Integrating uncertainties in short-term operational planning

Paul Javal

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à MINES ParisTech

Prise en compte des incertitudes dans la gestion prévisionnelle court-terme
Integrating uncertainties in short-term operational planning

Soutenue par
Paul JAVAL
Le 29 Novembre 2021

École doctorale n°84
ED STIC

Spécialité
Contrôle, Optimisation, Prospective

Composition du jury :

Andrew Keane Professeur, University College Dublin	*Rapporteur*
Tito Homem-de-Mello Professeur, Universite Adolfo Ibanez	*Rapporteur*
Annabella Astorino Chercheuse, ICAR-CNR	*Examinatrice*
Gilles Guerassimoff Professeur, École Nationale Supérieure des Mines de Paris	*Examinateur*
Nouredine Hadj-Said Professeur, Grenoble INP	*Président du jury*
Sophie Demassey Maître-Assistante, École Nationale Supérieure des Mines de Paris	*Directrice de thèse*
Welington Luis de Oliveira Chargé de Recherche, École Nationale Supérieure des Mines de Paris	*Directeur de thèse*

MINES ParisTech | PSL

ACKNOWLEDGEMENTS AND GRATITUDE

The acknowledgements section is very peculiar in that it is the only readable and understandable one to any anglophone reader in a Thesis, while having close to no rule to follow - except that of being produced on a support of comparable size to neighbouring pages. This curiosity naturally leads to a myriad of interesting and non-trivial questions. If it strikes you as being strange that I shall be asking myself lots of questions, then we probably do not know each other very well. In that case, please note I cannot be held responsible for the possibly low interest you may find in this section.

Now, looking back on PhD years, it turns out many people have participated in this life experience that is greater than plain "(maths) studies". Some might not know that a single discussion over a year ago would have me think about them when gathering my thoughts as I am writing down this section. I might have not even met with others. Some might guess I am grateful to them without knowing how much, starting with Sophie and Welington in this subsection. Aside from the inspiring scientific rigour and crisp thinking, I am deeply thankful for their ironclad calmness, patience and kindness. This is not only true in absolute terms, but also in "relative terms" as it remains true when hearing PhD discussions about supervisors that each and every student comes to hear throughout the years - which usually turn out to be cliché and cynical. Hugo and Bhargav, the other part of the supervision team at EDF, share the same qualities. I am sincerely thankful to them for their endless patience, and look up their ability of making complex projects work.

Once a first acceptable version of the manuscript was finished, it has then to be reviewed. For this substantial work, as well as for their kind words and interesting discussion, I am very thankful to members of my jury. Andrew, Tito, Annabella, Gilles and Nouredine made my last step in the PhD a pleasant one, and I shall keep some particular feelings for a long time. Their acceptance to be part of my jury is awe-inspiring.

This work was set up, managed and financed thanks to EDF R&D, a company where academic research is possible. Moreover, as a significant amount of researchers have gone through academic research in their life, they understand what is at stake in these years. I am very grateful to Wim for his explanations of a whole new field to me and admiring of his work which (unfortunately) took me years to grasp. Mélaine and Gauthier, as well as Matthieu had the unattractive task of long-term monitoring an extending PhD candidate, and I am much obliged for their work.

Teams in both Sophia Antipolis and Palaiseau had significant impacts on the improvement of daily work. Starting with Palaiseau and Charles, a gem to share the office with, whose incredible dad jokes (sometimes for kids over the age of majority) have finally made him one. Stepping out into the corridor, I am thankful to Arnaud , Vojislav, Keddy, Clément and his flourishing ideas, Pierre and his other flourishing ideas, Dominique who was patient enough to answer my weekly childish questions, Naji for his perseverance dealing with the OpenData, Thierry, Jules for his (relevant) situation analysis, Beatrice for her efficiency and straightforward guidance on our administrative tasks. I am sure

that the (incredibly) growing PhD team of MIRE, which went from 1 to (at least) 9, will be a great add to the whole department ! Moving to Sophia-Antipolis, I've had great pleasure in getting to meet and talk with the team: Edi, Gilles, Jean-Paul - a true "artist", just slight brushes of his and your problem goes from a blackbox to a path - Valentina, Valerie (students' shield against Adum), Catherine who had such an impact on CMA that you can still hear her accent, Alice who's totally living up to taking this role. Moreover, when visited by more philosophical questions as days were coming to an end, I am grateful to the other PhD students for seemingly endless discussions: it started with Gratien, Rémi, Ariane, Arnold, Mathieu and Gildas, and went on thanks to Mathilde, Carlos, Cindy, Lucas, Victor, Zixhuan, Rabab, Naïma, Yacine, Louis and Amir.

