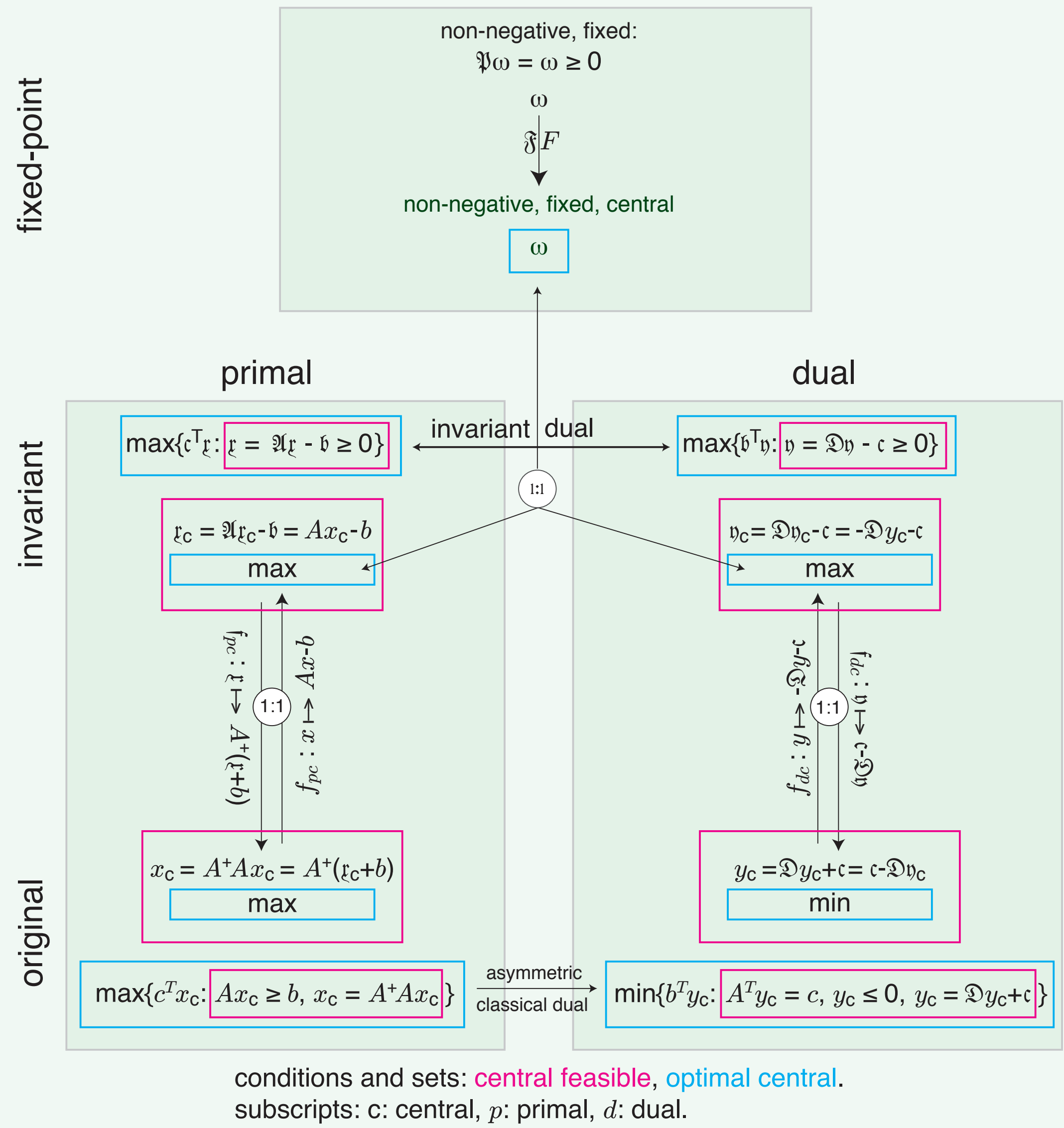
Optimization by the Fixed-Point Method

Edition 2.17 - Free Version



Jalaluddin Abdullah

Optimization by the Fixed-Point Method

©Jalaluddin Abdullah

March 16, 2016

Publication: Optimization by the Fixed-Point Method
2.17 - Free Version

Petaling Jaya, Selangor, Malaysia

A version of this book, available from the author at email address jalaludn@gmail.com, includes working hypertext links, 3D diagrams, links to Geogebra documents, and programming.

Foreword

The author's first became interested in the linear programming problem because of the publicity given to Kamarkar's method of solution, however the author believed that the problem is essentially one of linear algebra, and thus merited a solution as linear algebraic as possible, and that none of the modern approaches known to the author appeared "linear" enough.

While carrying out research for a Ph.D. at the University of Birmingham, the author came across the fixed-point method of Pyle and Cline and its elegant formulation by Bruni. After showing the method to be classificatory in nature, the author took this approach further, developing both a vector lattice and a regression algorithm for solution of the fixed-point problem and thus the linear optimisation problem.

The book is a continuation of the thesis research, putting the solution to the linear problem on a precise basis, and then generalising the fixed-point method to the non-linear, convex case, and providing necessary and sufficient conditions for an optimum, alternative to those of Kuhn and Tucker.

In this book division by zero is set to zero (this does not break any field axiom) which appears to be appropriate in the context of linear algebra; functions are written on the right in the manner of category theorists. The book has many hyperlinks, and some of the figures are three dimensional native to the PDF document, with accompanying links to three dimensional Geogebra code.

The book provides a precise solution to the linear problem, while the convex problem analysis may be regarded as work in progress.

Jalaluddin Abdullah
@ Julian Graham Morris
jalaludn@gmail.com
2015/08/11

Preface

In the name of God, the Benificent, the Merciful

This book is on the fixed-point method of solving linear and convex optimization problems. It begins with a brief history of the linear programming problem in Chapter 1, followed by required background theory in Chapter 2, introducing some general results from function theory, matrix algebra, regression theory, affine spaces and duality theory. Chapter 3 is on fixed-point theory; in this section of the chapter the notion of a swapping matrix is introduced - this matrix acts on idempotent symmetric matrices and is used in the construction of what is called a semi-unitary matrix which has linear properties in the region close to fixed-points associated with solutions to the LP problem. The fixed-points of the semi-unitary matrix and convergence to such fixed-points is investigated.

The classificatory stage of an investigation is regarded in the literature of research methodology as the first stage of a proper investigation however, since classification in the work of Pyle, Cline, and finally Bruni (where it finds its simplest and most concise expression) is implicit, it might have escaped operations researchers that this stage of the investigation of the LP problem has been completed rather elegantly, and that the resulting fixed-point formulation was therefore particularly worthy of further analysis. For this reason Bruni's results are carefully derived in Chapter 4, making explicit the classificatory nature of this approach. The duality theorem of von Neumann [23], and the Karush-Kuhn-Tucker Theorem [19, *op cit*] provide the basis for the derivation. Chapter 4 shows that the work of these authors can be regarded as creating equivalence classes of similar linear programs and a canonical representative for each such class.

In Chapter 5 the primal and dual invariant problems are combined to form a fixed-point problem - the problem of finding the non-negative fixed-points of an Hermitian matrix. This approach combines both the invariant primal and invariant dual in one balanced equation.

In Chapter 6 the key notion of proximality is introduced and it is shown that, in the context of the LP fixed-point problem, proximality implies linear behaviour; this lays the groundwork for the construction of a solution method which combines a converging series approach with regression to ensure termination after a finite number of steps. In the following Chapter 8 the solution method is applied to a number of LP problems.

Finally, Chapter 9 contains a generalization of the approach to non-linear convex optimization problems. Operational conditions for an optimum are derived which differ from the Kuhn-Tucker conditions.

Scheme functions for the computing used in the book are available for the commercial version.

Appendix A contains a spectral analysis of the product of projection matrices, being relevant to the convergence analysis of Chapter 3.2.1.

Appendix B is on the derivation of the pseudo-inverse of a transportation problem matrix.

The programming system DrRacket, is used throughout the book and all example programs run under it; the system is available for OS X, Windows, various Linux's, and as source code.

The book has grown out of research carried out in the Department of Economics at the University of Birmingham, England under Dr. Aart Heesterman leading to a thesis entitled *Fixed Point Algorithms for Linear Programming* [1]. A number of staff at Universiti Teknologi Malaysia assisted the author while writing an earlier version of the book and in this respect I am particularly grateful for the use of library facilities at Perpustakaan Sultanah Zanariah, to Dr. Yusof Yaakob (Publishing), Encik Aziz Yaakob in the Computer Centre, and to Hj. Mohammad Shah, Dr. Ali Abdul Rahman for checking an earlier draft, and Encik Muhammad Hisyam Lee Wee Yee in the Mathematics Department, for support in getting LaTeX running well, to Allahyarham Professor Mohd. Rashidi bin Md. Razali (late of the Mathematics Department) for encouragement, to Dr. Ibrahim Ngah and Abdul Razak and others who have built up and maintained the Resource Centre of Fakulti Alam Bina, and to all those who have developed and maintained the TeX typesetting system including Andrew Trevorrow for OzTeX, Richard Koch *et al* for TeXShop, and Christiaan M. Hofman Adam R. Maxwell and Michael O. McCracken for Skim.

Jalaluddin Abdullah @ Julian Graham Morris

March, 2016

jalaludn@gmail.com

Contents

Notation

Changes in notation between chapters are indicated in the introductions to chapters. Blocks of text are sometimes marked by brackets. These are used at the beginning and end points of sub-sections of proofs.

Matrices and operators are generally represented by uppercase letters, and vectors by lower-case letters which may be boldface.

Some symbols' meaning changes, generally from general to specific; such symbols have more than one entry in the following table.

Symbol	Description/Definition	Chapter	and Verse
$\mathbf{1}_i$	an $i \times 1$ vector of 1's	7	78
$\mathbf{1}$	a vector of 1's	6.4	72
$\overline{\alpha}$	complex conjugate	2.2.4	9
$\exists$	*there exists*	throughout	
$\Rightarrow$	*implies*	throughout	
$\subseteq$	*implies*	6.2	6.2.2
$\subseteq$	*subset of*	throughout	
$< f, g >$	function pair	2.1	6
A	LP coefficient matrix	1.1	1
$\mathfrak{A}$	invariant LP matrix	4.1	46
$\mathcal{A}$	reciprocal space $\mathcal{A} = P\mathcal{B}$	A.1	102
$\boldsymbol{b}$	LP constraint vector	1.1	1
$\mathfrak{b}$	invariant LP constraint vector	4.1	46
$\mathcal{B}$	reciprocal space $\mathcal{B} = K\mathcal{A}$	A.1	102
$\boldsymbol{c}$	LP objective vector	1.1	1
$\mathfrak{c}$	invariant LP objective vector	4.1	46
d	discriminant	5.2.3	5.10
$\mathfrak{D}$	invariant dual matrix	4.1	46
$\boldsymbol{g}_i$	orbit of $P\mathbf{1}$	3.2.1	E 3.11
$\mathfrak{g}_i$	orbit of $\mathfrak{P}\mathbf{1}$	6.4	E 6.14
G	P or K	A.1	105
$\tilde{G}$	K or P respectively	A.1	105
K	Karush matrix	3	30
$\overline{K}$	$K(I - S)$ oblique Karush matrix	3.1.4	E 3.8
$\mathfrak{K}_z$	specific Karush matrix	3	68
$\mathfrak{K}$	$\boldsymbol{z} \mapsto \mathfrak{K}_z\boldsymbol{z}$	6.1.4	E 6.13
$\overline{\mathfrak{K}}_z$	specific oblique Karush matrix	6.1.3	E 6.7
$\mathcal{K}$	space $\{Kx : \boldsymbol{x} \in \Re^m\}$	A.1	102
$\mathcal{L}$	row space	2.2.2	2.2.2
m	number of rows of A	1.1	1
n	number of columns of A	1.1	1
$\mathcal{N}$	null space	2.2.2	9
$\boldsymbol{p}$	non-negative fixed-point	6.4	72
P	fixed-point matrix	2.2.5	32
P_X	XX^+	2.2.6.3	E 2.2.23 a
P'_X	$I - XX^+$	2.2.6.3	E 2.2.23 b
$\mathfrak{P}$	specific fixed-point matrix	5.1	E 5.4
$\mathcal{P}$	space $\{Px : \boldsymbol{x} \in \Re^m\}$	A.1	102
$\overline{Q}$	$Q(I - S)$ oblique projection	3.1.1	E 3.3

$\mathcal{R}$	column space	2.2.2	2.2.2
S	swapping matrix	3.1.1	30
$\mathfrak{S}$	specific swapping matrix	5	67
U	unitary and semi-unitary definition	2.2.5	10
U	general semi-unitary matrix	3.1.4	E 3.1.9
$\mathfrak{U}_z$	specific semi-unitary matrix	6.1.4	D 6.8
$\mathfrak{U}$	$\boldsymbol{z} \mapsto \mathfrak{U}_z \boldsymbol{z}$	6.1.4	E 6.13
$\boldsymbol{\mathfrak{v}}_i$	converging orbit of $\boldsymbol{\mathfrak{P}1}$	6.4	E 6.15
V	averaging matrix	3.1.5	D 3.1.12
$\mathfrak{V}_z$	specific averaging matrix	6.1.4	E 6.9
$\mathfrak{V}$	$\boldsymbol{z} \mapsto \mathfrak{V}_z \boldsymbol{z}$	6.1.4	E 6.13
$\boldsymbol{x}$	primal solution	1.1	1
$\boldsymbol{\mathfrak{x}}$	invariant primal solution	4.3	E 4.5
$\vee$	logical OR	throughout	
$\wedge$	logical AND	throughout	
$\vee$	vector lattice operator	throughout	
$\wedge$	vector lattice operator	throughout	
$\boldsymbol{y}$	asymmetric dual solution	4.2	47
$\boldsymbol{\mathfrak{y}}$	invariant dual solution	4.3	48
$\boldsymbol{z}$	$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}$	throughout	
$\boldsymbol{\mathfrak{z}}$	$\begin{bmatrix} \boldsymbol{\mathfrak{x}} \\ \boldsymbol{\mathfrak{y}} \end{bmatrix}$	throughout	
superscript T	matrix transpose	2.2.4	9
superscript +	Moore-Penrose generalized inverse	2.2.6	10
superscript *	operator involution	2.2.4	9

List of Figures

Chapter 1

Introduction

1.1 Problem Statement

Operations research (OR) involves the study and control of activities and has thus been the object of considerable analysis throughout man's history. One of the important problems to have been delineated is the linear programming (LP) problem, which involves maximizing a linear expression subject to linear inequality constraints is:

Given an m by n matrix A, and vectors $\boldsymbol{b}$ and $\boldsymbol{c}$,

$$\textit{find the vectors } \boldsymbol{x} \textit{ which maximize } \mathbf{c}^T\boldsymbol{x}, \textit{ subject to } A\boldsymbol{x} \geq \boldsymbol{b}. \tag{1.1}$$

The more than or equal sign signifies that each LP constraint is of this form; thus we are concerned essentially with homogeneous systems of inequalities where each line of the matrix equation has a binary relation of the form $\geq$. The case of equality, where the i^{th} constraint is of the form $\boldsymbol{a}_i \cdot \boldsymbol{x} = \boldsymbol{b}_i$, where $\boldsymbol{a}_i$ denotes the i^{th} row of A, can be allowed for by including the additional line $\boldsymbol{a}_i \cdot \boldsymbol{x} \leq \boldsymbol{b}_i$, or by solving the equality for one of the variables and substituting it out, or (in the context of this book) by projecting it out - which may on occasions be the computationally most stable approach.

Note that in this book, "optimal" and "solution" both connote "feasible and the best possible" - the term "optimal feasible problem" is dispensed with and instead the term "bounded problem" is used in context, the thinking behind this being that if a vector is regarded as optimal then it is feasible, and it doesn't make good sense to talk about a bound on an empty set.

There are a number of formulations of the LP problem (including the one above) which can be shown to be equivalent:

$$\begin{array}{ll} \max\ \{\boldsymbol{c}^T\boldsymbol{x} : A\boldsymbol{x} \le \boldsymbol{b}\} & \text{(a)} \\ \min\ \{\boldsymbol{c}^T\boldsymbol{x} : A\boldsymbol{x} \ge \boldsymbol{b}\} & \text{(b)} \\ \max\ \{\boldsymbol{c}^T\boldsymbol{x} : A\boldsymbol{x} \le \boldsymbol{b}, \boldsymbol{x} \ge \boldsymbol{0}\} & \text{(c)} \\ \min\ \{\boldsymbol{c}^T\boldsymbol{x} : A\boldsymbol{x} \ge \boldsymbol{b}, \boldsymbol{x} \ge \boldsymbol{0}\} & \text{(d)} \\ \max\ \{\boldsymbol{c}^T\boldsymbol{x} : A\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \ge \boldsymbol{0}\} & \text{(e)} \\ \min\ \{\boldsymbol{c}^T\boldsymbol{x} : A^T\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \ge \boldsymbol{0}\} & \text{(f)} \end{array} \tag{1.2}$$

For details refer to [33, p. 91].

1.2 Historical Context

1.2.1 Origins

The origins of the linear programming problem date back to Johann Bernoulli in 1717, who was working on theoretical mechanics. He introduced the concept of a virtual velocity: Given a region R, the vector $\boldsymbol{y}$ is a virtual velocity in the point $\boldsymbol{x}^* \in R$ if there exists a curve C in R starting in $\boldsymbol{x}$ such that the half line $\{\boldsymbol{x}^* + \lambda\boldsymbol{y} : \lambda \ge 0\}$ is tangent to C. Assuming that

$$\begin{array}{l} \textit{for any virtual velocity } \boldsymbol{y}, \\ \textit{the vector } -\boldsymbol{y} \textit{ is also a virtual velocity} \end{array} \tag{1.3}$$

the virtual velocity principle states that a mass point at $\boldsymbol{x}^*$ subject to force $\boldsymbol{b}$ is in equilibrium iff $\boldsymbol{b}^T\boldsymbol{y} = 0$ (refer to Figure 1.1).

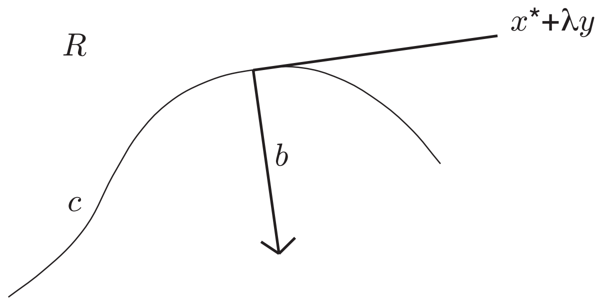


Figure 1.1: **Johann Bernoulli's Virtual Velocity**

Lagrange [20] observed that if the region is given by

$$R = \{\boldsymbol{x} : f_1(\boldsymbol{x}) = \cdots = f_n(\boldsymbol{x}) = 0\}\ , \tag{1.4}$$

where $f_i : \Re^n \mapsto \Re$ and the gradients $\bigtriangledown f_i,\ i = 1, \ldots, n$ are linearly independent, then condition 1.3 is satisfied and the principle of virtual velocity can be written as

$$\bigtriangledown f_1(\boldsymbol{x}^*)^T\boldsymbol{y} = \cdots = \bigtriangledown f_n(\boldsymbol{x}^*)^T\boldsymbol{y} = 0\ ,$$

where

$$\bigtriangledown f_i = \begin{bmatrix} \partial f_i / \partial x_1 \\ \vdots \\ \partial f_i / \partial x_n \end{bmatrix}.$$

As was observed by Fourier [11], the condition $A\boldsymbol{x} \leq \boldsymbol{b}$ defines a convex set, while the objective function $\boldsymbol{c}^T\boldsymbol{x}$ can be regarded as defining a set of parallel hyperplanes. The optimum solution has the hyperplane from this set with the maximum value of $\boldsymbol{c}^T\boldsymbol{x}$ which osculates the convex set of feasible solutions.

Fourier defined the feasible region along the lines of Equation 1.4, but allowed inequality constraints. He generalized the principle of virtual velocity to:

> A mass point in $\boldsymbol{x}^*$ subject to force $\boldsymbol{b}$, whose position is restricted to R, is in equilibrium iff $\boldsymbol{b}^T\boldsymbol{y} \leq 0$.

Thus Fourier's contribution to the analysis of optimization problems goes beyond simply contributing directly in the form of a solution algorithm, as he extended the work of Johann Bernoulli and Lagrange on optimization subject to continuously differentiable constraints to the case where a function is subject to inequality constraints. He also pioneered the link between the theory of inequalities and the theory of polyhedra, thus spanning the algebraic and geometric perspectives. Fourier appears to be the first to suggest the geometric idea of navigating between vertices along the edges of the convex polyhedron associated with the feasible solution set, until an optimum is reached.

The first solution of the linear programming problem is also due to Fourier; it is now known as Fourier-Motzkin elimination, and is of considerable analytic importance as a generalization of the procedure is used as a tool for proving results leading to Farkas' Lemma, and the transposition theorems of Gordan and Stiemke [19]. (Refer also to [34, Ch. 1])

Poussin in the years 1910-11 designed the algebraic analogue of Fourier's geometric method for solving the linear programming problem [26]. Dantzig in 1951 gave this algebra an efficient tableau form and called it the Simplex Method [9].

1.2.2 Inefficiency of the Simplex Method

The branch of mathematics which considers the efficiency of algorithms is called complexity analysis; we introduce a few simple notions from this branch at this point.

The basic arithmetic operations are taken to be addition, multiplication, and division. If the number of basic arithmetic operations required for an algorithm to compute the answer to a problem increases as a polynomial in the parameters which describe the size of the problem, then the problem is said to be *polynomial-time*.

Under certain conditions the Simplex Method is not efficient - it is possible for the current solution to do a “tour” of many or even all of the vertices of the polyhedron defined by the feasible set of solutions [18], or to cycle repeatedly through a sequence of vertices without reaching an optimum. In fact the Simplex Method is not a polynomial-time algorithm, and so in recent times (the last forty years or so) there has been research into new algorithms. To be fair, the problem of cycling can be removed, and the non polynomial-time nature appears not to affect certain “average” problems. For implementation of the simplex method one may refer to [12].

Through this renewed research a number of algorithms have been developed which are polynomial-time, including Khachian’s ellipsoid method [17], Karmarkar’s simplex method [16], and Renegar’s centreing algorithm [30].

In summary, interest in the LP problem has not abated, for a number of reasons, which are detailed:

1. The non polynomial-time nature of the Simplex Method: This classical and most common method of solution, is now known to be an inefficient algorithm, due to the publication in 1972 of an example where the algorithm tours all the vertices of the feasible set, showing that it is not a polynomial-time algorithm [18]. The desire to solve larger problems, and this discovery of deficiency in the Simplex Method have lead to new efforts to find efficient ways of solving LPs, and to a resurgence in analysis of the problem.

2. The ability to represent a number of OR problems as LP’s - for example an important early problem in OR, the transportation problem, can be written as a linear program.

3. The advent of parallel processing machines and the possibility that new algorithms may be suitable for hybrid analog-digital machines.

1.3 Applications of Linear Programming

The formulation of practical problems as linear programs dates back at least to work by G.B. Dantzig in 1947. He published his work *Programming in a Linear Structure* [8] the following year. Actual formulation of problems as linear programs can confidently be assumed to date back to at least the Second World War.

An example of problems which can be formulated as an LP is

1.3.1 The Product Mix Problem

Mashino Corporation produces mountain bikes in n variants $v_1, \cdots, v_n$. Each variant sells at a fixed price, leading to profits $s_1, \cdots, s_n$. To produce one unit of variant i, l_i man hours of labour are needed and m_i resource units.

The corporation is constrained in that it has only l man-hours available, and can deliver at most m units of material per hour.

We see that labour required to produce x_i units of variant i is $l_i x_i$, so total labour requirement is $l_1 x_1 + \cdots + l_n x_n \leq l$. Similarly material requirement is $m_1 x_1 + \cdots + m_n x_n \leq m$, and profit is $s_1 x_1 + \cdots + s_n x_n$. The company wishes to maximize profit, so the problem can be written concisely as

$$\begin{aligned}
&\textit{Find } x_1, \cdots, x_n \textit{ which maximize } s_1 x_1 + \cdots + s_n x_n \textit{ subject to} \\
&l_1 x_1 + \cdots + l_n x_n \leq l \textit{ and } m_1 x_1 + \cdots + m_n x_n \leq m, \qquad (1.5) \\
&\textit{where } x_1, \cdots, x_n \geq 0.
\end{aligned}$$

Setting $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \boldsymbol{x}$, $\begin{bmatrix} l_1 \cdots l_n \\ m_1 \cdots m_n \end{bmatrix} = A$, $\begin{bmatrix} \boldsymbol{l} \\ \boldsymbol{m} \end{bmatrix} = \boldsymbol{b}$, and $\begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} = \boldsymbol{c}$,

shows that the problem 1.5 can be formulated as a linear program.

A similar problem involves arriving at the most economical purchase decision for animal feed, given the dietary requirements of the animals and the prices and composition of the feed brands.

Chapter 2

Background Theory

2.1 Function-Pairs

A function is said to be surjective, or onto, if its image is equal to its codomain; a function is said to be bijective if it is one-to-one and onto - that is 1:1 and surjective.

Definition: 2.1.1 Two functions are said to form a **function-pair** if the domain of each contains the range of the other function.

We will write $< f, \mathfrak{f} >$ if f together with $\mathfrak{f}$ forms a function pair.

Definition: 2.1.2 A element is said to be **central** in the domain of one function of a function-pair if the element is in the range of the other function of the pair.

Definition: 2.1.3 The function pair $< f, \mathfrak{f} >$ is said to be **regular** if $f\mathfrak{f}f = f$ and $\mathfrak{f}f\mathfrak{f} = \mathfrak{f}$. (This usage is motivated by semigroup nomenclature. [7])

Definition: 2.1.4 Given a function-pair $< f, \mathfrak{f} >$ with $f : X \to \mathfrak{X}, \mathfrak{f} : \mathfrak{X} \to X, X_c = \mathfrak{X}\mathfrak{f}$ and $\mathfrak{X}_c = Xf$, functions of the form $f_c = f_{|X_c}$ with codomain $\mathfrak{X}_c$, and $\mathfrak{f}_c = \mathfrak{f}_{|\mathfrak{X}_c}$ with codomain X_c are called **central functions**, and function-pairs of the form $< f_c, \mathfrak{f}_c >$ are called **central function-pairs**.

Lemma 2.1.5 Given a regular function-pair $< f, \mathfrak{f} >$ with $f : X \to \mathfrak{X}, \mathfrak{f} : \mathfrak{X} \to X, X_c = \mathfrak{X}\mathfrak{f}$ and $\mathfrak{X}_c = Xf$,
(a) $f_c = f_{|X_c}$ is bijective,
(b) $\mathfrak{f}_c = \mathfrak{f}_{|\mathfrak{X}_c}$ is bijective,
(c) $X_c = \mathfrak{X}_c\mathfrak{f}$,
(d) $\mathfrak{X}_c = X_c f$, and
(e) f_c and $\mathfrak{f}_c$ are mutual inverses.
(f) $x \in X$ is central iff $x = xf\mathfrak{f}; X_c = Xf\mathfrak{f}$,
(g) $\mathfrak{x} \in \mathfrak{X}$ is central iff $\mathfrak{x} = \mathfrak{x}\mathfrak{f}f; \mathfrak{X}_c = \mathfrak{X}\mathfrak{f}f$,

Proof (a) f_c is 1:1: $(x_1, x_2 \in X_c) \wedge (x_1 f = x_2 f) \Rightarrow (x_1 = \mathfrak{x}_1\mathfrak{f}) \wedge (x_2 = \mathfrak{x}_2\mathfrak{f}) \wedge (x_1 f = x_2 f) \Rightarrow (x_1 = \mathfrak{x}_1\mathfrak{f}) \wedge (x_2 = \mathfrak{x}_2\mathfrak{f}) \wedge (\mathfrak{x}_1\mathfrak{f}f = \mathfrak{x}_2\mathfrak{f}f) \Rightarrow (x_1 = \mathfrak{x}_1\mathfrak{f}) \wedge (x_2 = \mathfrak{x}_2\mathfrak{f}) \wedge (\mathfrak{x}_1\mathfrak{f}f\mathfrak{f} = \mathfrak{x}_2\mathfrak{f}f\mathfrak{f}) \Rightarrow (x_1 = \mathfrak{x}_1\mathfrak{f}) \wedge (x_2 = \mathfrak{x}_2\mathfrak{f}) \wedge (\mathfrak{x}_1\mathfrak{f} = \mathfrak{x}_2\mathfrak{f}) \Rightarrow x_1 = x_2$; f_c is onto $\mathfrak{X}_c$: $X_c f_c = X_c f \subseteq Xf = \mathfrak{X}_c = Xf = Xf\mathfrak{f}f \subseteq \mathfrak{X}\mathfrak{f}f = X_c f = X_c f_c$, that is $X_c f_c = Xf = \mathfrak{X}_c$, so f_c is bijective. (b) The proof is similar to (a), (c) $X_c = \mathfrak{X}\mathfrak{f} = \mathfrak{X}\mathfrak{f}f\mathfrak{f} \subseteq Xf\mathfrak{f} = \mathfrak{X}_c\mathfrak{f} \subseteq \mathfrak{X}\mathfrak{f} = X_c$, so $X_c = \mathfrak{X}_c\mathfrak{f}$, (d) the proof is similar to (c). (e) For

$x \in X_c, x = \mathfrak{x}\mathfrak{f}\ \exists \mathfrak{x} \in \mathfrak{X} \Rightarrow x f_c \mathfrak{f}_c = \mathfrak{x}\mathfrak{f} f_c \mathfrak{f}_c = \mathfrak{x}\mathfrak{f} f \mathfrak{f}_c = \mathfrak{x}\mathfrak{f} f \mathfrak{f} = \mathfrak{x}\mathfrak{f} = x$, so $f_c \mathfrak{f}_c$ is the identity map on X_c, similarly it is found that $\mathfrak{f}_c f_c$ is the identity map on $\mathfrak{X}_c$, so f and $\mathfrak{f}$ are mutual inverses. (f) x is central $\Leftrightarrow x = \mathfrak{x}\mathfrak{f}\ \exists \mathfrak{x} \in \mathfrak{X} \Rightarrow (xf = \mathfrak{x}\mathfrak{f}f) \wedge (x = \mathfrak{x}\mathfrak{f})\ \exists \mathfrak{x} \in \mathfrak{X} \Rightarrow (xf\mathfrak{f} = \mathfrak{x}\mathfrak{f}f\mathfrak{f} = \mathfrak{x}\mathfrak{f}) \wedge (x = \mathfrak{x}\mathfrak{f})\ \exists \mathfrak{x} \in \mathfrak{X} \Rightarrow x = xf\mathfrak{f} \Rightarrow x = \mathfrak{x}\mathfrak{f}$ where $\mathfrak{x} = xf \Leftrightarrow x$ is central, that is x is central iff $x = xf\mathfrak{f}$. (g) proof is exactly similar to that of (f) □

Restating just (f) and (g) of Lemma 2.1.5:

Corollary: 2.1.6 Given the regular function-pair $< f, \mathfrak{f} >$ with $f : X \to \mathfrak{X}$ and $\mathfrak{f} : \mathfrak{X} \to X$,
(a) $\boldsymbol{x}$ is central iff $\boldsymbol{x} = \boldsymbol{x} f\mathfrak{f}$.
(b) $\boldsymbol{\mathfrak{x}}$ is central iff $\boldsymbol{\mathfrak{x}} = \boldsymbol{\mathfrak{x}}\mathfrak{f}f$.

In the following lemma we are interested in conditions expressed as sub-domains of a regular function-pair, where the image under either function of any point satisfying the condition corresponding to the sub-domain satisfies the condition corresponding to the sub-domain of the other function of the pair.

Lemma 2.1.7 Given the regular function-pair $< f, \mathfrak{f} >$ with $f : X \to \mathfrak{X}, \mathfrak{f} : \mathfrak{X} \to X$, $X_1 \subseteq X, \mathfrak{X}_1 \subseteq \mathfrak{X}, X_1 f \subseteq \mathfrak{X}_1$, and $\mathfrak{X}_1 \mathfrak{f} \subseteq X_1$,
(a) $(X_1 \cap \mathfrak{X}\mathfrak{f})f = \mathfrak{X}_1 \cap Xf$
(b) $(\mathfrak{X}_1 \cap Xf)\mathfrak{f} = X_1 \cap \mathfrak{X}\mathfrak{f}$
(c) $f_1 = f : X_1 \cap \mathfrak{X}\mathfrak{f} \to \mathfrak{X}_1 \cap Xf$ is a bijection
(d) $\mathfrak{f}_1 = \mathfrak{f} : \mathfrak{X}_1 \cap Xf \to X_1 \cap \mathfrak{X}\mathfrak{f}$ is a bijection
(e) f_1 and $\mathfrak{f}_1$ are mutual inverses.

Proof (a) and (b): $(X_1 \cap \mathfrak{X}\mathfrak{f})f \subseteq X_1 f \cap \mathfrak{X}\mathfrak{f}f \subseteq X_1 f \cap Xf \subseteq \mathfrak{X}_1 \cap Xf$,
that is

$$(X_1 \cap \mathfrak{X}\mathfrak{f})f \subseteq \mathfrak{X}_1 \cap Xf, \tag{2.1}$$

and similarly

$$(\mathfrak{X}_1 \cap Xf)\mathfrak{f} \subseteq X_1 \cap \mathfrak{X}\mathfrak{f}, \tag{2.2}$$

So

$$(X_1 \cap \mathfrak{X}\mathfrak{f})f\mathfrak{f} \subseteq (\mathfrak{X}_1 \cap Xf)\mathfrak{f},$$

and

$$(\mathfrak{X}_1 \cap Xf)\mathfrak{f}f \subseteq (X_1 \cap \mathfrak{X}\mathfrak{f})f$$

Now $f\mathfrak{f}$ and $\mathfrak{f}f$ are 1:1 on $\mathfrak{X}\mathfrak{f}$ and Xf respectively, so

$$\mathfrak{X}_1 \cap Xf \subseteq (X_1 \cap \mathfrak{X}\mathfrak{f})f, \tag{2.3}$$

and

$$X_1 \cap \mathfrak{X}\mathfrak{f} \subseteq (\mathfrak{X}_1 \cap Xf)\mathfrak{f}, \tag{2.4}$$

From (2.1) and (2.3)

$$(X_1 \cap \mathfrak{X}\mathfrak{f})f = \mathfrak{X}_1 \cap Xf,$$

and from (2.2) and (2.4)

$$(\mathfrak{X}_1 \cap Xf)\mathfrak{f} = X_1 \cap \mathfrak{X}\mathfrak{f},$$

(c): that f_1 is well defined follows from (a), that it is 1:1 follows from it being a restriction of f, that it is onto also follows from (a), so it is a bijection. (d) the proof is similar to that of (c). (e) follows from f and $\mathfrak{f}$ being mutual inverses. □

2.2 Matrices

Matrices are rectangular arrays of numbers, for example $\begin{bmatrix} 1 & 2 & 5 \\ -1 & 4 & 7 \end{bmatrix}$.

We define multiplication of a vector v by a matrix A as follows:

$$vM = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} v_1a_{11} + \cdots + v_ma_{m1} & v_1a_{12} + \cdots + v_ma_{m2} & \cdots & v_1a_{1n} + \cdots + v_ma_{mn} \end{bmatrix}$$

If we define the function $f_A : v \mapsto vA$ we find that f_A is a linear map from $\Re^m$ to $\Re^n$; moreover any linear map from $\Re^m$ to $\Re^n$ is of the form $f_X : v \mapsto vX$ for some $m \times n$ matrix X and the relationship $X \to f_X$ is a surjection from the set of $m \times n$ matrices to the set of linear maps from $\Re^m$ to $\Re^n$.

We define AB, for conformable matrices A and B, by

$$f_{AB} = f_A \circ f_B$$

With this definition of matrix multiplication it is immediately apparent that such multiplication is associative because it corresponds with function composition which is associative; with matrix inversion corresponding to function inversion it is also obvious that a matrix will only have an inverse if its corresponding linear function has an inverse.

2.2.1 The Identity

A right identity I_r satisfies $XI_r = X$ for all X. while a left identity satisfies $I_lX = X$ for all X. So if we have a left identity L and a right identity R then $LR = R$ and $LR = L$, so $L = R$. Thus the matrix I with 1's on its diagonal and zeros elsewhere is the unique identity.

2.2.2 Spaces

The **column space** of the $m \times n$ matrix A is the space of m dimensional column vectors

$$\mathcal{R}(A) = \{A\boldsymbol{x} : \boldsymbol{x} \in \Re^n\} \ .$$

The **row space** of the $m \times n$ matrix A is the set of n dimensional row vectors

$$\mathcal{L}(A) = \{\boldsymbol{x}A : \boldsymbol{x} \in \Re^m\} \ .$$

The **null space** of the $m \times n$ matrix A is the space of m dimensional column vectors

$$\mathcal{N}(A) = \{\boldsymbol{x} : \boldsymbol{x} \in \Re^n, A\boldsymbol{x} = 0 \ .\}$$

Not that $\mathcal{N}(A) = \{(I - A^+A)\boldsymbol{y} : \boldsymbol{y} \in \Re^n\}$, using the yet to be defined Moore-Penrose pseudo-inverse.

2.2.3 The Inverse

The matrix R is a right inverse of A if

$$AR = I \tag{2.5}$$

The matrix L is a left inverse of A if

$$LA = I \tag{2.6}$$

From Equation 2.5 we have $LAR = L$ and from Equation 2.6 we have $LAR = R$ and so $L = R$, thus if a matrix A has both a left and a right inverse then it has a unique inverse which we denote by A^{-1} and we say that A is **regular**. Note that for A and B regular $(B^{-1}A^{-1})AB = B^{-1}(A^{-1}A)B = B^{-1}B = I$ and $AB(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$, so $(AB)^{-1} = B^{-1}A^{-1}$ for A and B regular. Thus the set of regular matrices is closed under multiplication.

2.2.4 Involution

With $\overline{\alpha}$ denoting the complex conjugate of α - that is, for $a, b \in \Re, \overline{a + ib} = a - ib$, the matrix transpose, or more generally, an involution, $*$, satisfies

1. $(A^*)^* = A$,
2. $(A + B)^* = A^* + B^*$,
3. $(AB)^* = B^*A^*$,
4. $(\alpha A)^* = \overline{\alpha}A^*$.

A matrix X is said to be *Hermitian* if $X^* = X$ [31, p. 178]; a matrix X is said to be *idempotent* if $X^2 = X$. A matrix X is said to be *unitary* if $XX^* = X^*X = I$.

The involution used in this book is the matrix transpose, denoted by a superscript T.

We identify a column vector with a matrix having only one column; we define $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^T\boldsymbol{x}$.

Lemma 2.2.1 $\boldsymbol{y} = X\boldsymbol{y} \Leftrightarrow (\|\boldsymbol{y}\|^2 = \|X\boldsymbol{y}\|^2 \text{ and } \boldsymbol{y}^T\boldsymbol{y} = \boldsymbol{y}^T X\boldsymbol{y})$.

Proof $\boldsymbol{y} = X\boldsymbol{y} \Rightarrow (\|\boldsymbol{y}\|^2 = \|X\boldsymbol{y}\|^2) \wedge (\boldsymbol{y}^T\boldsymbol{y} = \boldsymbol{y}^T X\boldsymbol{y})$

$$\Rightarrow (\|\boldsymbol{y} - X\boldsymbol{y}\|^2 = \|\boldsymbol{y}\|^2 - 2\boldsymbol{y}^T X\boldsymbol{y} + \|X\boldsymbol{y}\|^2) \wedge (\|\boldsymbol{y}\|^2 = \|X\boldsymbol{y}\|^2) \wedge (\boldsymbol{y}^T\boldsymbol{y} = \boldsymbol{y}^T X\boldsymbol{y})$$

$$\Rightarrow \|\boldsymbol{y} - X\boldsymbol{y}\|^2 = \|\boldsymbol{y}\|^2 - 2\boldsymbol{y}^T\boldsymbol{y} + \|\boldsymbol{y}\|^2 \Rightarrow \|\boldsymbol{y} - X\boldsymbol{y}\|^2 = 0 \Rightarrow \boldsymbol{y} = X\boldsymbol{y} \ . \ \square$$

2.2.5 Unitary and Semi-Unitary Matrices

The matrix X is said to be *unitary* if $XX^T = X^TX = I$.

The unitary matrices form a group under matrix multiplication since
$(XY)(XY)^T = XYY^TX^T = XX^T = I$, and $(XY)^T(XY) = Y^TX^TXY = Y^TY = I$.

A matrix X is said to be **semi-unitary** w.r.t. an Hermitian idempotent Q if $XX^T = X^TX = Q$; X is said to be Q-unitary.[1]

Lemma 2.2.2
(a) X is Q-unitary $\Leftrightarrow X^T$ is Q-unitary.
(b) X is Q-unitary $\Rightarrow XQ = QX = X$.
(c) The set of Q-unitary matrices of the same dimension is closed under matrix multiplication.

Proof

(a) X is Q-unitary $\Leftrightarrow XX^T = X^TX = Q \Leftrightarrow X^TX = XX^T = Q \Leftrightarrow X^T$ is Q-unitary

(b) X is Q-unitary $\Rightarrow XX^T = X^TX = Q \Rightarrow (XX^TX = QX) \wedge (X^TX = Q)$
$\Rightarrow XQ = QX; (QX - X)(QX - X)^T = (QX - X)(X^TQ^T - X^T)$
$= (QX - X)(X^T - X^T) = QXX^TQ - QXX^T - XX^TQ + XX^T$
$= Q^3 - Q^2 - Q^2 + Q = Q - Q - Q + Q = 0 \Rightarrow (QX - X)(QX - X)^T = 0$
$\Rightarrow QX = X$.

(c) Suppose X_1 and X_2 are Q-unitary matrices of the same dimension,
then $(X_1X_2)^T(X_1X_2) = X_2^TX_1^TX_1X_2 = X_2^TQX_2 = X_2^TX_2Q = Q^2 = Q$,
and $(X_1X_2)(X_1X_2)^T = X_1X_2X_2^TX_1^T = X_1QX_1^T = QX_1X_1^T = Q^2 = Q$. $\square$

2.2.6 The Moore-Penrose Pseudo-Inverse

The Moore-Penrose pseudo-inverse (MPPI) is used extensively in the following chapters; it will mostly be referred to a simply "the inverse". Given a matrix X, define

$$X^+ = \lim_{\delta \to 0}\{X^T(XX^T + \delta^2 I)^{-1}\}.$$

[1] Naimark [21] introduces the similar notion of partially isometric operators.

It can be shown that X^+ is well-defined [4, p. 19, Theorem 3.4]; it is called the Moore-Penrose pseudo-inverse (MPPI) of X.

A non-constructive definition is as follows:

Definition: 2.2.3 $Y = X^+$ if and only if

$$XYX = X, \quad \text{(a)}$$
$$YXY = Y, \quad \text{(b)}$$
$$(XY)^T = XY, \text{ and} \quad \text{(c)}$$
$$(YX)^T = YX. \quad \text{(d)}$$

The characterization given by this equation set is much employed for verifying that the inverse has been correctly computed.

2.2.6.1 Uniqueness

If the Moore-Penrose inverse of a matrix exists then it is unique, since if we assume Y and Z are two Moore-Penrose inverses of X, then $Y = YXY = Y(XY)^T = YY^TX^T = YY^T(X^TZ^TX^T) = Y(Y^TX^T)(Z^TX^T) = Y(XY)^T(XZ)^T = Y(XY)(XZ) = Y(XYX)Z = YXZ = YXZXZ$
$= (YX)^T(ZX)Z = (YX)^T(ZX)^TZ = X^TY^TX^TZ^TZ = (X^TY^TX^T)Z^TZ$
$= (XYX)^TZ^TZ = X^TZ^TZ = (ZX)^TZ = ZXZ = Z.$

Remark: 2.2.4 If a matrix X has an inverse then this inverse is the MPPI since 1. $XX^{-1}X = X(X^{-1}X) = XI = X$. 2. $X^{-1}XX^{-1} = X^{-1}(XX^{-1}) = X^{-1}I = X^{-1}$. 3. $XX^{-1} = I$ which is symmetric. 4. $X^{-1}X = I$ which is symmetric. □

Lemma 2.2.5 If X^+ exists then $X^{+T} = X^{T+}$

Proof Let $Y = X^{+T}$ then
1. $X^TYX^T = X^TX^{+T}X^T = (XX^+X)^T = X^T$,
2. $YX^TY = X^{+T}X^TX^{+T} = (X^+XX^+)^T = X^{+T} = Y$
3. $(X^TY)^T = (X^TX^{+T})^T = X^+X = (X^+X)^T = X^TX^{+T} = X^TY$,
4. $(YX^T)^T = (X^{+T}X^T)^T = XX^+ = (XX^+)^T = X^{+T}X^T = YX^T$,
so Y is the inverse of X^T that is $X^{+T} = X^{T+}$. □

2.2.6.2 Existence

We first define the MPPI for the field of real numbers as follows:

Definition: 2.2.6 for $x \in \Re, x^+ = \begin{cases} 1/x & \text{if} \quad x \neq 0 \\ 0 & \text{if} \quad x = 0 \end{cases}$

Next, for diagonal matrices - that is square matrices with non-zero entries only on the diagonal we define the inverse

$$\begin{bmatrix} d_1 & 0 & \cdots & 0 & 0 \\ 0 & d_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & d_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & d_n \end{bmatrix}^+ = \begin{bmatrix} d_1^+ & 0 & \cdots & 0 & 0 \\ 0 & d_2^+ & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & d_{n-1}^+ & 0 \\ 0 & 0 & \cdots & 0 & d_n^+ \end{bmatrix}$$

which can be seen to satisfy the four non-constructive conditions for the MPPI.

Further, we establish that any symmetric matrix has an MPPI, since such a matrix can be written in the form $Y = ZDZ^{-1}$ where $Z^{-1} = Z^T$ and D is a diagonal matrix, in which case $Y^\dagger = ZD^+Z^{-1}$ is an MPPI of Y since

1. $YY^\dagger Y = (ZDZ^{-1})(ZD^+Z^{-1})(ZDZ^{-1}) = ZD(Z^{-1}Z)D^+(Z^{-1}Z)DZ^{-1}$
   $= ZDD^+DZ^{-1} = ZDZ^{-1} = Y,$
2. $Y^\dagger YY^\dagger = (ZD^+Z^{-1})(ZDZ^{-1})(ZD^+Z^{-1}) = ZD^+(Z^{-1}Z)D(Z^{-1}Z)D^+Z^{-1}$
   $= ZD^+DD^+Z^{-1} = ZD^+Z^{-1} = Y^\dagger,$
3. $YY^\dagger = (ZDZ^{-1})(ZD^+Z^{-1}) = ZD(Z^{-1}Z)D^+Z^{-1} = ZDD^+Z^{-1} = ZDD^+Z^T,$
   which is symmetric,
4. $Y^\dagger Y = (ZD^+Z^{-1})(ZDZ^{-1}) = ZD^+(Z^{-1}Z)DZ^{-1} = ZD^+DZ^{-1} = ZD^+DZ^T,$
   which is symmetric.

So if Y is symmetric then $Y = ZDZ^{-1}$, and we define $Y^+ = ZD^+Z^{-1}$, which we have shown to be the MPPI of Y.

Summing up, we have

Lemma 2.2.7 If Y is a symmetric matrix then
(a) $\exists Z$, with $Z^{-1} = Z^T$ such that $Y = ZDZ^{-1}$
(b) $YZ = ZD$
(c) $Y^+ = ZD^+Z^{-1}$

Finally, we prove

Lemma 2.2.8 $X(X^TX)^+X^TX = X.$

Proof

$$
\begin{aligned}
&(X - X(X^TX)^+X^TX)^T(X - X(X^TX)^+X^TX)\\
&= (X^T - X^TX(X^TX)^{+T}X^T)(X - X(X^TX)^+X^TX)\\
&\overset{L\,2.2.5}{=} (X^T - X^TX(X^TX)^{T+}X^T)(X - X(X^TX)^+X^TX)\\
&= (X^T - X^TX(X^TX)^+X^T)(X - X(X^TX)^+X^TX)\\
&= X^TX - X^TX(X^TX)^+X^TX - X^TX(X^TX)^+X^TX + X^TX(X^TX)^+X^TX(X^TX)^+X^TX\\
&= X^TX - X^TX - X^TX + X^TX = 0, \text{ so } (X - X(X^TX)^+X^TX)^T(X - X(X^TX)^+X^TX) = 0\\
&\text{and therefor } X(X^TX)^+X^TX = X,
\end{aligned}
$$

and we can now prove existence for the general matrix X :

Lemma 2.2.9 $X^+ = (X^TX)^+X^T$

Proof Set $X^\dagger = (X^TX)^+X^T$ noting that $X^\dagger$ is well-defined since X^TX is symmetric and therefor has an MPPI;

$XX^\dagger X = X(X^TX)^+X^TX \overset{L\,2.2.8}{=} X$;

$X^\dagger XX^\dagger = (X^TX)^+X^TX(X^TX)^+X^T = (X^TX)^+X^T = X^\dagger$;

$(XX^\dagger)^T = (X(X^TX)^+X^T)^T = X(X^TX)^{+T}X^T = X(X^TX)^{T+}X^T = X(X^TX)^+X^T = XX^\dagger$;

$(X^\dagger X)^T = ((X^TX)^+X^TX)^T = (X^TX)^+X^TX = X^\dagger X$.

So $X^\dagger$ satisfies the non-constructive conditions and is therefor the MPPI of X. $\square$

Thus the MPPI always exists for matrices over the field of real numbers.

Note that $X(X^TX)^+X^T$ is the projection onto the column space of X since $(X(X^TX)^+X^T)^2 = X(X^TX)^+X^TX(X^TX)^+X^T = X(X^TX)^+(X^TX)(X^TX)^+X^T = X(X^TX)^+X^T$ and $(X(X^TX)^+X^T)^T = X(X^TX)^{+T}X^T = X(X^TX)^{T+}X^T = X(X^TX)^+X^T$ from which it follows that $X(X^TX)^+X^T$ is an orthogonal projection; moreover $X(X^TX)^+X^TX = X$ implies the range of the projection is at least the column space of X, while a typical element of the range of the projection is of the form $X(X^TX)^+X^T\boldsymbol{y} = X\boldsymbol{z}$ which is in the column space of X, so the range of the projection $X(X^TX)^+X^T$ is precisely the column space of X.

2.2.6.3 Results

In matrix calculations we identify column vectors with $n \times 1$ matrices, row vectors with $1 \times n$ matrices, and real numbers with 1×1 matrices. For 1×1 matrices we have, from the previous lemma, $[0]^+ = [0]$; so for 1×1 matrix $[a]$ we have $[a]^+ = [a^+]$.[2]

The following result is of general interest:

Lemma 2.2.10 (a) $(-X)^+ = -X^+$
(b) $X^{++} = X$

Proof

(a)
1. $(-X)(-X^+)(-X) = -XX^+X = -X$
2. $(-X^+)(-X)(-X^+) = -X^+XX^+ = -X^+$
3. $((-X)(-X^+))^T = (XX^+)^T = XX^+ = (-X)(-X^+)$
4. $((-X^+)(-X))^T = (X^+X)^T = X^+X = (-X^+)(-X)$

(b) Let $Y = X^+$ then $Y = X^+$ and
1. $XYX = X$
2. $YXY = Y$
3. $(XY)^T = XY$
4. $(YX)^T = YX$

$\Leftrightarrow$ $Y = X^+$ and
1. $YXY = Y$
2. $XYX = X$
3. $(YX)^T = YX$
4. $(XY)^T = XY$

$\Leftrightarrow$ $Y = X^+$ and $X = Y^+$ $\Rightarrow X^{++} = X$.

Lemma 2.2.11 *(a)* $(XX^T)^+ = X^{T+}X^+$
(b) $(X^TX)^+ = X^+X^{T+}$

[2] It appears that $a^{-1} = 0$ is undecidable and may therefor be appended as an axiom to the field laws.

Proof

1. $(XX^T)(X^{T+}X^+) = X(X^TX^{T+})X^+ = X(X^TX^{+T})X^+$
$= X(X^+X)^TX^+ = X(X^+X)X^+ = (XX^+X)X^+ = XX^+$ which is symmetric.

2. $(X^{T+}X^+)(XX^T) = X^{T+}(X^+X)X^T = X^{T+}(X^+X)^TX^T$
$= X^{T+}(X^TX^{+T})X^T = X^{T+}(X^TX^{T+}X^T) = X^{T+}X^T$ which is symmetric.

3. From 1 we have $(XX^T)(X^{T+}X^+) = XX^+$
$\Rightarrow (XX^T)(X^{T+}X^+)(XX^T) = (XX^+)(XX^T) = (XX^+X)X^T = XX^T$.

4. Again from 1 we have $(XX^T)(X^{T+}X^+) = XX^+$
$\Rightarrow (X^{T+}X^+)(XX^T)(X^{T+}X^+) = (X^{T+}X^+)(XX^+) = X^{T+}(X^+XX^+) = X^{T+}X^+$. □

Lemma 2.2.12
(a) $(XX^T)^+X = X^{+T}$
(b) $X(X^TX)^+ = X^{+T}$
(c) $X^T(XX^T)^+ = X^+$

Proof
(a) Replace X with X^T in Lemma 2.2.9
(b) Transpose Lemma 2.2.9.
(c) Transpose result (a). □

Lemma 2.2.13
(a) $X^TXX^+ = X^T$,
(b) $X^+XX^T = X^T$
(c) $XX^+X^{T+} = X^{T+}$,

Proof
(a) $X^TXX^+ = X^T(XX^+)^T = X^TX^{+T}X^T = (XX^+X)^T = X^T$.
(b) $X^+XX^T = (X^+X)^TX^T = X^TX^{+T}X^T = (XX^+X)^T = X^T$.
(c) $XX^+X^{T+} = (XX^+)^TX^{T+} = (XX^+)^TX^{+T} = (X^+XX^+)^T = X^{+T} = X^{T+}$. □

Lemma 2.2.14 $X^+Y = 0 \Leftrightarrow X^TY = 0$.

Proof
$X^+Y = 0 \Rightarrow XX^+Y = 0 \Rightarrow (XX^+)^TY = 0 \Rightarrow X^{+T}X^TY = 0$
$\Rightarrow X^{T+}X^TY = 0 \Leftrightarrow X^TX^{T+}X^TY = 0 \Rightarrow X^TY = 0 \Rightarrow X^{T+}X^TY = 0$
$\Rightarrow (XX^+)^TY = 0 \Rightarrow XX^+Y = 0 \Rightarrow X^+XX^+Y = 0 \Rightarrow X^+Y = 0$. □

2.2.6.4 Computation

The following result is required to derive the pseudo-inverse of matrices which occur in Chapter 2.4 and in Chapter 8.

Theorem: 2.2.15 If $C_{m+1} = [C_m|\boldsymbol{c}_{m+1}]$ then

(a) $$C^+_{m+1} = \left[\frac{C^+_m[I - \boldsymbol{c}_{m+1}\boldsymbol{k}^T_{m+1}]}{\boldsymbol{k}^T_{m+1}}\right],$$

where

(b) $$\boldsymbol{k}_{m+1} = \frac{(I - C_mC^+_m)\boldsymbol{c}_{m+1}}{\|(I - C_mC^+_m)\boldsymbol{c}_{m+1}\|^2},$$

if $(I - C_mC^+_m)\boldsymbol{c}_{m+1} \neq 0$,

(c) $$= \frac{C^{+T}_mC^+_m\boldsymbol{c}_{m+1}}{1 + \|C^+_m\boldsymbol{c}_{m+1}\|^2},$$

otherwise.

Proof By verification of the four non-constructive conditions above (refer to [4]). □

Lemma 2.2.16 If X is unitary Hermitian ($XX^T = X^TX = I, X^T = X$) then $X^+ = X$

Proof 1. $XX^TX = XXX = X$; 2. $X^TXX^T = XXX = X$; 3. XX^T is symmetric; 4. X^TX is symmetric, so $X^+ = X^T = X$.

Lemma 2.2.17 If X is idempotent Hermitian ($X^2 = X, X^T = X$) then $X^+ = X$.

2.2.6.5 Solution of Linear Equations

The idea which is implicit here is that if a concise solution of the general linear equation appears to require the pseudo-inverse, then perhaps a concise (and precise) solution of the linear programming problem might also require the pseudo-inverse.

The general linear equation can be written in matrix form as

$$A\boldsymbol{x} = \boldsymbol{b}$$

where A is an $m \times n$ matrix, $\boldsymbol{x}$ is an $n \times 1$ matrix (that is an n-dimensional column vector) and $\boldsymbol{b}$ is an $m \times 1$ matrix (that is an m-dimensional column vector). Now we write

$$A = [\boldsymbol{a}_1 \, \boldsymbol{a}_2 \, \cdots \boldsymbol{a}_n]$$

so the above equation becomes

$$[\boldsymbol{a}_1 \cdots \boldsymbol{a}_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} ,$$

that is

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b} \ .$$

Thus we are asking whether there is a linear combination of the vectors $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$ equal to $\boldsymbol{b}$.

Alternatively: does $\boldsymbol{b}$ lie in the space generated by $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$? There are two possibilities

1. $AA^+\boldsymbol{b} = \boldsymbol{b}$, that is $\boldsymbol{b}$ is in the space generated by $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$; further
   (a) $A^+A = I$, in which case $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$ are linearly independent and there is only one solution $\boldsymbol{x}_s = A^+\boldsymbol{b} = (A^TA)^{-1}A^T\boldsymbol{b}$ (the inverse exists!) or
   (b) $A^+A \neq I$, in which case $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$ are linearly dependent and there are multiple solutions of the form
   $\boldsymbol{x}_s = A^+\boldsymbol{b} + \mathcal{N}(A) = A^+\boldsymbol{b} + (I - A^+A)\boldsymbol{y}, \boldsymbol{y} \in \Re^n$.
2. $AA^+\boldsymbol{b} \neq \boldsymbol{b}$, in which case $A\boldsymbol{x} = \boldsymbol{b} \wedge AA^+\boldsymbol{b} \neq \boldsymbol{b} \Rightarrow AA^+A\boldsymbol{x} = AA^+\boldsymbol{b} \wedge AA^+\boldsymbol{b} \neq \boldsymbol{b} \Rightarrow A\boldsymbol{x} = AA^+\boldsymbol{b} \wedge AA^+\boldsymbol{b} \neq \boldsymbol{b} \Rightarrow A\boldsymbol{x} \neq \boldsymbol{b}$, that is there are no solutions. It is, however, possible to obtain a regression solution $\hat{\boldsymbol{x}} = A^+\boldsymbol{b}$ for which $A\hat{\boldsymbol{x}}$ is close to $\boldsymbol{b}$ - an approach which dates back to Gauss.

