UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS REITOR JOÃO DAVID FERREIRA LIMA
POST-GRADUATION PROGRAM IN ELECTRICAL ENGINEERING

GUILHERME MATIUSSI RAMALHO

EXTENSIONS FOR PROBABILISTIC CONSTRAINED PROGRAMMING PROBLEMS: THE CASES OF NON-CONTINUOUS UNIT COMMITMENT AND BILINEAR ENERGY PORTFOLIO MANAGEMENT

Florianópolis

2019

Guilherme Matiussi Ramalho

EXTENSIONS FOR PROBABILISTIC CONSTRAINED PROGRAMMING PROBLEMS: THE CASES OF NON-CONTINUOUS UNIT COMMITMENT AND BILINEAR ENERGY PORTFOLIO MANAGEMENT

Doctoral Thesis submitted to the Post-graduation Program in Electrical Engineering of the Universidade Federal de Santa Catarina for obtaining the Grade of Doctor of Electrical Engineering.
Advisor
Universidade Federal de Santa Catarina: Prof. Dr. Erlon Cristian Finardi
Co-advisor
Électricité de France: HdR. Dr. Ir. Wim Stefanus van Ackooij

Florianópolis

2019

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Ramalho, Guilherme Matiussi
Extensions for probabilistic constrained programming problems: The cases of non-continuous unit commitment and bilinear energy portfolio management / Guilherme Matiussi Ramalho ; orientador, Erlon Cristian Finardi, coorientador, Wim Stefanus van Ackooij, 2019.
180 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2019.

Inclui referências.

1. Engenharia Elétrica. 2. Otimização. 3. Restrições Probabilísticas. 4. Problema de Unit Commitment. 5. Problema de Portfólio. I. Finardi, Erlon Cristian. II. van Ackooij, Wim Stefanus. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

Guilherme Matiussi Ramalho

EXTENSIONS FOR PROBABILISTIC CONSTRAINED PROGRAMMING PROBLEMS: THE CASES OF NON-CONTINUOUS UNIT COMMITMENT AND BILINEAR ENERGY PORTFOLIO MANAGEMENT

This Doctoral Thesis has been evaluated and approved by the examining committee composed of the following members:

Prof. Dr. André Luiz Diniz
CEPEL, Rio de Janeiro, Brasil

Prof. Dr. Bernardo Kulnig Pagnoncelli
Universidad Adolfo Ibáñez, Santiago, Chile

Prof. Dr. Claudia Alejandra Sagastizabal
Universidade Estadual de Campinas, Campinas, Brasil

We certify that this is the **original and final version** of the Doctoral Thesis which was deemed suitable to obtain the Grade of Doctor of Electrical Engineering.

Bartolomeu Ferreira Uchoa Filho:47636211491
Assinado de forma digital por Bartolomeu Ferreira Uchoa Filho:47636211491
Dados: 2019.11.28 09:48:06 -03'00'

Prof. Dr. Bartolomeu Ferreira Uchôa-Filho
Coordinator
Universidade Federal de Santa Catarina

Erlon Cristian Finardi:02036474918
Assinado de forma digital por Erlon Cristian Finardi:02036474918
Dados: 2019.11.27 09:46:22 -03'00'

Prof. Dr. Erlon Cristian Finardi
Advisor
Universidade Federal de Santa Catarina

Florianópolis, October 15 , 2019.

This thesis is dedicated to the women of my life: my (future) wife Renata, my mother Ivete and my aunt Izabel

ACKNOWLEDGEMENTS

Like many things in my life, the present thesis takes its origin in somewhat unexpected deviations from my life. It was 2015, I was already in the field that I wanted to work on, I mean, "something dealing with energy market", but a bit apprehensive with future outlook. The energy market was not in its best days due to various regulatory doubts. In a conversation with some friends, I was encouraged to study until the "bad tide" passed. At first glance, I was a bit upset with it, I had just finished my master's degree, and at this time, I wanted a little peace of mind. But that low voice always comes and haunts us: "Come on, why not?". Then, It began the search for a university, finally finding my space in the Federal University of Santa Catarina, 800 km from home, on the recommendation of my friend Gustavo Arfux.

Once in Santa Catarina, I was very well received by the UFSC Post-Graduation Electrical Engineering team, especially Labplan staff. First of all, I would like to thank Erlon Finardi, who agreed to be my advisor, and without whom this thesis would not have even started. Without their teachings and directions (and jokes of course), I would hardly have completed this work. He also organized the first joint work with Wim van Ackooij, funded by the SPARHTACUS project, where our desire to work with probabilistic constrained programming problems applied to energy planning began. It was also his idea to call Wim van Ackooij to co-adivise me in the thesis, which, to my surprise, he readily accepted. Here are my special thanks to my co-advisor Wim van Ackooij, who very patiently introduced, guided and corrected me (over and over again) in this beautiful and challenging field of research. I am very grateful to both of you, and for my part, I hope we can still work together on many projects.

This thesis would not have been possible without the letters of recommendation from my former advisors Oswaldo Valle da Costa and Carlo Sgarra. Thank you very much to both of you! This thesis is also due to the team of Norus (in my time Plan4) which first of all provided me with great friends, and also allowed to reconcile a challenging research project involving topics that I had never worked with the doctorate. The portfolio energy application present in this thesis was partially inspired by the problem proposed in the context of the CELESC/R&D 05697-0614/2016 project. In particular, my thanks go to Vitor Matos, who participated in the qualifying board and Paulo Larroyd who participated in the thesis defence. I am grateful for the friendship, endless discussions and barbecues!

The present work reached a higher level of quality after passing through the qualifying board, composed, besides the advisors, by Vitor Matos, André Diniz, Bernardo Pagnoncelli. Thank you for the patient reading and excellent recommendations. Special thanks to René Henrion, who accepted the challenge of being rapporteur for this thesis. Finally, my thanks go to Claudia Sagastizábal and Paulo Larroyd for participating in the final stage of the doctorate. Thank you for trying to understand me!

I would also like to thank the mentors who inspired me in the doctorate's curse. Firstly regarding the Labplan/UFSC's professors Mauro Rosa, Ildemar Decker, Diego Issicaba and Patrick Peng who helped me in the first steps taken in my doctorate. To the professors Erlon Finardi, Eduardo Camponogara, Claudia Sagastizábal, Welington de Oliveira and Juan Pablo Luna for the great classes in mathematical programming. Your lessons (and notes) inspire me up to now.

I thank all my friends for the partnership in the doctorate's track. Initially, to the Jiu-Jitsu Sensei Márcio Branco for enlighting me in new fields of knowledge, presenting me the resilience of the great champions and helping me withstand the stress of the doc-

torate. Now, comes the challenge of listing all Labplan/UFSC teammates. Sometimes in academic discussions and other times just talking nonsenses, they never denied me any help. Guilherme Fredo, Felipe Beltrán, Marcelo Cordova, Sidney Carvalho, Ivo Costa, Renata Pedrini, Kenny Vinente, Rodolfo Machado, Sandy Miranda, Fabricio Sperandio, Dante Sifuentes, Gabriel Bolacell, Daniel Tenfen, Fabricio Takigawa, Erika Petite , Kauana Palma, Murilo Scuzziato, Luciana Cabral, Marcelo Agostini, Thiago Sousa, Valmor Zimmer, Rodolfo Bialecki, Paulo Sehn, Pedro Vieira, Laura Eduarda and Lucas Venturi. Thanks for the great times, always ready to grab a beer!

Finally, I would like to give special thanks to my (future) wife Renata, I appreciate your persistence. Even so far away, we won! This mission would not have been fulfilled without the efforts of my mother Ivete, my father, Paulo, my aunt Izabel and my brother Ian. Thanks for the moments of support, understanding, patience and help. I love you!

Virtú contro a furore/Prederà l'arme, e fia el combatter corto/Ché l'antico valore/Nel brasiliani cor non è ancor morto.

(Adapted from Francesco Petrarca, 1304 - 1374)

RESUMO

O presente estudo explora métodos para utilizar conjuntos convexos em problemas com Restrições Probabísticas de forma eficiente a fim de resolver aplicações de otimização com precisão. No primeiro tópico da tese é proposto um método para resolver o problema de *Unit Commitment* hidrotérmico estocástico por meio da Programação Linear Inteira Mista. Este problema, em sua forma mais comum, procura definir para um dado sistema o cronograma de geração hidrotérmico de suas unidades, considerando o acoplamento temporal e espacial, com o objetivo de suprir a demanda e minimizando os custos de despacho. Do ponto de vista das incertezas que podem ser levadas em conta no modelo estão as fontes de energia alternativa, como a eólica e fotovoltaica, as afluências naturais e as demandas de energia. Para resolver o problema, propomos o modelo de otimização por Restrições Probabilísticas que define soluções obedecendo um limite de probabilidade. O problema de progamação contém funções (contínuas) analíticas de distribuição que caracterizam a abordagem por Restrições Proabilísticas. Neste quadro, para o primeiro tópico do trabalho, propõe-se um método de Restrições Probabilística por hiperplanos suportes. Esse método oferece soluções factíveis exatas quanto ao nível de probabilidade, lidando com conjuntos factíveis não contínuos que caracterizam o problema de *Unit Commitment* hidrotérmico estocástico linear inteiro misto. A performance do algoritmo proposto foi atestada por meio de comparações numéricas com estratégias alternativas de se lidar com problemas de Restrições Probabilísticas. A eficácia do método também foi testadas em um sistema de 46 barras, envolvendo 10.657 variáveis (1.512 binárias) e 25.618 restrições, mostrando que o método também é apropriado para sistemas complexos. No segundo tópico da tese, explora-se outra estrutura de Restrições de Probabilidade, caracterizada por restrições que possuem variáveis aleatórias que multiplicam as de decisão. A literatura define este tipo de estrutura como bilinear, comumente encontrada em aplicações de gestão de portfólio de ativos, por exemplo. O estudo apresenta a comprovação de que essas estruturas podem ser classificadas como um mapa localmente concavo generalizado, possibilitando o uso de uma combinação de algoritmos próprios de programação convexa, para assim resolver uma classe de problemas mais ampla. O algoritmo é testado em um problema de portfólio de energia de uma distribuidora, onde os riscos de ganhos e perdas de um cesta de contratos expostos aos preços de curto prazo foram controlados por uma Restrição Probabilísticas bilinear e bilateral.

Palavras-chave: Restrições Probabilísticas. *Unit Commitment* Hidrotérmico. Programação Linear Inteira Mista. Energia Renovável. Convexidade. Problema de Portfólio.

RESUMO EXPANDIDO

Introdução

As aplicações de programação estocástica em problemas de planejamento de energia têm muitas ramificações, dependendo, por exemplo, do escopo, prazo, discretização, tamanho e metodologia da solução. Na presente tese, analisamos duas dessas vertentes, explorando os procedimentos matemáticos para superar as limitações enfrentadas pelas aplicações em diferentes contextos. O primeiro tópico do trabalho resolve o despacho de geração em um problema de *Unit Commitment*, lidando com o problema de tratamento das descontinuidades dos conjuntos factíveis que distinguem essas aplicações. A principal contribuição deste tópico é um algoritmo que resolve um problema *Unit Commitment* hidrotérmico capaz de incorporar a incerteza do vento. Para esse fim, usamos restrições probabilísticas conjuntas nas restrições de balanço energético, incorporadas ao problema original por um método de hiperplano de suporte. O processo resulta na resolução de problemas de programação linear misto inteiros que são iterativamente complementados até a convergência do algoritmo. O segundo tópico da tese altera o escopo e concentra-se em estender a aplicação da programação matemática com restrições probabilísticas em problemas de gerenciamento de portfólio de contratos de energia. A abordagem proposta visa oferecer uma estratégia que minimize o risco de multas, mas também considere possíveis ganhos com o gerenciamento do portfólio, em um nível controlado de risco. Além disso, o método visa estender os conceitos de concavidade generalizada a estruturas bilineares de restrições bilaterais. Problemas bilineares, típicos de aplicações em portfólios, são caracterizados quando a variável de decisão multiplica a variável aleatória nas restrições do problema de otimização.

Objetivos

O presente trabalho de doutorado propõe duas contribuições principais. Pelo melhor conhecimento do autor, um método exato de otimização de problemas de programação lineares inteiros mistos com restrições probabilísticas que fornece iterativamente uma solução viável com *gap* de otimalidade ainda é desconhecido. Deste ponto de vista, uma tendência atual de estudos estão nos problemas de *Unit Commitment* como estratégia para se lidar com a aleatoriedade de curto prazo imposta pela introdução de energia renovável nos mercados de eletricidade. Supondo que o modelo estocástico que descreve a energia renovável seja definido a priori, o principal objetivo do primeiro tópico do trabalho é introduzir uma abordagem de restrições probabilísticas que ofereça soluções exatas e viáveis. A abordagem lida com o conjunto viável não contínuo que distingue os problemas de *Unit Commitment* estocásticos. Na mesma tendência de avaliações de convexidade de restrições probabilísticas, o segundo tópico do presente trabalho de doutorado visa estender estudos anteriores em mapas côncavos generalizados, possibilitando sua aplicação em uma classe mais ampla de problemas.

Metodologia

A tese se inicia com a proposição de um pequeno problema de planejamento energético estocástico com restrições probabilística, apelidado de *toy problem*, sendo resolvido por meio de um algoritmo de hiperplano suporte clássico da literatura. Junto à resolução do *toy problem* é discorrido sobre os instrumentos matemáticos necessários para a compreensão

dos algoritmos propostos pela tese. O problema de *Unit Commitment* hidrotérmico é inicialmente proposto a partir da formulação do problema de programação matemática que será resolvido. Deste último problema, as restrições de balanço energético são substituídas pelas restrições probabilísticas, introduzindo a estocasticidade. Após a sua formulação, são introduzidos os algoritmos utilizados para se resolver o problema proposto. Para se auferir a eficiência do método, são discutidos resultados computacionais de métodos alternativos acompanhados de diversas simulações de sensibilidade. Visando a melhora da perfomance do algoritmo proposto, sugere-se ainda novas metodologias a serem incorporadas ao algoritmo original. O segundo tópido da tese, que trata de problemas de restrições probabilísticas bilineares e bilaterais, e inicia conceituando-se as funções r-côncavas. Na sequência, são verificadas as condições para que as funções possam preservar a r-concavidade, de forma a serem utilizadas em algoritmos de programação matemática garantindo a sua convexidade local. Os conceitos apresentados anteriormente são utilizados para compor um algoritmo que envolve uma etapa preliminar de programação cônica de segunda ordem e um método iterativo de região de confiança, resolvendo problemas do programação quadrática. Finalmente, o algoritmo é aplicado para se resolver um problema reduzido de portfólio de uma distribuidora de energia hipotética.

Resultados e Discussão

Em referência ao primeiro tópico da presente tese, os resultados iniciam-se com a resolução de um problema de *Unit Commitment* composto por cinco barras. Para este problema de pequeno porte, observou-se que o algoritmo converge de forma exata para o nível de probabilidade requisitado após 73 iterações e um tempo computacional de 2942 segundos. Verifica-se também que, conforme o nível de probabilidade aumenta, de 90% a 97.5% o custo total também aumenta, ou seja, um despacho mais seguro em termos de risco é acompanhado de um maior comissionamento hidrotérmico. Os experimentos numéricos comparativos entre a metodologia proposta e os algoritmos baseados em restrições probabilísticas individuais demonstram que, apesar destes últimos terem níveis de probabilidade suficientemente altos, violações de restrição ocorrem quando se considera o horizonte de otimização total conjuntamente. Por exemplo, quando se configura o algoritmo com restrições probabilísticas individuais em probabilidades de 95%, nota-se que quando as restrições são tomadas conjuntamente, o nível de probabilidade auferido à posteriori é de 83.42%, para uma das barras do problema. Este resultado demonstra a falta de acurácia destes últimos algoritmos. Em seguida são feitas comparações numéricas para uma abordagem de restrições probabilísticas baseadas em amostra, utilizando-se um algoritmo *Big-M* simplificado. O objetivo é mensurar a qualidade da solução desses algoritmos em relação ao nível de probabilidade, custo total e tempo de convergência. Os resultados mostram que quando os algoritmos *Big-M* são utilizados, procedimentos heurísticos auxiliares mais avançados são necessários para se controlar o nível de probabilidade da solução obtida. Nota-se também que, ao se elevar muito a precisão, a resolução torna-se inviável computacionalmente. Por último, o algoritmo proposto na tese é testado para um problema composto por 46 barras. Para este problema a convergência é alcançada após 3852 segundos, demonstrando a sua viabilidade para sistemas de maior porte. Para o segundo tópico da tese é proposto um algoritmo composto para resolução de problemas portfólio com restrições probabilísticas. Os resultados mostram que a metodologia combinada proposta alcança soluções ótimas para problemas que anteriormente eram inviáveis quando empregados os métodos anteriores da literatura. Os tempos de convergência vão de 1 a 18 min., a depender do horizonte do problema, demonstrando a eficiência da metodologia.

Considerações Finais

Os resultados indicam que a falta de precisão em termos de nível de probabilidade conjunta dos algoritmos que resolvem problemas de restrições probabilísticas requer o uso de estratégias ad-hoc que dificultam sua aplicação a outros problemas. Isto abre um campo de pesquisa para o desenvolvimento de metodologias que exigem níveis precisão mais altos em relação aos riscos. Sobre esse problema discorremos no primeito tópico da tese. Quando comparado a outros algoritmos da literatura, verificou-se que o algoritmo proposto possui um interessante compromisso entre precisão e velocidade convergência, mesmo para problemas de maior escala. Nota-se também que a nova metodologia proposta abre linhas de pesquisas no sentido de melhorar o tempo de convergência do algoritmo, além de dar os embasamentos matemáticos necessários para que novas funções de distribuição, mais precisas com relação à variável randômica a ser modelada, sejam incorporadas ao modelo. Os objetivos traçados para o segundo tópico da tese, que trata de restrições probabilísticas bilineares e bilaterais, são endereçados inicialmente apresentando o embasamento matemático necessário para a generalização do método, de forma a alcançar um campo mais amplo de problemas que podem ser resolvidos nesta linha de pesquisa. Como aplicação, foi proposto um problema de pequeno porte de portifólio de energia. Como resultado, observamos que, para altos níveis de concavidade, como o caso log-côncavo, o problema resulta em muitos casos inviável. Essa restrição é apropriadamente contornada usando um nível mais baixo de r-concavidade, flexibilidade essa somente possível aplicando-se os desenvolvimentos da presente tese. Portanto, os recursos fornecidos pela metodologia proposta cumprem os objetivos traçados de generalização da aplicação de restrições probabilísticas para problemas bilineares e bilaterais.

Palavras-chave: Restrições Probabilísticas. *Unit Commitment* Hidrotérmico. Programação Linear Inteira Mista. Energia Renovável. Convexidade. Problema de Portfólio.

ABSTRACT

The present work explores methods to efficiently make use of convex sets in Probabilistic Constrained (PC) problems to solve energy optimization applications accurately. The first topic is related to a method to deal with the stochastic Hydrothermal Unit Commitment (HTUC) problem through Mixed-Integer Linear Programming (MILP). The standard stochastic HTUC problem intends to define the (hydrothermal) units' generation schedule for a given system, considering its time and space coupling, aiming at the load demand supply with minimal dispatch costs. Regarding the uncertainties, the alternative energy sources, like wind or photovoltaic, the natural river inflows and the load demand can possibly be taken into account. To solve the problem, we propose a probabilistically constrained (PC) optimization model finding solutions obeying a probability threshold. The programming problem contains analytic (continuous) distribution functions that distinguish the PC approach. In this framework, by the first topic of the work, we propose a supporting hyperplane PC method that offers exact feasible solutions in terms of the probability level, handling the non-continuous feasible set that distinguishes the stochastic HTUC MILP problems. We attest the performance of the proposed algorithm by numerical comparisons with commonly used strategies that handle with PC problems. The capabilities of the method are also tested in a 46-bus system, which involves 10,657 variables (1,512 binaries) and 25,618 constraints, showing that the method is also affordable for complex systems. In the second topic of the thesis, we explore a different structure of PC, when random variables multiply the decision ones. The literature defines this type of structure as bilinear, commonly found among asset portfolio applications, for instance. As a research contribution, we prove that such structures can be classified as a locally generalized concave mapping, turning possible the use of a combination of proper convex programming algorithms to solve a broader class of problems. We efficiently test the algorithm in a utility energy portfolio problem, where a two-sided bilinear PC controls the risks of gains and losses of a basket of contracts exposed to the short-term energy price.

Keywords: Probabilistic Constraints. Hydrothermal Unit Commitment. Mixed Integer Linear Programming. Renewable Energy. Convexity. Portfolio Problem.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

UC	Unit Commitment
ED	Economic Dispatch
OPF	Optimal Power Flow
MIP	Mixed Integer Programming
LP	Linear Programming
LR	Lagrangian Relaxation
HTUC	Hydrothermal Unit Commitment
ISO	Independent System Operator
MILP	Mixed-Integer Linear Programming
MINLP	Mixed-Integer Nonlinear Programming
PC	Probability Constraints, Probabilistic Constrained
HPF	Hydro Production Function
SAA	Sample Average Approximation
SOCP	Second Order Cone Programming
TR	Trust Region
QCQP	Quadratically Constrained Quadratic Program
CDF	Cumulative Distribution Function

iid	Independent identically distributed
SAA	Sample Average Approximation
TP	Toy Problem
ISO	Independent System Operator
PDF	Probability Density Function
iff	If and only if

CONTENTS

1 INTRODUCTION

Stochastic programming applications in energy planning problems has many branches depending on, for instance, the scope, time frame, discretization, size and solution methodology (van ACKOOIJ et al., 2018). Throughout the present thesis, we explore two of these branches, highlighting the mathematical procedures to overcome the limitations faced by the applications in different contexts. The first topic solves the generation dispatch in a power Unit Commitment (UC) problem, dealing with the discontinuity convexity issues that distinguish such applications. The second topic changes the scope and focuses on extending the uses of energy contract portfolio management by a generalization of the convexity statements that define the constraint structures.

The first topic focus on UC, a classic problem on energy generation planning. The most common approach aims at defining a schedule for the dispatch of generation units for a short-term period to satisfy a distributed load with minimal costs, taking into account the operational characteristics of the system and units in time and space. UC problems commonly work with discrete variables to define the commitment of production decisions, and thus, in the present work, when discrete variables are not taken into account for the short term schedule problem, in the present work we use the terminology Economic Dispatch (ED) problem or an Optimal Power Flow (OPF) problem (TAHANAN et al., 2015).

The UC problem is a branch of a broader field of study that comprises the energy planning studies in general. As pointed out in (MACEIRA et al., 2002), a common approach to deal with energy production planning in Brazilian system is to consider uncertainties in the long and medium-term problems and consider a more detailed deterministic modelling for the short-term. From the perspective of the short-term problem, the deterministic approach possibly succeeds for their purposes in the prior studies. Nevertheless, the increasing importance of renewable energy into the power systems turns unavoidable the necessity to handle their stochastic characteristics also in this framework.

With predominant hydro generation, the UC problem of the Brazilian system is distinguished by an accurate modelling of the hydropower system. Interesting approaches are given in (FINARDI; SILVA, 2006; FINARDI; SCUZZIATO, 2013, 2014) for purely hydroelectric based systems, and in (BELLONI et al., 2003; TAKIGAWA; FINARDI; SILVA, 2013) for a hydrothermal system. In these works, to deal with the complex hydro production functions and the combinatorial aspects of the problem, the authors use different Lagrangian Relaxation (LR) schemes and focus on comparisons between the approaches. Benders decomposition (BENDERS, 1962) methods for large scale systems in different planning horizons are found in (PINTO; BORGES; MACEIRA, 2013; SANTOS; DINIZ; BORGES, 2017). Alternative methods, as semidefinite programming relaxations to solve the UC problem are discussed in (PAREDES; MARTINS; SOARES, 2015) and Markovian stochastic dynamic programming in (SCARCELLI et al., 2014). Several test cases for involving systems are provided in (DINIZ, 2010).

As observed by those last works, in the Brazilian context a distinct branch of UC studies are the hydrothermal unit commitment (HTUC) problem, where the main task consists in defining the dispatch, taking into account the physical characteristics of units and transmission system. In general, the objective function (and some specific constraints) depend on the market regulatory framework. In a deregulated electricity market, where free market forces define the spot prices by the compromise between demand, generation and trading agents, the most common objective is related to the profit maximization of the generation company, generally based on the definition of bidding strategies. On the other hand, in a centralized dispatch framework, HTUC is used by the Independent System Operator (ISO) for minimizing the expected production costs and also for the day-ahead spot price definition.

Using the Brazilian framework, here we focus on the centralized dispatch, although the methodology proposed can be directly applied in a deregulated market setting. Due to its non-continuity, normally with complex nonlinear productions functions, several systems and units coupling constraints, the HTUC is normally addressed as a large-scale mixed-integer linear programming (MILP) or mixed-integer nonlinear programming

(MINLP) problem (REDONDO; CONEJO, 1999; CHANG et al., 2001; BORGHETTI et al., 2008; FRANGIONI; GENTILE; LACALANDRA, 2011; van ACKOOIJ; MALICK, 2016). As mentioned, in this context, a particular trend is the growing importance of wind power generation, which significantly increases the uncertainty in the system operation. Ignoring such uncertainty can lead to unacceptable risks. To cope with the undesired effects promoted by that variability, we handle the HTUC as a stochastic programming problem. In this context, a crucial aspect is how to manage the uncertainty, which in turn has implications on the form of stochastic HTUC, its potential accuracy, the expected cost of the computed schedules, and on the computational burden (TAHANAN et al., 2015). Indeed, the works (TAHANAN et al., 2015; van ACKOOIJ et al., 2018) offer extensive literature surveys on the evolution of deterministic/stochastic large scale UC problems. For an overview in deterministic hydro UC problem techniques, we refer to (TAKTAK; D'AMBROSIO, 2016).

As pointed by (FINARDI; TAKIGAWA; BRITO, 2016) and cited materials therein, the volatility related to wind power can dominate the volatility of demand, mainly when considered the bus where the wind farm is connected. Therefore, the main contribution of the first topic of the work is an algorithm that solves a HTUC problem capable of incorporating wind uncertainty. To this end, we use joint Probability Constraints (PC) (PRÉKOPA, 1995) on the load-requirement constraints, that are incorporated into the original problem by a supporting hyperplane method. The process results in the resolution of MILP problems that are iteratively complemented up to the convergence of the algorithm. We certify that the approach is capable of generating feasible solutions at each iteration, as well as a quality-certificate of such solution.

It is clear that the discontinuity faced by MILP problems sets is due to their binary variables. As discussed throughout the document, this discontinuity interferes on the direct application of classical supporting hyperplane algorithms (PRÉKOPA, 2003) since it is assumed continuous feasible sets for the PC formulation. The proposed supporting hyperplane algorithm of the present work handles the usual discontinuity of feasible sets of such stochastic HTUC problems by a convex combination strategy of the PC iterates. It is important to remark that in the current stage, the proposed method is valid for the

separated case under normality assumptions. The adopted procedure avoids the use of complex heuristics schemes to deal with the stochasticities of the problem, as the ones commonly found in Individual PC and Sample-based PC approach. The use of such heuristics are discussed in Sections 2.1 and 2.2 respectively.

In a similar procedure as done in (van ACKOOIJ et al., 2014) the present work is also concerned with comparing the level of accuracy obtained when employing Individual PC rather than joint PC. We provide numerical experiments to put into evidence that, despite taking a high enough probability level for the individual constraints, constraint violations do occur when considering the entire optimization horizon. In (van ACKOOIJ et al., 2014) it is verified that the insufficient accuracy of individual PC problems provides unacceptable risk solutions. The present work also confirms this statement into a MILP framework.

We also provide numerical comparisons to a sample-based PC approach, which is given by a simplified Big-M algorithm. The objective is to check the solution quality of these algorithms concerning the probability level, total cost and convergence time. The advantage of the proposed algorithm, confirmed by experiments, is that the PC is taken into account precisely in our case and leads to a feasible solution. This statement contrasts with Big-M and similar sample-based approaches that may need heuristics or more advanced procedures to control the probability level of the obtained solution.

As certified by numerical tests, depending on the number of samples of the sample-based algorithms, these methods can offer swift solutions in terms of CPU time. Therefore, it is essential to study ways to turn the supporting hyperplane model more competitive in terms of convergence time. To this end, the present work also provides some strategies to improve the algorithm convergence performance when compared with its original formulation. These strategies include modifications to the interpolation method and also ad-hoc strategies on the lower bound problem.

Another contribution of the present work is the suggestion of a new MILP formulation for the hydro production function (HPF), where the power production is regarded for the entire plant with still committing the dispatch by individual hydro units. This

strategy aims at reducing the number of necessary piecewise linear approximations of the HPF, and it is only possible if the outflow of the identical on-line turbines is considered equal. Comparative simulations to motivate its effectiveness are briefly evaluated in Section 4.6.

The second main topic of the present thesis document refers to the energy portfolio selection, a challenging problem that is commonly found in PC constrained applications (PAGNONCELLI; REICH; CAMPI, 2012). In its simplest version, it looks at maximizing the profit of a basket of assets with a controlled level the risk by defining the optimal amount of each component. Typically, the source of uncertainties is the asset prices, that when multiplied by their volumes gives the total monetary exposition of the agent. A common approach to deal with this type of problems are the methods based on samplings, as the Sample Average Approximation (SAA) methods e.g. (SHAPIRO, 2003; LUEDTKE; AHMED, 2008), scenario-based algorithms (CAMPI; GARATTI, 2011) and integer programming sample-based methods (LUEDTKE; AHMED; NEMHAUSER, 2010). These methods, despite the versatility in terms of the random variable probability distribution and the PC structure, suffers from the imprecision of the solution, regarding the probability level, and deep dependence on the sample size. In many cases, the method ensures, with a probability precision, an inner or outer solution to the problem, and the increasing of the precision turns the problem intractable (van ACKOOIJ; FINARDI; RAMALHO, 2018). To avoid such problem, in the present work, we pursue a classical approach of PC problems (PRÉKOPA, 1995), based on exact solutions using convex programming algorithms.

Recent works in portfolio selection problem in the Brazilian energy contracting case are discussed in (MATOS et al., 2017; RAMALHO et al., 2018). To solve the underlying problem, the authors developed a two-stage programming problem, providing adherence to the original problem by a sequential convex programming heuristic. In these cases, since the primary objective of the utility that manages the portfolio is to avoid losses and penalties, the possibility of gains are neglected in their model. The strategy adopted by the authors can also be explained by the fact that the search for profits could expose the utility to an unacceptable level of risk. The here proposed PC approach aims at offering

a strategy that minimizes the risk of penalties, but alternatively also turns possible gains with the portfolio management, in a controlled level of risk. Furthermore, the method aims at extending the concepts of generalized concavity to bilateral PC bilinear structures, extending the previous works (van ACKOOIJ, 2017; MINOUX; ZORGATI, 2017).

The main challenge of the previously cited works, which was clear from the proofs of the statements, are the algebraic complexities of the equations. This issue is also treated in the present work, which aims at finding straightforward conditions to ensure r-concavity to the mappings. The former studies that inspired the second topic of the thesis were (MINOUX; ZORGATI, 2016; van ACKOOIJ, 2017; MINOUX; ZORGATI, 2017). Section 5.1 sets basic notions of generalized concavity that are used along the chapter and motivates the problem. Aiming at turning possible the use of the convexity statements of the work in real problems, we propose a composed algorithm involving a Second Order Cone Programming (SOCP) step and an iterative Trust Region (TR) method that solves Quadratically Constrained Quadratic Program (QCQP) problems. The results show that the algorithm indeed achieves feasible optimal solutions to problems that were restricted, or infeasible, when employed the methods of the literature.

1.1 OBJECTIVES

1.1.1 General Objective

The present thesis document proposes two main contributions. To the best of the author's knowledge, an exact PC optimization method in MILP problems framework that provides iteratively a feasible solution and an optimality gap is still unknown. In this view, a particular current trend of studies in the UC problems are in strategies to deal with the short-term randomness imposed by the introduction of renewable energy in electricity markets. Assuming that the stochastic model that describes the renewable energy is defined a priori, the main objective of the first topic of the work is to introduce a supporting hyperplane PC approach that offers exact feasible solutions in terms of the probability level. The approach handles the non-continuous feasible set that distinguish

stochastic HTUC problems defining the convex set of the PC iteratively.

In the same trend of PC convexity evaluations, the second topic aims at extending previous studies in generalized concave mappings, turning possible its application in a broader class of PC problems. The studies focus on bilinear structures, i.e. when the decision variable multiplies the underlying random variable in the constraints of the programming problem.

1.1.2 Specific objectives

The specific objectives are:

- Propose an exact supporting hyperplane method for MILP with PC;
- Explore methods to improve the rate of convergence of the proposed algorithm, testing these methods by numerical experiments;
- Propose an efficient HPF formulation that can be used as an UC alternative for hydroelectric systems;
- Establish conditions to ensure that the mapping $g(x) = \mathbb{P}\{a \leq \tilde{\xi}^\top x \leq b\}$ is locally r-concave, for $\tilde{\xi} \sim \mathcal{N}(\mu, \Sigma)$;
- Propose an algorithm composed of a preliminary step that solves SOCP problems and afterward finds the solution by quadratic approximations of the former problem, using an iterative combined local TR-QCQP method.

1.2 AGENDA

The present thesis is organized as follows. A bibliography survey on applications of PC into UC problems is provided in Chapter 2. The mathematical background necessary to the application of the proposed algorithm, with a small scaled example, is furnished in Chapter 3. The underlying stochastic HTUC problem is presented in Chapter 4, showing details of the joint PC formulation associated with the load requirement constraints. This

chapter also describes the proposed supporting hyperplane method for mixed-integer PC problems and performs several comparative numerical experiments. The last part of Chapter 4 provides strategies to improve the performance of the algorithm, followed by the proposition of the HPF model.

We start the development of the second topic of the thesis in Chapter 5. The main statements of the chapter are made in Section 5.3 where are developed necessary and sufficient conditions and sufficient conditions to ensure generalized concavity properties to the underlying PC mapping. The sufficient conditions are extended to implied terms in Section 5.4, turning the former restrictions appropriate to be used as constraints within standard optimization problem methods. All the steps of the TR-based algorithm are presented in Chapter 6, followed by the application in a small scaled utility energy portfolio problem in Section 6.1. The final remarks of the thesis are made in the conclusion Chapter 7. Appendix B starts an analysis that motivates the use of alternative distributions and structures for the random vectors in PCs. Appendix C offers support for the analytical developments of Chapter 5.

2 UNIT COMMITMENT WITH PROBABILISTIC CONSTRAINTS

The first topic of the work focuses on the use of PC (PRÉKOPA, 1995) to handle wind power uncertainty in HTUC problems. The method in this framework is motivated since PCs allow, for instance, a reliable control on the trade-off between cost and accuracy based on the probability level (or risk level) of the selected solution. Moreover, the results are easy to understand and interpret, since the level of risk, chosen a priori, can be directly linked to the solution's schedule of generation and total cost. In a broad sense, a PC programming problem can be written in the following form:

$$\begin{aligned} \min_{\dot{x}\in X} \quad & c(\dot{x}) \\ \text{s.t.:} \quad & \mathbb{P}\left[h_e(\dot{x},\tilde{\xi}) \leq 0, e \in \mathcal{E}\right] \geq p \end{aligned} \tag{2.1}$$

where, $\dot{x}$ is the decision variable vector that belongs to the set X (definition of the constraints that distinguishes the UC problem) and $c(\cdot)$ is a cost function. In (2.1) the term $\mathbb{P}[h_e(\dot{x},\tilde{\xi}) \leq 0, e \in \mathcal{E}] \geq p$ is the representation of the PC, $\tilde{\xi}$ is the random vector, $\mathbb{P}$ is the probability measure, as described in Section A.1, $h_e(\cdot)$ is a function that handles the interaction between the decision and random vectors, e the individual index of the constraint and $\mathcal{E}$ is an index set. Finally, in (2.1) $p \in [0,1]$ is the probability level. The analytic description of each of these components is provided in Section 3.2.

In the context of problems that aim at making decisions dealing with uncertain data, an interpretation of such PC problem (2.1) is also referred as a less conservative version of the Robust Optimization problem (BEN-TAL; GHAOUI; NEMIROVSKI, 2009; BERTSIMAS; BROWN; CARAMANIS, 2011). In this approach, after assumptions over a full knowledge of uncertainty set and restrictions on the violation of the constraints that define the problem, one looks at finding suitable solutions whatever the unknown information is revealed. This last interpretation is inherently associated with finding worst-case-oriented solutions, then avoiding any acceptable risk assumption, with the drawback to be generally linked with expensive solutions, also known as the "price of robustness" in the literature (BERTSIMAS;

SIM, 2004). For a comprehensive two-stage adaptive robust UC problem, we refer to (BERTSIMAS et al., 2013) and an adjustable robust OPF dealing with renewable sources approach is found in (JABR, 2013). Here is essential to distinguish the differences from the present work PC proposed method, where the probability is adjusted to the desired level of exposition, turning the risk-return evaluation tangible.

We consider a stochastic HTUC in which we cope with the uncertainty of wind generation through the use of a joint PC. The methodology could be easily extended to deal with different sources of uncertain data, such as load demand variation or other intermittent generations. In the next sections is discussed the evolution of the applications of PC in optimization problems involving energy planning, focusing, when it is possible, in UC problems. Section 2.1 presents works for the most straightforward approach, the Individual PC, i.e. the cases commonly addressed by p-quantiles of the random variables. The last is followed by sampling approaches in Section 2.2, where discrete approximations substitute the use of analytic probability distribution functions. Finally, Section 2.3 depicts the status quo on strategies of PC when it is taken into account its original formulation in UC and energy planning problems in a broad sense.

2.1 UNIT COMMITMENT WITH INDIVIDUAL PROBABILISTIC CONSTRAINTS

In probabilistically constrained optimization problems, it is usual, rather than take the PC jointly, to consider each constraint separately. In these cases, one is aiming to deal with PC of type $\mathbb{P}[h_e(\dot{x}, \tilde{\xi}) \leq 0] \geq p$, for $e \in \mathcal{E}$ an index set, in (2.1), where e is the index of the individual constraint. The most common use is when $h_e(\cdot)$ is separable (i.e an affine function involving the random variable, decision variable and parameters) and there is a single random variable for each constraint, meaning that the PC can be written as an analytic distribution function involving quantiles, simplifying the implementation in the optimization problem. In cases in which the constraint contains affine functions of independent random variables, techniques as proposed in Appendix A.3 can also be applied.

The pioneering work (OZTURK; MAZUMDAR; NORMAN, 2004) applied individual PC in a thermal UC problem, where the probability level is changed iteratively to approximately solve a model with a joint PC in load requirement constraints. The strategy involves manipulations on the quantile of the distribution function. The method also uses LR to separate the dual problem into slave subproblems that define which generator must be online for each stage. With the on/off commitment established, the master primal problem sets the schedule of generation. After the convergence of the primal-dual problem, the solution is evaluated in a joint probability distribution pespective. This calculation is performed with the use of the code of Genz (GENZ, 1992). If the probability level is not met, given an acceptable deviation, heuristics update the quantile and the algorithm restarts to provide another schedule.

In (DING et al., 2010) the focus is on the thermal UC with wind generation farms. The paper considers uncertainties in demand, outages of generating units and wind farms production. A PC is defined to ensure that the load requirement constraint is met, given a probability level. A quantile approach is used to deal with the uncertainty of load, wind power and energy prices. For wind generation modelling purposes, rather than use standard Cumulative Distribution Function (CDF) distributions, such as Gaussian or Weibull, the authors used quantiles from an empirical table. The PC that considers outages of generation units brought combinatorial explosion issues to the problem. To overcome this, the authors studied the number of necessary units and the level of the normal moment order that could substitute the original discrete distribution with acceptable error. In the end, all PCs are transformed into affine constraints, using quantile approaches.

In a recent contribution, (WU et al., 2016) presented a UC system that deals with uncertainty on loads and wind power to provide load supply with high reliability. In this problem, PCs were used to model the loss of load probability, the wind power utilization and the transmission lines overloading. Through a two-stage stochastic programming problem, the dispatch is defined in the first stage. The second stage, for fixed scenarios of load and wind power, the objective is to minimize the penalty cost of load shedding and wasted wind power. The work has strategies on how to define correctly the quantiles since

it deals with truncated and correlated normal distributions. To manage the issue in a way that ensures the desired reliability level, the quantiles of a deterministic version of the PC are iteratively updated into the two-stage stochastic problem, using the scenarios of load and wind power from Monte-Carlo sampling to establish the quality of the solution.

By the presented studies, it is clear that if one aims at using individual PC to control the joint probability level of the problem, it is necessary to use heuristic schemes to manage the probability to the desired risk level. In many cases, it is hard to reproduce these heuristics, usually combined with strategies that set very high individual probability levels, which is a disadvantage of these methods. Another common characteristic is that those methods ensure only feasible solutions to the joint PC problem, and so the advantage of obtaining optimal costs is disregarded. On the other hand, the problem formulation for these cases are much easier to be dealt with, and one does not have to work with multidimensional probability functions embedded into the optimization problem. Here, another advantage of these approaches is the lower computational burden, providing faster solutions.

The accuracy of individual PC problems when compared to joint ones is checked for the load requirement constraint in Chapter 4. Similar comparisons for applications in hydro reservoir management are provided in (van ACKOOIJ et al., 2014).

2.2 PROBABILISTIC CONSTRAINTS WITH DISCRETIZED RANDOM VARIABLES

Another common approach for dealing with uncertainty using PC in UC framework are the ones based on discretized distributions by sample approximation methods. These procedures have their sources from SAA methods (SHAPIRO, 2003; PAGNONCELLI; AHMED; SHAPIRO, 2009), where the authors propose solution for the following expectation problem:

$$\min_{\dot{x}\in X}\{c(\dot{x}) := \mathbb{E}[h(\dot{x}, \tilde{\xi})]\} \tag{2.2}$$

where $h(\cdot)$ compounds the objective function and establishes the relationship between the decision variable $\dot{x}$ and the random (vector) $\tilde{\xi}$. By this method, rather than using the

real distribution of the random vector $\tilde{\xi}$, the strategy consists of solutions computation submitted to Monte Carlo realizations of the random vector. After assumptions on the generation of independent identically distributed (iid) and properties over the function $h(\cdot)$ (including convexity and measurability) the authors ensure that the solution x_N of the approximation problem $\min_{\dot{x}\in X} \{\hat{c}_Q(\dot{x}) = \frac{1}{Q}\sum_{q=1}^{Q} h(x, \xi_q)\}$ is a statistical estimator of the optimal solution of the original problem (2.2). In this last equation, Q is the number of samples, and ξ_q is a sample realization of the random variables. Evolutions on SAA methods by the use of importance sampling techniques are discussed in (BARRERA et al., 2016).

Similarly to the strategy used by SAA methods, which deal with the uncertainty by the objective function, the random variables can be moved to the constraints. An essential work to address these sample based methods is (LUEDTKE; AHMED, 2008). The authors propose a discretized form to deal with PC, where random samples are used to replace the direct use of the underlying random vector. The method introduces the conditions that the problem (2.1) can be approximated by:

$$\begin{aligned} \min_{\dot{x}\in X} \quad & c(\dot{x}) \\ \text{s.t.:} \quad & \tfrac{1}{Q}\sum_{q=1}^{Q} \mathbb{I}\left[h_e(\dot{x}, \tilde{\xi}_q) \leq 0\right] \geq p_{sb} \end{aligned} \tag{2.3}$$

where $\mathbb{I}(\cdot)$ is the indicator function that is equal to one when the constraint $h_e(\dot{x}, \tilde{\xi}_q) \leq 0$ is met and zero otherwise. On the method it is common the use of a higher probability level $p_{sb} > p$, (where p is the former one), that under fairly general assumptions over the funtions $c(\cdot)$, $h_e(\cdot)$ and the random vector $\tilde{\xi}$, the problem (2.3) offers inner and outer solutions to the original PC problem (2.1). In these approaches, it is usual to simulate the indicator function by the use of mixed-integer strategies, being necessary, therefore, the use of solvers that deal with these type of problems. It is important to mention that when p_{sb} is chosen in such a way that $p_{sb} = 1$, the method agrees with the scenario approximation robust variant (CALAFIORE; CAMPI, 2005).

Many recent applications of sample-based methods for UC problems are inspired by the cited works (SHAPIRO, 2003; CALAFIORE; CAMPI, 2005; LUEDTKE; AHMED, 2008).

The paper (WANG; GUAN; WANG, 2012) considers the uncertainty of wind power within an embedded PC problem into a two-stage model. In the first stage, the algorithm defines the wind farm commitment, being the second stage reponsible to measure the differences between the real generation, included the wind sampled scenarios, and the commited energy. The shortage is penalized. The evaluation is set for three different generation policies, defined by the type of used PC: (a) the first (less restrictive), establishes that the sum of utilization of the wind power for the 24 hours period is greater than a percentage of the total available wind power; (b) the second is equivalent to an individual PC for each period; (c) the third, more restrictive, takes into account a joint PC for all periods. The solution method involves heuristics in order to validate each candidate to optimal solution, defining iteratively upper and lower bounds to the original problem. A similar approach is considered in (ZHAO et al., 2014) where, combined with the PC that controls the occurrence of load imbalances, there is also an expected value constraint that defines the average amount of wind power usage for all periods.

The work (ZHANG et al., 2017) is an application of UC with uncertainty on the load and wind generation in a two-stage, PC framework. The authors argue that the method is adherent to the use of modern non-parametric wind forecast models. It is proposed a bilinear reformulation for the classic Big-M method, that is introduced in the MILP formulation by auxiliary variables and constraints (McCORMICK, 1976). Using a variant of Benders decomposition (BENDERS, 1962), feasibility cuts are added to the master problem to validate the first-stage decision variables into different scenarios of the second-stage problem. The results show that the proposed bilinear method is more efficient than the usual Big-M one. It is also stated that the use of the proposed variant of Benders decomposition is more efficient than solve an equivalent large stochastic MILP.

Those analyzed works and the cited materials therein evidence that PC in programming problems are regularly found in the literature to cope with UC problems. From this analysis, the discretized version of PC is becoming popular by its more explicit interpretation into the algorithms, since the PC is handled directly in the programming problems by the use of indicator functions. Aids the wide spreading of the method the

mainstream development of efficient MILP solvers, together with the versatility provided by the use of samples of random variables into optimization problems. A drawback on the use of the discretized version of PC is the fact that commonly the sample size that ensures the desired probability is sometimes huge, generating problems that in many cases are intractable in terms of the dimension of the problem. Another disadvantage is the current use of ad-hoc strategies to the specific problem, reducing the flexibility of the method for different applications.

We remark that there are methods to reduce the variance of the original random vector, as the linear controlling and importance sampling techniques (SHAPIRO; DENTCHEVA; RUSZCZYŃSKI, 2009), which commonly involves operations with auxiliary (distribution) functions in order to reduce the dispersion of the random vector. These techniques can be an interesting procedure to reduce the sample size, ensuring that the statistical properties of the random variable are well established in the optimization problem. The importance sampling technique in PC problems framework is discussed in (BARRERA et al., 2016).

2.3 JOINT PROBABILISTIC CONSTRAINTS USING ORIGINAL DISTRIBUTION FUNCTIONS

Initial applications of PC in energy planning optimization problems using the underlying probability distribution goes back to Prékopa's work on systems for controlling natural inflows, e.g. (PRÉKOPA; SZÁNTAI, 1978). This study considers multivariate random (normal and gamma) distributions for rivers streamflows, where graph theory is used to design the river system mathematically. The problem aims at offering a tool to develop a flood control system by using the reservoirs spread over the river. The central idea is to minimize the total building costs. The optimization problem is solved by a supporting hyperplane method (VEINOTT, 1967) defined over a convex polyhedral set, in a similar research line as presented in Section 3.1. Further backtesting processes to validate the solutions are provided in (PRÉKOPA; SZÁNTAI, 1979). For different applications of joint PC we refer to (SZÁNTAI, 1988).

The procedures introduced by Prékopa were extended in many works that deal with

hydro management for energy production planning. We find in literature, e.g. (PRÉKOPA, 1995), that strategies are commonly used in applications for hydro reservoir management taking into account medium and long-term horizons, with time periods defined weekly or even monthly, since the consequences of the policy strategies have cumulative properties as, for instance, the analysis of the level of water in dams over time. Recent applications in this field are provided in (van ACKOOIJ et al., 2011, 2014). In these last works, the authors focus on hydropower production, comparing different ways of modelling the PCs, and considering a 48-hours time frame. The hydro production function and the ED problem are defined in a way that the feasible set is ensured to be continuous and convex, and no other power generation sources are considered. For an application that establishes policies that respect probabilistic surfaces of risk aversion for dams' stored energy in a hydrothermal ED problem context, we refer to (RODRIGUES; DINIZ; PRADA, 2017).

In the line of research of multistage stochastic programming problems applied to energy management, a relatively recent topic in PC optimization problems deal with partially known random process (ANDRIEU; HENRION; RÖMISCH, 2009). In contrast with static PC, the decisions of these models are kept dynamic and depend both on the realizations of random variables (i.e. the past) and the still unknown ones. The method uses conditional distribution functions, where the proposed model considers continuous (joint) distribution functions and discretized decision variables. The application is considered for a single water reservoir in two-stage and three-stage settings, where the authors establish that the procedure can be efficiently treated up to three stages, due to the complexity of the discrete treatment.

In the present work, the proposed algorithm to deal with HTUC by the use of PCs under a non-continuous decision variable setting is presented in Chapter 4. A similar approach is proposed in (ARNOLD et al., 2014) in a short-term UC system with hydro and wind generation where the supporting hyperplane method is inserted into a branch-and-bound framework. In this approach, during the iterations of the branch-and-bound algorithm, the solution is currently tested in order to certify that the PC holds. If the PC is not met it is chosen heuristically between add a supporting hyperplane or perform a

branching for a fixed binary variable. The last procedure differs from algorithm proposed by the thesis, where the MILP algorithm is taken separately from the process of adding supporting hyperplanes. The advantage of the present process is that a single PC approximation cutting plane is added per iteration, based on a feasible solution regarding the remaining constraints of the UC problem. Therefore, the procedure tends to perform less calculations of gradients, being a remarkable characteristic since the computational cost is related with the dimension of the random variable. Another advantage of the proposed method is that, at each iteration of the algorithm, it offers an upper bound to the problem (and an optimality gap), and so a valid policy is always available.

The provided bibliography survey suggests that the use of PC (with different approaches) is a current trend in the study of UC problems. As stated, Individual PCs presented in Section 2.1 have the advantage of the simplicity of their characterization in the optimization problem. Nevertheless, commonly the risk associated with the provided optimal solutions are undervalued when the probability measure is taken jointly. Sample-based methods presented in Section 2.2 are a fast evolution trend, supported by the improvements of solvers that deal efficiently with their mixed integer-linearity. An advantage of the sample-based methods is the flexibility in the establishment of the uncertainty in the optimization problem. The use of sampled scenarios is an appealing alternative when the random variable distribution functions are discrete or hard to be defined analytically. The drawback is the lack of precision on their solutions in terms of probability level when regarding their real original continuous distributions.

The present document sets its efforts in the development of HTUC problems taking the PC modelled as in the works of Section 2.3. These algorithms have the common characteristic of working with the probability distributions precisely, i.e. the optimal solution respects an a priori defined probability level. A drawback of the method is the complexity involving the use of these distributions functions since it commonly involves working with multidimensional non-linear functions in the optimization problem. In the following chapters, we handle these functions by a cutting plane method that approximates the feasible set defined by the PC iteratively.

