

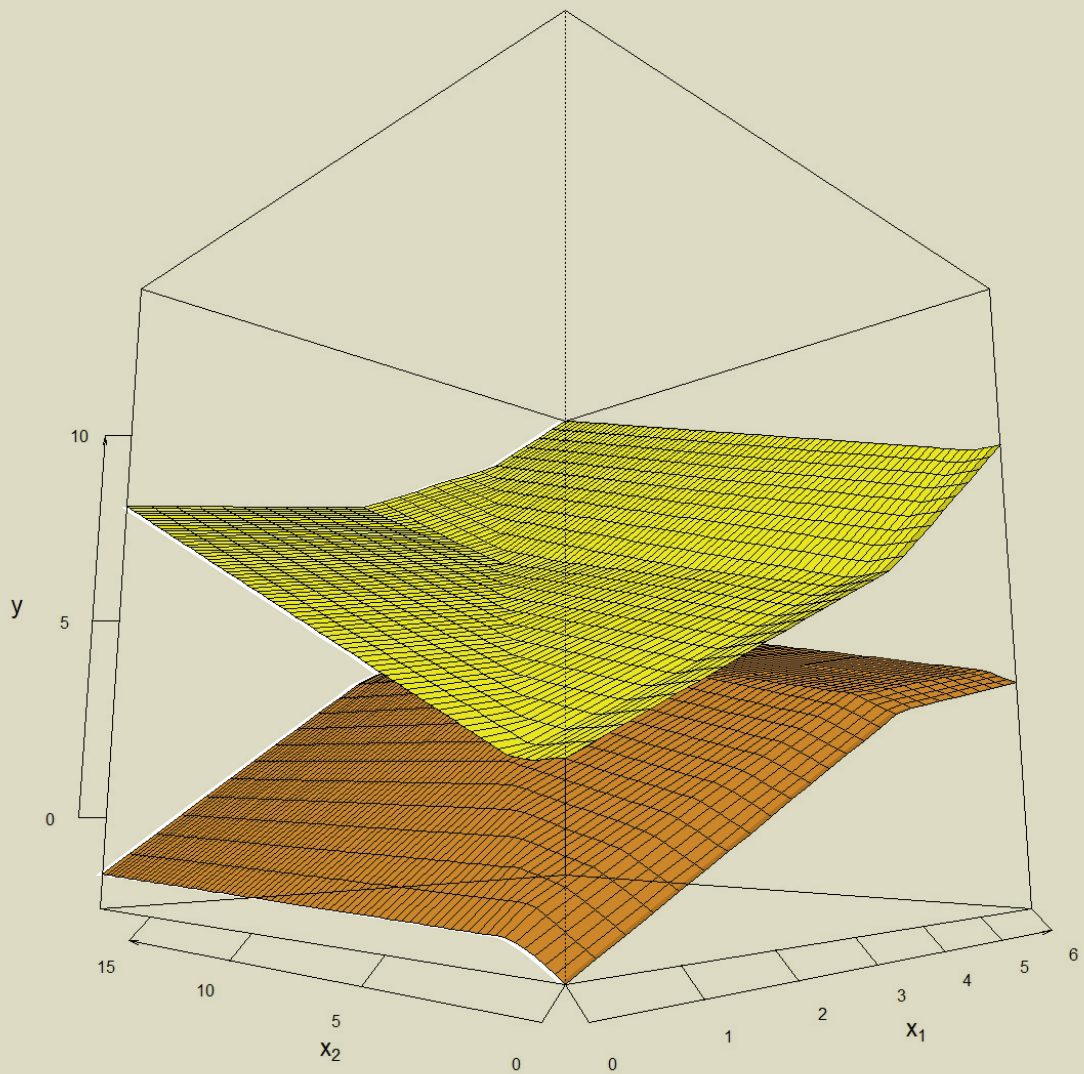


Chulalongkorn University Press

Operations Research

with Statistical Applications

Seksan Kiatsupaibul



Operations Research with Statistical Applications

Operations Research with Statistical Applications

Seksan Kiatsupaibul



Chulalongkorn University Press

2018

300.-

Seksan Kiatsupaibul

Operations Research with Statistical Applications / Seksan Kiatsupaibul

1. Operations research. 2. Operations research – Statistical methods.

658.4034

ISBN 978-974-03-3834-5

CUP 2357



www.ChulaPress.com

Knowledge to All

Copyright © 2018, by the Chulalongkorn University Press

First Printing 2018 400 Copies

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the publisher.