Case 1: $(AA^+)\boldsymbol{b} = \boldsymbol{b} \Rightarrow A(A^+\boldsymbol{b}) = \boldsymbol{b}$, so $A^+\boldsymbol{b}$ is certainly a solution. Further, if $\boldsymbol{y}$ is any other solution then $A\boldsymbol{y} = \boldsymbol{b} \Rightarrow AA^+A\boldsymbol{y} = AA^+\boldsymbol{b} \Rightarrow A\boldsymbol{y} = AA^+\boldsymbol{b} \Rightarrow A(\boldsymbol{y} - A^+\boldsymbol{b}) = \boldsymbol{0} \Rightarrow A^+A(\boldsymbol{y} - A^+\boldsymbol{b}) = \boldsymbol{0} \Rightarrow (I - A^+A)(\boldsymbol{y} - A^+\boldsymbol{b}) = \boldsymbol{y} - A^+\boldsymbol{b} \Rightarrow (I - A^+A)\boldsymbol{y} = \boldsymbol{y} - A^+\boldsymbol{b} \Rightarrow \boldsymbol{y} = A^+\boldsymbol{b} + (I - A^+A)\boldsymbol{y} \Rightarrow \boldsymbol{y} \in A^+\boldsymbol{b} + (I - A^+A)\boldsymbol{z}, \boldsymbol{z} \in \Re^n$. Conversely, $\boldsymbol{y} = A^+\boldsymbol{b} + (I - A^+A)\boldsymbol{z}, \exists \boldsymbol{z} \in \Re^n \Rightarrow A\boldsymbol{y} = AA^+\boldsymbol{b} + A(I - A^+A)\boldsymbol{z}, \exists \boldsymbol{z} \in \Re^n \Rightarrow A\boldsymbol{y} = AA^+\boldsymbol{b} \Rightarrow A\boldsymbol{y} = \boldsymbol{b}$. Thus in this case (i.e. $(AA^+)\boldsymbol{b} = \boldsymbol{b}$) the solutions are precisely of the form $A^+\boldsymbol{b} + (I - A^+A)\boldsymbol{z}, \exists \boldsymbol{z} \in \Re^n$. The exhaustive sub-cases (a) and (b) follow immediately.
Case 2. If there exists a solution $\boldsymbol{x}$ then $A\boldsymbol{x} = \boldsymbol{b} \Rightarrow (A^+A\boldsymbol{x} = A^+\boldsymbol{b}) \wedge (A\boldsymbol{x} = \boldsymbol{b}) \Rightarrow (AA^+A\boldsymbol{x} = AA^+\boldsymbol{b}) \wedge (A\boldsymbol{x} = \boldsymbol{b}) \Rightarrow (A\boldsymbol{x} = AA^+\boldsymbol{b}) \wedge (A\boldsymbol{x} = \boldsymbol{b}) \Rightarrow AA^+\boldsymbol{b} = \boldsymbol{b}$. Thus $AA^+\boldsymbol{b} \neq \boldsymbol{b}$ implies the problem has no solution. □

Specifically what is needed for our purposes is simply a summary of Case 1 above:

Lemma 2.2.18 $A\boldsymbol{x} = AA^+\boldsymbol{b} \Leftrightarrow \exists \boldsymbol{y} : \boldsymbol{x} = A^+\boldsymbol{b} + (I - A^+A)\boldsymbol{y}$

Proof
$$\begin{aligned}
&A\boldsymbol{x} = AA^+\boldsymbol{b}\\
&\Rightarrow \big(A\boldsymbol{x} = AA^+\boldsymbol{b}\big) \wedge \big(\boldsymbol{x} = A^+A\boldsymbol{x} + (I - A^+A)\boldsymbol{x}\big)\\
&\Rightarrow \boldsymbol{x} = A^+AA^+\boldsymbol{b} + (I - A^+A)\boldsymbol{x}\\
&\Rightarrow \boldsymbol{x} = A^+\boldsymbol{b} + (I - A^+A)\boldsymbol{x}\\
&\Rightarrow \exists \boldsymbol{y} : \boldsymbol{x} = A^+\boldsymbol{b} + (I - A^+A)\boldsymbol{y}\\
&\Rightarrow A\boldsymbol{x} = AA^+\boldsymbol{b}
\end{aligned}$$

Refer also to [15].

2.2.6.6 LP's with Equality Constraints

Consider the problem

$$\begin{array}{ll}
\textit{maximize} & \boldsymbol{c}^T\boldsymbol{x}\\
\textit{subject to} & A_1\boldsymbol{x} \geq \boldsymbol{b}_1\\
\textit{and} & A_2\boldsymbol{x} = \boldsymbol{b}_2
\end{array}$$

Since all solutions will be of the form $A_2^+\boldsymbol{b}_2 + (I - A_2^+A_2)\boldsymbol{y}$ we can write the objective function as

$$\boldsymbol{c}^T\big(A_2^+\boldsymbol{b}_2 + (I - A_2^+A_2)\boldsymbol{y}\big) = \big((I - A_2^+A_2)\boldsymbol{c}\big)^T\boldsymbol{y} + \boldsymbol{c}^TA_2^+\boldsymbol{b}_2$$

and the inequality as

$$A_1\big(A_2^+\boldsymbol{b}_2 + (I - A_2^+A_2)\boldsymbol{y}\big) \geq \boldsymbol{b}_1 \Rightarrow A_1A_2^+\boldsymbol{b}_2 + A_1(I - A_2^+A_2)\boldsymbol{y} \geq \boldsymbol{b}_1$$

$$\Rightarrow A_1(I - A_2^+A_2)\boldsymbol{y} \geq \boldsymbol{b}_1 - A_1A_2^+\boldsymbol{b}_2\ .$$

Thus the problem can be rewritten as

$$\begin{array}{ll}
\textit{maximize} & \boldsymbol{c}^T(I - A_2^+A_2)\boldsymbol{y} + \boldsymbol{c}^TA_2^+\boldsymbol{b}_2\\
\textit{subject to} & A_1(I - A_2^+A_2)\boldsymbol{y} \geq \boldsymbol{b}_1 - A_1A_2^+\boldsymbol{b}_2\ .
\end{array}$$

or, with the constant part of the objective function removed,

$$\begin{array}{ll}
\textit{maximize} & \boldsymbol{c}^T(I - A_2^+A_2)\boldsymbol{y}\\
\textit{subject to} & A_1(I - A_2^+A_2)\boldsymbol{y} \geq \boldsymbol{b}_1 - A_1A_2^+\boldsymbol{b}_2\ .
\end{array}$$

2.2.7 Orthogonal Projections

We say that a matrix P whose domain is vector space $\mathcal{V}$ and whose codomain is a subspace of $\mathcal{V}$, is a *projection* if it satisfies $P^2 = P$; if it is also symmetric then it is called an *orthogonal projection*.

Regarded as a linear transformation, X^+X is the orthogonal projection onto the row space $\mathcal{L}(X)$ of X, while XX^+ is the orthogonal projection onto the column space $\mathcal{R}(X)$ of X. The matrix XX^+ is symmetric and idempotent as, from the definition of the pseudo-inverse, $(XX^+)^2 = XX^+XX^+ = XX^+$ (since $XX^+X = X$), while property (c) above asserts the symmetry of XX^+.

Lemma 2.2.19 If P is an orthogonal projection then $\|\boldsymbol{x}\|^2 = \|P\boldsymbol{x}\|^2 + \|\boldsymbol{x} - P\boldsymbol{x}\|^2$.

Proof $\|\boldsymbol{x}\|^2 = \|P\boldsymbol{x} + (I-P)\boldsymbol{x}\|^2 = (P\boldsymbol{x})^TP\boldsymbol{x} + 2(P\boldsymbol{x})^T(I-P)\boldsymbol{x} + [(I-P)\boldsymbol{x}]^T(I-P)\boldsymbol{x}$
$= \|P\boldsymbol{x}\|^2 + 2\boldsymbol{x}^TP^T(I-P)\boldsymbol{x} + \|(I-P)\boldsymbol{x}\|^2 = \|P\boldsymbol{x}\|^2 + \|(I-P)\boldsymbol{x}\|^2$. □

Corollary: 2.2.20 If P is an orthogonal projection then $\|P\boldsymbol{x}\|^2 \leq \|\boldsymbol{x}\|^2$.

Proof Obvious from Lemma 2.2.19. □

Lemma 2.2.21 If P is an orthogonal projection then $\|P\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2 \Leftrightarrow P\boldsymbol{x} = \boldsymbol{x} \Leftrightarrow \boldsymbol{x}^TP\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{x}$.

Proof
$\|P\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2 \Rightarrow \|\boldsymbol{x} - P\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2 - 2\boldsymbol{x}^TP\boldsymbol{x} + \|P\boldsymbol{x}\|^2$
$= \|\boldsymbol{x}\|^2 - 2\boldsymbol{x}^TPP\boldsymbol{x} + \|\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2 - 2\|P\boldsymbol{x}\|^2 + \|\boldsymbol{x}\|^2$
$= \|\boldsymbol{x}\|^2 - 2\|\boldsymbol{x}\|^2 + \|\boldsymbol{x}\|^2 = 0 \Rightarrow P\boldsymbol{x} = \boldsymbol{x} \Rightarrow \boldsymbol{x}^TP\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{x} \Rightarrow \boldsymbol{x}^TPP\boldsymbol{x}$
$= \boldsymbol{x}^T\boldsymbol{x} \Rightarrow \|P\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2$.

Lemma 2.2.22 $(PX = X) \wedge (P\boldsymbol{x} = 0\ \forall \boldsymbol{x} \in <X>^\perp) \Rightarrow P = XX^+$

Proof $(PX = X) \wedge (P\boldsymbol{x} = 0\ \forall \boldsymbol{x} \in <X>^\perp) \Rightarrow (PXX^+ = XX^+) \wedge (P(I - XX^+)\boldsymbol{x} = 0\ \forall \boldsymbol{x} \in \Re^m)$
$\Rightarrow (PXX^+ = XX^+) \wedge (P(I - XX^+) = 0) \Rightarrow (PXX^+ = XX^+) \wedge (P = PXX^+) \Rightarrow P = XX^+$.

Definition: 2.2.23 (a) $P_X = XX^+$
(b) $P'_X = I - XX^+$

Lemma 2.2.24 $(P'_XY)^+X = 0$

Proof
$(P'_XY)^+X = (P'_XY)^+(P'_XY)(P'_XY)^+X = (P'_XY)^+((P'_XY)(P'_XY)^+)^TX$
$= (P'_XY)^+(P'_XY)^{+T}(P'_XY)^TX = (P'_XY)^+(P'_XY)^{+T}Y^TP'^T_XX$
$= (P'_XY)^+(P'_XY)^{+T}Y^TP'_XX = (P'_XY)^+(P'_XY)^{+T}Y^T0 = 0.$

Definition: 2.2.25 For $m \times n_1$ matrix X and $m \times n_2$ matrix Y,
$<X, Y> = \{X\boldsymbol{u} + Y\boldsymbol{v} : \boldsymbol{u} \in \Re^{n_1}, \boldsymbol{v} \in \Re^{n_2}\}$,

The following lemma, theorem and corollary are needed for analysis of the multiple residual operator $\lhd$ in Chapter 2.3.3.

Lemma 2.2.26 With the definitions 2.2.23, for $\Pi = P_X + P'_XY(P'_XY)^+$
(a) Π is idempotent symmetric.
(b) $\Pi X = X$
(c) $\Pi Y = Y$
(d) $\forall \boldsymbol{z} \in <X, Y>^\perp$: $\Pi \boldsymbol{z} = 0$

Proof (a) Π is plainly symmetric, while $\Pi^2 = \big(P_X + P'_X Y(P'_X Y)^+\big)^2$
$= P_X^2 + P_X P'_X Y(P'_X Y)^+ + P'_X Y(P'_X Y)^+ P_X + \big(P'_X Y(P'_X Y)^+\big)^2$
$= P_X + 0Y(P'_X Y)^+ + (P'_X Y)^{+T} Y^T P'_X P_X + P'_X Y(P'_X Y)^+$
$= P_X + (P'_X Y)^{+T} Y^T 0 + P'_X Y(P'_X Y)^+ = P_X + P'_X Y(P'_X Y)^+ = \Pi$
so Π is idempotent symmetric.

(b) $\Pi X = \big(P_X + P'_X Y(P'_X Y)^+\big)X = P_X X + P'_X Y(P'_X Y)^+ X$
$= P_X X + (P'_X Y)^{+T} Y^T P'_X X = X + (P'_X Y)^{+T} Y^T 0 = X$;

(c) $Y^T \Pi Y = Y^T\big(P_X + P'_X Y(P'_X Y)^+\big)Y = Y^T P_X Y + Y^T\big(P'_X Y(P'_X Y)^+\big)Y$
$= Y^T P_X Y + Y^T\big(P'_X Y(Y^T P'_X Y)^+ Y^T P'_X\big)Y = Y^T P_X Y + (Y^T P'_X Y)(Y^T P'_X Y)^+(Y^T P'_X Y)$
$= Y^T P_X Y + Y^T P'_X Y = Y^T(P_X + P'_X)Y = Y^T Y \quad\Rightarrow\quad Y^T \Pi Y = Y^T Y \quad\Rightarrow\quad Y^T(I-\Pi)Y = 0$
$\Rightarrow\quad ((I-\Pi)Y)^T(I-\Pi)Y = 0 \quad\Rightarrow\quad (I-\Pi)Y = 0 \quad\Rightarrow\quad \Pi Y = Y$;

(d) $\boldsymbol{z} \in < X, Y >^\perp \quad\Rightarrow\quad (X^T\boldsymbol{z} = 0 \ \wedge\ Y^T\boldsymbol{z} = 0) \quad\Rightarrow\quad (X^{+T}X^T\boldsymbol{z} = 0 \ \wedge\ Y^T\boldsymbol{z} = 0)$
$\Rightarrow\quad \big((XX^+)^T\boldsymbol{z} = 0 \ \wedge\ Y^T\boldsymbol{z} = 0\big) \quad\Rightarrow\quad \big((P_X)^T\boldsymbol{z} = 0 \ \wedge\ Y^T\boldsymbol{z} = 0\big)$
$\Rightarrow\quad (P_X\boldsymbol{z} = 0 \ \wedge\ Y^T\boldsymbol{z} = 0) \quad\Rightarrow\quad (P_X\boldsymbol{z} = 0 \ \wedge\ P'_X\boldsymbol{z} = \boldsymbol{z} \ \wedge\ Y^T\boldsymbol{z} = 0)$,
so $\Pi\boldsymbol{z} = \big(P_X + P'_X Y(P'_X Y)^+\big)\boldsymbol{z} = \big(P_X + P'_X Y(Y^T P'_X Y)^+ Y^T P'_X\big)\boldsymbol{z}$
$= P_X\boldsymbol{z} + P'_X Y(Y^T P'_X Y)^+ Y^T P'_X\boldsymbol{z} = 0 + P'_X Y(Y^T P'_X Y)^+ Y^T\boldsymbol{z}$
$= P'_X Y(Y^T P'_X Y)^+ 0 = 0 \quad\Rightarrow\quad \Pi\boldsymbol{z} = 0$, so $\Pi(< X, Y >^\perp) = \{0\}$. □

Theorem: 2.2.27 $P_{[X:Y]} = P_X + P'_X Y(P'_X Y)^+$

Proof As in Lemma 2.2.26 let $\Pi = P_X + P'_X Y(P'_X Y)^+$, then $\Pi X = X$ and $\Pi Y = Y$ so it follows that $\Pi[X : Y] = [X : Y]$; further since $\Pi\boldsymbol{z} = 0\ \forall\boldsymbol{z} \in < X, Y >^\perp \Leftrightarrow \Pi\boldsymbol{z} = 0\ \forall\boldsymbol{z} \in < [X : Y] >^\perp$ we have $(\Pi[X : Y] = [X : Y]) \wedge (\forall\boldsymbol{z} \in < [X : Y] >^\perp : \Pi\boldsymbol{z} = 0)$ so from Lemma 2.2.22 $\Pi = [X : Y][X : Y]^+$, that is $\Pi = P_{[X:Y]}$. □

Corollary: 2.2.28 $P'_{[X:Y]} = P'_X - P'_X Y(P'_X Y)^+$

Proof $P'_{[X:Y]} = I - P_{[X:Y]} = I - (P_X + P'_X Y(P'_X Y)^+) = (I - P_X) - P'_X Y(P'_X Y)^+ = P'_X - P'_X Y(P'_X Y)^+$. □

2.2.8 The Kronecker Product

The results of this section are needed for the analysis of the transportation model, which is covered in Chapter 7 and will be continued in the following chapters.

Given an $m \times n$ matrix $X = (x_{ij})$ and a $p \times q$ matrix $Y = (y_{k\ell})$, the Kronecker product $X \otimes Y$ is defined by the $mp \times nq$ matrix

$$(X \otimes Y)_{(i-1)p+k,(j-1)q+\ell} = x_{ij}y_{k\ell} \ ,$$

that is

$$X \otimes Y = \begin{pmatrix} x_{11}Y & \cdots & x_{1n}Y \\ \vdots & \ddots & \vdots \\ x_{m1}Y & \cdots & x_{mn}Y \end{pmatrix}$$

Note, for vectors $\boldsymbol{v}$ and $\boldsymbol{w}$,

$$\boldsymbol{v}^T \otimes \boldsymbol{w} = [v_1\boldsymbol{w}\ v_2\boldsymbol{w}\ \cdots v_m\boldsymbol{w}] = \boldsymbol{w}\boldsymbol{v}^T.$$
$$\boldsymbol{v} \otimes \boldsymbol{w}^T = [w_1\boldsymbol{v}\ w_2\boldsymbol{v}\ \cdots w_n\boldsymbol{w}] = \boldsymbol{v}\boldsymbol{w}^T.$$

From [25, p. 919] we have, for conformable matrices,

$$\begin{array}{ll}
\otimes \text{ is associative,} & \text{(a)}\\
X \otimes (Y+Z) = X \otimes Y + X \otimes Z, & \text{(b)}\\
(X+Y) \otimes Z = X \otimes Z + Y \otimes Z, & \text{(c)}\\
(W \otimes X)(Y \otimes Z) = (WY) \otimes (XZ), & \text{(d)}\\
(X \otimes Y)^T = X^T \otimes Y^T, & \text{(e)}\\
(X \otimes Y)^+ = X^+ \otimes Y^+ \text{ (Lemma 2.2.30)}, & \text{(f)}\\
\text{For X, Y square, } trace(X \otimes Y) = (traceX)(traceY), & \text{(g)}\\
rank(X \otimes Y) = (rankX)(rankY), \text{ and} & \text{(h)}\\
\text{Each of the sets of normal, Hermitian, positive definite,} & \text{(i)}\\
\text{and unitary matrices is closed under the Kronecker product.} &
\end{array} \tag{2.7}$$

Lemma 2.2.29 If $WY = YW$ and $XZ = ZX$ then $(W \otimes X)(Y \otimes Z) = (Y \otimes Z)(W \otimes X)$.

Proof $(W \otimes X)(Y \otimes Z) = (WY) \otimes (XZ) = (YW) \otimes (ZX) = (Y \otimes Z)(W \otimes X)$.

Lemma 2.2.30 $(X \otimes Y)^+ = X^+ \otimes Y^+$.

Proof We verify the four conditions for the pseudo-inverse:

1. $(X \otimes Y)(X^+ \otimes Y^+)(X \otimes Y) = (XX^+) \otimes (YY^+)(X \otimes Y) = (XX^+X) \otimes (YY^+Y) = X \otimes Y.$

2. $(X^+ \otimes Y^+)(X \otimes Y)(X^+ \otimes Y^+) = (X^+X \otimes Y^+Y)(X^+ \otimes Y^+) = (X^+XX^+ \otimes Y^+YY^+) = X^+ \otimes Y^+.$

3. $\{(X \otimes Y)(X^+ \otimes Y^+)\}^T = \{(XX^+) \otimes (YY^+)\}^T = (XX^+)^T \otimes (YY^+)^T = (XX^+) \otimes (YY^+) = (X \otimes Y)(X^+ \otimes Y^+).$

4. Similar to 3. □

2.3 The Gram-Schmidt Function

The aim of this section is to develop regression theory from an algebraic rather than geometric perspective for the reason that computers have been taught more algebra than geometry; in Appendix B this perspective will be seen to facilitate the development of an algorithm for affine minimization.

2.3.1 The Elementary Function

Our aim here is to make precise the vector function underlying Gram-Schmidt orthogonalization and to delineate some important properties of this function. We need this theory in Chapter 2.4 to develop an efficient algorithm for computing a solution to the *fixed-point problem* of Chapter 5.

2.3.1.1 Definition

Given vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ from the same space, with superscript "+" as in Definition 2.2.6, the regression function $\triangleright$ and residual (or Gram-Schmidt) function $\triangleleft$ are defined as:

Definition: 2.3.1 (a) $\boldsymbol{x} \triangleright \boldsymbol{y} = \boldsymbol{y}(\boldsymbol{y}^T\boldsymbol{y})^+\boldsymbol{y}^T\boldsymbol{x}$
(b) $\boldsymbol{x} \triangleleft \boldsymbol{y} = (I - \boldsymbol{y}(\boldsymbol{y}^T\boldsymbol{y})^+\boldsymbol{y}^T)\boldsymbol{x}$

Alternatively we have

$$\begin{aligned} \boldsymbol{x} \triangleright \boldsymbol{y} &= \boldsymbol{y}\boldsymbol{y}^+\boldsymbol{x} && \text{(a)} \\ \boldsymbol{x} \triangleleft \boldsymbol{y} &= (I - \boldsymbol{y}\boldsymbol{y}^+)\boldsymbol{x} && \text{(b)} \end{aligned} \qquad (2.8)$$

Remark: 2.3.2 $(\boldsymbol{x} \triangleleft \boldsymbol{y}) \cdot (\boldsymbol{x} \triangleright \boldsymbol{y}) = 0$.

Figure 2.1 sums up the relationship between $\boldsymbol{x} \triangleleft \boldsymbol{y}$ and $\boldsymbol{x} \triangleright \boldsymbol{y}$.

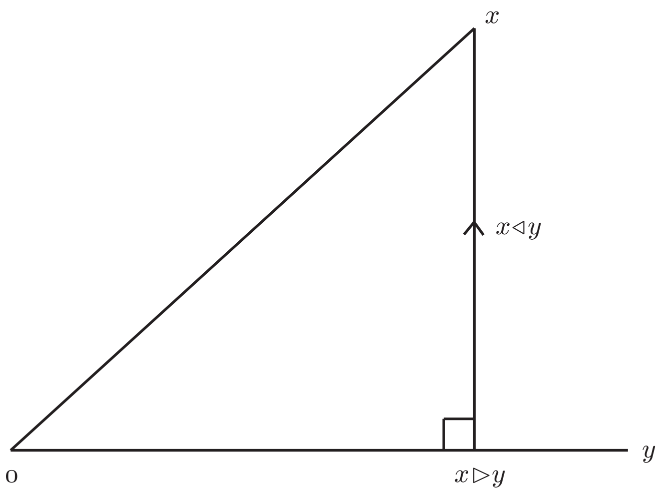


Figure 2.1: **Elementary Gram-Schmidt Function**

Remark: 2.3.3 $\boldsymbol{y} \cdot \boldsymbol{x} \triangleleft \boldsymbol{y} = \boldsymbol{y}^T(I - \boldsymbol{y}(\boldsymbol{y}^T\boldsymbol{y})^+\boldsymbol{y}^T)\boldsymbol{x} = 0$.

2.3.1.1.1 Properties of the Elementary Operator

Lemma 2.3.4

a. $(\lambda\boldsymbol{x}) \triangleleft \boldsymbol{y} = \lambda(\boldsymbol{x} \triangleleft \boldsymbol{y})$,
b. $\boldsymbol{x} \triangleleft (\lambda\boldsymbol{y}) = \boldsymbol{x} \triangleleft \boldsymbol{y}$, for $\lambda \neq 0$
c. $(\boldsymbol{x} + \boldsymbol{y}) \triangleleft z = \boldsymbol{x} \triangleleft z + \boldsymbol{y} \triangleleft z$,
d. $\boldsymbol{y} \triangleleft (\boldsymbol{x} \triangleleft \boldsymbol{y}) = \boldsymbol{y}$.
e. $\boldsymbol{x} \triangleleft \boldsymbol{y} = \boldsymbol{0} \Leftrightarrow \boldsymbol{x} = \lambda\boldsymbol{y} \; \exists\lambda \in \Re$
f. $\boldsymbol{x} \triangleleft \boldsymbol{0} = \boldsymbol{x}$
g. $\boldsymbol{0} \triangleleft \boldsymbol{x} = \boldsymbol{0}$

Proof: a. If $\boldsymbol{y} = \mathbf{0}$ then $(\lambda\boldsymbol{x}) \lhd \boldsymbol{y} = (\lambda\boldsymbol{x}) \lhd \mathbf{0} = \lambda\boldsymbol{x} = \lambda(\boldsymbol{x} \lhd \mathbf{0}) = \lambda(\boldsymbol{x} \lhd \boldsymbol{y})$; if $\boldsymbol{y} \neq \mathbf{0}$ then $(\lambda\boldsymbol{x}) \lhd \boldsymbol{y} = \lambda\boldsymbol{x} - \frac{(\lambda\boldsymbol{x})\cdot\boldsymbol{y}}{\boldsymbol{y}\cdot\boldsymbol{y}}\boldsymbol{y} = \lambda\left(\boldsymbol{x} - \frac{\boldsymbol{x}\cdot\boldsymbol{y}}{\boldsymbol{y}\cdot\boldsymbol{y}}\boldsymbol{y}\right) = \lambda(\boldsymbol{x} \lhd \boldsymbol{y})$; b. If $\boldsymbol{y} = \mathbf{0}$ then $\boldsymbol{x} \lhd (\lambda\boldsymbol{y}) = \boldsymbol{x} \lhd \mathbf{0} = \boldsymbol{x} \lhd \boldsymbol{y}$; if $\boldsymbol{y} \neq \mathbf{0}$ then $\boldsymbol{x} \lhd (\lambda\boldsymbol{y}) = \boldsymbol{x} - \frac{\boldsymbol{x}\cdot(\lambda\boldsymbol{y})}{(\lambda\boldsymbol{y})\cdot(\lambda\boldsymbol{y})}(\lambda\boldsymbol{y}) = \boldsymbol{x} - \frac{\lambda^2(\boldsymbol{x}\cdot\boldsymbol{y})}{\lambda^2\boldsymbol{y}\cdot\boldsymbol{y}}\boldsymbol{y} = \boldsymbol{x} - \frac{(\boldsymbol{x}\cdot\boldsymbol{y})\boldsymbol{y}}{\boldsymbol{y}\cdot\boldsymbol{y}} = \boldsymbol{x} \lhd \boldsymbol{y}$; c. If $\boldsymbol{z} = \mathbf{0}$ then $(\boldsymbol{x} + \boldsymbol{y}) \lhd \boldsymbol{z} = (\boldsymbol{x} + \boldsymbol{y}) \lhd \mathbf{0} = (\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{x} \lhd \mathbf{0} + \boldsymbol{y} \lhd \mathbf{0} = \boldsymbol{x} \lhd \boldsymbol{z} + \boldsymbol{y} \lhd \boldsymbol{z}$, otherwise $(\boldsymbol{x} + \boldsymbol{y}) \lhd \boldsymbol{z} = (\boldsymbol{x}+\boldsymbol{y}) - \frac{(\boldsymbol{x}+\boldsymbol{y})\cdot\boldsymbol{z}}{\boldsymbol{z}\cdot\boldsymbol{z}}\boldsymbol{z} = \boldsymbol{x}+\boldsymbol{y} - \frac{\boldsymbol{x}\cdot\boldsymbol{z}}{\boldsymbol{z}\cdot\boldsymbol{z}}\boldsymbol{z} - \frac{\boldsymbol{y}\cdot\boldsymbol{z}}{\boldsymbol{z}\cdot\boldsymbol{z}}\boldsymbol{z} = \boldsymbol{x} - \frac{\boldsymbol{x}\cdot\boldsymbol{z}}{\boldsymbol{z}\cdot\boldsymbol{z}}\boldsymbol{z} + \boldsymbol{y} - \frac{\boldsymbol{y}\cdot\boldsymbol{z}}{\boldsymbol{z}\cdot\boldsymbol{z}}\boldsymbol{z} = \boldsymbol{x} \lhd \boldsymbol{z} + \boldsymbol{y} \lhd \boldsymbol{z}$; d. If $\boldsymbol{x} = \mathbf{0}$ or $\boldsymbol{y} = \mathbf{0}$ the result obviously holds; otherwise $\boldsymbol{y} \lhd (\boldsymbol{x} \lhd \boldsymbol{y}) = \boldsymbol{x} - \frac{\boldsymbol{y}\cdot(\boldsymbol{x}\lhd\boldsymbol{y})}{(\boldsymbol{x}\lhd\boldsymbol{y})\cdot(\boldsymbol{x}\lhd\boldsymbol{y})}\boldsymbol{x} \lhd \boldsymbol{y} = \boldsymbol{y} - \frac{\mathbf{0}}{(\boldsymbol{x}\lhd\boldsymbol{y})\cdot(\boldsymbol{x}\lhd\boldsymbol{y})}\boldsymbol{x} \lhd \boldsymbol{y} = \boldsymbol{y}$; e. If $\boldsymbol{y} \neq \mathbf{0}$, then $\boldsymbol{x} \lhd \boldsymbol{y} = \mathbf{0} \Rightarrow \boldsymbol{x} - (\boldsymbol{x}\cdot\boldsymbol{y}/(\boldsymbol{y}\cdot\boldsymbol{y})\boldsymbol{y} = \mathbf{0} \Rightarrow \boldsymbol{x} = (\boldsymbol{x}\cdot\boldsymbol{y}/(\boldsymbol{y}\cdot\boldsymbol{y})\boldsymbol{y} \Rightarrow \boldsymbol{x} = \lambda\boldsymbol{y}$ if $\boldsymbol{y} = \mathbf{0}$, then $\boldsymbol{x} \lhd \boldsymbol{y} = \mathbf{0} \Rightarrow \boldsymbol{x} \lhd \mathbf{0} = \mathbf{0} \Rightarrow \boldsymbol{x} = \mathbf{0} \Rightarrow \boldsymbol{x} = \lambda\boldsymbol{y}$. Conversely, if $\boldsymbol{y} \neq \mathbf{0}$, then $\boldsymbol{x} = \lambda\boldsymbol{y} \Rightarrow \boldsymbol{x} \lhd \boldsymbol{y} = (\lambda\boldsymbol{y}) \lhd \boldsymbol{y} = \lambda(\boldsymbol{y} \lhd \boldsymbol{y}) = \lambda\mathbf{0} = \mathbf{0}$, and if $\boldsymbol{y} = \mathbf{0}$, then $\boldsymbol{x} = \lambda\boldsymbol{y} \Rightarrow \boldsymbol{x} \lhd \boldsymbol{y} = \boldsymbol{x} \lhd \mathbf{0} = \boldsymbol{x} = \lambda\boldsymbol{y} = \lambda\mathbf{0} = \mathbf{0}$; f. $\boldsymbol{x} \lhd \mathbf{0} = (I - \mathbf{0}\mathbf{0}^+)\boldsymbol{x} = I\boldsymbol{x} = \boldsymbol{x}$; g. $\mathbf{0} \lhd \boldsymbol{x} = (I - \boldsymbol{x}\boldsymbol{x}^+)\mathbf{0} = \mathbf{0}$.

2.3.2 The Binary Function

The binary function is introduced rather slowly. First we define the normalized vector

Definition: 2.3.5 $\breve{\boldsymbol{x}} = \|\boldsymbol{x}\|^+\boldsymbol{x}$

then note the logical equivalence:

Proposition: 2.3.6 $\breve{\boldsymbol{x}} = \|\boldsymbol{x}\|^+\boldsymbol{x} \Leftrightarrow \|\boldsymbol{x}\|\breve{\boldsymbol{x}} = \boldsymbol{x}$

Next $(\boldsymbol{u} \lhd \boldsymbol{v}) \lhd (\boldsymbol{w} \lhd \boldsymbol{v})$ is written in a more convenient form:

Lemma 2.3.7 $(\boldsymbol{u} \lhd \boldsymbol{v}) \lhd (\boldsymbol{w} \lhd \boldsymbol{v}) = (\boldsymbol{u} \lhd \breve{\boldsymbol{v}}) \lhd (\breve{\boldsymbol{w}} \lhd \breve{\boldsymbol{v}})$

Proof $(\boldsymbol{u} \lhd \boldsymbol{v}) \lhd (\boldsymbol{w} \lhd \boldsymbol{v}) = (\boldsymbol{u} \lhd \boldsymbol{v}) \lhd ((\|\boldsymbol{w}\|\breve{\boldsymbol{w}}) \lhd \boldsymbol{v}) = (\boldsymbol{u} \lhd \boldsymbol{v}) \lhd \|\boldsymbol{w}\|(\breve{\boldsymbol{w}} \lhd \boldsymbol{v})$

$$\begin{cases} \text{if } \ \boldsymbol{w} = \mathbf{0} : = (\boldsymbol{u} \lhd \boldsymbol{v}) \lhd \mathbf{0} = (\boldsymbol{u} \lhd \boldsymbol{v}) \lhd (\breve{\boldsymbol{w}} \lhd \boldsymbol{v}) \\ \text{if } \ \boldsymbol{w} \neq \mathbf{0} : = (\boldsymbol{u} \lhd \boldsymbol{v}) \lhd (\breve{\boldsymbol{w}} \lhd \boldsymbol{v}) \end{cases}$$

$= (\boldsymbol{u} \lhd \boldsymbol{v}) \lhd (\breve{\boldsymbol{w}} \lhd \boldsymbol{v})$

$$\begin{cases} \text{if } \ \boldsymbol{v} = \mathbf{0} : & = & (\boldsymbol{u} \lhd \breve{\boldsymbol{v}}) \lhd (\breve{\boldsymbol{w}} \lhd \breve{\boldsymbol{v}}) \\ \text{if } \ \boldsymbol{v} \neq \mathbf{0} : & \overset{L\,2.3.4\,b}{=} & (\boldsymbol{u} \lhd (\|\boldsymbol{v}\|\breve{\boldsymbol{v}})) \lhd (\breve{\boldsymbol{w}} \lhd (\|\boldsymbol{v}\|\breve{\boldsymbol{v}})) = (\boldsymbol{u} \lhd \breve{\boldsymbol{v}}) \lhd (\breve{\boldsymbol{w}} \lhd \breve{\boldsymbol{v}}) \end{cases}$$

$= (\boldsymbol{u} \lhd \breve{\boldsymbol{v}}) \lhd (\breve{\boldsymbol{w}} \lhd \breve{\boldsymbol{v}})$

Lemma 2.3.8 $(\boldsymbol{u} \lhd \boldsymbol{v}) \lhd (\boldsymbol{w} \lhd \boldsymbol{v}) = (\boldsymbol{u} \lhd \boldsymbol{w}) \lhd (\boldsymbol{v} \lhd \boldsymbol{w})$

Proof

$$\begin{aligned}
&(\boldsymbol{u} \lhd \boldsymbol{v}) \lhd (\boldsymbol{w} \lhd \boldsymbol{v}) = (\boldsymbol{u} \lhd \breve{\boldsymbol{v}}) \lhd (\breve{\boldsymbol{w}} \lhd \breve{\boldsymbol{v}}) = \big((I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{+})\boldsymbol{u}\big) \lhd \big((I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{+})\breve{\boldsymbol{w}}\big) \\
&= \big((I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})\boldsymbol{u}\big) \lhd \big((I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})\breve{\boldsymbol{w}}\big) = \Big(I - \big((I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})\breve{\boldsymbol{w}}\big)\big((I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})\breve{\boldsymbol{w}}\big)^{+}\Big)\Big(I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\Big)\boldsymbol{u} \\
&= \left(I - \frac{\big((I-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})\breve{\boldsymbol{w}}\big)\big((I-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})\breve{\boldsymbol{w}}\big)^{T}}{\big((I-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})\breve{\boldsymbol{w}}\big)^{T}\big((I-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})\breve{\boldsymbol{w}}\big)}\right)\Big(I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\Big)\boldsymbol{u} \\
&= \left(I - \frac{(\breve{\boldsymbol{w}}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})(\breve{\boldsymbol{w}}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^{T}}{(\breve{\boldsymbol{w}}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^{T}(\breve{\boldsymbol{w}}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})}\right)\Big(I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\Big)\boldsymbol{u} \\
&= \left(I - \frac{(\breve{\boldsymbol{w}}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})(\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})}{(\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})(\breve{\boldsymbol{w}}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})}\right)\Big(I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\Big)\boldsymbol{u} \\
&= \left(I - \frac{(\breve{\boldsymbol{w}}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})(\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})}{\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{w}}-\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}-\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}+\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}}\right)\Big(I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\Big)\boldsymbol{u} \\
&= \left(I - \frac{\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}+\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}}{1-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2}\right)\Big(I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\Big)\boldsymbol{u} \\
&= \Bigg(I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T} - \frac{\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}+\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}}{1-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2} \\
&\qquad + \frac{(\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}+\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T})\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}}{1-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2}\Bigg)\boldsymbol{u} \\
&= \Bigg(I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T} - \frac{\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}+\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}}{1-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2} \\
&\qquad + \frac{\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}+\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}}{1-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2}\Bigg)\boldsymbol{u} \\
&= \Bigg(I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T} - \frac{\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}+\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}}{1-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2} \\
&\qquad + \frac{\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}+\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}}{1-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2}\Bigg)\boldsymbol{u} \\
&= \left(I - \breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T} - \frac{\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}+\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}}{1-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2}\right)\boldsymbol{u} \\
&= \left(I - \frac{\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}+\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}+(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}}{1-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2}\right)\boldsymbol{u} \\
&= \left(I - \frac{\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}+\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}-\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}-\breve{\boldsymbol{v}}\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}}\breve{\boldsymbol{w}}^{T}}{1-(\breve{\boldsymbol{v}}^{T}\breve{\boldsymbol{w}})^2}\right)\boldsymbol{u}
\end{aligned}$$

which is symmetric in v and w and therefor equal to $(\boldsymbol{u} \lhd \boldsymbol{w}) \lhd (\boldsymbol{v} \lhd \boldsymbol{w})$. □

2.3.3 The Multiple Function

2.3.3.1 Definition

Let the brackets [ ] denote sequences, then we define

$$[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_i] \lhd \boldsymbol{y} = [\boldsymbol{x}_1 \lhd \boldsymbol{y}, \boldsymbol{x}_2 \lhd \boldsymbol{y}, \cdots, \boldsymbol{x}_i \lhd \boldsymbol{y}] \quad \text{for } i \geq 1 \tag{2.9}$$

We define the recursion

$$\boldsymbol{x} \lhd [\boldsymbol{y}] = \boldsymbol{x} \lhd \boldsymbol{y} \ , \tag{2.10}$$

$$\boldsymbol{x} \lhd [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_i] = (\boldsymbol{x} \lhd \boldsymbol{y}_i) \lhd \big([\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_{i-1}] \lhd \boldsymbol{y}_i\big) \quad \text{for } i \geq 2. \tag{2.11}$$

Note that Equation 2.11 is a recursive scheme.

2.3.3.2 Set Property

Lemma 2.3.8 states that $(\boldsymbol{x} \lhd \boldsymbol{y}) \lhd (\boldsymbol{z} \lhd \boldsymbol{y}) = (\boldsymbol{x} \lhd \boldsymbol{z}) \lhd (\boldsymbol{y} \lhd \boldsymbol{z})$; this means that the definition $\boldsymbol{x} \lhd \{\boldsymbol{y}, \boldsymbol{z}\} = (\boldsymbol{x} \lhd \boldsymbol{y}) \lhd (\boldsymbol{z} \lhd \boldsymbol{y})$ is well-defined, because the order of $\boldsymbol{y}$ and $\boldsymbol{z}$ does not affect the result. We now generalize this result as

Theorem: 2.3.9 $\boldsymbol{x}_1 \lhd [\boldsymbol{x}_2, \cdots, \boldsymbol{x}_n] = \boldsymbol{x}_1 \lhd [\boldsymbol{x}_{\pi 2}, \cdots, \boldsymbol{x}_{\pi n}]$ for π an arbitrary permutation of $[2, \cdots, n]$.

Proof Assume true for $n < m$ then

$$\boldsymbol{x}_1 \lhd [\boldsymbol{x}_2, \cdots, \boldsymbol{x}_m] \overset{2.11}{=} (\boldsymbol{x}_1 \lhd \boldsymbol{x}_m) \lhd [\boldsymbol{x}_2 \lhd \boldsymbol{x}_m, \cdots, \boldsymbol{x}_{m-1} \lhd \boldsymbol{x}_m]$$

which, under the above assumption,

$$\begin{aligned} &= (\boldsymbol{x}_1 \lhd \boldsymbol{x}_m) \lhd [\boldsymbol{x}_{\pi 2} \lhd \boldsymbol{x}_m, \cdots, \boldsymbol{x}_{\pi(m-1)} \lhd \boldsymbol{x}_m] \;\; \text{for arbitrary permutation } \pi \text{ on } \{2, \cdots, m-1\} \\ &\overset{2.11}{=} \boldsymbol{x}_1 \lhd [\boldsymbol{x}_{\pi 2}, \cdots, \boldsymbol{x}_{\pi(m-1)}, \boldsymbol{x}_m] \;\; \text{for arbitrary permutation } \pi \text{ on } \{2, \cdots, m-1\} \end{aligned}$$

Thus $\boldsymbol{x}_2$ to $\boldsymbol{x}_{m-1}$ can be permuted arbitrarily. Further

$$\begin{aligned} \boldsymbol{x}_1 \lhd [\boldsymbol{x}_2, \cdots, \boldsymbol{x}_m] &\overset{2.11}{=} (\boldsymbol{x}_1 \lhd \boldsymbol{x}_m) \lhd [\boldsymbol{x}_2 \lhd \boldsymbol{x}_m, \cdots, \boldsymbol{x}_{m-1} \lhd \boldsymbol{x}_m] \\ &\overset{2.11}{=} ((\boldsymbol{x}_1 \lhd \boldsymbol{x}_m) \lhd (\boldsymbol{x}_{m-1} \lhd \boldsymbol{x}_m)) \lhd [(\boldsymbol{x}_2 \lhd \boldsymbol{x}_m) \lhd (\boldsymbol{x}_{m-1} \lhd \boldsymbol{x}_m), \cdots, (\boldsymbol{x}_{m-2} \lhd \boldsymbol{x}_m) \lhd (\boldsymbol{x}_{m-1} \lhd \boldsymbol{x}_m)] \\ &\overset{L\,2.3.8}{=} ((\boldsymbol{x}_1 \lhd \boldsymbol{x}_{m-1}) \lhd (\boldsymbol{x}_m \lhd \boldsymbol{x}_{m-1})) \lhd [(\boldsymbol{x}_2 \lhd \boldsymbol{x}_{m-1}) \lhd (\boldsymbol{x}_m \lhd \boldsymbol{x}_{m-1}), \cdots, (\boldsymbol{x}_{m-2} \lhd \boldsymbol{x}_{m-1}) \lhd (\boldsymbol{x}_m \lhd \boldsymbol{x}_{m-1})] \\ &\overset{2.11}{=} (\boldsymbol{x}_1 \lhd \boldsymbol{x}_{m-1}) \lhd [(\boldsymbol{x}_2 \lhd \boldsymbol{x}_{m-1}), \cdots, (\boldsymbol{x}_{m-2} \lhd \boldsymbol{x}_{m-1}), (\boldsymbol{x}_m \lhd \boldsymbol{x}_{m-1})] \\ &\overset{2.11}{=} \boldsymbol{x}_1 \lhd [\boldsymbol{x}_2, \cdots, \boldsymbol{x}_{m-2}, \boldsymbol{x}_m, \boldsymbol{x}_{m-1}] \end{aligned}$$

so $\boldsymbol{x}_{m-1}$ and $\boldsymbol{x}_m$ can be permuted. Thus the terms $\boldsymbol{x}_2$ to $\boldsymbol{x}_m$ can be permuted arbitrarily. □

It now follows that the definitions

$$\begin{aligned} \boldsymbol{x} \lhd \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n\} &= \boldsymbol{x} \lhd [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n] \\ \boldsymbol{x} \rhd \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n\} &= \boldsymbol{x} - \boldsymbol{x} \lhd [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n] \end{aligned} \tag{2.12}$$

are well-defined, so from this point we may use the notation $\boldsymbol{x} \lhd \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n\}$.

Diagram 2.2 summarizes Theorem 2.3.9.

It follows from Theorem 2.3.9 that for arbitrary $\boldsymbol{y} \in \mathcal{V}$ we have the unique orthogonal decomposition

$$\boldsymbol{y} = XX^+\boldsymbol{y} + (I - XX^+)\boldsymbol{y} = \boldsymbol{y} \rhd \{\boldsymbol{x}_1 \cdots \boldsymbol{x}_n\} + \boldsymbol{y} \lhd \{\boldsymbol{x}_1 \cdots \boldsymbol{x}_n\} \tag{2.13}$$

where $XX^+\boldsymbol{y} \in < X >$ and $(I - XX^+)\boldsymbol{y} \in < X >^\perp$.

The final result is on two stage regression:

Theorem: 2.3.10 For $n \geq 3, \boldsymbol{x}_1 \lhd [\boldsymbol{x}_2, \cdots, \boldsymbol{x}_n] = (\boldsymbol{x}_1 \lhd [\boldsymbol{x}_2, \cdots, \boldsymbol{x}_{n-1}]) \lhd (\boldsymbol{x}_n \lhd [\boldsymbol{x}_2, \cdots, \boldsymbol{x}_{n-1}])$

Proof

$$\begin{aligned} &\boldsymbol{x}_1 \lhd [\boldsymbol{x}_2, \cdots, \boldsymbol{x}_{i-2}, \boldsymbol{x}_{i-1}, \boldsymbol{x}_i] \overset{T\,2.3.9}{=} P'_{X_{2,n}}\boldsymbol{x}_1 = P'_{[X_{2,n-1}:\boldsymbol{x}_n]}\boldsymbol{x}_1 \\ &\overset{C\,2.2.28}{=} (P'_{X_{2,n-1}} - (P'_{X_{2,n-1}}\boldsymbol{x}_n)(P'_{X_{2,n-1}}\boldsymbol{x}_n)^+)\boldsymbol{x}_1 = (I - (P'_{X_{2,n-1}}\boldsymbol{x}_n)(P'_{X_{2,n-1}}\boldsymbol{x}_n)^+)P'_{X_{2,n-1}}\boldsymbol{x}_1 \\ &\overset{D\,2.3.1\,b}{=} (P'_{X_{2,n-1}}\boldsymbol{x}_1) \lhd (P'_{X_{2,n-1}}\boldsymbol{x}_n) \overset{T\,2.3.9}{=} (\boldsymbol{x}_1 \lhd \{\boldsymbol{x}_2, \cdots, \boldsymbol{x}_{n-1}\}) \lhd (\boldsymbol{x}_n \lhd \{\boldsymbol{x}_2, \cdots, \boldsymbol{x}_{n-1}\}) \; . \end{aligned}$$

□

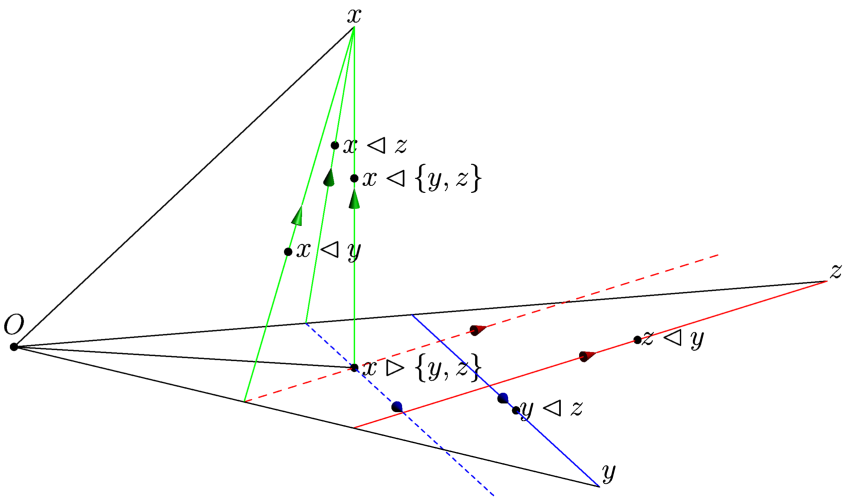


Figure 2.2: Two Variable Gram-Schmidt Operator

2.4 Affine Minimization

We compute the point of minimum norm in a linearly transformed affine set. This result is needed for computing a fixed-point in Chapter 2.4. Given the affine space $S = \{x : \boldsymbol{a}^T\boldsymbol{\mu} - 1 = 0\}$ which is transformed to GS by matrix G, we wish to find the point of minimum norm in GS - that is the problem

$$\text{minimize} \quad \|G\boldsymbol{\mu}\|^2 \quad \text{subject to} \quad \boldsymbol{a}^T\boldsymbol{\mu} - 1 = 0. \tag{2.14}$$

Note that

1. S is proper affine space since x must necessarily be non-zero.
2. GS might not be a proper affine space.
3. if $\boldsymbol{a} = 0$ the problem is infeasible, so only the case $\boldsymbol{a} \neq 0$, is considered.

Lemma 2.4.1 GS is a proper affine space iff $(I - G^+G)\boldsymbol{a} = \mathbf{0}$.

Proof $(I - G^+G)\boldsymbol{a} = \mathbf{0} \Rightarrow \boldsymbol{a} = G^+G\boldsymbol{a} \Rightarrow \boldsymbol{a}^T = \boldsymbol{a}^T(G^+G)^T \Rightarrow \boldsymbol{a}^T = \boldsymbol{a}^T G^+G$, so $\boldsymbol{\mu} \in S \Rightarrow \boldsymbol{a}^T\boldsymbol{\mu} = 1 \Rightarrow \boldsymbol{a}^T G^+ G\boldsymbol{\mu} = 1 \Rightarrow G\boldsymbol{\mu} \neq \mathbf{0}$, thus GS is a proper affine space; with $(I - G^+G)\boldsymbol{a} \neq \mathbf{0}$, define $\boldsymbol{\mu}_2 = \dfrac{(I - G^+G)\boldsymbol{a}}{\boldsymbol{a}^T(I - G^+G)\boldsymbol{a}}$ then $\boldsymbol{\mu}_2 \in S$ and $G\boldsymbol{\mu}_2 = 0$, thus GS is an improper affine space, and $\boldsymbol{\mu}_2$ is a solution to Problem 2.14 for the case where $(I - G^+G)\boldsymbol{a} \neq \mathbf{0}$.

2.5 Lagrangian Multiplier Method

Problem 2.14 leads to the Lagrangian[35]

$$\mathfrak{L} = \|G\boldsymbol{\mu}\|^2 + \lambda(\boldsymbol{a}^T\boldsymbol{\mu} - 1);$$

setting the derivative of $\mathcal{L}$ w.r.t. $\boldsymbol{\mu}$ equal to zero:

$$\partial\mathfrak{L}/\partial\boldsymbol{\mu} = 2G^TG\boldsymbol{\mu} + \lambda\boldsymbol{a} = 0 \tag{2.15}$$

Multiplying Equation 2.15 by G^{+T} yields $2G^{+T}G^TG\boldsymbol{\mu} + \lambda G^{+T}\boldsymbol{a} = 0 \Rightarrow 2(GG^+)^TG\boldsymbol{\mu} + \lambda G^{+T}\boldsymbol{a} = 0$ $\Rightarrow 2GG^+G\boldsymbol{\mu} + \lambda G^{+T}\boldsymbol{a} = 0 \Rightarrow 2G\boldsymbol{\mu} + \lambda G^{+T}\boldsymbol{a} = 0 \Rightarrow$

$$G\boldsymbol{\mu} = -\lambda G^{+T}\boldsymbol{a}/2 \tag{2.16}$$

Multiplying Equation 2.15 by $\boldsymbol{\mu}^T$ yields $2\boldsymbol{\mu}^TG^TG\boldsymbol{\mu} + \lambda\boldsymbol{\mu}^T\boldsymbol{a} = 0 \Rightarrow 2\boldsymbol{\mu}^TG^TG\boldsymbol{\mu} + \lambda = 0 \Rightarrow \Rightarrow \lambda =$ $-2(-\lambda G^{+T}\boldsymbol{a}/2)^T(-\lambda G^{+T}\boldsymbol{a}/2) = -\lambda^2(G^{+T}\boldsymbol{a})^TG^{+T}\boldsymbol{a}/2 = -\lambda^2\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}/2$
$\Rightarrow \lambda = -\lambda^2\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}/2 \Rightarrow \lambda = 0$ or

$$\lambda = \frac{-2}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} \tag{2.17}$$

2.5.1 Case 1: $(I - G^+G)\boldsymbol{a} = \mathbf{0}$

Since we have a solution for the case $(I - G^+G)\boldsymbol{a} \neq \mathbf{0}$, the case where

$$(I - G^+G)\boldsymbol{a} = \mathbf{0} \tag{2.18}$$

is considered.

Substituting Equation 2.17 into Equation 2.16:

$$G\boldsymbol{\mu} = -\frac{-2}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}}G^{+T}\boldsymbol{a}/2 = \frac{G^{+T}\boldsymbol{a}}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} \tag{2.19}$$

so

$$G^+G\boldsymbol{\mu} = \frac{G^+G^{+T}\boldsymbol{a}}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} \tag{2.20}$$

Taking

$$\boldsymbol{\mu}_1 = \frac{G^+G^{+T}\boldsymbol{a}}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} \overset{(L\,2.2.11\,b)}{=} \frac{(G^TG)^+\boldsymbol{a}}{\boldsymbol{a}^T(G^TG)^+\boldsymbol{a}} \tag{2.21}$$

we see that $\boldsymbol{\mu}_1$ is feasible since

$$\boldsymbol{a}^T\boldsymbol{\mu}_1 - 1 = \boldsymbol{a}^T\frac{(G^TG)^+\boldsymbol{a}}{\boldsymbol{a}^T(G^TG)^+\boldsymbol{a}} - 1 = 1 - 1 = 0$$

Now

$$G\boldsymbol{\mu}_1 = \frac{GG^+G^{+T}\boldsymbol{a}}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} = \frac{G^{+T}\boldsymbol{a}}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} \,. \tag{2.22}$$

and $\|G\boldsymbol{\mu}_1\|^2 \overset{2.22}{=} \left(\frac{G^{+T}\boldsymbol{a}}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}}\right)^T \left(\frac{G^{+T}\boldsymbol{a}}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}}\right) = \frac{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}}{(\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a})^2}$,

that is

$$\|G\boldsymbol{\mu}_1\|^2 = \frac{1}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} \,. \tag{2.23}$$

Given feasible $\mathfrak{y}_1$,

$$\boldsymbol{\mu}_1^TG^TG\mathfrak{y}_1 = (G\boldsymbol{\mu}_1)^TG\mathfrak{y}_1 \overset{2.22}{=} \left(\frac{G^{+T}\boldsymbol{a}}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}}\right)^T G\mathfrak{y}_1$$
$$= \frac{\boldsymbol{a}^TG^+G\mathfrak{y}_1}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} \overset{(2.18)}{=} \frac{\boldsymbol{a}^T\mathfrak{y}_1}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} = \frac{1}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}}$$

that is

$$\boldsymbol{\mu}_1^TG^TG\mathfrak{y}_1 = \frac{1}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} \tag{2.24}$$

so

$$\|G(\mathfrak{y}_1-\boldsymbol{\mu}_1)\|^2 = (\mathfrak{y}_1-\boldsymbol{\mu}_1)^TG^TG(\mathfrak{y}_1-\boldsymbol{\mu}_1) = \|G\mathfrak{y}_1\|^2 - 2\boldsymbol{\mu}_1^TG^TG\mathfrak{y}_1 + \|G\boldsymbol{\mu}_1\|^2$$
$$\overset{(2.24)}{=} \|G\mathfrak{y}_1\|^2 - \frac{2}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} + \|G\boldsymbol{\mu}_1\|^2 \overset{(2.23)}{=} \|G\mathfrak{y}_1\|^2 - \frac{2}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} + \frac{1}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}}$$
$$= \|G\mathfrak{y}_1\|^2 - \frac{1}{\boldsymbol{a}^TG^+G^{+T}\boldsymbol{a}} \overset{(2.23)}{=} \|G\mathfrak{y}_1\|^2 - \|G\boldsymbol{\mu}_1\|^2 \,,$$

that is

$$\|G(\mathfrak{y}_1-\boldsymbol{\mu}_1)\|^2 = \|G\mathfrak{y}_1\|^2 - \|G\boldsymbol{\mu}_1\|^2 \,, \tag{2.25}$$

and it follows that $\|G\mathfrak{y}_1\|^2 \geq \|G\boldsymbol{\mu}_1\|^2$, and that if $(I-G^+G)\boldsymbol{a} = 0$ then $\boldsymbol{\mu}_1$ is optimal and thus a solution.

2.5.2 Case 2: $(I - G^+G)\boldsymbol{a} \neq \mathbf{0}$

Set $\boldsymbol{\mu}_2 = \frac{(I-G^+G)\boldsymbol{a}}{\boldsymbol{a}^T(I-G^+G)\boldsymbol{a}}$, then $\boldsymbol{a}^T\boldsymbol{\mu}_2 = 1$ and $\|G\boldsymbol{\mu}_2\|^2 = \left\|\frac{G(I-G^+G)\boldsymbol{a}}{\boldsymbol{a}^T(I-G^+G)\boldsymbol{a}}\right\|^2 = 0$,
so $\boldsymbol{\mu}_2$ is a solution.

