Theorems and proofs for Nonlinear optimisation

Edvin Listo Zec
edvinli@student.chalmers.se

December 2014

Introduction

This text is written as an aid for those that are taking the course TMA947 Nonlinear optimisation 2014/15. It contains the recommended theorems and proofs from the year 2014/15 taken from the book An Introduction to Continuous Optimization (second edition) by Niclas Andréasson, Anton Evgrafov, Michael Patriksson, Emil Gustavsson and Magnus Önnheim.

Contents

1 Separation Theorem

Theorem

Suppose that the set $C \subseteq \mathbb{R}^n$ is nonempty, closed and convex. Suppose further that the point $\boldsymbol{y} \notin C$. Then it holds that $\exists$ a vector $\boldsymbol{\pi} \neq \boldsymbol{0}^n$ and $\alpha \in \mathbb{R}$ such that $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{y} > \alpha$ and $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{x} \leq \alpha$, $\forall \boldsymbol{x} \in C$.

Proof

Define a function $f : \mathbb{R}^n \to \mathbb{R}$ as $f(\boldsymbol{x}) := \|\boldsymbol{x} - \boldsymbol{y}\|^2/2$, $\boldsymbol{x} \in \mathbb{R}^n$. From Weierstrass' theorem we get that $\exists$ a minimum $\tilde{\boldsymbol{x}}$ of f over C. By the first-order optimality conditions, the minimum is characterised by the variational inequality:

$$(\boldsymbol{y} - \tilde{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x} - \tilde{\boldsymbol{x}}) \leq 0, \ \forall \boldsymbol{x} \in C \ (\text{since } -\nabla f(\tilde{\boldsymbol{x}}) = \boldsymbol{y} - \tilde{\boldsymbol{x}})$$

We now define $\boldsymbol{\pi} := \boldsymbol{y} - \tilde{\boldsymbol{x}}$. Since $\boldsymbol{y} \notin C$, $\boldsymbol{\pi}$ will be non-zero. Let $\alpha := (\boldsymbol{y} - \tilde{\boldsymbol{x}})^{\mathrm{T}}\tilde{\boldsymbol{x}}$. The above variational inequality gives now that $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{x} \leq \boldsymbol{\pi}^{\mathrm{T}}\tilde{\boldsymbol{x}} = \alpha$, $\forall \boldsymbol{x} \in C$, while $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{y} - \alpha = (\boldsymbol{y} - \tilde{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{y} - \tilde{\boldsymbol{x}}) = \|\boldsymbol{y} - \tilde{\boldsymbol{x}}\|^2 > 0$. $\square$

2 Farkas' Lemma

Theorem

Consider the systems:

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b}, \\ \boldsymbol{x} &\geq \boldsymbol{0}^n, \end{aligned} \tag{1}$$

and

$$\begin{aligned} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} &\leq \boldsymbol{0}^n, \\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} &> 0. \end{aligned} \tag{2}$$

If $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, then *exactly one* of the systems 1 and 2 has a feasible solution, and the other system is inconsistent.

Proof

Assume that system 1 has a solution $\boldsymbol{x}$, then

$$\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} > 0$$

But $\boldsymbol{x} \geq \mathbf{0}^n \Rightarrow \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \leq \mathbf{0}^n$ cannot hold $\Rightarrow$ system 2 is infeasible.

Assume now that 2 is infeasible and consider the linear program:

$$\begin{aligned} \text{maximise} \quad & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y} \\ \text{subject to} \quad & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \leq \mathbf{0}^n, \\ & \boldsymbol{y} \text{ free}, \end{aligned}$$

and its dual program:

$$\begin{aligned} \text{minimise} \quad & (\mathbf{0}^n)^{\mathrm{T}}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \\ & \boldsymbol{x} \geq \mathbf{0}^n. \end{aligned}$$

$\boldsymbol{y} = \mathbf{0}^m$ will be an optimal solution to the primal program, since system 2 is infeasible. The Strong Duality Theorem implies thus that there is an optimal solution to the dual program, which is feasible in system 1.

We have thus proved the equivalence

$$(1) \Leftrightarrow \neg(2).$$

Which is logically equivalent to

$$\neg(1) \Leftrightarrow (2).$$

This means that *exactly one* of the systems 1 and 2 has a solution. $\square$

3 Characterization of convex functions in C^1

Theorem

If $f \in C^1$ on an open convex set S, then:

(a) f is convex on $S \Leftrightarrow f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}),\ \forall \boldsymbol{x},\boldsymbol{y} \in S$.