3 MATHEMATICAL BACKGROUND FOR PROBABILISTIC CONSTRAINED OPTIMIZATION PROBLEMS

Among the classes of optimization problem approaches that deal with uncertainty, the present work focus on the use of PC programming. This chapter aims at introducing mathematical instruments used by PCs, necessary to a better comprehension of the here proposed methods.

In Section 3.1, we motivate the use of PC in programming problems by a simple numerical application. It is presented a classic algorithm using the instruments discussed in details in the subsequent sections. Section 3.2 depicts the two PC's structures used by the applications of the Thesis, specifically the affine and the bilinear ones. Section 3.3 presents the procedure to compute gradients of appropriate functions used to define the supporting hyperplanes. Section 3.4 discusses the used affine structures, important also in order to establish the uniformity of variables used in the PC.

Complementing the mathematical tools sections, elementary concepts used along the Thesis are in Appendix A. Appendix A.1 is a brief revision of essential analysis tools that currently are used in PC applications. Appendix A.2 presents notions of $r-$concave functions. Appendix A.4 is an important section that links the supporting hyperplane theory to the pattern of functions used by PCs. Appendix A.5 has some considerations about the calculation of multivariate distribution functions. An important step in the proposed algorithms are the interpolation procedure, explored in Appendix A.6, where we describe the used bisection algorithm and others.

3.1 ALGORITHM FOR A TOY PROBLEM

The present section aims to present an algorithm to solve a simplified thermal ED problem, designated as Toy Problem (TP). The main objective is to motivate its use, clarifying the application of the mathematical instruments described in the next sections of the chapter. These instruments, when used throughout the algorithm, are linked to

the appropriate section for better comprehension. To achieve this, we model a thermal-wind ED problem where it is asked to respect a lower threshold for the load requirement constraints.

The TP system is composed of a single bus, where a load is supplied by eight thermal plants and one wind farm. The dispatch is considered in a centralized framework, where the ISO aims at minimizing the expected production cost over a two-hours planning horizon. The operational characteristics of the thermal plants are presented in Table 1.

Table 1 – Thermal plants operational characteristics

Plant	gt_1	gt_2	gt_3	gt_4	gt_5	gt_6	gt_7	df
Max. gen. $[MW]$	455	162	162	130	455	55	55	∞
Un. Var. cost $[\$/MW]$	15	18	20	25	35	50	55	10^3

Source: Thesis results.

Where it is shown the maximum generation and the unitary variable costs for the thermal plants. The thermal costs are represented in its simplest form, by a fixed price that multiplies the productions. They possess null minimum generation limits. Furthermore, df represents the slack variable, the deficit power generation.

The wind farm generation is given by the bivariate normal random vector that establishes the temporal dependence over two stages. Its mean μ_{or} and covariance matrix Σ_{or} are given by:

$$\mu_{or} = \begin{bmatrix} 3.7 \cdot 10^2 \\ 4.1 \cdot 10^2 \end{bmatrix} \quad \Sigma_{or} = \begin{bmatrix} 5.7 \cdot 10^4 & 4.9 \cdot 10^4 \\ 4.9 \cdot 10^4 & 4.9 \cdot 10^4 \end{bmatrix}. \tag{3.1}$$

The load demand ld, in $[MW]$, for the two stages is given by the following vector:

$$ld = [1195.20 \quad 1394.40]^\top. \tag{3.2}$$

Considering these characteristics, we provided the ED model by the following programming problem:

$$\min \sum_{t=1}^{2} \left(\sum_{j=1}^{7} c_j \dot{gt}_{j,t} + c_{df} \dot{df}_t \right) \tag{3.3}$$

s.t.:

$$\sum_{j=1}^{7} \dot{gt}_{j,t} + \widetilde{gw}_t + \dot{df}_t = ld_t, \forall t \tag{3.4}$$

$$\dot{gt}_{j,t} \leq gt_j^{max}, \forall t, j \tag{3.5}$$

$$\dot{df}_t, \dot{gt}_{j,t} \geq 0, \forall t, j, \tag{3.6}$$

where $\dot{gt}_{j,t}$ is the decision variable representing the thermal generation of plant j in the period t, $\dot{df}_t$ is the deficit generation, c_j and c_{df} are the incremental thermal costs and the deficit cost parameters $[\$/MW]$. The random variables $\widetilde{gw}_t$ represent the wind farm generation $[MW]$. The parameter gt_j^{max} represents the maximum generation of a given plant. The objective function (3.3) is composed by thermal costs c_j and deficit cost c_{df} over the optimization horizon. The load requirement constraints is given by (3.4). Constraints (3.5) and (3.6) define limits on the decision variables.

The interpretation of the underlying problem (3.3)-(3.6) is that its solution depends on the random variables $\widetilde{gw}_t$. A simple procedure to take into account the wind farm generation would be setting a deterministic forecast value of $\widetilde{gw}_t$, turning them into fixed parameters gw_t. This procedure does not consider the possible deviations from the predicted values, thus a complete method that takes into account the stochasticity of the wind farm generation is the PC framework, where the problem (3.3)-(3.6) has the constraints (3.4) replaced by the PC counterpart. For more information about this type of PC, it has a complete description in the following Section 3.4.2 in (3.35). The TP via joint PC modelling is expressed in a compact form as:

$$\begin{aligned} \min_{\dot{x}} \quad & c^{\mathsf{T}} \dot{x} \\ \text{s.t.:} \quad & A_{in} \dot{x} \leq b_{in} \\ & \phi_{\tilde{\eta}}(\dot{x}) := \mathbb{P}[ld^{low} \leq h(\dot{x}) + \tilde{\eta}] \geq p, \end{aligned} \tag{3.7}$$

where the minimum threshold to supply is (freely) defined 10% greater than the predicted load demand, i.e. $ld^{low} = 1.1 \cdot ld$. In this example, matrix A_{in} and vector b_{in} are defined by (3.5) and (3.6). Under certain conditions the set $M(p) = \{\dot{x} \in \mathbb{R}^m | \mathbb{P}[ld^{low} \leq h(\dot{x}) + \tilde{\eta}] \geq p\}$ is convex. If these conditions are met, the optimization problem (3.7) is also convex. The conditions that make $M(p)$ convex regard the class of the function $h(\dot{x})$ and the distribution of $\tilde{\eta}$, that are met since $h(\dot{x})$ is affine and $\tilde{\eta}$ is normally distributed in the TP. Further discussions about convexity of the set $M(p)$ are made in Section 3.2.

The probability level for the problem (3.7) is initially fixed in $p = 90\%$. The decision variable vector $\dot{x} \in \mathbb{R}^m$, that combines all decision variables, is given by:

$$\dot{x} = [\dot{g}_{1,1}, \cdots, \dot{g}_{7,1}, \dot{df}_1, \dot{g}_{1,2}, \cdots, \dot{g}_{7,2}, \dot{df}_2]^\top.$$

We summarize the steps of the approach in Algorithm 1. The present version of the supporting hyperplane method for (continuous) linear problems is derived from the classic algorithm in (PRÉKOPA, 2003).

Algorithm 1 - Supporting Hyperplane Method for Continuous Problems

1: (Initialisation) Definition of correlation matrix R, the structure matrix $B \in \mathbb{R}^T \times \mathbb{R}^m$ and vector $b \in \mathbb{R}^T$. Definition of the Slater Point $x^S \in R^m$ (feasible solution for the problem (3.7) such that $\phi_{\tilde{\eta}}(\dot{x}) > p$). Definition of stopping tolerance $\epsilon > 0$ and counter $k \leftarrow 1$.

2: (Lower bound) Solve the optimization problem (3.7) replacing the PC by its piecewise linear approximation, a set of constraints, as follows:

$$\begin{array}{ll} \min\limits_{\dot{x} \in X} & c^\mathsf{T} \dot{x} \\ \text{s.t.} & A_{in} \dot{x} \leq b_{in} \\ & G_b(x_i) \cdot \dot{x} \leq g_b(x_i), i \in \mathcal{I}, \end{array} \tag{3.8}$$

providing the solution x_k^L, being the related objective function value L_k a lower bound to the original problem (3.7).

3: (Interpolation) By the bisection procedure compute the largest λ such that $x_k := \lambda x_k^L + (1 - \lambda) x^S$ satisfies $\phi(x_k) \approx p$. The solution x_k attains the PC, being a feasible solution for (3.7). Hence, we can define the solution $x_k^U = x_k$ and we are able to define the upper bound $U_k = c^\mathsf{T} x_k$.

4: (Stopping test) Check if $\frac{|U_k - L_k|}{|L_k|} < \varepsilon$ and then stop. The solution is the last feasible solution x_k^U. Otherwise go to the next step.

5: (Oracle call) Call an "oracle" to compute a gradient $\nabla \phi_{\tilde{\xi}}(x_k^U)$, defining $G_b(x_k^U)$ and $g_b(x_k^U)$, adding a new constraint to the set of cutting planes in Step 2.

6: (Return) Set $k \leftarrow k + 1$; Go to Step 2

In Algorithm 1 - Step 1 (Initialisation) is computed the affine structures. Essentially, the challenge is how to write the original PC $\mathbb{P}[ld^{low} \leq h(\dot{x}) + \tilde{\eta}] \geq p$ using the standard form $\mathbb{P}[\tilde{\xi} \leq B\dot{x} + b]$, where $\tilde{\xi} \sim \mathcal{N}(0, R)$, being $R = D \cdot \Sigma_{or} \cdot D'$ a correlation matrix. This convenient standard form is used in the PC algorithms and auxiliary functions of the chapter. A complete description of the procedures to obtain the matrix parameters from the original PC is discussed in the following Section 3.4.2.

The standardization starts evaluating the auxiliary matrix D by the coefficients of the original wind generation covariance matrix Σ_{or}. We skip the procedures to define D and the coefficients B and b, that are described in details in the following Section 3.4.2, Remark 3.4.4. The structure matrix $H_{or} \in \mathbb{R}^2 \times \mathbb{R}^{16}$ is defined as:

$$H_{or} = \begin{bmatrix} 1 \cdots 1 & 0 \cdots 0 \\ 0 \cdots 0 & 1 \cdots 1 \end{bmatrix},$$

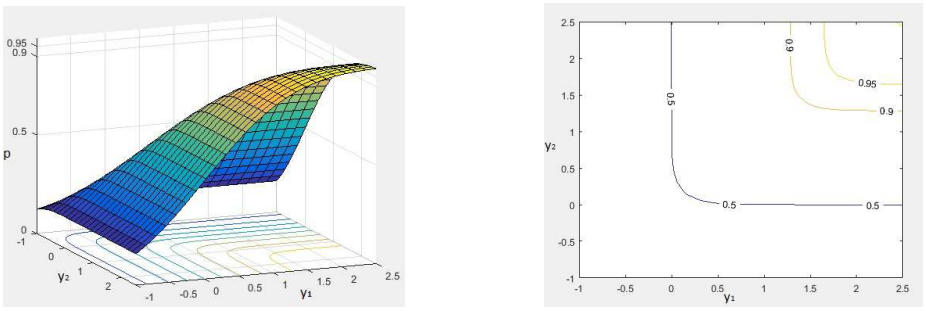
where the series $1 \cdots 1$ are composed of 8 values, one for each generation unit $\dot{g}_{m,t}$ plus the deficit load variable. By these matrix manipulations, rather than use the original wind farm generation distribution $\tilde{\eta}$, the affine transformations described above allow us to use its standard form $\tilde{\xi} \sim \mathcal{N}(0, R)$. The correlation matrix is given as follows:

$$R = \begin{bmatrix} 1 & 0.9284 \\ 0.9284 & 1 \end{bmatrix}. \tag{3.9}$$

The $\tilde{\xi}$'s CDF is shown in Figure 1(a), with the coloured probability level curves projection at $p = 0$ ($y_1 \times y_2$ axis). Some of these last contours are highlighted in Figure 1(b), where in the present example we pick the level curve $p = 90\%$.

Still in Step 1, we must define a Slater solution $x^S \in \mathbb{R}^m$. We can do it by forcing a high generation to the thermal units heuristically. This is done, for example, by the solution of the following optimization problem:

Figure 1 – Cumulative distribution function of the standard random variable $\tilde{\xi}$.

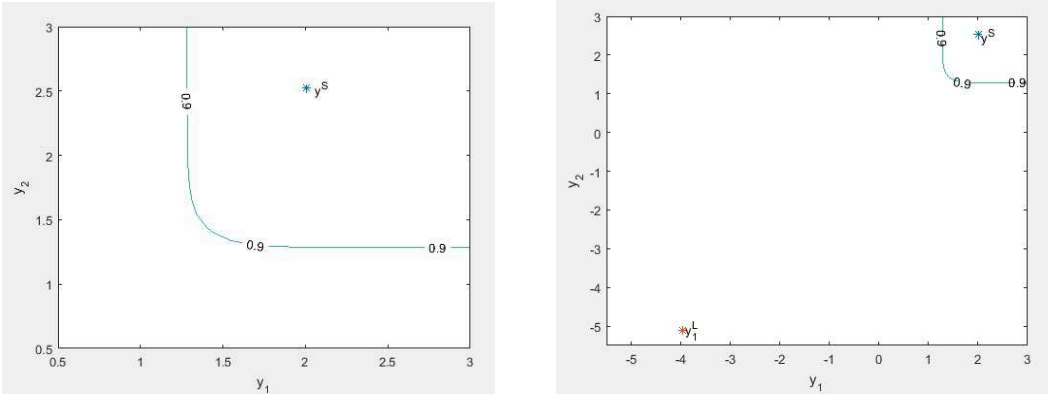


(a) Cumulative distribution function. (b) Probability level curve contours.
Source: Thesis results.

$$\begin{aligned} \min_{\dot{x}\in X} \quad & c^{\mathsf{T}}\dot{x} \\ \text{s.t.} \quad & A_{in}\dot{x} \leq b_{in} \\ & H_{or}\dot{x} = ld \cdot 1.5 - \mu_{or} \end{aligned} \quad . \tag{3.10}$$

In this auxiliary linear problem, we establish a higher demand to be met, considering the wind farm production as its mean. Solving problem (3.10), the optimal solution x^S, after the affine transformation $y^S = Bx^S + b$, is presented in Figure 2(a):

Figure 2 – First iteration solutions of Steps 1 and 2 of Algorithm 1.



(a) Slater point y^S. (b) Lower bound solution y_1^L.
Source: Thesis results.

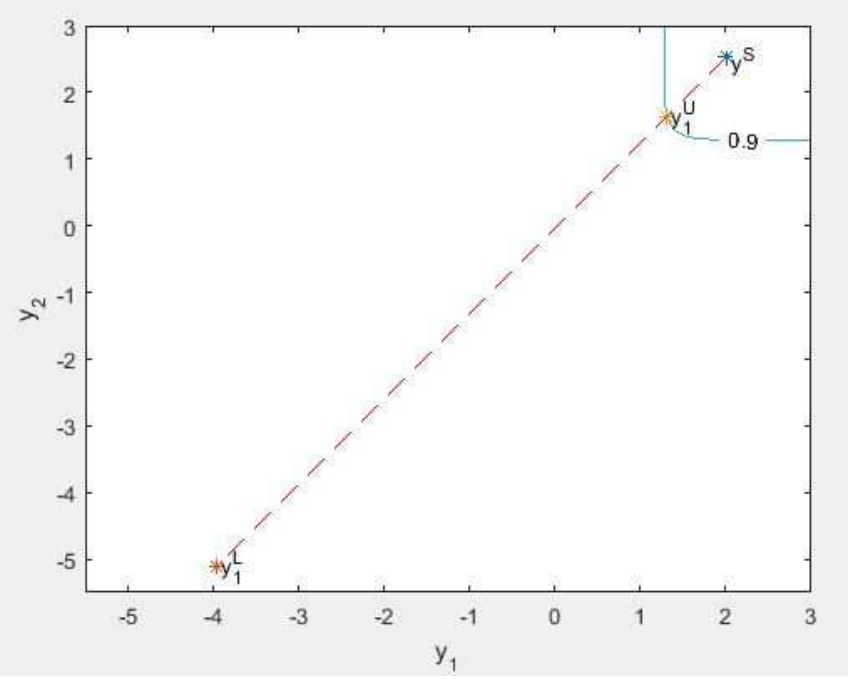
The probability level $p^S = \mathbb{P}[h(x^S) + \tilde{\eta} \geq ld^{low}] = 97.70\%$. This probability $p^S > p = 90\%$, hence higher than the desired minimum probability and so being a valid Slater solution. The last part of Step 1 (Initialisation) is the definition of the stopping tolerance $\epsilon = 10^{-3}$.

Following the Algorithm 1 in Step 2 (Lower bound), in the first iteration $k = 1$ the

linear programming problem (3.8) is solved unconstrained in terms of the load requirement constraints, i.e. the index set $\mathcal{I} = \emptyset$ and so no cuts are added to the problem. As a result, it will produce the solution x_1^L, with optimal value $L_1 = 0$ as a lower bound to problem (3.7). Figure 2(b) plots the lower bound solution $y_1^L = Bx_1^L + b$, that has probability $\mathbb{P}[h(x_1^L) + \tilde{\eta} \geq ld^{low}] \approx 0$ and solution $L_1 \approx 0$. Once obtained the Slater solution x^S and the first lower bound solution x_1^L we can proceed to the next Step 3 (Interpolation).

The Step 3 employs the bisection procedure. The procedure is detailed in Appendix A.6.1, by Algorithm 4. The scope is to find an iterate x_1 such that $\mathbb{P}[h(x^1) + \tilde{\eta} \geq ld^{low}] \approx 90\%$. In the first interpolation step, after 12 iterations of the bisection method reaches a $\lambda = 0.1159$, and so defining the iterate x_1. Since this iterate satisfies all constraints of problem (3.7), it is a (upper bound) solution $x_1^U = x_1$, with optimal value $U_1 = c^\mathsf{T} x_1^U = \$2.50 \cdot 10^5$, and probability level $p_1^U = 0.9000$. Figure 3 presents the result of the interpolation iterate x_1^U by a convex combination of solutions x^S and x_1^L. The dashed red line represents all the possible interpolation solutions between these two last points, noting that $y_1^U = Bx_1^U + b$ rests on the $p = 90\%$ level curve.

Figure 3 – Bisection Procedure finding the upper bound solution y_1^U

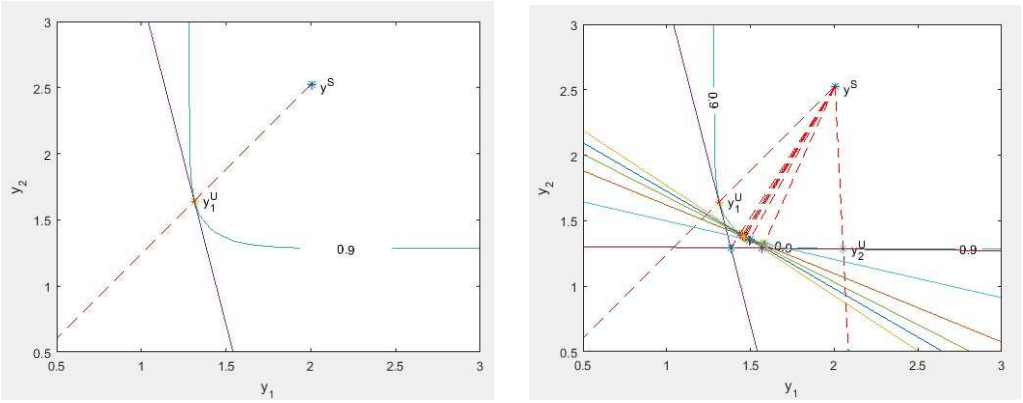


Source: Thesis results.

The next step of Algorithm 1 is the stopping test analysis. $\frac{|U_1 - L_1|}{|L_1|} = \frac{|2.50 \cdot 10^5 - 0|}{|0|} > \varepsilon$

, therefore the condition is not met and we can proceed to the Step 5 (Oracle call). In this step, using the fixed point x_1^U, it is defined the gradient $\nabla\phi_{\tilde{\xi}}(x_1^U)$. The procedure to compute gradients for PC are provided in the following Section 3.3, equation (3.16). The knowledge of the point x_1^U on the boundary of the PC feasible set and $\nabla\phi_{\tilde{\xi}}(x_1^U)$ are the necessary ingredients to calculate a local approximation of the PC feasible set limited by the supporting hyperplane at point x_1^U. For further material in supporting hyperplane theory see Appendix A.4. The supporting hyperplane representation in $y_1^U = Bx_1^U + b$ is presented in Figure 4(a). Note that as we are dealing with a bi-dimensional distribution function, the supporting hyperplane is defined as a line.

Figure 4 – Supporting Hyperplanes on the boundary of the PC feasible set of level $p = 90\%$, defining its local approximation.



(a) Supporting Hyperplane at point y_1^U

(b) Supporting Hyperplanes after 8 iterations of Algorithm 1

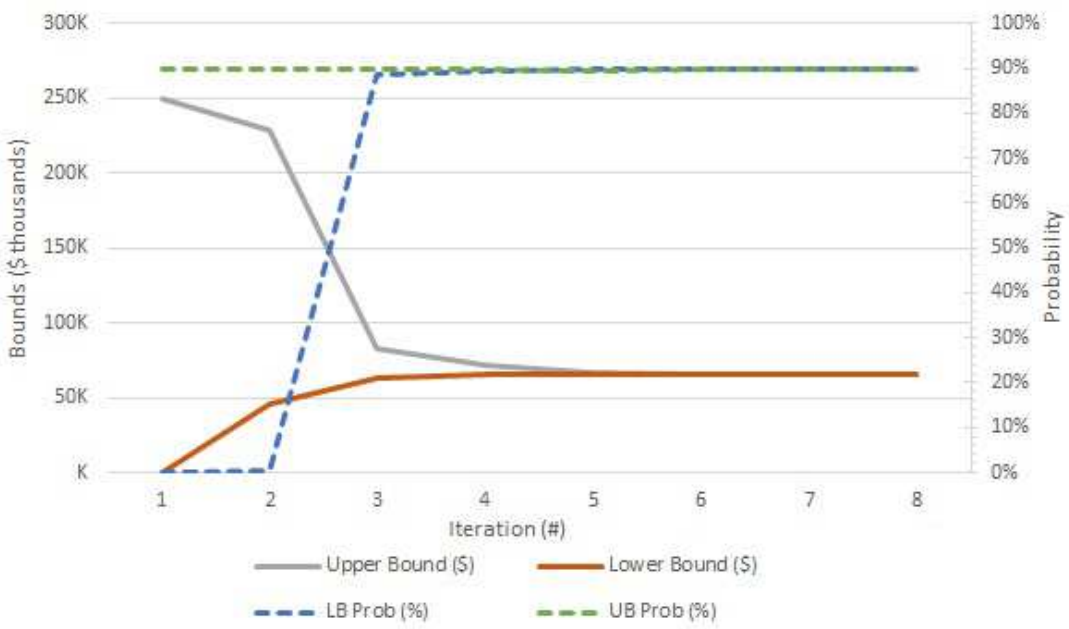
Source: Thesis results.

In the last part of Step 5, we must add constraints $G_b(x_i) \cdot \dot{x} \leq g_b(x_i)$ to the lower bound Problem (3.8) in Step 2. This is done calculating the cuts coefficients $G_b(x_1)$ and $g_b(x_1)$ employing Remark 3.4.2, described below in Section 3.3. By that, we are able to go to Step 6 (Return) setting $k \leftarrow 2$, returning to Step 2 (Lower Bound). We restart the loop solving (3.8) with the calculated cutting plane, and so i=1. We proceed with the steps of the algorithm, and after #8 iterations it converges, stopping in Step 4 (Stopping test) with the representation of the evaluated supporting hyperplanes given in Figure 4(b). Note that in the last figure, after obtaining y_1^U and y_2^U, the local search for further solutions are confined to the set limited by the two associated supporting hyperplanes.

As a result, all the subsequent lower bound solutions y_k^L are iteratively restrict to smaller space, as can be seen by the concentration of dashed lines. Essentially this is the core of the algorithm, i.e., incrementing iteratively the lower bound problem (3.8) defining a local approximation of the PC, up to the convergence.

The evolution through the algorithm iterations are presented in Figure 5. In this graph, it is shown that the algorithm converges to an upper bound $U_8 = \$65.28 \cdot 10^3$ and an equivalent lower bound $L_8 = \$65.28 \cdot 10^3$ reaching the optimality gap of $\frac{|U_8 - L_8|}{|L_8|} = \frac{|65.28 - 65.28|}{|65.28|} \approx 0$. Note that both, upper and lower probabilities, get close to $p = 90\%$ after the iteration #3, resulting in an efficient algorithm for this simplified example. The elapsed time, running the code in Matlab was 2.3 sec.

Figure 5 – Convergence of the supporting hyperplane PC problem



Source: Thesis results.

The optimal solution x_8^U is presented in Table 2 for the two periods. In this table we find that the thermoelectric generation are ranked by crescent incremental costs, as expected for an optimal solution, remembering that the costs are given in Table 1.

To establish the sensitivity of the model to the probability level, Table 3 presents comparison results for three probabilities 75%, 90% and 95%. The second and third columns display the total generation of the last thermal plant, in the merit order, for

Table 2 – Optimal thermal generation

Thermal plant	gt_1	gt_2	gt_3	gt_4	gt_5	gt_6	gt_7	df
Per. t=1 $[MW]$	455	162	162	130	384.37	0	0	0
Per. t=2 $[MW]$	455	162	162	130	455	55	12.51	0

Source: Thesis results.

periods $t = 1$ and $t = 2$, where we observe that a higher probability level induces more dispatch. This is only possible with the associated higher total costs. Note that, for the case $p = 95\%$ the model requires more generation than available for the second stage, and this situation is dealt with through the deficit cost slack variable, with total generation $df_2 = 17.76$ MW. We also verify that higher probability (risk) levels induce higher total costs. Due to the problem simplicity, the simulations do not offer trends or conclusions related to the elapsed time and number of iterations in relation to the risk level.

Table 3 – Probability Sensibility

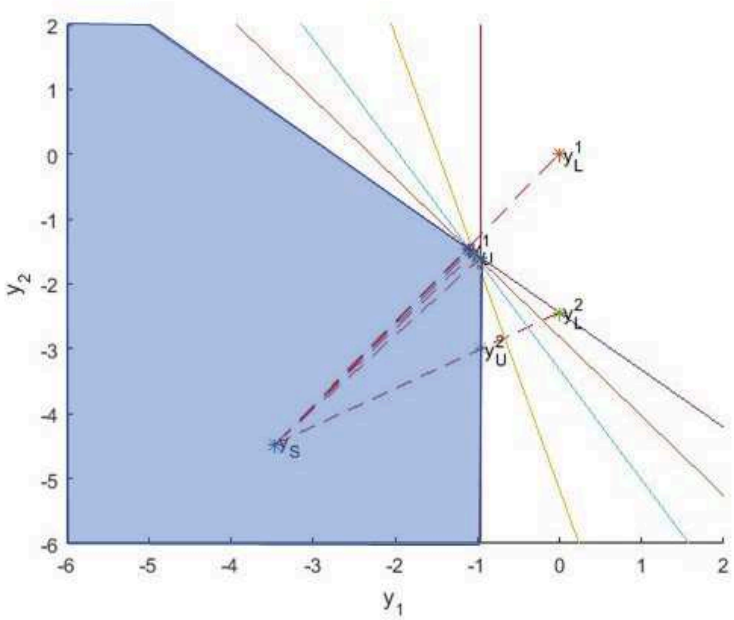
Prob.	L. Plant t = 1 [MW]	L. Plant t = 2 [MW]	Total Cost [$]	#It.	Elapsed Time [sec.]
75%	$gt_5 = 228.50$	$gt_5 = 400.22$	$54.50 \cdot 10^3$	8	2.83
90%	$gt_5 = 384.37$	$gt_7 = 12.51$	$65.28 \cdot 10^3$	8	2.3
95%	$gt_7 = 38.8$	$df_2 = 17.76$	$92.73 \cdot 10^3$	6	1.86

Source: Thesis results.

Up to now, the proposed analysis has only a lower threshold to the PC, as shown in equation (3.35) $\phi_{\tilde{\eta}}(\dot{x}) := \mathbb{P}[ld^{low} \leq h(\dot{x}) + \tilde{\eta}] \geq p$. Still, the strategy to obtain Slater solution described by problem (3.10) fits exclusively to the ones limited below. Contrasting, it is interesting to check the behaviour of the algorithm when it is requested a rectangular joint PC as $G_{\tilde{\eta}}(\dot{x}) := \mathbb{P}[ld^{low} \leq h(\dot{x}) + \tilde{\eta} \leq ld^{up}] \geq p$. At this point we are interested in set up visually how the solutions are interconnected during simulations. Therefore, Figure 6 is displayed.

This simple simulation in Figure 6 was performed setting $p = 50\%$ with $ld^{low} = 0.5 \cdot ld$ and $ld^{up} = 1.5 \cdot ld$. As checked in Section 3.4.1, each approximation iteration for the bilateral probability distribution is given by a single constraint. If we work with the linear transformation $\dot{y} = B\dot{x}$, only to make possible the visualization of the constructed

Figure 6 – Supporting hyperplane method for rectangle PC problem after 6 iterations



Source: Thesis results.

cutting planes in a graph, we check the feasible region $\dot{x}$ in light blue.

By this last figure, we certify that the feasible set defined by the cutting planes is pointed to "south-west", contrasting with Figure 4 and its "northeast" pattern. This indicates that for this specific case of PCs with bilateral probability distributions, the lower threshold constrains more the problem than the upper threshold. This situation happens because, as establish by Remark 3.4.3 bellow, the matrix A and B become negative in standardize process of the random vector $\tilde{\eta}$, since the random vector and the decision variable vector have both the same signs in the original equation. Since we are minimizing the programming problem, and the lower bound outer solutions are pointed to "northeast" in Figure 6, it is natural that the lower bounds of the probability distribution were active more often. We would have a different situation if the random vector and the decision variables had opposite signs in the original equation. In such case, we would expect a figure similar to Figure 4(b). Further discussions of this type of bilateral separable PC are made in Remark 3.4.3 in the following Section 3.3.

Consider the generic formulation of a PC optimization problem:

$$\begin{aligned} \min_{\dot{x}\in X} \quad & c(\dot{x}) \\ \text{s.t.} \quad & \mathbb{P}[h_e(\dot{x},\tilde{\xi}) \leq 0, e \in \mathcal{E}] \geq p, \end{aligned} \tag{3.11}$$

where $X \in \mathbb{R}^m \to \mathbb{R}$, a continuous set, $\tilde{\xi}$ is a T-dimensional random vector, $\mathbb{P}$ is probability measure induced by the random vector $\tilde{\xi}$ in $\mathbb{R}^T$, as established in Appendix A.1, $h_e : \mathbb{R}^m \times \mathbb{R}^T \to \mathbb{R}$, $e \in \mathcal{E}$ with $|\mathcal{E}| = E$ an index set, and $p \in (0, 1]$ is the probability level, set as parameter. In (3.11), the higher is the probability level p, more restrict is the feasible set defined by PC.

To ensure that we are dealing with a convex feasible set when working with PCs, we must analyse, besides the properties of the functions h_e, the properties of the random vector $\tilde{\xi}$. In this sense, Theorem 10.2.1 in (PRÉKOPA, 1995) states that if $h_e(\dot{x},\tilde{\xi})$ are quasi-concave functions in $\mathbb{R}^m \times \mathbb{R}^T$ with $\tilde{\xi}$ a continuous distribution with log-concave PDF, the mapping $\dot{x} \mapsto \mathbb{P}[h_e(\dot{x},\tilde{\xi}) \leq 0, e \in \mathcal{E}]$ is a log-concave function in $\mathbb{R}^m$. The last result are used in the next subsections, where we present the two structures of the function $h_e(\dot{x},\tilde{\xi})$ that are covered in the present Thesis.

3.2.1 Separable Affine Probabilistic Constrained Case

In the first topic of the thesis, motivated by Section 3.1 and developed in Chapter 4, we are interested in the case that $h_e(\dot{x},\tilde{\xi})$ has a separable affine structure, i.e. which we write in the convenient formulation where the decision variables are separated from the random vectors, as in $h_e(\dot{x},\tilde{\xi}) = \tilde{\xi} - h_e(\dot{x})$, where $h_e(\dot{x})$ is an affine function. The use of such structures is justified since we work with load requirement constraints for the first topic of the thesis, as will be clarified in the application Section 4.2, that comes up with a separable constraint function naturally. In this case, $h_e(\dot{x},\tilde{\xi})$ is affine in $\mathbb{R}^m \times \mathbb{R}^T$, so a concave function. We can conclude that the last function is also log-concave, applying the results discussed in Appendix A.2. Applying Theorem 10.2.1 in (PRÉKOPA, 1995), we find

that, if considered affine assumptions for $h_e(\cdot)$ and log-concavity for the density of $\tilde{\xi}$, the PC mapping in (3.11) is log-concave. It is well known that the domain of (log)-concave functions is convex, implying in a feasible set of the PC defined as:

$$M(p) = \{\dot{x} \in \mathbb{R}^m | \mathbb{P}[\tilde{\xi} \leq h_e(\dot{x}), e \in \mathcal{E}] \geq p\}, \tag{3.12}$$

being also convex. The convexity of feasible sets defined by log-concave PC is useful, for instance, in the supporting hyperplanes definitions, as shown in Appendix A.4. Provided the separated form of function h_e, we can define the CDF in such a way that:

$$F_{\tilde{\xi}}(h_e(\dot{x})) := \mathbb{P}[\tilde{\xi} \leq h_e(\dot{x}), e \in \mathcal{E}], \tag{3.13}$$

where this multidimensional distribution is taken jointly and so $|\mathcal{E}| = T$ has the same size of the $T-$dimensional random vector $\tilde{\xi}$. Regarding the probability measure $\mathbb{P}[\cdot]$, if it is requested that the distribution function holds for equalities, i.e $\mathbb{P}[\tilde{\xi} = h_e(\dot{x}), e \in \mathcal{E}]$, we are asking the measure for a 1-point set, a countable set or at most in a set of lower dimension $\mathbb{R}^{T'}$, where $T' < T$. For all these cases, we are actually dealing with null sets, as stated in Appendix A.1 and so $\mathbb{P}[\tilde{\xi} = h_e(\dot{x}), e \in \mathcal{E}] = 0$.

If $\tilde{\xi}$ has independent components, the CDF can be split into the marginal CDFs $F_{\tilde{\xi}}(h_e(\dot{x})) = F_{\tilde{\xi}_1}(h_1(\dot{x})) \cdot F_{\tilde{\xi}_2}(h_1(\dot{x})) \cdots F_{\tilde{\xi}_e}(h_e(\dot{x}))$, and the same procedure works for groups of independent components of $\tilde{\xi}$. In some cases, it would be necessary to deal with sum of stochastic variables. This is the case, for instance, when there are two or more random sources injecting energy in the same bus. We introduce how to deal with these problems in Appendix A.3.

The present material offers numerical experiments considering normally distributed random vectors, as set in Sections 3.1 and Chapter 4. The following remark checks the log-concavity of such structures:

Remark 3.2.1 *Let $\tilde{\xi}$ be a normally distributed random vector such that $\tilde{\xi} \in \mathbb{R}^T$ with mean μ and positive definite covariance matrix Σ. Considering those assumptions, $\tilde{\xi}$ has a log-concave PDF and CDF.*

Proof.

Given $y \in \mathbb{R}^T$, the probability density function of $\tilde{\xi}$ is given by:

$$f(y) = \frac{e^{-\frac{1}{2}(y-\mu)^\top \Sigma^{-1}(y-\mu)^\top}}{\sqrt{det(\Sigma)(2\pi)^T}}. \tag{3.14}$$

It is direct from this equation that:

$$\log\left(f(y)\right) = \log\left(\frac{1}{\sqrt{det(\Sigma)(2\pi)^T}}\right) + \log\left(e^{-\frac{1}{2}(y-\mu)^\top \Sigma^{-1}(y-\mu)^\top}\right) \Leftrightarrow$$
$$\Leftrightarrow \log\left(\frac{1}{\sqrt{det(\Sigma)(2\pi)^T}}\right) - \frac{(y-\mu)^\top \Sigma^{-1}(y-\mu)^\top}{2},$$

where $\log\left(\frac{1}{\sqrt{det(\Sigma)(2\pi)^T}}\right)$ *is a constant. Given that* Σ^{-1} *is positive definite, the quadratic term is convex, and by the negative sign, the whole expression is concave. It is known that if* $log(f(y))$ *is concave,* $f(y)$ *is log-concave. Applying Theorem 1, we conclude that its CDF is also log-concave.* □

In addition to the multivariate normal CDF, in (PRÉKOPA, 2003) and (DENTCHEVA, 2009) is provided a list of common distributions which the log-concavity or at least quasi-concavity property are established. Among them, we find, for instance:

- Uniform distribution: If is defined in a convex subset of $\mathbb{R}^T$, its PDF is log-concave.
- Gamma distribution: Defined properly its multivariate distribution function, the joint CDF of the components $\tilde{\xi}$ is log-concave.
- Wishart, Beta and Dirichlet distribution: With specific assumptions concerning the fixed parameters and variables, their PDFs are proved to be log-concave.
- Cauchy and Pareto distributions: With proper assumptions on their parameters, quasi-concave CDFs are defined.

The work (HENRION; STRUGAREK, 2008) defines conditions to ensure concavity using specific PDFs having independent random components and considering a broader class of functions $h_e(\cdot)$, where concavity holds for large enough probability levels. These functions define what is called the *eventual convex* sets, being extended in (HENRION;

STRUGAREK, 2011; van ACKOOIJ, 2015) to different classes of functions (Copulas). Among functions with independent components that preserve the eventual concavity, we find the Weibull PDF, commonly found to model the speed of wind in wind farms generation studies. A bibliography survey on the use of alternative distributions functions is also provided in the Appendix B.

3.2.2 Bilinear Probabilistic Constrained Case

In the second topic of the Thesis, developed in Chapter 5, we focus on the case that $h_e(\dot{x}, \tilde{\xi})$ has a single bilinear structure, i.e., structures that $e = 1$ and the random variables appear in a scalar multiplication with the decision ones, as in $h_e(x, \tilde{\xi}) = \tilde{\xi}^\top x$, for instance. Specifically, the scope is to contribute in the extension of the concepts that make the set $M(p) = \{x \in \mathbb{R}^m | \mathbb{P}[a \leq \tilde{\xi}^\top x \leq b] \geq p]\}$ (locally) convex, by looking for properties that makes the CDF $\mathbb{P}[a \leq \tilde{\xi}^\top x \leq b]$ locally $r-$concave, a generalization of the concept of concavity (HENRION; STRUGAREK, 2008). The $r-$concavity property in the PC context is discussed in Appendix A.2. The former studies in problems of such structures go back to the classical papers of (PANNE; POPP, 1963; KATAOKA, 1963), verifying that the set $M_u(p) = \{x \in \mathbb{R}^m | \mathbb{P}[\tilde{\xi}^\top x \leq b] \geq p]$ is convex for $p \geq 50\%$, where $\tilde{\xi} \sim \mathcal{N}(\mu, \Sigma)$ being $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ a nonsingular covariance matrix. The analysis of these structures are extended in (HENRION, 2007), where the former concepts applied to normal distributions are generalized to the class of elliptically symmetric distributions, and a full characterization of the critical probability level above which compactness and nonemptiness of the PC can be ensured.

3.3 GRADIENTS FOR AFFINE PROBABILITY CONSTRAINTS FUNCTIONS

The objective of the section is to provide the mathematics instruments to calculate gradients, enabling the use of supporting hyperplanes, as discussed in Appendix A.4. Supporting hyperplanes are used in the optimization algorithms of the first topic of the present work. The next approach is established in Section 6.6.4 (PRÉKOPA, 1995) and

for the case of multivariate standard normal distribution function in (PRÉKOPA, 1970). We define the short notation of the distribution $\phi_{\tilde{\xi}}(\dot{y}) := F_{\tilde{\xi}}(h_e(\dot{x}))$, where $\dot{y} = h_e(\dot{x})$ is an affine transformation with components as in Section 3.2. By the following analysis, we verify that the evaluation of the gradient $\nabla\phi(\dot{y})$ is reduced to the calculation of T CDFs, each of them with dimension $T-1$. As clarified below, these $(T-1)$ probability distributions have mean and covariance matrix dependent from the original $\phi(\dot{y})$. Lemma 1 in (van ACKOOIJ et al., 2010) summarizes the calculation of these gradient components.

Lemma 1 *Assumed $\phi_{\tilde{\xi}}(\dot{y})$ the CDF of random vector $\tilde{\xi} \sim \mathcal{N}(\mu, R) \in \mathbb{R}^T$, where R is a $\mathbb{R}^T \times \mathbb{R}^T$ non-singular correlation matrix. Then, the partial derivative of $\phi_{\tilde{\xi}}(\dot{y})$ with respect to the variable $\dot{y}_i$ is given by:*

$$\frac{\partial\phi_{\tilde{\xi}}(\dot{y})}{\partial\dot{y}_i} = \phi_{\tilde{\eta}}(\dot{y}_1, \ldots, \dot{y}_{i-1}, \dot{y}_{i+1}, \ldots, \dot{y}_T) \cdot f_{\tilde{\xi}_i}(\dot{y}_i), \tag{3.15}$$

where $\tilde{\eta} \sim \mathcal{N}(\hat{\mu}, \hat{R})$ is a $(T-1)$-dimensional normally distributed random vector with the jth component $\hat{\mu}_j = \mu_j + \sigma_{ii}^{-1}(\dot{y}_i - \mu_i)\sigma_i \forall j \in \{1, \ldots, i-1, i+1, \ldots, T\}$ and $\hat{R}$ being giving by $R - \rho_{ii}^{-1}\rho_i\rho_i^T$ deleting all the ith row and column components. The function $f_{\tilde{\xi}_i}$ is the marginal probability density function of the component $\tilde{\xi}_i$ of the T-dimensional random vector $\tilde{\xi}$.

Derives directly from the Lemma 1 the gradients for the distribution function $\phi_{\tilde{\xi}}(\dot{y})$, being given by the following equation:

$$\nabla\phi_{\tilde{\xi}}(\dot{y}) = \left(\frac{\partial\phi_{\tilde{\xi}}(\dot{y})}{\partial\dot{y}_1}, \ldots, \frac{\partial\phi_{\tilde{\xi}}(\dot{y})}{\partial\dot{y}_T}\right), \tag{3.16}$$

hence, a T-dimensional function with components derived from the calculus of $(T-1)$-dimensional CDFs and a marginal density function, as stated in the preamble. Up to now, we highlight that we are dealing with probability distributions of type $F_{\tilde{\xi}}(\dot{y}) := \mathbb{P}[\tilde{\xi} \leq \dot{y}, e \in \mathcal{E}]$. In the present work, it is also necessary to introduce gradient calculations for rectangle (also known as two-sided or bilateral) probability functions, i.e. for functions of the following form:

$$G_{\tilde{\xi}}(\dot{a},\dot{b}) := \mathbb{P}\left[\dot{a} \leq \tilde{\xi} \leq \dot{b}, e \in \mathcal{E}\right], \tag{3.17}$$

where it is defined the affine transformations $\dot{a} = h_a(\dot{x})$ and $\dot{b} = h_b(\dot{x})$, being $h_a(\cdot)$ and $h_b(\cdot)$ affine functions defined in Section 3.4.1. For these cases, we follow again (van ACKOOIJ et al., 2010) in the following Lemma:

Lemma 2 *With $\tilde{\xi}$ as in the Lemma 1, for each component of $\dot{a}_i$ and $\dot{b}_i$, where $i = 1, \ldots, n$, it follows:*

$$\begin{aligned} \frac{\partial G_{\tilde{\xi}}(\dot{a},\dot{b})}{\partial \dot{b}_i} &= G_{\tilde{\beta}}(\dot{a}'(i), \dot{b}'(i)) \cdot f_{\xi_i}(\dot{b}_i) \\ \frac{\partial G_{\tilde{\xi}}(\dot{a},\dot{b})}{\partial \dot{a}_i} &= -G_{\tilde{\alpha}}(\dot{a}'(i), \dot{b}'(i)) \cdot f_{\xi_i}(\dot{a}_i), \end{aligned} \tag{3.18}$$

where $\tilde{\beta} \sim \mathcal{N}(\hat{\mu}, \hat{R})$ is a $(T-1)$-dimensional normally distributed random vector with the jth component $\hat{\mu}_j = \mu_j + \sigma_{ii}^{-1}(\dot{a}_i - \mu_i)\sigma_i \forall j \in \{1, \ldots, i-1, i+1, \ldots, T\}$ (for $\dot{a}_i$ case) and the equivalent for $\dot{b}_i$. The formula components $\hat{R}$, $f_{\xi_i}(\cdot)$ are considered as in Lemma 1, and $\dot{a}'(i) = (\dot{a}_1, \ldots, \dot{a}_{i-1}, \dot{a}_{i+1}, \ldots, \dot{a}_T)$ and $\dot{b}'(i)$ in an equivalent procedure.

By means of Lemma 2, we compute the gradients for the rectangular probability distribution, using the following notation:

$$\begin{aligned} \nabla_{\dot{b}} G_{\tilde{\xi}}(\dot{a},\dot{b}) &= \left(\frac{\partial G_{\tilde{\xi}}(\dot{a},\dot{b})}{\partial \dot{b}_1}, \ldots, \frac{\partial G_{\tilde{\xi}}(\dot{a},\dot{b})}{\partial \dot{b}_T} \right), \\ \nabla_{\dot{a}} G_{\tilde{\xi}}(\dot{a},\dot{b}) &= \left(\frac{\partial G_{\tilde{\xi}}(\dot{a},\dot{b})}{\partial \dot{a}_1}, \ldots, \frac{\partial G_{\tilde{\xi}}(\dot{a},\dot{b})}{\partial \dot{a}_T} \right), \end{aligned} \tag{3.19}$$

where these gradient components are used in Section 3.4.1 in order to design the gradient for the bilateral probability function.

3.4 APPROXIMATION CONSTRAINTS DEFINITIONS

The gradients calculation are essential to construct piecewise approximations of the convex set defined by the log-concave PCs in optimization problems. This section aims at summarizing the considerations to link the knowledge of the previous Section 3.3 to a stochastic optimization problem.

3.4.1 Affine approximations of probabilistic constraints

Let us consider the separable structure of an affine PC, using the short notation $\phi_{\tilde{\xi}}(\dot{x}) = G_{\tilde{\xi}}(h_a(\dot{x}), h_b(\dot{x}))$ in (3.17):

$$\phi_{\tilde{\xi}}(\dot{x}) := \mathbb{P}\left[h_a(\dot{x}) \leq \tilde{\xi} \leq h_b(\dot{x})\right] \geq p, \tag{3.20}$$

where $h_a : \mathbb{R}^m \to \mathbb{R}^T$ and $h_b(\cdot)$ in an equivalent way, such that $h_a(\dot{x}) = A\dot{x} + a$ and $h_b(\dot{x}) = B\dot{x} + b$, being the matrices $A, B \in \mathbb{R}^m \times \mathbb{R}^T$ and the vectors $a, b \in \mathbb{R}^T$. We remark:

Remark 3.4.1 *The gradient of the bilateral probability distribution* (3.20) *is given by:*

$$\nabla\phi_{\tilde{\xi}}(\dot{x}) = \nabla_{h_a(\dot{x})}G_{\tilde{\xi}}(h_a(\dot{x}), h_b(\dot{x}))^\top A + \nabla_{h_b(\dot{x})}G_{\tilde{\xi}}(h_a(\dot{x}), h_b(\dot{x}))^\top B, \tag{3.21}$$

where $\nabla_{h_a(\dot{x})}G_{\tilde{\xi}}(h_a(\dot{x}), h_b(\dot{x}))^\top$, *and* $\nabla_{h_b(\dot{x})}G_{\tilde{\xi}}(h_a(\dot{x}), h_b(\dot{x}))^\top$ *are defined in Section 3.3, equation* (3.19).

Proof. From the chain rule for total derivatives, we obtain the partial derivative for (3.20):

$$\frac{\partial\phi_{\tilde{\xi}}(\dot{x})}{\partial\dot{x}_n} = \sum_{i=1}^{T} \frac{\partial G_{\tilde{\xi}}(h_a(\dot{x}), h_b(\dot{x}))}{\partial h_a(\dot{x})_i}\frac{\partial h_a(\dot{x})_i}{\partial\dot{x}_n} + \frac{\partial G_{\tilde{\xi}}(h_a(\dot{x}), h_b(\dot{x}))}{\partial h_b(\dot{x})_i} \cdot \frac{\partial h_b(\dot{x})_i}{\partial\dot{x}_n}, \tag{3.22}$$

noting that the function $h_a(\dot{x})$ has a dependent A component and a free component a, $\frac{\partial h_a(\dot{x})_i}{\partial\dot{x}_n}$ is given by matrix component $A_{i,n}$ where n is a column index and line i. Using then notation A_n representing the column n of A (and equivalent to B_n), and definitions of the gradient $\nabla_{\cdot}G_{\tilde{\xi}}(\cdot,\cdot)$ in (3.19), we find that:

$$\frac{\partial\phi_{\tilde{\xi}}(\dot{x})}{\partial\dot{x}_n} = \nabla_{h_a(\dot{x})}G_{\tilde{\xi}}(h_a(\dot{x}), h_b(\dot{x}))^\top A_n + \nabla_{h_b(\dot{x})}G_{\tilde{\xi}}(h_a(\dot{x}), h_b(\dot{x}))^\top B_n. \tag{3.23}$$

Finally, the gradient of $\phi_{\tilde{\xi}}(\dot{x})$ is given by:

$$\nabla\phi_{\tilde{\xi}}(\dot{x}) = \left(\frac{\partial\phi_{\tilde{\xi}}(\dot{x})}{\partial\dot{x}_1}, \cdots, \frac{\partial\phi_{\tilde{\xi}}(\dot{x})}{\partial\dot{x}_N}\right), \tag{3.24}$$

where (3.21) has a matrix notation. □

Considering the above definitions, we evaluate the approximation of the set restricted by the constraint (3.20) at point x_i, using $i \in \mathcal{I}$ an index set, in such a way that:

$$\left\langle \nabla\phi_{\tilde{\xi}}(x_i), (\dot{x} - x_i) \right\rangle \geq 0, i \in \mathcal{I}, \tag{3.25}$$

where the last approximation follow the supporting hyperplane concepts presented in Appendix A.4. In the optimization problem, we intend to, in practice, that constraints (3.25) were added in the structured way:

$$G(x_i) \cdot \dot{x} \leq g(x_i), i \in \mathcal{I}, \tag{3.26}$$

where $G(\cdot)$ and $g(\cdot)$ are defined in the following Remark 3.4.2.

Remark 3.4.2 *Let the approximation of feasible set defined by the PC at point x_i being given by* (3.26)*. Then, for:*

$$\begin{aligned} G(x_i) &= -\nabla\phi_{\tilde{\xi}}(\dot{x})^\top \\ g(x_i) &= -\nabla\phi_{\tilde{\xi}}(\dot{x})^\top \cdot x_i, \end{aligned} \tag{3.27}$$

(3.25) *holds true.*

Proof. The proof comes directly from opening in the primitives the dot product in (3.25). □

Considering $h_a(\dot{x}) = -\infty$, the usual CDF structure with an upper boundary $\phi_{\tilde{\xi}}(\dot{x}) = \mathbb{P}[\tilde{\xi} \leq h_b(\dot{x})]$ is established. In such case, by (3.16) we define the approximation of the feasible set of following joint PC:

$$\phi_{\tilde{\xi}}(\dot{x}) := \mathbb{P}[\tilde{\xi} \leq h_b(\dot{x})] \geq p, \tag{3.28}$$

by the set of supporting hyperplanes $\left\langle \nabla\phi_{\tilde{\xi}}(\dot{x}), h_b(\dot{x} - x_i) \right\rangle \geq 0, i \in \mathcal{I}$. This notation is useful since we can project the linear approximations directly from the probability

distributions, as in Figure 4 in Section 3.1. Opening the last dot product, one defines the constraints $G(x_i) \cdot \dot{x} \leq g(x_i), i \in \mathcal{I}$, as in constraints (3.26). For further information in how to proceed to construct the supporting hyperplanes based on the gradient $\nabla\phi_{\tilde{\xi}}(\dot{x})$, proceed to Appendix A.4.

