STOCHASTIC LINEAR PROGRAMMING
Models, Theory, and Computation

PETER KALL
University of Zürich/Switzerland

JÁNOS MAYER
University of Zürich/Switzerland

Recent titles in the

INTERNATIONAL SERIES IN OPERATIONS RESEARCH & MANAGEMENT SCIENCE

Frederick S. Hillier, Series Editor, *Stanford University*

Zhu/ *QUANTITATIVE MODELS FOR PERFORMANCE EVALUATION AND BENCHMARKING*

Ehrgott & Gandibleux/ *MULTIPLE CRITERIA OPTIMIZATION: State of the Art Annotated Bibliographical Surveys*

Bienstock/ *Potential Function Methods for Approx. Solving Linear Programming Problems*

Matsatsinis & Siskos/ *INTELLIGENT SUPPORT SYSTEMS FOR MARKETING DECISIONS*

Alpern & Gal/ *THE THEORY OF SEARCH GAMES AND RENDEZVOUS*

Hall/*HANDBOOK OF TRANSPORTATION SCIENCE - 2nd Ed.*

Glover & Kochenberger/ *HANDBOOK OF METAHEURISTICS*

Graves & Ringuest/ *MODELS AND METHODS FOR PROJECT SELECTION: Concepts from Management Science, Finance and Information Technology*

Hassin & Haviv/ *TO QUEUE OR NOT TO QUEUE: Equilibrium Behavior in Queueing Systems*

Gershwin et al/ *ANALYSIS & MODELING OF MANUFACTURING SYSTEMS*

Maros/ *COMPUTATIONAL TECHNIQUES OF THE SIMPLEX METHOD*

Harrison, Lee & Neale/ *THE PRACTICE OF SUPPLY CHAIN MANAGEMENT: Where Theory and Application Converge*

Shanthikumar, Yao & Zijm/ *STOCHASTIC MODELING AND OPTIMIZATION OF MANUFACTURING SYSTEMS AND SUPPLY CHAINS*

Nabrzyski, Schopf & Węglarz/ *GRID RESOURCE MANAGEMENT: State of the Art and Future Trends*

Thissen & Herder/ *CRITICAL INFRASTRUCTURES: State of the Art in Research and Application*

Carlsson, Fedrizzi, & Fullér/ *FUZZY LOGIC IN MANAGEMENT*

Soyer, Mazzuchi & Singpurwalla/ *MATHEMATICAL RELIABILITY: An Expository Perspective*

Chakravarty & Eliashberg/ *MANAGING BUSINESS INTERFACES: Marketing, Engineering, and Manufacturing Perspectives*

Talluri & van Ryzin/ *THE THEORY AND PRACTICE OF REVENUE MANAGEMENT*

Kavadias & Loch/*PROJECT SELECTION UNDER UNCERTAINTY: Dynamically Allocating Resources to Maximize Value*

Brandeau, Sainfort & Pierskalla/ *OPERATIONS RESEARCH AND HEALTH CARE: A Handbook of Methods and Applications*

Cooper, Seiford & Zhu/ *HANDBOOK OF DATA ENVELOPMENT ANALYSIS: Models and Methods*

Luenberger/ *LINEAR AND NONLINEAR PROGRAMMING, 2nd Ed.*

Sherbrooke/ *OPTIMAL INVENTORY MODELING OF SYSTEMS: Multi-Echelon Techniques, Second Edition*

Chu, Leung, Hui & Cheung/ *4th PARTY CYBER LOGISTICS FOR AIR CARGO*

Simchi-Levi, Wu & Shen/ *HANDBOOK OF QUANTITATIVE SUPPLY CHAIN ANALYSIS: Modeling in the E-Business Era*

Gass & Assad/ *AN ANNOTATED TIMELINE OF OPERATIONS RESEARCH: An Informal History*

Greenberg/ *TUTORIALS ON EMERGING METHODOLOGIES AND APPLICATIONS IN OPERATIONS RESEARCH*

Weber/ *UNCERTAINTY IN THE ELECTRIC POWER INDUSTRY: Methods and Models for Decision Support*

Figueira, Greco & Ehrgott/ *MULTIPLE CRITERIA DECISION ANALYSIS: State of the Art Surveys*

Reveliotis/ *REAL-TIME MANAGEMENT OF RESOURCE ALLOCATIONS SYSTEMS: A Discrete Event Systems Approach*

**** A list of the early publications in the series is at the end of the book ****

STOCHASTIC LINEAR PROGRAMMING
Models, Theory, and Computation

PETER KALL
University of Zürich/Switzerland

JÁNOS MAYER
University of Zürich/Switzerland

Springer

Peter Kall
University of Zurich
Switzerland

János Mayer
University of Zurich
Switzerland

Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available
from the Library of Congress.

ISBN 0-387-23385-7 e-ISBN 0-387-24440-9 Printed on acid-free paper.

Copyright © 2005 by Kluwer Academic Publishers.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science + Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now know or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks and similar terms, even if the are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11050001

springeronline.com

Contents

Notations ix
Preface 3

1. BASICS 7
   1 Introduction 7
   2 Linear Programming Prerequisites 13
      2.1 Algebraic concepts and properties 13
      2.2 Geometric interpretation 16
      2.3 Duality statements 19
      2.4 The Simplex Method 22
      2.5 The Dual Simplex Method 27
      2.6 Dual Decomposition 29
      2.7 Nested Decomposition 37
      2.8 Regularized Decomposition 53
      2.9 Interior Point Methods 56
   3 Nonlinear Programming Prerequisites 60
      3.1 Optimality Conditions 64
      3.2 Solution methods 66

2. SINGLE–STAGE SLP MODELS 75
   1 Introduction 75
   2 Models involving probability functions 92
      2.1 Basic properties 96
      2.2 Finite discrete distribution 98
      2.3 Separate probability functions 100
      2.3.1 Only the right–hand–side is stochastic 102
      2.3.2 Multivariate normal distribution 103

2.3.3 Stable distributions 111
2.3.4 A distribution–free approach 117
2.4 The independent case 120
2.5 Joint constraints: random right–hand–side 122
2.5.1 Generalized–concave probability measures 122
2.5.2 Generalized–concave distribution functions 132
2.5.3 Maximizing joint probability functions 136
2.6 Joint constraints: random technology matrix 136
2.7 Summary on the convex programming subclasses 141
3 Quantile functions, Value at Risk 144
4 Models based on expectation 146
4.1 Integrated chance constraints 149
4.1.1 Separate integrated probability functions 149
4.1.2 Joint integrated probability functions 154
4.2 A model involving conditional expectation 158
4.3 Conditional Value at Risk 159
5 Models built with deviation measures 166
5.1 Quadratic deviation 166
5.2 Absolute deviation 169
5.3 Quadratic semi–deviation 174
5.4 Absolute semi–deviation 177
6 Modeling risk and opportunity 178
7 Risk measures 180
7.1 Risk measures in finance 182
7.2 Properties of risk measures 184
7.3 Portfolio optimization models 189

3. MULTI–STAGE SLP MODELS 193
1 The general SLP with recourse 193
2 The two-stage SLP 198
2.1 Complete fixed recourse 201
2.2 Simple recourse 226
2.3 Some characteristic values for two-stage SLP's 243
3 The multi-stage SLP 248
3.1 MSLP with finite discrete distributions 249
3.2 MSLP with non-discrete distributions 255