(b) f is convex on $S \Leftrightarrow [\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})]^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{y}) \geq 0,\ \forall \boldsymbol{x},\boldsymbol{y} \in S$.

Proof

(a) [$\Rightarrow$] Take $\boldsymbol{x}^1,\boldsymbol{x}^2 \in S$ and $\lambda \in (0,1)$. Then we have that:

$$\lambda f(\boldsymbol{x}^1) + (1-\lambda) f(\boldsymbol{x}^2) \geq f(\lambda \boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2)$$

$$\Leftrightarrow$$

$$f(\boldsymbol{x}^1) - f(\boldsymbol{x}^2) \geq (1/\lambda)[f(\lambda\boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2) - f(\boldsymbol{x}^2)],$$

which holds since $\lambda > 0$. If we now let $\lambda \to 0^+$, the right-hand side of the above inequality goes to the directional derivative of f at $\boldsymbol{x}^2$ in the direction of $(\boldsymbol{x}^1 - \boldsymbol{x}^2)$. The limit thus becomes:

$$f(\boldsymbol{x}^1) - f(\boldsymbol{x}^2) \geq \nabla f(\boldsymbol{x}^2)^{\mathrm{T}}(\boldsymbol{x}^1 - \boldsymbol{x}^2).$$

The result follows.

[$\Leftarrow$] We have that:

$$f(\boldsymbol{x}^1) \geq f(\lambda\boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2) + (1-\lambda)\nabla f(\lambda\boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2)^{\mathrm{T}}(\boldsymbol{x}^1 - \boldsymbol{x}^2),$$
$$f(\boldsymbol{x}^2) \geq f(\lambda\boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2) + \lambda\nabla f(\lambda\boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2)^{\mathrm{T}}(\boldsymbol{x}^2 - \boldsymbol{x}^1).$$

We multiply the inequalities with λ and $(1-\lambda)$, respectively, and add them together. This yields the result sought.

(b)[$\Rightarrow$] We use (a) and the two inequalities:

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}), \quad \boldsymbol{x},\boldsymbol{y} \in S,$$
$$f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{y}), \quad \boldsymbol{x},\boldsymbol{y} \in S.$$

If we add them together, we get that $[\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})]^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{y}) \geq 0,\ \forall \boldsymbol{x},\boldsymbol{y} \in S$.

[$\Leftarrow$] We get from the mean-value theorem that

$$f(\boldsymbol{x}^2) - f(\boldsymbol{x}^1) = \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}^2 - \boldsymbol{x}^1), \tag{3}$$

where $\boldsymbol{x} = \lambda\boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2$, $\lambda \in (0,1)$. By assumption we have that $[\nabla f(\boldsymbol{x})] - \nabla f(\boldsymbol{x}^1)]^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^1) \geq 0$, so $(1-\lambda)[\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{x}^1)]^{\mathrm{T}}(\boldsymbol{x}^2 - \boldsymbol{x}^1) \geq 0$. From this it follows that $\nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}^2 - \boldsymbol{x}^1) \geq \nabla f(\boldsymbol{x}^1)^{\mathrm{T}}(\boldsymbol{x}^2 - \boldsymbol{x}^1)$. By using this inequality and (3), we get that $f(\boldsymbol{x}^2) \geq f(\boldsymbol{x}^1) + \nabla f(\boldsymbol{x}^1)^{\mathrm{T}}(\boldsymbol{x}^2 - \boldsymbol{x}^1)$.
This completes the proof. □

4 The Fundamental Theorem of global optimality

Theorem

Consider the problem:

$$\begin{aligned} &\text{minimise} && f(\boldsymbol{x}) \\ &\text{subject to} && \boldsymbol{x} \in S, \end{aligned}$$

If S is a convex set and f is a convex function on S, then every local minimum of f over S is also a global minimum.