3.4.2 Affine structures coefficients

In cutting plane approximations of PCs, it is essential to have a well-established procedure to define the coefficient components of the affine equations. As stated in Section 3.3, to apply Lemmas 1 and 2 it is required that $\tilde{\xi} \sim \mathcal{N}(\mu, R)$. Moreover, if $\tilde{\xi}$ is standardized in $\mu = 0$ (i.e. a zero-mean column vector), the application of these Lemmas demands fewer calculations. Let us consider the original random vector $\tilde{\eta} \sim \mathcal{N}(\mu_{or}, \Sigma_{or})$, where $\mu_{or} \in \mathbb{R}^T$ is the original mean of the underlying random vector and $\Sigma_{or} \in \mathbb{R}^T \times \mathbb{R}^T$ is a positive definite covariance matrix. The idea is, starting from the original PC, make clear the procedures to define a pattern linking to the cited Lemmas efficiently. The original joint PC is given by:

$$G_{\tilde{\eta}}(\dot{x}) := \mathbb{P}\left[ld^{low} \leq h(\dot{x}) + \tilde{\eta} \leq ld^{up}\right] \geq p, \tag{3.29}$$

where $ld^{low} \in \mathbb{R}^T$ and $ld^{up} \in \mathbb{R}^T$ are thresholds vectors as, for instance, the minimum and maximum accepted load demand deviations. If the PCs are defined as load requirement constraints of ED problems, for instance, it makes sense to work with $\tilde{\eta}$ as being wind farms random power generation. The affine decision variable function $h(\dot{x})$ is defined as follows:

$$h(\dot{x}) = H_{or} \cdot \dot{x}, \tag{3.30}$$

where the decision vector $\dot{x}$ is the dispatch of sources for a given load requirement constraint, without the wind generation. H_{or} is a structure matrix where each line defines the correspondent coefficient of the decision variables of $\dot{x}$ for a given period $t \in T$, with the following form (for example):

$$H_{or} = \begin{bmatrix} 1\dots1 & 0\dots0 & \dots & 0\dots0 \\ 0\dots0 & 1\dots1 & \dots & 0\dots0 \\ & \vdots & & \\ 0\dots0 & 0\dots0 & \dots & 1\dots1 \end{bmatrix} \tag{3.31}$$

$$\dot{x} = [\dot{g}_{1,1}, \dots, \dot{g}_{N,1}, \dots, \dot{g}_{1,T}, \dots, \dot{g}_{N,T}]^{\top}.$$

In this example, a group of ones $1\dots1$ has N components, $\dot{g}_{n,t}$ represents the generation of the unit n in period t and $H_{or} \in \mathbb{R}^{T\times N\cdot T}$, being N the total number of generation units and T the number of periods. Defining the covariance diagonal coefficient $\sigma_{t,t}$ of the covariance matrix Σ_{or}, we set the following auxiliary matrix D:

$$D = \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\sqrt{\sigma_{TT}}} \end{pmatrix}, \tag{3.32}$$

and it is well known the correlation matrix R of $\tilde{\eta}$ is given by:

$$R = D \cdot \Sigma_{or} \cdot D'. \tag{3.33}$$

Using these instruments, we propose the following remark:

Remark 3.4.3 *Standardizing the original random vector $\tilde{\eta} \sim \mathcal{N}(\mu_{or}, \Sigma_{or})$, we establish the equivalent version of the PC* (3.29)*, considering $h(\dot{x})$ as in* (3.30)*:*

$$\phi_{\tilde{\xi}}(\dot{x}) := \mathbb{P}\left[A\dot{x} + a \leq \tilde{\xi} \leq B\dot{x} + b\right] \geq p,$$

where $\tilde{\xi} \sim \mathcal{N}(0, R)$ and the coefficients are defined as:

$$\begin{aligned} A &= -D \cdot H_{or} \\ a &= D \cdot (ld^{low} - \mu_{or}) \\ B &= -D \cdot H_{or} \\ b &= D \cdot (ld^{up} - \mu_{or}). \end{aligned} \tag{3.34}$$

Proof.

$$\mathbb{P}[ld^{low} \leq H_{or} \cdot \dot{x} + \tilde{\eta} \leq ld^{up}] \Leftrightarrow \mathbb{P}[ld^{low} - H_{or} \cdot \dot{x} \leq \tilde{\eta} \leq ld^{up} - H_{or} \cdot \dot{x}]$$
$$\Leftrightarrow \mathbb{P}[D \cdot (ld^{low} - \mu_{or}) - D \cdot H_{or} \cdot \dot{x} \leq \tilde{\xi} \leq D \cdot (ld^{up} - \mu_{or}) - D \cdot H_{or} \cdot \dot{x}]$$
$$\Leftrightarrow \mathbb{P}[A\dot{x} + a \leq \tilde{\xi} \leq B\dot{x} + b],$$

where it is used the fact that $\tilde{\xi} = D \cdot (\tilde{\eta} - \mu_{or})$. □

Note that with the last statement we established the equivalences $h_a(\dot{x}) = A\dot{x} + a$ and $h_b(\dot{x}) = B\dot{x} + b$ of (3.20). The presented structure works if we want to limit the PC between the upper and lower thresholds. A similar procedure is performed when the PC respects a minimal threshold:

$$\phi_{\tilde{\eta}}(\dot{x}) := \mathbb{P}[ld^{low} \leq h(\dot{x}) + \tilde{\eta}] \geq p. \tag{3.35}$$

By the following remark we manage with this PC:

Remark 3.4.4 *Standardizing the original random vector $\tilde{\eta} \sim \mathcal{N}(\mu_{or}, \Sigma_{or})$, we establish the equivalent version of the PC* (3.35)*, considering $h(\dot{x})$ as in* (3.30)*:*

$$\phi_{\tilde{\xi}}(\dot{x}) := \mathbb{P}[\tilde{\xi} \leq B\dot{x} + b] \geq p,$$

where $\tilde{\xi} \sim \mathcal{N}(0, R)$ and the coefficients are defined as:

$$\begin{aligned} B &= D \cdot H_{or} \\ b &= D \cdot (\mu_{or} - ld^{low}). \end{aligned} \tag{3.36}$$

Proof.

$$\mathbb{P}[ld^{low} \leq h(\dot{x}) + \tilde{\eta}] \Leftrightarrow \mathbb{P}[-\tilde{\eta} \leq -ld^{low} + H_{or} \cdot \dot{x}]$$
$$\Leftrightarrow \mathbb{P}[\tilde{\xi} \leq D \cdot (\mu_{or} - ld^{low}) + D \cdot H_{or} \cdot \dot{x}] \quad ,$$
$$\Leftrightarrow \mathbb{P}[\tilde{\xi} \leq B\dot{x} + b]$$

where it is used the symmetry property of the standardized random vector $\tilde{\xi} \equiv -\tilde{\xi} = D \cdot (\mu_{or} - \tilde{\eta})$. □

The standardization provided by Remarks 3.4.3 and 3.4.4 generalizes the use of Remark 3.4.2 for affine separable structures of PC. This allows the iterative use of constraints

(3.26) in the PC optimization problem as piecewise linear approximations of the original PC (3.29).

4 THE EXACT SUPPORTING HYPERPLANE METHOD FOR SOLVING STOCHASTIC HTUC MILP PROBLEMS

This chapter discusses the first topic of Thesis, given by a stochastic HTUC system and depicts alternatives employing PC programming problems to solve it. The main contribution is the proposition of an algorithm that deals with the mixed-integer setting, used to describe a complete UC system. To handle the random variables, the model employs the original joint probability distribution functions that describe these variables, maintaining their correlation effects.

Section 4.1 defines the underlying MILP problem, replacing the load requirement constraint by the PC in Subsection 4.2.2. Section 4.3 describes the algorithms that are used to solve the proposed HTUC MILP problem. The comparative computational results among the methods are discussed in Section 4.4, with other sensibility simulations. Part of the formulation and experiments of these last sections are presented in the paper (van ACKOOIJ; FINARDI; RAMALHO, 2018). Additionally to this last work we provide in Section 4.5 suggestions of improvements to the original algorithm. The chapter ends in Section 4.6, where the used hydro production function model is compared to a standard model from the literature.

4.1 PROBLEM DEFINITION

The section starts with the employed set of indexes, parameters and nomenclatures employed to describe the system. In this formulation, we use the following notation: Continuous decision variables has the dot notation as in $\dot{x}$, the bar notation $\bar{x}$ denotes integers, random variables are $\tilde{x}$ and parameters has no sign x. Those indexes are followed by the underlying MILP problem description.

4.1.1 Indices

t	Index of time periods, where $t \in \mathcal{T}$ with $\lvert\mathcal{T}\rvert = T$.
b	Index of buses in the transmission system, where $b \in \mathcal{B}$ and $\mathcal{B}_{\mathcal{W}} \subset \mathcal{B}$ the subset of buses supplied by wind farms.
b'	Index of buses in the oriented transmission flow, where $b' \in \mathcal{B}_{+}^{(b)} \cup \mathcal{B}_{-}^{(b)}$ representing the set of buses that supplies b and drains from b, respectively.
s	Index associated with piecewise linear future cost function, where $s \in \mathcal{S}$.
i	Index of thermal plants, where $i \in \mathcal{I}$, and $\mathcal{I}^{(b)} \subset \mathcal{I}$ the subset of plants that supplies bus b.
l	Index of hydro plants, where $l \in \mathcal{L}$, and $\mathcal{L}^{(b)} \subset \mathcal{L}$ the subset restricted to b.
w	Index of wind farms, where $w \in \mathcal{W}$, and $\mathcal{W}^{(b)} \subset \mathcal{W}$ the subset restricted to b.
j	Index of units of the hydro plants, where $j \in \mathcal{J}_l$ and l is the referred hydro plant.
mo	Index of upstream hydro plants, where $mo \in \mathcal{M}_l$ and l is the downstream hydro plant.
n	Index of piecewise linear hydro production function, where $n \in \mathcal{N}_l$ and l is the referred hydro plant.
r	Index of piecewise linear penstock head function, where $r \in \mathcal{R}_l$ and l is the referred hydro plant.
k	Index of the proposed supporting hyperplane method iteration.
q	Index of samples used in the sample based method, where $q \in \mathcal{Q}$ with $\lvert\mathcal{Q}\rvert = Q$.

4.1.2 Decision Variables

$\overline{u}_{i,t}$	Commitment status of thermal unit i and period t, $\overline{u}_{i,t} =$ 1 if unit is online and 0 otherwise.
$\overline{z}_{j,l,t}$	Commitment status of hydro unit j, plant l and period t, $\overline{z}_{j,l,t} = 1$ if unit is online and 0 otherwise.
$\dot{gt}_{i,t}^{(b)}$	Power of thermal unit i and period t, located at bus b $[MW]$.
$\dot{gh}_{l,t}^{(b)}$	Hydro power output of plant l, period t, located at bus b.
$\dot{grh}_{l,t}^{(b)}$	Gross head of plant l, period t, located at bus b $[m]$.
$\dot{vs}_{j,l,t}$	Gross head auxiliary variable for hydro unit j, plant l and period t $[m]$.
$\dot{f}_t^{(b,b')}$	Power flow from bus b to bus b' in period t $[MW]$.
$\dot{df}_t^{(b)}$	deficit load at bus b and period t $[MW]$.
$\dot{v}_{l,t}$	Reservoir volume of plant l in the end of period t $[hm^3]$.
$\dot{q}_{j,l,t}$	Turbined outflow of unit j, plant l and period t $[m^3/s]$.
$\dot{s}_{l,t}$	Spillage of plant l and period t $[m^3/s]$.
$\dot{ql}_{l,t}$	Penstock head loss of hydro plant l in period t $[m]$.
$\dot{qct}_{l,t}$	Auxiliary turbined outflow control variable of hydro plant l in period t $[m^3/s]$.
$\dot{\theta}_t^{(b)}$	Voltage angle of bus b and period t in the DC power-flow model $[rad]$.
$\dot{K}_{i,t}$	Start-up cost of thermal unit i and period t [\$].
$\dot{\Theta}$	Expected future operational cost [\$].

4.1.3 Random Variables

$\widetilde{gw}_{w,t}^{(b)}$ Power of wind farm w and period t located at bus b, with variance-covariance matrix $\Sigma^{(b)}$ and mean $\mu^{(b)}$ $[MW]$.

4.1.4 Input Parameters

$y_{l,t}$	Incremental inflow of plant l and period t $[m^3/s]$.
V_l^{min}	Minimum (maximum) reservoir level of plant l $[hm^3]$.
$q_{j,l}^{min}$	Minimum (maximum) turbined outflow of unit j of plant l $[m^3/s]$.
s_l^{min}	Minimum (maximum) spillage of hydro plant l $[m^3/s]$.
gh_l^{min}	Minimum (maximum) power of hydro plant l $[MW]$.
grh_l^{min}	Minimum (maximum) gross head of plant l $[m]$.
gt_i^{min}	Minimum (maximum) power of thermal plant i $[MW]$.
X_k	Reactance of line k $[\Omega]$.
L_k	Power capacity of line k and period t $[MW]$.
SB	Conversion constant $[\frac{\Omega \cdot MW}{rad}]$.
$ld_t^{(b)}$	Load demand at bus b and period t $[MW]$.
$ld_b^{up(low)}$	T-dimensional vector with upper (lower) limits of the load at bus b $[MW]$.
CP_i	Start-up cost of unit i [\$].
c	Constant to convert m^3/s to hm^3 in one hour period $c = 60 * 60/10^6$.
$cd^{(b)}$	Cost of deficit parameter of bus b $[\$/MW]$.
c_{0i}, c_{1i}	Thermal cost parameters of unit i [\$], $[\$/MW]$.
p_{0l}, p_{1l}	Upstream water level function parameters of hydro plant l $[\frac{m}{hm^3}]$, $[m]$.
d_{0l}, d_{1l}	Downstream water level function parameters of hydro plant l $[\frac{m}{m^3/s}]$, $[m]$.
$h_{0l}^{(r)}$, $h_{1l}^{(r)}$	Penstock head loss function parameters of hydro plant l, at index r $[\frac{m}{m^3/s}]$, $[m]$.
$n_{0l}^{(n)}$, $n_{1l}^{(n)}$,$n_{2l}^{(n)}$	Power production function parameters of hydro plant l, at index k $[\frac{MW}{m^3/s}]$, $[\frac{MW}{m}]$, $[MW]$.

τ_{ml}	Water travelling time from hydro plant m to l.
$\tau^{+}_{j,l}$	Minimum up-time of unit j and hydro plant l $[h]$.
τ^{+}_{i} / τ^{-}_{i}	Minimum up/down-time of thermal unit i $[h]$.
Δ^{+}_{i} / Δ^{-}_{i}	Maximum ramp-up/ramp-down rates $[MW]$.
GT^{Δ}_{i}	Minimum production level in ramp functions of thermal plant i $[MW]$.
μ^{s}_{l}, z^{s}	Future cost function parameters of hydro plant l at index s $[\frac{\$}{hm^3}]$, $[\$]$.
In_0	Vector of initial condition parameters of the system.

4.1.5 Probabilistic Constraints Formulation

$\dot{x}$	Vector $\in X$ of size $m = m_1 + m_2$ of all decision variables, m_1 are continuous and m_2 binaries.
$\tilde{\xi}^{(b)}$	Standardized normal random variable with mean 0 and $R^{(b)} \in \mathbb{R}^T \times \mathbb{R}^T$ non-singular correlation matrix.
$a^{(b)}, b^{(b)}$	Auxiliary vectors $\in \mathbb{R}^T$.
$A^{(b)}, B^{(b)}$	Auxiliary matrices $\in \mathbb{R}^T \times \mathbb{R}^m$.
$\phi^{(b)}(\dot{x})$	Probabilistic constraint cumulative distribution function of bus b.
$\nabla\phi^{(b)}(x_i)$	Gradient vector at point x_i of the probabilistic constraint of bus b.

4.2 PROBLEM FORMULATION

This section presents the problem formulation addressed in the present work. In Section 4.3, the underlying nominal MILP given below is extended to include the specific PC approaches.

$$\min \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}}(c_{0i}\overline{u}_{i,t}+c_{1i}\dot{gt}_{i,t}+\dot{K}_{i,t})+\sum_{t\in\mathcal{T}}\sum_{b\in\mathcal{B}}cd^{(b)}\dot{df}_t^{(b)}+\dot{\Theta} \tag{4.1}$$

s.t.:

$$\sum_{i\in\mathcal{I}^{(b)}}\dot{gt}_{i,t}+\sum_{w\in\mathcal{W}^{(b)}}\widetilde{gw}_{w,t}+\sum_{l\in\mathcal{L}^{(b)}}\dot{gh}_{l,t}+$$
$$+\dot{df}_t^{(b)}+\sum_{b'\in\mathcal{B}_+^{(b)}\cup\mathcal{B}_-^{(b)}}\dot{f}_t^{(b,b')}=ld_t^{(b)},\forall b,t \tag{4.2a}$$

$$\dot{f}_t^{(b,b')}=\frac{(\dot{\theta}_t^{(b)}-\dot{\theta}_t^{(b')})\cdot SB}{X_{b,b'}},\forall b,t;b'\in\mathcal{B}_+^{(b)}\cup\mathcal{B}_-^{(b)} \tag{4.2b}$$

$$\frac{-\pi}{2}\leq\dot{\theta}_t^{(b)}\leq\frac{\pi}{2},\forall b,t \tag{4.2c}$$

$$-L_{b,b'}\leq\dot{f}_t^{(b,b')}\leq L_{b,b'},\forall b,t;b'\in\mathcal{B}_+^{(b)}\cup\mathcal{B}_-^{(b)} \tag{4.2d}$$

$$\dot{gt}_{i,t}\leq\dot{gt}_{i,t-1}+\overline{u}_{i,t-1}\Delta_i^+ +(1-\overline{u}_{i,t-1})GT_i^{\Delta},\forall t,i \tag{4.3a}$$

$$\dot{gt}_{i,t-1}\leq\dot{gt}_{i,t}+\overline{u}_{i,t}\Delta_i^- +(1-\overline{u}_{i,t})GT_i^{\Delta},\forall i,t \tag{4.3b}$$

$$\overline{u}_{i,t}\geq\overline{u}_{i,r}-\overline{u}_{i,r-1},\forall t,r\in[t-\tau_i^+ +1,t-1],\forall i \tag{4.3c}$$

$$\overline{u}_{i,t}\leq 1+(\overline{u}_{i,r}-\overline{u}_{i,r-1}),\forall t,r\in[t-\tau_i^- +1,t-1],\forall i \tag{4.3d}$$

$$\dot{K}_{i,t}\geq CP_i(\overline{u}_{i,t}-\overline{u}_{i,t-1}),\forall t,i \tag{4.3e}$$

$$gt_i^{min}\overline{u}_{i,t}\leq\dot{gt}_{i,t}\leq gt_i^{max}\overline{u}_{i,t},\forall i,t \tag{4.3f}$$

$$\dot{v}_{l,t+1}=\dot{v}_{l,t}+c[\dot{y}_{l,t}+\sum_{mo\in\mathcal{M}_l}(\sum_{j\in\mathcal{J}_mo}\dot{q}_{j,mo,t-\tau_{ml}}+$$
$$-\dot{s}_{mo,t-\tau_{mol}})-\sum_{j\in\mathcal{J}_l}\dot{q}_{j,l,t}-\dot{s}_{l,t}],\forall l,t \tag{4.4a}$$

(4.16a)-(4.16g) (HPF formulation) (4.4b)

$$\overline{z}_{j,l,t}\geq\overline{z}_{j,l,r}-\overline{z}_{j,l,r-1},\forall t,r\in[t-\tau_{j,l}^+ +1,t-1],\forall j,l \tag{4.4c}$$

$$\dot{\Theta}-\sum_{l\in\mathcal{L}}\mu_l^{(s)}\dot{v}_{l,T}\geq z^{(s)},\forall s \tag{4.4d}$$

$$V_l^{min}\leq\dot{v}_{l,t}\leq V_l^{max},\forall l,t \tag{4.4e}$$

$$s_l^{min}\leq\dot{s}_{l,t}\leq s_l^{max},\forall l,t \tag{4.4f}$$

$$gh_l^{min}\leq\dot{gh}_{l,t}\leq gh_l^{max},\forall l,t \tag{4.4g}$$

$$\dot{df}_t^{(b)},\dot{K}_{i,t},\dot{\Theta}\geq 0,\forall l,t,b \tag{4.5}$$

$$[\dot{v}_{l,0},\dot{q}_{j,l,0},\overline{z}_{j,l,0},\dot{s}_{l,0},\dot{gt}_{i,0},\overline{u}_{i,0}]\cdot\mathbf{1}=\mathbf{In_0},\forall l,j,i. \tag{4.6}$$

In (4.1)-(4.6), the objective function (4.1) consists of minimizing the operation cost for T periods of a day-ahead horizon, which is composed of fixed, variable and start-up thermal costs, load shedding cost and the expected future cost related to the usage of the water, this last representing boundary conditions that establish a cost to the outflow in the planning horizon. Constraints (4.2a) refer to the load requirement for each bus of the transmission system, represented as a classic DC power flow model. Given that the wind generation is present in (4.2a), the PC model replaces some of these constraints for handling the uncertainty, according to Subsection 4.2.2. The relation between the buses voltage angle and power flow is accomplished by (4.2b)-(4.2d). The variable $\dot{f}_t^{(b,b')}$ is positive in the direction of power flow from bus b to the bus b' and negative otherwise. Thermal units ramp-up and ramp-down rates are explicit in (4.3a)-(4.3b) while (4.3c)-(4.3d) express the minimum up-time and down-time. Thermal start-up costs are accounted in (4.3e), followed by limits to their production in (4.3f). Constraints (4.4a) detail the time and space coupling associated with the reservoirs operational characteristics. The set of constraints (4.4b) refers to the hydro production function, as discussed in 4.6. Constraints (4.4c) represents the minimum up-time of hydro units. A piecewise linear model for the expected future cost associated with the use of the water over the horizon is presented in (4.4d). Constraints (4.4e) - (4.5) define bounds on the variables, and finally, (4.6) represents the set of initial conditions, where $\mathbf{1}$ is the all-one vector of appropriate size and $\mathbf{In_0}$ is the vector of initial values themselves.

4.2.1 Hydro production function model

This section aims to detail the HPF model used by this work. The HPF is a nonlinear, nonconvex and discontinuous function that depends on forebay and tailrace levels, penstock head losses, turbine and generator efficiencies (GULLIVER; ARNDT, 1991). The main idea of the proposed HPF formulation is instead of using a piecewise linear approximations to the entire nonlinear productivity curve of each turbine, as addressed in, for instance (TONG; ZHAI; GUAN, 2013; LI et al., 2014; FINARDI; TAKIGAWA; BRITO,

2016), is to approximate the minor functions that compound the HPF and considers the power production of the entire plant rather than of each turbine. In practice, we represent a group of identical units by a single piecewise linear HPF.

By the bibliography survey, we clarify the differences between the proposed HPF model and the literature. In (CONEJO et al., 2002) it is proposed an aggregated model for each plant where the three dimensional HPF is simplified making use of two-dimensional curves that establish the relation between the power production and the outflow, fixing the reservoir volume to different levels. The authors use piecewise linear functions to represent the original nonlinear hill charts, where the linear partitions are select by the aid of binary variables. An advantage of the method is that it is not binding on assumptions of limiting the solutions on partitions of the hill chart where convex simplifications could be considered, as used by the present work. In this perspective, a drawback of the method in (CONEJO et al., 2002) is the increasing number of binary variables necessary to control the active block of the HPF slopes. Besides that, in our strategy where we aggregate by groups of identical HPF, it is considered the dependence on the single unit outflow and the total discharge (total outflow of the plant plus spillage) by the net head effect. As stated, in (CONEJO et al., 2002), the head effect is simplified by the different discretized curves, changing with the reservoir volume, then distinguishing from our approach.

In (LI et al., 2014) the focus is on a MILP model that deals with a hydro unit commitment for the Three Gorges Project (TGP) in China. The formulation considers the head variation over the operation and its effects on power generation. The units are considered individually heterogeneous and are considered as nonlinear power generation functions. To model the head effect for the unit power production, they use a linear function to approximate the forebay water level, in a similar procedure as proposed by the present work. However, the authors claim that TGP tailrace elevation cannot fallow the same linear approximation procedure due to local particularities, hence being necessary an ad-hoc procedure to define the tailrace elevation using sequential MILP executions up to the convergence. At this point, their model differs from the here proposed work, where additional MILP executions are not necessary since our model makes use of a linear

approximation (that could be easily extended to a piecewise linear formulation) to define the tailrace water level.

Rather than providing an analytic function to approximate the penstock head loss function, as the piecewise linear function of the here proposed model, in (LI et al., 2014) it is set constant, and so any dependence on the unit outflow was ruled out. The main difference between our model and TGP is a triangular tessellation technique that defines the values of the variables by a combination of predefined weighted points (vertex of the triangles). This model makes necessary the addition of three binary variables per hydro unit. The addition of binaries is a concern in our model since we try to reduce the number of binaries to aid the CPU execution time of the MILP problems.

Among the works that use aggregated HPF functions, in (PAREDES; MARTINS; SOARES, 2015) the authors assume the use of identical units efficiency and a quadratic function linking the power to the water discharge, maintaining the penstock head losses effects implicit to the function. In our approach, these losses are analytically considered, since piecewise linear functions can precisely model the convexity of these functions (quadratic in the unit outflow). In (BORGHETTI et al., 2008) it is used piecewise linear partitions, controlled by binary variables, in order to establish the relationship between power production and outflow. Each of these curves is related to a reservoir volume, that is used to estimate the head effected of the unit hill chart. This model differs from the here proposed since in our model, fewer binaries are necessary and the head effect is analytically considered in the MILP formulation.

In (GUEDES et al., 2017) it is proposed a MILP model that makes use of an interpolation procedure of a discretized values of the power efficiency function. The heuristic aims to enforce the turbines to work at the best operational points, known as “design points”. The model defers from the here proposed since the power production functions are not directly embedded into the MILP formulation. In (SANTO; COSTA, 2016) it is proposed a solution of a hydroelectric based UC system solved by an MINLP model that minimizes sources of production losses, namely, turbine operating points distant from ’design points’, tailrace elevation and penstock head. The nonlinearities of the model come

from a three-dimensional polynomial curve that establishes the relationship between the losses with unit production levels and the number of committed units. The proposed model differs from the here stated since forebay elevation is considered fixed in the optimization horizon, and spillage is disregarded in the head effect. These differences can also be justified when one observes that the main objective of this last work concerns at finding good operational points for the turbines, that differs from costs reduction in an integrated electric system, as the objective of the present chapter.

Simplifying the mathematical notation, the models in the present section consider plants in which all units are identical. Nevertheless, the notation can be easily extended for cases with different groups of identical units. Initially, we consider that the following piecewise model is available for each generating unit:

$$\dot{gu}h_{j,l,t} - n_{0l}^{(n)}\dot{qu}_{j,l,t} - n_{1l}^{(n)}\dot{h}_{j,l,t} - n_{2l}^{(n)} \leq 0, \forall n, j, l, t. \tag{4.7}$$

In (4.7) $\dot{gu}h_{j,l,t}$, $\dot{qu}_{j,l,t}$ and $\dot{h}_{j,l,t}$ are, respectively, the power (MW), the turbined outflow (m^3/s) and the net head (m) of unit j, plant l and period t. On the other hand,$n_{0l}^{(n)}$,$n_{1l}^{(n)}$, $n_{2l}^{(n)}$ are constants, and n is the index that represents the number of functions used in the piecewise linear model. In our work, (4.7) is obtained applying convex hull techniques over the operative zones of the unit. For instance, a standard nonlinear formulation is provided in (GULLIVER; ARNDT, 1991) where the unit HPF is defined as:

$$\dot{gu}h_{j,l,t} = G \cdot \dot{\eta}_{j,l,t} \cdot \dot{h}_{j,l,t} \cdot \dot{qu}_{j,l,t} \quad \forall n, j, l, t, \tag{4.8}$$

where G is a constant and $\dot{\eta}_{j,l,t}$ is the turbine efficiency of the hydro unit l. In this formulation, we consider only the loss of efficiency provided by the turbine, being disregarded the small losses coming from the generator group. The following curve gives the efficiency of the turbine:

$$\dot{\eta}_{j,l,t} = e_{0l} + e_{1l} \cdot \dot{qu}_{j,l,t} + e_{2l} \cdot \dot{h}_{j,l,t} + e_{3l} \cdot \dot{h}_{j,l,t} \cdot \dot{qu}_{j,l,t} +$$
$$+ e_{4l} \cdot \dot{qu}^2_{j,l,t} + e_{5l} \cdot \dot{h}^2_{j,l,t} \quad \forall n, j, l, t, \tag{4.9}$$

where $e_{0l}, \ldots, e_{5l}$ are constants. Combining (4.8) and (4.9) we have the nonlinear HPF, that is approximated locally by the set of piecewise linear functions (4.7). We rewrite (4.7) as follows:

$$\dot{guh}_{j,l,t} - n^{(n)}_{0l} \dot{qu}_{j,l,t} - n^{(n)}_{1l} (\dot{grh}_{l,t} - \dot{qul}_{j,l,t}) - n^{(n)}_{2l} \leq 0, \forall n, j, l, t, \tag{4.10}$$

where $\dot{grh}_{l,t}$ and $\dot{qul}_{j,l,t}$ are, respectively, the gross head (m) and the unit j penstock head loss (m) of plant l and period t. When it is considered identical units assumption, we easily check via classic optimality conditions that $\dot{qu}_{j,l,t}$ of online units must also be identical. As a result, we can join the inequalities (4.10) in a single group, resulting in the following piecewise linear model with $|\mathcal{N}_l|$ constraints:

$$\dot{gh}_{l,t} - n^{(n)}_{0l} \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} \cdot \dot{q}_{j,l,t} - n^{(n)}_{1l} \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} (\dot{grh}_{l,t} - \dot{qul}_{j,l,t}) +$$
$$- n^{(n)}_{2l} \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} \leq 0, \forall n, l, t, \tag{4.11}$$

where $\dot{gh}_{l,t}$ is the power of plant l and period t (MW) and $\overline{z}_{j,l,t}$ defines the on/off status of the unit j, plant l and period t. In (4.11) , when compared to (4.10), we use the notation $\dot{qu}_{j,l,t} = \overline{z}_{j,l,t} \cdot \dot{q}_{j,l,t}$ for the commitment of the unit. The nonlinearity due to the multiplication between two decision variables $\overline{z}_{j,l,t} \cdot \dot{grh}_{l,t}$, is addressed replacing them by the auxiliary variable $\dot{vs}_{j,l,t} = \overline{z}_{j,l,t} \cdot \dot{grh}_{l,t}$ and including the following constraints:

$$grh^{min}_{l,t} \overline{z}_{j,l,t} \leq \dot{vs}_{j,l,t} \leq grh^{max}_{l,t} \overline{z}_{j,l,t}, \forall j, l, t$$
$$grh^{min}_{l,t} (1 - \overline{z}_{j,l,t}) \leq \dot{grh}_{l,t} - \dot{vs}_{j,l,t} \leq grh^{max}_{l,t} (1 - \overline{z}_{j,l,t}), \forall j, l, t$$

where $grh_{l,t}^{min(max)}$ are the gross head limits. Hence, (4.11) can be rewritten as:

$$\dot{gh}_{l,t} - n_{0l}^{(n)} \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} \cdot \dot{q}_{j,l,t} - n_{1l}^{(n)} \sum_{j \in \mathcal{J}_l} \dot{vs}_{j,l,t} +$$
$$+ n_{1l}^{(n)} \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} \cdot \dot{qul}_{j,l,t} - n_{2l}^{(n)} \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} \leq 0, \forall n, l, t \tag{4.12a}$$

$$grh_{l,t}^{min} \overline{z}_{j,l,t} \leq \dot{vs}_{j,l,t} \leq grh_{l,t}^{max} \overline{z}_{j,l,t}, \forall j, l, t \tag{4.12b}$$

$$grh_{l,t}^{min}(1 - \overline{z}_{j,l,t}) \leq \dot{grh}_{l,t} - \dot{vs}_{j,l,t} \leq grh_{l,t}^{max}(1 - \overline{z}_{j,l,t}), \forall j, l, t. \tag{4.12c}$$

To address the nonlinearity $\overline{z}_{j,l,t} \cdot \dot{q}_{j,l,t}$, it is essential to remember that we are considering identical units and, therefore, the identical turbined outflow is enforced by the use of the auxiliary variables $qct_{l,t}$, and the inclusion of the following constraints to the model (4.12a)-(4.12c):

$$q_{j,l}^{min} \cdot \overline{z}_{j,l,t} \leq \dot{q}_{j,l,t} \leq q_{j,l}^{max} \cdot \overline{z}_{j,l,t} \forall j, l, t$$
$$q_{j,l}^{min}(\overline{z}_{j,l,t} - 1) \leq \dot{qct}_{l,t} - \dot{q}_{j,l,t} \leq q_{j,l}^{max}(1 - \overline{z}_{j,l,t}), \forall j, l, t,$$

where $q_{j,l}^{min(max)}$ are limits of the turbined outflow of units (m^3/s). We note that the last constraints also deal with the nonlinearity $\overline{z}_{j,l,t} \cdot \dot{q}_{j,l,t}$ and, as a result, we have:

$$\dot{gh}_{l,t} - n_{0l}^{(n)} \sum_{j \in \mathcal{J}_l} \dot{q}_{j,l,t} - n_{1l}^{(n)} \sum_{j \in \mathcal{J}_l} \dot{vs}_{j,l,t} + n_{1l}^{(n)} \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} \cdot \dot{qul}_{l,t} +$$
$$- n_{2l}^{(n)} \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} \leq 0, \forall n, l, t \tag{4.13a}$$

$$grh_{l,t}^{min} \overline{z}_{j,l,t} \leq \dot{vs}_{j,l,t} \leq grh_{l,t}^{max} \overline{z}_{j,l,t}, \forall j, l, t \tag{4.13b}$$

$$grh_{l,t}^{min}(1 - \overline{z}_{j,l,t}) \leq \dot{grh}_{l,t} - \dot{vs}_{j,l,t} \leq grh_{l,t}^{max}(1 - \overline{z}_{j,l,t}), \forall j, l, t \tag{4.13c}$$

$$q_{j,l}^{min} \cdot \overline{z}_{j,l,t} \leq \dot{q}_{j,l,t} \leq q_{j,l}^{max} \cdot \overline{z}_{j,l,t} \forall j, l, t \tag{4.13d}$$

$$q_{j,l}^{min}(\overline{z}_{j,l,t} - 1) \leq \dot{qct}_{l,t} - \dot{q}_{j,l,t} \leq q_{j,l}^{max}(1 - \overline{z}_{j,l,t}), \forall j, l, t. \tag{4.13e}$$

Next, we handle with $\overline{z}_{j,l,t} \cdot \dot{qul}_{l,t}$. Since the penstock head loss is typically addressed as a (convex) quadratic function (GULLIVER; ARNDT, 1991), we can approximate it by the

following piecewise linear function:

$$\dot{qul}_{l,t} - h_{0l}^{(r)} \cdot \dot{q}_{j,l,t} - h_{1l}^{(r)} \geq 0, \forall r, j, l, t,$$

where $h_{0l}^{(r)}$ and $h_{1l}^{(r)}$are constants, and r is the index that represents the number of functions used in the piecewise linear model composed of $|\mathcal{R}_l|$ constraints. Considering equal losses to online units, the same procedure as in (4.10) is used, which results in following penstock head loss of the plan l:

$$\dot{ql}_{l,t} - \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} \left(h_{0l}^{(r)} \cdot \dot{q}_{j,l,t} + h_{1l}^{(r)} \right) \geq 0, \forall r, l, t, \tag{4.14}$$

where $\dot{ql}_{l,t}$ represents the penstock head loss of the entire plant l. Given that the nonlinearity $\overline{z}_{j,l,t} \cdot \dot{q}_{j,l,t} = \dot{q}_{j,l,t}$ was already addressed previously, we join (4.14) to the (4.13a)-(4.13e) formulation, establishing:

$$\begin{aligned} &\dot{gh}_{l,t} - n_{0l}^{(n)} \sum_{j \in \mathcal{J}_l} \dot{q}_{j,l,t} - n_{1l}^{(n)} \sum_{j \in \mathcal{J}_l} \dot{vs}_{j,l,t} + n_{1l}^{(n)} \dot{ql}_{l,t} \\ &- n_{2l}^{(n)} \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} \leq 0, \forall n, l, t \end{aligned} \tag{4.15a}$$

$$\dot{ql}_{l,t} - h_{0l}^{(r)} \sum_{j \in \mathcal{J}_l} \dot{q}_{j,l,t} - h_{1l}^{(r)} \sum_{j \in \mathcal{J}_l} \overline{z}_{j,l,t} \geq 0, \forall r, l, t \tag{4.15b}$$

$$grh_{l,t}^{min} \overline{z}_{j,l,t} \leq \dot{vs}_{j,l,t} \leq grh_{l,t}^{max} \overline{z}_{j,l,t}, \forall j, l, t \tag{4.15c}$$

$$grh_{l,t}^{min}(1 - \overline{z}_{j,l,t}) \leq \dot{grh}_{l,t} - \dot{vs}_{j,l,t} \leq grh_{l,t}^{max}(1 - \overline{z}_{j,l,t}), \forall j, l, t \tag{4.15d}$$

$$q_{j,l}^{min} \cdot \overline{z}_{j,l,t} \leq \dot{q}_{j,l,t} \leq q_{j,l}^{max} \cdot \overline{z}_{j,l,t} \forall j, l, t \tag{4.15e}$$

$$q_{j,l}^{min}(\overline{z}_{j,l,t} - 1) \leq \dot{qct}_{l,t} - \dot{q}_{j,l,t} \leq q_{j,l}^{max}(1 - \overline{z}_{j,l,t}), \forall j, l, t. \tag{4.15f}$$

Finally, we need to define the net head variable $\dot{grh}_{l,t}$. To that end, the functions representing the forebay and tailrace levels can be approximated by a linear or piecewise linear model[1]. Indeed, in our formulation we express the $\dot{grh}_{l,t}$ by the following constraints,

[1]Those functions can also be obtained using Convex Hull techniques over classic polynomial formulations (GULLIVER; ARNDT, 1991)

which accounts on a linear model for the forebay and tailrace water levels[2]:

$$\dot{grh}_{l,t} - p_{0l}\dot{v}_{l,t} - p_{1l} + d_{0l}\left(\sum_{j\in\mathcal{J}_l}\overline{z}_{j,l,t}\cdot\dot{q}_{j,l,t} + \dot{s}_{l,t}\right) + d_{1l} = 0, \forall l, t,$$

p_{0l}, p_{1l}, d_{1l} and d_{1l} are constants, $\dot{v}_{l,t}$ is the reservoir volume (hm^3) and $\dot{s}_{l,t}$ is the spillage (m^3/s) in period t. In these last constraints we once again set that $\overline{z}_{j,l,t}\cdot\dot{q}_{j,l,t} = \dot{q}_{j,l,t}$, establishing the final formulation for the HPF model:

$$\begin{aligned}
&\dot{gh}_{l,t} - n_{0l}^{(n)}\sum_{j\in\mathcal{J}_l}\dot{q}_{j,l,t} - n_{1l}^{(n)}\sum_{j\in\mathcal{J}_l}\dot{vs}_{j,l,t} + n_{1l}^{(n)}\dot{ql}_{l,t} + \\
&- n_{2l}^{(n)}\sum_{j\in\mathcal{J}_l}\overline{z}_{j,l,t} \leq 0, \forall n, l, t && \text{(4.16a)}\\
&\dot{ql}_{l,t} - h_{0l}^{(r)}\sum_{j\in\mathcal{J}_l}\dot{q}_{j,l,t} - h_{1l}^{(r)}\sum_{j\in\mathcal{J}_l}\overline{z}_{j,l,t} \geq 0, \forall r, l, t && \text{(4.16b)}\\
&\dot{grh}_{l,t} - p_{0l}\dot{v}_{l,t} - p_{1l} + d_{0l}\left(\sum_{j\in\mathcal{J}_l}\dot{q}_{j,l,t} + \dot{s}_{l,t}\right) + d_{1l} = 0, \forall l, t && \text{(4.16c)}\\
&grh_{l,t}^{min}\overline{z}_{j,l,t} \leq \dot{vs}_{j,l,t} \leq grh_{l,t}^{max}\overline{z}_{j,l,t}, \forall j, l, t && \text{(4.16d)}\\
&grh_{l,t}^{min}(1-\overline{z}_{j,l,t}) \leq \dot{grh}_{l,t} - \dot{vs}_{j,l,t} \leq grh_{l,t}^{max}(1-\overline{z}_{j,l,t}), \forall j, l, t && \text{(4.16e)}\\
&q_{j,l}^{min}\cdot\overline{z}_{j,l,t} \leq \dot{q}_{j,l,t} \leq q_{j,l}^{max}\cdot\overline{z}_{j,l,t}\forall j, l, t && \text{(4.16f)}\\
&q_{j,l}^{min}(\overline{z}_{j,l,t}-1) \leq \dot{qct}_{l,t} - \dot{q}_{j,l,t} \leq q_{j,l}^{max}(1-\overline{z}_{j,l,t}), \forall j, l, t. && \text{(4.16g)}
\end{aligned}$$

4.2.2 Formulation with joint probability constraints

In the stochastic version of (4.1) - (4.6), the main idea is to consider wind power in (4.2a) as a random variable. Consider, for the brevity of notation, the wind generation random vector $\widetilde{gw}_w^{(b)} = [\widetilde{gw}_{w,1}^{(b)}, \ldots, \widetilde{gw}_{w,T}^{(b)}]$, where $\widetilde{gw}_w^{(b)} \in \mathbb{R}^T$ is a Gaussian random vector with the variance-covariance matrix $\Sigma^{(b)}$ and mean $\mu^{(b)}$, for each wind farm w at bus b. The vector related to thermal power is defined as $\dot{gt}_i^{(b)} = [\dot{gt}_{i,1}^{(b)}, \ldots, \dot{gt}_{i,T}^{(b)}]$ and likewise for the other variables. With this notation, for each bus $b \in \mathcal{B}_{\mathcal{W}}$, we define the function

[2]Linearity considerations for the forebay function are reasonable in the day ahead operation scheduling. However, it is also possible to include a piecewise linear model for this function, and especially for the tailrace case, since this last has a greater impact in the short-term horizon, when compared to the variation of the forebay level.

$h(\dot{x}, \widetilde{gw}_w^{(b)}) : \mathbb{R}^m \times \mathbb{R}^T \to \mathbb{R}^T$ as follows:

$$h(\dot{x}, \widetilde{gw}_w^{(b)}) = \sum_{i \in \mathcal{I}^{(b)}} \dot{gt}_i + \sum_{w \in \mathcal{W}^{(b)}} \widetilde{gw}_w + \sum_{l \in \mathcal{L}^{(b)}} \dot{gh}_l + \dot{df}^{(b)} + \\ - \sum_{b' \in \mathcal{B}_-^{(b)}} \dot{f}^{(b',b)} + \sum_{b' \in \mathcal{B}_+^{(b)}} \dot{f}^{(b,b')}, \tag{4.17}$$

where $\dot{x} \in X$ is of appropriate size m and denotes all decision variables in a generic notation.

To model the uncertainty, constraints (4.2a) become the following joint PC:

$$\mathbb{P}\left[\; ld_b^{low} \leq \;\; h(\dot{x}, \widetilde{gw}_w^{(b)}) \;\; \leq ld_b^{up} \; \right] \geq p, \forall b \in \mathcal{B}_{\mathcal{W}}, \tag{4.18}$$

where p is a user-defined probability representing the desired safety level (for instance, $p = 95\%$), ld_b^{low} and ld_b^{up} are $\mathbb{R}^T$ boundary parameters. Hence, we want to meet the load with a probability level of p, and the interpretation of equation (4.18) is to offer a policy $\dot{x}$ such that the probability of "deviation" of the load, induced by the uncertainty of $\widetilde{gw}_w^{(b)}$ is not smaller than p. This deviation should fall in between the lower ld_b^{low} and upper ld_b^{up} thresholds, for all the stages T (taken jointly). Discussions about the motivation on the use of bilateral inequalities (4.18) rather than a "direct interpretation" such as $\mathbb{P}\left[h(\dot{x}, \widetilde{gw}_w^{(b)}) = ld^{(b)}\right] \geq p$ are made in (van ACKOOIJ, 2013) for the separable case of $h(\cdot)$ and (HENRION, 2007; LUBIN; BIENSTOCK; VIELMA, 2016) for specific cases when the random variables multiply the decision variables. Indeed, if the random vector $\widetilde{gw}_w^{(b)}$ admits a probability density function concerning to the Lebesgue measure, (e.g., the Gaussian case), for a given policy $\dot{x} \in X$ the probability $\mathbb{P}\left[h(\dot{x}, \widetilde{gw}_w^{(b)}) = ld^{(b)}\right] = 0$ and, hence, no safe policy exists.

As observed in (4.17), the decision variables and random vector are kept in the separable structure $h(\dot{x}, \widetilde{gw}_w^{(b)}) = h(\dot{x}) + \widetilde{gw}_w^{(b)}$. For buses which are not supplied by wind farms, the load requirement constraint is assumed to hold deterministically, i.e. $h(\dot{x}) = ld^{(b)}$. Using the separable property, and following Section 3.4.2, after matrix manipulation, it is possible to rewrite (4.18) in a rectangle pattern which has a two-sided

decision variable dependent structure, i.e.:

$$\phi^{(b)}(\dot{x}) = \mathbb{P}\left[a^{(b)} + A^{(b)}\dot{x} \leq \tilde{\xi}^{(b)} \leq b^{(b)} + B^{(b)}\dot{x}\right] \geq p, \tag{4.19}$$

where $b \in \mathcal{B}_{\mathcal{W}}$ and $\tilde{\xi}^{(b)}$ is a random vector $\tilde{\xi}^{(b)} \sim \mathcal{N}(\mathbf{0}, R^{(b)})$ with $R^{(b)}$ the $\mathbb{R}^T \times \mathbb{R}^T$ a non-singular correlation matrix. Indeed, the vectors $a^{(b)}, b^{(b)} \in \mathbb{R}^T$ and $T \times m$ matrices $A^{(b)}$ and $B^{(b)}$ are algebraically manipulated to standardize the original $\widetilde{gw}_w^{(b)}$ into $\tilde{\xi}^{(b)}$.

Given that $\tilde{\xi}^{(b)}$ has a log-concave probability distribution, it follows from (PRÉKOPA, 1995, Thm 10.2.1), that the CDF $\phi^{(b)}(\dot{x})$ is log-concave and, as a consequence, the constraint (4.19) defines a convex set $M = [\dot{x} \in \mathbb{R}^m | \phi^{(b)}(\dot{x}) \geq p]$ for any probability level $p \in [0, 1]$. Based on these considerations, HTUC in problem (4.1) - (4.6) under wind power uncertainty via joint PC modelling is expressed in the following compact form:

$$\min_{\dot{x} \in \mathbb{R}^{m_1} \times \{0,1\}^{m_2}} \{c^{\mathsf{T}}\dot{x} \ : \text{s.t. } A^{\mathtt{in}}\dot{x} \leq b^{\mathtt{in}}, \phi^{(b)}(\dot{x}) \geq p, \forall b \in \mathcal{B}_{\mathcal{W}}\}, \tag{4.20}$$

where $A^{\mathtt{in}}$ and $b^{\mathtt{in}}$ represents the constraints (4.2a) - (4.6) except for the load requirement constraints (4.2a) supplied by wind farms, i.e. $b \notin \mathcal{B}_{\mathcal{W}}$. Additionally m_1 and m_2 are integers of appropriate size such that $m = m_1 + m_2$.

4.3 ALGORITHMS FOR MILP WITH PROBABILITY CONSTRAINTS

The feasible set of (4.20) is not convex since it is the intersection of a convex set with a partially discrete set X. Hence, the usage of a classic algorithm of stochastic programming with PC such as the supporting hyperplane method, e.g., (PRÉKOPA, 1995) is ruled out. In its original form, this method constructs iteratively an outer polyhedral approximation of the convex set defined by the joint PC, see Algorithm 1. The polyhedral approximation is built from the intersection of supporting half-spaces at iterate points $\dot{x}_k$ on the boundary of the PC feasible set, as discussed in Section A.4. An essential step in this method, is a bisection procedure that, starting from a Slater point x^S (i.e. $\phi(x^S) > p$) and a lower bound solution x_k^L, takes advantage of the convexity of the problem to find the feasible point $\dot{x}_k$ by a convex combination of these last two points. The main

difficulty in applying this classic algorithm in the proposed MILP is that, since the set X is nonconvex, due to its non-continuity, it is not ensured that the bisection procedure would find a feasible solution to the problem. However, to overcome this limitation, this work proposes a new algorithm that, using the information provided by the bisection procedure, defines an upper bound solution to (4.20). Moreover, at each iteration, an estimation of the optimality gap is available, and convergence of the algorithm to an optimal solution is also guaranteed. We provide the algorithm in the next section.

4.3.1 The exact supporting hyperplane method for MILP with probability constraints

We will employ the short notation $h_{\alpha}^{(b)}(\dot{x}) = a^{(b)} + A^{(b)}\dot{x}$ and $h_{\beta}^{(b)}(\dot{x}) = b^{(b)} + B^{(b)}\dot{x}$ to abbreviate elements in (4.19).

In Algorithm 2 the Slater point x^S can be obtained with a heuristic procedure or by the use of a sample method as explained in Section 4.3.3 with high enough probability $p_{sb} > p$. The gradient $\nabla\phi^{(b)}(\dot{x}_k)$ is computed by the procedures that are shown in Section 3.4.1, and the Genz code (GENZ, 1992) which evaluates the multivariate Gaussian distribution functions with high precision. Indeed, following (van ACKOOIJ et al., 2010), computing $\phi^{(b)}(\dot{x})$, $\nabla\phi(\dot{x}_k)$ can be analytically lead back to evaluating multivariate Gaussian distribution functions. The algorithm provides a feasible solution at each iteration, contrary to (ARNOLD et al., 2014) where only the outer approximation is used. Convergence of the algorithm is ensured by classic approaches of convex optimization and cutting plane methods (e.g., (van ACKOOIJ; OLIVEIRA, 2016)). Figure 7 illustrates two iterations of the algorithm, showing the relationship between the solutions. Light blue area and points (integer solutions) represent the non-continuous feasible set of the relaxed problem in Step 2. The purple area defines the convex set defined by the PC, not known a priori at the start of the algorithm. The red solid lines represent the cutting planes of the polyhedral approximations of the PC calculated by the Step 6, traced over $\dot{x}_k$ found by the bisection procedure between x^S and x_k^L in Step 3.