Despite my slow-paced answering manner when it comes to long-distance exchanges, I am very grateful to my friends that kept on calling/sending news. In a disordered enumeration, I am looking forward to getting back with Romain, Adrien, Thomas, Yo, MaPiotte, my personal research group, the once called groupe Brosse, the tennis team, Kantorovitch, ... I am also looking forward to getting to discover even more my "newer" friends in Florac.

Among people that probably have no idea I am grateful to them is Line Roald for what I interpreted as a very positive answer to one of my mail while battling in the middle of code swirls. Alain Damasio, whose work-mania quite interestingly visible in Bora Vocal seems to resemble mine, had a significant impact - a 50 year old man struggling to make the world different to be seen, struggling to bring the most people with him in almost a Tragic manner: thanks to him, now I believe the set of possible future can be expanded. France Culture for the best introductory material to many fascinating subjects, Tom Gauld, Le journal de la joie de vivre, Sci-Hub, CyclingAbout who made me travel by bike in my dreams...

Finally, as the embodiment of an inspirational freedom I cherish, while her name is freed from this page it should still be seen throughout this work.

"A fortiori, si vous évoquez la probabilité de la disparition de l'humanité dans les prochaines décades (*trois milliards d'* [humains], *trois milliards d'années d'évolution biologique . . .*) *: c'est trop énorme pour être vraiment concevable, c'est une abstraction. Absolument nul comme contenu émotif. Donc impossible à prendre au sérieux. On lutte pour des augmentations de salaires, pour la liberté de parole et la liberté de presse, la sécurité de l'emploi, contre la sélection à l'Université, contre la bourgeoisie, l'alcoolisme, la peine de mort, le cancer, le racisme, à la rigueur contre la guerre au Vietnam ou contre la guerre tout court. Mais l'annihilation de la vie sur la terre dépasse l'entendement de chacun de nous, c'est un "irréalisable" ; comment inciterait-il à l'action ?"*

Alexandre Grothendieck, *Responsabilité du savant dans le monde d'aujourd'hui*, 1970

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

In this work, we are interested in several different fields related to applied mathematics and the electrical grid. Attempts to build a bridge between parts of these two are fruitful in terms of results and new discoveries; on the other hand, such a combination of two, seemingly ever-expanding, research universes is a humbling exercise. We would like to contribute to this bridging effort by exploring a peculiar, yet key to the electrical grid, optimisation problem referred to as the *Optimal Power Flow*. The motivation stems from the evolving nature of electrical distribution in France on one side, and the thriving numerical methods for general optimisation problems on the other side.

The key ingredients to our work are the following three:

- the understanding and subsequent mathematical modelling of this Optimal Power Flow from a distribution point of view;
- the study of uncertainties in an optimisation context, more particularly the "Chance-constrained" one;
- the development and implementation of a tailored algorithm that is capable to tackle general optimisation problems.

A part of this thesis is dedicated to each of these items. The fourth part concludes our bridging attempt. As such, the first three parts can be taken independently from one another, even though we maintain consistency in the notations.

In the first part, we are interested in the electrical system and the necessary elements to grasp before moving to a mathematical model. For a reader coming from a more mathematical background, this could be of use if insight is needed on the material side and existing mathematical models. Without falling into endless turf wars, the electrical network is among the largest human-made systems. It requires a stringent continuous and endless management. It is both the output of tremendous historical investments and, *de facto*, the source of a vast majority of our activities. It has become a tool for, and a domain where, part of the necessary evolutions to reach the international environmental agreements will be held. This unique position of the electrical system, a continuum from the international interfaces as well as in virtually all remote villages, sheds light on to why so many evolutions are on the way and more are foreseen. We are interested in a particular element of this positioning that concentrates a significant amount of these changes, which is the distribution system. After a brief historical overview of its development in France, we discuss the actors in the deregulated electricity system interacting with one another through the electrical distribution grid. With continuing deregulation, the changes in the distribution grid are numerous, the number of actors ever increasing, and the roles and challenges facing existing actors evolving significantly. In such a context, we define the challenges from a Distributor's point of view as well as the criterion that discriminates what is a "good" solution to these challenges.

