

Nonlinear Optimization (TMA947)

Edoardo Mangia

Contents

1	Convexity	2
1.1	Farkas' Lemma	2
1.2	TODO Farkas' Lemma for an inequality system	2
1.3	Convexity Characterization in C^1	3
2	Primal Optimality Conditions	4
2.1	The Fundamental Theorem of Global Optimality	4
2.2	Necessary Optimality Conditions, C^1 case	4
2.3	Necessary and Sufficient Global Optimality Conditions	4
2.4	The Separation Theorem	5
3	Chapter 5: Primal-Dual Optimality Conditions	5
3.1	Karush-Kuhn-Tucker necessary conditions	5
3.2	Sufficiency of the Karush-Kuhn-Tucker Conditions for Convex Problems	5
4	Lagrangian Duality	6
4.1	Relaxation Theorem	6
4.2	Weak Duality Theorem	7
4.3	Global Optimality Conditions in the Absence of a Duality Gap	7
5	Linear Programming Models	7
5.1	Existence and Properties of Optimal Solutions	7
6	The Simplex Method	8
6.1	Finiteness of the Simplex Method	8
7	LP Duality and Sensitivity Analysis	9
7.1	Weak Duality Theorem	9
7.2	Strong Duality Theorem	9
7.3	Complementarity Slackness Theorem 1	10
7.4	Complementarity Slackness Theorem 2	10
8	Constrained Optimization	10
8.1	Global Convergence of a Penalty Method	10

1 Convexity

1.1 Farkas' Lemma

Theorem 1.1 (Farkas' Lemma). *Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then, exactly one of the systems*

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

and

$$\begin{aligned} A^T \pi &\leq 0^n, \\ b^T \pi &> 0, \end{aligned} \tag{2}$$

has a feasible solution, and the other system is inconsistent.

Farkas' Lemma says that:

a) Either b lies in the cone spanned by the columns of A , i.e.,

$$b = Ax, \quad \text{for some } x \geq 0,$$

b) or, b does not lie in the cone (and can then be separated from the cone).

Farkas' Lemma is crucial for LP theory and optimality conditions.

Proof. Assume that system (1) has a solution x , then

$$b^T y = x^T A^T y > 0$$

But $x \geq 0^n \Rightarrow A^T y \leq 0^n$ cannot hold \Rightarrow system (2) is infeasible.

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

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

and its dual problem:

$$\begin{aligned} &(0^n)^T x \\ &Ax = b, \\ &x \geq 0^n. \end{aligned}$$

$y = 0^m$ 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. ■

1.2 TODO Farkas' Lemma for an inequality system

Theorem 1.2 (Farkas' Lemma for an inequality system).

1.3 Convexity Characterization in C^1

Theorem 1.3 (Convexity characterization in C^1). *Let $f \in C^1$ on an open convex set S*

$$f \text{ is convex on } S \iff f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}), \text{ for all } x, \bar{x} \in S.$$

This theorem basically says that every tangent plane to the graph of f lies on, or below, the epigraph of f , or that each first-order approximation of f lies below f .

Another equivalent way of writing this is the following: $f \in C^1$ is convex on the open, convex set S if and only if

$$[\nabla f(x) - \nabla f(y)]^T(x - y) \geq 0$$

- the gradient of f is monotone on S , or

- the angle between $\nabla f(x) - \nabla f(y)$ and $x - y$ should be between $-\pi/2$ and $\pi/2$.

Another leaner way to write it is:

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

a) f is convex on $S \iff f(y) \geq f(x) + \nabla f(x)^T(y - x), \forall x, y \in S$.

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

Proof.

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

$$\begin{aligned} \lambda f(x^1) + (1 - \lambda)f(x^2) &\geq f(\lambda x^1 + (1 - \lambda)x^2) \\ &\iff \end{aligned}$$

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

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

$$f(x^1) - f(x^2) \geq \nabla f(x^2)^T(x^1 - x^2).$$

The result follows.

$[\Leftarrow]$ We have that:

$$f(x^1) \geq f(\lambda x^1 + (1 - \lambda)x^2) + (1 - \lambda)\nabla f(\lambda x^1 + (1 - \lambda)x^2)^T(x^1 - x^2), \quad f(x^2) \geq f(\lambda x^1 + (1 - \lambda)x^2) + \lambda\nabla f(\lambda x^1 + (1 - \lambda)x^2)^T(x^2 - x^1)$$

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:

$$\begin{aligned} f(y) &\geq f(x) + \nabla f(x)^T(y - x), \quad x, y \in S, \\ f(x) &\geq f(y) + \nabla f(y)^T(x - y), \quad x, y \in S. \end{aligned}$$

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

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

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

where $x = \lambda x^1 + (1 - \lambda)x^2, \lambda \in (0, 1)$.