4. ALGORITHMS 273
1 Introduction 273

2	Single–stage models with separate probability functions		273
	2.1	A guide to available software	275
3	Single–stage models with joint probability functions		276
	3.1	Numerical considerations	277
	3.2	Cutting plane methods	281
	3.3	Other algorithms	284
	3.4	Bounds for the probability distribution function	284
	3.5	Computing probability distribution functions	292
	3.5.1	A Monte–Carlo approach with antithetic variates	293
	3.5.2	A Monte–Carlo approach based on probability bounds	296
	3.6	Finite discrete distributions	299
	3.7	A guide to available software	299
	3.7.1	SLP problems with logconcave distribution functions	299
	3.7.2	Evaluating probability distribution functions	300
	3.7.3	SLP problems with finite discrete distributions	301
4	Single–stage models based on expectation		301
	4.1	Solving equivalent LP's	302
	4.2	Dual decomposition revisited	302
	4.3	Models with separate integrated probability functions	307
	4.4	Models involving CVaR–optimization	309
	4.5	Models with joint integrated probability functions	311
	4.6	A guide to available software	314
	4.6.1	Models with separate integrated probability functions	314
	4.6.2	Models with joint integrated probability functions	314
	4.6.3	Models involving CVaR	314
5	Single–stage models involving VaR		314
6	Single–stage models with deviation measures		315
	6.1	A guide to available software	316
7	Two–stage recourse models		316
	7.1	Decomposition methods	317
	7.2	Successive discrete approximation methods	318
	7.2.1	Computing the Jensen lower bound	320
	7.2.2	Computing the E–M upper bound for an interval	320
	7.2.3	Computing the bounds for a partition	323
	7.2.4	The successive discrete approximation method	326
	7.2.5	Implementation	331
	7.2.6	Simple recourse	337
	7.2.7	Other successive discrete approximation algorithms	341

	7.3	Stochastic algorithms	342
	7.3.1	Sample average approximation (SAA)	342
	7.3.2	Stochastic decomposition	348
	7.3.3	Other stochastic algorithms	352
	7.4	Simple recourse models	353
	7.5	A guide to available software	353
8	Multistage recourse models		356
	8.1	Finite discrete distribution	356
	8.2	Scenario generation	358
	8.2.1	Bundle–based sampling	360
	8.2.2	A moment–matching heuristic	361
	8.3	A guide to available software	367
9	Modeling systems for SLP		368
	9.1	Modeling systems for SLP	368
	9.2	SLP–IOR	369
	9.2.1	General issues	370
	9.2.2	Analyze tools and workbench facilities	371
	9.2.3	Transformations	372
	9.2.4	Scenario generation	372
	9.2.5	The solver interface	373
	9.2.6	System requirements and availability	374
References			375
Index			395

Notations

One–stage models: Joint chance constraints

$A, B, C, \cdots$	:	arrays (usually given real matrices)
$a, b, c, \cdots$	:	arrays (usually given real vectors)
$x, y, z, \cdots$	:	arrays (usually real or integer variable vectors)
$(\Omega, \mathcal{F}, P)$	:	probability space
$\mathbb{N}$	:	set of natural numbers
$(\mathbb{R}^r, \mathcal{B}^r)$	:	$\mathbb{R}^r$ endowed with the Borel σ-algebra $\mathcal{B}^r$
$\xi : \Omega \to \mathbb{R}^r$	:	random vector, i.e. Borel measurable mapping inducing the probability measure $\mathbb{P}_\xi$ on $\mathcal{B}^r$ according to $\mathbb{P}_\xi(M) = P(\xi^{-1}[M])\ \forall M \in \mathcal{B}^r$
$T(\xi), h(\xi)$	:	random array and random vector, respectively, defined as:
$T(\cdot) : \mathbb{R}^r \to \mathbb{R}^{m_2 \times n_1}$	:	$T(\xi) = T + \sum_{j=1}^{r} T^j \xi_j \,;\, T,\ T^j \in \mathbb{R}^{m_2 \times n_1}$ fix
$h(\cdot) : \mathbb{R}^r \to \mathbb{R}^{m_2}$	:	$h(\xi) = h + \sum_{j=1}^{r} h^j \xi_j \,;\, h,\ h^j \in \mathbb{R}^{m_2}$ fix
$\overline{\xi}$	:	expectation $\mathbb{E}_\xi[\xi] = \int_{\mathbb{R}^r} \xi \, \mathbb{P}_\xi(d\xi) = \int_\Omega \xi(\omega)\, dP$
$\overline{T}, \overline{h}$	:	expectations $\mathbb{E}_\xi[T(\xi)] = T(\overline{\xi})$ and $\mathbb{E}_\xi[h(\xi)] = h(\overline{\xi})$, respectively
$\widehat{\xi}$	:	realization of random ξ
$\widehat{T}, \widehat{h}$	:	realizations $T(\widehat{\xi}),\ h(\widehat{\xi})$, respectively

One–stage models: Separate chance constraints

$t_i(\cdot)$	:	i-th row of $T(\cdot)$
$h_i(\cdot)$	:	i-th component of $h(\cdot)$

Two–stage recourse models

$W(\xi), q(\xi)$	:	random array and random vector, respectively, defined as:

$W(\cdot) : \mathbb{R}^r \to \mathbb{R}^{m_2 \times n_2}$	:	$W(\xi) = W + \sum_{j=1}^{r} W^j \xi_j \,;\, W,\ W^j \in \mathbb{R}^{m_2 \times n_2}$
$q(\cdot) : \mathbb{R}^r \to \mathbb{R}^{n_2}$	:	$q(\xi) = q + \sum_{j=1}^{r} q^j \xi_j \,;\, q,\ q^j \in \mathbb{R}^{n_2}$
$\overline{W}$, $\overline{q}$	:	expectations $\mathbb{E}_\xi[W(\xi)] = W(\overline{\xi})$ and $\mathbb{E}_\xi[q(\xi)] = q(\overline{\xi})$, respectively

Multi–stage recourse models

$\xi : \Omega \to \mathbb{R}^R$	:	random vector $\xi = (\xi_2, \cdots, \xi_T)$ with $\xi_t : \Omega \to \mathbb{R}^{r_t},\ t = 2, \cdots, T$ and $\sum_{t=2}^{T} r_t = R$
$\zeta_t : \Omega \to \mathbb{R}^{R_t}$	:	the **state** of the process at stage t, defined as random vector $\zeta_t = (\xi_2, \cdots, \xi_t)$, $t \geq 2$, or else $\zeta_t = (\eta_1, \cdots, \eta_{R_t})$ with $R_t = \sum_{\tau=2}^{t} r_\tau$, with the corresponding marginal distribution of ξ
$A_{t\tau}(\cdot) : \mathbb{R}^{R_t} \to \mathbb{R}^{m_t \times n_\tau}$:		$A_{t\tau}(\zeta_t) = A_{t\tau} + \sum_{\kappa=2}^{t} \sum_{\nu=R_{\kappa-1}+1}^{R_\kappa} A_{t\tau\nu}\, \eta_\nu$, where $A_{t\tau},\ A_{t\tau\nu} \in \mathbb{R}^{m_t \times n_\tau}$ and $R_1 = 0$, with $1 \leq \tau \leq t$ and $2 \leq t \leq T$
$b_t(\cdot) : \mathbb{R}^{R_t} \to \mathbb{R}^{m_t}$	:	$b_t(\zeta_t) = b_t + \sum_{\kappa=2}^{t} \sum_{\nu=R_{\kappa-1}+1}^{R_\kappa} b_{t\nu}\, \eta_\nu$, where $b_t,\ b_{t\nu} \in \mathbb{R}^{m_t}$ and $2 \leq t \leq T$
$c_t(\cdot) : \mathbb{R}^{R_t} \to \mathbb{R}^{n_t}$	:	$c_t(\zeta_t) = c_t + \sum_{\kappa=2}^{t} \sum_{\nu=R_{\kappa-1}+1}^{R_\kappa} c_{t\nu}\, \eta_\nu$, where $c_t,\ c_{t\nu} \in \mathbb{R}^{n_t}$ and $2 \leq t \leq T$

Multi–stage recourse models: Discrete distribution

$\xi : \Omega \to \mathbb{R}^R$	:	random vector with discrete distribution $\{(\widehat{\xi}^s, q_s);\, s = 1, \cdots, S\}$, i.e. **scenarios** $\widehat{\xi}^s = (\widehat{\xi}_2^s, \cdots, \widehat{\xi}_T^s) = (\widehat{\eta}_1^s, \cdots, \widehat{\eta}_R^s)$ with $\mathbb{P}_\xi(\xi = \widehat{\xi}^s) = q_s$, $s \in \mathcal{S} := \{1, \cdots, S\}$
$\zeta_t : \Omega \to \mathbb{R}^{R_t}$	:	discrete set $\{\widehat{\zeta}_t^s = (\widehat{\xi}_2^s, \cdots, \widehat{\xi}_t^s);\ s \in \mathcal{S}\}$ of

states defining $k_t \geq 1$ different equivalence classes $U_t^\nu \subseteq \mathcal{S}$, with $s_i, s_j \in U_t^\nu \Leftrightarrow \widehat{\zeta}_t^{s_i} = \widehat{\zeta}_t^{s_j}$ and an associated set of different states at stage t which may be defined by $\mathcal{S}_t := \{\rho \mid \rho \text{ minimal in one of the } U_t^\nu\}$ as $\{\widehat{\zeta}_t^\rho \mid \rho \in \mathcal{S}_t\}$ with the distribution $\mathbb{P}_\xi(\zeta_t = \widehat{\zeta}_t^\rho) = \pi_{t\rho} = \sum_{s \in \mathcal{S}} \{q_s \mid \widehat{\zeta}_t^s = \widehat{\zeta}_t^\rho\}$ (see Fig. 1 with e.g. $\mathcal{S}_2 = \{1, 6, 11\}$)