The here used interpolation method is a classic bisection algorithm that starts from

Algorithm 2 - Supporting Hyperplane Method for MILP Problem

1: (Initialisation) Let $x^S \in \mathbb{R}^m$ a Slater point for (4.20). Set $k \leftarrow 1$ and the stopping tolerance $\varepsilon > 0$.

2: (Lower bound) Solve the MILP (4.20) replacing the joint PC by the following set of constraints:

$$\left\langle \nabla \phi^{(b)}(x_i), \dot{x} - x_i \right\rangle \geq 0, i = 1, \ldots, k-1, \forall b \in \mathcal{B}_{\mathcal{W}}$$

to compute x_k^L. The objective function value is a lower bound L_k on the optimal value of (4.20).

3: (Interpolation) Use a bisection procedure to compute the largest λ such that $\dot{x}_k := \lambda x_k^L + (1-\lambda)x^S$ satisfies $\phi^{(b)}(\dot{x}_k) \approx p$ for each bus $b \in \mathcal{B}_{\mathcal{W}}$. Thus $\dot{x}_k$ attains the PC defining the parameters $\alpha_k^{(b)} := h_\alpha^{(b)}(\dot{x}_k)$ and $\beta_k^{(b)} := h_\beta^{(b)}(\dot{x}_k)$, for $\forall b \in \mathcal{B}_{\mathcal{W}}$.

4: (Upper Bound) With $\alpha_k^{(b)}$ and $\beta_k^{(b)}$ as fixed parameters, solve the following problem:

$$\begin{aligned} \min_{\dot{x} \in \{x \in \mathbb{R}^{m_1} \times \{0,1\}^{m_2} \,:\, A^{\text{in}} x \leq b^{\text{in}}\}, \dot{\lambda}_v^{(b)} \geq 0, \sum_{v=1}^k \dot{\lambda}_v^{(b)} = 1, \forall b \in \mathcal{B}_{\mathcal{W}}} & \quad c^{\mathsf{T}} \dot{x} \\ \text{s.t.} - \sum_{v=1}^{k} \dot{\lambda}_v^{(b)} \alpha_v^{(b)} & \leq -h_\alpha^{(b)}(\dot{x}), \forall b \in \mathcal{B}_{\mathcal{W}} \\ + \sum_{v=1}^{k} \dot{\lambda}_v^{(b)} \beta_v^{(b)} & \leq +h_\beta^{(b)}(\dot{x}), \forall b \in \mathcal{B}_{\mathcal{W}} \end{aligned} \tag{4.21}$$

which provides a feasible solution x_k^U and an upper bound U_k for (4.20).

5: (Stopping test) If $\frac{|U_k - L_k|}{|L_k|} < \varepsilon$, stop. The solution is x_k^U. Otherwise, go to the next step.

6: (Oracle call): Call an "oracle" to compute gradients $\nabla \phi^{(b)}(\dot{x}_k)$, where $\dot{x}_k$ is the point of the last Interpolation Step 3. Using the gradient values, built new constraints and add them to the set of cutting planes of Step 2. Set $k \leftarrow k+1$ and return to Step 2.

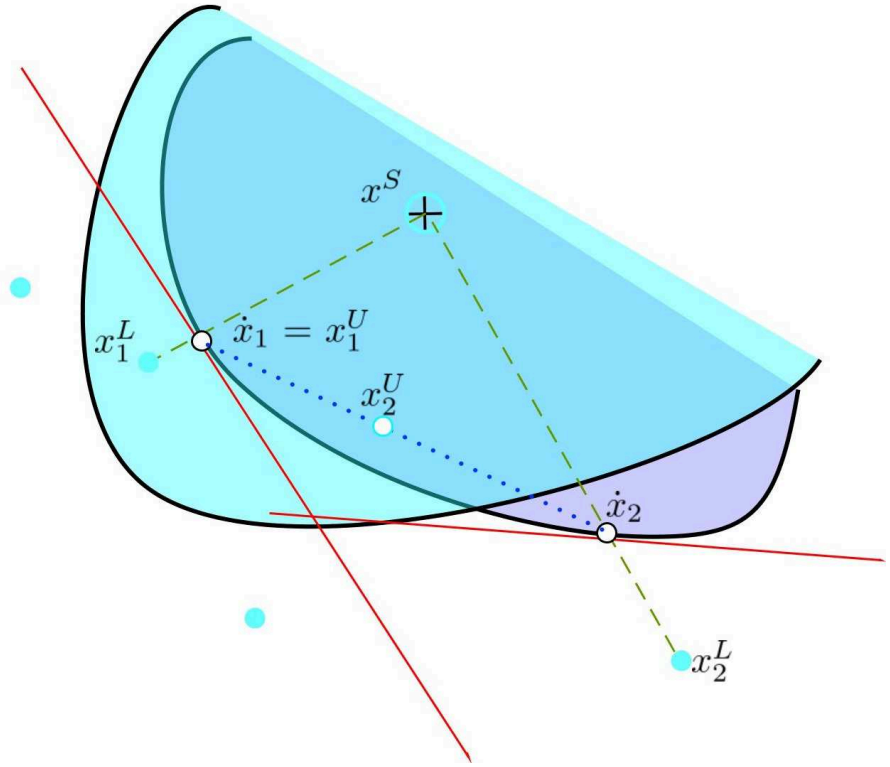
tracing a linear segment between the lower bound solution iterate x_k^L and the Slater x^S. This segment is divided into two equivalent sized sections in order to certify where the root is located, i.e. the $\dot{x}_k$ such that $\phi^{(b)}(\dot{x}_k) \approx p$. This information is provided by the evaluation of the underlying function on the limits of the segments, defining which segment passes through the root. Once the segment is found, it is again divided into two smaller sections and the bisection algorithm runs up to the convergence ($\phi^{(b)}(\dot{x}_k) \approx p$) and $\dot{x}_k$ is defined. Note that $\dot{x}_2$ found by the bisection procedure is not feasible, since it is not inside the blue area. However we still use it to compute and add a new supporting hyperplane. The blue pointed lines are the boundaries of the convex combination of the bisection procedure iterates $\dot{x}_k$, forming a convex set that limits the set of feasible points x_k^U in Step 4.

In Step 4 is interesting to realize that the coefficients $\alpha_v^{(b)}$, for $v = 1, \ldots, k$, represent solutions to the "lower" partition of the PC. When we establish the constraint $h_\alpha^{(b)}(\dot{x}) \leq \sum_{v=1}^{k} \dot{\lambda}_v^{(b)} \alpha_v^{(b)}$ in (4.21), we are asking that the PC constraint is met in the optimization problem at least for the probability p achieved by the interpolation steps. The same is done for the "upper" partition of the PC, by the coefficients $\beta_v^{(b)}$, for $v = 1, \ldots, k$. Since in problem (4.21) the original UC problem is defined by the constraints $A^{\mathtt{in}} x \leq b^{\mathtt{in}}$, the solution of (4.21) attains the UC constraints and the PC, being so a feasible solution for (4.20).

4.3.2 Problem using individual probability constraints

Aiming at comparing the results obtained by the Algorithm 2, this section presents a popular (approximation) problem replacing (4.19) by the associated individual PCs. In the present reformulation of the former problem, for each bus b, the joint PCs are approximated by the individual counterparts $\mathbb{P}\left[a_t^{(b)} + A_t^{(b)} \cdot \dot{x} \leq \tilde{\eta}_t^{(b)}\right] \geq p, \forall b \in \mathcal{B}_{\mathcal{W}}$ and $\mathbb{P}\left[\tilde{\eta}_t^{(b)} \leq b_t^{(b)} + B_t^{(b)} \cdot \dot{x}\right] \geq p, \forall b \in \mathcal{B}_{\mathcal{W}}, \forall t \in \mathcal{T}$. The index t in $a_t^{(b)}$ is the t-th component of the vector $a^{(b)} \in \mathbb{R}^T$ and the tth line of the matrix $A^{(b)} \in \mathbb{R}^T \times \mathbb{R}^m$. Note that $\tilde{\eta}_t^{(b)}$ is a standard Gaussian random one-dimensional variable, and hence any dependencies in $\tilde{\xi}^{(b)}$

Figure 7 – Iterations of the exact supporting hyperplane method for MILP with PC.



Source: Thesis results.

are ruled out. Now, we consider the following version of model (4.20):

$$\begin{aligned} &\min_{\dot{x}\in\{x\in\mathbb{R}^{m_1}\times\{0,1\}^{m_2}\,:A^{\text{in}}x\leq b^{\text{in}}\}} c^{\mathsf{T}}\dot{x} \\ &\text{s.t.} \quad a_t^{(b)} + A_t^{(b)}\cdot\dot{x} \leq F^{-1}(1-p), \forall b\in\mathcal{B}_{\mathcal{W}}, \forall t\in\mathcal{T} \\ &\quad -b_t^{(b)} - B_t^{(b)}\cdot\dot{x} \leq -F^{-1}(p), \forall b\in\mathcal{B}_{\mathcal{W}}, \forall t\in\mathcal{T}, \end{aligned} \tag{4.22}$$

where $F^{-1}(p)$ and $F^{-1}(1-p)$ represent the inverse of a standard Gaussian CDF at level p and $1-p$, respectively.

4.3.3 Sample based algorithm

Introducing another approach to compare the results, this section derives the popular sample-based (i.e., discretized) version of (4.20). Following (LUEDTKE; AHMED, 2008; LUEDTKE; AHMED; NEMHAUSER, 2010), the strategy starts with the generation of Q samples $\tilde{\xi}_q^{(b)}$ of size T, associated binary variables $\overline{z}_q$ and the big-M parameter M. The

discretized version of (4.20) is:

$$
\begin{aligned}
&\min_{\dot{x}\in\{x\in\mathbb{R}^{m_1}\times\{0,1\}^{m_2}\,:A^{\text{in}}x\leq b^{\text{in}}\},\overline{z}_q\in\{0,1\}^Q} c^{\mathsf{T}}\dot{x} \\
&\text{s.t.} \quad a_t^{(b)} + A_t^{(b)}\cdot\dot{x} \leq +\tilde{\xi}_{t,q}^{(b)} + \overline{z}_q\cdot M \\
&\quad - b_t^{(b)} - B_t^{(b)}\cdot\dot{x} \leq -\tilde{\xi}_{t,q}^{(b)} + \overline{z}_q\cdot M \\
&\quad \forall b\in\mathcal{B}_{\mathcal{W}}, \forall t\in\mathcal{T}, \forall q\in\mathcal{Q} \\
&\quad \sum_{q=1}^{Q}\overline{z}_q \leq Q\cdot(1-p_{sb}),
\end{aligned}
\tag{4.23}
$$

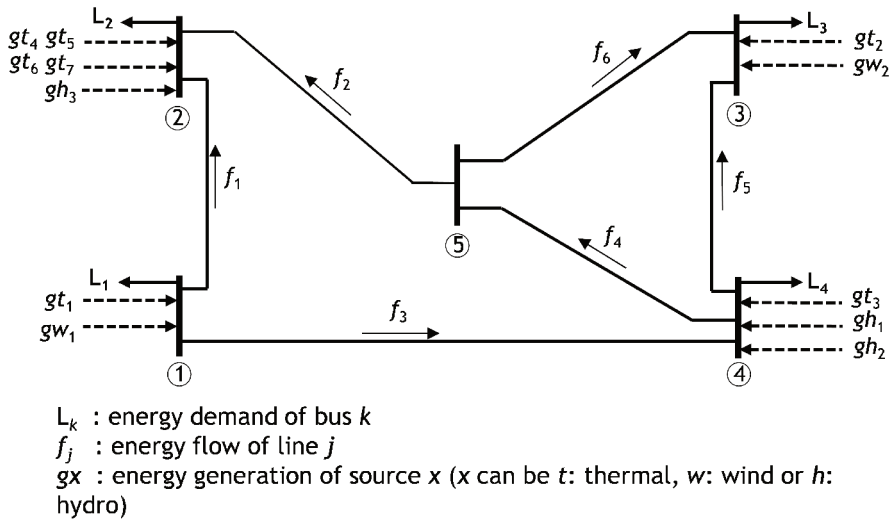
where p_{sb} is a higher probability level than the desired probability level p, as explained in the bibliography survey in Section 2.2 and in the simulation Section 4.4.2. It is important to highlight that (4.23) can be strengthened with additional inequalities, e.g., (KUCUKYAVUZ, 2012). However, since the objective is to perform a comparison, this work have opted for a simple variant.

4.4 COMPUTATIONAL RESULTS

The numerical results consider a day ahead horizon, discretized in hourly steps. The hydrothermal system, based on a reduced variant of the Brazilian system, is composed of 21 units (12 hydro units located in 3 reservoirs, 7 thermal and 2 wind farms). The transmission system has 6 lines, 5 buses and connects the generating units to 4 load demands. Figure 8 depicts the system, where hydro unit gh_1 is upstream to gh_2 with a fixed one hour water travelling delay. The total installed capacity of each generation type is: Hydro - 4335 MW, Thermal - 1392 MW and Wind - 500 MW.

The wind farms information comes from a proprietary database of Labplan-UFSC, which were rescaled to provide a proportion of wind generation similar to the Brazilian system (close to 8% of the total generation for the 5-bus system). The data comprises 3 months of hourly observations of generation, from June to August, of each wind farm. Since the two sets of information refer to wind farms located in distant parts of Brazil, they have a very low spatial correlation between them, close to 0.36, and so, for the sake of

Figure 8 – Power system schematic diagram.



Source: Thesis results.

direct interpretation, they were considered independent in the model. More important to be considered is the periodic correlation of the wind farms, i.e., the correlation between the sequence of hours of a single wind farm. These correlations range, for example, considering wind farm of bus 1, from 0.99 of two sequences of hours (hour 0 and hour 1) to 0.23 of hour 0 and hour 23 of wind. All those hourly correlations are considered in the model.

The computational results were performed on an Intel Xeon CPU with 2.60GHz, using 8 threads and 128 GB RAM. All MILPs were solved using Gurobi 7.5.2 called from C++. The normal probabilities in the probability level verifications and gradients calculations were carried out by the use of a Genz based code (GENZ, 1992). All MILPs were adjusted with 0.4% optimality gap tolerance. The PC rate of convergence was $\varepsilon = 0.5\%$, used as stopping criteria for simulations of algorithm in Section 4.3.1, and $M = 4$ was found sufficient for the algorithm in Section 4.3.3. All the Monte-Carlo out-of-sample verifications were performed with 1,000 wind power scenarios.

4.4.1 Supporting hyperplane method results

This section presents the algorithm steps and some numerical insights into its performance. The initial simulation was performed with a joint probability level of $p = 95\%$.

Following the steps of the Algorithm 2 presented in Section 4.3.1, the Slater point $x^S \in \mathbb{R}^m$ of Step 1 was empirically computed by solving the following MILP:

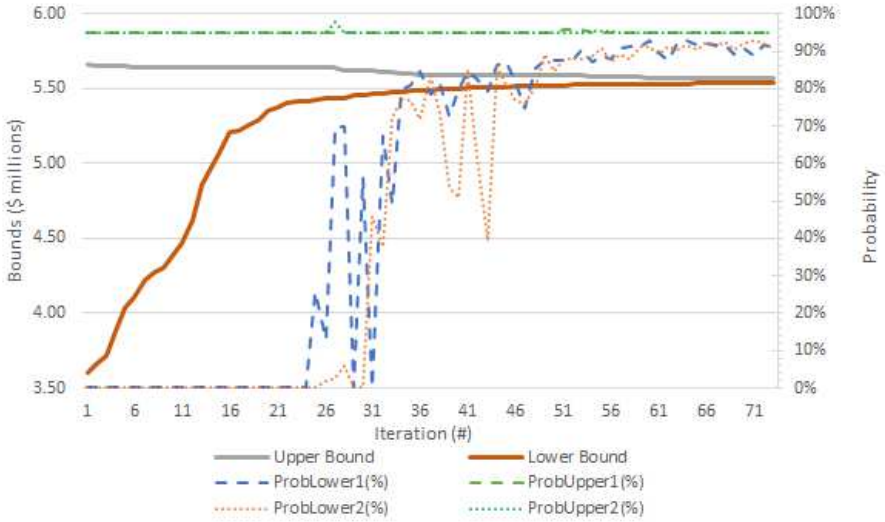
$$\begin{aligned} &\min_{\dot{x} \in \{x \in \mathbb{R}^{m_1} \times \{0,1\}^{m_2} \,:\, A^{\text{in}} x \leq b^{\text{in}}\}} c^{\mathsf{T}} \dot{x} \\ &\text{s.t.} \quad h(\dot{x}) = ld^{(b)} - \mu^{(b)}, \forall b \in \mathcal{B}_{\mathcal{W}}. \end{aligned} \tag{4.24}$$

The solution x^S of (4.24) evaluates the CDF $\phi^{(b)}(x^S)$ in 98.10% for bus 1 and 98.83% for bus 3. Step 2 of the algorithm is performed solving (4.20), observing that in the first iteration the set of cutting planes is still empty. The solution x_1^L evaluates the CDF $\phi^{(b)}(x_1^L) \approx 0\%$ for both buses 1 and 3 and a lower bound of $L_1 = \$3.60 \cdot 10^6$. In Step 3 the bisection procedure uses the solutions x_1^L and x^S. For bus 1 the bisection procedure converges after 10 iterations, resulting in a $\lambda \approx 0.07$ and $\phi^{(b)}(\dot{x}_1) \approx 95.00\%$. Bus 3 converges after 11 iterations, with $\lambda \approx 0.03$ and $\phi^{(b)}(\dot{x}_1)^{(b)} \approx 95.03\%$. Keeping these solutions, we compute $\alpha_1^{(b)}$ and $\beta_1^{(b)}$ for bus 1 and 3.

Step 4 solves the MILP, setting $\alpha_1^{(b)}$ and $\beta_1^{(b)}$ as parameters, defining the solution x_1^U, and providing an upper bound of $U_1 = \$5.66 \cdot 10^6$. The stopping test of Step 5 results in $(5.66 - 3.60)/3.60 \approx 57\% > \varepsilon = 0.5\%$, hence the algorithm does not stop. In Step 6, the oracle is called to enrich the cutting plane model for the subsequent next iteration of the algorithm. Convergence is reached in 49.03 *min.* after 73 iterations with an upper bound of $U_{73} = \$5.57 \cdot 10^6$ and a probability level of $\phi^{(b)}(x_{73}^U) \approx 95.00\%$ and 95.00% for bus 1 and 3, respectively. Figure 9 details the convergence of the algorithm, observed by the approximation between U_k and L_k, and the upper and lower joint probability of the solutions for both buses. This figure also shows that a feasible solution x_k^U is provided in all iterations (i.e., $p \geq 95\%$).

Table 4 presents the total time of each step, in % of total time. The L.Bound step comprises the sum of the total time spent to calculate all the Lower bound solutions in Step 2 of algorithm 2. The Bisection P. is the total time of Step 3, U.Bound is the total time of Step 4 and Gradients is the total time of Step 6. From this table we note that the two costly steps are the Lower Bound evaluation and the calculation of gradients,

Figure 9 – Supporting Hyperplane Results at probability level of 95%.



Source: Thesis results.

Table 4 – Supporting hyperplane method at p=95% - Time per step

Step	L. Bound	Bisection P.	U. Bound	Gradients
Time (%)	37.93	13.04	17.67	28.79

Source: Thesis results.

both are in common with the former Algorithm 1, what is a compelling result in the perspective that the new algorithm does not increase the computational burden sharply with the additional steps.

To understand the sensitivity of the proposed supporting hyperplane method, we have conducted a series of experiments varying the probability level p. The Slater point is the same for all simulations. The summary of the results is in Table 5, where the small deviations from the desired probability level are due to the precision level of the bisection procedure. This table reveals that with the increasing probability level, the cost to dispatch the system also increases, going from $\$5.50 \cdot 10^6$ at $p = 90\%$ to $\$5.64 \cdot 10^6$ at $p = 97.5\%$. The increasing cost stems from two contributions: more thermal generation reflected in a higher immediate cost and more hydroelectric generation with higher associated future costs captured by the expected future cost function. Table 5 also reveals that the algorithm tends to converge faster, with fewer iterations and a lower time per iteration when the probability increases. We explain this intuitively by the fact that the two MILPs that are solved per iteration are simpler when the probability level is

Table 5 – Sensitivity to probability level for the mixed integer supporting hyperplane method: 5-bus System

Probability level	90%	95%	97.5%
CPU time [sec]	3,463	2,942	2,321
# of iterations	72	73	66
Joint Prob. $bus = 1$	90.0%	95.0%	97.5%
Joint Prob. $bus = 3$	90.0%	95.0%	97.5%
Tot. Cost [$\$10^6$]	5.50	5.57	5.64
CC Converged Gap	$4.9 \cdot 10^{-3}$	$4.9 \cdot 10^{-3}$	$4.8 \cdot 10^{-3}$

Source: Thesis results.

Table 6 – Sensitivity of Individual PC Method

Probability level	90%	95%	97.5%	99%
CPU time [sec]	10.13	28.95	10.33	7.49
Joint Prob. $bus = 1$	63.75%	78.14%	87.33%	93.87%
Joint Prob. $bus = 3$	71.33%	83.42%	90.65%	95.51%
Tot. Cost [$\$10^6$]	5.33	5.41	5.48	5.55

Source: Thesis results.

higher, since the number of options to dispatch diminishes. In practice, the system has to dispatch more and more on-stack to reduce the risk level, with direct effects on the B&B algorithm of each MILP.

4.4.2 Comparative results with Individual Probabilistic Constraints

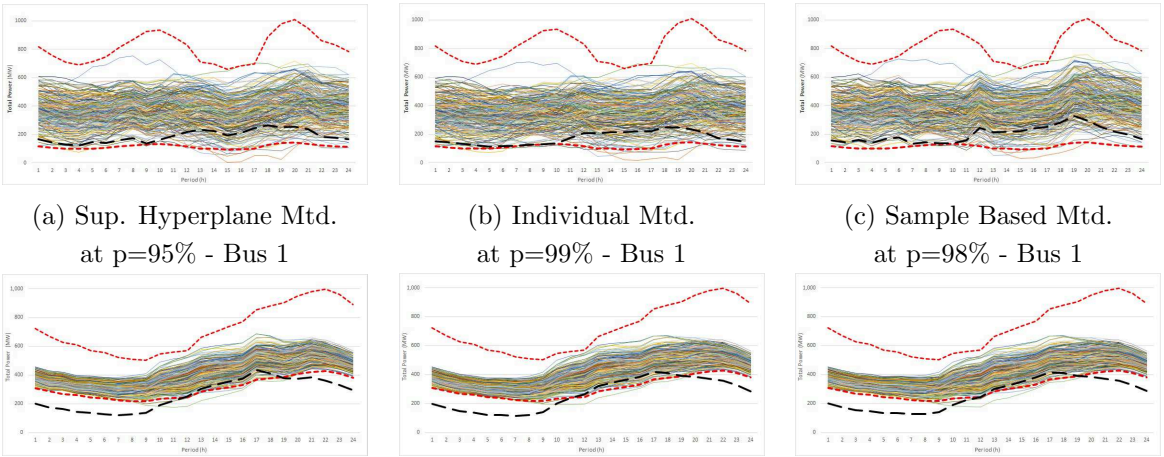
This section compares the model with individual PC with the proposed (joint) model. The results of sensitivity simulations are presented in Table 6.

Even though the requested individual probability level is, for instance, 95% it becomes apparent from Table 6 the actual probability level at buses 1 and 3 for all time steps is significantly lower than requested. Therefore, the solution appears to be of lower cost (close to 3% lower for $p = 95\%$), when compared to the solution given by the supporting hyperplane method, but this is only so since the PC is strongly violated. An alternative strategy is to use heuristics to set a higher individual probability level, for example, 99% in table, to reach the desirable joint probability level close to 95%. Nevertheless, it demands ad-hoc strategies for the specific application if the precision in joint probability level is desired, and optimality is hardly ensured. As an advantage of the individual method is the lower CPU time, which can make this solution a nice initial guess for the supporting

hyperplane method. The strategy of incorporating initial solutions to the exact joint PC algorithm is explored in Section 5.4.

In Figure 10 we show 250 (out of 1,000) a posteriori Monte-Carlo simulations of load constraints in bus $b = 1$ and $b = 3$. The red dashed lines are the load limits set as parameters, and the black dashed lines are the total net generation (including transmission) and not including the random wind generation. The solid coloured lines are the total generation, including wind samples. Table 8 summarizes the results of out-of-sample verifications of the supporting hyperplane method at probability level of $p = 95\%$ (the same simulation as in Table 5) giving results of 94.2% and 95.9% for buses $b = 1$ and $b = 3$, respectively. The same analysis for the individual case at probability of $p = 99\%$ in Table 6 results in 92.8% and 95.9%, reinforcing the topic of lack of precision of the individual method when the probability is taken jointly, especially for bus $b = 1$. In Figure 10 the spatial discrepancies of buses 1 and 3 appear distinctly through the different volatilities, which were the result of the specific variance-covariance matrices $\Sigma^{(b)}$.

Figure 10 – Out-of-sample simulations of the power balance between methods having obtained a similar a posteriori probability level. Dashed red lines are upper and lower load thresholds. Dashed black lines are the total power due to generation and transmission, excluding (random) wind generation. Coloured continuous lines are the whole generation including wind Monte-Carlo simulations.



(a) Sup. Hyperplane Mtd. at p=95% - Bus 1

(b) Individual Mtd. at p=99% - Bus 1

(c) Sample Based Mtd. at p=98% - Bus 1

(d) Sup. Hyperplane Mtd. at p=95% - Bus 3

(e) Individual Mtd. at p=99% - Bus 3

(f) Sample Based Mtd. at p=98% - Bus 3

Source: Thesis results.

Table 7 – Sensitivity to sample size and probability level of Sample Based Method [Results on average of 10 simulations]

# of samples	100	250	500	500
Probability level (p_{sb})	99%	99%	99%	98%
CPU time [sec]	53	121	442	4904
# Resampling tries	1.1	1.5	2.5	1.1
Joint Prob. $bus = 1$	91.68%	94.53%	96.64%	94.75%
Joint Prob. $bus = 3$	93.54%	96.03%	97.17%	95.69%
Tot. Cost [$\$10^6$]	5.56	5.58	5.63	5.57

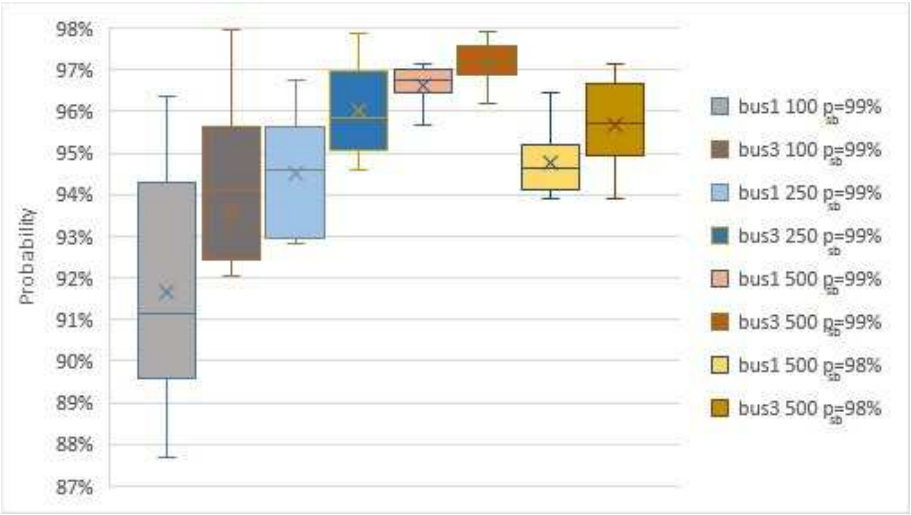
Source: Thesis results.

4.4.3 Comparative results with Sample Based Probabilistic Constraints

Following (LUEDTKE; AHMED, 2008), the sample based method uses a higher probability level p_{sb} (for instance, 99%) than the desired probability level, to have the confidence to achieve feasible solutions with a level close to $p = 95\%$. We have moreover evaluated the conservative theoretical minimum number of samples necessary to offer an upper bound with 90% of confidence and found this number to be $70,000$. Such a large sample size was found to be impractical for computations on standard computer workstations. One of the main difficulties observed in the here analysed sample-based method is to define the number of samples and the safe probability level p_{sb} to have confidence to obtain solutions close to desired probability level $p = 95\%$. In order to evaluate the practically needed number of samples we conducted a series of experiments with the method of Section 4.3.3. For a test with 1000 samples, the memory consumption went over 6 GB RAM in the dedicated server, which provoked memory issues and no solutions could be retrieved when tested in a typical computer workstation (8 GB RAM).

The summary of the results with the average values of 10 simulations of each case are presented in Table 7. We note that the number of resampled scenarios to get a feasible solution increases with the higher number of samples, going from 1.1 to 2.5 sampled scenarios tries, when the probability level (i.e, $p_{sb} = 99\%$) is maintained. The average CPU time tends to increase with the number of samples in the simulation. The table shows the average CPU time of the last simulation (not accounting for resampling simulation time). An additional observation is the difference in CPU time in the simulations with

Figure 11 – Box-plot of a posteriori probability of Sample Based Method simulations.



Source: Thesis results.

#500 samples when $p_{sb} = 99\%$ and $p_{sb} = 98\%$, where a slight change in the probability level changed the MILP time for convergence in 11× on average.

Figure 11 shows a box-plot chart where the probability value of the sample based solutions are evaluated. The range of possible probabilities diminishes with a larger number of samples. For instance, in 10 simulations of 100 samples at a probability level $p_{sb} = 99\%$ for bus $b = 1$ (bus1 100 $p_{sb} = 99\%$ in the figure) the a posteriori joint probability verification ranges from 87.7% to 96.4%, while for 500 samples at the same p_{sb} this range is much tighter. With this figure, we could assert empirically, for example, that the sample-based method with 500 scenarios and a $p_{sb} = 98\%$ lead to solutions with a joint probability level near the desired level $p = 95\%$ for both buses.

Comparing the overall results between the supporting hyperplane method of Table 5 with a probability level of $p = 95\%$ and Table 7 we observe that if high accuracy is desired, large sample size is required. Another strategy would be to evaluate many simulations with fewer samples, for instance, 250, and pick up the best solutions. In any case, some heuristic strategy is required to handle the solutions. In terms of total cost, both methods (in Tables 5 and 7), provide an optimal approximate cost of $\$5.57 \cdot 10^6$ when at close (a posteriori) probability levels. Figure 10 shows an out-of-sample simulation of the power balance of the three methods with similar *a posteriori* joint probability level of nearly

Table 8 – Comparative results between algorithms - Commitment strategy

Method	Supporting Hyperplane	Individual	Sample Based
Probability level	95%	99%	98%
Out-of-sample (Bus 1)	94.20%	92.80%	94.50%
Out-of-sample (Bus 3)	94.90%	95.90%	95.50%
Thermo Gen. [GW]	28.40	28.40	28.15
Hydro Gen. [GW]	45.55	45.34	46.00

Source: Thesis results.

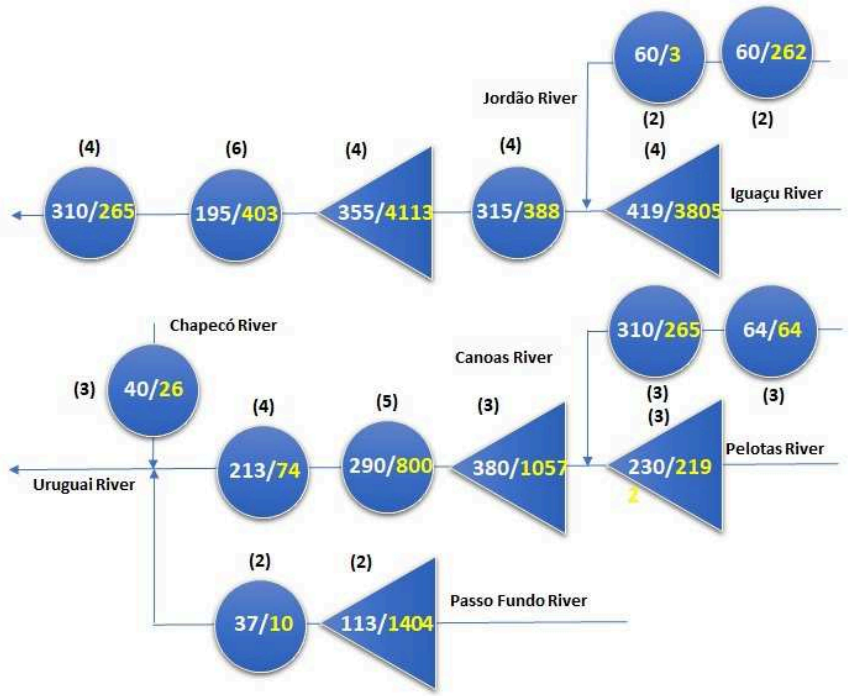
95%. Table 8 reveals that in terms of total thermoelectric and hydroelectric generation the three methods have similar results, with the individual PC slightly prioritizing thermal production rather than hydro, when compared with the others methods.

When examining Table 7, the sample-based method appears to be roughly faster than supporting hyperplane method, depending on the number of sampled scenarios and the probability level p_{sb} . Still, this hides the fact that due to sampling, the true probability level of the given solution is only known in *a posteriori* verification. Moreover, as reinforced by Figure 11, the results depend on the set of drawn samples, and sometimes even the model can be infeasible for a given set, yielding the need for resampling. In particular, if the desired probability level is fixed in $p = 95\%$ the solution can be slightly infeasible given a too optimistic view of the optimal cost (case of #100 samples in Table 7) or strictly feasible yielding a pessimistic view of the optimal cost (case of #500 samples and $p_{sb} = 99\%$). In this aspect, the method deeply differs from the proposed supporting hyperplane method, which allows one to set exactly the desired probability level. Moreover, the obtained solutions do not depend on "a drawn set of scenarios".

4.4.4 Case study: 46-bus system

In this section, we apply the algorithm to a large-scale computational instance, based on the Southern Brazilian system, with 16 hydro plants (52 units), 11 thermal plants, and two wind farms. The power capacities are 11.82, 3.47, and 0.5 GW for hydro, thermal and wind, respectively. The demand fluctuates around 10 and 12.5 GW over the 24-hours planning horizon, and it is distributed in an electrical network with 46 buses

Figure 12 – Hydro plants data - 46-Bus system: Triangles indicate reservoirs with regularization capacity and circles indicate run-of-river plants.



Source: Thesis results.

and 95 lines. For illustrative purposes, the hydro system is presented in Fig. 12. To understand it, consider the only plant in Chapecó River. This plant has 3 units, each one with 40 MW capacity, and a reservoir with 26 hm^3 useful volume. All units in the same plant are identical, and water travelling between two consecutive reservoirs is 1 or 2 hours depending on the distance. As a result, the MILP (4.1)-(4.6) has 10,657 variables (1,512 binaries) and 25,618 constraints. Due to the hydro predominance, the piecewise linear model of the HPF corresponds to approximately 64% of the constraints.

Table 9 presents the optimal solutions with different probability levels. As expected, this instance needed more CPU time for convergence, although, on average, the computational effort is 1.31 times greater in comparison with the small system instances. Table 10 presents the CPU time in each step of the algorithm, where approximately 60% is related to the solution of the upper and lower bounds MILP problems. As the number of wind farms is identical in the computational instances, similar CPU time is spent on computing the interpolation and gradients in each iteration. Based on these results, we see that the number of wind farms is not the bottleneck of the approach. If the dimension of random

Table 9 – Sensitivity to probability level for the mixed integer supporting hyperplane method : 46-bus system

Probability level	90%	95%	97.5%
CPU time [sec]	4,117	3,852	3,869
# of iterations	94	91	88
Joint Prob. $bus = 1$	89.99%	94.96%	97.47%
Joint Prob. $bus = 3$	90.04%	94.95%	97.50%
Tot. Cost [$\$10^6$]	1.07	1.10	1.13
CC Converged Gap	$4.7 \cdot 10^{-3}$	$4.7 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$

Source: Thesis results.

Table 10 – Supporting hyperplane method at p=95% - Tps - 46-bus system

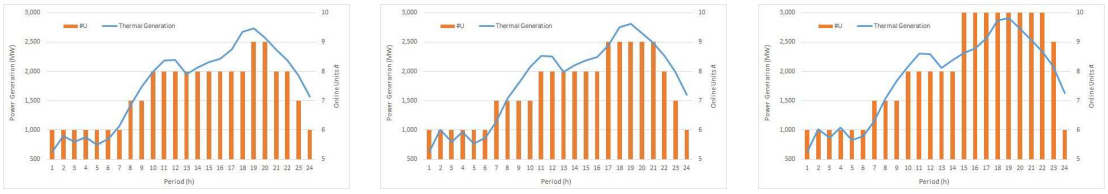
Step	L. Bound	Bisection P.	U. Bound	Gradients
Time (%)	38.00	11.56	21.68	26.38

Source: Thesis results.

variables is maintained, one would expect a linear time complexity in computing interpolation and gradients, since they are strictly related to the number of considered PC in the problem. This linearity in time cannot be ensured for the whole method since the lower and upper bound steps are concerned. However, the use of decomposition techniques could allow a far more efficient solution (SCUZZIATO; FINARDI; FRANGIONI, 2017).

In Figure 13, we present the optimal solutions for the 46-bus system in terms of the thermal dispatch with different levels of probability. From these, it is clear that the requirement to ensure that the level of probability holds induces higher generation for the thermal plants. Note that at probability $p = 90\%$ at most 9 thermal units are committed for just 2 hours, while at probability $p = 97.5\%$ 10 units are committed for 7 hours. The relationship between generation and demand for whole 46-bus system is shown in Fig. 14, where we note how the optimal dispatch meets the load demand, within thresholds that take into account the uncertainty of the wind farms generation. By those last figures, we observe that the hydro generation changes slightly from the level of 9 GW and the thermal generation sources adapt themselves to the required level of probability. Depending on the expected future cost of the water function, i.e., how the system is in terms of water availability, this flexibility can also happen for the hydro dispatch.

Figure 13 – Optimal thermal dispatch for the 46-bus system.



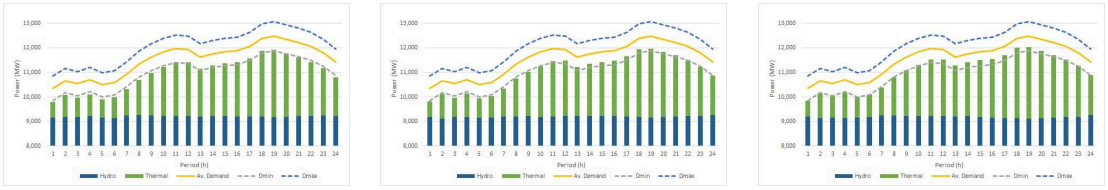
(a) Thermal dispatch at $p = 90\%$

(b) Thermal dispatch at $p = 95\%$

(c) Thermal dispatch at $p = 97.5\%$

Source: Thesis results.

Figure 14 – Relashionship between generation and demand for the 46-bus system.



(a) Relationship at $p = 90\%$

(b) Relationship at $p = 95\%$

(c) Relationship at $p = 97.5\%$

Source: Thesis results.

4.5 IMPROVING THE ALGORITHM'S PERFORMANCE

Although the algorithm performance as presented has already an interesting trade-off between the quality of the solution and computation time, this section aims to explore some strategies to improve the CPU time performance for the proposed supporting hyperplane method for MILP problems. The first strategy that we follow is the investigation on the influence of changing the bisection procedure in the interpolation step of Algorithm 2. After that, we check some strategies for dealing with the step of the lower bound MILP problem.

4.5.1 Interpolation Procedure

One of the most known procedures to find a root of a function is the bisection method. It is due to its versatility, easy implementation and guaranteed convergence. In Section A.6, we presented some alternative methods that are found in the literature. As shown in Table 4, the interpolation procedure takes around 13% of the total running time of the algorithm. The objective of this section is to study the performance of those

Table 11 – Interpolation method - Comparative results

	Prob.	Bisection	Regula Falsi	Ridders
CPU time [min]	90%	61.70	57.88	56.33
Iterations[#]	90%	72	71	75
Elapsed Time	90%	11.58%	11.00%	8.51%
Int. Inter. [in average #]	90%	9.19	7.80	2.86
Tot. Cost [$\$10^6$]	90%	5.50	5.49	5.50
CPU time [min]	95%	47.47	45.64	46.02
Iterations[#]	95%	73	75	74
Elapsed Time	95%	13.48%	17.22%	9.96%
Int. Inter. [in average #]	95%	7.94	9.61	2.77
Tot. Cost [$\$10^6$]	95%	5.57	5.57	5.57
CPU time [min]	97.5%	39.83	43.10	40.27
Iterations[#]	97.5%	66	70	72
Elapsed Time	97.5%	11.71%	19.95%	9.95%
Int. Inter. [in average #]	97.5%	6.33	11.93	2.53
Tot. Cost [$\$10^6$]	97.5%	5.64	5.64	5.64

Source: Thesis results.

methods when it substitutes the bisection method in Step 2 of Algorithm 2. Table 11 summarizes the comparative simulations for the 5-bus system.

From those simulations, we note that the Regula Falsi interpolation method seems to perform worse than the other methods when regarded the Elapsed Time (in % of total time) of the interpolation procedure, mainly for the cases of $p = 95\%$ and $p = 97.5\%$. This can be explained due to the fact that its heuristics to defined the bracket position seems to perform more iterations than the other methods. The simulations also make clear that the elapsed time necessary for the interpolation method, is consistently lower when applied the Ridder's method for all probability levels. Note that the optimal total cost does not change much between methods, but the total CPU time changes as the different points found by each root method have effects on the MILP Branch and Bound convergence for each iteration. This last situation "masks" the advantage on the use of the Ridders method when it is highlighted the total time, but if it is regarded only the interpolation relative time, the Ridder's presents, in average, 22.71% lower time for convergence than the bisection method and 41.00% lower time when compared with the Regula Falsi method.

As shown in Section A.6.3, a drawback on the use of the Ridder's method is that for each iteration of interpolation method it is necessary to compute the underlying function

(i.e. the probability distribution by the code of Genz) two times, when for the other methods it is necessary just once. If we consider that there is no reason to the Genz's code response varies between methods, since the evaluated function is the same in all cases, the use of Ridder's method starts to become more attractive when the interpolation procedure converges in less than the half of iterations of the others methods. As shown in Table 11, the number of iterations of the interpolation method at a probability level of 95%, when it is used the Ridder's method, is 2.77 iterations on average. The bisection and Regula Falsi methods converge 7.94 and 9.61 iterations on average, respectively. For both cases, they are higher than $2 \times 2.77 = 5.54$ iterations, and hence the faster rate of convergence of the the Ridder's method induces a faster convergence for our system. The same trend was noted for simulations with probability levels of 90% and 97.5%. These gains could justify the use of the Ridder's method on our problem, replacing the bisection procedure.

4.5.2 Heuristics on the lower bound problem

As suggested by Table 4, during simulations, the lower bound MILP problem is the step that demands the highest CPU time consumption on average, in the whole algorithm process. This section aims to investigate some strategies to reduce this proportion, or at least diminish the number of necessary lower bound MILP problem evaluations.

The first followed strategy is made noting that the problem that offers feasible solutions to the problem is the upper bound MILP. As the lower bound problem only limits the optimal value from below, the idea was to relax the binary variables, hence defining them in $[0, 1]$ on the lower bound MILP problem, solving so a LP problem. The objective here is to take advantage of the usual faster convergence of LP algorithms when compared to MILP ones. The advantage is understandable since the basic MILP problem algorithm, the Branch and Bound, is basically composed of solving several LP relaxed problems to provide bounds to the original problem (WOLSEY, 1998). On simulations tests with a probability level of 95%, even though the average time for an iteration of

the supporting hyperplane method has decreased, the generated cutting planes has poor quality when compared to one provided by the MILP lower bound problem. It makes the rate of convergence for the whole problem slower than the initially proposed algorithm, making the use of the LP relaxation not attractive, and so this strategy was abandoned.

Through Table 6, it is shown that, despite the lower quality of answers, the individual PC problem provides fast solutions, since it is solved only one MILP per running. The idea of the second tested heuristic is to precompute solutions using the individual PC algorithm, setting in the initial steps of the supporting hyperplane method, cutting planes based on these "good" initial solutions. It is done by computing Individual PC solutions for high enough probability levels. For instance, if the underlying problem has a desired probability level of 95%, pre-compute solutions using the individual problem for the correspondent quantile of the percentiles 95% or 97.5% (if individual solution not feasible to the joint one). These solutions are then plugged into the lower bound problem achieving lower bound solutions for the proposed supporting hyperplane MILP method. Fixing the Slater solution, one can use the interpolation method to compute feasible solutions for the PC, solution that is used to support a cutting plane in this iteration.

The idea behind the last strategy is to help the algorithm, in its starting iterations, to generate active cutting planes that would eventually be constructed only in advanced iterations, and then providing higher lower bounds on early iterations. In test simulations, the settlement of artificial solutions has provided high lower bound solutions at the start of the algorithm, but they did not stay active for long on the subsequent iterations, with the lower bound eventually decreasing over iterations. Therefore, this strategy, in its initial conception was not appealing either.

An alternative strategy that we pursue to this last one is, rather than use initial solutions to plug active constraints, to add the correspondent constraints of the individual PC problem in the lower bound problem of the MILP supporting hyperplane method. It is done by a precomputing step that defines the quantiles that will be used to parametrize the constraints, as set in Individual PC problem (4.22). The advantage of this procedure is that the solution of lower bound MILP problem is at least as high as the individual PC

Table 12 – Sensitivity to probability level with Individual PC added to the lower bound MILP problem

Probability level	90%	95%	97.5%
Ind. prob. level	95%	97.5%	99%
CPU time [sec]	413	754	745
# of iterations	11	23	26
Joint Probability $b = 1$	90.0%	95.1%	97.5%
Joint Probability $b = 3$	90.0%	95.0%	97.5%
Tot. Cost [$\$10^6$]	5.49	5.56	5.64
CC Converged Gap	$4.8 \cdot 10^{-3}$	$4.9 \cdot 10^{-3}$	$4.9 \cdot 10^{-3}$
HPF prod. PP1	-0.15 %	0.23 %	0.00 %
HPF prod. PP2	-1.11 %	-1.12 %	-1.06 %
HPF prod. PP3	-1.44 %	-1.49 %	-1.49 %

Source: Thesis results.

problem. Using a high enough individual probability level, it indeed makes the supporting hyperplane algorithm converge faster than its former formulation (2).

The summary of the results of the strategy of adding individual PC to the supporting hyperplane MILP method is in Table 12. The correspondent joint probability of the individual probability level is presented in Table 6, meaning that for the supporting hyperplane algorithm, the first iteration of the lower bound MILP problem has at least joint probability of 78.14% for Bus $b = 1$ and 83.42% for $b = 3$. Comparing the results with the initially proposed problem of Table 5 we note that for all cases the total CPU time decreased sharply without significant changes on the converged optimal total cost. For instance, for the joint probability level of 95% the total time decreased 74.4% going from 2942 sec. to 754 sec., with a difference in the total cost of 0.09%. These results support the use of this strategy to improve the algorithm performance in terms of time for convergence, maintaining the quality of the solution.

4.6 HYDRO PRODUCTION FUNCTION COMPARATIVE RESULTS

Motivating the use the proposed HPF formulation proposed in Section 4.2.1, we perform comparative simulations regarding a traditional HPF modelling, i.e. several linear piecewise approximation to the nonlinear HPF of each unit. In Table 13 we present a set of simulations using a convex hull technique that traces planes over points of the original nonlinear HPF function. For details of the piecewise linear HPF function, we

Table 13 – Sensitivity to probability level using the linear piecewise production function

Probability level	90%	95%	97.5%
Ind. prob. level	95%	97.5%	99%
CPU time [sec]	1010	1149	1505
# of iterations	21	27	29
Joint Probability $b = 1$	90.0%	95.1%	97.5%
Joint Probability $b = 3$	90.0%	95.0%	97.5%
Tot. Cost [$\$10^6$]	5.53	5.60	5.67
CC Converged Gap	$5.0 \cdot 10^{-3}$	$4.9 \cdot 10^{-3}$	$4.9 \cdot 10^{-3}$
HPF prod. PP1	-0.79 %	-0.79 %	-0.60 %
HPF prod. PP2	7.66 %	7.75 %	6.15 %
HPF prod. PP3	-6.03 %	-6.77 %	-6.11 %

Source: Thesis results.

refer to (DINIZ; MACEIRA, 2008). For classic algorithms of the convex hull method we refer to (GRAHAM, 1972; BARBER; HUHDANPAA, 1996).

Comparing the Table 12, where it was used the formulation of Section 4.2.1, with Table 13 we note that the time for convergence for the proposed model is, in general, lower for all simulations. The last part of those tables make the comparison considering fixed the policy of outflow ($\sum_{j \in \mathcal{J}_l} \dot{q}_{j,l,t} + \dot{s}_{l,t}$) of each plant for the optimal solution. Meaning, it is established how much the simulated hydropower production deviates from the nonlinear HPF (GULLIVER; ARNDT, 1991) of each plant (taken by unit), since this last is supposed to be the best approximation for the real plant production. Table 13 presents a much higher deviation from the nonlinear formulation in general, with worse responses for plants $l = 2$ and 3. This higher error rate is also reflected in the optimal total costs when we note that the piecewise linear approximation simulations have offered higher costs for all cases. An explanation to these higher costs is that the higher level of errors on the production level turns the production of the turbines less efficient, hence it is necessary a higher thermal generation level to meet the load requirement constraint (in PC form), reflecting in higher costs.

5 GENERALIZED CONCAVITY OF BILATERAL PROBABILISTIC CONSTRAINED FUNCTIONS

Throughout this chapter we aim at discussing the second topic of the Thesis. We explore a particular case of the PC optimization problem of the following type:

$$\min_{\dot{x}} \quad k_0(\dot{x}) \tag{5.1}$$

$$\text{s.t.} \quad A \cdot \dot{x} = b_{eq} \tag{5.2}$$

$$\mathbb{P}\left[a_e \leq h_e(\dot{x}, \tilde{\xi}) \leq b_e, e \in \mathcal{E}\right] \geq p, \tag{5.3}$$

where $\dot{x} \in \mathbb{R}^m$, is defined in a continuous set, $\tilde{\xi}$ is a n-dimensional random vector, $\mathbb{P}$ is the cumulative distribution function induced by the random vector $\tilde{\xi}$ on $\mathbb{R}^n$, $h_e : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, $e \in \mathcal{E}$ with $|\mathcal{E}| = E$ an index set, and $p \in [0, 1]$ is the probability level, set as a parameter. The inequality $a_e \leq h_e(\dot{x}, \tilde{\xi}) \leq b_e$ in (5.3), where $a_e, b_e \in \mathbb{R}$, is called two-sided (or bilateral) structure in the present work, that confronts with the (simpler) one-sided (or unilateral) $h_e(\dot{x}, \tilde{\xi}) \leq b_e$ version. A direct interpretation of (5.3) define that the higher is the probability level p, more restrict is the feasible set defined by the PC.

In the present chapter, we focus on the case that $h_e(\dot{x}, \tilde{\xi})$ has a single bilinear structure, i.e., structures that $e = 1$ and the random variables appear in a scalar multiplication with the decision ones, as in $h_e(x, \tilde{\xi}) = \tilde{\xi}^\top x$, for instance. The last structure was introduced in Section 3.2.2.