Proof

Suppose $\boldsymbol{x}^*$ is a local minimum, but not a global one. Consider then a vector $\bar{\boldsymbol{x}} \in S$, such that $f(\bar{\boldsymbol{x}}) < f(\boldsymbol{x}^*)$. Let $\lambda \in (0,1)$. The following holds by the convexity of S and f:

$$\lambda\bar{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^* \in S$$
$$f(\lambda\bar{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}^*) \leq \lambda f(\bar{\boldsymbol{x}}) + (1-\lambda)f(\boldsymbol{x}^*) < f(\boldsymbol{x}^*)$$

By choosing $\lambda > 0$ small enough, we get a contradiction to the local optimality of $\boldsymbol{x}^*$. $\square$

5 Necessary optimality conditions, C^1 case

Theorem

Suppose that $S \subseteq \mathbb{R}^n$ and that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is in C^1 around a point $\boldsymbol{x} \in S$ for which $f(\boldsymbol{x}) < +\infty$.

(a) If $\boldsymbol{x}^* \in S$ is a local minimum of f over S, then $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{p} \geq 0$ holds for every feasible direction $\boldsymbol{p}$ at $\boldsymbol{x}^*$.

(b) Suppose that S is convex and that $f \in C^1(S)$. If $\boldsymbol{x}^* \in S$ is a local minimum of f over S, then

$$\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \boldsymbol{x} \in S \tag{4}$$

holds.

Proof

(a) We begin by Taylor expanding f around $\boldsymbol{x}^*$:

$$f(\boldsymbol{x}^* + \alpha\boldsymbol{p}) = f(\boldsymbol{x}^*) + \alpha\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{p} + o(\alpha).$$

The proof is by contradiction. We know from a proposition that if there is a direction $\boldsymbol{p}$ for which it holds that $\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{p} < 0$, then $f(\boldsymbol{x}^* + \alpha\boldsymbol{p}) < f(\boldsymbol{x}^*)$ for all sufficiently small $\alpha > 0$. Here it suffices to state that $\boldsymbol{p}$ should be a feasible direction in order to reach a contradiction to the local optimality of $\boldsymbol{x}^*$.

(b) If S is convex, then it holds that $\forall \boldsymbol{x} \in S$, $\boldsymbol{p} := \boldsymbol{x} - \boldsymbol{x}^*$ is a feasible direction. Then (a) $\Rightarrow$ (4). $\square$

6 Necessary and sufficient global optimality conditions

Theorem

Suppose that $S \subseteq \mathbb{R}^n$ is nonempty and convex. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function such that $f \in C^1(S)$. Then:

$$\boldsymbol{x}^* \text{ is a global minimum of } f \text{ over } S \Leftrightarrow \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \boldsymbol{x} \in S \tag{5}$$

Proof

[⇒] This has been shown in the previous theorem, since a global minimum is a local minimum.
[⇐] The convexity of f yields that $\forall \boldsymbol{y} \in S$,

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}^*) \geq f(\boldsymbol{x}^*),$$

where the second inequality stems from the second part of (5). $\square$

7 Karush–Kuhn–Tucker necessary conditions

Theorem

Assume that at a given point $\boldsymbol{x}^* \in S$ Abadie's constraint qualification holds. If $\boldsymbol{x}^* \in S$ is a local minimum of f over S, then $\exists \boldsymbol{\mu} \in \mathbb{R}^m$ such that:

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}^*) &= \boldsymbol{0}^n \\ \mu_i g_i(\boldsymbol{x}^*) &= 0, \quad i = 1, \dots, m, \\ \boldsymbol{\mu} &\geq \boldsymbol{0}^m. \end{aligned} \tag{6}$$

In other words,

$$\left.\begin{array}{r} \boldsymbol{x}^* \text{ local minimum of } f \text{ over } S \\ \text{Abadie's CQ holds at } \boldsymbol{x}^* \end{array}\right\} \Rightarrow \exists \boldsymbol{\mu} \in \mathbb{R}^m : (6) \text{ holds.}$$

Proof

From theory we know that $\mathring{F}(\boldsymbol{x}^*) \cap T_S(\boldsymbol{x}^*) = \emptyset$, where

$$\mathring{F}(\boldsymbol{x}^*) := \{\boldsymbol{p} \in \mathbb{R}^n \mid \nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{p} < 0\}$$

and $T_S(\boldsymbol{x}^*)$ is the tangent cone. With $G(\boldsymbol{x}) := \{\boldsymbol{p} \in \mathbb{R}^n \mid \nabla g_i(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{p} < 0, i \in \mathcal{I}(\boldsymbol{x})\}$, where $\mathcal{I}(\boldsymbol{x})$ denotes the index set of active inequality constraints at $\boldsymbol{x} \in \mathbb{R}^n$, we get due to our assumptions that $\mathring{F}(\boldsymbol{x}^*) \cup G(\boldsymbol{x}^*) = \emptyset$.