By assumption we have that

$$[\nabla f(x) - \nabla f(x^1)]^T(x - x^1) \geq 0, \text{ so } (1 - \lambda)[\nabla f(x) - \nabla f(x^1)]^T(x^2 - x^1) \geq 0.$$

From this it follows that $\nabla f(x)^T(x^2 - x^1) \geq \nabla f(x^1)^T(x^2 - x^1)$.

By using this inequality and (3), we get that $f(x^2) \geq f(x^1) + \nabla f(x^1)^T(x^2 - x^1)$.

This completes the proof. ■

2 Primal Optimality Conditions

2.1 The Fundamental Theorem of Global Optimality

Theorem 2.1 (The Fundamental Theorem of Global Optimality). *Consider the problem:*

$$\begin{array}{ll} \text{minimise} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in S, \end{array}$$

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

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

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

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

2.2 Necessary Optimality Conditions, C^1 case

Theorem 2.2 (Necessary Optimality Conditions, C^1 case). *Suppose that $S \subseteq \mathbb{R}^n$ and that $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is in C^1 around a point $x \in S$ for which $f(x) < +\infty$.*

a) *If $x^* \in S$ is a local minimum of f over S , then*

$$\nabla f(x^*)^T p \geq 0$$

holds for every feasible direction p at x^ .*

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

$$\nabla f(x^*)^T (x - x^*) \geq 0, \quad x \in S \tag{2}$$

Proof. .

a) We begin by Taylor expanding f around x^* :

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

The proof is by contradiction. We know from a proposition that if there is a direction p for which it holds that $\nabla f(x^*)^T p < 0$, then $f(x^* + \alpha p) < f(x^*)$ for all sufficiently small $\alpha > 0$.

Here it suffices to state that p should be a feasible direction in order to reach a contradiction to the local optimality of x^* .

b) If S is convex, then it holds that $\forall x \in S, p := x - x^*$ is a feasible direction.

Then a) \Rightarrow (2). ■

2.3 Necessary and Sufficient Global Optimality Conditions

Theorem 2.3 (Necessary and Sufficient Global Optimality Conditions). *Suppose that S is convex and that f in C^1 is convex. Then*

$$x^* \text{ is a global minimum of } f \text{ over } S \iff x \text{ is stationary}$$

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

Proof. .

\Rightarrow This has been shown in the previous theorem, since a global minimum is a local minimum.

\Leftarrow The convexity of f yields that $\forall y$ in S ,

$$f(y) \geq f(x^*) + \nabla f(x^*)^T (y - x^*) \geq f(x^*),$$

where the second inequality stems from the second part of (5). ■

2.4 The Separation Theorem

Theorem 2.4 (The Separation Theorem). *Suppose that $S \subseteq \mathbb{R}^n$ is closed and convex, and that the point y does not lie in S . Then there exists a vector $\pi \neq 0$ and a scalar $\alpha \in \mathbb{R}$ such that $\pi^T y > \alpha$, and $\pi^T x \leq \alpha$ for all $x \in S$.*

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

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

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

3 Chapter 5: Primal-Dual Optimality Conditions

3.1 Karush-Kuhn-Tucker necessary conditions

Theorem 3.1 (Karush-Kuhn-Tucker necessary conditions). *Assume that at a given point $x^* \in S$ Abadie's constraint qualification holds. If $x^* \in S$ is a local minimum of f over S , then $\exists \mu \in \mathbb{R}^m$ such that:*

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

In other words,

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

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

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

and $T_S(x^*)$ is the tangent cone.

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

Now construct a matrix A with columns $\nabla g_i(x^*), i \in \mathcal{I}(x)$. Then, the system $A^T p < 0^{|\mathcal{I}(x^*)|}$ and $-\nabla f(x^*)^T p > 0$ has no solutions. $|\mathcal{I}(x^*)|$ denotes the cardinality of the set $\mathcal{I}(x^*)$.

By Farkas' Lemma the system $A\xi = -\nabla f(x^*), \xi \geq 0^{|\mathcal{I}(x^*)|}$ has a solution.

Define the vector $\mu_{\mathcal{I}(x^*)} = \xi$, and $\mu_i = 0$ for $i \notin \mathcal{I}(x^*)$. Then μ verifies the KKT conditions in (6). ■

3.2 Sufficiency of the Karush-Kuhn-Tucker Conditions for Convex Problems

Theorem 3.2 (Sufficiency of the Karush-Kuhn-Tucker Conditions for Convex Problems). *Assume that the problem*

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

with the feasible set S given by:

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

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 $x^* \in S$ the KKT conditions:

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

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

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

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

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

and using the affinity of the functions h_j we get that

$$-\nabla h_j(x^*)^T(x - x^*) = h_j(x^*) - h_j(x) = 0, \quad \forall j = 1, \dots, l.$$