A necessary and sufficient condition to ensure the local concavity of a given individual and one-sided Gaussian PC is provided in (MINOUX; ZORGATI, 2016), which also develop a sufficient condition for the bilateral case. The proofs in this last work are based on the search of the conditions that make the Hessian matrix of the probability function $f(x) = \mathbb{P}[\tilde{\xi}^\top x \leq b]$ be negative semi-defined, being as a result locally concave. The authors extend their works in (MINOUX; ZORGATI, 2017) defining necessary and sufficient conditions to ensure that the bilateral is locally concave. Finally, in (van ACKOOIJ, 2017) the unilateral case is extended in terms of the conditions to ensure that the unilateral case of

the probability function is r-concave, providing as a result a concept that potentially enlarges the convex set $M(p)$ when compared to the concave proof in (MINOUX; ZORGATI, 2016), indeed an interesting property in applications of PC in programming problems. Further discussions on the challenges of dealing with PC in programming problems are found in the former classical works (BRASCAMP; LIEB, 1976; PRÉKOPA, 1995; RINOTT, 1976; DENTCHEVA, 2009), active lines of research involving convexity of PC constraints are also found in (HENRION; STRUGAREK, 2008, 2011; ZADEH; KHORRAM, 2012; van ACKOOIJ, 2015; van ACKOOIJ; MALICK, 2018; FARSHBAF-SHAKER; HENRION; HOMBERG, 2018). We commonly find those bilinear structures in problems that deal with assets portfolio definition, where x, for instance, is the vector representing the amount of underlying asset through time and $\tilde{\xi}$ are the random variables representing its future prices. This pattern of constraint is extended in the next sections, showing its potentialities in a real based energy portfolio problem.

5.1 PRELIMINARIES, NOTATIONS AND MOTIVATION

The present section makes use of r-concave functions, a generalization of concave functions that plays an important role to define convexity of feasible sets for PC problems. Following the definition of generalized concave functions, as explored in Section A.2 by the Definition A.2.1 regarding r-concave mappings. As checked in (DENTCHEVA, 2009, Lemma 4.8), an important property in such mapping $r \mapsto m_r(a, b, \lambda)$ is that it is a nondecreasing and continuous one. As a result, for $r_1 \leq r_2$, if f is r_2-concave, it is also r_1-concave.

So the work extends generalized concavity properties for bilateral probability distributions $\mathbb{P}\{a \leq \tilde{\xi}^\top x \leq b\}$, where $\tilde{\xi} \sim \mathcal{N}(\mu, \Sigma)$ being the mean $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ a nonsingular covariance matrix. To generalize the approach, we make use of standardized normal distributions, this way we define the following functions:

$$g(x) := \mathbb{P}[a \leq \tilde{\xi}^\top x \leq b] = \Phi(h_b(x)) + \Phi(h_a(x)) - 1 \qquad f_r(x) := g^r(x) \tag{5.4}$$

where,

$$\Phi(x) := \int_{-\infty}^{x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du \tag{5.5}$$

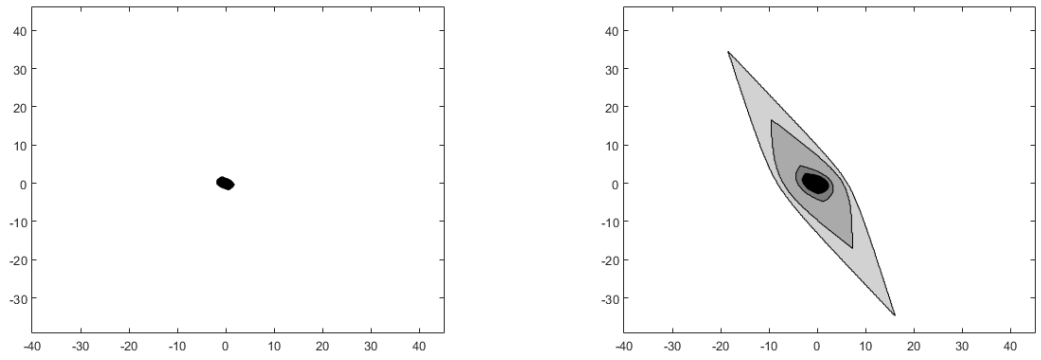
$$h_a(x) := \frac{\mu^\top x - a}{\sqrt{x^\top \Sigma x}} \qquad\qquad h_b(x) := \frac{b - \mu^\top x}{\sqrt{x^\top \Sigma x}}. \tag{5.6}$$

The advantage of the development of generalized concepts of concavity for such probabilistic functions as $g(x)$ is straightforward when one observes the effects in the level sets defined by the constraints (5.3). Indeed, as demonstrated in (van ACKOOIJ, 2017) for the unilateral case of $g(x)$, computable generalization of $r-$concave functions implies in larger (locally) convex feasible sets. A backlash of these generalizations is the fact that, differently from concave functions, $r-$concavity does not preserve the additive property. As a result, the sum of two $r-$concave functions does not define necessarily a $r-$concave function, turning a direct application of the concepts in (van ACKOOIJ, 2017) to bilateral functions nontrivial. Nevertheless, we generalize the concavity properties of the mapping $g(x)$, and the capabilities of the method are motivated, for instance, by the comparison Figure 15. In Figure 15(a) the small black image results from the application of the conditions of Theorem 1 in (MINOUX; ZORGATI, 2017) to Example 5.4.3 below to ensure local convexity of the level set $\{x : g(x) \geq p\}$. In contrast, we apply the here proposed conditions in Figure 15(b), for $\bar{p} = 100\%$, where $\bar{p}$ is a parameter linked with the threshold p, and the advantage in terms of enlargement of the ensured convex set is clear.

5.2 FUNCTIONS DEFINITIONS AND DERIVATIVES

Throughout the present chapter, it is used the differentiations of functions (5.4)-(5.6). Firstly, we offer the ones for $h_\bullet(x)$ functions, where $\nabla_x h_\bullet(x)$ and $\nabla_x^2 h_\bullet(x)$ are the

Figure 15 – Comparison between methods. Lighter colour areas corresponds to different lower r-concavity levels.



(a) Reference. (b) $\bar{p} = 100\%$.

Source: Thesis results.

gradient and Hessian matrix respectively:

$$\nabla_x h_a(x) = \frac{-1}{\sqrt{x^\top \Sigma x}}(-\mu + \frac{(\mu^\top x - a)\Sigma x}{x^\top \Sigma x}), \tag{5.7}$$

$$\begin{aligned}\nabla_x h_a(x)\nabla_x h_a(x)^\top = &\frac{1}{x^\top \Sigma x}(\mu\mu^\top - \frac{(\mu^\top x - a)}{x^\top \Sigma x}[\mu x^\top \Sigma + \Sigma x \mu^\top] \\ &+ \left(\frac{(\mu^\top x - a)}{x^\top \Sigma x}\right)^2 \Sigma x x^\top \Sigma),\end{aligned} \tag{5.8}$$

$$\nabla_x^2 h_a(x) = \frac{1}{(x^\top \Sigma x)^{\frac{3}{2}}}\left[-\Sigma x \mu^\top - \mu(\Sigma x)^\top - (\mu^\top x - a)\Sigma + 3\frac{(\mu^\top x - a)}{(x^\top \Sigma x)}\Sigma x x^\top \Sigma\right]. \tag{5.9}$$

$$\nabla_x h_b(x) = \frac{-1}{\sqrt{x^\top \Sigma x}}(\mu + \frac{(b - \mu^\top x)\Sigma x}{x^\top \Sigma x}), \tag{5.10}$$

$$\begin{aligned}\nabla_x h_b(x)\nabla_x h_b(x)^\top = &\frac{1}{x^\top \Sigma x}(\mu\mu^\top + \frac{(b - \mu^\top x)}{x^\top \Sigma x}[\mu x^\top \Sigma + \Sigma x \mu^\top] \\ &+ \left(\frac{(b - \mu^\top x)}{x^\top \Sigma x}\right)^2 \Sigma x x^\top \Sigma),\end{aligned} \tag{5.11}$$

$$\nabla_x^2 h_b(x) = \frac{1}{(x^\top \Sigma x)^{\frac{3}{2}}}\left[\Sigma x \mu^\top + \mu(\Sigma x)^\top - (b - \mu^\top x)\Sigma + 3\frac{(b - \mu^\top x)}{(x^\top \Sigma x)}\Sigma x x^\top \Sigma\right]. \tag{5.12}$$

And the cross products:

$$\nabla_x h_b(x)\nabla_x h_a(x)^\top = \frac{1}{x^\top \Sigma x}\left(-\mu\mu^\top + \frac{(\mu^\top x - a)}{x^\top \Sigma x}\mu(\Sigma x)^\top - \frac{(b-\mu^\top x)}{x^\top \Sigma x}\Sigma x \mu^\top\right) + \frac{1}{x^\top \Sigma x}\left(\frac{(\mu^\top x - a)(b-\mu^\top x)}{(x^\top \Sigma x)^2}\Sigma x x^\top \Sigma\right), \tag{5.13}$$

$$\nabla_x h_a(x)\nabla_x h_b(x)^\top = \frac{1}{x^\top \Sigma x}\left(-\mu\mu^\top - \frac{(b-\mu^\top x)}{x^\top \Sigma x}\mu(\Sigma x)^\top + \frac{(\mu^\top x - a)}{x^\top \Sigma x}\Sigma x \mu^\top\right) + \frac{1}{x^\top \Sigma x}\left(\frac{(\mu^\top x - a)(b-\mu^\top x)}{(x^\top \Sigma x)^2}\Sigma x x^\top \Sigma\right). \tag{5.14}$$

For the $g(x)$ function, we calculate the following differentiations:

$$\nabla g(x) = \nabla\Phi(h_b(x)) + \nabla\Phi(h_a(x)) = \frac{e^{-\frac{1}{2}h_b^2(x)}}{\sqrt{2\pi}}\nabla h_b(x) + \frac{e^{-\frac{1}{2}h_a^2(x)}}{\sqrt{2\pi}}\nabla h_a(x), \tag{5.15}$$

$$\nabla^2 g(x) = \frac{e^{-\frac{1}{2}h_b(x)^2}}{\sqrt{2\pi}}(-h_b(x)\nabla h_b(x)\nabla h_b(x)^\top + \nabla^2 h_b(x)) + \frac{e^{-\frac{1}{2}h_a(x)^2}}{\sqrt{2\pi}}(-h_a(x)\nabla h_a(x)\nabla h_a(x)^\top + \nabla^2 h_a(x)). \tag{5.16}$$

We define $f_r(x) := g^r(x)$. And we extend its derivatives:

$$\nabla f_r(x) = r g^{r-1}(x)\nabla g(x) = r g^{r-1}(x)\left(\frac{e^{-\frac{1}{2}h_a^2(x)}}{2\pi}\nabla h_a(x) + \frac{e^{-\frac{1}{2}h_b^2(x)}}{2\pi}\nabla h_b(x)\right). \tag{5.17}$$

And so defining:

$$\nabla f_{r_a}(x) := r g^{r-1}(x)\frac{e^{-\frac{1}{2}h_a^2(x)}}{2\pi}\nabla h_a(x), \tag{5.18}$$

$$\nabla f_{r_b}(x) := r g^{r-1}(x)\frac{e^{-\frac{1}{2}h_b^2(x)}}{2\pi}\nabla h_b(x). \tag{5.19}$$

Then, $\nabla f_r(x) = \nabla f_{r_a}(x) + \nabla f_{r_b}(x)$. Establishing $\nabla^2 f_r(x) = \nabla^2 f_{r_a}(x) + \nabla^2 f_{r_b}(x)$, where:

$$\nabla^2 f_{r_a}(x) = rg^{r-1}(x)\frac{e^{-\frac{h_a^2(x)}{2}}}{\sqrt{2\pi}}\left(\frac{(r-1)e^{-\frac{h_a^2(x)}{2}}}{g(x)\sqrt{2\pi}}\nabla h_a(x)\nabla h_a^\top(x) - h_a(x)\nabla h_a(x)\nabla h_a^\top(x)\right)$$
$$+ rg^{r-1}(x)\frac{e^{-\frac{h_a^2(x)}{2}}}{\sqrt{2\pi}}\left(\nabla^2 h_a(x) + \frac{(r-1)e^{-\frac{h_b^2(x)}{2}}}{g(x)\sqrt{2\pi}}\nabla h_b(x)\nabla h_a^\top(x)\right), \tag{5.20}$$

$$\nabla^2 f_{r_b}(x) = rg^{r-1}(x)\frac{e^{-\frac{h_b^2(x)}{2}}}{\sqrt{2\pi}}\left(\frac{(r-1)e^{-\frac{h_b^2(x)}{2}}}{g(x)\sqrt{2\pi}}\nabla h_b(x)\nabla h_b^\top(x) - h_b(x)\nabla h_b(x)\nabla h_b^\top(x)\right)$$
$$+ rg^{r-1}(x)\frac{e^{-\frac{h_b^2(x)}{2}}}{\sqrt{2\pi}}\left(\nabla^2 h_b(x) + \frac{(r-1)e^{-\frac{h_a^2(x)}{2}}}{g(x)\sqrt{2\pi}}\nabla h_a(x)\nabla h_b^\top(x)\right). \tag{5.21}$$

5.3 CONVEXITY ANALYSIS

This section provides conditions to certificate that the function $g(x)$ preserves locally $r-$concave properties. Initially, in Section 5.3.1, it is developed necessary and sufficient conditions, followed by Section 5.3.2 which uses more straightforward sufficient conditions that are appropriate to be applied in the subsequent sections.

5.3.1 A necessary and sufficient condition

In the following theorem, we define a necessary and sufficient condition to ensure the convexity for the feasible set of the underlying mapping $g(x)$.

Theorem 5.3.1 *Let $\Phi : \mathbb{R} \to [0,1]$ be the one dimensional Gaussian CDF. For $n \geq 1$, let $b \in \mathbb{R}$, $\mu \in \mathbb{R}^n$ and positive definite $n \times n$ matrix Σ be given. Considering $b - a > 0$ for any x such that $\mu^\top x \in [a, b]$, and for any $r < 0$ the mapping $g(x)$ defined in (5.4) is locally r-concave around x, if and only if the following condition holds:*

$$(C) \quad \mu^\top \Sigma^{-1}\mu \leq g_r - 2h_r\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} + c_r^2\left(\mu^\top \Sigma^{-1}\mu - \frac{(\mu^\top x)^2}{x^\top \Sigma x}\right), \tag{5.22}$$

where, considering $h_a = h_a(x)$, $h_b = h_b(x)$, $\pi_a = \frac{e^{-\frac{h_a^2}{2}}}{g(x)\sqrt{2\pi}}$ and $\pi_b = \frac{e^{-\frac{h_b^2}{2}}}{g(x)\sqrt{2\pi}}$, the variables

g_r, h_r and c_r are established below:

$$g_r = \frac{r(\pi_a^2 h_a^2 + \pi_b^2 h_b^2) - (\pi_a h_a + \pi_b h_b)^2 + \pi_a h_a(2 - h_a^2) + \pi_b h_b(2 - h_b^2)}{\pi_a h_a + \pi_b h_b - (r-1)(\pi_a^2 + \pi_b^2) + 2\pi_a \pi_b}, \tag{5.23}$$

$$h_r = \frac{(\pi_a h_a(-(r-1)\pi_a + h_a + \pi_b) + \pi_b h_b((r-1)\pi_b - h_b - \pi_a) - \pi_a + \pi_b)}{\pi_a h_a + \pi_b h_b - (r-1)(\pi_a^2 + \pi_b^2) + 2\pi_a \pi_b}, \tag{5.24}$$

$$c_r^2 = 3 + \frac{(r(\pi_a^2 h_a^2 + \pi_b^2 h_b^2) - (\pi_a h_a - \pi_b h_b)^2 - h_a^3 \pi_a - h_b^3 \pi_b)}{\pi_a h_a + \pi_b h_b} \tag{5.25}$$

$$+ \frac{(\pi_a h_a(-(r-1)\pi_a + h_a + \pi_b) + \pi_b h_b((r-1)\pi_b - h_b - \pi_a) - \pi_a + \pi_b)^2}{((\pi_a h_a + \pi_b h_b) - (r-1)(\pi_a^2 + \pi_b^2) + 2\pi_a \pi_b)) \cdot (\pi_a h_a + \pi_b h_b)}. \tag{5.26}$$

Proof.

From the calculations performed in Section 5.2, where the gradients $\nabla h_a(x)$, $\nabla h_b(x)$ and the Hessian matrix $\nabla^2 h_a(x)$, $\nabla^2 h_b(x)$ and $\nabla^2 f_r(x)$ are established, we define the auxiliary matrix $\nabla^2 f_r(x) = H_r := H_{r_a} + H_{r_b} + H_{r_{ab}}$. Where:

$$H_{r_a}(x) = rg^r(x)\frac{e^{-\frac{h_a^2(x)}{2}}}{g(x)\sqrt{2\pi}}\left(\frac{(r-1)e^{-\frac{h_a^2(x)}{2}}}{g(x)\sqrt{2\pi}}\nabla h_a(x)\nabla h_a^\top(x) - h_a(x)\nabla h_a(x)\nabla h_a^\top(x) + \nabla^2 h_a(x)\right), \tag{5.27}$$

$$H_{r_b}(x) = rg^r(x)\frac{e^{-\frac{h_b^2(x)}{2}}}{g(x)\sqrt{2\pi}}\left(\frac{(r-1)e^{-\frac{h_b^2(x)}{2}}}{g(x)\sqrt{2\pi}}\nabla h_b(x)\nabla h_b^\top(x) - h_b(x)\nabla h_b(x)\nabla h_b^\top(x) + \nabla^2 h_b(x)\right), \tag{5.28}$$

$$H_{r_{ab}}(x) = rg^r(x)\frac{e^{-\frac{h_a^2(x)}{2}}}{g(x)\sqrt{2\pi}}\frac{e^{-\frac{h_b^2(x)}{2}}}{g(x)\sqrt{2\pi}}\left((r-1)\nabla h_b(x)\nabla h_a^\top(x) + (r-1)\nabla h_a(x)\nabla h_b^\top(x)\right). \tag{5.29}$$

Using the definitions $s_r(h_\bullet) := (r-1)\frac{e^{-\frac{1}{2}h_\bullet^2}}{g(x)\sqrt{2\pi}}$, this implies that $s_r(h_a) = (r-1)\pi_a$ and $s_r(h_b) = (r-1)\pi_b$ and we replace these definitions in the following mappings:

$$H_{r_a}(x) = rg^r(x)\pi_a\left((r-1)\pi_a\nabla h_a(x)\nabla h_a^\top(x) - h_a(x)\nabla h_a(x)\nabla h_a^\top(x) + \nabla^2 h_a(x)\right), \tag{5.30}$$

$$H_{r_b}(x) = rg^r(x)\pi_b\left((r-1)\pi_b\nabla h_b(x)\nabla h_b^\top(x) - h_b(x)\nabla h_b(x)\nabla h_b^\top(x) + \nabla^2 h_b(x)\right), \tag{5.31}$$

$$H_{r_{ab}}(x) = rg^r(x)\pi_a\pi_b\left((r-1)\nabla h_b(x)\nabla h_a^\top(x) + (r-1)\nabla h_a(x)\nabla h_b^\top(x)\right). \tag{5.32}$$

Defining the following notation for the values inside the brackets, as in:

$$\hat{H}_{r_a} := \left((r-1)\pi_a\nabla h_a(x)\nabla h_a^\top(x) - h_a(x)\nabla h_a(x)\nabla h_a^\top(x) + \nabla^2 h_a(x)\right),$$

we rewrite the matrix H_r:

$$H_r = rg^r(x)\left(\pi_a\hat{H}_{r_a} + \pi_b\hat{H}_{r_b} + \pi_a\pi_b\hat{H}_{r_{ab}}\right). \tag{5.33}$$

Opening in their primitives, as in Section 5.2, and after rearrangements we have:

$$\begin{aligned}\hat{H}_{r_a} &= \frac{(s_r(h_a)-h_a)}{x^\top\Sigma x}\left[\mu\mu^\top - \frac{(\mu^\top x-a)}{x^\top\Sigma x}[\mu x^\top\Sigma + \Sigma x\mu^\top] + \left(\frac{(\mu^\top x-a)}{x^\top\Sigma x}\right)^2\Sigma x x^\top\Sigma\right] \\ &+ \frac{1}{(x^\top\Sigma x)^{\frac{3}{2}}}\left[-\Sigma x\mu^\top - \mu(\Sigma x)^\top - (\mu^\top x-a)\Sigma + 3\frac{(\mu^\top x-a)}{(x^\top\Sigma x)}\Sigma x x^\top\Sigma\right], \end{aligned}\tag{5.34}$$

$$\begin{aligned}\hat{H}_{r_b} &= \frac{(s_r(h_b)-h_b)}{x^\top\Sigma x}\left[\mu\mu^\top + \frac{(b-\mu^\top x)}{x^\top\Sigma x}[\mu x^\top\Sigma + \Sigma x\mu^\top] + \left(\frac{(b-\mu^\top x)}{x^\top\Sigma x}\right)^2\Sigma x x^\top\Sigma\right] \\ &+ \frac{1}{(x^\top\Sigma x)^{\frac{3}{2}}}\left[\Sigma x\mu^\top + \mu(\Sigma x)^\top - (b-\mu^\top x)\Sigma + 3\frac{(b-\mu^\top x)}{(x^\top\Sigma x)}\Sigma x x^\top\Sigma\right], \end{aligned}\tag{5.35}$$

$$\hat{H}_{r_{ab}} = \frac{1}{x^\top\Sigma x}\left(-2\mu\mu^\top + \frac{(\mu^\top x-a)-(b-\mu\top x)}{x^\top\Sigma x}[\mu(\Sigma x)^\top + \Sigma x\mu^\top] + \frac{2(\mu^\top x-a)(b-\mu^\top x)}{(x^\top\Sigma x)^2}\Sigma x^\top x\Sigma\right). \tag{5.36}$$

Considering that the positive semidefiniteness of (5.33) is ensured, for $r \leq 0$, if $\bar{H}_r$ below is negative semidefinite:

$$\bar{H}_r = \frac{x^\top\Sigma x}{\pi_a h_a + \pi_b h_b}\Sigma^{-\frac{1}{2}}\left(\pi_a\hat{H}_{r_a} + \pi_b\hat{H}_{r_b} + \pi_a\pi_b\hat{H}_{r_{ab}}\right)\Sigma^{-\frac{1}{2}}. \tag{5.37}$$

Making the necessary rearrangements, we get:

$$\bar{H}_r = \bar{\gamma}\cdot\frac{\Sigma^{\frac{1}{2}}xx^\top\Sigma^{\frac{1}{2}}}{(x^\top\Sigma x)} + \bar{\beta}\cdot\left[\frac{\Sigma^{-\frac{1}{2}}\mu x^\top\Sigma^{\frac{1}{2}} + \Sigma^{\frac{1}{2}}x\mu^\top\Sigma^{-\frac{1}{2}}}{(x^\top\Sigma x)^{\frac{1}{2}}}\right] + \bar{\alpha}\cdot\Sigma^{-\frac{1}{2}}\mu\mu^\top\Sigma^{-\frac{1}{2}} - \bar{\sigma}\cdot I. \tag{5.38}$$

Using $s_r(h_a) = (r-1)\pi_a$ and $s_r(h_b) = (r-1)\pi_b$ we make the simplifications:

$$\begin{aligned}
\bar{\alpha} &= \frac{s_r(h_a)\pi_a - h_a\pi_a + s_r(h_b)\pi_b - h_b\pi_b - 2\pi_a\pi_b}{(\pi_a h_a + \pi_b h_b)} \\
&= (-1 + \frac{(r-1)(\pi_a^2 + \pi_b^2) - 2\pi_a\pi_b}{\pi_a h_a + \pi_b h_b}), \qquad (5.39) \\
\bar{\beta} &= \frac{-s_r(h_a)h_a\pi_a + h_a^2\pi_a - \pi_a + s_r(h_b)h_b\pi_b - h_b^2 + \pi_b + \pi_a\pi_b(h_a - h_b)}{\pi_a h_a + \pi_b h_b} \\
&= \frac{\pi_a h_a(-(r-1)\pi_a + h_a + \pi_b) + \pi_b h_b((r-1)\pi_b - h_b - \pi_a) - \pi_a + \pi_b}{\pi_a h_a + \pi_b h_b}, \qquad (5.40) \\
\bar{\gamma} &= \frac{(((r-1)\pi_a - h_a)h_a^2 + 3h_a)\pi_a + (((r-1)\pi_b - h_b)h_b^2 + 3h_b)\pi_b + 2h_a h_b \pi_a \pi_b}{\pi_a h_a + \pi_b h_b} \\
&= 3 + \frac{(r(\pi_a^2 h_a^2 + \pi_b^2 h_b^2) - (\pi_a h_a - \pi_b h_b)^2 - h_a^3\pi_a - h_b^3\pi_b)}{\pi_a h_a + \pi_b h_b}, \qquad (5.41) \\
\bar{\sigma} &= \frac{\pi_a h_a + \pi_b h_b}{\pi_a h_a + \pi_b h_b} = 1. \qquad (5.42)
\end{aligned}$$

Using the convention $V = \Sigma^{\frac{1}{2}}x$ and $W = \Sigma^{-\frac{1}{2}}\mu$, we write (5.38):

$$\bar{H}_r = \bar{\gamma}\frac{VV^\top}{\|V\|^2} + \bar{\beta}\frac{[WV^\top + VW^\top]}{\|V\|} + \bar{\alpha}WW^\top - I. \qquad (5.43)$$

Considering $Z = d_r\frac{V}{\|V\|} + f_r W$ and $Y = c_r\frac{V}{\|V\|}$, we write (5.43) as follows:

$$\bar{H}_r = -ZZ^\top + YY^\top - I, \qquad (5.44)$$

where $ZZ^\top = d_r^2\frac{VV^\top}{\|V\|^2} + d_r f_r\frac{[VW^\top + WV^\top]}{\|V\|} + f_r^2 WW^\top$ and $YY^\top = c_r^2\frac{VV^\top}{\|V\|^2}$. From (5.44), we have:

$$\bar{H}_r = (c_r^2 - d_r^2)\frac{VV^\top}{\|V\|^2} + (-d_r f_r)\frac{[VW^\top + WV^\top]}{\|V\|} - f_r^2 WW^\top - I. \qquad (5.45)$$

And so, comparing (5.43) with (5.45), we establish the values of the auxiliary variables f_r, d_r and c_r.

For f_r:

$$-f_r^2 = \left(-1 + \frac{(r-1)(\pi_a^2+\pi_b^2) - 2\pi_a\pi_b}{\pi_a h_a + \pi_b h_b}\right) \Leftrightarrow f_r^2 = 1 - \frac{(r-1)(\pi_a^2+\pi_b^2) - 2\pi_a\pi_b}{\pi_a h_a + \pi_b h_b}$$
$$f_r = \sqrt{1 - \frac{(r-1)(\pi_a^2+\pi_b^2) - 2\pi_a\pi_b}{\pi_a h_a + \pi_b h_b}}. \tag{5.46}$$

For d_r:

$$\begin{aligned}
-d_r \cdot \sqrt{1 - \frac{(r-1)(\pi_a^2+\pi_b^2) - 2\pi_a\pi_b}{\pi_a h_a + \pi_b h_b}} &= \frac{\pi_a h_a(-(r-1)\pi_a + h_a + \pi_b) + \pi_b h_b((r-1)\pi_b - h_b - \pi_a) - \pi_a + \pi_b}{\pi_a h_a + \pi_b h_b} \\
d_r^2 \cdot \left(1 - \frac{(r-1)(\pi_a^2+\pi_b^2) - 2\pi_a\pi_b}{\pi_a h_a + \pi_b h_b}\right) &= \frac{(\pi_a h_a(-(r-1)\pi_a + h_a + \pi_b) + \pi_b h_b((r-1)\pi_b - h_b - \pi_a) - \pi_a + \pi_b)^2}{(\pi_a h_a + \pi_b h_b)^2} \\
d_r^2 &= \frac{(\pi_a h_a(-(r-1)\pi_a + h_a + \pi_b) + \pi_b h_b((r-1)\pi_b - h_b - \pi_a) - \pi_a + \pi_b)^2}{((\pi_a h_a + \pi_b h_b) - (r-1)(\pi_a^2+\pi_b^2) + 2\pi_a\pi_b)) \cdot (\pi_a h_a + \pi_b h_b)} \\
d_r &= \sqrt{\frac{(\pi_a h_a(-(r-1)\pi_a + h_a + \pi_b) + \pi_b h_b((r-1)\pi_b - h_b - \pi_a) - \pi_a + \pi_b)^2}{((\pi_a h_a + \pi_b h_b) - (r-1)(\pi_a^2+\pi_b^2) + 2\pi_a\pi_b)) \cdot (\pi_a h_a + \pi_b h_b)}}.
\end{aligned} \tag{5.47}$$

For c_r:

$$c_r^2 = 3 + \frac{(r(\pi_a^2 h_a^2 + \pi_b^2 h_b^2) - (\pi_a h_a - \pi_b h_b)^2 - h_a^3\pi_a - h_b^3\pi_b)}{\pi_a h_a + \pi_b h_b} + d_r^2. \tag{5.48}$$

Considering the procedures clarified in Appendix C.1, we know that in order to have (5.45) negative semidefinite the highest eigenvalue of the matrix (5.45) must be given by the following equation:

$$\|Y\|^2 - \|Z\|^2 + \|Y\|^2 \times \|Z\|^2 - (Z^\top Y)^2 \geq 1, \tag{5.49}$$

where:

$$\|Y\|^2 = c_r^2 \frac{V^\top V}{\|V\|^2} = c_r^2, \tag{5.50}$$

$$\|Z\|^2 = d_r^2 + 2d_r f_r \frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} + f_r^2 \mu \Sigma^{-1} \mu, \tag{5.51}$$

$$(Z^\top Y)^2 = d_r^2 c_r^2 + 2d_r f_r c_r^2 \frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} + f_r^2 c_r^2 \frac{(\mu^\top x)^2}{x^\top \Sigma x}. \tag{5.52}$$

So we are able to discuss the left side of (5.49):

$$\begin{aligned}
\|Y\|^2 - \|Z\|^2 + \|Y\|^2 \times \|Z\|^2 - (Z^\top Y)^2 \quad &= c_r^2 - d_r^2 - 2d_r f_r \tfrac{\mu^\top x}{\sqrt{x^\top \Sigma x}} - f_r^2 \mu^\top \Sigma^{-1} \mu + d_r^2 c_r^2 + \\
&+ 2d_r f_r c_r^2 \tfrac{\mu^\top x}{\sqrt{x^\top \Sigma x}} \\
&+ f_r^2 c_r^2 \mu^\top \Sigma^{-1} \mu - d_r^2 c_r^2 - 2d_r f_r c_r^2 \tfrac{\mu^\top x}{\sqrt{x^\top \Sigma x}} - f_r^2 c_r^2 \tfrac{(\mu^\top x)^2}{x^\top \Sigma x} \\
&= c_r^2 - d_r^2 - 2d_r f_r \tfrac{\mu^\top x}{\sqrt{x^\top \Sigma x}} - f_r^2 \mu^\top \Sigma^{-1} \mu + \\
&+ f_r^2 c_r^2 \mu^\top \Sigma^{-1} \mu - f_r^2 c_r^2 \tfrac{(\mu^\top x)^2}{x^\top \Sigma x} \\
&= c_r^2 - d_r^2 - 2d_r f_r \tfrac{\mu^\top x}{\sqrt{x^\top \Sigma x}} - f_r^2 \mu^\top \Sigma^{-1} \mu + \\
&+ f_r^2 c_r^2 \left(\mu^\top \Sigma^{-1} \mu - \tfrac{(\mu^\top x)^2}{x^\top \Sigma x} \right).
\end{aligned} \tag{5.53}$$

After algebraic manipulations, we observe that if we divide both sides of (5.49) by $\frac{1}{f_r^2}$ from (5.53) we get:

$$\frac{c_r^2 - d_r^2 - 1}{f_r^2} - 2\frac{d_r}{f_r} \frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} - \mu^\top \Sigma^{-1} \mu + c_r^2 \left(\mu^\top \Sigma^{-1} \mu - \frac{(\mu^\top x)^2}{x^\top \Sigma x} \right) \geq 0. \tag{5.54}$$

Defining g_r and h_r, after some algebraic manipulation:

$$g_r := \frac{c_r^2 - d_r^2 - 1}{f_r^2} = \frac{r(\pi_a^2 h_a^2 + \pi_b^2 h_b^2) - (\pi_a h_a + \pi_b h_b)^2 + \pi_a h_a (2 - h_a^2) + \pi_b h_b (2 - h_b^2)}{\pi_a h_a + \pi_b h_b - (r-1)(\pi_a^2 + \pi_b^2) + 2\pi_a \pi_b}, \tag{5.55}$$

$$h_r := \frac{d_r}{f_r} = \frac{(\pi_a h_a(-(r-1)\pi_a + h_a + \pi_b) + \pi_b h_b((r-1)\pi_b - h_b - \pi_a) - \pi_a + \pi_b)}{\pi_a h_a + \pi_b h_b - (r-1)(\pi_a^2 + \pi_b^2) + 2\pi_a \pi_b}. \tag{5.56}$$

Inequality (5.54) is rewritten as follows:

$$\mu^\top \Sigma^{-1} \mu \leq g_r - 2h_r \frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} + c_r^2 \left(\mu^\top \Sigma^{-1} \mu - \frac{(\mu^\top x)^2}{x^\top \Sigma x} \right), \tag{5.57}$$

where we reach condition C in (5.22).

□

5.3.2 Sufficient condition

We also achieve a milder condition.

Lemma 5.3.1 *Let $\Phi : \mathbb{R} \to [0,1]$ be the one dimensional Gaussian CDF. For $n \geq 1$, let $b \in \mathbb{R}$, $\mu \in \mathbb{R}^n$ and positive definite $n \times n$ matrix Σ be given. Considering $b - a > 0$ for any x such that $\mu^\top x \in [a,b]$, and for any $r < 0$ the mapping $g(x) = \mathbb{P}\{a \leq \tilde{\xi}^\top x \leq b\} = \Phi(\frac{b-\mu^\top x}{\sqrt{x^\top \Sigma x}}) - \Phi(\frac{a-\mu^\top x}{\sqrt{x^\top \Sigma x}})$ is locally r-concave around x, if the following conditions hold simultaneously:*

$$(C1) \quad \mu^\top \Sigma^{-1} \mu \leq \theta_a \left(\delta_r(\theta_a) - 1\right) - 2\theta_a^{\frac{1}{2}} \delta_r(\theta_a) \left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right) + \left(\delta_r(\theta_a) + 2\right) \left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2, \quad (5.58)$$

$$(C2) \quad \mu^\top \Sigma^{-1} \mu \leq \theta_b \left(\delta_r(\theta_b) - 1\right) + 2\theta_b^{\frac{1}{2}} \delta_r(\theta_b) \left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right) + \left(\delta_r(\theta_b) + 2\right) \left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2, \quad (5.59)$$

$$(C3) \quad \mu^\top \Sigma^{-1} \mu \leq \left(\frac{\mu^\top x}{\sqrt{x \Sigma x}}\right)^2. \quad (5.60)$$

where $\theta_a = h_a^2(x) = \left(\frac{\mu^\top x - a}{\sqrt{x^\top \Sigma x}}\right)^2$, $\theta_b = h_b^2(x) = \left(\frac{b - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2$, $\delta_r(\theta_a) = \theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a) - 1$ and $s_r(\theta_a) = (r-1)\frac{e^{-\frac{1}{2}\theta_a}}{\sqrt{2\pi} g(x)}$. For $\delta_r(\theta_b)$ and $s_r(\theta_b)$ we proceed in a equivalent way.

Proof.

We define $\nabla^2 f_r(x) = \nabla^2 f_{r_a}(x) + \nabla^2 f_{r_b}(x) := H_a + H_b + H_{ab}$. Where:

$$H_a(x) = r g^{r-1}(x) \frac{e^{-\frac{h_a^2(x)}{2}}}{\sqrt{2\pi}} \left(\frac{(r-1) e^{-\frac{h_a^2(x)}{2}}}{g(x)\sqrt{2\pi}} \nabla h_a(x) \nabla h_a^\top(x) - h_a(x) \nabla h_a(x) \nabla h_a^\top(x) + \nabla^2 h_a(x) \right), \quad (5.61)$$

$$H_b(x) = r g^{r-1}(x) \frac{e^{-\frac{h_b^2(x)}{2}}}{\sqrt{2\pi}} \left(\frac{(r-1) e^{-\frac{h_b^2(x)}{2}}}{g(x)\sqrt{2\pi}} \nabla h_b(x) \nabla h_b^\top(x) - h_b(x) \nabla h_b(x) \nabla h_b^\top(x) + \nabla^2 h_b(x) \right), \quad (5.62)$$

$$H_{ab}(x) = r g^{r-1}(x)(r-1) \frac{e^{-\frac{h_a^2(x)}{2}}}{g(x)\sqrt{2\pi}} \frac{e^{-\frac{h_b^2(x)}{2}}}{\sqrt{2\pi}} \left(\nabla h_b(x) \nabla h_a^\top(x) + \nabla h_a(x) \nabla h_b^\top(x) \right). \quad (5.63)$$

The idea of the proof is to verify for which case $\nabla^2 f_r(x)$ is positive semidefinite for $r < 0$. We will analyse this with a sufficient condition, such that each of the matrices $H_a, H_b, H_{ab} \succeq 0$. This is done splitting the analysis in the individual terms:

- Term $H_a(x)$:

Working with $H_a(x)$, we know that its positive definiteness depends on the negative definiteness of:

$$\hat{H}_a(x) = \left(\frac{(r-1)e^{-\frac{h_a^2(x)}{2}}}{g(x)\sqrt{2\pi}} \nabla h_a(x) \nabla h_a^\top(x) - h_a(x) \nabla h_a(x) \nabla h_a^\top(x) + \nabla^2 h_a(x) \right). \tag{5.64}$$

We make the definitions $s_r(\theta_a) = (r-1)\frac{e^{-\frac{1}{2}\theta_a}}{\sqrt{2\pi}g(x)}$, $h_a = \frac{\mu^\top x - a}{\sqrt{x^\top \Sigma x}}$ and $\theta_a = h_a^2(x)$. And opening $\hat{H}_a(x)$ on their primitives, we have:

$$\begin{aligned}\hat{H}_a(x) = \frac{(s_r(\theta_a) - h_a)}{x^\top \Sigma x} \left[\mu\mu^\top - \frac{(\mu^\top x - a)}{x^\top \Sigma x} [\mu x^\top \Sigma + \Sigma x \mu^\top] + \left(\frac{(\mu^\top x - a)}{x^\top \Sigma x} \right)^2 \Sigma x x^\top \Sigma \right] \\ + \frac{1}{(x^\top \Sigma x)^{\frac{3}{2}}} \left[-\Sigma x \mu^\top - \mu(\Sigma x)^\top - (\mu^\top x - a)\Sigma + 3\frac{(\mu^\top x - a)}{(x^\top \Sigma x)} \Sigma x x^\top \Sigma \right].\end{aligned} \tag{5.65}$$

We use $\hat{H}_a(x)$ to define the auxiliary matrix $\bar{H}_a(x)$, that has equivalent definiteness than $\hat{H}_a(x)$. Therefore, $\bar{H}_a(x) = \frac{x^\top \Sigma x}{h_a(x)} \Sigma^{-\frac{1}{2}} \hat{H}_a(x) \Sigma^{-\frac{1}{2}}$ and we establish:

$$\begin{aligned}\bar{H}_a(x) = (\frac{s_r(\theta_a)}{h_a} - 1) \left[\Sigma^{-\frac{1}{2}} \mu\mu^\top \Sigma^{-\frac{1}{2}} - \frac{h_a^2}{\mu^\top x - a} [\Sigma^{-\frac{1}{2}} \mu x^\top \Sigma^{\frac{1}{2}} + \Sigma^{\frac{1}{2}} x \mu^\top \Sigma^{-\frac{1}{2}}] + \left(\frac{(\mu^\top x - a)}{x^\top \Sigma x} \right)^2 \Sigma^{\frac{1}{2}} x x^\top \Sigma^{\frac{1}{2}} \right] \\ + \left[-\frac{\Sigma^{\frac{1}{2}} x \mu^\top \Sigma^{-\frac{1}{2}}}{\mu^\top x - a} - \frac{\Sigma^{-\frac{1}{2}} \mu x^\top \Sigma^{\frac{1}{2}}}{\mu^\top x - a} - I + \frac{3\Sigma^{\frac{1}{2}} x x^\top \Sigma^{\frac{1}{2}}}{(x^\top \Sigma x)} \right],\end{aligned} \tag{5.66}$$

$$\begin{aligned}\bar{H}_a(x) = (\frac{s_r(\theta_a)}{h_a} - 1)\Sigma^{-\frac{1}{2}} \mu^\top \mu \Sigma^{-\frac{1}{2}} + (\frac{-s_r(\theta_a)h_a + h_a^2 - 1}{(\mu^\top x - a)})[\Sigma^{-\frac{1}{2}} \mu x^\top \Sigma^{\frac{1}{2}} + \Sigma^{\frac{1}{2}} x \mu^\top \Sigma^{-\frac{1}{2}}] \\ + \frac{h_a s_r(\theta_a) - h_a^2(\theta) + 3}{x^\top \Sigma x} (\Sigma^{\frac{1}{2}} x x^\top \Sigma^{\frac{1}{2}}) - I.\end{aligned} \tag{5.67}$$

Making the substitutions $V := \Sigma^{\frac{1}{2}} x$, $W := \Sigma^{-\frac{1}{2}} \mu$, $\|V\|^2 := x^\top \Sigma x$, we have:

$$\bar{H}_a(x) = (\frac{s_r(\theta_a)}{h_a} - 1) WW^\top + (\frac{-s_r(\theta_a)h_a + h_a^2 - 1}{(\mu^\top x - a)})[WV^\top + VW^\top] + \frac{h_a s_r(\theta_a) - h_a^2(\theta) + 3}{\|V\|^2} (VV^\top) - I. \tag{5.68}$$

Defining $Z_a = \alpha(\theta_a)\frac{V}{\|V\|} + \gamma(\theta_a)W$ and $Y_a = \beta(\theta_a)\frac{V}{\|V\|}$, we establish the following

coefficients:

$$\gamma(\theta_a) = \sqrt{1 - \frac{s_r(\theta_a)}{h_a}}, \tag{5.69}$$

$$-\alpha(\theta_a)\sqrt{\frac{h_a - s_r(\theta_a)}{h_a}} = \frac{(-s_r(\theta_a)h_a + h_a^2 - 1)}{h_a\|V\|} \Leftrightarrow \alpha(\theta_a) = \frac{s_r(\theta_a)h_a - h_a^2 + 1}{\sqrt{h_a^2 - h_a s_r(\theta_a)}}, \tag{5.70}$$

$$\beta(\theta_a) = \sqrt{h_a s_r(\theta_a) - h_a^2 + 3 + \alpha^2(\theta_a)} = \sqrt{\frac{h_a^2 - h_a s_r(\theta_a) + 1}{h_a^2 - h_a s_r(\theta_a)}}. \tag{5.71}$$

And so we rewrite (5.68) in terms of Z_a and Y_a, being $\bar{H}_a = -Z_a Z_a^\top + Y_a Y_a^\top - I$:

$$\begin{aligned}
\bar{H}_a(x) &= -\left(\frac{\gamma^2(\theta_a)}{\|V\|^2}VV^\top - \alpha(\theta_a)\gamma(\theta_a)[VW^\top + WV^\top] + \gamma^2(\theta_a)WW^\top\right) + \frac{\beta^2(\theta_a)}{\|V\|^2}VV^\top - I, \\
\bar{H}_a(x) &= -\gamma^2(\theta_a)WW^\top - \alpha(\theta_a)\gamma(\theta_a)[VW^\top + WV^\top] + \frac{\beta^2(\theta_a) - \alpha^2(\theta_a)}{\|V\|^2}VV^\top - I.
\end{aligned} \tag{5.72}$$

Comparing (5.68) and (5.72) it is clear that the coefficients (5.69)-(5.71) hold. From the results from Appendix C.1, we observe that the negative semi-definiteness of (5.72) is guaranteed if the highest eigenvalue obey:

$$\|Y_a\|^2 - \|Z_a\|^2 + \|Y_a\|^2 \times \|Z_a\|^2 - (Z_a^\top Y_a)^2 \leq 1. \tag{5.73}$$

It is clear that:

$$\begin{aligned}
\|Y_a\|^2 &= \beta^2(\theta_a), \\
\|Z_a\|^2 &= \alpha^2(\theta_a) + 2\gamma(\theta_a)\alpha(\theta_a)\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} + \gamma^2(\theta_a)\mu^\top \Sigma^{-1}\mu, \\
Z_a^\top Y_a &= \alpha(\theta_a)\beta(\theta_a) + \frac{\gamma(\theta_a)\beta(\theta_a)W^\top V}{\sqrt{x^\top \Sigma x}}.
\end{aligned}$$

After algebraic manipulations, we get:

$$\beta^2(\theta_a) - \alpha^2(\theta_a) + \mu^\top \Sigma^{-1}\mu\gamma^2(\theta_a)(\beta^2(\theta_a) - 1) - 2\gamma(\theta_a)\alpha(\theta_a)\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} + \\ -\frac{\gamma^2(\theta_a)\beta^2(\theta_a)(\mu^\top x)^2}{x^\top \Sigma x} \leq 1.$$

Opening in their primitives from (5.69)-(5.71), we have:

$$h_a s_r(\theta_a) - h_a^2 + 3 + \alpha^2(\theta_a) - \alpha^2(\theta_a) + \mu^\top \Sigma^{-1}\mu \left(1 - \frac{s_r(\theta_a)}{h_a}\right)(\beta^2(\theta_a) - 1) + \\ +2\sqrt{1 - \frac{s_r(\theta_a)}{h_a}}\frac{s_r(\theta_a h_a - h_a^2 + 1)}{\sqrt{h_a^2 - h_a s_r(\theta_a)}}\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} - \left(1 - \frac{s_r(\theta_a)}{h_a}\right)\beta^2(\theta_a)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right) \leq 1. \quad (5.74)$$

Making the substitution $h_a(x) = \sqrt{\theta_a}$, and after some algebraic manipulation then:

$$\theta_a^{\frac{1}{2}} s_r(\theta_a) - \theta_a + 3 + \mu^\top \Sigma^{-1}\mu \left(\frac{\theta_a^{\frac{1}{2}} - s_r(\theta_a)}{\theta_a^{\frac{1}{2}}}\right)(\beta^2(\theta_a) - 1) - \\ +2\left(\frac{s_r(\theta_a)\theta_a^{\frac{1}{2}} - \theta_a + 1}{\theta_a^{\frac{1}{2}}}\right)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right) - \left(\frac{\theta_a^{\frac{1}{2}} - s_r(\theta_a)}{\theta_a^{\frac{1}{2}}}\right)\beta^2(\theta_a)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2 \leq 1. \quad (5.75)$$

Multiplying both sides by θ_a:

$$\theta_a^{\frac{3}{2}} s_r(\theta_a) - \theta_a^2 + 3\theta_a + \mu^\top \Sigma^{-1}\mu \left(\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a)\right)(\beta^2(\theta_a) - 1) + \\ -2\left(\theta_a s_r(\theta_a) - \theta_a^{\frac{1}{2}} + \theta_a^{-\frac{1}{2}}\right)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right) - \left(\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a)\right)\beta^2(\theta_a)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2 \leq \theta_a. \quad (5.76)$$

Making the substitutions from $\beta(\theta_a)$ as in (5.71) we have the following equivalences:

$$\left(\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a)\right)\left(\frac{\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta) + 1}{\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a)} - 1\right) = 1,$$

$$\left(\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a)\right)\frac{\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta) + 1}{\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a)} = \theta_a - \theta_a^{\frac{1}{2}} s_r(\theta) + 1.$$

Which simplifies (5.76) in the following sequence:

$$\begin{aligned}\theta_a^{\frac{3}{2}} s_r(\theta_a) - \theta_a^2 + 3\theta_a + \mu^\top \Sigma^{-1}\mu - 2\left(\theta_a s_r(\theta_a) - \theta_a^{\frac{1}{2}} + \theta_a^{-\frac{1}{2}}\right)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right) - \\ + \left(\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a) + 1\right)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2 \leq \theta_a,\\ -\theta_a\left(\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a) - 2\right) + \mu^\top \Sigma^{-1}\mu + 2\theta_a^{\frac{1}{2}}\left(\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a) - 1\right)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right) - \\ + \left(\theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a) + 1\right)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2 \leq 0.\end{aligned}$$

Finally, defining $\delta_r(\theta_a) = \theta_a - \theta_a^{\frac{1}{2}} s_r(\theta_a) - 1$:

$$\mu^\top \Sigma^{-1}\mu \leq \theta_a\left(\delta_r(\theta_a) - 1\right) - 2\theta_a^{\frac{1}{2}}\delta_r(\theta_a)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right) + \left(\delta_r(\theta_a) + 2\right)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2. \tag{5.77}$$

from last inequality we reach condition $C1$ (5.58).

- Term $H_b(x)$:

The negative semidefinitess of $H_b(x)$ is verified in Lemma 1 (van ACKOOIJ, 2017) in similar procedures as for $H_a(x)$, if the following inequality is verified:

$$\mu^\top \Sigma^{-1}\mu \leq \theta_b\left(\delta_r(\theta_b) - 1\right) + 2\theta_b^{\frac{1}{2}}\delta_r(\theta_b)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right) + \left(\delta_r(\theta_b) + 2\right)\left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2, \tag{5.78}$$

where $\theta_b = h_b^2(x) = \left(\frac{b - \mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2$, $\delta_r(\theta_b) = \theta_b - \theta_b^{\frac{1}{2}} s_r(\theta_b) - 1$ and $s_r(\theta_b) = (r-1)\frac{e^{-\frac{1}{2}\theta_b}}{\sqrt{2\pi} g(x)}$.

From last inequality we reach condition $C2$ (5.59).

- Term $H_{ab}(x)$:

Working with $H_{ab}(x)$, we define $\hat{H}_{ab}(x) = \nabla h_b(x)\nabla h_a^\top(x) + \nabla h_a(x)\nabla h_b^\top(x)$, and noting that for $r < 0$ the term $r(r-1)g^{r-1}(x)\frac{e^{-\frac{h_a^2(x)}{2}}}{g(x)\sqrt{2\pi}} \cdot \frac{e^{-\frac{h_b^2(x)}{2}}}{\sqrt{2\pi}}$ is always positive. For $r \in [0,1]$, the term $r(r-1)$ is nonpositive and so in order to obtain $H_{ab}(x)$ negative semi-definite we must have $\hat{H}_{ab}(x)$ positive definite. Hence, the objective here is to verify in which cases $\hat{H}_{ab}(x)$ is indeed positive definite. Opening $\hat{H}_{ab}(x)$ in its primitives we obtain:

$$\hat{H}_{ab}(x) = \frac{1}{x^\top \Sigma x}\left(-2\mu\mu^\top + \frac{(\mu^\top x - a) - (b - \mu\top x)}{x^\top \Sigma x}[\mu(\Sigma x)^\top + \Sigma x \mu^\top] + \frac{2(\mu^\top x - a)(b - \mu^\top x)}{(x^\top \Sigma x)^2}\Sigma x^\top x \Sigma\right). \tag{5.79}$$

If we make $\bar{H}_{ab} = x^\top \Sigma x \cdot \Sigma^{-\frac{1}{2}}\hat{H}_{ab}\Sigma^{-\frac{1}{2}}$, defining $V := \Sigma^{\frac{1}{2}}x$ and $W := \Sigma^{-\frac{1}{2}}\mu$ then we have:

$$\bar{H}_{ab} = -2WW^\top + \frac{(\mu^\top x - a) - (b - \mu^\top x)]}{x^\top \Sigma x}[WV^\top + VW^\top] + \frac{2(\mu^\top x - a)(b - \mu^\top x)}{(x^\top \Sigma x)^2}VV^\top. \tag{5.80}$$

Considering the two vectors:

$$Z_{ab} = z_1 \cdot \frac{V}{\|V\|} + z_2 \cdot W \ , \quad Y_{ab} = z_3 \cdot \frac{V}{\|V\|}. \tag{5.81}$$

where, using the following notation $\alpha = \mu^\top x - a$ and $\beta = b - \mu^\top x$ we establish:

$$\bar{H}_{ab} = -2WW^\top + \frac{(\alpha - \beta)}{x^\top \Sigma x}[WV^\top + VW^\top] + \frac{2\alpha\beta}{(x^\top \Sigma x)^2}VV^\top. \tag{5.82}$$

Considering $\bar{H}_{ab} = -Z_{ab}Z_{ab}^\top + Y_{ab}Y_{ab}^\top$, after some algebraic manipulation we have:

$$\bar{H}_{ab} = -z_2 WW^\top + \frac{(-z_1 z_2)}{\|V\|}[WV^\top + VW^\top] + (z_3^2 - z_1^2)\frac{VV^\top}{(\|V\|)^2}, \tag{5.83}$$

using the vectors in (5.81), we use the notation $z_1 = -\frac{(\alpha-\beta)}{2}$, $z_2 = \sqrt{2}$ and $z_3 = \sqrt{2\alpha\beta + \frac{(\alpha-\beta)^2}{2}}$.