Even with only a glimpse into the first part, it rapidly becomes evident that we have to face an optimisation problem with uncertainties. Adding a slight overview of optimisation related to electrical grids leads us to a better characterization of the mathematical

object we have to deal with: without approximations, we have a nonconvex nonsmooth optimisation problem. Nonconvex because of the electrical grid model, nonsmooth because of the uncertainties. This guides us to the second part, which provides a look into DoC (Difference-of-Convex) programming. The DoC framework is a wide and flexible one, providing useful structural properties for functions that are DoC. It includes all linear, convex and sufficiently (sub)smooth functions. According to some precursors of DoC programming, virtually all engineering optimisation problems can be cast, or approximated with a user-defined precision, as a DoC one (Indeed, every extended real-valued lower-semicontinuous function can be approximated by not only DoC, but actually piecewise affine DoC of the kind max-max [256]). This motivates the selection of this framework to model our problem as defined in the first part. Observing the recent progress in numerical methods for nonconvex nonsmooth optimisation (namely, so-called *bundle methods*) we propose a new algorithm we refer to as PBMDC^2 for Proximal Bundle Method for DoC-constrained DoC-problems. PBMDC^2 builds upon recent literature on DoC programming, and enables one to solve very general problems with, in particular, "difficult" constraints. Such problems would previously have been tackled by penalization techniques that are deemed sensible to user parametrisation, the CCP (Convex-Concave Procedure) (see [197]) or an extension of the DCA (Difference-of-Convex Algorithm) (see [183]) which are supposedly less efficient numerically when compared to bundle methods. Let us clarify that our DoC approach is not included in the global optimisation scheme as defined in [285] [1]. Although we indeed provide a numerically sound nonconvex optimisation algorithm with an optimality certificate for the achieved point, we do not investigate its possible global optimality property. This choice is motivated by our operational objective, with a limited amount of time and resources, and justifies that we leave out global optimisation algorithms, as branch-and-bound. PBMDC^2 proves to be efficient in practise, and provides an alternative to penalization techniques or approximations of DoC-constrained DoC problems.

Thanks to PBMDC^2, we are able to numerically solve a broad class of nonsmooth nonconvex optimisation problems. Now appears another necessary step, namely to cast our model with uncertainty into a DoC-constrained DoC problem. Optimisation under uncertainties also is a hot topic in applied mathematics. Several frameworks to model these uncertainties co-exist, and after a brief overview we discuss one of them in detail: chance-constrained programming. In this framework, uncertainties are modelled with probabilistic functions which constitute a vast and quickly expanding mathematical subfield. Even if it is often applied to the Energy field, it cannot be reduced to it, which leads us to a somewhat general study of chance-constraints. As a matter of fact, several interesting questions appear at the crossroads of DoC and chance-constrained programming, the former being a reasonable candidate framework in which to exploit the latter which is in general nonconvex and nonsmooth. In the third part, we address some of these questions. Some are more theoretical in a way, as the study of probabilistic functions' continuity and variations under appropriate and relatively general assumptions, while others are more practical, as how a chance-constraint can be approximated by a DoC function. The take-away from this part is twofold: a variational approach of probabilistic functions, and the missing link between uncertainty modelling and a DoC model for the

[1] "Global optimization is concerned with finding global solutions to nonconvex optimization problems", Preface to the First Edition

OPF (Optimal Power Flow).

Summing our developments from the first three parts, we present in the fourth one a DoC form of an OPF with uncertainties modelled as chance-constraints. Exhibiting a DoC formulation for a function is not trivial in general, and we thus have to prove the validity of the considered DoC decompositions. We also address the question of data: due to privacy and intellectual property-related concerns, one cannot easily have access to real data. Moreover, from an operational planning perspective, an interesting case would be one with scenarios that forecast electrical constraints. How do we obtain realistic data and forecasts, which realisation would lead to a constrained electrical setting? Enedis OpenData initiative that started in 2015, answers part of this question: it is now possible to retrieve real data, averaged and processed in order to comply to French and European privacy-related laws. In particular, forecasts of production and consumption, as well as repartitions of producers and loads on a given legal area are accessible. We then tackle the second part of this question by proposing a heuristic whose input is the data sourced from Enedis Open Data and whose output is a readily usable for our operational planning problem. This dataset is comprised of loads and producers, scenarios of forecasts, all defined according to distributions extracted from the Enedis OpenData. We then present results on these realistically set-up electrical grids of different DoC-tailored algorithms.