Now construct a matrix $\boldsymbol{A}$ with columns $\nabla g_i(\boldsymbol{x}^*)$, $i \in \mathcal{I}(\boldsymbol{x})$. Then, the system $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{p} < \boldsymbol{0}^{|\mathcal{I}(\boldsymbol{x}^*)|}$ and $-\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{p} > 0$ has no solutions. $|\mathcal{I}(\boldsymbol{x}^*)|$ denotes the cardinality of the set $\mathcal{I}(\boldsymbol{x}^*)$. By Farkas' Lemma the system $\boldsymbol{A\xi} = -\nabla f(\boldsymbol{x}^*)$, $\boldsymbol{\xi} \geq \boldsymbol{0}^{|\mathcal{I}(\boldsymbol{x}^*)|}$ has a solution. Define the vector $\boldsymbol{\mu}_{\mathcal{I}(\boldsymbol{x}^*)} = \boldsymbol{\xi}$, and $\mu_i = 0$ for $i \notin \mathcal{I}(\boldsymbol{x}^*)$. Then $\boldsymbol{\mu}$ verifies the KKT conditions in (6). □

8 Sufficiency of the Karush–Kuhn–Tucker conditions for convex problems

Theorem

Assume that the problem

$$\begin{aligned} &\text{minimise} && f(\boldsymbol{x}) \\ &\text{subject to} && \boldsymbol{x} \in S, \end{aligned}$$

with the feasible set S given by:

$$S := \{\boldsymbol{x} \in \mathbb{R}^n \mid g_i(\boldsymbol{x}) < 0, \quad i = 1, \ldots, m; \; h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, l\}$$

is convex, i.e. the objective function f as well as the functions g_i are convex and the function h_j are affine. Assume further that for $\boldsymbol{x}^* \in S$ the KKT conditions:

$$\begin{aligned} \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \tilde{\mu}_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{l} \tilde{\lambda}_j \nabla h_j(\boldsymbol{x}^*) &= \boldsymbol{0}^n, \\ \tilde{\mu}_i g_i(\boldsymbol{x}^*) &= 0, \quad i = 1, \ldots, m, \\ \tilde{\boldsymbol{\mu}} &\geq \boldsymbol{0}^m. \end{aligned} \tag{7}$$

are satisfied. Then $\boldsymbol{x}^*$ is a globally optimal solution of the above problem. In other words,

$$\left.\begin{array}{r} \text{the above problem is convex} \\ \text{the KKT conditions (7) hold at } \boldsymbol{x}^* \end{array}\right\} \Rightarrow \boldsymbol{x}^* \text{ global minimum in the above problem.}$$

Proof

Choose an arbitrary $\boldsymbol{x} \in S$. Then by the convexity of the functions g_i it holds that

$$-\nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \geq g_i(\boldsymbol{x}^*) - g_i(\boldsymbol{x}) = -g_i(\boldsymbol{x}) \geq 0, \quad \forall i \in \mathcal{I}(\boldsymbol{x}^*), \tag{8}$$

and using the affinity of the functions h_j we get that

$$-\nabla h_j(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*) = h_j(\boldsymbol{x}^*) - h_j(\boldsymbol{x}) = 0, \quad \forall j = 1,\dots,l. \tag{9}$$

By using the convexity of the objective function, the first two parts of (7), the non-negativity of μ_i and equations (8) and (9) we obtain the inequality

$$\begin{aligned} f(\boldsymbol{x}) - f(\boldsymbol{x}^*) &\geq \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*) \\ &= -\sum_{i\in\mathcal{I}(\boldsymbol{x}^*)} \mu_i \nabla g_i(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*) - \sum_{j=1}^{l} \lambda_j \nabla h_j(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*) \\ &\geq 0. \end{aligned}$$

Since $\boldsymbol{x}$ was arbitrary, $\boldsymbol{x}^*$ solves the problem. □

9 Relaxation theorem

Definitions

Given the problem to find

$$\begin{aligned} f^* := \underset{x}{\text{infimum}} \quad & f(\boldsymbol{x}), \\ \text{subject to} \quad & \boldsymbol{x} \in S, \end{aligned} \tag{10}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a given function and $S \subseteq \mathbb{R}^n$, we define a relaxation to (10) to be a problem of the following form: find

$$\begin{aligned} f_R^* := \underset{x}{\text{infimum}} \quad & f_R(\boldsymbol{x}), \\ \text{subject to} \quad & \boldsymbol{x} \in S_R, \end{aligned} \tag{11}$$

where $f_R : \mathbb{R}^n \to \mathbb{R}$ is a function with the property that $f_R < f$ on S, and where $S \subseteq S_R$.