In order to establish the conditions for $\bar{H}_{ab}$ being positive semi-definite, from the Appendix C.1 we know that the eigenvalue must observe the following condition:

$$\lambda^- = \frac{\|Y_{ab}\|^2 - \|Z_{ab}\|^2}{2} - \frac{\sqrt{(\|Y_{ab}\|^2 + \|Z_{ab}\|^2)^2 - 4(Z_{ab}^\top Y_{ab})^2}}{2} \geq 0,$$

that after algebraic manipulations, reduces to the following inequality:

$$\|Y_{ab}\|^2\|Z_{ab}\|^2 - (Z_{ab}^\top Y_{ab})^2 \leq 0. \tag{5.84}$$

Determining the factors, we get:

$$\|Y_{ab}\|^2 = z_3^2 \tag{5.85}$$

$$\|Z_{ab}\|^2 = z_1^2 + 2z_1z_2(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}) + z_2^2\mu^\top\Sigma^{-1}\mu \tag{5.86}$$

$$Z_{ab}^\top Y_{ab} = z_1z_3 + z_2z_3\frac{W^\top V}{\sqrt{x^\top \Sigma x}}. \tag{5.87}$$

Its clear from (5.84) and simplifying:

$$\begin{aligned}
&z_3^2\left(z_1^2 + 2z_1z_2(\frac{\mu^\top x}{\sqrt{x^\top\Sigma x}}) + z_2^2\mu^\top\Sigma^{-1}\mu\right) - \left(z_1z_3 + z_2z_3\frac{W^\top V}{\sqrt{x^\top\Sigma x}}\right)\left(z_1z_3 + z_2z_3\frac{W^\top V}{\sqrt{x^\top\Sigma x}}\right) \leq 0,\\
&z_3^2z_1^2 + 2z_1z_2z_3^2(\frac{\mu^\top x}{\sqrt{x^\top\Sigma x}}) + z_2^2z_3^2\mu^\top\Sigma^{-1}\mu - z_1^2z_3^2 - 2z_1z_2z_3^2(\frac{\mu^\top x}{\sqrt{x^\top\Sigma x}}) - z_2^2z_3^2\left(\frac{\mu^\top x}{\sqrt{x^\top\Sigma x}}\right)^2 \leq 0,\\
&\mu^\top\Sigma^{-1}\mu \leq \left(\frac{\mu^\top x}{\sqrt{x^\top\Sigma x}}\right)^2.
\end{aligned} \tag{5.88}$$

With the last inequality and the above developments of the terms $H_a(x)$ and $H_b(x)$, we establish the conditions (5.58)-(5.60). □

5.4 IMPLIED CONDITIONS

This section aims at developing conditions that turn possible the application of the statements of Section 5.3.2. The final objective is to define thresholds, $\hat{\theta}_a$ and $\hat{\theta}_b$ that entails the sufficient conditions (5.58)-(5.60). In view of this, we propose the following implied condition:

Proposition 5.4.1 *If it holds:*

$$(D1) \qquad \delta_r(\theta_a) \geq 2, \quad \delta_r(\theta_b) \geq 2, \quad \mu^\top \Sigma^{-1} \mu \leq \left(\frac{\mu^\top x + a}{\sqrt{x^\top \Sigma x}} \right)^2,$$
$$\mu^\top \Sigma^{-1} \mu \leq \left(\frac{b + \mu^\top x}{\sqrt{x^\top \Sigma x}} \right)^2 \text{ and } \mu^\top \Sigma^{-1} \mu \leq \left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} \right)^2, \tag{5.89}$$

then conditions established in (5.58)- (5.60) *from Lemma 5.3.1 are ensured.*

Proof. From (van ACKOOIJ, 2017) we know that after algebraic manipulations in (5.59) we verify that $\delta_r(\theta_b) \geq 2$ and $\mu^\top \Sigma^{-1} \mu \leq \left(\frac{b+\mu^\top x}{\sqrt{x^\top \Sigma x}} \right)^2$ implies (5.59). Considering $\theta_a = \left(\frac{\mu^\top x - a}{\sqrt{x^\top \Sigma x}} \right)^2$, we replace $\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} = \sqrt{\theta_a} + \frac{a}{\sqrt{x^\top \Sigma x}}$ in (5.58) and after manipulations we establish that:

$$\mu^\top \Sigma^{-1} \mu \leq \left(\frac{2a}{\sqrt{x^\top \Sigma x}} + \sqrt{\theta_a} \right)^2 + (\theta_a - \sqrt{\theta_a} s_r(\theta_a) - 3) \frac{a^2}{x^\top \Sigma x}. \tag{5.90}$$

In order to ensure the positiveness of the second term in (5.90), we hold that $\theta - \sqrt{\theta_a} s_r(\theta_a) - 3 = \delta_r(\theta_a) - 2 \geq 0$, then $\delta_r(\theta_a) \geq 2$. Keeping the last inequality, from the first term in (5.90) we establish that $\mu^\top \Sigma^{-1} \mu \leq \left(\frac{a+\mu^\top x}{\sqrt{x^\top \Sigma x}} \right)^2$ ensuring that condition (5.58) holds. □ Intending to use the results of Proposition 5.4.1 in programming problems, such as in Section 6.1, we define the thresholds $\hat{\theta}_a$ and $\hat{\theta}_b$ on which the first to conditions of (5.89) hold, i.e. if $\hat{\theta}_a \leq \theta_a$ and $\hat{\theta}_b \leq \theta_b$ then $\delta_r(\theta_a) \geq 2, \delta_r(\theta_b) \geq 2$ are ensured. In this view, it is useful to define:

Definition 5.4.1 *From the definitions of Lemma 5.3.1, we define the following auxiliary*

functions:

$$s_{r,p}(t) = (r-1)\frac{e^{-\frac{1}{2}t}}{\sqrt{2\pi}p}, \tag{5.91}$$

$$\delta_{r,p}(t) = t - t^{\frac{1}{2}}s_{r,p}(t) - 1, \tag{5.92}$$

where t takes place of θ_a and θ_b indiscriminately.

Taking the definitions of $\delta_r(\theta_a)$ and $s_r(\theta_a)$ (and for equivalence the θ_b case) from Lemma 5.3.1 we observe that $\delta_r(\theta_a)$ is not only dependent of parameter a but also b by the CDF $g(x) := \Phi(h_b(x)) + \Phi(h_a(x)) - 1$. In applications of PC problems, it is commonly requested that $1 \geq g(x) \geq p$, where p is a minimal probability limitation. Regarding the function $\delta_r(\theta_a)$ (and θ_b) and its dependence on both parameters a and b, we dispose the following proposition:

Lemma 5.4.1 *Giving the definitions from Lemma 5.3.1 and Definition 5.4.1 we establish the limits over the function $\delta_r(t)$ as $\delta_{r,p}(t) \geq \delta_r(t) \geq \delta_{r,1}(t)$, where p is a minimum probability threshold.*

Proof. Since $r \leq 1$ and $t \in \mathbb{R}^+$, it is easily verified that $\delta_{r,pr}(t) \geq \delta_r(t) \geq \delta_{r,1}(t)$ by the dominance $s_{r,1}(t) \geq s_r(t) \geq s_{r,pr}(t)$ and the positiveness of the term $-t^{\frac{1}{2}}s_{r,p}(t)$ in (5.92). □

The next results derive properties for the inequality $\delta_{r,p}(t) \geq 2$, especially in terms of defining the roots of the equality $\delta_{r,p}(t) = 2$.

Lemma 5.4.2 *The function $\delta_{r,p}(t)$, for $t \in \mathbb{R}^+$, has a single inflection point. It occurs at $t = 1 + \sqrt{2}$ and it is independent from r and p. For fixed p and r, in the parametrization over t, we distinguish the two cases:*

1. *if $\delta'_{r,p}(1+\sqrt{2}) \geq 0$ or $r = 1$: Then $\delta_{r,p}(t) = 2$ occurs just once.*

2. *if $\delta'_{r,p}(1+\sqrt{2}) < 0$: Then $\delta_{r,p}(t) = 2$ occurs at most 3 times.*

For a fixed p, for $r \geq r_{ch}(p)$ we are in case (1) and for $r < r_{ch}(p)$ we are in case (2), where $r_{ch}(p) := -\frac{\sqrt{1+\sqrt{2}}}{\sqrt{2}} 2\sqrt{2\pi} \cdot p \cdot e^{\frac{(1+\sqrt{2})}{2}} + 1$. Trivially, we also have the change of pattern for a fixed r, which changes for probability $p_{ch}(r) = \frac{(1-r)\sqrt{2}}{\sqrt{(1+\sqrt{2})2\pi} \cdot 2e^{\frac{(1+\sqrt{2})}{2}}}$.

Proof. We check by the differentiation of the function $\delta_{r,p}(t)$. Indeed, we observe by the first and second derivatives:

$$\delta'_{r,p}(t) = 1 + \frac{(r-1)e^{-\frac{1}{2}t}}{2\sqrt{2\pi}p} \left(\frac{t-1}{\sqrt{t}} \right), \tag{5.93}$$

$$\delta''_{r,p}(t) = \frac{(r-1)e^{-\frac{1}{2}t}}{2\sqrt{2\pi}p} \left(-\frac{1}{2}t^{\frac{1}{2}} + t^{-\frac{1}{2}} + \frac{1}{2}t^{-\frac{3}{2}} \right). \tag{5.94}$$

On the search for the roots of (5.94), $\delta''_{r,p}(t) = 0$, for $r < 1$ one observes that $t = 1+\sqrt{2}$ is the only root for $t \in \mathbb{R}^+$, being the single extreme point of $\delta'_{r,p}(t)$. For $r = 1$, $\delta''_{r,p}(t) = 0 \ \forall t$. Considering the nontrivial case, i.e. $p > 0$ and $r > 1$, since $\lim_{t\to 0} \delta'_{r,p}(t) = \infty$ and $\lim_{t\to\infty} \delta'_{r,p}(t) = 1$ (and also $\delta'''_{r,p}(t) > 0$) the extreme point $\delta'_{r,p}(1+\sqrt{2})$ is minima, and related to the function $\delta_{r,p}(t)$, depending on r and p, we face two possibilities:

1. $\delta'_{r,p}(1+\sqrt{2}) \geq 0$ or $r = 1$: Hence $\delta'_{r,p}(t) \geq 0 \ \forall t$ and $\delta_{r,p}(t)$ is a non-decreasing function with at most one extreme point (at $t = 1+\sqrt{2}$). Since $\lim_{t\to 0} \delta_{r,p}(t) < 2$ and $\lim_{t\to\infty} \delta_{r,p}(t) > 2$ so it crosses $\delta_{r,p}(t) = 2$ just once.

2. $\delta'_{r,p}(1+\sqrt{2}) < 0$: Hence $\delta_{r,p}(t)$ has two extreme points. We can define the two extreme points of $\delta_{r,p}(t)$ as $t_d < 1+\sqrt{2} < t_u$, where t_d establishes a local maximum and t_u a local minimum. If $\delta_{r,p}(t_d) < 2$ or $\delta_{r,p}(t_u) > 2$ than $\delta_{r,p}(t) = 2$ occurs just once. If $\delta_{r,p}(t_d) = 2$ or $\delta_{r,p}(t_u) = 2$ than $\delta_{r,p}(t) = 2$ occurs twice. If $\delta_{r,p}(t_d) > 2$ and $\delta_{r,p}(t_u) < 2$ than $\delta_{r,p}(t) = 2$ occurs three times.

In the second case, the assertion of local maximum and minimum is easily proved if we study the function $\delta'_{r,p}(t)$. Since $\lim_{t\to 0} \delta'_{r,p}(t) = \infty$ and $\delta'_{r,p}(1+\sqrt{2}) < 0$ on the extreme point t_d the derivative of $\delta_{r,p}(t)$ changes its signal from positive to a negative value, defining a local maximum. An equivalent thought is used to qualify t_u as a local minimum.

The two last assertions of the Lemma are proved using (5.93), when we establish $\delta'_{r,p}(1+\sqrt{2}) = 0$. We easily check that for $r(p) = -\frac{\sqrt{1+\sqrt{2}}}{\sqrt{2}} 2\sqrt{2\pi} \cdot p \cdot e^{\frac{1}{2}(1+\sqrt{2})} + 1$ the function $\delta_{r,p}(t)$ changes from pattern 1 to pattern 2 in the two possibilities. The last assertion of the lemma is easily checked isolating p instead of r. □

Lemma 5.4.2 is useful on the track of finding the limits $\theta_a \geq \hat{\theta}_a$ and $\theta_b \geq \hat{\theta}_b$ on which $\delta_r(\theta_a) \geq 2$ and $\delta_r(\theta_b) \geq 2$ of Proposition 5.4.1 holds. From the functions of Lemma 5.4.1 we define:

Definition 5.4.2 *We define* $\hat{t}_{p,r} := \max\{t : \delta_{r,p}(t) = 2\}$. *In this sense we also define* $\hat{\theta}_{a,p,r} := \max\{\theta_a : \delta_{r,p}(\theta_a) = 2\}$ *and* $\hat{\theta}_{b,p,r} := \max\{\theta_b : \delta_{r,p}(\theta_b) = 2\}$ *for the singular cases.*

By the Definition 5.4.2, we remark:

Remark 5.4.1 *Using the properties of the functions* $\delta_{r,p}(t)$ *studied in Lemma 5.4.2 regarding* $\hat{t}_{p,r}$ *and the relationship* $\delta_{r,p}(t) \geq \delta_r(t) \geq \delta_{r,1}(t)$ *of Lemma 5.4.1 we develop the relationship* $\hat{\theta}_{a,p,r} \leq \hat{\theta}_a \leq \hat{\theta}_{a,1,r} \leq \theta_a$ *(and for equivalence the same for* $\hat{\theta}_b$*). Indeed all the proof is based on the fact that* $\delta_{r,p}(t)$ *for* $\hat{t}_{p,r} \leq t$ *is non-decreasing, therefore* $\delta_{r,p}(\theta_a) \geq 2$ *(and for equivalence* $\delta_{r,p}(\theta_b) \geq 2$*) is ensured.*

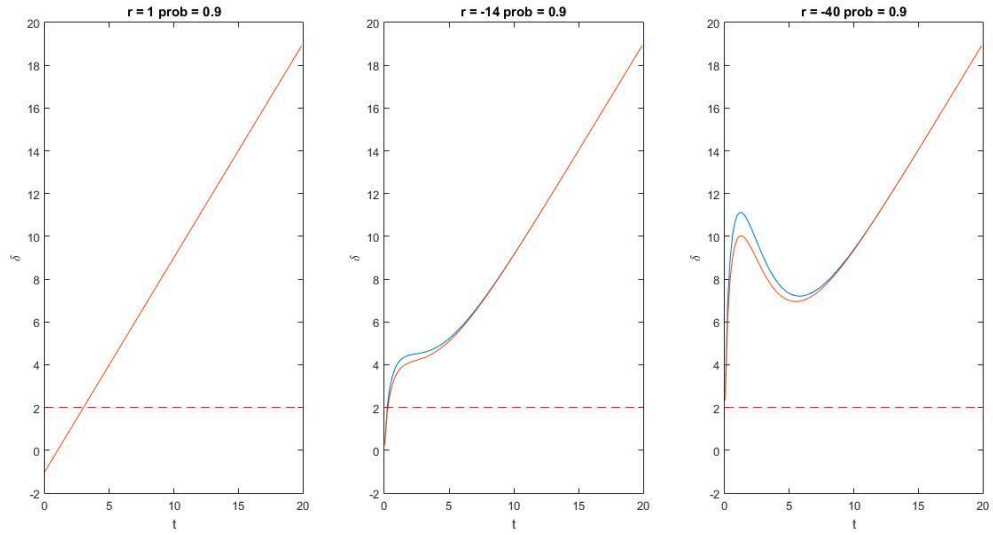
Aiming at deriving estimations for thresholds θ_a and θ_b in order to use, for instance, in programming problems, one defines the values $\hat{\theta}_{a,\bar{p},r}$ and $\hat{\theta}_{b,\bar{p},r}$, being more conservative, for $\bar{p} = 1$ or more flexible for $\bar{p} = p$, where p is the original probability threshold from the programming problem. We observe that only the case $\bar{p} = 1$ ensures a conservative feasibility in relation to the former thresholds $\hat{\theta}_a$ and $\hat{\theta}_b$. The procedure to define these limits are illustrated by Example 5.4.3.

Example 5.4.3 *To understand the results, we exemplify the mapping* $\mathbb{P}\{a \leq \tilde{\xi}^\top x \leq b\}$ *for* $\tilde{\xi}$ *defined as a multivariate normal distribution* $\mathcal{N}(\mu, \Sigma)$. *Considering the coefficients* $a = -15$, $b = 10$, μ *and* Σ *are displayed below:*

$$\mu = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 9 & 3 \\ 3 & 16 \end{bmatrix}.$$

Aiming at defining the feasible set of the mapping, we use Lemma 5.4.2 to drive a dichotomy procedure to define the maximum t such that the condition $\delta_{r,p}(t) = 2$ is met. Example of functions for different r parameters are defined in Figure 16. As stated in Lemma 5.4.2, we note that $r_{ch}(90\%) = -15.57$, implying that for $r \geq r_{ch}(90\%)$, $\delta_{r,p}(t)$ is a non-decreasing (cases $r = 1$ and $r = -14$ in Figure 16) and so by a dichotomy procedure on $\delta_{r,p}(t)$ function changing the t, the t for which $\delta_{r,p}(t) = 2$ is unique. Since for $r < r_{ch}(p)$ the function $\delta_{r,p}(t)$ has two extreme points, with a simple dichotomy procedure starting from $t = 0$ we cannot ensure that there is no greater t such that $\delta_{r,p}(t) = 2$. Nevertheless, as stated in the proof of Lemma 5.4.2, looking for the extreme point t_u (by a second dichotomy procedure, for example) we have all the ingredients to ensure that $\delta_{r,p}(t) = 2$ is unique or not, and in the negative case, we find the greatest $t \in \{t : t \geq t\_u\}$ such that $\delta_{r,p}(t) = 2$. Visually we observe that for $r = -40$ in Figure 16 we also have a single $\delta_{r,p}(t) = 2$.

Figure 16 – $\delta_{r,p}(t)$ behavior for $p = 90\%$ (in blue) and $p = 100\%$ (in red), changing the r parameter. The red dashed line refers to $\delta_{r,p}(t) = 2$.

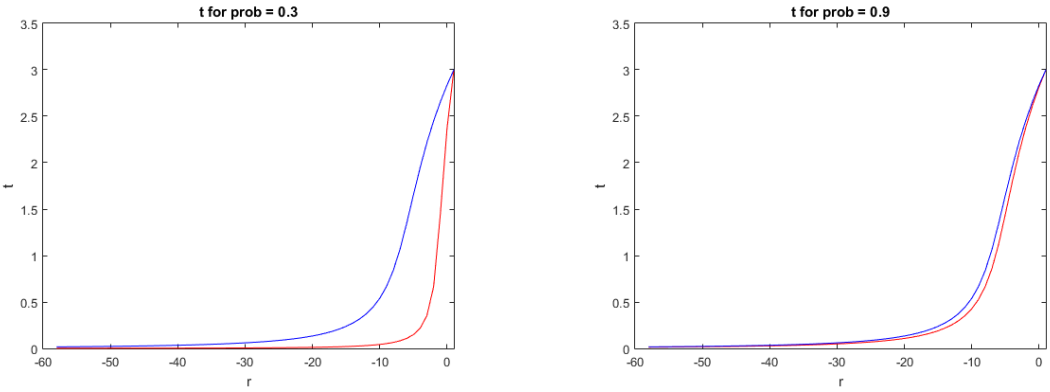


Source: Thesis results.

Checking Figure 16 we also observe that for a lower r, the t such that $\delta_{r,p}(t) = 2$ decreases (if just one root is ensured), implying in a broader feasible set. We present the relationship between the r parameter and t is for two probability levels in Figure 17.

We are now able to distinguish the set defined by the intersection of the two first

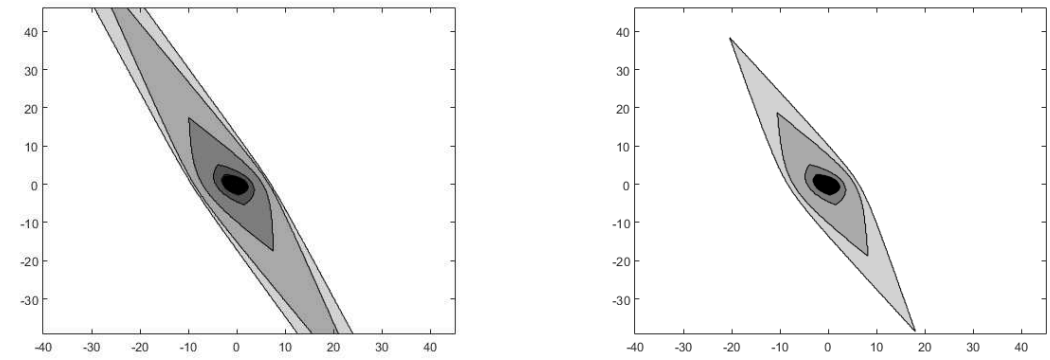
Figure 17 – Cumulative distribution function of the standard random variable $\tilde{\xi}$.



(a) p = 30% (in red) p = 100% (in blue). (b) p = 90% (in red) p = 100% (in blue).
Source: Thesis results.

conditions of Proposition 5.4.1, i.e., by $X = \{x : \theta_a \geq \hat{\theta}_a \quad and \quad \theta_b \geq \hat{\theta}_b\}$, *where* $\hat{\theta}_a$ *and* $\hat{\theta}_b$ *are the highest roots such that* $\delta_r(\theta_a) = 2$ *and* $\delta_r(\theta_b) = 2$. *In order to have approximations of* $\hat{\theta}_a$ *and* $\hat{\theta}_b$ *we make use of Remark 5.4.1, defining thresholds where we guarantee that the limits are sufficiently met. Graphically, we show the feasible sets for different* r *parameter levels in Figure 18. By these figures, considering that the inequalities* $\hat{\theta}_a \leq \hat{\theta}_{a,1,r} \leq \theta_a$ *and* $\hat{\theta}_b \leq \hat{\theta}_{b,1,r} \leq \theta_b$ *hold, we note that the initially presented Figure 15(b) defines the feasible set* $X = \{x : \theta_a \geq \hat{\theta}_{a,1,r} \quad and \quad \theta_b \geq \hat{\theta}_{b,1,r}\}$ *and as a result the first two inequalities* $\delta_r(\theta_a) \geq 2, \delta_r(\theta_b) \geq 2$ *of condition* $D1$ *of Proposition 5.4.1.*

Figure 18 – Feasible sets for different r parameters. The most inner (dark) area corresponds to $r = 1$ and the most outer (light) area corresponds to $r = -59$.



(a) p = 30% (b) p = 90%
Source: Thesis results.

In Figure 18(a) and 18(b) we note the feasible sets established by the thresholds defined by $X = \{x : \theta_a \geq \hat{\theta}_{a,p,r} \quad and \quad \theta_b \geq \hat{\theta}_{b,p,r}\}$ *for* $p = 30\%$ *and* 90%. *Comparing*

those figures with Figure 15(b) we note the feasible sets that are potentially missed, by the loss of generalization when contrasted with former thresholds $\hat{\theta}_a$ and $\hat{\theta}_b$. So, we note that the loss of information when using the mapping $\delta_{r,p}(t)$ instead of the former $\delta_r(t)$ for high levels of p tends to be marginal.

The next results aim at simplifying Proposition 5.4.1 in the view of applications, preparing its use in Algorithm 3 of Chapter 6.

Remark 5.4.2 *If $[a,b] \subset \mathbb{R}^+$ and $\mu^\top x \in [a,b]$, then condition* (5.89) *can be simplified by:*

$$(D2) \quad \delta_r(\theta_a) \geq 2, \quad \delta_r(\theta_b) \geq 2, \ \textbf{\textit{and}} \ \mu^\top \Sigma^{-1} \mu \leq \left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} \right)^2. \tag{5.95}$$

Proof. As $0 \leq a \leq b$, and from $\mu^\top \Sigma^{-1} \mu \leq \left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} \right)^2$, we observe:

$$\sqrt{\mu^\top \Sigma^{-1} \mu} \leq \frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} \leq \frac{\mu^\top x + a}{\sqrt{x^\top \Sigma x}} \leq \frac{\mu^\top x + b}{\sqrt{x^\top \Sigma x}}. \tag{5.96}$$

So, $\mu^\top \Sigma^{-1} \mu \leq \left(\frac{\mu^\top x}{\sqrt{x^\top \Sigma x}} \right)^2$ trivially implies the other inequalities. □

Proposition 5.4.2 *If $a \in \mathbb{R}$, $b \geq 0$ and it holds:*

$$(D3) \quad \frac{\theta_a \left(\delta_{r,p}(\theta_a) - 2 \right)}{\left(\delta_{r,p}(\theta_a) + 2 \right)} \geq \mu^\top \Sigma^{-1} \mu, \tag{5.97}$$

$$(D4) \quad \frac{\theta_b \left(\delta_{r,p}(\theta_b) - 2 \right)}{\left(\delta_{r,p}(\theta_b) + 2 \right)} \geq \mu^\top \Sigma^{-1} \mu. \tag{5.98}$$

then conditions established in (5.58)- (5.60) *from Lemma 5.3.1 are ensured.*

Proof. Condition (5.98) comes directly from the procedures applied in (van ACKOOIJ, 2017, Corollary 1)(MINOUX; ZORGATI, 2016, Corollary 1) for the function $\delta_{r,p}(t)$ as assigned in Definition 5.4.1 and ensures the condition (5.59). In the same line, finding the minimal value in terms of the quadratic function over a on the right side of inequality (5.90), we achieve (5.97), ensuring the former condition (5.58). Remembering that

$\theta_b = h_b^2(x) = \left(\frac{b-\mu^\top x}{\sqrt{x^\top \Sigma x}}\right)^2 = \left(\frac{\mu^\top x - b}{\sqrt{x^\top \Sigma x}}\right)^2$, and since $\delta_{r,p}(\theta_b) \geq -1$, raising both sides of (5.98) to the power of $\frac{1}{2}$ and replacing θ_b it follows:

$$\sqrt{\mu^\top \Sigma^{-1} \mu} \leq \frac{\mu^\top x - b}{\sqrt{x^\top \Sigma x}} \sqrt{\frac{\delta_{r,p}(\theta_b) - 2}{\delta_{r,p}(\theta_b) + 2}} \leq \frac{\mu^\top x}{\sqrt{x^\top \Sigma x}}, \tag{5.99}$$

where the last inequality comes from the fact that $\sqrt{\frac{\delta_{r,p}(\theta_b)-2}{\delta_{r,p}(\theta_b)+2}} \leq 1$ and $b \geq 0$. This last inequality implies the former condition (5.60). □

Defining $\vartheta_{r,p}(t) := \frac{t(\delta_{r,p}(t)-2)}{(\delta_{r,p}(t)+2)}$, where t is indistinctly defined $t = \theta_a$ or θ_b, then we remark:

Remark 5.4.3 *For a fixed p in the mapping $\vartheta_{r,p}(t)$, we can use the results established in Lemma 3 and Lemma 4 in (van ACKOOIJ, 2017) to ensure that exists a unique solution for the equation $\vartheta_{r,p}(t) = \gamma$, where $\gamma \geq 0$. Making use of a dichotomy procedure, then we can define the solution $\hat{t}_{p,r}$, considering $\gamma = \mu^\top \Sigma^{-1} \mu$. By the properties of the mapping $\delta_{r,p}(t)$ in Remark 5.4.1, we observe that for $p \leq p'$ we can ensure $\frac{(\delta_{r,p}(\theta_a)-2)}{(\delta_{r,p}(\theta_a)+2)} \geq \frac{\left(\delta_{r,p'}(\theta_a)-2\right)}{\left(\delta_{r,p'}(\theta_a)+2\right)}$. A direct implication for that is the relationship $\hat{\theta}_{a,p,r} \leq \hat{\theta}_a \leq \hat{\theta}_{a,1,r} \leq \theta_a$ (and for equivalence the same for $\hat{\theta}_b$), where p is a minimum probability threshold. Still making use of Lemma 4 in (van ACKOOIJ, 2017), being $\hat{t}_{p,r}$ a strictly decreasing sequence in r, the results are easily extended to prove that $\lim_{r \downarrow -\infty} \hat{t}_{p,r} = \mu^\top \Sigma^{-1} \mu$.*

We use the results related to Remark 5.4.3 in Section 6.1 in order to define a convex feasible set for the proposed problem. Indeed, working with the thresholds $\hat{\theta}_{a,p,r}$ and $\hat{\theta}_{b,p,r}$ in a way that conditions (5.97) and (5.98) hold, we have all the necessary ingredients to ensure that the conditions of Lemma 5.3.1 are met, and so the mapping $g(x) = \mathbb{P}\{a \leq \tilde{\xi}^\top x \leq b\}$ is locally r-concave around x. The next section proposes an algorithm that appropriately uses the achievements of the chapter.

6 TRUST REGION BASED ALGORITHM FOR BILINEAR PC PROBLEMS

In this chapter we aim at solving the former PC programming problem defined in (5.1)-(5.3), where we consider $g(x) = \mathbb{P}\{a \leq \tilde{\xi}^\top x \leq b\}$. Using $f_r(x) = g^r(x)$, for $r < 0$, raising both sides of (5.3) to the power of r, we define:

$$k_r(x) := f_r(x) - p^r \leq 0. \tag{6.1}$$

As discussed in Section 5.3, respecting conditions the function $g(x)$ is locally r-concave around x, being so the function $f_r(x)$ locally convex. It is easily seen that this property also implies the local convexity for $k_r(x)$. The idea is to work with the constraint $k_r(x) \leq 0$ in the programming problem, approximately reformulating it by a second-order Taylor's expansion in a barrier method (BOYD; VANDENBERGHE, 2006).We observe that:

$$\nabla k_r(x) = \nabla f_r(x) = r g^{r-1}(x) \left(\frac{e^{-\frac{1}{2}h_a^2(x)}}{2\pi} \nabla h_a(x) + \frac{e^{-\frac{1}{2}h_b^2(x)}}{2\pi} \nabla h_b(x) \right), \tag{6.2}$$

$$\nabla^2 k_r(x) = \nabla^2 f_{r_a}(x) + \nabla^2 f_{r_b}(x). \tag{6.3}$$

where $\nabla^2 f_{r_a}(x) + \nabla^2 f_{r_b}(x)$ are developed in (5.20)-(5.21):

We define the auxiliary functions to the barrier method:

$$\begin{aligned} \phi(x) &:= -\log(-k_r(x)), \quad \nabla\phi(x) := \frac{\nabla k_r(x)}{-k_r(x)}, \\ \nabla\phi^2(x) &:= \frac{\nabla k_r(x) \nabla k_r(x)^\top}{k_r^2(x)} + \frac{\nabla^2 k_r(x)}{-k_r(x)}. \end{aligned} \tag{6.4}$$

Considering the optimization problem (5.1)-(5.3), we know that the barrier method aims at solving the following problem:

$$\min_{\dot{x}} \quad t \cdot k_0(\dot{x}) + \phi(\dot{x}) \tag{6.5}$$

$$\text{s.t.} \quad A \cdot \dot{x} = b_{eq}, \tag{6.6}$$

where $t > 0$ is a parameter that changes iteratively and sets the accuracy of the approx-

imation. For each iteration of the barrier method, we solve iteratively problem (6.5) - (6.6) by the following Taylor approximation of the function $\phi(x)$ around $\hat{x}$, established as a QCQP problem:

$$\min_{\dot{v}} \quad t * k_0(\hat{x} + \dot{v}) + \phi(\hat{x}) + \nabla\phi(\hat{x})\top\dot{v} + \frac{1}{2}\dot{v}^\top\nabla^2\phi(\hat{x})\dot{v} \tag{6.7}$$

$$\text{s.t.} \quad A(\hat{x} + \dot{v}) = b_{eq} \tag{6.8}$$

$$\|\dot{v}\| \leq \delta \tag{6.9}$$

where the constraint (6.9) controls the local search of the programming problem. The algorithm that iteratively solves (6.7)-(6.9) is based on a Trust Region method as discussed in (DENNIS; LI; TAPIA, 1989). We summarize the method in Algorithm (3).

Algorithm 3 - TR algorithm for bilinear $r-$concave PC programming problem

1: (Preliminary) For fixed p in (5.3) and r in (6.1), find a conservative probability threshold p_r^{min}.
2: (Initialisation) Let x^S a Slater point for (5.1)-(5.3). Set the fixed values $\mu > 0$, $\rho, c_0 \in (0,1)$, update $t \leftarrow t^0$, $\hat{x} \leftarrow x^S$ and stopping tolerances $\varepsilon_b > 0$ and $\varepsilon_{tr} > 0$.
3: (Barrier loop) If $1/t < \varepsilon_b$, then *quit*. Otherwise, $\delta \leftarrow \delta^0$, $ver_2 = 1$ and *continue*.
4: (TR loop) If $ver_2 < \varepsilon_{tr}$, then *quit*. Otherwise, *continue*.
5: (Computation) Solve the programming problem (6.7) - (6.9) around $\hat{x}$. Getting the candidate solution v^*.
6: (TR update) Define: $ared := \phi(\hat{x}) - \phi(\hat{x} + v^*)$, $pred := \nabla\phi(\hat{x})\top v^* + \frac{1}{2}v^{*\top}\nabla^2\phi(\hat{x})v^*$ and $ver = \frac{ared}{pred}$.
   If $ver \notin \mathbb{R}$ or $ver \leq c_0$ or $g(\hat{x} + v^*) \leq p_r^{min}$, (Failed) then $\delta \leftarrow \delta \cdot \rho$ and $\hat{x} \leftarrow \hat{x}$.
   otherwise, (Succeed) then $\hat{x} \leftarrow \hat{x} + v^*$. $ver_2 = \|v^*\|$
   Return to Step 4.
7: (Barrier update) $t \leftarrow t \cdot \mu$ and return to Step 3.

The Step 1-Preliminary in Algorithm 3 is solved by a SOCP problem, being the SOC constraints defined in Remark 5.4.3.

$$\min_{\dot{x}} \quad k_0(\dot{x}) \tag{6.10}$$

$$\text{s.t.} \quad A \cdot \dot{x} = b_{eq} \tag{6.11}$$

$$\hat{\theta}_{a,p,r} \leq \theta_a(\dot{x}) \tag{6.12}$$

$$\hat{\theta}_{b,p,r} \leq \theta_b(\dot{x}) \tag{6.13}$$

For each r, the solution $\hat{x}_{cv}$ of (6.10)-(6.13) defines a conservative probability $p_r^{min} = g(\hat{x}_{cv})$ on which the convexity of the feasible set of (5.1)-(5.3) is locally observed. This solution furnish us a minimal probability threshold $p \geq p_r^{min}$, that must be met as a necessary condition along the next steps of the algorithm.

The strictly feasible solution x^S of Step 2 - Initialization can be found choosing a sufficient high $r \leq 1$ parameter in (6.10)-(6.13), since, as observed in Example 5.4.3, as much higher the r, narrower is the associated feasible set (and higher is p_r^{min}). The loop that starts in Step 3 - Barrier loop is an interior-point method that produces $1/t-$suboptimal central path solutions based in the version of (BOYD; VANDENBERGHE, 2006), that goes back to (FIACCO; MCCORMICK, 1987).

The TR loop starts at Step 4 - TR loop and ends at Step 6 - TR update, solve the programming problem iteratively (6.7)-(6.9), an approximation of the former problem (5.1)-(5.3), inside the radius defined by the parameter δ. Briefly, the algorithm is based on performing trial iterations that verifies, in a posteriori evaluation, if the proposed predicted reduction (identified as *pred*) is not much higher (or even smaller) then the actual reduction (*ared*). We make this verification in Step 6 - TR update. When the predicted reduction is not accepted (depending on c_0), or $ver \notin \mathbb{R}$ (due to the update $\hat{x}+v^*$ leading to imaginary values in $\phi(\hat{x}+v^*) = -\log(-k_r(\hat{x}+v^*))$), or even the solution leads to a lower probability $g(\hat{x}+v^*)$ then p_r^{min}(i.e. where the convexity is not ensured) then the radius δ is reduced by ρ and it is performed a new trial solution v^* in Step 5 - Computation. If the predicted reduction is accepted, then the fixed solution is updated $\hat{x} \leftarrow \hat{x} + v^*$. The TR loop is performed up to the moment that the vector solution v^*

is small enough, measured by ε_{tr}. When this solution meets the last criterion, then the barrier method parameter t is updated by μ in the last Step 7 - Barrier update, and so restarting the loop in Step 3 - Barrier loop. The solution $x^* = \hat{x}$ is achieved when Step 3 condition holds.

The convergence of the TR algorithm is extensively discussed in (DENNIS; LI; TAPIA, 1989). The central path performed by the Barrier method converges to the optimal point, as observed in (BOYD; VANDENBERGHE, 2006), since the algorithm starts its search in a feasible convex set and the criterion of convexity is always met by the inclusion of the constraint (6.9) which turns possible, at each iteration, to identify if it is a valid solution.

6.1 ENERGY PORTFOLIO APPLICATION

As an application, we work with a simplified version of a portfolio problem faced by energy utility companies. In general, utility companies in Brazil intend to reduce the penalization that they are exposed if they retained an undercontrated energy position, keeping under control the short-term exposition to the spot price of energy. Its energy consumption is backed by pre-contracted energy, measured by the value vp_t, for each stage t, and for new candidate contracts which have their amounts $\dot{pr}_d$, for each product d, defined as the output of the programming problem. The underlying programming problem is defined as follows:

$$\min_{\dot{pn},\dot{pr},\dot{xp}} \quad ct^\top \dot{pn} \tag{6.14}$$

$$\text{s.t.} \quad vp_t + \sum_{p \in P} \dot{pr}_d \cdot dt_{d,t} + \dot{xp}_t = dm_t \quad \forall t \in T \tag{6.15}$$

$$\mathbb{P}\{a \leq \dot{xp}^\top (pm - \widetilde{pc}) \leq b\} \geq p \tag{6.16}$$

$$\dot{pn}_t - \dot{xp}_t \geq 0, \quad \dot{pn}_t \geq 0 \quad \forall t \in T, \quad \dot{pr}_p \geq 0 \quad \forall d \in D \tag{6.17}$$

where the decision variables are: $\dot{pr}_d \in \mathbb{R}_+$ the amount (in units) of contracted product d, $\dot{xp} \in \mathbb{R}^T$ the amount (in MWh) of exposition by overcontracted (if negative) or undercontracted (if positive) energy and $\dot{pn} \in \mathbb{R}^T_+$ amount (in MWh) of penalization by the

undercontracting exposition for each stage $t = 1, \ldots, T$. The parameters are: $ct \in \mathbb{R}^T_+$ the penalization cost vector (in \$/MWh), vp_t the amount of pre-contracted energy (in MWh), at each stage t, already in the portfolio of the distribution company. $dt_{d,t} \in \mathbb{R}^T$ is the energy distribution profile (in MWh) over the period t, dm_t the energy load demand (in MWh), a, b upper and lower financial thresholds (in \$), $pm \in \mathbb{R}^T$ is the fixed average energy price in the company portfolio, and p the probability or desired level of risk. The energy spot price is defined as a normal random variable $\tilde{pc} \sim \mathcal{N}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^T$ and $\Sigma \in \mathbb{R}^{T \times T}$ is a positive semidefinite covariance matrix.

The objective function (6.14) minimizes the total cost of penalization of the company. The balance equations (6.15) establish that for each stage of time $t = 1, \ldots, T$, the sum of the pre-contracted energy, the contracted energy within the optimization horizon and the short-term exposition is equal to the energy demand. In these last equations, note that for a given product p the vector dt_p can assume negative terms, simulating energy selling products. The PC (6.16) is based on the fact that we are in a minimization problem (since we are reducing the risk of penalization), hence it is viewed in the opposite way of a common portfolio problem. In view of that, considering an overcontracted position for a time t, $\dot{xp}_t$ is negative and the company is selling energy to the spot market. In this case, the company gains money if the spot price $\widetilde{pc}$ is greater than the average price paid for the energy of its portfolio pm. The opposite happens in the undercontracted position, since in this case the company sells energy to its consumers by the average price of its portfolio buying it from the spot market. In this constraint, a is the lower accepted exposition (typically, but not restricted, to be negative) and b is the upper accepted exposition (usually positive). The first constraint in (6.17) defines how much energy is undercontracted, regarding to evaluate the penalization properly.

Simulations were performed in a common workstation (Intel Core i7 1.9 GHz, 8GB Ram), solving the SOCP and QCQP problems in Matlab equipped with CVX toolbox (M.C., 2018). The scenarios are based on two time horizons $t = 5$ and $t = 10$ years. The $d = 1, \ldots, 5$ possible products, the pre-contracted portfolio vp_t and the portfolio average prices pm were freely setted. The random prices $\tilde{pc}$ were estimated from Brazilian

Table 14 – Main results of the portfolio application

Sim.	Dem.	Horz. [yrs]	$p^{min}_{r=-\infty}$	$p_{r=0}$	p^{max}	Steps	Conv. [min]	Cost [$]
1	# 1	5	48.04 %	95.56 %	96.04 %	30	1.1	101.75
2	# 1	10	62.57 %	NaN %	91.39 %	31	15.4	238.23
3	# 2	5	48.04 %	NaN %	90.00 %	39	1.6	165.28
4	# 2	10	67.36 %	93.48 %	93.92 %	37	18.5	350.43

Source: Thesis results.

spot price historical data (18 years of weekly provided prices). For all simulations, it was fixed the minimal probability threshold $p = 85\%$ in (6.16). We tested the demand profiles dm_t for two different scenarios. The simulations main results are provided in Table 14, where the second column indicates the demand pattern, defined by parameter dm_t, the third column is the considered horizon, in years, meaning that, for instance, the problem has 5 stages for the first simulation. The next columns are sorted as $p^{min}_{r=-\infty}$ is the minimum conservative probability, $p_{r=0}$ is the minimum probability threshold for the log-concave case, pr_{max} is the probability to find Slater solution x^S in Step 2 - Initialization in Algorithm 3. The Steps column is the number of steps necessary for convergence of the algorithm (for both Barrier and TR loops altogether), followed by the convergence CPU time, in minutes, and the final total cost due to penalizations.

The #1 demand pattern has a regular rate of growth, turning the problem of portfolio definition easier and leading to fewer chances of short term expositions and penalizations. As a result, we have less total costs for both time horizons, when compared to the demand pattern #2. The convergence CPU time is intrinsically related to the number of stages, since the calculation of gradients and Hessians, as in (6.4), become computationally costly for higher dimensions. The probability $p^{min}_{r=-\infty}$ is defined as a result of the SOCP (6.10)-(6.13) using the procedures suggested in Remark 5.4.3 to define $\hat{\theta}_{a,p=85\%,r=-\infty}$ and $\hat{\theta}_{b,p=85\%,r=-\infty}$. Comparing $p^{min}_{r=-\infty}$ with $p_{r=0}$ we note how low can be the minimal conservative threshold, showing the advantages achieved by the use of the proposed method. Furthermore, for simulations 2 and 3 no feasible solutions were found for the log-concave case $p_{r=0}$, showing how restrictive are the solutions provided so far.

7 CONCLUSIONS

As a motivation for the first topic of the thesis, in Chapter 2, we started discussing the status quo on the use of stochastic programming with PC to solve UC problems. By this analysis, it is clear that the use of PC applications in the short-term problem, for instance, stochastic HTUC with use of renewable energy sources, as the wind generation, is a mainstream field of research. Nevertheless, it was found that the provided algorithms commonly uses strategies that work with individual PC or discrete approximations of the underlying random vector. The lack of accuracy in terms of joint probability level of these last algorithms makes necessary the use of ad-hoc strategies that hinders their applications to other (similar) problems. This situation opens a field of research to develop a methodology that one applies to a more general set of MILP problems, with probability levels handled precisely.

In Chapter 3, we offer a brief introduction to some mathematical instruments used by standard PC problems algorithms. The chapter starts with a simple application of a classic PC problem algorithm, in order to clarify how the instruments that are present throughout the chapter are linked and how it performs over iterations. It was emphasised the assumed properties of the constraints and the relationship between decision and random variables of the approach. It was also introduced the "hidden" motivation behind the use of the supporting hyperplanes in the algorithms. Section A.6 of this chapter introduces alternative interpolation algorithms that were tested by numeric simulations in the subsequent chapter. Visual insights, by figures, are also provided.

After the introductory chapters the present work pursued five main objectives. The first objective is accomplished by Chapter 4, where we have proposed a new variant of the supporting hyperplane method for mixed integer PC optimization problems. We have moreover benchmarked this algorithm on instances from stochastic UC with an alternative model based on using individual PC and an alternative solution method based

on the popular sample-based methodology. We illustrated that the solution coming from the model with individual PC does not satisfy the requested probability level. The sample based approach was found to be highly sensitive on the number of scenarios, in particular, to obtain feasible solutions. The model was found infeasible in several experiments, and resampling had to be used. Moreover, when the number of scenarios was too large (here 1000), a conventional computer workstation presented memory overflows. Aiming at verifying the impact of the method in a more complex system, a test with a 46-bus system was performed. The algorithm achieved convergence in a reasonable time, close to 1 hour, showing that scalability of the bus system is not necessarily a bottleneck of the method.

When comparing to other steps of the algorithm, the time for computing gradients and values of the probability functions were not very expensive during simulations. Nevertheless, we did observe the high impact on the MILP sub-problems resolution times. The latter depends strongly on the requested MILP optimality and the state of the simulated system itself. This scenario motivates the second objective of this work, that was to offer strategies to reduce its computation time. The first idea was to test the effect on changing the interpolation procedure. This last experiment found that the algorithm could offer faster optimal solutions, maintaining its quality, if the Ridder's method of root finding is used. Another strategy that was followed, and tested by numerical results, is the use of individual PCs into the lower bound problem of the proposed algorithm. This simple strategy makes the convergence speed up, reducing the CPU time to achieve optimal solutions in about 74.4% in some occasions, when compared to the original algorithm proposition, justifying its use.

Still in the domain of UC Problems, the third objective of the work intended to offer a practical engineering contribution by the proposition of a new HPF formulation. In Section 4.2.1 all the proposed formulation is described, with numerical experiments in Section 4.6, showing that it has brought gains when compared to linear piecewise counterparts. Simulations performed to different probability levels has shown advantages on its use by a faster rate of convergence and more accuracy on the power generations when

both methodologies are benchmarked to the nonlinear HPF simulations. Disadvantages of its use are the need for binary variables, making the model not proper for LP problems. Another disadvantage is that eventually ad-hoc calculations for the definition of the minors piecewise functions, as the joint turbine function $\dot{gh}_{l,t}$, the upstream and downstream levels and the hydraulic losses. These functions depend on pre-calculations regarding the operational region where the problem will work. In an equivalence between accuracy and flexibility, our formulation is less adaptable than the linear piecewise one since this last is typically used maintaining the set of approximations regardless of the initial state.

The second main topic of the thesis starts in Chapter 5, with the presentation of the fourth objective where it is discussed conditions to ensure that the probabilistic function $\mathbb{P}\{a \leq \tilde{\xi}^\top x \leq b\} \geq p$ is locally r-concave. Initially, hard to verify necessary and sufficient conditions are provided. Subsequently, in the view of making use of these conditions for real PC problems, more straightforward conditions are stated. These last conditions are extended to implied ones, being so reduced to thresholds that are appropriate to be stated as SOCP constraints.

By the achievements of the last chapter, the fifth and final objective of the thesis is the proposition of a TR algorithm that implies a SOCP problem resolution and iterative QCQP problems, aiming at exploring the defined convex feasible set, and through this, solving the former PC problem exactly. A real reduced case of use, involving an energy portfolio problem, was problematized, that was efficiently solved by the proposed algorithm. As a result, we observe that for high levels of concavity, such as 0-concave (log-concave) case, the problem results in many instances infeasible, but this restriction is appropriately bypassed using a lower r-concavity level. So, the features provided in this chapter accomplishes the previous condition.

For the first topic of the thesis, suggestions for extensions of the present work are in consider less generic wind power random model, using appropriate distributions. Indeed in this thesis the mathematical convenience of using normality assumptions predominated. Some challenges on dealing with these appropriate distributions are traced in Appendix B. In this line of research, an involving and compelling trend is in the use

of Copulas in eventual convex feasible sets, as motivated in Appendix Section B.1. Another path is on the study of the performance of the algorithm when imposed to more complex systems, requiring possibly the use of decomposition methods as provided in (TAKIGAWA; FINARDI; SILVA, 2013; van ACKOOIJ; MALICK, 2016; SCUZZIATO; FINARDI; FRANGIONI, 2017). There are also interesting lines opened by the dynamic PC programming (ANDRIEU; HENRION; RÖMISCH, 2009), offering perspectives that approximate the use of PC problems to standard multi-stage stochastic programming strategies (SHAPIRO; DENTCHEVA; RUSZCZYŃSKI, 2009).

For the second topic of the present thesis, that find the conditions to extend the two-sided bilinear PC structures to the class of local general concave functions, future works can be addressed on evolutions of the method to joint bilinear PC problems, where the challenge is on working with harsh algebraic chain rule manipulations. Another possibility is the extension of the statements to a more general class of probability distributions, such as the symmetric ones.

REFERENCES

ANDRIEU, L.; HENRION, R.; RÖMISCH, W. A model for dynamic chance constraints in hydro power management. **Preprint 536, DFG Research Center MATHEON "Mathematics for key technologies"**, p. 21, 2009.

ARNOLD, T. et al. A mixed-integer stochastic nonlinear optimization problem with joint probabilistic constraints. **Pacific Journal of Optimization**, v. 10, p. 5–20, 2014.

BARBER, D. P. D. C. B.; HUHDANPAA, H. The quickhull algorithm for convex hulls. **ACM Transactions on Mathematical Software (TOMS)**, v. 22, n. 4, p. 469–483, 1996. Available at: <http://dl.acm.org/citation.cfm?id=235821>.