Distributed by Chulalongkorn University Book Center

Phayathai Road, Pathumwan, Bangkok 10330, Thailand

Branch Sala Prakiew Tel. 0-2218-7000-3 Fax. 0-2255-4441

Siam Square Tel. 0-2218-9881-2 Fax. 0-2254-9495

Naresuan University (Phitsanulok) Tel. 0-5526-0162-4 Fax. 0-5526-0165

Suranaree University (Nakhonratchasima) Tel. 0-4421-6131-4 Fax. 0-4421-6135

Burapha University (Chonburi) Tel. 0-3839-4855-9 Fax. 0-3839-3239

Chulachomklao Royal Military Academy (Nakhon-Nayok) Tel. 0-3739-3023 Fax. 0-3739-3023

University of Phayao Tel. 0-5446-6799-800 Fax. 0-5446-6798

Chamchuri Square (4th Floor) Tel. 0-2160-5300-1 Fax. 0-2160-5304

Huamark Tel. 0-2374-1378 Fax. 0-2374-1377

Rajamangala University of Technology Isan Tel.0-4423-0000 Fax.0-4423-3052

Call Center Tel. 0-2255-4433 <http://chulabook.com> and Networks

Wholesale Huamark (Intersection Huamark) Tel. 0-2374-1374-5 Fax. 0-2374-1377

Editorial Department: Thipawan Holasut

Artwork: Seksan Kiatsupabul

Cover: Khomsan Kiatsupaibul

Printed by: Chulalongkorn University Printing House Tel. 0-2218-3549-50 Fax. 0-2218-3551

Preface

This book has been developed from lecture notes for a course in operations research at the Department of Statistics, Chulalongkorn Business School. The objective is to provide the audience with basic tools to analyze decision making problems under uncertainty. Since the group of statistics students is a primary target of this book, a number of examples in statistical applications are also included.

The presentation in each chapter usually starts with a simple decision making case study, followed by a sequence of logical arguments that leads to solution methods. Each case study is kept in its simplest form and does not attempt to be over realistic. This style of presentation is designed so that the audience can grasp the problem structure as quickly as possible, then use it to induce an insight into the solution method. More efforts are spent on the solution methods. The case studies also play a key role as tokens that remind the audience of the solution concepts.

The first three chapters, as in other operations research books, are devoted to linear programming formulations. A linear program is very rich in its structure and has always been a good place to start. Most mathematical programming concepts are introduced in this chapter and real world applications are related. The simplex method, a notable practical solution method for a linear program, is postponed to Chapter 4. This is done so that the author can have more space to elaborate and put more emphasis on the solution aspects of operations research. Chapter 5 introduces the concept of duality in linear programming. In Chapter 6, the presentation departs from a traditional operations research textbook by discussing an application of linear programming to statistics. Quantile regression is specifically discussed. In Chapter 7, uncertainty is introduced into decision making problems. In this chapter, the author shows how linear programming, deterministic in its nature, can address the problem with uncertainty. A quadratic program in a form of a Markowitz model is also introduced as an alternative to addressing uncertainty. In Chapter 8, convex programming and its solution techniques are introduced. Examples on how to fit a logistic regression model in statistics by convex programming solution techniques are given. Chapters 9 and 10 are dedicated to sequential decision making problems. Chapter 9 discusses deterministic sequential decision models. Chapter 10 deals with stochastic sequential decision models. Chapters 11 and 12 discuss some stochastic systems. Chapter 11 introduces basic queueing theory and sets the stage for the simulation techniques in Chapter 12. In the appendix, the prerequisites on vector

calculus and probability are provided.