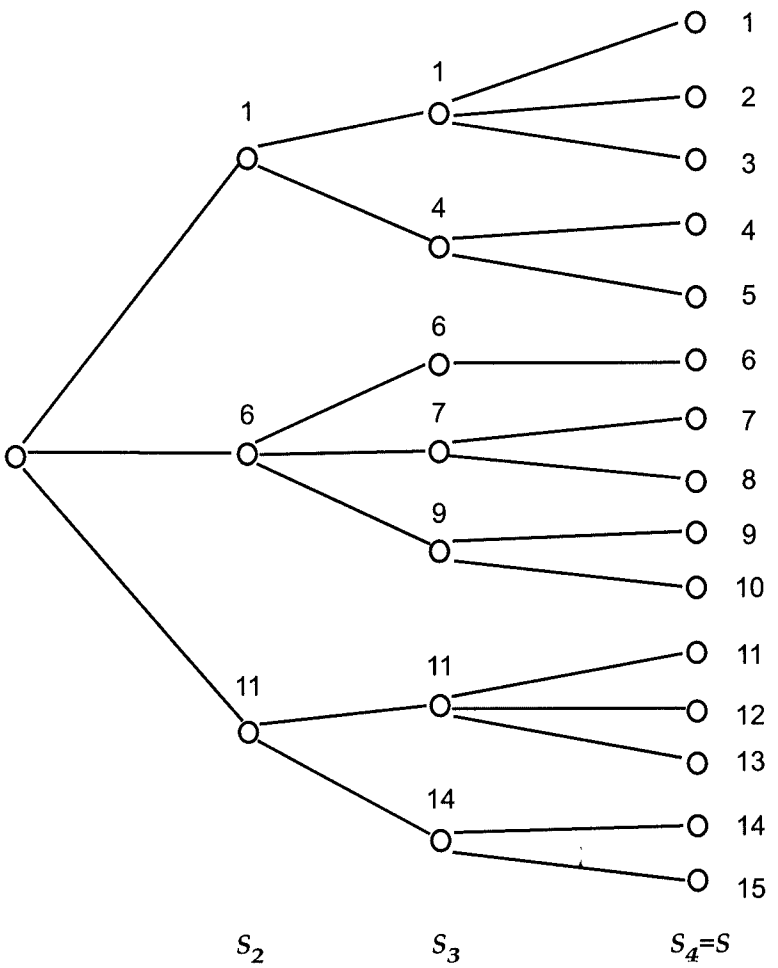

Figure 1. Scenario tree: Assigning states to nodes.

Multi–stage recourse models: The scenario tree

$(\mathcal{N}, \mathcal{A})$	:	tree with nodes $\mathcal{N} \subset \mathbb{N}$, where $n = 1$ is the (unique) root and $\lvert\mathcal{N}\rvert = \sum_{t=2}^{T} \lvert\mathcal{S}_t\rvert + 1$
t_n	:	the stage to which $n \in \mathcal{N}$ belongs; there is a bijection $(t_{\{\cdot\}}, \rho(\cdot)) : \{\mathcal{N} - \{1\}\} \to \bigcup_{t=2}^{T} \{(t, \mathcal{S}_t)\}$ such that $n \leftrightarrow (t_n, \rho(n)),\ n \geq 2$; hence we assign with any node $n \geq 2$
$\widehat{\zeta}^n$	:	$\widehat{\zeta}^n = \widehat{\zeta}_{t_n}^{\rho(n)}$ with $\{\widehat{\zeta}_{t_n}^{\rho(n)},\ \rho(n) \in \mathcal{S}_{t_n}\}$ uniquely determined by $n \in \mathcal{N}$ (state in node n)
$\mathcal{D}(t) \subset \mathcal{N}$	:	set of nodes in stage $t,\ 1 \leq t \leq T$
h_n	:	the parent node of node $n \in \mathcal{N},\ n \geq 2$ (immediate predecessor)
$\mathcal{H}(n) \subset \mathcal{N}$	:	set of nodes in the path from $n \in \mathcal{N}$ to the root, ordered by stages, including n (history of n)
$\mathcal{S}(n) \subset \{1, \cdots, S\}$	:	$\mathcal{S}(n) = \{s \in \mathcal{S} \mid \widehat{\zeta}_{t_n}^{s} = \widehat{\zeta}^n\}$, i.e. the index set of those scenarios, for which the scenario path contains $n \in \mathcal{N}$. $\mathcal{S}(n)$ and the related set of scenarios are called the **scenario bundle** of the corresponding node n
p_n	:	probability of $\mathcal{S}(n)$: $p_n = \mathbb{P}_\xi(\zeta_{t_n} = \widehat{\zeta}^n) = \pi_{t_n\,\rho(n)}$
$\mathcal{C}(n) \subset \mathcal{N}$	:	set of children (immediate successors) of n
$\mathcal{G}_s(n) \subseteq \mathcal{N}$	:	future of node n along scenario $s \in \mathcal{S}(n)$, including n (and hence $\mathcal{G}_s(n) = \emptyset$ if $s \notin \mathcal{B}(n)$)
$\mathcal{G}(n) \subseteq \mathcal{N}$	:	$\mathcal{G}(n) = \bigcup_{s \in \mathcal{B}(n)} \mathcal{G}_s(n)$, the future of $n \in \mathcal{N}$

To Helene and Ilona

Preface

The beginning of stochastic programming, and in particular stochastic linear programming (SLP), dates back to the 50's and early 60's of the last century. Pioneers who—at that time—contributed to the field, either by identifying SLP problems in particular applications, or by formulating various model types and solution approaches for dealing adequately with linear programs containing random variables in their right–hand–side, their technology matrix, and/or their objective's gradient, have been among others (in alphabetical order):

E.M.L. Beale [10], proposing a quadratic programming approach to solve special simple recourse stochastic programs;
A. Charnes and W.W. Cooper [38], introducing a particular stochastic program with chance constraints;
G.B. Dantzig [43], formulating the general problem of linear programming with uncertain data and
G.B. Dantzig and A. Madansky [47], discussing at an early stage the possibility to solve particular two-stage stochastic linear programs;
G. Tintner [287], considering stochastic linear programming as an appropriate approach to model particular agricultural applications; and
C. van de Panne and W. Popp [293], considering a cattle feed problem modeled with probabilistic constraints.

In addition we should mention just a few results and methods achieved before 1963, which were not developed in connection with stochastic programming, but nevertheless turned out to play an essential role in various areas of our field. One instance is the Brunn-Minkowski inequality based on the investigations of H. Brunn [32] in 1887 and H. Minkowski [206] in 1897, which comes up in connection with convexity statements for probabilistic constraints, as mentioned e.g. in Prékopa [234]. Furthermore, this applies in particular to the discussion about bounds on distribution functions, based on inequalities published by G. Boole in 1854 and by C.E. Bonferroni in 1937 (for the references

see Prékopa [234]), and on the other hand, about bounds on the expectation of a convex function of a random variable, leading to a lower bound by the inequality of J.L. Jensen [128], and to the Edmundson–Madansky upper bound due to H.P. Edmundson [71] and A. Madansky [183].

Among the concepts of solution approaches, developed until 1963 for linear or nonlinear programming problems, the following ones, in part after appropriate modifications, still serve as basic tools for dealing with SLP problems:

Besides Dantzig's simplex method and the Dantzig–Wolfe decomposition, described in detail in G.B. Dantzig [44], the dual decomposition proposed by J.F. Benders [12], cutting plane methods as introduced by J.E. Kelley [159], and feasible direction methods proposed and discussed in detail by G. Zoutendijk [311], may be recognized even within today's solution methods for various SLP problems. Of course, these methods and in particular their implementations have been revised and improved meanwhile, and in addition we know of many new solution approaches, some of which will be dealt with in this book.