BARRERA, J. et al. Chance-constrained problems and rare events: an importance sampling approach. **Mathematical Programming**, v. 157, n. 1, p. 153–189, May 2016. ISSN 1436-4646.

BELLONI, A. et al. Bundle relaxation and primal recovery in unit-commitment problems. the brazilian case. **Annals of Operations Research**, v. 120, n. 1-4, p. 21–44, 2003.

BEN-TAL, A.; GHAOUI, L. E.; NEMIROVSKI, A. **Robust Optimization**. [S.l.]: Princeton University Press, 2009. 564 p.

BENDERS, J. Partitioning procedures for solving mixed-variables programming problems. **Numerische Mathematik**, v. 4, n. 1, p. 238–252, 1962.

BERTSIMAS, D.; BROWN, D.; CARAMANIS, C. Theory and applications of robust optimization. **SIAM Review**, v. 53, n. 3, p. 464–501, 2011.

BERTSIMAS, D. et al. Adaptive robust optimization for the security constrained unit commitment problem. **IEEE Transactions on Power Systems**, v. 28, n. 1, p. 52–63, March 2013.

BERTSIMAS, D.; SIM, M. The price of robustness. **Operations Research**, v. 52, n. 1, p. 35–53, February 2004.

BORGHETTI, A. et al. A MILP approach for short-term hydro scheduling and unit commitment with head-dependent reservoir. **IEEE Transactions on Power Systems**, v. 23, n. 3, p. 1115–1124, 2008.

BOYD, S.; VANDENBERGHE, L. Convex optimization. **Available at http://www.stanford.edu/ boyd/cvxbook**, ISBN 0 521 83378 7, 2006.

BRASCAMP, H.; LIEB, E. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log-concave functions and with an application to the diffusion equations. **Journal of Functional Analysis**, v. 22, p. 366–389, 1976.

CALAFIORE, G. C.; CAMPI, M. C. Uncertain convex programs: Randomized solutions and confidence levels. **Mathematical Programming**, v. 102, n. 1, p. 25–46, 2005.

CAMPI, M. C.; GARATTI, S. A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality. **Journal of Optimization Theory and Applications**, v. 148, n. 2, p. 257–280, 2011.

CAROTHERS, N. L. **Real analysis**. [S.l.]: Cambridge University Press, 2000. ISBN 978-0-521-49749-7 978-0-521-49756-5.

CHANG, G. et al. Experiences with mixed integer linear programming based approaches on short-term hydro scheduling. **IEEE Transactions on Power systems**, v. 16, n. 4, p. 743–749, 2001.

CONEJO, A. et al. Self-scheduling of a hydro producer in a pool-based electricity market. **IEEE Transactions on Power Systems**, v. 17, n. 4, p. 1265–1272, 2002. ISSN 0885-8950. Available at: <http://ieeexplore.ieee.org/document/1137622/>.

DENNIS, J.; LI, S.-B.; TAPIA, R. A unified approach of trust-region methods for nonsmooth optimization. **Mathematical Programming**, p. 1–41, 1989.

DENTCHEVA, D. Optimisation models with probabilistic constraints. In: SHAPIRO, A.; DENTCHEVA, D.; RUSZCZYŃSKI, A. (Ed.). **Lectures on Stochastic Programming. Modeling and Theory**. [S.l.]: SIAM and MPS, Philadelphia, 2009, (MPS-SIAM series on optimization, v. 9). p. 87–154.

DING, X. et al. Studies on stochastic unit commitment formulation with flexible generating units. **Electric Power Systems Research**, v. 80, p. 130–141, 2010.

DINIZ, A.; MACEIRA, M. A four-dimensional model of hydro generation for the short-term hydrothermal dispatch problem considering head and spillage effects. **IEEE Transactions on Power Systems**, v. 23, n. 3, p. 1298–1308, 2008.

DINIZ, A. L. Test cases for unit commitment and hydrothermal scheduling problems. IEEE, p. 1–8, jul. 2010. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5589757>.

DINIZ, A. L.; HENRION, R. On probabilistic constraints with multivariate truncated gaussian and lognormal distributions. **Energy Systems**, p. 1–19, 2016.

DRITSAS, I. **Stochastic Optimization - Seeing the Optimal for the Uncertain**. INTECH, 2011. 476 p. Available at: <http://www.intechopen.com/books/show/title/stochastic-optimization-seeing-the-optimal-for-the-uncertain>.

ELSTRODT, J. **Maß und Integrationstheorie**. 7th. ed. [S.l.]: Springer-Verlag, 2011. 451 p.

FARSHBAF-SHAKER, M. H.; HENRION, R.; HOMBERG, D. Optimization methods applied for solving the short-term hydrothermal coordination problem. **Set-Valued and Variational Analysis**, v. 26, p. 821–841, 2018.

FIACCO, A.; MCCORMICK, G. **Nonlinear Programming: Sequential Unconstrained Minimization Techniques**. 2nd. ed. [S.l.]: SIAM, 1987. 226 p. (Classics in Applied Mathematics).

FINARDI, E.; SCUZZIATO, M. Hydro unit commitment and loading problem for day-ahead operation planning problem. **Electrical Power and Energy Systems**, v. 44, p. 7–16, 2013.

FINARDI, E.; SCUZZIATO, M. A comparative analysis of different dual problems in the lagrangian relaxation context for solving the hydro unit commitment problem. **Electric Power Systems Research**, v. 107, p. 221–229, 2014.

FINARDI, E.; SILVA, E. D. Solving the hydro unit commitment problem via dual decomposition and sequential quadratic programming. **IEEE Transactions on Power Systems**, v. 21, n. 2, p. 835–844, 2006.

FINARDI, E.; TAKIGAWA, F. Y. K.; BRITO, B. H. Assessing solution quality and computational performance in the hydro unit commitment problem considering different mathematical programming approaches. **Electric Power Systems Research**, To Appear, p. 1–22, 2016.

FRANGIONI, A.; GENTILE, C.; LACALANDRA, F. Sequential Lagrangian-MILP Approaches for Unit Commitment Problems. **International Journal of Electrical Power and Energy Systems**, v. 33, p. 585–593, 2011.

GENZ, A. Numerical computation of multivariate normal probabilities. **J. Comp. Graph Stat.**, v. 1, p. 141–149, 1992.

GRAHAM, R. Efficient algorith for determining the convex hull of a finite planar set. **Information Processing Letters 1**, 1972.

GUEDES, L. S. M. et al. A unit commitment algorithm and a compact MILP model for short-term hydro-power generation scheduling. **IEEE Transactions on Power Systems**, v. 32, n. 5, p. 3381–3390, 2017. ISSN 0885-8950, 1558-0679. Available at: <http://ieeexplore.ieee.org/document/7790824/>.

GULLIVER, J.; ARNDT, R. **Hydropower Engineering Handbook**. [S.l.]: McGraw-Hill, 1991.

HENRION, R. Introduction to chance constraint programming. **Tutorial paper for the Stochastic Programming Community HomePage, http://www.wias-berlin.de/people/henrion/publikat.html**, 2004.

HENRION, R. Structural properties of linear probabilistic constraints. **Optimization: A Journal of Mathematical Programming and Operations Research**, v. 56, n. 4, p. 425–440, August 2007.

HENRION, R.; STRUGAREK, C. Convexity of chance constraints with independent random variables. **Computational Optimization and Applications**, v. 41, p. 263–276, 2008.

HENRION, R.; STRUGAREK, C. Convexity of chance constraints with dependent random variables: the use of copulae. In: BERTOCCHI, M.; CONSIGLI, G.; DEMPSTER, M. (Ed.). **Stochastic Optimization Methods in Finance and Energy: New Financial Products and Energy Market Strategies**. [S.l.]: Springer-Verlag New York, 2011, (International Series in Operations Research and Management Science). p. 427–439.

JABR, R. A. Adjustable robust OPF with renewable energy sources. **IEEE Transactions on Power Systems**, IEEE, v. 28, n. 4, p. 4741–4751, 2013.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 6th ed. ed. [S.l.]: Pearson Prentice Hall, 2007. OCLC: ocm70867129. ISBN 978-0-13-187715-3.

KATAOKA, S. A stochastic programming model. **Econometrica**, v. 31, p. 181–196, 1963.

KUCUKYAVUZ, S. On mixing sets arising in chance-constrained programming. **Mathematical Programming**, v. 132, n. 1-2, p. 31–56, 2012.

LI, X. et al. Hydro unit commitment via mixed integer linear programming: A case study of the three gorges project, china. **IEEE Transactions on Power Systems**, v. 29, n. 3, p. 1232–1241, maio 2014. ISSN 0885-8950, 1558-0679.

LUBIN, M.; BIENSTOCK, D.; VIELMA, J. P. Two-sided linear chance constraints and extensions. **Arxiv 1507.01995**, p. 1–19, 2016.

LUEDTKE, J.; AHMED, S. A sample approximation approach for optimization with probabilistic constraints. **SIAM Journal on Optimization**, v. 19, p. 674–699, 2008.

LUEDTKE, J.; AHMED, S.; NEMHAUSER, G. An integer programming approach for linear programs with probabilistic constraints. **Mathematical Programming**, v. 122, n. 2, p. 247–272, 2010.

MACEIRA, M. E. P. et al. Chain of Optimizaton Models for Setting the Energy Dispatch and Spot Price in the Brazilian System. In: . [S.l.: s.n.], 2002. Session 43, Paper 1, Page 1.

MATOS, V. D. et al. Assessment of electricity distribution companies risks in the brazilian energy market framework. **24rd International Conference on Electricity Distribution**, v. 24, 2017.

M.C., B. S. G. **The CVX User's Guide**. [S.l.: s.n.], 2018. 23 p. (Release 2.1).

McCORMICK, G. P. Computability of global solutions to factorable nonconvex programs: Part i - convex underestimating problems. **Mathematical programming**, v. 10, n. 1, p. 147–175, 1976.

MINOUX, M.; ZORGATI, R. Convexity of gaussian chance constraints and of related probability maximization problems. **Computational Statistics**, v. 31, n. 1, p. 387–408, 2016.

MINOUX, M.; ZORGATI, R. Global probability maximization for a gaussian bilateral inequality in polynomial time. **Journal of Global Optimization**, v. 68, n. 4, 2017.

NELSEN, R. B. **An Introduction to Copulas**. 2nd. ed. [S.l.]: Springer-Verlag New York, 2006. 272 p. (Springer Series in Statistics).

OZTURK, U.; MAZUMDAR, M.; NORMAN, B. A solution to the stochastic unit commitment problem using chance constrained programming. **IEEE Transactions on Power Systems**, v. 19, n. 3, p. 1589–1598, August 2004.

PAGNONCELLI, B.; AHMED, S.; SHAPIRO, A. Sample average approximation method for chance constrained programming: Theory and applications. **J. Optim. Theory Appl**, v. 142, p. 399–416, 2009.

PAGNONCELLI, B. K.; REICH, D.; CAMPI, M. C. Risk-return trade-off with the scenario approach in practice: A case study in portfolio selection. **Journal of Optimization Theory and Applications**, v. 155, n. 2, p. 707–722, Nov 2012. ISSN 0022-3239, 1573-2878.

PANNE, C. van de; POPP, W. Minimum-cost cattle feed under probabilistic protein constraints. **Managment Science**, v. 9, p. 405–430, 1963.

PAREDES, M.; MARTINS, L. S. A.; SOARES, S. Using semidefinite relaxation to solve the day-ahead hydro unit commitment problem. **IEEE Transactions on Power Systems**, v. 30, n. 5, p. 2695–2705, 2015. ISSN 0885-8950, 1558-0679. Available at: <http://ieeexplore.ieee.org/document/6919349/>.

PATA, V. **Appunti del Corso di ANALISI REALE E FUNZIONALE**. 2007.

PINTO, R. J.; BORGES, C. L. T.; MACEIRA, M. E. P. An efficient parallel algorithm for large scale hydrothermal system operation planning. **IEEE Transactions on Power Systems**, v. 28, n. 4, p. 4888–4896, 2013. ISSN 0885-8950, 1558-0679.

PRÉKOPA, A. On probabilistic constrained programming. In: KUHN, H. (Ed.). **Proceedings of the Princeton Symposium on Math. Prog.** [S.l.: s.n.], 1970. v. 28, p. 113–138.

PRéKOPA, A. Boole-bonferroni inequalities and linear programing. **Operations Research**, v. 36, p. 145–162, 1988.

PRÉKOPA, A. **Stochastic Programming**. [S.l.]: Kluwer, Dordrecht, 1995.

PRÉKOPA, A. Probabilistic programming. In: RUSZCZYŃSKI, A.; SHAPIRO, A. (Ed.). **Stochastic Programming**. [S.l.]: Elsevier, Amsterdam, 2003, (Handbooks in Operations Research and Management Science, v. 10). p. 267–351.

PRÉKOPA, A.; SZÁNTAI, T. Flood control reservoir system design using stochastic programming. **Math. Programming Study**, v. 9, p. 138–151, 1978.

PRÉKOPA, A.; SZÁNTAI, T. On optimal regulation of a storage level with application to the water level regulation of a lake. **European Journal of Operations Research**, v. 3, p. 175–189, 1979.

RAMALHO, G. M. et al. Um metodo de otimizacao estocastica para auxiliar companhias distribuidoras no gerenciamento de seus portfolios de energia. **CIDEL 2018**, Sep 2018.

REDONDO, N.; CONEJO, A. Short-term hydro-thermal coordination by lagrangian relaxation: solution of the dual problem. **IEEE Transactions on Power Systems**, v. 14, p. 89–95, 1999.

RIDDERS, C. A new algorithm for computing a single root of a real continuous function. **IEEE Transactions on circuits and systems**, v. 26, n. 11, p. 979–980, 1979.

RINOTT, Y. On the convexity of measures. **Annals of Probability**, v. 4, p. 1020–1026, 1976.

RODRIGUES, L. F.; DINIZ, A. L.; PRADA, R. B. Avaliação do uso de restrições probabilísticas para a superfície de aversão a risco no problema de planejamento de médio prazo da operação hidrotérmica. In: **XXIV SNPTEE SEMINÁRIO NACIONAL DE PRODUçÃO E TRANSMISSÃO DE ENERGIA ELÉTRICA**. [S.l.: s.n.], 2017. p. 9.

RUSZCZYŃSKI, A.; SHAPIRO, A. **Stochastic Programming**. [S.l.]: Elsevier, Amsterdam, 2003. (Handbooks in Operations Research and Management Science, v. 10).

SANTO, T. D.; COSTA, A. S. Hydroelectric unit commitment for power plants composed of distinct groups of generating units. **Electric Power Systems Research**, v. 137, p. 16–25, 2016. ISSN 03787796. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0378779616300840>.

SANTOS, T. N.; DINIZ, A. L.; BORGES, C. L. T. A new nested benders decomposition strategy for parallel processing applied to the hydrothermal scheduling problem. **IEEE Transactions on Smart Grid**, v. 8, n. 3, p. 1504–1512, 2017. ISSN 1949-3053, 1949-3061.

SCARCELLI, R. O. et al. Aggregated inflows on stochastic dynamic programming for long term hydropower scheduling. In: . [S.l.]: IEEE, 2014. p. 1–6. ISBN 978-1-4799-5904-4.

SCUZZIATO, M. R.; FINARDI, E. C.; FRANGIONI, A. Comparing spatial and scenario decomposition for stochastic hydrothermal unit commitment problems. **IEEE Transactions on Sustainable Energy, DOI 10.1109/TSTE.2017.2781908**, 2017. ISSN 1949-3029, 1949-3037. Available at: <http://ieeexplore.ieee.org/document/8171768/>.

SHAPIRO, A. **Monte Carlo sampling methods. Chapter 6 in (RUSZCZYŃSKI; SHAPIRO, 2003)**. [S.l.]: Elsevier, 2003. 353-425 p. (Handbooks in Operations Research and Management Science, v. 10).

SHAPIRO, A.; DENTCHEVA, D.; RUSZCZYŃSKI, A. **Lectures on Stochastic Programming. Modeling and Theory**. [S.l.]: SIAM and MPS, Philadelphia, 2009. (MPS-SIAM series on optimization, v. 9).

STEWART, J. **Calculus**. 7th ed. ed. [S.l.]: Brooks/Cole, Cengage Learning, 2012. OCLC: ocn651906312. ISBN 978-0-538-49781-7 978-0-8400-5818-8.

SZÁNTAI, T. **Numerical evaluation of probabilities concerning multi-dimensional probability distributions**. Thesis (PhD) — Hungarian Academy of Sciences, 1985.

SZÁNTAI, T. A computer code for solution of probabilistic-constrained stochastic programming problems. **In (Y. Ermoliev and R.J.-B. Wets eds.): *Numerical Techniques for Stochastic Optimization***, p. 229–235, 1988.

TAHANAN, M. et al. Large-scale unit commitment under uncertainty: a literature survey. **4OR**, v. 13, n. 2, p. 115–171, 2015.

TAKIGAWA, F.; FINARDI, E.; SILVA, E. da. A decomposition strategy to solve the short-term hydrothermal scheduling based on lagrangian relaxation. **Journal of Algorithms and Optimization**, v. 1, n. 1, p. 13–24, 2013.

TAKTAK, R.; D'AMBROSIO, C. An overview on mathematical programming approaches for the deterministic unit commitment problem in hydro valleys. **Energy Systems**, To appear, 2016.

TONG, B.; ZHAI, Q.; GUAN, X. An MILP based formulation for short-term hydro generation scheduling with analysis of the linearization effects on solution feasibility. **IEEE Transactions on Power Systems**, v. 28, n. 4, p. 3588–3599, 2013.

van ACKOOIJ, W. **Chance Constrained Programming: with applications in Energy Management**. 250 p. Thesis (PhD) — École Centrale Paris, December 2013. Available at: <https://tel.archives-ouvertes.fr/tel-00978519/>.

van ACKOOIJ, W. Eventual convexity of chance constrained feasible sets. **Optimization (A Journal of Math. Programming and Operations Research)**, v. 64, n. 5, p. 1263–1284, 2015.

van ACKOOIJ, W. Convexity statements for linear probability constraints with gaussian technology matrices and copulæ correlated rows. **Researchgate**, p. 1–16, 2017.

van ACKOOIJ, W.; FINARDI, E. C.; RAMALHO, G. M. An exact solution method for the hydrothermal unit commitment under wind power uncertainty with joint probability constraints. **IEEE Transactions on Power Systems**, p. 6487 – 6500, 2018. Available at: <https://ieeexplore.ieee.org/document/8387479/>.

van ACKOOIJ, W. et al. On probabilistic constraints induced by rectangular sets and multivariate normal distributions. **Mathematical Methods of Operations Research**, v. 71, n. 3, p. 535–549, 2010.

van ACKOOIJ, W. et al. **Chance Constrained Programming and Its Applications to Energy Management. In (DRITSAS, 2011) (Chapter 13)**. INTECH, 2011. 291-320 p. Available at: <http://www.intechopen.com/articles/show/title/chance-constrained-programming-and-its-applications-to-energy-management>.

van ACKOOIJ, W. et al. Joint chance constrained programming for hydro reservoir management. **Optimization and Engineering**, v. 15, p. 509–531, 2014.

van ACKOOIJ, W. et al. Large-scale unit commitment under uncertainty: an updated literature survey. **Annals of Operations Research**, v. 271, p. 11–85, 12 2018.

van ACKOOIJ, W.; MALICK, J. Decomposition algorithm for large-scale two-stage unit-commitment. **Annals of Operations Research**, v. 238, n. 1, p. 587–613, 2016.

van ACKOOIJ, W.; MALICK, J. Eventual convexity of probability constraints with elliptical distributions. **Mathematical Programming**, p. 1–18, Jan 2018.

van ACKOOIJ, W.; MINOUX, M. A characterization of the subdifferential of singular Gaussian distribution functions. **Set Valued and Variational Analysis**, v. 23, n. 3, p. 465–483, 2015.

van ACKOOIJ, W.; OLIVEIRA, W. de. Convexity and optimization with copulæ structured probabilistic constraints. **Optimization: A Journal of Mathematical Programming and Operations Research**, v. 65, n. 7, p. 1349–1376, 2016.

VEINOTT, A. The supporting hyperplane method for unimodal programming. **Operations Research**, v. 15, p. 147–152, 1967.

WANG, Q.; GUAN, Y.; WANG, J. A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output. **IEEE Transactions on Power Systems**, v. 27, n. 1, p. 206–215, 2012.

WOLSEY, L. A. **Integer Programming**. 1. ed. [S.l.]: Wiley, 1998. ISBN 9780471283669.

WU, Z. et al. A Solution to the Chance-Constrained Two-Stage Stochastic Program for Unit Commitment With Wind Energy Integration. **IEEE Transactions on Power Systems**, v. 31, n. 6, p. 4185–4196, nov. 2016. ISSN 0885-8950, 1558-0679. Available at: <http://ieeexplore.ieee.org/document/7384775/>.

ZADEH, Z. M.; KHORRAM, E. Convexity of chance constrained programming problems with respect to a new generalized concavity notion. **Annals of Operations Research**, v. 196, n. 1, p. 651–662, 2012.

ZHANG, Y. et al. Chance-constrained two-stage unit commitment under uncertain load and wind power output using bilinear benders decomposition. **IEEE Transactions on Power Systems**, 2017. ISSN 0885-8950, 1558-0679. Available at: <http://ieeexplore.ieee.org/document/7822944/>.

ZHAO, C. et al. Expected value and chance constrained stochastic unit commitment ensuring wind power utilization. **IEEE Transactions on Power Systems**, v. 29, n. 6, p. 2696–2705, nov. 2014. ISSN 0885-8950, 1558-0679. Available at: <http://ieeexplore.ieee.org/document/6810870/>.

APPENDIX A – Mathematical Background

In this appendix we review some mathematical tools used throughout the document.

A.1 NOTIONS OF MEASURE THEORY

In this section, we offer a non-self-contained introduction to notions of measure theory that we use currently along this work. For further introdutory material in measure theory, we refer to (PATA, 2007) and for a complete material on real and functional analysis we refer to (CAROTHERS, 2000; ELSTRODT, 2011). Given a metric space $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{M}}, \boldsymbol{\lambda})$, where $\boldsymbol{\Omega}$ is a topological space and $\boldsymbol{\mathcal{M}}$ is a $\boldsymbol{\sigma}$-algebra over $\boldsymbol{\Omega}$. The function $\boldsymbol{\lambda : \mathcal{M} \to [0, \infty]}$ is a measure in this metric space iff it agrees with *Countable Additivity*, and so for all countable collections of pairwise disjoint sets $\boldsymbol{E_n \in \mathcal{M}}$, we have $\boldsymbol{\lambda(\cup_n E_n) = \sum_n \lambda(E_n)}$, where $\boldsymbol{\cup_n}$ represents a countable union of sets. We also have the following properties for the measures:

- $\boldsymbol{\lambda(\emptyset) = 0}$;
- $\boldsymbol{\lambda}$ is finitely additive, i.e. given $\boldsymbol{E, F \in \mathcal{M}}$ with $\boldsymbol{E \subset F}$, then $\boldsymbol{\lambda(E) \leq \lambda(F)}$;
- $\boldsymbol{\lambda}$ is monotone, i.e. given $\boldsymbol{E, F \in \mathcal{M}}$ with $\boldsymbol{E \subset F}$, if $\boldsymbol{\lambda(E) < \infty}$, then $\boldsymbol{\lambda(F - E) = \lambda(F) - \lambda(E)}$;
- $\boldsymbol{\lambda(E_n) \to \lambda(E)}$ if $\boldsymbol{E = \cup E_n}$, $\boldsymbol{E_n \in \mathcal{M}}$;
- if $\boldsymbol{\lambda(E \cap F) < \infty}$, then $\boldsymbol{\lambda(E \cup F) = \lambda(E) + \lambda(F) - \lambda(E \cap F)}$.

One important concept to introduce is the *Lebesgue measure*. Given a metric space $\boldsymbol{(\mathbb{R}, \mathcal{L}(\mathbb{R}), \lambda)}$, where $\boldsymbol{\mathcal{L}(\mathbb{R})}$ is a $\boldsymbol{\sigma}$-algebra of $\mathbb{R}$, an *interval* is given by $\boldsymbol{I = (a, b) \subset \mathbb{R}}$, and we define $\boldsymbol{\ell(I) = b - a}$ as the *length*. Considering $\boldsymbol{E \in \mathcal{L}(\mathbb{R})}$, the *external measure* in $\mathbb{R}$ is the measure that respects the property $\boldsymbol{\lambda(E) = \inf_{E \subset \cup_n I_n} \sum_n \ell(I_n)}$. The meaning is that the external measure of $\boldsymbol{E}$, or $\boldsymbol{\lambda(E)}$, is the infimum of the union of countable intervals that contains $\boldsymbol{E}$.

Following (PATA, 2007), we define some useful terms in respect to measurability of sets.

Definition A.1.1 *Given a metric space* $\boldsymbol{(\Omega, \mathcal{M}, \lambda)}$ *as defined above, if* $\boldsymbol{\mathcal{E}}$ *is a collection of subsets of* $\boldsymbol{\Omega}$*, there is the smallest* $\boldsymbol{\sigma}$*-algebra* $\boldsymbol{\mathcal{M}^*}$ *over* $\boldsymbol{\Omega}$ *such that* $\boldsymbol{\mathcal{E} \subset \mathcal{M}^*}$*. This* $\boldsymbol{\sigma}$*-algebra is said generated by* $\boldsymbol{\mathcal{E}}$*.*

Definition A.1.2 ***(Borel set)*** *The* $\boldsymbol{\sigma}$*-algebra generated by the open sets of* $\boldsymbol{\Omega}$ *is called the Borel* $\boldsymbol{\sigma}$*-algebra and it is indicated by* $\boldsymbol{\mathcal{B}(\Omega)}$*. The elements of* $\boldsymbol{\mathcal{B}(\Omega)}$ *are said Borel sets, and so the open and closed sets, the intersection of countable open sets and union of countable closed sets are Borel sets.*

The family of sets $\boldsymbol{E \in \mathcal{L}(\mathbb{R})}$ that respects the condition $\boldsymbol{\lambda(T) = \lambda(T \bigcap E) + \lambda(T \bigcap E^c)}$, for $\boldsymbol{\forall T \in \mathcal{L}(\mathbb{R})}$ forms a $\boldsymbol{\sigma}$-algebra of $\mathbb{R}$, called the $\boldsymbol{\sigma}$-algebra of Lebesgue and indicated simply by $\boldsymbol{\mathcal{L}}$. The settlement of the external measure $\boldsymbol{\lambda}$ over $\boldsymbol{\mathcal{L}}$ is called the *Lebesgue measure*. Some properties of Lebesgue measures are:

- It is complete over a set, i.e. for $\boldsymbol{E \subset N}$ and $\boldsymbol{\lambda(N) = 0}$ implies that $\boldsymbol{E \in \mathcal{L}}$, and $\boldsymbol{\lambda(E) = 0}$;

- Invariant to translation, i.e., $\boldsymbol{\lambda(E + a) = \lambda(E)}$ with $\boldsymbol{E} \subset \mathbb{R}$ and $\boldsymbol{a} \in \mathbb{R}$;
- Agrees with the interval measure $\boldsymbol{\lambda(a, b) = b - a}$ with $\boldsymbol{a}$ and $\boldsymbol{b} \in \mathbb{R}$;
- Measures all the Borel sets (over $\mathbb{R}$), so all the $\boldsymbol{\sigma}$-algebras of open sets;
- Assigns a finite measure for all the bounded sets;
- Assigns a null measure for all countable sets.

As presented, the Lebesgue measure can be understood as the standard procedure of assigning length to a set. Using the theory for higher dimension sets $\in \mathbb{R}^{\boldsymbol{n}}$, one defines areas or volumes of subsets in Euclidean spaces.

Another important concept is of *null sets*. A subset $\boldsymbol{N} \subset \mathbb{R}$ has null Lebesgue measure and is considered to be a null set in $\mathbb{R}$ if, given any positive number $\boldsymbol{\epsilon}$, there is a sequence $\boldsymbol{I_n}$ of intervals in $\mathbb{R}$ such that $\boldsymbol{N}$ is contained in the union of the $\boldsymbol{I_n}$ and the total length of the union is less than $\boldsymbol{\epsilon}$. This condition can be generalised to $\mathbb{R}^{\boldsymbol{n}}$, using n-cubes instead of intervals. For instance:

- With respect to $\mathbb{R}^{\boldsymbol{n}}$, all 1-point sets are null, and therefore all countable sets are null. In particular, the set Q of rational numbers is a null set, despite being dense in $\mathbb{R}$;
- All the subsets of $\mathbb{R}^{\boldsymbol{n}}$ whose dimension are smaller than $\boldsymbol{n}$ have null Lebesgue measure in $\mathbb{R}^{\boldsymbol{n}}$. For instance straight lines are null sets in $\mathbb{R}^{\boldsymbol{2}}$;
- Sard's lemma: the set of critical values of smooth functions has measure zero.

Finally, linked to the idea of Lebesgue measure we have the probability measure. Consider the metric (probability) space ($\boldsymbol{\Omega}$,$\boldsymbol{\mathcal{M}}$, $\mathbb{P}$) and the random vector $\tilde{\boldsymbol{\xi}}$ that takes its value in $\tilde{\boldsymbol{\xi}} : \boldsymbol{\Omega} \rightarrow \mathbb{R}^{\boldsymbol{T}}$, where $\boldsymbol{\Omega}$ is the *space of events*. We define the *probability measure* induced by a random variable $\tilde{\boldsymbol{\xi}}$ in such a way that: $\boldsymbol{P_{\tilde{\xi}}(B) := \mathbb{P}[w \in \Omega : \tilde{\xi}(\omega) \in B]}$ for any $\boldsymbol{B} \in \boldsymbol{\mathcal{B}}(\mathbb{R}^{\boldsymbol{T}})$, where $\boldsymbol{\mathcal{B}}(\mathbb{R}^{\boldsymbol{T}})$ is the Borel $\boldsymbol{\sigma}$-algebra. A direct property of probability measures is that the measure of the whole set in which it is defined is $\boldsymbol{1}$, i.e. if $\mathbb{R}^{\boldsymbol{T}}$ is the whole probability space, $\boldsymbol{P_{\tilde{\xi}}(\mathbb{R}^T) = 1}$. This notation and concept of probability measure $\mathbb{P}$ will be used throughout this document.

A.2 RELATIONS BETWEEN $\boldsymbol{R}-$CONCAVE PROBABILITY DISTRIBUTION FUNCTIONS

This section presents some concepts about r-concave functions, which is a generalization of concave functions and plays a vital role to establish the convexity of feasible sets of PC problems. Following the definition of generalized concave functions (DENTCHEVA, 2009):

Definition A.2.1 *Let* $\boldsymbol{f(x) \geq 0}$ *be a function defined on a convex set* $\boldsymbol{\Omega} \subset \mathbb{R}^{\boldsymbol{n}}$. *This function is said to be* $\boldsymbol{r-}$*concave, where* $\boldsymbol{r} \in [-\boldsymbol{\infty}, +\boldsymbol{\infty}]$ *if for* $\forall \boldsymbol{x, y} \in \boldsymbol{\Omega}$ *and* $\forall \boldsymbol{\lambda} \in [\boldsymbol{0, 1}]$ *it holds:*

$$\boldsymbol{f(\lambda x + (1 - \lambda)y) \geq m_r(f(x), f(y), \lambda)}, \tag{A.1}$$

where the map $\boldsymbol{m_r : \mathbb{R}_+ \times \mathbb{R}_+ \times [0, 1]}$ *is given by:*

• *if* $\boldsymbol{a \cdot b = 0}$*:*

$$\boldsymbol{m_r(a, b, \lambda) = 0,} \tag{A.2}$$

• *else if* $\boldsymbol{a > 0}$, $\boldsymbol{b > 0}$, $\boldsymbol{0 \leq \lambda \leq 1}$*:*

$$\boldsymbol{m_r(a, b, \lambda)} = \begin{cases} \boldsymbol{a^\lambda b^{1-\lambda}} & \boldsymbol{if \quad r = 0,} \\ \boldsymbol{\max\{a, b\}} & \boldsymbol{if \quad r = \infty,} \\ \boldsymbol{\min\{a, b\}} & \boldsymbol{if \quad r = -\infty,} \\ \boldsymbol{(\lambda a^r + (1-\lambda) b^r)^{1/r}} & \boldsymbol{otherwise.} \end{cases} \tag{A.3}$$

One observes that when $\boldsymbol{r = 0}$, $\boldsymbol{f(x)}$ is said to be *log-concave*, i.e. $\boldsymbol{log(f(x))}$ is a concave function. When $\boldsymbol{r = -\infty}$, $\boldsymbol{f(x)}$ is said to be *quasi-concave* and $\boldsymbol{r = 1}$ is the classic definition of *concavity*. From this definition, it is clear that for $\boldsymbol{r > 0}$ the mapping $\boldsymbol{f^r}$ is concave and $\boldsymbol{r < 0}$ it is convex. For example, we can observe that the function $\boldsymbol{h(x) = \frac{1}{x^2}}$ is not concave, but it is $\boldsymbol{-1}$-concave, since $\boldsymbol{h^{-1}(x) = x^2}$ is convex.

The mapping $\boldsymbol{r \mapsto m_r(a, b, \lambda)}$ is nondecreasing and continuous, as proved by Lemma 4.8 in (DENTCHEVA, 2009). As a result, a direct consequence of this lemma is the fact that if a function is concave, it will also be log-concave and quasi-concave. The opposite is not valid, i.e. a quasi-concave function is not necessarily log-concave and so on.

Expanding these concepts to probability measures, a probability measure $\mathbb{P}$ is said to be $\boldsymbol{r}-$concave if, for any Borel measurable sets $\boldsymbol{A, B \subset \Omega}$, we can assert the following inequality, for all $\boldsymbol{\lambda \in [0, 1]}$:

$$\boldsymbol{P(\lambda A + (1-\lambda)B) \geq m_r(P(A), P(B), \lambda),} \tag{A.4}$$

where the sum over sets is the Minkowski sum, i.e., is the set defined by $\boldsymbol{\lambda A + (1-\lambda)B = \{\lambda x + (1-\lambda)y : x \in A, y \in B\}}$. A random vector $\boldsymbol{\tilde{\xi} \in \mathbb{R}^T}$ has an $\boldsymbol{r}-$concave distribution if the probability measure $\mathbb{P}_{\tilde{\xi}}$ induced by $\boldsymbol{\tilde{\xi}}$ on $\mathbb{R}^T$ is $\boldsymbol{r}-$concave (PRÉKOPA, 2003). From this definition, (DENTCHEVA, 2009) establishes the relation between the concavity of the PDF $\boldsymbol{f_{\tilde{\xi}}(\cdot)}$ and its CDF $\boldsymbol{F_{\tilde{\xi}}(\cdot)}$. This is summarized by the following theorem.

Theorem 1 *The probability measure* $\mathbb{P}$ *defined on the set* $\boldsymbol{\Omega \in \mathbb{R}^T}$ *is* $\boldsymbol{\gamma}$*-concave with* $\boldsymbol{\gamma \in [-\infty, 1/T]}$ *iff its PDF with respect to the Lebesgue measure on* $\boldsymbol{\Omega}$ *is* $\boldsymbol{r}-$*concave, with* $\boldsymbol{r}$ *defined as follows:*

$$\boldsymbol{r} = \begin{cases} \boldsymbol{\gamma/(1-T\gamma)} & \boldsymbol{if \quad \gamma \in (-\infty, 1/T),} \\ \boldsymbol{-1/T} & \boldsymbol{if \quad \gamma = -\infty,} \\ \boldsymbol{+\infty} & \boldsymbol{if \quad \gamma = 1/T.} \end{cases}$$

This important theorem states that, for instance, if the PDF of a given random vector $\boldsymbol{\tilde{\xi}}$ is log-concave then it implies that its CDF will also be log-concave. The relationship established by Theorem 1 is remarkable since it usually is much easier to verify the $\boldsymbol{r}-$concavity of PDFs than that of CDFs.

A.3 OPERATIONS WITH RANDOM VARIABLES

Eventually, it is necessary to deal with operations of random variables. It is the case of , for instance, when one must include two renewable energies in the same bus. To

motivate such analysis, we consider the random vectors $\tilde{\boldsymbol{\xi}}_{\mathbf{1}}, \ldots, \tilde{\boldsymbol{\xi}}_{\boldsymbol{I}} \in \mathbb{R}^{\boldsymbol{T}}$. Considering normal random distributions assumptions, in (JOHNSON; WICHERN, 2007) we find:

Proposition A.3.1 *Let $\tilde{\boldsymbol{\xi}}_{\mathbf{1}}, \ldots, \tilde{\boldsymbol{\xi}}_{\boldsymbol{i}}, \ldots, \tilde{\boldsymbol{\xi}}_{\boldsymbol{I}} \in \mathbb{R}^{\boldsymbol{T}}$ be independent, with $\tilde{\boldsymbol{\xi}}_{\boldsymbol{i}}$ distributed as $\boldsymbol{N}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma})$ (Note the covariance matrix $\boldsymbol{\Sigma}$ is fixed). The following random vector:*

$$\boldsymbol{V} = \boldsymbol{c_1}\boldsymbol{\xi_1} + \ldots + \boldsymbol{c_I}\boldsymbol{\xi_I},$$

is distributed as $\boldsymbol{N}(\sum_{i=1}^{\boldsymbol{I}} \boldsymbol{c_i}\boldsymbol{\mu_i}, (\sum_{i=1}^{\boldsymbol{I}} \boldsymbol{c_i^2})\boldsymbol{\Sigma})$.

The last proposition allows us to work with affine operations of normal random vectors in PCs.

For log-concave distributions, as also stated in (van ACKOOIJ, 2015), we observe that in Theorem 4.2.3 (PRÉKOPA, 1995) is verified that the convolution of log-concave densities results in a log-concave density. If we consider the individual $\boldsymbol{e}$-component of the random vectors $\tilde{\boldsymbol{\xi}}_{\mathbf{1}\boldsymbol{e}}$ and $\tilde{\boldsymbol{\xi}}_{\mathbf{2}\boldsymbol{e}}$ following log-concave distributions, in this sense we can ensure that the marginal CDF $\boldsymbol{F_e}(\boldsymbol{h_e}(\dot{\boldsymbol{x}})) = \mathbb{P}[\tilde{\boldsymbol{\xi}}_{\mathbf{1}\boldsymbol{e}} + \tilde{\boldsymbol{\xi}}_{\mathbf{2}\boldsymbol{e}} \leq \boldsymbol{h_e}(\dot{\boldsymbol{x}})]$, is a log-concave function if $\boldsymbol{h_e}(\cdot)$ is quasi-concave. This last case can potentially be explored with the use of Copulas, as the cases provided in Appendix B.2.

A.4 APPROXIMATION OF CONVEX SETS BY SUPPORTING HYPERPLANES

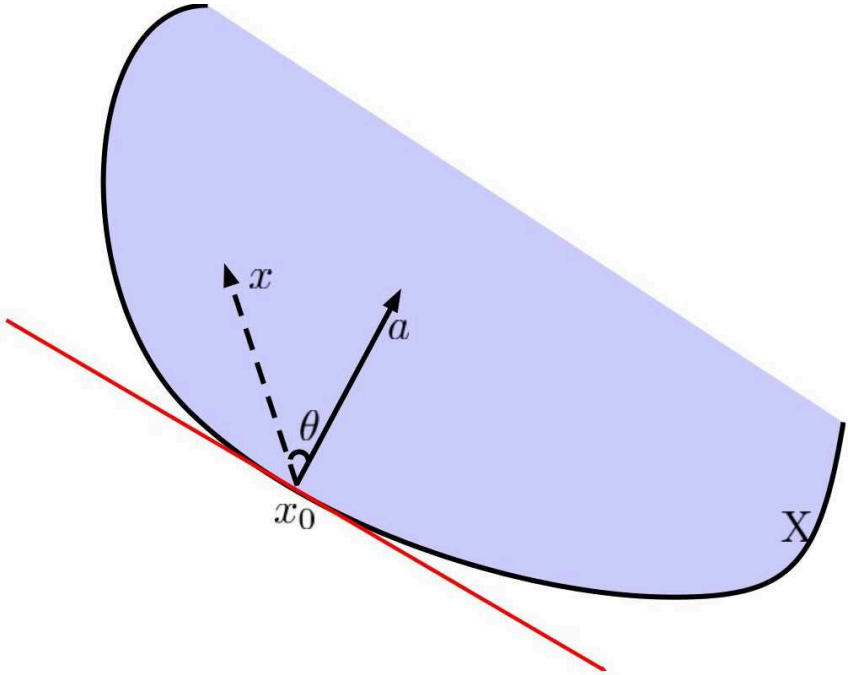
A fundamental concept to the optimization algorithms of the present work is the *supporting hyperplane* to a given set, a classic tool of convex programming. In (BOYD; VANDENBERGHE, 2006) the supporting hyperplane is defined as follows:

Definition A.4.1 *Let the convex set $\boldsymbol{X} \subseteq \mathbb{R}^{\boldsymbol{T}}$ and a point $\boldsymbol{x_0} \in \mathbb{R}^{\boldsymbol{T}}$ in the boundary of set $\boldsymbol{X}$. If we have $\boldsymbol{a} \neq \mathbf{0}$ such that $\boldsymbol{a}^\top(\boldsymbol{x} - \boldsymbol{x_0}) \geq \mathbf{0}, \forall \boldsymbol{x} \in \boldsymbol{X}$, then the hyperplane defined by $\{\boldsymbol{x} | \boldsymbol{a}^\top(\boldsymbol{x} - \boldsymbol{x_0}) = \mathbf{0}\}$ is called the supporting hyperplane to $\boldsymbol{X}$ at the point $\boldsymbol{x_0}$.*

The last definition is equivalent to say that there is a hyperplane tangent to the set $\boldsymbol{X}$ (red line in Figure 19) at the point $\boldsymbol{x_0}$ that separates the space into two half spaces. One half-space containing the set $\boldsymbol{X}$ and the other without intersections with it. The vector $\boldsymbol{a}$ is the normal vector that defines the hyperplane at the point $\boldsymbol{x_0}$. Definition A.4.1 states that if we can find $\boldsymbol{a}$ such that the dot product $\boldsymbol{a}^\top(\boldsymbol{x} - \boldsymbol{x_0}) = \|\boldsymbol{a}\|\|(\boldsymbol{x} - \boldsymbol{x_0})\| \cos\boldsymbol{\theta} \geq \mathbf{0}$, hence $\boldsymbol{x} \in \boldsymbol{X}$ is such that $\frac{-\boldsymbol{\pi}}{\mathbf{2}} \leq \boldsymbol{\theta} \leq \frac{\boldsymbol{\pi}}{\mathbf{2}}$ for all $\boldsymbol{x}$, the supporting hyperplane is established. A direct result of the supporting hyperplane theorem (BOYD; VANDENBERGHE, 2006) is that for a nonempty convex set $\boldsymbol{X}$ and any point $\boldsymbol{x_0}$ on its boundary, there exists a supporting hyperplane to $\boldsymbol{X}$ at $\boldsymbol{x_0}$. This concept is frequently used throughout Section 3.1 and the proposed MILP PC in Chapter 4.

Regarding the supporting hyperplane concept, one challenge consists in determining the procedure to evaluate the normal vector $\boldsymbol{a}$, which defines the supporting hyperplane of a set for a given point. For a calculus intuition, we follow (STEWART, 2012). For a given function $\boldsymbol{\phi} : \mathbb{R}^{\boldsymbol{T}} \to \mathbb{R}$, and the expression $\boldsymbol{\phi}(\dot{\boldsymbol{x}}) = \boldsymbol{p}$, where $\boldsymbol{p} \in \mathbb{R}$, we define the level curve surface $\boldsymbol{S}$ by the function $\boldsymbol{F} : \mathbb{R}^{\boldsymbol{T}} \times \mathbb{R} \to \mathbb{R}$ by $\boldsymbol{F}(\dot{\boldsymbol{x}}, \boldsymbol{p}) = \boldsymbol{\phi}(\dot{\boldsymbol{x}}) - \boldsymbol{p}$. Here $\dot{\boldsymbol{x}}$ is a $\mathbb{R}^{\boldsymbol{T}}$-vector $\dot{\boldsymbol{x}} = (\dot{\boldsymbol{x}}_{\mathbf{1}}, \dot{\boldsymbol{x}}_{\mathbf{2}}, \ldots, \dot{\boldsymbol{x}}_{\boldsymbol{T}})^\top$. Choosing a generic level curve as $\boldsymbol{F}(\dot{\boldsymbol{x}}, \boldsymbol{p}) = \mathbf{0}$, we assume the point $(\boldsymbol{x_0}, \boldsymbol{p_0})$, where $\boldsymbol{x_0} = (\boldsymbol{x_{01}}, \ldots, \boldsymbol{x_{0T}})^\top$ is fixed, belonging to the domain of $\boldsymbol{F}(\cdot)$, and lying on the surface $\boldsymbol{S}$.

Figure 19 – Supporting hyperplane to set X at the point $\boldsymbol{x_0}$.



Source: Thesis results.

Assuming also that there is a parametric curve over $\boldsymbol{S}$ described by the vector function $\boldsymbol{c(q)} = (\dot{\boldsymbol{x}}(\boldsymbol{q}), \boldsymbol{p(q)}) = (\dot{\boldsymbol{x}}_1(\boldsymbol{q}), \ldots, \dot{\boldsymbol{x}}_T(\boldsymbol{q}), \boldsymbol{p(q)})^\top$, where $\boldsymbol{q} \in \mathbb{R}$, and the functions $\dot{\boldsymbol{x}}_t(\boldsymbol{q})$ and $\boldsymbol{p(q)}$ are generic functions that describe the trajectory of $\boldsymbol{c(q)}$ through $\boldsymbol{S}$. We want to certificate that the derivative of the vector function $\boldsymbol{c(q)}$ is a tangent vector to the surface $\boldsymbol{S}$.

Definition A.4.2 *(Tangent vector) If we consider the generic parameters $\boldsymbol{q_0} \in \mathbb{R}$ and $\boldsymbol{q_1} = \boldsymbol{q_0} + \boldsymbol{h}$, where $\boldsymbol{h}$ is a scalar, by construction it is clear that $\boldsymbol{y} = \boldsymbol{c(q_1)} - \boldsymbol{c(q_0)}$ is a vector that passes through $\boldsymbol{c(q_1)}$ and $\boldsymbol{c(q_0)}$. The distance between $\boldsymbol{c(q_1)}$ and $\boldsymbol{c(q_0)}$ decreases when $\boldsymbol{h}$ gets smaller, and so intuitively we approach to a tangent line of $\boldsymbol{c(q_0)}$. The vector $\frac{\boldsymbol{y}}{\boldsymbol{h}}$ has the same direction of the vector $\boldsymbol{y}$. Taking $\boldsymbol{h} \to \boldsymbol{0}$, we have the ingredients to find the derivative of $\boldsymbol{c(q)}$ on $\boldsymbol{q_0}$:*

$$\frac{\boldsymbol{dc(q_0)}}{\boldsymbol{dq}} = \lim_{\boldsymbol{h \to 0}} \frac{\boldsymbol{c(q_0 + h)} - \boldsymbol{c(q_0)}}{\boldsymbol{h}}, \tag{A.5}$$

we define (A.5) *as the tangent vector $\boldsymbol{c'(q)}$ to the surface $\boldsymbol{S}$ at $\boldsymbol{q_0}$.*

Assume that $\boldsymbol{c(q)}$ passes through $(\boldsymbol{x(q_0), p(q_0)}) = (\boldsymbol{x_0, p_0})$ at $\boldsymbol{q} = \boldsymbol{q_0}$. Since the curve $\boldsymbol{C}$ respects the function $\boldsymbol{F(\cdot)}$ over $\boldsymbol{S}$ it satisfies the original equation $\boldsymbol{F(c(q))} = \boldsymbol{0}$. Assuming continuity and differentiability assumptions for the functions $\boldsymbol{\phi(\cdot)}$ and $\boldsymbol{c(\cdot)}$, we assume that exists a family of points $\hat{\boldsymbol{q}}$ such that:

$$(\boldsymbol{\nabla \phi(\dot{x}), -1})^\top \cdot \boldsymbol{c'(q)}|_{\boldsymbol{q \in \hat{q}}} = \frac{\boldsymbol{\partial \phi}}{\boldsymbol{d\dot{x}_1}} \frac{\boldsymbol{d\dot{x}_1}}{\boldsymbol{dq}}|_{\boldsymbol{q \in \hat{q}}} + \cdots + \frac{\boldsymbol{\partial \phi}}{\boldsymbol{d\dot{x}_T}} \frac{\boldsymbol{d\dot{x}_T}}{\boldsymbol{dq}}|_{\boldsymbol{q \in \hat{q}}} - \frac{\boldsymbol{\partial p}}{\boldsymbol{dp}} \frac{\boldsymbol{dp}}{\boldsymbol{dq}}|_{\boldsymbol{q \in \hat{q}}} = \boldsymbol{0}, \tag{A.6}$$

where $\boldsymbol{\nabla \phi(\dot{x})}$ is the gradient vector of $\boldsymbol{\phi(\dot{x})}$. If we are on the point $(\boldsymbol{x_0, p_0}) \in \hat{\boldsymbol{q}}$ on the surface $\boldsymbol{S}$, consider the tangent vector to the parametric curve $\boldsymbol{C}$ at $\boldsymbol{q_0}$, i.e. $\boldsymbol{c'(q_0)}$. Equation (A.6) states that the dot product between the vector $\boldsymbol{(\nabla \phi(x_0), -1)}$ and the vector

$\boldsymbol{c'(q_0)}$ equals $\mathbf{0}$. If none of them is zero, this is only possible if they are perpendicular. From that, it allows us to conclude that $\boldsymbol{(\nabla\phi(x_0), -1)}$ is normal to the level surface $\boldsymbol{S}$ at point $\boldsymbol{(x_0, p_0)}$. Keeping this normal vector, we construct a tangent hyperplane $\boldsymbol{T_{(x_0,p_0)}(x, p)}$ at $\boldsymbol{(x_0, p_0)}$, using its standard equation (STEWART, 2012):

$$\boldsymbol{(\nabla\phi(\dot{x}_0), -1)^\top \cdot (\dot{x}_1 - x_{01}, \ldots, \dot{x}_T - x_{0T}, p - p_0) = 0,} \tag{A.7}$$

where the vector $\boldsymbol{(\dot{x}_1 - x_{01}, \ldots, \dot{x}_T - x_{0T}, p - p_0) = (\dot{x}, p) - (x_0, p_0)}$. Recovering the supporting hyperplane definition, it is clear that any point of the tangent hyperplane can always be described by a linear combination of the tangent vector $\boldsymbol{c'(q_0)}$ chosen an appropriate curve $\boldsymbol{C}$ that passes through $\boldsymbol{(x_0, p_0)}$. If we compare (A.7) to Definiton A.4.1, we note that for $\boldsymbol{a = (\nabla\phi(\dot{x}_0), -1)}$, and $\boldsymbol{T_{(x_0,p_0)}(x, p)}$ is the supporting hyperplane of the set $\boldsymbol{\{(\dot{x}, p) : (\dot{x}, p) \in X \times \mathbb{R}\}}$ at the point $\boldsymbol{(x_0, p_0)}$. Applying the dot product to the last term in (A.7), it results:

$$\boldsymbol{\nabla\phi(\dot{x}_0)^\top \cdot (\dot{x} - x_0) = p - p_0,} \tag{A.8}$$

note that, on plugging $\boldsymbol{p = p_0}$ we are still in a point of the supporting hyperplane $\boldsymbol{T_{(x_0,p_0)}(\dot{x}, p)}$ where $\boldsymbol{\phi(\dot{x}) = p_0}$. We also certify that $\boldsymbol{p \geq p_0 \Leftrightarrow \phi(\dot{x}_0)^\top \cdot (\dot{x} - x_0) \geq 0}$.