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

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

Since x was arbitrary, x^* solves the problem. ■

4 Lagrangian Duality

4.1 Relaxation Theorem

Theorem 4.1 (Relaxation Theorem). *Given the problem to find*

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

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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

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

where $f_R : \mathbb{R}^n \rightarrow \mathbb{R}$ 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, x_R^* , for which it holds that

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

then 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

$$f(x_R^*) = f_R(x_R^*) \leq f_R(x) \leq f(x), \quad x \in S,$$

from which the result follows. ■

4.2 Weak Duality Theorem

Theorem 4.2 (Weak Duality Theorem). *Consider the following two problems:*

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

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

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

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

Let x and $\boldsymbol{\mu}$ be feasible in problems (12) and (13), respectively.

Then $q(\boldsymbol{\mu}) \leq f(x)$ and in particular, $q^* < f^*$.

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

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

so

$$q^* = \supremum_{\boldsymbol{\mu} > 0^m} q(\boldsymbol{\mu}) \leq \infimum_{x \in X: g(x) \leq 0^m} f(x) = f^*$$

The result follows. ■

4.3 Global Optimality Conditions in the Absence of a Duality Gap

Theorem 4.3 (Global Optimality Conditions in the Absence of a Duality Gap). *The vector $(x^*, \boldsymbol{\mu}^*)$ is a pair of primal optimal solution and Lagrange multiplier vector iff*

$$\begin{aligned} \boldsymbol{\mu}^* &\geq 0^m, \quad (\text{Dual Feasibility}) \\ x^* &\in \arg \min_{x \in X} L(x^*, \boldsymbol{\mu}^*), \quad (\text{Lagrangian Optimality}) \\ x^* &\in X, \quad g(x^*) \leq 0^m, \quad (\text{Primal Feasibility}) \\ \mu_i^*(x^*) &= 0, \quad i = 1, \dots, m, \quad (\text{Complementary Slackness}) \end{aligned} \tag{14}$$

Proof. Suppose that the pair $(x^*, \boldsymbol{\mu}^*)$ satisfies (14), then we have from the Dual feasibility that the Lagrangian problem to minimise $L(x, \boldsymbol{\mu}^*)$ over $x \in X$ is a Lagrangian relaxation of (12).

We also have that x^* solves (12) according to the Lagrangian optimality. The Primal feasibility shows that x^* is feasible in (12), and the Complementary slackness implies that $L(x^*, \boldsymbol{\mu}^*) = f(x^*)$.

We now use the Relaxation Theorem which gives us that x^* is optimal in (12), which implies that $\boldsymbol{\mu}^*$ is a Lagrange multiplies vector. Contrarily, if $(x^*, \boldsymbol{\mu}^*)$ 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. ■

5 Linear Programming Models

5.1 Existence and Properties of Optimal Solutions

Theorem 5.1 (Existence and Properties of Optimal Solutions). *Let $P := \{x \in \mathbb{R}^n \mid Ax = b; x \geq 0^n\}$ and $V := \{v^1, \dots, v^k\}$ be the set of extreme points of P . Further let $C := \{x \in \mathbb{R}^n \mid Ax = 0^m; x \geq 0^n\}$ and $D := \{d^1, \dots, d^r\}$ be the set of extreme directions of C . Consider the linear program:*

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

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 $c^T d^j \geq 0 \forall d^j \in D$.

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

Proof. .

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

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

for some $\alpha_1, \dots, \alpha_k \geq 0$ such that $\sum_i \alpha_i = 1$, and $\beta_1, \dots, \beta_r \geq 0$. Then it follows that

$$c^T x = \sum_{i=1}^k \alpha_i c^T v^i + \sum_{j=1}^r \beta_j c^T d^j. \quad (17)$$

If we let 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 $\sum_i \alpha_i = 1$. The second term is lower bounded as x varies over P iff $c^T d^j \geq 0 \forall d^j \in D$, since otherwise we could let $\beta_j \rightarrow +\infty$ for an index j with $c^T d^j < 0$, and get that $z \rightarrow -\infty$.

If $c^T d^j \geq 0 \forall d^j \in D$, then it is clearly optimal to choose $\beta_j = 0$ for $j = 1, \dots, r$. It remains to search for the optimal solution in the convex hull of V .