2.5.3 Summary

Summarizing, we have the "centred" solution:

$$\boxed{\begin{array}{ll} \boldsymbol{\mu}_1 = \dfrac{(G^TG)^+\boldsymbol{a}}{\boldsymbol{a}^T(G^TG)^+\boldsymbol{a}} & \text{if } (I-G^+G)\boldsymbol{a} = \mathbf{0} \\ \boldsymbol{\mu}_2 = \dfrac{(I-G^+G)\boldsymbol{a}}{\boldsymbol{a}^T(I-G^+G)\boldsymbol{a}} & \text{if } (I-G^+G)\boldsymbol{a} \neq \mathbf{0} \end{array}} \,. \tag{2.26}$$

Note

1. that $(I - G^+G)\boldsymbol{a} = \mathbf{0} \Rightarrow G^+G\boldsymbol{a} \neq \mathbf{0} \Rightarrow G^{+T}G^+G\boldsymbol{a} \neq \mathbf{0} \Rightarrow G^{+T}G^TG^{+T}\boldsymbol{a} \neq \mathbf{0} \Rightarrow G^{+T}\boldsymbol{a} \neq \mathbf{0} \Rightarrow \boldsymbol{a}^TG^+G^{+T}\boldsymbol{a} \neq 0$, so $\boldsymbol{\mu}_1$ is computable; we are dealing with an proper affine space.

2. $\boldsymbol{\mu}_2$ is also plainly computable; $G\boldsymbol{\mu}_2 = \mathbf{0}$ so we are dealing with an improper affine space.

We posit the general solution sets:

$$\boxed{\begin{array}{ll} S_1 = \left\{ \dfrac{(G^TG)^+\boldsymbol{a}}{\boldsymbol{a}^T(G^TG)^+\boldsymbol{a}} + (I - G^+G)\boldsymbol{z} : \boldsymbol{z} \in \Re^n \right\} & \text{if } \ (I - G^+G)\boldsymbol{a} = 0 \\ S_2 = \left\{ \dfrac{(I - G^+G)\boldsymbol{a}}{\boldsymbol{a}^T(I - G^+G)\boldsymbol{a}} + (I - G_a^+G_a)\boldsymbol{z} : \boldsymbol{z} \in \Re^n \right\} & \text{if } \ (I - G^+G)\boldsymbol{a} \neq 0 \ . \\ \text{where} \quad G_a = \begin{bmatrix} G \\ \boldsymbol{a}^T \end{bmatrix} & \end{array}} \tag{2.27}$$

2.5.4 Case 1

Further, we see that $\|G\mathfrak{y}_1\|^2 = \|G\boldsymbol{\mu}_1\|^2 \overset{(2.25)}{\Leftrightarrow} \|G(\mathfrak{y}_1 - \boldsymbol{\mu}_1)\|^2 = 0 \Leftrightarrow G\mathfrak{y}_1 = G\boldsymbol{\mu}_1 \Leftrightarrow \exists z : \mathfrak{y}_1 = \boldsymbol{\mu}_1 + (I - G^+G)z$, so the set of solutions is given precisely by Equation 2.27.

2.5.5 Case 2

The vector $\boldsymbol{\mu}_2$ clearly a solution; if $\mathfrak{y}_2$ is also a solution then

$$\|G(\mathfrak{y}_2 - \boldsymbol{\mu}_2)\|^2 = (\mathfrak{y}_2 - \boldsymbol{\mu}_2)^TG^TG(\mathfrak{y}_2 - \boldsymbol{\mu}_2) = \|G\mathfrak{y}_2\|^2 - 2\boldsymbol{\mu}_2^TG^TG\mathfrak{y}_2 + \|G\boldsymbol{\mu}_2\|^2 = 0$$

Now $\|G(\mathfrak{y}_2 - \boldsymbol{\mu}_2)\|^2 = 0 \Leftrightarrow G\mathfrak{y}_2 = G\boldsymbol{\mu}_2 \Leftrightarrow \exists z : \mathfrak{y}_2 = \boldsymbol{\mu}_2 + (I - G^+G)z$; additionally, $\boldsymbol{a}^T\boldsymbol{\mu}_2 = 1 \wedge \boldsymbol{a}^T\mathfrak{y} = 1 \Rightarrow \boldsymbol{a}^T\mathfrak{y} - \boldsymbol{a}^T\boldsymbol{\mu}_2 = 0 \Rightarrow \exists z : \mathfrak{y} = \boldsymbol{\mu}_2 + (I - a^+a)z$. Combining the conditions $\exists z : \mathfrak{y}_2 = \boldsymbol{\mu}_2 + (I - G^+G)z$ and $\exists z : \mathfrak{y} = \boldsymbol{\mu}_2 + (I - a^+a)z$ yields the condition $\exists z : \mathfrak{y}_2 = \boldsymbol{\mu}_2 + (I - G_a^+G_a)z$, so the set of solutions is given precisely by Equation 2.27.

2.6 Farkas' Lemma

Farkas' Lemma is the basis for the duality results in Chapter 4 and the fixed-point approach of Chapter 5. A preliminary lemma is needed before stating and proving Farkas' Lemma

Lemma 2.6.1 Let $\mathfrak{C}$ be a closed, convex, non-empty set in $\Re^n$ and $\boldsymbol{b} \notin \mathfrak{C}$. Let $\boldsymbol{p}$ be the closest point in $\mathfrak{C}$ to $\boldsymbol{b}$ then, for any $\boldsymbol{z} \in \mathfrak{C}$, $(\boldsymbol{b} - \boldsymbol{p})^T(\boldsymbol{z} - \boldsymbol{p}) \leq 0$.

Proof Define $\boldsymbol{p}(\lambda) = \boldsymbol{p} + \lambda(\boldsymbol{z} - \boldsymbol{p})$ and consider the square of the distance $d(\boldsymbol{b}, \boldsymbol{p}(\lambda))$ from $\boldsymbol{b}$ to $\boldsymbol{p}(\lambda)$: $d(\boldsymbol{b}, \boldsymbol{p}(\lambda))^2 = \|\boldsymbol{b} - \boldsymbol{p}(\lambda)\|^2 = (\boldsymbol{b} - (\boldsymbol{p} + \lambda(\boldsymbol{z} - \boldsymbol{p})))^T(\boldsymbol{b} - (\boldsymbol{p} + \lambda(\boldsymbol{z} - \boldsymbol{p})))$
Differentiating $d(\boldsymbol{b}, \boldsymbol{p}(\lambda))^2$ w.r.t. λ yields $2(\boldsymbol{p} - \boldsymbol{z})^T(\boldsymbol{b} - (\boldsymbol{p} + \lambda(\boldsymbol{z} - \boldsymbol{p}))) = 2(\lambda(\boldsymbol{z} - \boldsymbol{p})^T(\boldsymbol{z} - \boldsymbol{p}) - (\boldsymbol{b} - \boldsymbol{p})^T(\boldsymbol{z} - \boldsymbol{p}))$ and at $\lambda = 0$ the derivative is $-2(\boldsymbol{b} - \boldsymbol{p})^T(\boldsymbol{z} - \boldsymbol{p}) \geq 0$ since $d(\boldsymbol{b}, \boldsymbol{p})$ is at the minimum within $\mathfrak{C}$, so $(\boldsymbol{b} - \boldsymbol{p})^T(\boldsymbol{z} - \boldsymbol{p}) \leq 0$ for any $\boldsymbol{z}$ in $\mathfrak{C}$. □

Lemma 2.6.2 Given A and b, precisely one of the following statements is true:
1. $A\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}$ has a solution,
2. $A^T\boldsymbol{y} \geq \boldsymbol{0}$ and $\boldsymbol{y}^T\boldsymbol{b} < 0$ has a solution,

The reason that both 1. and 2. cannot occur is that if the conditions are both met then $\boldsymbol{x}^T A^T\boldsymbol{y} \geq 0$ since $\boldsymbol{x} \geq \boldsymbol{0}$ and $A^T\boldsymbol{y} \geq \boldsymbol{0}$ while $\boldsymbol{x}^T A^T\boldsymbol{y} = (A\boldsymbol{x})^T\boldsymbol{y} = \boldsymbol{b}^T\boldsymbol{y} < 0$.

It remains to prove that one or the other condition holds, and to this end first assume there does not exist $\boldsymbol{x}$ such that $A\boldsymbol{x} = b, \boldsymbol{x} \geq 0$. Let $\mathfrak{C} = \{A\boldsymbol{x} : \boldsymbol{x} \geq \boldsymbol{0}\}$ then $\boldsymbol{b} \notin \mathfrak{C}$. Let $\boldsymbol{p} = A\boldsymbol{w}, \boldsymbol{w} \geq \boldsymbol{0}$ be the closest point in $\mathfrak{C}$ to $\boldsymbol{b}$ then

$$\forall \boldsymbol{z} \in \mathfrak{C} : (\boldsymbol{b} - A\boldsymbol{w})^T(\boldsymbol{z} - A\boldsymbol{w}) \leq \boldsymbol{0}$$

that is

$$\forall x \geq \boldsymbol{0} : (\boldsymbol{b} - Aw)^T(A\boldsymbol{x} - A\boldsymbol{w}) \leq \boldsymbol{0}\ . \tag{2.28}$$

Define $\boldsymbol{y} = p - \boldsymbol{b} = Aw - \boldsymbol{b}$ then 2.28 can be written

$$\forall \boldsymbol{x} \geq \boldsymbol{0} : (\boldsymbol{x} - \boldsymbol{w})^T A^T\boldsymbol{y} \geq 0\ . \tag{2.29}$$

Let $\boldsymbol{e} \geq \boldsymbol{0}$ be an arbitrary vector, and take $\boldsymbol{x} = \boldsymbol{w} + \boldsymbol{e} \geq \boldsymbol{0}$ then 2.29 can be written

$$\forall \boldsymbol{x} \geq \boldsymbol{0} : \boldsymbol{e}^T A^T\boldsymbol{y} \geq 0\ , \tag{2.30}$$

which implies $A^T\boldsymbol{y} \geq \boldsymbol{0}$.

Further, from the definition of $\boldsymbol{y}$,

$$\boldsymbol{y}^T\boldsymbol{b} = \boldsymbol{y}^T(\boldsymbol{p} - \boldsymbol{y}) = \boldsymbol{y}^T\boldsymbol{p} - \boldsymbol{y}^T\boldsymbol{y} \tag{2.31}$$

while setting $\boldsymbol{x} = 0$ in 2.28 yields

$$-(\boldsymbol{b} - A\boldsymbol{w})^T A\boldsymbol{w} \leq 0 \Rightarrow -(\boldsymbol{b} - \boldsymbol{p})^T\boldsymbol{p} \leq 0 \Rightarrow \boldsymbol{y}^T\boldsymbol{p} \leq 0 \tag{2.32}$$

and it follows from 2.31, $\boldsymbol{y} \neq 0$, and 2.32 that $\boldsymbol{y}^T\boldsymbol{b} < 0$, so statement 2. is true.

2.7 Exercises

The exercises may been carried out using the program DrRacket which can be obtained from the Racket website. After opening DrRacket, under "Language" select “Swindle without CLOS” and checking the box “Case sensitive”, click on Scheme Functions. The usage of the scheme functions is as follows

1. (a) Compute the pseudo-inverse of the matrix $B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.

   (b) Compute BB^+; what is the result. (Hint: Use the function **mat*mat**.)

Table 2.1: Lisp Function Usage

name	example	semantics
vector	'(1 2 3)	VECTOR
vec+	(vec+ '(1 2 3) '(4 5 6))	addition
vec-	(vec- '(1 2 3) '(4 5 6))	subtraction
s*vec	(s*vec 4.5 '(1 2 3))	scalar mult.
vec/s	(vec/s '(1 2 3) 7)	scalar div.
matrix	'((1 2 3) (4 5 6))	MATRIX (row list)
idmat	(idmat 4)	identity
mat+	(mat+ '((1 2 3) (2 3 4)) '((6 7 4) (2 -1 7)))	addition
mat-	(mat- '((1 2 3) (2 3 4)) '((6 7 4) (2 -1 7)))	subtraction
s*mat	(s*mat 2.5 '((6 7 4) (-5 -1 6)))	scalar mult.
mat/s	(s*mat '((6 7 4) (-5 -1 6)) 2.5)	scalar div.
mat*mat	(mat+ '((1 2 3) (2 3 4)) '((6 7) (4 2) (-1 7)))	multiplication
transpose	(transpose '((1 2 3) (2 3 4)))	transpose
ginverse	(ginverse '((1 2 3) (4 5 6)))	Moore-Penrose Pseudo-Inverse

(c) Compute B^+B; check that this product is symmetric and idempotent.

(d) What is the trace of B^+B?

2. (a) Compute the matrix $\mathfrak{A} = AA^+$, where $A = \begin{bmatrix} 5 & 6 \\ 4 & -2 \\ 7 & -6 \end{bmatrix}$,

   (b) Compute the matrix $I - \mathfrak{A}$. (Hint: Use the functions **idmat** and **mat-**.)

   (c) Compute $\mathfrak{A}(I - \mathfrak{A}), \mathfrak{A}^2$, and $(I - \mathfrak{A})^2$.

3. (a) Check that the pseudo-inverse is correctly computed for matrix $\mathfrak{A}$. (use **nonconstructive**)

   (b) Compute the pseudo-inverse of the matrix $C = \begin{bmatrix} 1.0 & 2.0 & 3.0 \\ 4.0 & 5.0 & 6.0 \end{bmatrix}$ and compare the result with the pseudo-inverse of B.

4. Define $X \lhd Y = \{x_1, \cdots, x_i\} \lhd \{y_1, \cdots, y_j\} = \{\{x_1 \lhd \{y_1, \cdots, y_j\}, \cdots, x_i \lhd \{y_1, \cdots, y_j\}\}$.
   Prove that $(X \lhd Y) \lhd (Z \lhd Y) = (X \lhd Z) \lhd (Y \lhd Z)$.

Chapter 3

The General Fixed-Point Problem

A fixed-point of a function is a point which is mapped to itself; in the case of a square matrix, say $P, \boldsymbol{x}$ is said to be a fixed-point of P if $P\boldsymbol{x} = \boldsymbol{x}$.

In this section we introduce orthogonal projection matrices P and K, and a swapping-matrix S. From these matrices we construct a matrix U which is unitary in nature. We characterize the fixed-points of U and show how they can be computed by a converging series approach, a regression approach, and by a vector lattice approach. In Chapter 6 the developed theory will be used in the specific context of matrix $\mathfrak{P}$ of Chapter 5, Karush matrix $\mathfrak{K}_z$ and the specific swapping matrix $\mathfrak{S}$ which are introduced in Chapter 6. The specific $\mathfrak{P}, \mathfrak{K}_z$ and $\mathfrak{S}$ correspond in their general properties with those of P, K and S respectively and relate directly to the solution of the fixed-point problem; thus the general results of this section apply in the later specific context of the following chapters.

3.1 Theory

3.1.1 Swapping Matrices

S is called a *swapping-matrix* if S is unitary Hermitian; we say that S swaps Q (or Q is swapped by S) if Q is an Hermitian idempotent (i.e. an orthogonal projection), and

$$SQS = I - Q \; ; \tag{3.1}$$

If S is a swapping-matrix which swaps Q then from the above definition

$$\begin{array}{lcll} Q^2 & = & Q & \text{(a)} \\ Q^T & = & Q & \text{(b)} \\ S^2 & = & I & \text{(c)} \\ S^T & = & S & \text{(d)} \end{array} \tag{3.2}$$

since (a) and (b) follow from Q being an symmetric idempotent, (c) follows since S unitary implies $SS^T = I \Rightarrow S^2 = I$ since S is symmetric, (d) follows since S is Hermitian and in the real context this means symmetric.

Lemma 3.1.1 If S swaps Q then
(a) $S(I-Q)S = Q$
(b) $QS + SQ = S$
(c) $QSQ = 0$
(d) $(I-S)Q(I-S) = I-S$
(e) $(I+S)Q(I+S) = I+S$

Proof:
(a) $S(I-Q)S = S^2 - SQS = I - (I-Q) = Q$
(b) $S(I-Q)S = Q \Rightarrow SQS = I - Q \Rightarrow S^2QS = S(I-Q) \Rightarrow QS = S - SQ \Rightarrow QS + SQ = S$
(c) $Q^2 = Q \Rightarrow Q(I-Q) = 0 \Rightarrow QSQS = 0 \Rightarrow QSQS^2 = 0 \Rightarrow QSQ = 0$
(d) $(I-S)Q(I-S) = (Q-SQ)(I-S) = Q - QS - SQ + SQS$
$= Q - (QS+SQ) + (I-Q) = Q - S + I - Q = I - S$
(e) $(I+S)Q(I+S) = Q + (SQ+QS) + SQS = Q + S + I - Q = I + S$. □

Lemma 3.1.2 For unitary Hermitian S, $(SX)^+ = X^+S$.

Proof: By verifying the four non-constructive conditions for the pseudo-inverse:

1. $SXX^+SSX = SXX^+X \overset{2.2.3\,a}{=} SX$,
2. $X^+SSXX^+S = X^+XX^+S \overset{2.2.3\,b}{=} X^+S$,
3. $(SXX^+S)^T = S^T(XX^+)^TS^T \overset{2.2.3\,c}{=} SXX^+S$,
4. $(X^+SSX)^T = (X^+X)^T \overset{2.2.3\,d}{=} X^+X = X^+SSX$.

3.1.2 Projections and Oblique Projectors

Define

$$\overline{Q} = Q(I-S) \tag{3.3}$$

then we have

Lemma 3.1.3
(a) $\overline{Q}^2 = \overline{Q}$
(b) $\overline{Q}^T\overline{Q} = I - S$
(c) $\overline{Q}\,\overline{Q}^T = 2Q$

Proof:
(a) $\overline{Q}^2 = (Q(I-S))^2 = Q(I-S)Q(I-S) \overset{3.1.1\,d}{=} Q(I-S) = \overline{Q}$,
(b) $\overline{Q}^T\overline{Q} = (Q(I-S))^TQ(I-S) = (I-S)^TQ^TQ(I-S)$
$= (I-S)QQ(I-S) = (I-S)Q(I-S) \overset{3.1.1\,d}{=} I - S$.
(c) $\overline{Q}\,\overline{Q}^T = Q(I-S)(Q(I-S))^T = Q(I-S)^2Q = 2Q(I-S)Q = 2Q^2 = 2Q$. □

Note that $\overline{Q}$ is what Afriat [3] calls an *oblique projector*.

3.1.3 Operations on Vectors

Notionally the vectors are considered to be column vectors on which conjugation is defined; for matrix X and vectors u and v we require

$$\begin{aligned} X\boldsymbol{u}\cdot\boldsymbol{v} &= \boldsymbol{u}\cdot X^T\boldsymbol{v} && \text{(a)} \\ \boldsymbol{u}\cdot\boldsymbol{v} &= \boldsymbol{v}\cdot\boldsymbol{u} && \text{(b)} \end{aligned} \tag{3.4}$$

Lemma 3.1.4 If S swaps Q and $Q\boldsymbol{z} = \boldsymbol{z}$ then $\boldsymbol{z} \perp S\boldsymbol{z}$, that is $\boldsymbol{z}^T S\boldsymbol{z} = 0$.

Proof: $\boldsymbol{z}^T S\boldsymbol{z} = (Q\boldsymbol{z})^T S(Q\boldsymbol{z}) = \boldsymbol{z}^T Q^T SQ\boldsymbol{z} = \boldsymbol{z}^T QSQ\boldsymbol{z} \overset{3.1.1\,c}{=} \boldsymbol{z}^T 0\boldsymbol{z} = 0$. □

Lemma 3.1.5 $Q\boldsymbol{z} = \boldsymbol{z} \Leftrightarrow \overline{Q}\boldsymbol{z} = \boldsymbol{z}$.

Proof: $Q\boldsymbol{z} = \boldsymbol{z} \Rightarrow (Q\boldsymbol{z} - QSQ\boldsymbol{z} = \boldsymbol{z}) \wedge (Q\boldsymbol{z} = \boldsymbol{z}) \Rightarrow (Q^2\boldsymbol{z} - QSQ\boldsymbol{z} = \boldsymbol{z}) \wedge (Q\boldsymbol{z} = \boldsymbol{z}) \Rightarrow (Q(I - S)Q\boldsymbol{z} = \boldsymbol{z}) \wedge (Q\boldsymbol{z} = \boldsymbol{z}) \Rightarrow (Q(I - S)\boldsymbol{z} = \boldsymbol{z}) \wedge (Q\boldsymbol{z} = \boldsymbol{z}) \Rightarrow (\overline{Q}\boldsymbol{z} = \boldsymbol{z}) \wedge (Q\boldsymbol{z} = \boldsymbol{z}) \Rightarrow \overline{Q}\boldsymbol{z} = \boldsymbol{z}$, while $\overline{Q}\boldsymbol{z} = \boldsymbol{z} \Rightarrow (Q(I-S)\boldsymbol{z} = \boldsymbol{z}) \wedge (\overline{Q}\boldsymbol{z} = \boldsymbol{z}) \Rightarrow (Q^2(I-S)\boldsymbol{z} = Q\boldsymbol{z}) \wedge (\overline{Q}\boldsymbol{z} = \boldsymbol{z}) \Rightarrow (Q(I-S)\boldsymbol{z} = Q\boldsymbol{z}) \wedge (\overline{Q}\boldsymbol{z} = \boldsymbol{z}) \Rightarrow (\overline{Q}\boldsymbol{z} = Q\boldsymbol{z}) \wedge (\overline{Q}\boldsymbol{z} = \boldsymbol{z}) \Rightarrow Q\boldsymbol{z} = \boldsymbol{z}$. □

3.1.4 The General P-Unitary Matrix U

From this point we assume the existence of general idempotent symmetric K and P, both swapped by S. Referring forward, in Chapter 6 it will become apparent that the specific future rôle we have in mind for P is as the specific fixed-point matrix $\mathfrak{P}$ given by Equation 5.4, while that for K is the matrix $\mathfrak{K}_z$ given by Equation 6.7 which forces non-negativity and orthogonality, and thus eventually the invariant complementary slackness condition of Chapter 4.3, Equation 4.3.2 to hold.

With matrices K and P swapped by S, we have $PKSP = P(KS + SK - SK)P \overset{3.1.1\,b}{=} P(S - SK)P = PSP - PSKP \overset{3.1.1\,c}{=} -PSKP$, that is:

$$PKSP = -PSKP \; , \tag{3.5}$$

so $PK(I + S)P = PKP + PKSP = PKP - PSKP = P(K - SK)P = P(I - S)KP$, Another similar result is obtained by transposition, so

$$\begin{aligned} PK(I+S)P &= P(I-S)KP & \text{(a)} \\ P(I+S)KP &= PK(I-S)P & \text{(b)} \end{aligned} \tag{3.6}$$

Results 3.6′ dual to Equation set 3.6, where P and K are exchanged, also obtain.

For idempotent B define the B *component* of vector a or matrix A to be $a^T Ba$ or $A^T BA$ respectively.

Now $\boldsymbol{z}^T PKSP\boldsymbol{z} \overset{3.5}{=} \boldsymbol{z}^T(-PSKP)\boldsymbol{z} = -\boldsymbol{z}^T PSKP\boldsymbol{z} = -\boldsymbol{z}^T PKSP\boldsymbol{z}$, since P, K and S are Hermitian. Thus $\boldsymbol{z}^T PKSP\boldsymbol{z} = -\boldsymbol{z}^T PKSP\boldsymbol{z}$, which implies

$$\boldsymbol{z}^T PKSP\boldsymbol{z} = 0 \; , \; \text{for arbitrary } \boldsymbol{z} \tag{3.7}$$

and, of course, the dual result (3.7'): $\boldsymbol{z}^T KPSK\boldsymbol{z} = 0$, for arbitrary $\boldsymbol{z}$. Thus

Lemma 3.1.6 $\boldsymbol{z} = P\boldsymbol{z} \Rightarrow \boldsymbol{z}^T KS\boldsymbol{z} = 0$. □

Consistent with Equation 3.3 a, we define the general *oblique complementary slackness matrix*

$$\overline{K} = K(I - S) \tag{3.8}$$

and prove that $\boldsymbol{z} \to \overline{K}\boldsymbol{z}$ leaves the P component of fixed-point $\boldsymbol{z}$ unchanged:

Lemma 3.1.7 If $P\boldsymbol{z} = \boldsymbol{z}$ then
(a) $(\overline{K}\boldsymbol{z})^T\overline{K}\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}$,
(b) $\boldsymbol{z}^T\overline{K}\boldsymbol{z} = \boldsymbol{z}^T K\boldsymbol{z}$.

Proof: (a) $(\overline{K}\boldsymbol{z})^T\overline{K}\boldsymbol{z} = \boldsymbol{z}^T\overline{K}^T\overline{K}\boldsymbol{z} \overset{L\,3.1.3\,b}{=} \boldsymbol{z}^T(I-S)\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z} - \boldsymbol{z}^T S\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z} - \boldsymbol{z}^T PSP\boldsymbol{z} \overset{3.1.1\,c}{=} \boldsymbol{z}^T\boldsymbol{z}$,
(b) $\boldsymbol{z}^T\overline{K}\boldsymbol{z} = \boldsymbol{z}^T K(I-S)\boldsymbol{z} = \boldsymbol{z}^T K\boldsymbol{z} - \boldsymbol{z}^T KS\boldsymbol{z} = \boldsymbol{z}^T K\boldsymbol{z} - \boldsymbol{z}^T PKSP\boldsymbol{z} \overset{L\,3.1.6}{=} \boldsymbol{z}^T K\boldsymbol{z}$.

Corollary: 3.1.8 If $P\boldsymbol{z} = \boldsymbol{z}$ then $\boldsymbol{z}^T\overline{K}\boldsymbol{z} \leq \boldsymbol{z}^T\boldsymbol{z}$.

Proof: $\boldsymbol{z}^T\overline{K}\boldsymbol{z} \overset{L\,3.1.7\,b}{=} \boldsymbol{z}^T K\boldsymbol{z} \overset{L\,2.2.20}{=} \boldsymbol{z}^T K\boldsymbol{z} \leq \boldsymbol{z}^T\boldsymbol{z}$. □

We now define the matrix U and show that it is P-unitary:

Definition: 3.1.9 U=P(I+S)K(I-S)P ,

where K, P and S are as defined at Chapter 3.1.4 - that is S swaps both K and P. Note that $PU = UP = U$, a result which will be used without reference.

Theorem: 3.1.10 U is P-unitary.

Proof

$$U^TU \overset{D\,3.1.9}{=} \{P(I+S)K(I-S)P\}^T\{P(I+S)K(I-S)P\}$$
$$= P(I-S)K(I+S)PP(I+S)K(I-S)P \overset{3.2\,a}{=} P(I-S)K(I+S)P(I+S)K(I-S)P$$
$$\overset{3.1.1\,e}{=} P(I-S)K(I+S)K(I-S)P \overset{3.1.1\,c}{=} P(I-S)K^2(I-S)P$$
$$\overset{3.2\,a}{=} P(I-S)K(I-S)P \overset{3.1.1\,d}{=} P(I-S)P$$
$$\overset{3.1.1\,c}{=} P^2 \overset{3.2\,a}{=} P.$$

Similarly $UU^T = P$. □

In other words, with $R(P) = \{P\boldsymbol{z} : \boldsymbol{z} \in \Re^{2m}\}, U$ when restricted to domain $R(P)$, is a unitary function, so any vector $\boldsymbol{z}$ in the range of P (i.e. $P\boldsymbol{z} = \boldsymbol{z}$) will be mapped by U to a vector with the same norm - that is $P\boldsymbol{z} = \boldsymbol{z} \Rightarrow \|U\boldsymbol{z}\|^2 = (U\boldsymbol{z})^T(U\boldsymbol{z}) = \boldsymbol{z}^T U^T U\boldsymbol{z} = \boldsymbol{z}^T P\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z} = \|\boldsymbol{z}\|^2$.

One final result is needed for the next section:

Lemma 3.1.11 $\boldsymbol{z} = U\boldsymbol{z} \Leftrightarrow (\boldsymbol{z} = P\boldsymbol{z}) \wedge (\boldsymbol{z}^T U\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z})$.

Proof: $(\boldsymbol{z} = P\boldsymbol{z}) \wedge (\boldsymbol{z}^T U\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z})$

$\Leftrightarrow (\|\boldsymbol{z} - U\boldsymbol{z}\|^2 = \boldsymbol{z}^T\boldsymbol{z} - \boldsymbol{z}^T U\boldsymbol{z} - \boldsymbol{z}^T U^T\boldsymbol{z} + \boldsymbol{z}^T U^T U\boldsymbol{z}) \wedge (\boldsymbol{z} = P\boldsymbol{z}) \wedge (\boldsymbol{z}^T U\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z})$

$\Leftrightarrow (\|\boldsymbol{z} - U\boldsymbol{z}\|^2 = \boldsymbol{z}^T\boldsymbol{z} - 2\boldsymbol{z}^T U\boldsymbol{z} + \boldsymbol{z}^T U^T U\boldsymbol{z}) \wedge (\boldsymbol{z} = P\boldsymbol{z}) \wedge (\boldsymbol{z}^T U\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z})$

$\Leftrightarrow (\|\boldsymbol{z} - U\boldsymbol{z}\|^2 = \boldsymbol{z}^T\boldsymbol{z} - 2\boldsymbol{z}^T\boldsymbol{z} + \boldsymbol{z}^T U^T U\boldsymbol{z}) \wedge (\boldsymbol{z} = P\boldsymbol{z}) \wedge (\boldsymbol{z}^T U\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z})$

$\overset{T\,3.1.10}{\Leftrightarrow} (\|\boldsymbol{z} - U\boldsymbol{z}\|^2 = \boldsymbol{z}^T\boldsymbol{z} - 2\boldsymbol{z}^T\boldsymbol{z} + \boldsymbol{z}^T P\boldsymbol{z}) \wedge (\boldsymbol{z} = P\boldsymbol{z}) \wedge (\boldsymbol{z}^T U\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z})$

$\Leftrightarrow (\|\boldsymbol{z} - U\boldsymbol{z}\|^2 = \boldsymbol{z}^T\boldsymbol{z} - 2\boldsymbol{z}^T\boldsymbol{z} + \boldsymbol{z}^T\boldsymbol{z}) \wedge (\boldsymbol{z} = P\boldsymbol{z}) \wedge (\boldsymbol{z}^T U\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z})$

$\Leftrightarrow (\|\boldsymbol{z} - U\boldsymbol{z}\|^2 = 0) \wedge (\boldsymbol{z} = P\boldsymbol{z}) \wedge (\boldsymbol{z}^T U\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z})$

$\Leftrightarrow \boldsymbol{z} = U\boldsymbol{z}$. □

3.1.5 Characterization

Define the averaging matrix

Definition: 3.1.12 V = (P + U)/2 .

Note that $PV = VP = V$ and this result will be used without reference.

With reference to Figure 3.1 we see that $V\boldsymbol{z}$ is sandwiched between $\boldsymbol{z}$ and $U\boldsymbol{z}$, and that

Lemma 3.1.13 $\boldsymbol{z} = V\boldsymbol{z} \Leftrightarrow \boldsymbol{z} = U\boldsymbol{z}$

Proof: $\boldsymbol{z} = V\boldsymbol{z} \Leftrightarrow (\boldsymbol{z} = (P+U)\boldsymbol{z}/2) \wedge (\boldsymbol{z} = P\boldsymbol{z}) \Leftrightarrow (\boldsymbol{z} = U\boldsymbol{z}) \wedge (\boldsymbol{z} = P\boldsymbol{z}) \Leftrightarrow \boldsymbol{z} = U\boldsymbol{z}$. □

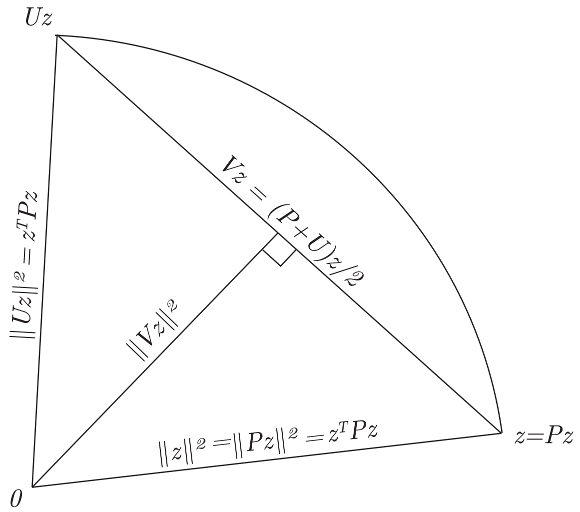


Figure 3.1: **Equivalent Fixed-Point Conditions**

In view of Lemma 3.1.5 we have

Corollary: 3.1.14 $K\boldsymbol{z} = \boldsymbol{z} \Leftrightarrow \overline{K}\boldsymbol{z} = \boldsymbol{z}$ □

Lemma 3.1.15 $\boldsymbol{z}$ is a fixed-point of U iff $\boldsymbol{z}$ is a fixed-point of P and K.

Proof: $U\boldsymbol{z} = \boldsymbol{z} \overset{3.1.11}{\Leftrightarrow} (\boldsymbol{z}^T U\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}) \wedge (P\boldsymbol{z} = \boldsymbol{z}) \overset{3.1.9}{\Leftrightarrow} (\boldsymbol{z}^T P(I+S)K(I-S)P\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}) \wedge (P\boldsymbol{z} = \boldsymbol{z})$
$\Leftrightarrow (\boldsymbol{z}^T PKP\boldsymbol{z} + \boldsymbol{z}^T PSKP\boldsymbol{z} - \boldsymbol{z}^T PKSP\boldsymbol{z} - \boldsymbol{z}^T PSKSP\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}) \wedge (P\boldsymbol{z} = \boldsymbol{z})$
$\Leftrightarrow (\boldsymbol{z}^T PKP\boldsymbol{z} - \boldsymbol{z}^T PSKSP\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}) \wedge (P\boldsymbol{z} = \boldsymbol{z})$
$\Leftrightarrow (\boldsymbol{z}^T K\boldsymbol{z} - \boldsymbol{z}^T SKS\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}) \wedge (P\boldsymbol{z} = \boldsymbol{z})$
$\overset{3.1}{\Leftrightarrow} (\boldsymbol{z}^T K\boldsymbol{z} - \boldsymbol{z}^T(I-K)\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}) \wedge (P\boldsymbol{z} = \boldsymbol{z})$
$\Leftrightarrow (\boldsymbol{z}^T K\boldsymbol{z} - \boldsymbol{z}^T\boldsymbol{z} + \boldsymbol{z}^T K\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}) \wedge (P\boldsymbol{z} = \boldsymbol{z})$
$\Leftrightarrow (\boldsymbol{z}^T K\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}) \wedge (P\boldsymbol{z} = \boldsymbol{z})$
$\overset{L\,2.2.21}{\Leftrightarrow} (K\boldsymbol{z} = \boldsymbol{z}) \wedge (P\boldsymbol{z} = \boldsymbol{z})$. □

Lemma 3.1.16 $\boldsymbol{z}$ is a fixed-point of V iff $\|\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2$.

Proof: $\boldsymbol{z} = V\boldsymbol{z} \overset{L\,3.1.13}{\Leftrightarrow} \boldsymbol{z} = U\boldsymbol{z} \overset{L\,2.2.1}{\Leftrightarrow} (\boldsymbol{z}^T\boldsymbol{z} = \boldsymbol{z}^T U\boldsymbol{z}) \wedge (\boldsymbol{z} = P\boldsymbol{z})$

$$\Leftrightarrow 2\boldsymbol{z}^T\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z} + \boldsymbol{z}^T U\boldsymbol{z} \wedge \boldsymbol{z} = P\boldsymbol{z} \Leftrightarrow \boldsymbol{z}^T\boldsymbol{z} = (\boldsymbol{z}^T\boldsymbol{z} + \boldsymbol{z}^T U\boldsymbol{z})/2 \wedge \boldsymbol{z} = P\boldsymbol{z}$$

$$\Leftrightarrow \boldsymbol{z}^T\boldsymbol{z} = (2\boldsymbol{z}^T\boldsymbol{z} + 2\boldsymbol{z}^T U\boldsymbol{z})/4 \wedge \boldsymbol{z} = P\boldsymbol{z} \Leftrightarrow \boldsymbol{z}^T\boldsymbol{z} = (\boldsymbol{z}^T\boldsymbol{z} + \boldsymbol{z}^T U^T\boldsymbol{z} + \boldsymbol{z}^T U\boldsymbol{z} + \boldsymbol{z}^T U^T U\boldsymbol{z})/4 \wedge \boldsymbol{z} = P\boldsymbol{z}$$

$$\Leftrightarrow \boldsymbol{z}^T\boldsymbol{z} = ((I+U)\boldsymbol{z}/2)^T((I+U)\boldsymbol{z}/2) \wedge \boldsymbol{z} = P\boldsymbol{z} \Leftrightarrow \|\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2 \ . \ \square$$

Summing up Corollary 3.1.14 and the previous three lemmas we have:

Theorem: 3.1.17
$\boldsymbol{z}$ is a fixed-point of U
$\Leftrightarrow$ $\boldsymbol{z}$ is a fixed-point of V
$\Leftrightarrow$ $\boldsymbol{z}$ is a fixed-point of P and K
$\Leftrightarrow$ $\boldsymbol{z}$ is a fixed-point of P and $\overline{K}$
$\Leftrightarrow \|\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2$.

This theorem is illustrated by Figure 3.2.

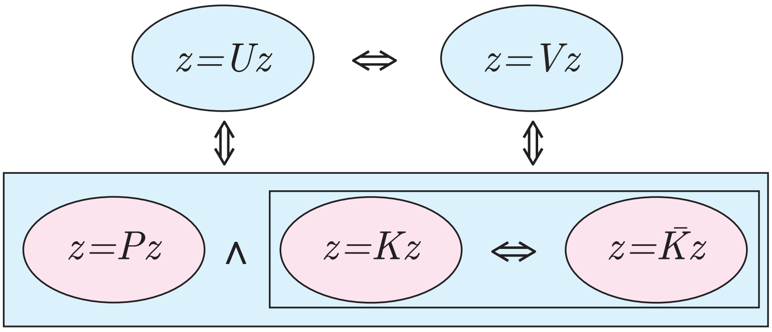


Figure 3.2: **Relationship Between Fixed-Point Conditions**

3.1.6 Convergence

We show that $\{V^n\boldsymbol{z}\}$ is a Cauchy sequence by showing that

1. the series is strictly decreasing in norm, and
2. the distance between successive terms is strictly decreasing and, moreover,
3. the distance between successive terms is bounded by a decreasing geometric series.

Lemma 3.1.18 $\|P\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2 + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2$

Proof: $\|P\boldsymbol{z}\|^2 = \|V\boldsymbol{z} + P\boldsymbol{z} - V\boldsymbol{z}\|^2$
$= \|V\boldsymbol{z}\|^2 + 2(V\boldsymbol{z})^T(P\boldsymbol{z} - V\boldsymbol{z}) + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2 + \boldsymbol{z}^T(P+U^T)/2(P\boldsymbol{z} - (P+U)\boldsymbol{z}/2) + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2$
$= \|V\boldsymbol{z}\|^2 + (\boldsymbol{z}^T(P+U^T)/2)(P\boldsymbol{z} - U\boldsymbol{z})/2 + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2 + \boldsymbol{z}^T(P+U^T)(P-U)\boldsymbol{z}/4 + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2$
$= \|V\boldsymbol{z}\|^2 + \boldsymbol{z}^T(P - U + U^T - P)\boldsymbol{z}/4 + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2 + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2$. $\square$

Corollary: 3.1.19 (a) $\|\boldsymbol{z}\|^2 = \|(I-P)\boldsymbol{z}\|^2 + \|V\boldsymbol{z}\|^2 + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2$
(b) $\|\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2 \Leftrightarrow \boldsymbol{z} = V\boldsymbol{z}$

Proof: (a) $\|\boldsymbol{z}\|^2 = \|(I-P)\boldsymbol{z}+P\boldsymbol{z}\|^2 = \|(I-P)\boldsymbol{z}\|^2 + \|P\boldsymbol{z}\|^2 \overset{L\,3.1.18}{=} \|(I-P)\boldsymbol{z}\|^2 + \|V\boldsymbol{z}\|^2 + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2$.
(b) $\|\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2 \overset{a}{\Rightarrow} \|(I-P)\boldsymbol{z}\|^2 + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2 = 0 \Rightarrow (\boldsymbol{z}=P\boldsymbol{z}) \wedge (P\boldsymbol{z}=V\boldsymbol{z}) \Rightarrow \boldsymbol{z}=V\boldsymbol{z} \Rightarrow \|\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2$.

Corollary: 3.1.20 $\|\boldsymbol{z} - V\boldsymbol{z}\|^2 = \|\boldsymbol{z}\|^2 - \|V\boldsymbol{z}\|^2$

Proof: (C 3.1.19) $\Rightarrow \|\boldsymbol{z}\|^2 - \|V\boldsymbol{z}\|^2 = \|(I-P)\boldsymbol{z}\|^2 + \|P\boldsymbol{z} - V\boldsymbol{z}\|^2$

$$\Rightarrow \|\boldsymbol{z}\|^2 - \|V\boldsymbol{z}\|^2 = \|(I-P+P-V)\boldsymbol{z}\|^2 = \|(I-V)\boldsymbol{z}\|^2$$

$$\Rightarrow \|\boldsymbol{z}\|^2 - \|V\boldsymbol{z}\|^2 = \|\boldsymbol{z} - V\boldsymbol{z}\|^2 \;. \;\square$$

Figure 3.1 and Pythagoras' Theorem make the above lemma obvious.

The series $\|\{V^n\boldsymbol{z}\|^2\}$ decreases strictly monotonically:

Lemma 3.1.21 (a) $\|\boldsymbol{z}\|^2 \geq \|P\boldsymbol{z}\|^2 \geq \|V\boldsymbol{z}\|^2$
(b) $\|\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2$ iff $\boldsymbol{z} = V\boldsymbol{z}$

Proof: (a) $\|\boldsymbol{z}\|^2 \overset{C\,2.2.20}{\geq} \|P\boldsymbol{z}\|^2 \overset{L\,3.1.18}{=} \|V\boldsymbol{z}\|^2 + \|Pv - V\boldsymbol{z}\|^2 \geq \|V\boldsymbol{z}\|^2$.
(b) $\|\boldsymbol{z}\|^2 = \|V\boldsymbol{z}\|^2 \Leftrightarrow \|\boldsymbol{z}\|^2 - \|V\boldsymbol{z}\|^2 = 0 \overset{C\,3.1.20}{\Leftrightarrow} \|\boldsymbol{z} - V\boldsymbol{z}\|^2 = 0 \Leftrightarrow \boldsymbol{z} = V\boldsymbol{z}$.

The series $\|\{V^{n+1}\boldsymbol{z} - V^n\boldsymbol{z}\|^2\}$ decreases strictly monotonically:

Lemma 3.1.22 With $P\boldsymbol{z} = \boldsymbol{z}$,
(a) $\|\boldsymbol{z} - V\boldsymbol{z}\|^2 \geq \|V\boldsymbol{z} - V^2\boldsymbol{z}\|^2$
(b) $\|\boldsymbol{z} - V\boldsymbol{z}\|^2 = \|V\boldsymbol{z} - V^2\boldsymbol{z}\|^2 \Leftrightarrow \boldsymbol{z} = V\boldsymbol{z}$

Proof: (a) $\|\boldsymbol{z} - V\boldsymbol{z}\|^2 \overset{L\,3.1.21\,a}{\geq} \|V(\boldsymbol{z} - V\boldsymbol{z})\|^2 = \|V\boldsymbol{z} - V^2\boldsymbol{z}\|^2$
(b) $\|\boldsymbol{z}-V\boldsymbol{z}\|^2 = \|V\boldsymbol{z}-V^2\boldsymbol{z}\|^2 \overset{L\,3.1.21\,b}{\Leftrightarrow} \boldsymbol{z}-V\boldsymbol{z} = U(\boldsymbol{z}-V\boldsymbol{z}) \Leftrightarrow P\boldsymbol{z}-(U+P)\boldsymbol{z}/2 = U(P\boldsymbol{z}-(U+P)\boldsymbol{z}/2) \Leftrightarrow U\boldsymbol{z} = U^2\boldsymbol{z} \Leftrightarrow U^TU\boldsymbol{z} = U^TU^2\boldsymbol{z} \Leftrightarrow P\boldsymbol{z} = PU\boldsymbol{z} \Leftrightarrow P\boldsymbol{z} = U\boldsymbol{z} \Leftrightarrow 2P\boldsymbol{z} = (P+U)\boldsymbol{z} \Leftrightarrow P\boldsymbol{z} = V\boldsymbol{z} \Leftrightarrow \boldsymbol{z} = V\boldsymbol{z}$. $\square$

The distance between successive terms is bounded by a decreasing geometric series:

To establish this we first show that V and V^T commute, and also compute their product:

Lemma 3.1.23 $V^TV = VV^T = (U^T + 2P + U)/4 = PKP$

Proof: $VV^T = ((P+U)/2)((P+U)/2)^T = (P+U)(P+U^T)/4 = (P+U+U^T+UU^T)/4 = (P+U+U^T+U^TU)/4 = ((P+U)/2)^T((P+U)/2) = V^TV = ((P+U)/2)^T(P+U)/2 = (P+U^T+U+P)/4 = (2P+P(I+S)K(I-S)P+P(I-S)K(I+S)P)/4 = (2P+PKP+PSKP-PKSP-PSKSP+PKP-PSKP+PKSP-PSKSP)/4 = (2P+2PKP-2PSKSP)/4 = (2P+2PKP-2P(I-K)P)/4 = (2P+2PKP-2P+2PKP)/4 = PKP$. $\square$

The following results connect the developed theory with the theory of *reciprocal spaces* [3], and the product of the idempotents P and K.

We note that

$$(V^T)^i V^i = (PK)^i P = P(KP)^i, \tag{3.9}$$

since $(V^T)^i V^i \overset{L\,3.1.23}{=} (V^T V)^i = (PKP)^i = (PK)^i P = P(KP)^i$, so for $\boldsymbol{z} = P\boldsymbol{z}$, $\|V^i \boldsymbol{z}\|^2 = (V^i \boldsymbol{z})^T V^i \boldsymbol{z} = \boldsymbol{z}^T (V^T)^i V^i \boldsymbol{z} = \boldsymbol{z}^T (PK)^i P\boldsymbol{z} = \boldsymbol{z}^T (PK)^i \boldsymbol{z} = \boldsymbol{z}^T (KP)^i \boldsymbol{z}$, that is

$$\text{For } \boldsymbol{z} = P\boldsymbol{z}, \|V^i \boldsymbol{z}\|^2 = \boldsymbol{z}^T (PK)^i \boldsymbol{z} = \boldsymbol{z}^T (KP)^i \boldsymbol{z} \ . \tag{3.10}$$

In Appendix A we consider reciprocal vectors and spaces; we show by spectral decomposition of the matrix PK that the convergence in norm of the sequence $(PK)^n x$ is better than geometric, so it follows that the series $\{V^n \boldsymbol{z}\}$ is a Cauchy sequence, whose limit is a fixed point of U.

3.1.7 Fixed-Point Component

Here we develop some results on the P-unitary matrix U and give them a geometric interpretation.

Lemma 3.1.24 (a) The P component of U^n (i.e $(U^n)^T P U^n$) remains constant at P
(b) The K component of U^n remains constant at PKP
(that is $(U^n)^T P (U^n) = P$ and $U^{nT} K U^n = PKP$)

Proof: (a) $U^T P U = U^T U \overset{T\,3.1.10}{=} P$, further, $(U^n)^T P U^n = (U^n)^T U^n = (U^{n-1})^T U^T U U^{n-1} = (U^{n-1})^T P U^{n-1} = (U^{n-1})^T U^{n-1} = \cdots = U^T U = P$.

(b) For $n = 1$, $U^T K U = U^T PKPU \overset{D\,3.1.9}{=} \{P(I+S)K(I-S)P\}^T \, PKP \, \{P(I+S)K(I-S)P\}$

$= P(I-S)K(I+S)P^2 KP^2 (I+S)K(I-S)P = P(I-S)[K(I+S)PK]P(I+S)K(I-S)P$

$\overset{3.6\,b'}{=} P(I-S)[KP(I-S)K]P(I+S)K(I-S)P = P(I-S)KP(I-S)[KP(I+S)K](I-S)P$

$\overset{3.6\,a'}{=} P(I-S)KP(I-S)[K(I-S)PK](I-S)P = P(I-S)KP[(I-S)K(I-S)]PK(I-S)P$

$\overset{3.1.1\,d}{=} P(I-S)K[P(I-S)P]K(I-S)P$

$\overset{3.1.1\,c}{=} P(I-S)KP^2 K(I-S)P = [P(I-S)KP][PK(I-S)P]$

$\overset{3.6}{=} PK[(I+S)P(I+S)]KP \overset{3.1.1\,e}{=} PK(I+S)KP$

$\overset{3.1.1c}{=} PK^2 P \overset{3.2\,a}{=} PKP$.

For $n \geq 2, U^{nT} K U^n = U^{(n-1)T} U^T K U U^{n-1} = U^{(n-1)T} U^T PKPU U^{n-1}$

$= U^{(n-1)T} PKPU^{n-1} = \cdots = PKP.$ □

3.2 Computational Methods

3.2.1 Averaging

Theorem 3.1.17 is the basis for this method.

With $\mathbf{1} = [1 \cdots 1]^T$ we consider the sequence $\{g_i\}$ where

$$\begin{aligned} g_1 &= P\mathbf{1} \ , && \text{(a)} \\ g_{i+1} &= Ug_i \ , && \text{(b)} \end{aligned} \tag{3.11}$$

and the sequence $\{v_i\}$ where

$$v_{i+1} = Vv_i \ , \quad v_1 = P\mathbf{1} \ .$$

Note that

$$v_1 = g_1 \tag{3.12}$$

Since U is P-unitary it follows that $\|g_n\|$ is constant; on the other hand we show that $\{v_i\}$ converges:

$$\|v_i - v_{i+j}\| = \|\sum_{k=0}^{j-1}(v_{i+k} - v_{i+k+1})\|$$

$$= \|\sum_{k=0}^{j-1} V^k(v_i - v_{i+1})\| \leq \sum_{k=0}^{j-1} \|V^k(v_i - v_{i+1})\|$$

$$\leq \sum_{k=0}^{j-1} \sqrt{(v_i - v_{i+1})^T V^{kT} V^k (v_i - v_{i+1})}$$

$$\overset{3.10}{=} \sum_{k=0}^{j-1} \sqrt{(v_i - v_{i+1})^T (PK)^k (v_i - v_{i+1})}$$

$$\leq \sum_{k=0}^{j-1} \sqrt{\|v_i - v_{i+1}\| \lambda^{2k} \|v_i - v_{i+1}\|}$$

where λ is the largest eigenvalue of PK other than unity (in Appendix A we show that the eigenvalues of PK lie in the interval [-1,1]).

$$\leq \left(\sum_{k=0}^{\infty} \lambda^k\right) \|v_i - v_{i+1}\| = \frac{1}{1-\lambda}\|v_i - v_{i+1}\|$$

That is $\|v_i - v_{i+j}\| \leq \frac{1}{1-\lambda}\|v_i - v_{i+1}\|$, so

$$\|v_i - v_{i+j}\|^2 \leq \frac{1}{(1-\lambda)^2}\|v_i - v_{i+1}\|^2$$

and so, using Corollary 3.1.20,

$$\|v_i - v_{i+j}\|^2 = \frac{1}{(1-\lambda)^2}(\|v_i\|^2 - \|v_{i+1}\|^2) \tag{3.13}$$

and since $\|v_i\|^2$ is a Cauchy sequence, it follows that $\{v_i\}$ is also a Cauchy sequence, which necessarily converges; we set $v_\infty = \lim_{i\to\infty} v_i$ and obviously $\forall i \geq 0 : Uv_\infty = v_\infty$.

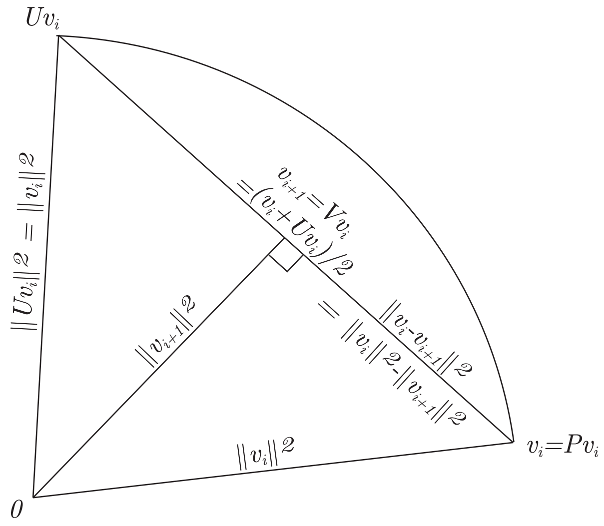


Figure 3.3: **Distance Between Successive Vectors**

3.2.2 Affine Regression

An affine sub-space of a vector space V is a set of the form $\boldsymbol{x}+S$ where S is a sub-space of V. Alternatively we can define an affine space as a set which is closed under the binary operation $+_\lambda$ for all λ, where $\boldsymbol{x} +_\lambda \boldsymbol{y} = \lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}$.

An affine sub-space of an inner product space has a unique point of minimum norm, since if we choose and arbitrary point a and form an infinite sequence of vectors of strictly decreasing norm beginning with a they will all lie in $A = \{\boldsymbol{x} : \|\boldsymbol{x}\| \leq \|a\|\}$ and since A is closed and bounded and hence compact this sequence will have an accumulation point which, in view of the continuity of the norm, will have minimum norm, so there exists a point of minimum norm. Now suppose there are two vectors of minimum norm, say w_1 and w_2, then the point $(w_1 + w_2)/2$ which is also in the affine space has a smaller norm than both w_1 and w_2, contradicting their minimality; therefor there is a unique point of minimum norm.

Let g_1 be a vector in the range of P, and $\boldsymbol{p}$ be a fixed-point of U and thus also in the range of P such that $g_1 \cdot \boldsymbol{p} > 0$. (in Chapter 6.4 we construct a vector which satisfies this requirement for a non-negative fixed-point of the specific unitary operator U introduced in Chapter 6.1.4). We consider the affine space

$$\mathcal{A}_j = \left\{ \sum_{i=1}^{j} \lambda_i g_i : \sum_{i=1}^{j} \lambda_i = 1 \right\} .$$

Obviously $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \mathcal{A}_3 \cdots$, moreover it can be shown that $\mathcal{A}_{i+1} = \mathcal{A}_i \Rightarrow \mathcal{A}_{i+2} = \mathcal{A}_i$, so we define

$$\mathcal{A} = \mathcal{A}_j \quad \textit{where} \quad j \quad \textit{is such that} \quad \mathcal{A}_j = \mathcal{A}_{j+1}$$

Take $g_1 = \boldsymbol{p}_1 + q_1$, where $\boldsymbol{p}_1 = g_1 \rhd \boldsymbol{p}$ and $q_1 = g_1 \lhd \boldsymbol{p}$ which is an orthogonal decomposition in view of Equation 2.13, and since $g_1 \cdot \boldsymbol{p} > 0$. So $\mathcal{A}$ is not equal to the range of P since each of its elements is of the form $\boldsymbol{p}_1 + r$, where r is orthogonal to $\boldsymbol{p}$, and thus $\mathcal{A}$ is a proper affine space. It follows that the vector of smallest norm in $\mathcal{A}$ (not necessarily $\boldsymbol{p}$) is a fixed-point of U ; this vector may be computed following the approach of Chapter 2.4.

Note that $\boldsymbol{v}_1 = g_1$, so $\boldsymbol{v}_1 \in \mathcal{A}$, also that $\boldsymbol{v}_i \in \mathcal{A} \Rightarrow \boldsymbol{v}_i \in \mathcal{A}_j\, \exists j \Rightarrow \boldsymbol{v}_{i+1} \in \mathcal{A}_{j+1} \Rightarrow \boldsymbol{v}_{i+1} \in \mathcal{A}$; since $\mathcal{A}$ is compact it follows that $\boldsymbol{v}_\infty \in \mathcal{A}$.

Now $(\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty = (\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T P \boldsymbol{v}_\infty = (\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T U^T U \boldsymbol{v}_\infty = (\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T U^T \boldsymbol{v}_\infty = (U\boldsymbol{v}_i - U\boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty = (U\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty$ that is

$$(\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty = (U\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty \ .$$

Averaging this equation and the tautology $(\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty = (\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty$ yields

$$(\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty = (\boldsymbol{v}_{i+1} - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty$$

from which it follows that

$$\forall i \geq 1 : (\boldsymbol{v}_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty = 0 \ .$$

Further, $(g_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty = (g_i - \boldsymbol{v}_\infty)^T P \boldsymbol{v}_\infty = (g_i - \boldsymbol{v}_\infty)^T U^T U \boldsymbol{v}_\infty = [U(g_i - \boldsymbol{v}_\infty)]^T U \boldsymbol{v}_\infty = (g_{i+1} - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty$ that is

$$(g_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty = (g_{i+1} - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty \ .$$

Finally, since $g_1 = \boldsymbol{v}_1$, it follows that $(g_1 - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty = (\boldsymbol{v}_1 - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty$ and so

$$\forall i \geq 1 : (g_i - \boldsymbol{v}_\infty)^T \boldsymbol{v}_\infty = 0 \ .$$

Thus $\boldsymbol{v}_\infty$ is the unique point of minimum norm in the affine space $\mathcal{A}$.

From these computations it also follows that $g_n = U^n g_1$ orbits $\boldsymbol{v}_\infty$ at a fixed distance from $\boldsymbol{v}_\infty$ in the affine space $\mathcal{A}$, as shown in Figure 3.4.

The above calculations form the basis for an algorithm for computing a fixed-point as follows.

Let

$$G_i = [g_1 \cdots g_i] \tag{3.14}$$

then we minimize $\|G_i \boldsymbol{x}\|^2$ subject to $\mathbf{1}_i{}^T \boldsymbol{x} = 1$. The solution to this minimization problem was given by Equation 2.22, yielding

$$\boldsymbol{\sigma}_i = G_i x_s = \frac{G_i^{+T} \mathbf{1}_i}{\left\|G_i^{+T} \mathbf{1}_i\right\|^2} = \frac{G_i (G_i^T G_i)^+ \mathbf{1}_i}{\mathbf{1}_i{}^T (G_i^T G_i)^+ \mathbf{1}_i} \tag{3.15}$$

Note that in the above formulae the pseudo-inverse is used, but the ordinary inverse may obtain if $i \leq q$.

