Two-Stage Stochastic Semidefinite Programming: Theory, Algorithms, and Application to AC Power Flow under Uncertainty

Von der Fakultät für Mathematik der
Universität Duisburg-Essen
zur Erlangung des akademischen Grades eines
Dr. rer. nat.

genehmigte Dissertation

von

Herrn Dipl.-Math. Tobias Wollenberg
Geb. am 5. Januar 1986 in Oberhausen

Referent: Prof. Dr. Rüdiger Schultz
Korreferent: Prof. Dr. Alexander Martin

Datum der Disputation: 8. September 2016

Zusammenfassung

Betrachtet man reale Entscheidungsprobleme, die also der Wirklichkeit entstammen, so ist man fast immer mit Unsicherheiten und Risiken konfrontiert. Für konkrete Optimierungsprobleme äußert sich dies sowohl in Form von ungewissen Parametern in den Eingangsdaten, als auch durch eine unzureichende Kenntnis über die Systembeschreibung selbst. Handelt es sich um zufallsbehaftete Eingangsdaten, dessen Verteilung bekannt ist, so stellt die Stochastische Optimierung eine Vielzahl von Modellen bereit – allesamt mit dem Ziel sich gegen Unsicherheiten und Risiken abzusichern. Die am Häufigsten verwendeten stochastischen Modelle sind zweistufige Modelle. Diese gestatten folgende Kompensationsstrategie: Eine Erststufenentscheidung wird getroffen bevor das Zufallsereignis eintritt. Nach Realisierung des Zufalls können Korrekturmaßnahmen (zweite Stufe) ergriffen werden, welche häufig, aber nicht immer, als "Kompensation" verstanden werden.

Die vorliegende Arbeit behandelt Semidefinite Programme, dessen Parameter nicht mit Sicherheit bekannt sind. Der Zulässigkeitsbereich dieser Optimierungsprobleme entsteht aus dem Durchschnitt affiner oder auch allgemeinerer Gleichungen mit dem Kegel der symmetrisch und positiv semidefiniten Matrizen. Die Zielfunktion kann relativ allgemein sein, wird aber häufig, wie es auch in dieser Arbeit der Fall ist, als linear angenommen.

Es werden risikoneutrale und risikoaverse zweistufige stochastische semidefinite Optimierungsprobleme mit jeweils stetiger und gemischt-ganzzahliger Kompensation betrachtet. Wir analysieren die Struktur dieser stochastischen Optimierungsprobleme, leiten dekompositionsbasierte Lösungsverfahren her und wenden unsere Resultate auf das Problem der optimalen Kraftwerkseinsatzplanung in Wechselstromnetzen an.

Ferner beschäftigt sich diese Arbeit mit der deterministischen Kraftwerkseinsatzplanung in Wechselstromnetzen. Neben den traditionellen technischen Bedingungen an die einzelnen Kraftwerke wird auch die Physik des Wechselstroms berücksichtigt. Um global optimale Lösungen zu erhalten wird eine auf Semidefinite Programmierung (SDP) basierende Lösungsstrategie benutzt. Dieser Ansatz resultiert in einem umfangreichen semidefiniten Programm, welches zusätzlich diskrete Entscheidungsvariablen enthält. Da selbst die SDP Relaxierung dieses Optimierungsproblems zu groß ist um es mittels gängiger SDP Löser auf einmal zu lösen, wird eine effiziente und zuverlässige Methode benötigt. Es wird ein Algorithmus basierend auf dem Dekompositionsprinzip von Benders vorgeschlagen.

Ausgehend vom Energiebedarf (Last) und der Einspeisung der erneuerbaren Energien als Unsicherheitsquelle, wird ein zweistufiges stochastisches Optimierungsproblem formuliert. Das Ziel ist es, einen Kraftwerkseinsatzplan zu finden, der wirtschaftlich effektiv und robust gegenüber Veränderungen in den Daten ist. Es werden die Auswirkungen des risikoneutralen und risikoaversen Ansatzes auf die stochastische Lösung untersucht und miteinander verglichen. Um die resultierenden zweistufigen Programme zu lösen wird das Wechselstromnetz mit Hilfe des SDP Ansatzes approximiert. Dies führt zu zweistufigen stochastischen gemischt-ganzzahligen semidefiniten Programmen mit spezieller Struktur. Als Lösungsmethoden wurden die L-shaped Methode und die duale Dekomposition verwendet.

Teile der vorliegenden Arbeit wurden bei dem Journal RAIRO Recherche Operationnelle eingereicht.

Abstract

In real life decision problems, one almost always is confronted with uncertainty and risk. For practical optimization problems this is manifested by unknown parameters within the input data, or, an inexact knowledge about the system description itself. In case the uncertain problem data is governed by a known probability distribution, stochastic programming offers a variety of models hedging against uncertainty and risk. Most widely employed are two-stage models, who admit a recourse structure: The first-stage decisions are taken before the random event occurs. After its outcome, a recourse (second-stage) action is made, often but not always understood as some "compensation".

In the present thesis, the optimization problems that involve parameters which are not known with certainty are semidefinite programming problems. The constraint sets of these optimization problems are given by intersections of the cone of symmetric, positive semidefinite matrices with either affine or more general equations. Objective functions, formally, may be fairly general, although they often are linear as in the present thesis.

We consider risk neutral and risk averse two-stage stochastic semidefinite programs with continuous and mixed-integer recourse, respectively. For these stochastic optimization problems we analyze their structure, derive solution methods relying on decomposition, and finally apply our results to unit commitment in alternating current (AC) power systems.

Furthermore, deterministic unit commitment in AC power transmission systems is addressed. Beside traditional unit commitment constraints, the physics of power flow are included. To gain globally optimal solutions a recent semidefinite programming (SDP) approach is used which leads to large-scale semidefinite programs with discrete variables on top. As even the SDP relaxation of these programs is too large for being handled in an all-at-once manner by general SDP

solvers, it requires an efficient and reliable method to tackle them. To this end, an algorithm based on Benders decomposition is proposed.

With power demand (load) and in-feed from renewables serving as sources of uncertainty, two-stage stochastic programs are set up heading for unit commitment schedules which are both cost-effective and robust with respect to data perturbations. The impact of different, risk neutral and risk averse, stochastic criteria on the shapes of the optimal stochastic solutions will be examined. To tackle the resulting two-stage programs, we propose to approximate AC power flow by semidefinite relaxations. This leads to two-stage stochastic mixed-integer semidefinite programs having a special structure. To solve the latter, the L-shaped method and dual decomposition have been applied and compared.

Parts of this thesis have been submitted to the journal RAIRO Recherche Operationnelle.

Keywords: Stochastic Programming, Semidefinite Programming, Decomposition Methods, Risk Aversion, Unit Commitment, AC Power Flow.

Acknowledgements

I would like to thank all those people who have helped me, motivated me, and supported me throughout my research project and therefore made this thesis possible in the first place.

First and foremostly, I would like to express my deep appreciation and gratitude to my supervisor, Prof. Dr. Rüdiger Schultz, for mentoring me, guiding me, and letting me scientifically explore on my own. For the most part, it was up to him that I started my dissertation project. As a student I attended one of his advanced seminars where he asked me to write my diploma thesis under his supervision. After successful cooperation he encouraged me and offered me the possibility to write my thesis in his research group – and I am really glad he did. I owe him most of my knowledge about stochastic programming and mathematical optimization. Without his valuable suggestions and our discussions this thesis would not be what it is today. Thanks to him, I had the opportunity to present my ongoing research at different conferences around the world which certainly was a great experience for me. I got useful feedback and had the chance to meet researchers from all over the world and exchange thoughts with them.

My sincere thanks go to my friends and colleagues of the working group for "Optimization and Algorithmic Discrete Mathematics" at Duisburg-Essen University for the amicable working atmosphere. I thank all of them for their advices and suggestions, the helpful scientific discussions, and a lot of little things making life more beautiful. The completion of this thesis definitely would have been more difficult without them. I had a really great time working in this group.

In particular, I would like to say thank you to my parents who have done everything in all these years to transform my dreams into successes. They always believed in me, they continuously supported me, and they gave me the necessary time and freedom to become who I am. Without them I certainly would not be the person I am today.

Furthermore, I would like to thank my beloved little sister, Dr. Nadine Wollenberg, who supported and motivated me so many times. Already during our studies it was a great pleasure for me to prepare for the exams together. It is a real gift to have such a sister with whom I could discuss mathematical issues and share all my problems.

Many thanks are owed to my wife, Vera, for her great understanding and cheering-up during all the ups and downs of my PhD project. She is always there for me when I need her. She supported me wherever she can. I am so lucky for having her in my life.

Last but not least I gratefully acknowledge the support of the Faculty of Mathematics of the University of Duisburg-Essen and the German Research Foundation (DFG) within the collaborative research center TRR 154 “Mathematical Modeling, Simulation and Optimization Using the Example of Gas Networks”.

Tobias Wollenberg

April, 2016

Contents

List of Figures

List of Tables

Abbreviations

AC	**A**lternating **C**urrent
DC	**D**irect **C**urrent
IEEE	**I**nstitute of **E**lectrical and **E**lectronics **E**ngineers
KKT	**K**arush-**K**uhn-**T**ucker
LP	**L**inear **P**rogram
MILP	**M**ixed-**I**nteger **L**inear **P**rogram
NP	**N**on-deterministic **P**olynomial-time
OPF	**O**ptimal **P**ower **F**low
PDE	**P**artial **D**ifferential **E**quation
PSH	**P**umped-**S**torage **H**ydroelectricity
QC	**Q**uadratic **C**onvex
QCQP	**Q**uadratically **C**onstraint **Q**uadratic **P**rogram
QP	**Q**uadratic **P**rogram
SDP	**S**emidefinite **P**rogram
SOCP	**S**econd **O**rder **C**one **P**rogram

Symbols

Sets

$\mathbb{R}^n$	real column vector of dimension n
$\mathbb{R}^n_+$	nonnegative real column vector of dimension n
$\mathbb{R}^n_-$	nonpositive real column vector of dimension n
$\mathbb{Z}^n$	integer column vector of dimension n
$\mathbb{Z}^n_+$	nonnegative integer column vector of dimension n
$\mathbb{C}^n$	complex column vector of dimension n
$\mathcal{S}^n$	the set of symmetric matrices in $\mathbb{R}^{n\times n}$
$\mathcal{S}^n_+$	the set of symmetric positive semidefinite matrices in $\mathbb{R}^{n\times n}$
$\mathcal{S}^n_{++}$	the set of symmetric positive definite matrices in $\mathbb{R}^{n\times n}$
$\mathbb{H}^n$	the set of Hermitian matrices in $\mathbb{C}^{n\times n}$
$\mathbb{H}^n_+$	the set of positive semidefinite Hermitian matrices in $\mathbb{C}^{n\times n}$
$B^n_1(0)$	unit ball in $\mathbb{R}^n$
Ω	a sample space (set of all possible outcomes)
$\mathcal{F}$	a set of events
$\mathbb{P}$	the set of Borel probability measures
$\mathcal{N}$	the set of net buses
$\mathcal{L}$	the set of net flow lines

$\mathcal{N}(k)$	the set of adjacent net buses to bus $k \in \mathcal{N}$
$\mathcal{G}$	the set of generators
$\{1, \ldots, I\}$	the set of coal fired blocks
$\{1, \ldots, R\}$	the set of gas turbines
$\{1, \ldots, H\}$	the set of pumped-storage plants
$\{1, \ldots, T\}$	the set of discrete time intervals
$\mathcal{U}_I$	the set of feasible decisions for coal fired blocks
$\mathcal{U}_R$	the set of feasible decisions for gas turbines
$\mathcal{U}_H$	the set of feasible decisions for pumped-storage plants

Data

p_{D_k}	active power demand at bus $k \in \mathcal{N}$
q_{D_k}	reactive power demand at bus $k \in \mathcal{N}$
$P_k^{\min}$	minimum output of active power at generator $k \in \mathcal{N}$
$P_k^{\max}$	maximum output of active power at generator $k \in \mathcal{N}$
$Q_k^{\min}$	minimum output of reactive power at generator $k \in \mathcal{N}$
$Q_k^{\max}$	maximum output of reactive power at generator $k \in \mathcal{N}$
$W_h^{\max}$	maximum active power consumption
	of pumped-storage plant $h \in H$
$\overline{W}_h^{\max}$	maximum reactive power consumption
	of pumped-storage plant $h \in H$
c_{k0}	start-up costs for thermal generator $k \in \mathcal{G} \backslash H$
c_{k1}	linear cost coefficient for thermal generator $k \in \mathcal{G} \backslash H$
c_{k2}	quadratic cost coefficient for thermal generator $k \in \mathcal{G} \backslash H$

τ_i	minimum downtime for generator $k \in \mathcal{G} \backslash H$
$l_h^{\max}$	maximum upper dam fills in active power
	for pumped-storage plant $h \in H$
l_h^{in}	initial fill in active power of the upper dam
	to pumped-storage plant $h \in H$
l_h^{end}	final fill in active power of the upper dam
	to pumped-storage plant $h \in H$
η_h	pumping efficiency factor for pumped-storage plant $h \in H$
y_{lm}	(complex) admittance between the buses l and m
y_{kk}	(complex) admittance-to-ground at bus $k \in \mathcal{N}$
g_{lm}	conductance on transmission line $(l, m) \in \mathcal{L}$
b_{lm}	susceptance on transmission line $(l, m) \in \mathcal{L}$
b_{lm}^0	shunt element on transmission line $(l, m) \in \mathcal{L}$
$Y = G + jB$	net corresponding nodal admittance matrix
$S_{lm}^{\max}$	maximum apparent power bound
	on transmission line $(l, m) \in \mathcal{L}$
$P_{lm}^{\max}$	maximum active power bound
	on transmission line $(l, m) \in \mathcal{L}$
$\Delta V_{lm}^{\max}$	maximum difference in voltage magnitude
	between bus l and m

Variables

$u_k^t \in \{0, 1\}$	off/on switching decision for generator $k \in \mathcal{G} \backslash H$
	at time interval t

$r_k^t \in \{0, 1\}$	auxiliary variable indicating whether $k \in \mathcal{G} \backslash H$ is switched from non-operating into operating mode at time interval t
$l_h^t \in \mathbb{R}_+$	fill in active power of the upper dam at pumped-storage plant $h \in H$ at time interval t
$s_k^t \in \mathbb{C}$	apparent power at bus $k \in \mathcal{N}$ at time interval t
$p_k^t \in \mathbb{R}$	active power at net bus $k \in \mathcal{N}$ at time interval t
$q_k^t \in \mathbb{R}$	reactive power at net bus $k \in \mathcal{N}$ at time interval t
$w_h^t \in \mathbb{R}$	consumed active power at pumped-storage plant $h \in H$ at time interval t
$\overline{w}_h^t \in \mathbb{R}$	consumed reactive power at pumped-storage plant $h \in H$ at time interval t
$V_k^t \in \mathbb{C}$	complex voltage at bus $k \in \mathcal{N}$ at time interval t
$E_k^t \in \mathbb{R}$	real part of the complex voltage at bus $k \in \mathcal{N}$ at time interval t
$F_k^t \in \mathbb{R}$	imaginary part of the complex voltage at bus $k \in \mathcal{N}$ at time interval t
$\lvert V_k^t \rvert \in \mathbb{R}$	voltage magnitude at bus $k \in \mathcal{N}$ at time interval t
$\theta_k^t \in \mathbb{R}$	voltage angle at bus $k \in \mathcal{N}$ at time interval t

Notations

A^T	transpose of the matrix (vector) A
$A \succeq B$	$A - B \in \mathcal{S}_+$, i.e., $A - B$ is positive semidefinite
$\mathrm{tr}(A) = \sum_{i=1}^n a_{ii}$	the trace of the square matrix A
$A \bullet B$	the trace of the matrix product $A^T B$

$\text{rank}(A)$	largest collection of linearly independent columns/rows of A
$\mathfrak{R}(z)$	real part of the complex vector $z \in \mathbb{C}^n$
$\mathfrak{I}(z)$	imaginary part of the complex vector $z \in \mathbb{C}^n$
$e_i \in \mathbb{R}^n$	i-th standard basis vector in $\mathbb{R}^n$
$I_n \in \mathbb{R}^{n \times n}$	the identity matrix of size $n \times n$
μ	a measure
ξ	an $\mathbb{R}^s$-valued random variable
$\mathbb{E}$	the expectation
$\mathcal{R}$	a risk measure
Φ	optimal value function
$\mathbb{P}(A)$	probability of event A
f'	the derivative of the function f
∂	subdifferential
$\lceil x \rceil$	the ceiling function (round up), i.e., $q = \lceil x \rceil$ is the smallest integer that is not less than x
$\lfloor x \rfloor$	the floor function (round down), i.e., $q = \lfloor x \rfloor$ is the largest integer that does not exceed x
$\text{round}(x)$	the element-wise rounding of the vector x
$\text{ri}(S)$	the relative interior of the set S
$\Vert \cdot \Vert$	a norm
$\Vert \cdot \Vert_2$	the Euclidean norm: $\Vert x\Vert_2 := \sqrt{x_1^2 + ... + x_n^2}$ for $x \in \mathbb{R}^n$
$\Vert \cdot \Vert_\infty$	the maximum norm: $\Vert x\Vert_\infty := \max(\vert x_1\vert, ..., \vert x_n\vert)$ for $x \in \mathbb{R}^n$
$\Vert \cdot \Vert_F$	the Frobenius norm: $\Vert A\Vert_F := \sqrt{A \bullet A}$ for $A \in \mathbb{R}^{n \times n}$

Dedicated to my beloved parents and family.
For their endless love, support and encouragement.

Chapter 1

Introduction

In recent years, research in stochastic programming has moved into various new directions. This concerns both theory and applications. Without aiming at completeness, one could mention on the theory side: risk aversion with risk measures or stochastic orders, stochastic programs in mixed-integer, semidefinite, bilevel, or partial differential equation (PDE) constrained optimization as well as scenario tree construction and reduction. Fields of applications include finance, logistics, and energy optimization in the broadest sense.

The present thesis contributes to this development with a clear accent on pushing ahead mathematical foundations at the interface of stochastic optimization and semidefinite programming. Special attention is paid to the appropriate treatment of risk. Optimal power flow under uncertainty in AC networks serves as the major field of industrial application. The thesis draws on seminal work in risk neutral semidefinite stochastic programming [1], [2], on recent progress in power flow optimization [3], and on risk aversion by forming objective functions involving risk measures [4], [5].

To capture risk aversion in a minimization context, we resort to an intuitive measure which is the probability of a random quantity to exceed a preassigned critical level. This measure, called excess or exceedance probability, has been analyzed in two-stage stochastic linear mixed-integer stochastic programming in [5]. In reliability analysis it has a role in various fields of engineering of which a more recent one is seismic risk analysis [6]. Here, the total cost of damage and retrofit caused by an earthquake is the random quantity of interest, and risk is measured by the probability of this quantity to exceed a threshold value.

Our motivation to investigate risk aversion from the viewpoint of excess probabilities comes from the simultaneous treatment of unit commitment and AC load flow under uncertainty of power demand and in-feed from renewables in power management. The "geographic split of the two", meaning that the locations where electricity is produced from renewables and the locations where electricity is consumed are distant apart, has given transportation via the electric grid increased importance. Therefore, it is reasonable to expect that previous unit commitment models neglecting the grid at all or using DC approximations of the AC load flow are too coarse.

Another recent development, this time in power flow methodology, has spurred our interest in incorporating risk aversion into power flow optimization models. In [3] the authors formulate AC load flow by means of convex semidefinite constraints and some rank condition (for optimal power flow problems semidefinite relaxation techniques were first applied in [7]). With a fixed commitment of generating units and for a fixed point in time, they solve the dual to the mentioned rank constrained program. When heading for a primal solution, a good many times, their proposed solution approach has the ability to retrieve the relaxed rank-one condition[1], such that it enables the opportunity to solve (non-convex) power flow problems to global optimality.

Motivated by the uncertain parameters typically developing in time, we study unit commitment over some time horizon. We extend existing unit commitment models by putting simultaneous consideration of AC load flow and stochastic uncertainty on top of the model. Together with the semidefinite programming approach in [3], this will lead to two-stage mixed-integer stochastic semidefinite programs whose structure will be analyzed and for which potential decomposition algorithms will be presented.

1.1 Stochastic Programming

When considering real world optimization problems in areas such as finance, engineering, telecommunication, or medicine, one often is faced with an inexact knowledge about the system description itself as well as its input data (such as consumer demand, production costs, resources, and capacities). For making "good"

[1]In the dual of the dual semidefinite program.

or "acceptable" decisions under this lack of information stochastic programming (which traces its roots to Dantzig [8]) is qualified. It provides several approaches for modeling and tackling programs involving uncertain data. For an introduction into basic aspects of stochastic programming, we refer to the books by Prékopa [9], Birge and Louveaux [10], Ruszczyński and Shapiro [11], and Shapiro, Dentcheva and Ruszczyński [12]. A collection of real applications using stochastic programming is presented in [13].

The Benefit of Stochastic Programming Models: Stochastic programming models enable managing and measuring the risk of decisions [4]. Several simulation studies confirm (cf. [13], [14], [15]) that solutions of stochastic programs, which hedge against uncertainty, are much more useful than of deterministic ones (whose solutions, in many cases, could be very misleading). Solutions of stochastic programs are normally never optimal after the fact. However, in practice, they are hardly ever really bad and fairly robust with respect to the changes in the data.

A class of random programs considered in stochastic programming reads:

$$\text{“min”}\{g_0(x,\xi) : g_i(x,\xi) \le 0,\ i = 1,...,m,\ x \in X\}. \tag{1.1}$$

Here, ξ is a random vector on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^s$, $g_i(x,\cdot) : \Omega \to \mathbb{R}$, $i = 0,...,m$ are random variables themselves, and $X \subseteq \mathbb{R}^n$ is some feasible region.

A common adoption in stochastic programming that will be assumed throughout this thesis is to suppose that the uncertainty is exogenous. This means that all decisions taken by the decision maker do not affect future outcomes of the underlying random variables. Practically relevant stochastic programs with endogenous uncertainty that could be managed are presented in [16] and [17], for instance.

The information constraint that the probability distribution of $\mathbb{P}$ does not depend on the decisions $x \in X$ is referred to as the nonanticipativity constraint. This condition admits the principle of taking a decision on x before knowing the realization of ξ. It can be written in diverse forms, which may be beneficial in several situations, and leads to the well known dual decomposition methods[2].

[2]An overview of dual methods for two-stage and multi-stage stochastic programming is given in [18].

It is noted that (1.1) is not well-posed, namely, as long as ξ is unknown, it exhibits for every fixed $\overline{x} \in X$ another random variable $g_0(\overline{x}, \xi)$, such that "min" in (1.1) can be seen as selecting "the best" member among this family of random variables that fulfills the constraints $g_i(\overline{x}, \xi) \leq 0, \forall i$ which are also not clear at all.

Robust Optimization: Assuming that $\Omega \subset \mathbb{R}^s$, the most conservative view for instance would lead to the (robust) worst-case model

$$\min\{g(x) := \max_{\xi \in \Omega} g_0(x, \xi) : g_i(x, \xi) \leq 0,\ i = 1, ..., m, \forall \xi \in \Omega,\ x \in X\}. \quad (1.2)$$

A comprehensive account of robust optimization is given in the book by Ben-Tal, El Ghaoui and Nemirovski [19]. This worst case (non-stochastic) approach, however, could be "too restrictive" if the set Ω is large. If worst comes to the worst, (1.2) may be inconsistent.

Probabilistic Constrained Optimization: A possible remedy is chance constraints (or probabilistic constraints) as introduced by Charnes and Cooper [20], i.e., the probability that $g_i(x, \xi)$ exceeds the prescribed target level $\eta := 0$ is less than a significance level $\alpha_i \in (0, 1)$:

$$\mathbb{P}[\{\xi \in \Omega : g_i(x, \xi) > \eta\}] \leq \alpha_i. \quad (1.3)$$

Typically, the significance level α_i is chosen quite small, such that (1.3), which is equivalent to $1 - \alpha_i \leq \mathbb{P}[\{\xi \in \Omega : g_i(x, \xi) \leq \eta\}]$, request the system to be consistent with high probability (for a detailed account on probabilistic constrained optimization see [21] and [22]).

Ranking by Expectation: Moreover, instead of evaluating $g_0(x, \xi)$ for $x \in X$ by its worst outcome for $\xi \in \Omega$, one could minimize its expectation $\mathbb{E}_\xi[g_0(x, \xi)]$. This approach can be justified by the law of large numbers. It states that by performing the process a large number of times for a given decision $x \in X$, the average of total costs will converge with probability one to the expectation $\mathbb{E}_\xi[g_0(x, \xi)]$. This means that the minimum obtained (expected value of all decisions) will be optimal on average.

Ranking by Risk Measures: Ranking by expectation alone has the drawback of neglecting the variability at all, such that a decision x may be chosen whose associated random variable takes unfavorable values "too often". Instead of taking into account the expected value alone, an option is to consider mean-risk models [23], leading to the following objective function

$$\mathbb{E}_\xi[g_0(x,\xi)] + \mathcal{R}_\xi[g_0(x,\xi)].$$

This function prefers decisions x whose risk of adverse costs is not "too high" in the sense that the risk measure $\mathcal{R}_\xi[g_0(x,\xi)]$ makes it "acceptable" for the decision maker. Regarding this risk functional, a variety of choices is possible, here, one could briefly mention the value-at-risk and conditional value-at-risk [24] or the exceedance probability as defined in (1.3).

Stochastic Dominance Constraints: Furthermore, instead of ranking random variables by statistical parameters, stochastic dominance constraints [25], [26], which induce a partial order in the space of real random variables, are an option for risk-averse decision making (the study of stochastic dominance constraints in terms of optimization was initiated by Dentcheva and Ruszczyński in [27] and [28]). The idea is to shift the random variables through the constraints and "accept" those whose risk is preferable compared to some given benchmark. Preferring smaller outcomes to larger ones, the concept of stochastic dominance states that x dominates y to first degree ($x \succeq_1 y$) if

$$F_x(\eta) \geq F_y(\eta), \quad \forall \eta \in \mathbb{R},$$

where $F_x(\eta) := \mathbb{P}(\{\xi \in \Omega : f(x,\xi) \leq \eta\})$ denotes the distribution function of the random variable $f(x,\xi)$. Further, x dominates y to second degree ($x \succeq_2 y$) if

$$F_x^{(2)}(\eta) \geq F_y^{(2)}(\eta), \quad \forall \eta \in \mathbb{R},$$

where $F_x^{(2)}(\eta) := \int_{-\infty}^{\eta} F_x(\xi) d\xi$, for $\eta \in \mathbb{R}$. Stochastic dominance constraints may be employed in portfolio optimization [29] where benchmarks arise naturally by stock indices. For deeper insights into comparison methods for random variables, we refer to the book by Müller and Stoyan [30].

Two-stage Stochastic Programming: A further way to tackle random programs (1.1) is to apply two-stage stochastic programming. This approach can be pursued if the problem description admits some recourse (or compensation) strategy. Instead of claiming that $g_i(x, \xi) \le 0, \forall i, \forall \xi \in \Omega$, recourse actions that possesses additional costs are accepted or can be naturally modeled due to the problem characteristics. In doing so, a multifunction $\mathcal{G} : \mathbb{R}^n \times \Omega \rightrightarrows \mathbb{R}^m$ could be introduced that yields the following two-stage random program

$$\text{“}\min_{x \in X}\text{”} \left\{ f(x, \xi) := \min\nolimits_{y \in \mathcal{G}(x,\xi)} g(x, y, \xi) \right\}. \tag{1.4}$$

Here, we have two types of decision variables which are naturally classified as first-stage and second-stage decision variables. In the first-stage, before the realization of the underlying random variable becomes known, we decide on $x \in X$[3]. Then, in the second stage, after having selected $\overline{x} \in X$, and after observing the outcome of the underlying random variable, the optimal corrective (recourse) action is taken, i.e., $y(\overline{x}, \xi) \in \arg\min_{y \in \mathcal{G}(\overline{x},\xi)} g(\overline{x}, y, \xi)$. Due to the information constraint

$$\text{decide } x \quad \to \quad \text{observe } \xi \quad \to \quad \text{decide } y(x, \xi) \tag{1.5}$$

first-stage variables are often referred to as here-and-now decisions/solutions and second-stage variables are referred to as wait-and-see decisions/solutions.

This thesis: In this thesis, we will limit ourselves to two-stage programs (1.4) in which linking constraints between first-stage variables x and second-stage variables y are linear and all remaining relationships are modeled in terms of (linear) semidefinite programming. For semidefinite programming, an accessible introduction is Chapter 2 of [31] as well as [32], whereas the recent state-of-the-art can be obtained from [33]. Two-stage semidefinite programs are introduced in Section 2.1.

Among the criteria for ranking $f(x, \xi)$ we have selected ranking by expectation and risk aversion via exceedance probabilities (1.3). The exceedance probability was chosen since in reliability analysis it has a role in various fields of engineering of

[3]In order to do so, these decision variables should be nonanticipative, i.e., independent on future observations.

which a more recent one is seismic risk analysis [6]. Being more precise, two-stage stochastic programs of the following form are considered:

(1) risk neutral:

$$\min\left\{\underbrace{\mathbb{E}_\xi[f(x,\xi)]}_{:=\mathcal{Q}_{\mathbb{E}}(x)} : x \in X\right\}, \tag{1.6}$$

(2) risk averse:

$$\min\left\{\underbrace{\mathbb{E}_\xi[f(x,\xi)]}_{:=\mathcal{Q}_{\mathbb{E}}(x)} + \rho \cdot \underbrace{\mathbb{P}_\xi[\{f(x,\xi) > \eta\}]}_{:=\mathcal{Q}_{\mathbb{P}_\eta}(x)} : x \in X\right\}, \tag{1.7}$$

where $\rho \geq 0$ is a prefixed parameter.

1.2 Contributions

The contribution of this thesis can be divided into three parts. The first part concerns theory and decomposition methods for two-stage stochastic semidefinite programs. Basic properties and characteristics of two-stage stochastic semidefinite programs with continuous recourse and mixed-integer recourse, respectively, are discussed (see Chapter 2). To solve those stochastic programs efficiently, decomposition algorithms, dual decomposition [34] and the L-shaped method [35], are presented (see Chapter 3).

The second part of this thesis concerns unit commitment with AC load flow. Unit commitment including optimal AC power flow (AC unit commitment) is formulated in Chapter 4. A modeling contribution is the approximation of AC unit commitment via mixed-integer semidefinite programming (see Chapter 5). To tackle the resulting large-scale mixed-integer semidefinite programs, a solution approach based on Benders decomposition is proposed (see Chapter 5).

The third part of this thesis covers AC unit commitment under uncertainty of both power demand and in-feed of renewables (this problem is addressed in Chapter 6). Here, two-stage stochastic programming is employed for planning a unit commitment schedule under uncertainty, leading us to risk neutral and risk averse two-stage stochastic unit commitment problems (see Section 6.1). Approximating further AC load flow by semidefinite relaxations leads to two-stage mixed-integer semidefinite programs, for which a decomposition algorithm is presented. For two-stage stochastic unit commitment, dual decomposition [34] is improved significantly by introducing feasibility cuts (see Section 6.1.3).

The efficiency of the proposed algorithms is confirmed by various case studies (see Chapter 7). Moreover, the value of the stochastic solution (Section 7.4) and the impact of different, risk neutral and risk averse, stochastic criteria on the shapes of the optimal solutions (Section 7.5) is illustrated.

Chapter 2

Two-Stage Stochastic Semidefinite Programming

This chapter provides basic properties and characteristics of two-stage stochastic semidefinite programs with continuous and mixed-integer recourse, respectively. Risk neutral (expectation based) and risk averse (excess probability mean-risk) models are formulated and investigated. In case where the underlying probability measure follows a finite discrete probability distribution, we derive deterministic equivalents. This leads to large-scale mixed-integer semidefinite programs for which primal and dual decomposition methods are introduced.

Stochastic (mixed-integer) linear programs have been intensively studied in the literature – their properties as well as their structure have been analyzed indepth. For a comprehensive discussion of stochastic linear programs and its theory, we refer to the textbook by Kall and Mayer [36] whereas an overview of stochastic integer programming is given in [37].

Since semidefinite programming differs from linear programming in some essential aspects, the results from stochastic (mixed-integer) linear programming cannot be adopted without further ado. Whilst certain parts of the analysis of stochastic semidefinite programs will follow the traditional lines of argumentation in stochastic linear programming fairly closely, others fail or have to be extended. This chapter will place accent on those topics.

2.1 Two-Stage Stochastic Semidefinite Programs

Consider the two-stage random program (1.4). For that, let us assume that the feasible set of first-stage decisions and the feasible set of second-stage decisions, respectively, are (non-empty) spectrahedra, i.e., intersections of solution sets of affine matrix inequalities with the cone of positive semidefinite matrices. In order to meet a given random vector $z(\xi)$, assume further that first-stage actions $x \in X \subseteq \mathcal{S}^n_+$ (the set $\mathcal{S}^n_+$ denotes the n-dimensional positive semidefinite cone) may be compensated by

$$Tx + Wy = z(\xi), \quad y \in \mathcal{S}^m_+.$$

Here, $z(\xi)$ should be a random vector on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^s$ whose distribution does not depend on x. With linear compensation costs $q^T y$ for the selected (second-stage) variables

$$y(x, \xi) \in \{y \in \mathcal{S}^m_+ : Wy = z(\xi) - Tx\},$$

as well as linear first-stage costs $c^T x$ for $x \in X$, this yields the following two-stage (semidefinite) random program[1]:

$$\text{“}\min_{x,y}\text{”} \Big\{ c^T x + q^T y : Wy = z(\xi) - Tx,\ y \in \mathcal{S}^m_+,\ x \in X \Big\}. \tag{2.1}$$

Here, the second-stage variable y can be interpreted as compensating or recourse action, which, in the seminal stochastic programming literature, led to the notion of a stochastic program with recourse. To emphasize the two-stage character of decision making we rewrite (2.1) as:

$$\text{“}\min_{x}\text{”} \Big\{ c^T x + \underbrace{\text{“}\min_{y}\text{”} \big\{ q^T y : Wy = z(\xi) - Tx,\ y \in \mathcal{S}^m_+ \big\}}_{=:\Phi(z(\xi) - Tx)} : x \in X \Big\}. \tag{2.2}$$

Here, the function

$$\Phi : \mathbb{R}^s \to \mathbb{R}, \qquad t \mapsto \min \big\{ q^T y : Wy = t,\ y \in \mathcal{S}^m_+ \big\}, \tag{2.3}$$

[1]For the two-stage random program (1.4) this implies $g(x, y, \xi) := c^T x + q^T y$ and $\mathcal{G}(x, \xi) := \{y : Wy = z(\xi) - Tx,\ y \in \mathcal{S}^m_+\}$.

is the optimal-value function of the inner (second-stage) semidefinite program seen as a parametric optimization problem with parameter t.

Remark 2.1. Note that c plus the rows $t_1, ..., t_s$ in the (technology) matrix T and q plus the rows $w_1, .., w_s$ in the (recourse) matrix W can expected to be symmetric matrices[2] in $\mathcal{S}^n$ and $\mathcal{S}^m$, respectively (see Remark A.26).

Stochastic Linear Programming: In stochastic linear programming the second-stage value function (2.3) becomes a parametric linear program:

$$t \mapsto \min \left\{ q^T y : Wy = t,\ y \in \mathbb{R}^m_+ \right\}. \tag{2.4}$$

This means $\Phi(t) := \min \left\{ q^T y : Wy = t,\ y \in \mathbb{R}^m_+ \right\}$. By linear programming theory, this function is real valued on $\mathbb{R}^s$ iff the following holds:

(1) $W(\mathbb{R}^m_+) = \mathbb{R}^s$,

(2) $\{u \in \mathbb{R}^s : W^T u \leq q\} \neq \emptyset$.

Further, for $t \in W(\mathbb{R}^m_+)$, non-emptiness of $\{u \in \mathbb{R}^s : W^T u \leq q\}$ and linear programming duality provides

$$\min\{q^T y : Wy = t, y \in \mathbb{R}^m_+\} = \max\{t^T u : W^T u \leq q\} = \max\nolimits_{j=1,...,N} d_j^T t,$$

where $d_1, ..., d_N$ denote the vertices of the polyhedron $\{u \in \mathbb{R}^s : W^T u \leq q\}$. Hence, in assuming (1) and (2), (2.4) is piecewise linear convex and Lipschitz continuous. The latter is due to $\max_{j=1,...,N} d_j^T t \leq \max_{j=1,...,N} \|d_j\| \|t\|$.

Optimal value functions that are induced by mixed-integer recourse, i.e., by (2.4) in which some variables are required to be integers, are investigated among others by Bank and Mandel in [38], and by Blair and Jeroslow in [39] and [40].

2.1.1 Basic Properties

As in stochastic linear programming (cf. [41]), we have basic assumptions ensuring that model ingredients are well-defined. More specifically, assume what is called

[2]The vectors $c, t_1, ..., t_s \in R^{n^2}$ and $q, w_1, ..., w_s \in \mathbb{R}^{m^2}$ arise by vectorizing corresponding symmetric matrices in $\mathcal{S}^n$ and $\mathcal{S}^m$, respectively. A matrix is converted into a column vector by stacking its columns on top of each other. The linear transformation that serves this purpose for $A \in \mathcal{S}^n$ is denoted by $\mathrm{vec}(A)$.

complete recourse in stochastic programming

$$W(\mathcal{S}_+^m) = \mathbb{R}^s. \tag{2.5}$$

This serves the purpose to have a non-empty feasible set for the second-stage optimization problem for any right-hand side. Moreover, it guarantees that future outcomes of the random variable do not affect the set of feasible first-stage decisions. If, furthermore

$$M_D := \{u \in \mathbb{R}^s : W^T u \prec q\} \neq \emptyset, \tag{2.6}$$

then, by duality (cf. Theorem A.29) the second-stage problem is always solvable and attainment of its minimum is guaranteed. Here, "$\prec$" is the Löwner partial order: $A \succ 0$ denotes that A is symmetric and positive definite, while $A \succeq 0$ denotes that A is symmetric and positive semidefinite. The expressions $A \succ B$ and $A \succeq B$ are seen as $A - B \succ 0$ and $A - B \succeq 0$, respectively.

The cone of positive semidefinite matrices is not finitely generated (see Appendix A) and thus not polyhedral. Therefore, its image under a linear map is not always closed [42]. Without the existence of a Slater point ($M_D \neq \emptyset$) it might happen that the minimum in (2.3) is not attained for almost all t (see Example 2.1). If so, well-posedness of (2.2) cannot be guaranteed.

Example 2.1. *Consider for $t \in \mathbb{R}$ the program*

$$\min \left\{ \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} y : \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} y = t,\ y \in \mathcal{S}_+^2 \right\}.$$

Selecting for $\bar{t} \in \mathbb{R}$, $\bar{y} = \begin{bmatrix} |\bar{t}| & \bar{t} \\ \bar{t} & |\bar{t}| \end{bmatrix} \in \mathcal{S}_+^2$, then $\begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \bar{y} = \bar{t}$, such that $\begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \left(\mathcal{S}_+^2\right) = \mathbb{R}$. Moreover, the dual to the above program reads

$$\max \left\{ t \cdot u : \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \cdot u \preceq \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \right\},$$

which is feasible for $u = 0$, only. Thus, the dual optimal value is 0, attained at $u = 0$. Further, as the primal is strictly feasible for any right-hand side $t \in \mathbb{R}$ (select $y_{11} = y_{22} = |t| + 1, y_{12} = y_{21} = t$), we obtain by duality that the infimum of *the primal has to be zero as well. Furthermore, for $t \in \mathbb{R}$,*

$$\begin{bmatrix} y_{11} & \frac{t}{2} \\ \frac{t}{2} & y_{22} \end{bmatrix} \succeq 0 \quad \Longleftrightarrow \quad y_{11} \geq 0, \quad y_{22} \geq 0, \quad y_{11}y_{22} - \left(\frac{t}{2}\right)^2 \geq 0,$$

yields the lower bound $y_{11} \geq \frac{t^2}{4y_{22}}$. For arbitrary $t \in \mathbb{R}\backslash\{0\}$ this bound is zero if $y_{22} \to \infty$, so that the primal optimal value is not attained.

Remark 2.2. Instead of claiming complete recourse (2.5) we might require:

$$\begin{aligned} &\forall x \in X \Rightarrow \{y \in \mathcal{S}_+^m : Wy = z - Tx\} \neq \emptyset \quad \text{a.s.,} \\ \Leftrightarrow \quad &\forall x \in X \text{ holds: } \left[W^T u \succeq 0 \;\Rightarrow\; (z - Tx)^T u \geq 0 \text{ a.s.}\right], \end{aligned}$$

where the equivalence is due to the semidefinite Farkas lemma (cf. Lemma A.34), Lemma A.35, and $M_D \neq \emptyset$. Since this requirement is a joint restriction on x and the range of the underlying random variable, simultaneously, it could be difficult to verify. Therefore, in applications, it is often preferred to assume (2.5), which is a requirement on the operator W, only.

If we assume non-emptiness of the set $M_D := \{u \in \mathbb{R}^s : W^T u \prec q\}$, then compactness of its closure is equivalent to complete recourse (2.5).

Lemma 2.3. *Assume $M_D := \{u \in \mathbb{R}^s : W^T u \prec q\} \neq \emptyset$. Then $M_D^{\preceq} := \{u \in \mathbb{R}^s : W^T u \preceq q\}$ is compact if and only if $W(\mathcal{S}_+^m) = \mathbb{R}^s$.*

Proof. According to the Heine-Borel theorem, the set $M_D^{\preceq}$ (as a subset of the euclidean space $\mathbb{R}^s$) is compact if it is closed and bounded.

$M_D^{\preceq}$ is closed due to the continuity of the eigenvalue[3] and the fact that a matrix is positive semidefinite if and only if all its eigenvalues are non-negative.

In order to show that $M_D^{\preceq}$ is bounded if $W(\mathcal{S}_+^m) = \mathbb{R}^s$, let us assume that there exists a sequence $(v_n)_{n\in\mathbb{N}} \in M_D^{\preceq}$ with $\|v_n\| \to \infty$. If we further define $\widetilde{u}_n := v_n/\|v_n\|$, then $\|\widetilde{u}_n\| \to 1$, such that there is a subsequence converging to some $\widetilde{u} \neq 0$. The latter limit satisfies $W^T\widetilde{u} \preceq 0$ which is seen as follows: By $v_n \in M_D^{\preceq}$

[3] The eigenvalues are the solutions of the characteristic polynomial and therefore depend continuously on the matrix elements.

we have $q - W^T v_n \succeq 0, \forall n \in \mathbb{N}$. Dividing by $\|v_n\|$ yields

$$\underbrace{\lim_{n\to\infty} \frac{1}{\|v_n\|} q - W^T \frac{v_n}{\|v_n\|}}_{=-W^T\widetilde{u}} \succeq 0.$$

Now, $\overline{u} \in M_D^{\preceq}$ implies $\overline{u} + \alpha\widetilde{u} \in M_D^{\preceq}$ for $\alpha \geq 0$. Therefore,

$$\widetilde{u}^T(\overline{u} + \alpha\widetilde{u}) = \widetilde{u}^T\overline{u} + \alpha\|\widetilde{u}\|^2 \to \infty, \quad \text{for } \alpha \to \infty,$$

verifying

$$\sup\left\{\widetilde{u}^T u : W^T u \preceq q\right\} = \infty.$$

By duality, the primal feasible set $\left\{y \in \mathcal{S}_+^m : Wy = \widetilde{u}\right\}$ then has to be empty which contradicts the assumption $W(\mathcal{S}_+^m) = \mathbb{R}^s$.

Let vice versa $M_D^{\preceq}$ be compact. Then once again by duality for arbitrary $\overline{t} \in \mathbb{R}^s$, there exists $u^* \in M_D^{\preceq}$ with

$$\min\{q^T y : Wy = \overline{t},\ y \in \mathcal{S}_+^m\} = u^{*T}\overline{t},$$

which implies $\overline{t} \in W(\mathcal{S}_+^m)$ and thus $W(\mathcal{S}_+^m) = \mathbb{R}^s$. $\square$

Remark 2.4. If complete recourse and $M_D \neq \emptyset$ is assumed, then, due to compactness of $M_D^{\preceq}$ (Lemma 2.3) attainment of the supremum

$$\sup\{t^T u : W^T u \preceq q\}$$

is guaranteed (for arbitrary $t \in \mathbb{R}^s$).

2.1.1.1 The Optimal Value Function

In program (2.2) each first-stage decision $x \in X$ defines another random variable $c^T x + \Phi(z(\xi) - Tx)$. Criteria to select the "best" random variable among

$$(c^T x + \Phi(z(\xi) - Tx))_{x\in X},$$

are offered by stochastic programming. As for almost all of them the optimal value function (2.3) appears as an integrand of a suitable integral, it plays a fundamental role when analyzing the structure of stochastic programs. This section discusses some of its fundamental properties.

Lemma 2.5. *Assume $W(\mathcal{S}_+^m) = \mathbb{R}^s$ and $M_D := \{u \in \mathbb{R}^s : W^T u \prec q\} \neq \emptyset$. Then the optimal value function $\Phi : \mathbb{R}^s \to \mathbb{R},\ t \mapsto \min\left\{q^T y\ :\ Wy = t, y \in \mathcal{S}_+^m\right\}$ is finite, convex and Lipschitz continuous on $\mathbb{R}^s$.*

Proof. Due to $W(\mathcal{S}_+^m) = \mathbb{R}^s$ and $M_D := \{u \in \mathbb{R}^s : W^T u \prec q\} \neq \emptyset$, strong duality holds true, i.e.

$$(\Phi(t) =)\ \min\{q^T y : Wy = t,\ y \in \mathcal{S}_+^m\} = \max\nolimits_{u \in M_D^{\preceq}} t^T u,$$

for all $t \in \mathbb{R}^s$. Moreover, in view of Lemma 2.3, the set $M_D^{\preceq}$ is compact which implies that the right-hand side stays bounded for each t and so Φ is finite.

Further, for arbitrary $\lambda \in [0,1]$ and $t_1, t_2 \in \mathbb{R}^s$ the following inequality is valid

$$\begin{aligned}
\Phi(\lambda t_1 + (1-\lambda)t_2) &= \min\{q^T y : Wy = \lambda t_1 + (1-\lambda)t_2,\ y \in \mathcal{S}_+^m\} \\
&= \max\{(\lambda t_1 + (1-\lambda)t_2)^T u : W^T u \preceq q\} \\
&\leq \lambda \max\{t_1^T u : W^T u \preceq q\} + (1-\lambda)\max\{t_2^T u : W^T u \preceq q\} \\
&= \lambda\Phi(t_1) + (1-\lambda)\Phi(t_2),
\end{aligned}$$

which proves the asserted convexity of Φ.

To establish Lipschitz continuity, let $t_1, t_2 \in \mathbb{R}^s$ be arbitrary and fixed. Then by Weierstrass' theorem, there exists $\overline{v}_1, \overline{v}_2 \in M_D^{\preceq}$ such that

$$\begin{aligned}
\Phi(t_1) &= \min\{q^T y : Wy = t_1, y \in \mathcal{S}_+^m\} = \max\{t_1^T u : W^T u \preceq q\} = t_1^T \overline{v}_1, \\
\Phi(t_2) &= \min\{q^T y : Wy = t_2, y \in \mathcal{S}_+^m\} = \max\{t_2^T u : W^T u \preceq q\} = t_2^T \overline{v}_2.
\end{aligned}$$

Moreover, the following estimates are valid

$$\begin{aligned}
\Phi(t_1) - \Phi(t_2) &= t_1^T \overline{v}_1 - t_2^T \overline{v}_2 \geq t_1^T \overline{v}_2 - t_2^T \overline{v}_2 \geq \|\overline{v}_2\| \cdot \|t_1 - t_2\|, \\
\Phi(t_1) - \Phi(t_2) &= t_1^T \overline{v}_1 - t_2^T \overline{v}_2 \leq t_1^T \overline{v}_1 - t_2^T \overline{v}_1 \leq \|\overline{v}_1\| \cdot \|t_1 - t_2\|,
\end{aligned}$$

and thus

$$|\Phi(t_1) - \Phi(t_2)| \leq \max\{\|\overline{v}_1\|, \|\overline{v}_2\|\} \cdot \|t_1 - t_2\|.$$

Putting now $L_\Phi := \max_{u \in M_D^{\preceq}} \|u\|$ we obtain for arbitrary $t_1, t_2 \in \mathbb{R}^s$:

$$|\Phi(t_1) - \Phi(t_2)| \leq L_\Phi \cdot \|t_1 - t_2\|$$

and the proof is complete. □

Remark 2.6. From basic results in convex analysis (cf. [43]) it is well known that a finite valued convex function is differentiable almost everywhere and so the optimal value function Φ (provided $W(\mathcal{S}_+^m) = \mathbb{R}^s$ and $M_D \neq \emptyset$). Moreover, since Φ is convex, it is also known, that its directional derivative at a point t with respect to a vector $\bar{t}$,

$$\Phi'(t; \bar{t}) := \lim_{h \to 0} \frac{\Phi(t + h\bar{t}) - \Phi(t)}{h}, \tag{2.7}$$

exists for any $t, \bar{t} \in \mathbb{R}^s$.