The aim of this volume is to draw a bow from solution methods of (deterministic) mathematical programming, being of use in SLP as well, through theoretical properties of various SLP problems which suggest in many cases the design of particular solution approaches, to solvers, understood as implemented algorithms for the solution of the corresponding SLP problems.

Obviously we are far from giving a complete picture on the present knowledge and computational possibilities in SLP. First we had to omit the area of stochastic integer programming (SILP), since following the above concept would have implied to give first a survey on those integer programming methods used in SILP; this would go beyond the limits of this volume. However the reader may get a first flavour of SILP by having a look for instance into the articles of W.K. Klein Haneveld, L. Stougie, and M.H. van der Vlerk [168], W. Römisch and R. Schultz [256], M.H. van der Vlerk [299], and the recent survey of S. Sen [268].

And, as the second restriction, in presenting detailed descriptions we have essentially confined ourselves to those computational methods for solving SLP problems belonging to one of the following categories:

Either information on the numerical efficiency of a corresponding solver is reported in the literature based on reasonable test sets (not just three examples or less!) and the solver is publicly available;

or else, corresponding solvers have been attached to our model management system SLP-IOR, either implemented by ourselves or else provided by their authors, such that we were able to gain computational experience on the methods presented, based on running the corresponding solvers on randomly generated test batteries of SLP's with various characteristics like problem size, matrix

entries density, probability distribution, range and sign of problem data, and some others.

Finally, we owe thanks to many colleagues for either providing us with their solvers to link them to SLP-IOR, or for their support in implementing their methods by ourselves. Further, we gratefully acknowledge the critical comments of Simon Siegrist at our Institute. Obviously, the remaining errors are the sole responsibility of the authors. Last but not least we are indebted to the publisher for an excellent cooperation. This applies in particular to the publisher's representative, Gary Folven, to whom we are also greatly obliged for his patience.

Zürich, September 2004 Peter Kall and János Mayer

Chapter 1

BASICS

1. Introduction

Linear programs have been studied in many aspects during the last 50 years. They have shown to be appropriate models for a wide variety of practical problems and, at the same time, they became numerically tractable even for very large scale instances. As standard formulations of linear programs (LP) we find problems like

$$\left.\begin{array}{lrcl} & \min c^{\mathrm{T}}x & & \\ \text{subject to} & Ax & \propto & b \\ & l \leq x & \leq & u, \end{array}\right\} \tag{1.1}$$

with the matrix $A \in \mathbb{R}^{m\times n}$, the objective's gradient $c \in \mathbb{R}^n$, the right–hand–side $b \in \mathbb{R}^m$, and the lower and upper bounds $l \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$, respectively. If some x_i is unbounded below and/or above, this corresponds to $l_i = -\infty$ and/or $u_i = \infty$. A, b, c, l, u are assumed to be known fixed data in the above model. The relation '$\propto$' is to be replaced row-wise by one of the relations '$\leq$', '$=$', or '$\geq$'. Then the task is obviously to find the—or at least one—optimal feasible solution $x \in \mathbb{R}^n$. Alternatively, we often find also the LP-formulation

$$\left.\begin{array}{lrcl} & \min c^{\mathrm{T}}x & & \\ \text{subject to} & Ax & \propto & b \\ & x & \geq & 0, \end{array}\right\} \tag{1.2}$$

under the analogous assumptions as above. For these two LP types it holds obviously that, given a problem of one type, it may be reformulated into an equivalent problem of the other type. More precisely,

— given the LP in the formulation (1.2), by introducing the lower bounds $l = (0, \cdots, 0)^{\mathrm{T}}$ and the upper bounds $u = (\infty, \cdots, \infty)^{\mathrm{T}}$ (in computations rather markers $u = (M, \cdots, M)^{\mathrm{T}}$ with a sufficiently large number M, e.g. $M = 10^{20}$, just to indicate unboundedness), the problem is trivially of the type (1.1); and

— having the LP of type (1.1), introducing variables $x^+ \in \mathbb{R}^n_+$, $x^- \in \mathbb{R}^n_+$, inserting $x = x^+ - x^-$, $x^+ \geq 0, x^- \geq 0$, introducing the *slack variables* $y \in \mathbb{R}^n_+$ and $z \in \mathbb{R}^n_+$, and restating the conditions $l \leq x \leq u$ equivalently as

$$\begin{array}{rrrcl} x^+ - x^- & -y & & = & l \\ x^+ - x^- & & +z & = & u \\ & y & & \geq & 0 \\ & & z & \geq & 0, \end{array}$$

the problem is transformed into the type (1.2).

In the same way it follows that every LP may be written as

$$\left.\begin{array}{rrcl} & \min c^{\mathrm{T}} x & & \\ \text{subject to} & Ax & = & b \\ & x & \geq & 0, \end{array}\right\} \tag{1.3}$$

i.e. as a special variant of (1.2).

Numerical methods known to be efficient in solving LP's belong essentially to one of the following classes:

— Pivoting methods, in particular the simplex and/or the dual simplex method;

— interior point methods for LP's with very sparse matrices;

— decomposition, dual decomposition and regularized decomposition approaches for LP's with special block structures of their coefficient matrices A.

In real life problems the fundamental assumption for linear programming, that the problem entries—except for the variables x—be known fixed data, does often happen not to hold. It either may be the case that (some of) the entries are constructed as statistical estimates from some observed real data, i.e. from some samples, or else that we know from the model design that they are random variables (like capacities, demands, productivities or prices). The standard approach to replace these random variables by their mean values—corresponding to the choice of statistical estimates mentioned before—and afterwards to solve the resulting LP may be justified only under special conditions; in general, it can easily be demonstrated to be dramatically wrong.

Assume, for instance, as a model for a diet problem the LP

$$\left.\begin{array}{lrcl} & \min c^{\mathrm{T}}x & & \\ \text{s. t.} & Ax & \geq & b \\ & Tx & \geq & h \\ & x & \geq & 0, \end{array}\right\} \tag{1.4}$$

where x represents the quantities of various foodstuffs, and c is the corresponding price vector. The constraints reflect chemical or physiological requirements to be satisfied by the diet. Let us assume that the elements of A and b are fixed known data, i.e. deterministic, whereas at least some of the elements of T and/or h are random with a known joint probability distribution, which is not influenced by the choice of the decision x. Further, assume that the realizations of the random variables in T and h are not known before the decision on the diet x is taken, i.e. before the consumption of the diet. Replacing the random T and h by their expectations $\overline{T}$ and $\overline{h}$ and solving the resulting LP

$$\left.\begin{array}{lrcl} & \min c^{\mathrm{T}}x & & \\ \text{s. t.} & Ax & \geq & b \\ & \overline{T}x & \geq & \overline{h} \\ & x & \geq & 0, \end{array}\right\} \tag{1.5}$$

can result in a diet $\hat{x}$ violating the constraints in (1.4) very likely and hence with a probability much higher than feasible for the diet to serve successfully its medical purpose. Therefore, the medical experts would rather require a decision on the diet which satisfies all constraints jointly with a rather high probability, as 95% say, such that the problem to solve were

$$\left.\begin{array}{lrcl} & \min c^{\mathrm{T}}x & & \\ \text{s. t.} & Ax & \geq & b \\ & P(Tx \geq h) & \geq & 0.95 \\ & x & \geq & 0, \end{array}\right\} \tag{1.6}$$

a *stochastic linear program (SLP) with joint probabilistic constraints.* Here we had at the starting point the LP (1.4) as model for our diet problem. However, the (practical) requirement to satisfy—besides the deterministic constraints $Ax \geq b$—also the reliability constraint $P(Tx \geq h) \geq 0.95$, yields with (1.6) a nonlinear program (NLP). This is due to the fact, that in general the probability function $G(x) := P(Tx \geq h)$ is clearly nonlinear.