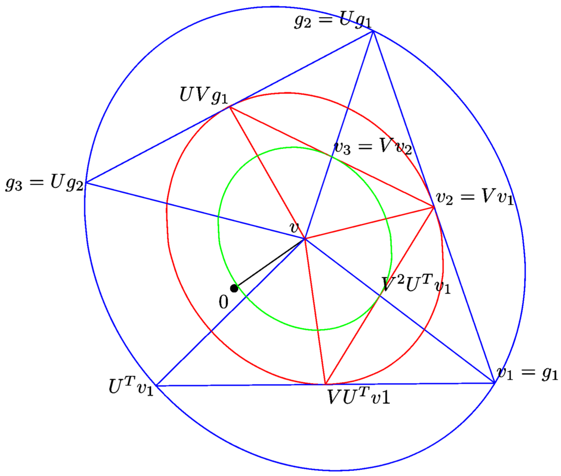


Figure 3.4: Orbit of a Vector

3.2.3 Ordered Spaces

Another method of solution of the general fixed-point problem involves using a vector lattice approach, which is detailed in this section. The results up to Corollary 3.2.5 may be found in the literature on vector lattices.

A *lattice*[1] is a triple $(L, \vee, \wedge)$ where $\vee : L^2 \to L$, and $\wedge : L^2 \to L$. These binary functions are normally written between their operands, and are required to satisfy

1. $x \vee x = x$; $x \wedge x = x$, (nilpotency)
2. $x_1 \vee x_2 = x_2 \vee x_1$; $x_1 \wedge x_2 = x_2 \wedge x_1$, (commutativity)
3. $x_1 \vee (x_2 \vee x_3) = (x_1 \vee x_2) \vee x_3$; $x_1 \wedge (x_2 \wedge x_3) = (x_1 \wedge x_2) \wedge x_3$, (associativity)
4. $x_1 \vee (x_1 \wedge x_2) = x_1$; $x_1 \wedge (x_1 \vee x_2) = x_1$, (absorptivity)

If, in addition

$$x_1 \vee (x_2 \wedge x_3) = (x_1 \vee x_2) \wedge (x_1 \vee x_3),$$

or

$$x_1 \wedge (x_2 \vee x_3) = (x_1 \wedge x_2) \vee (x_1 \wedge x_3),$$

[1] Some of the treatment here is an extension of that in [5].

holds, then the other can be shown to hold and the lattice is said to be *distributive.*

We can associate an *order relation* with a lattice by

$$x_1 \geq x_2 \ \textit{if}\ x_1 \vee x_2 = x_1.$$

or, if we have an order relation and $sup(x_1, x_2)$ and $inf(x_1, x_2)$ both exist for all x_1 and x_2, we take $x_1 \vee x_2 = sup(x_1, x_2)$ and $x_1 \wedge x_2 = inf(x_1, x_2)$, and can show that these functions satisfy the above four defining properties for a lattice.

3.2.3.1 Vector Lattices

We have an *ordered vector space* if

$$\forall \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \in L: \ \boldsymbol{x}_1 \geq \boldsymbol{x}_2 \Rightarrow \boldsymbol{x}_1 + \boldsymbol{x}_3 \geq \boldsymbol{x}_2 + \boldsymbol{x}_3,$$

and

$$\forall \alpha \in \Re^+ \cup \{0\} \ \textit{and}\ \boldsymbol{x}_1, \boldsymbol{x}_2 \in L: \ \boldsymbol{x}_1 \geq \boldsymbol{x}_2 \Rightarrow \alpha \boldsymbol{x}_1 \geq \alpha \boldsymbol{x}_2, \quad .$$

If the order on an ordered vector space defines a lattice then we have (as a definition) a *vector lattice.*

Note that we give priority to multiplication over $\vee$ and $\wedge$, and priority to $\vee$ and $\wedge$ over unary and binary $+$ and $-$. Thus, for example $-\boldsymbol{x}_1 \vee \boldsymbol{x}_2 = -(\boldsymbol{x}_1 \vee \boldsymbol{x}_2), \boldsymbol{x}_1 + \boldsymbol{x}_2 \wedge \boldsymbol{x}_3 = \boldsymbol{x}_1 + (\boldsymbol{x}_2 \wedge \boldsymbol{x}_3)$, and $\alpha \boldsymbol{x}_1 \vee \boldsymbol{x}_2 = (\alpha \boldsymbol{x}_1) \vee \boldsymbol{x}_2$.

For a vector lattice the following results hold

$$\begin{array}{lcll} \boldsymbol{x}_1 \vee \boldsymbol{x}_2 & = & -(-\boldsymbol{x}_1) \wedge (-\boldsymbol{x}_2) & \text{(a)} \\ \boldsymbol{x}_1 \vee \boldsymbol{x}_2 + \boldsymbol{x}_3 & = & (\boldsymbol{x}_1 + \boldsymbol{x}_3) \vee (\boldsymbol{x}_2 + \boldsymbol{x}_3) & \text{(b)} \\ \alpha(\boldsymbol{x}_1 \vee \boldsymbol{x}_2) & = & (\alpha \boldsymbol{x}_1) \vee (\alpha \boldsymbol{x}_2) & \text{(c)} \end{array} \tag{3.16}$$

These results and a proof of (b) (which provides the flavour for proving the others) are given in [5]. Dual results where $\vee$ is replaced by $\wedge$, and *vice versa* hold and will be denoted by a $'$ in the equation designator: Thus, for example,

$$\boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \quad = \quad -(-\boldsymbol{x}_1) \vee (-\boldsymbol{x}_2). \qquad \text{(a}'\text{)}$$

We define the positive and negative parts of a vector, and the absolute value of a vector as follows:

$$\begin{array}{lcll} \boldsymbol{x}^{\vee} & = & \boldsymbol{x} \vee 0 & \text{(a)} \\ \boldsymbol{x}^{\wedge} & = & \boldsymbol{x} \wedge 0 \ \ \textit{and} & \text{(b)} \\ |\boldsymbol{x}| & = & \boldsymbol{x}^{\vee} - \boldsymbol{x}^{\wedge} & \text{(c)} \end{array} \tag{3.17}$$

Lemma 3.2.1 $\boldsymbol{x} = \boldsymbol{x}^{\vee} + \boldsymbol{x}^{\wedge}$.

Proof $\boldsymbol{x} - \boldsymbol{x}^{\vee} = \boldsymbol{x} - \boldsymbol{x} \vee 0 = \boldsymbol{x} + [-(\boldsymbol{x} \vee 0)] = \boldsymbol{x} + (-\boldsymbol{x}) \wedge 0 = (\boldsymbol{x} - \boldsymbol{x}) \wedge (\boldsymbol{x} + 0) = 0 \wedge \boldsymbol{x} = \boldsymbol{x}^{\wedge}$. □

Thus $\boldsymbol{x} - \boldsymbol{x}^{\vee} = \boldsymbol{x}^{\wedge}$, which implies $\boldsymbol{x} = \boldsymbol{x}^{\vee} + \boldsymbol{x}^{\wedge}$. $\square$

Lemma 3.2.2 $\boldsymbol{x}_1 + \boldsymbol{x}_2 = \boldsymbol{x}_1 \vee \boldsymbol{x}_2 + \boldsymbol{x}_1 \wedge \boldsymbol{x}_2$.

Proof From the previous lemma,

$\boldsymbol{x}_1 - \boldsymbol{x}_2 = (\boldsymbol{x}_1 - \boldsymbol{x}_2)^{\vee} + (\boldsymbol{x}_1 - \boldsymbol{x}_2)^{\wedge}$

$\Rightarrow \boldsymbol{x}_1 + \boldsymbol{x}_2 = (\boldsymbol{x}_1 - \boldsymbol{x}_2) \vee 0 + \boldsymbol{x}_2 + (\boldsymbol{x}_1 - \boldsymbol{x}_2) \wedge 0 + \boldsymbol{x}_2$

$\overset{3.16\,b}{\Rightarrow} \boldsymbol{x}_1 + \boldsymbol{x}_2 = \boldsymbol{x}_1 \vee \boldsymbol{x}_2 + \boldsymbol{x}_1 \wedge \boldsymbol{x}_2$. $\square$ [2]

3.2.3.2 Hilbert Lattices

Consider a vector lattice on which a norm is defined - that is a map $\| \; \| : V \to \Re$; we require

$$\||\boldsymbol{x}|\| = \|\boldsymbol{x}\|$$

in which case the norm is said to be a *lattice norm*, and the lattice is a *normed vector lattice*. if the lattice is complete with respect to the norm then it is said to be a *Banach lattice.*

Consider a normed vector lattice on which an inner product is defined and for which the norm is defined by $\|\boldsymbol{x}_1\| = \sqrt{\boldsymbol{x}_1 \cdot \boldsymbol{x}_1}$; if the vector space is complete for the topology induced by the norm (i.e. a Banach lattice) then such a lattice is called a *Hilbert lattice*. The inner product of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ will be denoted by $\boldsymbol{x}_1 \cdot \boldsymbol{x}_2$, or where the vectors are cartesian by $\boldsymbol{x}_1^T \boldsymbol{x}_2$. If $\boldsymbol{x}_1 \cdot \boldsymbol{x}_2 = 0$ then we will say that $\boldsymbol{x}_1$ is orthogonal to $\boldsymbol{x}_2$, or $\boldsymbol{x}_1 \perp \boldsymbol{x}_2$.

Note: we require $\boldsymbol{x}_1, \boldsymbol{x}_2 \geq 0 \Rightarrow \boldsymbol{x}_1 \cdot \boldsymbol{x}_2 \geq 0$ and the existence of a swapping-matrix S which is *consonant* with the lattice order, that is $\boldsymbol{x}_1 \geq \boldsymbol{x}_2 \Rightarrow S\boldsymbol{x}_1 \geq S\boldsymbol{x}_2$ from which it follows that

$$\begin{aligned} S(\boldsymbol{x}_1 \vee \boldsymbol{x}_2) &= (S\boldsymbol{x}_1) \vee (S\boldsymbol{x}_2) && \text{(a)} \\ S(\boldsymbol{x}_1 \wedge \boldsymbol{x}_2) &= (S\boldsymbol{x}_1) \wedge (S\boldsymbol{x}_2) && \text{(b)} \end{aligned} \tag{3.18}$$

Lemma 3.2.3 For a normed vector lattice $\boldsymbol{x}^{\vee} \perp \boldsymbol{x}^{\wedge}$.

Proof We have $\||\boldsymbol{x}|\| = \|\boldsymbol{x}\| \;\; \Rightarrow \;\; \||\boldsymbol{x}|\|^2 = \|\boldsymbol{x}\|^2$

$$\Rightarrow \;\; |\boldsymbol{x}|^T |\boldsymbol{x}| = \boldsymbol{x}^T \boldsymbol{x} \overset{L\,3.2.1}{=} (\boldsymbol{x}^{\vee} + \boldsymbol{x}^{\wedge}) \cdot (\boldsymbol{x}^{\vee} + \boldsymbol{x}^{\wedge})$$

$$\overset{3.17\,c}{\Rightarrow} (\boldsymbol{x}^{\vee} - \boldsymbol{x}^{\wedge}) \cdot (\boldsymbol{x}^{\vee} - \boldsymbol{x}^{\wedge}) = (\boldsymbol{x}^{\vee} + \boldsymbol{x}^{\wedge}) \cdot (\boldsymbol{x}^{\vee} + \boldsymbol{x}^{\wedge})$$

$$\Rightarrow \;\; \boldsymbol{x}^{\vee} \cdot \boldsymbol{x}^{\vee} - 2\boldsymbol{x}^{\vee} \cdot \boldsymbol{x}^{\wedge} + \boldsymbol{x}^{\wedge} \cdot \boldsymbol{x}^{\wedge} = \boldsymbol{x}^{\vee} \cdot \boldsymbol{x}^{\vee} + 2\boldsymbol{x}^{\vee} \cdot \boldsymbol{x}^{\wedge} + \boldsymbol{x}^{\wedge} \cdot \boldsymbol{x}^{\wedge}$$

$$\Rightarrow \;\; -2\boldsymbol{x}^{\vee} \cdot \boldsymbol{x}^{\wedge} = 2\boldsymbol{x}^{\vee} \cdot \boldsymbol{x}^{\wedge} \;\; \Rightarrow \;\; \boldsymbol{x}^{\vee} \cdot \boldsymbol{x}^{\wedge} = 0 \;\; \Rightarrow \;\; \boldsymbol{x}^{\vee} \perp \boldsymbol{x}^{\wedge} \; . \; \square$$

Lemma 3.2.4 $\boldsymbol{x}_1 \vee \boldsymbol{x}_2 \cdot \boldsymbol{x}_1 \wedge \boldsymbol{x}_2 = \boldsymbol{x}_1 \cdot \boldsymbol{x}_2$

[2] A more general approach is given in [5, p. 197].

Proof $\boldsymbol{x}_1 \vee \boldsymbol{x}_2 \cdot \boldsymbol{x}_1 \wedge \boldsymbol{x}_2 = (\boldsymbol{x}_1 - \boldsymbol{x}_2 + \boldsymbol{x}_2) \vee \boldsymbol{x}_2 \cdot (\boldsymbol{x}_1 - \boldsymbol{x}_2 + \boldsymbol{x}_2) \wedge \boldsymbol{x}_2$
$= ((\boldsymbol{x}_1 - \boldsymbol{x}_2) \vee 0 + \boldsymbol{x}_2) \cdot ((\boldsymbol{x}_1 - \boldsymbol{x}_2) \wedge 0 + \boldsymbol{x}_2) = ((\boldsymbol{x}_1 - \boldsymbol{x}_2)^{\vee} + \boldsymbol{x}_2) \cdot ((\boldsymbol{x}_1 - \boldsymbol{x}_2)^{\wedge} + \boldsymbol{x}_2)$
$\overset{L\,3.2.3}{=} (\boldsymbol{x}_1 - \boldsymbol{x}_2)^{\vee} \cdot \boldsymbol{x}_2 + (\boldsymbol{x}_1 - \boldsymbol{x}_2)^{\wedge} \cdot \boldsymbol{x}_2 + \boldsymbol{x}_2 \cdot \boldsymbol{x}_2 = ((\boldsymbol{x}_1 - \boldsymbol{x}_2)^{\vee} + (\boldsymbol{x}_1 - \boldsymbol{x}_2)^{\wedge}) \cdot \boldsymbol{x}_2 + \boldsymbol{x}_2 \cdot \boldsymbol{x}_2$
$\overset{L\,3.2.1}{=} (\boldsymbol{x}_1 - \boldsymbol{x}_2) \cdot \boldsymbol{x}_2 + \boldsymbol{x}_2 \cdot \boldsymbol{x}_2 = \boldsymbol{x}_1 \cdot \boldsymbol{x}_2 - \boldsymbol{x}_2 \cdot \boldsymbol{x}_2 + \boldsymbol{x}_2 \cdot \boldsymbol{x}_2 = \boldsymbol{x}_1 \cdot \boldsymbol{x}_2.$ □

Corollary: 3.2.5 For a normed vector lattice $\boldsymbol{x}_1 \vee \boldsymbol{x}_2 \perp \boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \Leftrightarrow \boldsymbol{x}_1 \perp \boldsymbol{x}_2$.

3.2.3.3 Positive Fixed-Points

Suppose S swaps P then, given fixed-point $\boldsymbol{z} = P\boldsymbol{z}$, from Lemma 3.1.4 $\boldsymbol{z} \perp S\boldsymbol{z}$, so the decomposition $\boldsymbol{z} + S\boldsymbol{z} \overset{L\,3.2.2}{=} \boldsymbol{z} \vee S\boldsymbol{z} + \boldsymbol{z} \wedge S\boldsymbol{z}$ is an orthogonal decomposition in view of Corollary 3.2.5. Multiplying each side of this equation by the idempotent P we have the decomposition

$$\boldsymbol{z} = P(\boldsymbol{z} \vee S\boldsymbol{z}) + P(\boldsymbol{z} \wedge S\boldsymbol{z}),$$

which is also an orthogonal decomposition in view of the following

Lemma 3.2.6 $\boldsymbol{z} = P\boldsymbol{z}$ and $(\boldsymbol{z} \vee S\boldsymbol{z}) \perp (\boldsymbol{z} \wedge S\boldsymbol{z}) \Rightarrow P(\boldsymbol{z} \vee S\boldsymbol{z}) \perp P(\boldsymbol{z} \wedge S\boldsymbol{z})$.

Proof:

$SP(\boldsymbol{z} \vee S\boldsymbol{z}) = SPS(\boldsymbol{z} \vee S\boldsymbol{z}) \overset{3.1}{=} (I - P)(\boldsymbol{z} \vee S\boldsymbol{z}) = (\boldsymbol{z} \vee S\boldsymbol{z}) - P(\boldsymbol{z} \vee S\boldsymbol{z}),$
so $SP(\boldsymbol{z} \vee S\boldsymbol{z}) = (\boldsymbol{z} \vee S\boldsymbol{z}) - P(\boldsymbol{z} \vee S\boldsymbol{z}),$
and taking the inner product of this equation w.r.t. $\boldsymbol{z} \wedge S\boldsymbol{z}$,
$(\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot SP(\boldsymbol{z} \vee S\boldsymbol{z}) = (\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot (\boldsymbol{z} \vee S\boldsymbol{z}) - (\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P(\boldsymbol{z} \vee S\boldsymbol{z})$
$\overset{C\,3.2.5}{\Rightarrow} (\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot SP(\boldsymbol{z} \vee S\boldsymbol{z}) = -(\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P(\boldsymbol{z} \vee S\boldsymbol{z})$
$\overset{3.2\,d}{\Rightarrow} (\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot S^T P(\boldsymbol{z} \vee S\boldsymbol{z}) = -(\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P(\boldsymbol{z} \vee S\boldsymbol{z})$
$\overset{3.4\,a}{\Rightarrow} S(\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P(\boldsymbol{z} \vee S\boldsymbol{z}) = -(\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P(\boldsymbol{z} \vee S\boldsymbol{z})$
$\overset{3.18\,b}{\Rightarrow} (\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P(\boldsymbol{z} \vee S\boldsymbol{z}) = -(\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P(\boldsymbol{z} \vee S\boldsymbol{z})$
$\Rightarrow (\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P(\boldsymbol{z} \vee S\boldsymbol{z}) = 0 \;\; \Rightarrow (\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P^2(\boldsymbol{z} \vee S\boldsymbol{z}) = 0$
$\Rightarrow (\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P^T P(\boldsymbol{z} \vee S\boldsymbol{z}) = 0 \overset{3.4}{\Rightarrow} P(\boldsymbol{z} \wedge S\boldsymbol{z}) \cdot P(\boldsymbol{z} \vee S\boldsymbol{z}) = 0$
$\Rightarrow P(\boldsymbol{z} \vee S\boldsymbol{z}) \perp P(\boldsymbol{z} \wedge S\boldsymbol{z}).$

Summing up:

$$\boldsymbol{z} = P\boldsymbol{z} \overset{L\,3.1.4}{\Rightarrow} \boldsymbol{z} \perp S\boldsymbol{z} \overset{C\,3.2.5}{\Rightarrow} (\boldsymbol{z} \vee S\boldsymbol{z}) \perp (\boldsymbol{z} \wedge S\boldsymbol{z})$$

$$\overset{L\,3.2.6}{\Rightarrow} P(\boldsymbol{z} \vee S\boldsymbol{z}) \perp P(\boldsymbol{z} \wedge S\boldsymbol{z})$$

$$\overset{C\,3.2.5}{\Rightarrow} P(\boldsymbol{z} \wedge S\boldsymbol{z}) \;\vee\; P(\boldsymbol{z} \vee S\boldsymbol{z}) \;\perp\; P(\boldsymbol{z} \wedge S\boldsymbol{z}) \;\wedge\; P(\boldsymbol{z} \vee S\boldsymbol{z}) \;.$$

Note also that

$$P(\boldsymbol{z} \vee S\boldsymbol{z}) \;\vee\; P(\boldsymbol{z} \wedge S\boldsymbol{z}) \;+\; P(\boldsymbol{z} \vee S\boldsymbol{z}) \;\wedge\; P(\boldsymbol{z} \wedge S\boldsymbol{z})$$

$$= \; P(\boldsymbol{z} \vee S\boldsymbol{z}) \;+\; P(\boldsymbol{z} \wedge S\boldsymbol{z}) \;=\; P(\boldsymbol{z} \vee S\boldsymbol{z} + \boldsymbol{z} \wedge S\boldsymbol{z})$$

$$= \; P(\boldsymbol{z} + S\boldsymbol{z}) \;=\; \boldsymbol{z} \;;$$

so we have the orthogonal decomposition

$$\begin{aligned} \boldsymbol{z} &= \boldsymbol{z}_1 + \boldsymbol{z}_2 \,, \quad \text{where} && \text{(a)} \\ \boldsymbol{z}_1 &= P(\boldsymbol{z} \vee S\boldsymbol{z}) \;\vee\; P(\boldsymbol{z} \wedge S\boldsymbol{z}) \,, \quad \text{and} && \text{(b)} \\ \boldsymbol{z}_2 &= P(\boldsymbol{z} \vee S\boldsymbol{z}) \;\wedge\; P(\boldsymbol{z} \wedge S\boldsymbol{z}) \,. && \text{(c)} \end{aligned} \qquad (3.19)$$

Since $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are orthogonal,

$$\|\boldsymbol{z}_1 - \boldsymbol{z}_2\| \;=\; \|\boldsymbol{z}\| \,, \qquad (3.20)$$

and so

$$\|P(\boldsymbol{z}_1 - \boldsymbol{z}_2)\| \;\leq\; \|\boldsymbol{z}\| \,,$$

and so,

$$\|P\boldsymbol{z}_1 - P\boldsymbol{z}_2\| \;\leq\; \|\boldsymbol{z}\| \,.$$

Let $\boldsymbol{p}$ be a non-negative fixed-point of P, that is $\boldsymbol{p} \geq 0$ and $P\boldsymbol{p} = \boldsymbol{p}$. We now show that the $\boldsymbol{p}$-proportion of $P\boldsymbol{z}_1 - P\boldsymbol{z}_2$ is at least as great as that of $\boldsymbol{z}$.

$\ll$ To this end note that, from Equation 3.19, $\boldsymbol{p} \cdot (P\boldsymbol{z}_1 + P\boldsymbol{z}_2) = \;\boldsymbol{p} \cdot P(\boldsymbol{z}_1 + \boldsymbol{z}_2) = \;(P^T\boldsymbol{p}) \cdot (\boldsymbol{z}_1 + \boldsymbol{z}_2) = (P\boldsymbol{p}) \cdot (\boldsymbol{z}_1 + \boldsymbol{z}_2) = \;\boldsymbol{p} \cdot \boldsymbol{z}$; that is

$$p \cdot (P\boldsymbol{z}_1 + P\boldsymbol{z}_2) = \boldsymbol{p} \cdot \boldsymbol{z} \,. \qquad (3.21)$$

Also note that $p \cdot P\boldsymbol{z}_1 = \boldsymbol{p} \cdot P\big(P(\boldsymbol{z} \vee S\boldsymbol{z}) \vee P(\boldsymbol{z} \wedge S\boldsymbol{z})\big) = (P^Tp) \cdot \big(P(\boldsymbol{z} \vee S\boldsymbol{z}) \vee P(\boldsymbol{z} \wedge S\boldsymbol{z})\big) = (P\boldsymbol{p}) \cdot \big(P(\boldsymbol{z} \vee S\boldsymbol{z}) \vee P(\boldsymbol{z} \wedge S\boldsymbol{z})\big) = \boldsymbol{p} \cdot \big(P(\boldsymbol{z} \vee S\boldsymbol{z}) \vee P(\boldsymbol{z} \wedge S\boldsymbol{z})\big) \geq \boldsymbol{p} \cdot \big(P(\boldsymbol{z} \vee S\boldsymbol{z})\big) = (P^Tp) \cdot (\boldsymbol{z} \vee S\boldsymbol{z}) = (P\boldsymbol{p}) \cdot (\boldsymbol{z} \vee S\boldsymbol{z}) = \boldsymbol{p} \cdot (\boldsymbol{z} \vee S\boldsymbol{z}) \geq \boldsymbol{p} \cdot \boldsymbol{z}$, that is

$$\boldsymbol{p} \cdot (P\boldsymbol{z}_1) \geq \boldsymbol{p} \cdot \boldsymbol{z} \;; \qquad (3.22)$$

From Equations 3.21 and 3.22 it follows that $p \cdot (P\boldsymbol{z}_2) \leq 0$ and that

$$\boldsymbol{p} \cdot (P\boldsymbol{z}_1 - P\boldsymbol{z}_2) \geq \boldsymbol{p} \cdot \boldsymbol{z} \;; \qquad (3.23)$$

In view of Equation 3.23, the p-quantity in $P\boldsymbol{z}_1 - P\boldsymbol{z}_2$ is at least equal to that of $\boldsymbol{z}$. and from Equation 3.20 it follows that the p proportion of $P\boldsymbol{z}_1 - P\boldsymbol{z}_2$ is strictly greater than that of $\boldsymbol{z}$, unless $P\boldsymbol{z}_1 - P\boldsymbol{z}_2 = \boldsymbol{z}_1 - \boldsymbol{z}_2$; but $\boldsymbol{z}_1 - \boldsymbol{z}_2 \geq 0$, so this would mean we had reached a non-negative fixed-point. $\gg$

Thus the recursion

$$\boldsymbol{z}_1 = P1$$

$$\boldsymbol{z}_{n+1} = P\Big(P(\boldsymbol{z}_n \vee S\boldsymbol{z}_n) \;\vee\; P(\boldsymbol{z}_n \wedge S\boldsymbol{z}_n) - P(\boldsymbol{z}_n \vee S\boldsymbol{z}_n) \;\wedge\; P(\boldsymbol{z}_n \wedge S\boldsymbol{z}_n)\Big)$$

converges to a positive fixed-point of P.

Chapter 4

The Original and Invariant Problems

The aim of this chapter is to define the original and invariant problems, and show that they are equivalent.

The invariant problems are introduced then primal and dual function-pairs and invariant linear programming problems, and then we relate the invariant and original problems using these function-pairs.

In the following computations the superscript + denotes the Moore-Penrose pseudo-inverse which was defined in Chapter 2.2.6.

4.1 The Invariant Framework

The major use of the variables detailed here is the construction of the invariant problems in Chapter 4.3

Define

$$\begin{array}{lcll} \mathfrak{A} & = & AA^{+} & \text{(a)} \\ \mathfrak{D} & = & I - AA^{+} & \text{(b)} \\ \mathfrak{b} & = & \mathfrak{D}b & \text{(c)} \\ \mathfrak{c} & = & A^{T+}\boldsymbol{c}. & \text{(d)} \end{array} \qquad (4.1)$$

Note that $\mathfrak{A}$ and $\mathfrak{D}$ are symmetric idempotents in view of the latter two non-constructive conditions for the pseudo-inverse. It is a straightforward matter, using the definitions above and the nonconstructive characterization of the Moore-Penrose pseudo-inverse, to show that

$$\begin{array}{lcll|lcll} \mathfrak{A}^T & = & \mathfrak{A} & \text{(a)} & \mathfrak{A}\mathfrak{b} & = & \mathbf{0} & \text{(f)} \\ \mathfrak{A}^2 & = & \mathfrak{A} & \text{(b)} & \mathfrak{D}\mathfrak{b} & = & \mathfrak{b} & \text{(g)} \\ \mathfrak{D}^T & = & \mathfrak{D} & \text{(c)} & \mathfrak{A}\mathfrak{c} & = & \mathfrak{c} & \text{(h)} \\ \mathfrak{D}^2 & = & \mathfrak{D} & \text{(d)} & \mathfrak{D}\mathfrak{c} & = & \mathbf{0} & \text{(i)} \\ \mathfrak{A}\mathfrak{D} = \mathfrak{D}\mathfrak{A} & = & \mathbf{0} & \text{(e)} & \mathfrak{b}^T\mathfrak{c} & = & \mathbf{0} & \text{(j)} \end{array} \qquad (4.2)$$

4.2 The Original Primal and Dual

An object is a dual of another object if it is the mirror image in some sense of that object. The most important property of this image is that the image of the image of an object be the original object itself. We generally require this of any dual - that the dual of the dual be the identity map. In this chapter we deal with a pair of problems, one of which is an asymmetric "dual" and doesn't satisfy this image property; nevertheless the term is used as closely related problems are dual in nature.

We define the binary operator $\leq$ on two sets $A, B \subseteq \Re$, the real numbers, as follows: $A \leq B$ means that for all $x \in A$ and $y \in B,\ x \leq y$.

An asymmetric dual is given in [14], and Schrijver [33, p. 95] gives the form

$$max\{\boldsymbol{c}^T\boldsymbol{x} : A\boldsymbol{x} \leq \boldsymbol{b}\} = min\{\boldsymbol{b}^T\boldsymbol{y} : A^T\boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \geq \boldsymbol{0}\}, \quad \textit{if either exists.} \tag{4.3}$$

which can be shown to be equivalent to

$$max\{\boldsymbol{c}^T\boldsymbol{x} : A\boldsymbol{x} \geq \boldsymbol{b}\} = min\{\boldsymbol{b}^T\boldsymbol{y} : A^T\boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \leq \boldsymbol{0}\}, \quad \textit{if either exists.} \tag{4.4}$$

whose right hand side has the form of asymmetric dual corresponding to Problem 1.1 - the left hand side.

Referring to Equation 4.4, the maximization problem is called the *original primal* and the minimization problem is called the *original dual*; together they comprise the *original dual pair*. We introduce the concept of quasi-boundedness as it is necessary for the construction of the function pairs which relate the original and invariant LP's:

Definition: 4.2.1 Problem 1.1 is said to be quasi-bounded if $A^+A\boldsymbol{c} = \boldsymbol{c}$.

Now if Problem 1.1 is feasible then there exists a solution, say $\boldsymbol{x}_s$ such that $A\boldsymbol{x}_s \geq b$. Now $A\boldsymbol{x}_s \geq \boldsymbol{b} \Rightarrow A[\boldsymbol{x}_s + \lambda(I - A^+A)\boldsymbol{c}] \geq \boldsymbol{b}$, so $\boldsymbol{x}_s + \lambda(I - A^+A)\boldsymbol{c}$ is also feasible and its objective value is $\boldsymbol{c}^T[\boldsymbol{x}_s + \lambda(I - A^+A)\boldsymbol{c}] = \boldsymbol{c}^T\boldsymbol{x}_s + \boldsymbol{c}^T\lambda(I - A^+A)\boldsymbol{c} = \boldsymbol{c}^T\boldsymbol{x}_s + \lambda\boldsymbol{c}^T(I - A^+A)\boldsymbol{c} = \boldsymbol{c}^T\boldsymbol{x}_s + \lambda\left\|(I - A^+A)\boldsymbol{c}\right\|^2$, which is unbounded unless $\left\|(I - A^+A)\boldsymbol{c}\right\|^2 = 0 \Rightarrow (I - A^+A)\boldsymbol{c} = 0 \Rightarrow A^+A\boldsymbol{c} = \boldsymbol{c}$, so we have

Lemma 4.2.2 Problem 1.1 is feasible bounded $\Rightarrow$ Problem 1.1 is quasi-bounded.

Note that

1. this means quasi-boundedness is a necessary condition for a feasible bounded problem
2. usually A is of full rank and $m \geq n$, so $A^+A = I$, which implies $A^+A\boldsymbol{c} = \boldsymbol{c}$, and so such problem is quasi-bounded.

Lemma 4.2.3 Problem 1.1 is quasi-bounded $\Leftrightarrow A^T\mathfrak{c} = \boldsymbol{c} \Leftrightarrow \exists\boldsymbol{y}$ such that $A^T\boldsymbol{y} = \boldsymbol{c}$.

Proof Problem 1.1 is quasi-bounded $\Rightarrow A^T\mathfrak{c} \overset{4.1\,d}{=} A^T(A^{T+}\boldsymbol{c}) = (A^TA^{T+})\boldsymbol{c} = (A^TA^{+T})\boldsymbol{c} = (A^+A)^T\boldsymbol{c} = A^+A\boldsymbol{c} \overset{D\,4.2.1}{=} \boldsymbol{c} \Rightarrow A^T\mathfrak{c} = \boldsymbol{c} \Rightarrow (A^T\mathfrak{c} = \boldsymbol{c}) \wedge (A^+AA^T\mathfrak{c} = A^+A\boldsymbol{c}) \Rightarrow (A^T\mathfrak{c} = \boldsymbol{c}) \wedge (A^T\mathfrak{c} = A^+A\boldsymbol{c}) \Rightarrow A^+A\boldsymbol{c} = \boldsymbol{c} \overset{D\,4.2.1}{\Rightarrow}$ Problem 1.1 is quasi-bounded. So Problem 1.1 is quasi-bounded $\Leftrightarrow A^T\mathfrak{c} = \boldsymbol{c}$. Further, $A^T\mathfrak{c} = \boldsymbol{c} \Rightarrow \exists \boldsymbol{y}$ such that $A^T\boldsymbol{y} = \boldsymbol{c} \Rightarrow (A^T\boldsymbol{y} = \boldsymbol{c}) \wedge (\mathfrak{A}\boldsymbol{y} = AA^+\boldsymbol{y}) \Rightarrow (A^T\boldsymbol{y} = \boldsymbol{c}) \wedge (\mathfrak{A}\boldsymbol{y} = (AA^+)^T\boldsymbol{y}) \Rightarrow (A^T\boldsymbol{y} = \boldsymbol{c}) \wedge (\mathfrak{A}\boldsymbol{y} = A^{+T}\boldsymbol{c} \overset{4.1\,d}{=} \mathfrak{c}) \Rightarrow (A^T\boldsymbol{y} = \boldsymbol{c}) \wedge (A^T\mathfrak{A}\boldsymbol{y} = A^T\mathfrak{c}) \Rightarrow (A^T\boldsymbol{y} = \boldsymbol{c}) \wedge (A^T\boldsymbol{y} = A^T\mathfrak{c}) \Rightarrow A^T\mathfrak{c} = \boldsymbol{c}$; so $A^T\mathfrak{c} = \boldsymbol{c} \Leftrightarrow \exists\boldsymbol{y}$ such that $A^T\boldsymbol{y} = \boldsymbol{c}$. □

From Lemmas 4.2.2 and 4.2.3 we have

Lemma 4.2.4 Problem 1.1 is feasible bounded $\Rightarrow A^T\mathfrak{c} = \boldsymbol{c}$. □

4.3 The Invariant Problems

4.3.1 Construction of the Central and Peripheral Forms

From the original problems we compute what is called the invariant problems - so-called as this formulation of the LP problem is invariant under one-to-one transformations of the solution space of the original dual pair of problems. Refer to [1] for a more intuitive development.

Substituting $\mathfrak{A}$ for A, $\mathfrak{b}$ for $\boldsymbol{b}$ and $\mathfrak{c}$ for $\boldsymbol{c}$ in Equation 4.3

$$\max\{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} \geq \mathfrak{b}\} = \min\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{A}\mathfrak{y} = \mathfrak{c}, \mathfrak{y} \leq 0\}, \textit{ if either exists.} \tag{4.5}$$

The left hand side of this equation is called the **peripheral invariant primal** while the right hand side is called the **asymmetric invariant dual**, and we see that it is feasible iff the asymmetric dual is feasible since $A^T\boldsymbol{y} = \boldsymbol{c} \quad \Leftrightarrow \quad \mathfrak{A}\boldsymbol{y} = \mathfrak{c}$.

Proof $A^T\boldsymbol{y} = \boldsymbol{c} \Rightarrow (A^{+T}A^T\boldsymbol{y} = A^{+T}\boldsymbol{c}) \wedge (A^T\mathfrak{c} = \boldsymbol{c}) \Rightarrow ((AA^+)^T\boldsymbol{y} = A^{T+}\boldsymbol{c}) \wedge (A^T\mathfrak{c} = \boldsymbol{c}) \Rightarrow (AA^+\boldsymbol{y} = \mathfrak{c}) \wedge (A^T\mathfrak{c} = \boldsymbol{c}) \Rightarrow (\mathfrak{A}\boldsymbol{y} = \mathfrak{c}) \wedge (A^T\mathfrak{c} = \boldsymbol{c}) \Rightarrow (A^T\mathfrak{A}\boldsymbol{y} = A^T\mathfrak{c}) \wedge (A^T\mathfrak{c} = \boldsymbol{c}) \Rightarrow (A^T\boldsymbol{y} = A^T\mathfrak{c}) \wedge (A^T\mathfrak{c} = \boldsymbol{c}) \Rightarrow A^T\boldsymbol{y} = \boldsymbol{c}$.

The asymmetric dual is, however, required in symmetric form so we proceed as follows:

Now $\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{A}\mathfrak{y} = \mathfrak{c}, \mathfrak{y} \leq 0\} = \{\mathfrak{b}^T\mathfrak{y} : (I - \mathfrak{D})\mathfrak{y} = \mathfrak{c}, \mathfrak{y} \leq 0\} = \{\mathfrak{b}^T\mathfrak{y} : \mathfrak{y} - \mathfrak{D}\mathfrak{y} = \mathfrak{c}, \mathfrak{y} \leq 0\} = \{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}(-\mathfrak{y}) - \mathfrak{c} = -\mathfrak{y}, -\mathfrak{y} \geq 0\} = -\{\mathfrak{b}^T(-\mathfrak{y}) : \mathfrak{D}(-\mathfrak{y}) - \mathfrak{c} = -\mathfrak{y}, -\mathfrak{y} \geq 0\} = -\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq 0\}$, that is

$$\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{A}\mathfrak{y} = \mathfrak{c}, \mathfrak{y} \leq 0\} = -\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq 0\} \tag{4.6}$$

So

$$\min\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{A}\mathfrak{y} = \mathfrak{c}, \mathfrak{y} \leq 0\} = \min\big(-\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq 0\}\big)$$

$$= -\max\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq 0\} \quad \textit{if either maximum exists.}$$

and (4.5) can be written

$$\max\{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} \geq \mathfrak{b}\} = -\max\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq 0\} \quad \textit{if either maximum exists.} \tag{4.7}$$

or

$$\max\{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} \geq \mathfrak{b}\} + \max\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq 0\} = 0 \quad \textit{if either maximum exists.} \tag{4.8}$$

The first maximum problem of 4.8 has already been labelled as the peripheral invariant primal; consistent with Definition 2.1.2, Lemma 2.1.5(g) and Equations 4.12 and 4.14, the second is called the **central invariant dual**. We summarize:

The central primal dual forms are, respectively

$$\max\{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} - \mathfrak{b} = \mathfrak{x} \geq \mathbf{0}\} \quad \text{and} \quad \max\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq \mathbf{0}\}$$

while the peripheral primal and dual forms are, respectively

$$\max\{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} \geq \mathfrak{b}\} \quad \text{and} \quad \max\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} \geq \mathfrak{c}\}.$$

Note that with the primal and dual function-pairs specified in Chapter 4.4.1, the use here of the term "central" is consistent with Definition 2.1.2 of Chapter 2.

4.3.2 Equivalence of the Central and Peripheral Forms

The set of feasible objective values each of the central invariant problems is equal to the feasible objective values of the corresponding peripheral invariant problem:

Lemma 4.3.1 (a) $\{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} - \mathfrak{b} = \mathfrak{x} \geq \mathbf{0}\} = \{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} \geq \mathfrak{b}\}$
(b) $\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq \mathbf{0}\} = \{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} \geq \mathfrak{c}\}$

Proof (a): $\{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} - \mathfrak{b} = \mathfrak{x} \geq \mathbf{0}\} \subseteq \{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} \geq \mathfrak{b}\} = \{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} - \mathfrak{b} \geq \mathbf{0}\} = \{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}(\mathfrak{A}\mathfrak{x} - \mathfrak{b}) - \mathfrak{b} = \mathfrak{A}\mathfrak{x} - \mathfrak{b} \geq \mathbf{0}\} = \{\mathfrak{c}^T(\mathfrak{A}\mathfrak{x} - \mathfrak{b}) : \mathfrak{A}(\mathfrak{A}\mathfrak{x} - \mathfrak{b}) - \mathfrak{b} = \mathfrak{A}\mathfrak{x} - \mathfrak{b} \geq \mathbf{0}\} \subseteq \{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} - \mathfrak{b} = \mathfrak{x} \geq \mathbf{0}\}$. (b): proof is analogous to proof of (a). □

This means that we need only consider central points - that is invariant primal points satisfying $\mathfrak{A}\mathfrak{x} - \mathfrak{b} = \mathfrak{x}$, and invariant dual points satisfying $\mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y}$.

From (4.8) and Lemma 4.3.1 (b) we have

Lemma 4.3.2 $\max\{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} \geq \mathfrak{b}\} + \max\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} \geq \mathfrak{c}\} = 0$.
if either maximum exists.

and this lemma is illustrated by Figure 4.1.

From (4.8) and Lemma 4.3.1 (a) we have

Lemma 4.3.3 $\max\{\mathfrak{c}^T\mathfrak{x} : \mathfrak{A}\mathfrak{x} - \mathfrak{b} = \mathfrak{x} \geq \mathbf{0}\} + \max\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq \mathbf{0}\} = 0$,
if either maximum exists.

Remark: 4.3.4 If these programs are feasible bounded then the bound is attained
- a requirement for Theorem 5.2.11 - however the proof of this is a little technical.

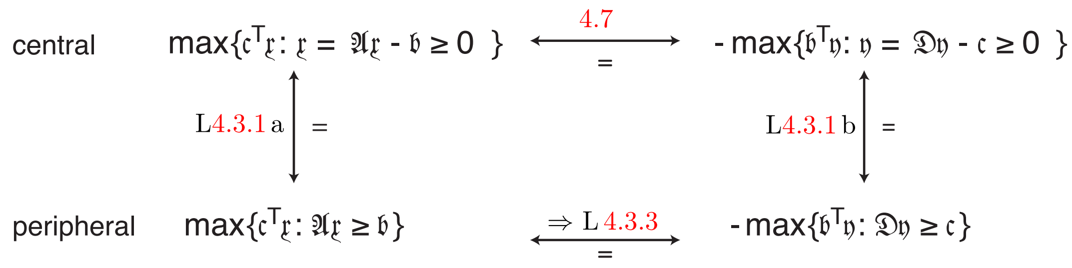


Figure 4.1: **The Invariant Duals**

4.4 Equivalence of the Original and Invariant Problems

We show that there is, effectively, a one-to-one correspondence between solutions to the original and invariant problems. Under this correspondences original residuals are equated with invariant residuals, original feasible solutions are shown to correspond to invariant feasible solutions, original optimal solutions to invariant optimal solutions, and original objective values are shown to differ by a fixed quantity from invariant objective values (with a sign change for the dual problems). Original central points are shown to be in one-to-one correspondence with invariant central points and original central solutions are shown to be in one-to-one correspondence with invariant central solutions We conclude by showing that complementary slackness corresponds to optimality, in the context of the invariant primal. The condition that the original problem be quasi-bounded is necessary in order that the one-to-one relationship exist between the original and invariant problems established in this section. The lower half of Figure 4.3 shows the original dual pair while the upper half shows the invariant dual pair; the functions which define the effectively one-to-one correspondence are indicated.

4.4.1 The Primal and Dual Function-Pairs

The original and invariant problems are related using two function-pairs as follows:

the primal function-pair $< f_p, \mathfrak{f}_p >$ where

$$
\begin{aligned}
& f_p : \Re^n \to \Re^m,\ \boldsymbol{x} \mapsto A\boldsymbol{x} - \boldsymbol{b}, \text{ and} && \text{(a)} \\
& \mathfrak{f}_p : \Re^m \to \Re^n,\ \mathfrak{x} \mapsto A^+(\mathfrak{x} + \boldsymbol{b}) && \text{(b)}
\end{aligned}
\tag{4.9}
$$

and the dual function-pair $< f_d, \mathfrak{f}_d >$ where

$$
\begin{aligned}
& f_d : \Re^m \to \Re^m,\ \boldsymbol{y} \mapsto -\mathfrak{D}\boldsymbol{y} - \mathfrak{c} \text{ and} && \text{(a)} \\
& \mathfrak{f}_d : \Re^m \to \Re^m,\ \mathfrak{y} \mapsto \mathfrak{c} - \mathfrak{D}\mathfrak{y} && \text{(b)}
\end{aligned}
\tag{4.10}
$$

For the primal mappings we compute the compositions $\boldsymbol{x} f_p \mathfrak{f}_p \overset{4.9\,a}{=} (A\boldsymbol{x} - \boldsymbol{b})\mathfrak{f}_p \overset{4.9\,b}{=} A^+((A\boldsymbol{x} - \boldsymbol{b}) + \boldsymbol{b}) = A^+A\boldsymbol{x}$, and $\mathfrak{x}\mathfrak{f}_p f_p \overset{4.9\,b}{=} (A^+(\mathfrak{x} + \boldsymbol{b}))f_p \overset{4.9\,a}{=} A(A^+(\mathfrak{x} + \boldsymbol{b})) - \boldsymbol{b} = AA^+(\mathfrak{x} + \boldsymbol{b}) - \boldsymbol{b} = \mathfrak{A}\mathfrak{x} + \mathfrak{A}\boldsymbol{b} - \boldsymbol{b} = \mathfrak{A}\mathfrak{x} - \mathfrak{b}$, that is

$$\boldsymbol{x} f_p \mathfrak{f}_p = A^+A\boldsymbol{x} \tag{4.11}$$

$$\mathfrak{x}\mathfrak{f}_p f_p = \mathfrak{A}\mathfrak{x} - \mathfrak{b} \tag{4.12}$$

and, for the dual, the compositions $\boldsymbol{y} f_d \mathfrak{f}_d \overset{4.10\,a}{=} (-\mathfrak{D}\boldsymbol{y}-\mathfrak{c})\mathfrak{f}_d \overset{4.10\,b}{=} \mathfrak{c}-\mathfrak{D}(-\mathfrak{D}\boldsymbol{y}-\mathfrak{c}) \overset{4.2}{=} \mathfrak{c}-(-\mathfrak{D}\boldsymbol{y}) = \mathfrak{c}+\mathfrak{D}\boldsymbol{y}$, and $\mathfrak{y}\mathfrak{f}_d f_d \overset{4.10\,b}{=} (\mathfrak{c}-\mathfrak{D}\mathfrak{y}) f_d \overset{4.10\,a}{=} -\mathfrak{D}(\mathfrak{c}-\mathfrak{D}\mathfrak{y}) - \mathfrak{c} \overset{4.2}{=} \mathfrak{D}\mathfrak{y} - \mathfrak{c}$, that is

$$\boldsymbol{y} f_d \mathfrak{f}_d = \mathfrak{D}\boldsymbol{y} + \mathfrak{c} \tag{4.13}$$

$$\mathfrak{y}\mathfrak{f}_d f_d = \mathfrak{D}\mathfrak{y} - \mathfrak{c} \tag{4.14}$$

With respect to Figure 4.3, the original and invariant primal are shown in green on the left, the original and invariant dual are shown in blue on the right, and the peripheral and central feasible sets w.r.t. the function-pairs for the original and invariant problems are indicated respectively by the red and cyan boxes; optimal conditions and sets are indicated by green boxes.

For the primal mappings we compute $\boldsymbol{x} f_p \mathfrak{f}_p f_p \overset{4.11}{=} (A^+A\boldsymbol{x}) f_p \overset{4.9\,a}{=} A(A^+A\boldsymbol{x}) - \boldsymbol{b} = A\boldsymbol{x} - \boldsymbol{b} \overset{4.9\,a}{=} \boldsymbol{x} f_p$, and $\mathfrak{x}\mathfrak{f}_p f_p \mathfrak{f}_p \overset{4.12}{=} (\mathfrak{A}\mathfrak{x} - \mathfrak{b})\mathfrak{f}_p \overset{4.9\,b}{=} A^+((\mathfrak{A}\mathfrak{x}-\mathfrak{b}) + \boldsymbol{b}) = A^+\mathfrak{A}\mathfrak{x} - A^+\mathfrak{b} + A^+\boldsymbol{b} = A^+\mathfrak{x} - A^+\mathfrak{A}\mathfrak{b} + A^+\boldsymbol{b} = A^+(\mathfrak{x}+\boldsymbol{b}) \overset{4.9\,b}{=} \mathfrak{x}\mathfrak{f}_p$, that is

$$f_p \mathfrak{f}_p f_p = f_p \tag{4.15}$$

$$\mathfrak{f}_p f_p \mathfrak{f}_p = \mathfrak{f}_p \tag{4.16}$$

and for the dual mappings $\boldsymbol{y} f_d \mathfrak{f}_d f_d \overset{4.13}{=} (\mathfrak{D}\boldsymbol{y}+\mathfrak{c}) f_d \overset{4.10\,a}{=} -\mathfrak{D}(\mathfrak{D}\boldsymbol{y}+\mathfrak{c}) - \mathfrak{c} = -\mathfrak{D}\boldsymbol{y} - \mathfrak{c} \overset{4.10\,a}{=} \boldsymbol{y} f_d$, and $\mathfrak{y}\mathfrak{f}_d f_d \mathfrak{f}_d \overset{4.14}{=} (\mathfrak{D}\mathfrak{y}-\mathfrak{c})\mathfrak{f}_d \overset{4.10\,b}{=} \mathfrak{c} - \mathfrak{D}(\mathfrak{D}\mathfrak{y}-\mathfrak{c}) = \mathfrak{c} - \mathfrak{D}\mathfrak{y} \overset{4.10\,b}{=} \mathfrak{y}\mathfrak{f}_d$, that is

$$f_d \mathfrak{f}_d f_d = f_d \tag{4.17}$$

$$\mathfrak{f}_d f_d \mathfrak{f}_d = \mathfrak{f}_d \tag{4.18}$$

So the function-pairs $< f_p, \mathfrak{f}_p >$ defined by 4.9 and $< f_d, \mathfrak{f}_d >$ defined by 4.10 are regular so the results of Chapter 2.1 obtain and, in particular Lemma 2.1.5 applies, giving

Theorem: 4.4.1
(a) Central mappings $f_{pc}, \mathfrak{f}_{pc}, f_{dc}$ and $\mathfrak{f}_{dc}$ exist and are bijective.
(b) The primal and dual central function-pairs $< f_{pc}, \mathfrak{f}_{pc} >$ and $< f_{dc}, \mathfrak{f}_{dc} >$ comprise mutually inverse functions.
(c) The conditions $\boldsymbol{x} = A^+A\boldsymbol{x}, \mathfrak{x} = \mathfrak{A}\mathfrak{x} - \mathfrak{b}, \boldsymbol{y} = \mathfrak{D}\boldsymbol{y} + \mathfrak{c}$, and $\mathfrak{y} = \mathfrak{D}\mathfrak{y} - \mathfrak{c}$ define precisely the central points of $X, \mathfrak{X}, Y$, and $\mathfrak{Y}$ respectively.

Proof (a) and (b) follow since the function-pairs $< f_p, \mathfrak{f}_p >$ and $< f_d, \mathfrak{f}_d >$ satisfy the conditions for Lemma 2.1.5; (c) follows from 2.1.5 (f) and (g), or Corollary 2.1.6, together with (4.11), (4.12), (4.13) and (4.14) respectively. ▭

4.4.2 Equivalence of Residuals

For the primal and dual function-pairs the maps leave residuals unchanged (except for a sign change for the dual problems), as can be seen from the following

Lemma 4.4.2
(a) $\mathfrak{A}(\boldsymbol{x} f_p) - \mathfrak{b} = A\boldsymbol{x} - \boldsymbol{b}$
(b) $A(\mathfrak{x}\mathfrak{f}_p) - \boldsymbol{b} = \mathfrak{A}\mathfrak{x} - \mathfrak{b}$
(c) $\mathfrak{D}(\boldsymbol{y} f_d) - \mathfrak{c} = -\boldsymbol{y}$
(d) $\mathfrak{y}\mathfrak{f}_d = -(\mathfrak{D}\mathfrak{y} - \mathfrak{c})$

Proof
(a) $\mathfrak{A}(\boldsymbol{x}f_p) - \mathfrak{b} \overset{4.12}{=} \boldsymbol{x}f_p(\mathfrak{f}_p f_p) \overset{4.15}{=} \boldsymbol{x}f_p \overset{4.9\,a}{=} A\boldsymbol{x} - \boldsymbol{b}$
(b) $A(\mathfrak{x}\mathfrak{f}_p) - \boldsymbol{b} \overset{4.9\,a}{=} (\mathfrak{x}\mathfrak{f}_p)f_p = \mathfrak{x}(\mathfrak{f}_p f_p) \overset{4.12}{=} \mathfrak{A}\mathfrak{x} - \mathfrak{b}$
(c) $\mathfrak{D}(\boldsymbol{y}f_d) - \mathfrak{c} \overset{4.10\,a}{=} \mathfrak{D}(-\mathfrak{D}\boldsymbol{y} - \mathfrak{c}) - \mathfrak{c} = -\mathfrak{D}\boldsymbol{y} - \mathfrak{c}$
$= (AA^+)^T\boldsymbol{y} - \boldsymbol{y} - \mathfrak{c} = A^{T+}A^T\boldsymbol{y} - \boldsymbol{y} - \mathfrak{c} \overset{4.5}{=} A^{T+}\boldsymbol{c} - \boldsymbol{y} - \mathfrak{c} \overset{4.1\,d}{=} -\boldsymbol{y}$
(d) $\mathfrak{y}\mathfrak{f}_d \overset{4.10\,b}{=} \mathfrak{c} - \mathfrak{D}\mathfrak{y} = -(\mathfrak{D}\mathfrak{y} - \mathfrak{c})$. □

The residuals of a peripheral point and its corresponding central point are identical:

Lemma 4.4.3
(a) $\mathfrak{A}\mathfrak{x} - \mathfrak{b} = \mathfrak{A}(\mathfrak{A}\mathfrak{x} - \mathfrak{b}) - \mathfrak{b}$
(b) $\mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{D}(\mathfrak{D}\mathfrak{y} - \mathfrak{c}) - \mathfrak{c}$

Proof
(a) $\mathfrak{A}(\mathfrak{A}\mathfrak{x} - \mathfrak{b}) - \mathfrak{b} \overset{4.12}{=} (\mathfrak{A}\mathfrak{x} - \mathfrak{b})\mathfrak{f}_p f_p \overset{4.12}{=} \mathfrak{x}\mathfrak{f}_p f_p \mathfrak{f}_p f_p \overset{4.15}{=} \mathfrak{x}\mathfrak{f}_p f_p = \mathfrak{A}\mathfrak{x} - \mathfrak{b}$.
(b) $\mathfrak{D}(\mathfrak{D}\mathfrak{y} - \mathfrak{c}) - \mathfrak{c} \overset{4.14}{=} (\mathfrak{D}\mathfrak{y} - \mathfrak{c})\mathfrak{f}_p f_p \overset{4.14}{=} \mathfrak{y}\mathfrak{f}_p f_p \mathfrak{f}_p f_p \overset{4.17}{=} \mathfrak{y}\mathfrak{f}_p f_p = \mathfrak{D}\mathfrak{y} - \mathfrak{c}$. □

4.4.3 Feasible Vectors Correspond

From Lemma 4.4.2 we see that the function-pairs $< f_p, \mathfrak{f}_p >$ and $< f_d, \mathfrak{f}_d >$ induce a correspondence between feasible vectors:

Corollary: 4.4.4
(a) $A\boldsymbol{x} \geq b \Leftrightarrow \mathfrak{A}(\boldsymbol{x}f_p) \geq \mathfrak{b}$
(b) $\mathfrak{A}\mathfrak{x} \geq \mathfrak{b} \Leftrightarrow A(\mathfrak{x}\mathfrak{f}_p) \geq b$
(c) $\mathfrak{D}(\boldsymbol{y}f_d) \geq \mathfrak{c} \Leftrightarrow \boldsymbol{y} \leq \boldsymbol{0}$ if $A^T\boldsymbol{y} = \boldsymbol{c}$
(d) $(A^T(\mathfrak{y}\mathfrak{f}_d) = \boldsymbol{c} \;\wedge\; \mathfrak{y}\mathfrak{f}_d \leq \boldsymbol{0}) \Leftrightarrow \mathfrak{D}\mathfrak{y} \geq \mathfrak{c}$ if Problem 1.1 is quasi-bounded

thus the feasible points of the original and invariant problems correspond, and it follows from Lemma 2.1.7 that feasible central points of the original and invariant problems are in one-to one correspondence.

4.4.4 Equivalence of Objective Values

Next we relate the objective values of points and their images, showing that the map from original primal to invariant primal shifts the objective value by a fixed amount, and the map from original dual to invariant dual involves a sign change as well as a shift of the objective value by a fixed amount:

The objective values of vectors are related by the primal and dual function-pairs as follows:

Lemma 4.4.5
(a) $\mathfrak{c}^T\boldsymbol{x}f_p = \boldsymbol{c}^T\boldsymbol{x} - \boldsymbol{b}^T\mathfrak{c}$
(b) $\boldsymbol{c}^T\mathfrak{x}\mathfrak{f}_p = \mathfrak{c}^T\mathfrak{x} + \boldsymbol{b}^T\mathfrak{c}$
(c) $\mathfrak{b}^T\boldsymbol{y}f_d = \mathfrak{b}^T\mathfrak{c} - \mathfrak{b}^T\boldsymbol{y}$
(d) $\boldsymbol{b}^T\mathfrak{y}\mathfrak{f}_d = \boldsymbol{b}^T\mathfrak{c} - \mathfrak{b}^T\mathfrak{y}$

Proof
(a) $\mathfrak{c}^T\boldsymbol{x}f_p \overset{4.9a}{=} \mathfrak{c}^T(A\boldsymbol{x} - \boldsymbol{b}) = \mathfrak{c}^TA\boldsymbol{x} - \mathfrak{c}^T\boldsymbol{b}$
$= (A^{T+}\boldsymbol{c})^TA\boldsymbol{x} - \mathfrak{c}^T\boldsymbol{b} = \boldsymbol{c}^TA^+A\boldsymbol{x} - \mathfrak{c}^T\boldsymbol{b} \overset{D\,4.2.1}{=} \boldsymbol{c}^T\boldsymbol{x} - \boldsymbol{b}^T\mathfrak{c}$
(b) $\boldsymbol{c}^T(\mathfrak{x}\mathfrak{f}_p) \overset{4.9b}{=} \boldsymbol{c}^T(A^+(\mathfrak{x} + \boldsymbol{b})) = (A^{+T}\boldsymbol{c})^T(\mathfrak{x} + \boldsymbol{b})$
$= (A^{T+}\boldsymbol{c})^T(\mathfrak{x} + \boldsymbol{b}) \overset{4.1\,d}{=} \mathfrak{c}^T(\mathfrak{x} + \boldsymbol{b}) = \mathfrak{c}^T\mathfrak{x} + \boldsymbol{b}^T\mathfrak{c}$
(c) $\mathfrak{b}^T\boldsymbol{y}f_d \overset{4.10\,a}{=} (\mathfrak{D}b)^T\boldsymbol{y}f_d = \boldsymbol{b}^T\mathfrak{D}(\boldsymbol{y}f_d) \overset{L\,4.4.2\,c}{=} \boldsymbol{b}^T(-\boldsymbol{y} + \mathfrak{c}) = \boldsymbol{b}^T\mathfrak{c} - \boldsymbol{b}^T\boldsymbol{y}$
(d) $\boldsymbol{b}^T\mathfrak{y}\mathfrak{f}_d \overset{4.10\,b}{=} \boldsymbol{b}^T(\mathfrak{c} - \mathfrak{D}\mathfrak{y}) \overset{4.1\,c}{=} \boldsymbol{b}^T\mathfrak{c} - \mathfrak{b}^T\mathfrak{y}$. □

So the effect of the functions is to shift objective values by a constant amount, preceded by a sign change in the case of the dual problems - thus the functions map optimal solutions to optimal solutions.