Differentiability results for (2.3) can be concluded from Lemma 2.5.

Corollary 2.7. *Assume $W(\mathcal{S}_+^m) = \mathbb{R}^s$ and $M_D := \{u \in \mathbb{R}^s : W^T u \prec q\} \neq \emptyset$. Then the optimal value function (2.3) is differentiable almost everywhere. The subdifferential of Φ at t is a closed, convex, and bounded set, given by*

$$\partial\Phi(t) = \arg\max\{u^T t : u \in M_D^{\preceq}\}. \tag{2.8}$$

Moreover, Φ is differentiable at t if and only if $\partial\Phi(t)$ is a singleton.

Proof. According to [44], the subdifferential of Φ at t is a closed, convex, and bounded set, given by $\arg\max\{u^T t : u \in M_D^{\preceq}\}$, and Φ is differentiable at t if and only if $\partial\Phi(t)$ is a singleton. □

Remark 2.8. Corollary 2.7 implies that the dual to the second-stage program (2.3) generally has a unique maximum.

Remark 2.9. With the aid of the chain rule (whose proof depends on the separating hyperplane theorem [43]),

$$\partial\Phi(z - Tx) = -T^T \arg\max\{u^T(z - Tx) : u \in M_D^{\preceq}\}$$

for $x \in \mathcal{S}_+^n$.

2.1.1.2 The Expected Value Functional

Ranking the random variables $c^T x + \Phi(z(\xi) - Tx)$ in (2.2) by its expected value leads to the following (risk neutral) stochastic program:

$$\min \left\{ \mathcal{Q}_{\mathbb{E}}(x) : Ax = b,\ x \in \mathcal{S}^n_+ \right\}, \tag{2.9}$$

where $\mathcal{Q}_{\mathbb{E}}(x) : \mathcal{S}^n \to \mathbb{R}$ is the expected value function defined by

$$\mathcal{Q}_{\mathbb{E}}(x) := \mathbb{E}_\xi[c^T x + \Phi(z(\xi) - Tx)]. \tag{2.10}$$

Properties of (2.10) will now be discussed[4]. For convenience, let us define the image measure $\mu := \mathbb{P} \circ z^{-1}$ acting on $\mathbb{R}^s$. Then, the expected value function (2.10) may be written as

$$\mathcal{Q}_{\mathbb{E}}(x) = c^T x + \int_{\mathbb{R}^s} \Phi(z - Tx)\ \mu(dz). \tag{2.11}$$

In order that program (2.9) is well-defined, the integral on the right-hand side in (2.11) has to be real-valued. To this end, a further assumption (in addition to complete recourse and existence of a Slater point) is needed. The next theorem shows that this is accomplished by assuming that the image measure μ has finite first moment, i.e.,

$$\int_{\mathbb{R}^s} \|z\| \mu(dz) < \infty.$$

Theorem 2.10. *Assume $W(\mathcal{S}^m_+) = \mathbb{R}^s$, $M_D := \{u \in \mathbb{R}^s : W^T u \prec q\} \neq \emptyset$, and $\int_{\mathbb{R}^s} \|z\| \mu(dz) < \infty$. Then the expected value function (2.11) is real-valued, convex, and Lipschitz continuous on $\mathcal{S}^n_+$.*

Proof. According to Lemma 2.5, the optimal value function Φ is finite as well as continuous and, thus, μ-measurable. Further, for any fixed $x \in \mathcal{S}^n_+$:

$$\begin{aligned} |\mathcal{Q}_{\mathbb{E}}(x)| &= |c^T x + \int_{\mathbb{R}^s} \Phi(z - Tx)\mu(dz)| \\ &\leq |c^T x| + \int_{\mathbb{R}^s} |\Phi(z - Tx)|\ \mu(dz). \end{aligned}$$

[4]Basic properties of the expected recourse function of a stochastic linear program were examined by Wets in [45] and [46], and by Walkup and Wets in [47] and [48].

By strong duality and Lemma 2.3 with Weierstrass' theorem, there exists $\widetilde{v} \in M_D^{\preceq}$ such that

$$\begin{aligned}\int_{\mathbb{R}^s} |\Phi(z - Tx)|\mu(dz) &\leq \int_{\mathbb{R}^s} \|\widetilde{v}\| \cdot \|(z - Tx)\|\mu(dz) \\ &\leq \|\widetilde{v}\| \cdot \int_{\mathbb{R}^s} \|z\|\mu(dz) + \|\widetilde{v}\| \cdot \|T\| \cdot \|x\|.\end{aligned}$$

Hence,

$$|\mathcal{Q}_{\mathbb{E}}(x)| \leq \Big(\|c\| + \|\widetilde{v}\| \cdot \|T\|\Big) \cdot \|x\| + \|\widetilde{v}\| \cdot \int_{\mathbb{R}^s} \|z\| \ \mu(dz).$$

By assumption, the latter integral is finite which implies that $Q_{\mathbb{E}}$ is real-valued on $\mathcal{S}^n_+$. Furthermore, with the aid of Lemma 2.5, for any fixed $z \in \mathbb{R}^s$, the optimal value function $\Phi(z - Tx)$ is convex in x. Thus, for $\lambda \in [0, 1]$ and $x_1, x_2 \in \mathcal{S}^n_+$:

$$\begin{aligned} & \int_{\mathbb{R}^s} \Phi(z - T(\lambda x_1 + (1-\lambda)x_2))\mu(dz) \\ \leq \quad & \int_{\mathbb{R}^s} \Big[\lambda\Phi(z - Tx_1) + (1-\lambda)\Phi(z - Tx_2)\Big]\mu(dz) \\ = \quad & \lambda \int_{\mathbb{R}^s} \Phi(z - Tx_1)\mu(dz) + (1-\lambda)\int_{\mathbb{R}^s} \Phi(z - Tx_2)\mu(dz),\end{aligned}$$

which proves the asserted convexity of $Q_{\mathbb{E}}$.

As a result of the Lipschitz continuity of Φ, we obtain for $x_1, x_2 \in \mathcal{S}^n_+$:

$$\begin{aligned}|\mathcal{Q}_{\mathbb{E}}(x_1) - \mathcal{Q}_{\mathbb{E}}(x_2)| &\leq \|c\| \cdot \|x_1 - x_2\| + \left|\int_{\mathbb{R}^s} \Phi(z - Tx_1) - \Phi(z - Tx_2)\mu(dz)\right| \\ &\leq \|c\| \cdot \|x_1 - x_2\| + \int_{\mathbb{R}^s} L_\Phi \cdot \|z - Tx_1 - (z - Tx_2)\| \ \mu(dz) \\ &\leq \|c\| \cdot \|x_1 - x_2\| + \|T\| \cdot \|x_1 - x_2\| \cdot L_\Phi \cdot \underbrace{\int_{\mathbb{R}^s} \mu(dz)}_{=1}.\end{aligned}$$

For the latter inequality, note that the Lipschitz constant L_Φ is independent of z (see proof of Lemma 2.5). Hence, $\mathcal{Q}_{\mathbb{E}}$ is Lipschitz continuous with Lipschitz constant $\overline{L} := \|c\| + L_\Phi \cdot \|T\|$, and the proof is complete. □

Remark 2.11. Under the conditions of Theorem 2.10, (2.9) is minimization of a real convex function over a convex set. Thus, all powerful tools of convex optimization (see [49]) may be applied.

Due to [50] and Theorem 2.10, the directional derivative of $\mathcal{Q}_{\mathbb{E}}$ at x with respect to the vector $\overline{x}$ exists for any $x, \overline{x} \in \mathcal{S}^n$ and is given by

$$\mathcal{Q}'_{\mathbb{E}}(x; \overline{x}) = c^T \overline{x} + \int_{\mathbb{R}^s} \Phi'(z - Tx; \overline{x})\ \mu(dz). \tag{2.12}$$

Note that $\Phi'(z(\xi) - Tx; \overline{x})$ as a pointwise limit of measurable functions, is measurable for every $x, \overline{x} \in \mathcal{S}^n$. Further, in similar way to traditional stochastic linear programming, differentiability of $\mathcal{Q}_{\mathbb{E}}$ can be adopted (cf. [41]).

Lemma 2.12. *Assume $W(\mathcal{S}^m_+) = \mathbb{R}^s$, $M_D \neq \emptyset$, and $\int_{\mathbb{R}^s} \|z\|\mu(dz) < \infty$. If for $x_0 \in \mathcal{S}^n_+$ the set $\arg\max\{u^T(z - Tx_0) : u \in M_D^{\preceq}\}$ is a singleton for μ-almost all $z \in \mathbb{R}^s$, then $\mathcal{Q}_{\mathbb{E}}$ is differentiable at x_0.*

Proof. In view of Corollary 2.7, for μ-almost all $z \in \mathbb{R}^s$, $\Phi(z - Tx)$ is differentiable with derivative $\Phi'(z - Tx) = -T^T u(x, z)$, where $u(x, z) = \arg\max\{(z - Tx)^T u : u \in M_D^{\preceq}\}$. The latter derivative, in particular, is measurable. If we consider for $x_0 \in \mathcal{S}^n_+$ the function

$$g(x) = \frac{\Phi(z - Tx) - \Phi(z - Tx_0) - \Phi'(z - Tx_0)(x - x_0)}{\|x - x_0\|},$$

then by assumption $\lim_{x \to x_0} g(x) = 0$. Moreover, $\|g(x)\| \leq 2L_\Phi \|T\|$ (see Lemma 2.5) such that $2L_\Phi\|T\|$ which is independent of z, is an integrable majorant for g. Hence, by Lebesgue's dominated convergence theorem

$$\lim_{x \to x_0} \frac{\int_{\mathbb{R}^s} \Phi(z - Tx) - \Phi(z - Tx_0)\mu(dz) - \int_{\mathbb{R}^s} \Phi'(z - Tx_0)(x - x_0)\mu(dz)}{\|x - x_0\|}$$
$$= \int_{\mathbb{R}^s} \underbrace{\lim_{x \to x_0} g(x)}_{=0}\, \mu(dz),$$

i.e., $\mathcal{Q}_{\mathbb{E}}$ is differentiable at x_0 with derivative $c + \int_{\mathbb{R}^s} \Phi'(z - Tx_0)\mu(dz)$. □

Remark 2.13. In case the random variable z is absolutely continuous and the operator T is injective, $\mathcal{Q}_{\mathbb{E}}$ is differentiable everywhere.

Similar to formula (2.12), subdifferentiation and integration may be exchanged (cf. Proposition 2.2 in [50]):

$$\partial_x \left(\int_{\mathbb{R}^s} \Phi(z - Tx)\mu(dz) \right) = \int_{\mathbb{R}^s} \partial_x \ \Phi(z - Tx)\mu(dz), \tag{2.13}$$

where the expression on the right-hand side is defined by the following set:

$$\{\overline{x} \in \mathcal{S}^n : \overline{x} = \int_{\mathbb{R}^s} \overline{x}(z)\mu(dz), \quad \overline{x}(z) \text{ measurable}, \quad \overline{x}(z) \in \partial_x \Phi(z - Tx) \text{ a.e.}\}.$$

Corollary 2.14. *Assume $W(\mathcal{S}^m_+) = \mathbb{R}^s$, $M_D \neq \emptyset$ and that the underlying random variable $z(\xi)$ follows a finite discrete probability distribution with realizations z_ξ and probabilities π_ξ, $\xi = 1, ..., S$. Then, for any $x_0 \in \mathcal{S}^n$,*

$$\partial \mathcal{Q}_{\mathbb{E}}(x_0) = \sum_{\xi=1}^{S} \pi_\xi \partial\Phi(z_\xi - Tx_0), \tag{2.14}$$

where $\partial\Phi(z_\xi - Tx_0) = -T^T \arg\max\{(z_\xi - Tx_0)^T u : W^T u \preceq q\}$.

Proof. It follows immediately from (2.13) and Remark 2.9. □

Remark 2.15. Because of the Moreau-Rockafellar Theorem (see, for instance, [43]), of which (2.14) is a particular case, the assumptions in Corollary 2.14 can be softened. Instead of complete recourse, i.e., $W(\mathcal{S}^m_+) = \mathbb{R}^s$, it is sufficient to assume that $\Phi(z_\xi - Tx_0)$ is finite for all $\xi = 1, ..., S$. Furthermore, if each function $\Phi(z_\xi - T\cdot)$ is polyhedral[5] (as it is the case in stochastic linear programming), the regularity condition $M_D \neq \emptyset$ may be omitted as well.

Although, by means of the Hausdorff metric, convex bodies (subsets of the euclidean space that are non-empty, compact, and convex) and so bounded spectrahedra can be approximated arbitrarily close by polytopes (cf. [51]), not all results from stochastic linear programming can be transferred to the semidefinite case. As the next example will show, strong convexity (on unbounded sets) of

$$\widetilde{\mathcal{Q}}_{\mathbb{E}}(\xi) := \int_{\mathbb{R}^s} \Phi(z - \xi)\mu(dz),$$

[5] An extended real valued function $g : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, \infty\}$ is called polyhedral if its epigraph is a convex closed polyhedron, and $g(x)$ is finite for at least one x.

which can be attained in stochastic linear programming, (see [52]) cannot be derived, no matter what is claimed on the probability measure μ.

Example 2.2. *Let $q = I_3$ and $W : \mathcal{S}^3_+ \to \mathbb{R}^2$ be given by the matrices*

$$W_1 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}.$$

Since $q - Wu$ is positive semidefinite if all of its principal minors are non-negative, $u \in M_D \subseteq \mathbb{R}^2$ is equivalent to

$$\begin{vmatrix} 1 & 0 & u_1 \\ 0 & 1 & u_2 \\ u_1 & u_2 & 1 \end{vmatrix} = 1 - u_1^2 - u_2^2 \geq 0.$$

Hence,

$$\Phi(z - \xi) = \max_{u \in B_1(0)} u^T(z - \xi) = \|z - \xi\|_2.$$

Recall that $\widetilde{\mathcal{Q}}_{\mathbb{E}}$ is said to be strongly convex on $V \subseteq \mathbb{R}^2$ if there exists $\kappa > 0$ such that for all $\xi_1, \xi_2 \in V$ and all $\lambda \in [0,1]$,

$$\widetilde{\mathcal{Q}}_{\mathbb{E}}(\lambda\xi_1 + (1-\lambda)\xi_2) \leq \lambda\widetilde{\mathcal{Q}}_{\mathbb{E}}(\xi_1) + (1-\lambda)\widetilde{\mathcal{Q}}_{\mathbb{E}}(\xi) - \kappa\lambda(1-\lambda)\|\xi_1 - \xi_2\|_2^2$$

or equivalently if $\widetilde{\mathcal{Q}}_{\mathbb{E}}(\xi) - \frac{\kappa}{2}\|\xi\|_2^2$ is a convex function. Assume now that there exists a fixed $\kappa > 0$ such that the latter function is convex. By setting $\lambda = \frac{1}{2}$ this would yield: $\widetilde{\mathcal{Q}}_{\mathbb{E}}(\frac{1}{2}(\xi_1 + \xi_2)) - \frac{1}{2}\widetilde{\mathcal{Q}}_{\mathbb{E}}(\xi_1) - \frac{1}{2}\widetilde{\mathcal{Q}}_{\mathbb{E}}(\xi_2) \leq \frac{\kappa}{8}\|\xi_1 + \xi_2\|_2^2 - \frac{\kappa}{4}\left(\|\xi_1\|_2^2 + \|\xi_2\|_2^2\right)$, i.e.,

$$\begin{aligned} & \int_{\mathbb{R}^2} \|z - \frac{1}{2}(\xi_1 + \xi_2)\|_2 \mu(dz) - \frac{1}{2}\left(\int_{\mathbb{R}^2} \|z - \xi_1\|_2 \mu(dz) + \int_{\mathbb{R}^2} \|z - \xi_2\|_2 \mu(dz)\right) \\ \leq \; & \frac{\kappa}{8}\|\xi_1 + \xi_2\|_2^2 - \frac{\kappa}{4}\left(\|\xi_1\|_2^2 + \|\xi_2\|_2^2\right). \end{aligned} \tag{2.15}$$

As

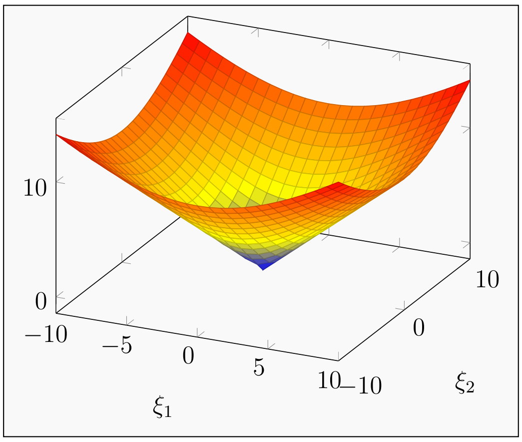
$$\int_{\mathbb{R}^2} \|z - \xi_1\|_2 \mu(dz) + \int_{\mathbb{R}^2} \|z - \xi_2\|_2 \mu(dz) \leq \int_{\mathbb{R}^2} \|z\|_2 \mu(dz) + \|\xi_1\|_2 + \|\xi_2\|_2,$$

the left-hand side in (2.15) is always greater than $-(\|\xi_1\|_2 + \|\xi_2\|_2)$. *Dividing the right-hand side in (2.15) by this term and multiplying both sides with* $\frac{8}{\kappa}$ *results in*

$$\frac{8}{\kappa} \geq \frac{\|\xi_1 + \xi_2\|_2^2 - 2\left(\|\xi_1\|_2^2 + \|\xi_2\|_2^2\right)}{\|\xi_1\|_2 + \|\xi_2\|_2} \geq \frac{\|\xi_1\|_2^2 - 3\|\xi_2\|_2^2}{\|\xi_1\|_2 + \|\xi_2\|_2}. \tag{2.16}$$

Hence, there has to exist $\kappa > 0$ *such that (2.16) holds true for all* $\xi_1, \xi_2 \in V$. *Since the set* V *is assumed to be unbounded, it contains an unbounded sequence* $(\xi_n)_{n\in\mathbb{N}}$ *with* $\|\xi_n\|_2 \to \infty$. *For the latter sequence and fixed* $\bar{\xi} \in V$ *we obtain* $\lim\limits_{n\to\infty} \left(\|\xi_n\|_2^2 - 3\|\bar{\xi}\|_2^2\right) / \left(\|\xi_n\|_2 + \|\bar{\xi}\|_2\right) = \lim\limits_{n\to\infty} \|\xi_n\|_2 = \infty$. *This implies that the above functional* $\widetilde{\mathcal{Q}}_{\mathbb{E}}$ *cannot be strongly convex on unbounded sets.*

FIGURE 2.1: Plot of the function $\Phi(z - \xi) = \|z - \xi\|_2$ at $z = (0, 0)$.



2.1.1.3 The Excess Probability Functional

In this subsection the excess probability functional $\mathcal{Q}_{\mathbb{P}_\eta} : \mathcal{S}^n \to \mathbb{R}$ defined by

$$\mathcal{Q}_{\mathbb{P}_\eta}(x) = \mu\Big[\big\{z \in \mathbb{R}^s \ : \ c^T x + \Phi(z - Tx) > \eta\big\}\Big] \tag{2.17}$$

will be discussed. For its analysis, a few prerequisites are needed (cf. [5]).

Definition 2.16. A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ is said to be lower semi-continuous at $x_0 \in X$ if $\liminf_{x\to x_0} f(x) \geq f(x_0)$. It is upper semi-continuous at $x_0 \in X$ if $\limsup_{x\to x_0} f(x) \leq f(x_0)$. The function f is called lower (upper) semi-continuous if it is lower (upper) semi-continuous at every point of its domain.

Note that a function is continuous if and only if it is upper and lower semi-continuous.

Definition 2.17. Given a sequence $(M_n)_{n\in\mathbb{N}}$ of measurable sets in $\mathbb{R}^s$, the limes inferior $\liminf_{n\to\infty} M_n$ and the limes superior $\limsup_{n\to\infty} M_n$ are defined as the sets of all points belonging to all but a finite number of the M_n and to infinitely many M_n, respectively.

The semi-continuity of probability measures, which can be followed from Fatou's lemma (Lemma B.8), yields the following inequalities

$$\mu\left(\liminf_{n\to\infty} M_n\right) \leq \liminf_{n\to\infty} \mu(M_n) \leq \limsup_{n\to\infty} \mu(M_n) \leq \mu\left(\limsup_{n\to\infty} M_n\right), \tag{2.18}$$

for measurable sequences $(M_n)_{n\in\mathbb{N}}$.

Theorem 2.18. *Assume $W(\mathcal{S}^m_+) = \mathbb{R}^s$ and $M_D \neq \emptyset$. Then $\mathcal{Q}_{\mathbb{P}_\eta}(x) : \mathcal{S}^n_+ \to \mathbb{R}$ with*

$$\mathcal{Q}_{\mathbb{P}_\eta}(x) = \mu\Big[\big\{z \in \mathbb{R}^s \ : \ c^T x + \Phi(z - Tx) > \eta\big\}\Big]$$

is real-valued and lower semi-continuous on $\mathcal{S}^n_+$. If for some $x \in \mathcal{S}^n_+$, it holds $\mu(\{z \in \mathbb{R}^s \ : \ c^T x + \Phi(z - Tx) = \eta\}) = 0$, then $\mathcal{Q}_{\mathbb{P}_\eta}$ is continuous at x.

Proof. In order to facilitate notation, we introduce for $x \in \mathcal{S}^n_+$ the sets

$$\begin{aligned} M(x) &:= \{z \in \mathbb{R}^s \ : \ c^T x + \Phi(z - Tx) > \eta\}, \\ M_e(x) &:= \{z \in \mathbb{R}^s \ : \ c^T x + \Phi(z - Tx) = \eta\}. \end{aligned}$$

Assuming $W(\mathcal{S}^m_+) = \mathbb{R}^s$ and $M_D \neq \emptyset$, the optimal value function Φ is continuous (cf. Theorem 2.5) and therefore $M(x)$ is μ-measurable for all $x \in \mathcal{S}^n_+$. Hence, the excess probability measure $\mathcal{Q}_{\mathbb{P}_\eta}(x) = \mu(M(x))$ is real-valued.
Next, we will exhibit that for all $x \in \mathcal{S}^n_+$ the following implications are valid:

$$M(x) \subseteq \liminf_{x_n \to x} M(x_n) \subseteq \limsup_{x_n \to x} M(x_n) \subseteq M(x) \cup M_e(x).$$

To this, let $\overline{z} \in M(x)$. Then, due to the continuity of Φ we have

$$\liminf_{x_n \to x} \left(c^T x + \Phi(\overline{z} - Tx_n)\right) = c^T x + \Phi(\overline{z} - Tx) > \eta.$$

In other words, there has to be $n_0 \in \mathbb{N}$ such that

$$c^T x_n + \Phi(\overline{z} - T x_n) > \eta \quad \text{for } n \geq n_0.$$

The latter implies $\overline{z} \in M(x_n)$ for all $n \geq n_0$, i.e.,

$$M(x) \subseteq \liminf_{x_n \to x} M(x_n).$$

Further, if we assume $\overline{z} \in \limsup_{x_n \to x} M(x_n) \backslash M(x)$. Then there has to exist an infinite subset $\widetilde{\mathbb{N}}$ of $\mathbb{N}$ such that $c^T x + \Phi(\overline{z} - Tx) \leq \eta$ and

$$c^T x_n + \Phi(\overline{z} - T x_n) > \eta, \quad \forall n \in \widetilde{\mathbb{N}}.$$

Passing to the limit in the latter inequality yields $c^T x + \Phi(\overline{z} - Tx) = \eta$, and therefore $\overline{z} \in M_e(x)$. Thus, in view of (2.18) we have for all $x \in \mathcal{S}^n_+$:

$$\begin{aligned} \mathcal{Q}_{\mathbb{P}_\eta}(x) = \mu(M(x)) &\leq \mu(\liminf_{x_n \to x} M(x_n)) \\ &\leq \liminf_{x_n \to x} \mu(M(x_n)) \; = \; \liminf_{x_n \to x} \mathcal{Q}_{\mathbb{P}_\eta}(x_n), \end{aligned}$$

i.e., $\mathcal{Q}_{\mathbb{P}_\eta}$ is lower semi-continuous. In case $\mu(M_e(x)) = 0$ this can be extended to

$$\begin{aligned} \mathcal{Q}_{\mathbb{P}_\eta}(x) &= \mu(M(x)) = \mu(M(x) \cup M_e(x)) \\ &\geq \mu\left(\limsup_{x_n \to x} M(x_n)\right) \geq \limsup_{x_n \to x} \mu(M(x_n)) = \limsup_{x_n \to x} \mathcal{Q}_{\mathbb{P}_\eta}(x_n), \end{aligned}$$

which is the upper semi-continuity of $\mathcal{Q}_{\mathbb{P}_\eta}$ at x. Hence, $\mathcal{Q}_{\mathbb{P}_\eta}$ is continuous at x if $\mu(M_e(x)) = 0$. □

The above continuity properties are due to the continuity of the optimal value function Φ (derived in Theorem 2.5). It is noted that the proof of Theorem 2.18 follows completely analogous to [5].

Theorem 2.18 turns to the question of a global continuity result for $\mathcal{Q}_{\mathbb{P}_\eta}$. Provided μ is absolutely continuous, then, in view of the proof of Theorem 2.18, sufficient conditions for the set $M_e(x)$ to be included in a set with Lebesgue measure zero come to the fore. This, for instance, is the case if the normal cone to

M_D at zero, i.e.,

$$N_{M_D}(0) := \{z \in \mathbb{R}^s : \langle z, u \rangle \leq 0, \ \forall u \in M_D\}$$

has Lebesgue measure zero. The latter is guaranteed if all available vertices[6] belonging to M_D are non-zero.

Remark 2.19. If the feasible set X is compact, then, the continuity of $\mathcal{Q}_{\mathbb{E}}$ and the lower continuity of $\mathcal{Q}_{\mathbb{P}_\eta}$ imply that $\min\{\mathcal{Q}_{\mathbb{E}}(x) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(x) : x \in X\}$ is well-posed (in the sense that its minimum is finite and attained).

Remark 2.20. Following [53] and [54], stochastic dominance relations can be associated with mean-risk models (1.7). It can be shown that the mean-risk functional $\mathcal{Q}_{\mathbb{E}}(x) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(x)$ is α-consistent with first degree stochastic dominance (cf. [5]). The latter means that $\alpha \geq 0$ and $f(x, \xi) \succeq_1 f(y, \xi)$ implies

$$\mathcal{Q}_{\mathbb{E}}(x) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(x) \leq \mathcal{Q}_{\mathbb{E}}(y) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(y).$$

2.2 Two-Stage Stochastic Semidefinite Programs with Mixed-Integer Recourse

In this section, we consider two-stage programs (2.2) in which some or all variables are integers, with the restriction that relations among integer variables as well as relationships between integer variables and continuous variables are linear. Further, it is assumed that the feasible set of second-stage integer variables is bounded. This leads to the following two-stage random program

$$\text{“min”} \left\{ c^T x + c'^T x' + \Phi(z(\xi) - Tx) : Ax + A'x' = b, \ x \in \mathcal{S}^n_+, \ x' \in \mathbb{Z}^{n'} \right\}, \quad (2.19)$$

where

$$\Phi(t) := \min\{q^T y + q'^T y' : Wy + W'y' = t, \ y \in \mathcal{S}^m_+, \ y' \in \mathbb{Z}^{m'}_+, \ \|y'\|_\infty \leq K\}. \quad (2.20)$$

[6] A vertex of a convex set is an extreme point of that set whose normal cone is full-dimensional.

Here, it is assumed that all ingredients have conformable dimensions. We call random programs defined this way two-stage semidefinite programs with mixed-integer recourse.

Remark 2.21. It is well known that any bounded integer variable

$$y'_{LB} \leq y' \leq y'_{UB}$$

can be expressed via binary variables, $u_1, ..., u_N$, as

$$y' = y'_{LB} + u_1 + 2u_2 + ... + 2^{N-1}u_{N-1},$$

where $N = \lfloor \ln_2(y'_{UB} - y'_{LB}) \rfloor + 1$ (cf. [55]). Therefore, (2.20) can be equivalently expressed by the following mixed-binary semidefinite program

$$\min\{q^T y + q''^T y'' : Wy + W''y'' = t,\ y \in \mathcal{S}^m_+,\ y'' \in \{0,1\}^{m''}\}.$$

Remark 2.22. Introducing integer variables into the second stage, as in (2.20), has a great impact on the structural properties of the optimal value function Φ. For instance, in contrast to purely continuous variables, the optimal value function of a mixed-integer linear program is in general neither convex nor continuous (cf. Chapter 4 of [11]).

Theorem 2.23. *Assume complete mixed-integer recourse*

$$W(\mathcal{S}^m_+) + W'(\mathbb{Z}^{m'}_+ \cap [0, K]^{m'}) = \mathbb{R}^s, \tag{2.21}$$

and that the dual to the SDP relaxation of (2.20) contains a point $u \in \mathbb{R}^s$ with $W^T u \prec q$, $W'u \leq q'$, then the optimal value function Φ defined by mixed-integer (semidefinite) recourse (2.20) is well-defined on $\mathbb{R}^s$.

Proof. We have to show that $\Phi(t) \in \mathbb{R}$ holds for all $t \in \mathbb{R}^s$. To this end, let $t \in \mathbb{R}^s$. The set

$$\{y \in \mathcal{S}^m_+, y' \in \mathbb{Z}^{m'}_+, \|y'\|_\infty \leq K : Wy + W'y' = t\} \tag{2.22}$$

is a finite union of spectrahedral sets of the form

$$\{(y, y') : Wy = t - W'y', y \in \mathcal{S}^m_+\},$$

where y' is a non-negative integer vector bounded by $\|y'\|_\infty \le K$. By (2.21), the mixed-integer semidefinite program (2.20) is feasible and thus any of these spectrahedra has to be non-empty. Let us denote by $\mathbb{Z}_+^{m'}(t)$ the set of all points $y' \in \mathbb{Z}_+^{m'}, \|y'\|_\infty \le K$ for which $\{y \in \mathcal{S}_+^m : Wy = t - W'y'\}$ is non-empty, i.e.,

$$\mathbb{Z}_+^{m'}(t) := \{y' \in \mathbb{Z}_+^{m'}, \|y'\|_\infty \le K : Wy' \in t - W(\mathcal{S}_+^m)\}.$$

Then,

$$\Phi(t) = \min_{y' \in \mathbb{Z}_+^{m'}(t)} \left[q'^T y' + \min\{q^T y : Wy = t - W'y',\ y \in \mathcal{S}_+^m\} \right]. \tag{2.23}$$

Further, due to the assumption $\{u \in \mathbb{R}^s : W^T u \prec q, W'^T u \le q'\} \neq \emptyset$ and by duality (see Theorem A.29), attainment of the infimum in

$$\inf\{q^T y : Wy = t - W'y',\ y \in \mathcal{S}_+^m\}$$

is guaranteed for each $y' \in \mathbb{Z}_+^{m'}(t)$. The latter implies that the right-hand side of formula (2.23) is real valued (due to finiteness of $\mathbb{Z}_+^{m'}(t)$). Hence, $\Phi(t)$ is well-defined on $\mathbb{R}^s$. □

Remark 2.24. In stochastic integer (linear) programming, the existence theorem for mixed-integer linear programs [38], [56], [57] implies the following: If

- the matrices W and W' are rational,
- complete mixed-integer recourse $W(\mathbb{R}_+^m) + W'(\mathbb{Z}_+^{m'}) = \mathbb{R}^s$ is valid,
- and sufficiently expensive recourse

  $$\{u \in \mathbb{R}^s : W^T u \le q,\ W'^T u \le q'\} \neq \emptyset$$

  holds true,

then, the optimal value function of the mixed-integer program

$$\min\{q^T y + q'^T y' : Wy + W'y' = t,\ y \in \mathbb{R}_+^m,\ y' \in \mathbb{Z}_+^{m'}\}$$

is real-valued on $\mathbb{R}^s$ (see [39] and [40] for details).

It is easy to see that for mixed-integer recourse (2.20), the complete mixed-integer recourse property is satisfied iff $W(\mathcal{S}_+^m) = \mathbb{R}^s$, which due to Lemma 2.3 is equivalent to compactness of $M_D^{\preceq} = \{u \in \mathbb{R}^s : W^T u \preceq q\}$.

Let us denote by $Y' \subset \mathbb{Z}_+^{m'}$ the finite set of integers $\{y' \in \mathbb{Z}_+^{m'} : \|y'\|_\infty \leq K\}$. If complete mixed-integer recourse and $M_D \neq \emptyset$ holds true, then, the right-hand side in formula (2.23) becomes

$$\min_{y' \in Y'} \left[q'^T y' + \max_{u \in M_D^{\preceq}} (t - W'y')^T u \right].$$

This means that Φ is made up by a pointwise minimum of convex functions (see Theorem 2.5).

Theorem 2.25. *Assume* $W(\mathcal{S}_+^m) + W'(\mathbb{Z}_+^{m'} \cap [0, K]^{m'}) = \mathbb{R}^s$ *and* $M_D = \{u \in \mathbb{R}^s : W^T u \prec q,\ W'^T u \leq q'\} \neq \emptyset$. *Then there exist positive constants* β *and* γ *such that*

$$|\Phi(t_1) - \Phi(t_2)| \leq \beta \|t_1 - t_2\| + \gamma. \tag{2.24}$$

Proof. Let $t_1, t_2 \in \mathbb{R}^s$. Then

$$|\Phi(t_1) - \Phi(t_2)| = |q'^T(y_1' - y_2') + (t_1 - W'y_1')^T u_1 - (t_2 - W'y_2')^T u_2|, \tag{2.25}$$

where (y_1', u_1) and (y_2', u_2) are optimal points in (2.23) for t_1 and t_2, respectively. In assuming complete mixed-integer recourse and that $M_D \neq \emptyset$, the set $M_D^{\preceq}$ has to be compact. Therefore, the right-hand side in (2.25) can be estimated as follows

$$\begin{aligned} &|q'^T(y_1' - y_2') + (t_1 - W'y_1')^T u_1 - (t_2 - W'y_2')^T u_2| \\ &\leq (\|q'\| + \beta \cdot \|W'\|) \cdot \|y_1' - y_2'\| + \beta \cdot \|t_1 - t_2\|, \end{aligned}$$

where $\beta := \max_{u \in M_D^{\preceq}} \|u\|$. As the norm of the vectors $y_1', y_2' \in \mathbb{Z}_+^{m'} \cap [0, K]^{m'}$ is bounded by $\sqrt{m'}K$, we obtain

$$|\Phi(t_1) - \Phi(t_2)| \leq \beta \|t_1 - t_2\| + \gamma,$$

with $\gamma := (\|q'\| + \beta \cdot \|W'\|) \cdot 2\sqrt{m'}K$, which completes the proof. $\square$

Unbounded Second-stage Integers: Assume that the second-stage mixed-integer semidefinite program (2.20) is modified as follows:

$$\min\{q^T y + q'^T y' : Wy + W'y' = t,\ y \in \mathcal{S}_+^m,\ y' \in \mathbb{Z}_+^{m'}\}. \tag{2.26}$$

According to the proof of Theorem 2.23, if complete mixed-integer recourse $W(\mathcal{S}_+^m) + W'(\mathbb{Z}_+^{m'}) = \mathbb{R}^s$, and $\{u \in \mathbb{R}^s : W^T u \prec q, W'^T u \leq q'\} \neq \emptyset$ holds true, program (2.26) may be written as

$$\min_{y' \in \mathbb{Z}_+^{m'}(t)} \left[q'^T y' + \max\{(t - W'y')^T u : W^T u \preceq q\} \right], \tag{2.27}$$

where $\mathbb{Z}_+^{m'}(t) := \{y' \in \mathbb{Z}_+^{m'} : Wy' \in t - W(\mathcal{S}_+^m)\}$. This means that the corresponding optimal value function is made up by a pointwise minimum of convex functions whose domains of definition are spectrahedra arising as shifts of the cone $W(\mathcal{S}_+^m)$. Note that due to our basic assumption $W(\mathcal{S}_+^m) + W'(\mathbb{Z}_+^{m'}) = \mathbb{R}^s$, the cone $W(\mathcal{S}_+^m)$ is full-dimensional.

Lemma 2.26. *Assume $W(\mathcal{S}_+^m) + W'(\mathbb{Z}_+^{m'}) = \mathbb{R}^s$ and that $M_D = \{u \in \mathbb{R}^s : W^T u \prec q,\ W'^T u \leq q'\} \neq \emptyset$. Then $M_D^{\preceq} := \{u \in \mathbb{R}^s : W^T u \preceq q,\ W'^T u \leq q'\}$ is compact.*

Proof. The proof is completely analogous to the proof of Lemma 2.3. □

Without the restriction that all integers in (2.20) are linear[7] and bounded, the optimal value function Φ not necessarily has to be well-defined. This is demonstrated by the next example.

Example 2.3. *Consider the following mixed-integer semidefinite program*

$$\min\{q^T y : Wy = t,\ y_{11} \geq 1,\ y_{11}, y_{12} \in \mathbb{Z}_+,\ y \in \mathcal{S}_+^5\}, \tag{2.28}$$

where $q = e_{25}$, and

$$W = \begin{bmatrix} vec(W_1) & vec(W_2) & vec(W_3) & vec(W_4) \end{bmatrix}^T$$

[7]This means, relations among integer variables are linear.

with

$$W_1 = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$W_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad W_4 = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

For such program, it holds $W\left(\left\{y \in \mathcal{S}^5_+ : y_{11} \geq 0,\ y_{11}, y_{12} \in \mathbb{Z}_+ \backslash \{0\}\right\}\right) = \mathbb{R}^s$. *Indeed, let* $t \in \mathbb{R}^4$*, then, for*

$$y = \begin{bmatrix} y_{11} & y_{12} & * & * & * \\ y_{21} & y_{22} & * & * & * \\ * & * & y_{33} & y_{34} & * \\ * & * & y_{43} & y_{44} & * \\ * & * & * & * & y_{55} \end{bmatrix} \in \mathcal{S}^5,$$

the equation $Wy = t$ *is fulfilled if*

$$\begin{bmatrix} vec(W_1)^T y \\ vec(W_2)^T y \\ vec(W_3)^T y \\ vec(W_4)^T y \end{bmatrix} = \begin{bmatrix} 2y_{11} - y_{22} \\ y_{55} + y_{12} - 2y_{33} \\ y_{11} - y_{34} \\ y_{12} - y_{44} \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix} \quad \Leftrightarrow \quad \left\{ \begin{aligned} & y_{11} = 1/2 \cdot (t_1 + y_{22}), \\ & y_{55} = t_2 + 2y_{33} - y_{12}, \\ & y_{34} = 1/2 \cdot (t_1 + y_{22}) - t_3, \\ & y_{44} = y_{12} - t_4, \end{aligned} \right\}$$

holds. Selecting now

$$\begin{aligned}
\widetilde{y}_{11} &= \lceil |t_1| \rceil + \lceil |t_4| \rceil + 1,\\
\widetilde{y}_{12} &:= \lceil |t_4| \rceil + 1,\\
\widetilde{y}_{22} &:= 2(\lceil |t_1| \rceil - 1/2 \cdot t_1 + \lceil |t_4| \rceil + 1),\\
\widetilde{y}_{33} &:= (\lceil |t_1| \rceil + \lceil |t_4| \rceil + 1 + |t_3|)^2 + |t_2|,\\
\widetilde{y}_{34} &:= \lceil |t_1| \rceil + \lceil |t_4| \rceil + 1 - t_3,\\
\widetilde{y}_{44} &:= \lceil |t_4| \rceil + 1 - t_4,
\end{aligned}$$

then, $\widetilde{y}_{11}$ *and* $\widetilde{y}_{12}$ *are integers that are greater or equal to* 1*. Furthermore, we have*

$$\begin{aligned}
\widetilde{y}_{22} &= 2(\lceil |t_1| \rceil - 1/2 \cdot t_1 + \lceil |t_4| \rceil + 1) \ \geq \ \lceil |t_4| \rceil + 1,\\
\widetilde{y}_{33} &= (\lceil |t_1| \rceil + \lceil |t_4| \rceil + 1 + |t_3|)^2 + |t_2| \ \geq \ \widetilde{y}_{34}^2 + |t_2|,\\
\widetilde{y}_{44} &= \lceil |t_4| \rceil + 1 - t_4 \geq 1,\\
\widetilde{y}_{55} &\geq t_2 + 2|t_2| + 1,
\end{aligned}$$

as well as

$$det\left(\begin{bmatrix} \widetilde{y}_{11} & \widetilde{y}_{12} \\ \widetilde{y}_{12} & \widetilde{y}_{22} \end{bmatrix}\right) \geq (\lceil |t_1| \rceil + \lceil |t_4| \rceil + 1) \cdot (\lceil |t_4| \rceil + 1) - (\lceil |t_4| \rceil + 1)^2 \geq 0,$$

$$det\left(\begin{bmatrix} \widetilde{y}_{33} & \widetilde{y}_{34} \\ \widetilde{y}_{34} & \widetilde{y}_{44} \end{bmatrix}\right) \geq (\widetilde{y}_{34}^2 + |t_2|) \cdot \widetilde{y}_{44} - \widetilde{y}_{34}^2 \geq |t_2|.$$

Therefore, the matrix

$$\widetilde{y} := diag\left(\begin{bmatrix} \widetilde{y}_{11} & \widetilde{y}_{12} \\ \widetilde{y}_{12} & \widetilde{y}_{22} \end{bmatrix}, \begin{bmatrix} \widetilde{y}_{33} & \widetilde{y}_{34} \\ \widetilde{y}_{34} & \widetilde{y}_{44} \end{bmatrix}, \widetilde{y}_{55}\right)$$

is positive semidefinite, fulfilling $\widetilde{y}_{11} \geq 1$*,* $y_{11}, y_{12} \in \mathbb{Z}_+$*, and* $W\widetilde{y} = t$*. Since* t *was arbitrary, we obtain* $W\left(\{y \in \mathcal{S}^5_+ : y_{11} \geq 0,\ y_{11}, y_{12} \in \mathbb{Z}_+ \backslash \{0\}\}\right) = \mathbb{R}^s$*.*

The dual to the SDP relaxation of (2.28) reads

$$\max\{t^T u + u_5 : W^T u + e_1 u_5 \preceq q,\ u_5 \geq 0\}.$$

Here, $u \in \mathbb{R}^5$ complies with $W^T u + e_1 u_5 \preceq q$ and $u_5 \geq 0$ if

$$\begin{bmatrix} -2u_1 - u_3 - u_5 & -1/2 \cdot (u_2 + u_4) & & & & \\ -1/2 \cdot (u_2 + u_4) & u_1 & & & & \\ & & 2u_2 & 1/2 \cdot u_3 & & \\ & & 1/2 \cdot u_3 & u_4 & & \\ & & & & -u_2 + 1 & \\ & & & & & u_5 \end{bmatrix} \succeq 0.$$

The latter matrix inequality is satisfied for $u_1 = u_2 = u_3 = u_4 = u_5 = 0$, such that the dual to the SDP relaxation of (2.28) has a non-empty feasible set.

Hence, for arbitrary $t \in \mathbb{R}^4$, (2.28) is primal and dual feasible, and, moreover, all matrices are rational. However, program (2.28) is not well-defined for each $t \in \mathbb{R}^4$ as its minimum is not attained for $\widetilde{t} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$. To see this, set $t = \widetilde{t}$ in (2.28). Doing so, (2.28) turns into

$$\min \left\{ \alpha : \begin{bmatrix} x & y & & & & \\ y & 2x & & & & \\ & & 1/2 \cdot (\alpha + y) & x & & \\ & & x & y & & \\ & & & & \alpha & \\ & & & & & x - 1 \end{bmatrix} \succeq 0, \quad x, y \in \mathbb{Z} \right\}. \tag{2.29}$$

Now, so that (2.29) becomes feasible, α has to be greater or equal to zero. This implies that (2.29) is bounded below by 0. Moreover, $\alpha = 0$ is feasible, if the matrix

$$\begin{bmatrix} x & y & & \\ y & 2x & & \\ & & 1/2 \cdot y & x \\ & & x & y \end{bmatrix}$$

is positive semidefinite. The latter holds for any $x, y \geq 0$ satisfying $2x^2 - y^2 = 0$.

As x has to be greater or equal to 1, this implies $\sqrt{2} = \frac{y}{x}$. We define the sequence $\alpha_n := 2 \cdot \left(\frac{y_n}{x_n}\right)^{-2} - 1$ with $x_n \geq 1$, $x_n, y_n \in \mathbb{Z}_+$, $2x_n^2 - y_n^2 \leq 0$, and $\frac{y_n}{x_n} \to \sqrt{2}$. This sequence always stays feasible for (2.29) and its objective value gets arbitrarily close to 0. However, the value 0 will never be attained as $x, y \in \mathbb{Z}$.

Furthermore, it is noted that the semidefinite cone $\mathcal{S}_+^m$ is not polyhedral and therefore its image under a linear map is not necessarily closed. This is illustrated in Example 2.4.

Example 2.4. *Consider the linear map $W : \mathcal{S}^2 \to \mathbb{R}^2$ that is defined by*

$$Wy = \begin{bmatrix} W_1^T y \\ W_2^T y \end{bmatrix} = \begin{bmatrix} y_{11} \\ y_{12} \end{bmatrix},$$

where

$$W_1 = vec\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right), \quad W_2 = vec\left(\begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}\right).$$

Then, the image of the semidefinite cone

$$\mathcal{S}_+^2 := \left\{(y_{11}\ y_{12}\ y_{22})^T \in \mathbb{R}^3 : y_{11} \geq 0,\ y_{11}y_{22} - y_{12}^2 \geq 0\right\}$$

under W is not closed. In fact, the image of the matrix sequence defined by

$$y_n := \begin{bmatrix} \frac{1}{n} & 1 \\ 1 & n \end{bmatrix} \in \mathcal{S}_+^2, \quad 1 \leq n \in \mathbb{N},$$

under the (linear) map W converges to $\lim_{n\to\infty} Wy_n = \lim_{n\to\infty} \begin{bmatrix} \frac{1}{n} & 1 \end{bmatrix}^T = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$, which therefore lies in the closure of $W(\mathcal{S}_+^2)$. However, the latter point cannot be contained in $W(\mathcal{S}_+^2)$ because a matrix satisfying $y_{11} = 0$ and $y_{12} = 1$ cannot be positive semidefinite.

Lemma 2.27. *Assume that either one of the following conditions holds:*

(i) $\{u \in \mathbb{R}^s : W^T u \succ 0\} \neq \emptyset$.

(ii) $\{y \in \mathcal{S}_+^m : Wy = 0\} = \{0\}$.

Then $W(\mathcal{S}_+^m)$ is closed.

Proof. Let us assume that (i) holds true. Consider the set

$$C := \left\{ y \in \mathcal{S}_+^m : \langle W^T \widetilde{u}, y \rangle = 1 \right\},$$

where $\widetilde{u}$ should satisfy $W^T \widetilde{u} \succ 0$. The set C is closed as the intersection of two closed sets. Furthermore, it can be shown that it is bounded. Indeed, let us assume that C is unbounded, containing $y_n \in C$ with $\|y_n\| \to \infty$. If we further define $\widetilde{y}_n = y_n / \|y_n\|$, then $\|\widetilde{y}_n\| \to 1$, such that there exists a subsequence converging to some $0 \neq \widetilde{y} \in \mathcal{S}_+^m$ with

$$\langle W^T \widetilde{u}, \widetilde{y} \rangle = \lim_{n \to \infty} \underbrace{\langle W^T \widetilde{u}, \widetilde{y}_n \rangle}_{=1} / \|y_n\| = 0,$$

which due to

$$\langle W^T \widetilde{u}, \widetilde{y} \rangle \geq \lambda_{min}(W^T \widetilde{u}) \cdot \lambda_{max}(\widetilde{y})$$

contradicts our assumption $W^T \widetilde{u} \succ 0$. Thus, C is compact. Moreover, because of $\langle W^T \widetilde{u}, y \rangle = \langle \widetilde{u}, Wy \rangle = 1$, $W(C)$ does not contain the origin. Since the conical hull is closed for non-empty and compact sets that does not contain the origin (see, for instance, Proposition 1.4.7 in Chapter III. of [58]), $\operatorname{cone}(W(C))$ is closed. Finally, as the set C contains the system

$$\{y : y = (1/\alpha) \cdot xx^T,\ x \in \mathbb{R}^m,\ \|x\| = 1,\ \alpha = x^T \left(W^T \widetilde{u} \right) x > 0\},$$

which forms a generating system for $\mathcal{S}_+^m$, $\operatorname{cone}(W(C)) = W(\mathcal{S}_+^m)$.

