to first order, if $\nabla f(x)^T d < 0$.

The direction 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

$$c_1(x) + \nabla c_1(x)^T d \geq 0. \tag{11}$$

In determining whether a direction d exists that satisfies both (9) and (11), we consider the following two cases:

- Case I: x lies *strictly inside the circle*;
- Case II: x lies *on the boundary of the circle*.

Example - A Single Inequality Constraint

Case I: If x lies strictly inside the circle, i.e. $c_1(x) > 0$ holds.

In this case, *any* vector d satisfies the condition (11), provided only that its length is sufficiently small.

Whenever $\nabla f(x) \neq 0$, we can obtain a direction d that satisfies both (9) and (11) by setting

$$d = \alpha \nabla f(x)$$

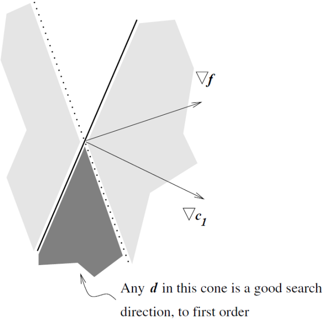
for any positive scalar α sufficiently small. The only situation in which such a direction fails to exist is when

$$\nabla f(x) = 0.$$

Example - A Single Inequality Constraint

Case II: If x lies on the boundary of the circle, i.e. $\nabla f(x) = 0$ holds.

The conditions (9) and (11) therefore become $\nabla c_1(x)^T d \geq 0, \nabla f(x)^T d < 0$. The first of these conditions defines an open half-space, while the second defines a closed half-space, as illustrated in following figure.



It is clear from this figure that the two regions fail to intersect only when $\nabla f(x)$ and $\nabla c_1(x)$ point in the same direction, that is, when $\nabla f(x) = \lambda_1 \nabla c_1(x)$, for some $\lambda_1 \geq 0$.

Example - A Single Inequality Constraint

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

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

where we also require that

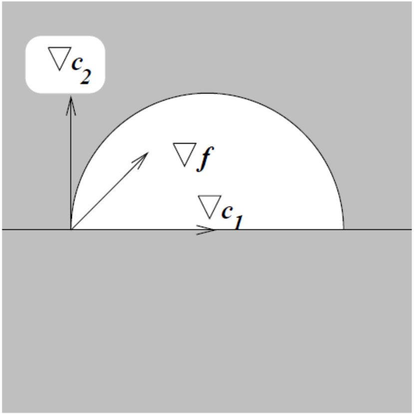
$$\lambda_1^* c_1(x^*) = 0.$$

Example - Two Inequality Constraints

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

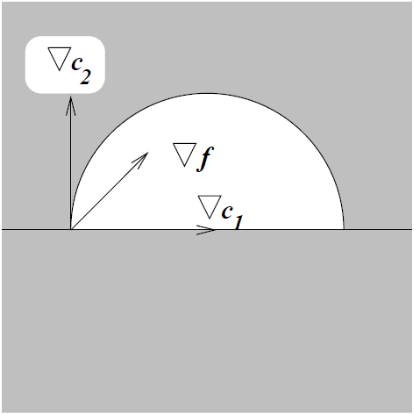
Example - Two Inequality Constraints

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



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 solution lies at $(-\sqrt{2}, 0)^T$, a point at which both constraints are active.

Example - Two Inequality Constraints

We include an additional term $\lambda_i c_i(x)$ in the Lagrangian for each additional constraint, so the definition of $\mathcal{L}$ becomes

$$L(x, \lambda) = f(x) - \lambda_1 c_1(x) - \lambda_2 c_2(x),$$

where $\lambda = (\lambda_1, \lambda_2)^T$ is the vector of Lagrange multipliers. Them