4.4.5 Solutions Correspond

This follows from Corollary 4.4.4 and Lemma 4.4.5. In detail:

For the primal problems, with bx_1 and bx_2 feasible, we have $\boldsymbol{c}^T\boldsymbol{x}_1 \stackrel{4.4.5\,a}{=} \mathfrak{c}^T(\boldsymbol{x}_1 f_p) + \mathfrak{c}^T\boldsymbol{b}$ and $\boldsymbol{c}^T\boldsymbol{x}_2 \stackrel{4.4.5\,a}{=} \mathfrak{c}^T(\boldsymbol{x}_2 f_p) + \mathfrak{c}^T\boldsymbol{b}$, so $\boldsymbol{c}^T\boldsymbol{x}_1 \geq \boldsymbol{c}^T\boldsymbol{x}_2 \Rightarrow \mathfrak{c}^T(\boldsymbol{x}_1 f_p) + \mathfrak{c}^T\boldsymbol{b} \geq \mathfrak{c}^T(\boldsymbol{x}_2 f_p) + \mathfrak{c}^T\boldsymbol{b} \Rightarrow \mathfrak{c}^T(\boldsymbol{x}_1 f_p) \geq \mathfrak{c}^T(\boldsymbol{x}_2 f_p)$ and it follows, since f_p is onto, that if $\boldsymbol{x}_1$ is maximal feasible then $\boldsymbol{x}_1 f_p$ is also, showing that solutions to the original primal are mapped to the full set of central solutions to the invariant primal. Similarly $\mathfrak{c}^T\mathfrak{x}_1 \geq \mathfrak{c}^T\mathfrak{x}_2 \Rightarrow \boldsymbol{c}^T(\mathfrak{x}_1\mathfrak{f}) \geq \boldsymbol{c}^T(\mathfrak{x}_2\mathfrak{f})$ which shows that solutions to the invariant primal are mapped to the full set of central solutions to the original primal. Central solutions to the original and invariant primals, of course, correspond one to one.

For the dual problems we have $\boldsymbol{b}^T\boldsymbol{y}_1 \stackrel{4.4.5\,c}{=} \boldsymbol{b}^T\mathfrak{c} - \mathfrak{b}^T(y_1 f_d)$ and $\boldsymbol{b}^T\boldsymbol{y}_2 \stackrel{4.4.5\,c}{=} \boldsymbol{b}^T\mathfrak{c} - \mathfrak{b}^T(y_2 f_d)$, so $\boldsymbol{b}^T\boldsymbol{y}_1 \leq \boldsymbol{b}^T\boldsymbol{y}_2 \Rightarrow \boldsymbol{b}^T\mathfrak{c} - \mathfrak{b}^T(y_1 f_d) \leq \boldsymbol{b}^T\mathfrak{c} - \mathfrak{b}^T(y_2 f_d) \Rightarrow \mathfrak{b}^T(y_1 f_d) \geq \mathfrak{b}^T(y_2 f_d)$ and it follows, since f_d is onto, that if $\boldsymbol{x}_1$ is minimal feasible then $\boldsymbol{x}_1 f_d$ is maximal feasible, showing that solutions to the original primal are mapped to central solutions to the invariant primal. Similarly $\mathfrak{c}^T\mathfrak{x}_1 \geq \mathfrak{c}^T\mathfrak{x}_2 \Rightarrow \boldsymbol{c}^T(\mathfrak{x}_1\mathfrak{f}) \geq \boldsymbol{c}^T(\mathfrak{x}_2\mathfrak{f})$ which shows that solutions to the invariant primal are mapped to central solutions to the original primal. Central solutions to the original and invariant duals, of course, correspond one to one.

4.5 Summary

We have shown that the solution of the original linear program 1.1 and its dual can, provided the original primal is quasi-bounded, be transformed to finding the solutions to the invariant primal and dual, and from the invariant primal and dual to the central invariant primal and dual - that is:

Table 4.1: Primal-Dual Systems

Original		Invariant		Central Invariant
$A\boldsymbol{x} \geq \boldsymbol{b}$ $A^T\boldsymbol{y} = \boldsymbol{c}\boldsymbol{y} \leq \boldsymbol{0}$ $\boldsymbol{c}^T\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{y} = 0$ $A^+A\boldsymbol{c} = \boldsymbol{c}$	$\leftrightarrow$	$\mathfrak{A}\mathfrak{x} \geq \mathfrak{b}$ $\mathfrak{D}\mathfrak{y} \geq \mathfrak{c}$ $\mathfrak{c}^T\mathfrak{x} + \mathfrak{b}^T\mathfrak{y} = 0$ $A^+A\boldsymbol{c} = \boldsymbol{c}$	$\leftrightarrow$	$\mathfrak{A}\mathfrak{x} - \mathfrak{b} = \mathfrak{x} \geq \boldsymbol{0}$ $\mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq \boldsymbol{0}$ $\mathfrak{c}^T\mathfrak{x} + \mathfrak{b}^T\mathfrak{y} = 0$ $A^+A\boldsymbol{c} = \boldsymbol{c}$

The results of this chapter are summarized by Figures 4.2 and 4.3, establishing

1. a many to one relationship between

   (a) all original primal solutions on the one hand and all central original primal solutions on the other.

   (b) all original (asymmetric) dual solutions on the one hand and all central original (asymmetric) dual solutions on the other.

   (c) all invariant primal solutions on the one hand and all central invariant primal solutions on the other.

   (d) all invariant dual solutions on the one hand and all central invariant dual solutions on the other.

2. a one to one relationship between

   (a) central solutions to the original primal on the one hand and central solutions to the invariant primal on the other

   (b) central solutions to the original dual on the one hand and central solutions to the invariant dual on the other

In the next chapter we derive the LP fixed-point problem from the invariant primal and dual.

	primal		*dual*
invariant	$\mathfrak{X} = \{\mathfrak{x} : \mathfrak{A}\mathfrak{x} \geq \mathfrak{b}\}$	$\xrightarrow{\text{inv. dual}}$ / $\xleftarrow{\text{inv. dual}}$	$\mathfrak{Y} = \{\mathfrak{y} : \mathfrak{D}\mathfrak{y} \geq \mathfrak{c}\}$
	$\downarrow \mathfrak{f}_p f_p$		$\downarrow \mathfrak{f}_d f_d$
invariant central	$\mathfrak{X}_c = \{\mathfrak{x} : \mathfrak{A}\mathfrak{x} - \mathfrak{b} = \mathfrak{x} \geq \mathbf{0}\}$	$\xrightarrow{\text{inv. dual}}$ / $\xleftarrow{\text{inv. dual}}$	$\mathfrak{Y}_c = \{\mathfrak{y} : \mathfrak{D}\mathfrak{y} - \mathfrak{c} = \mathfrak{y} \geq \mathbf{0}\}$
	$f_p \uparrow \downarrow \mathfrak{f}_p$		$f_d \uparrow \downarrow \mathfrak{f}_d$
original central	$X_c = \{\boldsymbol{x} : A\boldsymbol{x} \geq \boldsymbol{b}, \boldsymbol{x} = A^{+}A\boldsymbol{x}\}$	$\xrightarrow{asym.dual}$ / $\xleftarrow{asym.dual}$	$Y_c = \{y : A^T\boldsymbol{y} = \boldsymbol{c}, \mathfrak{D}\boldsymbol{y} + \mathfrak{c} = \boldsymbol{y} \leq \mathbf{0}\}$
	$f_p \mathfrak{f}_p \uparrow$		$f_d \mathfrak{f}_d \uparrow$
original	$X = \{\boldsymbol{x} : A\boldsymbol{x} \geq \boldsymbol{b}\}$	$\xrightarrow{asym.dual}$ / $\xleftarrow{asym.dual}$	$Y = \{y : A^T\boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \leq \mathbf{0}\}$

Figure 4.2: Mappings in Detail

However, given the clarity brought to the analysis by the notion of centrality, Figure 4.3 may be simplified to Figure 4.4.

4.6 Examples

4.6.1 A Two Variable Problem

Consider the problem of maximizing $3x_1 + 4x_2$ subject to $x_1 + x_2 \leq 3$, $x_1 - x_2 \leq 1.5$, and $2x_1 + x_2 \geq 4$. The problem can be seen to be feasible and bounded by referring to the following diagram in which the set of feasible solutions is represented by the green triangular area, the constraints are represented by the lines forming the boundary of this area, and the objective function is represented by the indicated line which should be moved until it osculates the hatched area.

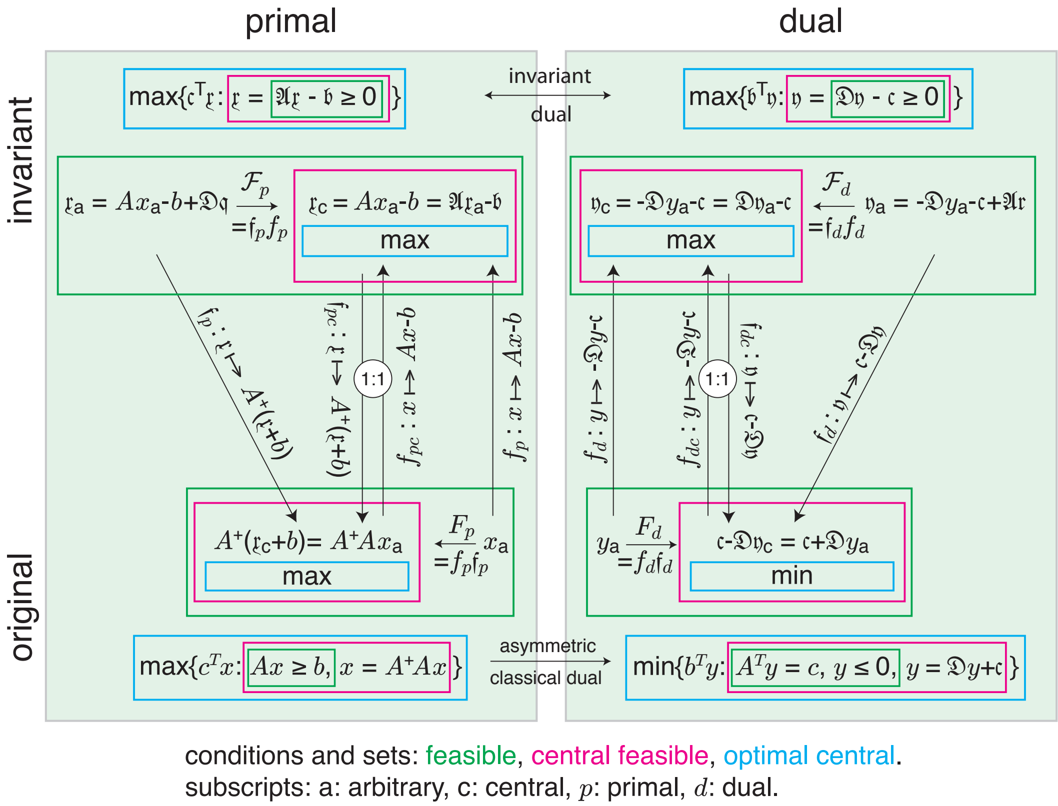


Figure 4.3: **Equivalence of Original and Invariant Problems**

The problem can be written in matrix form as

$$\textit{maximize} \begin{bmatrix} 3 \\ 4 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \textit{ subject to} \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \geq \begin{bmatrix} -3 \\ -1.5 \\ 4 \end{bmatrix}.$$

So $A = \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 2 & 1 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} -3 \\ -1.5 \\ 4 \end{bmatrix}$, and $\boldsymbol{c} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

The Moore-Penrose pseudo-inverse can be computed as follows:

$$A^+ = \left\{ \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 2 & 1 \end{bmatrix}^T \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 2 & 1 \end{bmatrix} \right\}^+ \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 2 & 1 \end{bmatrix}^T$$

$$= \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}^{-1} \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 2 & 1 \end{bmatrix}^T = \begin{bmatrix} 3 & -2 \\ -2 & 6 \end{bmatrix} \begin{bmatrix} -1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix} /14$$

$$= \begin{bmatrix} -1 & -5 & 4 \\ -4 & 8 & 2 \end{bmatrix} /14,$$

and

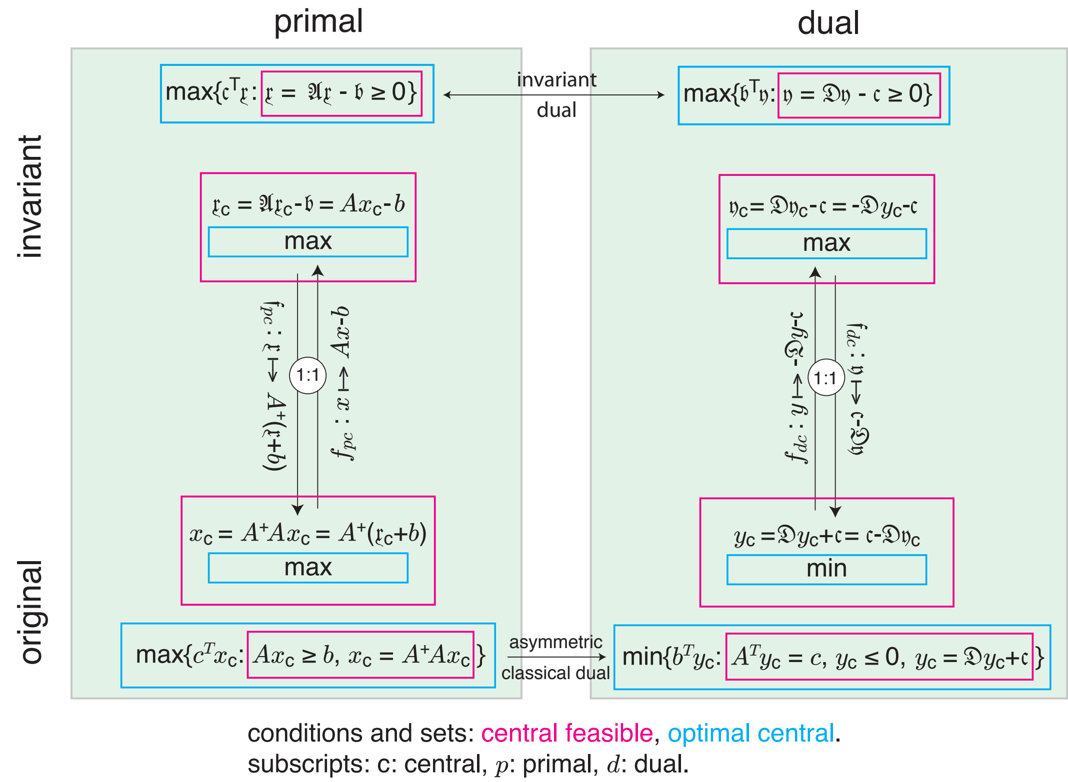


Figure 4.4: **Equivalence of Original and Invariant Central Problems**

$$AA^+ = \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & -5 & 4 \\ -4 & 8 & 2 \end{bmatrix} /14 = \begin{bmatrix} 5 & -3 & -6 \\ -3 & 13 & -2 \\ -6 & -2 & 10 \end{bmatrix} /14.$$

$$\mathfrak{b} = (I - AA^+)b = \begin{bmatrix} 9 & 3 & 6 \\ 3 & 1 & 2 \\ 6 & 2 & 4 \end{bmatrix} \begin{bmatrix} -3 \\ -1.5 \\ 4 \end{bmatrix} /14 = \begin{bmatrix} -7.5 \\ -2.5 \\ -5.0 \end{bmatrix} /14$$

$$\mathfrak{c} = A^{T+}\boldsymbol{c} = \begin{bmatrix} -1 & -4 \\ -5 & 8 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} /14 = \begin{bmatrix} -19 \\ 17 \\ 20 \end{bmatrix} /14.$$

Thus the invariant primal is

$$\textit{maximize} \quad \begin{bmatrix} -19 \\ 17 \\ 20 \end{bmatrix}^T \begin{bmatrix} \mathfrak{x}_1 \\ \mathfrak{x}_2 \\ \mathfrak{x}_1 \end{bmatrix} /14 \quad s.t. \quad \begin{bmatrix} 5 & -3 & -6 \\ -3 & 13 & -2 \\ -6 & -2 & 10 \end{bmatrix} \begin{bmatrix} \mathfrak{x}_1 \\ \mathfrak{x}_2 \\ \mathfrak{x}_1 \end{bmatrix} /14 \geq \begin{bmatrix} -7.5 \\ -2.5 \\ -5.0 \end{bmatrix} /14.$$

These computations can be effected using the function **invprob**, provided in the module `solve-lp.rkt`. Note that `linalg1.rkt` provides a number of linear algebra functions which are required by `solve-lp.rkt` which in turn provides the function **invprob** which computes the invariants $\mathfrak{A}$, $\mathfrak{b}$, $\mathfrak{c}$.

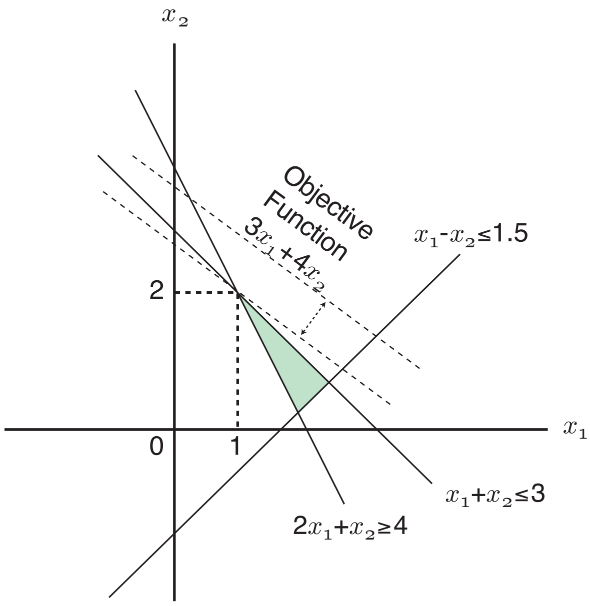


Figure 4.5: **Feasible Set and Objective Function**

4.7 Exercises

1. Prove that if the problem *maximize* $\mathbf{c}^T\boldsymbol{x}$ *subject to* $A\boldsymbol{x} \leq b$ is feasible bounded then

$$max\{\mathbf{c}^T\boldsymbol{x} : A\boldsymbol{x} \leq b\} = \boldsymbol{b}^T\mathfrak{c} - max\{\mathfrak{b}^T\mathfrak{y} : \mathfrak{D}\mathfrak{y} \leq \mathfrak{c}\} \ .$$

   Hint: look at Lemma 4.4.5.

2. What is the rank of the matrix AA^+ in Section 4.6.1?

   What is its trace?

   Why are they equal?

3. Prove: $\left(I_{m,n} \otimes \mathbf{1}_n^{T+}\right)\boldsymbol{s} = (I_{m,n}\boldsymbol{s}) \otimes \mathbf{1}_n^{T+}$.

4. Prove: $\left(\mathbf{1}_m^{T+} \otimes I_{n,m}\right)\boldsymbol{d} = \mathbf{1}_m^{T+} \otimes (I_{n,m}\boldsymbol{d})$.

Chapter 5

The LP Fixed-Point Problem

In this chapter the invariant primal and dual LP problems are represented as a fixed-point problem. Such representation and the subsequent gradient projection method of solution is the work of Rosen, Pyle, Cline, Bruni *et al* [32, 27, 28, 6]; here the representation is developed with precision.

We represent the invariant problems as the problem of finding the non-negative fixed-points of an Hermitian matrix $\mathfrak{P}$ which satisfies the properties of the general P of Chapter 3. Thus we make use of the results of that chapter to solve the fixed-point problem.

In this chapter we

1. first specify the fixed-point problem,
2. carry out some taxonomic analysis of points related to the fixed-point problem in order to establish a precise correspondence between solutions to the invariant problem and the fixed-point problem.
3. consider the degenerate case - that is the case of a zero central fixed-point.
4. consider the possibility that a different problem has been solved.

5.1 Specification

From Lemma 4.3.2 the conditions for the solution of the invariant primal-dual pair can be written as

$$\begin{array}{ll} \mathfrak{A}\mathfrak{x} - \mathfrak{b} & \geq \mathbf{0} \\ \mathfrak{D}\mathfrak{y} - \mathfrak{c} & \geq \mathbf{0} \\ \mathfrak{c}^T\mathfrak{x} + \mathfrak{b}^T\mathfrak{y} & = 0 \end{array} \tag{5.1}$$

and from Lemma 4.3.3 the conditions for the solution of the central invariant primal-dual pair can be written as

$$\begin{array}{lll} \mathfrak{A}\mathfrak{x} - \mathfrak{b} & = \mathfrak{x} & \geq \mathbf{0} \\ \mathfrak{D}\mathfrak{y} - \mathfrak{c} & = \mathfrak{y} & \geq \mathbf{0} \\ \mathfrak{c}^T\mathfrak{x} + \mathfrak{b}^T\mathfrak{y} & = & 0 \end{array} \tag{5.2}$$

We now define the ordinary and central versions of the fixed point problem and show their respective equivalence to the above conditions.

Define the following short forms:

$$
\begin{aligned}
\Pi_{\mathfrak{A}} &= \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} && \text{(a)} \\
\mathfrak{j} &= \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} && \text{(b)} \\
\mathfrak{k} &= \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} && \text{(c)}
\end{aligned}
\qquad (5.3)
$$

and the *fixed-point matrix*

$$
\begin{aligned}
\mathfrak{P} &= \Pi_{\mathfrak{A}} + \mathfrak{j}\mathfrak{j}^+ - \mathfrak{k}\mathfrak{k}^+ && \text{(a)} \\
& \quad \text{or equivalently} \\
\mathfrak{P} &= \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} + \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ && \text{(b)}
\end{aligned}
\qquad (5.4)
$$

Symmetry of $\mathfrak{P}$ follows from the symmetry of its components $\Pi_{\mathfrak{A}}, \mathfrak{j}\mathfrak{j}^+$ and $\mathfrak{k}\mathfrak{k}^+$

while $\mathfrak{P}^2 \overset{5.4a}{=} \left(\Pi_{\mathfrak{A}} + \mathfrak{j}\mathfrak{j}^+ - \mathfrak{k}\mathfrak{k}^+\right)^2$

$$= \Pi_{\mathfrak{A}}^2 + \Pi_{\mathfrak{A}}\mathfrak{j}\mathfrak{j}^+ - \Pi_{\mathfrak{A}}\mathfrak{k}\mathfrak{k}^+ + \mathfrak{j}\mathfrak{j}^+\Pi_{\mathfrak{A}} + \mathfrak{j}\mathfrak{j}^+\mathfrak{j}\mathfrak{j}^+ - \mathfrak{j}\mathfrak{j}^+\mathfrak{k}\mathfrak{k}^+ - \mathfrak{k}\mathfrak{k}^+\Pi_{\mathfrak{A}} - \mathfrak{k}\mathfrak{k}^+\mathfrak{j}\mathfrak{j}^+ + \mathfrak{k}\mathfrak{k}^+\mathfrak{k}\mathfrak{k}^+$$

$$= \Pi_{\mathfrak{A}} + 0 - \mathfrak{k}\mathfrak{k}^+ + 0 + \mathfrak{j}\mathfrak{j}^+ - 0 - \mathfrak{k}\mathfrak{k}^+ - 0 + \mathfrak{k}\mathfrak{k}^+ = \Pi_{\mathfrak{A}} + \mathfrak{j}\mathfrak{j}^+ - \mathfrak{k}\mathfrak{k}^+ = \mathfrak{P} \ ,$$

in other words $\mathfrak{P}$ is an *orthogonal projection.*

Further, $\mathfrak{P}\mathfrak{j} = \left(\Pi_{\mathfrak{A}} + \mathfrak{j}\mathfrak{j}^+ - \mathfrak{k}\mathfrak{k}^+\right)\mathfrak{j} = \Pi_{\mathfrak{A}}\mathfrak{j} + \mathfrak{j}\mathfrak{j}^+\mathfrak{j} - \mathfrak{k}\mathfrak{k}^+\mathfrak{j} = \mathbf{0} + \mathfrak{j} + \mathbf{0} = \mathfrak{j}$
while $\mathfrak{P}\mathfrak{k} = \left(\Pi_{\mathfrak{A}} + \mathfrak{j}\mathfrak{j}^+ - \mathfrak{k}\mathfrak{k}^+\right)\mathfrak{k} = \Pi_{\mathfrak{A}}\mathfrak{k} + \mathfrak{j}\mathfrak{j}^+\mathfrak{k} - \mathfrak{k}\mathfrak{k}^+\mathfrak{k} = \mathfrak{k} + \mathbf{0} - \mathfrak{j} = \mathbf{0}$, that is,

$$
\begin{aligned}
\mathfrak{P}\mathfrak{j} &= \mathfrak{j} \ , && \text{(a)} \\
\mathfrak{P}\mathfrak{k} &= 0. && \text{(b)}
\end{aligned}
\qquad (5.5)
$$

A point $\mathfrak{z}$ is said to be a **non-negative fixed-point** (NNFP) if $\mathfrak{P}\mathfrak{z} = \mathfrak{z} \geq \mathbf{0}$.

Now, with $\mathfrak{z} = \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix}, \mathfrak{P}\mathfrak{z} = \mathfrak{z} \geq \mathbf{0}$

$$\Leftrightarrow \left(\begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} + \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \right) \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} = \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \geq \mathbf{0}$$

$$\Leftrightarrow \begin{array}{l} \left(\begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} + \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \right) \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} = \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \geq \mathbf{0} \quad \text{and} \\ \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^T \left(\begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} + \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \right) \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} = \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^T \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \end{array}$$

$$\Leftrightarrow \left(\begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} + \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \right) \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} = \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \geq \mathbf{0} \quad \wedge \quad \mathfrak{c}^T\mathfrak{x} + \mathfrak{b}^T\mathfrak{y} = 0$$

$$\Leftrightarrow \begin{bmatrix} \mathfrak{A}\mathfrak{x} \\ \mathfrak{D}\mathfrak{y} \end{bmatrix} + \left(\begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \right) \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} = \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \geq \mathbf{0} \quad \wedge \quad \mathfrak{c}^T\mathfrak{x} + \mathfrak{b}^T\mathfrak{y} = 0$$

and with $\lambda = -\begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix}$ and $\begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} = \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} / \lambda$, if $\lambda > 0$

$$\begin{aligned} & \mathfrak{A}\boldsymbol{\xi} - \mathfrak{b} = \boldsymbol{\xi} \geq \mathbf{0} \\ \Leftrightarrow\ & \mathfrak{D}\boldsymbol{\zeta} - \mathfrak{c} = \boldsymbol{\zeta} \geq \mathbf{0} \\ & \mathfrak{c}^T\boldsymbol{\xi} + \mathfrak{b}^T\boldsymbol{\zeta} = 0 \ , \end{aligned}$$

which is System 5.2.

5.2 Correspondence

By considering the taxonomy of central points, quasi-optimal points and fixed-points (both defined in this section) we establish a correspondence between solutions to the fixed-point problem and solutions to the invariant problem, including the central form thereof.

5.2.1 The Primal-Dual Function Pair

We define the primal-dual function-pair $< F_p, \mathfrak{F}_p >$ by

$$\begin{aligned} F &: \Re^{2m} \to \Re^{2m}, \quad \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \mapsto \begin{bmatrix} \boldsymbol{x} f_p \\ \boldsymbol{y} f_d \end{bmatrix}, \text{ and} && \text{(a)} \\ \mathfrak{F} &: \Re^{2m} \to \Re^{2m}, \quad \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} \mapsto \begin{bmatrix} \boldsymbol{\xi}\mathfrak{f}_p \\ \boldsymbol{\zeta}\mathfrak{f}_d \end{bmatrix} && \text{(b)} \end{aligned} \tag{5.6}$$

where f_p and $\mathfrak{f}_p$ are given by Equations 4.9, and f_d and $\mathfrak{f}_d$ by 4.10.

In view of Equations 4.15, 4.16, 4.17 and 4.18 the function pair $< F_p, \mathfrak{F}_p >$ is regular:

$$\boldsymbol{z}F\mathfrak{F}F = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} F\mathfrak{F}F = \begin{bmatrix} \boldsymbol{x}f_p \\ \boldsymbol{y}f_d \end{bmatrix} \mathfrak{F}F = \begin{bmatrix} \boldsymbol{x}f_p\mathfrak{f}_p \\ \boldsymbol{y}f_d\mathfrak{f}_d \end{bmatrix} F = \begin{bmatrix} \boldsymbol{x}f_p\mathfrak{f}_p f_p \\ \boldsymbol{y}f_d\mathfrak{f}_d f_d \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}f_p \\ \boldsymbol{y}f_d \end{bmatrix} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} F = \boldsymbol{z}F$$

and

$$\boldsymbol{\omega}\mathfrak{F}F\mathfrak{F} = \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} \mathfrak{F}F\mathfrak{F} = \begin{bmatrix} \boldsymbol{\xi}\mathfrak{f}_p \\ \boldsymbol{\zeta}\mathfrak{f}_d \end{bmatrix} F\mathfrak{F} = \begin{bmatrix} \boldsymbol{\xi}\mathfrak{f}_p f_p \\ \boldsymbol{\zeta}\mathfrak{f}_d f_d \end{bmatrix} \mathfrak{F} = \begin{bmatrix} \boldsymbol{\xi}\mathfrak{f}_p f_p\mathfrak{f}_p \\ \boldsymbol{\zeta}\mathfrak{f}_d f_d\mathfrak{f}_d \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi}\mathfrak{f}_p \\ \boldsymbol{\zeta}\mathfrak{f}_d \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} \mathfrak{F} = \boldsymbol{\omega}\mathfrak{F} \ .$$

Centrality for original $\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}$ and invariant $\begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix}$ solutions is defined in the context of the invariant primal-dual function pair, so

Definition: 5.2.1 $\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}$ is central if $\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} F\mathfrak{F} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}$

$\boldsymbol{\omega}$ is central if $\boldsymbol{\omega}\mathfrak{F}F = \boldsymbol{\omega}$.

and original and invariant central vectors are precisely of the form $\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} F\mathfrak{F}$ and $\begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} \mathfrak{F}F$, that is

$$\boldsymbol{w}F\mathfrak{F} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} F\mathfrak{F} = \begin{bmatrix} A\boldsymbol{x} - \boldsymbol{b} \\ -\mathfrak{D}\boldsymbol{y} - \mathfrak{c} \end{bmatrix}$$

and

$$\boldsymbol{\omega}\mathfrak{F}F = \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} \mathfrak{F}F = \begin{bmatrix} \xi\mathfrak{f}_p f_p \\ \zeta\mathfrak{f}_d f_d \end{bmatrix} = \begin{bmatrix} (A^+(\xi + b))f_p \\ (\mathfrak{c} - \mathfrak{D}\zeta)f_d \end{bmatrix} = \begin{bmatrix} AA^+(\xi + b) - b \\ -\mathfrak{D}(\mathfrak{c} - \mathfrak{D}\zeta) - \mathfrak{c} \end{bmatrix}$$

$$= \begin{bmatrix} \mathfrak{A}\xi - \mathfrak{D}b \\ \mathfrak{D}\zeta - \mathfrak{c} \end{bmatrix} = \begin{bmatrix} \mathfrak{A}\xi - \mathfrak{b} \\ \mathfrak{D}\zeta - \mathfrak{c} \end{bmatrix} = \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} = \Pi_{\mathfrak{A}}\boldsymbol{\omega} - \mathfrak{j}$$

leading to the operational definition of "central" for an invariant vector pair:

Definition: 5.2.2 $\boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix}$ is said to be central if

$\boldsymbol{\omega} = \Pi_{\mathfrak{A}}\boldsymbol{\omega} - \mathfrak{j}$ or, equivalently,

$\begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} = \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}$. or, equivalently,

$\mathfrak{A}\xi - \mathfrak{b} = \boldsymbol{\xi}$ and $\mathfrak{D}\boldsymbol{\zeta} - \mathfrak{c} = \boldsymbol{\zeta}$.

Note: 5.2.3 The set of central vectors is convex, moreover not just every point in the line segment joining two central points is central, so too is every point in the line joining the two.

For partitioned points of the form $\boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix}$, where $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are m dimensional vectors, we say that $\boldsymbol{\omega}$ is a **feasible-point** if $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are feasible for the invariant primal and dual respectively, and that $\boldsymbol{\omega}$ is a **solution-point** if $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are, respectively, solutions to the invariant primal and dual.. Note that, consistent with Theorem 4.4.1 and Definition 5.2.1, Definition 5.2.2 says that $\boldsymbol{\omega}$ is a *central point* if $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are central - that is if $\mathfrak{A}\xi - \mathfrak{b} = \boldsymbol{\xi}$ and $\mathfrak{D}\boldsymbol{\zeta} - \mathfrak{c} = \boldsymbol{\zeta}$.

Definition: 5.2.4 We say that $\boldsymbol{\omega}$ is a **quasi-optimal** point if $\mathfrak{c}^T\boldsymbol{\xi} + \mathfrak{b}^T\boldsymbol{\zeta} = 0$,

Note: 5.2.5 The set of quasi-optimal vectors is convex, moreover not just every point in the line segment joining two quasi-optimal points is central, so too is every point in the line joining the two.

5.2.2 Fixed-Points

Note that $\begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}$ is a fixed point that is

$$\mathfrak{P}\begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} = \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} . \tag{5.7}$$

Fixed-points are quasi-optimal:

Lemma 5.2.6 $\boldsymbol{\omega}$ is a fixed point $\Rightarrow \boldsymbol{\omega}$ is quasi-optimal.

Proof $\boldsymbol{\omega} = \mathfrak{P}\boldsymbol{\omega} \Rightarrow \mathfrak{k}^T\boldsymbol{\omega} = \mathfrak{k}^T\mathfrak{P}\boldsymbol{\omega}$ $\overset{5.5\,b}{\Rightarrow} \mathfrak{k}^T\boldsymbol{\omega} = 0^T\boldsymbol{\omega} \Rightarrow \mathfrak{k}^T\boldsymbol{\omega} = 0$.

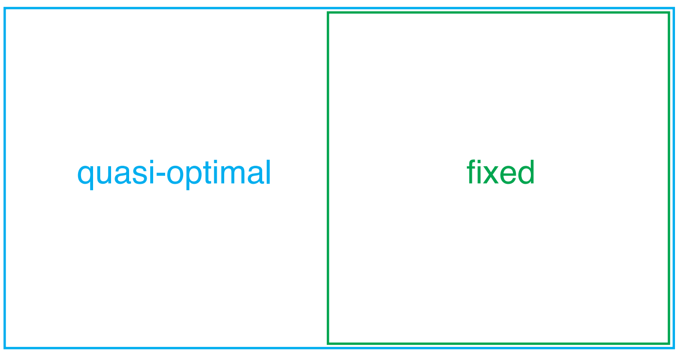


Figure 5.1: **Fixed-Points are Quasi-Optimal**

5.2.3 Central Points

We consider central points (refer to Theorem 4.4.1 c and Definition 5.2.2), in the context of quasi-optimal points (Chapter 5.1), and fixed-points (Chapter 5.1).

Note that $\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}$ is central $\Leftrightarrow \begin{bmatrix}\mathfrak{A} & 0\\ 0 & \mathfrak{D}\end{bmatrix}\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix} - \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix} = \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix} \Leftrightarrow -\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix} = \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix} \Leftrightarrow \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix} = \mathbf{0}$, that is:

$$\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix} \text{ is central } \Leftrightarrow \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix} = \mathbf{0} \tag{5.8}$$

For central point $\boldsymbol{\omega} = \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} = \begin{bmatrix}\mathfrak{A}\boldsymbol{\xi} - \mathfrak{b}\\ \mathfrak{D}\boldsymbol{\zeta} - \mathfrak{c}\end{bmatrix} \Rightarrow \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}^+ \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} = \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}^+ \begin{bmatrix}\mathfrak{A}\boldsymbol{\xi} - \mathfrak{b}\\ \mathfrak{D}\boldsymbol{\zeta} - \mathfrak{c}\end{bmatrix} = -\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}^+ \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}$,
that is

$$\text{for central } \boldsymbol{\omega} = \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix}: \quad \begin{aligned} \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}^+ \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} &= -\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}^+ \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix} \quad \text{or. equivalently,} && \text{(a)}\\ \mathbf{j}^+\boldsymbol{\omega} &= -\mathbf{j}^+\mathbf{j} && \text{(b)} \end{aligned} \tag{5.9}$$

This leads to the definition of the **discriminant**:

$$d(\omega) = -\mathbf{j}^+\omega, \tag{5.10}$$

It follows from 5.9 that for central ω, $d(\omega) \in \{0, 1\}$, and $d(\omega) = 0 \Leftrightarrow \mathfrak{b} = \mathfrak{c} = 0$. This is a very important result as it implies that all central solutions to the LP fixed-point problem are non-zero unless both $\mathfrak{b}$ and $\mathfrak{c}$ are zero.

Lemma 5.2.7 $\exists \boldsymbol{\omega} = \mathbf{0}, \boldsymbol{\omega}$ central $\Leftrightarrow \mathfrak{b} = \mathfrak{c} = \mathbf{0}$.

Proof: $\boldsymbol{\omega} = \mathbf{0} \overset{5.10}{\Rightarrow} d(\boldsymbol{\omega}) = \mathbf{0} \Leftrightarrow \mathfrak{b} = \mathfrak{c} = \mathbf{0}$; conversely, if $\mathfrak{b} = \mathfrak{c} = \mathbf{0}$ then $\boldsymbol{\omega} = \mathbf{0}$ is central.

Definition: 5.2.8 A fixed-point problem is said to be degenerate if $d(\omega) = 0$

Lemma 5.2.9 $\boldsymbol{\omega}$ is quasi-optimal and central $\Rightarrow \boldsymbol{\omega}$ is fixed.

Proof $\boldsymbol{\omega}$ quasi-optimal and central $\Rightarrow \boldsymbol{\omega} = \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} \overset{D\,5.2.2}{=} \begin{bmatrix}\mathfrak{A}\boldsymbol{\xi} - \mathfrak{b}\\ \mathfrak{D}\boldsymbol{\zeta} - \mathfrak{c}\end{bmatrix}$, and $\begin{bmatrix}\mathfrak{c}\\ \mathfrak{b}\end{bmatrix}^T \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} \overset{D\,5.2.1}{=} 0$

$$\Rightarrow \mathfrak{P}\boldsymbol{\omega} = \mathfrak{P}\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} = \left(\begin{bmatrix}\mathfrak{A} & 0\\ 0 & \mathfrak{D}\end{bmatrix} + \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}^+ - \begin{bmatrix}\mathfrak{c}\\ \mathfrak{b}\end{bmatrix}\begin{bmatrix}\mathfrak{c}\\ \mathfrak{b}\end{bmatrix}^+\right)\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}\mathfrak{c}\\ \mathfrak{b}\end{bmatrix}^T \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} = 0$$

$$\Rightarrow \mathfrak{P}\boldsymbol{\omega} = \begin{bmatrix}\mathfrak{A} & 0\\ 0 & \mathfrak{D}\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} + \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}^+\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} - \begin{bmatrix}\mathfrak{c}\\ \mathfrak{b}\end{bmatrix}\begin{bmatrix}\mathfrak{c}\\ \mathfrak{b}\end{bmatrix}^+\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}\mathfrak{c}\\ \mathfrak{b}\end{bmatrix}^T \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} = 0$$

$$\Rightarrow \mathfrak{P}\boldsymbol{\omega} = \begin{bmatrix}\mathfrak{A} & 0\\ 0 & \mathfrak{D}\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} + \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}^+\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} - \begin{bmatrix}\mathfrak{c}\\ \mathfrak{b}\end{bmatrix}^{+T}\begin{bmatrix}\mathfrak{c}\\ \mathfrak{b}\end{bmatrix}^T\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}\mathfrak{c}\\ \mathfrak{b}\end{bmatrix}^T \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} = 0$$

$$\Rightarrow \mathfrak{P}\boldsymbol{\omega} = \begin{bmatrix}\mathfrak{A} & 0\\ 0 & \mathfrak{D}\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix} + \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}^+\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\zeta}\end{bmatrix}$$

$$\Rightarrow \mathfrak{P}\boldsymbol{\omega} \overset{5.9}{=} \begin{bmatrix}\mathfrak{A}\boldsymbol{\xi}\\ \mathfrak{D}\boldsymbol{\zeta}\end{bmatrix} + \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}\left(-\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}^+\begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix}\right) = \begin{bmatrix}\mathfrak{A}\boldsymbol{\xi}\\ \mathfrak{D}\boldsymbol{\zeta}\end{bmatrix} - \begin{bmatrix}\mathfrak{b}\\ \mathfrak{c}\end{bmatrix} = \begin{bmatrix}\mathfrak{A}\boldsymbol{\xi} - \mathfrak{b}\\ \mathfrak{D}\boldsymbol{\zeta} - \mathfrak{c}\end{bmatrix} = \boldsymbol{\omega},$$

$\Rightarrow \boldsymbol{\omega}$ is a fixed-point. □

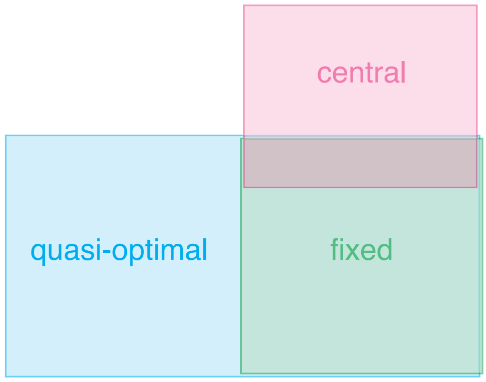


Figure 5.2: **Central Quasi-Optimal Points are Fixed**

From Lemmas 5.2.6 and 5.2.9 we have

Lemma 5.2.10 $\boldsymbol{\omega}$ is a central fixed-point $\Leftrightarrow$ $\boldsymbol{\omega}$ is a central quasi-optimal point,

and from this lemma it follows that $\boldsymbol{\omega}$ is a non-negative central fixed-point $\Leftrightarrow$ $\boldsymbol{\omega}$ is a non-negative, central quasi-optimal point and thus, since $\boldsymbol{\omega}$ is a non-negative, central quasi-optimal point $\Leftrightarrow$ $\boldsymbol{\omega}$ is a central solution to the invariant problem we have

Theorem: 5.2.11 $\boldsymbol{\omega}$ is a non-negative central fixed-point $\Leftrightarrow$
$\boldsymbol{\omega}$ is a central solution to the invariant problem.

This is the main result of this chapter as it establishes the equivalence of non-negative central solutions to the fixed-point problem and central solutions to the invariant problem.

We sum up the relation between the fixed-point, invariant and original problems with Figure 5.3.

5.3 The Solution in Detail

5.3.1 Degeneracy

From 5.9(a) we see that for **degenerate** central fixed-point solution

$$\begin{bmatrix} \boldsymbol{\mathfrak{x}} \\ \boldsymbol{\mathfrak{y}} \end{bmatrix} = \begin{bmatrix} \mathfrak{A}\mathfrak{x} - \mathfrak{b} \\ \mathfrak{D}\mathfrak{y} - \mathfrak{c} \end{bmatrix} = \mathbf{0} \Rightarrow \begin{bmatrix} \boldsymbol{\mathfrak{b}} \\ \boldsymbol{\mathfrak{c}} \end{bmatrix} = \mathbf{0} \Rightarrow \mathfrak{D}\boldsymbol{b} = \mathbf{0} \;\wedge\; A^{T+}\boldsymbol{c} = \mathbf{0} \Rightarrow \boldsymbol{b} = AA^{+}\boldsymbol{b} \;\wedge\; A^{T+}\boldsymbol{c} = \mathbf{0}$$

yielding the invariant programs

$$\max \mathbf{0}^T \boldsymbol{\mathfrak{x}} \; s.t. \; \boldsymbol{\mathfrak{x}} = \mathfrak{A}\boldsymbol{\mathfrak{x}} \geq \mathbf{0} \quad \text{and} \quad \max \mathbf{0}^T \boldsymbol{\mathfrak{y}} \; s.t. \; \boldsymbol{\mathfrak{y}} = \mathfrak{D}\boldsymbol{\mathfrak{y}} \geq \mathbf{0}$$

that is, find the invariant primal and dual sets

$$\{\boldsymbol{\mathfrak{x}} : \boldsymbol{\mathfrak{x}} = \mathfrak{A}\boldsymbol{\mathfrak{x}} \geq \mathbf{0}\} \quad \text{and} \quad \{\boldsymbol{\mathfrak{y}} : \boldsymbol{\mathfrak{y}} = \mathfrak{D}\boldsymbol{\mathfrak{y}} \geq \mathbf{0}\}$$

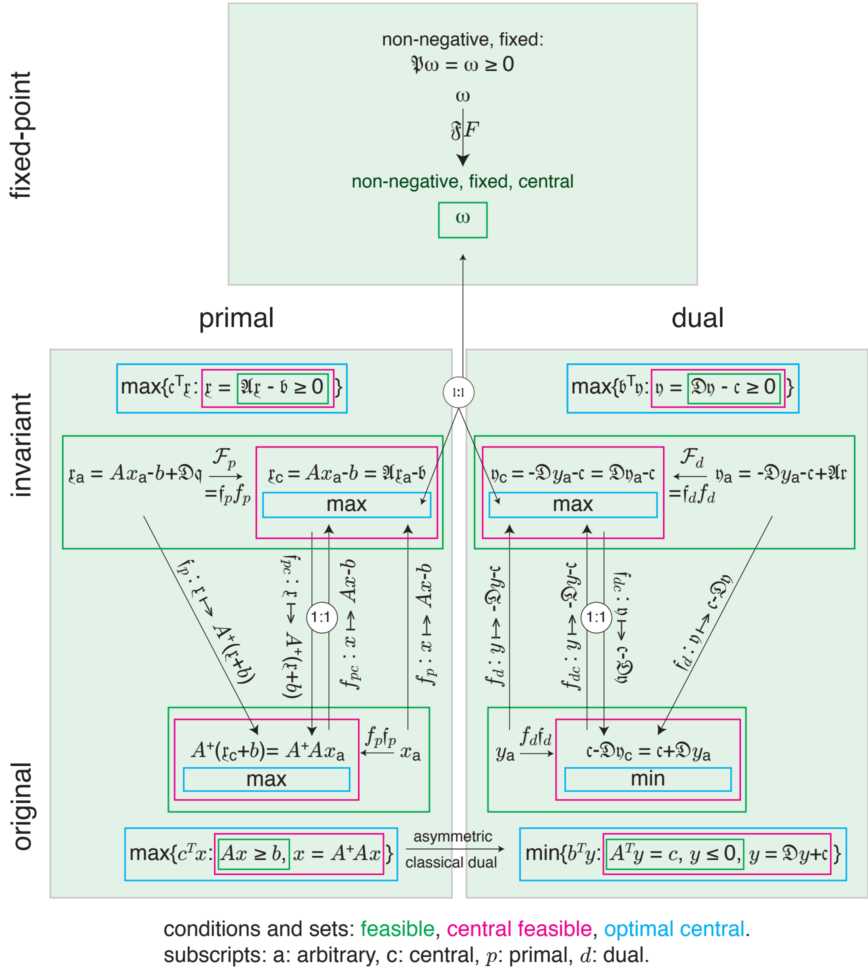


Figure 5.3: **Original, Invariant and Fixed-Point Problem Relationship**

Alternatively, considering $\mathfrak{b}$ and $\mathfrak{c}$ separately,

if $\mathfrak{c} = \mathbf{0}$ then $A^{T+}\boldsymbol{c} = \mathbf{0} \Rightarrow A^T A^{T+}\boldsymbol{c} = \mathbf{0} \Rightarrow (A^+A)^T\boldsymbol{c} = \mathbf{0} \Rightarrow A^+A\boldsymbol{c} = \mathbf{0}$, but from Definition 4.2.1 and Lemma 4.2.2, Problem 1.1 is feasible bounded $\Rightarrow A^+A\boldsymbol{c} = \boldsymbol{c}$, so $\boldsymbol{c} = \mathbf{0}$ if the problem is feasible bounded, so $max\{\boldsymbol{c}^T\boldsymbol{x} : A\boldsymbol{x} \geq \boldsymbol{b}\}$ becomes $max\{0 : A\boldsymbol{x} \geq \boldsymbol{b}\}$, and $min\{\boldsymbol{b}^T\boldsymbol{y} : A^T\boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \geq \mathbf{0}\}$ becomes min $\{\boldsymbol{b}^T\boldsymbol{y} : A^T\boldsymbol{y} = \mathbf{0}, \boldsymbol{y} \geq \mathbf{0}\}$ but since from 4.5 $\boldsymbol{b}^T\boldsymbol{y} = 0$, we have the original primal and dual problems

$$\text{find } \{x : A\boldsymbol{x} \geq \boldsymbol{b}\} \text{ and } \{\boldsymbol{y} : \boldsymbol{b}^T\boldsymbol{y} = 0, A^T\boldsymbol{y} = \mathbf{0}, \boldsymbol{y} \geq \mathbf{0}\};$$

if $\mathfrak{b} = \mathbf{0} \Rightarrow (I - AA^+)\boldsymbol{b} = \mathbf{0} \Rightarrow \boldsymbol{b} = AA^+\boldsymbol{b}$ so $min\{\boldsymbol{b}^T\boldsymbol{y} : A^T\boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \geq \mathbf{0}\}$ becomes $min\{(AA^+\boldsymbol{b})^T\boldsymbol{y} : A^T\boldsymbol{y} = \boldsymbol{c}, \boldsymbol{y} \geq \mathbf{0}\}$ - that is $min\{\boldsymbol{b}^T A^{+T} A^T\boldsymbol{y} : A^T\boldsymbol{y} = \mathbf{0}, \boldsymbol{y} \geq \mathbf{0}\}$ which becomes $min\{\mathbf{0} : A^T\boldsymbol{y} = \mathbf{0}, \boldsymbol{y} \geq \mathbf{0}\}$, so the problem becomes finding $\{\boldsymbol{y} : \boldsymbol{b}^T\boldsymbol{y} = 0, A^T\boldsymbol{y} = \mathbf{0}, \boldsymbol{y} \geq \mathbf{0}\} = \{\boldsymbol{y} : \boldsymbol{b}^T AA^+\boldsymbol{y} = 0, A^T\boldsymbol{y} = \mathbf{0}, \boldsymbol{y} \geq \mathbf{0}\} = \{\boldsymbol{y} : A^T\boldsymbol{y} = \mathbf{0}, \boldsymbol{y} \geq \mathbf{0}\}$, while the original primal maximum must be $\mathbf{0}$ and so we have the problem of finding $\{\boldsymbol{x} : \boldsymbol{c}^T\boldsymbol{x} = 0, A\boldsymbol{x} \geq \boldsymbol{b}\}$. Summarizing, we have the original primal and dual problems

$$\text{find } \{\boldsymbol{x} : \boldsymbol{c}^T\boldsymbol{x} = 0, A\boldsymbol{x} \geq \boldsymbol{b}\} \text{ and } \{\boldsymbol{y} : A^T\boldsymbol{y} = \mathbf{0}, \boldsymbol{y} \geq \mathbf{0}\}.$$

Analogously with the simplification of Figure 4.3 to Figure 4.4 we simplify Figure 5.3 to Figure 5.4.

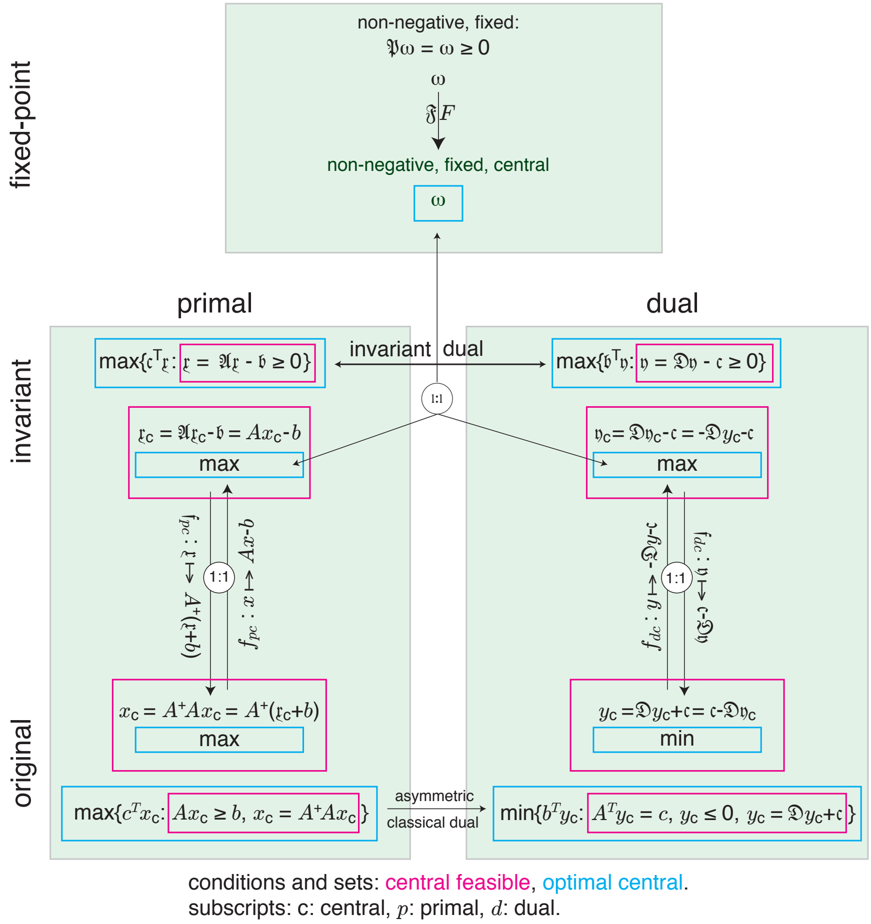


Figure 5.4: **Original, Invariant and Fixed-Point Central Problem Relationship**

5.4 The Alternative Solution

The same fixed-point problem can result from an alternative original problem, the reason being that changing the signs of A $\boldsymbol{b}$ and $\boldsymbol{c}$ affects the signs of $\mathfrak{b}$ and $\mathfrak{c}$; if the signs for $\mathfrak{b}$ and $\mathfrak{c}$ both remain the same, or both change to negative then the fixed-point matrix remain the same. So we investigate which sign changes of A, $\boldsymbol{b}$ and $\boldsymbol{c}$ cause a sign change or no sign change in both $\mathfrak{b}$ and $\mathfrak{c}$:

The possibilities are

$$
\begin{array}{rrr|rrr|c|lrlrl}
A & \boldsymbol{b} & \boldsymbol{c} & \mathfrak{A} & \mathfrak{b} & \mathfrak{c} & + & \max & \boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} & A\boldsymbol{x} \geq & \boldsymbol{b} \\
A & \boldsymbol{b} & -\boldsymbol{c} & \mathfrak{A} & \mathfrak{b} & -\mathfrak{c} & & \max & -\boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} & A\boldsymbol{x} \geq & \boldsymbol{b} \\
A & -\boldsymbol{b} & \boldsymbol{c} & \mathfrak{A} & -\mathfrak{b} & \mathfrak{c} & & \max & \boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} & A\boldsymbol{x} \geq & -\boldsymbol{b} \\
A & -\boldsymbol{b} & -\boldsymbol{c} & \mathfrak{A} & -\mathfrak{b} & -\mathfrak{c} & - & \max & -\boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} & A\boldsymbol{x} \geq & -\boldsymbol{b} \\
-A & \boldsymbol{b} & \boldsymbol{c} & \mathfrak{A} & \mathfrak{b} & -\mathfrak{c} & & \max & \boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} & -A\boldsymbol{x} \geq & \boldsymbol{b} \\
-A & \boldsymbol{b} & -\boldsymbol{c} & \mathfrak{A} & \mathfrak{b} & \mathfrak{c} & + & \max & -\boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} & -A\boldsymbol{x} \geq & \boldsymbol{b} \\
-A & -\boldsymbol{b} & \boldsymbol{c} & \mathfrak{A} & -\mathfrak{b} & -\mathfrak{c} & - & \max & \boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} & -A\boldsymbol{x} \geq & -\boldsymbol{b} \\
-A & -\boldsymbol{b} & -\boldsymbol{c} & \mathfrak{A} & -\mathfrak{b} & \mathfrak{c} & & \max & -\boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} & -A\boldsymbol{x} \geq & -\boldsymbol{b}
\end{array}
$$

of which the following problems yield the same fixed-point matrix:

$$\begin{array}{lrl} \max & \boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} \quad A\boldsymbol{x} \geq b \\ \max & -\boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} \quad A\boldsymbol{x} \geq -b \\ \max & -\boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} \; -A\boldsymbol{x} \geq b \\ \max & \boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} \; -A\boldsymbol{x} \geq -b \end{array}$$

which may be rewritten as, respectively,

$$\begin{array}{lrl} \max & \boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} \quad A\boldsymbol{x} \geq b \\ \max & -\boldsymbol{c}^T(-\boldsymbol{x}), & \textit{subject to} \quad A(-\boldsymbol{x}) \geq -b \\ \max & -\boldsymbol{c}^T(-\boldsymbol{x}), & \textit{subject to} \; -A(-\boldsymbol{x}) \geq b \\ \max & \boldsymbol{c}^T\boldsymbol{x}, & \textit{subject to} \; -A\boldsymbol{x} \geq -b \end{array}$$

which in turn may be rewritten as, respectively,

$$\begin{array}{l} \max\ \boldsymbol{c}^T\boldsymbol{x},\ \textit{subject to}\ A\boldsymbol{x} \geq b \\ \max\ \boldsymbol{c}^T\boldsymbol{x},\ \textit{subject to}\ A\boldsymbol{x} \leq b \\ \max\ \boldsymbol{c}^T\boldsymbol{x},\ \textit{subject to}\ A\boldsymbol{x} \geq b \\ \max\ \boldsymbol{c}^T\boldsymbol{x},\ \textit{subject to}\ A\boldsymbol{x} \leq b \end{array}$$

finally yielding two problems:

$$\begin{array}{ll} \max\ \boldsymbol{c}^T\boldsymbol{x},\ \textit{subject to}\ A\boldsymbol{x} \geq b & \text{original} \\ \max\ \boldsymbol{c}^T\boldsymbol{x},\ \textit{subject to}\ A\boldsymbol{x} \leq b & \text{alternative} \end{array}$$

Note that, although the alternative problem has a non-negative fixed-point, except for the already analyzed degenerate case $\begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} = \mathbf{0}$; it does not have a non-negative central fixed-point, and the solution method does not provide such spurious "solution".