If (ii) is valid, then by Theorem 3.10 in [59], $W(\mathcal{S}_+^m)$ is closed. □

Remark 2.28. The assumptions made in Lemma 2.27 could be softened. Instead of claiming (i), a sufficient condition for $W(\mathcal{S}_+^m)$ to be closed (cf. [60]) is

$$\operatorname{ri} \left(\mathcal{S}_+^m \cap \{ z \in \mathcal{S}^m : z = W^T u,\ u \in \mathbb{R}^s \} \right) \neq \emptyset,$$

where ri denotes the relative interior[8]. Furthermore, if there exists a matrix

$$\overline{y} = \begin{bmatrix} 0 & 0 \\ 0 & \overline{y}_{22} \end{bmatrix} \in \mathcal{S}^m_+$$

with $\overline{y}_{22} \succ 0$ fulfilling $W\overline{y} = 0$. Then, $W(\mathcal{S}^m_+)$ is closed if for all matrices belonging to the set $\{z \in \mathcal{S}^m : z = W^T u,\ u \in \mathbb{R}^s\}$ and are of the shape

$$\begin{bmatrix} \overline{z}_{11} & \overline{z}_{12} \\ \overline{z}_{12}^T & 0 \end{bmatrix},$$

it follows $\overline{z}_{12} = 0$ (cf. [61]).

Theorem 2.29. *Assume* $W(\mathcal{S}^m_+) + W'(\mathbb{Z}^{m'}_+) = \mathbb{R}^s$, $M_D = \{u \in \mathbb{R}^s : W^T u \prec q,\ W'^T u \leq q'\} \neq \emptyset$, *and that* $W(\mathcal{S}^m_+)$ *is closed*[9]*. Moreover, assume that the optimal value function (2.27) is well-defined. Then, (2.27) is lower semi-continuous on* $\mathbb{R}^s$*.*

Proof. Let $t_n \to t$. Then, due to complete mixed-integer recourse, there has to exist a sufficiently large $n_0 \in \mathbb{N}$ and $y' \in \mathbb{Z}^{m'}_+$ such that $t_n - W'y' \in W(\mathcal{S}^m_+)$ for all $n \geq n_0$. Moreover, as $W(\mathcal{S}^m_+)$ is assumed to be closed, we obtain $t - W'y' \in W(\mathcal{S}^m_+)$. This yields $\mathbb{Z}^{m'}_+(t_n) \subseteq \mathbb{Z}^{m'}_+(t)$ for all $n \geq n_0$. Hence,

$$\begin{aligned}
\liminf_{t_n \to t} \Phi(t_n) &= \liminf_{t_n \to t} \min_{y' \in \mathbb{Z}^{m'}_+(t_n)} \left[q'^T y' + \max\{(t_n - W'y')^T u : W^T u \preceq q\} \right] \\
&\geq \min_{y' \in \mathbb{Z}^{m'}_+(t)} \left[q'^T y' + \max\{(t - W'y')^T u : W^T u \preceq q\} \right] \\
&= \Phi(t),
\end{aligned}$$

which is the asserted lower semi-continuity of Φ at t. □

In order to obtain the lower semi-continuity of the functional Φ, we cannot sacrifice the closedness condition of $W(\mathcal{S}^m_+)$.

[8] For any nonempty convex set $C \subseteq \mathbb{R}^n$ its relative interior can be defined as $\text{ri}(C) := \{x \in C : \forall y \in C\ \exists \lambda > 1 : \lambda x + (1 - \lambda) y \in C\}$.

[9] Assume, for instance, that one of the conditions in Lemma 2.27 holds.

Example 2.5. *Let W_1, W_2 as in Example 2.4, and define*

$$W_3 = vec\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right), \quad W' = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

Consider the following optimal value function

$$\Phi(t) = \min\left\{ \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} y' : \begin{bmatrix} W_1^T \\ W_2^T \\ W_3^T \end{bmatrix} y + W'y' = t,\ y \in \mathcal{S}_+^2,\ y' \in \mathbb{Z}_+^3 \right\}. \tag{2.30}$$

It is easy to see, that complete mixed-integer recourse holds true. Moreover, for the vector $\widetilde{u} \in \mathbb{R}^3$ with $\widetilde{u}_1 = \widetilde{u}_3 = -1$ and $\widetilde{u}_2 = 0$ it holds

$$q - W^T\widetilde{u} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \succ 0,$$

and

$$W'^T\widetilde{u} - q' = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}^T - \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T \leq 0,$$

such that we in addition have $M_D = \{u \in \mathbb{R}^s : W^T u \prec q, W'^T u \leq q'\} \neq \emptyset$. However, the optimal value function (2.30) is not lower semi-continuous. Indeed, if we fix the second and third component of $t \in \mathbb{R}^3$ to one, i.e., $t_2 = t_3 = 1$, the above program becomes

$$\begin{aligned} & \min\{y_1' : \ y_1' = y_{11} - t_1,\ y_{12} = y_{21} = 1,\ y_{22} = 1 - y_2' + y_3' \\ & \qquad\qquad y_{11} \geq 0,\ y_{22} \geq 0,\ y_{11}y_{22} \geq y_{12}^2,\ y_1', y_2', y_3' \in \mathbb{Z}_+ \ \} \\ = \ & \min\{y_1' : y_1' = y_{11} - t_1,\ y_{11} \geq 0,\ y_{11}y_{22} \geq 1,\ y_1', y_{22} \in \mathbb{Z}_+\} \\ = \ & \begin{cases} 0, & \text{for} \quad t_1 > 0, \\ \lfloor |t_1| \rfloor + 1, & \text{for} \quad t_1 \leq 0. \end{cases} \end{aligned}$$

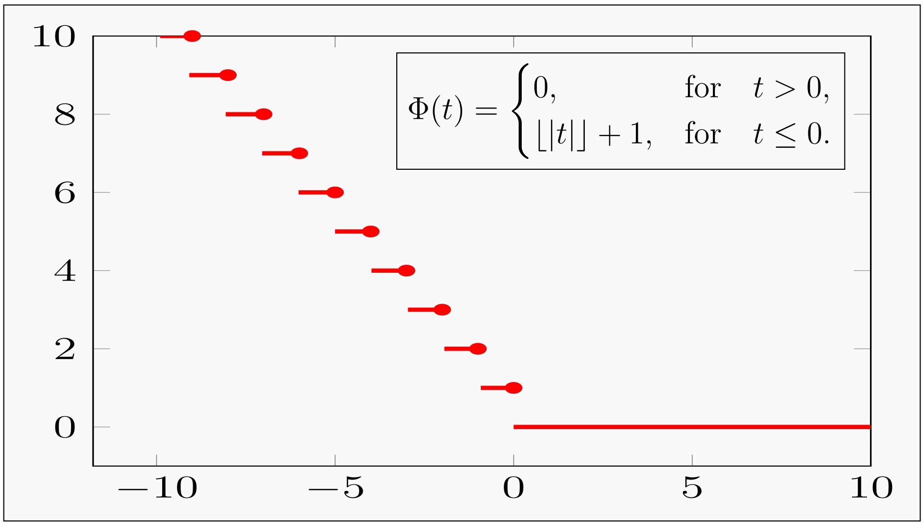
If now the above value function Φ were lower semi-continuous, so would its restriction to the affine hyperplane $\{t \in \mathbb{R}^3 : t_2 = t_3 = 1\}$. Consider the sequence

$t_n = \left[-1+\frac{1}{n}, 1, 1\right]^T \to t = [-1,1,1]^T$, *then*

$$\liminf_{n\to\infty} \Phi(t_n) = \liminf_{n\to\infty} \lfloor 1 - \frac{1}{n} \rfloor + 1 = 1 < 2 = \Phi(t).$$

Hence, Φ *fails to be lower semi-continuous.*

FIGURE 2.2: Plot of the optimal value function introduced in Example 2.5 restricted to the affine hyperplane $\{t \in \mathbb{R}^3 : t_2 = t_3 = 1\}$.



2.2.1 The Expected Value Functional

In this section we consider the expected value functional that is induced by mixed-integer recourse (2.20). This leads us to the following function

$$\mathcal{Q}_{\mathbb{E}}(x) = c^T x + \int_{\mathbb{R}^s} \Phi(z - Tx)\, \mu(dz), \tag{2.31}$$

where

$$\Phi(t) := \min\{q^T y + q'^T y' : Wy + W'y' = t,\; y \in \mathcal{S}_+^m,\; y' \in \mathbb{Z}_+^{m'},\; \|y'\|_\infty \leq K\}.$$

Remark 2.30. Note that the integrand Φ in (2.31) is lower semi-continuous (see Theorem 2.29) and thus, in particular, measurable.

Due to estimate (2.24) and as a consequence of Fatou's lemma, lower semi-continuity for (2.31) can be derived[10].

Theorem 2.31. *Assume* $W(\mathcal{S}_+^m) + W'(\mathbb{Z}_+^{m'} \cap [0, K]^{m'}) = \mathbb{R}^s$*,* $M_D = \{u \in \mathbb{R}^s : W^T u \prec q\} \neq \emptyset$*, and* $\int_{\mathbb{R}^s} \|z\| \mu(dz) < \infty$*. Then the expected value function*

$$\mathcal{Q}_{\mathbb{E}}(x) = c^T x + \int_{\mathbb{R}^s} \Phi(z - Tx)\ \mu(dz)$$

is lower semi-continuous on $\mathbb{R}^s$*.*

Proof. In view of Theorem 2.29, the integrand Φ in (2.31) is lower semi-continuous on $\mathbb{R}^s$ and thus measurable. Furthermore, according to Theorem 2.25 and due to $\Phi(0) = 0$, the following inequality holds true

$$\begin{aligned}\Phi(z - Tx)\ &\geq \Phi(0) - |\Phi(z - Tx) - \Phi(0)| \\ &\geq -\beta\|z - Tx\| - \gamma \\ &\geq -\beta\|z\| - \beta\|T\|\|x\| - \gamma.\end{aligned}$$

Since it is assumed that the image measure $\mu = \mathbb{P} \circ z^{-1}$ has finite first moment, it holds: for any sequence $(x_n)_{n\in\mathbb{N}} \in \mathcal{S}^n$ with $x_n \to x_0$ and $r := \max_{n\in\mathbb{N}} \|x_n\|$, the function

$$g(z) := -\beta\|z\| - \beta r\|T\| - \gamma$$

is an integrable minorant for the functions $g_n(z) := \Phi(z - Tx_n), n \in \mathbb{N}$. Hence, Fatou's lemma together with the semi-continuity of Φ yields the following estimate

$$\begin{aligned}\mathcal{Q}_{\mathbb{E}}(x) - c^T x\ &= \int_{\mathbb{R}^s} \Phi(z - Tx)\ \mu(dz) \\ &\leq \int_{\mathbb{R}^s} \liminf_{n\to\infty} \Phi(z - Tx_n)\ \mu(dz) \\ &\leq \liminf_{n\to\infty} \int_{\mathbb{R}^s} \Phi(z - Tx_n)\ \mu(dz) = \liminf_{n\to\infty} \mathcal{Q}_{\mathbb{E}}(x_n) - c^T x_n,\end{aligned}$$

i.e., $\mathcal{Q}_{\mathbb{E}}$ is lower semi-continuous on $\mathcal{S}^n$. □

Remark 2.32. The first continuity result for the expected value functional for mixed-integer linear recourse is due to Stougie [62]. Further properties, such as

[10]To show this, the general methodology as described in Chapter 1 of [11] has been applied.

sufficient conditions for lower semi-continuity and Lipschitz continuity on bounded subsets, are presented by Schultz in [63] and [64].

Theorem 2.33. *Assume* $W(\mathcal{S}_+^m) + W'(\mathbb{Z}_+^{m'} \cap [0, K]^{m'}) = \mathbb{R}^s$*,* $M_D = \{u \in \mathbb{R}^s : W^T u \prec q\} \neq \emptyset$*, and* $\int_{\mathbb{R}^s} \|z\|\mu(dz) < \infty$*. If, furthermore,* $\mu(E(x)) = 0$*, where*

$$E(x) := \{z \in \mathbb{R}^s : \Phi \textit{ is discontinuous at } z - Tx\},$$

then $\mathcal{Q}_\mathbb{E}$ *as defined in (2.31) is continuous at* x*.*

Proof. Let $(x_n)_{n\in\mathbb{N}} \in \mathcal{S}^n$ with $x_n \to x_0$. As the set $E(x)$ is assumed to have μ-measure zero, we have

$$\lim_{n\to\infty} \Phi(z - Tx_n) = \Phi(z - Tx_0), \quad \text{for } \mu\text{-almost all } z \in \mathbb{R}^s.$$

Further, in view of the proof of Theorem 2.31, we obtain that the function

$$\beta\|\cdot\| + \beta r\|T\| + \gamma,$$

with $r := \max_{n\in\mathbb{N}} \|x_n\|$, provides an integrable majorant for the functions $|\Phi(\cdot - Tx_n)|$. Hence, Lebesgue's dominated convergence theorem implies

$$\lim_{n\to\infty} \int_{\mathbb{R}^s} \Phi(z - Tx_n)\, \mu(dz) = \int_{\mathbb{R}^s} \Phi(z - Tx_0)\, \mu(dz),$$

which yields the desired continuity of $\mathcal{Q}_\mathbb{E}$. □

Remark 2.34. For stochastic mixed-integer linear programs, the set of discontinuity points of Φ is contained in a countable union of hyperplanes, which has Lebesgue measure zero. Therefore, in this case, $\mathcal{Q}_\mathbb{E}$ is continuous if μ has a density.

2.2.2 The Excess Probability Functional

Let us introduce for $x \in \mathcal{S}_+^n$ the sets

$$\begin{aligned} M(x) &:= \{z \in \mathbb{R}^s \;:\; c^T x + \Phi(z - Tx) > \eta\}, \\ M_e(x) &:= \{z \in \mathbb{R}^s \;:\; c^T x + \Phi(z - Tx) = \eta\}, \\ M_d(x) &:= \{z \in \mathbb{R}^s \;:\; \Phi \text{ is discontinuous at } z - Tx\}, \end{aligned}$$

where Φ should be the optimal value function as defined in (2.20).

Theorem 2.35. *Assume $W(\mathcal{S}^m_+) + W'(\mathbb{Z}^{m'}_+ \cap [0,K]^{m'}) = \mathbb{R}^s$ and $M_D = \{u \in \mathbb{R}^s : W^T u \prec q\} \neq \emptyset$. Then $\mathcal{Q}_{\mathbb{P}_\eta}(x) : \mathcal{S}^n_+ \to \mathbb{R}$, with*

$$\mathcal{Q}_{\mathbb{P}_\eta}(x) = \mu\Big[\big\{z \in \mathbb{R}^s \ : \ c^T x + \Phi(z - Tx) > \eta\big\}\Big],$$

is real-valued and lower semi-continuous on $\mathcal{S}^n_+$. If, furthermore, for some $x \in \mathcal{S}^n_+$, it holds $\mu(M_e(x) \cup M_d(x)) = 0$, then $\mathcal{Q}_{\mathbb{P}_\eta}$ is continuous at x.

Proof. As $W(\mathcal{S}^m_+) + W'(\mathbb{Z}^{m'}_+ \cap [0,K]^{m'}) = \mathbb{R}^s$ holds true if and only if $W(\mathcal{S}^m_+) = \mathbb{R}^s$, Theorem 2.29 provides us with the lower semi-continuity of the optimal value function Φ. Consequently, the set $M(x)$ is μ-measurable for all $x \in \mathcal{S}^n_+$, implying that $\mathcal{Q}_{\mathbb{P}_\eta}$ is real-valued on $\mathcal{S}^n_+$.

In order to obtain the asserted lower semi-continuity of $\mathcal{Q}_{\mathbb{P}_\eta}$ (which will result from the lower semi-continuity of Φ and the lower semi-continuity of probability measures), we can proceed similar to the proof of Theorem 2.18. To this end, let $\overline{z} \in M(x)$. Then, the lower semi-continuity of Φ yields

$$\liminf_{x_n \to x} \big(c^T x + \Phi(\overline{z} - Tx_n)\big) \geq c^T x + \Phi(\overline{z} - Tx) > \eta.$$

According to this, there has to exist $n_0 \in \mathbb{N}$ such that

$$c^T x_n + \Phi(\overline{z} - Tx_n) > \eta \quad \text{ for } n \geq n_0,$$

or, in other words, $\overline{z} \in M(x_n)$ for all $n \geq n_0$. The latter implies

$$M(x) \subseteq \liminf_{x_n \to x} M(x_n).$$

Hence, in view of the semi-continuity of probability measures (2.18) the following holds for all $x \in \mathcal{S}^n_+$:

$$\begin{aligned}\mathcal{Q}_{\mathbb{P}_\eta}(x) = \mu(M(x)) &\leq \mu(\liminf_{x_n \to x} M(x_n)) \\ &\leq \liminf_{x_n \to x} \mu(M(x_n)) \ = \ \liminf_{x_n \to x} \mathcal{Q}_{\mathbb{P}_\eta}(x_n),\end{aligned}$$

i.e., $\mathcal{Q}_{\mathbb{P}_\eta}$ is lower semi-continuous on $\mathcal{S}^n_+$.

Furthermore, let $\overline{z} \in \limsup_{x_n \to x} M(x_n) \backslash M(x)$. If so, there must be an infinite subset $\widetilde{\mathbb{N}}$ of $\mathbb{N}$ such that

$$c^T x_n + \Phi(\overline{z} - Tx_n) > \eta, \ \forall n \in \widetilde{\mathbb{N}} \qquad \text{and} \qquad c^T x + \Phi(\overline{z} - Tx) \leq \eta. \tag{2.32}$$

Now, exactly two cases are possible: Either the optimal value function Φ is discontinuous at $\overline{z} - Tx$, or it is continuous at $\overline{z} - Tx$. The former implies $\overline{z} \in M_d(x)$ whereas the latter and (2.32) yield $c^T x + \Phi(\overline{z} - Tx) = \eta$, i.e., $\overline{z} \in M_e(x)$. Thus, in case $\mu(M_e(x) \cup M_d(x)) = 0$, we obtain the following estimate

$$\begin{aligned}
\mathcal{Q}_{\mathbb{P}_\eta}(x) = \mu(M(x)) &= \mu\Big(M(x) \cup M_e(x) \cup M_d(x)\Big) \\
&\geq \mu\left(\limsup_{x_n \to X} M(x_n)\right) \geq \limsup_{x_n \to x} \mu(M(x_n)) = \limsup_{x_n \to x} \mathcal{Q}_{\mathbb{P}_\eta}(x_n).
\end{aligned}$$

Note that the latter is the upper semi-continuity of $\mathcal{Q}_{\mathbb{P}_\eta}$ at x. Hence, together with the lower semi-continuity of $\mathcal{Q}_{\mathbb{P}_\eta}$ (derived in the first part of this proof) this yields the continuity of $\mathcal{Q}_{\mathbb{P}_\eta}$ at x. $\square$

2.3 Deterministic Equivalents for Finite Discrete Probability Distributions

Let us assume that the underlying random variable $z(\xi)$ follows a finite discrete probability distribution with realizations (scenarios) z_ξ and probabilities $\pi_\xi, \xi = 1, ..., S$. Then, for two-stage continuous (linear semidefinite) recourse (2.1), the excess probability mean-risk model

$$\min_{x \in X} \left\{ \underbrace{\mathbb{E}_\xi \left[c^T x + \Phi(z(\xi) - Tx) \right]}_{:= \mathcal{Q}_{\mathbb{E}}(x)} + \rho \cdot \underbrace{\mathbb{P}_\xi \left[\left\{ c^T x + \Phi(z(\xi) - Tx) > \eta \right\} \right]}_{:= \mathcal{Q}_{\mathbb{P}_\eta}(x)} \right\}, \tag{2.33}$$

where

$$\Phi(t) := \min\left\{ q^T y : Wy = t, \ y \in \mathcal{S}_+^m \right\},$$

adopts a block structure as revealed in the following theorem.

Theorem 2.36. *Assume* $W(\mathcal{S}_+^m) = \mathbb{R}^s$, $M_D := \{u \in \mathbb{R}^s : W^T u \prec q\} \neq \emptyset$, *and that* X *is compact. Then there exists a constant* $M > 0$ *such that the excess probability* *mean-risk model (2.33) can be equivalently restated as*

$$\begin{aligned} \min \quad & c^T x + \sum\nolimits_{\xi=1}^{S} \pi_\xi q^T y_\xi + \rho \cdot \sum\nolimits_{\xi=1}^{S} \pi_\xi \theta_\xi \\ s.t. \quad & Tx + W y_\xi = z_\xi, \\ & c^T x + q^T y_\xi - M\theta_\xi \leq \eta, \\ & x \in X, \quad y_\xi \in \mathcal{S}_+^m, \quad \theta_\xi \in \{0,1\}, \qquad \xi = 1,..,S. \end{aligned} \tag{2.34}$$

Proof. Before showing the equivalence of the mentioned models, let us confirm that compactness of X yields existence of the required constant M. Indeed, let

$$M > \sup\{c^T x + \Phi(z_\xi - Tx) : x \in X, \xi = 1,...,S\} - \eta. \tag{2.35}$$

To see that the supremum on the right is bounded consider for each $\xi = 1,...,S$ the estimate

$$\sup_{x \in X} c^T x + \Phi(z_\xi - Tx) \leq \sup_{x \in X} \|c\| \cdot \|x\| + \sup_{x \in X} \max_{u \in M_D^{\preceq}} (z_\xi - Tx)^T u. \tag{2.36}$$

Since $M_D^{\preceq}$ is compact (due to our assumptions and Lemma 2.3), so is $X \times M_D^{\preceq}$. Moreover, as both $\|c\| \cdot \|x\|$ and $(z_\xi - Tx)^T u$ are continuous functions in (x,u), finiteness in (2.35) follows via (2.36) from Weierstrass' theorem.

Now let us turn to the equivalence of the models (2.33) and (2.34). Let $\overline{x}$ be an optimal solution to (2.33) and assume there is a feasible (x^*, y^*, θ^*) to (2.34) whose objective value in (2.33) is less than $\mathcal{Q}_{\mathbb{E}}(\overline{x}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(\overline{x})$.

By the definition of Φ (cf. (2.3)), it holds $\Phi(z_\xi - Tx^*) \leq q^T y^*, \ \forall \omega$. This yields

$$\mathcal{Q}_{\mathbb{E}}(x^*) = c^T x^* + \sum\nolimits_{\xi=1}^{S} \pi_\xi \Phi(z_\xi - Tx^*) \leq c^T x^* + \sum\nolimits_{\xi=1}^{S} \pi_\xi q^T y^*$$

and, moreover, the following implication holds:

$$\theta_\xi^* = 0 \quad \Rightarrow \quad c^T x^* + \Phi(z_\xi - Tx^*) \leq \eta.$$

Thus, we obtain the inclusion

$$\{\xi : c^T x^* + \Phi(z_\xi - Tx^*) > \eta\} \subseteq \{\xi : \theta_\xi^* = 1\},$$

which yields $\mathcal{Q}_{\mathbb{P}_\eta}(x^*) \leq \sum_{\xi=1}^S \pi_\xi \theta_\xi^*$. Altogether, we get

$$\begin{aligned}\mathcal{Q}_{\mathbb{E}}(x^*) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(x^*) &\leq c^T x^* + \sum\nolimits_{\xi=1}^S \pi_\xi q^T y^* + \sum\nolimits_{\xi=1}^S \pi_\xi \theta_\xi^* \\ &< \mathcal{Q}_{\mathbb{E}}(\overline{x}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(\overline{x}),\end{aligned}$$

contradicting the optimality of $\overline{x}$ in (2.33), i.e. the optimal value of (2.34) is, in any case, an upper bound.

Furthermore, to see equality, let vice versa $\overline{x}$ be optimal in (2.33). Set

$$\overline{y}_\xi \in \arg\min\{q^T y_\xi : W y_\xi = z_\xi - T\overline{x},\ y_\xi \in \mathcal{S}_+^m\},$$

and

$$\overline{\theta}_\xi = \begin{cases} 0 & \text{if} \quad c^T\overline{x} + q^T\overline{y}_\xi - \eta \leq 0, \\ 1 & \text{otherwise,} \end{cases}$$

for $\xi = 1, ..., S$. Then

$$\mathcal{Q}_{\mathbb{E}}(\overline{x}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(\overline{x}) = c^T\overline{x} + \sum\nolimits_{\xi=1}^S \pi_\xi q^T \overline{y}_\xi + \rho \cdot \sum\nolimits_{\xi=1}^S \pi_\xi \overline{\theta}_\xi,$$

where in addition $(\overline{x}, \overline{y}, \overline{\theta})$ is feasible to (2.34). This completes the proof. □

Remark 2.37. The structure of (2.34) has similarity to two-stage chance-constrained models. While (2.34) penalizes the violation of $f(x, \xi) := c^T x + \Phi(z(\xi) - Tx) > \eta$ by a multiple of its probability, chance-constrained models limit the choice of first-stage actions $x \in X$ by postulating that the probability $\mathbb{P}[\{\omega : f(x, \xi) > \eta\}]$ is less or equal to a given threshold. A recent work considering two-stage chance-constrained models is [65].

Remark 2.38. In case the underlying random variable does not have finite support, one often is led to approximations by finite discrete distributions (as numerical methods take only those). For instance, if the resulting program does not allow for an analytic solution. Moreover, note that exact computation of expectations and probabilities is beyond the present numerical capabilities for a large class of distributions, where this especially includes multivariate continuous ones.

For mixed-integer recourse (2.20) the result from Theorem 2.36 holds true as well.

Corollary 2.39. *Assume* $W(\mathcal{S}_+^m) + W'(\mathbb{Z}^{m'} \cap [0, K]^{m'}) = \mathbb{R}^s$*,* $M_D := \{u \in \mathbb{R}^s : W^T u \prec q,\ W'^T u \leq q'\} \neq \emptyset$*, and that* X *is compact. Then there exists a constant*

$M > 0$ *such that the excess probability mean-risk model with mixed-integer recourse (2.20) can be equivalently expressed by*

$$\begin{aligned} \min \quad & c^T x + \sum\nolimits_{\xi=1}^{S} \pi_\xi \left(q^T y_\xi + q'^T y'_\xi \right) + \rho \cdot \sum\nolimits_{\xi=1}^{S} \pi_\xi \theta_\xi \\ s.t. \quad & Tx + W y_\xi + W' y'_\xi = z_\xi, \\ & c^T x + q^T y_\xi + q'^T y'_\xi - M \theta_\xi \le \eta, \\ & x \in X, \quad y_\xi \in \mathcal{S}_+^m, \quad y'_\xi \in \mathbb{Z}_+^{m'} \cap [0, K]^{m'}, \quad \theta_\xi \in \{0,1\}, \qquad \xi = 1, .., S. \end{aligned} \tag{2.37}$$

Proof. Theorem 2.25 provides us with the following estimate

$$\sup_{x \in X} c^T x + \Phi(z_\xi - Tx) \le \sup_{x \in X} \|c\| \cdot \|x\| + \beta \cdot \|T\| \cdot \|x\| + \beta \cdot \|z_\xi\| + \gamma, \tag{2.38}$$

where $\beta = \max_{u \in M_D^{\preceq}} \|u\|$ and $\gamma = (\|q'\| + \beta \cdot \|W'\|) \cdot 2\sqrt{m'} K$ (it is noted that due to our assumptions and Lemma 2.26, $M_D^{\preceq}$ is compact). Hence, estimation (2.36) holds true as well. The remaining part of the proof is completely analogous to the proof of Theorem 2.36. □

Remark 2.40. In establishing the stochastic programs (2.34) and (2.37), respectively, the most challenging part is to find a discrete probability distribution that reasonably approximates the true probability measure through which the stochastic program is formulated. In problems of practical interest, a good many times, there is a sample from historical time series available, which then can be used to gain such an approximation. The justification of using the empirical sample lies in the fact that those samples may converge to the true underlying probability distribution if the sample size increases.

Remark 2.41. Note that the incorporation of more and more scenarios leads to huge programs. Therefore, it is eligible to minimize the number of scenarios in order to keep the computational complexity low. Approximations of the underlying finite probability measure by a measure that has smaller support can be obtained by scenario reduction[11] [66], [67]. Doing so, the quality of a scenario approximation depends on the distance between the original probability and the scenario

[11] The idea of scenario reduction is the approximation of the underlying finite probability measure by a measure having smaller support in terms of a probability metric which can be associated to the stochastic program.

probability. Quantitative stability results that establish the relation between distances of probability models on one side and distances between optimal values or solutions on the other side are presented in [68], [69], [70], [71], and [72].

Chapter 3

Decomposition Methods

To facilitate notation, let us first neglect the risk measure functional $\mathcal{Q}_{\mathbb{P}^\eta}$ in (2.33) and consider the risk neutral model (2.9), i.e.,

$$\min\{\mathcal{Q}_{\mathbb{E}}(x) : Ax = b,\ x \in \mathcal{S}^n_+\}.$$

Assume that the underlying random variable follows a finite discrete probability distribution with realizations (scenarios) z_ξ and probabilities $\pi_\xi, \xi = 1, ..., S$. Then, according to Theorem 2.36 and by using the traditional notation in semidefinite programming, (2.9) can be equivalently expressed by:

$$\begin{array}{lllll} \min & C \bullet \boldsymbol{X} + \sum_{\xi=1}^{S} \pi_\xi H \bullet \boldsymbol{Y}_\xi & & & \\ \text{s.t.} & \mathcal{T}\boldsymbol{X} + \mathcal{W}\boldsymbol{Y}_1 & & & = z_1, \\ & \quad\vdots & \ddots & & \vdots \\ & \mathcal{T}\boldsymbol{X} & & + \mathcal{W}\boldsymbol{Y}_S & = z_S, \\ & \boldsymbol{X} \in \mathcal{X},\ \boldsymbol{Y}_1 \succeq 0,\ \dots, \boldsymbol{Y}_S \succeq 0. & & & \end{array} \tag{3.1}$$

Here, the term $C \bullet X$ denotes the trace of the matrix product $C^T X$ (i.e., the sum of the diagonal elements of the square matrix $C^T X$), $\mathcal{X}$ is a spectrahedra (i.e., $\mathcal{X}$ is given by $\{X \in \mathcal{S}^n_+ : A_i \bullet X = b_i, i = 1, ..., m\}$), and $\mathcal{T} : \mathcal{S}^n \to \mathbb{R}^s$ and $\mathcal{W} : \mathcal{S}^m \to \mathbb{R}^s$ are linear operators defined by $\mathcal{T}X = [T_1 \bullet X, ..., T_s \bullet X]^T$ and $\mathcal{W}Y = [W_1 \bullet Y, ..., W_s \bullet Y]^T$, respectively.

Since, with growing number of scenarios, the dimension of (3.1) quickly becomes too large for being handled in an all-at-once manner by general SDP solvers, decomposition methods come to the fore. Mehrotra and Özevin [73] propose an extension of [74] to semidefinite programs, leading to a Benders decomposition based interior point method. While the latter works well if there are no integer variables, it fails with integer requirements to second-stage variables. Some of the unit commitment problems studied in this thesis, however, contain substantial numbers of Boolean decision variables in the second stage. For such programs, we therefore will apply Lagrangian relaxation of the nonanticipativity condition which is due to Carøe and Schultz [34].

For risk neutral two-stage stochastic (linear) semidefinite programs (3.1) that do not contain second-stage integers, another algorithm, which is known as the L-shaped method[1] [35], is presented. This method is a very reliable and computationally efficient technique for solving two-stage programs with linear recourse. In contrast to dual decomposition [34], it is able to share information between the decomposed scenario programs in a higher-level master-problem.

The dual decomposition algorithm that is based on Lagrangian relaxation of nonanticipativity is presented in Section 3.1. How to apply Benders decomposition (the L-shaped method) to two-stage stochastic semidefinite programs with continuous recourse is described in Section 3.2.

3.1 Dual Decomposition

Dual decomposition methods are due to Lagrangian duality (cf. [76]). In the stochastic programming literature, diverse algorithms based on duality were proposed. Just to mention a few of them: we have scenario decomposition [77], [78], progressive hedging [79], and augmented Lagrangian methods [80] for stochastic programs with continuous recourse, and an extension of progressive hedging [81] and scenario decomposition [34] for stochastic programs with mixed-integer recourse. Since, in this thesis, the only dual decomposition method that will be applied is that of Carøe and Schultz [34], we refer to this method as dual decomposition.

As mentioned, to solve problem (3.1), we will pursue the dual decomposition method of Carøe and Schultz [34]. Here, Lagrangian relaxation of the nonanticipative first-stage decisions is recommended, which then leads to decomposition

[1]The L-shaped method is the stochastic version of the Benders decomposition approach [75].

into smaller subproblems of tractable dimensions. In implementing this idea, we will closely follow [82]. To this end, we introduce an additional matrix variable $\boldsymbol{X}^*$ plus copies $\boldsymbol{X}_\xi, \xi = 1, ..., S$ of the first-stage variable $\boldsymbol{X}$, and add the requirements

$$\boldsymbol{X}_\xi - \boldsymbol{X}^* = 0, \qquad \xi = 1, ..., S. \tag{3.2}$$

In doing so, we obtain the following equivalent reformulation of (3.1):

$$\begin{array}{lllll}
\min & \sum_{\xi=1}^{S} C \bullet \boldsymbol{X}_\xi + H_\xi \bullet \boldsymbol{Y}_\xi & & & \\
\text{s.t.} & \mathcal{T}\boldsymbol{X}_1 + \mathcal{W}\boldsymbol{Y}_1 & & = z_1, & \\
 & & \ddots & \vdots & \\
 & & \mathcal{T}\boldsymbol{X}_S + \mathcal{W}\boldsymbol{Y}_S & = z_S, & \\
 & & \boldsymbol{X}_\xi - \boldsymbol{X}^* & = 0, & \xi = 1, ..., S, \\
 & \boldsymbol{X}_1 \in \mathcal{X},\ \boldsymbol{Y}_1 \succeq 0, \quad ... & , \boldsymbol{X}_S \in \mathcal{X},\ \boldsymbol{Y}_S \succeq 0. & &
\end{array}$$

Here, we have tacitly denoted $H_\xi := \pi_\xi H$. Relaxing nonanticipativity (3.2) leads to S independent subproblems, each corresponding to a particular scenario. In context of semidefinite programming, we arrive at the following Lagrangian function:

$$L\big(\boldsymbol{X}_1, ..., \boldsymbol{X}_S, \boldsymbol{X}^*, \boldsymbol{Y}_1, ..., \boldsymbol{Y}_S\big) := \sum_{\xi=1}^{S} L_\xi\big(\boldsymbol{X}_\xi, \boldsymbol{X}^*, \boldsymbol{Y}_\xi\big), \tag{3.3}$$

with $L_\xi\big(\boldsymbol{X}_\xi, \boldsymbol{X}^*, \boldsymbol{Y}_\xi\big) := C \bullet \boldsymbol{X}_\xi + H_\xi \bullet \boldsymbol{Y}_\xi + \Lambda_\xi \bullet \big(\boldsymbol{X}_\xi - \boldsymbol{X}^*\big)$. Thus, with the dual function

$$D\big(\boldsymbol{\Lambda}\big) := \min \left\{ \sum_{\xi=1}^{S} L_\xi\big(\boldsymbol{X}_\xi, \boldsymbol{X}^*, \boldsymbol{Y}_\xi\big) \ : \ \begin{array}{ll} \mathcal{T}\boldsymbol{X}_\xi + \mathcal{W}\boldsymbol{Y}_\xi = z_\xi, & \xi = 1, ..., S \\ \boldsymbol{X}_\xi \in \mathcal{X}, \quad \boldsymbol{Y}_\xi \succeq 0, & \xi = 1, ..., S \end{array} \right\},$$

we obtain the associated Lagrangian dual:

$$\max \left\{ D\big(\boldsymbol{\Lambda}_1, ..., \boldsymbol{\Lambda}_S\big) : \boldsymbol{\Lambda}_\xi \in \mathcal{S}^{m_1}, \xi = 1, ..., S \right\}. \tag{3.4}$$

Now, as the auxiliary variable $\boldsymbol{X}^*$ is unconstrained, its coefficients must cancel out when forming the sum for $\xi = 1, ...S$, i.e., $\sum_{\xi=1}^{S} \boldsymbol{\Lambda}_\xi = 0$. We further mention that the dual function is separable, i.e., by determining

$$D_\xi\big(\boldsymbol{\Lambda_\xi}\big) := \min_{\boldsymbol{X}_\xi, \boldsymbol{Y}_\xi} \left\{ C \bullet \boldsymbol{X}_\xi + H_\xi \bullet \boldsymbol{Y}_\xi + \boldsymbol{\Lambda}_\xi \bullet \boldsymbol{X}_\xi : \begin{array}{r} \mathcal{T}\boldsymbol{X}_\xi + \mathcal{W}\boldsymbol{Y}_\xi = z_\xi, \\ \boldsymbol{X}_\xi \in \mathcal{X}, \quad \boldsymbol{Y}_\xi \succeq 0 \end{array} \right\}, \tag{3.5}$$

for $\xi = 1, ..., S$, we obtain $D\big(\boldsymbol{\Lambda}_1, ..., \boldsymbol{\Lambda}_S\big) = \sum_{\xi=1}^S D_\xi\big(\boldsymbol{\Lambda}_\xi\big)$. Using this notation, the Lagrangian dual, arising by Lagrangian relaxation of the nonanticipativity condition (3.2), can be expressed by

$$\max \left\{ \sum_{\xi=1}^S D_\xi\big(\boldsymbol{\Lambda}_\xi\big) : \sum_{\xi=1}^S \boldsymbol{\Lambda}_\xi = 0 \right\},$$

and this in turn is equivalent to

$$\max_{\theta, \Lambda} \left\{ \sum_{\xi=1}^S \boldsymbol{\theta}_\xi : \sum_{\xi=1}^S \boldsymbol{\Lambda}_\xi = 0, \quad \boldsymbol{\theta}_\xi \le D_\xi\big(\boldsymbol{\Lambda}_\xi\big),\ \xi = 1, ..., S \right\}. \tag{3.6}$$

In order to solve the problem above, we will apply proximal bundle methods (cf. [83]). The basic idea is to approximate the constraints in (3.6) by cutting planes and adding a regularization term to the objective. Following this idea, one has to solve the subsequent program at each iteration K:

$$\begin{aligned} \max_{\theta, \Lambda} \quad & \sum_{\xi=1}^S \boldsymbol{\theta}_\xi - \frac{1}{2}\tau \sum_{\xi=1}^S \left\| \boldsymbol{\Lambda}_\xi - \Lambda_\xi^+ \right\|_F^2 \\ \text{s.t.} \quad & \sum_{\xi=1}^S \boldsymbol{\Lambda}_\xi = 0 \\ & \boldsymbol{\theta}_\xi \le D_\xi\big(\Lambda_\xi^{(k)}\big) + X_\xi^{(k)} \bullet \left(\boldsymbol{\Lambda}_\xi - \Lambda_\xi^{(k)} \right), \quad \xi = 1, ..., S, \quad k = 1, ..., K. \end{aligned} \tag{3.7}$$

Taking into account that $-D_\xi$ is convex for all $\xi = 1, ..., S$, $-X_\xi^{(k)}$ is selected as a member of the subdifferential $\partial[-D_\xi](\Lambda_\xi^{(k)})$, given by

$$\left\{ \boldsymbol{X} \in \mathcal{S}^{m_1} : D_\xi(\boldsymbol{\Lambda}) - D_\xi(\Lambda_\xi^{(k)}) + \boldsymbol{X} \bullet (\boldsymbol{\Lambda} - \Lambda_\xi^{(k)}) \le 0,\ \forall \boldsymbol{\Lambda} \in \mathcal{S}^{m_1} \right\},$$

where $(-1)\partial[-D_\xi](\Lambda_\xi^{(k)})$ coincides with the $\boldsymbol{X}_\xi$-part of the optimal solution set to program (3.5). The Point $\big(\Lambda_1^+, ..., \Lambda_S^+\big)$ is the current proximal center, fulfilling $\sum_{\xi=1}^S \Lambda_\xi^+ = 0$, and τ is some regularization parameter which can be adjusted at each iteration.

Remark 3.1. Note that (3.7) is a quadratic program (QP), i.e., it can be tackled by well-established algorithms. Here, one could briefly mention active set strategies, trust region methods, conjugate gradient methods, and interior point methods (see [84], for instance).

Finally, we arrive at the following decomposition method:

ALGORITHM 3.1.1: Decomposition based proximal bundle method.

Initialize:	Accuracy parameter $\epsilon > 0$; $m = 0.1$; $K := 1$; $\tau = 1$;
	set for $\xi = 1, ..., S$, $\Lambda_\xi^+ := 0$ as well as $\Lambda_\xi^{(K)} := 0$;
	solve $D_\xi(\Lambda_\xi^{(K)}), \xi = 1, ..., S$, save optimal solution $X_\xi^{(K)}$;
	and put $curObj := \sum_{\xi=1}^S D_\xi(\Lambda_\xi^{(K)})$.
Step 1.	Solve (3.7), obtaining optimal θ_ξ^* and Λ_ξ^*, for $\xi = 1, ..., S$.
Step 2.	Let $v = \left(\sum_\xi^S \theta_\xi^*\right) - curObj$.
	If $v/(1 + \lvert curObj \rvert) < \epsilon$ terminate; else continue.
Step 3.	$K := K + 1$;
	solve $D_\xi(\Lambda_\xi^*), \xi = 1, ..., S$, save its optimal value $D_\xi(\Lambda_\xi^{(K)})$
	as well as its optimal solution $X_\xi^{(K)}$;
	$newObj := \sum_{\xi=1}^S D_\xi(\Lambda_\xi^{(K)})$;
	$u := 2\tau(1 - (newObj - curObj)/v)$,
	$\tau := \min(\max(u, \tau/10, 10^{-4}), 10\tau)$;
	if $(newObj - curObj > m \cdot v)$,
	then, update $\Lambda_\xi^+ := \Lambda_\xi^*$ and $curObj := newObj$;
	go to Step 1.

3.1.1 Dual Decomposition for Two-Stage Stochastic Mixed-Integer Semidefinite Programs

First, it is shown that the risk averse programs (2.34) and (2.37) can be equivalently transformed into (3.1) with the exception that it contains additional integer requirements to some second-stage variables. To this end, we introduce for $\xi = 1, ..., S$ the second-stage variables

$$\boldsymbol{Y}_{EP\xi} := \operatorname{diag}(\boldsymbol{Y}_\xi, \boldsymbol{\theta}_\xi, \boldsymbol{s}_\xi),$$

and extend the linear matrix operator $\mathcal{T}$ to the linear operator $\mathcal{T}_{EP} : \mathcal{S}^{m_1} \to \mathbb{R}^{s+1}$, defined by $\mathcal{T}_{EP}\boldsymbol{X} = [(\mathcal{T}\boldsymbol{X})^T, C \bullet \boldsymbol{X}]^T$. Moreover, let us define the linear operator $\mathcal{W}_{EP} : \mathcal{S}^{m_2} \to \mathbb{R}^{s+1}$, given by the following modified recourse matrices:

$$W_{EP_1} := \mathrm{diag}(W_1, 0, 0),\ ...,\ W_{EP_s} := \mathrm{diag}(W_s, 0, 0),\ W_{EP_{s+1}} := \mathrm{diag}(H, -M, 1).$$

Finally, by setting $H_{EP\xi} := \mathrm{diag}(\pi_\xi H, \rho, 0)$ and $z_{EP\xi} := (z_\xi^T, \eta)^T$ for $\xi = 1, ..., S$, we obtain that (2.34) is indeed equivalent to:

$$\begin{array}{ll} \min & C \bullet \boldsymbol{X} + \sum_{\xi=1}^{S} H_{EP\xi} \bullet \boldsymbol{Y}_{EP\xi} \\ \text{s.t.} & \mathcal{T}_{EP}\boldsymbol{X} + \mathcal{W}_{EP}\boldsymbol{Y}_{EP\xi} = z_{EP\xi}, \qquad \forall\xi, \\ & \boldsymbol{X} \in \mathcal{X}, \quad \boldsymbol{Y}_{EP\xi} \succeq 0, \qquad\qquad \forall\xi, \\ & \boldsymbol{Y}_{EP\xi}(m_2+1, m_2+1) \in \{0,1\},\ \forall\xi, \end{array} \tag{3.8}$$

which obviously has the same structure as the risk neutral model (3.1), in the sense that there are no constraints involving second-stage variables from different scenarios. Analogously for (2.37).

Tackling the non-convex program (3.8) by the proposed proximal bundle methods may result in a solution that does not meet the nonanticipativity condition. If so, the solution to the Lagrangian dual (3.6) provides us a lower bound. To measure the quality of this lower bound, upper bounds in terms of feasible points are required. Since, the previously relaxed constraints (3.2) are quite simple, namely, we have to make all first-stage copies identical, ideas for heuristics come up straightforwardly. For instance, one may pick from $X_\xi^{opt}, \xi = 1, ..., S$ a candidate, by averaging over them all and rounding to integers (see [85] for further ideas for heuristics).

If the resulting gap (the gap between the generated feasible point and the Lagrangian dual) is unsatisfactory, we recommend the embedding into a Branch-and-Bound scheme, where the underlying two-stage stochastic program is understood as a non-convex global minimization problem.

The proposed dual decomposition algorithm for excess probability mean-risk models (2.34) and (2.37) can be summarized in the subsequent Branch-and-Bound framework.

ALGORITHM 3.1.2: Dual decomposition for two-stage stochastic mixed-integer linear semidefinite programs.

Initialize:	Let **P** be the list of current problems.
	Denote for $P \in \mathbf{P}$ by $\varphi_{\mathrm{LD}}(P)$ its Lagrangian lower bound
	(obtained by the proximal bundle method from Section 3.1).
	Put $\overline{\varphi} = +\infty$ and add the underlying problem to the list **P**.
Step 1.	If $\mathbf{P} = \emptyset$ then $\overline{X}$ with $\overline{\varphi} = \mathcal{Q}_{\mathbb{E}}(\overline{X}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(\overline{X})$ is optimal;
	else, go to Step 2.
Step 2.	Select and delete from **P** a problem $P \in \mathbf{P}$.
	Solve its Lagrangian dual with the proximal bundle method
	from Section 3.1. If $\varphi_{\mathrm{LD}}(P)$ is $+\infty$, go to Step 1;
	otherwise, go to Step 3.
Step 3.	If $\varphi_{\mathrm{LD}}(P) \geq \overline{\varphi}$, then go to Step 1.
Step 3.1	If the scenario solutions $X_\xi^{opt}, \xi = 1, ..., S$, are identical and
	$\mathcal{Q}_{\mathbb{E}}(X_\xi^{opt}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(X_\xi^{opt}) < \overline{\varphi}$, then $\overline{\varphi} := \mathcal{Q}_{\mathbb{E}}(X_\xi^{opt}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(X_\xi^{opt})$,
	store X_ξ^{opt}. Delete from **P** all problems P' with $\varphi_{\mathrm{LD}}(P') \geq \overline{\varphi}$;
	go to Step 1.
Step 3.2	If the scenario solutions differ, then run feasibility heuristic.
	Set $\widehat{X} = \sum_{\xi=1}^{S} \pi_\xi X_\xi^{opt}$. Check if $\widehat{X}$ is feasible and fulfills
	$\mathcal{Q}_{\mathbb{E}}(\widehat{X}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(\widehat{X}) < \overline{\varphi}$. If so, then $\overline{\varphi} := \mathcal{Q}_{\mathbb{E}}(\widehat{X}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(\widehat{X})$,
	store $\widehat{X}$, and delete from **P** all problems P' with $\varphi_{\mathrm{LD}}(P') \geq \overline{\varphi}$;
	go to Step 4.
Step 4.	Select the component of $\boldsymbol{X}$, that differs the most in the set
	$\left\{\pi_\xi X_\xi^{opt} : \xi = 1, .., S\right\}$, i.e., $(i,j) \in \arg\max_{(i,j)} \| \sum_{\xi=1}^{S} \pi_\xi (X_\xi^{opt})_{ij} \|$
	and add two new problems to **P** which arise from P by adding
	the constraints $\boldsymbol{X}_{ij} \leq \lfloor \widehat{X}_{ij} \rfloor$ and $\boldsymbol{X}_{ij} \geq \lfloor \widehat{X}_{ij} \rfloor + 1$ (if $\boldsymbol{X}_{ij}$ is an
	integer component), or $\boldsymbol{X}_{ij} \leq \widehat{X}_{ij} - \epsilon$ and $\boldsymbol{X}_{ij} \geq \widehat{X}_{ij} + \epsilon$ (if $\boldsymbol{X}_{ij}$
	is a continuous variable), where $\epsilon > 0$ is a tolerance parameter;
	go to Step 1.

Remark 3.2. If the feasible set of first-stage decisions $\mathcal{X}$ is bounded and all components of the variables $\boldsymbol{X} \in \mathcal{X}$ are required to be integers, then, obviously, the above algorithm will terminate in a finite number of steps. In case where some components of $\boldsymbol{X}$ are continuous variables, a stopping criterion has to be built-in that avoids endless branching on such variables.

3.2 Benders Decomposition (L-shaped Method) for Stochastic Semidefinite Programs with Continuous Recourse

Related to stochastic linear programming the Benders decomposition approach [75], which in stochastic programming is known as the L-shaped method, has its seeds in solving expectation-based two-stage models [35], [86]. A nested version of [75] can be used for tackling multistage stochastic linear [87] and quadratic programs [88], respectively. Moreover, an extension of the L-shaped method [35] to stochastic integer programs is introduced by Laporte and Louveaux in [89].