As another example, let some production problem be formulated as

$$\left.\begin{array}{lrcl} & \min c^{\mathrm{T}}x & & \\ \text{s. t.} & Ax & = & b \\ & Tx & = & h \\ & x & \geq & 0, \end{array}\right\} \tag{1.7}$$

where T and h may contain random variables (productivities, demands, capacities, etc.) with a joint probability distribution (independent again of the choice of x), and the decision on x has to be taken before the realization of the random variables is known. Consequently, the decision x will satisfy the constraints $Ax = b,\ x \geq 0$; but after the observation of the random variables' realization it may turn out that $Tx \neq h$, i.e. that part of the target (like satisfying the demand for some of the products, capacity constraints, etc.) is not properly met. However, it may be necessary—by a legal commitment, the strong intention to maintain goodwill, or similar reasons—to compensate for the deficiency, i.e. for $h - Tx$, after its observation. One possibility to cope with this obligation may be the introduction of recourse by defining the constraints $Wy = h - Tx,\ y \geq 0$, for instance as model of an emergency production process or simply as the measurement of the absolute values of the deficiencies (represented by $W = (I, -I)$, with I the identity matrix). Let us assume W to be deterministic, and assume the recourse costs to be given as linear by $q^{\mathrm{T}}y$, say. Obviously we want to achieve this compensation with minimal costs. Hence we have the recourse problem

$$\left.\begin{array}{lrcl} & Q(x;T,h) := \min q^{\mathrm{T}}y & & \\ \text{s. t.} & Wy & = & h - Tx \\ & y & \geq & 0. \end{array}\right\} \tag{1.8}$$

For any x, feasible to the *first stage constraints* $Ax = b,\ x \geq 0$, the recourse function, i.e. the optimal value $Q(x;T,h)$ of the *second stage problem* (1.8), depends on T and h and is therefore a random variable. In many applications, e.g. in cases where the production plan x has to be implemented periodically (daily or weekly, for instance), it may be meaningful to choose x in such a way that the average overall costs, i.e. the sum of the first stage costs $c^{\mathrm{T}}x$ and the expected recourse costs $\mathbb{E}\,Q(x;T,h)$, are minimized. Hence we have the problem

$$\left.\begin{array}{lrcl} & \min c^{\mathrm{T}}x + \mathbb{E}\,Q(x;T,h) & & \\ \text{s. t.} & Ax & = & b \\ & x & \geq & 0, \end{array}\right\} \tag{1.9}$$

a *two-stage stochastic linear program (SLP) with fixed recourse.*

Also in this case, although our starting point was the LP (1.7), the resulting problem (1.9) will be an NLP if the random variables in T and h have a

continuous-type joint distribution (i.e. a distribution defined by a density function).

If, however, the random variables in T and h have a joint discrete distribution, defined by the realizations (T^j, h^j) with the probabilities p_j, $j = 1, \cdots, S$ (with $p_j > 0$ and $\sum_{j=1}^{S} p_j = 1$), problem (1.9) is easily seen to be equivalent to

$$\left.\begin{array}{lllll} & \min c^{\mathrm{T}}x + \sum_{j=1}^{S} p_j q^{\mathrm{T}} y^j & & & \\ \text{s. t.} & Ax & & = & b \\ & T^j x & +Wy^j & = & h^j,\ j = 1, \cdots, S \\ & x & & \geq & 0 \\ & & y^j & \geq & 0, \end{array}\right\} \tag{1.10}$$

such that under the discrete distribution assumption we get an LP again, with the special data structure indicated in Fig. 1.1.

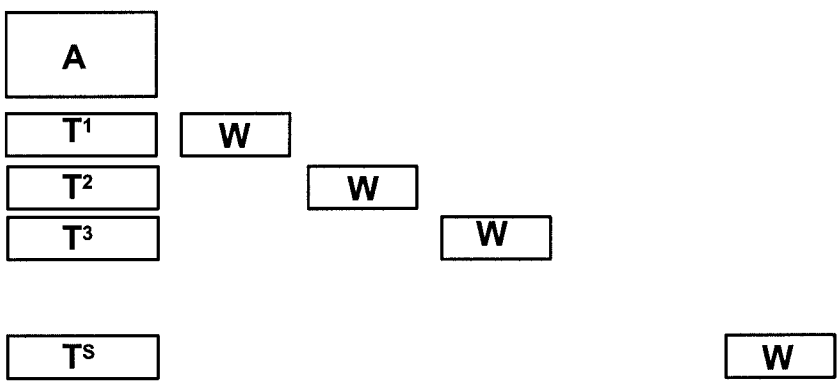

Figure 1.1. Dual decomposition structure.

In applications we observe an increasing need to deal with a generalization of the two-stage SLP with recourse (1.9) and (1.10), respectively. At this point we just give a short description as follows: In a first stage, a decision x_1 is chosen to be feasible with respect to some deterministic first stage constraints. Later on, after the realization of a random vector ξ_2, a deficiency in some second stage constraints has to be compensated for by an appropriate recourse decision $x_2(\xi_2)$. Then after the realization of a further random vector ξ_3, the former decisions x_1 and $x_2(\xi_2)$ may not be feasible with respect to some third stage constraints, and a further recourse decision $x_3(\xi_2, \xi_3)$ is needed, and so on, until a final stage T is reached. Again, we assume that, besides the first stage costs $c_1^{\mathrm{T}} x_1$, the recourse decisions $x_t(\zeta_t)$, $t \geq 2$, imply additional linear

costs $c_t^{\mathrm{T}} x_t(\zeta_t)$, where $\zeta_t = (\xi_2, \cdots, \xi_t)$. Then the *multi-stage SLP with fixed recourse* is formulated as

$$\left.\begin{array}{ll}
\min \left\{ c_1^{\mathrm{T}} x_1 + \mathbb{E}_{\zeta_T} \left[\sum_{t=2}^{T} c_t^{\mathrm{T}} x_t(\zeta_t) \right] \right\} & \\
\text{subject to} & \\
A_{11} x_1 & = b_1 \\
A_{t1}(\zeta_t) x_1 + \sum_{\tau=2}^{t} A_{t\tau}(\zeta_t) x_\tau(\zeta_\tau) & = b_t(\zeta_t), \text{ a.s., } t = 2, \cdots, T \\
x_1 \geq 0, \; x_t(\zeta_t) & \geq 0, \qquad \text{a.s., } t = 2, \cdots, T,
\end{array}\right\} \tag{1.11}$$

where, in general, we shall assume $A_{tt}(\zeta_t)$, $t \geq 2$, the matrices on the diagonal, to be deterministic, i.e. $A_{tt}(\zeta_t) \equiv A_{tt}$. It will turn out that, for general probability distributions, this problem—an NLP again—is much more difficult than the two-stage SLP (1.9), and methods to approximate a solution are just at their beginning phase, at best. However, under the assumption of discrete distributions of the random vectors ζ_t, problem (1.11) can also be reformulated into an equivalent LP, which in general is of (very) large scale, but again with a special data structure to be of use for solution procedures.

From this short sketch of the subject called SLP, which is by far not complete with respect to the various special problem formulations to be dealt with, we may already conclude that a basic toolkit of linear and nonlinear programming methods cannot be waived if we want to deal with the computational solution of SLP problems. To secure the availability of these resources, in the following sections of this chapter we shall remind to basic properties of and solution methods for LP's and NLP's as they are used or referred to in the SLP context, later on.

In Chapter 2, we present various Single–stage SLP models (like e.g. problem (1.6) on page 9) and discuss their theoretical properties, relevant for their computational tractability, as convexity statements, for instance.

In Chapter 3 follows an anlogous discussion of Multi–stage SLP models (like problem (1.9) in particular, and problem (1.11) in general), focussed among others on properties allowing for the construction of particular approximation methods for computing (approximate) solutions.

For some of the models discussed before, Chapter 4 will present solution methods, which have shown to be efficient in extensive computational experiments.

2. Linear Programming Prerequisites

In this section we briefly present the basic concepts in linear programming and, for various types of solution methods, the conceptual algorithms.