Chapter 6

Fixed-Point Problem Solution

In this chapter an algorithm is developed for solution of the fixed-point problem following the scheme of Chapter 3.

We continue with $\mathfrak{P}$ as specified in Chapter 5 and introduce a matrix $\mathfrak{S}$ which swaps $\mathfrak{P}$. We introduce the function $\mathfrak{K}_z$ and show that $\mathfrak{S}$ also swaps $\mathfrak{K}_z$ and thus $\mathfrak{K}_z$ serves as a specific example of the gewneral idempotent symmetric K. We then construct $\mathfrak{U}_z$ and $\mathfrak{V}_z$ in a manner analogous to Chapter 2.2.5, thus the results of Chapter 3 apply for the specific matrices $\mathfrak{S}, \mathfrak{P}, \mathfrak{K}_z, \mathfrak{U}_z$ and $\mathfrak{V}_z$ taken together.

6.1 Construction of the Specific $\mathfrak{P}$-Unitary Matrix $\mathfrak{U}$

6.1.1 Specific S

Define the $2m \times 2m$ matrix

$$\mathfrak{S} = \begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix}, \tag{6.1}$$

then $\mathfrak{S}$ is a swapping-matrix since $\mathfrak{S}^2 = I_{2m}$, and $\mathfrak{S}^T = \mathfrak{S}$ imply $\mathfrak{S}\mathfrak{S}^T = \mathfrak{S}^T\mathfrak{S} = I$, so $\mathfrak{S}$ is unitary Hermitian.

We have

$$\mathfrak{S}\begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix}\mathfrak{S} = \begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix}\begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix}\begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix} = \begin{bmatrix} 0 & \mathfrak{D} \\ \mathfrak{A} & 0 \end{bmatrix}\begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix} = \begin{bmatrix} \mathfrak{D} & 0 \\ 0 & \mathfrak{A} \end{bmatrix} = I_{2m} - \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \tag{6.2}$$

Further note that

$$\mathfrak{S}\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{bmatrix} \tag{6.3}$$

and

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}^+ \mathfrak{S} \overset{L\,3.1.2}{=} \left(\mathfrak{S}\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}\right)^+ = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{bmatrix}^+ \tag{6.4}$$

6.1.2 Specific P

Our specific P is given by Equation 5.4 - that is $\mathfrak{P} = \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} + \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+$ so, using Equations 6.2, 6.3 and 6.4,

$$\begin{aligned} \mathfrak{S}\mathfrak{P}\mathfrak{S} &= \mathfrak{S} \left\{ \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} + \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \right\} \mathfrak{S} \\ &= \begin{bmatrix} \mathfrak{D} & 0 \\ 0 & \mathfrak{A} \end{bmatrix} + \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ = I - \mathfrak{P} \ . \end{aligned} \tag{6.5}$$

Since $\mathfrak{P}$ is an Hermitian idempotent, $\mathfrak{S}$ swaps $\mathfrak{P}$; we also have the following complementary slackness results for fixed-points of $\mathfrak{P}$: $\boldsymbol{\omega}_r^T \mathfrak{S} \boldsymbol{\omega}_s = (\mathfrak{P}\boldsymbol{\omega}_r)^T \mathfrak{S}\mathfrak{P}\boldsymbol{\omega}_s = \boldsymbol{\omega}_r^T \mathfrak{P}^T \mathfrak{S}\mathfrak{P}\boldsymbol{\omega}_s = \boldsymbol{\omega}_r^T \mathfrak{P}\mathfrak{S}\mathfrak{P}\boldsymbol{\omega}_s, \overset{3.1.1\,c}{=} 0$, that is

$$\boldsymbol{\omega}_r^T \mathfrak{S} \boldsymbol{\omega}_s = 0 \quad \textit{for fixed-points } \boldsymbol{\omega}_r \textit{ and } \boldsymbol{\omega}_s \textit{ of } \mathfrak{P}. \tag{6.6}$$

and

$$\xi_s^T \zeta_s = \begin{bmatrix} \xi_s \\ \zeta_s \end{bmatrix}^T \mathfrak{S} \begin{bmatrix} \xi_s \\ \zeta_s \end{bmatrix} /2 \overset{6.1.2}{=} 0 \ .$$

6.1.3 Specific K

Here we continue within the context of Section 2.2.5, specifying K as a diagonal matrix which forces the orthogonality (i.e. complementary slackness condition) to hold. From this specific K we construct a specific unitary matrix and an averaging matrix in a manner exactly analogous to Definitions 3.1.9 and 3.1.12 respectively.

Define

$$i' = \text{mod}_{2m}(i + m)$$

Given a vector $\boldsymbol{z}$ of length $2m$ we define the *specific Karush matrix* as the $2m$ by $2m$ diagonal matrix $\mathfrak{K}_z$ by

$$(\mathfrak{K}_z)_{ii} = \begin{cases} 1 & \textit{if} \quad z_i > z_{i'} \\ 0 & \textit{if} \quad z_i < z_{i'} \\ 1 & \textit{if} \quad z_i = z_{i'} \quad \textit{and} \quad i < i' \\ 0 & \textit{if} \quad z_i = z_{i'} \quad \textit{and} \quad i > i' \end{cases} \tag{6.7}$$

Thus $\mathfrak{K}_z$ is an Hermitian idempotent which has the value 1 at one of the indices i, i', and the value 0 at the other index.

Lemma 6.1.1 $\mathfrak{S}$ swaps $\mathfrak{K}_z$.

Proof From (6.7) we see that $\mathfrak{K}_z$ is an Hermitian idempotent matrix; further, we can write $\mathfrak{K}_z = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$ where D_1 and D_2 are diagonal matrices and $D_1 + D_2 = I$. Further,

$$\mathfrak{S}\mathfrak{K}_z\mathfrak{S} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} = \begin{bmatrix} 0 & D_2 \\ D_1 & 0 \end{bmatrix} \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$$

$$= \begin{bmatrix} D_2 & 0 \\ 0 & D_1 \end{bmatrix} = \begin{bmatrix} I - D_1 & 0 \\ 0 & I - D_2 \end{bmatrix} = I - \mathfrak{K}_z . \square$$

6.1.4 Specific U and V

Analogous to Equation 3.1.9, we define

$$\mathfrak{U}_z = \mathfrak{P}(I + \mathfrak{S})\mathfrak{K}_z(I - \mathfrak{S})\mathfrak{P}, \tag{6.8}$$

and analogous to Definition 3.1.12 the averaging matrix

$$\mathfrak{V}_z = (\mathfrak{P} + \mathfrak{U}_z)/2 . \tag{6.9}$$

Note that $\mathfrak{U}_z$ is $\mathfrak{P}$-unitary in view of Theorem 3.1.10.

Analogous to the definition of the general oblique Karush matrix $\overline{K}$ given by Equation 3.8 in Chapter 2.2.5, define

$$\overline{\mathfrak{K}}_z = \mathfrak{K}_z(I - \mathfrak{S}) . \tag{6.10}$$

Lemma 6.1.2 $\overline{\mathfrak{K}}_z \boldsymbol{z} = ((I - \mathfrak{S})\boldsymbol{z}) \vee \mathbf{0} \geq \mathbf{0}$.

Proof $(\overline{\mathfrak{K}}_z \boldsymbol{z})_i = (\mathfrak{K}_z(I - \mathfrak{S})\boldsymbol{z})_i = (\mathfrak{K}_z)_i \cdot (\boldsymbol{z} - \boldsymbol{z}') = (\mathfrak{K}_z)_{ii}(z_i - z_{i'})$

$$= \begin{cases} z_i - z_{i'} & \mathit{if} \quad z_i > z_{i'} \\ 0 & \mathit{if} \quad z_i < z_{i'} \\ z_i - z_{i'} & \mathit{if} \quad z_i = z_{i'} \quad \mathit{and} \quad i < i' \\ 0 & \mathit{if} \quad z_i = z_{i'} \quad \mathit{and} \quad i > i' \end{cases} = \begin{cases} z_i - z_{i'} & \mathit{if} \;\; z_i > z_{i'} \\ 0 & \mathit{otherwise} \end{cases} = ((I - \mathfrak{S})\boldsymbol{z}) \vee \mathbf{0}$$

So

$$\overline{\mathfrak{K}}_z \boldsymbol{z} = \mathfrak{K}_z(I - \mathfrak{S})\boldsymbol{z} = ((I - \mathfrak{S})\boldsymbol{z}) \vee \mathbf{0} = (\boldsymbol{z} - \mathfrak{S}\boldsymbol{z})/2 + |\boldsymbol{z} - \mathfrak{S}\boldsymbol{z}|/2 . \tag{6.11}$$

Further, assuming $\boldsymbol{z} = \mathfrak{P}\boldsymbol{z}$,

$$\mathfrak{U}_z \boldsymbol{z} = \mathfrak{P}(I + \mathfrak{S})\mathfrak{K}_z(I - \mathfrak{S})\boldsymbol{z} \overset{6.10}{=} \mathfrak{P}(I + \mathfrak{S})\overline{\mathfrak{K}}_z \boldsymbol{z} = \mathfrak{P}(I + \mathfrak{S})[(\boldsymbol{z} - \mathfrak{S}\boldsymbol{z})/2 + |\boldsymbol{z} - \mathfrak{S}\boldsymbol{z}|/2]$$

$$= \mathfrak{P}(I + \mathfrak{S})|\boldsymbol{z} - \mathfrak{S}\boldsymbol{z}|/2 = \mathfrak{P}|\boldsymbol{z} - \mathfrak{S}\boldsymbol{z}| ,$$

so we have the simple computational form

$$\boxed{\mathfrak{U}_z \boldsymbol{z} = \mathfrak{P}|\boldsymbol{z} - \mathfrak{S}\boldsymbol{z}| \; , \quad \mathit{for} \; \boldsymbol{z} = \mathfrak{P}\boldsymbol{z} \; .} \tag{6.12}$$

Note that Nguyen [24, p. 34], defines the function $A(x) = (|x| + x)/2$ which is shown to be the proximity map onto the non-negative cone $\{x \in \Re^{2m} : x \geq \mathbf{0}\}$. He shows that repeated application of $A\mathfrak{P}$ or of $\mathfrak{P}A$ converges to a non-negative fixed-point of $\mathfrak{P}$.

Define

$$\begin{aligned}\mathfrak{K} &: \boldsymbol{z} \mapsto \mathfrak{K}_z \boldsymbol{z} \\ \mathfrak{U} &: \boldsymbol{z} \mapsto \mathfrak{U}_z \boldsymbol{z} \\ \mathfrak{V} &: \boldsymbol{z} \mapsto \mathfrak{V}_z \boldsymbol{z}\end{aligned} \tag{6.13}$$

then we see from Equation 6.12 that $\mathfrak{K}$, $\mathfrak{U}$ and $\mathfrak{V}$ are continuous functions; note that they are not linear functions.

6.2 Proximality

We introduce the notion of proximality of vectors and show that, in the context of the LP fixed-point problem, proximality implies linear behaviour; this lays some of the groundwork for a solution method.

Definition: 6.2.1 The vector $\boldsymbol{p}$ is said to be *proximal* to $\boldsymbol{q}$ if
$\forall i: \ ((I-\mathfrak{S})\boldsymbol{q})_i > 0 \ \Rightarrow \ ((I-\mathfrak{S})\boldsymbol{p})_i \geq 0$,
while $\boldsymbol{p}$ is said to be proximal to $\boldsymbol{q}$ for component i if
$((I-\mathfrak{S})\boldsymbol{q})_i > 0 \Rightarrow ((I-\mathfrak{S})\boldsymbol{p})_i \geq 0$.

We define the i^{th} *component-pair* of a $2m$-dimensional vector x to be the pair $(x_i, x_{i'})$; a $2m$ dimensional vector x is said to have a *zero component-pair* at i if $x_i = x_{i'} = 0$.

Note that proximality is reflexive since a vector is proximal to itself, and symmetric in view of

Lemma 6.2.2 *$\boldsymbol{p}$ is proximal to $\boldsymbol{q}$ $\Leftrightarrow$ $\boldsymbol{q}$ is proximal to $\boldsymbol{p}$.*

Proof $\boldsymbol{p}$ is proximal to $\boldsymbol{q}$
implies $\forall i: \ ((I-\mathfrak{S})\boldsymbol{q})_i > 0 \Rightarrow ((I-\mathfrak{S})\boldsymbol{p})_i \geq 0$
implies $\forall i: \ ((I-\mathfrak{S})\boldsymbol{p})_i < 0 \Rightarrow ((I-\mathfrak{S})q)_i \leq 0$ *using* $(\sim (A \subseteq B)) \subseteq (\sim B \subseteq \sim A)$
implies $\forall i: \ ((I-\mathfrak{S})\boldsymbol{p})_{i'} > 0 \Rightarrow ((I-\mathfrak{S})q)_{i'} \geq 0$
implies $\forall i': \ ((I-\mathfrak{S})\boldsymbol{p})_{i'} > 0 \Rightarrow ((I-\mathfrak{S})q)_{i'} \geq 0$
implies $\forall i: \ ((I-\mathfrak{S})\boldsymbol{p})_i > 0 \Rightarrow ((I-\mathfrak{S})q)_i \geq 0\}$
implies q is proximal to $\boldsymbol{p}$. □

This means we may say "$\boldsymbol{p}$ and $\boldsymbol{q}$ are proximal" or "$\boldsymbol{p}$ is proximal with $\boldsymbol{q}$" rather than $\boldsymbol{p}$ is proximal to $\boldsymbol{q}$; proximality is reflexive and symmetric, however it is not transitive.

Remark: 6.2.3 It follows from the definition of proximality that the set of vectors proximal with a particular vector is a closed set.

Theorem: 6.2.4 The set $\{\boldsymbol{x} : \boldsymbol{x}$ and $\boldsymbol{p}$ are proximal$\}$ is a convex cone containing the zero vector.

Proof For $\boldsymbol{x}$ and $\boldsymbol{y}$ both proximal with $\boldsymbol{p}$, for all i
$\big((I-\mathfrak{S})\boldsymbol{p}\big)_i > 0 \Rightarrow \big((I-\mathfrak{S})\boldsymbol{x}\big)_i \geq 0 \wedge \big((I-\mathfrak{S})\boldsymbol{y}\big)_i \geq 0$
$\Rightarrow \big((I-\mathfrak{S})(\lambda\boldsymbol{x}+\mu\boldsymbol{y})\big)_i \geq 0$ for $\lambda, \mu \geq 0$
$\Rightarrow \lambda\boldsymbol{x}+\mu\boldsymbol{y}$ is proximal with $\boldsymbol{p}$ for $\lambda, \mu \geq 0$.

Lemma 6.2.5 If $\boldsymbol{x}$ and $\boldsymbol{y}$ are mutually proximal then any two elements of $[\boldsymbol{x}, \boldsymbol{y}]$ are mutually proximal.

Lemma 6.2.6 If $\boldsymbol{\omega}_r$ and $\boldsymbol{\omega}_s$ are non-negative fixed-points of $\mathfrak{P}$ then $\boldsymbol{\omega}_r$ and $\boldsymbol{\omega}_s$ are mutually proximal.

Proof $\boldsymbol{\omega}_r$ and $\boldsymbol{\omega}_s$ non-negative fixed $\Rightarrow \boldsymbol{\omega}_r^T \mathfrak{S} \boldsymbol{\omega}_s = \boldsymbol{\omega}_r^T \mathfrak{P} \mathfrak{S} \mathfrak{P} \boldsymbol{\omega}_s \overset{3.1.1\,c}{=} 0$, so $\forall i : \omega_{ri}\omega_{si'} = 0 \Rightarrow \forall i : \omega_{ri} > 0 \Rightarrow \omega_{si'} = 0$ but $\forall i : \omega_{ri} > 0 \Leftrightarrow ((I - \mathfrak{S})\omega_r)_i > 0$, and $\forall i : \omega_{si'} = 0 \Leftrightarrow ((I - \mathfrak{S})\omega_s)_i \geq 0$, so $\forall i : ((I - \mathfrak{S})\omega_r)_i > 0 \Rightarrow ((I - \mathfrak{S})\omega_s)_i \geq 0$, that is $\boldsymbol{\omega}_r$ and $\boldsymbol{\omega}_s$ are mutually proximal.

Note that the set of non-negative fixed points is a convex cone.

6.3 Orbits

We define a sequence of vectors $\{\mathfrak{g}_i\}$ which orbits the fixed-points of $\mathfrak{P}$ and a sequence of vectors $\{\mathfrak{v}_i\}$ which converges to a fixed-point of $\mathfrak{P}$. An algorithm which combines the averaging function $\mathfrak{V}$ and affine regression is proposed and it is shown that this algorithm terminates.

The first scheme we adopt is to form the sequence $\{\mathfrak{v}_i\}$ analogous in construction to the sequence $\{v_i\}$ of Sub-section 3.2.1. We know that this sequence converges, but usually the convergence is slow, so to accelerate the process we modify the scheme by using the affine regression method detailed in Chapter 3.2.2, however this is still not computationally satisfactory as we would need to compute n points and then regress without being sure that the points were proximal; to overcome this problem an incremental affine regression algorithm is developed in Appendix B; if the regression does not yield a non-negative fixed-point then proximality has not been reached and we need to proceed further with the series $\{v_i\}$.

We may use the following result:

Lemma 6.3.1 $\psi = \begin{bmatrix} \mathfrak{A}\xi \\ \mathfrak{D}\zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}$ is a central fixed-point.

Proof fixed:

$$\mathfrak{P}\psi = \left(\begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} + \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \right) \left(\begin{bmatrix} \mathfrak{A}\xi \\ \mathfrak{D}\zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \right)$$

$$= \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \begin{bmatrix} \mathfrak{A}\xi \\ \mathfrak{D}\zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}$$

$$+ \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ \begin{bmatrix} \mathfrak{A}\xi \\ \mathfrak{D}\zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}^+ \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}$$

$$- \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \mathfrak{A}\xi \\ \mathfrak{D}\zeta \end{bmatrix} + \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} + \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}$$

$$= \begin{bmatrix} \mathfrak{A}\xi \\ \mathfrak{D}\zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \mathbf{0} \quad + \mathbf{0} - \mathbf{0} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \quad - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} + \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} + \mathbf{0}$$

$$= \begin{bmatrix} \mathfrak{A}\xi \\ \mathfrak{D}\zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} = \psi$$

central:

$$\begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \psi - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} = \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \left(\begin{bmatrix} \mathfrak{A}\xi \\ \mathfrak{D}\zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} \right) - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}$$

$$= \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \begin{bmatrix} \mathfrak{A}\xi \\ \mathfrak{D}\zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{A} & 0 \\ 0 & \mathfrak{D} \end{bmatrix} \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix}$$

$$= \begin{bmatrix} \mathfrak{A}\xi \\ \mathfrak{D}\zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix} \begin{bmatrix} \mathfrak{c} \\ \mathfrak{b} \end{bmatrix}^+ \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \begin{bmatrix} \mathfrak{b} \\ \mathfrak{c} \end{bmatrix} = \psi \; . \; \square$$

6.4 Fixed-Point Analysis

Here we build on the fixed-point theory of Chapter 3, laying the basis for the affine regression algorithm of the following section.

Definitions: Let $\boldsymbol{p}$ be a non-negative fixed-point of $\mathfrak{P}$ then, with $\rhd$ as defined by Equation 2.8 a, the vector $\boldsymbol{p}_z = \boldsymbol{z} \rhd \boldsymbol{p}$ is the **fixed-point component** of $\boldsymbol{z}$ w.r.t $\boldsymbol{p}$. With $\lhd$ as defined by Equation 2.8 b, the **residual component** of $\boldsymbol{z}$ w.r.t $\boldsymbol{p}$ is defined as the vector $\boldsymbol{p}^z = \boldsymbol{z} \lhd \boldsymbol{p}$. The scalar $\|\boldsymbol{p}_z\| \,/\, \|\boldsymbol{z}\|$ is the **fixed-point proportion** of $\boldsymbol{p}$ in $\boldsymbol{z}$.

We start with a vector which necessarily contains a positive component of a non-negative fixed-point of $\mathfrak{P}$, if such fixed-point exists; a suitable choice is $\mathfrak{v}_1 = \mathfrak{P}\mathbf{1}_{2m}$, where $\mathbf{1}_{2m}$ is a $2m-$dimensional vector each of whose entries is unity, then we consider the sequences $\{\mathfrak{g}_i\}$ and $\{\mathfrak{v}_i\}$ where

$$\mathfrak{g}_{i+1} = \mathfrak{U}\mathfrak{g}_i, \quad \mathfrak{g}_1 = \mathfrak{P}\mathbf{1} \tag{6.14}$$

and

$$\mathfrak{v}_{i+1} = \mathfrak{V}\mathfrak{v}_i, \quad \mathfrak{v}_1 = \mathfrak{P}\mathbf{1} \tag{6.15}$$

and note that if a non-negative fixed-point of $\mathfrak{P}$, say $\mathfrak{p}$, exists then $\mathfrak{p}^T\mathfrak{u}_1 = \mathfrak{p}^T\mathfrak{P}\mathbf{1} = \mathfrak{p}^T\mathbf{1} \geq \mathbf{0}$, so $\mathfrak{u}_1$ contains a non-negative component of $\mathfrak{p}$.

The first result is that the fixed-points of $\mathfrak{U}$ are precisely the non-negative fixed-points of $\mathfrak{P}$.

Theorem: 6.4.1 $\mathfrak{U}\boldsymbol{z} = \boldsymbol{z} \Leftrightarrow \boldsymbol{z}$ is a non-negative fixed-point of $\mathfrak{P}$.

Proof: $\mathfrak{U}\boldsymbol{z} = \boldsymbol{z} \overset{6.13}{\Leftrightarrow} \mathfrak{U}_z\boldsymbol{z} = \boldsymbol{z} \overset{T\,3.1.17}{\Leftrightarrow} (\mathfrak{P}\boldsymbol{z} = \boldsymbol{z} \text{ and } \mathfrak{K}_z\boldsymbol{z} = \boldsymbol{z}) \overset{L\,3.1.5}{\Leftrightarrow} (\mathfrak{P}\boldsymbol{z} = \boldsymbol{z} \text{ and } \overline{\mathfrak{K}}_z\boldsymbol{z} = \boldsymbol{z}) \overset{L\,6.1.2}{\Rightarrow} \mathfrak{P}\boldsymbol{z} = \boldsymbol{z} \geq 0$. Conversely, $\mathfrak{P}\boldsymbol{z} = \boldsymbol{z} \geq 0 \Rightarrow \big(\mathfrak{P}\boldsymbol{z} = \boldsymbol{z} \geq 0\big) \wedge \big(\overline{\mathfrak{K}}_z\boldsymbol{z} \overset{6.11}{=} (\boldsymbol{z} - \mathfrak{S}\boldsymbol{z}) \vee 0 \geq \boldsymbol{z} - \mathfrak{S}\boldsymbol{z}\big) \Rightarrow \big(\mathfrak{P}\boldsymbol{z} = \boldsymbol{z} \geq 0\big) \wedge \big(\boldsymbol{z}^T\overline{\mathfrak{K}}_z\boldsymbol{z} \geq \boldsymbol{z}^T(\boldsymbol{z} - \mathfrak{S}\boldsymbol{z}) \overset{L\,3.1.4}{=} \boldsymbol{z}^T\boldsymbol{z}\big) \overset{C\,3.1.8}{\Rightarrow} \big(\mathfrak{P}\boldsymbol{z} = \boldsymbol{z} \geq 0\big) \wedge \big(\boldsymbol{z}^T\overline{\mathfrak{K}}_z\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}\big) \Rightarrow \big(\mathfrak{P}\boldsymbol{z} = \boldsymbol{z} \geq 0\big) \wedge \big(\boldsymbol{z}^T\overline{\mathfrak{K}}_z\boldsymbol{z} = \boldsymbol{z}^T\boldsymbol{z}\big) \wedge \big(\|\boldsymbol{z} - \overline{\mathfrak{K}}_z\boldsymbol{z}\|^2 = \boldsymbol{z}^T\boldsymbol{z} - 2\boldsymbol{z}^T\overline{\mathfrak{K}}_z\boldsymbol{z} + (\overline{\mathfrak{K}}_z\boldsymbol{z})^T\overline{\mathfrak{K}}_z\boldsymbol{z}\big) \Rightarrow \big(\mathfrak{P}\boldsymbol{z} = \boldsymbol{z} \geq 0\big) \wedge \big(\|\boldsymbol{z} - \overline{\mathfrak{K}}_z\boldsymbol{z}\|^2 = \boldsymbol{z}^T\boldsymbol{z} - 2\boldsymbol{z}^T\boldsymbol{z} + (\overline{\mathfrak{K}}_z\boldsymbol{z})^T\overline{\mathfrak{K}}_z\boldsymbol{z}\big) \Rightarrow \big(\mathfrak{P}\boldsymbol{z} = \boldsymbol{z} \geq 0\big) \wedge \big(\|\boldsymbol{z} - \overline{\mathfrak{K}}_z\boldsymbol{z}\|^2 \overset{L\,3.1.7\,a}{=} \boldsymbol{z}^T\boldsymbol{z} - 2\boldsymbol{z}^T\boldsymbol{z} + \boldsymbol{z}^T\boldsymbol{z} = 0\big) \Rightarrow (\mathfrak{P}\boldsymbol{z} = \boldsymbol{z}) \wedge (\overline{\mathfrak{K}}_z\boldsymbol{z} = \boldsymbol{z}) \overset{C\,3.1.14}{\Rightarrow} (\mathfrak{P}\boldsymbol{z} = \boldsymbol{z}) \wedge (\mathfrak{K}_z\boldsymbol{z} = \boldsymbol{z}) \overset{L\,3.1.15}{\Rightarrow} \mathfrak{U}_z\boldsymbol{z} = \boldsymbol{z} \Rightarrow \mathfrak{U}\boldsymbol{z} = \boldsymbol{z}$.

We next show that, for any non-negative fixed-point $\mathfrak{p}$ of $\mathfrak{P}$, the component of this fixed-point in any arbitrary fixed-point $\boldsymbol{z}$ (i.e. not necessarily non-negative) increases unless $\boldsymbol{z}$ is proximal to $\mathfrak{p}$:

Theorem: 6.4.2 If $\mathfrak{p}$ is a non-negative fixed-point of $\mathfrak{P}$ and $\mathfrak{z}$ is any fixed-point of $\mathfrak{P}$ then $\mathfrak{p} \cdot \mathfrak{U}\mathfrak{z} \geq \mathfrak{p} \cdot \mathfrak{z}$; further, $\mathfrak{p} \cdot \mathfrak{U}\mathfrak{z} = \mathfrak{p} \cdot \mathfrak{z} \Leftrightarrow \mathfrak{z}$ is proximal to $\mathfrak{p}$.

Proof $\mathfrak{p}\cdot\mathfrak{U}\mathfrak{z} \overset{6.13}{=} \mathfrak{p}\cdot\mathfrak{U}_{\mathfrak{z}}\mathfrak{z} \overset{6.12}{=} \mathfrak{p}\cdot\mathfrak{P}|\mathfrak{z}-\mathfrak{S}\mathfrak{z}| = \mathfrak{p}\cdot|\mathfrak{z}-\mathfrak{S}\mathfrak{z}| \geq \mathfrak{p}\cdot(\mathfrak{z}-\mathfrak{S}\mathfrak{z}) = \mathfrak{p}\cdot\mathfrak{z}-\mathfrak{p}\cdot\mathfrak{S}\mathfrak{z}$ $= \mathfrak{p}\cdot\mathfrak{z}-(\mathfrak{P}\mathfrak{p})\cdot\mathfrak{S}\mathfrak{P}\mathfrak{z} = \mathfrak{p}\cdot\mathfrak{z}-\mathfrak{p}^T\mathfrak{P}\mathfrak{S}\mathfrak{P}\mathfrak{z} \overset{3.1.1c}{=} \mathfrak{p}\cdot\mathfrak{z}$; further, in the preceding calculation we have $\mathfrak{p}\cdot|\mathfrak{z}-\mathfrak{S}\mathfrak{z}| \geq \mathfrak{p}\cdot(\mathfrak{z}-\mathfrak{S}\mathfrak{z})$; equality will occur in this expression iff $\mathfrak{p}_i > 0 \Rightarrow (\mathfrak{z})_i-(\mathfrak{z})_{i'} \geq 0\ \forall i$, which is precisely the condition that $\mathfrak{z}$ is proximal to $\mathfrak{p}$. □

Thus fixed-point content of $\{\mathfrak{g}_i\}$ monotonic strictly increasing until proximality is reached, then it remains constant unless $\{\mathfrak{g}_i\}$ becomes non-proximal again (which can, but eventually won't, happen).

Corollary: 6.4.3 For $\mathfrak{p}$ a non-negative fixed-point of $\mathfrak{P}$ and $\mathfrak{z}$ any fixed-point of $\mathfrak{P}$, $\mathfrak{p}\cdot\mathfrak{V}\mathfrak{z} \geq \mathfrak{p}\cdot\mathfrak{z}$, moreover $\mathfrak{p}\cdot\mathfrak{V}\mathfrak{z} = \mathfrak{p}\cdot\mathfrak{z} \Leftrightarrow \mathfrak{z}$ is proximal to $\mathfrak{p}$

Proof $\mathfrak{p}\cdot\mathfrak{V}\mathfrak{z} = \mathfrak{p}\cdot(I+\mathfrak{U})\mathfrak{z}/2 = \mathfrak{p}\cdot\mathfrak{z}/2+\mathfrak{p}\cdot\mathfrak{U}\mathfrak{z}/2 \overset{T\,6.4.2}{\geq} \mathfrak{p}\cdot\mathfrak{z}/2+\mathfrak{p}\cdot\mathfrak{z}/2 = \mathfrak{p}\cdot\mathfrak{z}$. The second part of the lemma also follows on applying the second part of Theorem 6.4.2 to the calculations.

Corollary: 6.4.4 Let $\mathfrak{p}$ be a non-negative fixed-point of $\mathfrak{P}$ then $\mathfrak{p}\cdot\mathfrak{v}_{i+1} \geq \mathfrak{p}\cdot\mathfrak{v}_i$ moreover $\mathfrak{p}\cdot\mathfrak{v}_{i+1} = \mathfrak{p}\cdot\mathfrak{v}_i \Leftrightarrow \mathfrak{v}_i$ is proximal to $\mathfrak{p}$.

Proof $\mathfrak{p}\cdot\mathfrak{v}_{i+1} = \mathfrak{p}\cdot\mathfrak{V}\mathfrak{v}_i \overset{C\,6.4.3}{\geq} \mathfrak{p}\cdot\mathfrak{v}_i$. □

Let $\mathfrak{v}_\infty = \lim_{i\to\infty}\mathfrak{V}^i\mathbf{1}_{2m}$, and $\boldsymbol{z} = \mathfrak{P}\mathbf{1}_{2m}$, then

$$\mathfrak{v}_\infty\cdot\mathfrak{V}^i\mathbf{1}_{2m} \geq \mathfrak{v}_\infty\cdot\mathfrak{P}\mathbf{1}_{2m} \Rightarrow \mathfrak{v}_\infty\cdot\lim_{i\to\infty}\mathfrak{V}^i\mathbf{1}_{2m} \geq \mathfrak{v}_\infty\cdot\mathbf{1}_{2m}$$

$$\Rightarrow 1 \geq \frac{\mathfrak{v}_\infty\cdot\mathbf{1}_{2m}}{\mathfrak{v}_\infty\cdot\mathfrak{v}_\infty} \Rightarrow \mathfrak{v}_\infty \geq \frac{\mathfrak{v}_\infty\cdot P\mathbf{1}_{2m}}{\mathfrak{v}_\infty\cdot\mathfrak{v}_\infty}\mathfrak{v}_\infty \Rightarrow \mathfrak{v}_\infty \geq \mathbf{1}_{2m} \rhd \mathfrak{v}_\infty$$

It is possible for the sequence $\mathfrak{g}_i$ to move from proximality to non-proximality however eventually it will end up permanently proximal to all the fixed-points. The situation is described by Figure 6.1.

Although just as for $\{\mathfrak{u}_i\}$ if a non-trivial non-negative fixed-point exists then the fixed-point content of $\{\mathfrak{v}_i\}$ only increases until proximality is attained, $\{\mathfrak{v}_i\}$ decreases in norm and thus its fixed-point proportion continues to increase. By a slight extension of the argument of Chapter 3.2.1 we can prove that the sequence $\{\mathfrak{v}_i\}$ converges to a non-negative fixed-point, and that this limit will be non-trivial.

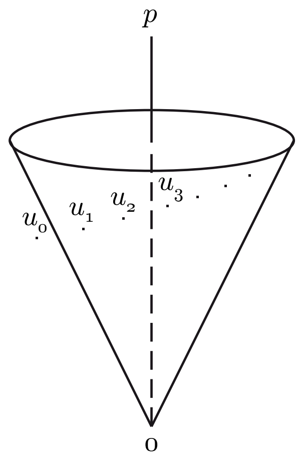


Figure 6.1: **Convergence to Proximality**

6.5 Infeasibility Detection

We prove that if there is a solution to the fixed-point problem then the solution will have a component-pair containing a zero entry with the other entry greater than or equal to unity; it then follows the norm of a vector in the sequence $\{\mathfrak{v}_i\}$ lies between unity and $\sqrt{m}$. If the norm falls below unity then the problem is infeasible or feasible unbounded.

From Corollary 6.4.4 for $\boldsymbol{p}$ a non-negative fixed-point of $\mathfrak{U}$, $\boldsymbol{p}\cdot\mathfrak{v}_\infty \geq \boldsymbol{p}\cdot\mathfrak{v}_1$. Thus $\boldsymbol{p}\cdot\mathfrak{v}_\infty \geq \boldsymbol{p}\cdot\mathfrak{v}_1 = \boldsymbol{p}\cdot\mathfrak{P}\mathbf{1}_{2m} = (\mathfrak{P}^T\boldsymbol{p})\cdot\mathbf{1}_{2m} = (\mathfrak{P}\boldsymbol{p})\cdot\mathbf{1}_{2m} = \boldsymbol{p}\cdot\mathbf{1}_{2m}$, that is

$$\boldsymbol{p}\cdot\mathfrak{v}_\infty \geq \boldsymbol{p}\cdot\mathbf{1}_{2m}$$

Since $\mathfrak{v}_\infty$ is a non-negative fixed-point of $\mathfrak{U}$ we can set $\boldsymbol{p} = \mathfrak{v}_\infty$ in this equation arriving at

$$\mathfrak{v}_\infty\cdot\mathfrak{v}_\infty \geq \mathfrak{v}_\infty\cdot\mathbf{1}_{2m} \Rightarrow 1 \geq \frac{\mathbf{1}_{2m}\cdot\mathfrak{v}_\infty}{\mathfrak{v}_\infty\cdot\mathfrak{v}_\infty} \Rightarrow \mathfrak{v}_\infty \geq \frac{\mathbf{1}_{2m}\cdot\mathfrak{v}_\infty}{\mathfrak{v}_\infty\cdot\mathfrak{v}_\infty}\mathfrak{v}_\infty \geq \mathbf{0}$$

which, applying Equation 2.8 a,

$$\Rightarrow \mathfrak{v}_\infty \geq \mathbf{1}_{2m} \rhd \mathfrak{v}_\infty \geq \mathbf{0}\ .$$

This result is illustrated by Figure 6.2.

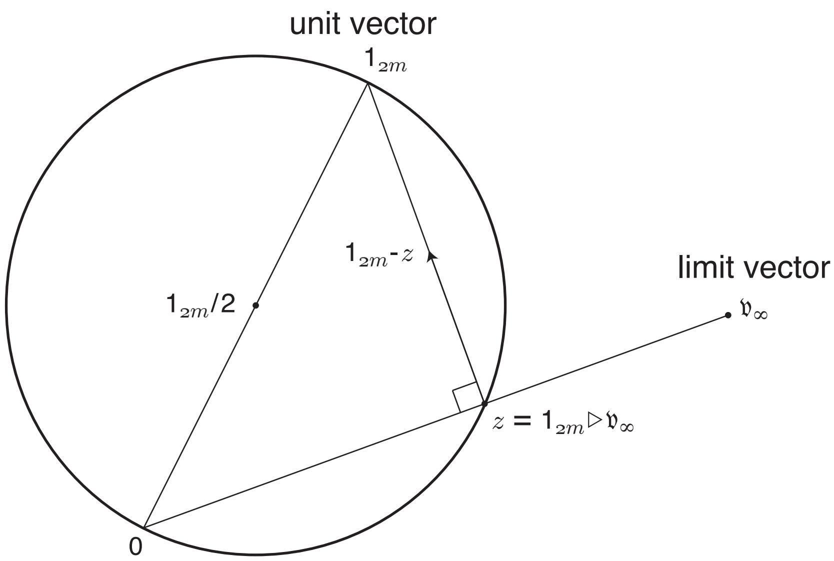


Figure 6.2: **The Unit and Limit Vectors**

We set $\boldsymbol{z} = \mathbf{1}_{2m} \rhd \mathfrak{v}_\infty$, so z is the projection of the vector $\mathbf{1}_{2m}$ onto the vector $\mathfrak{v}_\infty$, and

$$\|\mathfrak{v}_\infty\| \geq \|z\|\ . \tag{6.16}$$

We have, assuming $v_\infty \neq 0$,

$$(\mathbf{1}_{2m} \lhd \mathfrak{v}_\infty)\cdot(\mathbf{1}_{2m} \rhd \mathfrak{v}_\infty) \overset{R\,2.3.2}{=} 0 \Rightarrow (\mathbf{1}_{2m} - \boldsymbol{z})^T\boldsymbol{z} = 0 \quad\Rightarrow\quad \sum_{i=1}^{2m}(1-z_i)z_i = 0 \quad\Rightarrow\quad \sum_{i=1}^{2m} z_i^2 = \sum_{i=1}^{2m} z_i$$

$$\Rightarrow\quad \sum_{i=1}^{2m}(z_i - 0.5 + 0.5)^2 = \sum_{i=1}^{2m} z_i \quad\Rightarrow\quad \sum_{i=1}^{2m}(z_i - 0.5)^2 + \sum_{i=1}^{2m}(z_i - 0.5) + \sum_{i=1}^{2m} 0.25 = \sum_{i=1}^{2m} z_i$$

$$\begin{aligned}
&\Rightarrow \quad \sum_{i=1}^{2m}(z_i-0.5)^2+\sum_{i=1}^{2m}z_i-m+m/2=\sum_{i=1}^{2m}z_i\\
&\Rightarrow \quad \sum_{i=1}^{2m}(z_i-0.5)^2=m/2\\
&\Rightarrow \quad \big(\exists i\in\{1,\cdots,2m\}:(z_i-0.5)^2>1/4\big)\vee\big(\forall i\in\{1,\cdots,2m\}:z_i\in\{0,1\}\big)\\
&\Rightarrow \quad \big(\exists i\in\{1,\cdots,2m\}:z_i-0.5>1/2\vee z_i-0.5<-1/2\big)\vee\big(\forall i\in\{1,\cdots,2m\}:z_i\in\{0,1\}\big)\\
&\Rightarrow \quad \big(\exists i\in\{1,\cdots,2m\}:z_i>1\big)\vee\big(\forall i\in\{1,\cdots,2m\}:z_i\in\{0,1\}\big)\\
&\Rightarrow \quad \|z\|>1\overset{6.16}{\Rightarrow}\|\mathfrak{v}_\infty\|>1\ .
\end{aligned}$$

Now from 3.1.21 $\|\mathfrak{v}_i\|$ is a decreasing sequence, so $\|\mathfrak{v}_i\|<1\Rightarrow\|\mathfrak{v}_\infty\|<1$, thus we have a test for infeasibility/feasible unboundedness: if for some i, $\|\mathfrak{v}_i\|<1$ then the problem has no solution - it is infeasible or unbounded.

6.6 The Algorithm

The algorithm relies on the process of averaging leading eventually to proximality.

We know that if we have a g_1 which is proximal then $g_i=\mathfrak{U}_{g_1}^{i-1}g_1$ orbits a fixed-point. We don't worry whether or not g_1 to g_i are proximal - it is the proximality of the binding pattern which generates the sequence which matters: we simply use the binding pattern from g_1 to determine $\mathfrak{U}$ and keep this same $\mathfrak{U}$ as we generate $g_2,\cdots,g_i$, even if the binding pattern of subsequent g's changes, and apply affine regression. The regression is performed step by step in a manner similar to the conjugate gradient method (refer to 3.2.2) and at each step the resulting vector is checked to see if the negative proportion has decreased; if it fails to decrease, the previous vector together with its binding pattern becomes the starting point for a new affine regression. Note that the first step in affine regression is equivalent to applying the averaging matrix.

6.7 Conclusion

In summary it has been shown that it is possible to solve the linear programming problem by transforming it to an invariant form, then to a fixed-point problem, and then solving the resulting fixed-point problem using linear and linear inequality transformations. The approach is a full solution to the LP problem, handling degenerate cases in a manner which is transparent from the linear algebraic point of view. A converging algorithm has been described which handles even the most problematic linear programs.

6.8 Exercises

1. Is the set of vectors proximal to a vector, say $\boldsymbol{p}$, convex?

2. Is this set open or is it closed?

Chapter 7

The Transportation Problem

The transportation problem can be regarded as the problem of minimizing the cost of transporting a commodity from m sources to $n-1$ destinations, where the cost of moving a commodity unit from source i to destination j is b_{ij} currency units. The problem was formulated by F.H. Hitchcock in 1941 [13]. The reader may refer to [29, Ch. 3] for the mathematical description of, and classical solutions to, this problem.

7.1 Specification

From an economic viewpoint it is not necessary to assume that total supply is equal to total demand as one may exceed the other in a disequilibrium situation. With w_{ij} commodity units transported from source i to destination j we have the *unbalanced* transportation problem

minimize	$\sum_{i=1}^{m} \sum_{j=1}^{n-1} b_{ij} w_{ij}$	*subject to*
$\sum_{j=1}^{n-1} w_{ij} \leq s_i$	$i = 1, \cdots, m$	*(supply constraints)*
$\sum_{i=1}^{m} w_{ij} \geq d_j$	$j = 1, \cdots, n-1$	*(demand constraints)*
$w_{ij} \geq 0$	$\forall\ i, j.$	

If supply exceeds demand then a dummy destination (with index $j = n$) is introduced with demand equal to the excess, then the condition $\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j$ will obtain, and the system can be written as

minimize	$\sum_{i=1}^{m} \sum_{j=1}^{n} b_{ij} w_{ij}$	*subject to*
$\sum_{j=1}^{n} w_{ij} = s_i$	$i = 1, \cdots, m$	*(supply constraints)*
$\sum_{i=1}^{m} w_{ij} = d_j$	$j = 1, \cdots, n$	*(demand constraints)*
$w_{ij} \geq 0$	$\forall\ i, j,$	

which is the *standard* transportation problem. If demand exceeds supply a dummy source with supply equal to the excess demand is added.

The above formulation can be written as

$$\text{minimize } \boldsymbol{b}^T\boldsymbol{w} \text{ subject to } A_\tau^T\boldsymbol{w} = \boldsymbol{c},\ \ \boldsymbol{w} \geq \boldsymbol{0}, \qquad \text{(a)}$$

where

$$A_\tau = \left[\, I_m \otimes \mathbf{1}_n \middle| \mathbf{1}_m \otimes I_n \,\right] \qquad \text{(b)}$$

$$\boldsymbol{b}_\tau = \begin{bmatrix} b_{11} \\ \vdots \\ b_{1n} \\ \vdots \\ b_{m1} \\ \vdots \\ b_{mn} \end{bmatrix} \quad \text{or, in matrix form,} \quad B_\tau = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} \qquad \text{(c)}$$

$$(7.1)$$

$$\boldsymbol{c}_\tau = \begin{bmatrix} s \\ d \end{bmatrix} \qquad \text{(d)}$$

and

$$\boldsymbol{w} = \begin{bmatrix} w_{11} \\ \vdots \\ w_{1n} \\ \vdots \\ w_{m1} \\ \vdots \\ w_{mn} \end{bmatrix} \quad \text{or, in matrix form,} \quad W_\tau = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \cdots & w_{mn} \end{bmatrix} \qquad \text{(e)}$$

where I_m is an $m \times m$ matrix, $\mathbf{1}_i$ is an $i \times 1$ vector of 1's, $\otimes$ is the Kronecker product (considered in detail in Section 2.2.8), s is the $m-$dimensional supply vector, and d is the $n-$dimensional demand vector.

7.2 Invariant Form

Here the invariant form theory of Chapter 4 is applied to the transport problem 7.1. Note that some references are made to calculations in Appendix 7.3.

Formulation 7.1 has, in view of Equation 4.3, the dual LP

$$max\ \boldsymbol{c}_\tau^T\boldsymbol{x}\ s.t.\ A_\tau\boldsymbol{x} \leq \boldsymbol{b}_\tau,$$

which has the form given by Equation 1.2 (a).

Now from the calculations in Appendix 7.3, from Equation 7.8:

$$A_\tau^+ = \left[\frac{I_{m,n} \otimes \mathbf{1}_n^+}{\mathbf{1}_m^+ \otimes I_{n,m}}\right] \ ,$$

and from Equation 7.9

$$\mathfrak{A}_\tau = I_{m,n} \otimes \mathbf{1}_n\mathbf{1}_n^+ + \mathbf{1}_m\mathbf{1}_m^+ \otimes I_{n,m}\ ,$$

where

$$I_{m,n} = I_m - \frac{m\mathbf{1}_m\mathbf{1}_m^+}{m+n} \text{ and } I_{n,m} = I_n - \frac{n\mathbf{1}_n\mathbf{1}_n^+}{m+n}$$

Further, referring to Appendix 7.3 ,[1]

$$\begin{aligned}\mathfrak{D}_\tau &\overset{7.11}{=} (I_m - \mathbf{1}_m\mathbf{1}_m^+)\otimes(I_n - \mathbf{1}_n\mathbf{1}_n^+) = \big(I_m(I_m - \mathbf{1}_m\mathbf{1}_m^+)\big)\otimes\big((I_n - \mathbf{1}_n\mathbf{1}_n^+)I_n\big)\\ &\overset{2.7\,d}{=} \big(I_m\otimes(I_n - \mathbf{1}_n\mathbf{1}_n^+)\big)\big((I_m - \mathbf{1}_m\mathbf{1}_m^+)\otimes I_n\big)\\ &\overset{L\,2.2.29}{=} \big((I_m - \mathbf{1}_m\mathbf{1}_m^+)\otimes I_n\big)\big(I_m\otimes(I_n - \mathbf{1}_n\mathbf{1}_n^+)\big)\end{aligned} \tag{7.2}$$

So

$$\mathfrak{b}_\tau = \mathfrak{D}_\tau b_\tau \overset{7.2}{=} \big((I_m - \mathbf{1}_m\mathbf{1}_m^+)\otimes I_n\big)\big(I_m\otimes(I_n - \mathbf{1}_n\mathbf{1}_n^+)\big)\boldsymbol{b}_\tau \tag{7.3}$$

Note that, with $\{b_{ij}\}_\tau$ in matrix form B_τ, the matrices $(I_m - \mathbf{1}_m\mathbf{1}_m^+)\otimes I_n$ and $I_m\otimes(I_n - \mathbf{1}_n\mathbf{1}_n^+)$ effect column and row mean correction respectively (note that these processes commute) since the RHS of Equation 7.3 corresponds to

$$\mathfrak{B}_\tau = (I_m - \mathbf{1}_m\mathbf{1}_m^+)B_\tau(I_n - \mathbf{1}_n\mathbf{1}_n^+) \tag{7.4}$$

and we are multiplying B_τ on right and left by idempotent matrices which effect row and column mean correction respectively; further it is obvious from the associativity of matrix multiplication that these mean corrections "commute" and, since the matrices are idempotent, only need to be applied once.

The calculation of $\mathfrak{c}_\tau$ is as follows: $A_\tau^{T+} = \begin{bmatrix} I_{m,n}\otimes\mathbf{1}_n^{T+} & \mathbf{1}_m^{T+}\otimes I_{n,m}\end{bmatrix}$ (refer to Appendix 7.3) so

$$\begin{aligned}\mathfrak{c}_\tau &\overset{4.1\,d}{=} A_\tau^{T+}\boldsymbol{c}_\tau = \begin{bmatrix} I_{m,n}\otimes\mathbf{1}_n^{T+} & \mathbf{1}_m^{T+}\otimes I_{n,m}\end{bmatrix}\begin{bmatrix}\boldsymbol{s}\\ \boldsymbol{d}\end{bmatrix}\\
&= \big(I_{m,n}\otimes\mathbf{1}_n^{T+}\big)\,\boldsymbol{s} + \big(\mathbf{1}_m^{T+}\otimes I_{n,m}\big)\,\boldsymbol{d} = \big(I_{m,n}\otimes\mathbf{1}_n^{T+}\big)(\boldsymbol{s}\otimes 1) + \big(\mathbf{1}_m^{T+}\otimes I_{n,m}\big)(1\otimes\boldsymbol{d})\\
&\overset{2.7\,d}{=} (I_{m,n}\boldsymbol{s})\otimes\mathbf{1}_n^{T+} + \mathbf{1}_m^{T+}\otimes(I_{n,m}\boldsymbol{d})\\
&= \left((I_m - \frac{m\mathbf{1}_m\mathbf{1}_m^+}{m+n})\boldsymbol{s}\right)\otimes\mathbf{1}_n^{T+} + \mathbf{1}_m^{T+}\otimes\left((I_n - \frac{n\mathbf{1}_n\mathbf{1}_n^+}{m+n})\boldsymbol{d}\right) .\\
&= \left(\boldsymbol{s} - \frac{m\mathbf{1}_m\mathbf{1}_m^+}{m+n}\boldsymbol{s}\right)\otimes\mathbf{1}_n^{T+} + \mathbf{1}_m^{T+}\otimes\left(\boldsymbol{d} - \frac{n\mathbf{1}_n\mathbf{1}_n^+}{m+n}\boldsymbol{d}\right)\\
&= \left(\boldsymbol{s} - \Sigma\boldsymbol{s}\frac{\mathbf{1}_m}{m+n}\right)\otimes\mathbf{1}_n^{T+} + \mathbf{1}_m^{T+}\otimes\left(\boldsymbol{d} - \Sigma\boldsymbol{d}\frac{\mathbf{1}_n}{m+n}\right)\\
&= \boldsymbol{s}\otimes\mathbf{1}_n^{T+} + \mathbf{1}_m^{T+}\otimes\boldsymbol{d} - \Sigma\boldsymbol{s}\frac{\mathbf{1}_m}{m+n}\otimes\mathbf{1}_n^{T+} - \Sigma\boldsymbol{d}\,\mathbf{1}_m^{T+}\otimes\frac{\mathbf{1}_n}{m+n}\\
&= \boldsymbol{s}\otimes\mathbf{1}_n^{T+} + \mathbf{1}_m^{T+}\otimes\boldsymbol{d} - \Sigma\boldsymbol{s}\frac{\mathbf{1}_{mn}}{n(m+n)} - \Sigma\boldsymbol{d}\frac{\mathbf{1}_{mn}}{m(m+n)}\\
&= \boldsymbol{s}\otimes\mathbf{1}_n^{T+} + \mathbf{1}_m^{T+}\otimes\boldsymbol{d} - \Sigma\boldsymbol{s}\left(\frac{\mathbf{1}_{mn}}{n(m+n)} + \frac{\mathbf{1}_{mn}}{m(m+n)}\right)\end{aligned}$$

[1] It appears that row and column mean correction only necessarily commute after global mean correction, and either Lemma 2.2.29 or its interpretation is incorrect. Resolve!

$$= \boldsymbol{s} \otimes \mathbf{1}_n^{T+} \ + \ \mathbf{1}_m^{T+} \otimes \boldsymbol{d} \ - \ \Sigma \boldsymbol{s}\, \mathbf{1}_{mn}^{T+} = \boldsymbol{s} \otimes \mathbf{1}_n^{T+} \ + \ \mathbf{1}_m^{T+} \otimes \boldsymbol{d} \ - \ \Sigma \boldsymbol{s}\, \mathbf{1}_m^{T+} \otimes \mathbf{1}_n^{T+}$$

that is

$$\mathfrak{c}_\tau = \boldsymbol{s} \otimes \mathbf{1}_n^{T+} \ + \ \mathbf{1}_m^{T+} \otimes \boldsymbol{d} \ - \ \Sigma \boldsymbol{s}\, \mathbf{1}_m^{T+} \otimes \mathbf{1}_n^{T+} \tag{7.5}$$

so

$$\mathfrak{C}_\tau = \ \boldsymbol{s} \mathbf{1}_n^{+} \ + \ \mathbf{1}_m^{T+} \boldsymbol{d}^T \ - \ \Sigma \boldsymbol{s}\, \mathbf{1}_m^{T+} \mathbf{1}_n^{+} \tag{7.6}$$

Note that

1. $\Sigma \boldsymbol{d}$ could just as well have been used as $\Sigma \boldsymbol{s}$ in the previous equation,
2. this result for $\mathfrak{C}_\tau$ ($\mathfrak{c}_\tau$ in matrix form) is succinct and quite sufficient as we only compute it once.

So

$$\begin{aligned}
\mathfrak{c}_\tau^T \mathfrak{c}_\tau &= (\boldsymbol{s} \otimes \mathbf{1}_n^{T+} + \mathbf{1}_m^{T+} \otimes \boldsymbol{d} - \Sigma \boldsymbol{s}\, \mathbf{1}_m^{T+} \otimes \mathbf{1}_n^{T+})^T (\boldsymbol{s} \otimes \mathbf{1}_n^{T+} + \mathbf{1}_m^{T+} \otimes \boldsymbol{d} - \Sigma \boldsymbol{s}\, \mathbf{1}_m^{T+} \otimes \mathbf{1}_n^{T+}) \\
&= (\boldsymbol{s}^T \otimes \mathbf{1}_n^{+} + \mathbf{1}_m^{+} \otimes \boldsymbol{d}^T - \Sigma \boldsymbol{s}\, \mathbf{1}_m^{+} \otimes \mathbf{1}_n^{+})(\boldsymbol{s} \otimes \mathbf{1}_n^{T+} + \mathbf{1}_m^{T+} \otimes \boldsymbol{d} - \Sigma \boldsymbol{s}\, \mathbf{1}_m^{T+} \otimes \mathbf{1}_n^{T+}) \\
&= \|\boldsymbol{s}\|^2/n + \bar{s}\bar{d} - \Sigma \boldsymbol{s}\bar{s}/n + \bar{s}\bar{d} + \|\boldsymbol{d}\|^2/m - \Sigma \boldsymbol{s}\bar{d}/m - \Sigma \boldsymbol{s}\bar{s}/n - \Sigma \boldsymbol{s}\bar{d}/m + (\Sigma \boldsymbol{s})^2/(mn) \\
&= \|\boldsymbol{s}\|^2/n + \bar{s}\bar{d} - \bar{s}\bar{d} + \bar{s}\bar{d} + \|\boldsymbol{d}\|^2/m - \bar{s}\bar{d} - \bar{s}\bar{d} - \bar{s}\bar{d} + \bar{s}\bar{d} \\
&= \|\boldsymbol{s}\|^2/n - \bar{s}\bar{d} + \|\boldsymbol{d}\|^2/m
\end{aligned}$$

Now

$$\mathfrak{P}_\tau \overset{5.4}{=} \begin{bmatrix} \mathfrak{A}_\tau & 0 \\ 0 & \mathfrak{D}_\tau \end{bmatrix} + \begin{bmatrix} \mathfrak{b}_\tau \\ \mathfrak{c}_\tau \end{bmatrix} \begin{bmatrix} \mathfrak{b}_\tau \\ \mathfrak{c}_\tau \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{c}_\tau \\ \mathfrak{b}_\tau \end{bmatrix} \begin{bmatrix} \mathfrak{c}_\tau \\ \mathfrak{b}_\tau \end{bmatrix}^+$$

however $\mathfrak{P}_\tau$ is not computed; what is required is

$$\begin{aligned}
\mathfrak{P}_\tau \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} &= \left(\begin{bmatrix} \mathfrak{A}_\tau & 0 \\ 0 & \mathfrak{D}_\tau \end{bmatrix} + \begin{bmatrix} \mathfrak{b}_\tau \\ \mathfrak{c}_\tau \end{bmatrix} \begin{bmatrix} \mathfrak{b}_\tau \\ \mathfrak{c}_\tau \end{bmatrix}^+ - \begin{bmatrix} \mathfrak{c}_\tau \\ \mathfrak{b}_\tau \end{bmatrix} \begin{bmatrix} \mathfrak{c}_\tau \\ \mathfrak{b}_\tau \end{bmatrix}^+ \right) \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \\
&= \begin{bmatrix} \mathfrak{A}_\tau \mathfrak{x} \\ \mathfrak{D}_\tau \mathfrak{y} \end{bmatrix} + \left(\begin{bmatrix} \mathfrak{b}_\tau \\ \mathfrak{c}_\tau \end{bmatrix}^+ \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \right) \begin{bmatrix} \mathfrak{b}_\tau \\ \mathfrak{c}_\tau \end{bmatrix} - \left(\begin{bmatrix} \mathfrak{c}_\tau \\ \mathfrak{b}_\tau \end{bmatrix}^+ \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \right) \begin{bmatrix} \mathfrak{c}_\tau \\ \mathfrak{b}_\tau \end{bmatrix}
\end{aligned}$$

that is,

$$\mathfrak{P}_\tau \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} = \begin{bmatrix} \mathfrak{A}_\tau \mathfrak{x} \\ \mathfrak{D}_\tau \mathfrak{y} \end{bmatrix} + \left(\begin{bmatrix} \mathfrak{b}_\tau \\ \mathfrak{c}_\tau \end{bmatrix}^T \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \right) \begin{bmatrix} \mathfrak{b}_\tau \\ \mathfrak{c}_\tau \end{bmatrix}^{+T} - \left(\begin{bmatrix} \mathfrak{c}_\tau \\ \mathfrak{b}_\tau \end{bmatrix}^T \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \right) \begin{bmatrix} \mathfrak{c}_\tau \\ \mathfrak{b}_\tau \end{bmatrix}^{+T} \tag{7.7}$$

where $\mathfrak{A}_\tau \mathfrak{x}$ is computed by subtracting row and mean corrected $\mathfrak{x}$ from $\mathfrak{x}$, $\mathfrak{b}_\tau$ is computed using 7.3/7.4, $\mathfrak{c}_\tau$ is computed using 7.5/7.6, and $\mathfrak{D}_\tau \mathfrak{y}$ is computed by row mean correction and then column mean correction of $\mathfrak{y}$.