We will give a brief sketch of how the L-shaped method [35] can be adopted to two-stage stochastic semidefinite programs with continuous recourse. For this purpose, let us consider the expectation-based model (2.9), i.e.,

$$\min\big\{\mathcal{Q}_{\mathbb{E}}(x) : Ax = b,\ x \in \mathcal{S}_+^n\big\},$$

and suppose that all variables are continuous. Assume further that the underlying random variable $z(\xi)$ has a finite support. Then, by using the traditional notation in semidefinite programming this program may be written as

$$\min\left\{C \bullet \boldsymbol{X} + \boldsymbol{\theta}\ :\ \boldsymbol{\theta} \geq \sum\nolimits_{\xi=1}^{S} \pi_\xi \Phi(z_\xi - \mathcal{T}\boldsymbol{X}),\quad \boldsymbol{X} \in \mathcal{X}\right\}, \tag{3.9}$$

where

$$\Phi(z_\xi - \mathcal{T}\boldsymbol{X}) = \min\{H \bullet \boldsymbol{Y} : \mathcal{W}\boldsymbol{Y} = z_\xi - \mathcal{T}\boldsymbol{X},\ \boldsymbol{Y} \in \mathcal{S}_+^m\},$$

and $\mathcal{X} := \big\{\boldsymbol{X} \in \mathcal{S}_+^n : A_i \bullet \boldsymbol{X} = b_i,\ i = 1, ..., m\big\}$. Now, provided the constraint qualification $M_D = \{\boldsymbol{u} \in \mathbb{R}^s : W^T\boldsymbol{u} \prec H\} \neq \emptyset$ holds, we know the subgradient of the function $\sum_{\xi=1}^{S} \pi_\xi \Phi(z_\xi - \mathcal{T}\boldsymbol{X})$ in $\overline{X} \in \mathcal{S}^n$. Namely, if, for $\xi = 1, ..., S$,

$\Phi\left(z_\xi - \mathcal{T}\overline{X}\right)$ is finite, it holds[2]

$$\partial\left(\sum_{\xi=1}^{S}\pi_\xi\Phi\left(z_\xi - \mathcal{T}\overline{X}\right)\right) = -\sum_{\xi=1}^{S}\pi_\xi\mathcal{T}^T u_\xi^*, \tag{3.10}$$

where

$$u_\xi^* \in \arg\max\left\{\left(z_\xi - \mathcal{T}\overline{X}\right)^T \boldsymbol{u} : W^T\boldsymbol{u} \preceq H\right\}.$$

Note that (3.10) implies that subgradients of $\sum_{\xi=1}^{S}\pi_\xi\Phi\left(z_\xi - \mathcal{T}\boldsymbol{X}\right)$ can be computed by solving the decomposed second stage problems

$$\max\left\{\left(z_\xi - \mathcal{T}\overline{X}\right)^T \boldsymbol{u} : W^T\boldsymbol{u} \preceq H\right\}.$$

Further, by the definition of the subgradient the following estimate holds true

$$\mathcal{Q}_{\mathbb{E}}(\boldsymbol{X}) \geq \underbrace{\mathcal{Q}_{\mathbb{E}}\left(\overline{X}\right) - \left(C - \sum_{\xi=1}^{S}\pi_\xi\mathcal{T}^T u_\xi^*\right)\bullet\overline{X}}_{=:\alpha} + \left(C - \underbrace{\sum_{\xi=1}^{S}\pi_\xi\mathcal{T}^T u_\xi^*}_{=:B}\right)\bullet\boldsymbol{X}, \tag{3.11}$$

such that $\alpha + (C - B)\bullet\boldsymbol{X}$ forms a supporting hyperplane of $\mathcal{Q}_{\mathbb{E}}$ at $\overline{X}$. The latter insight can now be used to get an increasingly better approximation of the expected value function $\mathcal{Q}_{\mathbb{E}}$. Namely, by generating an adequate number of supporting hyperplanes (3.11), wisely,

$$\min\left\{C\bullet\boldsymbol{X} + \boldsymbol{\theta}\ :\ \boldsymbol{\theta} \geq \alpha_k - B_k\bullet\boldsymbol{X},\ k = 1, ..., K,\quad \boldsymbol{X}\in\mathcal{X}\right\} \tag{3.12}$$

will become a suitable approximation of (3.9).

If complete recourse $\mathcal{W}(\mathcal{S}_+^m) = \mathbb{R}^s$ holds true, then, starting with

$$X_1 \in \arg\min\left\{C\bullet\boldsymbol{X} : \boldsymbol{X}\in\mathcal{X}\right\},$$

the following update procedure

$$\begin{aligned}
\alpha_{k+1} &:= \mathcal{Q}_{\mathbb{E}}\left(\overline{X}\right) - \left(C - \sum_{\xi=1}^{S}\pi_\xi\mathcal{T}^T u_\xi^{(k+1)}\right)\bullet X_{k+1},\\
B_{k+1} &:= \sum_{\xi=1}^{S}\pi_\xi\mathcal{T}^T u_\xi^{(k+1)},
\end{aligned}$$

[2]This is due to Corollary 2.14 and Remark 2.15.

where

$$X_{k+1} \in \arg\min \{C \bullet \boldsymbol{X} + \boldsymbol{\theta} \ : \ \boldsymbol{\theta} \geq \alpha_k - B_k \bullet \boldsymbol{X}, \ k = 1, ..., K, \quad \boldsymbol{X} \in \mathcal{X}\}$$

and

$$u_\xi^{(k+1)} \in \arg\max \left\{ (z_\xi - \mathcal{T} X_{k+1})^T \boldsymbol{u} : W^T \boldsymbol{u} \preceq H \right\} \tag{3.13}$$

leads to such an approximation.

Remark 3.3. Observe that for a given iterate X_{k+1}, the value $\alpha_{k+1} - B_{k+1} \bullet X_{k+1}$ represents its costs of recourse. Since adding the cut

$$\boldsymbol{\theta} \geq \alpha_{k+1} - B_{k+1} \bullet \boldsymbol{X}$$

to program (3.12) leads to a better approximation of the cost function $\mathcal{Q}_{\mathbb{E}}$, it is referred to as an optimality cut.

If complete recourse does not hold, i.e., $\mathcal{W}(\mathcal{S}_+^m) \neq \mathbb{R}^s$, approximating (3.9) by using optimality cuts (3.11) alone will not work properly as (3.13) might contain an improving ray. In such situation, the set of feasible first-stage decisions need to be approximated as well. The latter is realized by cutting off iterates that are (second-stage) infeasible. We describe how this works in principle. To this end, let us assume that $\widehat{X} \in \mathcal{X}$ is second-stage infeasible. Consequently, there has to exist a scenario $\widehat{\xi} \in \{1, ..., S\}$, such that

$$\left\{ \boldsymbol{Y} \in \mathcal{S}_+^m : \mathcal{W}\boldsymbol{Y} = z_{\widehat{\xi}} - \mathcal{T}\widehat{X} \right\} = \emptyset.$$

By the semidefinite Farkas lemma (Lemma A.34), there has to exist

$$\widehat{u} \in \mathbb{R}^s \text{ with } \mathcal{W}^T \widehat{u} \preceq 0 \text{ and } \left(z_{\widehat{\xi}} - \mathcal{T}\widehat{X} \right)^T \widehat{u} > 0. \tag{3.14}$$

On the other hand, for any $\boldsymbol{X}$ that is feasible for (3.9), there exists $\boldsymbol{Y} \in \mathcal{S}_+^m$ with $\mathcal{W}\boldsymbol{Y} = z_{\widehat{\xi}} - \mathcal{T}\boldsymbol{X}$. Therefore, for all such points, we have

$$\left(z_{\widehat{\xi}} - \mathcal{T}\boldsymbol{X} \right)^T \widehat{u} = (\mathcal{W}\boldsymbol{Y})^T \widehat{u} = \underbrace{(\mathcal{W}^T \widehat{u})}_{\preceq 0} \bullet \underbrace{\boldsymbol{Y}}_{\succeq 0} \preceq 0.$$

Hence,

$$\underbrace{z_{\hat{\xi}}^T \widehat{u}}_{=:\gamma_\xi} - (\underbrace{\mathcal{T}^T \widehat{u}}_{=:D_\xi}) \bullet \boldsymbol{X} \leq 0 \tag{3.15}$$

induces a linear cut that is valid for all feasible points, and invalid for the (infeasible) point $\widehat{X}$. In doing so, the required point $\widehat{u} \in \mathbb{R}^s$ can be taken from the following set of maximizer

$$\arg\max \left\{ \left(z_{\hat{\xi}} - \mathcal{T}\widehat{x} \right)^T \boldsymbol{u} : \mathcal{W}^T \boldsymbol{u} \preceq 0, \ \|\boldsymbol{u}\|_\infty \leq 1 \right\}.$$

Furthermore, it is much more efficient to approximate the functions $\Phi(z_\xi - \mathcal{T}\boldsymbol{X})$ in (3.9) separately (if complete recourse does not hold). Otherwise, bounds on the costs of recourse can be made only for first-stage points that are feasible for all scenarios. To this, let us rewrite (3.9) as follows

$$\min \left\{ C \bullet \boldsymbol{X} + \sum\nolimits_{\xi=1}^{S} \pi_\xi \boldsymbol{\theta}_\xi \ : \ \boldsymbol{\theta}_\xi \geq \Phi\left(z_\xi - \mathcal{T}\boldsymbol{X}\right), \ \xi = 1, ..., S, \quad \boldsymbol{X} \in \mathcal{X} \right\}. \tag{3.16}$$

Now, let $\Phi(z_\xi - \mathcal{T}X_k)$ be finite for some iterate X_k. Then,

$$\partial\left(\Phi\left(z_\xi - \mathcal{T}X_k\right)\right) = -\mathcal{T}^T u_\xi^{(k)},$$

where

$$u_\xi^{(k)} \in \arg\max \left\{ \left(z_\xi - \mathcal{T}X_k\right)^T \boldsymbol{u} : \mathcal{W}^T \boldsymbol{u} \preceq H \right\}.$$

Further, the following holds

$$\Phi\left(z_\xi - \mathcal{T}\boldsymbol{X}\right) \geq \underbrace{\left(\Phi\left(z_\xi - \mathcal{T}X_k\right) + \left(\mathcal{T}^T u_\xi^{(k)}\right) \bullet X_k \right)}_{=:\alpha_\xi^{(k)}} - \Big(\underbrace{\mathcal{T}^T u_\xi^{(k)}}_{=:B_\xi^{(k)}}\Big) \bullet \boldsymbol{X}. \tag{3.17}$$

Thus, the supporting hyperplane

$$\alpha_\xi^{(k)} - B_\xi^{(k)} \bullet \boldsymbol{X}$$

of $\Phi\left(z_\xi - \mathcal{T}\boldsymbol{X}\right)$ at X_k can be used to approximate the function $\Phi\left(z_\xi - \mathcal{T}\boldsymbol{X}\right)$.

Now, the L-shaped method for stochastic semidefinite programs can be summarized in the following algorithmic framework.

ALGORITHM 3.2.1: The L-shaped method for two-stage stochastic linear semidefinite programs

Initialize: Accuracy parameter $\epsilon > 0$; $K := 1$; $UB := \infty$

Step 1. Solve $\min \left\{ C \bullet \boldsymbol{X} + \sum_{\xi=1}^{S} \pi_\xi \boldsymbol{\theta}_\xi \ : \ \boldsymbol{X} \in \mathcal{X}, \ \boldsymbol{\theta} \geq 0 \right\}$, save optimal solution X_1, and put $LB := C \bullet X_1$.

Step 1. For $\xi = 1, ..., S$, try to solve $\max \left\{ (z_\xi - \mathcal{T} X_K)^T \boldsymbol{u} : \mathcal{W}^T \boldsymbol{u} \preceq H \right\}$.
If for $\xi \in \{1, ..., S\}$, $\max \left\{ (z_\xi - \mathcal{T} X_K)^T \boldsymbol{u} : \mathcal{W}^T \boldsymbol{u} \preceq H \right\}$ is finite, store the cut $\boldsymbol{\theta}_\xi \geq \alpha_\xi^{(k)} - B_\xi^{(k)} \bullet \boldsymbol{X}$ as defined in (3.17).
Else, solve $\max \left\{ \left(z_{\hat{\xi}} - \mathcal{T} \widehat{X} \right)^T \boldsymbol{u} : \mathcal{W}^T \boldsymbol{u} \preceq 0, \ \|\boldsymbol{u}\|_\infty \leq 1 \right\}$ and store the cut $0 \geq \gamma_\xi^{(k)} - D_\xi^{(k)T} \bullet \boldsymbol{X}$ as defined in (3.15).

Step 2. If for all $\xi \in \{1, ..., S\}$, $\max \left\{ (z_\xi - \mathcal{T} X_K)^T \boldsymbol{u} : \mathcal{W}^T \boldsymbol{u} \preceq H \right\}$ was finite, update upper bound: $UB = C \bullet X_K + \sum_{\xi=1}^{S} \pi_\xi \left(\alpha_\xi^{(k)} - B_\xi^{(k)} \bullet X_K \right)$.

Step 3. Add all generated cuts from Step 1 to

$$
\begin{aligned}
\min \quad & C \bullet \boldsymbol{X} + \textstyle\sum_{\xi=1}^{S} \pi_\xi \boldsymbol{\theta}_\xi \\
\text{s.t.} \quad & \boldsymbol{X} \in \mathcal{X}, \\
& \boldsymbol{\theta}_\xi \geq \alpha_\xi^{(k)} - B_\xi^{(k)} \bullet \boldsymbol{X}, \quad \xi \in F(k), \quad k = 1, ..., K, \\
& 0 \geq \gamma_\xi^{(k)} - D_\xi^{(k)} \bullet \boldsymbol{X}, \quad \xi \in F^C(k), \quad k = 1, ..., K.
\end{aligned}
$$

Here, $F(k)$ denotes the set of all $\xi \in \{1, ..., S\}$ for which $\max \left\{ (z_\xi - \mathcal{T} X_k)^T \boldsymbol{u} : \mathcal{W}^T \boldsymbol{u} \preceq H \right\}$ is finite, whereas $F^C(k)$ denotes its complement. Solve this program, set LB to its optimal value, and X_{K+1} to its optimal solution.

Step 4. Check termination criterion. If $\frac{|UB - LB|}{|LB|} < \epsilon$ terminate.
Else, $K := K + 1$ and go to Step 1.

Remark 3.4. If the set of feasible first-stage decisions is finite discrete, then Algorithm 3.2.1 terminates in a finite number of steps for any $\epsilon > 0$ and even for $\epsilon = 0$ (see [90] and [91]).

Chapter 4

Unit Commitment

We consider an AC power system that interconnects various power production units (such as coal fired blocks, gas turbines, pumped-storage units, and wind parks) to consumers. For some preassigned planning horizon, the challenge is to provide "optimal service" to the consumers in economically efficient, technologically feasible, and operationally reliable manner.

From the mathematical optimization perspective, these three targets concern main branches of current research. Economic aspects, usually addressed under the key words of power dispatch and unit commitment, lead into large-scale mixed-integer (linear) optimization (see [92] for instance). While here linearity often provides an acceptable compromise for model precision, this no longer holds true for the technological aspects capturing generation and transmission of electricity subject to the physical laws and engineering constraints. Jointly, these features are addressed as optimal power flow. As an additional difficulty, one faces the non-linearity inevitably arising in its non-convex fashion. Finally, the reliability issue, in the widest sense, leads into optimization under uncertainty with robust and stochastic optimization as major lines of development.

Given the breadth of topics with seminal contributions dating back for 50 years and more, e.g., the first model for optimal power flow due to [93], there is a vast literature on the above themes. Therefore, we here confine ourselves to refer to the recent very useful primer [94] and the excellent bibliographical review in [95] and [96]. Although all three papers mainly circle around different aspects of optimal power flow, coverage of the economic aspects and the uncertainty issue is substantial as well.

4.1 Basic Traditional Unit Commitment Model

To begin with, we introduce principal characteristics of the unit commitment part of our full model. Drawing on [97] and [98], the presentation is fairly detailed, mainly to be self-contained, but also to introduce the quite complex notation needed subsequently.

Throughout, boldfaced symbols in mathematical formulas represent variables, symbols in normal font represent problem data.

Consider a power grid with the set of buses $\mathcal{N} := \{1, ..., n\}$, the set of generators $\mathcal{G} \subseteq V$, and the set of flow lines $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$. Assume that for $(l, m) \in \mathcal{L}$, we also have $(m, l) \in \mathcal{L}$. The set of all generator buses $\mathcal{G}$ decomposes into coal fired blocks, attached gas turbines, and installed pumped-storage hydroelectricity (PSH) units, denoted by $i = 1, ..., I$, $r = 1, ..., R$, and $h = 1, ..., H$, respectively. Wind power is modeled by positive in-feed at wind farm buses, such that these units are not considered as controllable production devices. We will optimize over a time horizon which is discretized into finitely many hourly planning intervals $t = 1, ..., T$. The Boolean decision variables

$$\mathbf{u}_i^t \in \{0, 1\}, \quad i = 1, ..., I, \quad t = 1, ..., T, \tag{4.1}$$

then indicate whether the coal fired block i is off- or on-line during time interval t. Analogously, there are the variables $\mathbf{u}_r^t \in \{0, 1\}, r = 1, ..., R; t = 1, ..., T$ for the gas turbines as well as the non-negative continuous variables

$$\begin{array}{lll} \mathbf{p}_i^t, \mathbf{q}_i^t, & i = 1, ..., I, & t = 1, ..., T, \\ \mathbf{p}_r^t, \mathbf{q}_r^t, & r = 1, ..., R, & t = 1, ..., T, \\ \mathbf{p}_h^t, \mathbf{q}_h^t, \quad \mathbf{w}_h^t, \overline{\mathbf{w}}_h^t, & h = 1, ..., H, & t = 1, ..., T, \end{array}$$

representing the output levels, in both active and reactive power, for the coal fired thermal units, the gas turbines, the pumped-storage units in generation and in pumping modes.

For each of the coal fired units and gas turbines we assume a quadratic cost function with given non-negative coefficients accounting for the fuel cost in terms of active power generation, i.e. the fuel costs and thus the objective to be minimized is given by

$$\sum\nolimits_{t=1}^{T} \left(\sum\nolimits_{k \in I \cup R} f_k(\mathbf{p}_k^t, \mathbf{u}_k^t, \mathbf{r}_k^t) \right), \tag{4.2}$$

with

$$f_k(\mathbf{p}_k^t, \mathbf{u}_k^t, \mathbf{r}_k^t) = \mathbf{u}_k^t \left(c_{k2} \left(\mathbf{p}_k^t\right)^2 + c_{k1}\mathbf{p}_k^t \right) + c_{k0}\mathbf{r}_k^t, \tag{4.3}$$

where

$$\mathbf{r}_k^t = \max\{\mathbf{u}_k^t - \mathbf{u}_k^{t-1}, 0\}. \tag{4.4}$$

Further, we denote by $P_i^{\min}$, $P_i^{\max}$, $Q_i^{\min}$, $Q_i^{\max}$, $P_r^{\max}$, $Q_r^{\max}$, $P_h^{\max}$, $Q_h^{\max}$, $W_h^{\max}$, $\overline{W}_h^{\max}$, the minimal and maximal outputs of the particular power production units. All outputs have to be within these bounds, where the natural lower bound for pumped-storage units (in generation and pumping mode) is zero. This yields for all $t = 1, ..., T$:

$$\begin{aligned} P_k^{\min} \cdot \mathbf{u}_k^t \leq \mathbf{p}_k^t \leq P_k^{\max} \cdot \mathbf{u}_k^t, \qquad \forall k \in I \cup R, \\ Q_k^{\min} \cdot \mathbf{u}_k^t \leq \mathbf{q}_k^t \leq Q_k^{\max} \cdot \mathbf{u}_k^t, \qquad \forall k \in I \cup R, \end{aligned} \tag{4.5}$$

and

$$\begin{aligned} -W_h^{\min} \leq \mathbf{p}_h^t - \mathbf{w}_h^t \leq P_h^{\max}, \qquad \forall h \in H, \\ -\overline{W}_h^{\min} \leq \mathbf{q}_h^t - \overline{\mathbf{w}}_h^t \leq Q_h^{\max}, \qquad \forall h \in H. \end{aligned} \tag{4.6}$$

Remark 4.1. Since the $\mathbf{p}_k^t$ variables are constrained to be zero when $\mathbf{u}_k^t = 0$, the multiplication $\mathbf{u}_k^t \cdot \left(c_{k2} \left(\mathbf{p}_k^t\right)^2 + c_{k1}\mathbf{p}_k^t \right)$ is unnecessary, and the objective (4.3) becomes convex:

$$f_k(\mathbf{p}_k^t, \mathbf{u}_k^t, \mathbf{r}_k^t) = c_{k2} \left(\mathbf{p}_k^t\right)^2 + c_{k1}\mathbf{p}_k^t + c_{k0}\mathbf{r}_k^t.$$

Beside the production bounds (4.5), the coal fired blocks must adhere to minimum downtimes to avoid excessive thermal strains. These are expressed by the following inequalities:

$$\begin{aligned} \mathbf{u}_i^{t-1} - \mathbf{u}_i^t \leq 1 - \mathbf{u}_i^l, \qquad & i = 1, ..., I, \quad t = 2, ..., T-1, \\ & l = t+1, ..., \min\{t + \tau_i - 1, T\}, \end{aligned} \tag{4.7}$$

where τ_i represents the required downtime.

Furthermore, there are variables $\mathbf{l}_h^t$, $h = 1, ..., H; t = 1, ..., T$, specifying the fill (in active power) of the upper dam at pumped-storage unit h at the end of time interval t. At all times, the (non-negative) fill levels must not exceed the maximum fills $l_h^{\max}$, and, together with generation and pumping, the fill has to meet the following balances:

$$\begin{aligned} &\mathbf{l}_h^0 = l_h^{\text{in}}, \quad \mathbf{l}_h^T = l_h^{\text{end}}, \\ &\mathbf{l}_h^t = \mathbf{l}_h^{t-1} - \left(\mathbf{p}_h^t - \eta_h \mathbf{w}_h^t\right) \leq l_h^{\max}, \quad h = 1, ..., H,\ t = 1, ..., T, \end{aligned} \tag{4.8}$$

where l_h^{in}, l_h^{end} are the initial and final fills, respectively, and $0 \leq \eta_h < 1$ indicates the pumping efficiency.

4.2 AC Load Flow Extension

Turning attention to AC load flow, for every network bus $k \in \mathcal{N}$, we consider its apparent power[1]

$$\mathbf{s}_k^t = \mathbf{p}_k^t + j\mathbf{q}_k^t$$

at time $t = 1, ..., T$, where $\mathbf{p}_k^t$ denotes its active and $\mathbf{q}_k^t$ its reactive power, respectively. The apparent power is subject to Kirchhoff's first law, i.e. at any node in an electrical network, the sum of currents flowing into that node is equal to the sum of currents flowing out of it:

$$\mathbf{p}_k^t = \sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{p}_{kl}^t, \qquad \forall k \in \mathcal{N}, \quad t = 1, ..., T \tag{4.9}$$

$$\mathbf{q}_k^t = \sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{q}_{kl}^t, \qquad \forall k \in \mathcal{N}, \quad t = 1, ..., T \tag{4.10}$$

where $\mathbf{p}_{kl}^t$ and $\mathbf{q}_{kl}^t$ are the active and reactive power, respectively, transferred from k to the rest of the network through line $(k, l) \in \mathcal{L}$, and $\mathcal{N}(k)$ denotes the set of all buses directly connected to k. The apparent power $\mathbf{s}_k^t$ can also be written as the difference between complex in-feed $\mathbf{s}_{G_k}^t := \mathbf{p}_{G_k}^t + j\mathbf{q}_{G_k}^t$ and complex load $\mathbf{s}_{D_k}^t := \mathbf{p}_{D_k}^t + j\mathbf{q}_{D_k}^t$, such that together with (4.9) and (4.10) we arrive at the

[1] Here, j denotes the imaginary unit. This is to avoid confusion with the unit for electrical current.

following power balance equations:

$$\mathbf{p}_{G_k}^t - \sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{p}_{kl}^t = p_{D_k}^t, \qquad \forall k \in \mathcal{G}, \qquad t = 1, ..., T, \tag{4.11}$$

$$\mathbf{q}_{G_k}^t - \sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{q}_{kl}^t = q_{D_k}^t, \qquad \forall k \in \mathcal{G}, \qquad t = 1, ..., T, \tag{4.12}$$

$$- \sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{p}_{kl}^t = p_{D_k}^t, \qquad \forall k \in \mathcal{N} \backslash \mathcal{G}, \quad t = 1, ..., T, \tag{4.13}$$

$$- \sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{q}_{kl}^t = q_{D_k}^t, \qquad \forall k \in \mathcal{N} \backslash \mathcal{G}, \quad t = 1, ..., T, \tag{4.14}$$

where, the active and reactive electrical load $\{(p_D^t, q_D^t) : t = 1, ..., T\}$ in terms of demand and in-feed of renewables is given in advance and has to be covered (exactly).

To represent the energy flows, one possibility, for others see [94], is to select rectangular coordinates for voltage, $\mathbf{V}_k^t = \mathbf{E}_k^t + j\mathbf{F}_k^t \in \mathbb{C}$ at every bus $k \in \mathcal{N}$. Then, there needs to be at least one slack bus with specified voltage magnitude

$$|\mathbf{V}_k^t| = \sqrt{(\mathbf{E}_k^t)^2 + (\mathbf{F}_k^t)^2}$$

and angle $\boldsymbol{\theta}_k^t = \arg(\mathbf{V}_k^t)$. It is used to balance apparent power, in such a way that it compensates system losses by emitting and absorbing active power and reactive power to and from the system, respectively. In selecting the slack bus, it is important to ensure that a powerful bus[2] is chosen, which can absorb all uncertainties arising from the system. Here, we pick bus $1 \in \mathcal{N}$ as slack bus and claim $\theta_1^t = 0$.

Furthermore, in order to represent the energy flows along the lines, it takes the introduction of some parameters: For lines $(l, m) \in \mathcal{L}$, the complex parameter $y_{lm} := g_{lm} + jb_{lm}$ is referred to as the admittance between the nodes l and m. This definition is extended to all $l \neq m$ by putting y_{lm} equal to zero, whenever bus l and m are not directly linked. The parameter y_{kk} denotes the admittance-to-ground at $k \in \mathcal{N}$, it is defined as the sum over all connected line admittances and shunt admittances b_{lm}^0. With these agreements, the admittance matrix

$$Y = G + jB \in \mathbb{C}^{n \times n},$$

[2]Generally, a load bus or the most powerful generator bus is chosen as slack bus.

defined by y_{ll} for diagonal elements and $-y_{lm}$ otherwise, represents the core of the underlying AC grid as it reflects all of its characteristics. Note that for $(l,m) \in \mathcal{L}$ the given conductances $g_{lm} \in \mathbb{R}_+$, susceptances $b_{lm} \in \mathbb{R}_-$, and shunts $b^0_{lm} \in \mathbb{R}_+$ specify the line transmission capabilities. In doing so, the existing transformers are implicitly taken into account, since due to their existence, transmission capabilities (conductances, susceptances, and shunts) will be improved, such that the corresponding parameters can be readjusted. Moreover, observe that Y is sparse with just $n + 2|\mathcal{L}|$ nonzero entries.

Without going into detail, using the above notations, a fairly accurate approximation of the steady-state behavior of the energy flows along the lines $(l,m) \in \mathcal{L}$ can be modeled by the following quadratic expressions (cf. [99] and [94]):

$$\mathbf{p}^t_{lm} = g_{lm}\left((\mathbf{E}^t_l)^2 + (\mathbf{F}^t_l)^2\right) - g_{lm}\left(\mathbf{E}^t_l\mathbf{E}^t_m + \mathbf{F}^t_l\mathbf{F}^t_m\right) - b_{lm}\left(\mathbf{F}^t_l\mathbf{E}^t_m - \mathbf{E}^t_l\mathbf{F}^t_m\right), \quad (4.15)$$
$$\mathbf{q}^t_{lm} = b_{lm}\left(\mathbf{E}^t_l\mathbf{E}^t_m + \mathbf{F}^t_l\mathbf{F}^t_m\right) - g_{lm}\left(\mathbf{F}^t_l\mathbf{E}^t_m - \mathbf{E}^t_l\mathbf{F}^t_m\right) - \bar{b}_{lm}\left((\mathbf{E}^t_l)^2 + (\mathbf{F}^t_l)^2\right), \quad (4.16)$$

where $\bar{b}_{lm} := b_{lm} + b^0_{lm}$.

For the grid, we claim that voltage magnitudes $|\mathbf{V}^t_k|$ have to be within particular bounds

$$V^{\min}_k \leq \sqrt{(\mathbf{E}^t_k)^2 + (\mathbf{F}^t_k)^2} \leq V^{\max}_k, \qquad \forall k \in \mathcal{N}, \quad (4.17)$$

where we have $V^{\min}_1 = V^{\max}_1$ at the slack bus, and that lines $(l,m) \in \mathcal{L}$ may not be overstrained, i.e., power flow is limited by the maximum transmission capacities $S^{\max}_{lm}, P^{\max}_{lm}, \Delta V^{\max}_{lm} \in \mathbb{R}_+$:

$$(\mathbf{p}^t_{lm})^2 + (\mathbf{q}^t_{lm})^2 \leq (S^{\max}_{lm})^2, \qquad \forall (l,m) \in \mathcal{L}, \quad (4.18)$$
$$(\mathbf{p}^t_{lm})^2 \leq (P^{\max}_{lm})^2, \qquad \forall (l,m) \in \mathcal{L}, \quad (4.19)$$
$$\sqrt{(\mathbf{E}^t_l - \mathbf{E}^t_m)^2 + (\mathbf{F}^t_l - \mathbf{F}^t_m)^2} \leq \Delta V^{\max}_{lm}, \qquad \forall (l,m) \in \mathcal{L}. \quad (4.20)$$

It may happen that some of the constraints (4.17)-(4.20) are not needed in certain modeling situations. Then, the vacuous constraints can be removed by setting their upper/lower bounds to $\pm\infty$. Moreover, it is noted that (4.19) is equivalent to $\mathbf{p}^t_{lm} \leq P^{\max}_{lm}$ which is due to the fact that $\mathbf{p}^t_{lm} + \mathbf{p}^t_{ml} \geq 0$ and $P^{\max}_{lm} = P^{\max}_{ml}$.

Hence, by taking into account AC load flow (in steady-state), we arrive at the following unit commitment model:

$$\begin{aligned} \min \quad & \sum\nolimits_{t=1}^{T} \left(\sum\nolimits_{k \in I \cup R} \mathbf{u}_k^t \left(c_{k2} \left(\mathbf{p}_k^t \right)^2 + c_{k1} \mathbf{p}_k^t \right) + c_{k0} \mathbf{r}_k^t \right) \\ \text{s.t.} \quad & (4.4), (4.5), (4.6), (4.7), (4.8), \text{ and } (4.11)\text{-}(4.20). \end{aligned} \tag{4.21}$$

The inclusion of AC power flow leads us to a (mixed-integer) nonlinear program. Whenever these constraints enter into an optimization problem, its feasible set becomes non-convex and the problem itself NP-hard [3], [100].

Chapter 5

Semidefinite Programming for AC Unit Commitment

The non-convex AC power flow constraints (4.11)-(4.20) have been intensively studied in the literature and a multitude of algorithms have been proposed for solving optimization problems, taking into account these nonlinear restrictions [101], [102]. Most of these solution methods are based on solving the corresponding Karush-Kuhn-Tucker (KKT) conditions and thus at best guarantee local optimality.

Rather than to work with the equations (4.15) and (4.16) directly, these are relaxed and approximated, respectively. The DC[1] (direct current) Power Flow model [94], for instance, assumes that the difference of voltage angles is zero, that all voltage magnitudes are equal to one, and that the reactive power may be neglected (these assumptions may be justified under normal system operating). Doing so, AC load flow becomes linear, making this approach extremely popular in practice (see for instance [103]).

The DC power flow model being lossless, including these losses at least approximately will improve the model. In [104], [105], and [106], the DC model is refined by inclusion of Ohmic losses. These are modeled by trigonometric equations becoming relaxed to inequalities for computations. The relaxation is such that it overestimates losses and leads to convexity of the constraint set. Numerical optimization procedures heading for the minimization of losses then have the

[1] The DC power flow approximation of AC power flow is named so because the resulting equations equal DC power flow.

tendency to drive the overestimation back to zero, thus fulfilling the inequality as an equation.

In recent years, several convex relaxations were proposed, which are tight under certain conditions and thus provide a significantly better approximation of AC power flow than the DC approach and its extensions. These include Second Order Cone (SOC) [107], SDP [3], Convex-DistFlow (CDF) [108], and Quadratic Convex (QC) [109] relaxations. A comprehensive comparison of these relaxations is presented by Coffrin, Hijazi, and Hentenryck in [110]. We give a brief summary of these relaxations in Appendix C.

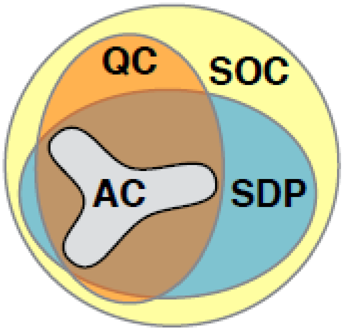


Figure 5.1: A Venn diagram of the solution sets for various AC power flow relaxations (cf. [110]).

A wide class of AC power flow models is presented in [3], where the convexification via semidefinite programming relaxations may lead to globally optimal solutions. This approach works for all IEEE benchmark systems (cf. [111]), provided a small resistance (10^{-5} per unit) is added to every transformer that originally is assumed to have zero resistance. Further, it has been confirmed in [110] that the SDP relaxation is the tightest relaxation among the mentioned relaxations. However, it is noted that the SDP relaxation does not work for all power grids – its limitations are examined in [112] as well as in [113]. Moreover, we refer to very recent publications on sufficient conditions on when the SDP approximation of optimal power flow (OPF) finally enables solution of the original problem and when not – we refer for details to the doctoral thesis [114] and survey paper [115].

To solve the introduced deterministic unit commitment problem (4.21), we suggest a combination of the semidefinite programming approach with a traditional Benders decomposition. Tackling these programs by a form of a Benders algorithm can also be found in a recent work by Amjady and Ansari [116]. The basic idea is to separate the restrictions to the generators from the nonlinear conditions to the power grid, such that the latter can be tackled by the mentioned semidefinite approach.

This chapter is organized as follows: Semidefinite programming relaxations for AC unit commitment (4.21) are introduced in Section 5.1. A strategy for finding globally optima of AC optimal power flow problems is specified in Section 5.2. Finally, Section 5.3.1 presents our proposed Benders decomposition approach for tackling AC unit commitment (4.21).

5.1 Semidefinite Relaxations

A popular approach to approximate challenging non-convex quadratically constrained quadratic programs (QCQPs) is based on semidefinite programming (cf. [117]). This relaxation technique particularly has drawn attention to the optimization community due to a seminal work for the maximum cut problem by Goemans and Williamson [118]. The basic principle of this approach is the introduction of the matrix variable

$$X := xx^T$$

for given $x \in \mathbb{R}^n$. Doing so, the matrix $X \in \mathbb{R}^{n\times n}$ is symmetric and positive semidefinite, where, for $1 \leq i \leq j \leq n$, its entry X_{ij} depicts the product $x_i x_j$. Thus, if we consider the quadratic term $x^T Ax$, it can be transformed into

$$x^T Ax = \mathrm{tr}(x^T Ax) = \mathrm{tr}(Ax^T x) = A \bullet X,$$

so that what previously was quadratic in x now becomes linear in X. Applying this to any program involving quadratic functions, it may be rewritten as one involving linear functions plus the non-convex constraint $X = xx^T$. Hence, if we further introduce the augmented matrix

$$\widehat{X} := \begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}^T = \begin{bmatrix} 1 & x^T \\ x & X \end{bmatrix},$$

any QCQP can be reformulated as a rank constrained linear semidefinite program:

$$\text{(QCQP)} \quad \text{"} \Longleftrightarrow \text{"} \quad \min\{C \bullet \widehat{X} : \mathcal{A}\widehat{X} = b,\ \widehat{X} \succeq 0,\ \operatorname{rank}(\widehat{X}) = 1\}. \tag{5.1}$$

The programs on the right-hand side in (5.1) are just as hard to solve as the initial QCQPs, however, the fundamental complexity in solving them now only lies in the non-convex rank constraint $\operatorname{rank}(\widehat{X}) = 1$. If this constraint is dropped, these programs turn into (convex) linear semidefinite programs, and the latter can be solved numerically reliable and efficient (in polynomial time) by using interior-point algorithms[2] [120].

After having solved the semidefinite relaxation of (5.1), the essential issue is how (if possible) a feasible point in (5.1) can be generated that has (nearly) the same objective value as the relaxed program. If this is possible, the initial program is solved to (nearly) global optimality. Note that, generally, this is not achievable, since, otherwise, we would have solved a NP-hard problem in polynomial time.

5.1.1 Semidefinite Relaxations for AC Unit Commitment

In this section, the mixed-integer QCQP (4.21) is relaxed to a mixed-integer semidefinite program[3]. This relaxation is obtained by the procedure outlined above: First, (4.21) is transformed into a rank constraint mixed-integer semidefinite program, and then relaxation of the non-convex rank constraints will yield a mixed-integer semidefinite programming approximation of (4.21). For this purpose, it takes the introduction of the voltage vector

$$\mathbf{x}_V^t = \begin{bmatrix} \Re(\mathbf{V}^t) & \Im(\mathbf{V}^t) \end{bmatrix}^T = \begin{bmatrix} \mathbf{E}^t & \mathbf{F}^t \end{bmatrix}^T \in \mathbb{R}^{2n},$$

for each time interval $t \in T$. The outer (or tensor) product

$$\mathbf{X}_V^t := \mathbf{x}_V^t (\mathbf{x}_V^t)^T \in \mathcal{S}^{2n}$$

[2]In this thesis, we resort to the solver SeDuMi by Jos F. Sturm [119]. This solver is able to exploit sparsity and therefore is suitable for solving large-scale optimization problems. Moreover, SeDuMi tends to produce low rank solutions.

[3]For optimal power flow problems this approach was first presented in [7].

then contains, for $l, m = 1, .., n$, all of the (voltage) products $\mathbf{E}_l^t\mathbf{E}_m^t$, $\mathbf{E}_l^t\mathbf{F}_m^t$, $\mathbf{F}_l^t\mathbf{F}_m^t$. Moreover, if we define, for $i, j = 1, ..., n$, the matrices[4] $A_{ij} := \frac{1}{2}e_ie_j^T + \frac{1}{2}e_je_i^T \in \mathcal{S}^{2n}$, where $e_1, ..., e_{2n}$ are the standard basis vectors in $\mathbb{R}^{2n}$, then, due to

$$A_{ij} \bullet \mathbf{X}_V^t = \begin{cases} \mathbf{E}_i^t\mathbf{E}_j^t, \text{ for } 1 \le i, j \le n, \\ \mathbf{E}_i^t\mathbf{F}_j^t, \text{ for } 1 \le i \le n,\ n+1 \le j \le 2n \text{ or } n+1 \le i \le 2n,\ 1 \le j \le n, \\ \mathbf{F}_i^t\mathbf{F}_j^t, \text{ for } n+1 \le i, j \le 2n, \end{cases}$$

we know how to select each element in $\mathbf{X}_V^t$. Thus, the active and reactive powers on transmission lines, respectively, i.e., (4.15) and (4.16), can be rewritten as:

$$\mathbf{p}_{lm}^t = \underbrace{\left[g_{lm}(A_{ll} + A_{\widetilde{l}\widetilde{l}} - A_{lm} - A_{\widetilde{l}\widetilde{m}}) - b_{lm}(A_{\widetilde{l}m} - A_{l\widetilde{m}})\right]}_{=:Y_{lm}} \bullet \mathbf{X}_V^t, \tag{5.2}$$

$$\mathbf{q}_{lm}^t = \underbrace{\left[b_{lm}(A_{lm} + A_{\widetilde{l}\widetilde{m}}) - g_{lm}(A_{\widetilde{l}m} - A_{l\widetilde{m}}) - \bar{b}_{lm}(A_{ll} + A_{\widetilde{l}\widetilde{l}})\right]}_{=:\overline{Y}_{lm}} \bullet \mathbf{X}_V^t, \tag{5.3}$$

where $\widetilde{l} = l + n$ and $\widetilde{m} = m + n$. This implies

$$\sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{p}_{kl}^t = \sum\nolimits_{l \in \mathcal{N}(k)} Y_{kl} \bullet \mathbf{X}_V^t = \left(\sum\nolimits_{l \in \mathcal{N}(k)} Y_{kl}\right) \bullet \mathbf{X}_V^t,$$

and

$$\sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{q}_{kl}^t = \sum\nolimits_{l \in \mathcal{N}(k)} \overline{Y}_{kl} \bullet \mathbf{X}_V^t = \left(\sum\nolimits_{l \in \mathcal{N}(k)} \overline{Y}_{kl}\right) \bullet \mathbf{X}_V^t.$$

Hence, defining $Y_k := \left(\sum_{l \in \mathcal{N}(k)} Y_{kl}\right)$ and $\overline{Y}_k := \left(\sum_{l \in \mathcal{N}(k)} \overline{Y}_{kl}\right)$, the power balance equations (4.11)-(4.14) may be equivalently stated as

$$\mathbf{p}_{G_k}^t - p_{D_k}^t = Y_k \bullet \mathbf{X}_V^t, \quad \forall k \in \mathcal{G}, \qquad t = 1, ..., T, \tag{5.4}$$

$$\mathbf{q}_{G_k}^t - q_{D_k}^t = \overline{Y}_k \bullet \mathbf{X}_V^t, \quad \forall k \in \mathcal{G}, \qquad t = 1, ..., T, \tag{5.5}$$

$$-p_{D_k}^t = Y_k \bullet \mathbf{X}_V^t, \quad \forall k \in \mathcal{N} \backslash \mathcal{G}, \qquad t = 1, ..., T, \tag{5.6}$$

$$-q_{D_k}^t = \overline{Y}_k \bullet \mathbf{X}_V^t, \quad \forall k \in \mathcal{N} \backslash \mathcal{G}, \qquad t = 1, ..., T. \tag{5.7}$$

[4] For $i = j$, we have $A_{ij} = A_{ii} = I_{2n}$, and, for $i \neq j$, all entries of the matrix A_{ij} are zero except for the elements $a_{ij} = a_{ji} = \frac{1}{2}$.

In exactly the same way, the voltage magnitude bounds turn from (4.17) to

$$(V_k^{\min})^2 \leq M_k \bullet \mathbf{X}_V^t \leq (V_k^{\max})^2, \qquad \forall k \in \mathcal{N}, \quad t = 1, ..., T, \tag{5.8}$$

whereas the network line limitations (4.18)-(4.20) become

$$\begin{aligned}
(Y_{lm} \bullet \mathbf{X}_V^t)^2 + (\overline{Y}_{lm} \bullet \mathbf{X}_V^t)^2 \leq (S_{lm}^{\max})^2, \qquad &\forall (l,m) \in \mathcal{L}, \quad t = 1, ..., T, &(5.9)\\
Y_{lm} \bullet \mathbf{X}_V^t \leq P_{lm}^{\max}, \qquad &\forall (l,m) \in \mathcal{L}, \quad t = 1, ..., T, &(5.10)\\
M_{lm} \bullet \mathbf{X}_V^t \leq (\Delta V_{lm}^{\max})^2, \qquad &\forall (l,m) \in \mathcal{L}, \quad t = 1, ..., T. &(5.11)
\end{aligned}$$

Here, the matrices M_k and M_{lm} were defined by $M_k := A_{kk} + A_{\widetilde{k}\widetilde{k}}$ and $M_{lm} := \left[A_{ll} + A_{mm} - 2A_{lm} + A_{\widetilde{l}\widetilde{l}} + A_{\widetilde{m}\widetilde{m}} - 2A_{\widetilde{l}\widetilde{m}}\right]$, respectively.

Furthermore, inserting $\mathbf{p}_{G_k}^t = Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t$ and $\mathbf{q}_{G_k}^t = \overline{Y}_k \bullet \mathbf{X}_V^t + q_{D_k}^t$ into the power production bounds (4.5) yield

$$\begin{aligned}
P_k^{\min} \cdot \mathbf{u}_k^t \leq Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t \leq P_k^{\max} \cdot \mathbf{u}_k^t,\\
Q_k^{\min} \cdot \mathbf{u}_k^t \leq \overline{Y}_k \bullet \mathbf{X}_V^t + q_{D_k}^t \leq Q_k^{\max} \cdot \mathbf{u}_k^t.
\end{aligned} \tag{5.12}$$

Since by (5.12), $\mathbf{p}_{G_k}^t = Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t$ is constrained to be zero when $\mathbf{u}_k^t = 0$, the objective (4.2) may be written as

$$\sum\nolimits_{t=1}^{T} \left(\sum\nolimits_{k \in I \cup R} c_{k2} \left(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t\right)^2 + c_{k1} \left(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t\right) + c_{k0} \mathbf{r}_k^t\right). \tag{5.13}$$

For network buses with connected pumped-storage units, the active power in-feed/output[5] and the reactive power in-feed/output may be expressed by $\mathbf{p}_h^t - \mathbf{w}_h^t = Y_h \bullet \mathbf{X}_V^t + p_{D_h}^t$ and $\mathbf{q}_h^t - \overline{\mathbf{w}}_h^t = \overline{Y}_h \bullet \mathbf{X}_V^t + q_{D_h}^t$, respectively. Therefore, the bounds in (4.6) can be equivalently described by:

$$\begin{aligned}
-W_h^{\max} \leq Y_h \bullet \mathbf{X}_V^t + p_{D_h}^t \leq P_h^{\max},\\
-\overline{W}_h^{\max} \leq \overline{Y}_h \bullet \mathbf{X}_V^t + q_{D_h}^t \leq Q_h^{\max}.
\end{aligned} \tag{5.14}$$

[5]Depending on whether the pumped-storage unit operates in power production or pumping mode.

Altogether, AC unit commitment (4.21) is equivalent to the subsequent mixed-integer rank constraint semidefinite program:

$$
\begin{aligned}
\min \quad & \textstyle\sum_{t=1}^{T} \left(\sum_{k\in\mathcal{G}\backslash H} c_{k2}(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t)^2 + c_{k1}(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t) + c_{k0} \cdot \mathbf{r}_k^t \right) \\
\text{s.t.} \quad & \left.\begin{aligned} \mathbf{u}_k^t P_k^{\min} \leq\ & Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t \leq \mathbf{u}_k^t P_k^{\max}, \quad \forall k \in \mathcal{G}\backslash H, \\ \mathbf{u}_k^t Q_k^{\min} \leq\ & \overline{Y}_k \bullet \mathbf{X}_V^t + q_{D_k}^t \leq \mathbf{u}_k^t Q_k^{\max}, \quad \forall k \in \mathcal{G}\backslash H, \end{aligned}\right\} \begin{array}{l}\text{output bounds}\\ \text{and load coverage}\\ \text{at non-renewables}\end{array} \\
& \left.\begin{aligned} -W_h^{\max} \leq\ & Y_h \bullet \mathbf{X}_V^t + p_{D_h}^t \leq P_h^{\max}, \quad \forall h \in H, \\ -\overline{W}_h^{\max} \leq\ & \overline{Y}_h \bullet \mathbf{X}_V^t + q_{D_h}^t \leq Q_h^{\max}, \quad \forall h \in H, \end{aligned}\right\} \begin{array}{l}\text{output bounds and}\\ \text{satisfaction of load at}\\ \text{pumped-storage plants}\end{array} \\
& \left.\begin{aligned} Y_n \bullet \mathbf{X}_V^t + p_{D_n}^t = 0, \quad \forall n \in \mathcal{N}\backslash\mathcal{G}, \\ \overline{Y}_n \bullet \mathbf{X}_V^t + q_{D_n}^t = 0, \quad \forall n \in \mathcal{N}\backslash\mathcal{G}, \end{aligned}\right\} \begin{array}{l}\text{load coverage at wind farms}\\ \text{and non-generator buses}\end{array} \\
& (V_k^{\min})^2 \leq M_k \bullet \mathbf{X}_V^t \leq (V_k^{\max})^2, \quad \forall k \in \mathcal{N}, \Big\} \ \text{voltage magnitude bounds} \\
& \left.\begin{aligned} (Y_{lm} \bullet \mathbf{X}_V^t)^2 + (\overline{Y}_{lm} \bullet \mathbf{X}_V^t)^2 &\leq (S_{lm}^{\max})^2, && \forall (l,m) \in \mathcal{L}, \\ Y_{lm} \bullet \mathbf{X}_V^t &\leq P_{lm}^{\max}, && \forall (l,m) \in \mathcal{L}, \\ M_{lm} \bullet \mathbf{X}_V^t &\leq (\Delta V_{lm}^{\max})^2, && \forall (l,m) \in \mathcal{L}, \end{aligned}\right\} \begin{array}{l}\text{line}\\ \text{limitations}\end{array} \\
& \left.\begin{aligned} \mathbf{u}_i^{t-1} - \mathbf{u}_i^t \leq 1 - \mathbf{u}_i^l, \quad & \forall i \in I, \quad t = 2, ..., T-1, \\ & l = t+1, ..., \min\{t + \tau_i - 1, T\}, \end{aligned}\right\} \begin{array}{l}\text{min. down-}\\ \text{times for}\\ \text{thermal units}\end{array} \\
& \left.\begin{aligned} & \mathbf{l}_h^0 = l_h^{\text{in}}, \quad \mathbf{l}_h^T = l_h^{\text{end}}, && \forall h \in H, \\ & \mathbf{l}_h^t = \mathbf{l}_h^{t-1} - (\mathbf{p}_h^t - \eta_j \mathbf{w}_h^t) \leq l_h^{\max}, && \forall h \in H, \\ & \mathbf{p}_h^t - \mathbf{w}_h^t = Y_h \bullet \mathbf{X}_V^t + p_{D_h}^t, && \forall h \in H, \\ & \mathbf{p}_h^t \geq 0, \quad \mathbf{w}_h^t \geq 0, , && \forall h \in H, \end{aligned}\right\} \begin{array}{l}\text{holding of max. dam fills}\\ \text{plus considering of inter-}\\ \text{connections in pumped-}\\ \text{storage units}\end{array} \\
& \mathbf{r}_k^t = \max\{\mathbf{u}_k^t - \mathbf{u}_k^{t-1}, 0\}, \quad \mathbf{u}_k^t \in \{0,1\}, \quad \forall k \in \mathcal{G}\backslash H, \\
& \operatorname{rank}(\mathbf{X}_V^t) = 1, \qquad \forall t \in \{1, ..., T\}.
\end{aligned} \tag{5.15}
$$

Note that all functions in (5.15) are linear in $\mathbf{X}_V^t$ except the objective (5.13) and the apparent power limitations (5.9), which are quadratic in $\mathbf{X}_V^t$. The latter functions, however, can be linearized by Schur's complement (see Theorem A.4): given a symmetric matrix

$$X := \begin{bmatrix} A & B \\ B^T & C \end{bmatrix},$$

with $C \succ 0$, then the Schur complement $S := A - BC^{-1}B^T$ of block C is positive semidefinite if and only if X is positive semidefinite.