As mentioned on page 8 we may use the following standard formulation of an LP:

$$\left.\begin{array}{rrcl} & \min c^{\mathrm{T}}x & & \\ \text{s. t.} & Ax & = & b \\ & x & \geq & 0. \end{array}\right\} \tag{2.1}$$

With A being an $(m \times n)$-matrix, and b and c having corresponding dimensions, we know from linear algebra that the system of equations

$$Ax = b \text{ is solvable if and only if } \operatorname{rank}(A, b) = \operatorname{rank}(A).$$

Therefore, solvability of the system $Ax = b$ implies that

- either $\operatorname{rank}(A) = m$,
- or the system contains redundant equations which may be omitted, such that for the remaining system $\tilde{A}x = \tilde{b}$ we have the same set of solutions as for the original system, and that, for the $(m_1 \times n)$-matrix $\tilde{A}$, $m_1 < m$, the condition $\operatorname{rank}(\tilde{A}) = m_1$ holds.

Observing this well known fact, we henceforth assume without loss of generality, that $\operatorname{rank}(A) = m$ $(\leq n)$ for the $(m \times n)$-matrix A.

2.1 Algebraic concepts and properties

Solving the LP (2.1) obviously requires to find an extreme (minimal in our formulation) value of a linear function on a feasible set described as the intersection of a linear manifold, $\{x \mid Ax = b\}$, and finitely many halfspaces, $\{x \mid x_j \geq 0\}$, $j = 1, \cdots, n$, suggesting that this problem may be discussed in algebraic terms.

DEFINITION 2.1 *Any feasible solution $\hat{x}$ of (2.1) is called a feasible basic solution if, for $I(\hat{x}) = \{i \mid \hat{x} > 0\}$, the set $\{A_i,\ i \in I(\hat{x})\}$ of columns in A is linearly independent.*

According to this definition, for any feasible basic solution $\hat{x}$ of (2.1) holds

$$\hat{x}_i > 0 \text{ for } i \in I(\hat{x}),\ \hat{x}_j = 0 \text{ for } j \notin I(\hat{x}), \text{ and } \sum_{i \in I(\hat{x})} A_i \hat{x}_i = b.$$

Furthermore, with $|I(\hat{x})|$ being the cardinality of this set (i.e. the number of its elements), if $|I(\hat{x})| < m$ such that the basic solution $\hat{x}$ contains less than

m strictly positive components, then due to our rank assumption on A there is a subset $I_B(\hat{x})$ with $I_B(\hat{x}) \supset I(\hat{x})$ and $|I_B(\hat{x})| = m$ such that the column set $\{A_i,\ i \in I_B(\hat{x})\}$ is linearly independent or equivalently, that the $(m \times m)$-matrix $B = (A_i \mid i \in I_B(\hat{x}))$ is nonsingular. Introducing, with $I_B(\hat{x}) = \{i_1, \cdots, i_m\}$ and $I_N(\hat{x}) = \{1, \cdots, n\} \setminus I_B(\hat{x}) = \{j_1, \cdots, j_{n-m}\}$, the vectors $x^{\{B\}} \in \mathbb{R}^m$—the *basic variables*—and $x^{\{N\}} \in \mathbb{R}^{n-m}$—the *non-basic variables*—according to

$$x_k^{\{B\}} = x_{i_k},\ i_k \in I_B(\hat{x}) \text{ for } k = 1, \cdots, m;$$

$$x_l^{\{N\}} = x_{j_l},\ j_l \in I_N(\hat{x}) \text{ for } l = 1, \cdots, n-m,$$

then, with the $(m \times (n-m))$-matrix $N = (A_j \mid j \in I_N(\hat{x}))$ the system $Ax = b$ is, up to a possible rearrangement of columns and variables, equivalent to the system

$$Bx^{\{B\}} + Nx^{\{N\}} = b.$$

Therefore, up to the mentioned rearrangement of variables, the former feasible basic solution $\hat{x}$ corresponds to $(\hat{x}^{\{B\}} = B^{-1}b \geq 0,\ \hat{x}^{\{N\}} = 0)$, and the submatrix B of A is called a *feasible basis* . With the same rearrangement of the components of the vector c into the two vectors $c^{\{B\}}$ and $c^{\{N\}}$ we may rewrite problem (2.1) as

$$\begin{array}{lrcrcl} \min & c^{\{B\}\mathrm{T}} x^{\{B\}} & + & c^{\{N\}\mathrm{T}} x^{\{N\}} & & \\ \text{s. t.} & Bx^{\{B\}} & + & Nx^{\{N\}} & = & b \\ & x^{\{B\}} & & & \geq & 0 \\ & & & x^{\{N\}} & \geq & 0. \end{array}$$

Solving the system of equations for $x^{\{B\}}$ we get $x^{\{B\}} = B^{-1}b - B^{-1}Nx^{\{N\}}$ such that—with $\gamma_B := c^{\{B\}\mathrm{T}} B^{-1} b$ the objective value of the feasible basic solution $(\hat{x}^{\{B\}} = B^{-1}b \geq 0,\ \hat{x}^{\{N\}} = 0)$—problem (2.1) is equivalent to

$$\left.\begin{array}{lrcrcl} \min & \gamma_B + \left(c^{\{N\}\mathrm{T}} - c^{\{B\}\mathrm{T}} B^{-1} N\right) x^{\{N\}} & & & & \\ \text{s. t.} & x^{\{B\}} & = & B^{-1}b - B^{-1}Nx^{\{N\}} & \geq & 0 \\ & & & x^{\{N\}} & \geq & 0. \end{array}\right\} \tag{2.2}$$

For computational purposes (2.2) is usually represented by the *simplex tableau*

$$\begin{array}{|c|c|}\hline \zeta & d^{\mathrm{T}} \\ \hline \beta & D \\ \hline \end{array} = \begin{array}{|c|ccc|}\hline \zeta & \delta_1 & \cdots & \delta_{n-m} \\ \hline \beta_1 & \alpha_{11} & \cdots & \alpha_{1n-m} \\ \vdots & \vdots & & \vdots \\ \beta_m & \alpha_{m1} & \cdots & \alpha_{mn-m} \\ \hline \end{array} \tag{2.3}$$

such that the objective and the equality constraints of (2.2) are rewritten as

$$\left.\begin{array}{lcl} z & := & \zeta \;-\; d^{\mathrm{T}} x^{\{N\}} \\ x^{\{B\}} & = & \beta \;-\; D\, x^{\{N\}} \end{array}\right\} \tag{2.4}$$

with $\zeta = \gamma_B = c^{\{B\}^{\mathrm{T}}} B^{-1} b$, $\beta = (\beta_1, \cdots, \beta_m)^{\mathrm{T}} = B^{-1} b$, and furthermore

$$D = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n-m} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn-m} \end{pmatrix} = B^{-1} N$$

and

$$d^{\mathrm{T}} = (\delta_1, \cdots, \delta_{n-m}) = \left(c^{\{B\}^{\mathrm{T}}} B^{-1} N - c^{\{N\}^{\mathrm{T}}}\right) = \left(c^{\{B\}^{\mathrm{T}}} D - c^{\{N\}^{\mathrm{T}}}\right).$$

Although not written down explicitly, we assume that also for the reformulation (2.3) and (2.4) the nonnegativity constraints $x^{\{B\}} \geq 0$, $x^{\{N\}} \geq 0$ have to hold.

To justify the simplex algorithm as a solution method for (2.1) the following statements are essential.

PROPOSITION 2.1 *Provided that the LP (2.1) is feasible, i.e. that the feasible set* $\mathcal{B} := \{x \mid Ax = b,\ x \geq 0\} \neq \emptyset$, *there exists at least one feasible basic solution.*

PROPOSITION 2.2 *If the LP (2.1) is solvable with the optimal value* $\hat{\gamma}$, *then there exists at least one feasible basis* $\hat{B}$, *yielding* $c^{\{\hat{B}\}^{\mathrm{T}}} \hat{B}^{-1} b = \hat{\gamma}$.