Theorem

(a) [*relaxation*] $f_R^* \leq f^*$

(b) [*infeasibility*] If (11) is infeasible, then so is (10).

(c) [*optimal relaxation*] If the problem (11) has an optimal solution, $\boldsymbol{x}_R^*$, for which it holds that

$$\boldsymbol{x}_R^* \in S \text{ and } f_R(\boldsymbol{x}_R^*) = f(\boldsymbol{x}_R^*),$$

then $\boldsymbol{x}_R^*$ is an optimal solution to (10) as well.

Proof

The result in (a) is obvious, since every solution feasible in (10) is also feasible in (11) and has a lower objective value in the latter problem. The result in (b) follows for similar reasons. For the result in (c), we note that

$$f(\boldsymbol{x}_R^*) = f_R(\boldsymbol{x}_R^*) \leq f_R(\boldsymbol{x}) \leq f(\boldsymbol{x}), \quad \boldsymbol{x} \in S,$$

from which the result follows. □

10 Weak Duality Theorem

Theorem

Consider the following two problems:

$$\begin{aligned} f^* := \underset{x}{\text{infimum}} \quad & f(\boldsymbol{x}), \\ \text{subject to} \quad & \boldsymbol{x} \in X, \\ & g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m, \end{aligned} \tag{12}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ are given functions, and $X \subseteq \mathbb{R}^n$. For this problem we assume that $-\infty < f^* < \infty$.

$$\begin{aligned} \underset{\mu}{\text{maximise}} \quad & q(\boldsymbol{\mu}), \\ \text{subject to} \quad & \boldsymbol{\mu} \geq \mathbf{0}^m, \end{aligned} \tag{13}$$

where $q(\boldsymbol{\mu}) := \underset{x \in X}{\text{infimum}}\ L(\boldsymbol{x},\boldsymbol{\mu})$, called the Lagrangian dual function.

Let $\boldsymbol{x}$ and $\boldsymbol{\mu}$ be feasible in problems (12) and (13), respectively. Then $q(\boldsymbol{\mu}) \leq f(\boldsymbol{x})$ and in particular, $q^* < f^*$.

Proof

$\forall \boldsymbol{\mu} > \mathbf{0}^m$ and $\boldsymbol{x} \in X$ with $\boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}^m$,

$$q(\boldsymbol{\mu}) = \underset{z \in X}{\text{infimum}}\ L(\boldsymbol{z}, \boldsymbol{\mu}) \leq f(\boldsymbol{x}) + \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{x}) \leq f(\boldsymbol{x}),$$

so

$$q^* = \underset{\boldsymbol{\mu} > \mathbf{0}^m}{\text{supremum}}\ q(\boldsymbol{\mu}) \leq \underset{\boldsymbol{x} \in X : \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}^m}{\text{infimum}}\ f(\boldsymbol{x}) = f^*$$

The result follows. □

11 Global optimality conditions in the absence of a duality gap

Theorem

The vector $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ is a pair of primal optimal solution and Lagrange multiplier vector iff

$$\begin{aligned} \boldsymbol{\mu}^* \geq \mathbf{0}^m, &\quad \textit{(Dual feasibility)} \\ \boldsymbol{x}^* \in \underset{\boldsymbol{x}\in X}{\arg\min}\, L(\boldsymbol{x}^*, \boldsymbol{\mu}^*), &\quad \textit{(Lagrangian optimality)} \\ \boldsymbol{x}^* \in X, \quad \boldsymbol{g}(\boldsymbol{x}^*) \leq \mathbf{0}^m, &\quad \textit{(Primal feasibility)} \\ \mu_i^*(\boldsymbol{x}^*) = 0, \quad i = 1, \ldots, m,. &\quad (\textit{Complementary slackness}) \end{aligned} \tag{14}$$

Proof

Suppose that the pair $(\boldsymbol{x}^*,\boldsymbol{\mu}^*)$ satisfies (14), then we have from the Dual feasibility that the Lagrangian problem to minimise $L(\boldsymbol{x},\boldsymbol{\mu}^*)$ over $\boldsymbol{x} \in X$ is a Lagrangian relaxation of (12). We also have that $\boldsymbol{x}^*$ solves (12) according to the Lagrangian optimality. The Primal feasiblity shows that $\boldsymbol{x}^*$ is feasible in (12), and the Complementary slackness implies that $L(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}^*)$. We now use the Relaxation Theorem which gives us that $\boldsymbol{x}^*$ is optimal in (12), which implies that $\boldsymbol{\mu}^*$ is a Lagrange multiplies vector.