The main objective of this work is to provide a brief overview of an energy related optimisation problem, from an industrial ambition then on to mathematical developments necessary to present a solution before its implementation. The variety of subjects at hand in this process also implies that there is a similarly vast variety of open questions. The final chapter is dedicated to a summary of the major questions that remain to be addressed and future challenges.

INSERT 1: Note on some concept-abuses.
Throughout this work, several words will be used interchangeably for simplicity even though, depending on the chosen point of view, they are not strictly equivalent. For instance *consumer* and *load* are used to the same end though the latter emphasizes the "technical" side of the former. In fact a consumer is a load with a consumption contract. Here a list of similar concept-abuses used that we believe do not weaken the understanding:

- *producer - generator*
- *grid - network*
- *nodes - buses*
- *arcs - lines*

From now on, when unspecified, the "electrical grid", "grid" or "network" refers to the electrical distribution grid which is the central element of our study.
In Parts I and IV we use the name of the class to refer to an instance of this class. For example, the Optimal Power Flow is a class of problems but, when unambiguous, we write "solving the OPF" in place of "solving an instance of the OPF".

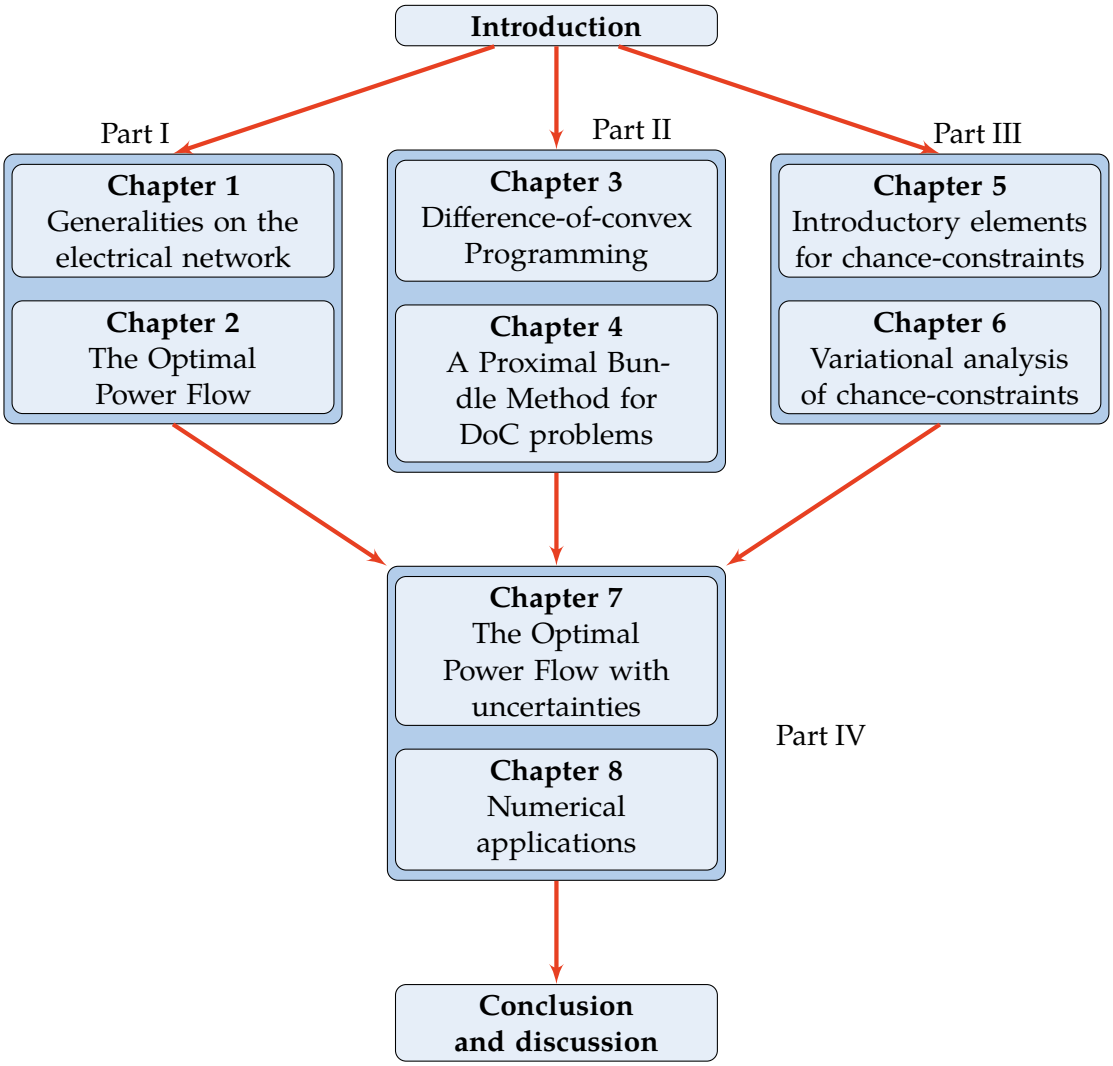


FIGURE 1: Organisation of the thesis.