Schur's complement involves that

$$(S_{lm}^{\max})^2 - (Y_{lm} \bullet \mathbf{X}_V^t)^2 - (\overline{Y}_{lm} \bullet \mathbf{X}_V^t)^2 \geq 0$$

is equivalent to

$$\begin{bmatrix} (S_{lm}^{\max})^2 & Y_{lm} \bullet \mathbf{X}_V^t & \overline{Y}_{lm} \bullet \mathbf{X}_V^t \\ Y_{lm} \bullet \mathbf{X}_V^t & 1 & 0 \\ \overline{Y}_{lm} \bullet \mathbf{X}_V^t & 0 & 1 \end{bmatrix} \succeq 0. \tag{5.16}$$

Accordingly, for $(l, m) \in \mathcal{L}$, (5.16) and thus (5.9) holds true if and only if there exists a positive semidefinite matrix $\mathbf{Z}_{lm}^t \in \mathcal{S}_+^3$ such that the following holds:

$$\begin{aligned}
&\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \bullet \mathbf{Z}_{lm} = (S_{lm}^{\max})^2, &&\begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \bullet \mathbf{Z}_{lm} = Y_{lm} \bullet \mathbf{X}_V^t, \\
&\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \bullet \mathbf{Z}_{lm} = 1, &&\begin{bmatrix} 0 & 0 & 1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 0 \end{bmatrix} \bullet \mathbf{Z}_{lm} = \overline{Y}_{lm} \bullet \mathbf{X}_V^t, \\
&\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \bullet \mathbf{Z}_{lm} = 0, &&\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \bullet \mathbf{Z}_{lm} = 1.
\end{aligned} \tag{5.17}$$

The quadratic objective function (5.13) can be linearized in the same way. Indeed, by introducing for $k \in \mathcal{G} \backslash H$ and $t \in \{1, ..., T\}$ auxiliary variables $\mathbf{a}_k^t$, (5.13) may be expressed by $\sum_{t=1}^{T} \sum_{k \in \mathcal{G} \backslash H} (\mathbf{a}_k^t + c_{k0} \cdot \mathbf{r}_k^t)$ provided the following constraints are complied with:

$$\mathbf{a}_k^t \geq c_{k2}(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t)^2 + c_{k1}(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t), \quad \forall k \in \mathcal{G}\backslash H. \tag{5.18}$$

The latter conditions now can be linearized via Schur's complement. To this end, introduce for $k \in \mathcal{G}\backslash H$ and $t \in \{1, ..., T\}$ positive semidefinite matrices $\mathbf{A}_k^t \in \mathcal{S}_+^2$, then, (5.18) is equivalent to

$$\begin{aligned} &\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \bullet \mathbf{A}_k + c_{k1} Y_k \bullet \mathbf{X}_V^t - \mathbf{a}_k = -c_{k1} p_{D_k}^t, \qquad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \bullet \mathbf{A}_k = 1, \\ &\begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix} \bullet \mathbf{A}_k - \sqrt{c_{k2}} Y_k \bullet \mathbf{X}_V^t = \sqrt{c_{k2}} p_{D_k}^t. \end{aligned} \tag{5.19}$$

Finally, the SDP relaxation of (5.15), i.e., the program that results from (5.15) by relaxing the rank constraints $\mathrm{rank}(\mathbf{X}_V^t) = 1$, may be equivalently expressed by the subsequent mixed-integer linear semidefinite program.

$$\begin{aligned} \min \quad & \textstyle\sum_{t=1}^T \sum_{k \in \mathcal{G}\backslash H} (\mathbf{a}_k^t + c_{k0} \cdot \mathbf{r}_k^t) \\ s.t. \quad & (4.7), (4.8), (5.6), (5.7), (5.8), (5.10), \\ & (5.11), (5.12), (5.14), (5.17), (5.19), \\ & \mathbf{r}_k^t = \max\{\mathbf{u}_k^{t-1} - \mathbf{u}_k^t, 0\}, \quad \forall k \in \mathcal{G}\backslash H, \quad t = 1, ..., T, \\ & \mathbf{u}_k^t \in \{0, 1\}, \quad \mathbf{A}_k^t \in \mathcal{S}_+^2, \quad \forall k \in \mathcal{G}\backslash H, \quad t = 1, ..., T, \\ & \mathbf{Z}_{lm}^t \in \mathcal{S}_+^3, \qquad \forall (l, m) \in \mathcal{L}, \quad t = 1, ..., T, \\ & \mathbf{X}_V^t \in \mathcal{S}_+^{2n}, \qquad t = 1, ..., T. \end{aligned} \tag{5.20}$$

The relaxation of the rank constraints permits overload any time at any network bus (cf. [121]). This may be beneficial in stressed network situations, and therefore could result in an infeasible commitment/dispatch decision. Still, it has been shown to be tight for tree networks [122], [123], [124] and for cyclic networks if every cycle contains a line with a controllable phase shifter [121]. Doing so, it is demonstrated in [122] that the rank of the SDP solution is upper bounded by the tree-width of the underlying network[6], which in practice is expected to be small.

It can be observed that all data matrices introduced above, i.e., $Y_k, \overline{Y}_k, M_k$, for $k \in \mathcal{N}$ and $Y_{lm}, \overline{Y}_{lm}, M_{lm}$ for $(l,m) \in \mathcal{L}$ are of following principal shape[7]:

$$\begin{bmatrix} A & B \\ -B & A \end{bmatrix}, \tag{5.21}$$

where $A, B \in \mathcal{S}^N$. Indeed, according to (5.2),

$$\begin{aligned} Y_{lm} = \quad & \frac{g_{lm}}{2}\left[\left(2e_le_l^T - e_le_m^T - e_me_l^T\right) + \left(2e_{n+l}e_{n+l}^T - e_{n+l}e_{n+m}^T - e_{n+m}e_{n+l}^T\right)\right] \\ - \; & \frac{b_{lm}}{2}\left[\left(e_{n+l}e_m^T - e_{n+m}e_l^T\right) + \left(e_me_{n+l}^T - e_le_{n+m}^T\right)\right] \\ = \quad & \frac{g_{lm}}{2}\begin{bmatrix} 2e_le_l^T - e_le_m^T - e_me_l^T & 0_{n\times n} \\ 0_{n\times n} & 2e_le_l^T - e_le_m^T - e_me_l^T \end{bmatrix} \\ - \; & \frac{b_{lm}}{2}\begin{bmatrix} 0_{n\times n} & e_me_l^T - e_le_m^T \\ -\left(e_me_l^T - e_le_m^T\right) & 0_{n\times n} \end{bmatrix} \\ = \quad & \begin{bmatrix} \widetilde{A}_{lm} & \widetilde{B}_{lm} \\ -\widetilde{B}_{lm} & \widetilde{A}_{lm} \end{bmatrix}, \end{aligned}$$

where $\widetilde{A}_{lm} = \frac{g_{lm}}{2}\left(2e_le_l^T - e_le_m^T - e_me_l^T\right)$ and $\widetilde{B}_{lm} = -\frac{b_{lm}}{2}\left(e_me_l^T - e_le_m^T\right)$. Analogously, due to (5.3), we have

$$\overline{Y}_{lm} = \begin{bmatrix} \overline{A}_{lm} & \overline{B}_{lm} \\ -\overline{B}_{lm} & \overline{A}_{lm} \end{bmatrix},$$

where $\overline{A}_{lm} = \frac{b_{lm}}{2}\left(e_le_m^T + e_me_l^T\right) - \bar{b}_{lm}e_le_l^T$ and $\overline{B}_{lm} = -\frac{g_{lm}}{2}\left(e_me_l^T - e_le_m^T\right)$. Moreover, as the matrices $Y_k, \overline{Y}_k, k \in \mathcal{N}$ are formed by a sum of specific Y_{lm} and specific $\overline{Y}_{lm}$, respectively, they inherit the structure (5.21) from them. Furthermore, by definition, we have

$$M_k = \begin{bmatrix} e_ke_k^T & 0_{n\times n} \\ 0_{n\times n} & e_ke_k^T \end{bmatrix}, \quad \text{for } k \in \mathcal{N},$$

[6]Note that this implies that the SDP relaxation is tight for tree networks.
[7]This property has been observed in [3].

and

$$M_{lm} = \begin{bmatrix} M'_{lm} & 0_{n\times n} \\ 0_{n\times n} & M'_{lm} \end{bmatrix}, \quad \text{for } (l,m) \in \mathcal{L},$$

where $M'_{lm} := e_l e_l^T + e_m e_m^T - e_l e_m^T - e_m e_l^T$. Hence, all data matrices in (5.15) are of the form (5.21).

5.2 A Strategy for Solving AC Optimal Power Flow Problems

If we apply the reformulation described in the previous section to (4.21) and further fix all generator switching decisions as well as the active and reactive power outputs/consumptions, respectively, at all pumped-storage plants, then, coupling over time disappears, and the model decomposes into T independent problems which are closely related to the continuous OPF problem. For fixed $t \in \{1, ..., T\}$, the resulting decoupled program is given by:

$$\begin{aligned}
\min \quad & \textstyle\sum_{k\in\mathcal{G}\setminus H} \mathbf{a}_k^t \\
\text{s.t.} \quad & P_k^{\min} \le Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t \le P_k^{\max}, \quad \forall k \in \mathcal{N}, \\
& Q_k^{\min} \le \overline{Y}_k \bullet \mathbf{X}_V^t + q_{D_k}^t \le Q_k^{\max}, \quad \forall k \in \mathcal{N}, \\
& (V_k^{\min})^2 \le M_k \bullet \mathbf{X}_V^t \le (V_k^{\max})^2, \quad \forall k \in \mathcal{N}, \\
& \begin{bmatrix} \mathbf{a}_k^t - c_{k1}(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t) & \sqrt{c_{k2}}(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t) \\ \sqrt{c_{k2}}(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t) & 1 \end{bmatrix} \succeq 0, \ \forall k \in \mathcal{G}\setminus H, \\
& \begin{bmatrix} (S_{lm}^{\max})^2 & Y_{lm} \bullet \mathbf{X}_V^t & \overline{Y}_{lm} \bullet \mathbf{X}_V^t \\ Y_{lm} \bullet \mathbf{X}_V^t & 1 & 0 \\ \overline{Y}_{lm} \bullet \mathbf{X}_V^t & 0 & 1 \end{bmatrix} \succeq 0, \quad \forall (l,m) \in \mathcal{L}, \\
& Y_{lm} \bullet \mathbf{X}_V^t \le P_{lm}^{\max}, \quad \forall (l,m) \in \mathcal{L}, \\
& M_{lm} \bullet \mathbf{X}_V^t \le (\Delta V_{lm}^{\max})^2, \quad \forall (l,m) \in \mathcal{L}, \\
& \operatorname{rank}(\mathbf{X}_V^t) = 1, \quad \mathbf{X}_V^t \succeq 0.
\end{aligned} \tag{5.22}$$

Here, the definition of $P^{\min}$, $P^{\max}$, $Q^{\min}$, and $Q^{\max}$ has to be extended to all network buses by setting $P_k^{\min} = P_k^{\max} = Q_k^{\min} = Q_k^{\max} = 0$ for $k \in \mathcal{N}\backslash\mathcal{G}$ as well as for those $k \in \mathcal{G}\backslash H$ for which $\mathbf{u}_k^t$ is zero. Further, note that if the switching decisions are determined, so are the arising start-up costs. Therefore, the latter costs can be disregarded as they will simply appear as a constant. The dual to the decomposed OPF (5.22) then is given by (cf. [3]):

$$\max_{\boldsymbol{\lambda}^t \geq 0} \left\{ \begin{aligned} h(\boldsymbol{\lambda}^t, \boldsymbol{r}^t) : \quad & A(\boldsymbol{\lambda}^t, \boldsymbol{r}^t) \succeq 0, \\ & \begin{bmatrix} 1 & \boldsymbol{r}_{k,1}^t \\ \boldsymbol{r}_{k,1}^t & \boldsymbol{r}_{k,2}^t \end{bmatrix} \succeq 0, \qquad \forall k \in \mathcal{G}\backslash H, \\ & \begin{bmatrix} \boldsymbol{r}_{lm,1}^t & \boldsymbol{r}_{lm,2}^t & \boldsymbol{r}_{lm,3}^t \\ \boldsymbol{r}_{lm,2}^t & \boldsymbol{r}_{lm,4}^t & \boldsymbol{r}_{lm,5}^t \\ \boldsymbol{r}_{lm,3}^t & \boldsymbol{r}_{lm,5}^t & \boldsymbol{r}_{lm,6}^t \end{bmatrix} \succeq 0, \quad \forall (l,m) \in L \end{aligned} \right\}, \tag{5.23}$$

with

$$\begin{aligned} A(\boldsymbol{\lambda}^t, \boldsymbol{r}^t) := \quad & \sum\nolimits_{k \in N} \left(\overline{\boldsymbol{\lambda}}_k^t - \underline{\boldsymbol{\lambda}}_k^t\right) Y_k + \left(\overline{\boldsymbol{\gamma}}_k^t - \underline{\boldsymbol{\gamma}}_k^t\right) \overline{Y}_k + \left(\overline{\boldsymbol{\mu}}_k^t - \underline{\boldsymbol{\mu}}_k^t\right) M_k \\ & + \sum\nolimits_{k \in \mathcal{G}\backslash H} (c_{k1} + 2\sqrt{c_{k2}} r_{k,1}^t) Y_k \\ & + \sum\nolimits_{(l,m) \in L} (\boldsymbol{\lambda}_{lm}^t + 2\boldsymbol{r}_{lm,2}^t) Y_{lm} + 2\boldsymbol{r}_{lm,3}^t \overline{Y}_{lm} + \boldsymbol{\mu}_{lm}^t M_{lm}, \end{aligned} \tag{5.24}$$

$$\begin{aligned} h(\boldsymbol{\lambda}^t, \boldsymbol{r}^t) := \quad & \sum\nolimits_{k \in N} \underline{\boldsymbol{\lambda}}_k^t P_k^{\min} - \overline{\boldsymbol{\lambda}}_k^t P_k^{\max} + \left(\overline{\boldsymbol{\lambda}}_k^t - \underline{\boldsymbol{\lambda}}_k^t\right) p_{D_k}^t \\ & + \sum\nolimits_{k \in N} \underline{\boldsymbol{\gamma}}_k^t Q_k^{\min} - \overline{\boldsymbol{\gamma}}_k^t Q_k^{\max} + \left(\overline{\boldsymbol{\gamma}}_k^t - \underline{\boldsymbol{\gamma}}_k^t\right) q_{D_k}^t \\ & + \sum\nolimits_{k \in N} \underline{\boldsymbol{\mu}}_k^t \left(V_k^{\min}\right)^2 - \overline{\boldsymbol{\mu}}_k^t \left(V_k^{\max}\right)^2 \\ & - \sum\nolimits_{(l,m) \in L} \boldsymbol{\lambda}_{lm}^t P_{lm}^{\max} + \boldsymbol{\mu}_{lm}^t \left(\Delta V_{lm}^{\max}\right)^2 + \boldsymbol{r}_{lm,1}^t \left(S_{lm}^{\max}\right)^2 + \boldsymbol{r}_{lm,4}^t + \boldsymbol{r}_{lm,6}^t, \end{aligned}$$

where

$$\boldsymbol{\lambda}^t = (\overline{\boldsymbol{\lambda}}_k^t, \underline{\boldsymbol{\lambda}}_k^t, \overline{\boldsymbol{\gamma}}_k^t, \underline{\boldsymbol{\gamma}}_k^t, \overline{\boldsymbol{\mu}}_k^t, \underline{\boldsymbol{\mu}}_k^t, \boldsymbol{\lambda}_{lm}^t, \boldsymbol{\mu}_{lm}^t),$$

and

$$\boldsymbol{r}^t = (\boldsymbol{r}_{k,1}^t, \boldsymbol{r}_{k,2}^t, \boldsymbol{r}_{lm,1}^t, \boldsymbol{r}_{lm,2}^t, \boldsymbol{r}_{lm,3}^t, \boldsymbol{r}_{lm,4}^t, \boldsymbol{r}_{lm,5}^t, \boldsymbol{r}_{lm,6}^t).$$

Now, provided (5.23) is solvable, let us denote by $(\lambda^t_{\text{opt}}, r^t_{\text{opt}})$ any optimal solution to this program. Assume further that X^t_{opt} is primal optimal (i.e., it solves the dual to (5.23)) and that Slater's condition is satisfied[8]. Then, due to strong duality, $A(\lambda^t_{\text{opt}}, r^t_{\text{opt}}) \bullet X^t_{opt} = 0$. This equation is valid if and only if the product of the symmetric and positive semidefinite matrices $A(\lambda^t_{\text{opt}}, r^t_{\text{opt}})$ and X^t_{opt} vanishes. Hence, writing the symmetric matrix X^t_{opt} by using its eigenvalue decomposition $P^t\Lambda^t P^{tT} = \sum_{i=1}^{2N} \lambda^t_i p^t_i p^{tT}_i$, the following equations have to hold true:

$$A(\lambda^t_{\text{opt}}, r^t_{\text{opt}})p^t_i = 0, \quad \text{for those } i \in \{1, ..., 2N\} \text{ for which } \lambda_i \neq 0. \tag{5.25}$$

This means that all of the concerned orthogonal eigenvectors (eigenvectors to nonzero eigenvalues of X^t_{opt}) must belong to the kernel of $A(\lambda^t_{\text{opt}}, r^t_{\text{opt}})$. If now the latter were of dimension one, the primal would have a rank-one solution. Hence, there were a zero duality gap between the OPF and its SDP relaxation.

The same result is obtained when the kernel of $A(\lambda^t_{\text{opt}}, r^t_{\text{opt}})$ has dimension less than or equal to 2 (cf. [3]). Indeed, in view of Section 5.1.1, the matrix $A(\lambda^t_{\text{opt}}, r^t_{\text{opt}})$ as a weighted sum of the matrices $Y_k, \overline{Y}_k, M_k, Y_{lm}, \overline{Y}_{lm}, M_{lm}$ has the following block structure:

$$A(\lambda^t_{\text{opt}}, r^t_{\text{opt}}) = \begin{bmatrix} \overline{A}(\lambda^t_{\text{opt}}, r^t_{\text{opt}}) & B(\lambda^t_{\text{opt}}, r^t_{\text{opt}}) \\ -B(\lambda^t_{\text{opt}}, r^t_{\text{opt}}) & \overline{A}(\lambda^t_{\text{opt}}, r^t_{\text{opt}}) \end{bmatrix}.$$

This implies: if the kernel of $A(\lambda^t_{\text{opt}}, r^t_{\text{opt}})$ includes

$$p^t = \begin{bmatrix} p^{t\,T}_1 & p^{t\,T}_2 \end{bmatrix}^T,$$

then it also includes

$$\begin{bmatrix} -p^{t\,T}_2 & p^{t\,T}_1 \end{bmatrix}^T.$$

As these two vectors are orthogonal, they must be the two eigenvectors to the zero eigenvalue of $A(\lambda^t_{\text{opt}}, r^t_{\text{opt}})$. Therefore, for primal optimal X^t_{opt}, the following holds

$$X^t_{opt} = \lambda^t_1 \begin{bmatrix} p^t_1 \\ p^t_2 \end{bmatrix} \begin{bmatrix} p^{t\,T}_1 & p^{t\,T}_2 \end{bmatrix} + \lambda^t_2 \begin{bmatrix} -p^t_2 \\ p^t_1 \end{bmatrix} \begin{bmatrix} -p^{t\,T}_2 & p^{t\,T}_1 \end{bmatrix}.$$

[8] It is shown in [3] that Slater's condition is always satisfied for (5.23).

Further, due to the fact that the trace of a skew-symmetric and symmetric matrix is equal to zero, it could be observed that

$$\begin{aligned}\begin{bmatrix} A & B \\ -B & A \end{bmatrix} \bullet \begin{bmatrix} p_1^t \\ p_2^t \end{bmatrix} \begin{bmatrix} {p_1^t}^T & {p_2^t}^T \end{bmatrix} &= A \bullet p_1^t {p_1^t}^T + A \bullet p_2^t {p_2^t}^T \\ &= \begin{bmatrix} A & B \\ -B & A \end{bmatrix} \bullet \begin{bmatrix} -p_2^t \\ p_1^t \end{bmatrix} \begin{bmatrix} -{p_2^t}^T & {p_1^t}^T \end{bmatrix}. \end{aligned} \tag{5.26}$$

Hence, the rank-one matrix

$$\overline{X}_V^t = (\lambda_1^t + \lambda_2^t) \begin{bmatrix} p_1^t \\ p_2^t \end{bmatrix} \begin{bmatrix} {p_1^t}^T & {p_2^t}^T \end{bmatrix}$$

is globally optimal for the original OPF (it satisfies all of its constraints and produces the same objective value as X_{opt}^t). Summing up, this leads to the following corollary (cf. [3]).

Corollary 5.1. *Assume that $(\lambda_{opt}^t, r_{opt}^t)$ is an optimal solution to (5.23) and that*

$$dim\left(ker\left(A(\lambda_{opt}^t, r_{opt}^t)\right)\right) \leq 2. \tag{5.27}$$

Then for any nonzero vector p^t in the null space of $A(\lambda_{opt}^t, r_{opt}^t)$, there exist two real-valued scalars λ_1^t and λ_2^t such that

$$\overline{X}_V^t = (\lambda_1^t + \lambda_2^t) p^t {p^t}^T$$

is a global optimum of the corresponding OPF problem.

Remark 5.2. The scalars λ_1^t and λ_2^t in Corollary 5.1 may be obtained by considering two linear equations. The voltage angle at our slack bus being zero introduces one such equation whereas the second one can be formed by identifying the active voltage constraints.

Remark 5.3. The verification whether for a given symmetric matrix a high/low rank condition is satisfied or not causes some numerical issues concerning the decision when eigenvalues are regarded as zero-eigenvalues. This especially is made difficult when the matrix is large-scale and has almost identical entries.

In summary, the following strategy for finding a global optimum of the underlying OPF problem can be applied (see [3]):

ALGORITHM 5.2.1: Algorithmic framework for solving OPF problems.

Step 1.	Compute a solution $(\lambda^t_{\text{opt}}, r^t_{\text{opt}})$ of (5.23).
Step 2.	If $h(\lambda^t_{\text{opt}}, r^t_{\text{opt}})$ is $+\infty$, terminate – the OPF is infeasible. Else, go to Step 3.
Step 3.	Find the multiplicity ψ of the zero eigenvalue of $A(\lambda^t_{\text{opt}}, r^t_{\text{opt}})$.
Step 4.	If $\psi > 2$, the solution of (5.23) depicts a lower bound for (5.22).
Step 5.	If $\psi \leq 2$, then a (globally) optimal solution to the OPF (5.22) can be constructed via Corollary 5.1 and Remark 5.2.

Note that, beside the nice feature to convexify NP-hard OPF problems, applying the above SDP approach leads to an enormous inflation of problem size since it squares the number of voltage variables. Indeed, when considering large network instances with a huge number of buses, this approach yields an SDP with an enormous number of variables[9]. Nevertheless, according to [3], for all IEEE benchmark systems, the SDP approach works very well – meaning that these problems could be solved within a few seconds. For larger network instances tree decomposition techniques have been proposed (see [125] and [126]) in order to break down the large-scale semidefinite constraint into small-sized constraints.

Furthermore, having solved the dual (5.23), and it turned out that $\psi > 2$, estimates for the present duality gap, i.e., the gap between the OPF (5.22) and its dual (5.23), can only be made if feasible points to (5.22) are available. Penalization methods [122], [127], a good many times, provide feasible points for (5.22) that are nearly (globally) optimal. The idea is to add some penalty term to the SDP relaxation that pushes the solution to have low-rank (at best, this will result in a rank-one solution). For instance, it is shown in [128] that, under certain conditions minimizing the nuclear norm[10] leads to guaranteed minimum-rank solutions. Suitable penalty terms for SDP relaxations of OPF problems are presented

[9]Note that, as to date SDP solver are less mature than other nonlinear solvers, such that this drawback poses a serious issue.

[10]The nuclear norm of a matrix A is defined as $\|A\|_* = \operatorname{tr}\left(\sqrt{A^* A}\right) = \sum_{i=1}^{r} \sigma_i(A)$, where $\sigma_1(A), ..., \sigma_r(A)$ are the singular values of A.

by Madani, Sojoudi, and Lavaei in [127]. There, it is recommended to penalize reactive power consumption and line losses, respectively. Moreover, an adaptive penalization algorithm is proposed. The latter (iteratively) identifies problematic lines by a (graph-theoretic) convex program and incorporates the loss over those into the objective as a regularization term.

5.3 A Decomposition Algorithm for Solving the AC Unit Commitment SDP Relaxation

As the SDP relaxation of (4.21) results in a very large-scale semidefinite program with discrete variables on top, current SDP solvers[11] are not really capable for solving it. Therefore, rather than tackle (5.20) in an all-at-once manner, we propose to solve it by using Benders decomposition techniques. In doing so, (5.20) will decompose into significantly smaller (linear) SDP subproblems which then might be handled efficiently by current SDP solvers. This decomposition approach is introduced in the following section.

5.3.1 Semidefinite Programming Based Benders Decomposition for Solving AC Unit Commitment Problems

Adopting semidefinite relaxation techniques to tackle AC unit commitment problems (4.21) leads us to mixed-integer linear semidefinite programs having a special block structure (the transformation into the required SDP format is outlined in Section 5.1.1). Each block corresponds to a specific time interval $t \in \{1, ..., T\}$ at which those are only coupled by (4.7) and (4.8). Instead of considering all constraints as well as all decision variables simultaneously, the splitting of them via generalized Benders decomposition (see [129]) will enable partitioning into tractable smaller programs. Here, we will separate (4.7) and (4.8) from flow conservation and net limitations, and the Boolean switching variables and pumped-storage operations from the dispatch decisions. In doing so, our proposed algorithm which employs a cutting-plane approach, will result in a mixed-integer linear

[11] Scalability remains an open question for SDP solvers, and, moreover, SDP solvers supporting discrete variables are very recent and by far have not the scientific maturity of MIP solvers.

programming (MILP) master problem and two sets of continuous linear SDP subproblems.

As in [116], the first step of our proposed Benders decomposition algorithm treats the following MILP master problem:

$$\begin{aligned} \mu_M := \min \quad & \sum_{t=1}^{T} \left(\sum_{k \in \mathcal{G}\setminus H} c_{k0} \cdot \mathbf{r}_k^t + \boldsymbol{\eta}_{Obj}^t \right) \\ \text{s.t.} \quad & (4.6), (4.7), \text{ and } (4.8), \end{aligned} \tag{5.28}$$

where $\mathcal{G}\setminus H$ is the set of all thermal generators, and $\boldsymbol{\eta}_{Obj}^t$ are additional nonnegative variables introduced for the objective cuts measuring the power production costs for a feasible binary on/off assignment of the on-line thermal units. Note that (5.28) takes into account all constraints regarding unit commitment switching decisions as well as active power generation at pumped-storage units. The model relaxes the nonlinear conditions to the grid.

After having solved the master problem (5.28), its solution forms the input to a first set of subproblems. These subproblems emerge from both fixing the solution to (5.28) in (5.15) and relaxing the rank constraints $\text{rank}(\mathbf{X}_V^t) = 1$. This implies decoupling of time intervals and decomposition into the following $t = 1, \ldots, T$ SDP subproblems:

$$\begin{aligned} \mu_{Obj}^t := \min_{\mathbf{X}_V^t \in \mathfrak{N}} & \sum_{k \in \mathcal{G}\setminus H} c_{k2}(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t)^2 + c_{k1}(Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t) \\ \text{s.t.} \quad & \mathbf{u}_k^t \cdot P_k^{\min} \le \quad Y_k \bullet \mathbf{X}_V^t + p_{D_k}^t \quad \le \mathbf{u}_k^t \cdot P_k^{\max}, \quad \forall k \in \mathcal{G}\setminus H, \\ & \mathbf{u}_k^t \cdot Q_k^{\min} \le \quad \overline{Y}_k \bullet \mathbf{X}_V^t + q_{D_k}^t \quad \le \mathbf{u}_k^t \cdot Q_k^{\max}, \quad \forall k \in \mathcal{G}\setminus H, \\ & \qquad Y_h \bullet \mathbf{X}_V^t + p_{D_h}^t \quad = \mathbf{p}_h^t - \mathbf{w}_h^t, \quad \forall h \in H, \\ & -W_h^{\max} \le \quad \mathbf{p}_h^t - \mathbf{w}_h^t \quad \le P_h^{\max}, \quad \forall h \in H, \\ & -\overline{W}_h^{\max} \le \quad \overline{\mathbf{Y}}_h \bullet \mathbf{X}_V^t + q_{D_h}^t \quad \le Q_h^{\max}, \quad \forall h \in H, \\ & \qquad Y_n \bullet \mathbf{X}_V^t + p_{D_n}^t \quad = 0, \quad \forall n \in N\setminus\mathcal{G}, \\ & \qquad \overline{Y}_n \bullet \mathbf{X}_V^t + q_{D_n}^t \quad = 0, \quad \forall n \in N\setminus\mathcal{G}, \\ & \mathbf{u}_k^t = \widetilde{\mathrm{u}}_k^t, \ \forall k \in \mathcal{G}\setminus H, \quad \mathbf{p}_h^t - \mathbf{w}_h^t = \widetilde{\mathrm{p}}_{D_h}^t, \ \mathbf{p}_h^t, \mathbf{w}_h^t \ge 0, \ \forall h \in H, \end{aligned} \tag{5.29}$$

where, for reasons of brevity, we have denoted by $\mathfrak{N}$ the set of those symmetric positive semidefinite matrices $\mathbf{X}_V^t \in \mathcal{S}_+^{2n}$ fulfilling (5.8)-(5.11). This set describes the physical limits of the underlying grid with respect to voltage magnitude bounds (5.8) as well as line limitations (5.9)-(5.11) provided one claims in addition that $\mathbf{X}_V^t \in \mathfrak{N}$ has rank-one. Moreover, $\widetilde{\mathrm{u}}_k^t$ and $\widetilde{\mathrm{p}}_{D_h}^t := \widetilde{\mathrm{p}}_h^t - \widetilde{\mathrm{w}}_h^t$ denote the optimal solution to (5.28) delivering switching decisions for the installed thermal units and power output/consumption at pumped-storage units, respectively (here, the nonlinear objective once again can be linearized as described in (5.19)). If for $t \in \{1, ..., T\}$ its corresponding first subproblem becomes feasible, the following objective cut is added to (5.28):

$$\boldsymbol{\eta}_{Obj}^t \geq \mu_{Obj}^t + \sum\nolimits_{k \in \mathcal{G}} \lambda_{Obj,k}^t (\mathbf{u}_k^t - \widetilde{\mathrm{u}}_k^t) + \sum\nolimits_{h \in H} \lambda_{Obj,h}^t (\mathbf{p}_h^t - \mathbf{w}_h^t - \widetilde{\mathrm{p}}_{D_h}^t), \quad (5.30)$$

where $\lambda_{Obj,k}^t$ and $\lambda_{Obj,h}^t$ are the optimal dual variables with respect to the inserted constraints $\mathbf{u}_k^t = \widetilde{\mathrm{u}}_k^t$, $\forall k \in \mathcal{G} \backslash H$ and $Y_h \bullet \mathbf{X}_V^t + d_{D_h}^t = \widetilde{\mathrm{p}}_{D_h}^t$, $\forall h \in H$, respectively.

If otherwise, for $t \in T$ its associated first subproblem is infeasible, its infeasibility in terms of active power bounds at generator buses, voltage restrictions at net nodes, and network line limitations is measured by an appropriate second subproblem. To this end, the non-negative auxiliary variables $\boldsymbol{z}_k^t, \boldsymbol{v}_n^t, \boldsymbol{p}_{lm}^t, \boldsymbol{m}_{lm}^t$, as well as $\boldsymbol{s}_{lm}^t$ are introduced, to reflect the violation of active power production bounds in (4.5) as well as the failure of the network limitations in (4.17), (4.18), (4.19), and (4.20) by means of the inequalities:

- **violation of active power production bounds:**

$$\begin{aligned}
\boldsymbol{z}_k^t &\geq \mathbf{u}_k^t P_k^{\min} - (Y_k \bullet \mathbf{X}_V^t + d_{D_k}^t), && \forall k \in \mathcal{G} \backslash H, && (5.31)\\
\boldsymbol{z}_k^t &\geq (Y_k \bullet \mathbf{X}_V^t + d_{D_k}^t) - \mathbf{u}_k^t P_k^{\max}, && \forall k \in \mathcal{G} \backslash H, && (5.32)
\end{aligned}$$

- **failure of network limitations:**

$$\begin{aligned}
\boldsymbol{v}_n^t &\geq (V_n^{\min})^2 - M_k \bullet \mathbf{X}_V^t, && \forall n \in \mathcal{N}, && (5.33)\\
\boldsymbol{v}_n^t &\geq M_k \bullet \mathbf{X}_V^t - (V_n^{\max})^2, && \forall n \in \mathcal{N}, && (5.34)\\
\boldsymbol{p}_{lm}^t &\geq Y_{lm} \bullet \mathbf{X}_V^t - P_{lm}^{\max}, && \forall (l,m) \in \mathcal{L}, && (5.35)\\
\boldsymbol{m}_{lm}^t &\geq M_{lm} \bullet \mathbf{X}_V^t - (\Delta V_{lm}^{\max})^2, && \forall (l,m) \in \mathcal{L}, && (5.36)\\
\boldsymbol{s}_{lm}^t &\geq \big(Y_{lm} \bullet \mathbf{X}_V^t\big)^2 + \big(\overline{Y}_{lm} \bullet \mathbf{X}_V^t\big)^2 - (S_{lm}^{\max})^2, && \forall (l,m) \in \mathcal{L}. && (5.37)
\end{aligned}$$

This leads to the following set of second subproblems[12]:

$$\begin{aligned}
\mu^t_{Feas} := \min\ & \sum\nolimits_{i\in I} \boldsymbol{z}^t_k + \sum\nolimits_{n\in N} \boldsymbol{v}^t_n + \sum\nolimits_{(l,m)\in L} (\boldsymbol{p}^t_{lm} + \boldsymbol{m}^t_{lm} + \boldsymbol{s}^t_{lm}) \\
\text{s.t. }\ & (5.31), (5.32), (5.33), (5.34), (5.35), (5.36), \text{ and } (5.37), \\
& \mathbf{u}^t_k \cdot Q^{\min}_k \le \ \overline{Y}_k \bullet \mathbf{X}^t_V + q^t_{D_k} \ \le \mathbf{u}^t_k \cdot Q^{\max}_k, \forall k \in \mathcal{G}\backslash H, \\
& -W^{\max}_h \le \ \overline{Y}_h \bullet \mathbf{X}^t_V + q^t_{D_h} \ \le Q^{\max}_h, \qquad \forall h \in H, \\
& Y_n \bullet \mathbf{X}^t_V + p^t_{D_n} \ = 0, \quad \forall n \in N\backslash(\mathcal{G}\cup H), \\
& \overline{Y}_n \bullet \mathbf{X}^t_V + q^t_{D_n} \ = 0, \quad \forall n \in N\backslash(\mathcal{G}\cup H), \\
& \mathbf{u}^t_k = \widetilde{\mathrm{u}}^t_k, \forall k \in \mathcal{G}, \quad Y_h \bullet \mathbf{X}^t_V + p^t_{D_h} = \widetilde{\mathrm{p}}^t_{D_h}, \forall h \in H, \\
& \boldsymbol{z}^t_k,\ \boldsymbol{v}^t_n,\ \boldsymbol{p}^t_{lm},\ \boldsymbol{m}^t_{lm},\ \boldsymbol{s}^t_{lm} \ge 0, \quad \mathbf{X}^t_V \succeq 0.
\end{aligned} \tag{5.38}$$

After solving this linear SDP, the subsequent feasibility cut[13] is added to (5.28):

$$0 \ge \mu^t_{Feas} + \sum\nolimits_{k\in\mathcal{G}} \lambda^t_{Feas,k}(\mathbf{u}^t_k - \widetilde{\mathrm{u}}^t_k) + \sum\nolimits_{h\in H} \lambda^t_{Feas,h}(\mathbf{p}^t_h - \mathbf{w}^t_h - \widetilde{\mathrm{p}}^t_h). \tag{5.39}$$

Now, our suggested algorithm can be summarized by the following steps:

ALGORITHM 5.3.1: Algorithmic framework for solving AC unit commitment.

Initialize:	Accuracy parameter $\epsilon > 0$; set $\varphi_{UB} := \infty$, and $\varphi_{LB} := 0$; solve (5.28) with $\boldsymbol{\eta}^t_{Obj} = 0$ and obtain $\widetilde{\mathrm{u}}$ plus $\widetilde{\mathrm{p}}_H$ from its solution.
Step 1.	Update lower bound $\varphi_{LB} := \mu_M - \sum_{t=1}^T \widetilde{\eta}^t_{Obj}$; solve (for $t = 1, ..., T$) the first set of subproblems (5.29); if subproblem t becomes feasible, keep μ^t_{Obj}, $\lambda^t_{Obj,k}$, $\lambda^t_{Obj,h}$; else, solve second subproblem (5.38) and keep μ^t_{Feas}, $\lambda^t_{Feas,k}$, $\lambda^t_{Feas,h}$.
Step 2.	If all first set subproblems become feasible, update upper bound: $\varphi_{UB} := \mu_M + \sum_{t=1}^T \mu^t_{Obj} - \sum_{t=1}^T \widetilde{\eta}^t_{Obj}$; if $\frac{\lvert\varphi_{UB}-\varphi_{LB}\rvert}{\lvert\varphi_{LB}\rvert} < \epsilon$ go to Step 4;
Step 3.	Add all generated cuts to the master problem; solve this new master problem and update $\mu_M, \widetilde{\eta}_{Obj}, \widetilde{\mathrm{u}}$ plus $\widetilde{\mathrm{p}}_H$; go to Step 1.
Step 4.	Try to retrieve the rank-one conditions (as described in Section 5.2).

[12]Note that (5.37) could be linearized (cf. (5.17)).

[13]This cut will guarantee that the current commitment will be cut off from the feasible region of our master problem (5.28).

Remark 5.4. Since we are interested in the commitment and dispatch decisions only, it is not necessary to determine the exact voltage at each network bus. Namely, due to the strategy outlined in Section 5.2, it is sufficient to check whether, for any time $t \in \{1, ..., T\}$, the matrix $A(\lambda^t_{\mathrm{opt}}, r^t_{\mathrm{opt}})$ (as defined in (5.24)) has not more than 2 vectors in its kernel (see Corollary 5.1).

Chapter 6

Unit Commitment Under Uncertainty

To manage uncertainty[1] in power systems, basically, differentiation can be made between reserve requirements[2] (implicit reserve) [130], [131] and stochastic programming (explicit reserve) [132], [133], [134]. A simultaneous consideration of both implicit reserve and explicit reserve is proposed in [135]. While spinning reserve constraints take uncertainty into account merely implicit and may result in a policy that is not efficient (at least economically), stochastic programming provides explicit uncertainty models which will commit sufficient capacity for all uncertainties that were taken into account. For this reason, and, moreover, due to the fact that solutions of stochastic programs, which guard against uncertainty, are quite robust with respect to changes in the data, the stochastic programming approach has been selected (stochastic programming as a tool for decision making under uncertainty is also motivated in Chapter 2 of this thesis).

In the literature, various stochastic programming approaches are proposed to tackle unit commitment under uncertainty. For instance, Takriti, Birge, and Long [133] use progressive hedging [79] in order to solve two-stage stochastic unit commitment problems[3]. After configuration and tuning [136], this method is capable to solve large-scale stochastic unit commitment problems with a large number of

[1]In power systems, the major source of uncertainty results from the loads.

[2]Operating reverse constraints are widely employed in the industry. They should make sure that even under major forecasting errors the power systems will not be pushed to its limit.

[3]In their paper the transmission system is not taken into account.

scenarios within reasonable time[4].

Furthermore, in power planning, multi-stage stochastic programming [12], became more and more established in recent years. For the scope adopted in this thesis with its elaborate model of power flow, however, the multi-stage approach still seems premature, at least computationally. Here, two-stage models still pose challenging research questions.

6.1 Unit Commitment Under Uncertainty Using Two-Stage Stochastic Programming

In this section we will focus on planning a unit commitment schedule under uncertainty of both power demand and output of renewables. Hence, the uncertainties, at time interval t occur at the nodes (buses) and concern the active and reactive (apparent) power, denoted by $p^t(\xi)$ and $q^t(\xi)$ for $t = 1, ..., T$, respectively.

We assume that $z(\xi) = (p^t(\xi), q^t(\xi))$ is a random variable whose probability distribution is known at the beginning of the optimization horizon. The latter, alone, already is non-trivial, and obtaining meaningful probability distributions from statistical data is a field of active research in stochastic programming and beyond, see [137], for instance. In our case we will adopt a finite event space where the realizations (scenarios) and their probabilities are obtained from recorded load profiles of the past.

Concerning their operational flexibility, the on/off decisions of the coal fired thermal units are the most inertial ones. Even when making decisions with respect to a rather coarse, hourly time discretization, for instance, it is not possible to follow a random load-and-renewables profile by on/off determinations of thermal blocks alone. This observation leads to modeling the switching decisions of the coal fired thermal units as first-stage variables. The second-stage is formed by the remaining short term on/off decision for gas turbines and by the operation levels of the on-line thermal and pumped-storage units.

[4]Using progressive hedging, the modest-scale WECC-240 test system with 100 scenarios is solved in 15 minutes to an optimality gap of 1.5%. The WECC-240 test instance represents a simplified version of the US western interconnect and includes 240 buses, 140 generators, and 448 transmission elements.

Denoting then by $\mathbf{u}_I = \{\mathbf{u}_i^t\}_{i\in I}$ and $\mathbf{u}_R = \{\mathbf{u}_r^t\}_{i\in R}$ the Boolean vectors for switching decisions of coal fired blocks as well as gas turbines, and in addition by $\mathcal{U}_I$ and $\mathcal{U}_R$ their feasible sets, this leads to a random two-stage optimization problem of the following principal shape:

$$\text{“min”}\left\{g(\mathbf{u}_I) + h(\mathbf{u}_R, \mathbf{p}) \;:\; \begin{array}{ll} \mathcal{W}(\mathbf{p}^t, \mathbf{l}^t, \mathbf{X}^t) = z^t(\xi), & \forall t, \\ T_I\mathbf{u}_I^t + W_R\mathbf{u}_R^t + W_G\big[\mathbf{p}^t\;\mathbf{q}^t\big]^T \le b, & \forall t, \\ \mathbf{X}^t \in \mathfrak{N}_+, \;\; \operatorname{rank}(\mathbf{X}^t) = 1, & \forall t, \\ \mathbf{u}_I \in \mathcal{U}_I, \;\; \mathbf{u}_R \in \mathcal{U}_R, \;\; \mathbf{l} \in \mathcal{U}_H & \end{array}\right\}, \tag{6.1}$$

where $T_I, T_R, \mathcal{W}, D$, and g, h are appropriate operators describing the constraints and the objective in (5.15), respectively. Here, the variables $\mathbf{p}^1, ..., \mathbf{p}^T \in \mathbb{R}_+^{|I|+|R|}$ have been introduced in order to simplify notation. The latter variables represent the produced active power at the network buses $\mathcal{G}\backslash H$. It is assumed that the equations

$$\mathbf{p}_k^t = Y_k \bullet \mathbf{X}^t + p_k^t(\xi), \quad k \in \mathcal{G}\backslash H,$$

are part of $\mathcal{W}(\mathbf{p}^t, \mathbf{l}^t, \mathbf{X}^t) = z^t(\xi)$. This is to make randomness only affecting the right-hand side of the constraints. Furthermore, the vector $\mathbf{l} = (\mathbf{l}^1, ..., \mathbf{l}^T)$ contains the pumped-storage hydroelectricity variables and $\mathcal{U}_H$ depicts its feasible set. Moreover, $\mathfrak{N}_+$ denotes the set[5]

$$\{\mathbf{X}^t \in \mathcal{S}_+^{2n} : \mathbf{X}^t \text{ fulfills the network limits } (5.8) - (5.11)\}.$$

Relaxing the non-convex rank conditions in (6.1) we arrive at the following two-stage (mixed-integer) semidefinite random program:

$$\text{“min”}\left\{g(\mathbf{u}_I) + \Phi\left(\underbrace{\begin{bmatrix} z(\xi) \\ b \end{bmatrix}}_{:=z_b(\xi)} - \underbrace{\begin{bmatrix} 0 \\ T_I\mathbf{u}_I \end{bmatrix}}_{:=T\mathbf{u}_I}\right) : \mathbf{u}_I \in \mathcal{U}_I\right\}, \tag{6.2}$$

[5]Note that this set is equivalent to a spectrahedra (see Subsection 5.1.1).

where

$$\Phi(\mathbf{d}) := \min\left\{ h(\mathbf{u}_R, \mathbf{p}) : \begin{array}{ll} \mathcal{W}(\mathbf{p}^t, \mathbf{l}^t, \mathbf{X}^t) = \mathbf{d}_1^t, & \forall t, \\ W_R \mathbf{u}_R^t + W_G \big[\mathbf{p}^t \ \mathbf{q}^t\big]^T \leq \mathbf{d}_2^t, & \forall t, \\ \mathbf{u}_R \in \mathcal{U}_R, \ \ \mathbf{l} \in \mathcal{U}_H, \ \ \mathbf{X}^t \in \mathfrak{N}_+, & \forall t \end{array} \right\}.$$

Defining random variables $f(\mathbf{u}_I, z(\xi)) := g(\mathbf{u}_I) + \Phi(z_b(\xi) - T\mathbf{u}_I),\ \mathbf{u}_I \in \mathcal{U}_I$, the random program (6.1) turns into minimization over a family of random variables. Among these random variables, we now would like to choose the "best" prospect. To this, different, risk neutral and risk averse, (stochastic) selection criteria are applied, and compared later on[6].

6.1.1 Risk Neutral Approach

Raking the random variables $f(\mathbf{u}_I, z(\xi)), \mathbf{u}_I \in \mathcal{U}_I$ by its expectation leads to the following risk neutral stochastic program:

$$\min\Big\{ \underbrace{\mathbb{E}[f(\mathbf{u}_I, z(\xi))]}_{=\mathcal{Q}_{\mathbb{E}}(\mathbf{u}_I)} : \mathbf{u}_I \in \mathcal{U}_I \Big\}. \tag{6.3}$$

Note that all functions involved may be represented by linear ones (cf. Subsection 5.1.1), such that (6.3) can be seen as a two-stage stochastic mixed-integer linear semidefinite program. Basic properties and a detailed discussion about the characteristics of such programs are presented in Chapter 2. Since it is assumed that the underlying measure follows a finite discrete probability distribution, (6.3) possess the following block structure (see Theorem 2.36):

$$\begin{array}{lll} \min & g(\mathbf{u}_I) + \sum_{\xi=1}^{S} \pi_\xi h(\mathbf{u}_\xi, \mathbf{p}_\xi) & \\ \text{s.t.} & \mathcal{W}\left(\mathbf{p}_\xi^t, \mathbf{l}_\xi^t, \mathbf{X}_\xi^t\right) = z_\xi^t, & \forall t,\ \forall \xi, \\ & T_I \mathbf{u}_I^t + W_R \mathbf{u}_\xi^t + W_G \big[\mathbf{p}_\xi^t \ \mathbf{q}_\xi^t\big]^T \leq b, & \forall t,\ \forall \xi, \\ & \mathbf{u}_I \in \mathcal{U}_I, \ \ \mathbf{u}_\xi \in \mathcal{U}_R, \ \ \mathbf{l}_\xi \in \mathcal{U}_H, \ \ \mathbf{X}_\xi^t \in \mathfrak{N}_+, & \forall t,\ \forall \xi. \end{array} \tag{6.4}$$

[6] In Subsection 7.5, the benefit of risk averse approach (compared to the risk neutral one) is discussed.