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

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

Further let

$$a \in \arg \min_{i \in \{1, \dots, k\}} c^T v^i.$$

Then,

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

that is, the extreme point v^a is a global minimum. ■

6 The Simplex Method

6.1 Finiteness of the Simplex Method

Theorem 6.1 (Finiteness of the Simplex Method). *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,

$$\mu^* = \min_{i \in \{k | (B^{-1}N_j)_k > 0\}} \frac{(B^{-1}b)_i}{(B^{-1}N_j)_i} > 0.$$

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. ■

7 LP Duality and Sensitivity Analysis

7.1 Weak Duality Theorem

Theorem 7.1 (Weak Duality Theorem). *Consider the following two problems:*

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

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

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

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

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

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

$$q(\boldsymbol{\mu}) = \infimum L(x, \boldsymbol{\mu}) \leq f(x) + \boldsymbol{\mu}^T g(x) \leq f(x),$$

so

$$q^* = \supremum_{\boldsymbol{\mu} \geq \mathbf{0}^m} q(\boldsymbol{\mu}) \leq \infimum_{x \in X: g(x) \leq \mathbf{0}^m} f(x) = f^*$$

The result follows. ■

7.2 Strong Duality Theorem

Theorem 7.2 (Strong Duality Theorem). *Consider the primal linear program (P)*

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

and its dual linear program (D):

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

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

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, $x^* = x_B^T, x_N^T)^T$, to (P). We construct an optimal solution to (D):

$$(y^*)^T := c_B^T B^{-1}.$$

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

$$b^T y^* = b^T (B^{-1})^T c_B = c_B^T B^{-1} b = c_B^T x_B = c^T x^*,$$

so by a Corollary we have that y^* is an optimal solution to (D). ■

7.3 Complementarity Slackness Theorem 1

Theorem 7.3 (Complementarity Slackness Theorem 1). *Let x be a feasible solution to (P) and y a feasible solution to (D). Then*

$$\left. \begin{array}{l} x \text{ optimal to (P)} \\ y \text{ optimal to (D)} \end{array} \right\} \Leftrightarrow x_j(c_j - A_{\cdot j}^T y) = 0, \quad j = 1, \dots, n, \quad (18)$$

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

Proof. If x and y are feasible we have that

$$c^T x \geq (A^T y)^T x = y^T A x = b^T y. \quad (19)$$

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

$$c^T x = (A^T y)^T x \Leftrightarrow x^T (c - A^T y) = 0.$$

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

7.4 Complementarity Slackness Theorem 2

Theorem 7.4 (Complementarity Slackness Theorem 2). *Often the Complementary Slackness Theorem is stated for the primal-dual pair given by:*

$$\begin{array}{ll} \text{maximise} & c^T x, \\ \text{subject to} & Ax \leq b, \\ & x \geq 0^n, \\ \text{minimise} & b^T y, \\ \text{subject to} & A^T y \geq c, \\ & y \geq 0^m, \end{array}$$

Let x and y be feasible solutions to (20) and (21), respectively. Then x and y are optimal to (20) and (21), respectively, iff

$$x_j(c_j - y^T A_{\cdot j}) = 0, \quad j = 1, \dots, n, \quad y_i(A_{i \cdot} x - b_i) = 0, \quad i = 1, \dots, m,$$

where $A_{\cdot j}$ is the j^{th} column of A and $A_{i \cdot}$ the i^{th} row of A .

Proof. The proof is similar to that of Complementarity Slackness Theorem (1). ■

8 Constrained Optimization

8.1 Global Convergence of a Penalty Method

Theorem 8.1 (Global Convergence of a Penalty Method). *Consider the constrained problem:*

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

where $S \subseteq \mathbb{R}^n$ is a nonempty, closed set and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a given differentiable function. Assume that this problem has optimal solutions. Then every limit point of the sequence $\{x_\nu^*\}, \nu \rightarrow +\infty$, of globally optimal solutions to

$$\min_{x \in \mathbb{R}^n} f(x) + \nu \tilde{\chi}_S(x) \quad (23)$$

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

$$\left. \begin{array}{l} x_\nu^* \text{ globally optimal in (23)} \\ x_\nu^* \rightarrow x^* \text{ as } \nu \rightarrow +\infty \end{array} \right\} \Rightarrow x^* \text{ globally optimal in (22)}.$$

Proof. Let x^* denote an arbitrary globally optimal solution to (22). From the inequality

$$f(x_\nu^*) + \nu \tilde{\chi}_S(x_\nu^*) \leq f(x^*), \quad (24)$$

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

$$0 \leq \nu \tilde{\chi}_S(x_\nu^*) \leq f(x^*) - f(x_1^*).$$

Thus $\tilde{\chi}_S(x_\nu^*) \rightarrow 0$ as $\nu \rightarrow +\infty$ and every limit point of the sequence $\{x_\nu^*\}$ must be feasible in (22), due to the continuity of $\tilde{\chi}_S$. Now let \hat{x} denote an arbitrary limit point of $\{x_\nu^*\}$, that is,

$$\lim_{k \rightarrow \infty} x_{\nu_k}^* = \hat{x},$$

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

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

where the last inequality follows from (24).

However, due to the feasibility of \hat{x} in (22) the reverse inequality $f(x^*) \leq f(\hat{x})$ must also hold. The two inequalities combined imply the required claim. \blacksquare