Remark 1. Difference-of-convex functions/programming are known in the literature as DC functions/programming. However, in optimal power flow problems, DC has a different meaning: Direct Current. For this reason, throughout this thesis, we abbreviate "difference-of-convex" by DoC and reserve DC for Direct Current. ▷

CONTRIBUTION OF THE THESIS

The contributions of this thesis can be classified into three distinct categories. Parts II to IV are each dedicated to a type of contribution.

DIFFERENCE-OF-CONVEX PROGRAMMING - PART II

In the Optimisation field, we have explored the subfield of Difference-of-Convex (DoC) programming which is a class of optimisation problems. This class combines *generality* as a significant amount of real-life continuous problems belong to it, and *strong structural*

properties which can be leveraged to design efficient algorithms. This combination of the two aspects of DoC programming is at the centre of recent works in this field. Our contribution in this domain are the following:

1. We propose a novel algorithm named PBMDC^2 for a *generalized*-DoC programs, whereas a majority of previous works considered special cases of DoC programs. This has been published in [9], and is detailed in Part II.
2. We prove that PBMDC^2 is efficient on six different academic problems, two deterministic and four stochastic.

Probabilistic functions - Part III

Uncertainties are an important subfield in Optimisation, as they naturally occur in every physical model while usually significantly increasing the difficulty of these models. One framework that addresses uncertainties is *Stochastic programming*. Within this framework, a thriving method to model constraints subject to uncertainties in called *Chance-constraints*. In this setting, one wants to control the probability of violation of a constraint. Although chance-constraints, which involve probabilistic functions, have attracted a lot of attention in particular since the 2010, it still is a challenge to compute information on their variations, or even functional values. Following the lead of [19], we have studied the variational properties of some probabilistic functions. Our contributions in this domain, published in [17], are the following:

1. We present a general approximation of the first order information of the probabilistic functions we study.
2. We propose a computationally tractable formulation for this approximation.
3. Apart from some special pathological points (whose set is of zero measure), we show that this approximation is indeed exact.

The Optimal Power Flow with uncertainties - Part IV

The Optimal Power Flow (OPF) is a mathematical optimisation problem based on an electrical network. The objective of an OPF is to provide an optimal point while ensuring that flows of power are within some specified bounds. Since the 1960's, it has been at the centre of a large field of Operational Research. Current challenges on the OPF include the development of methods to take uncertainties into account. Building upon Parts II and III, we study the chance-constrained OPF in Part IV. Our contributions in this domain are the following:

1. We propose a novel two-step framework for the OPF under uncertainties jointly considered, that relies on DoC programming.
2. We present four different models which are OPF taking uncertainties into account one of which is a chance-constrained OPF.
3. We have conducted numerical experiments as a *proof-of-concept* to compare the four models and several DoC algorithms.

Part I

Operational Planning

INTRODUCTION

This first part presents the industrial objective at EDF R&D, where study of the electrical distribution networks is a significant activity. Several key concepts are defined or introduced which will guide the set up of the mathematical model of part IV.
The first chapter is aimed at a general understanding of the historical part of the electrical system and the new, fast-evolving current stakes and challenges. This non-technical section is mainly intended to stress out the underlying reality of the particularly vast and astonishingly expanding technical literature on electrical grid matters. An emphasis is given to the definition of new actors and the nature of the bilateral relations. This leads to an abstract and high level model of the electrical grid and its actors. This model underlies the developments of Part IV.
The second chapter provides mathematical background on the technical side of the electrical system, namely the OPF. As visible by the huge literature on this subject since the formalisation of the OPF in the 1960s, this optimisation problem has stimulated a lot of research teams and a large number of different models, solving methodologies have been proposed to solve it. As we are interested in mathematical methods of the OPF, we leave out of our discussion a vast facet of literature that mainly includes heuristics methods. We stress out that this decision should not lead a reader to disregard heuristic approaches to such numerically difficult optimisation problems as the OPF. As it turns out, such methods can be useful in practice.