Notice that for a fixed first-stage commitment $u_I \in \mathcal{U}_I$, (6.4) decomposes into the following scenario-specific mixed-integer semidefinite programs

$$\begin{aligned} \min \quad & \pi_\xi h(\mathbf{u}_\xi, \mathbf{p}_\xi) \\ \text{s.t.} \quad & \mathcal{W}\left(\mathbf{p}_\xi^t, \mathbf{l}_\xi^t, \mathbf{X}_\xi^t\right) = z_\xi^t, && \forall t, \\ & W_R \mathbf{u}_\xi^t + W_G \left[\mathbf{p}_\xi^t \; \mathbf{q}_\xi^t\right]^T \leq b - T_I \mathbf{u}_I^t, && \forall t, \\ & \mathbf{u}_\xi \in \mathcal{U}_R, \;\; \mathbf{l}_\xi \in \mathcal{U}_H, \;\; \mathbf{X}_\xi^t \in \mathfrak{N}_+, && \forall t. \end{aligned} \tag{6.5}$$

This is a characteristic of two-stage stochastic programs and the starting point for decomposition algorithms such as dual decomposition and the L-shaped method (see Chapter 3 for a detailed description of these algorithms). While dual decomposition is able to tackle two-stage programs with mixed-integer recourse and thus can be used to solve (6.4) for general settings, the L-shaped method, as it is introduced in Section 3.2, solely works for continuous recourse (i.e., the L-shaped method just works in case $\mathcal{U}_R = \emptyset$).

Applying dual decomposition [34] to (6.4), the decomposed programs (6.5) are SDP relaxations of unit commitment problems of the form (4.21) and thus they are of the shape (5.20). Hence, they can be solved via the deterministic Benders algorithm from Section 5.3.1. Together, this results in a decomposition algorithm (deterministic Benders decomposition for unit commitment) that is contained in a decomposition algorithm (scenario decomposition). This is described in detail in Subsection 6.1.3. There, we also introduce feasibility cuts for dual decomposition[7] in order to make this algorithm much more efficient.

In case $\mathcal{U}_R = \emptyset$, the L-shaped method, as it is introduced in Section 3.2, can be employed to solve (6.4). If further the approach from Section 5.3.1 is applied to tackle the decomposed scenario programs (6.5), this results in a nested Benders decomposition [87], [18]. This method is more efficient than dual decomposition [34] as it combines information about the individual scenarios in terms of cooperating objective and feasibility cuts.

[7]Note that dual decomposition [34] in its original form is simply a Branch-and-Bound algorithm, which branches on the first-stage variables. Therefore, it tends to be very slow.

6.1.2 Risk Averse Approach

The approach (6.3) is able to compensate scenarios with high costs by those with lower costs. This may be less tolerable in cases where rare, but extreme, events occur. For instance, if there is a very limited number of scenarios with great power demand. In such situations risk averse decision making is rational.

A possible risk averse decision strategy is provided by risk aversion via excess probability mean-risk models [5]. The latter approach aims at finding efficient solutions of (6.2) by solving the following mean-risk models:

$$\min\left\{\underbrace{\mathbb{E}[f(\mathbf{u}_I, z(\xi))]}_{=\mathcal{Q}_{\mathbb{E}}(\mathbf{u}_I)} + \rho \cdot \underbrace{\mathbb{P}[f(\mathbf{u}_I, z(\xi)) > \eta]}_{=\mathcal{Q}_{\mathbb{P}_\eta}(\mathbf{u}_I)} : \mathbf{u}_I \in \mathcal{U}_I\right\}, \tag{6.6}$$

i.e., among the random variables $\{f(\mathbf{u}_I, z(\xi))\}_{\mathbf{u}_I \in \mathcal{U}_I}$ those were selected that are Pareto efficient as for a given value of the mean they minimize the excess probability $\mathbb{P}[f(\mathbf{u}_I, z(\xi)) > \eta]$ and for a given value of the excess probability risk measure they minimize the mean $\mathbb{E}[f(\mathbf{u}_I, z(\xi))]$.

A detailed discussion about basic properties and characteristics of (6.6) is presented in Chapter 2. As it is assumed that the underlying probability distribution has a finite event space, (6.6) is equivalent to the following mixed-integer semidefinite program (see Theorem 2.36):

$$\begin{array}{lll} \min & g(\mathbf{u}_I) + \sum_{\xi=1}^{S} \pi_\xi h(\mathbf{u}_\xi, \mathbf{p}_\xi) + \rho \cdot \sum_{\xi=1}^{S} \pi_\xi \boldsymbol{\theta}_\xi & \\ \text{s.t.} & \mathcal{W}\left(\mathbf{p}_\xi^t, \mathbf{l}_\xi^t, \mathbf{X}_\xi^t\right) = z_\xi^t, & \forall t,\ \forall \xi, \\ & T_I \mathbf{u}_I^t + W_R \mathbf{u}_\xi^t + W_G \left[\mathbf{p}_\xi^t\ \mathbf{q}_\xi^t\right]^T \le b, & \forall t,\ \forall \xi, \\ & g(\mathbf{u}_I) + h(\mathbf{u}_\xi, \mathbf{p}_\xi) - M\boldsymbol{\theta}_\xi \le \eta, & \forall \xi, \\ & \mathbf{u}_I \in \mathcal{U}_I,\ \ \mathbf{u}_\xi \in \mathcal{U}_R,\ \ \mathbf{l}_\xi \in \mathcal{U}_H,\ \ \mathbf{X}_\xi^t \in \mathfrak{N}_+, & \forall t,\ \forall \xi. \end{array} \tag{6.7}$$

Here, the big-M constant M may be selected as

$$T \cdot \left(\sum_{k \in \mathcal{G} \setminus H} c_{k2} \left(P_k^{\max}\right)^2 + c_{k1} P_k^{\max} + c_{k0} \right).$$

In line with (6.4), program (6.7) decomposes into the following mixed-integer linear semidefinite programs, provided the first-stage decision $\mathbf{u}_I \in \mathcal{U}_I$ is fixed:

$$\begin{aligned} \min \quad & \pi_\xi h(\mathbf{u}_\xi, \mathbf{p}_\xi) + \rho \cdot \pi_\xi \boldsymbol{\theta}_\xi \\ \text{s.t.} \quad & \mathcal{W}\left(\mathbf{p}_\xi^t, \mathbf{l}_\xi^t, \mathbf{X}_\xi^t\right) = z_\xi^t, && \forall t, \\ & W_R \mathbf{u}_\xi^t + W_G \big[\mathbf{p}_\xi^t \; \mathbf{q}_\xi^t\big]^T \leq b - T_I \mathbf{u}_I^t, && \forall t, \\ & h(\mathbf{u}_\xi, \mathbf{p}_\xi) - M\boldsymbol{\theta}_\xi \leq \eta - g(\mathbf{u}_I), \\ & \mathbf{u}_\xi \in \mathcal{U}_R, \;\; \mathbf{l}_\xi \in \mathcal{U}_H, \;\; \mathbf{X}_\xi^t \in \mathfrak{N}_+, && \forall t. \end{aligned} \tag{6.8}$$

The decomposed programs (6.8) are essentially of the form (4.21) except that the violation of $h(\mathbf{u}_\xi, \mathbf{p}_\xi) \leq \eta - g(\mathbf{u}_I)$ is penalized by $\rho \cdot \pi_\xi$. Hence, they belong to the class of problems that can be handled by the Benders decomposition approach from Section 5.3.1. For this purpose, simply the objective and the constraints in the master problem (5.28) need to be modified and added by

$$\left[\sum\nolimits_{t=1}^{T} \left(\sum\nolimits_{k \in \mathcal{G} \setminus H} c_{k0} \cdot \mathbf{r}_k^t + \boldsymbol{\eta}_{Obj}^t\right)\right] + \rho \cdot \boldsymbol{\theta}$$

and

$$\left[\sum\nolimits_{t=1}^{T} \left(\sum\nolimits_{k \in \mathcal{G} \setminus H} c_{k0} \cdot \mathbf{r}_k^t + \boldsymbol{\eta}_{Obj}^t\right)\right] + c_I^T \mathbf{u}_I - \eta \leq M\boldsymbol{\theta},$$

respectively. As, in any case, (6.7) contains second-stage integers, the L-shaped method cannot be applied. However, dual decomposition still works.

6.1.3 Dual Decomposition for Two-Stage Stochastic Unit Commitment

Applying dual decomposition [34] to tackle (6.4) and (6.7), the decomposed programs (6.5) and (6.8), respectively, are of the shape (5.20). Hence, they can be tackled by the deterministic Benders decomposition approach from Section 5.3.1. If the latter is used for solving (6.5) and (6.8), respectively, we obtain the following dual decomposition algorithm:

ALGORITHM 6.1.1: Dual decomposition for two-stage stochastic unit commitment.

Initialize	Let $\mathbf{P}$ be the list of current problems. Denote for $P \in \mathbf{P}$ by $\varphi_{\mathrm{LD}}(P)$ its Lagrangian lower bound[8]. This bound is obtained by the proximal bundle method presented in Section 3.1, where the decomposed programs $D_\xi(\mathbf{\Lambda}_\xi)$ are solved by Algorithm 5.3.1. Put $\overline{\varphi} = +\infty$ and add the underlying problem to the list $\mathbf{P}$.
Step 1	If $\mathbf{P} = \emptyset$ then $\overline{u}$ with $\overline{\varphi} = \mathcal{Q}_{\mathbb{E}}(\overline{u}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(\overline{u})$ is optimal; Else, go to Step 2.
Step 2	Select and delete from the list $\mathbf{P}$ a problem $P \in \mathbf{P}$ and solve its Lagrangian dual. If $\varphi_{\mathrm{LD}}(P)$ is $+\infty$, go to Step 1; otherwise, go to Step 3.
Step 3	If $\varphi_{\mathrm{LD}}(P) \geq \overline{\varphi}$, then go to Step 1.
Step 3.1	The scenario solutions $u_\xi^{opt}, \xi = 1, ..., S$, gained by solving the Lagrangian dual are identical[9], i.e., $u_1^{opt} = ... = u_S^{opt}$. If further $\mathcal{Q}_{\mathbb{E}}(u_1^{opt}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(u_1^{opt}) < \overline{\varphi}$, then $\overline{\varphi} := \mathcal{Q}_{\mathbb{E}}(u_1^{opt}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(u_1^{opt})$ and $\overline{u} := u_1^{opt}$. Delete from $\mathbf{P}$ all problems P' with $\varphi_{\mathrm{LD}}(P') \geq \overline{\varphi}$; go to Step 1.
Step 3.2	If the solutions $u_\xi^{opt}, \xi = 1, ..., S$ differ, then run a feasibility heuristic. If its outcome $\widehat{u}$ is feasible and $\mathcal{Q}_{\mathbb{E}}(\widehat{u}) + \rho \cdot \mathcal{Q}_{\mathbb{P}_\eta}(\widehat{u}) < \overline{\varphi}$, then $\overline{\varphi} := Q_{\mathbb{E}}(\widehat{u})$ and $\overline{u} := \widehat{u}$. Delete from $\mathbf{P}$ all problems P' with $\varphi_{\mathrm{LD}}(P') \geq \overline{\varphi}$; go to Step 4.
Step 4	Select a component $(\mathbf{u}_I)_i$ of $\mathbf{u}_I$ and add two new problems to $\mathbf{P}$ which arise from P by adding the constraints $(\mathbf{u}_I)_i = 0$ and $(\mathbf{u}_I)_i = 1$, respectively; go to Step 1.

Remark 6.1. The above algorithm is a Branch-and-Bound algorithm that branches on the finite number of first-stage decisions $\mathbf{u}_I \in \mathcal{U}_I \subseteq \{0,1\}^{|I|}$. Hence, it will terminate after a finite number of steps.

[8]Here, the Lagrangian lower bound means the lower bound that is obtained by Lagrangian relaxation of the nonanticipativity constraint (3.2).

[9]This implies that u_1^{opt} is feasible for (6.7).

6.1.3.1 Selected Feasibility Heuristics

With the aid of feasibility heuristics, promising first-stage decisions that are feasible for the overall stochastic program may be generated. If such feasible points are available (and provide a good upper bound), the number of nodes that have to be examined in the underlying Branch-and-Bound tree can be reduced considerably[10].

Carøe and Schultz [34] use the solution of the Lagrangian dual (3.4) as a starting point for heuristics. Since the relaxed constraints (3.2) are quite simple (all introduced first-stage copies have to be equal), ideas for heuristics are obvious. To retrieve nonanticipativity (3.2) the first-stage copies have to be made identical.

In context of unit commitment, the subsequent feasibility heuristics are suggested (see [85] for further heuristics). Starting from the solution of the Lagrangian dual $u^{opt} = (u_1^{opt}, ..., u_S^{opt})$, the following first-stage determinations are computed and checked for feasibility:

(i) Average over all candidates $u_\xi^{opt}, \xi = 1, ..., S$ and round to integers[11]:

$$\widehat{u} = \text{round}\left(\sum\nolimits_{\xi=1}^{S} \pi_\xi u_\xi^{opt}\right).$$

(ii) Whenever there exists $\xi \in \{1, ..., S\}$ for which $(u_\xi^{opt})_i$ is equal to one[12], set $\widehat{u}_i^t = 1$, else $\widehat{u}_i^t = 0$:

$$\widehat{u} = \left\lceil \frac{1}{S} \sum\nolimits_{\xi=1}^{S} u_\xi^{opt} \right\rceil,$$

where $\lceil \cdot \rceil$ denotes the element-wise up rounding of the vector $\frac{1}{S}\sum_{\xi=1}^{S} u_\xi^{opt}$.

6.1.3.2 Feasibility Cuts

Due to feasibility heuristics, dual decomposition [34] finds good upper bounds rather quickly. However, the lower bound obtained from Lagrangian relaxation of nonanticipativity often is very poor. This is due to the fact that each decomposed scenario program is considered separately (no sharing of information between scenarios). The Lagrangian dual (3.4) provides for each scenario specific program an

[10] Due to the presence of improved upper bounds a lot of nodes can be pruned.
[11] Here, round depicts the element-wise rounding of the vector $\sum_{\xi=1}^{S} \pi_\xi u_\xi^{opt}$.
[12] This means that, for the scenario solution u_ξ^{opt}, generator $i \in I$ is on-line at time t.

individual first-stage determination. While the latter is feasible for its corresponding scenario it might be infeasible for others. This may result in big duality gaps and, consequently, in a large Branch-and-Bound tree.

For two-stage unit commitment, the root node Lagrangian dual (3.4) tends to provide scenario specific first-stage solutions that are almost always infeasible[13]. The reason for this is that start-up costs, which form a great part of the total costs, can be saved significantly by selecting first-stage determinations individually. On this account, (3.4) yields a very poor lower bound. This is most evident when the scenario corresponding load profiles vary considerably.

To receive improved lower bounds, feasibility cuts are added to each of the decomposed scenario programs. The added cuts can be divided into two groups: cuts that are added right at the beginning of dual decomposition and cuts that are added in each iteration.

First group of feasibility cuts: Note that for any scenario at any time step, there is a sufficient number of on-line first-stage generators to produce the required active power. In each time step, this in particular, holds true for the most power-consuming scenario. Therefore, we can add the following redundant constraints to the set of feasible first-stage decisions $\mathcal{U}_I$:

$$P_D^t(\xi) - \sum\nolimits_{r \in \mathcal{R}} P_r^{\max} - \sum\nolimits_{h \in \mathcal{H}} P_h^{\max} \leq \sum\nolimits_{i \in \mathcal{I}} \mathbf{u}_i^t \cdot P_i^{\max}, \quad t = 1, ..., T, \tag{6.9}$$

where $P_D^t(\xi) := \max_\xi \left(\sum_{k \in \mathcal{N}} p_{D_k}^t(\xi) \right)$.

Remark 6.2. Due to (6.9), a lot of switching decisions become inferior, already at the root node. Consequently, significantly less nodes have to be pruned by infeasibility (in context of Branch-and-Bound).

As already mentioned, dual decomposition [34] is not able to share information between the decomposed scenarios. We aim to change this by iteratively adding further feasibility cuts.

[13]This means that, for some of the scenarios, a schedule of first-stage generators is determined that is not feasible for all scenarios.

Second group of feasibility cuts: While executing Algorithm 6.1.1, this group of cuts will ensure that a wide range of impermissible first-stage actions does not occur twice. This is provided by the following routine:

- Having solved (3.4) at a current node of the Branch-and-Bound tree, we consider the individual scenario solutions $u_1^{opt}, ..., u_S^{opt}$. For each of them, we denote by $\mathcal{I}_{off}(\xi) \subset \mathcal{I}$ the set of (first-stage) generators that are switched off during the entire optimization horizon (note that this set can be empty).

- Next, for $\xi = 1, ..., S$, it is tested whether or not it is feasible to remain the generators in $\mathcal{I}_{off}(\xi)$ switched off. To this end, the following has to be checked for any $\xi \in \{1, ..., S\}$:

$$M\left(\overline{\xi}\right) \cap \{\mathbf{u}_I \in \mathcal{U}_{\mathcal{I}} : \mathbf{u}_i = 0, \text{ for } i \in \mathcal{I}_{off}(\xi)\} \neq \emptyset, \quad \text{for } \overline{\xi} \in \{1, ..., S\} \backslash \{\xi\},$$

  where[14]

$$M\left(\overline{\xi}\right) := \left\{ \mathbf{u}_I \in \mathcal{U}_{\mathcal{I}} : \begin{array}{ll} \mathcal{W}\left(\mathbf{p}_\xi^t, \mathbf{l}_\xi^t, \mathbf{X}_\xi^t\right) = z_\xi^t, & \forall t, \\ W_R \mathbf{u}_\xi^t + W_G \left[\mathbf{p}_\xi^t \; \mathbf{q}_\xi^t\right]^T \leq b - T_I \mathbf{u}_I^t, & \forall t, \\ \mathbf{u}_\xi \in \mathcal{U}_R, \;\; \mathbf{l}_\xi \in \mathcal{U}_H, \;\; \mathbf{X}_\xi^t \in \mathfrak{N}_+, & \forall t \end{array} \right\}.$$

- If for some $\overline{\xi} \in \{1, ..., S\} \backslash \{\xi\}$,

$$M\left(\overline{\xi}\right) \cap \{\mathbf{u}_I \in \mathcal{U}_{\mathcal{I}} : \mathbf{u}_i = 0, \text{ for } i \in \mathcal{I}_{off}(\xi)\} = \emptyset,$$

  a feasibility cut, which ensures that at least one of the generators in $\mathcal{I}_{off}(\xi)$ is switched on, is added to the set of feasible first-stage decisions $\mathcal{U}_{\mathcal{I}}$.

[14] $M\left(\overline{\xi}\right)$ depicts the feasible set of first-stage decisions for scenario $\overline{\xi}$.

Chapter 7

Computational Results

This chapter presents a series of computational results for the solution of the unit commitment problems (5.15), (6.4), and (6.7) with the aim of demonstrating the efficiency of the solution approaches presented in the previous chapters. Section 7.1 starts with the introduction of the problem instances. Here, the power system data has been taken from the IEEE power systems test case archive [111]. For some of the test instances, the infrastructure of the benchmark systems has been modified as pumped-storage hydroelectricity plants are added to a couple of network buses. Moreover, this section describes the way in which load profiles and load profile scenarios, respectively, are generated.

The performance of Algorithm 5.3.1, which has been proposed for solving deterministic AC unit commitment problems of the form (5.15), is investigated in Section 7.2. To solve two-stage programs (6.4) with $\mathcal{U}_R = \emptyset$, Algorithm 3.2.1 and Algorithm 6.1.1 have been suggested. Their effectiveness is studied in Section 7.3.1.1 and Section 7.3.1.2, respectively. In doing so, their total computational times are compared to each other in Section 7.3.1.2.

In case of second-stage integers[1], we have utilized Algorithm 6.1.1 to tackle the risk neutral and risk averse two-stage programs (6.4) and (6.7), respectively. Numerical test results for the risk neutral models (6.4) with $\mathcal{U}_R \neq \emptyset$ are reported in Section 7.3.1.3, whereas the computational results for Algorithm 6.1.1 for the solution of the risk-averse two-stage programs (6.7) are presented in Section 7.3.2.

Section 7.4 will highlight the value of the stochastic solution. Finally, this chapter concludes by discussing the benefit of the risk averse approach (Section 7.5).

[1]Note that in this case Algorithm 3.2.1 cannot be applied.

The computational tests have been performed using MATLAB R2015a with MILP's and QP's solved by CPLEX Studio 12.51 [138] (connector to MATLAB). For the arising SDPs we have employed SeDuMi version 1.3 [119]. A simple hardware set (laptop) has been used, namely consisting of an Intel(R) Core(TM) i7-2640M CPU @ 2.80GHz 2.80GHz processor with 4 GB of RAM running under Windows 7 Professional (Service Pack 1).

We would point out that in all numerical tests listed below, the obtained solutions satisfy the kernel condition[2] (5.27). This implies that the relaxed rank-one conditions may be recovered (see Section 5.2). Thus, all considered unit commitment problems were solved to the declared duality gaps.

Furthermore, we mention that for the considered benchmark systems, our implemented semidefinite programming based OPF solver, i.e., the solver that is used for solving the subproblems (5.29), provides the same solution as the implemented algorithms by Madani, Asharphijuo, and Lavaei [126] and Molzahn et al. [125].

7.1 Power System Data

To exhibit the efficiency of our proposed decomposition approach, it is tested with the well-known 14-bus IEEE network[3], 30-bus IEEE network[4], 39-bus IEEE (New England) test system[5], and 57-bus IEEE network[6]. For the 14-bus IEEE network and 30-bus IEEE network, this data is summarized in Appendix D.

We adopt a daily planning horizon subdivided into 24 equidistant time intervals. Starting from the IEEE load data as a reference, we first developed a practically relevant deterministic load profile for an individual day.

In order to generate scenarios (to obtain a finite discrete probability distribution), the preassigned daily load profile (p, q) has been perturbed as follows: with S denoting the number of realizations, scenarios

$$(p_\omega, q_\omega) = (p, q) + n_\omega \cdot (p, q), \quad \omega = 1, ..., S \tag{7.1}$$

[2]Here, eigenvalues smaller than 10^{-5} were considered as zero-eigenvalues, i.e., $0 \equiv [0, 10^{-5}]$.
[3]The 14-bus IEEE system consists of 5 generators, 20 transmission lines, and 11 loads.
[4]The 30-bus IEEE system consists of 6 generators, 41 transmission lines, and 24 loads.
[5]The 39-bus IEEE system consists of 10 generators, 46 transmission lines, and 29 loads.
[6]The 57-bus IEEE system consists of 7 generators, 80 transmission lines, and 50 loads.

are formed, where n_ω is a random number sampled from the normal distribution. In doing so, two normal distributions were considered, one with mean $\mu = 0$ and standard deviation $\sigma = 0.2$ and one with $\mu = 0$ and $\sigma = 0.5$. Figure 7.1 displays scenarios obtained in this way (the slightly thicker curve in this figure represents the preassigned daily load profile whose perturbation resulted in the scenarios).

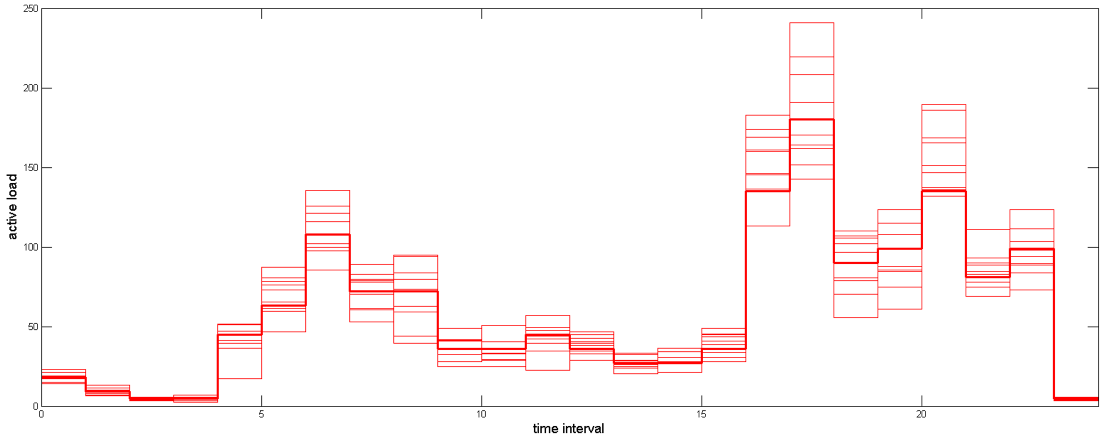


FIGURE 7.1: Scenario generation: active load scenarios which are obtained from perturbing the deterministic load profile (thicker curve).

In our tests, we have started with two basic deterministic network infrastructures: power systems with purely thermal generation and with pumped-storage plants added.

Stochastic expansions of these models were obtained by introducing random load values and assigning roles to variables making them members of the first and second stages, respectively. In this way, decisions in the first stage comprise on/off switching for the coal fired blocks. Variables in the second stage represent output levels of the coal fired units, switching decisions and output levels at gas turbines, and, if present, pumping and generation modes in the pumped storage plants. We also mention that start-up costs in both stages are modeled via suitable Boolean indicator variables.

7.2 Deterministic Models

Table 7.1 reports our computational results for the (deterministic) Benders decomposition approach from Section 5.3.1. Here, the deterministic load data equals the expected value of the random data, i.e., $(p, q) = \mathbb{E}_\omega[(p_\omega, q_\omega)]$. The stopping criterion has been set to 10^{-2}, i.e., in all of the tests listed below the overall mixed-integer SDP is solved to less then a 1% optimality gap.

TABLE 7.1: Computational results for the Benders decomposition approach from Section 5.3.1.

Model	Network	Generators	PSH Units	Variables (Binaries)	Constr.	Iter.	CPU	Gen. Cuts	Costs
D1	IEEE 14	5	0	11496 (120)	3319	14	132.68s	336	278328.52
D2	IEEE 14	5	1	11569 (120)	3465	21	184.27s	504	277770.19
D3	IEEE 14	5	2	11642 (120)	3611	35	329.35s	840	277814.05
D4	IEEE 30	6	0	46656 (144)	5250	11	204.15s	264	17597.26
D5	IEEE 30	6	1	46729 (144)	5396	18	324.57s	432	17597.61
D6	IEEE 30	6	2	46802 (144)	5442	20	372.09s	480	17604.62
D7	NE 39	10	0	77976 (240)	7694	4	94.71s	96	918751.08
D8	NE 39	10	1	78049 (240)	7840	1	22.73s	24	557948.54
D9	NE 39	10	2	78122 (240)	7986	1	22.18s	24	557948.54
D10	IEEE 57	7	0	161568 (168)	8237	19	944.08s	456	1117780.35
D11	IEEE 57	7	1	161641 (168)	8383	11	521.80s	264	988369.13
D12	IEEE 57	7	2	161714 (168)	8529	13	570.24s	312	988301.96

Starting from left, the following information is "encoded" in the columns of Table 7.1: model number, IEEE benchmark network, numbers of generators and pumped-storage hydroelectricity units, as well as resulting numbers of continuous and integer (binary) variables, plus constraints. The remaining four columns display the numbers of iterations, CPU time, cuts generated, and optimal costs.

Our Benders decomposition approach has been significantly improved by adding the constraints

$$\sum\nolimits_{k\in\mathcal{N}} p^t_{D_k} \leq \sum\nolimits_{k\in\mathcal{G}} \mathbf{u}^t_k \cdot P^{\max}_k + \sum\nolimits_{h\in\mathcal{H}} (\mathbf{p}^t_h - \mathbf{w}^t_h), \quad t = 1, ..., T, \tag{7.2}$$

to the master problem (5.28). For each time interval, these requirements guarantee that there is a sufficient number of on-line thermal generators to produce the required active power. Doing so, a lot of switching decisions become inferior, already in the master problem (5.28), such that they do not have to be cut off by solving the (computationally expensive) semidefinite subproblems (5.38).

Due to the presence of pumped-storage hydroelectricity stations more (binary) on/off switching decisions become feasible. As a consequence, the number of iterations and the total computational time increase. Nevertheless, the number of iterations remains quite small.

7.3 Stochastic Models

7.3.1 Risk Neutral Models

In this section, we present computational results for the (risk neutral) two-stage stochastic unit commitment problems (6.3). These programs were solved by the L-shaped method and by dual decomposition for networks exclusive of gas turbines (no second-stage integers), and by dual decomposition for networks including gas turbines. In the former case (no second-stage integers), the CPU times of the two methods are compared to each other.

7.3.1.1 Stochastic Benders Decomposition

In this section numerical results for Algorithm 3.2.1 are presented. Recall that this algorithm works only in case of continuous recourse. For two-stage stochastic unit commitment (6.3) this means that all generator switching decisions must be first-stage variables. Hence, this approach can be applied only if the system does not incorporate gas turbines.

TABLE 7.2: Computational results for the stochastic Benders decomposition algorithm (L-shaped method) for normally distributed load with mean $\mu = 0$ and standard deviation $\sigma = 0.2$.

Model	Network	Scena-rios	PSH	Var.	Constr.	Iter.	CPU (SeDuMi)	SeDuMi Calls	Gen. Cuts	Costs
E1	IEEE14	10	0	113760	30040	11	879 (868)	2713	2640	291022.02
E2	IEEE14	10	1	114610	31500	13	1134 (1112)	3499	3120	280561.06
E3	IEEE14	10	2	115340	32960	24	2370 (2285)	6259	5760	280561.79
E4	IEEE14	50	0	568920	148800	9	4330 (3919)	11682	10800	313346.08
E5	IEEE14	50	1	572570	156100	9	4576 (4140)	11633	18000	315479.80
E6	IEEE14	50	2	576220	163400	21	12398 (9873)	26983	25200	305817.80
E7	IEEE30	10	0	465264	48720	6	1114 (1108)	1454	1440	18087.58
E8	IEEE30	10	1	465994	50180	11	2142 (2116)	2742	2640	17891.83
E9	IEEE30	10	2	466724	50640	13	2618 (2582)	3223	3120	17891.42
E10	IEEE30	50	0	2325744	241920	6	5495 (5287)	7253	7200	18659.08
E11	IEEE30	50	1	2329394	249220	6	5458 (5220)	7347	7200	18655.58
E12	IEEE30	50	2	2333044	251520	6	5551 (5300)	7347	7200	18655.90
E13	IEEE57	10	0	465264	48720	14	6874 (6865)	3405	3360	810438.44
E14	IEEE57	10	1	465994	50180	12	5270 (5234)	2925	2880	795226.48
E15	IEEE57	10	2	466724	50640	15	6302 (6235)	3698	3600	795430.91
E16	IEEE57	50	0	2325744	241920	12	29535 (28579)	14784	14400	822369.14
E17	IEEE57	50	1	2329394	249220	11	23732 (22916)	13584	13200	806768.30
E18	IEEE57	50	2	2333044	251520	12	26019 (24926)	14784	14400	806758.92

Table 7.2 and Table 7.3 exhibit our computational results for Algorithm 3.2.1. They are structured as follows: the leftmost seven columns correspond to those of Table 7.1 with the exception that now scenarios have to be listed (in column three). Columns eight to ten correspond to the CPU time with SeDuMi share in brackets, the number of SeDuMi calls, generated cuts, and the optimal costs.

For each stochastic program listed, the overall (mixed-integer) semidefinite relaxation has been solved to less than a 1% optimality gap. The tables above show that most time (on average 87.04%) is spent solving the arising semidefinite subproblems by SeDuMi. Essentially, this is due to the vast number of cuts that have to be generated to solve the scenario problems. For instance, in case of 50 scenarios and the assumed partition into 24 time steps, in each iteration, 1200 cuts have to be generated (i.e., at least the same number of SDPs need to be solved).

TABLE 7.3: Computational results for the stochastic Benders decomposition algorithm (L-shaped method) for normally distributed load with mean $\mu = 0$ and standard deviation $\sigma = 0.5$.

Model	Network	Scena-rios	PSH	Var.	Constr.	Iter.	CPU (SeDuMi)	SeDuMi Calls	Gen. Cuts	Costs
E19	IEEE14	10	0	113760	30040	10	881 (870)	2766	2400	283843.55
E20	IEEE14	10	1	114610	31500	10	870 (857)	2806	2400	283831.99
E21	IEEE14	10	2	115340	32960	14	1413 (1379)	3924	3360	283831.99
E22	IEEE14	50	0	568920	148800	11	5714 (5119)	11589	13200	301460.66
E23	IEEE14	50	1	572570	156100	12	6457 (5683)	15996	15600	301460.66
E24	IEEE14	50	2	576220	163400	11	5961 (5247)	14964	13200	301460.66
E25	IEEE30	10	0	465264	48720	5	905 (900s)	1219	1200	19533.07
E26	IEEE30	10	1	465994	50180	6	1068 (1059)	1477	1440	19533.07
E27	IEEE30	10	2	466724	50640	6	1101 (1090)	1477	1440	19533.07
E28	IEEE30	50	0	2325744	241920	6	5445 (5236)	7390	7200	17906.23
E29	IEEE30	50	1	2329394	249220	6	5515 (5280)	7610	7200	17906.23
E30	IEEE30	50	2	2333044	251520	5	5632 (5382)	7610	7200	17906.23
E31	IEEE57	10	0	465264	48720	9	4405 (4385)	2227	2160	537418.91
E32	IEEE57	10	1	465994	50180	16	6644 (6574)	3910	3840	521393.09
E33	IEEE57	10	2	466724	50640	22	9303 (9144)	5375	5280	521395.89
E34	IEEE57	50	0	2325744	241920	12	29630 (28698)	15141	14400	768945.58
E35	IEEE57	50	1	2329394	249220	11	24190 (23392)	13941	13200	753199.30
E36	IEEE57	50	2	2333044	251520	11	24589 (23686)	13942	13200	752819.76

7.3.1.2 Dual Decomposition for Power Systems Exclusive of Gas Turbines (No Second-stage Integers)

Table 7.4 and Table 7.5 present numerical results for Algorithm 6.1.1. These tables are organized as follows: the model number, the IEEE benchmark system, the number of scenarios, and the number of pumped-storage hydroelectricity stations are listed in column one to four. Further, in the remaining four columns, the number of SDPs solved, the CPU time for Algorithm 6.1.1 (with SeDuMi share in brackets), the CPU time for Algorithm 3.2.1 (with SeDuMi share in brackets), and the optimal costs[7] are displayed.

Table 7.4: Power systems exclusive of gas turbines. Computational results for the dual decomposition algorithm for normally distributed load with standard deviation $\sigma = 0.2$.

Model	Network	Scena-rios	PSH	SeDuMi Calls	CPU (SeDuMi)	CPU Time L-Shaped Method	Costs
E1	IEEE14	10	0	5743	2128 (2083)	879 (868)	292565.64
E2	IEEE14	10	1	10276	3853 (3718)	1134 (1112)	293056.19
E3	IEEE14	10	2	17285	6545 (6128)	2370 (2285)	293156.36
E4	IEEE14	50	0	20730	7668 (7523)	4330 (3919)	314167.29
E5	IEEE14	50	1	49298	16119 (15526)	4576 (4140)	314174.97
E6	IEEE14	50	2	59774	22530 (21373)	12398 (9873)	314279.54
E7	IEEE30	10	0	7668	5705 (5661)	1114s (1108)	18079.13
E8	IEEE30	10	1	9028	6066 (5938)	2142s (2116)	17821.18
E9	IEEE30	10	2	4323	3276 (3229)	2618s (2582)	17822.27
E10	IEEE30	50	0	15252	11235 (11127)	5495s (5287)	18643.85
E11	IEEE30	50	1	16623	11247 (11057)	5458s (5220)	18645.24
E12	IEEE30	50	2	16840	11464 (11257)	5551s (5300)	18645.03

The execution of Algorithm 6.1.1 requires to solve almost identical scenario specific unit commitment problems again and again (see Chapter 6). In order to improve computational efficiency of this method, we stored the generated Benders cuts and reused them in each iteration. This has decreased CPU time considerably. Having solved a scenario specific unit commitment problem once, which takes minutes, it will be solved in a few seconds in future iterations. Cut deletion, clearly, would have been an option. This has not been pursued, since the share of solving master problems (including all previous cuts) amounted to a mere 3% or less of the total computation time.

TABLE 7.5: Power systems exclusive of gas turbines. Computational results for the dual decomposition algorithm for normally distributed load with standard deviation $\sigma = 0.5$.

Model	Network	Scena-rios	PSH	SeDuMi Calls	CPU (SeDuMi)	CPU Time L-Shaped Method	Costs
E19	IEEE14	10	0	6018	2167 (2118)	881 (870)	285349.35
E20	IEEE14	10	1	8217	2990 (2892)	870 (857)	285381.58
E21	IEEE14	10	2	13946	4937 (4693)	1413 (1379)	285360.81
E22	IEEE14	50	0	21077	7607 (7445)	5714 (5119)	302676.35
E23	IEEE14	50	1	34231	12749 (12342)	6457 (5683)	302806.67
E24	IEEE14	50	2	38888	14873 (14137)	5961 (5247)	301742.02
E25	IEEE30	10	0	2177	1585 (1568)	905 (900)	19533.00
E26	IEEE30	10	1	17173	11988 (11495)	1068 (1059)	20119.91
E27	IEEE30	10	2	5181	3515 (3436)	1101 (1090)	20119.91
E28	IEEE30	50	0	13238	9630 (9529)	5445 (5236)	17906.23
E29	IEEE30	50	1	16256	10648 (10460)	5515 (5280)	17913.58
E30	IEEE30	50	2	16305	11153 (10949)	5632 (5382)	17912.77

In contrast to Algorithm 3.2.1, where all generated cuts can be combined, dual decomposition considers each scenario individually. As expected, the computational time increases by using this method (compare column 6 and column 7 in the tables above). However, due to the introduced feasibility cuts (see Subsection 6.1.3.2), regarding total CPU time dual decomposition is competitive when compared with Algorithm 3.2.1. We mention that without these cuts, none of the programs listed, could be solved within the given time of 24 hours[8].

7.3.1.3 Dual Decomposition for Power Systems Incorporating Gas Turbines (Second-stage Integers)

Solving stochastic programs with integers in the second stage is in fact the strength of dual decomposition. For such stochastic programs, Algorithm 3.2.1 cannot be applied as cuts obtained from different scenarios could not be combined simply.

[7] The programs have been solved to less than a 1% optimality gap.
[8] The termination criterion has been set to a 1% optimality gap.

Table 7.6 displays numerical results for Algorithm 6.1.1 for risk neutral stochastic programs (6.4) with $\mathcal{U}_R \neq \emptyset$. Its columns correspond to those of Table 7.2 with the exception that the number of iterations and the number of generated cuts are not listed.

TABLE 7.6: Power systems incorporating gas turbines. Computational results for the dual decomposition algorithm for normally distributed load with standard deviation $\sigma = 0.2$.

Model	Network	Scenarios	Generators	PSH	Var.	Constr.	SeDuMi Calls	CPU (SeDuMi)	Costs
E37	IEEE 14	10	5(3)	0	114408	30040	6350	2135 (2083)	258080.00
E38	IEEE 14	10	5(3)	1	115258	31500	6402	2144 (2084)	247850.00
E39	IEEE 14	10	5(3)	2	115988	32960	9772	3572 (3424)	240280.00
E40	IEEE 14	50	5(3)	0	572448	148800	16864	5992 (5905)	281230.00
E41	IEEE 14	50	5(3)	1	576098	156100	26261	9815 (9556)	278960.00
E42	IEEE 14	50	5(3)	2	579748	163400	56706	18353 (17589)	271023.66
E43	IEEE 30	10	6(4)	0	466128	48720	9969	6955 (6765)	18077.00
E44	IEEE 30	10	6(4)	1	466858	50180	3771	2770 (2736)	17816.00
E45	IEEE 30	10	6(4)	2	467588	50640	3698	2818 (2783)	17816.00
E46	IEEE 30	50	6(4)	0	2330448	241920	6926	4762 (4736)	17896.00
E47	IEEE 30	50	6(4)	1	2334098	249220	14126	10086 (9934)	18606.00
E48	IEEE 30	50	6(4)	2	2337748	251520	14368	10681 (10509)	18607.00

Expectedly, the computational time for dual decomposition reduces by shifting generators to the second-stage, since this results in a clearly smaller Branch-and-Bound tree. Furthermore, shifting generators to the second-stage provides more flexibility regarding the choice of (feasible) first-stage decisions. This is reflected by considerably lower costs.

7.3.2 Risk Averse Models

This section summarizes some of our numerical results for solving risk averse excess probability mean-risk models (6.6) by using dual decomposition (cf. Section 3.1).

The results are reported in Table 7.7 which is organized just like Table 7.6. In column one, EP refers to the Excess Probability mean-risk model (6.6), whereas the subsequent numerals are related to the corresponding risk neutral models. The latter means that if ρ is set to 0 in (6.6), then, EP1 equals E1, EP3 equals E3, and so on.

TABLE 7.7: Computational results for the risk averse model.

Model	Network	Scenarios	Generators	PSH	Var.	Constr.	SeDuMi Calls	CPU (SeDuMi)	Costs
EP1	IEEE 14	10	5(0)	0	113770	30050	5292	1462 (1394)	296117.52
EP37	IEEE 14	10	5(3)	0	114418	30050	13618	3237 (3136)	287906.38
EP3	IEEE 14	50	5(0)	0	568970	148850	20730	5399 (5212)	319247.29
EP40	IEEE 14	50	5(3)	0	572498	148850	56654	12265 (12000)	281702.41
EP5	IEEE 30	10	6(0)	0	465274	48730	3448	1935 (1908)	18494.48
EP43	IEEE 30	10	6(4)	0	466138	48730	23132	10773 (10609)	18150.40
EP7	IEEE 30	50	6(0)	0	2325794	241970	15252	6369 (6294)	19167.85
EP48	IEEE 30	50	6(4)	0	2330498	241970	21251	12733 (12563)	19052.32

In all of the tests listed above, the threshold η (cf. model (6.6)) has been set to the optimal value of the risk neutral model (6.3). The penalty parameter ρ has been fixed to 10^5 in the models EP1, EP37, EP3, and EP40 – in the models EP5, EP43, EP7, and EP48, it as been set to 10^4.

Observe that the total times for solving the risk averse models are approximately the same as for the risk neutral ones except for the test instances containing gas turbines, here, the total time has roughly doubled. This is due to the complexity of the single-scenario unit commitment problems. We merely add just one big-M constraint to each of them and penalize its violation, such that, in this case, (6.7) is not that much harder to solve than (6.4).

7.4 The Value of The Stochastic Solution

When exploring the usefulness of the stochastic programming approach in (6.1), achieving second stage feasibility turned out difficult. In fact, in our instances

E1-E6, none of the optimal solutions to the averaged models has been feasible for all individual scenarios. This is reflected by Table 7.8, where column 2 displays infeasibility regarding both numbers of scenarios and accumulated probability.

Furthermore, column 2 of Table 7.8 reveals that deterministic solutions could be very misleading as they may be infeasible in a considerable number of cases. By contrast, the solutions of the stochastic programs E1-E6 are quite robust with respect to changes in the data. Moreover, calculating the gap between the wait-and-see (WS) solution[9] (the WS solution for E1-E6 is displayed in column 3 of Table 7.8) and the optimal value of E1-E6, it turned out, that, after the fact, the stochastic solutions are not optimal, but fairly good. These gaps and in addition the expected value of perfect information[10] (EVPI) are reported in column 5 and 4 of Table 7.8, respectively.

TABLE 7.8: Evaluation of the solutions of the expected value models.

Model	Infeasibility	Wait-and-See	EVPI	Gap
E1	3 (19.1%)	249098.16	43467.48	17.45%
E2	4 (24.8%)	240916.12	52140.07	21.64%
E3	3 (19.1%)	234096.86	59059.5	25.22%
E4	9 (15.0%)	273349.90	40817.39	14.93%
E5	19 (36.7%)	271728.20	42446.77	15.62%
E6	9 (15.0%)	271111.36	43168.18	15.92%

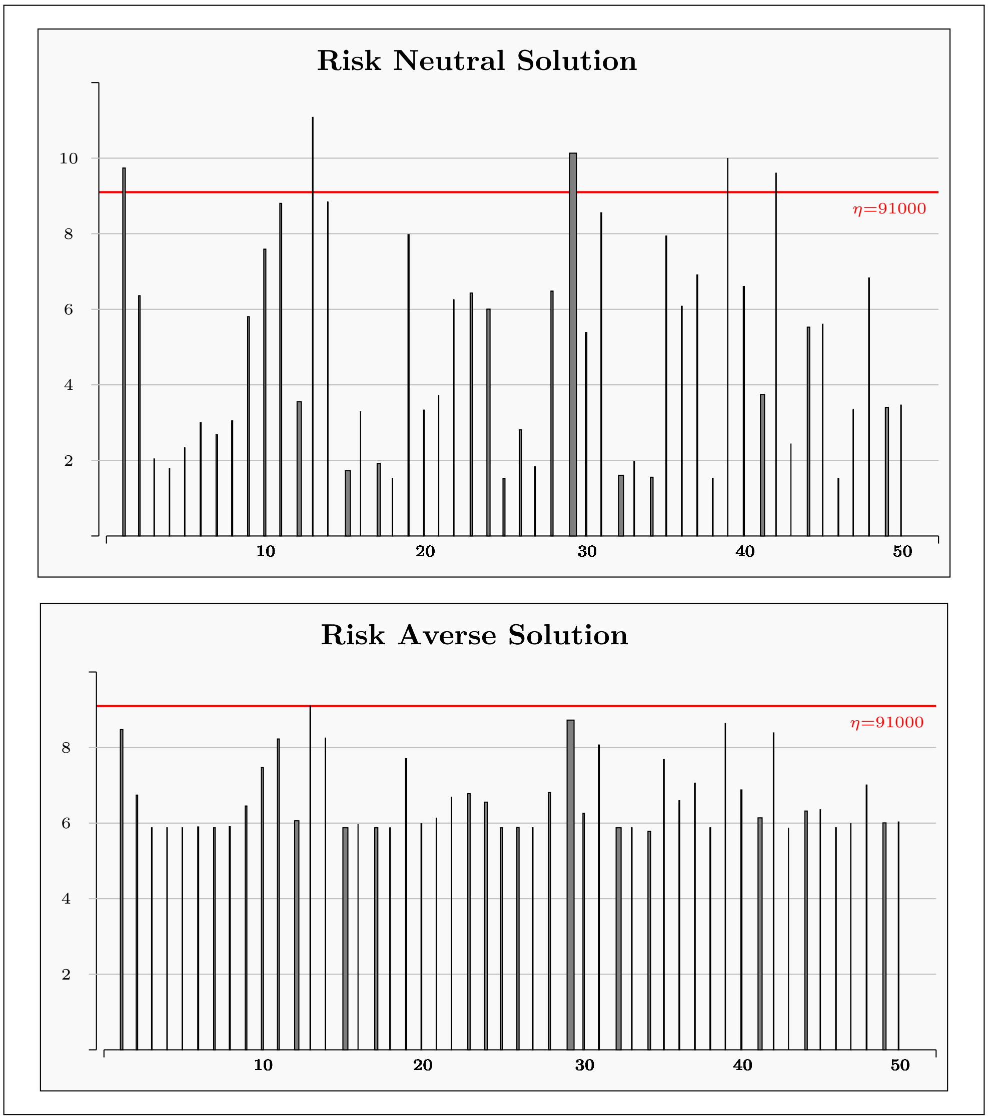
7.5 The Benefit of the Risk Averse Approach

In order to push the effect of the risk averse approach, compared to the risk neutral one, we have considered instances whose scenarios were less power consuming. To this end, we just divided the scenario load profiles (from the results above) by four. Than much more first-stage solutions become feasible – as more generators may be in off-state.

[9] The expected value $\mathbb{E}\left[\min_{\mathbf{X}\in\mathcal{X}} f(\mathbf{X},\omega)\right]$ is called wait-and-see solution.

[10] The value $\min_{\mathbf{X}\in\mathcal{X}} \mathbb{E}\left[f(\mathbf{X},\omega)\right] - \mathbb{E}\left[\min_{\mathbf{X}\in\mathcal{X}} f(\mathbf{X},\omega)\right]$ is called the expected value of perfect information.

FIGURE 7.2: Impact of different stochastic criteria on the shapes of the optimal solutions. The bar charts show objective values for each of the single-scenario problems for the IEEE 14 bus test instance with 50 scenarios (2 coal fired blocks, and 3 gas turbines).