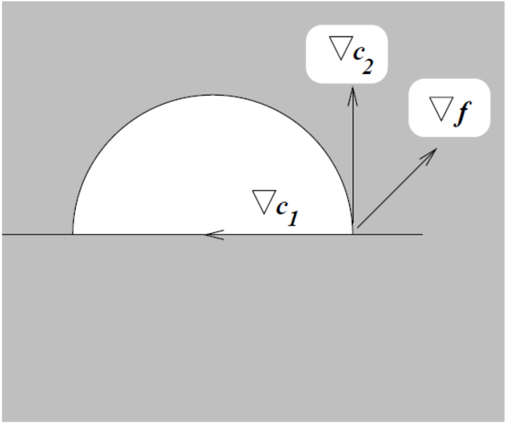
$$\nabla_x L(x^*, \lambda^*) = 0, \text{for some } \lambda^* \geq 0,$$

By applying the complementarity condition to both inequality constraints, we obtain

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

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

Undecided Directions from First Order Information

The first-order conditions-the KKT conditions-tell us how the first derivatives of f and the active constraints c_i are related at a solution x^*. When these conditions are satisfied, a move along any vector w from $\mathcal{F}(x^*)$

- either increases the first-order approximation to the objective function, i.e. $w^T \nabla f(x^*) > 0$,
- or else keeps this value the same, i.e. $w^T \nabla f(x^*) = 0$.

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.

Critical Cone

Definition

Given $\mathcal{F}(x^*)$ and some Lagrange multiplier vector λ^* satisfying the KKT conditions (6), we define a subset the *critical cone* $\mathcal{C}(x^*, \lambda^*)$ as follows:

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

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

From the fact that $\lambda_i^* = 0, \forall i \in \mathcal{I} \setminus \mathcal{A}(x^*)$, it follows that

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

Hence from the KKT conditions we have that

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

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

Second-Order Conditions

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 for directions in the critical cone.

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 $w^T \nabla f(x^*) = 0$.

Since we are discussing second derivatives, stronger smoothness assumptions are needed here than in the previous discussions:

f **and** c_i**,** $i \in \mathcal{E} \cup \mathcal{I}$**, are all assumed to be twice continuously differentiable.**

Second-Order Necessary Conditions

Theorem

Suppose that x^* is a local solution of (1) and that the LICQ condition is satisfied. Let λ^* be a Lagrange multiplier vector such that the KKT conditions (6) are satisfied. Then

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0, \forall w \in \mathcal{C}(x^*, \lambda^*). \tag{12}$$

This theorem shows that if x^* is a local solution, then the Hessian of the Lagrangian has nonnegative curvature along critical directions.

Second-Order Sufficient Conditions

Theorem

Suppose that for some feasible point $x^* \in \Re^n$, there is a Lagrange multiplier vector λ^* such that the KKT conditions (6) are satisfied. Suppose also that

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w > 0, \forall w \in \mathcal{C}(x^*, \lambda^*), w \neq 0. \tag{13}$$

Then x^* is a strict local solution for (1).

Second-Order Conditions and Projected Hessians

When the multiplier λ^* that satisfies the KKT conditions (6) is unique and strict complementarity holds, the definition of $\mathcal{C}(x^*, \lambda^*)$ reduces to

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

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

In other words, $\mathcal{C}(x^*, \lambda^*)$ is the null space of the matrix whose rows are the active constraint gradients at x^*. Define the matrix Z with full column rank whose columns span the space $\mathcal{C}(x^*, \lambda^*)$, that is

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

Second-Order Conditions and Projected Hessians

The second order necessary condition (12) can be restated as

$$u^T Z^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z u \geq 0, \forall u,$$

or, more succinctly,

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

Second-Order Conditions and Projected Hessians

The second order necessary condition (12) can be restated as

$$u^T Z^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) Z u \geq 0, \forall u,$$

or, more succinctly,

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

Similarly, the second order sufficient condition (13) can be restated as

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

Second-Order Conditions and Projected Hessians

The second order necessary condition (12) can be restated as

$$u^T Z^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) Z u \geq 0, \forall u,$$

or, more succinctly,

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

Similarly, the second order sufficient condition (13) 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.