DEFINITION 2.2 *Assume that* $\operatorname{rank}(A) = m$. *If for a feasible basis* B *and the corresponding feasible basic solution* $\hat{x}$ *with* $(\hat{x}^{\{B\}} = B^{-1} b, \hat{x}^{\{N\}} = 0)$ *it happens that* $|I(\hat{x})| < m$, *i.e. that less than* m *of the basic variables are strictly positive, then the basic solution* $\hat{x}$ *is called degenerate.*

Finally, if we have a feasible basis B such that $d^{\mathrm{T}} \leq 0$, than obviously this basis is optimal, i.e. $(\hat{x}^{\{B\}} = \beta,\ \hat{x}^{\{N\}} = 0)$ solves (2.1), since by (2.4) $z = \zeta - d^{\mathrm{T}} x^{\{N\}} \geq \zeta \quad \forall x^{\{N\}} \geq 0$. On the other hand, assume that (2.1) is solvable, and that in addition all feasible basic solutions are nondegenerate. Then for an optimal feasible basis B, existing due to Prop. 2.2, $d^{\mathrm{T}} \leq 0$ has to hold due to the following argument:

If, for any feasible basis, $d_j > 0$ would hold for some $j \in \{1, \cdots, n-m\}$, due to $\beta > 0$ by the assumed nondegeneracy, we could choose $x^{\{N\}} = \tau e_j$ (e_j the j-th unit vector in $\mathbb{R}^{n-m}$) with some $\tau > 0$, such that according to (2.4) would follow

$$x^{\{B\}} = \beta - \tau D e_j = \beta - \tau D_j \geq 0 \text{ and } z = \zeta - \tau d^{\mathrm{T}} e_j = \zeta - d_j < \zeta.$$

Hence, the basis at hand would not be optimal.

Even without the nondegeneracy assumption the above optimality condition, also known as the *simplex criterion*, can be shown to hold true.

PROPOSITION 2.3 *The LP (2.1) is solvable if and only if there exists an optimal feasible basis B such that the condition*

$$d^{\mathrm{T}} = \left(c^{\{B\}^{\mathrm{T}}} B^{-1} N - c^{\{N\}^{\mathrm{T}}}\right) = \left(c^{\{B\}^{\mathrm{T}}} D - c^{\{N\}^{\mathrm{T}}}\right) \leq 0 \qquad (2.5)$$

is satisfied.

The proof of the above statements may be found in the literature, e.g. in Dantzig [44], Maros [191], or Vanderbei [295].

2.2 Geometric interpretation

Besides the algebraic formulation of LP's, it is sometimes intuitively helpful to have in mind their geometric interpretation. To this end we need the concepts of a *convex polyhedron* and of a *convex polyhedral cone.*

DEFINITION 2.3 *Given finitely many vectors $x^{(1)}, \cdots, x^{(r)} \in \mathbb{R}^n$, then their convex hull*

$$\begin{aligned} \mathcal{P} &= \operatorname{conv}\{x^{(1)}, \cdots, x^{(r)}\} \\ &:= \{x \mid x = \sum_{j=1}^{r} \lambda_j x^{(j)} \text{ with } \sum_{j=1}^{r} \lambda_j = 1,\ \lambda_j \geq 0\ \forall j\} \end{aligned}$$

is called a convex polyhedron, and their positive hull

$$\mathcal{C} = \operatorname{pos}\{x^{(1)}, \cdots, x^{(r)}\} := \{y \mid y = \sum_{j=1}^{r} \mu_j x^{(j)} \text{ with } \mu_j \geq 0\ \forall j\}$$

is called a convex polyhedral cone.
Finally, $\mathcal{P}+\mathcal{C} = \{z \mid z = x+y : x \in \mathcal{P},\ y \in \mathcal{C}\}$ is called a convex polyhedral set.

To generate the polyhedron $\hat{\mathcal{P}}$ of Fig. 2.1, the elements $x^{(6)}$ and $x^{(7)}$ are obviously redundant, i.e. omitting these elements would result in the same polyhedron $\hat{\mathcal{P}}$, whereas no one of the elements $x^{(1)}, \cdots, x^{(5)}$ can be deleted without changing the polyhedron essentially. The simple reason is that a polyhedron is uniquely determined by its vertices.

DEFINITION 2.4 *Given a convex polyhedron $\mathcal{P}$, an element $y \in \mathcal{P}$ is a vertex if and only if there are no two further elements $v, w \in \mathcal{P}$ such that $v \neq y \neq w$ and $y = \lambda v + (1-\lambda)w,\ \lambda \in (0,1)$.*

Similarly, for a convex polyhedral cone not all of the generating elements mentioned in Def. 2.3 might be really needed to represent the cone. More

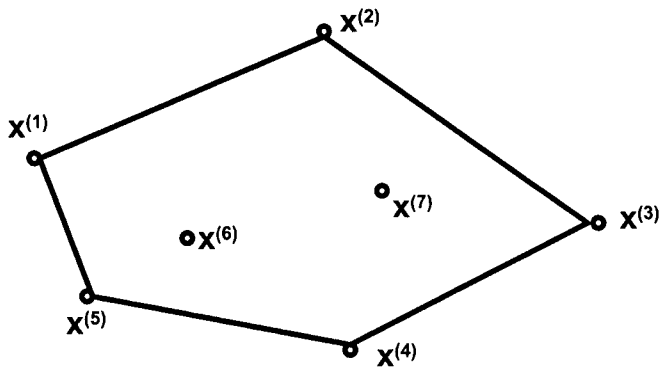

Figure 2.1. Polyhedron $\hat{\mathcal{P}} = \operatorname{conv}\{x^{(1)}, \cdots, x^{(7)}\}$.

precisely, whenever one of the generating elements equals a nonnegative linear combination of the other generating elements, it can be deleted without changing the cone.

With the LP (2.1) the set $\mathcal{C} = \{y \mid Ay = 0,\ y \geq 0\}$ can be associated.

PROPOSITION 2.4 *The set $\mathcal{C} = \{y \mid Ay = 0,\ y \geq 0\}$ is a convex polyhedral cone,*

generated either trivially by $\{0\}$, if $\mathcal{C} = \{0\}$,

or, if $\exists y \in \mathcal{C} :\ y \neq 0$, generated for instance by $\{y^{(1)}, \cdots, y^{(s)}\}$, the set of feasible basic solutions of the system

$$\begin{aligned} Ay &= 0 \\ e^{\mathrm{T}}y &= 1, \text{ where } e^{\mathrm{T}} = (1, \cdots, 1), \\ y &\geq 0. \end{aligned}$$

With these concepts we may describe the feasible set

$$\mathcal{B} = \{x \mid Ax = b,\ x \geq 0\} \tag{2.6}$$

as follows:

PROPOSITION 2.5 *For the feasible set $\mathcal{B} \neq \emptyset$ holds*

$$\mathcal{B} = \mathcal{P} + \mathcal{C} = \{z \mid z = x + y \text{ with } x \in \mathcal{P} \text{ and } y \in \mathcal{C}\},$$

where $\mathcal{C} = \{y \mid Ay = 0,\ y \geq 0\}$ and $\mathcal{P} = \operatorname{conv}\{x^{(1)}, \cdots, x^{(r)}\}$, with $\{x^{(1)}, \cdots, x^{(r)}\}$ being the set of feasible basic solutions of $\mathcal{B}$.

The set of feasible basic solutions of $\mathcal{B}$ can be shown to coincide with the set of vertices of $\mathcal{P}$ (and $\mathcal{B}$). The proofs of these statements may be found in the standard LP literature, or else in Kall–Wallace [152].

DEFINITION 2.5 *For any nonempty set $\mathcal{M} \subset \mathbb{R}^n$ its polar cone is the set*

$$\mathcal{M}^P := \{z \in \mathbb{R}^n \mid z^{\mathrm{T}} x \leq 0 \, \forall x \in \mathcal{M}\}.$$

An obvious consequence of this definition is

PROPOSITION 2.6 *For any nonempty set $\mathcal{M} \subset \mathbb{R}^n$ its polar cone $\mathcal{M}^P \subset \mathbb{R}^n$ is a closed convex cone, i.e. $\mathcal{M}^P \neq \emptyset$ is a closed set such that for any two $z^{(i)} \in \mathcal{M}^P$, $i = 1, 2$, holds $\lambda_1 z^{(1)} + \lambda_2 z^{(2)} \in \mathcal{M}^P \, \forall \lambda_i \geq 0$. In particular, for any convex polyhedral cone $\mathcal{C}$ its polar cone $\mathcal{C}^P$ is a convex polyhedral cone as well.*

Proof: Obviously, $0 \in \mathcal{M}^P$ and hence $\mathcal{M}^P \neq \emptyset$ is a convex cone. For $\{z^{(\nu)} \in \mathcal{M}^P, \, \nu \in \mathbb{N}\}$ converging to $\hat{z}$ we have for any arbitrary $\hat{x} \in \mathcal{M}$ that ${z^{(\nu)}}^{\mathrm{T}} \hat{x} \leq 0 \, \forall \nu \in \mathbb{N}$ and hence $\hat{z}^{\mathrm{T}} \hat{x} = \lim_{\nu \to \infty} {z^{(\nu)}}^{\mathrm{T}} \hat{x} \leq 0$, such that $\hat{z} \in \mathcal{M}^P$, i.e. $\mathcal{M}^P$ is closed.