Contrarily, if $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ is a pair of optimal primal solution and Lagrange multiplier vector, then they are primal and dual feasible, respectively. The Lagrangian optimality and Complementary slackness follow from the Theorem of Lagrange multipliers and global optima. □

12 Existence and properties of optimal solutions

Theorem

Let $P := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{Ax} = \boldsymbol{b}; \boldsymbol{x} \geq \mathbf{0}^n\}$ and $V := \{\boldsymbol{v}^1, \ldots, \boldsymbol{v}^k\}$ be the set of extreme points of P. Furhter let $C =: \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{Ax} = \mathbf{0}^m; \boldsymbol{x} \geq \mathbf{0}^n\}$ and $D := \{\boldsymbol{d}^1, \ldots, \boldsymbol{d}^r\}$ be the set of extreme directions of C. Consider the linear program:

$$\begin{aligned} \text{minimise} \quad & z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\ \text{subject to} \quad & \boldsymbol{x} \in P. \end{aligned} \tag{15}$$

(a) This problem has a finite optimal solution iff P is nonempty and z is lower bounded on P, i.e. P is nonempty and $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{d}^j \geq 0 \,\forall\, \boldsymbol{d}^j \in D$.

(b) If the problem has a finite optimal solution, then $\exists$ an optimal solution among the extreme points.

Proof

(a) Let $\boldsymbol{x} \in P$. Then it follows from the Representation Theorem that

$$\boldsymbol{x} = \sum_{i=1}^{k} \alpha_i \boldsymbol{v}^i + \sum_{j=1}^{r} \beta_j \boldsymbol{d}^j, \tag{16}$$

for some $\alpha_1, \ldots, \alpha_k \geq 0$ such that $\sum_{i=1}^{k} \alpha_i = 1$, and $\beta_1, \ldots, \beta_r \geq 0$. Then it follows that

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = \sum_{i=1}^{k} \alpha_i \boldsymbol{c}^{\mathrm{T}} \boldsymbol{v}^i + \sum_{j=1}^{r} \beta_j \boldsymbol{c}^{\mathrm{T}} \boldsymbol{d}^j. \tag{17}$$

If we let $\boldsymbol{x}$ vary over P, the value of z will correspond to variations of the weights α_i and β_j. The first term in the right-hand side of (17) is bounded as $\sum_{i=1}^{k} \alpha_i = 1$. The second term is lower bounded as $\boldsymbol{x}$ varies over P iff $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{d}^j \geq 0$ holds $\forall \boldsymbol{d}^j \in D$, since otherwise we could let $\beta \to +\infty$ for an index j with $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{d}^j < 0$, and get that $z \to -\infty$. If $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{d}^j \geq 0 \; \forall \boldsymbol{d}^j \in D$, then it is clearly optimal to choose $\beta_j = 0$ for $j = 1, \ldots, r$. It remains to search for the optimal solution in the convex hull of V.

(b) Assume that $\boldsymbol{x} \in P$ is an optimal solution and let $\boldsymbol{x}$ be represented as in (16). From the above we have that we can choose $\beta_1 = \cdots = \beta_r = 0$, so we can assume that

$$\boldsymbol{x} = \sum_{i=1}^{k} \alpha_i \boldsymbol{v}^i.$$

Further let

$$a \in \underset{i \in \{1, \ldots, k\}}{\text{arg minimum}} \; \boldsymbol{c}^{\mathrm{T}} \boldsymbol{v}^i.$$

Then,

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{v}^a = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v}^a \sum_{i=1}^{k} \alpha_i = \sum_{i=1}^{k} \alpha_i \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v}^a \leq \sum_{i=1}^{k} \alpha_i \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v}^i = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},$$

that is, the extreme point $\boldsymbol{v}^a$ is a global minimum. $\square$

13 Finiteness of the Simplex method

Theorem

If all of the basic feasible solutions (BFS) are non-degenerate, then the simplex algorithm terminates after a finite number of iterations.