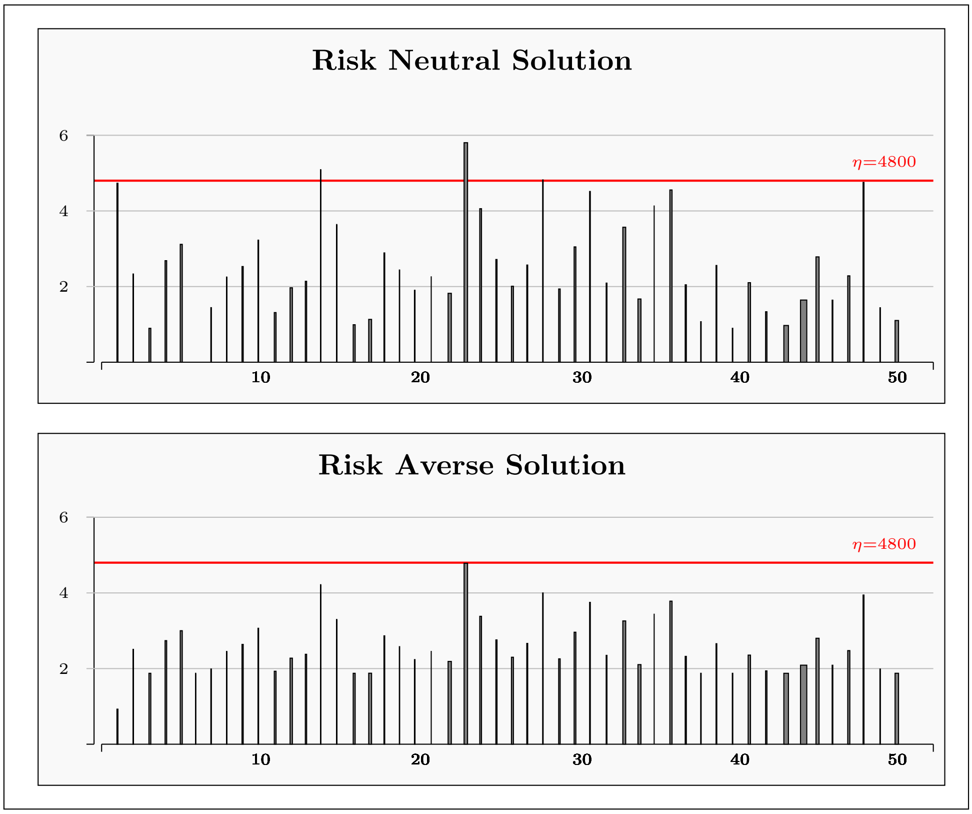
The bar charts in Figure 7.2 and Figure 7.3 illustrate the impact of different, risk neutral and risk averse, stochastic criteria on the shapes of the optimal solutions. They display the objective values of the single scenarios for the solutions to the 14-bus IEEE network[11] and the 30-bus IEEE network[12], respectively, both with 50 scenarios, each with the risk neutral and risk averse solution. Doing so, each individual column symbolizes one of the 50 scenarios where its particular

[11]Consisting of two coal fired blocks and three gas turbines.
[12]Consisting of two coal fired blocks and four gas turbines.

height and width refers to its objective value and its corresponding probability, respectively.

The expectation based model minimizes the sum over all single-scenario programs (where each scenario is weighted by its probability). This implies that scenarios with high costs may be compensated by scenarios with lower costs. In doing so, variability is neglected at all. Hence, this may result in a solution whose associated random variable fluctuates a lot and takes unfavorable values “too often”. These drawbacks can be observed in Figure 7.2 and Figure 7.3, respectively. For the IEEE 14 bus network (see Figure 7.2), for instance, the solution of the expectation based model varies much and incurs power production costs higher than 91000 in five of the fifty scenarios (corresponding probability is 14.1%). The expected

FIGURE 7.3: Impact of different stochastic criteria on the shapes of the optimal solutions. The bar charts show objective values for each of the single-scenario problems for the IEEE 30 bus test instance with 50 scenarios (2 coal fired blocks, and 4 gas turbines).



value amounts to 49253.70. In contrast, despite the fact that the expected value of the excess probability mean-risk solution is 66073.50, its single-scenario objectives do not vary that much and none of them exceeds the threshold $\eta = 91000$. This concludes our remarks on the benefit of the risk averse approach.

Chapter 8

Concluding Remarks

In this thesis, we have analyzed the structure of two-stage stochastic semidefinite programs with continuous recourse and mixed-integer recourse, respectively. Fundamental properties of these stochastic programs have been derived and decomposition methods to tackle them have been presented.

Furthermore, we have brought together unit commitment in AC transmission systems with risk averse stochastic optimization employing semidefinite programming. The latter recently was boosted by rank relaxations of semidefinite programs that lead to (globally) solvable optimal power flow problems. More specifically, relaxations of rank-one conditions could be recuperated for certain classes of electricity networks including among others popular IEEE OPF test instances.

Our focus has been to explore the potential of the recent findings in power flow when addressed under data uncertainty. The computations in this thesis confirm in principal that such a model extension remains computationally doable provided proper decomposition techniques are integrated into the algorithmic treatment.

In this context, we have introduced feasibility cuts for the dual decomposition method proposed by Carøe and Schultz [34], making this algorithm much more efficient and thus applicable for two-stage unit commitment problems.

Appendix A

Selected Facts of Semidefinite Programming

The set of $m \times n$ real matrices may be interpreted as a vector space in $\mathbb{R}^{n \cdot m}$ which is equipped with the inner product

$$\langle A, B \rangle = \operatorname{tr}(B^T A) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}, \tag{A.1}$$

for $A, B \in \mathbb{R}^{m \times n}$. Here, the trace $\operatorname{tr}(B^T A)$ also written by $A \bullet B$, is the sum of the diagonal elements of the square matrix $B^T A \in \mathbb{R}^{n \times n}$.

Note that the trace operator is a linear function and that for a product of matrices, the trace is invariant under cyclic permutations, i.e., the following holds true:

$$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB).$$

Further, the norm associated with the inner product (A.1) is the Frobenius norm,

$$\|A\|_F = \sqrt{A \bullet A}.$$

Definition A.1. A square matrix $A \in \mathbb{R}^{n \times n}$ is called symmetric if it is equal to its transpose, i.e., $A = A^T$. The set of symmetric matrices having dimension n will be denoted by $\mathcal{S}^n$. Further, a matrix $A \in \mathbb{R}^{n \times n}$ is called skew-symmetric if it satisfies[1] $A = -A^T$.

[1] Note that the diagonal of a skew-symmetric matrix has to be zero.

Equipped with the inner product (A.1) the set of symmetric matrices and the set of skew-symmetric matrices become a vector space of dimension $\binom{n+1}{2}$ and $\binom{n}{2}$, respectively. Observe that the orthogonal complement of the space of symmetric matrices is the space of skew-symmetric matrices. This implies that the symmetric matrices and the skew-symmetric matrices together span $\mathbb{R}^{n\times n}$. In other words, any matrix $A \in \mathbb{R}^{n\times n}$ can be decomposed into its symmetric part $(A + A^T)/2$ and its skew-symmetric part $(A - A^T)/2$,

$$A = \frac{A + A^T}{2} + \frac{A - A^T}{2}.$$

For symmetric matrices the following properties hold:

- All eigenvalues of a symmetric matrix $A \in \mathcal{S}^n$ are real.
- The trace of a matrix $A \in \mathcal{S}^n$ is equal to the sum of its eigenvalues.
- The rank of a symmetric matrix equals its nonzero eigenvalues.
- By forming for $A \in \mathcal{S}^n$ the real orthogonal matrix P which as its columns has the eigenvectors of A, then,

  $$PAP^T = \Lambda,$$

  where Λ is a diagonal matrix having the eigenvalues of A on its main diagonal[2].

Definition A.2. A symmetric matrix $A \in S^n$ is called positive semidefinite (this is also written by $A \in \mathcal{S}^n_+$ or $A \succeq 0$) if

$$x^T Ax \geq 0, \quad \text{for all } x \in \mathbb{R}^n.$$

$A \in S^n$ is called positive definite (also written by $A \in \mathcal{S}^n_{++}$ or $A \succ 0$) if

$$x^T Ax > 0, \quad \text{for all } x \in \mathbb{R}^n \backslash \{0\}.$$

Corollary A.3. *For positive (semi-) definite matrices the following statements hold true:*

[2]This implies that any symmetric matrix can be diagonalized by an orthonormal matrix.

(i) Any principal submatrix of a positive (semi-) definite matrix is again positive (semi-) definite. In particular, all diagonal elements of a positive definite matrix must be positive.

(ii) For $A \in \mathcal{S}^n_+$ there is always a diagonal element a_{ii} among the elements of largest absolute value, i.e.,

$$\exists i \in \{1, ..., n\} : a_{ii} = \max\Big\{|a_{ij}| : i, j \in \{1, ..., n\}\Big\}.$$

(iii) If $A \in \mathcal{S}^n_+$ and $a_{ii} = 0$ for some $i \in \{1, ..., n\}$, then,

$$a_{ij} = 0, \qquad \text{for all } j \in \{1, ..., n\}.$$

Proof. Follows immediately from Definition A.2. □

The next theorem gives a characterization of the positive (semi-) definiteness of a matrix via the positive (semi-) definiteness of the so-called Schur complement[3].

Theorem A.4. ***(Schur Complement)*** *Let $A \in \mathcal{S}^m_{++}, C \in \mathcal{S}^n$, and $B \in \mathbb{R}^{m\times n}$. Then, the following equivalences hold:*

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succ 0 \quad \Leftrightarrow \quad C - B^T A^{-1} B \ \succ\ 0,$$

and

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \quad \Leftrightarrow \quad C - B^T A^{-1} B \ \succeq\ 0.$$

Proof. See [49]. □

Theorem A.4 can be used for recognizing and factorizing positive (semi-) definite matrices. The latter becomes significant in modeling semidefinite programming problems.

A sufficient condition for a given matrix to be positive definite is due to the concept of diagonal dominance.

[3]The Schur complement is named after Issai Schur who used it to prove Schur's lemma.

Definition A.5. (Diagonal Dominance) A matrix $A \in \mathbb{R}^{n \times n}$ is strictly diagonally dominant if

$$|a_{ii}| > \sum_{i \neq j} |a_{ij}|, \quad \text{for } i = 1, ..., n.$$

Remark A.6. For strictly diagonally dominant matrices whose diagonal entries are all positive, it can be shown that positive definiteness follows. Note that diagonal dominance is easy to check. Therefore, in the context of semidefinite programming, it is often used to generate initial feasible points.

Lemma A.7. *For a given symmetric matrix $A \in \mathcal{S}^n$ the following statements are equivalent:*

(i) A is positive semidefinite.

(ii) All eigenvalues of A are greater or equal to zero.

(iii) $\exists C \in \mathbb{R}^{m \times n}$ such that $A = C^T C$. For any such C, rank(C) =rank(A).

Proof. See [31]. □

The fact that a positive semidefinite matrix may be written in terms of $C^T C$ is crucial in many proofs. An interpretation of this factorization is to view the columns of C as vectors v_i. Each element a_{ij} of A then is the scalar product $\langle v_i, v_j \rangle$ of the vectors v_i and v_j. In this context, the matrix A is referred to as the Gram matrix of the vectors $v_1, ..., v_n$. This factorization is not unique – there are several algorithmic possibilities to construct one.

Checking Positive Semidefiniteness: For a positive semidefinite matrix any leading principal minor is non-negative. However, the latter is not sufficient for being positive semidefinite[4]. Namely, a symmetric matrix is positive semidefinite if and only if the determinant of every (not just leading) principal submatrix is non-negative (see Chapter 7 of [140]).

Furthermore, due to Schur's complement the following recursive test can be used in order to check whether a matrix $A \in \mathcal{S}^n$ is positive semidefinite or not:

$$A_n = \begin{bmatrix} a_n & b_n^T \\ b_n & C_n \end{bmatrix} \quad \Leftrightarrow \quad \begin{cases} \text{either} & a_n > 0 \text{ and } A_{n-1} = C_n - \frac{1}{a_n} b_n b_n^T \succeq 0, \\ \text{or} & a_n = 0, b_n = 0 \text{ and } A_{n-1} = C_n \succeq 0. \end{cases}$$

[4]If all leading principal minors are positive, then the matrix is positive definite (Sylvester's criterion [139]).

This implies that the positive semidefiniteness can be checked in $n^3/3 + O(n^2)$ arithmetic operations.

Theorem A.8. ***(Cholesky Decomposition)*** *Let $A \in \mathcal{S}^n_{++}$. Then there exists a unique lower triangular matrix $L \in \mathbb{R}^{n\times n}$ such that $A = LL^T$.*

Proof. See [31]. □

In case of positive semidefinite matrices the Cholesky decomposition holds with the exception that the decomposition in general is not unique.

Remark A.9. The Cholesky decomposition is primarily used for the numerical solution of linear systems. For $A \in \mathcal{S}^n_{++}$, the equation $Ax = b$ can be solved, by first computing $A = LL^T$, then $Ly = b$, and, finally, $L^T x = y$.

A.1 The Cone $\mathcal{S}^n_+$

This section is a survey of the main properties of the set of positive semidefinite matrices $\mathcal{S}^n_+$, interpreted as a subset of $\mathcal{S}^n$.

Definition A.10. A set $C \subseteq \mathbb{R}^n$ is a cone if it is closed under non-negative multiplication and addition, i.e.,

$$x, y \in C \;\Rightarrow\; \lambda(x+y) \in C, \quad \forall \lambda \geq 0.$$

A cone C is pointed if $C \cap (-C) = \{0\}$.

Lemma A.11. *The set of symmetric and positive semidefinite matrices $\mathcal{S}^n_+$ is a full dimensional, closed pointed (convex) cone in $\mathbb{R}^{\binom{n+1}{2}}$.*

Proof. See [31]. □

Note that the set of positive definite matrices $\mathcal{S}^n_{++}$ cannot be a cone as it does not contain the null matrix. Further, it is easy to see that $\mathcal{S}^n_{++}$ is the interior of the cone $\mathcal{S}^n_+$. The boundary of $\mathcal{S}^n_+$ consists of those positive semidefinite matrices that have at least one zero eigenvalue.

Definition A.12. For a cone C, the polar cone is defined as the set

$$C^* := \Big\{ y : \langle x, y \rangle \geq 0, \; \forall x \in C \Big\}.$$

The polar cone C^* (of the cone C) can be seen as the set of tight valid linear inequalities for C (or, equivalently, as the set of tangent planes to C). For this reason, C^* is also referred to as the dual cone to C.

Lemma A.13. *Let $A, B \in \mathcal{S}_+^n$. Then, the following holds true: $A \bullet B \geq 0$. Moreover, $A \bullet B = 0$ if and only if $AB = 0$.*

Proof. See [31]. □

Lemma A.14. ***(Self-duality)*** *$\mathcal{S}_+^n = \mathcal{S}_+^{n*}$.*

Proof. See [31]. □

Cones that satisfy $C = C^*$ are called self-polar or self-dual. The above lemma (Lemma A.14) is equivalent to the following theorem.

Theorem A.15. ***(Fejer's Trace Theorem)*** *For $A \in \mathcal{S}^n$ the following relation holds: A is positive semidefinite if and only if $A \bullet B \geq 0$ for all $B \in \mathcal{S}_+^n$.*

Proof. Follows immediately from Lemma A.14. □

Definition A.16. A (convex) set $F \subset C$ is called a face of the convex set C, if for any $x, y \in C$ and $\alpha \in (0, 1)$, $\alpha x + (1 - \alpha)y \in F$, then, it follows $x, y \in F$.

Theorem A.17. *The faces of $\mathcal{S}_+^n$ are*

(i) the trivial faces $\emptyset$ and the set containing the zero matrix $\{0\}$,

(ii) or, they are generated by a rank k matrix $P \in \mathbb{R}^{n\times k}$ in the form

$$F = \left\{X : X = PWP^T,\ W \in \mathcal{S}_+^k\right\}.$$

Proof. See [31]. □

Remark A.18. In (ii) of Theorem A.17 the columns of the matrix P span a subspace of $\mathbb{R}^n$. All eigenvectors that belong to nonzero eigenvalues of matrices in this face are restricted to this subspace.

Remark A.19. According to Theorem A.17,

$$\{\lambda \cdot xx^T : \lambda \geq 0\}$$

is a face for any $x \in \mathbb{R}^n$, which can not be expressed as the convex combination of smaller faces. Consequently, the set

$$\left\{X : X = xx^T,\ x \in \mathbb{R}^n,\ \|x\| = 1\right\}$$

forms a minimal generating system for $\mathcal{S}_+^n$[5]. Hence, in contrast to polyhedral cones, $\mathcal{S}_+^n$ cannot be generated by a finite set.

Remark A.20. The faces of $\mathcal{S}_+^n$ have dimension $\binom{k+1}{2}$. Thus, going from a large face to a smaller one there is a considerable jump in dimension.

The cone of semidefinite matrices induces a partial order on the set of symmetric matrices, the so called Löwner partial order.

Definition A.21. (Löwner Partial Order) For $A, B \in \mathcal{S}^n$, the following is defined: $A \succeq B$ if $(A - B) \in \mathcal{S}_+^n$, and $A \succ B$ if $(A - B) \in \mathcal{S}_{++}^n$.

In connection with differentiability properties and convexity it often helps to transform matrices into vector form. The latter can be realized by the so called vec operator which stacks the columns of a given matrix one underneath the other.

Definition A.22. Let $A \in \mathbb{R}^{m\times n}$ and $A_{\bullet j}$ its j-th column, then

$$\text{vec}(A) := \begin{bmatrix} A_{\bullet 1}^T & \ldots & A_{\bullet n}^T \end{bmatrix}^T \in \mathbb{R}^{mn\times 1}.$$

Remark A.23. The vec operator is defined for any matrix, not just for square matrices. Note that vec(A) =vec(B) does not imply $A = B$, unless A and B are matrices of the same order.

Remark A.24. For the set $\mathcal{S}^n$, which is isomorphic to $\mathbb{R}^{\binom{n+1}{2}}$, the so called svec operator may be introduced. This operator is the symmetric analogue to the vec operator. For $A \in \mathcal{S}^n$, it is defined as follows

$$\text{svec}(A) := \left[a_{11}, \sqrt{2}a_{21}, ..., \sqrt{2}a_{n1}, a_{22}, \sqrt{2}a_{32}, ..., a_{nn}\right]^T.$$

Here, the factor $\sqrt{2}$ ensures that $\text{svec}(A)^T \cdot \text{svec}(B) = A \bullet B$, for $A, B \in \mathcal{S}^n$.

[5]Note that any $A \in \mathcal{S}_+^n$ can be written as a non-negative linear combination of n positive semidefinite matrices having rank one, i.e., $A = \sum_{i=1}^n \lambda_i x_i x_i^T$, where $\lambda_1, ..., \lambda_n \geq 0$.

A.2 Semidefinite Programming

What is Semidefinite Programming? Semidefinite programming (SDP) is an extension of linear programming. It is minimization/maximization of a linear function over intersections of affine equations with the cone of symmetric and positive semidefinite matrices. It may be interpreted as linear programming over the cone of symmetric and positive semidefinite matrices.

Compared to standard linear programming where the vector $x \in \mathbb{R}^n$ of variables is optimized over the non-negative orthant $x \geq 0$, linear semidefinite programming optimizes a matrix variable $X \in \mathcal{S}^n$ over the cone of symmetric and positive semidefinite matrices.

Why Semidefinite Programming? Semidefinite programs naturally evolve from problems whose data is given by matrices. Furthermore, semidefinite programming has been very fruitful in different fields of optimization:

- Semidefinite programming can be used to find approximate solutions in systems and control theory (see Chapter 14 of [32]).
- It provides tight relaxations for several combinatorial optimization problems (for instance, for the max cut problem [118]).
- It can be employed to approximate challenging non-convex quadratically constraint quadratic programs [117].
- Furthermore, it has diverse applications in robust optimization (see [141], for instance) and eigenvalue optimization.

Definition A.25. (Semidefinite Program in Primal Standard Form) For given matrices $C, A_1, ..., A_m \in \mathcal{S}^n$ and a given vector $b \in \mathbb{R}^m$, a semidefinite program in primal standard form reads

$$\min\{C \bullet X : A_i \bullet X = b_i,\ i = 1, ..., m,\ X \succeq 0\}. \tag{A.2}$$

Collecting the constraints $A_i \bullet X$ in the linear Operator $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ with

$$\mathcal{A}X := [A_1 \bullet X, \ldots, A_m \bullet X]^T,$$

this can be written in compact from as

$$\min\{C \bullet X : \mathcal{A}X = b,\ X \succeq 0\}.$$

Remark A.26. As for any symmetric matrix $A \in \mathcal{S}^n$ and a given skew-symmetric matrix $B \in \mathbb{R}^{n\times n}$ the following holds true

$$A \bullet B = \sum\nolimits_{i<j} a_{ij}b_{ij} + \sum\nolimits_{i>j} a_{ij}(-b_{ij}) = 0,$$

the matrices $C, A_1, ..., A_m$ in (A.2) do not need to be symmetric.

A.2.1 Duality Theory

In order to derive the dual to (A.2), it takes the adjoint operator to $\mathcal{A}$. By definition, it is the linear operator $\mathcal{A}^T : \mathbb{R}^m \to \mathcal{S}^n$ satisfying

$$(\mathcal{A}X)^T y = X \bullet \mathcal{A}^T y, \qquad \text{for all } X \in \mathcal{S}^n \text{ and } y \in \mathbb{R}^m.$$

As

$$(\mathcal{A}X)^T y = \sum\nolimits_{i=1}^m y_i(A_i \bullet X) = X \bullet \sum\nolimits_{i=1}^m y_iA_i = X \bullet \mathcal{A}^T y,$$

we obtain that

$$\mathcal{A}^T y = \sum\nolimits_{i=1}^m y_iA_i.$$

Lifting the primal equality constraints into the objective by means of a Lagrange multiplier $y \in \mathbb{R}^m$, the primal (A.2) reads

$$\inf_{X\succeq 0} \sup_{y\in\mathbb{R}^m} C \bullet X + y^T(b - \mathcal{A}X).$$

Now, the dual of (A.2) arises by interchanging inf and sup (cf. [76]). Note that the objective value of the dual program cannot exceed the objective value of the primal program (see [43]), i.e.,

$$\begin{aligned} \inf_{X\succeq 0} \sup_{y\in\mathbb{R}^m} C \bullet X + y^T(b - \mathcal{A}X) &\geq \sup_{y\in\mathbb{R}^m} \inf_{X\succeq 0} C \bullet X + y^T(b - \mathcal{A}X) \\ &= \sup_{y\in\mathbb{R}^m} \inf_{X\succeq 0} y^T b + \left(C - \mathcal{A}^T y\right) \bullet X. \end{aligned} \tag{A.3}$$

So that the supremum on the right hand side in (A.3) exceeds $-\infty$, the inner minimization over $X \succeq 0$ has to be finite for some $\widehat{y} \in \mathbb{R}^m$. According to Fejer's trace theorem (see Theorem A.15), the latter requires that $C - \mathcal{A}^T\widehat{y}$ is positive semidefinite. By introducing a slack matrix $Z = C - \mathcal{A}^T y$, $Z \succeq 0$, we obtain the following standard formulation of the dual to (A.2):

$$\max\left\{b^T y : \mathcal{A}^T y + Z = C,\ Z \succeq 0\right\}. \tag{A.4}$$

Due to (A.3), the gap between any dual feasible solution (y, Z) and a given primal feasible solution X is

$$C \bullet X - b^T y = \left(\mathcal{A}^T y + Z\right) \bullet X - (\mathcal{A}X)^T y = Z \bullet X \geq 0. \tag{A.5}$$

Here, the latter inequality follows from Lemma A.13. The property illustrated in (A.5) is called weak duality. If $Z \bullet X = 0$ (or, due to Lemma A.13, equivalently $ZX = 0$), then the primal-dual pair is optimal.

In linear programming, the existence of a primal-dual pair that satisfies primal feasibility, dual feasibility and complementarity is guaranteed if both the feasible set of the primal and the feasible set of the dual are non-empty (cf. [142]).

Duality in semidefinite programming is somewhat more complex. For instance, although the primal SDP and its dual are feasible, it may happen that their objective values do not coincide. This is demonstrated in Example A.1 (cf. [31]).

Example A.1. *Consider the following primal semidefinite program*

$$\min \left\{ \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \bullet X : \begin{array}{l} \begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \bullet X = 1, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \bullet X = 0, \\ \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \bullet X = 0, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \bullet X = 0, \quad X \succeq 0 \end{array} \right\}.$$

Dualizing the above program yields

$$\max\left\{y_1 : \quad Z = \begin{bmatrix} -y_2 & \frac{1+y_1}{2} & -y_3 \\ \frac{1+y_1}{2} & 0 & -y_4 \\ -y_3 & -y_4 & -y_1 \end{bmatrix} \succeq 0 \right\}.$$

For the primal: Since $x_{11} = 0$ and $X \in \mathcal{S}^3_+$, it is necessary that $x_{12} = 0$ as well. Therefore, the primal optimal value has to be zero.
For the dual: From $z_{22} = 0$ and $Z \in \mathcal{S}^3_+$, we obtain $z_{12} = 0$, and thus $y_1 = -1$. The latter implies that the dual optimal value is -1.

Consequently, the duality gap between any primal-dual pair is 1.

The duality gap in Example A.1 is due to the dualization procedure which ignores the actual geometry of the primal feasible set. In this example the primal feasible set is contained in a face of the semidefinite cone that has the following form[6] (cf. [31]):

$$F = \{PWP^T : W \succeq 0\} \quad \text{with } P^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

If we now replace $X \succeq 0$ by $X \in F$ in the primal, it will remain unchanged. Considering further the dual to $\inf_{X \in F} \sup_{y \in \mathbb{R}^m} C \bullet X + (b - \mathcal{A}X)^T y$, we obtain

$$\sup_{y \in \mathbb{R}^m} \inf_{X \in F} b^T y + \left(C - \mathcal{A}^T y\right) \bullet X \;=\; \max\left\{y_1 : \quad Z = \begin{bmatrix} 0 & -y_4 \\ -y_4 & -y_1 \end{bmatrix} \succeq 0 \right\}.$$

For the latter dual, the optimal solution is attained for $y_1 = 0$. Thus, the duality gap has disappeared. Here, the reduction of the primal cone to its non-redundant part has increased the freedom of the dual cone in such a way that the dual could reach the same objective value as the primal.

[6]In this particular example, this follows from Theorem A.17 and the fact that any feasible matrix $X \in \mathcal{S}^3_+$ need to have a zero eigenvalue with eigenvector $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$.

Definition A.27. A point $\overline{X} \in \{X \in \mathcal{S}_+^n : \mathcal{A}X = b\}$ is strictly feasible for the primal program (A.2) if it satisfies $\overline{X} \succ 0$. For the dual (A.4), a pair $(\overline{y}, \overline{Z}) \in \mathbb{R}^m \times \mathcal{S}_+^n$ fulfilling $\overline{Z} = C - \mathcal{A}^T\overline{y} \succeq 0$ is called strictly feasible if $\overline{Z} \succ 0$.

Remark A.28. The existence of a strictly feasible point in the primal makes sure that the primal cone is non-redundant and cannot be restricted to one of its faces while leaving the original program unchanged.

Theorem A.29. ***(Strong Duality)*** *Assume that there exists a strictly feasible solution $(\overline{y}, \overline{Z})$ for the dual program (A.4) and let*

$$\begin{aligned} p^* &= \inf\{C \bullet X : \mathcal{A}X = b,\ X \succeq 0\} \quad \textit{and} \\ d^* &= \sup\{b^T y : \mathcal{A}^T y + Z = C,\ Z \succeq 0\}. \end{aligned}$$

Then, $p^* = d^*$. If further p^* is finite, then, it is attained for some $\overline{X} \in \{X \succeq 0 : \mathcal{A}X = b\}$.

Proof. See [31]. □

Remark A.30. In semidefinite programming a certain zero duality gap cannot be guaranteed without a constraint qualification. The regularity condition assumed in Theorem A.29 (existence of a strictly feasible point) is referred to as Slater's condition. This condition is sufficient for strong duality in general convex programming (see [43]).

Remark A.31. If the considered semidefinite program is not strictly feasible, however, we know the minimal face that contains its feasible region, then, the projection onto this face will lead to zero duality gap. In this context, the required minimal cone can be generated if a point in the relative interior of the feasible set is known (see [31]).

Another approach that guarantees zero duality gap between the primal and dual pair is presented in [143]. Here, an extended dual (this dual is called the extended Lagrange-Slater dual) is introduced that yields a zero duality gap.

Corollary A.32. *Let p^* and d^* be defined as in Theorem A.29. Then, the following is valid.*

(i) If the primal program (A.2) is strictly feasible with p^* finite, then $p^* = d^*$ and this value is attained for the dual program (A.4).

(ii) *If the dual (A.4) is strictly feasible with d^* finite, then $p^* = d^*$ is attained for the primal (A.2).*

(iii) *If the primal (A.2) and the dual (A.4) are both strictly feasible, then $p^* = d^*$ is attained for both problems.*

Remark A.33. Note that the existence of a strictly feasible solution for the primal/dual (which guarantees zero duality gap) does not ensure that the infimum/-supremum of the primal/dual will be attained (see [31]).

To determine whether the feasible set of the primal program (A.2) is empty or non-empty, the following version of the semidefinite Farkas lemma is practical.

Lemma A.34. ***(Semidefinite Farkas Lemma)*** *Assume that the set $\{X \in \mathcal{S}_+^n : \mathcal{A}X = b\}$ is closed. Then either there exists an $X \in \mathcal{S}_+^n$ with $\mathcal{A}X = b$ or there exists $y \in \mathbb{R}^m$ such that $\mathcal{A}^T y \succeq 0$ and $b^T y < 0$.*

Proof. See Theorem 2.4 in [144]. □

Lemma A.35. *Assume that there exists $\overline{y} \in \mathbb{R}^m$ such that $\overline{Z} = \mathcal{A}^T \overline{y} \succ 0$. Then the set $\{X \in \mathcal{S}_+^n : \mathcal{A}X = b\}$ is closed.*

Proof. See Lemma 2 in [145]. □

An extension of Lemma A.34, which works without the closedness of $\{X \in \mathcal{S}_+^n : \mathcal{A}X = b\}$, is presented in [146]. There it is shown that the system $\{X \in \mathcal{S}_+^n : \mathcal{A}X = b\}$ is consistent if and only if

$$\sum\nolimits_{i=1}^{m} u_i A_i + \lambda I_n \succeq 0,\ \lambda \geq 0 \ \Rightarrow\ b^T u + \lambda \delta_0 \geq 0 \text{ for some } \delta_0 > 0.$$

Furthermore, in order to check if the primal has a strictly feasible solution, the subsequent lemma is helpfully.

Lemma A.36. ***(Positive Definite Farkas Lemma)*** *The set $\{X \in \mathcal{S}_{++}^n : \mathcal{A}X = b\}$ is non-empty if and only if $y^T b > 0$ holds for each and every vector $y \in \mathbb{R}^m \backslash \{0\}$ such that $\sum_{i=1}^m y_i A_i \succeq 0$.*

Proof. See [147]. □

Remark A.37. According to Lemma A.36, the spectrahedra $\{X \in \mathcal{S}_+^n : \mathcal{A}X = b\}$ has an empty interior if and only if $\min\left\{b^T y : \mathcal{A}^T y \succeq 0,\ \|y\| = 1\right\} \leq 0$.

A.3 Semidefinite Relaxations for QCQPs

Quadratically constrained quadratic programming is optimization of a quadratic objective function over constraints that are given by quadratic functions. Generally, these kind of optimization problems have the following form:

$$\min\{x^T P_0 x + p_0^T x : x^T P_i x + p_i^T x + r_i \leq 0, i = 1, ..., m\}. \tag{A.6}$$

Let us define the matrices

$$A_0 = \begin{bmatrix} 0 & \frac{1}{2}p_0^T \\ \frac{1}{2}p_0 & P_0 \end{bmatrix} \quad \text{and} \quad A_i = \begin{bmatrix} r_i & \frac{1}{2}p_i^T \\ \frac{1}{2}p_i & P_i \end{bmatrix}, \text{ for } i = 1, ..., m.$$

Then, (A.6) allows the following lifted representation (cf. [117]):

$$\min \left\{ A_0 \bullet X : A_i \bullet X \leq 0, i = 1, ..., m,\ X = \begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}^T \right\},$$

where the latter program is equivalent to the subsequent rank-constrained semidefinite program

$$\min \left\{ A_0 \bullet X : A_i \bullet X \leq 0, i = 1, ..., m,\ X_{11} = 1,\ X \succeq 0,\ \operatorname{rank}(X) = 1 \right\}. \tag{A.7}$$

Relaxing the non-convex rank constraint in (A.7) yields the following (linear) SDP

$$\min \left\{ A_0 \bullet X : A_i \bullet X \leq 0, i = 1, ..., m,\ X_{11} = 1,\ X \succeq 0 \right\}. \tag{A.8}$$

In this section we will present some conditions under which the semidefinite relaxation (A.8) is guaranteed to be tight for (A.7). Normally, as (A.6) may be non-convex, tightness of (A.8) for (A.6) cannot be expected. This is illustrated by the subsequent example.

Example A.2. *Consider the program* $\min\{-x^2 + 2x : x^2 \leq 1, x \geq 0\}$*, whose optimal value is zero attained at* $x = 0$*. Its SDP relaxation is*

$$\min_{Z \in \mathcal{S}^2_+} \left\{ \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & -1 \end{bmatrix} \bullet Z : \left(e_2 e_2^T\right) \bullet Z \leq 1, \quad \left(e_1 e_2^T + e_2 e_1^T\right) \bullet Z \geq 0, \quad Z_{11} = 1 \right\},$$

which returns an optimal value -1 *attained at*

$$\overline{Z} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

i.e., the SDP relaxation is not tight.

The SDP Relaxation is Tight for Convex QCQPs. As a convex relaxation, the SDP relaxation (A.8) at least should be tight for convex QCQPs. In fact, this holds and can be seen as follows: A quadratic function $f(x) = x^T P x + q^T x$ is convex if and only if $P \succeq 0$. Hence, (A.6) is convex if and only if the matrices $P_0, P_1, ..., P_m$ are all positive semidefinite. Let

$$X^* = \begin{bmatrix} 1 & (x^*)^T \\ x^* & Y^* \end{bmatrix}$$

be a minimizer for the SDP relaxation (A.8). Since $X^* \succeq 0$, Schur's complement (see Lemma A.4) implies $Y^* \succeq x^*(x^*)^T$. If we introduce

$$\overline{X} = \begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}^T = \begin{bmatrix} 1 & (x^*)^T \\ x^* & x^*(x^*)^T \end{bmatrix},$$

then, we have $X^* \succeq \overline{X}$. Moreover, as $P_i \succeq 0$, for $i = 0, 1, ..., m$, we obtain[7]

$$A_i \bullet (X^* - \overline{X}) \geq 0 \text{ for } i = 0, 1, ..., m.$$

[7] See Lemma A.13.

In other words, $\overline{X}$ is feasible for (A.8) and produces the same objective value as the minimizer X^*[8]. As $\overline{X}$ has rank-one, it is feasible and thus optimal for (A.7). Hence, the semidefinite relaxations (A.8) are indeed tight for convex QCQPs.

Remark A.38. Minimum attainment is not a necessary condition for zero duality gap. If QCQP is convex, then certainly its SDP relaxation admits zero duality gap, however, attainment of its minimum is not guaranteed.

The SDP Relaxation for Non-convex QCQPs. Unless all P_i are positive semidefinite, the QCQP (A.6) is non-convex. In such case, computing the global optima is difficult. However, one way to approximate this value is provided by the SDP relaxation (A.8). If, moreover, the set of minimizer of (A.8) include a rank-one solution, then the SDP relaxation even provides the global optima.

To decide whether (A.8) has a rank-one solution or not, is equivalent to the issue of finding the points with lowest rank in a spectrahedra. A general result for finding low rank solutions of linear matrix inequalities is due to Pataki (cf. [148] and [149]).

Theorem A.39. *If $\mathcal{B} \subseteq \mathcal{S}^n$ is an affine subspace such that $\mathcal{S}^n_+ \cap \mathcal{B} \neq \emptyset$ and*

$$dim(\mathcal{B}) \geq \binom{n+1}{2} - \binom{r+2}{2} + 1,$$

then there is a matrix $X \in \mathcal{S}^n_+ \cap \mathcal{B}$ such that $rank(X) \leq r$.

Proof. See Theorem 6.1 in [150]. □

Remark A.40. The bound in Theorem A.39 is sharp. This means that if $r < n$, then one can find a subspace $\mathcal{B} \subseteq \mathcal{S}^n$ such that $\mathcal{S}^n_+ \cap \mathcal{B} \neq \emptyset$, $\dim(\mathcal{B}) = \binom{n+1}{2} - \binom{r+2}{2}$ and for every matrix $X \in \mathcal{S}^n_+ \cap \mathcal{B}$ it holds $\operatorname{rank}(X) > r$ (see [150]).

Under boundedness of $\mathcal{S}^n_+ \cap \mathcal{B}$, the result of Theorem A.39 can be generalized as follows (cf. [151]).

Theorem A.41. *Let $r > 0$, let $n \geq r+2$, and let $\mathcal{B} \subseteq \mathcal{S}^n$ be an affine subspace such that the intersection $\mathcal{S}^n_+ \cap \mathcal{B}$ is non-empty, bounded, and $dim(\mathcal{B}) = \binom{n+1}{2} - \binom{r+2}{2}$. Then there is a matrix $X \in \mathcal{S}^n_+ \cap \mathcal{B}$ such that $rank(X) \leq r$.*

[8] Feasibility follows from $A_i \bullet \overline{X} \leq A_i \bullet X^* \leq 0$, $i = 1, ..., m$ and optimality from $A_0 \bullet \overline{X} \leq A_0 \bullet X^*$.

Proof. See Section 3 in [151]. □

Corollary A.42. *Consider a semidefinite program of dimension $n \geq 3$ having less or equal to 2 affine linear constraints, i.e.,*

$$\min\{C \bullet X : A_i \bullet X = b_i, i = 1, 2,\ X \in \mathcal{S}_+^n\}. \tag{A.9}$$

If there exists a Slater point in the dual to (A.9), then, (A.9) has a rank-one solution.

Proof. The existence of a Slater point guarantees that (A.9) is bounded and that its minimum is attained for some $\overline{X} \in \{X \in \mathcal{S}_+^n : A_i \bullet X = b_i, i = 1, 2\}$ (see Theorem A.29). Further, if we consider the affine subspace

$$\mathcal{B} = \{X \in \mathcal{S}^n : A_1 \bullet X = b_1,\ A_2 \bullet X = b_2,\ C \bullet X = \overline{b}\},$$

where $\overline{b} := C \bullet \overline{X}$, then $\dim(\mathcal{B}) = \dim(\mathcal{S}^n) - 3 = \binom{n+1}{2} - \binom{r+2}{2}$.

In order to apply Theorem A.41 it remains to show that $\mathcal{S}_+^n \cap \mathcal{B}$ is bounded. From convex analysis (cf. [43]) it is well known that a convex set is bounded if and only if it has a trivial recession cone. This implies that $\mathcal{S}_+^n \cap \mathcal{B}$ is bounded if and only if

$$\operatorname{rec}(\mathcal{S}_+^n \cap \mathcal{B}) = \{X \in \mathcal{S}_+^n : A_1 \bullet X = 0,\ A_2 \bullet X = 0,\ C \bullet X = 0\} = \{0\}.$$

Let $Y \in \operatorname{rec}(\mathcal{S}_+^n \cap \mathcal{B})$ and $Z = C - A_1 u_1 - A_2 u_2 \succ 0$ the assumed Slater point, then,

$$Z \bullet Y = -u_1(A_1 \bullet Y) - u_2(A_2 \bullet Y) + C \bullet Y = 0. \tag{A.10}$$

Writing further the positive semidefinite matrix Y as $\sum_{i=1}^n \lambda_i v_i v_i^T$, it can be deduced from (A.10) and the positive definiteness of Z that all eigenvalues of Y must be equal to zero. Indeed,

$$0 = Z \bullet Y = Z \bullet \left(\sum\nolimits_{i=1}^n \lambda_i v_i v_i^T\right) = \sum\nolimits_{i=1}^n \underbrace{\lambda_i}_{\geq 0} \underbrace{v_i^T Z v_i}_{>0} \quad \Rightarrow \quad \lambda_i = 0, i = 1, ..., n.$$

Hence, according to Theorem A.41, the set of minimizer $\mathcal{B}$ contains a matrix that has rank-one. □

Appendix B

Selected Facts of Convex Analysis and Probability Theory

Definition B.1. (Convex Set) A set $A \subseteq \mathbb{R}^n$ is said to be convex, if

$$\forall x, y \in A,\ \lambda \in (0,1): \quad (1-\lambda)x + \lambda y \in A.$$

Definition B.2. (Convex Function) A function $f : \mathbb{R}^n \supseteq S \to \mathbb{R} \cup \{\pm\infty\}$ is said to be convex, if its epigraph

$$\text{epi} f := \{(x, \alpha) \in S \times \mathbb{R} : f(x) \leq \alpha\}$$

is convex.

Definition B.3. (Subgradient) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued convex function. A vector $g \in \mathbb{R}^n$ is called a subgradient of f at a point x_0 if for any $x \in \mathbb{R}^n$ one has

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle.$$

The set of all subgradients at x_0 is called the subdifferential[1] of f at x_0 and is denoted by $\partial f(x_0)$.

Remark B.4. The subdifferential is always a non-empty convex compact set.

[1]The subdifferential on convex functions was introduced by Moreau and Rockafellar.

Theorem B.5. ***(Moreau-Rockafellar Theorem)*** *Let $f_1, ..., f_n$ be real-valued convex functions on $\mathbb{R}^n$. Then*

$$\partial f_1(x) + ... + \partial f_n(x) \subset \partial(f_1 + ... + f_n)(x)$$

for every $x \in \mathbb{R}^n$. If all functions $f_1, ..., f_n$, except possibly one, are continuous at a point x_0, then

$$\partial f_1(x) + ... + \partial f_n(x) = \partial(f_1 + ... + f_n)(x)$$

for all $x \in \mathbb{R}^n$.

Proof. See [59]. □

Definition B.6. (Measure Space) A measure space is a triple $(\Omega, \mathcal{F}, \mu)$, in which the three components are:

(1) **(Set)** A set Ω.

(2) **(Sigma-algebra)** A σ-algebra $\mathcal{F}$ over Ω, i.e., a collection of subsets of Ω that includes the empty set, is closed under complement, and closed under union (or intersection) of countably many subsets.

(3) **(Measure)** A function (measure) $\mu : \mathcal{F} \to \mathbb{R} \cup \{\pm\infty\}$ that is

- non-negative, i.e., $\mu(A) \geq 0,\ \forall A \in \mathcal{F}$,
- countable additive, i.e., for all countable collections $\{A_i\}_{i=1}^{\infty}$ of pairwise disjoint sets in $\mathcal{F} : \mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$,
- and for which $\mu(\emptyset) = 0$.

Definition B.7. (Measurable Function) If $(\Omega_x, \mathcal{F}_x)$ and $(\Omega_y, \mathcal{F}_y)$ are two measurable spaces, then a function $f : \Omega_x \to \Omega_y$ is called measurable if for every Ω_Y-measurable set $B \in \mathcal{F}_Y$, the inverse image is Ω_X-measurable, i.e., $f^{(-1)}(B) \in \mathcal{F}_X$.

A probability measure is a measure with total measure one, i.e., $\mu(\Omega) = 1$. If the measure μ is a probability measure, the triple (Ω, Σ, μ) is called probability space. A (real-valued) random variable X is a real-valued measurable function on a probability space, i.e., $X : \Omega \to \mathbb{R}$. The random variable is discrete, if it takes on only a countable number of distinct values.

Lemma B.8. ***(Fatou's Lemma)*** *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of non-negative measurable functions on a measure space $(\Omega, \mathcal{F}, \mu)$. Define the function $f : \Omega \to [0, \infty]$ almost everywhere pointwise limit by*

$$f(x) = \liminf_{n\to\infty} f_n(x), \quad x \in \Omega.$$

Then f is measurable and

$$\int_\Omega f d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

Proof. See [59]. □

Lemma B.9. ***(Lebesgue's Dominated Convergence Theorem)*** *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of real-valued measurable functions on a measure space $(\Omega, \mathcal{F}, \mu)$. Suppose that the sequence converges pointwise almost everywhere to a function f and is dominated by some integrable function g in the sense that*

$$|f_n(x)| \leq g(x)$$

for all $n \in \mathbb{N}$ and all $x \in \Omega$. Then f is integrable and

$$\lim_{n\to\infty} \int_\Omega |f_n - f| d\mu = 0.$$

Proof. See [152]. □

Appendix C

Convex Relaxations for AC Power Flow

All approaches that were proposed to convexify AC power flow (4.11)-(4.20) are based on the same idea. They all relax the non-convex voltage products

$$\boldsymbol{V}_l^t(\boldsymbol{V}_m^t)^* = (\mathbf{E}_l^t + j\mathbf{F}_l^t) \cdot (\mathbf{E}_m^t - j\mathbf{F}_m^t), \tag{C.1}$$

and solely differ in the way they convexify them.

As the non-convexities in (4.11)-(4.20) arise exclusively from the products (C.1), they can be segregated by substituting new variables

$$\boldsymbol{W}_{lm}^t = \boldsymbol{V}_l^t(\boldsymbol{V}_m^t)^*. \tag{C.2}$$

In doing so, the only source of non-convexity then lies in the latter constraints. Now, the DC [103], SOCP [107], QC [110], and SDP [3] approach vary in the way they convexify (C.2). This chapter gives a brief overview of the different ideas.

C.1 The DC Approximation

The DC approximation [153], [154], [155], [156] of AC power flow comes from a number of approximations of the AC equations (4.11)-(4.20). These approximations result from several assumptions which under normal system operating conditions could be made. If these assumptions hold true, then the non-convex energy flows along lines, (4.15) and (4.16), can be linearized quite accurately. The latter leads to a great simplification of the underlying optimization problem, making the DC approximation very popular in practice.

Assume that

(1) for each transmission line $(l,m) \in \mathcal{L}$, the susceptance b_{lm} is large relative to the conductance g_{lm}, i.e. $|g_{lm}| \ll |b_{lm}|$;

(2) the phase angle difference is small, i.e. $\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_m^t \approx 0$, $\forall (l,m) \in \mathcal{L}$;

(3) the voltage magnitudes $|\boldsymbol{V}_k^t|, k \in \mathcal{N}$ are close to 1 and do not vary significantly;

(4) and reactive power flow can be neglected.

Then, the following holds: Due to

$$\mathbf{E}_l^t\mathbf{E}_m^t + \mathbf{F}_l^t\mathbf{F}_m^t = \Re\left((\mathbf{E}_l^t + j\mathbf{F}_l^t)(\mathbf{E}_m^t - j\mathbf{F}_m^t)\right) = |\boldsymbol{V}_l^t||\boldsymbol{V}_m^t|\cos\left(\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_m^t\right), \quad \text{(C.3)}$$
$$\mathbf{F}_l^t\mathbf{E}_m^t - \mathbf{F}_m^t\mathbf{E}_l^t = \Im\left((\mathbf{E}_l^t + j\mathbf{F}_l^t)(\mathbf{E}_m^t - j\mathbf{F}_m^t)\right) = |\boldsymbol{V}_l^t||\boldsymbol{V}_m^t|\sin\left(\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_m^t\right), \quad \text{(C.4)}$$

the right-hand side in equation (4.15) can be equivalently expressed as

$$g_{lm}|\boldsymbol{V}_l^t|^2 - g_{lm}|\boldsymbol{V}_l^t||\boldsymbol{V}_m^t|\cos\left(\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_m^t\right) - b_{lm}|\boldsymbol{V}_l^t||\boldsymbol{V}_m^t|\sin\left(\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_m^t\right). \quad \text{(C.5)}$$

If (2) and (3) are assumed, then,

$$|\boldsymbol{V}_l^t|^2 \approx |\boldsymbol{V}_l^t| \cdot |\boldsymbol{V}_m^t| \cdot \cos\left(\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_m^t\right).$$

In case (1) holds true as well, the b_{lm} term highly dominates in (C.5). As a consequence the g_{lm} terms largely cancel out. Moreover, since sine is almost linear around the origin, assumption (2) implies $\sin\left(\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_m^t\right) \approx \boldsymbol{\theta}_l^t - \boldsymbol{\theta}_m^t$. Hence, in assuming (1)-(3), real power on transmission lines can be approximated by

$$p_{lm} \approx -b_{lm}\left(\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_m^t\right). \quad \text{(C.6)}$$

If, further, reactive power flow is neglected (4), then, AC power flow (4.11)-(4.16) may be approximated by the following DC equations

$$\mathbf{p}_{G_k}^t + \sum\nolimits_{l\in\mathcal{N}(k)} b_{lk}\left(\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_k^t\right) = p_{D_k}^t, \qquad \forall k \in \mathcal{G}, \qquad t = 1, ..., T, \tag{C.7}$$

$$\sum\nolimits_{l\in\mathcal{N}(k)} b_{lk}\left(\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_k^t\right) = p_{D_k}^t, \qquad \forall k \in \mathcal{N}\backslash\mathcal{G}, \quad t = 1, ..., T. \tag{C.8}$$

Furthermore, the network limitations (4.17)-(4.20) reduces to

$$|b_{lk}\left(\boldsymbol{\theta}_l^t - \boldsymbol{\theta}_k^t\right)| \leq P_{lm}^{\max}. \tag{C.9}$$

Remark C.1. Assumption (1) implies that all system branch resistances are negligible. This means that the transmission system is lossless. The latter may hold in geographically small networks where losses are marginal. However, it fails in large transmission system in which network buses are widely dispersed [155]. Moreover, due to (2) and (3), the DC approximation may be too imprecise for power systems having regions with large voltage differences.