The author would like to thank Professor Robert L. Smith for his guidance in operations research over the years at the University of Michigan and up until now. The author would also like to thank Professor Anthony Hayter who provided insightful comments regarding applications of operations research to problems in his research collaboration with the author. Professor Hayter also helped read through the manuscript, for which the author is very grateful. The early draft of this text has been used in the fourth year undergraduate classes at the Department of Statistics, Chulalongkorn University with graduate students as the teaching assistants. In particular, two graduate students, Pariyakorn Maneekul and Chutimon Sindhuprama, have been very helpful in reviewing the text and checking exercises. Special thanks to my brother, Khomsan Kiatsupaibul, who helped design the book cover. The author thanks Chulalongkorn Business School for its remarkable working environment. It was the main factor that made this small project possible. Any shortcomings of this book belong to the author alone.

Seksan Kiatsupaibul
December, 2018
Chulalongkorn Business School
Bangkok, Thailand

Contents

Preface

1	Introduction to Operations Research	1
1.1	Introduction	1
1.2	Operations Research	1
1.3	Optimization Model	2
1.4	A Case Study	4
2	Linear Programming	7
2.1	Introduction	7
2.2	A Case Study	7
2.3	A Linear Programming Model	8
2.3.1	Identify the Alternatives	9
2.3.2	Identify the Objective	10
2.3.3	Identify the Constraints	10
2.3.4	Linear Programming Model	12
2.4	Assumptions of a Linear Programming Model	14
2.5	Solution: Graphical Methods	14
2.6	Sensitivity Analysis	20
2.6.1	Changes in the Objective Function	21
2.6.2	Changes in the RHS	22
2.7	Reduced Cost	23
2.8	More Case Studies	24
3	Transformations in Linear Programming	31
3.1	Introduction	31
3.2	The Standard Form of a Linear Program	31
3.3	Transformation of Some Special Problems to LP	36
3.4	Notes	43

4	Solution Methods for Linear Programs	45
4.1	Introduction	45
4.2	Configuration of an Optimal Solution	45
4.3	Simplex Method	55
4.3.1	Checking for Optimality	55
4.3.2	Moving from the Current BFS to a Better BFS	57
4.3.3	Algorithm	59
4.4	Two Phase Method	62
4.5	Notes	64
5	Duality in Linear Programming	67
5.1	Introduction	67
5.2	Two-person Zero-sum Matrix Games	67
5.3	Linear Programming for a Matrix Game	68
5.4	Duality in Linear Programming	70
5.5	Complementary Slackness	74
5.6	Notes	75
6	Optimization Concepts in Linear Regression Modeling	77
6.1	Introduction	77
6.2	Stock Return Behavior	77
6.3	Data Fitting Models	78
6.3.1	Least Squares Model	80
6.3.2	Least Absolute Deviations Model	81
6.4	Statistics and Optimization	83
6.5	Quantile Regression Model	87
7	Optimization Under Uncertainty	91
7.1	Introduction	91
7.2	A Case Study	91
7.3	Without Uncertainty	93
7.4	Scenario Based Model	94
7.5	Risk Minimization Model	98
7.6	Notes	100
8	Convex Programming	105
8.1	Introduction	105
8.2	Optimality Condition	106
8.3	Unconstrained Convex Program	107
8.4	Solution Methods for Unconstrained Programs	110
8.4.1	Gradient Methods	110
8.4.2	Newton's Methods	111

8.4.3	Gradient Methods Versus Newton's Methods	115
8.5	Equality Constrained Convex Optimization	115
8.6	Inequality Constrained Convex Optimization	118
8.7	Notes	120
9	Dynamic Programming	127
9.1	Introduction	127
9.2	Shortest Path Problems	127
9.3	Critical Path Method	134
9.4	Dynamic Programming	142
9.5	Formal Treatment of Dynamic Programming	147
9.6	Some Dynamic Programming Examples	151
9.7	The Curse of Dimensionality	154
10	Stochastic Sequential Decision Models	159
10.1	Introduction	159
10.2	Stochastic Decision Trees	159
10.3	Bayesian Decision Analysis	165
10.4	Markov Decision Process	168
10.5	The Economic Order Quantity Model	169
10.6	Solving the Water Festival Case	172
10.7	Markov Decision Model	178
10.8	Notes	180
11	Queuing Theory	185
11.1	Introduction	185
11.2	Waiting Line Costs and a Decision Model	185
11.3	Queuing Model	187
11.4	Comments on Simple Queues	200
11.5	Multi-Server Queuing System	201
11.6	Notes	202
12	Statistical Simulation	205
12.1	Introduction	205
12.2	A Case Study	205
12.3	A Mathematical Model	206
12.4	Pseudo-Random Number	207
12.5	Simulation of the VR Theater	209
12.6	Simulation of a Random Variable	212
12.7	Simulation of Joint Random Variables	214
12.7.1	The Rejection Method	214
12.7.2	Hit-and-run Algorithms	215