Proof

A BFS is non-degenerate $\Rightarrow$ it has exactly m positive components, and hence has a unique associated basis. In this case, in the minimum ratio test,

$$\mu^* = \underset{i \in \{k \mid (\boldsymbol{B}^{-1}\boldsymbol{N}_j)_k > 0\}}{\text{minimum}} \frac{(\boldsymbol{B}^{-1}\boldsymbol{b})_i}{(\boldsymbol{B}^{-1}\boldsymbol{N}_j)_i} > 0.$$

Therefore, at each iteration the objective value decreases and hence a BFS that has appeared once can never reappear. Further, from a Corollary it follows that the number of extreme points, hence the number of BFS, is finite. $\square$

14 Strong Duality Theorem

Consider the primal linear program (P):

$$\begin{aligned} \text{minimise} \quad & z = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \\ & \boldsymbol{x} \geq \boldsymbol{0}^n, \end{aligned}$$

and its dual linear program (D):

$$\begin{aligned} \text{maximise} \quad & w = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}, \\ \text{subject to} \quad & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \leq \boldsymbol{c}. \\ & \boldsymbol{y} \text{ free}, \end{aligned}$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}, \boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{c} \in \mathbb{R}^n$.

Theorem

If the primal problem (P) and the dual problem (D) have feasible solutions, then there exists optimal solutions to (P) and (D), and their optimal objective function values are equal.

Proof

Since the dual (D) is feasible it follows from the Weak Duality Theorem that the objective function value of (P) is bounded from below. Hence the Theorem of Existence and properties of optimal solutions implies that there exists an optimal BFS, $\boldsymbol{x}^* = (\boldsymbol{x}_B^{\mathrm{T}}, \boldsymbol{x}_N^{\mathrm{T}})^{\mathrm{T}}$, to (P). We construct an optimal solution to (D):

$$(\boldsymbol{y}^*)^{\mathrm{T}} := \boldsymbol{c}_B^{\mathrm{T}}\boldsymbol{B}^{-1}.$$

Since $\boldsymbol{x}^*$ is an optimal BFS the reduced costs of the non-basic variables are non-negative, which gives that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}^* \leq \boldsymbol{c}$. Hence $\boldsymbol{y}^*$ is feasible to (D). Further we have that

$$\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}^* = \boldsymbol{b}^{\mathrm{T}}(\boldsymbol{B}^{-1})^{\mathrm{T}}\boldsymbol{c}_B = \boldsymbol{c}_B^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{b} = \boldsymbol{c}_B^{\mathrm{T}}\boldsymbol{x}_B = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^*,$$

so by a Corollary we have that $\boldsymbol{y}^*$ is an optimal solution to (D). $\square$

15 Complementary Slackness Theorem (I)

Theorem

Let $\boldsymbol{x}$ be a feasible solution to (P) and $\boldsymbol{y}$ a feasible solution to (D). Then

$$\left.\begin{array}{r}\boldsymbol{x} \text{ optimal to (P)}\\ \boldsymbol{y} \text{ optimal to (D)}\end{array}\right\} \Leftrightarrow x_j(c_j - \boldsymbol{A}_{\cdot j}^{\mathrm{T}}\boldsymbol{y}) = 0, \quad j = 1, \ldots, n, \tag{18}$$

where $\boldsymbol{A}_{\cdot j}$ is the j^{th} column of $\boldsymbol{A}$.

Proof

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are feasible we have that

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \geq (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y})^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}. \tag{19}$$

Further, by the Strong Duality Theorem and the Weak Duality Theorem, $\boldsymbol{x}$ and $\boldsymbol{y}$ are optimal iff $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}$, so (19) holds with equality, i.e.,

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y})^{\mathrm{T}}\boldsymbol{x} \Leftrightarrow \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}) = 0.$$

Since $\boldsymbol{x} \geq \boldsymbol{0}^n$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \leq \boldsymbol{c}$, $\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}) = 0$ is equivalent to each term in the sum being zero, that is (17) holds. $\square$