Outline

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. To put it another way, λ_i indicates how hard f is "pushing" or "pulling" against the particular constraint c_i.

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. To put it another way, λ_i indicates how hard f is "pushing" or "pulling" against the particular constraint c_i.

For the inactive constraint $i \notin \mathcal{A}(x^*)$, since $c_i(x^*) > 0$, the solution x^* and function value $f(x^*)$ are quite indifferent to whether this constraint is present or not. If we perturb c_i by a tiny amount, it will still be inactive and x^* will still be a local solution of the optimization problem. Since $\lambda_i^* = 0$ from (6e), the Lagrange multiplier indicates accurately that constraint i is not significant.

Lagrange Multipliers and Sensitivity

For the active constraint i $i \in \mathcal{A}(x^*)$, let us perturb the right-hand-side of this constraint a little, requiring, say, that $c_i(x) \geq -\epsilon\|\nabla c_i(x^*)\|$ instead of $c_i(x) \geq 0\|$. Suppose that ϵ is sufficiently small that the perturbed solution $x^*(\epsilon)$ still has the same set of active constraints, and that the Lagrange multipliers are not much affected by the perturbation. We then find that

$$-\epsilon\|\nabla c_i(x^*)\| = c_i(x^*(\epsilon)) - c_i(x^*) \approx (x^*(\epsilon) - x^*)^T \nabla c_i(x^*),$$

and

$$0 = c_j(x^*(\epsilon)) - c_i(x^*) \approx (x^*(\epsilon) - x^*)^T \nabla c_j(x^*), \forall j \in \mathcal{A}(x^*) \text{ with } j \neq i.$$

The value of $f(x^*(\epsilon))$, meanwhile, can be estimated with the help of (6a). We have

$$\begin{aligned} f(x^*(\epsilon)) - f(x^*) &\approx (x^*(\epsilon) - x^*)^T \nabla f(x^*) \\ &= \sum_{j \in \mathcal{A}(x^*)} \lambda_j^* (x^*(\epsilon) - x^*)^T \nabla c_j(x^*) \\ &\approx -\epsilon\|\nabla c_i(x^*)\|\lambda_i \end{aligned}$$

Lagrange Multipliers and Sensitivity

By taking limits, we see that the family of solutions $x^*(\epsilon)$ satisfies

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

A sensitivity analysis of this problem would conclude that

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

Lagrange Multipliers and Sensitivity

This discussion motivates the definition below, which classifies constraints according to whether or not their corresponding Lagrange multiplier is zero.

Definition

Let x^* be a solution of the problem (1), and suppose that the KKT conditions (6) are satisfied. We say that an inequality constraint c_i is strongly active or binding if $i \in \mathcal{A}(x^*)$ and $\lambda_i^* > 0$ for some Lagrange multiplier λ satisfying (6). We say that c_i is weakly active if $i \in \mathcal{A}(x^*)$ and $\lambda_i^* = 0$ for all λ^* satisfying (6).

Outline

Primal and Dual Problem

Consider

$$\min_{x\in\Re^n} f(x) \text{ s.t. } c(x) \geq 0, \tag{14}$$

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

Primal and Dual Problem

Consider

$$\min_{x \in \Re^n} f(x) \text{ s.t. } c(x) \geq 0, \tag{14}$$

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

The Lagrangian function with Lagrange multiplier vector $\lambda \in \Re^m$ is

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x).$$

We define the dual objective function $q : \Re^m \to \Re$ as follows:

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

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

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

The dual problem to (14) is defined as follows:

$$\max_{\lambda \in \Re^m} q(\lambda) \text{ s.t. } \lambda \geq 0. \tag{15}$$

Weak Duality

Theorem

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

Proof:

$$\mathcal{L}(x,(1-\alpha)\lambda^0+\alpha\lambda^1) = (1-\alpha)\mathcal{L}(x,\lambda^0)+\alpha\mathcal{L}(x,\lambda^1), \forall\lambda^0,\lambda^1 \in \Re^m, x \in \Re^n, \alpha \in [0,1].$$

$$q(x,(1-\alpha)\lambda^0+\alpha\lambda^1) \geq (1-\alpha)q(x,\lambda^0)+\alpha q(x,\lambda^1).$$