RÉSUMÉ EN FRANÇAIS

Cette première partie présente l'objectif industriel de EDF R&D, où l'étude des réseaux de distribution électrique est une activité importante. Plusieurs concepts clés sont définis ou introduits, qui guideront la mise en place du modèle mathématique de la partie IV.
Le premier chapitre vise à une compréhension générale de la partie historique du système électrique et des nouveaux enjeux et défis actuels en pleine évolution. Cette partie non technique est principalement destinée à souligner la réalité sous-jacente de la littérature technique sur les réseaux électriques, particulièrement vaste et en grande expansion. L'accent est mis sur la définition des nouveaux acteurs et la nature des relations bilatérales. Ceci conduit à un modèle abstrait et de haut niveau du réseau électrique et de ses acteurs. Ce modèle est à la base des développements de Crefpart:4.
Le deuxième chapitre fournit un contexte mathématique sur l'aspect technique du système électrique, à savoir l'OPF. Comme le montre l'énorme littérature consacrée à ce sujet depuis la formalisation de l'OPF dans les années 1960, ce problème d'optimisation a été au coeur des travaux de nombreuses équipes de recherche. Il en résulte qu'un grand nombre de modèles et de méthodes de résolution a été proposé pour le résoudre. Comme nous nous intéressons aux méthodes mathématiques de l'OPF, nous laissons de côté une vaste facette de la littérature qui comprend principalement des méthodes heuristiques. Nous soulignons que cette décision ne doit pas conduire le lecteur à négliger les approches heuristiques de problèmes d'optimisation numériquement difficiles comme l'OPF. Il s'avère que de telles méthodes peuvent être utiles en pratique.

THE FRENCH ELECTRICAL NETWORK AND MEANS OF EXPLOITATION

1

France's electrical network is the result of over 150 years of development, whose full study embraces a vast diversity of fields (including Economic / Organisational History, Sociology, Architecture, Politics, Physics, Mathematics,...). Its evolution is far from being to sole output of series of electro-technical choices, that would have been guided by new scientific breakthroughs. As a result, the management and development of the electrical system requires a strong understanding of its technical limits, which bear a material side (an electric line has an upper bound on intensity transit) and an immaterial one (actors on the grid have boundaries on their responsibilities). Leading an optimisation study on the electrical system is thus intrinsically linked to a grassroots knowledge of the material characteristics, their physical models on one part, and the mix of actors at stake on the other. The latter is a fast evolving field: often times the electric system is dated back to the end of the 21st century, and though electrons remain the same, their management is not. For instance, an ongoing change that began in the 1990s concerns the liberalisation of parts of the electric system. This change stemmed from Chile before starting to be implemented at the end of this decade in European countries. The system was separated into two categories: "regulated" and "deregulated" as a consequence. Distribution remained a regulated sector, whereas production and supply were deregulated. This separation implies that new actors were created for each of these responsibilities that previously were uniquely integrated. This example alone is sufficient to characterise what "fast" actually means: in a 20 year time period, paradigms of exploitation and future development of a complex human-made system, comprising around 1.5 million of kilometres of lines that transferred 548.6TWh of electrical energy in 2018 [257] in France alone, have changed.

This work is motivated by the new role of an actor in the electrical grid, namely the Distribution System Operator (DSO). Since the beginning of the 21st century its role has evolved with the electric network, pressed by three main forces: deregulation, ecological transition in which falls the energy one, and the development of new or "smart" technologies. Though the underlying motivations behind this evolution are numerous (political, social...), they all impact the distribution sector. Liberalisation increases the number of actors on the grid and modify the interactions between DSOs and other actors in the system; the transition strongly influences decisions on acceptable actions, as the choice of generation sources; new technologies bring more capabilities to the DSO, which comes with more challenges on how to adequately use them.