12.7.3 Simulate Joint Normal Random Variables	217
12.8 Simulation Process	218
12.9 Simulation of $M/U/1$ Queue	219
12.10 Simulation in Optimization	221
12.10.1 Pure Random Search	222
12.10.2 Simulation Method for MDP	224
12.11 Notes	227
A Convexity, Vector Calculus and Related Concepts	231
A.1 Introduction	231
A.2 Convexity	231
A.3 Vector Calculus	235
A.4 Line Search Algorithms	238
A.4.1 Leibniz's Rule	244
B Probability	245
B.1 Introduction	245
B.2 Probability Space	245
B.2.1 Sample Space	246
B.2.2 Events and Event Set	246
B.2.3 Probability	248
B.3 Probability Assignment	250
B.3.1 Uniform Probability	250
B.3.2 Independence	251
B.4 Random Variables	252
B.4.1 Definition and Simple Examples	252
B.4.2 Probability Mass Function (P.M.F)	254
B.4.3 Cumulative Distribution Function (C.D.F)	257
B.5 Expectation and Variance	259
B.6 Continuous Random Variables	261
B.7 Basic Probability Distributions	263
B.7.1 Discrete Probability Distributions	264
B.7.2 Continuous Probability Distributions	265
B.8 Joint Random Variables	266
B.9 Independent Random Variables	270
B.10 Covariance	271
B.11 Sum of Random Variables	272
B.12 Conditional Probability and Expectation	276
B.13 Stochastic Processes	278
B.14 Statistical Forecasting	279
Bibliography	283

Index	286
Author Bio	291

List of Figures

2.1	A solution represented by a point.	15
2.2	An equality constraint represented by a straight line.	16
2.3	An strict inequality constraint represented by an area.	16
2.4	An inequality constraint represented by a half plane.	17
2.5	The feasible region.	18
2.6	An objective line at the value of 5,000,000.	19
2.7	Solving a linear program by graphical method.	19
2.8	Changes of a coefficient of the objective function.	21
3.1	Panel (a) and Panel (b) illustrate, respectively, a convex and a concave piecewise linear functions. Both functions contain two breakpoints x_j^1 and x_j^2 where $j = 1, 2$	41
4.1	A BFS corresponds to a corner of the feasible region.	54
4.2	Geometric interpretation of the simplex method.	62
6.1	A scatter plot between the profit growths and the stock returns of 38 of the largest companies listed in Thailand stock market in 2016.	79
6.2	Two fitted lines to the scatter plot between the profit growths and the stock returns of 38 large companies listed in Thailand's stock market in 2016.	83
6.3	Three fitted lines to the scatter plot between the profit growths and the stock returns, based on the quantile regression models.	88
8.1	A contour plot of the negative log-likelihood function of the logistic regression model fitted to the data set in Table 8.1. The zigzag solid line illustrates the path of the solutions generated by the gradient method, starting with $\beta_0 = [0, 0]^T$ and ending at the optimal solution $\beta^* = [6.03, -0.19]^T$	113
8.2	The solid line illustrates the path of the solutions generated by Newton's method, starting at $\beta_0 = [0, 0]^T$ and ending at the optimal solution $\beta^* = [6.03, -0.19]^T$	114