Remark C.2. Outside the normal operating conditions (1)-(4), the accuracy of the DC approximation is not valid and an open point of discussion [155], [157], [158].

For the active power part in (4.11)-(4.16), the DC approximation provides a reasonably accurate approximation [155]. Moreover, as (C.7)-(C.9) are mixed-integer linear constraints, their embedding into mixed-integer programming models[1] has become very attractive which is mainly because the computational efficiency of LP and MIP solvers has significantly improved over the last two decades.

Remark C.3. Approximating the AC equations (4.11)-(4.20) by DC power flow (C.7)-(C.9) greatly simplifies the unit commitment problem (4.21). All constraints become mixed-integer linear and thus the hole program mixed-integer (convex) quadratic.

C.2 The Second Order Cone Relaxation

The second order cone relaxation [107] convexifies each of the non-convex constraints (C.2) separately. It takes the square of the absolute value of (C.2), changes

[1]Thus for a variety of optimization applications in power system operations.

the order of multiplication, and then relaxes the resulting equality constraint into an inequality one, i.e.,

$$\begin{aligned}
\boldsymbol{W}^t_{lm} &= \boldsymbol{V}^t_l(\boldsymbol{V}^t_m)^* \\
\Rightarrow \quad \boldsymbol{W}^t_{lm}(\boldsymbol{W}^t_{lm})^* &= \boldsymbol{V}^t_l(\boldsymbol{V}^t_m)^*(\boldsymbol{V}^t_l)^*\boldsymbol{V}^t_m \\
\Leftrightarrow \quad |\boldsymbol{W}^t_{lm}|^2 &= \boldsymbol{W}^t_{ll}\boldsymbol{W}^t_{mm} \\
\Rightarrow \quad |\boldsymbol{W}^t_{lm}|^2 &\leq \boldsymbol{W}^t_{ll}\boldsymbol{W}^t_{mm}.
\end{aligned} \tag{C.10}$$

Since the latter inequality can be equivalently expressed as

$$\left\| \begin{bmatrix} 2 \cdot \boldsymbol{W}^t_{lm} \\ \boldsymbol{W}^t_{ll} - \boldsymbol{W}^t_{mm} \end{bmatrix} \right\|_2 \leq \boldsymbol{W}^t_{ll} + \boldsymbol{W}^t_{mm}, \tag{C.11}$$

it is a rotated second-order cone constraint [49], [159].

The above relaxation is shown to be tight for load flow problems in radial distribution systems (cf. [107]). However, it may fail to hold for meshed networks.

Remark C.4. Applying the above SOCP approach to relax AC power flow in (4.21) would yield a mixed-integer SOCP problem. For these programs, powerful solvers are available (for instance, CLPEX and Gurobi [160] support SOCP problems).

C.3 The Quadratic Convex Relaxation

In order to obtain a convex relaxation of (4.11)-(4.20), the quadratic convex (QC) relaxation[2] [109] uses convex envelopes of the non-convex functions in (C.2). This approach is known as McCormick's relaxation [161].

Instead of expressing the voltage variables in rectangular coordinates, the QC relaxation works with its representation in polar form $\boldsymbol{V}^t_k = \boldsymbol{U}^t_k e^{j\theta_k}, k \in \mathcal{N}$. Using

[2]The QC relaxation was introduced to have stronger links between the voltage variables.

this notation, (C.2) may be written as follows (cf. (C.3) and (C.4)):

$$\begin{aligned}
\boldsymbol{W}^t_{ll} &= (\boldsymbol{U}^t_l)^2, \\
\boldsymbol{W}^t_{mm} &= (\boldsymbol{U}^t_m)^2, \\
\Re(\boldsymbol{W}^t_{lm}) &= \boldsymbol{U}^t_l \boldsymbol{U}^t_m \cos(\boldsymbol{\theta}^t_l - \boldsymbol{\theta}^t_m), \\
\Im(\boldsymbol{W}^t_{lm}) &= \boldsymbol{U}^t_l \boldsymbol{U}^t_m \sin(\boldsymbol{\theta}^t_l - \boldsymbol{\theta}^t_m).
\end{aligned}$$

Now, the QC relaxation relaxes the equations above by taking tight convex envelopes of their nonlinear terms (square, product of two variables, sine, and cosine) as it exploits the operational limits for $\boldsymbol{U}^t_l, \boldsymbol{U}^t_m, \boldsymbol{\theta}^t_l, \boldsymbol{\theta}^t_m$. This yields the following convex relaxation

$$\begin{aligned}
\boldsymbol{W}^t_{kk} &= \left\langle (\boldsymbol{U}^t_k)^2 \right\rangle^T, & \forall k \in \mathcal{N}, \\
\Re(\boldsymbol{W}^t_{lm}) &= \left\langle \left\langle \boldsymbol{U}^t_l \boldsymbol{U}^t_m \right\rangle^M \left\langle \cos(\boldsymbol{\theta}^t_l - \boldsymbol{\theta}^t_m) \right\rangle^C \right\rangle^M, & \forall (l,m) \in \mathcal{L}, \\
\Im(\boldsymbol{W}^t_{lm}) &= \left\langle \left\langle \boldsymbol{U}^t_l \boldsymbol{U}^t_m \right\rangle^M \left\langle \sin(\boldsymbol{\theta}^t_l - \boldsymbol{\theta}^t_m) \right\rangle^S \right\rangle^M, & \forall (l,m) \in \mathcal{L},
\end{aligned}$$

where $\langle x^2 \rangle^T$, $\langle \sin(x) \rangle^S$, $\langle \cos(x) \rangle^C$, and $\langle xy \rangle^M$ are the tight convex envelopes of the functions square, sine, cosine, and the product of two variables, respectively. We refer to [109] and [110] for details.

C.4 The Semidefinite Programming Relaxation

The SDP approach [3] relaxes the non-convex constraints (C.2) all at the same time by introducing the matrix variable $\boldsymbol{W}^t = \boldsymbol{V}^t(\boldsymbol{V}^t)^*$. Since a complex matrix $A \in \mathbb{C}^{n \times n}$ can be written as zz^*, for some $z \in \mathbb{C}$, if and only if it is Hermitian[3], positive semidefinite and has rank one, the following equivalence holds true

$$\boldsymbol{W}^t = \boldsymbol{V}^t(\boldsymbol{V}^t)^* \quad \Leftrightarrow \quad \boldsymbol{W}^t \in \mathbb{H}^n_+ \text{ and } \operatorname{rank}(\boldsymbol{W}^t) = 1. \tag{C.12}$$

[3]A square matrix with complex entries is Hermitian if it is equal to its own conjugate transpose.

This implies that $\boldsymbol{W}^t = \boldsymbol{V}^t(\boldsymbol{V}^t)^*$ can be replaced by requiring that $\boldsymbol{W}^t$ is Hermitian, positive semidefinite and has rank one. In doing so, the non-convexity is isolated in the rank constraint. Relaxing this non-convex rank-constraint yields a convex linear semidefinite program, which under certain conditions is shown to be tight [121]. For instance, the SDP relaxation is tight for tree networks [123], and for cyclic networks after the angle constraints are relaxed [124]. Moreover, it is shown to be tight (cf. [121]), if each circle contains at least one controllable phase shifter and load over-satisfaction is allowed at each network bus[4]. The latter means that the power balance equations (4.11)-(4.14) may be modified as follows

$$\begin{aligned}
\mathbf{p}^t_{G_k} - \sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{p}^t_{kl} &\geq p^t_{D_k}, \qquad \forall k \in \mathcal{G}, \qquad t = 1, ..., T,\\
\mathbf{q}^t_{G_k} - \sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{q}^t_{kl} &\geq q^t_{D_k}, \qquad \forall k \in \mathcal{G}, \qquad t = 1, ..., T,\\
- \sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{p}^t_{kl} &\geq p^t_{D_k}, \qquad \forall k \in \mathcal{N} \backslash \mathcal{G}, \quad t = 1, ..., T,\\
- \sum\nolimits_{l \in \mathcal{N}(k)} \mathbf{q}^t_{kl} &\geq q^t_{D_k}, \qquad \forall k \in \mathcal{N} \backslash \mathcal{G}. \quad t = 1, ..., T.
\end{aligned}$$

Remark C.5. The SDP relaxation admits active/reactive powers to disappear, which may be beneficial in stressed network situations and therefore could result in a duality gap.

Remark C.6. The introduction of the real vector

$$\boldsymbol{v}^t := \left[\Re\left(\boldsymbol{V}^t\right)^T \ \Im\left(\boldsymbol{V}^t\right)^T\right]^T = \left[\boldsymbol{E}^T \ \boldsymbol{F}^T\right]^T \in \mathbb{R}^{2n}$$

admits to operate with real matrices instead of complex ones. In doing so, with suitable matrices $A_{lm}, B_{lm} \in \mathcal{S}^{2n}$,

$$\begin{aligned}
\boldsymbol{W}^t_{lm} = \boldsymbol{V}^t_l(\boldsymbol{V}^t_m)^* &= \left(\mathbf{E}^t_l\mathbf{E}^t_m + \mathbf{F}^t_l\mathbf{F}^t_m\right) + j\left(\mathbf{F}^t_l\mathbf{E}^t_m - \mathbf{E}^t_l\mathbf{F}^t_m\right)\\
&= A_{lm} \bullet \boldsymbol{v}^t\left(\boldsymbol{v}^t\right)^T + jB_{lm} \bullet \boldsymbol{v}^t\left(\boldsymbol{v}^t\right)^T,
\end{aligned}$$

[4]The main idea of load over-satisfaction is the following. It is expected that whenever a power network operates under a normal condition, then, the solution of the underlying power flow problem remains intact or changes insignificantly under this assumption.

such that (C.2) is equivalent to $\boldsymbol{W}^t \in \mathcal{S}_+^{2n}$, $\mathrm{rank}(\boldsymbol{W}^t) = 1$ and

$$\begin{aligned} \boldsymbol{W}^t_{kk} &= A_{kk} \bullet \boldsymbol{W}^t, \\ \Re\left(\boldsymbol{W}^t_{lm}\right) &= A_{lm} \bullet \boldsymbol{W}^t, \\ \Im\left(\boldsymbol{W}^t_{lm}\right) &= B_{lm} \bullet \boldsymbol{W}^t. \end{aligned}$$

We refer to the tutorials [162] and [163] for the main theoretical results on SDP relaxations of OPF developed in the last few years. Here, [162] discusses the relationship between different convex relaxations (DC, SOCP, and SDP) of OPF and [163] presents sufficient conditions guaranteeing that these relaxations are exact.

C.5 Some Remarks

The relationship between the SOCP relaxation and SDP relaxation is investigated in [162] and [163]. Moreover, for the QC relaxation it can be shown that it is tighter than the SOCP relaxation and neither dominates nor is dominated by the SDP relaxation (cf. [110]).

The diverse relaxations presented in Appendix C vary considerably in their computational effort. As SDP solvers are less mature than SOCP solvers and, in addition, the SOCP relaxation has clearly fewer variables[5] (compared to the SDP relaxation), it is obvious that the SOCP relaxation can be solved much more faster and more reliable. Furthermore, from the computational prospective, it is shown in [110] that the QC relaxation (which adds additional constraints to the SOCP formulation) compares favorably well with the SOCP relaxation. There, it has also been shown that the SDP relaxation provides the tightest bounds among the mentioned relaxations (it often yields duality gaps below one percent).

The following is confirmed in [110]: In case the phase angle difference bounds are tight, the QC relaxation brings significant benefits in accuracy over the SOCP relaxation. Moreover, their solutions are extremely close to the SDP solutions and

[5]Due to the sparsity of AC power networks, this size reduction can lead to significant memory and computational savings.

it may happen that they are even better. As their solution time is much faster, the QC relaxation is a computational attractive alternative to the SDP approach.

In this context, we mention that the major drawback of the SDP relaxation are the vast number of variables needed. For realistic network settings with a huge number of buses, this approach yields a very high-dimensional SDP. This issue is addressed in [164]. There, the sparsity[6] of the large-scale SDPs is exploited. Implemented solvers that are based on this idea are presented in [125] and [126].

[6]For real world power systems, the sparsity results from the sparsity of the net corresponding graph.

Appendix D

Data

D.1 IEEE Test Systems and Modifications

D.1.1 Thermal Generators

Table D.1 and Table D.2 display the generator data for the IEEE 14 bus test system[1] bus and the IEEE 30 bus test system[2], respectively. Starting from the left the columns indicate for each thermal unit: bus type, minimum active power bound, maximum active power bound, minimum reactive power bound, and maximum reactive power bound. Further, the remaining columns list the start-up costs, the linear active power production costs, the nonlinear active power production costs, and the minimum downtimes.

TABLE D.1: Thermal generator data for the IEEE 14 bus test system.

Bus No.	Type	P^{min} (MW)	P^{max} (MW)	Q^{min} (MVAR)	Q^{max} (MVAR)	c_0	c_1	c_2	Minimum Downtime
1	3	25	250	0	10	46136.25	20	0.04303	2
2	2	14	140	-40	50	46200.00	20	0.25000	2
3	2	10	100	0	40	24600.00	40	0.01000	2
6	2	12	120	-6	24	29664.00	40	0.01000	2
8	2	4.5	45	-6	24	10921.50	40	0.01000	2

[1]The IEEE 14 bus test case represents a portion of the American Electric Power System (in the Midwestern US) as of February, 1962.

[2]The IEEE 30 bus test case represents a portion of the American Electric Power System (in the Midwestern US) as of December, 1961.

Here, the IEEE data has been modified as follows: In contrast to the original data, we now impose nonzero minimum active power production bounds, nonzero start-up costs and minimum downtimes for each generator. In doing so, for each generator, the minimum active power production bound comply with 10% of the maximum active power the generator can deliver to the network. The start-up costs for each thermal unit are calculated as follows: $6 \cdot (c_1 P^{max} + c_2 (P^{max})^2)$. This equals the costs of full use of the existing capacity over six time periods.

Table D.2: Thermal generator data for the IEEE 30 bus test system.

Bus No.	Type	P^{min}	P^{max}	Q^{min}	Q^{max}	c_0	c_1	c_2	Min Downtime
1	3	8	80	-20	150	1728.000	2	0.0200	2
2	2	8	80	-20	60	1512.000	1.75	0.0175	2
13	2	4	40	-15	44.7	960.000	3	0.0250	2
22	2	5	50	-15	62.5	1237.500	1	0.0625	2
23	2	3	30	-10	40	675.000	3	0.0250	2
29	2	5.5	55	-15	48.7	1223.145	3.25	0.0083	2

D.1.2 Pumped-storage Hydroelectricity Plants

In some of our test instances, PSH plants have been added. For the IEEE 14 bus test system and the IEEE 30 bus test system, this data is depicted in Table D.3 and Table D.4, respectively.

Table D.3: PSH data for the IEEE 14 bus test system.

Bus No.	Type	P^{min}	P^{max}	Q^{min}	Q^{max}	Max Fill (MW)	Initial Fill (MW)	Final Fill (MW)	Pumping Eff. Factor
10	4	-30	30	-30	30	300	100	100	0.5
12	4	-30	30	-30	30	300	100	100	0.5

Table D.3 and Table D.4 are structured as follows: Column one refers to the bus number through which the PSH is connected to the grid. In column 2 to 6 we have: bus type (4 refers to a PSH bus), minimum active power consumption, maximum active power production, minimum reactive power consumption, and maximum reactive power production. Here, the minimum active/reactive power corresponds to the active/reactive power that can be consumed by the installed PSH (in pumping mode). Further, Column 7 to 9 specify the maximum fill, the initial fill and the final fill in active power of the corresponding upper dam. Finally, the last column displays the pumping efficiency factor as introduced in (4.8).

Table D.4: PSH data for the IEEE 30 bus test system.

Bus No.	Type	P^{min}	P^{max}	Q^{min}	Q^{max}	Max Fill (MW)	Initial Fill (MW)	Final Fill (MW)	Pumping Eff. Factor
10	4	-30	30	-30	30	300	100	100	0.5
20	4	-30	30	-30	30	300	100	100	0.5

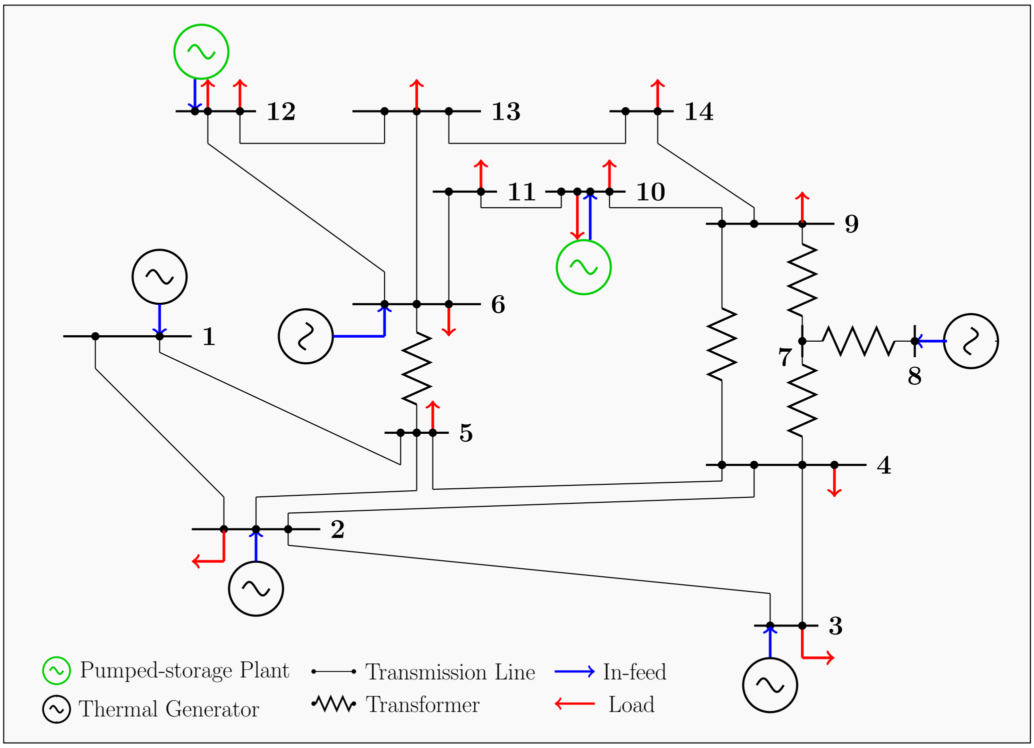


Figure D.1: Sketch of the IEEE 14 bus test system with added PSH plants.

D.1.3 Bus and Line Data

The bus and line data corresponds to the data that is provided by the power systems test case archive [111]. With the exception that we have used the given load data to generate a bunch of random load profiles, this data has not been changed in our test instances.

D.1.3.1 Bus Data

Table D.5: Bus data for the IEEE 14 bus test system.

Bus No.	Type	Load		Voltage Magnitude Bounds	
		Active Load	Reactive Load	Min Voltage Magnitude (p.u.)	Max Voltage Magnitude (p.u.)
1	3	0	0	0.94	1.06
2	2	21.7	12.7	0.94	1.06
3	2	94.2	19.1	0.94	1.06
4	0	47.8	-3.9	0.94	1.06
5	0	7.6	1.6	0.94	1.06
6	2	11.2	7.5	0.94	1.06
7	0	0	0	0.94	1.06
8	2	0	0	0.94	1.06
9	0	29.5	16.6	0.94	1.06
10	0	9.0	5.8	0.94	1.06
11	0	3.5	1.8	0.94	1.06
12	0	6.1	1.6	0.94	1.06
13	0	13.8	5.8	0.94	1.06
14	0	14.9	5.0	0.94	1.06

Table D.5 and Table D.6, which are structurally identical, illustrate the bus data of the IEEE 14 bus test system and the IEEE 30 bus test system, respectively. These tables have six columns which are as follows (from left to right): bus number, bus type (where, 3 = slack bus, 2 = thermal generator, and 0 = load), active load (in MW), reactive load (in MVAR), minimum voltage magnitude, and maximum voltage magnitude.

Table D.6: Bus data for the IEEE 30 bus test system.

Bus No.	Type	Load		Voltage Magnitude Bounds	
		Active Load	Reactive Load	Min Voltage Magnitude (p.u.)	Max Voltage Magnitude (p.u.)
1	3	0.0	0.0	0.95	1.05
2	2	21.7	12.7	0.95	1.1
3	0	2.4	1.2	0.95	1.05
4	0	7.6	1.6	0.95	1.05
5	0	0.0	0.0	0.95	1.05
6	0	0.0	0.0	0.95	1.05
7	0	22.8	10.9	0.95	1.05
8	0	30.0	30.0	0.95	1.05
9	0	0.0	0.0	0.95	1.05
10	0	5.8	2.0	0.95	1.05
11	0	0.0	0.0	0.95	1.05
12	0	11.2	7.5	0.95	1.05
13	2	0.0	0.0	0.95	1.1
14	0	6.2	1.6	0.95	1.05
15	0	8.2	2.5	0.95	1.05
16	0	3.5	1.8	0.95	1.05
17	0	9.0	5.8	0.95	1.05
18	0	3.2	0.9	0.95	1.05
19	0	9.5	3.4	0.95	1.05
20	0	2.2	0.7	0.95	1.05
21	0	17.5	11.2	0.95	1.05
22	2	0.0	0.0	0.95	1.1
23	2	3.2	1.6	0.95	1.1
24	0	8.7	6.7	0.95	1.05
25	0	0.0	0.0	0.95	1.05
26	0	3.5	2.3	0.95	1.05
27	2	0.0	0.0	0.95	1.1
28	0	0.0	0.0	0.95	1.05
29	0	2.4	0.9	0.95	1.05
30	0	10.6	1.9	0.95	1.05

D.1.3.2 Line Data

Table D.7 displays the line data of the IEEE 14 bus test system. This table is organized as follows: The endpoints of each edge (line) is given by the ordered pair (from bus, to bus) and these points are depicted in column one and column two, respectively. Further, in the remaining four columns, the line resistance, the line reactance, the line charging, and the transformer tap ratio are listed.

TABLE D.7: Line data for the IEEE 14 bus test system.

From Bus	To Bus	Resistance (p.u.)	Reactance (p.u.)	Line charging (p.u.)	tap ratio
1	2	0.01938	0.05917	0.0528	1
1	5	0.05403	0.22304	0.0492	1
2	3	0.04699	0.19797	0.0438	1
2	4	0.05811	0.17632	0.0374	1
2	5	0.05695	0.17388	0.034	1
3	4	0.06701	0.17103	0.0346	1
4	5	0.01335	0.04211	0.0128	1
4	7	0	0.20912	0	0.978
4	9	0	0.55618	0	0.969
5	6	0	0.25202	0	0.932
6	11	0.09498	0.1989	0	1
6	12	0.12291	0.25581	0	1
6	13	0.06615	0.13027	0	1
7	8	0	0.17615	0	1
7	9	0	0.11001	0	1
9	10	0.03181	0.08450	0	1
9	14	0.12711	0.27038	0	1
10	11	0.08205	0.19207	0	1
12	13	0.22092	0.19988	0	1
13	14	0.17093	0.34802	0	1

Table D.8 shows the line data of the IEEE 30 bus test system – its columns corresponds to those in Table D.7.

Table D.8: Line data for the IEEE 30 bus test system.

From Bus	To Bus	Resistance (p.u.)	Reactance (p.u.)	Line charging (p.u.)	tap ratio
1	2	0.0192	0.0575	0.0528	1
1	3	0.0452	0.1652	0.0408	1
2	4	0.0570	0.1737	0.0368	1
3	4	0.0132	0.0379	0.0084	1
2	5	0.0472	0.1983	0.0418	1
2	6	0.0581	0.1763	0.0374	1
4	6	0.0119	0.0414	0.0090	1
5	7	0.0460	0.1160	0.0204	1
6	7	0.0267	0.0820	0.0170	1
6	8	0.0120	0.0420	0.0090	1
6	9	0.0	0.2080	0.0	0.978
6	10	0.0	0.5560	0.0	0.969
9	11	0.0	0.2080	0.0	1
9	10	0.0	0.1100	0.0	1
4	12	0.0	0.2560	0.0	0.932
12	13	0.0	0.1400	0.0	1
12	14	0.1231	0.2559	0.0	1
12	15	0.0662	0.1304	0.0	1
12	16	0.0945	0.1987	0.0	1
14	15	0.2210	0.1997	0.0	1
16	17	0.0524	0.1923	0.0	1
15	18	0.1073	0.2185	0.0	1
18	19	0.0639	0.1292	0.0	1
19	20	0.0340	0.0680	0.0	1
10	20	0.0936	0.2090	0.0	1
10	17	0.0324	0.0845	0.0	1
10	21	0.0348	0.0749	0.0	1
10	22	0.0727	0.1499	0.0	1
21	22	0.0116	0.0236	0.0	1
15	23	0.1000	0.2020	0.0	1
22	24	0.1150	0.1790	0.0	1
23	24	0.1320	0.2700	0.0	1
24	25	0.1885	0.3292	0.0	1
25	26	0.2544	0.3800	0.0	1
25	27	0.1093	0.2087	0.0	1
28	27	0.0	0.3960	0.0	0.968
27	29	0.2198	0.4153	0.0	1
27	30	0.3202	0.6027	0.0	1
29	30	0.2399	0.4533	0.0	1
8	28	0.0636	0.2000	0.0428	1
6	28	0.0169	0.0599	0.0130	1

Bibliography

[1] K. A. Ariyawansa and Y. Zhu. Stochastic semidefinite programming: A new paradigm for stochastic optimization. *4OR: A Quarterly Journal of Operations Research*, 4:239–253, 2004.

[2] S. Mehrotra and M. G. Özevin. On the implementation of interior point decomposition algorithms for two-stage stochastic conic programs. *SIAM Journal on Optimization*, 19:1846–1880, 2009.

[3] J. Lavaei and S. H. Low. Zero duality gap in optimal power flow problem. *IEEE Transactions on Power Systems*, 27:92–107, 2011.

[4] G. C. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk*. World Scientific Publishing, 2007.

[5] R. Schultz and S. Tiedemann. Risk aversion via excess probabilities in stochastic programs with mixed-integer recourse. *SIAM Journal on Optimization*, 14:115–138, 2003.

[6] T. Haukaas. Probablistic models, methods, and decisions in earthquake engineering. In G. Deodatis, B. R. Ellingwood, and D. M. Frangopol, editors, *Safety, Reliability, Risk and Life-Cycle Performance of Structures & Infrastructures*, pages 47–66. CRC Press, 2014.

[7] X. Bai, H. Wei, K. Fujisawa, and Y. Wang. Semidefinite programming for optimal power flow problems. *International Journal of Electirc Power & Energy Systems*, 30:383–392, 2008.

[8] G. B. Dantzig. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.

[9] A. Prékopa. *Stochastic Programming.* Mathematics and Its Applications. Springer, 1995.

[10] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming.* International Series in Operations Research and Financial Engineering. Springer; Second Edition, 2011.

[11] A. Ruszczyński and A. Shapiro, editors. *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science.* Elsevier, 2003.

[12] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming, Modeling and Theory.* Society for Industrial and Applied Mathematics, 2009.

[13] S. Wallace and W. T. Ziemba. *Applications of Stochastic Programming.* MPS-SIAM Series on Optimization. SIAM, 2005.

[14] P. Kall and S. W. Wallace. *Stochastic Programming.* John Wiley & Sons, second edition, 2003.

[15] J. R. Birge. Models and model value in stochastic programming. *Annals of Operations Research*, 59:1–18, 1995.

[16] V. Goel and I. E. Grossmann. A class of stochastic programs with decision dependet uncertainty. *Mathematical Programming*, 108:355–394, 2006.

[17] T. W. Jonsbråten, R. J.-B. Wets, and D. L. Woodruff. A class of stochastic programs with decision dependet random elements. *Annals of Operations Research*, 82:83–106, 1998.

[18] A. Ruszczyński. Decomposition methods in stochastic programming. *Mathematical Programming*, 79(1-3):333–353, 1997.

[19] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization.* Princeton Series in Applied Mathematics. Princeton University Press, 2009.

[20] A. Charnes and W. W. Cooper. Chance-constrained programming. *Management Science*, 6:73–79, 1959.

[21] S. Uryasev. *Probabilistic Constrained Optimization: Methodology and Applications*, volume 49 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 2000.

[22] A. Prékopal. Probablistic programming. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 267–345. Elsevier, 2003.

[23] H. M. Markowitz, G. P. Todd, and W. F. Sharpe. *Mean-variance analysis in portfolio choice and capital markets*, volume 66. John Wiley & Sons, 2000.

[24] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26:1443–1471, 2002.

[25] G. A. Whitmore and M. C. Findlay. *Stochastic dominance: an approach to decision-making under risk*. Lexington Books, 1978.

[26] H. Levy. Stochastic dominance and expected utility: survey and analysis. *Management Science*, 38(4):555–593, 1992.

[27] D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14:548–566, 2003.

[28] D. Dentcheva and A. Ruszczyński. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Programming*, 99(2):329–350, 2004.

[29] D. Dentcheva and A. Ruszczyński. Portfolio optimization with stochastic dominance constraints. *Journal on Banking and Finance*, 30:433–451, 2006.

[30] A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, INC., 2002.

[31] C. Helmberg. Semidefinite programming for combinatorial optimization. *ZIB-Report ZR 00-34, Konrad-Zuse-Zentrum Berlin*, 2000.

[32] H. Wolkowitz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Progamming: Theory, Algorithms and Applications*, volume 27 of *International Series in Operations Research & Management Science*. Kluwer Academic Publishers, 2000.

[33] M. F. Anjos and J. B. Lasserre, editors. *Handbook on Semidefinite, Conic and Polynomial Optimization*, volume 166 of *International Series in Operations Research & Management Science*. Springer, 2012.

[34] C. C. Carøe and R. Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24:37–45, 1999.

[35] R. M. Van Slyke and R. J.-B. Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4):638–663, 1969.

[36] P. Kall and J. Mayer. *Stochastic Linear Programming Models, Theory, and Computation*, volume 156 of *International Series in Operations Research & Management Science*. Springer; Second Edition, 2011.

[37] F. V. Louveaux and R. Schultz. Stochastic integer programming. *Handbooks in operations research and management science*, 10:213–266, 2003.

[38] B. Bank and R. Mandel. *Parametric integer optimization*, volume 39. Akademie-Verlag, 1988.

[39] C. E. Blair and R. G. Jeroslow. The value function of a mixed integer program: I. *Discrete Mathematics*, 19:121–138, 1977.

[40] C. E. Blair and R. G. Jeroslow. The value function of a mixed integer program: II. *Discrete Mathematics*, 25:17–19, 1979.

[41] P. Kall. On approximations and stability in stochastic programming. In J. Guddat, H. T. Jongen, B. Kummer, and F. Nožička, editors, *Parametric Optimization and Related Topics*, pages 387–407. Akademie-Verlag, Berlin, 1987.

[42] A. Auslender. Closedness criteria for the image of a closed set by a linear operator. *Numerical Functional Analysis and Optimization*, 17:503–515, 1996.

[43] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[44] A. Ruszczynski. *Nonlinear Optimization*. Princeton University Press, 2006.

[45] R. J.-B. Wets. Programming under uncertainty: the equivalent convex program. *SIAM Journal on Applied Mathematics*, 14(1):89–105, 1966.

[46] R. J.-B. Wets. Stochastic programs with fixed recourse: the equivalent deterministic program. *SIAM Review*, 16(3):309–339, 1974.

[47] D. Walkup and R. J.-B. Wets. Stochastic programs with recourse. *SIAM Journal on Applied Mathematics*, 15:1299–1314, 1967.

[48] D. Walkup and R. J.-B. Wets. Stochastic programs with recourse ii: On the continuity of the objective. *SIAM Journal on Applied Mathematics*, 17: 98–103, 1969.

[49] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[50] D. P. Bertsekas. Stochastic optimization problems with nondifferentiable cost functionals. *Journal of Optimization Theory and Applications*, 12:218–231, 1973.

[51] S. R. Lay. *Convex Sets and Their Applications*. Dover Publications, 2007.

[52] R. Schultz. Strong convexity in stochastic programs with complete recourse. *Journal of Computational and Applied Mathematics*, 56:3–22, 1994.

[53] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: Semideviations as risk measures. *European Journal of Operational Research*, 116(1):33–50, 1999.

[54] W. Ogryczak and A. Ruszczynski. Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization*, 13(1):60–78, 2002.

[55] H. P. Williams. Logic and integer programming. In *International Series in Operations Research & Management Science*. Springer, 2009.

[56] L. A. Wolsey and G. L. Nemhauser. *Integer and combinatorial optimization*. Wiley, 1988.

[57] R. R. Meyer. On the existence of optimal solutions to integer and mixed-integer programming problems. *Mathematical Programming*, 7(1):223–235, 1974.

[58] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I.* A Series of Comprehensive Studies in Mathematics. Springer, 1996.

[59] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[60] G. Pataki. On the closedness of the linear image of a closed convex cone. *Mathematics of Operations Research*, 32(2):395–412, 2007.

[61] G. Pataki. Bad semidefinite programs: they all look the same. *preprint http://arxiv. org/abs/1112.1436*, 7, 2011.

[62] L. Stougie. *Design and Analysis of Algorithms for Stochastic Integer Programming.* PhD thesis, Center for Mathematics and Computer Sience, Amsterdam, 1985.

[63] R. Schultz. Continuity properties of expectation functions in stochastic integer programming. *Mathematics of Operations Research*, 18:578–589, 1993.

[64] R. Schultz. On structure and stability in stochastic programs with random technology matrix and complete integer recourse. *Mathematical Programming*, 70:73–89, 1995.

[65] X. Liu, S. Küçükyavuz, and J. Luedtke. Decomposition algorithms for two-stage chance-constrained programs. *Mathematical Programming*, pages 1–25, 2014.

[66] H. Heitsch and W. Römisch. Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications*, 24(2-3):187–206, 2003.

[67] J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming. *Mathematical Programming*, 95(3):493–511, 2003.

[68] S. M. Robinson and R. J.-B. Wets. Stability in two-stage stochastic programming. *SIAM Journal on Control and Optimization*, 25:1409–1416, 1987.

[69] W. Römisch and R. Schultz. Stability analysis for stochastic programs. *Annals of Operations Research*, 30:241–266, 1991.

[70] R. Schultz. Some aspects of stability in stochastic programming. *Annals of Operations Research*, 100:55–84, 2000.

[71] W. Römisch. Stability of stochastic programming problems. *Handbooks in Operations Research and Management Science*, 10:483–554, 2003.

[72] H. Heitsch, W. Römisch, and C. Strugarek. Stability of multistage stochastic programs. *SIAM Journal on Optimization*, 17:511–525, 2006.

[73] S. Mehrotra and M. G. Özevin. Decomposition-based interior point methods for two-stage stochastic semidefinite programming. *SIAM Journal on Optimization*, 18:206–222, 2007.

[74] G. Zhao. A log-barrier method with Benders decomposition for solving two-stage stochastic linear programs. *Mathematical Programming*, 90:507–536, 2001.

[75] J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4:238–252, 1962.

[76] C. Lemaréchal. Lagrangian relaxation. In M. Jünger and D. Naddef, editors, *Computational Combinatorial Optimization*, volume 2241 of *Lecture Notes in Computer Science*, pages 112–156. Springer, 2001.

[77] J. M. Mulvey and A. Ruszczyński. A new scenario decomposition method for large-scale stochastic optimization. *Operations Research*, 43(3):477–490, 1995.

[78] A. J. Berger, J. M. Mulvey, and A. Ruszczynski. An extension of the dqa algorithm to convex stochastic programs. *SIAM Journal on Optimization*, 4 (4):735–753, 1994.

[79] R. T. Rockafellar and R. J.-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16(1): 119–147, 1991.

[80] C. H. Rosa and A. Ruszczyński. On augmented Lagrangian decomposition methods for multistage stochastic programs. *Annals of Operations Research*, 64(1):289–309, 1996.

[81] A. Løkketangen and D. L. Woodruff. Progressive hedging and tabu search applied to mixed integer (0, 1) multistage stochastic programming. *Journal of Heuristics*, 2(2):111–128, 1996.

[82] M. Lubin, K. Martin, C. Petra, and B. Sandikçi. On parallelizing dual decomposition in stochastic integer programming. *Operations Research Letters*, 41:252–258, 2013.

[83] A. Ruszczyński. A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming*, 35:309–333, 1986.

[84] J. Nocedal and S. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2006.

[85] A. Märkert and R. Gollmer. *User's Guide to ddsip – A C Package for the Dual Decomposition of Two-Stage Stochastic Programs with Mixed-Integer Recourse*, 2015. `https://www.uni-due.de/~hn215go/software/ddsip-man.pdf`.

[86] P. Kall. Computational methods for solving two-stage stochastic linear programming problems. *Zeitschrift für angewandte Mathematik und Physik ZAMP*, 30(2):261–271, 1979.

[87] J. R. Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research*, 33(5):989–1007, 1985.

[88] F. V. Louveaux. A solution method for multistage stochastic programs with recourse with application to an energy investment problem. *Operations Research*, 28(4):889–902, 1980.

[89] G. Laporte and F. V. Louveaux. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13 (3):133–142, 1993.

[90] A. M. Geoffrion. Generalized Benders decomposition. *Journal of Optimization Theory and Applications*, 10(4):237–260, 1972.

[91] N. V. Sahinidis and I. E. Grossmann. Convergence properties of generalized Benders decomposition. *Computers & Chemical Engineering*, 15(7):481–491, 1991.

[92] K. Aoki, M. Itoh, T. Satoh, K. Nara, and M. Kanezashi. Optimal long-term unit commitment in large scale systems including fuel constrained thermal and pumped-storage hydro. *IEEE Transactions on Power Systems*, 4(3): 1065–1073, 1989.

[93] J. Carpentier. Contribution to the economic dispatch problem. *Bulletin Society Francaise Electriciens*, 3:431–447, 1962.

[94] S. Frank and S. Rebennack. A primer on optimal power flow: Theory, formulation, and practical examples. Working Paper, Colorado School of Mines, 2012.

[95] S. Frank, I. Steponavice, and S. Rebennack. A primer on optimal power flow: A bibliographic survey (i) – formulations and deterministic methods. *Energy Systems*, 3:221–258, 2012a.

[96] S. Frank, I. Steponavice, and S. Rebennack. A primer on optimal power flow: A bibliographic survey (ii) – non-deterministic and hybrid methods. *Energy Systems*, 3:259–289, 2012b.

[97] C. C. Carøe and R. Schultz. A two-stage stochastic program for unit commitment under uncertainty in a hydro-thermal power system. *ZIB-Preprint SC 98-11, Konrad-Zuse-Zentrum Berlin*, 1998.

[98] R. Gollmer, M. P. Nowak, W. Römisch, and R. Schultz. Unit commitment in power generation - a basic model and some extensions. *Annals of Operations Research*, 96:167–189, 2000.

[99] J. Zhu. *Optimization of Power System Operation*. IEEE Press Series on Power Engineering. John Wiley & Sons, INC., 2009.

[100] K. Lehmann, A. Grastien, and P. Van Hentenryck. AC-feasibility on tree networks is NP-hard. *arXiv preprint arXiv:1410.8253*, 2014.

[101] J. A. Momoh, R. Adapa, and M. E. El-Hawary. A review of selected optimal power flow literature to 1993 part i: Nonlinear and quadratic programming approaches. *IEEE Transactions on Power Systems*, 14:96–104, 1999.

[102] J. A. Momoh, R. Adapa, and M. E. El-Hawary. A review of selected optimal power flow literature to 1993 part ii: Nonlinear and quadratic programming approaches. *IEEE Transactions on Power Systems*, 14:105–111, 1999.

[103] D. Bienstock and S. Mattia. Using mixed-integer programming to solve power grid blackout problems. *Discrete Optimization*, 4(1):115–141, 2007.

[104] P. Sánchez-Martin and A. Ramos. Modeling transmission ohmic losses in a stochastic bulk production cost model. Instituto de Investigación Tecnológica, Universidad Pontificia Comillas, Madrid, 1997.

[105] P. Sánchez-Martin, A. Ramos, and J. F. Alonso. Probablistic midterm transmission planning in a liberalized market. *IEEE Transactions on Power Systems*, 20:2135–2142, 2005.

[106] S. Kuhn. *Betriebsoptimierung von elektrischen Energieerzeugungsanlagen und Übertragungssystemen bei unvollständiger Information.* PhD thesis, University of Duisburg-Essen, 2011.

[107] R. Jabr. Radial distribution load flow using conic programming. *IEEE Transactions on Power Systems*, 21(3):1458–1459, 2006.

[108] M. Farivar, C. R. Clarke, S. H. Low, and K. M. Chandy. Inverter var control for distribution systems with renewables. In *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 457–462. IEEE, 2011.

[109] H. L. Hijazi, C. Coffrin, and P. Van Hentenryck. Convex quadratic relaxations of mixed-integer nonlinear programs in power systems. *NICTA, Canberra, ACT Australia, Tech. rep*, 2013.

[110] C. Coffrin, H. L. Hijazi, and P. Van Hentenryck. The QC relaxation: Theoretical and computational results on optimal power flow. *arXiv preprint arXiv:1502.07847*, 2015.

[111] *Power Systems Test Case Archive.* UWEE University of Washington Electrical Engineering, http://www.ee.washington.edu/research/pstca/, 2014.

[112] B. C. Lesieutre, D. K. Molzahn, A. R. Borden, and C. L. DeMarco. Examining the limits of the application of semidefinite programming to power flow problems. In *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1492–1499, 2011.

[113] W. A. Bukhsh, A. Grothey, K. McKinnon, and P. Trodden. Local solutions of optimal power flow. *IEEE Transactions on Power Systems*, 28:4780–4788, 2013.

[114] D. K. Molzahn. *Application of Semidefinite Optimization Techniques to Problems in Electric Power Systems*. PhD thesis, University of Wisconsin–Madison, 2013.

[115] D. Mehta, D. K. Molzahn, and K. Turitsyn. Recent advances in computational methods for the power flow equations. *arXiv preprint arXiv:1510.00073*, 2015.

[116] N. Amjady and M. R. Ansari. Hydrothermal unit commitment with AC constraints by a new solution method based on Benders decomposition. *Energy Conversion and Management*, 65:57–65, 2013.

[117] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, 27: 20–34, 2010.

[118] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problem using semi-definite programming. *Journal of the ACM*, 42:1115–1145, 1955.

[119] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11:625–653, 1999.

[120] C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz. An interior-point method for semidefinite programming. *SIAM Journal on Optimization*, 6(2):342–361, 1996.

[121] S. Sojoudi and J. Lavaei. Physics of power networks makes hard optimization problems easy to solve. In *Power and Energy Society General Meeting*, pages 1–8. IEEE, 2012.

[122] R. Madani, M. Ashraphijuo, and J. Lavaei. Promises of conic relaxation for contingency-constrained optimal power flow problem. In *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, pages 1064–1071. IEEE, 2014.

[123] S. Bose, D. F. Gayme, S. Low, and K. M. Chandy. Optimal power flow over tree networks. In *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pages 1342–1348. IEEE, 2011.

[124] B. Zhang and D. Tse. Geometry of feasible injection region of power networks. In *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pages 1508–1515. IEEE, 2011.

[125] D. K. Molzahn, J. T. Holzer, B. C. Lesieutre, and C. L. DeMarco. Implementation of a large-scale optimal power flow solver based on semidefinite programming. *IEEE Transactions on Power Systems*, 28(4):3987–3998, 2013.

[126] R. Madani, M. Asharphijuo, and J. Lavaei. *SDP Solver of Optimal Power Flow User's Manual*, 2014. `http://www.columbia.edu/~rm3122/OPF_Solver_Guide.pdf`.

[127] R. Madani, S. Sojoudi, and J. Lavaei. Convex relaxation for optimal power flow problem: Mesh networks. *IEEE Transactions on Power Systems*, 30(1): 199–211, 2015.

[128] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52 (3):471–501, 2010.

[129] A. M. Geoffrion. Generalized Benders decomposition. *Journal of Optimization Theory and Applications*, 10:237–260, 1972.

[130] H. B. Gooi, D. P. Mendes, K. R. W. Bell, and D. S. Kirschen. Optimal scheduling of spinning reserve. *IEEE Transactions on Power Systems*, 14 (4):1485–1492, 1999.

[131] R. N. Allan. *Reliability evaluation of power systems.* Springer Science & Business Media, 2013.

[132] P. Carpentier, G. Gohen, J.-C. Culioli, and A. Renaud. Stochastic optimization of unit commitment: a new decomposition framework. *IEEE Transactions on Power Systems*, 11(2):1067–1073, 1996.

[133] S. Takriti, J. R. Birge, and E. Long. A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems*, 11(3):1497–1508, 1996.

[134] M. P. Nowak, R. Schultz, and M. Westphalen. A stochastic integer programming model for incorporating day-ahead trading of electricity into hydro-thermal unit commitment. *Optimization and Engineering*, 6(2):163–176, 2005.

[135] P. Ruiz, C. R. Philbrick, E. Zak, K. W. Cheung, and P. W. Sauer. Uncertainty management in the unit commitment problem. *IEEE Transactions on Power Systems*, 24(2):642–651, 2009.

[136] S. M. Ryan, R. J.-B. Wets, D. L. Woodruff, C. Silva-Monroy, and J.-P. Watson. Toward scalable, parallel progressive hedging for stochastic unit commitment. In *Power and Energy Society General Meeting (PES)*, pages 1–5. IEEE, 2013.

[137] H. Heitsch and W. Römisch. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118:371–406, 2009.

[138] IBM ILOG Cplex. V12.5.1 user's manual, 2015.

[139] G. T. Gilbert. Positive definite matrices and Sylvester's criterion. *The American Mathematical Monthly*, 98(1):44–46, 1991.

[140] F. Zhang. *Matrix theory: basic results and techniques*. Springer Science & Business Media, 2011.

[141] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

[142] V. Chvátal. *Linear Programming*. W. H. Freeman, 1983.

[143] M. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Mathematical Programming*, 77:129–162, 1997.

[144] A. Ben-Israel. Linear equations and inequalities on finite dimensional, real or complex, vector spaces: A unified theory. *Journal of Mathematical Analysis and Applications*, 27:367–389, 1969.

[145] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5: 13–51, 1995.

[146] J. B. Lasserre. A new Farkas lemma for positive semidefinite matrices. *IEEE Transactions on Automatic Control*, 40:1131–1133, 1995.

[147] J. Dattorro. *Convex Optimization & Euclidean Distance Geometry*. Meboo Publishing USA, 2005.

[148] G. Pataki. On the facial structure of cone-LPs and semidefinite programming. Technical report, Management Sience Research Report MSRR-#595, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, 1994.

[149] G. Pataki. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Mathematics of Operations Research*, 23:339–358, 1998.

[150] I. Pólik and T. Terlaky. A survey of the s-lemma. *SIAM Review*, 49:371–418, 2007.

[151] A. I. Bravinok. A remark on the rank of positive semidefinite matrices subject to affine constraints. *Discrete & Computational Geometry*, 25:23–31, 2001.

[152] H. L. Royden and P. Fitzpatrick. *Real Analysis*, volume 198. Macmillan New York, 1988.

[153] B. Stott and O. Alsaç. Fast decoupled load flow. *IEEE Transactions on Power Apparatus and Systems*, PAS-93(3):859–869, 1974.

[154] O. Alsac, J. Bright, M. Prais, and B. Stott. Further developments in LP-based optimal power flow. *IEEE Transactions on Power Systems*, 5(3): 697–711, 1990.

[155] B. Stott, J. Jardim, and O. Alsaç. DC power flow revisited. *IEEE Transactions on Power Systems*, 24(3):1290–1300, 2009.

[156] A. J. Wood and B. F. Wollenberg. *Power Generation, Operation, and Control*. John Wiley & Sons, 2012.

[157] K. Purchala, L. Meeus, D. Van Dommelen, and R. Belmans. Usefulness of DC power flow for active power flow analysis. In *Power Engineering Society General Meeting*, pages 454–459. IEEE, 2005.

[158] C. Coffrin, P. Van Hentenryck, and R. Bent. Approximating line losses and apparent power in AC power flow linearizations. In *Power and Energy Society General Meeting*, pages 1–8. IEEE, 2012.

[159] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

[160] Gurobi Optimization. Gurobi optimizer reference manual. *URL: http://www. gurobi. com*, 2012.

[161] G. P. McCormick. Computability of global solutions to factorable nonconvex programs: Part i—convex underestimating problems. *Mathematical Programming*, 10(1):147–175, 1976.

[162] S. H. Low. Convex relaxation of optimal power flow, part i: Formulations and equivalence. *arXiv preprint arXiv:1405.0766*, 2014.

[163] S. H. Low. Convex relaxation of optimal power flow, part ii: Exactness. *arXiv preprint arXiv:1405.0814*, 2014.

[164] R. Jabr. Exploiting sparsity in SDP relaxations of the OPF problem. *IEEE Transactions on Power Systems*, 2(27):1138–1139, 2012.