INSERT 2: Remarks on other service networks in France.
The yearly investments in the electrical grid have to be compared to investments in other French networks, for instance the water and the telecommunication ones. The electrical grid cost is covered by revenue from a quadrennial national energy price called TURPE. From the TURPE-5 [173], Enedis has a yearly operating cost closing 15b€ with about 4.3b€ of grid investments. RTE, roughly in charge of the remaining 5% of the electrical grid comprised of very high voltage lines and international connections, plans to invest 20b€ over the period $2021-2030$. On the other hand, it would be interesting to have a macroeconomic order of magnitude of how much we benefit from the grid and these investments: closely related to this question is the cost of outage, or "blackout". As a matter of fact there exists an online tool developed to the evaluation of a blackout cost [249] (counting only the direct cost of energy not delivered). A 1 hour French metropolitan blackout in February 2021 would be over 600m€. In this perilous evaluation exercise, another way of computing the cost of a national outage of one hour could lie in the report from the European Commission [276]. This report defines the value of loss-load (VoLL), and finds that the French VoLL is 12€ per kWh for households, and 34€ per kWh for non-households. Now, broadly speaking, French non-households clients have a 262TWh yearly consumption [225, Figure 1] leading to a national hourly cost of $1.017.10^9$€. As an attempt to put these numbers into perspective, from [279] we note that the annual investments in France averaged between 2009-2013 is shy of 2.3b€ (only including investments and developments of the water network, excluding the sanitation processes for fair comparison with the electrical grid for which we have not included the power generation costs). In [190], we get that the amount of yearly investments in the telecommunication network (comprised of all types of networks, including mobile) is about to reach 10b€[a].

[a] This amount includes the new developments of very high speed telecommunication, which in fact is the operators' largest expense.

This chapter aims at providing a general insight on the electrical grid, and on the DSO. From an initial presentation of the grid, our objective is to make explicit the role and objectives of the DSO, and describe the nature of its relations with other actors. To this end, the chapter is divided into five different sections. The fist one recalls basic considerations on the electrical grid. Main actors on the network are detailed in the second section, where we briefly overview their main give objectives. The third section presents some historical facts on the development of electricity in France. Most important recent evolutions in this development are detailed in the fourth section, while the last section presents Enedis' role and its modes of action to meet its objectives.

1.1 CURRENT MATERIAL ORGANISATION OF THE FRENCH ELECTRICAL GRID

The French electrical network is nearly exclusively exploited as a three-phase AC system. Although recent advances in power electronics made DC lines a technico-economically competitive options on some large transmission projects (see for instance the INELFE project, a 65 kilometre High Voltage Direct Current line between Spain and France [158]), the subject of interest here will be the MV grid.

Currently, the French electrical network is classified into different categories based on the voltage level. Roughly speaking, from major production centres of electricity to low-voltage end-users, the continuous grid is separated into the following classes:

- a substation interfacing large and centralized production units with the rest of the grid;
- extra high voltage (EHV) lines: the *transmission* grid and European interconnections;
- EHV to high voltage (HV) substations;
- HV lines: the *sub-transmission* grid;
- HV to medium voltage (MV) substations;
- MV lines: the *distribution* grid;
- MV to low voltage (LV) substations;
- LV lines: the low-voltage *distribution* grid;
- LV clients;

Following this classification, two sets commonly arise in the French structure. The first one consists of EHV/HV (*Haute tension B*); the second one consists of MV (*Haute Tension A* and LV (*Basse Tension*).

Here are the English-French translations for voltage levels:

English	French	Voltages
Extra-high voltage (EHV)	Haute-tension B3 (HTB3)	$350\text{kV} \leq \text{U} \leq 500\text{kV}$
High voltage (HV)	Haute-tension B2/B1 (HTB2/HTB1)	$50\text{kV} \leq \text{U} \leq 350\text{kV}$
Medium voltage (MV)	Haute-tension A (HTA)	$1\text{kV} \leq \text{U} \leq 50\text{kV}$
Low voltage (LV)	Basse-tension (BT)	$\text{U} \leq 50\text{kV}$

Table 1.1: English to French nomenclature of voltage levels.

The rest of this section aims at pointing out key facts that differentiate HV from MV and below. Differences have historically been huge on the material side, as well as the selection of variables to monitor. For instance, actors on the HV grid usually adopt a “system view”, while actors on the MV grid usually have a “network view”.

This difference of views reflects the differences of objectives affected to each actors, and is interesting if one wants to study the numerous recent works on better cooperation and services exchanges between HV and MV grids/actors (see for instance [65, 129] for academic work, or European projects TDX-Assist [286] as an example of European funded initiative for better exchanges between both grids).