If $\mathcal{C}$ is a convex polyhedral cone generated by $\{d^{(1)}, \cdots, d^{(r)}\}$, with the matrix $D = (d^{(1)}, \cdots, d^{(r)})$ the polar cone of $\mathcal{C}$ is given as $\mathcal{C}^P = \{z \mid D^{\mathrm{T}} z \leq 0\}$ which, in analogy to Prop. 2.4, is a convex polyhedral cone. □

According to Proposition 2.5, using the set of feasible basic solutions $\{x^{(1)}, \cdots, x^{(r)}\}$, i.e. the vertices of $\mathcal{P}$, and the generating set $\{y^{(1)}, \cdots, y^{(s)}\}$ of $\mathcal{C}$ as described in Prop. 2.4, the LP (2.1) can now be rewritten as

$$\left.\begin{array}{lrcl} \min \displaystyle\sum_{i=1}^{r} \lambda_i \, c^{\mathrm{T}} x^{(i)} + \sum_{j=1}^{s} \mu_j \, c^{\mathrm{T}} y^{(j)} & & & \\ \text{s. t.} \displaystyle\sum_{i=1}^{r} \lambda_i & = & 1 & \\ \lambda_i & \geq & 0 \; \forall i & \\ \mu_j & \geq & 0 \; \forall j. & \end{array}\right\} \tag{2.7}$$

This representation implies the following extension of Prop. 2.2.

PROPOSITION 2.7 *Provided that $\mathcal{B} \neq \emptyset$, the LP (2.1) is solvable if and only if $c^{\mathrm{T}} y \geq 0 \; \forall y \in \mathcal{C}$, i.e. $-c \in \mathcal{C}^P$; in this case an optimal solution can be chosen as a vertex $x^{(i_0)}$ of $\mathcal{B}$ (a feasible basic solution of $\mathcal{B}$) such that $c^{\mathrm{T}} x^{(i_0)} = \min_{i \in \{1, \cdots, r\}} c^{\mathrm{T}} x^{(i)}$.*

Proof: The assumption, that $c^{\mathrm{T}} y \geq 0 \, \forall y \in \mathcal{C}$, is equivalent to the requirement that $c^{\mathrm{T}} y^{(j)} \geq 0$, $j = 1, \cdots, s$. If this condition is violated for at least one $y^{(j)}$

(e.g. for j_1), then according to (2.7) for $\mu_{j_1} \to \infty$ follows for the objective $z \to -\infty$, such that the LP is unsolvable.
If, on the other hand, the condition is satisfied, then—to solve (2.7)—we would choose $\mu_j = 0 \; \forall j$, which implies the assertion immediately. □

As a consequence we get

PROPOSITION 2.8 *If $\mathcal{B} \neq \emptyset$, and if $c^{\mathrm{T}}x \geq \gamma \; \forall x \in \mathcal{B}$ for some $\gamma \in \mathbb{R}$, then the LP $\min\{c^{\mathrm{T}}x \mid x \in \mathcal{B}\}$ is solvable.*

Proof: For any fixed $\hat{x} \in \mathcal{B}$ and an arbitrary $y \in \mathcal{C}$ it holds true that $\hat{x} + \mu y \in \mathcal{B} \; \forall \mu > 0$, and by assumption we have $c^{\mathrm{T}}\hat{x} + \mu c^{\mathrm{T}}y \geq \gamma$, which implies that $c^{\mathrm{T}}y \geq 0$ is satisfied for each $y \in \mathcal{C}$; hence the assertion follows from Prop. 2.7. □

2.3 Duality statements

To the *primal* LP in its standard formulation

$$\left.\begin{array}{rrcl} & \min c^{\mathrm{T}}x & & \\ \text{s. t.} & Ax & = & b \\ & x & \geq & 0 \end{array}\right\} \tag{2.1}$$

another LP, called its *dual*, is assigned as

$$\left.\begin{array}{rrcl} & \max b^{\mathrm{T}}u & & \\ \text{s. t.} & A^{\mathrm{T}}u & \leq & c. \end{array}\right\} \tag{2.8}$$

The technical rules according to which the dual LP (2.8) is constructed from the primal LP (2.1) may roughly be stated as follows: To the equality constraints $Ax = b$ in (2.1) correspond the free variables $u \in \mathbb{R}^m$ in (2.8); to the non-negative variables $x \in \mathbb{R}^n_+$ correspond the inequality contraints $A^{\mathrm{T}}u \leq c$ with the transpose of A as the matrix of coefficients; the right–hand–side b of the primal program yields the objective's gradient of the dual program, whereas the objective's gradient c of the primal LP turns into the right–hand–side of the dual LP; finally, to the minimization in (2.1) corresponds the maximization in (2.8).

Rewriting (2.8) into the standard form, we want to solve the problem

$$\begin{array}{rrrrcl} & \gamma := \max\{b^{\mathrm{T}}u^+ - b^{\mathrm{T}}u^-\} = -\min\{-b^{\mathrm{T}}u^+ + b^{\mathrm{T}}u^-\} & & & & \\ \text{s.t.} & A^{\mathrm{T}}u^+ & -A^{\mathrm{T}}u^- & +v & = & c \\ & u^+, & u^-, & v & \geq & 0. \end{array}$$

To this LP we assign analogously the dual LP

$$\begin{array}{rrcl} & -\max c^{\mathrm{T}} z & & \\ \text{s. t.} & Az & \leq & -b \\ & -Az & \leq & b \\ & z & \leq & 0 \end{array}$$

which, using $x := -z$, yields

$$\begin{array}{rrcl} & -\max -c^{\mathrm{T}} x = \min c^{\mathrm{T}} x & & \\ \text{s. t.} & Ax & = & b \\ & x & \geq & 0 \end{array}$$

coinciding with (2.1) again. Hence, the dual of the dual LP is the primal program again and we therefore can speak of a pair of dual LP's.

There are further relations between the primal and the dual LP which are less obvious. First, we have the *weak duality theorem.*

PROPOSITION 2.9 *For any pair of feasible solutions $\tilde{x}$ and $\tilde{u}$ of (2.1) and (2.8), respectively, it holds that $b^{\mathrm{T}}\tilde{u} \leq c^{\mathrm{T}}\tilde{x}$.*

Proof: According to the assumed feasibilities $A\tilde{x} = b$, $\tilde{x} \geq 0$, and $A^{\mathrm{T}}\tilde{u} \leq c$ it follows that

$$b^{\mathrm{T}}\tilde{u} = (A\tilde{x})^{\mathrm{T}}\tilde{u} = \tilde{x}^{\mathrm{T}}(A^{\mathrm{T}}\tilde{u}) \leq \tilde{x}^{\mathrm{T}}c.$$

□

Moreover, there is the following relation between pairs of dual LP's.

PROPOSITION 2.10 *If both of the dual LP's (2.1) and (2.8) are feasible, then both of them are solvable.*

Proof: Let $\hat{u}$ be feasible for (2.8). Then, by the weak duality theorem, $c^{\mathrm{T}}x \geq b^{\mathrm{T}}\hat{u}\ \forall x \in \mathcal{B}$. Hence Prop. 2.8 yields the solvability of (2.1). The solvability of (2.8) follows analogously. □

Finally, we have the *strong duality theorem.*

PROPOSITION 2.11 *If the primal problem is solvable, then so is the dual problem, and the optimal values of the two problems coincide.*

Proof: According to Prop. 2.3 the LP (2.1) is solvable if and only if there exists an optimal feasible basis B such that the simplex criterion (2.5)

$$d^{\mathrm{T}} = \left(c^{\{B\}^{\mathrm{T}}} B^{-1} N - c^{\{N\}^{\mathrm{T}}}\right) = \left(c^{\{B\}^{\mathrm{T}}} D - c^{\{N\}^{\mathrm{T}}}\right) \leq 0$$