7.3 Appendix: A_τ^+

With A_τ given by Equation 7.1, we wish to show that

$$\begin{aligned}
A_\tau^+ &= \left[\frac{I_{m,n} \otimes \mathbf{1}_n^+}{\mathbf{1}_m^+ \otimes I_{n,m}}\right] && \text{(a)} \\
\text{where} & \\
I_{m,n} &= I_m - \frac{m\mathbf{1}_m\mathbf{1}_m^+}{m+n}\,, && \text{(b)} \\
I_{n,m} &= I_n - \frac{n\mathbf{1}_n\mathbf{1}_n^+}{m+n}, && \text{(c)}
\end{aligned} \tag{7.8}$$

7.3.1 Non-Constructive Verification

Set $A^\dagger = \left[\frac{I_{m,n} \otimes \mathbf{1}_n^+}{\mathbf{1}_m^+ \otimes I_{n,m}}\right]$ then $A_\tau A_\tau^\dagger = \left[\, I_m \otimes \mathbf{1}_n \middle| \mathbf{1}_m \otimes I_n \,\right] \left[\frac{I_{m,n} \otimes \mathbf{1}_n^+}{\mathbf{1}_m^+ \otimes I_{n,m}}\right]$, that is

$$\begin{aligned}
A_\tau A_\tau^\dagger &= I_{m,n} \otimes \mathbf{1}_n\mathbf{1}_n^+ + \mathbf{1}_m\mathbf{1}_m^+ \otimes I_{n,m} \\
&= \left(I_m - \frac{m\mathbf{1}_m\mathbf{1}_m^+}{m+n}\right) \otimes \mathbf{1}_n\mathbf{1}_n^+ + \mathbf{1}_m\mathbf{1}_m^+ \otimes \left(I_n - \frac{n\mathbf{1}_n\mathbf{1}_n^+}{m+n}\right) \\
&= I_m \otimes \mathbf{1}_n\mathbf{1}_n^+ - \tfrac{m}{m+n}\mathbf{1}_m\mathbf{1}_m^+ \otimes \mathbf{1}_n\mathbf{1}_n^+ + \mathbf{1}_m\mathbf{1}_m^+ \otimes I_n - \tfrac{n}{m+n}\mathbf{1}_m\mathbf{1}_m^+ \otimes \mathbf{1}_n\mathbf{1}_n^+
\end{aligned}$$

that is,

$$A_\tau A_\tau^\dagger = I_m \otimes \mathbf{1}_n\mathbf{1}_n^+ + \mathbf{1}_m\mathbf{1}_m^+ \otimes I_n - \mathbf{1}_m\mathbf{1}_m^+ \otimes \mathbf{1}_n\mathbf{1}_n^+ \quad \text{which is symmetric} \tag{7.9}$$

$$A_\tau^\dagger A_\tau = \left[\frac{I_{m,n} \otimes \mathbf{1}_n^+}{\mathbf{1}_m^+ \otimes I_{n,m}}\right] \left[\, I_m \otimes \mathbf{1}_n \middle| \mathbf{1}_m \otimes I_n \,\right] = \left[\begin{array}{c|c} I_{m,n} \otimes \mathbf{1}_n^+\mathbf{1}_n & I_{m,n}\mathbf{1}_m \otimes \mathbf{1}_n^+ \\ \hline \mathbf{1}_m^+ \otimes I_{n,m}\mathbf{1}_n & \mathbf{1}_m^+\mathbf{1}_m \otimes I_{n,m} \end{array}\right]$$

$$= \left[\begin{array}{c|c} I_{m,n} & I_{m,n}\mathbf{1}_m \otimes \mathbf{1}_n^+ \\ \hline \mathbf{1}_m^+ \otimes I_{n,m}\mathbf{1}_n & I_{n,m} \end{array}\right] = \left[\begin{array}{c|c} I_{m,n} & \left(I_m - \frac{m\mathbf{1}_m\mathbf{1}_m^+}{m+n}\right)\mathbf{1}_m \otimes \mathbf{1}_n^+ \\ \hline \mathbf{1}_m^+ \otimes \left(I_n - \frac{n\mathbf{1}_n\mathbf{1}_n^+}{m+n}\right)\mathbf{1}_n & I_{n,m} \end{array}\right]$$

$$= \left[\begin{array}{c|c} I_{m,n} & \frac{n}{m+n}\mathbf{1}_m \otimes \mathbf{1}_n^+ \\ \hline \frac{m}{m+n}\mathbf{1}_m^+ \otimes \mathbf{1}_n & I_{n,m} \end{array}\right] \quad \text{which is symmetric}$$

$$A_\tau A_\tau^\dagger A_\tau = (A_\tau A_\tau^\dagger)A_\tau \stackrel{7.9}{=} (I_m \otimes \mathbf{1}_n\mathbf{1}_n^+ + \mathbf{1}_m\mathbf{1}_m^+ \otimes I_n - \mathbf{1}_m\mathbf{1}_m^+ \otimes \mathbf{1}_n\mathbf{1}_n^+)\left[\, I_m \otimes \mathbf{1}_n \middle| \mathbf{1}_m \otimes I_n \,\right]$$

$$= \left[\, I_m \otimes \mathbf{1}_n + \mathbf{1}_m\mathbf{1}_m^+ \otimes \mathbf{1}_n - \mathbf{1}_m\mathbf{1}_m^+ \otimes \mathbf{1}_n \;\middle|\; \mathbf{1}_m \otimes \mathbf{1}_n\mathbf{1}_n^+ + \mathbf{1}_m \otimes I_n - \mathbf{1}_m \otimes \mathbf{1}_n\mathbf{1}_n^+ \,\right]$$

$$= \left[\, I_m \otimes \mathbf{1}_n \;\middle|\; \mathbf{1}_m \otimes I_n \,\right] = A_\tau$$

So $A_\tau^\dagger$ satisfies the four non-constructive conditions is therefor the generalized inverse of A_τ.

7.3.2 A Constructive Solution

A constructive method of finding the MPPI of A_τ relies on finding the inverse of $A_\tau A_\tau^T$:

$$A_\tau A_\tau^T = \left[\, I_m \otimes \mathbf{1}_n \middle| \mathbf{1}_m \otimes I_n \,\right] \left[\, I_m \otimes \mathbf{1}_n \middle| \mathbf{1}_m \otimes I_n \,\right]^T = \left[\, I_m \otimes \mathbf{1}_n \middle| \mathbf{1}_m \otimes I_n \,\right] \begin{bmatrix} I_m \otimes \mathbf{1}_n^T \\ \hline \mathbf{1}_m^T \otimes I_n \end{bmatrix}$$

$$= I_m \otimes \mathbf{1}_n \mathbf{1}_n^T + \mathbf{1}_m \mathbf{1}_m^T \otimes I_n$$

$$= (I_m - \mathbf{1}_m \mathbf{1}_m^+ + \mathbf{1}_m \mathbf{1}_m^+) \otimes \mathbf{1}_n \mathbf{1}_n^T + \mathbf{1}_m \mathbf{1}_m^T \otimes (I_n - \mathbf{1}_n \mathbf{1}_n^+ + \mathbf{1}_n \mathbf{1}_n^+)$$

$$= (I_m - \mathbf{1}_m \mathbf{1}_m^+) \otimes \mathbf{1}_n \mathbf{1}_n^T + \mathbf{1}_m \mathbf{1}_m^+ \otimes \mathbf{1}_n \mathbf{1}_n^T + \mathbf{1}_m \mathbf{1}_m^T \otimes \mathbf{1}_n \mathbf{1}_n^+ + \mathbf{1}_m \mathbf{1}_m^T \otimes (I_n - \mathbf{1}_n \mathbf{1}_n^+)$$

$$= \Big((I_m - \mathbf{1}_m \mathbf{1}_m^+) \otimes n \mathbf{1}_n \mathbf{1}_n^+\Big) + \Big((m+n) \mathbf{1}_m \mathbf{1}_m^+ \otimes \mathbf{1}_n \mathbf{1}_n^+\Big) + \Big(m \mathbf{1}_m \mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n \mathbf{1}_n^+)\Big)$$

and $A_\tau A_\tau^T$ has been decomposed into three mutually orthogonal components, so

$$(A_\tau A_\tau^T)^+ = (I_m - \mathbf{1}_m \mathbf{1}_m^+) \otimes \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^+ + \frac{1}{m+n} \mathbf{1}_m \mathbf{1}_m^+ \otimes \mathbf{1}_n \mathbf{1}_n^+ + \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n \mathbf{1}_n^+)$$

and $A_\tau^+ = A_\tau^T (A_\tau A_\tau^T)^+$

$$= \begin{bmatrix} n I_m \otimes \mathbf{1}_n^+ \\ \hline \mathbf{1}_m^+ \otimes m I_n \end{bmatrix} \Big((I_m - \mathbf{1}_m \mathbf{1}_m^+) \otimes \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^+ + \frac{1}{m+n} \mathbf{1}_m \mathbf{1}_m^+ \otimes \mathbf{1}_n \mathbf{1}_n^+ + \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n \mathbf{1}_n^+) \Big)$$

$$= \begin{bmatrix} (I_m - \mathbf{1}_m \mathbf{1}_m^+) \otimes \mathbf{1}_n^+ + \frac{n}{m+n} \mathbf{1}_m \mathbf{1}_m^+ \otimes \mathbf{1}_n^+ + \frac{n}{m} \mathbf{1}_m \mathbf{1}_m^+ \otimes \mathbf{0} \\ \hline \mathbf{0} \otimes \frac{m}{n} \mathbf{1}_n \mathbf{1}_n^+ + \frac{m}{m+n} \mathbf{1}_m^+ \otimes \mathbf{1}_n \mathbf{1}_n^+ + \mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n \mathbf{1}_n^+) \end{bmatrix}$$

$$= \begin{bmatrix} (I_m - \mathbf{1}_m \mathbf{1}_m^+) \otimes \mathbf{1}_n^+ + \frac{n}{m+n} \mathbf{1}_m \mathbf{1}_m^+ \otimes \mathbf{1}_n^+ \\ \hline \frac{m}{m+n} \mathbf{1}_m^+ \otimes \mathbf{1}_n \mathbf{1}_n^+ + \mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n \mathbf{1}_n^+) \end{bmatrix} = \begin{bmatrix} \Big(I_m - \frac{m}{m+n} \mathbf{1}_m \mathbf{1}_m^+ \Big) \otimes \mathbf{1}_n^+ \\ \hline \mathbf{1}_m^+ \otimes \Big((I_n - \frac{n}{m+n} \mathbf{1}_n \mathbf{1}_n^+ \Big) \end{bmatrix} = \begin{bmatrix} I_{m,n} \otimes \mathbf{1}_n^+ \\ \hline \mathbf{1}_m^+ \otimes I_{n,m} \end{bmatrix}$$

7.3.3 Commutativity

$$\mathfrak{D}_\tau = I_{mn} - A_\tau A_\tau^+ \overset{7.9}{=} I_{mn} - (I_m \otimes \mathbf{1}_n \mathbf{1}_n^+ + \mathbf{1}_m \mathbf{1}_m^+ \otimes I_n - \mathbf{1}_m \mathbf{1}_m^+ \otimes \mathbf{1}_n \mathbf{1}_n^+)$$

$$= I_m \otimes I_n - I_m \otimes \mathbf{1}_n \mathbf{1}_n^+ - \mathbf{1}_m \mathbf{1}_m^+ \otimes I_n + \mathbf{1}_m \mathbf{1}_m^+ \otimes \mathbf{1}_n \mathbf{1}_n^+ \tag{7.10}$$

$$= (I_m - \mathbf{1}_m \mathbf{1}_m^+) \otimes (I_n - \mathbf{1}_n \mathbf{1}_n^+) \ , \text{ so}$$

$$\mathfrak{D}_\tau = (I_m - \mathbf{1}_m \mathbf{1}_m^+) \otimes (I_n - \mathbf{1}_n \mathbf{1}_n^+) \ . \tag{7.11}$$

7.4 Exercises

Provide detail for subsection 7.3.1.

Chapter 8

Computations

In this chapter computer programming of the theory developed in Chapters 4 to 6 is used to check the theory.

8.1 The Invariant Form

Here we compute the invariant form for an LP and check some of its properties which were deduced in Chapter 4. We use function `invprob` contained in the document `solve-lp.rkt`. The function `invprob` takes as inputs the data A, $\boldsymbol{b}$ and $\boldsymbol{c}$; it is invoked by typing `(invprob A b c)`.

We test this function for the specific LP

$$\textit{maximize} \quad 3x_1 + 4x_2$$

$$\textit{subject to} \quad \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \geq \begin{bmatrix} -3 \\ -3/2 \\ 4 \end{bmatrix}$$

with the specific test module **invar-check.rkt**, we obtain the following output:

```
defining LP:
A = ((-1 -1) (-1 1) (2 1))
b = (-3 -3/2 4)
c = (3 4)
computing the invariant LP:
**********************************************************
checking that the computed invariant form fits the theory:
**********************************************************
𝔄 =((5/14 -3/14 -3/7) (-3/14 13/14 -1/7) (-3/7 -1/7 5/7))
𝔟 =(-15/28 -5/28 -5/14)
𝔠 =(-19/14 17/14 10/7)
Checking necessary condition for the problem to be feasible bounded,
i.e. that A^T 𝔠 − c = 0 (Lemma 4.2.3): A^T 𝔠 − c =(0 0)
Checking that 𝔄 − 𝔄^T = 0, i.e. that 𝔄 is symmetric (Lemma 4.2(a)):
𝔄 − 𝔄^T = ((0 0 0) (0 0 0) (0 0 0))
and 𝔄 − 𝔄^2 = 0, i.e. that 𝔄 is idempotent (Equation 4.2(b)):
𝔄 − 𝔄^2 =((0 0 0) (0 0 0) (0 0 0))
Checking that 𝔄𝔟 = 0 (Equation 4.2(f)): 𝔄𝔟 =(0 0 0)
```

Checking that $\mathfrak{A}\mathfrak{c} = \mathfrak{c}$ (Equation 4.2(h)): $\mathfrak{A}\mathfrak{c} - \mathfrak{c}$ =(0 0 0)
Checking that $\mathfrak{b}^T\mathfrak{c} = 0$,(Equation 4.2(j)): $\mathfrak{b}^T\mathfrak{c}$ =0

These results confirm some of the elementary theory for the invariant form of the specific original LP above.

8.2 Test Problems

Here, for specific linear programs, we check that repeated application of the averaging matrix V of Definition 3.1.12 produces a sequence of vectors which converges to a fixed-point, yielding a solution to the respective linear program. Note that this recursion is not efficient (just as the theory suggests) and for this reason the number of applications has been set to a high value (100).

The functions `LP` and `Lattice` are defined in the module `solve-lp.rkt`, have respectively the form `(LP A b c siter verbose variant out_port)` and `(Lattice A b c siter verbose variant out_port)`, and are run using the `problem-xx.rkt` modules.

A, b and $\boldsymbol{c}$ are as previously defined, `siter` is the number of iterations, and `verbose` is a logical variable which determines the amount of output from the function `LP`, and `variant` is a logical variable - if $\#f$ then $e = (111...)$ is used as the starting vector for the recursion, and if $\#t$ then $-e$ is used.

8.2.1 Problem 1

This is the problem which was introduced in Chapter 4.6.1. It is run with the document `problem − 1.rkt`. Output is in the document `problem − 1.out`; the solution to the invariant problem is very close to (0 2.5 0 5 0 1), and the solution to the original problem is very close to (1 2). Output is

```
********************************************************************
******************************** Problem 1 *************************
********************************************************************
********************************************************************
*********** FIXED-POINT METHOD: AVERAGING APPROACH *****************
********************************************************************
Original Data:
A = ((-1 -1) (-1 1) (2 1))
b = (-3.0 -1.5 4.0)
c = (3.0 4.0)
********************************************************************
Invariant Data:
𝔄=((5/14 -3/14 -3/7) (-3/14 13/14 -1/7) (-3/7 -1/7 5/7))
trace(𝔄) =2
𝔟=(-0.5357142857142858 -0.1785714285714285 -0.357142857142857)
𝔠=(-1.357142857142857 1.2142857142857142 1.4285714285714284)
𝔷=(1 1 1 1 1 1)
********************************************************************
Necessary but not sufficient check for a feasible bounded problem
(ref. 4.2.1):A⁺Ac − c = 0.0 (should be zero)
****************************** INVARIANT SYSTEM *******************
Checking that 𝔷_c is a fixed-point: 𝔷_c − 𝔓𝔷_c =1.313370261723731e-31
```

```
(should be zero)
Checking that the solution to the invariant problems is central
(should be non-negative):
(4.042322032660195e-13 2.50000000003347 -1.734151711119125e-11
    5.000000000018385 6.128431095930864e-12 1.0000000000122569)
The solution to the invariant problems is non-negative.
discriminant (d) = 0.9999999999924527 (>= form)
Checking for optimality of the computed invariant central solution:
    𝔠^T𝔵 + 𝔟^T𝔶 = -1.27675647831893e-15
(should be zero)
*******************************************************************
complementary slackness condition: 𝔵_c^T𝔶_c =7.216449660031206e-16
(should be zero)
objective value of invariant primal is 𝔠^T𝔵 = 3.0357142857296058
estimated objective value of original primal is = 11.00000000001532
*******************************************************************
*******************************************************************
***************************** ORIGINAL SYSTEM *********************
Checking for optimality of the original problems:
c^Tx - b^Ty =   -1.7763568394002505e-15   (should be zero)
***************************** primal ******************************
solution to original primal is
(0.9999999999830628 2.000000000016533)
objective value of original primal is 11.00000000001532
Checking (should be non-negative):
Ax - b =(4.04121180963557e-13 2.50000000003347 -1.7341683644644945e-11)
***************************** dual ********************************
solution to original dual is
(-5.000000000018385 -6.128431095930864e-12 -1.0000000000122569)
Checking for feasibility: (A^Ty = c)
A^Ty - c =  (0.0 0.0)
Objective value of original dual is 11.000000000015321
(should be the same as objective value of original primal)
*******************************************************************
*************** FIXED-POINT METHOD: LATTICE APPROACH ***************
*******************************************************************
Checking for optimality
𝔠^T𝔵c_t + 𝔟^Tζ_t =1.1102230246251565e-16(should be zero)
********
discriminant (d) = 0.9999999999999997 ==> (>= form).
solution to invariant problem is (2.775557561562894e-17 2.500000000000001
    2.211672250271514e-16 5.000000000000002 -1.6653345369377358e-16
    1.0000000000000002)
solution to original problem is
(0.9999999999999996 2.0000000000000004)
Ax - b =(0.0 2.500000000000001 -4.440892098500626e-16)
```

8.2.2 Problem 2

This is the same as Problem 1, however the redundant constraint $0x_1 + 0x_2 \geq 0$ has been added . It is run with the document `problem − 2.rkt`. Output is

```
*******************************************************************
******************************** Problem 2 ************************
*******************************************************************
*******************************************************************
```

*********** FIXED-POINT METHOD: AVERAGING APPROACH ****************
****
Original Data:
A = ((-1 -1) (-1 1) (2 1) (0 0))
b = (-3.0 -1.5 4.0 0.0)
c = (3.0 4.0)
****
Invariant Data:
$\mathfrak{A}$ =((5/14 -3/14 -3/7 0) (-3/14 13/14 -1/7 0) (-3/7 -1/7 5/7 0) (0 0 0 0))
$trace(\mathfrak{A})$ =2
$\mathfrak{b}$ =(-0.5357142857142858 -0.1785714285714285 -0.357142857142857 0.0)
$\mathfrak{c}$ =(-1.357142857142857 1.2142857142857142 1.4285714285714284 0)
$\mathfrak{z}$ =(1 1 1 1 1 1 1 1)
****
Necessary but not sufficient check for a feasible bounded problem
(ref. 4.2.1): $A^{+}Ac - c =$ 0.0 (should be zero)
****************************** INVARIANT SYSTEM *******************
Checking that $\mathfrak{z}_c$ is a fixed-point: $\mathfrak{z}_c - \mathfrak{P}\mathfrak{z}_c$ =1.313370261723731e-31
(should be zero)
Checking that the solution to the invariant problems is central
(should be non-negative):
(4.042322032660195e-13 2.50000000003347 -1.734151711119125e-11 -0.0
    5.000000000018385 6.128431095930864e-12 1.0000000000122569
    3.724793454370768)
The solution to the invariant problems is non-negative.
discriminant (d) = 0.9999999999924527 (>= form)
Checking for optimality of the computed invariant central solution:
    $\mathfrak{c}^T\mathfrak{x} + \mathfrak{b}^T\mathfrak{y}$ =-1.27675647831893e-15
(should be zero)
****
complementary slackness condition: $\mathfrak{x}_c^T\mathfrak{y}_c$ =7.216449660031206e-16
(should be zero)
objective value of invariant primal is $\mathfrak{c}^T\mathfrak{x}$ = 3.0357142857296058
estimated objective value of original primal is = 11.00000000001532
****
****
****************************** ORIGINAL SYSTEM ********************
Checking for optimality of the original problems:
$c^Tx - b^Ty =$ -1.7763568394002505e-15 (should be zero)
****************************** primal *****************************
solution to original primal is
(0.9999999999830628 2.000000000016533)
objective value of original primal is 11.00000000001532
Checking (should be non-negative):
$Ax - b$ =(4.04121180963557e-13 2.50000000003347 -1.7341683644644945e-11 -0.0)
****************************** dual *******************************
solution to original dual is
(-5.000000000018385 -6.128431095930864e-12 -1.0000000000122569
    -3.724793454370768)
Checking for feasibility: ($A^Ty = c$)
$A^Ty - c =$ (0.0 0.0)
Objective value of original dual is 11.000000000015321
(should be the same as objective value of original primal)
****
************** FIXED-POINT METHOD: LATTICE APPROACH ***************
****
Checking for optimality
$\mathfrak{c}^T\mathfrak{x}c_t + \mathfrak{b}^T\zeta_t$ =1.1102230246251565e-16(should be zero)

```
********
discriminant (d) = 0.9999999999999997 ==> (>= form).
solution to invariant problem is (2.775557561562894e-17 2.500000000000001
    2.211672250271514e-16 0.0 5.000000000000002 -1.6653345369377358e-16
    1.0000000000000002 3.483566542109277)
solution to original problem is
(0.9999999999999996 2.0000000000000004)
Ax - b =(0.0 2.500000000000001 -4.440892098500626e-16 0.0)
```

In both cases the solution to the original problem is very close to (1 2), which is the same as Problem 1. Note however that for Problem 2 the averaging and lattice solutions to the fixed-point problem problem differ and this leads us to the construction of a zero component-pair solution as follows:

We see that there is the fixed-point

$$\boldsymbol{p}_1^T = \begin{pmatrix} 0 & 2.5 & 0 & 0 & 5 & 0 & 1 & 3.7247934543707664 \end{pmatrix}$$

from the averaging method, and the fixed-point

$$\boldsymbol{p}_2^T = \begin{pmatrix} 0 & 2.5 & 0 & 0 & 5 & 0 & 1 & 3.483566542109277 \end{pmatrix}$$

from the lattice method.

Taking a particular linear combination of $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ we obtain the non-central fixed-point

$$\boldsymbol{p}_3^T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and, subtracting a multiple of $\boldsymbol{p}_3$ from $\boldsymbol{p}_1$, we obtain the fixed-point

$$\boldsymbol{p}_4^T = \begin{pmatrix} 0 & 2.5 & 0 & 0 & 5 & 0 & 1 & 0 \end{pmatrix}$$

The non-negative fixed-point $\boldsymbol{p}_4$ has the zero component-pair $(\boldsymbol{p}_{4,4}, p_{4,8}) = (0, 0)$, so zero component-pairs do exist, but we haven't managed to find one directly during the solution of a real life problem. Vector $\boldsymbol{p}_3$ has three zero component-pairs, however it doesn't yield a solution as it isn't central.

8.2.3 Problem 3

The third test linear program is from Ecker & Kupferschmid [10, Exercise 2.3].

Minimize $x_1 - x_2$ subject to

$$x_1 + 2x_2 \geq 4\ , \quad x_2 \leq 4\ , \quad 3x_1 - 2x_2 \leq 0\ , \quad x_2 \geq 0$$

Refer to Figure 8.1 for a graphical solution.

The program is run using the module `problem-3.rkt`. Output is:

```
*****************************************************************
****************************** Problem 3 ************************
*****************************************************************
```

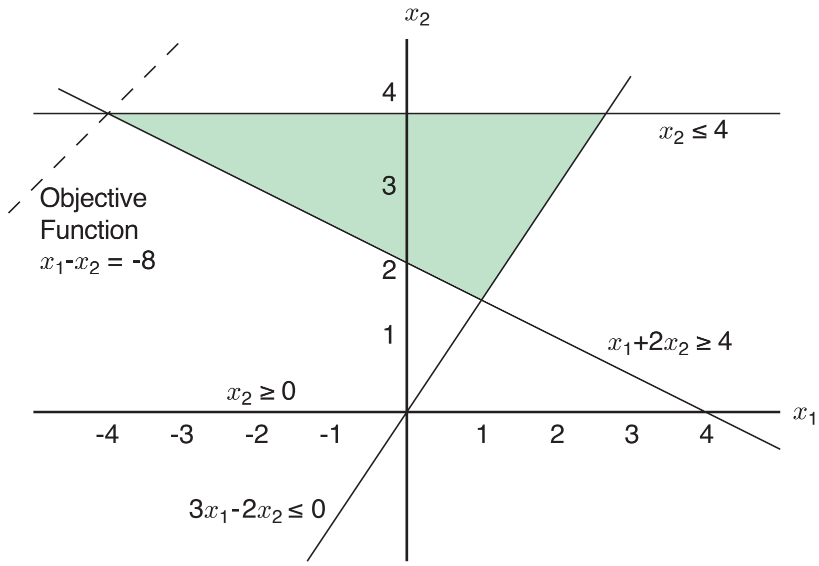


Figure 8.1: **A Two-Variable Linear Programming Problem**

```
*****************************************************************
*********** FIXED-POINT METHOD: AVERAGING APPROACH ***************
*****************************************************************
Original Data:
A = ((1 2) (0 -1) (-3 2) (0 1))
b = (4.0 -4.0 0.0 0.0)
c = (-1.0 1.0)
*****************************************************************
Invariant Data:
```

$\mathfrak{A}$ =((11/14 -2/7 -1/14 2/7) (-2/7 5/42 -2/21 -5/42) (-1/14 -2/21 41/42 2/21)
(2/7 -5/42 2/21 5/42))

$trace(\mathfrak{A})$ =2

$\mathfrak{b}$ =(-0.2857142857142857 -2.380952380952381 -0.09523809523809523
-1.619047619047619)

$\mathfrak{c}$ =(0.07142857142857142 -0.07142857142857142 0.35714285714285715
0.07142857142857142)

$\mathfrak{z}$ =(1 1 1 1 1 1 1 1)

```
*****************************************************************
Necessary but not sufficient check for a feasible bounded problem
```

(ref. 4.2.1): $A^+Ac - c =$ 0.0 (should be zero)

```
****************************** INVARIANT SYSTEM *******************
```

Checking that $\mathfrak{z}_c$ is a fixed-point: $\mathfrak{z}_c - \mathfrak{P}\mathfrak{z}_c$ =1.327337088180114e-29
(should be zero)

```
Checking that the solution to the invariant problems is central
(should be non-negative):
(5.551115123125783e-16 4.440892098500626e-16 19.999999999999993
    3.9999999999999996 0.9999999999999998 2.999999999999999 -1.6653345369377348e
    -16 9.71445146547012e-17)
The solution to the invariant problems is non-negative.
discriminant (d) = 0.9999999999999998 (>= form)
Checking for optimality of the computed invariant central solution:
```

$\mathfrak{c}^T\mathfrak{x} + \mathfrak{b}^T\mathfrak{y}$ =-1.41421266232014e-16
(should be zero)

```
*****************************************************************
```

complementary slackness condition: $\mathfrak{x}_c^T\mathfrak{y}_c$ =-1.054711873393898e-15

```
(should be zero)
objective value of invariant primal is 𝔠^T𝔵 = 7.428571428571426
estimated objective value of original primal is = 7.999999999999997
*****************************************************************
*****************************************************************
***************************** ORIGINAL SYSTEM ********************
Checking for optimality of the original problems:
c^T x - b^T y =  8.881784197001252e-16  (should be zero)
***************************** primal *****************************
solution to original primal is
(-3.9999999999999987 3.999999999999999)
objective value of original primal is 7.999999999999998
Checking (should be non-negative):
Ax - b =(-4.440892098500626e-16 8.881784197001252e-16 19.999999999999993
   3.999999999999999)
****************************** dual ******************************
solution to original dual is
(-0.9999999999999998 -2.999999999999999 1.6653345369377348e-16
   -6.938893903907228e-17)
Checking for feasibility: (A^T y = c)
A^T y - c =  (-2.220446049250313e-16 -2.220446049250313e-16)
Objective value of original dual is 7.999999999999997
(should be the same as objective value of original primal)
*****************************************************************
************** FIXED-POINT METHOD: LATTICE APPROACH **************
*****************************************************************
Checking for optimality
𝔠^T 𝔵c_t + 𝔟^T ζ_t =-4.4499550259345506e-23(should be zero)
********
discriminant (d) = 1.0 ==> (>= form).
solution to invariant problem is (8.542985108748319e-23 -5.66894392301607e-23
   20.0 4.0 1.0 3.0 -2.549062647602563e-24 3.3904929888762546e-23)
solution to original problem is
(-4.0 4.0)
Ax - b =(0.0 0.0 20.0 4.0)
```

Again a correct solution has been found.

8.2.4 Problem 4

This is a feasible unbounded problem; it is run with the document `problem − 4.rkt`. Output is in the document `problem − 4.out`.

```
*****************************************************************
*********************** Problem 4 ********************************
*****************************************************************
*****************************************************************
********** FIXED-POINT METHOD: AVERAGING APPROACH ****************
*****************************************************************
Original Data:
A = ((1 0 0) (-1 0 0) (0 1 0) (0 -1 0))
b = (-1.0 -1.0 -1.0 -1.0)
c = (1.0 1.0 1.0)
*****************************************************************
Invariant Data:
𝔄 =((1/2 -1/2 0 0) (-1/2 1/2 0 0) (0 0 1/2 -1/2) (0 0 -1/2 1/2))
```

```
trace(𝔄) =2
𝔟 =(-1.0 -1.0 -1.0 -1.0)
𝔠 =(0.5 -0.5 0.5 -0.5)
𝔷 =(1 1 1 1 1 1 1 1)
*********************************************************************
Necessary but not sufficient check for a feasible bounded problem
(ref. 4.2.1):A⁺Ac − c =  1.0 (should be zero)
** warning: problem is infeasible or unbounded; stopping **
*********************************************************************
*************** FIXED-POINT METHOD: LATTICE APPROACH ***************
*********************************************************************
**** warning: If feasible the problem is unbounded; stopping ****
```

. Unboundedness is detected.

8.2.5 Problem 5

This is an infeasible problem; it is run with the document `problem − 5.rkt`. Output is in the document `problem − 5.out`.

```
*********************************************************************
************************ Problem 5 *********************************
*********************************************************************
*********************************************************************
*********** FIXED-POINT METHOD: AVERAGING APPROACH *****************
*********************************************************************
Original Data:
A = ((1 0 0) (0 1 0) (0 0 1) (0 0 -1))
b = (1.0 1.0 1.0 0.0)
c = (1.0 1.0 1.0)
*********************************************************************
Invariant Data:
𝔄 =((1 0 0 0) (0 1 0 0) (0 0 1/2 -1/2) (0 0 -1/2 1/2))
trace(𝔄) =3
𝔟 =(0 0 0.5 0.5)
𝔠 =(1.0 1.0 0.5 -0.5)
𝔷 =(1 1 1 1 1 1 1 1)
*********************************************************************
Necessary but not sufficient check for a feasible bounded problem
(ref. 4.2.1):A⁺Ac − c =  0.0 (should be zero)
****************************** INVARIANT SYSTEM *******************
Checking that 𝔷c is a fixed-point:  𝔷c − 𝔓𝔷c =5.940423917354408e-32
(should be zero)
Checking that the solution to the invariant problems is central
(should be non-negative):
(1.1102230246251565e-16 1.1102230246251565e-16 -0.9999999999999999
    -1.1102230246251565e-16 -1.0 -1.0 -1.1102230246251565e-16
    0.9999999999999999)
The solution to the invariant problems is not non-negative.
The solution to the invariant problems is invalid.
Stopping.
*********************************************************************
*************** FIXED-POINT METHOD: LATTICE APPROACH ***************
*********************************************************************
Checking for optimality
𝔠ᵀ𝔵cₜ + 𝔟ᵀζₜ =0.0(should be zero)
```

```
********
discriminant (d) = -1.0 ==> (<= form).
solution to invariant problem is (0.0 0.0 0.0 -1.0 -1.0 -1.0 -1.0 0.0)
solution to original problem is
(1.0 1.0 1.0)
```
$Ax - b =$ (0.0 0.0 0.0 -1.0)

The “solution” is invalid.

8.3 Conclusion

The theory for both the unitary and lattice approaches works in practice.

Chapter 9

Convex Optimization

The aim is to develop conditions for solving convex optimization problems, by the strategy of linearization.

9.1 The Problem

We consider the general problem

$$\begin{array}{lll} \textit{maximize} & f(\boldsymbol{x}), \boldsymbol{x} \in \Re^n & \text{(a)} \\ \textit{subject to} & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0} \in \Re^m. & \text{(b)} \end{array} \tag{9.1}$$

Here f is a scalar function of the unknown n-dimensional vector $\boldsymbol{x}$, and $\boldsymbol{g}$ is an m-dimensional vector function of $\boldsymbol{x}$. The objective function f and the inequality constraint $\boldsymbol{g}$ are required to be continuously differentiable concave functions.[1]

9.2 Linearization of the Problem

Given a column vector x and a scalar function, y, we introduce the vector derivative or gradient of y as the column vector function $y_x = \partial y/\partial \boldsymbol{x} = \nabla_x y$ defined by

$$(y_x)_i = \partial y/\partial x_i \ . \tag{9.2}$$

For the function $h : \Re^n \to \Re$, define $\hat{h} = h(\hat{\boldsymbol{x}})$ and $\hat{\nabla} h \ = h_{\boldsymbol{x}}|\hat{\boldsymbol{x}}$, i.e. $\hat{\nabla} h$ is the n dimensional vector whose j^{th} term is $\partial h(x)/\partial x_j|\,\hat{x}$. Note that $\hat{\nabla}(-h) = (-h)_{\boldsymbol{x}}|\hat{\boldsymbol{x}} = -\ h_{\boldsymbol{x}}|\hat{\boldsymbol{x}} = -\hat{\nabla} h$.

A differentiable function h on a convex domain is convex if and only if

$$h(\boldsymbol{x}) \geq h(\hat{\boldsymbol{x}}) + \hat{\nabla} h^T(\boldsymbol{x} - \hat{\boldsymbol{x}}),$$

while a differentiable function h on a convex domain is concave if and only if

$$(-h)(\boldsymbol{x}) \geq (-h)(\hat{\boldsymbol{x}}) + \hat{\nabla}(-h)^T(\boldsymbol{x} - \hat{\boldsymbol{x}}) \Leftrightarrow -h(\boldsymbol{x}) \geq -h(\hat{\boldsymbol{x}}) - \hat{\nabla} h^{\,T}(\boldsymbol{x} - \hat{\boldsymbol{x}})$$

[1] The usual formulation is involves minimizing $f(\boldsymbol{x})$ subject to $\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$. where f and g are convex functions.

$$\Leftrightarrow h(\boldsymbol{x}) \leq h(\hat{\boldsymbol{x}}) + \hat{\triangledown} h^T(\boldsymbol{x} - \hat{\boldsymbol{x}}) \Leftrightarrow h(\boldsymbol{x}) \leq (h(\hat{\boldsymbol{x}}) - \hat{\triangledown} h^T \hat{\boldsymbol{x}}) + \hat{\triangledown} h^T \boldsymbol{x},$$

so

Lemma 9.2.1 A differentiable function h on a convex domain is concave $\Leftrightarrow h(\boldsymbol{x}) \leq h(\hat{\boldsymbol{x}}) + \hat{\triangledown} h^T(\boldsymbol{x} - \hat{\boldsymbol{x}}) \Leftrightarrow h(\boldsymbol{x}) \leq (h(\hat{\boldsymbol{x}}) - \hat{\triangledown} h^T \hat{\boldsymbol{x}}) + \hat{\triangledown} h^T \boldsymbol{x}$

and we can approximate the objective function f by $\hat{f} = f(\hat{\boldsymbol{x}}) + \hat{\triangledown} f^T(\boldsymbol{x} - \hat{\boldsymbol{x}})$ with a "constant" component, $f(\hat{\boldsymbol{x}}) - \hat{\triangledown} f^T \hat{\boldsymbol{x}}$, plus a linear component $\hat{\triangledown} f^T \boldsymbol{x}$ whereby $\hat{\triangledown} f$ becomes our "$\boldsymbol{c}$".

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are both column vectors, then we form the *Jacobian* matrix $\boldsymbol{y}_{\boldsymbol{x}}$ defined by

$$(\boldsymbol{y}_{\boldsymbol{x}})_{ij} = \partial y_i / \partial x_j. \tag{9.3}$$

Given an estimate, $\hat{\boldsymbol{x}}$, of the solution, write the Jacobian $\hat{G} = \boldsymbol{g}_{\boldsymbol{x}} | \hat{\boldsymbol{x}}$ - that is $\hat{G}$ is the $m \times n$ Jacobian matrix whose i^{th} row is $\partial g_i(\boldsymbol{x}) / \partial \boldsymbol{x}^T = \hat{\triangledown} g_i^T$ evaluated at $\hat{\boldsymbol{x}}$ - and approximating $\boldsymbol{g}$ by

$$\hat{\boldsymbol{g}}(\boldsymbol{x}) = \boldsymbol{g}(\hat{\boldsymbol{x}}) + \hat{G}(\boldsymbol{x} - \hat{\boldsymbol{x}}), \tag{9.4}$$

so

$$\hat{\boldsymbol{g}}(\boldsymbol{x}) \geq \boldsymbol{g}(\hat{\boldsymbol{x}}) \quad \Leftrightarrow \quad \hat{G}(\boldsymbol{x} - \hat{\boldsymbol{x}}) \geq \boldsymbol{0}$$

we see that the affine plane described by $\hat{\triangledown} g_i^T(\boldsymbol{x} - \hat{\boldsymbol{x}}) = 0$ osculates the (in general) curved surface $g_i(\boldsymbol{x}) = g_i(\hat{\boldsymbol{x}})$ since for concave $\boldsymbol{g}, \boldsymbol{g}(\boldsymbol{x}) \overset{L\,9.2.1}{\leq} \boldsymbol{g}(\hat{\boldsymbol{x}}) + \hat{G}(\boldsymbol{x} - \hat{\boldsymbol{x}}) \overset{9.4}{=} \hat{\boldsymbol{g}}(\boldsymbol{x})$, so $\hat{\boldsymbol{g}}(\boldsymbol{x}) \geq \boldsymbol{g}(\boldsymbol{x})$, and $\hat{\boldsymbol{g}}(\boldsymbol{x}) = \boldsymbol{g}(\hat{\boldsymbol{x}}) \overset{9.4}{\Leftrightarrow} \hat{G}(\boldsymbol{x} - \hat{\boldsymbol{x}}) = \boldsymbol{0} \quad \Leftrightarrow \quad \hat{\triangledown} g_i^T(\boldsymbol{x} - \hat{\boldsymbol{x}}) = 0 \;\; \forall i$.
Maximize $\hat{\boldsymbol{c}}^T \boldsymbol{x}$ *subject to* $\hat{A}\boldsymbol{x} \geq \hat{\boldsymbol{b}}$
where

$$\begin{aligned} \hat{A} &= \hat{G} && \text{(a)} \\ \hat{\boldsymbol{b}} &= \hat{G}\hat{\boldsymbol{x}} - \hat{\boldsymbol{g}} && \text{(b)} \\ \hat{\boldsymbol{c}} &= \hat{\triangledown} f && \text{(c)} \end{aligned} \tag{9.5}$$

is a feasible linearization of 9.1.

We also require 9.1(b) as there is no guarantee that the solution set for the problem has not been increased by linearization; to ensure this we may periodically apply Newton's method using the current binding constraints. (refer to Figure 9.5).

In summary, since for any value of $\hat{x}$ the feasible set of the convex problem is a subset of the feasible set of the linearized problem, if the linearized problem has a solution which is in the feasible set of the convex problem then this solution is also a solution for the convex problem. Figure 9.1 is an example of this possibility.

9.3 Qualitative Analysis

We consider the possibilities that convex problem has a non-empty solution set and its linearized form has a solution set which is

1. empty for some $\hat{x}$ - in this case so too is the feasible set of the convex problem.
2. precisely the same, and bounded; refer to Figure 9.1
3. a proper superset thereof, and bounded; refer to Figures 9.2 and 9.4
4. unbounded due to the feasible set being unbounded; refer to Figure 9.3

The developed fixed-point algorithm should detect the infeasibility of the first case, for case two it should converge to a correct solution, for cases three and four it should converge to an incorrect solution however the binding constraints having been identified Newton's method may be employed.

Considering these possibilities, cases 3 and 4 are problematic and will be given particular attention.

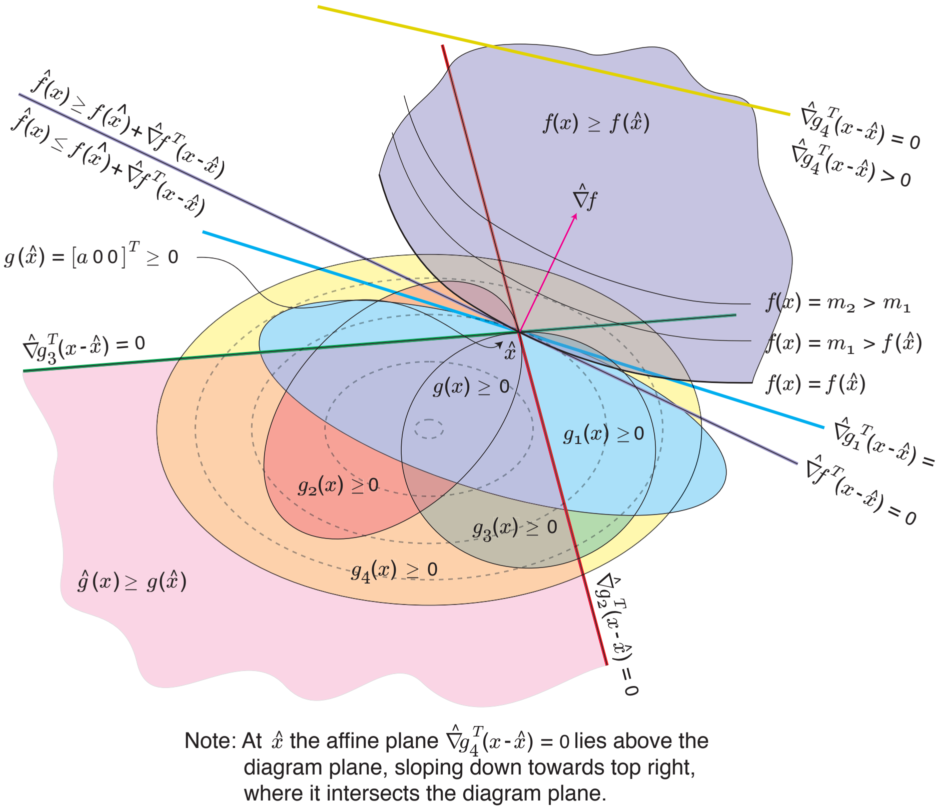


Figure 9.1: **The Same Solution Set Case: 2D Cut Through 3D Solution Space**

9.4 Solution Conditions

Analogous to Definition 4.2.1, at the solution point we require

$$\hat{G}^{+}\hat{G}\hat{\nabla} f = \hat{\nabla} f \quad . \tag{9.6}$$

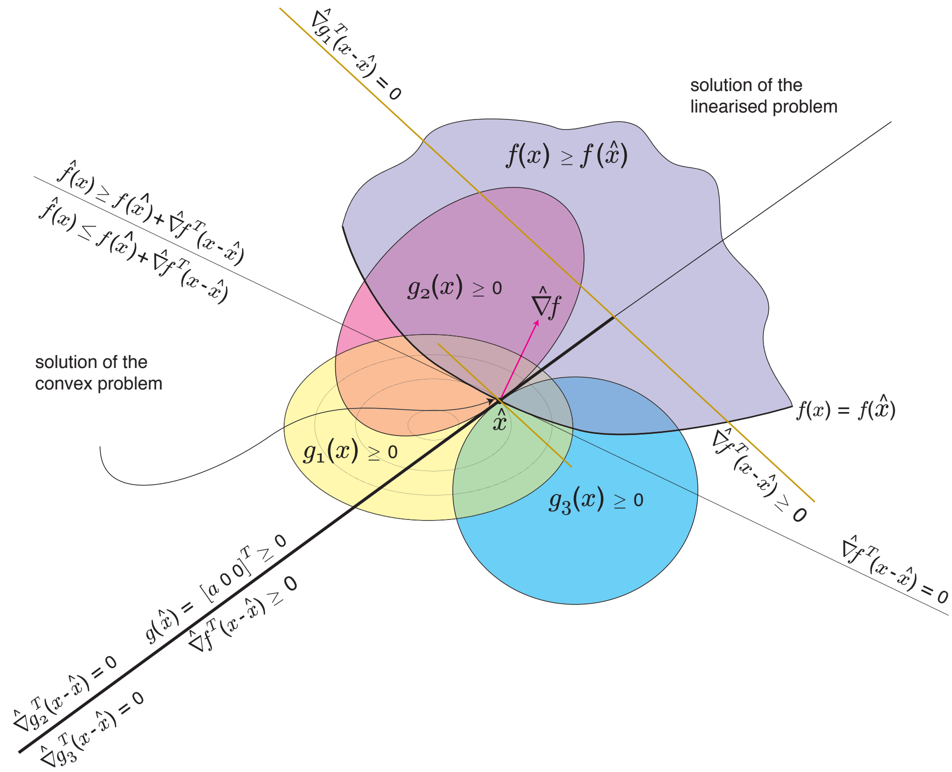


Figure 9.2: **The Unusual Case with Bounded Linearization**

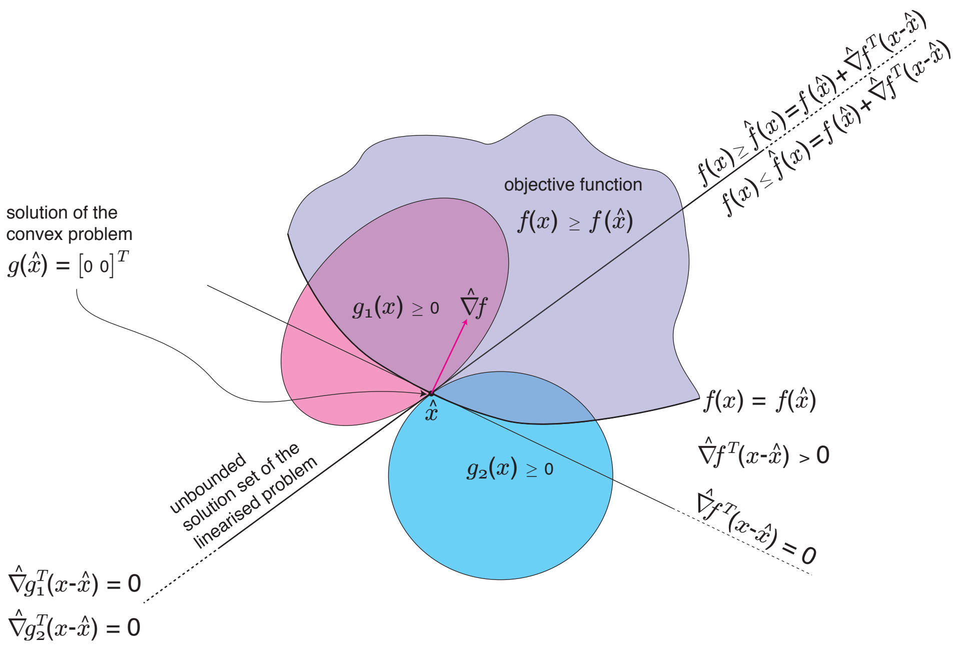


Figure 9.3: **The Unusual Case with Unbounded Linearization**

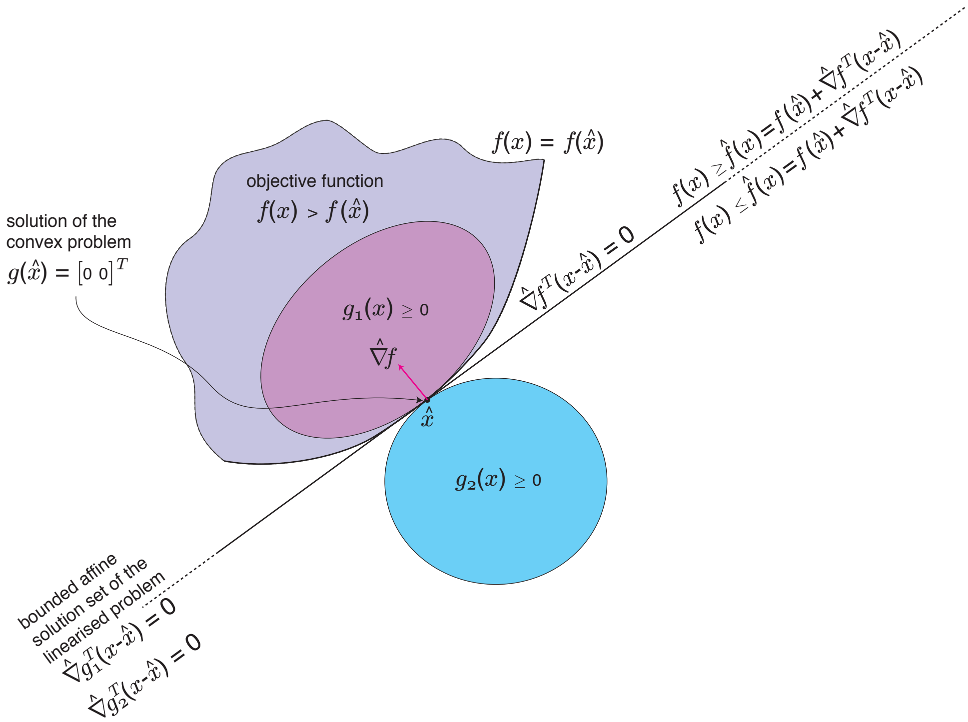


Figure 9.4: **The Highly Unusual Case**

As with LP's case we can construct invariant primal and dual LP's, and fixed-point problem

$$\textit{find} \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \textit{satisfying} \ \hat{\mathfrak{P}} \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} = \begin{bmatrix} \mathfrak{x} \\ \mathfrak{y} \end{bmatrix} \geq 0$$

Now

$$\hat{\mathfrak{A}} \overset{4.1\,a}{=} \hat{A}\hat{A}^{+} \overset{9.5\,a}{=} \hat{G}\hat{G}^{+} \tag{9.7}$$

$$\hat{\mathfrak{b}} \overset{4.1\,c}{=} (I - \hat{\mathfrak{A}})\hat{\boldsymbol{b}} \overset{9.5\,b}{=} (I - \hat{\mathfrak{A}})(\hat{G}\hat{\boldsymbol{x}} - \hat{\boldsymbol{g}}) \overset{9.7}{=} (I - \hat{G}\hat{G}^{+})(\hat{G}\hat{\boldsymbol{x}} - \hat{\boldsymbol{g}}) = -(I - \hat{G}\hat{G}^{+})\hat{\boldsymbol{g}},$$

$$\hat{\mathfrak{c}} \overset{4.1\,d}{=} \hat{A}^{T+}\hat{\boldsymbol{c}} \overset{9.5\,a}{=} \hat{G}^{T+}\hat{\boldsymbol{c}} \overset{9.5\,c}{=} \hat{G}^{T+}\hat{\nabla} f$$

$$\hat{\mathfrak{D}} \overset{4.1\,b}{=} I - \hat{A}\hat{A}^{+} \overset{9.5\,a}{=} I - \hat{G}\hat{G}^{+}$$

Summarizing, we have

$$\begin{aligned} \hat{\mathfrak{A}} &= \hat{G}\hat{G}^{+} && \text{(a)} \\ \hat{\mathfrak{b}} &= -(I - \hat{G}\hat{G}^{+})\hat{\boldsymbol{g}} && \text{(b)} \\ \hat{\mathfrak{c}} &= \hat{G}^{T+}\hat{\nabla} f && \text{(c)} \\ \hat{\mathfrak{D}} &= I - \hat{G}\hat{G}^{+} && \text{(d)} \end{aligned} \tag{9.8}$$

Collecting relevant equations and properties, including the conditions of Equation 5.1, with the additional $\hat{g} \geq 0$ to override the increase in the solution set due to linearization.

$$\begin{array}{lll} \hat{\mathfrak{x}} = \hat{\mathfrak{A}}\hat{\mathfrak{x}} - \hat{\mathfrak{b}} & \geq & 0 \quad \text{(a)} \\ \hat{\mathfrak{y}} = \hat{\mathfrak{D}}\hat{\mathfrak{y}} - \hat{\mathfrak{c}} & \geq & 0 \quad \text{(b)} \\ \hat{\mathfrak{c}}^T\hat{\mathfrak{x}} + \hat{\mathfrak{b}}^T\hat{\mathfrak{y}} & = & 0 \quad \text{(c)} \\ \hat{g} & \geq & 0 \quad \text{(d)} \\ \hat{\mathfrak{x}} & = & \hat{A}\hat{x} - \hat{b} \quad \text{(e)} \\ \hat{x} & = & \hat{A}^+(\hat{\mathfrak{x}} + \hat{b}) \quad \text{(f)} \end{array} \tag{9.9}$$

Note: $(e) \wedge (f) \Rightarrow \hat{\mathfrak{x}} = \hat{\mathfrak{A}}\hat{\mathfrak{x}} - \hat{\mathfrak{b}}$.

Note that these conditions are

1. necessary and sufficient if the problem is one of convex optimization.
2. operational since $\hat{\mathfrak{x}}$ and $\hat{\mathfrak{y}}$ can be constructed using the fixed-point approach,
3. free of any constraint qualification such as quasi-normality,
4. distinct from the Karush-Kuhn-Tucker conditions as condition 9.9(c) is scalar in nature while the corresponding KKT stationarity condition is a vector condition.

9.5 The Algorithm

We adopt the following recursive scheme for the usual case:

1. Choose an initial $\hat{\boldsymbol{x}}$ (try $\hat{\boldsymbol{x}} = \mathbf{0}$) as a solution to the primal,
2. Compute $\hat{A}$, $\hat{b}$ and $\hat{c}$, using Equation 9.5.
3. Compute $\hat{\mathfrak{A}}$, $\hat{\mathfrak{b}}$, $\hat{\mathfrak{c}}$ and $\hat{\mathfrak{D}}$ using Equation 9.8. (consistent with Equation 4.1); optionally, $\hat{\mathfrak{b}}$ and $\hat{\mathfrak{c}}$ may be normalized. In detail, compute:
   (a) $\hat{G}^+$ either by direct inversion or by updating using the iterative method of Ben-Israel and Cohen: $\hat{G}^\dagger_{i+1} = 2\hat{G}^\dagger_i - \hat{G}^\dagger_i\hat{G}\hat{G}^\dagger_i$, where $\hat{G}^\dagger_0$ is the MPPI of the previous $\hat{G}$,
   (b) $\hat{\mathfrak{A}} = \hat{G}\hat{G}^+$, $\hat{\mathfrak{b}} = -(I - \hat{G}\hat{G}^+)\hat{\boldsymbol{g}}$, $\hat{\mathfrak{c}} = \hat{G}^{T+}\hat{\nabla} f$ and $\hat{\mathfrak{D}} = I - \hat{\mathfrak{A}}$,
4. Compute $\hat{\mathfrak{P}}$ from $\hat{\mathfrak{A}}$, $\hat{\mathfrak{b}}$ and $\hat{\mathfrak{c}}$ using Equation 5.4 .
5. Fixed-point estimate: initial $\hat{\mathfrak{z}} = \hat{\mathfrak{P}}\mathbf{1}$; update with
   (a) $\hat{\mathfrak{z}}_{new} = \hat{\mathfrak{P}}|\hat{\mathfrak{z}}_{old} - \mathfrak{S}\hat{\mathfrak{z}}_{old}|$, or

   (b) when binding pattern becomes stable
      (i) perform affine regression (refer to Chapter 3.2.2), or
      (ii) apply Newton's Method using the binding constraints.