8.3	The solid line illustrates the path of the solutions generated by the barrier method, starting at $\beta_0 = [1.00, -0.01]^T$ and passing through each optimal solution with respect to the increasing t parameter from 10 to 1,000,000. The rectangle given by the thick line is the feasible region defined by the four regularization constraints in (8.24).	121
8.4	A non-convex multi-modal function on the interval $[0, 1]$ with multiple local optimal solutions.	123
9.1	Diamond network.	128
9.2	A path in the diamond network.	129
9.3	Illustration of the principle of optimality.	129
9.4	Backward recursive, the first step.	130
9.5	Backward recursive, the second step.	131
9.6	Backward recursive, the third step.	131
9.7	Backward recursive, the final step.	132
9.8	A shortest path in the diamond network.	133
9.9	Groups of nodes that require the same number of basic operations.	133
9.10	Acyclic network for the project scheduling problem.	135
9.11	An application of the forward recursive to the project scheduling network.	135
9.12	An application of the backward recursive to the project scheduling network.	136
9.13	Subtraction from the project completion time.	136
9.14	Project completion time when the completion time of task C is reduced from 5 to 4.	138
9.15	The completion time cannot be reduced further from 15 by task C time reduction.	138
9.16	An acyclic network representing the online organizational survey project detailed in Table 9.5.	140
9.17	Solutions to the online organizational survey CPM.	141
9.18	Decision tree for the equipment replacement problem.	143
9.19	Dynamic programming network for the equipment replacement problem.	145
9.20	Decision tree before node aggregation.	147
9.21	Dynamic programming network after node aggregation.	148
9.22	Solution to the dynamic programming network of the equipment replacement problem.	150
9.23	Decision tree for the resource allocation problem.	152
9.24	Dynamic programming network for the resource allocation problem.	152
9.25	Dynamic programming network for the streamlined equipment replacement model.	154
9.26	An illustration of the dynamic programming network for the traveling salesman problem. It demonstrates the state counts.	155

10.1	Decision tree for the technological change case study when there is no uncertainty.	161
10.2	Stochastic decision tree for the technological change case study.	162
10.3	Solution to the stochastic decision tree for the technological change case study.	163
10.4	Decision tree for clean technology A and B	166
10.5	Decision tree for clean technology A and B with a market survey.	166
10.6	Bayesian decision tree for clean technology A and B with a market survey.	168
10.7	Dynamics of the economic order quantity model.	170
10.8	The trade-off between the holding cost and the ordering cost in the economic order quantity model.	171
10.9	An MDP acyclic network diagram for the water festival problem.	173
10.10	An MDP acyclic network diagram for the water festival problem with all transition probability.	175
10.11	A completed MDP acyclic network diagram for the water festival problem.	177
10.12	A general MDP acyclic network diagram.	179
11.1	Trade-off between the service cost and the waiting cost.	186
11.2	Simple queue diagram.	188
11.3	Number of arrivals versus time.	195
11.4	Number of departures versus time.	195
11.5	Boxes represent periods of customers staying in the system.	196
11.6	Boxes represent number of customers in the system in each unit of time.	196
11.7	Two feedback queue policies.	204
12.1	Illustration of transformation a random number by a c.d.f.	212
12.2	Points drawn uniformly on the square $[-1, 1] \times [-1, 1]$. The ones in the unit ball centered at the origin are uniformly distributed on the ball.	216
12.3	A typical simulation process.	219
12.4	A simple MDP to illustrate the use of simulation to solve a stochastic sequential decision problem.	225
A.1	Convex sets.	232
A.2	Non-convex sets.	232
A.3	Convex functions.	233
A.4	Concave functions.	233
A.5	Depending on $f(\mu)$, either the subinterval $[a, \lambda)$ or the subinterval $(\mu, b]$ will be removed in the next iteration. The golden ratio $\alpha^* = x/y = (-1 + \sqrt{5})/2$ keeps the ratio of the remaining subintervals constant, regardless of which subinterval will be removed in the next iteration, i.e. $z/(x - z) = (y - x)/x = z/(y - x)$	241

A.6	Approximate function with $\alpha = 0.3$ for an inexact line search algorithm according to Armijo's Rule.	244
B.1	$p_X(k)$ representation by a graph.	256
B.2	The cumulative distribution (c.d.f.) $F_X(x)$ in Example B.18 represented by a graph.	258
B.3	A balance beam representing an expectation.	259
B.4	Area under p.d.f of X is the probability.	262
B.5	Plot of time series data of SET index during December 2009.	281

Chapter 1

Introduction to Operations Research

1.1 Introduction

This book is written for an introductory course in operations research. Students will learn about optimization techniques for solving decision making problem. In addition, fitting a statistical model is here categorized as a decision problem that we would also like to solve.