16 Complementary Slackness Theorem (II)

Often the Complementary Slackness Theorem is stated for the primal-dual pair given by:

$$\begin{aligned} \text{maximise} \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}. \\ & \boldsymbol{x} \geq \boldsymbol{0}^n, \end{aligned} \tag{20}$$

$$\begin{aligned} \text{minimise} \quad & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}, \\ \text{subject to} \quad & \boldsymbol{A}^{\mathrm{T}} \geq \boldsymbol{c}. \\ & \boldsymbol{y} \geq \boldsymbol{0}^m, \end{aligned} \tag{21}$$

Theorem

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be feasible solutions to (20) and (21), respectively. Then $\boldsymbol{x}$ and $\boldsymbol{y}$ are optimal to (20) and (21), respectively, iff

$$\begin{aligned} x_j(c_j - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}_{.j}) = 0, &\quad j = 1,\dots,n, \\ y_i(\boldsymbol{A}_{i.}\boldsymbol{x} - b_i) = 0, &\quad i = 1,\dots,m, \end{aligned}$$

where $\boldsymbol{A}_{.j}$ is the j^{th} column of A and $\boldsymbol{A}_{i.}$ the i^{th} row of $\boldsymbol{A}$. The proof is similar to that of Complementary Slackness Theorem (I).

17 Global convergence of a penalty method

Theorem

Consider the constrained problem:

$$\begin{aligned} &\text{minimise} \quad f(\boldsymbol{x}) \\ &\text{subject to} \quad \boldsymbol{x} \in S, \end{aligned} \tag{22}$$

where $S \subseteq \mathbb{R}^n$ is a nonempty, closed set and $f : \mathbb{R}^n \to \mathbb{R}$ is a given differentiable function.

Assume that this problem has optimal solutions. Then every limit point of the sequence $\{\boldsymbol{x}_\nu^*\}, \nu \to +\infty$, of globally optimal solutions to

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimise}} f(\boldsymbol{x}) + \nu\tilde{\chi}_S(\boldsymbol{x}) \tag{23}$$

is globally optimal in the problem (22). In other words,

$$\left.\begin{aligned} \boldsymbol{x}_\nu^* \text{ globally optimal in (23)} \\ \boldsymbol{x}_\nu^* \to \boldsymbol{x}^* \text{ as } \nu \to +\infty \end{aligned}\right\} \Rightarrow \boldsymbol{x}^* \text{ globally optimal in (22).}$$

Proof

Let $\boldsymbol{x}^*$ denote an arbitrary globally optimal solution to (22). From the inequality

$$f(\boldsymbol{x}_\nu^*) + \nu\tilde{\chi}_S(\boldsymbol{x}_\nu^*) \leq f(\boldsymbol{x}^*), \tag{24}$$

and from a Lemma (penalization constitutes a relaxation) we obtain uniform bounds on the penalty term $\nu\tilde{\chi}_S(\boldsymbol{x}_\nu^*)\ \forall\, \nu \geq 1$:

$$0 \leq \nu\tilde{\chi}_S(\boldsymbol{x}_\nu^*) \leq f(\boldsymbol{x}^*) - f(\boldsymbol{x}_1^*).$$

Thus $\tilde{\chi}_S(\boldsymbol{x}_\nu^*) \to 0$ as $\nu \to +\infty$ and every limit point of the sequence $\{\boldsymbol{x}_\nu^*\}$ must be feasible in (22), due to the continuity of $\tilde{\chi}_S$.

Now let $\hat{\boldsymbol{x}}$ denote an arbitrary limit point of $\{\boldsymbol{x}_\nu^*\}$, that is,

$$\lim_{k\to\infty} \boldsymbol{x}_{\nu_k}^* = \hat{\boldsymbol{x}},$$

for some sequence $\{\nu_k\}$ converging to infinity. Then, we have the following chain of inequalities:

$$f(\hat{\boldsymbol{x}}) = \lim_{k\to+\infty} f(\boldsymbol{x}_{\nu_k}^*) \leq \lim_{k\to+\infty} \{f(\boldsymbol{x}_{\nu_k}^*) + \nu_k \tilde{\chi}_S(\boldsymbol{x}_{\nu_k}^*)\} \leq f(\boldsymbol{x}^*),$$

where the last inequality follows from (24). However, due to the feasiblity of $\hat{\boldsymbol{x}}$ in (22) the reverse inequality $f(\boldsymbol{x}^*) \leq f(\hat{\boldsymbol{x}})$ must also hold. The two inequalities combined imply the required claim. $\square$