6. Centralize the fixed-point estimate $\hat{\mathfrak{z}} = \begin{bmatrix} \hat{\mathfrak{x}} \\ \hat{\mathfrak{y}} \end{bmatrix}$, by computing $\begin{bmatrix} \hat{\mathfrak{A}}\hat{\mathfrak{x}} - \hat{\mathfrak{b}} \\ \hat{\mathfrak{D}}\hat{\mathfrak{y}} - \hat{\mathfrak{c}} \end{bmatrix}$.

7. Consistent with Equation 4.9 (b), compute $\hat{\boldsymbol{x}} = \hat{A}^{+}(\hat{\mathfrak{x}} + \hat{b}) = \hat{G}^{+}(\hat{\mathfrak{x}} - (I - \hat{G}\hat{G}^{+})\hat{g}) = \hat{G}^{+}\hat{\mathfrak{x}}$, as an estimated solution for the original problem.

8. Stop if conditions 9.9 are met, otherwise go to step 2.

Note:

1. The Ben Israel and Cohen method might be generalized to

$$\hat{G}^{\dagger}_{i+1} = (m\hat{G}^{\dagger}_{i} - n\hat{G}^{\dagger}_{i}\hat{G}\hat{G}^{\dagger}_{i})/(m-n)$$

   also note that the method requires $\hat{G}^{\dagger}_{i}\hat{G}$ to be symmetric - a condition which we cannot satisfy, so we might try adding

$$\hat{G}^{\dagger}_{i+1} = (m\hat{G}^{\dagger}_{i}(\hat{G}\hat{G}^{\dagger}_{i})^{T} + n(\hat{G}^{\dagger}_{i}\hat{G})^{T}\hat{G}^{\dagger}_{i})/(m+n)$$

   or merging to obtain

$$\hat{G}^{\dagger}_{i+1} = \frac{m_1\hat{G}^{\dagger}_{i} + m_2\hat{G}^{\dagger}_{i}\hat{G}\hat{G}^{\dagger}_{i} + m_3\hat{G}^{\dagger}_{i}(\hat{G}\hat{G}^{\dagger}_{i})^{T} + m_4(\hat{G}^{\dagger}_{i}\hat{G})^{T}\hat{G}^{\dagger}_{i}}{m_1 + m_2 + m_3 + m_4}$$

   Proposed constraints are $m_1 \geq 2, m_2 < 0, m_1 + m_2 = 1, m_3 = m_4$.

2. If the scheme for the non-aberrant case fails to converge then set $\boldsymbol{c} = 0$ to determine feasible points, determine the affine space in which the feasible points lie, and recast the problem within this affine space.

9.6 Application

Here we apply the theory for the non-linear case to solving two problems. The algorithm works well for the first case but fails for the second, suggesting that the approach requires further refinement involving Step 5b above.

9.6.1 Non-Linear Problem I

Maximize $x + y + z$ subject to $x^2 + y^2 + z^2 \leq 1$, $x \geq -0.5$, $x \leq 0.5$.

With reference to Figure 9.5, when the objective function $x + y + z$ is restricted to the value 1.7247 we get a hyperplane of solutions some of which are depicted by the triangular blue area which osculates the

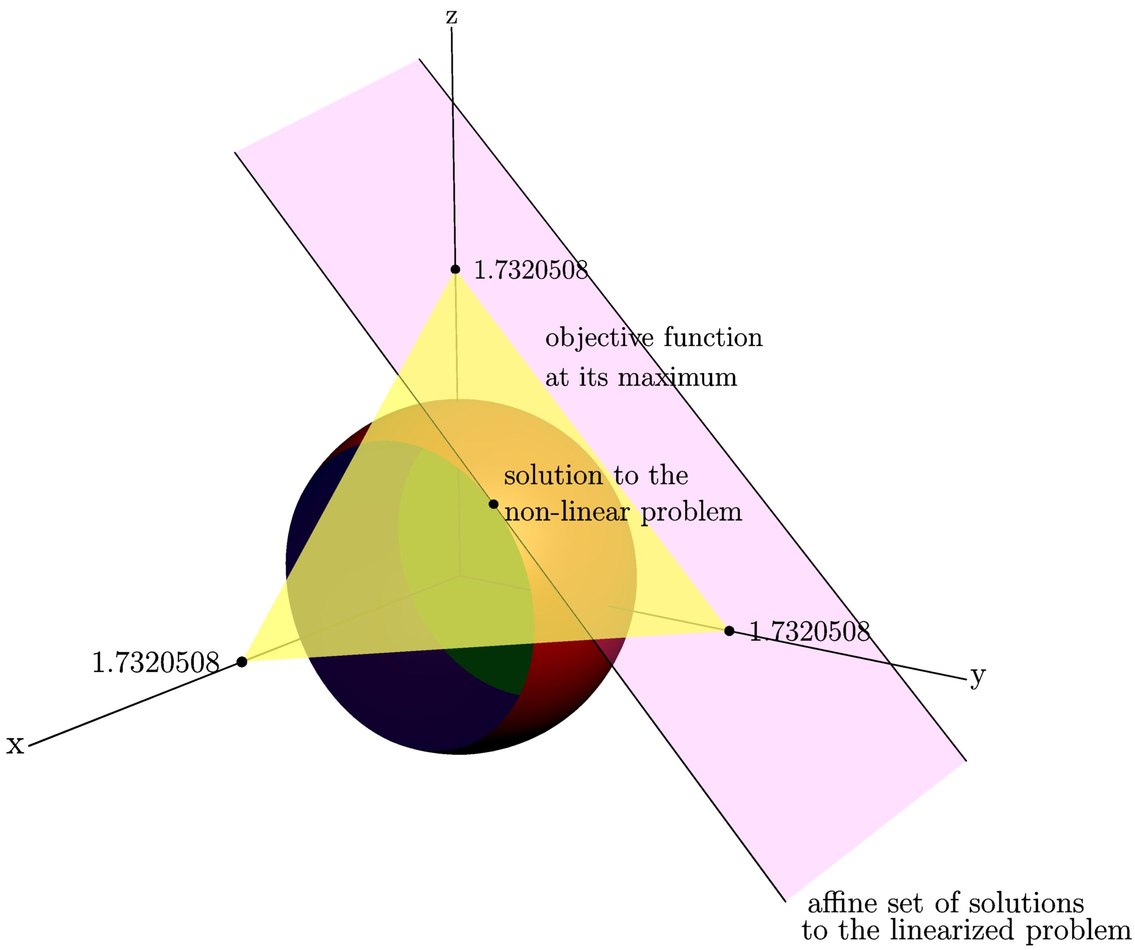


Figure 9.5: **Non-Linear Problem I**

feasible set (which is coloured dark green, green and red) at the unique solution $x = 0.5$, $y = 0.61235$, $z = 0.61235$ (indicated by a red dot).

The recursive linearization algorithm when applied to this problem gives this result, however this is more by luck than good management as there are in fact multiple solutions to the linearized problem, as indicated in Figure 9.5 by the line of solutions to the linearized problem; this is is an atypical problem.

9.6.2 Non-Linear Problem II

We apply the approach to the following problem (refer to [10, p.301]):

$$\begin{aligned} \textit{maximize}\ & 3x_1 - x_2^2/2 \\ \textit{subject to}\ & x_1^2 + x_2^2 \leq 1 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{aligned}$$

What we find is that when using the recursive linearization algorithm the solution doesn't converge properly. We can see why this happens by referring to Figure 9.6. The figure shows the feasible and optimal solutions to the limiting linearized problem. The optimal solutions to the limiting linearized problem form the set $\{(x_1, x_2) : x_1 = 1; x_2 \geq 0\}$, so there is no guarantee that the algorithm will converge to $(x_1, x_2) = (1, 0)$. To overcome this difficulty we should apply the Newton-Raphson algorithm, using the binding constraints, during convergence of the linearized form to a solution after Step 7 of the algorithm; this problem is also atypical.

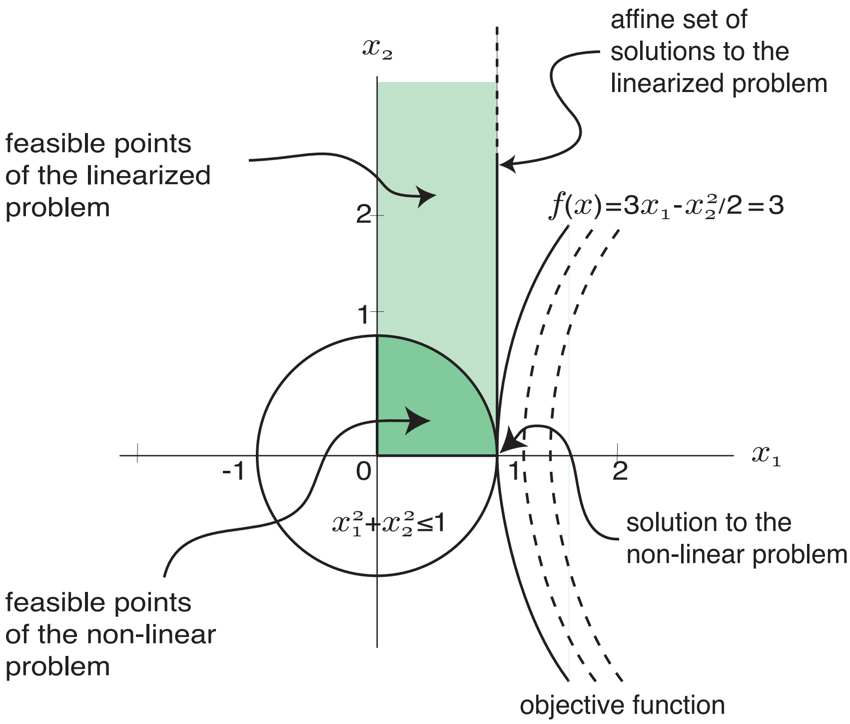


Figure 9.6: **Non-Linear Problem II**

Incorporating Newton's method is consistent with the research detailed in [22].

9.7 Conclusion

Both problems are atypical. The solution to the first problem was correct due to the symmetry of the starting guess. The solution reached for the second problem was only to the linearized problem, however the algorithm also successfully identified the slack and binding constraints, without incorporation of Newton's Method. So, generally speaking, in order to solve atypical non-linear optimization problems Newton's Method needs to be incorporated into the solution algorithm.

Appendices

Appendix A

Convergence

Here we consider the problem of convergence in greater depth. Specifically, the convergence of the averaging matrix V is investigated.

Note Equation 3.9, which states that $(V^T)^i V^i = (PK)^i P = P(KP)^i$. This implies that we can analyze the spectral properties of say PK in order to gain knowledge of the convergence properties of V.

Afriat's [2] theory of reciprocal spaces is used to investigate the spectral structure of PK and KP, leading to a better understanding of the P-unitary matrix U, through the development of some associated algebra. It is shown that the eigenvalues of PK (excluding those associated with fixed-points) are all in $(0, 1)$. Note that if the greatest of these is near unity then, after a number of iterations, neither $(PK)^n h$ nor $V^n h$ will converge quickly to a fixed-point of both P and K, which suggests that the affine regression approach may be required.

A.1 Reciprocal Vectors and Spaces

Following Afriat [2], spaces $\mathcal{A}$ and $\mathcal{B}$ are said to be *reciprocal spaces* in spaces $\mathcal{P}$, $\mathcal{K}$ if they are orthogonal projections of each other in $\mathcal{P}$, $\mathcal{K}$; thus if P and K are the symmetric idempotents defining the orthogonal projections onto $\mathcal{P}$ and $\mathcal{K}$ respectively, then $\mathcal{A} = P\mathcal{B}$, $\mathcal{B} = K\mathcal{A}$. Vectors which span *reciprocal rays* give rise to a *reciprocal pair* of vectors in the relevant spaces; if the vectors are of the same length we say they are a *balanced pair*. The *coefficient of inclination* between the reciprocal spaces $\mathcal{P}$ and $\mathcal{K}$ is $R = trace(PK)$. (Afriat uses e and f to denote a pair of symmetric idempotents - this notation is more common when analyzing more general structures.)

An eigenvalue in $(0, 1)$ is said to be *proper*; an eigenvector having a proper eigenvalue is also said to be *proper*; a reciprocal vector associated with a proper eigenvalue is said to be proper, and the subspace generated by the set of all proper eigenvalues of PK and KP is called the proper subspace and is denoted by $\mathcal{H}$. Note also that if $x \in \mathcal{P}$ then $Kx = 0 \Rightarrow (I - K)x = x, \Rightarrow SKSx = x, \Rightarrow KSx = Sx$.

Bearing in mind the possibility that ρ may be a complex number, let $PKx_P = \rho^2 x_P$. if $\rho \neq 0$ define

$x_K = Kx_P/\rho$, then $Px_K = PKx_P/\rho = \rho^2 x_P/\rho = \rho x_P$; so $x_P = Px_K/\rho$. Further, $Px_K = \rho x_P \Rightarrow KPx_K = \rho K x_P, = \rho^2 x_K$. Summing up, there is the *balanced reciprocal scheme*

$$\begin{array}{llll} PKx_P & = & \rho^2 x_P, & \text{(a)} \\ KPx_K & = & \rho^2 x_K, & \text{(b)} \\ Kx_P & = & \rho x_K, \quad \text{and} & \text{(c)} \\ Px_K & = & \rho x_P. & \text{(d)} \end{array} \qquad \text{(A.1)}$$

Note that, from Equation A.1 (c), $x_K^T K x_P = \rho x_K^T \boldsymbol{x}_K \Rightarrow x_K^T \boldsymbol{x}_P = \rho x_K^T \boldsymbol{x}_K$, (since $Kx_K = x_K$). Similarly $x_P^T \boldsymbol{x}_K = \rho x_P^T \boldsymbol{x}_P$; it follows that $\|x_P\| = \|x_K\|$, and $cos(x_P, x_K) = x_P^T \boldsymbol{x}_K/(\|x_P\|\|x_K\|) = x_P^T \boldsymbol{x}_K/\|x_P\|^2$, $= \rho$. Thus given the eigenvector x_P of PK, with eigenvalue ρ^2, take $x_K = Kx_P/\rho$, then x_P, x_K is a pair of reciprocals in $\mathcal{P}$, $\mathcal{K}$. Moreover, since they are of the same length we say they are a *balanced reciprocal pair.*

Lemma A.1.1 An eigenvector of PK and an eigenvector of KP with distinct eigenvalues are orthogonal.

Proof Let x_P be an eigenvector of PK with eigenvalue ρ^2, and x'_K be an eigenvector of KP with eigenvalue $\rho^{'2}$, distinct from ρ^2, and assume without loss of generality that $\rho^{'2} \neq 0$, then

$$\rho^{'2} x_K^{'T} x_P = (KPx'_K)^T \boldsymbol{x}_P = x_K^{'T} PKx_P = \rho^2 x_K^{'T} x_P.$$

Thus $(\rho^2 - \rho^{'2}) x_K^{'T} x_P = 0, \Rightarrow x_K^{'T} x_P = 0$. □

Lemma A.1.2 Two eigenvectors of PK with distinct eigenvalues are orthogonal.

Proof Let x_P and x'_K be eigenvectors of PK with eigenvalues ρ^2 and $\rho^{'2}$ respectively. Assume without loss of generality that $\rho^{'2} \neq 0$. Now

$$x_P^{'T} x_P = \{PKx_P/\rho^{'2}\}^T \boldsymbol{x}_P = x_P^{'T} KPx_P/\rho^{'2} = x_P^{'T} Kx_P/\rho^{'2} = \{Kx'_P\}^T \boldsymbol{x}_P/\rho^{'2}$$

$$= \{\rho' x'_K\}^T \boldsymbol{x}_P/\rho^{'2} = x_K^{'T} x_P/\rho' = 0 \quad \text{(by Lemma A.1.1)} \; . \; \square$$

Summing up we have

Theorem: A.1.3 Eigenvectors of PK and KP with distinct eigenvalues are orthogonal.

The angle between x_P and x_K is given by $cos(x_P, x_K) = \rho$; the coefficient of inclination is ρ^2. If the scheme has ρ positive then it is said to be an acute reciprocal scheme, and if it has ρ negative then it is said to be obtuse. Let $(x_P|x_K)$ denote a reciprocal pair x_P, x_K in $\mathcal{P}$, $\mathcal{K}$; let $[x_P|x_K]$ denote a balanced reciprocal pair x_P, x_K in $\mathcal{P}$, $\mathcal{K}$, and let $\{x_P|x_K\}$ denote a normalized balanced reciprocal pair x_P, x_K in $\mathcal{P}$, $\mathcal{K}$; define $\alpha(x_P|x_K) = (\alpha x_P|\alpha x_K)$ and regard two pairs $(x_P|x_K)$ and $(x'_P|x'_K)$ as equivalent if $(x_P|x_K) = \beta(x'_P|x'_K)$ for some $\beta > 0$.

Define the "sum" vector $s = x_P + x_K$ and the "difference" vector $d = x_P - x_K$; note that $s \perp d$. Now consider the symmetric positive definite matrix $W = (P + K)/2$; it can be shown that $Ws = \frac{1+\rho}{2} s$, and

$Wd = \frac{1-\rho}{2}d$. Thus a pair of eigenvalues and vectors has been delineated for W; since this is a symmetric positive definite matrix it follows that ρ is real; further, since P and K are symmetric idempotents, $\rho \in [-1, 1]$. It is now apparent that the eigenvalues, ρ^2 of PK and of KP are all positive real numbers in the closed interval $[0, 1]$.

The above discussion can be tightened to the following

Lemma A.1.4 ρ^2 is an eigenvalue of $PK \Leftrightarrow (1 \pm \rho)/2$ is an eigenvalue of $W = (P + K)/2$.

Proof: The forward implication has already been covered. For the reverse implication let x be an eigenvector of W with eigenvalue $(1+\rho)/2$; this implies $(P+K)x = (1\pm\rho)x$, $\Rightarrow P(P+K)x = (1\pm\rho)Px$, $\Rightarrow Px + PKx = (1 \pm \rho)Px$, $\Rightarrow PKx = \pm\rho Px$,

$$KPKx = \pm\rho KPx, \tag{A.2}$$

and, similarly,

$$PKPx = \pm\rho PKx. \tag{A.3}$$

Thus from equations A.2 and A.3, $PKPKx = \pm\rho PKPx = \rho^2 PKx$; similarly $KPKPx = \rho^2 KPx$. So ρ^2 is an eigenvalue of PK and of KP.□

It follows that the eigenvalues of PK all lie in $[0, 1]$.

Define $\eta = \sqrt{1-\rho^2}$, $\tau = \eta^{-1}$, and

$$\begin{array}{lcll} \xi_P & = & \tau PSx_K, \text{ and} & (a) \\ \xi_K & = & \tau KSx_P, & (b) \end{array} \tag{A.4}$$

Note that

$$\eta^2 = 1 - \rho^2. \tag{A.5}$$

Lemma A.1.5 The map $\mathcal{B} : [x_R|x_K] \mapsto [\xi_P|\xi_K] = \tau(PSx_K|KSx_P)$ is well-defined, $\mathcal{B}^2$ is the identity map, and $\mathcal{B}$ is one to one and onto for the set of all reciprocal pairs, and when restricted to the set of normalized reciprocals in $\mathcal{P}$, $\mathcal{K}$.

Proof Now $\xi_P = \tau PSx_K$, and $\xi_K = \tau KSx_P$, so $PK\xi_P = \tau PKPSx_K = \tau PKPSKx_K = -\tau PKSPKx_K$ $= \tau PSKPKx_K = \tau PSKPx_K = \rho^2\tau PSx_K = \rho^2\xi_P$; similarly $KP\xi_K = \rho^2\xi_K$. Also $K\xi_P = \tau KPSx_K$ $= \tau KPSKx_K = -\tau KSPKx_K = -\tau KSPx_K = -\rho\tau KSx_P = -\rho\xi_K$; similarly $P\xi_K = -\rho\xi_P$. □

In summary

$$\begin{array}{lcll} PK\xi_P & = & \rho^2\xi_P, & (a) \\ KP\xi_K & = & \rho^2\xi_K, & (b) \\ K\xi_P & = & -\rho\xi_K, \text{ and} & (c) \\ P\xi_K & = & -\rho\xi_P\ . & (d) \end{array} \tag{A.6}$$

Thus we have a balanced reciprocal pair and may write $[\xi_P|\xi_K]$ and so B is well-defined. Since $\|\xi_P\|^2 = \|\tau P S x_K\|^2 = \tau^2 x_K^T S P S x_K = \tau^2 x_K^T (I-P) x_K = \tau^2(1-\rho^2)\|x_K\|^2 = \|x_K\|^2$, and similarly $\|\xi_K\|^2 = \|x_P\|^2$, we have $\{x_P|x_K\} \Rightarrow \{\xi_P|\xi_K\}$. Further, for any pair $[x_P|x_K]$,

$$\mathcal{B}^2[x_P|x_K] = \tau\mathcal{B}[PSx_K|KSx_P]$$

$$= \tau^2[PSKSx_P|KSPSx_K] = \tau^2[P(I-K)x_P|K(I-P)x_K]$$

$$= \tau^2[x_P - PKx_P|x_K - KPx_K] = \tau^2(1-\rho^2)[x_P|x_K]$$

$$= [x_P|x_K];$$

that is $\mathcal{B}^2$ is the identity map, from which it follows that $\mathcal{B}$ is one to one and onto. □

Thus a new pair, $[\xi_P|\xi_K]$, of balanced reciprocals in $\mathcal{P}$, $\mathcal{K}$ has been constructed, corresponding to the same eigenvalue, ρ^2, of PK and KP. However $[x_P|x_K]$ and $[\xi_P|\xi_K]$ have opposite angularity. If the original pair is normalized then the new pair will also be.

Let G stand for P *or* K; let $\mathcal{G}$ stand for $\mathcal{P}$ *or* $\mathcal{K}$; and let $\tilde{P} = K$, $\tilde{K} = P$, then the first of the following equations is a direct consequence of Equation A.4, while the later is the result of a short computation:

$$\begin{aligned} GSx_{\tilde{G}} &= \eta\xi_G, && \text{(a)} \\ GS\xi_{\tilde{G}} &= \eta x_G. && \text{(b)} \end{aligned} \qquad \text{(A.7)}$$

Lemma A.1.6 (a) $x_G^T\xi_G = 0$, (b) $x_G^T S\xi_{\tilde{G}} = \eta$.

Proof:

(a) $x_G^T\xi_G = x_G^T(\tau GSx_{\tilde{G}}) = \tau x_G^T GS\tilde{G}x_G/\rho = \tau x_G^T GS\tilde{G}Gx_G/\rho \overset{3.7 or 3.7'}{=} 0$.

(b) $x_G^T S\xi_{\tilde{G}} = x_G^T GS\xi_{\tilde{G}} = x_G^T(\eta x_G) = \eta x_G^T \boldsymbol{x}_G = \eta$.

□

A.2 Spectra of Products of Idempotents

Spectral decomposition is an important source of articulation in analyzing multivariate problems: Using the method of reciprocals we establish a near orthogonality of bases for PK and KP. Essentially we show that only eigenvectors associated with the same eigenvalue are oblique - the rest are orthogonal.

Let $[x_P|x_K]$ and $[x'_P|x'_K]$ be balanced reciprocal pairs; if $x_P \perp x'_P$, $x_P \perp x'_K$, $x_K \perp x'_P$, and $x_K \perp x'_K$, then write $[x_P|x_K] \perp [x'_P|x'_K]$; we say that the reciprocal pairs are orthogonal. If $a \perp x_P$ and $a \perp x_K$ write $a \perp [x_P|x_K]$.

Lemma A.2.1 If $[x_P|x_K]$, $g \in \mathcal{G}$, and $g \perp x_G$ then $g \perp [x_P|x_K]$.

Proof $g \perp x_G \Rightarrow g^T \boldsymbol{x}_G = 0, \Rightarrow g^T GS\tilde{G}Sx_G = 0 \Rightarrow g^T S\tilde{G}Sx_G = 0 \Rightarrow g^T(I - \tilde{G})x_G = 0 \Rightarrow g^T \tilde{G}x_G = 0$ $\Rightarrow g^T(\rho x_{\tilde{G}}) = 0 \Rightarrow g \perp x_{\tilde{G}}$. Thus $g \perp [x_P|x_K]$. □

We specialize this result to what will be called the *Separation Lemma*:

Lemma A.2.2 $x'_G \perp x_G \Leftrightarrow x'_G \perp x_{\tilde{G}}$.

Proof $x'_G \perp x_G \Leftrightarrow x_G^{'T} x_G = 0 \Leftrightarrow x_G^{'T} \rho x_G = 0 \Leftrightarrow x_G^{'T} G x_{\tilde{G}} = 0 \Leftrightarrow x_G^{'T} x_{\tilde{G}} = 0 \Leftrightarrow x'_G \perp x_{\tilde{G}} = 0$. □

Corollary: A.2.3 $x'_G \perp x_G \Leftrightarrow x'_{\tilde{G}} \perp x_G$.

Proof: Swap x with x' in the lemma. □

Corollary: A.2.4 $x'_G \perp x_{\tilde{G}} \Leftrightarrow x'_{\tilde{G}} \perp x_{\tilde{G}}$.

Proof: Set G equal to $\tilde{G}$ in Corollary A.2.3. □

Thus, from Corollaries A.2.3 and A.2.4 we have the *Four Musketeers Lemma*:

Lemma A.2.5 $x'_G \perp x_G \Leftrightarrow x'_{\tilde{G}} \perp x_G \Leftrightarrow x'_G \perp x_{\tilde{G}} \Leftrightarrow x'_{\tilde{G}} \perp x_{\tilde{G}}$.

Continuing, there is the *Transposition Lemma*:

Lemma A.2.6 $x'_G \perp x_{\tilde{G}} \Leftrightarrow \xi'_{\tilde{G}} \perp \xi_G$.

Proof $x'_G \perp x_{\tilde{G}} \Leftrightarrow x_G^{'T} x_{\tilde{G}} = 0 \Leftrightarrow x_G^{'T} \tilde{G}S\xi_G = 0 \Leftrightarrow x_G^{'T} S\tilde{G}\xi_G = 0 \Leftrightarrow \xi_{\tilde{G}}^{'T} \xi_G = 0 \Leftrightarrow \xi'_{\tilde{G}} \perp \xi_G$. □

From these results it is straightforward to establish the following *Four Musketeers Theorem*:

Theorem: A.2.7 The spaces $< x_P, x_K, \xi_P, \xi_K >$ and $< x'_P, x'_K, \xi'_P, \xi'_K >$ are orthogonal if and only if $x'_P \perp < x_P, \xi_P >$.

This theorem shows that mutually orthogonal pairs of balanced reciprocal pairs can be constructed - that is we can construct

$$\{[x_P^{(1)}|x_K^{(1)}], [\xi_P^{(1)}|\xi_K^{(1)}]\}, \{[x_P^{(2)}|x_K^{(2)}], [\xi_P^{(2)}|\xi_K^{(2)}]\} \cdots ,$$

with

$$\{[x_P^{(i)}|x_K^{(i)}], [\xi_P^{(i)}|\xi_K^{(i)}]\}, \{[x_P^{(j)}|x_K^{(j)}], [\xi_P^{(j)}|\xi_K^{(j)}]\} \quad \forall i, j, i \neq j.$$

as long as we can find an eigenvector (with eigenvalue in $(0, 1)$) which is perpendicular to those already delineated.

Proof The above lemmas show that repeated roots do not cause problems in the construction of such pairs; furthermore note that the complete set of delineated eigenvectors of $P+K$ is a mutually orthogonal set, as $P + K$ is positive definite, so eigenvectors of PK (or KP) associated with different eigenvalues must be orthogonal. □

A.3 Conclusion

By spectral decomposition of the matrix PK we have shown that the convergence of the sequence $(PK)^n x$ is better than geometric, however no upper bound other than unity has been found for the coefficient of convergence.

Appendix B

Incremental Affine Regression

Equation 3.15 is computational theory and naturally we wish to take advantage of Theorem 2.2.15 to lighten the computational load. It is apparent from Chapter 6 that we are searching for the correct slack and binding conditions in a context where extra structure obtains; an algorithm which takes advantage of this extra structure is developed here since it is not practical to perform an affine regression each time proximality (defined in Chapter 6.2) may have been reached. So a method of regression is developed which is analogous to the conjugate gradient method - that is, it is incremental.

B.1 Regression Formulation

With $G_i = [\boldsymbol{g}_1, \cdots, \boldsymbol{g}_i]$, where the $\boldsymbol{g}_i$ are given by Equation 3.11, write

$$\gamma_i = G_i^{+T}\mathbf{1}_i \ , \tag{B.1}$$

then, if $(I - G_i^+ G_i)\mathbf{1}_i = \mathbf{0}$, the affine regression solution given by Equation 3.15 is $\boldsymbol{\sigma}_i = \gamma_i / \|\gamma_i\|^2$ Further, Theorem 2.2.15 in the present context may be written

Theorem: B.1.1 If $G_{i+1} = [G_i | \boldsymbol{g}_{i+1}]$ then

$$\text{(a)} \quad G_{i+1}^+ = \begin{bmatrix} G_i^+[I - \boldsymbol{g}_{i+1}k_{i+1}^T] \\ k_{i+1}^T \end{bmatrix}, \quad \text{where}$$

$$\text{(b)} \quad k_{i+1} = \frac{(I - G_i G_i^+)\boldsymbol{g}_{i+1}}{\left\|(I - G_i G_i^+)\boldsymbol{g}_{i+1}\right\|^2} \quad \text{if} \quad (I - G_i G_i^+)\boldsymbol{g}_{i+1} \neq \mathbf{0} \ .$$

then, multiplying by $\mathbf{1}_{i+1}$,

$$\begin{aligned} G_{i+1}^{+T}\mathbf{1}_{i+1} &= (I - \boldsymbol{k}_{i+1}\boldsymbol{g}_{i+1}^T)G_i^{+T}\mathbf{1}_i + \boldsymbol{k}_{i+1}, \\ \text{where} \quad \boldsymbol{k}_{i+1} &= \frac{(I - G_i G_i^+)\boldsymbol{g}_{i+1}}{\left\|(I - G_i G_i^+)\boldsymbol{g}_{i+1}\right\|^2}, \quad \text{if } (I - G_i G_i^+)\boldsymbol{g}_{i+1} \neq \mathbf{0}, \end{aligned} \tag{B.2}$$

B.2 Recursive Formulation

Here only the first possibility (Theorem 2.2.15 b) is important while the second possibility (Theorem 2.2.15 c) is the stopping condition.

In view of Equation B.1, Theorem B.1.1 can be rewritten as

$$\begin{aligned} \gamma_{i+1} &= (I - \boldsymbol{k}_{i+1}\boldsymbol{g}_{i+1}^T)\gamma_i + \boldsymbol{k}_{i+1} = \gamma_i + (1 - \boldsymbol{g}_{i+1}^T\gamma_i)\boldsymbol{k}_{i+1} \quad \text{where} \\ \boldsymbol{k}_{i+1} &\overset{2.13}{=} \frac{\boldsymbol{g}_{i+1} \lhd \{\boldsymbol{g}_1, \cdots, \boldsymbol{g}_i\}}{\|\boldsymbol{g}_{i+1} \lhd \{\boldsymbol{g}_1, \cdots, \boldsymbol{g}_i\}\|^2} \ \text{ if } \ \boldsymbol{g}_{i+1} \lhd \{\boldsymbol{g}_1, \cdots, \boldsymbol{g}_i\} \neq \boldsymbol{0}\ . \end{aligned} \tag{B.3}$$

Define, for $i < j$,

$$\begin{aligned} \boldsymbol{y}_{i,j} &= \boldsymbol{g}_i \lhd \{\boldsymbol{g}_{i+1}, \cdots, \boldsymbol{g}_j\} && \text{(a)} \\ \boldsymbol{z}_{j,i} &= \boldsymbol{g}_j \lhd \{\boldsymbol{g}_i, \cdots, \boldsymbol{g}_{j-1}\} && \text{(b)} \end{aligned} \tag{B.4}$$

then Equation B.4 can be written

$$\begin{aligned} \gamma_{i+1} &= \gamma_i + (1 - \boldsymbol{g}_{i+1}^T\gamma_i)\boldsymbol{k}_{i+1} \quad \text{where} \\ \boldsymbol{k}_{i+1} &= \frac{\boldsymbol{z}_{i+1,1}}{\|\boldsymbol{z}_{i+1,1}\|^2} \ \text{ if } \ \boldsymbol{z}_{i+1,1} \neq \boldsymbol{0}, \end{aligned} \tag{B.5}$$

There remains the problem of computing $\boldsymbol{k}_i$ for the special case where Equation 3.11 obtains; we proceed by noting that, in view of Theorem 2.3.10, Definition B.4 implies

$$\begin{aligned} \boldsymbol{y}_{1,j} &= \boldsymbol{g}_1 \lhd \{\boldsymbol{g}_2, \cdots, \boldsymbol{g}_j\} = (\boldsymbol{g}_1 \lhd \{\boldsymbol{g}_2, \cdots, \boldsymbol{g}_{j-1}\}) \lhd (\boldsymbol{g}_j \lhd \{\boldsymbol{g}_2, \cdots, \boldsymbol{g}_{j-1}\}) && \text{(a)} \\ \boldsymbol{z}_{j,1} &= \boldsymbol{g}_j \lhd \{\boldsymbol{g}_1, \cdots, \boldsymbol{g}_{j-1}\} = (\boldsymbol{g}_j \lhd \{\boldsymbol{g}_2, \cdots, \boldsymbol{g}_{j-1}\}) \lhd (\boldsymbol{g}_1 \lhd \{\boldsymbol{g}_2, \cdots, \boldsymbol{g}_{j-1}\}) && \text{(b)} \end{aligned} \tag{B.6}$$

that is

$$\begin{aligned} \boldsymbol{y}_{1,j} &= \boldsymbol{y}_{1,j-1} \lhd \boldsymbol{z}_{j,2} && \text{(a)} \\ \boldsymbol{z}_{j,1} &= \boldsymbol{z}_{j,2} \lhd \boldsymbol{y}_{1,j-1} && \text{(b)} \end{aligned} \tag{B.7}$$

So there are the starting conditions

$$\begin{aligned} \boldsymbol{y}_{1,2} &= \boldsymbol{g}_1 \lhd \boldsymbol{g}_2 && \text{(a)} \\ \boldsymbol{z}_{2,1} &= \boldsymbol{g}_2 \lhd \boldsymbol{g}_1 && \text{(b)} \end{aligned} \tag{B.8}$$

and the recursion

$$\begin{aligned} &\text{For } \ i \geq 3 \\ \boldsymbol{y}_{1,i} &= \boldsymbol{y}_{1,i-1} \lhd \boldsymbol{z}_{i,2} && \text{(c)} \\ \boldsymbol{z}_{i,1} &= \boldsymbol{z}_{i,2} \lhd \boldsymbol{y}_{1,i-1} && \text{(d)} \end{aligned} \tag{B.9}$$

Now for $j \geq 2$, $\boldsymbol{z}_{i,j} = U\boldsymbol{z}_{i-1,j-1}$, so with U taking precedence over $\lhd$ (i.e. $Ux \lhd \boldsymbol{y} = (Ux) \lhd \boldsymbol{y}$) we may write Equation B.9 as

$$\begin{aligned} &\text{For } \ i \geq 3 \\ \boldsymbol{y}_{1,i} &= \boldsymbol{y}_{1,i-1} \lhd U\boldsymbol{z}_{i-1,1} && \text{(a)} \\ \boldsymbol{z}_{i,1} &= U\boldsymbol{z}_{i-1,1} \lhd \boldsymbol{y}_{1,i-1} && \text{(b)} \end{aligned} \tag{B.10}$$

Thus we have the recursion:

Starting ($i = 2$):

$$\begin{array}{lll} i & := & 2 \\ \boldsymbol{y}_{1,i} & := & \boldsymbol{g}_1 \lhd \boldsymbol{g}_2, \\ \boldsymbol{z}_{i,1} & := & \boldsymbol{g}_2 \lhd \boldsymbol{g}_1, \\ \boldsymbol{k}_i & := & \boldsymbol{z}_{i,1}/\|\boldsymbol{z}_{i,1}\|^2 \\ \boldsymbol{\gamma}_i & := & [\boldsymbol{g}_1\, \boldsymbol{g}_2]^{+T}[1\,1]^T \\ \boldsymbol{\sigma}_i & := & (\boldsymbol{g}_1 + \boldsymbol{g}_2)/2 \end{array} \tag{B.11}$$

Summary:

definition				check
		$i = 2$	$i \geq 3$	
i	$:=$	2	$i+1$	
$\boldsymbol{y}_{1,i-1}$	$:=$	$\sim$	$\boldsymbol{y}_{1,i}$	
$\boldsymbol{z}_{i-1,1}$	$:=$	$\sim$	$\boldsymbol{z}_{i,1}$	
$\boldsymbol{\gamma}_{i-1}$	$:=$	$\sim$	$\boldsymbol{\gamma}_i$	
$\boldsymbol{y}_{1,i}$	$:=$	$\boldsymbol{g}_1 \lhd \boldsymbol{g}_2$	$\boldsymbol{y}_{1,i-1} \lhd U\boldsymbol{z}_{i-1,1}$	$\boldsymbol{g}_1 \lhd \{\boldsymbol{g}_2, \cdots, \boldsymbol{g}_i\}$
$\boldsymbol{z}_{i,1}$	$:=$	$\boldsymbol{g}_2 \lhd \boldsymbol{g}_1$	$U\boldsymbol{z}_{i-1,1} \lhd \boldsymbol{y}_{1,i-1}$	$\boldsymbol{g}_i \lhd \{\boldsymbol{g}_1, \cdots, \boldsymbol{g}_{i-1}\}$
$\boldsymbol{k}_i$	$:=$	$\boldsymbol{z}_{i,1}/\|\boldsymbol{z}_{i,1}\|^2$	$\boldsymbol{z}_{i,1}/\|\boldsymbol{z}_{i,1}\|^2, \boldsymbol{z}_{i,1} \neq \boldsymbol{0}$	
$\boldsymbol{\gamma}_i$	$:=$	$[\boldsymbol{g}_1\, \boldsymbol{g}_2]^{+T}[1\,1]^T$	$\boldsymbol{\gamma}_{i-1} + (1 - \boldsymbol{g}_i^T\boldsymbol{\gamma}_{i-1})\boldsymbol{k}_i$	$G_i^{+T}\boldsymbol{1}_i$
$\boldsymbol{\sigma}_i$	$:=$	$(\boldsymbol{g}_1 + \boldsymbol{g}_2)/2$	$\boldsymbol{\gamma}_i/\|\boldsymbol{\gamma}_i\|^2$	

(B.12)

The first part of the recursion forms the variable (parallel) part of a "do" construct in LISP, the stopping condition is $\boldsymbol{z}_{i,1} = \boldsymbol{0}$, and the second part of the recursion forms the sequential body of the "do".

Appendix C

An Inverse Computation

With reference to Chapter 7, the inverse of A_τ may be derived as follows using a result in Albert [4]:

$$[U\vdots V]^+ = \begin{bmatrix} U^+ - U^+VJ \\ J \end{bmatrix}$$

where

$$C = (I - UU^+)V$$

$$K = (I + (U^+V(I - C^+C))^T(U^+V(I - C^+C)))^{-1}$$

and

$$J = C^+ + (I - C^+C)KV^TU^{+T}U^+(I - VC^+) \ .$$

In the present case

$$C = \big(I - (I_m \otimes \mathbf{1}_n)(I_m \otimes \mathbf{1}_n)^+\big)\big(\mathbf{1}_m \otimes I_n\big) = \mathbf{1}_m \otimes I_n - (I_m \otimes \mathbf{1}_n)(I_m \otimes \mathbf{1}_n)^+(\mathbf{1}_m \otimes I_n)$$

$$= \mathbf{1}_m \otimes I_n - \mathbf{1}_m \otimes \mathbf{1}_n\mathbf{1}_n^+ = \mathbf{1}_m \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+)$$

$$\Rightarrow C^+ = \mathbf{1}_m^+ \otimes (I - \mathbf{1}_n\mathbf{1}_n^+)$$

$$\Rightarrow C^+C = (\mathbf{1}_m^+\mathbf{1}_m) \otimes (I - \mathbf{1}_n\mathbf{1}_n^+) = I - \mathbf{1}_n\mathbf{1}_n^+$$

$$\Rightarrow I - C^+C = \mathbf{1}_n\mathbf{1}_n^+ \ .$$

$$K = \Big(I + \big((I_m \otimes \mathbf{1}_n)^+(\mathbf{1}_m \otimes I_n)(I - C^+C)\big)^T\big((I_m \otimes \mathbf{1}_n)^+(\mathbf{1}_m \otimes I_n)(I - C^+C)\big)\Big)^{-1}$$

$$= \Big(I + \big((\mathbf{1}_m \otimes \mathbf{1}_n^+)(I - C^+C)\big)^T\big((\mathbf{1}_m \otimes \mathbf{1}_n^+)(I - C^+C)\big)\Big)^{-1}$$

$$= \Big(I + \big((\mathbf{1}_m \otimes \mathbf{1}_n^+)(\mathbf{1}_n\mathbf{1}_n^+)\big)^T\big((\mathbf{1}_m \otimes \mathbf{1}_n^+)(\mathbf{1}_n\mathbf{1}_n^+)\big)\Big)^{-1}$$

$$= \big(I + (\mathbf{1}_m \otimes \mathbf{1}_n^+)^T(\mathbf{1}_m \otimes \mathbf{1}_n^+)\big)^{-1} = \big(I + (\mathbf{1}_m^T \otimes \mathbf{1}_n^{+T})(\mathbf{1}_m \otimes \mathbf{1}_n^+)\big)^{-1}$$

$$= (I + \mathbf{1}_m^T\mathbf{1}_m \otimes \mathbf{1}_n^{+T}\mathbf{1}_n^+)^{-1} = (I + m \otimes \mathbf{1}_n^{+T}\mathbf{1}_n^+)^{-1} = (I + m\mathbf{1}_n^{+T}\mathbf{1}_n^+)^{-1}$$

$$= (I + \frac{m}{n}\mathbf{1}_n\mathbf{1}_n^+)^{-1} = I - \frac{m}{m+n}\mathbf{1}_n\mathbf{1}_n^+$$

and

$$\begin{aligned}
J &= C^+ + (I - C^+C)K(\mathbf{1}_m \otimes I_n)^T(I_m \otimes \mathbf{1}_n)^{+T}(I_m \otimes \mathbf{1}_n)^+\big(I - (\mathbf{1}_m \otimes I_n)C^+\big) \\
&= C^+ + (I - C^+C)K(\mathbf{1}_m^T \otimes I_n)(I_m \otimes \mathbf{1}_n^{+T})(I_m \otimes \mathbf{1}_n^+)\big(I - (\mathbf{1}_m \otimes I_n)C^+\big) \\
&= C^+ + \big(I - C^+C\big)K\big(\mathbf{1}_m^T \otimes (\mathbf{1}_n^{+T}\mathbf{1}_n^+)\big)\big(I - (\mathbf{1}_m \otimes I_n)C^+\big) \\
&= C^+ + \big(I - C^+C\big)\big(\mathbf{1}_m^T \otimes (K\mathbf{1}_n^{+T}\mathbf{1}_n^+)\big)\big(I - (\mathbf{1}_m \otimes I_n)C^+\big) \\
&= C^+ + \big(I - C^+C\big)\big(\mathbf{1}_m^T \otimes (K\mathbf{1}_n\mathbf{1}_n^+/n)\big)\big(I - (\mathbf{1}_m \otimes I_n)C^+\big) \\
&= C^+ + \big(I - C^+C\big)\big(\mathbf{1}_m^T \otimes (\mathbf{1}_n\mathbf{1}_n^+/(m+n))\big)\big(I - (\mathbf{1}_m \otimes I_n)C^+\big) \\
&= \mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+) + \big(\mathbf{1}_n\mathbf{1}_n^+\big)\big(\mathbf{1}_m^T \otimes (\mathbf{1}_n\mathbf{1}_n^+/(m+n))\big)\big(I - (\mathbf{1}_m \otimes I_n)(\mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+))\big) \\
&= \mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+) + \big(\mathbf{1}_m^T \otimes (\mathbf{1}_n\mathbf{1}_n^+/(m+n))\big)\big(I - (\mathbf{1}_m \otimes I_n)(\mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+))\big) \\
&= \mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+) + \big(\mathbf{1}_m^T \otimes (\mathbf{1}_n\mathbf{1}_n^+/(m+n))\big)\big(I - \mathbf{1}_m^+ \otimes ((\mathbf{1}_m \otimes I_n)(I_n - \mathbf{1}_n\mathbf{1}_n^+))\big) \\
&= \mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+) + \big(\mathbf{1}_m^T \otimes (\mathbf{1}_n\mathbf{1}_n^+/(m+n))\big)\big(I - \mathbf{1}_m^+ \otimes ((\mathbf{1}_m \otimes I_n(I_n - \mathbf{1}_n\mathbf{1}_n^+)))\big) \\
&= \mathbf{1}_m^+ \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+) + \big(\mathbf{1}_m^T \otimes (\mathbf{1}_n\mathbf{1}_n^+/(m+n))\big)\big(I - \mathbf{1}_m^+ \otimes \mathbf{1}_m \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+)\big) \\
&= \mathbf{1}_m^+ \otimes I_n - \mathbf{1}_m^+ \otimes \mathbf{1}_n\mathbf{1}_n^+ + \frac{m}{m+n}\big(\mathbf{1}_m^+ \otimes (\mathbf{1}_n\mathbf{1}_n^+)\big)\big(I - (\mathbf{1}_m\mathbf{1}_m^+) \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+)\big) \\
&= \mathbf{1}_m^+ \otimes I_n - \mathbf{1}_m^+ \otimes \mathbf{1}_n\mathbf{1}_n^+ + \frac{m}{m+n}\big(\mathbf{1}_m^+ \otimes (\mathbf{1}_n\mathbf{1}_n^+)\big) - \frac{m}{m+n}\big(\mathbf{1}_m^+ \otimes (\mathbf{1}_n\mathbf{1}_n^+)\big)\big((\mathbf{1}_m\mathbf{1}_m^+) \otimes (I_n - \mathbf{1}_n\mathbf{1}_n^+)\big) \\
&= \mathbf{1}_m^+ \otimes I_n - \mathbf{1}_m^+ \otimes \mathbf{1}_n\mathbf{1}_n^+ + \frac{m}{m+n}\big(\mathbf{1}_m^+ \otimes (\mathbf{1}_n\mathbf{1}_n^+)\big) = \mathbf{1}_m^+ \otimes I_n - \frac{n}{m+n}\big(\mathbf{1}_m^+ \otimes (\mathbf{1}_n\mathbf{1}_n^+)\big) \\
&= \mathbf{1}_m^+ \otimes \Big(I_n - \frac{n}{m+n}\mathbf{1}_n\mathbf{1}_n^+\Big) .
\end{aligned}$$

So

$$\big[\, I_m \otimes \mathbf{1}_n \big| \mathbf{1}_m \otimes I_n \,\big]^+ = \left[\frac{(I_m \otimes \mathbf{1}_n)^+ - (I_m \otimes \mathbf{1}_n)^+(\mathbf{1}_m \otimes I_n)J}{J}\right]$$

$$= \left[\frac{I_m \otimes \mathbf{1}_n^+ - (\mathbf{1}_m \otimes \mathbf{1}_n^+)J}{J}\right] = \left[\frac{I_m \otimes \mathbf{1}_n^+ - (\mathbf{1}_m \otimes \mathbf{1}_n^+)\mathbf{1}_m^+ \otimes \Big(I_n - \frac{n}{m+n}\mathbf{1}_n\mathbf{1}_n^+\Big)}{\mathbf{1}_m^+ \otimes \Big(I_n - \frac{n}{m+n}\mathbf{1}_n\mathbf{1}_n^+\Big)}\right]$$

$$= \left[\frac{I_m \otimes \mathbf{1}_n^+ - \mathbf{1}_m\mathbf{1}_m^+ \otimes \mathbf{1}_n^+\Big(I_n - \frac{n}{m+n}\mathbf{1}_n\mathbf{1}_n^+\Big)}{\mathbf{1}_m^+ \otimes \Big(I_n - \frac{n}{m+n}\mathbf{1}_n\mathbf{1}_n^+\Big)}\right] = \left[\frac{I_m \otimes \mathbf{1}_n^+ - \mathbf{1}_m\mathbf{1}_m^+ \otimes \Big(\mathbf{1}_n^+ - \frac{n}{m+n}\mathbf{1}_n^+\Big)}{\mathbf{1}_m^+ \otimes \Big(I_n - \frac{n}{m+n}\mathbf{1}_n\mathbf{1}_n^+\Big)}\right]$$

$$= \left[\frac{I_m \otimes \mathbf{1}_n^+ - \mathbf{1}_m\mathbf{1}_m^+ \otimes \Big(\frac{m}{m+n}\mathbf{1}_n^+\Big)}{\mathbf{1}_m^+ \otimes \Big(I_n - \frac{n}{m+n}\mathbf{1}_n\mathbf{1}_n^+\Big)}\right] = \left[\frac{\Big(I_m - \frac{m}{m+n}\mathbf{1}_m\mathbf{1}_m^+\Big) \otimes \mathbf{1}_n^+}{\mathbf{1}_m^+ \otimes \Big(I_n - \frac{n}{m+n}\mathbf{1}_n\mathbf{1}_n^+\Big)}\right] = \left[\frac{I_{m,n} \otimes \mathbf{1}_n^+}{\mathbf{1}_m^+ \otimes I_{n,m}}\right] .$$

where

$$I_{m,n} = I_m - \frac{m}{m+n}\mathbf{1}_m\mathbf{1}_m^+ \quad \text{and} \quad I_{n,m} = I_n - \frac{n}{m+n}\mathbf{1}_n\mathbf{1}_n^+$$

Bibliography

[1] J. Abdullah. *Fixed Point Algorithms for Linear Programming.* PhD thesis, Department of Economics, University of Birmingham, England, May 1992. iv, 48

[2] S.N. Afriat. On the latent vectors and characteristic values of products of pairs of symmetric idempotents. *Quarterly Journal of Mathematics*, 7(2):76–78, 1956. 102

[3] S.N. Afriat. Orthogonal and oblique projectors and the characteristics of pairs of vector spaces. *Proceedings of the Cambridge Philosophical Society*, 53:800–816, 1957. 31, 37

[4] A.A. Albert. *Regression and the Moore-Penrose Pseudoinverse.* Academic Press, New York, 1972. 11, 14, 111

[5] C.D. Aliprantis and O. Burkinshaw. *Principles of Real Analysis (Second Edition).* Academic Press, New York, 1990. 41, 42, 43

[6] A.J. Bruni. *Nonnegative, Nontrivial Fixed Points of Orthogonal Projections.* PhD thesis, Department of Mathematics, University of Houston, 1985. 58

[7] A.H. Clifford and G.B. Preston. *The Algebraic Theory of Semigroups*, volume 1. The American Mathematical Society, 1961. 6

[8] G. B. Dantzig. Programming in a linear structure. Technical report, U.S. Air Force Comptroller, USAF, Washington, D.C., 1948. 4

[9] G.B. Dantzig. Maximization of a linear function of variables subject to linear inequalities. In Tj.C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 339–347. Wiley, New York, 1951. 3

[10] Joseph G. Ecker and Michael Kupferschmid. *Introduction to Operations Research.* John Wiley & Sons, 1988. 87, 99

[11] J.B.J. Fourier. reported in *Analyse des travaux de l'Academie Royale des Sciences, pendant l'annee 1824.* 3

[12] A.R.G. Heesterman. *Matrices and Simplex Algorithms.* D. Reidel Pub. Co., 1982. 4

[13] Robert C. James and Glenn James, editors. *Mathematics Dictionary (Fourth Edition)*. Van Nostrand, 1976. 77

[14] Paul A. Jensen and Jonathan F. Bard. *Operations Research Models and Methods*. Wiley, New York, 2002. 47

[15] W. Kahan. The reduced row-echelon form is unique. *The Web*, 1998. 16

[16] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984. 4

[17] L.G. Khachian. A polynomial algorithm in linear programming. *Soviet Mathematics Doklady*, 20:191–194, 1979. 4

[18] V. Klee and G.J. Minty. How good is the simplex algorithm? In O. Shisha, editor, *Inequalities III*. Academic Press, New York, 1972. 4

[19] H.W. Kuhn. Solvability and consistency for linear equations and inequalities. *American Mathematical Monthly*, 63:217–232, 1956. iii, 3

[20] J.L. de la Grange. *Mechanique analitique*. Vve Desaint, Paris, 1788. 2

[21] M.A. Naimark. *Normed Rings*. P. Noordhoff N. V., Groningen, The Netherlands, 1964. Translated from the first Russian edition, by Leo F. Boron. 10

[22] Stephen G. Nash. SUMT (revisited). Web, July 22 1997. Operations Research and Engineering Dept., George Mason University, Fairfax, VA 22030. 100

[23] J. Von Neumann and O. Morgenstern. *The Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, New Jersey, U.S.A., 1944. iii

[24] T.M. Nguyen. *Applications of Generalized Inverse to Circulant Matrices, Intersection Projections, and Linear Programming*. PhD thesis, University of Houston, 1982. 69

[25] Carl E. Pearson, editor. *Handbook of Applied Mathematics, Selected Results and Methods*. Van Nostrand Reinhold, New York, second edition, 1983. 19

[26] Ch. de la Vallée Poussin. Sur la methode de l'approximation minimum. *Annales de la Societe Scientifique de Bruxelles*, 35(2), 1910–11. 3

[27] L.D. Pyle. The generalized inverse in linear programming - basic structure. *SIAM Journal on Applied Mathematics*, 23(3), 1972. 58

[28] L.D. Pyle and R.E. Cline. The generalized inverse in linear programming - interior gradient projection methods. *SIAM Journal on Applied Mathematics*, 24(4), 1973. 58

[29] A. Ravindran, Don T. Phillips, and James J. Solberg. *Operations Research Principles and Practice*. John Wiley and Sons, 1987. 77

[30] James Renegar. A polynomial-time algorithm, based on newton's method, for linear programming. *Mathematical Programming*, 40:59–93, 1988. 4

[31] Charles E. Rickart. *General Theory of Banach Algebras.* Van Nostrand, Princeton NJ, 1960. 10

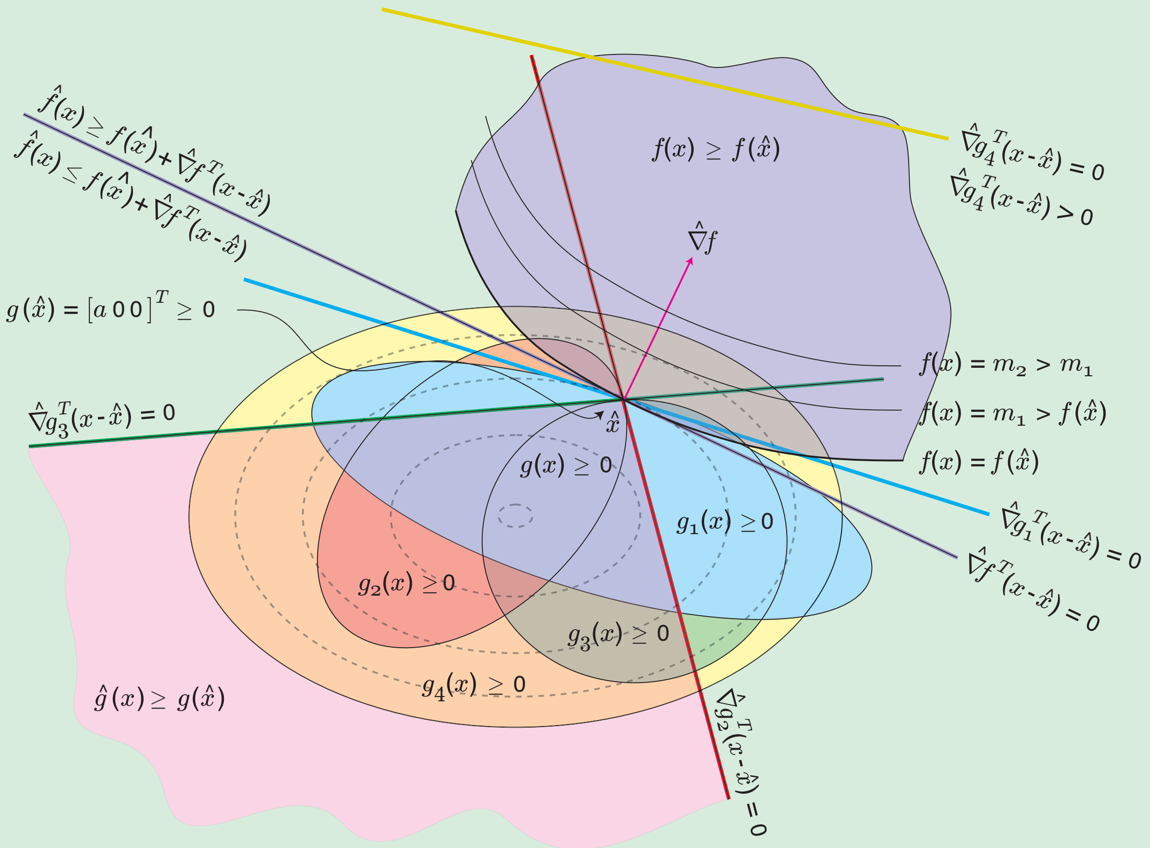
[32] J.B. Rosen. The gradient projection method for non-linear programming. *Journal of the Society for Industrial and Applied Mathematics*, 8(1):181–217, 1960. 58

[33] A. Schrijver. *Theory of Linear and Integer Programming.* Wiley, New York, 1986. 2, 47

[34] Josef Stoer and Cristoph Witzgall. *Convexity and Optimization in Finite Dimensions I.* Springer-Verlag, Berlin, 1970. 3

[35] Wikipedia. Lagrange multipliers. https://en.wikipedia.org/wiki/Lagrange_multipliers. 25

Index



Note: At $\hat{x}$ the affine plane $\hat{\nabla} g_4^T(x-\hat{x}) = 0$ lies above the diagram plane, sloping down towards top right, where it intersects the diagram plane.