Students should be aware that there exists an interplay between operations research and statistics. Operations research requires probability and statistics for modeling and analyzing uncertainty in its decision problems. In turn, statistics utilizes optimization techniques in operations research to fit its models. In this book, we also investigate this interplay between operations research and statistics.

The topics covered in this book include optimization models, linear programming, convex programming, dynamic programming, stochastic systems and simulation. In general, the presentation starts with a problem or a case study. Then a relevant operations research model is introduced to describe the problem, followed by model analyses and solution techniques. The presentation ends with a conclusion and insights obtained from the solution.

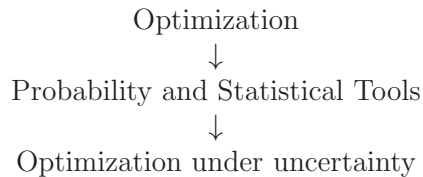
Students are encouraged to use a spreadsheet program such as Excel to implement the models discussed. A hands-on experience building a spreadsheet model to solve a problem will enhance the students' understanding of the subject.

1.2 Operations Research

Operations research is a discipline that emphasizes developing and employing quantitative techniques to solve decision making problems. In the past, such quantitative analysis was restricted to problems with simple structures, due to limited computational resources. However, today information technology opens ways for computationally intensive quantitative techniques to be implemented. Novel computationally intensive tools have been

developed to take advantage of these newly available resources. As a result, many complex problems that in the past were considered intractable have been re-considered and solved. For students, today easy-to-use spreadsheet programs, such as Excel, have these quantitative tools built into their system. Solving a complex problem is no longer out of reach.

The ultimate goal of this course is for students to be able to make a practical decision in the face of uncertainty. To achieve this goal, optimization techniques for decision models without uncertainty, the so-called deterministic decision model, are first introduced. The element of uncertainty is subsequently added through stochastic modeling. In combination, we obtain more sophisticated models and techniques to solve stochastic decision making problems that are more practical and realistic. The diagram below summarizes this process of development in this book.



Optimization is the template of our decision making problems. When we make a decision, we select the one that provides the best result.

1.3 Optimization Model

A decision problem is composed of *choices* and an *objective*. The *objective* refers to a decision criterion or a performance measure given a choice. The choices are bounded by constraints or policies.

Example 1.1 *A company has two projects - project A Go Big and project B Play Safe. The management has to choose the one that offers the highest total profit.*

Choices : Project A or B
Objective : Total profit
Constraints : Between only A and B

Here we decide between two choices.

Example 1.2 *A production manager decides the number of units of a product to be produced in order to meet demand within the capacity and specification at the lowest total cost.*

Choices : Producing 1,000 units, 1,001 units, ... of a product

Objective : Total cost

Constraints : Meet demand, within the limited resources

In this example, we decide among thousands of choices.

Each choice is associated with one value of the objective. In Example 1.1, suppose Project A offers THB 10 million of profit, while Project B offers THB 7.5 million of profit. The objectives corresponding to Project A and Project B are THB 10 million and THB 7.5 million, respectively. In Example 1.2, suppose producing 1,000 units incurs the cost of THB 10,000, while producing 2,000 units incurs THB 15,000. The objectives corresponding to producing 1,000 units and producing 2,000 units are THB 10,000 and THB 15,000, respectively.

Note that the objective defined for each choice must be a number, enabling us to compare the objectives among choices. To see the necessity of this restriction, consider the following situation. In a production line, suppose we are choosing between two technologies to be implemented in our manufacturing process. We want to produce a good quality product at a low cost. Both quality and cost are measured in grades: 1 (good), 2 (average), 3 (bad). Suppose we have only technology *A* and technology *B* to choose from. Technology *A*'s quality grade is 1, the cost grade is 3. Technology *B*'s quality is 3, the cost is 1. If we take both quality and cost as our objective, which technology would we choose? This simple example shows the necessity of having the objective as a scalar. For example, if we adopt the quality as the objective, we can then easily choose *A* as the technology of choice.

In case we have more than one performance measure to consider, a possible way is to select one as the objective and put the rest into constraints. Another way is to combine all of those measures into one measure with a mathematical transformation, such as an average or a linear combination.