Weak Duality

Theorem

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

Proof:

$$\mathcal{L}(x,(1-\alpha)\lambda^0+\alpha\lambda^1) = (1-\alpha)\mathcal{L}(x,\lambda^0)+\alpha\mathcal{L}(x,\lambda^1), \forall \lambda^0,\lambda^1 \in \Re^m, x \in \Re^n, \alpha \in [0,1].$$

$$q(x,(1-\alpha)\lambda^0+\alpha\lambda^1) \geq (1-\alpha)q(x,\lambda^0)+\alpha q(x,\lambda^1).$$

Theorem

For any $\bar{x}$ feasible for (14) and any $\bar{\lambda} \geq 0$, we have $q(\bar{\lambda}) \leq f(\bar{x})$.

Proof:

$$q(\bar{\lambda}) = \inf_x f(x) - \bar{\lambda}^T c(x) \leq f(\bar{x}) - \bar{\lambda}^T c(\bar{x}) \leq f(\bar{x}).$$

.

How the dual problem is related to primal problem?

The KKT conditions (6) specialized to (14) are as follows:

$$
\begin{aligned}
\nabla f(\bar{x}) - \nabla c(\bar{x})\bar{\lambda} &= 0, && && \text{(16a)} \\
c(\bar{x}) &\geq 0, && && \text{(16b)} \\
\bar{\lambda} &\geq 0, && && \text{(16c)} \\
\bar{\lambda}_i c_i(\bar{x}) &= 0, && i = 1, \cdots, m. && \text{(16d)}
\end{aligned}
$$

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

How the dual problem is related to primal problem?

The KKT conditions (6) specialized to (14) are as follows:

$$\nabla f(\bar{x}) - \nabla c(\bar{x})\bar{\lambda} = 0, \tag{16a}$$
$$c(\bar{x}) \geq 0, \tag{16b}$$
$$\bar{\lambda} \geq 0, \tag{16c}$$
$$\bar{\lambda}_i c_i(\bar{x}) = 0, \qquad i = 1, \cdots, m. \tag{16d}$$

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

The following result shows that optimal Lagrange multiplier for (14) are solutions of the dual problem (15) under certain conditions.

Theorem

Suppose that $\bar{x}$ is a solution of (14) 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 is a solution of (15)

How the dual problem is related to primal problem?

The following theorem shows that solutions to the dual problem can sometimes be used to derive solutions to the original problem.

Theorem

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 (14) at which LICQ holds. Suppose that $\hat{\lambda}$ solves (15) 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 (14)), and $f(\bar{x}) = \mathcal{L}(\hat{x}, \hat{\lambda})$

Wolfe Dual

A slightly different form of duality that is convenient for computations, knows the *Wolfe dual*, can be stated as follows:

$$\max_{x,\lambda} \quad \mathcal{L}(x,\lambda) \tag{17a}$$

$$s.t. \quad \nabla_x \mathcal{L}(x,\lambda) = 0, \lambda \geq 0. \tag{17b}$$

Wolfe Dual

A slightly different form of duality that is convenient for computations, knows the *Wolfe dual*, can be stated as follows:

$$\max_{x,\lambda} \quad \mathcal{L}(x,\lambda) \tag{17a}$$

$$s.t. \quad \nabla_x \mathcal{L}(x,\lambda) = 0, \lambda \geq 0. \tag{17b}$$

The following results explains the relationship of the Wolfe dual to (1).

Theorem

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 (14) at which LICQ holds. Then $(\bar{x}, \bar{\lambda})$ solves the problem (17).

Examples

Please give the dual problem to the following problems:

- Linear Programming:

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

- Quadratic Programming:

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

Thanks for your attention!