Considering an appropriate function $\boldsymbol{\phi(\dot{x})}$, we can assume convexity to the domain $\boldsymbol{\dot{x} \in X}$. In the case of a log-concave function with domain set $\boldsymbol{X = M(p_0)}$, as defined in(3.12) at a probability level $\boldsymbol{p = p_0}$. In this case, we may also say that the vector $\boldsymbol{(\nabla\phi(x_0), -1)}$ defines a hyperplane that supports the epigraph of function $\boldsymbol{\phi(\cdot)}$ at the point $\boldsymbol{(x_0, p_0)}$, and if it happens for any point $\boldsymbol{(\dot{x}, p)}$ in the boundary of $\boldsymbol{X}$, we ensure that $\boldsymbol{X}$ is convex.

For a given point $\boldsymbol{x_0}$, suppose that $\boldsymbol{\phi(\dot{x}_0)}$ is known and finite, and we know how to compute gradients $\boldsymbol{\nabla\phi(\dot{x}_0)}$, but the only knowledge that we have about the domain set $\boldsymbol{X}$ is that it is convex by a direct result of (3.12). Choosing $\boldsymbol{p \geq p_0}$, by the inequality $\boldsymbol{\nabla\phi(\dot{x}_0)^\top \cdot (\dot{x} - x_0) \geq 0}$ we are in practice defining a half-space of $\boldsymbol{\mathbb{R}^T \supset X}$ limited by the supporting hyperplane $\boldsymbol{T_{(x_0,p_0)}(\dot{x}, p)}$ at point $\boldsymbol{(x_0, p_0)}$. This half-space can be regarded as a local approximation of the domain $\boldsymbol{X}$ by the use of the supporting hyperplane as in Definition A.4.1.

Maintaining the probability level $\boldsymbol{p_0}$, we use similar procedures as described in this section to compute another local approximation (half-space) of $\boldsymbol{X}$ for different a different point $\boldsymbol{x_1}$. This provides us a new supporting hyperplane $\boldsymbol{T_{(x_1,p_0)}(\dot{x}, p)}$. Then, we can request that the $\boldsymbol{\dot{x}}$ respects inequalities limited by $\boldsymbol{T_{(x_0,p_0)}(\dot{x}, p)}$ and by $\boldsymbol{T_{(x_1,p_0)}(\dot{x}, p)}$, defining the following set:

$$\boldsymbol{X_{apx}(p_0, x_0, x_1) = \{\dot{x} : \nabla\phi(\dot{x}_0)^\top \cdot (\dot{x} - x_0) \geq 0, \nabla\phi(\dot{x}_1)^\top \cdot (\dot{x} - x_1) \geq 0\}.} \tag{A.9}$$

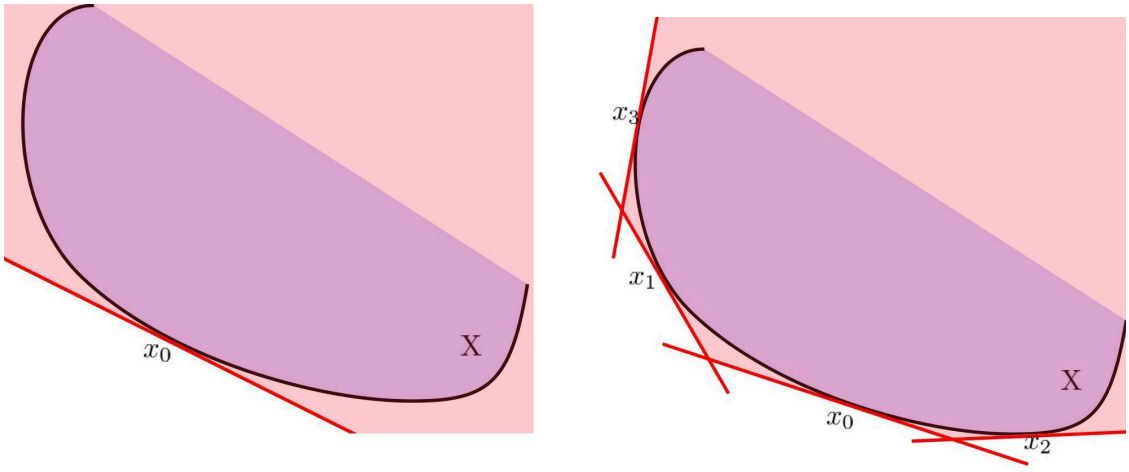
This procedure can be repeated for more iterates $\boldsymbol{x_2, \ldots, x_i}$, where $\boldsymbol{i \in \mathcal{I}}$ is an index, creating the family of sets $\boldsymbol{X_{apx}(p_0, x_0, \ldots, x_i)}$. As a result, the sets

$\boldsymbol{X_{apx}(p_0, x_0, \cdots, x_i)}$ are iterative local approximations to the original convex set $\boldsymbol{X}$, for a fixed $\boldsymbol{p_0}$. In practice, the set $\boldsymbol{X_{apx}(p_0, x_0, \cdots, x_I)}$ represents the intersection of half-spaces limited by $\boldsymbol{T_{(x_i,p_0)}(\dot{x}, p)}$, for $\boldsymbol{i = 1, \ldots, I}$.

We illustrate the above procedure by the schematic representations in Figures 20(a) and 20(b). The set $\boldsymbol{X_{apx}(p_0, x_0)}$ is coloured in light red in Figure 20(a) as a local approximation of $\boldsymbol{X}$ limited by $\boldsymbol{T_{(x_0,p_0)}(\dot{x}, p)}$. Figure 20(b) we have the set $\boldsymbol{X_{apx}(p_0, x_0, x_1, x_2, x_3)}$ as an approximation of $\boldsymbol{X}$ made by the intersection of half-spaces limited by four sup-

porting hyperplanes at points $\boldsymbol{x_0}, \boldsymbol{x_1}, \boldsymbol{x_2}$ and $\boldsymbol{x_3}$, keeping fixed $\boldsymbol{p_0}$. The applications of these iterative approximations in the context of the present work are explored in Section 3.4.

Figure 20 – Representation of the approximation of the convex set $\boldsymbol{X}$ by the intersection of half-spaces limited by supporting hyperplanes.



(a)One supporting hyperplane (b)Four supporting hyperplanes

Source: Thesis results.

A.5 CALCULATION OF MULTIVARIATE DISTRIBUTIONS

To deal with joint PC, it is essential to be equipped with methods to evaluate multivariate CDFs for a given point. As establish in Section 3.3, these instruments are used to calculate the probability level itself, and also for gradients evaluations ($\boldsymbol{\nabla \phi(\cdot)}$, $\boldsymbol{\nabla G(\cdot, \cdot)}$), since for multivariate normal random vectors these calculations are reduced to lower dimensional normal CDFs. The calculation of distribution functions is a complex field of study by itself, and the present section aims at pointing out remarks on this complexity and the instruments used by the here discussed algorithms.

In (HENRION, 2004) is traced two lines of research in methods of CDF calculations. One is on the use of simulation with bounding techniques (e.g.(PRéKOPA, 1988)). Another approach is on the use of numerical computation of multiple integrals, in the line of research of (GENZ, 1992).

The last approach, that was used in the main applications of the thesis, is of particular interest by the precision and the capability of computing, in reasonable time, multivariate normal distributions in rectangular sets, as found in (3.17). The method is based on a sequence of transformations that, starting from the multivariate normal distribution function integral defined in the original set, transforms it into an integral over a unit hypercube. Dealing with a $\boldsymbol{T}$-dimensional random vector, the basic algorithm requires the generation of $\boldsymbol{T - 1}$ Monte-Carlo uniformly distributed random samples for each iteration.

In the work (PRÉKOPA, 1995), pointing back to (SZÁNTAI, 1985) and (PRÉKOPA; SZÁNTAI, 1978), it is discussed an algorithm to calculate bidimensional multivariate normal distributions, based on series expansions and Hermite polynomials. These methods are used to calculate unilateral distribution functions in the form of (3.13). Combin-

ing these methods with combinatorial strategies, as pointed out in (van ACKOOIJ et al., 2010), it is also possible to derive the calculation of rectangular distribution functions as in equation (3.17). This is done by a composition of unilateral CDFs of form (3.13). The drawback here is the combinatorial explosion, of an order of $\boldsymbol{2^T}$ terms (where $\boldsymbol{T}$ is the dimension of the random vector), reducing the effectiveness of the technique to small dimension sizes.

A.6 INTERPOLATION STEP - ROOTS ESTIMATION

The final fundamental step in the here used PC algorithms is the interpolation step. Once fixed the desired probability level, this procedure aims to find solutions targeting this probability level, using an iterative process. Hence, this section will introduce the bisection procedure, a popular technique that takes its origin from early numerical calculus, and it is still frequently used in algorithms which root-finding is requested.

This section also presents algorithms for alternative procedures to find solutions given the desired probability, where numerical comparisons are provided in Section 4.5.1. In the context of programming PC applications, in the present section we use the following convention of functions and parameters: $\boldsymbol{\phi_{\tilde{\xi}}(\dot{x})}$ is the distribution function of $\tilde{\boldsymbol{\xi}}$, $\boldsymbol{p}$ is the (minimum) probability level (e.g. $\boldsymbol{p = 95\%}$), the mapping $\boldsymbol{g_{\tilde{\xi}} : \mathbb{R}^m \to \mathbb{R}}$ is defined as $\boldsymbol{g_{\tilde{\xi}}(\dot{x}) := \phi_{\tilde{\xi}}(\dot{x}) - p}$, $\boldsymbol{x^L}$ is a lower bound solution (i.e. $\boldsymbol{\phi_{\tilde{\xi}}(x^L) \leq p}$) and $\boldsymbol{x^S}$ is a Slater solution (i.e. $\boldsymbol{\phi_{\tilde{\xi}}(x^S) > p}$). The next subsections present the used algorithms for the bisection method, the Regula Falsi and Ridders method. The section ends with remarks on the application of other bracketing methods.

A.6.1 Bisection method

The former basic algorithm is, for each iteration, divide the section between two iterates into two equivalent sized segments and certify where the root is located. This information is provided by the evaluation of the underlying function $\boldsymbol{g_{\tilde{\xi}}(\cdot)}$ in the extremes and analysing in which segment the function changes its sign. The procedure is provided by the Algorithm 4.

The bisection method is an accurate procedure, with guaranteed linear convergence. A property of the method is that it depends on the evaluation of the function just once per iteration, this is an advantage when the function that one is dealing with requires a hard computational burden, as our case with multivariate CDFs.

A.6.2 Regula Falsi method

Another standard algorithm that belongs to the bracketing methods is the Regula Falsi. It is very similar to the bisection procedure, but rather than use the middle point at each iteration, it defines the next iterate by the point (solution) where a secant line, or an affine function, between the last two iterates cross the abscissa (or the domain basis). The algorithm is defined in Algorithm 5.

Note that the control at stage **4** (Reset) always certificate that the two brackets have opposite signs, ensuring the convergence of the method.

Algorithm 4 - Bisection method

1: (Initialisation) Definition of the lower limit $\boldsymbol{x_a = x^L}$ with $\boldsymbol{g_{\tilde{\xi}}(x_a) < 0}$. Definition of upper limit $\boldsymbol{x_b = x^S}$ with $\boldsymbol{g_{\tilde{\xi}}(x_b) > 0}$. Definition of $\boldsymbol{\lambda = 0.5}$, $\boldsymbol{\lambda_{aux} = 0.5}$. Definition of stopping criteria $\boldsymbol{\epsilon}$ and counter $\boldsymbol{k = 0}$.
2: (Iterate)Definition of iterate $\boldsymbol{x_t = \frac{x_a + x_b}{2}}$; Calculate $\boldsymbol{g_{\tilde{\xi}}(x_t)}$; $\boldsymbol{\lambda_{aux} \leftarrow \lambda_{aux}/2}$;
3: (Stopping test)
If $\boldsymbol{\frac{|g_{\tilde{\xi}}(x_t)|}{p} < \epsilon}$ **:** STOP, the answer is $\boldsymbol{x_t}$ with a $\boldsymbol{\lambda}$ parameter;
Else: go to next step;
4: (Reset)
If $\boldsymbol{(g_{\tilde{\xi}}(x_t) \cdot g_{\tilde{\xi}}(x_a) \leq 0)}$
$\boldsymbol{x_b = x_t;} \quad \boldsymbol{g_{\tilde{\xi}}(x_b) = g_{\tilde{\xi}}(x_t);} \quad \boldsymbol{\lambda \leftarrow \lambda + \lambda_{aux}}$
Else If $\boldsymbol{(g_{\tilde{\xi}}(x_t) \cdot g_{\tilde{\xi}}(x_b) \leq 0)}$
$\boldsymbol{x_a = x_t;} \quad \boldsymbol{g_{\tilde{\xi}}(x_a) = g_{\tilde{\xi}}(x_t);} \quad \boldsymbol{\lambda \leftarrow \lambda - \lambda_{aux}}$
5: (Return) $\boldsymbol{k \leftarrow k + 1;}$ Go to Step **2**

Algorithm 5 - Regula Falsi method

1: (Initialisation) Definition of the lower limit $\boldsymbol{x_{t-1} = x^L}$ with $\boldsymbol{g_{\tilde{\xi}}(x_{t-1})}$. Definition of upper limit $\boldsymbol{x_t = x^S}$ with $\boldsymbol{g_{\tilde{\xi}}(x_t)}$. Definition of stopping criteria $\boldsymbol{\epsilon}$ and counter $\boldsymbol{k = 0}$.
2: (Iterate)Definition of iterate:
$\boldsymbol{x_{t+1} = x_{t-1} + \frac{g_{\tilde{\xi}}(x_{t-1})(x_t - x_{t-1})}{g_{\tilde{\xi}}(x_t) - g_{\tilde{\xi}}(x_{t-1})}}$; Calculate $\boldsymbol{g_{\tilde{\xi}}(x_{t+1})}$;
3: (Stopping test) If $\boldsymbol{\frac{|g_{\tilde{\xi}}(x_{t+1})|}{p} < \epsilon}$ **:** STOP, the answer is $\boldsymbol{x_{t+1}}$;
Else : go to next Step;
4: (Reset)
If $\boldsymbol{(g_{\tilde{\xi}}(x_{t-1}) \cdot g_{\tilde{\xi}}(x_{t+1}) \leq 0)}$
$\boldsymbol{x_t = x_{t+1};} \quad \boldsymbol{g_{\tilde{\xi}}(x_t) = g_{\tilde{\xi}}(x_{t+1})}$
Else If $\boldsymbol{(g_{\tilde{\xi}}(x_t) \cdot g_{\tilde{\xi}}(x_{t+1}) \leq 0)}$
$\boldsymbol{x_{t-1} = x_{t+1};} \quad \boldsymbol{g_{\tilde{\xi}}(x_{t-1}) = g_{\tilde{\xi}}(x_{t+1});}$
5: (Return) $\boldsymbol{k \leftarrow k + 1;}$ Go to Step **2**

A.6.3 Ridders method

The Ridders' method is due to C. Ridders (RIDDERS, 1979). Similarly to the Regula Falsi, rather than use an affine function between the two brackets, it uses an exponential function. It is also computed the bisection iterate, using so, at each iteration, two iterates, the bisection and the modified Regula Falsi ones.

Algorithm 6 - Ridders method

1: (Initialisation) Definition of the lower limit $\boldsymbol{x_t = x^L}$ with $\boldsymbol{g_{\tilde{\xi}}(x_t)}$. Definition of upper limit $\boldsymbol{x_{qt} = x^S}$ with $\boldsymbol{g_{\tilde{\xi}}(x_{qt})}$. Definition of stopping criteria $\boldsymbol{\epsilon}$ and counter $\boldsymbol{k = 0}$.

2: (Iterate)Definition of iterate:
$\boldsymbol{x_{base} = \frac{x_t + x_{qt}}{2}}$; Calculate $\boldsymbol{g_{\tilde{\xi}}(x_{base})}$;
If $\boldsymbol{(g_{\tilde{\xi}}(x_t) \geq 0)}$: $\boldsymbol{x_{t+1} = x_{base} + \frac{(x_{base} - x_t) g_{\tilde{\xi}}(x_{base})}{\sqrt[2]{g_{\tilde{\xi}}(x_{base})^2 - g_{\tilde{\xi}}(x_t) g_{\tilde{\xi}}(x_{qt})}}}$
Else: $\boldsymbol{x_{t+1} = x_{base} - \frac{(x_{base} - x_t) g_{\tilde{\xi}}(x_{base})}{\sqrt[2]{g_{\tilde{\xi}}(x_{base})^2 - g_{\tilde{\xi}}(x_t) g_{\tilde{\xi}}(x_{qt})}}}$
Calculate $\boldsymbol{g_{\tilde{\xi}}(x_{t+1})}$;

3: (Stopping test)
If $\boldsymbol{\frac{|g_{\tilde{\xi}}(x_{t+1})|}{p} < \epsilon}$ **:** STOP, the answer is $\boldsymbol{x_{t+1}}$;
Else : go to next Step;

4: (Reset)
If $\boldsymbol{(g_{\tilde{\xi}}(x_{base}) \cdot g_{\tilde{\xi}}(x_{t+1}) \leq 0)}$
$\boldsymbol{x_{qt} = x_{base}; g_{\tilde{\xi}}(x_q t) = g_{\tilde{\xi}}(x_{base})}$
Else If $\boldsymbol{(g_{\tilde{\xi}}(x_t) \cdot g_{\tilde{\xi}}(x_{t+1}) \leq 0)}$
$\boldsymbol{x_{qt} = x_t; g_{\tilde{\xi}}(x_{qt}) = g_{\tilde{\xi}}(x_t);}$
Else If $\boldsymbol{(g_{\tilde{\xi}}(x_{qt}) \cdot g_{\tilde{\xi}}(x_{t+1}) \leq 0)}$
$\boldsymbol{x_{qt} = x_{qt}; g_{\tilde{\xi}}(x_{qt}) = g_{\tilde{\xi}}(x_{qt});}$

5: (Return) $\boldsymbol{k \leftarrow k + 1;}$ Go to Step **2**

As the Regula Falsi and bisection, the Riders' has ensured convergence. In a unidimensional analysis, if the function is well represented by a third order Taylor's expansion, in (RIDDERS, 1979) the author establish that the rate of convergence is at least quadratic. Since in the present experiments frequently we are dealing with log-concave functions, it is expected a faster convergence than the bisection procedure. A drawback is the fact that for each iteration of the interpolation method it is necessary to compute two function solutions to compare them. If the function evaluation is not an enough cheap process, the use of the method may not be worthwhile.

A.6.4 Others Root Finding Methods

By the numerical simulations, we checked that the secant method, presented below, did not converge for some iterations, even though the underlying function was relatively well behaved. In the following scheme, we present some common root finding methods and limitations faced on their applications in the present context:

- Secant Method: In Regula Falsi and Ridder's methods, there is a stabilization step that selects the bracket where the change of the sign of the underlying function

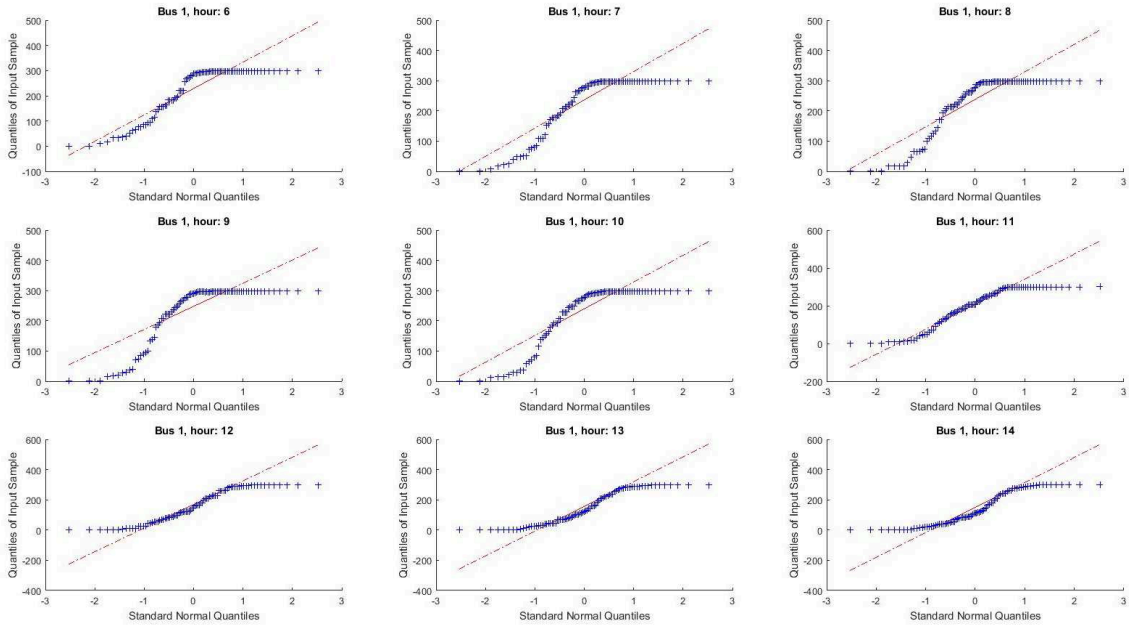
is verified. By this procedure, one is defining where the root is set, and so the search is reduced to a smaller domain. The secant method does not have this stabilization step and so eventually, depending on the behaviour of the function, the secant algorithm can select a portion of the domain where the root is not localized, reflecting in the non-convergence of the method. When applying the secant method in our model, numerical simulations show that convergence issues do occur for some iterations, inducing an infinite loop where the convergence test was not met, and so compromising the entire simulation.

- Newton Method: Since it is based on Taylor's expansion, it requires the calculation of gradients or partial derivatives of higher level (Jacobian matrix). As explained in Section 3.3, gradients calculation involves the evaluation of at most $\boldsymbol{T}$ similar functions of lower dimension of the underlying function. As the process to obtain the function value is relatively expensive (CDFs calculations), the use of gradients in the interpolation method would make it impractical.

- Muller's Method: Similarly to the secant method, rather than use a secant function over the underlying function to get the next iterates, it uses a quadratic equation to interpolate the function in **3** points. It was developed for one-dimensional function interpolation, so its use in the present framework would require additional developments to compute the required parameters for each iteration. Even though the method seems to be promising, these studies escape from the focus of the present work.

APPENDIX B – The Use of Alternative Distributions

Figure 21 – QQ plots from 6:00 to 14:59 of Bus 1.



Source: Thesis results.

The model to the wind generation random vector $\widetilde{\boldsymbol{gw}}_{\boldsymbol{w}}^{(\boldsymbol{b})}$ was defined in Section 4.2.2, where it is established normality assumptions to these variables. To motivate the analysis on the adherence of the used wind generation data set to hypothesis of normality, we offer sketches where we present quantile-quantile plots from the random variables. In Figure 21, we display nine QQ-plots from the random variables that represent the generation hour by hour from 6:00 to 14:59 of Bus 1 of the 5-bus system. We also reproduce the procedure to Bus 3 in Figure 22. The remaining hours have similar behaviour.

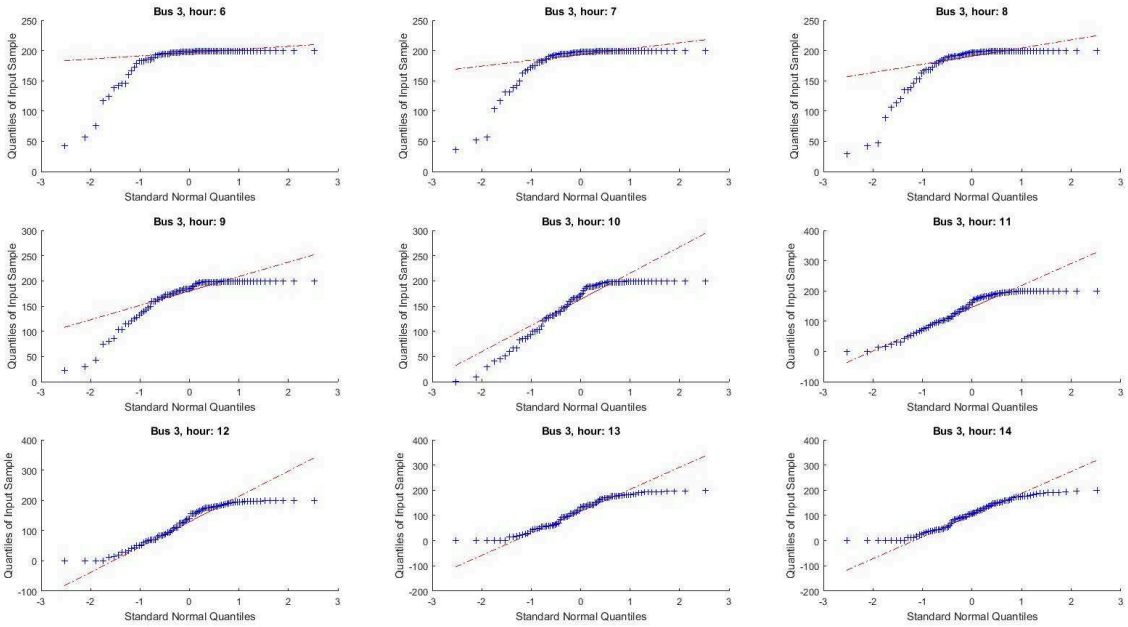
The interpretation of Figures 21 and 22 is that as better the blue signals (representing the quantiles of the sample) are supported by the red line (quantile of a standard Normal distribution), more we have evidence that we can not reject the normality hypothesis of the random vector $\widetilde{\boldsymbol{gw}}_{\boldsymbol{w}}^{(\boldsymbol{b})}$. In practice, one accepts the normality hypothesis if the blue signals are over the red line. The first observation of these figures is that the extremes of the blue signal shapes are limited by the maximum and minimum wind generation sources, and they are not well modelled by standard Gaussian approaches. A direct solution should be on the use of truncated distributions, what could be done imposing the wind random variable $\widetilde{\boldsymbol{gw}}_{\boldsymbol{w}}^{(\boldsymbol{b})}$ to be restricted to a polyhedral set, $\boldsymbol{b^{low}} \leq \boldsymbol{A_1}\widetilde{\boldsymbol{gw}}_{\boldsymbol{w}}^{(\boldsymbol{b})} \leq \boldsymbol{b^{up}}$, where $\boldsymbol{b^{low}}$ and $\boldsymbol{b^{up}}$ define the minimum and maximum limits to the wind farm generation and $\boldsymbol{A_1}$ is an identity matrix. In a such case, the PC (4.18) is modified as follows:

$$\mathbb{P}\left[\begin{array}{c} \boldsymbol{ld_b^{low}} \leq \boldsymbol{h}(\dot{\boldsymbol{x}}, \widetilde{\boldsymbol{gw}}_{\boldsymbol{w}}^{(\boldsymbol{b})}) \leq \boldsymbol{ld_b^{up}} \\ \boldsymbol{b^{low}} \leq \boldsymbol{A_1}\widetilde{\boldsymbol{gw}}_{\boldsymbol{w}}^{(\boldsymbol{b})} \leq \boldsymbol{b^{up}} \end{array}\right] \geq \boldsymbol{p} \cdot \mathbb{P}[\boldsymbol{b^{low}} \leq \boldsymbol{A_1}\widetilde{\boldsymbol{gw}}_{\boldsymbol{w}}^{(\boldsymbol{b})} \leq \boldsymbol{b^{up}}]. \tag{B.1}$$

As discussed in (DINIZ; HENRION, 2016), convexity for this PC still holds, but eventually, it would be necessary to work with singular distributions, and so in such case inclusion of technical procedures are required in order to achieve (sub)gradients for the non-smooth probabilistic constraints, as shown in (van ACKOOIJ; MINOUX, 2015).

In both Figures 21 and 22, checking that the red line supports the blue signals, we

Figure 22 – QQ plots from 6:00 to 14:59 of Bus 3.



Source: Thesis results.

may note that in the main partition of the distribution there are good adherences from hour 11 to hour 14 for both buses. Nevertheless, mainly for Bus 1, from 6:00 to 10:59, the wind did not behave so well, which could motivate the use of alternative distribution functions to describe them. In the context of defining alternative better distributions, we must observe some points, as follows:

1. The first one is regarding the concavity of the alternative distribution. This concern is to ensure that the PC defines a feasible convex set. Section 3.2 has more comments about this issue;

2. If it is well established its multivariate distribution function, and if it disposes of efficient methods to evaluate the CDF for a given point, as shown in Section A.5 to the Normal counterparts;

3. If the distribution disposes of appropriate formulas for its gradients. These are necessary, for example, to establish the cutting planes that bounds the PC feasible sets (Step vi of Algorithm 2);

4. If one aims at working with different distributions in a single joint PC, how to establish analytically the relationship between these random variables in a single CDF. One method that offers some flexibility in this sense is by the use of Copulas, as stated in Appendix Section B.2.

All those points are challenging lines of research in the stochastic programming community by themselves. The Section B.1 aims at providing a brief discussion on Copulas, that are important in the following Section B.2 that are related to the items **1** and **4** above regarding the use of alternative distributions to describe uncertainties in the PC programming problem.

B.1 NOTIONS ON COPULAS

This section provides a non self-contained introduction to Copulas. For a complete one we refer to (NELSEN, 2006). Here we follow the lines traced by (HENRION; STRUGAREK, 2011).

Definition B.1.1 *A copula is defined as a distribution function (CDF)* $\boldsymbol{C} : [\mathbf{0},\mathbf{1}]^{\boldsymbol{E}} \rightarrow [\mathbf{0},\mathbf{1}]$ *of a random vector that has its marginals uniformly distributed in* $[\mathbf{0},\mathbf{1}]$.

A fundamental concept in copulas theory is provided by the Sklar theorem, which states that for any CDF $\boldsymbol{F} : \mathbb{R}^{\boldsymbol{E}} \rightarrow [\mathbf{0},\mathbf{1}]$, with marginal components $\boldsymbol{F_e}, \boldsymbol{e} \in \boldsymbol{\mathcal{E}}, |\boldsymbol{\mathcal{E}}| = \boldsymbol{E}$, there exists a copula $\boldsymbol{C}$ for $\boldsymbol{F}$ such that:

$$\boldsymbol{F(x) = C(F_1(x_1), \ldots, F_E(x_E)), \forall x \in \mathbb{R}^E.} \tag{B.2}$$

In the case of a continuous CDF $\boldsymbol{F_e}$, then the copula $\boldsymbol{C}$ is uniquely defined:

$$\boldsymbol{C(u) = F(F_1^{-1}(u_1), \ldots, F_E^{-1}(u_E)); \quad F_e^{-1}(t) := \inf_{F_e(r) \geq t} r.} \tag{B.3}$$

This theorem gives us the instruments to define a distribution function by its copula counterpart, adjusting it to fit to a given distribution. Examples of copula:

- Independent Copula: $\boldsymbol{C(u) = \prod_{i=1}^{E} u_i}$;
- Maximum Copula: $\boldsymbol{C(u) = \min_{e \in \mathcal{E}} \; u_e}$;
- Normal Copula: $\boldsymbol{C^{\Sigma}(u) = \Phi^{\Sigma}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_E))}$, where $\boldsymbol{\Phi^{\Sigma}}$ is the normal CDF with zero mean and correlation matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{\Phi}$ is the one-dimensional standard Gaussian CDF;
- Archimedean Copula: $\boldsymbol{C(u) = \psi^{-1}(\sum_{e=1}^{E} \psi(u_e))}$, where $\boldsymbol{\psi}$ is the *generator of the Archimedean Copula*, a continuous decreasing function $\boldsymbol{\psi : [0,1] \rightarrow \mathbb{R}^{+}}$, with $\boldsymbol{\psi(1) = 0}$.

The independent Copula is a joint CDF with independent components. An interesting property of the Maximum Copula is that it dominates any other copula C, in such a way that $\boldsymbol{C(u) \leq \min_{e \in \mathcal{E}} \; u_e}$. Special cases of Archimedean Copulas are: The independent copula that has generator $\boldsymbol{\psi = -log(t)}$; The *Clayton* copula that has generator $\boldsymbol{\psi(t) = \theta^{-1}(t^{-\theta} - 1)}$, with $\boldsymbol{\theta > 0}$; The *Gambel* copula that has generator $\boldsymbol{\psi(t) = (-log(t))^{\theta}}$, with $\boldsymbol{\theta \geq 1}$.

B.2 GENERALIZED CONCAVITY AND EVENTUAL CONVEXITY

As discussed in Chapter 3, especially in Section 3.2, to ensure that the PC, e.g. $\mathbb{P}[\boldsymbol{h_e(\dot{x}, \tilde{\xi}) \leq 0, e \in \mathcal{E}}] \geq \boldsymbol{p}$, defines a convex feasible set, one must to analyse the concavity of the map $\boldsymbol{h_e(\cdot)}$ and the distribution of the random variable $\tilde{\boldsymbol{\xi}}$. As far the research of the author goes, the pioneering work (HENRION; STRUGAREK, 2008) defines conditions to ensure concavity for specific PDFs with independent random components, where concavity holds for large enough probability levels. In their paper, it is considered

separated structures as the particular case $\boldsymbol{h(\dot{x}, \tilde{\xi}) = h(\dot{x}) - \tilde{\xi}}$, allowing the feasible set of PC to be described directly from the distribution function of $\tilde{\boldsymbol{\xi}}$, such that, $\boldsymbol{M(p) = \{\dot{x} \in \mathbb{R}^m | F(h(x)) \geq p\}}$. The authors look at finding the properties that make possible ensuring convexity of the set $\boldsymbol{M(p)}$ for all $\boldsymbol{p \leq p^*}$, with $\boldsymbol{p^* < 1}$. The question imposed here is how to establish the convexity of the set $\boldsymbol{M(p)}$ when we can not guarantee the concavity of the components $\boldsymbol{h_e(\dot{x})}$ and the variable $\tilde{\boldsymbol{\xi}}$ distribution function.

Still in (HENRION; STRUGAREK, 2008), relating to general $\boldsymbol{r-}$concavity Definitions A.2.1, $\boldsymbol{r \in \mathbb{R}}$, the authors define the $\boldsymbol{r-}$*decreasing functions* as the continuous functions $\boldsymbol{f : \mathbb{R} \to \mathbb{R}}$ that have $\boldsymbol{t^r f(t)}$ strictly decreasing for all $\boldsymbol{t > t^*}$, for some $\boldsymbol{t^* > 0}$. The central result of the work is following theorem:

Theorem B.2.1 *Considering* $\boldsymbol{M(p)}$ *as in* (3.12)*, for* $\boldsymbol{e \in \mathcal{E}}$*,* $\boldsymbol{F_e}$ *the CDF of* $\tilde{\boldsymbol{\xi}}_{\boldsymbol{e}}$*,* $\boldsymbol{t_e^*}$ *the inferior limit of* $\boldsymbol{(r_e + 1)}$*-decreasing PDF* $\boldsymbol{f_e}$ *and the additional assumptions:*

1. $\boldsymbol{\exists r_e > 0}$ *such that components* $\boldsymbol{h_e}$ *are (-*$\boldsymbol{r_e}$*)-concave.*
2. *The components* $\tilde{\boldsymbol{\xi}}_{\boldsymbol{e}}$ *of* $\tilde{\boldsymbol{\xi}}$ *are independently distributed with* $\boldsymbol{(r_e + 1)}$*-decreasing densities* $\boldsymbol{f_e}$*.*

Then, $\boldsymbol{M(p)}$ *is convex for all* $\boldsymbol{p > p^* := \max\{F_e(t_e^*) | e \in \mathcal{E}\}}$*.*

This theorem is based on the relationship of the $\boldsymbol{(r + 1)}$-decreasing density $\boldsymbol{f_e}$ to the concavity of the mapping $\boldsymbol{F(z^{-1/r})}$ for $\boldsymbol{z \in (0, (t^*)^{-r})}$. The property of allowing convexity to the set $\boldsymbol{M(p)}$ depending on the level of the probability level $\boldsymbol{p}$ is called *eventual convexity* in the literature. Among distributions with independent components that preserve r-decreasing densities (with correspondent $\boldsymbol{t^*}$ values), we find Normal, Exponential, Gamma, $\boldsymbol{\chi}$ and $\boldsymbol{\chi^2}$. Moreover, the authors verify that Weibull PDF, very commonly found to model the speed of wind in wind farms generation studies, belongs to this class of functions.

The same authors of the previous work extend it in (HENRION; STRUGAREK, 2011) providing conditions to establish concavity properties when instead of imposing random variables with independent components it is allowed correlated structures through the use of copulas. One advantage in the use of copulas is the possibility to set correlation between random variables based on their marginal distributions. Then, it is discussed the convexity of the set:

$$\boldsymbol{M(p) = \{x \in \mathbb{R}^E | C(F_1(h_1(x)), \ldots, F_E(h_E(x))) \geq p\},} \tag{B.4}$$

where $\boldsymbol{h_e : \mathbb{R}^E \to \mathbb{R}}$, $\boldsymbol{F_e}$ is a one dimensional marginal CDF of the distribution function $\boldsymbol{F}$, and $\boldsymbol{C}$ is a Copula as in Section B.1. The use of (B.4) is quite flexible and for general applications $\boldsymbol{F_e}$ it is known exactly. $\boldsymbol{C}$ can be a Copula that well establish, heuristically, the relationship between the marginals. To present the main result of the paper, the authors introduce the following definition:

Definition B.2.2 *A Copula* $\boldsymbol{C : [0, 1]^E \to [0, 1]}$ *is called* $\boldsymbol{\log \exp}$*-concave on the set* $\boldsymbol{\prod_{e=1}^E [q_e, 1)}$*, for* $\boldsymbol{q \in (0, 1)^E}$ *if the mapping* $\boldsymbol{\tilde{C} : \mathbb{R}^E \to \mathbb{R}}$ *such that* $\boldsymbol{\tilde{C}(u) = \log C(e^{u_1}, \ldots, e^{u_E})}$ *is concave on* $\boldsymbol{\prod_{e=1}^E [\log q_e, 0)}$*.*

Borrowing the definitions from Theorem B.2.1, the main result of (HENRION; STRUGAREK, 2011) is given by the following theorem:

Theorem B.2.3 *Considering* (B.4), *for* $\boldsymbol{e = 1, \ldots, E}$, *we assume that:*

1.$\exists \boldsymbol{r_e > 0}$ *such that components* $\boldsymbol{h_e}$ *are (-*$\boldsymbol{r_e}$*)-concave.*

2.The marginal CDFs $\boldsymbol{F_e}$ *are related to* $\boldsymbol{(r_e + 1)}$*-decreasing densities* $\boldsymbol{f_e}$.

3.The Copula $\boldsymbol{C}$ *is log exp-concave on the set* $\prod_{e=1}^{E}[F_e(t_e^*), 1)$.

Then, $\boldsymbol{M(p)}$ *is convex for all* $\boldsymbol{p > p^* := \max\{F_e(t_e^*)|e \in \mathcal{E}\}}$.

For the application of the Theorem B.2.3, the first two assumptions are directly verified from the maps $\boldsymbol{h_e(\cdot)}$ and density $\boldsymbol{f_e}$, where many examples for $\boldsymbol{(r_e+1)}$-decreasing densities $\boldsymbol{f_e}$ are provided in Table 1 of (HENRION; STRUGAREK, 2008). Less trivial is to ensure the third assumption, i.e. Copula $\boldsymbol{C}$ being $\boldsymbol{\log \exp}$-concave. In (HENRION; STRUGAREK, 2011) it is checked that the independent, maximum and Gambel Copula are $\boldsymbol{\log \exp}$-concave. Nevertheless, the Clayton Copula is not.

In the same line of research, in (van ACKOOIJ, 2015) the author extends the Theorem B.2.3 for a larger family of Copulas $\boldsymbol{C}$. The extension of the $\boldsymbol{\log - \exp}$-concave Copula C is provided by the concept of $\boldsymbol{\delta - \gamma}-$concave Copula. This last is defined as follows:

Definition B.2.4 *Let* $\boldsymbol{\gamma \in \mathbb{R}}$ *and the set* $\boldsymbol{X(\gamma)}$ *defined as* $\boldsymbol{X(\gamma) = [0,1]^E}$ *for* $\boldsymbol{\gamma > 0}$, $\boldsymbol{X(0) = (-\infty, 0]^E}$ *and* $\boldsymbol{X(\gamma) = [1, \infty)^E}$ *for* $\boldsymbol{\gamma < 0}$. *Let* $\boldsymbol{\sigma \in [-\infty, \infty]}$ *be equally given. We define the Copula* $\boldsymbol{C[0,1]^E \to [0,1]}$ *a* $\boldsymbol{\gamma - \delta}$*-concave Copula, if :*

- *the mapping* $\boldsymbol{C(u^{1/\gamma})}$ *is* $\boldsymbol{\delta}$*-concave , if* $\boldsymbol{u \in X(\gamma)}$ *and* $\boldsymbol{\gamma \neq 0}$;
- *the mapping* $\boldsymbol{C(e^u)}$ *is* $\boldsymbol{\delta}$*-concave , if* $\boldsymbol{u \in X(0)}$ *and* $\boldsymbol{\gamma = 0}$.

By this last definition, we verify that the $\boldsymbol{\log - \exp}$-concave definition in B.2.2 is in effect a $\boldsymbol{0 - 0}$-concave mapping. Still in (van ACKOOIJ, 2015), the author extendeds this definition by the concept of $\boldsymbol{\delta}$-$\boldsymbol{\gamma}$ -$\boldsymbol{q}$-concave Copula $\boldsymbol{C}$, where the $\boldsymbol{q}$ turns out to be the parameter that defines the set $\boldsymbol{X(q, \gamma)}$ where $\boldsymbol{u}$ takes place. Then, the Definition B.2.4 has its sets redefined, such that: $\boldsymbol{X(q,\gamma) = \prod_{e \in \mathcal{E}}[q_e^\gamma, 1]}$ for $\boldsymbol{\gamma > 0}$, $\boldsymbol{X(q, 0) = \prod_{e \in \mathcal{E}}[\log(q_e), 0]}$ and $\boldsymbol{X(q,\gamma) = \prod_{e \in \mathcal{E}}[1, q_e^\gamma]}$ for $\boldsymbol{\gamma < 0}$. The main result of (van ACKOOIJ, 2015) is given by the following theorem, which is indeed an extension of Theorem B.2.3:

Theorem B.2.5 *Considering* (B.4), *for* $\boldsymbol{e = 1, \ldots, E}$, *we assume that:*

1.$\exists \boldsymbol{r_e > 0}$ *such that components* $\boldsymbol{h_e}$ *are* $\boldsymbol{r_e}$*-concave;*

2.The marginal CDFs $\boldsymbol{F_e}$ *are* $\boldsymbol{\gamma_e}-$*concave, for* $\boldsymbol{\gamma_e \in (-\infty, \infty], b_e > 0}$ *on the following sets: For* $\boldsymbol{r_e < 0}$ *and* $\boldsymbol{(0, b_e^{r_e}] \ni z \mapsto F_e(z^{1/r_e})}$*; For* $\boldsymbol{r_e = 0}$ *and* $\boldsymbol{[\log b_e, \infty) \ni z \mapsto F_e(\exp z)}$*; For* $\boldsymbol{r_e > 0}$ *and* $\boldsymbol{[b_e^{r_e}, \infty) \ni z \mapsto F_e(z^{1/r_e})}$*;*

3.The Copula $\boldsymbol{C}$ *is* $\boldsymbol{\delta - \gamma}$*-concave or* $\boldsymbol{\delta - \gamma - F(b)}$*-concave for* $\boldsymbol{\gamma \leq \gamma_e \leq \infty}$.

Then, $\boldsymbol{M(p)}$ *is convex for all* $\boldsymbol{p > p^* := \max\{F_e(b_e)|e \in \mathcal{E}\}}$.

In this theorem proof is stated that in the case which $\boldsymbol{r_e \geq 0}$, $\boldsymbol{F_e}$ being $\boldsymbol{\gamma_e}$-concave everywhere (and so it is not limited for some $\boldsymbol{b_i}$) and $\boldsymbol{C}$ being a $\boldsymbol{\delta}-\boldsymbol{\gamma}-$concave Copula, the set $\boldsymbol{M(p)}$ is convex for all $\boldsymbol{p \in [0,1]}$. Using the monotonicity of the Copula, one achieves a even sharpen (i.e. lower) $\boldsymbol{p^* := C(F_1(b_1), \ldots, F_n(b_n))}$. Complementing (HENRION; STRUGAREK, 2011) that has shown that the Clayton Copula is not $\boldsymbol{\log - \exp}$-concave, in (van ACKOOIJ, 2015) it is shown that this Copula is indeed $\boldsymbol{\gamma - \delta}$-concave for all $\boldsymbol{\gamma > 0}$ and $\boldsymbol{\delta \leq -\theta < 0}$.

APPENDIX C – Support to bilinear structures

C.1 EIGENVALUES OF SQUARE MATRIX

To establish the eigenvalues of the matrix $\boldsymbol{YY}^\top - \boldsymbol{ZZ}^\top$, we first observe all the eigenvalues and eigenvectors of the matrix $\boldsymbol{YY}^\top$. To achieve that we verify:

$$\begin{aligned}\boldsymbol{YY}^\top\boldsymbol{a} = \boldsymbol{\lambda a} &\Leftrightarrow \boldsymbol{Y}^\top\boldsymbol{YY}^\top\boldsymbol{a} = \boldsymbol{Y}^\top\boldsymbol{\lambda a} \\ &\Leftrightarrow \|\boldsymbol{Y}\|^2\boldsymbol{Y}^\top\boldsymbol{a} = \boldsymbol{Y}^\top\boldsymbol{\lambda a} \Leftrightarrow \boldsymbol{\lambda} = \|\boldsymbol{Y}\|^2.\end{aligned}$$

And so one eigenvalue of $\boldsymbol{YY}^\top$ is $\|\boldsymbol{Y}\|^2$. Verifying that:

$$\boldsymbol{YY}^\top\boldsymbol{a} = \boldsymbol{\lambda a} \Leftrightarrow \boldsymbol{YY}^\top\boldsymbol{Y} = \boldsymbol{\lambda Y} \Leftrightarrow \boldsymbol{Y}\|\boldsymbol{Y}\|^2 = \boldsymbol{\lambda Y}.$$

We note that the eigenvector related to the eigenvalue $\|\boldsymbol{Y}\|^2$ is the vector $\boldsymbol{Y}$. Also noting that considering an orthogonal vector $\boldsymbol{b}$ to $\boldsymbol{Y}$, we have $\boldsymbol{Y}^\top\boldsymbol{b} = \boldsymbol{0}$ and so $\boldsymbol{YY}^\top\boldsymbol{b} = \boldsymbol{\lambda b} = \boldsymbol{0}$ and then orthogonal vectors are eigenvectors with trivial eigenvalues $\boldsymbol{\lambda} = \boldsymbol{0}$. If $\boldsymbol{YY}^\top$ is a $\boldsymbol{n} \times \boldsymbol{n}$ matrix, finding $\boldsymbol{n-1}$ orthogonal vectors to $\boldsymbol{Y}$ we define a base to $\boldsymbol{YY}^\top$ and so the only non-trivial eigenvector is $\boldsymbol{Y}$ (and their parallel correspondents). Equivalent discussion is made for the matrix $\boldsymbol{ZZ}^\top$. From that, we observe that non-trivial eigenvalues of the matrix $\boldsymbol{YY}^\top - \boldsymbol{ZZ}^\top$ are given on the vector space defined by the bases made of the vectors $\boldsymbol{Y}$ and $\boldsymbol{Z}$. So, we have the following equivalence:

$$(\boldsymbol{YY}^\top - \boldsymbol{ZZ}^\top)(\boldsymbol{aY} + \boldsymbol{bZ}) = \boldsymbol{\lambda}(\boldsymbol{aY} + \boldsymbol{bZ}).$$

Opening $(\boldsymbol{YY}^\top - \boldsymbol{ZZ}^\top)(\boldsymbol{aY} + \boldsymbol{bZ})$, and being the scalars $\boldsymbol{Y}^\top\boldsymbol{Z} = \boldsymbol{Z}^\top\boldsymbol{Y}$:

$$\begin{aligned}(\boldsymbol{YY}^\top - \boldsymbol{ZZ}^\top)(\boldsymbol{aY} + \boldsymbol{bZ}) &= \boldsymbol{aYY}^\top\boldsymbol{Y} + \boldsymbol{bYY}^\top\boldsymbol{Z} - \boldsymbol{aZZ}^\top\boldsymbol{Y} - \boldsymbol{bZZ}^\top\boldsymbol{Z} \\ &= \|\boldsymbol{Y}\|^2\boldsymbol{aY} + \boldsymbol{bZ}^\top\boldsymbol{YY} - \boldsymbol{aZ}^\top\boldsymbol{YZ} - \|\boldsymbol{Z}\|^2\boldsymbol{bZ}, \\ (\boldsymbol{YY}^\top - \boldsymbol{ZZ}^\top)(\boldsymbol{aY} + \boldsymbol{bZ}) &= (\|\boldsymbol{Y}\|^2\boldsymbol{a} + \boldsymbol{Z}^\top\boldsymbol{Yb})\boldsymbol{Y} + (-\|\boldsymbol{Z}\|^2\boldsymbol{b} - \boldsymbol{Z}^\top\boldsymbol{Ya})\boldsymbol{Z} \\ &= \boldsymbol{\lambda}(\boldsymbol{aY} + \boldsymbol{bZ}).\end{aligned}$$

The last equality is equivalent to the following system of equalities:

$$\begin{aligned}\|\boldsymbol{Y}\|^2\boldsymbol{a} + \boldsymbol{Z}^\top\boldsymbol{Yb} &= \boldsymbol{\lambda a}, \\ -\|\boldsymbol{Z}\|^2\boldsymbol{b} - \boldsymbol{Z}^\top\boldsymbol{Ya} &= \boldsymbol{\lambda b}.\end{aligned}$$

That can be rewritten in following matrix:

$$\begin{pmatrix} \|\boldsymbol{Y}\|^2 & \boldsymbol{Z}^\top\boldsymbol{Y} \\ -\boldsymbol{Z}^\top\boldsymbol{Y} & -\|\boldsymbol{Z}\|^2 \end{pmatrix} \cdot [\boldsymbol{a} \quad \boldsymbol{b}]^\top = \boldsymbol{\lambda}[\boldsymbol{a} \quad \boldsymbol{b}]^\top. \tag{C.1}$$

We are in grade to verify the eigenvalues of that matrix. We observe that the characteristic polynomial is given by:

$$\boldsymbol{\lambda}^2 + \boldsymbol{\lambda}(\|\boldsymbol{Z}\|^2 - \|\boldsymbol{Y}\|^2) - \|\boldsymbol{Y}\|^2\|\boldsymbol{Z}\|^2 + (\boldsymbol{Z}^\top\boldsymbol{Y})^2 = \boldsymbol{0}$$

From the Bhaskara formula we find the highest eigenvalue is:

$$\boldsymbol{\lambda^+ = \frac{\|Y\|^2 - \|Z\|^2 + \sqrt{(\|Y\|^2 + \|Z\|^2)^2 - 4(Z^\top Y)^2}}{2}},$$

and the lowest:

$$\boldsymbol{\lambda^- = \frac{\|Y\|^2 - \|Z\|^2 - \sqrt{(\|Y\|^2 + \|Z\|^2)^2 - 4(Z^\top Y)^2}}{2}}.$$