The decision model discussed above can be put into the following template.

$$\begin{array}{ll} \text{Optimize the Objective} & \\ \text{s.t. Choices within Constraints} & \end{array} \quad (1.1)$$

The notation *s.t.* abbreviates *subject to*. Optimization can be either maximization or minimization. In what follows, the notations max and min stand for *maximize* and *minimize*, respectively.

Example 1.3 *From Example 1.1, we can write*

$$\begin{array}{ll} \max & \text{Profit of a project} \\ \text{s.t.} & \text{Choosing between project A and project B} \end{array}$$

From Example 1.2, we can write

$$\begin{array}{ll} \min & \text{Cost of producing } x \text{ units of a product} \\ \text{s.t.} & \text{Producing } x \text{ units such that} \\ & (i) \text{ meet demand} \\ & (ii) \text{ meet capacity} \end{array}$$

To analyze the problem quantitatively, we have to transform all elements in the above template into numbers. Since the objective is already a number, we only need to transform the choices. In a problem setup, since an optimal choice has not yet been discovered, it is a variable(s) called a *decision variable(s)*. A constraint is a mathematical expression that associates the decision variables to satisfy a policy or a limitation. The following case study will elaborate this point.

1.4 A Case Study

Case 1.1 (Puppy Universal) *Puppy Universal, a manufacturer producing pet foods, plans to launch a new product, aiming at a premium market. The plan requires new investment of THB 50,000 on production fundamentals. To produce one unit of the new product, THB 1.0 is required as a production cost. The product can be sold at THB 2.0 per unit. If the plan succeeds, it will enhance the share of the company in the market of premium grade products. The company wants to decide a suitable production quantity.*

Since the management wants to know the most suitable quantity, the choices are obvious: units of the product to be produced. Assume that we have only three choices, producing 10,000 units, 50,000 units and 100,000 units. Which one is the best? To put it into a mathematical notation, let Q represent a possible choice, called decision variable. The variable can take value of 10,000, 50,000 or 100,000. The problem is to find the best Q .

To determine the best choice, we need an objective. As a business unit, a sensible objective is profit.

$$\max \quad \text{Profit} = (2 - 1)Q - 50,000$$

From this problem setup, the best choice is to produce 100,000 units which will give 150,000 profit.

However, the possible choices Q are not necessarily limited by those three numbers. It can be 0, 1, 2, 3 and so on. In such a case the number of choices are infinite. If we can

produce in subunits, Q can be $0.1, 0.2, \dots$, which makes the number of choices becomes even larger. In general, we may allow any real number to be our possible choice.

One usually needs policies to reflect physical and economic limitations. Let c be the highest number of units the production line can accommodate, and let s be the expected demand in the market. We then have the following limitations.

$$\begin{aligned} Q &\leq c \\ Q &\leq s \end{aligned}$$

Another sensible policy can be that the profit gained from selling the product must cover the fixed cost. Therefore,

$$(2 - 1)Q \geq 50,000$$

or

$$Q \geq 50,000$$

These are the constraints that were mentioned earlier.

Combining all the elements together, we can write the problem mathematically as follows:

Program 1.1

$$\begin{aligned} \max \quad & (2 - 1)Q - 50,000 \\ \text{s.t.} \quad & (2 - 1)Q \geq 50,000 \\ & Q \leq c \\ & Q \leq s \\ & Q \geq 0 \end{aligned}$$

Note the last constraint $Q \geq 0$. This is called a non-negativity constraint that imposes another physical limitation that we cannot produce a negative number of products.

The optimization model written in this mathematical form is called a *mathematical program*. (The term “program” was used in the military community to mean a schedule to be met.) Some readers might recognize that Program 1.1 has a lot of redundancy, but we do not concern ourselves with this at the moment.

To summarize, Q represents choices, called the decision variable. The objective here is the profit, which is a function of Q . Observe that it is a linear function. The constraints are the limitations imposed on the decision maker. They are expressions written in terms of Q . The constraints we have seen are divided into two parts, *the Left hand side (LHS)* and *the Right Hand Side (RHS)*. LHS relates Q to a quantity representing a measure. RHS is the limitation on that measure. Observe that LHS is also a linear function. This specific mathematical program is called a linear program. Optimization problems do not have to be linear, but problems that are linear are considered to be a