1.1.1 HTB grid

HTB is referred to as the *transmission* network: it is meshed grid, exploited as such on the French metropolitan area. It comprises about 106 000 kilometres of mainly aerial (for 94% of them), and partly underground power lines (for the rest), which, "beyond the «node to node» transmission, [...] has to organize the large-scale pooling of different sources of production" [261]. For this purpose, it is connected to the major centres of production (Nuclear power stations, some hydroelectricity and thermal centres of production), as well as interfaced with other European transmission networks and the HTA/BT French grid through 2 200 sub-stations (*Poste Source*). Phase-to-phase voltage levels are maintained quite high: 400kV, 225kV, and 90 or 63kV. A major evolution in these grids concerns the increasing proportion of underground lines, with an upcoming renovation cycle as the average age of its elements is 50 years. Digitization, emphasis on connection to offshore wind-farms, and doubling the interconnection capacities with neighbouring countries are also part of future steps.

From this material point of view, the role of the transmission network is three-fold:

- provide a connection service for centralised power plants and large renewable energy generation systems;
- transport electricity with a low level losses to the consumption areas;
- supply large clients.

1.1.2 HTA/BT grid

On the lower end of voltages is the HTA/BT network, called the *distribution* grid: its operating voltages range from 1kV to 50kV for the HTA (a majority of lines have a 20kV set point, but some are exploited at 15 or 33kV), and below 1kV for BT (voltage set points of 400V or 230V). With a total length of about 1 400 000 kilometres, it is interfaced with the transmission network through substations, and with end users. These users can directly be connected to the HTA grid if their installed or contracted power capacity exceeds 250kVA, otherwise they will be connected at a BT level. The interface between HTA and BT is a low-voltage substation, and around 800 000 of them are to be found in France [96].

The distribution network follows the decrease of tension from production to consumption already seen on the transmission network. The role of distribution network is thus to deliver energy to the end consumer. A key element of recent evolutions on the electrical grid is that end consumers can also produce electricity. Moreover, the vast majority of upcoming wind and solar energy producers are to be connected to the distribution network. As a consequence, on top of being the largest grid, the distribution network also is the one likely to experience the most changes in the future.

1.1.3 Some insights

The rationale behind the exploitation of the electrical system in a decreasing stair-valued series of voltage set-points can be attributed to simple physical laws. Transporting electricity over long distances implies power losses, due to the Joule effect among others. This corresponds to the loss of power due to natural conversion of electric current into heat as it flows through a component with a resistance. This power loss value is commonly modelled as being proportional to the square of the electrical current, the proportionality factor traversed by the current: $P_{loss} = RI^2$. On the other hand, the definition of transmitted power from the entrance of the element to its output is the product of electrical current by voltage. Consequently, in order to lower the Joule losses, one should seek to lower the current flow through the element which, if a constant power delivery is desired, amounts to increasing the voltage. Additionally, a higher voltage implies a lower voltage drop along lines, and a higher value of transmitted power. All in all, from a minimisation of power losses perspective, it is more interesting to transmit power using higher voltage. The downsides include greater attention to networks' configurations: lines have to be higher if aerial or better insulated when underground, to prevent from the ionisation risk; additionally, the inter-phase space also has to be increased. This leads to a higher investment cost when setting up a high voltage network compared to a low voltage one. Based on these facts, the technico-economic solution is a multi-layer voltage network: a very high voltage to transport electricity on long distances, and sequentially lower the voltage as the grid reaches the end user's connection.

1.1.4 Network vs System view

Different visions of the electric power system coexist among actors. In anticipation of the information presented further below, we briefly present an overview of the difference between a "network" and a "system" point of view, being two dual visions. The TSO is related to the former one, whereas the latter is linked to both the TSO and DSO.

The *System* view refers to the electrical grid as viewed from above, where the electrical dispatch is closely monitored in order to continuously maintain a load and generation match. This is solely the role of the TSO. On the dual side, the *Network* point of view encompasses the physical and material dimensions of the electrical grid: in this context, the objective is its safe and secure exploitation. It includes, for example, the knowledge of bounds on state variables, or human management required to manoeuvre controllable elements on the grid. This is incumbent to both the TSO and DSO, as they are responsible of the maintenance in operating conditions on their respective networks.

1.2 DESCRIPTION OF CURRENT ACTORS

An increasing number of actors are interfaced with the grid in France. At present time, management and interaction with and through the electrical network is established for:

- the Transmission System Operator (TSO), which is in France *Réseau de Transport d'Electricité*;
- Distributors

- Regulator
- Aggregator
- Grid user
- Balance Responsible Parties

1.2.1 TSO

The entity in charge of the Transmission grid is the Transmission System Operator. In France, there is one TSO (RTE), compared to Germany where four TSOs work on different geographical areas. Its missions also specified in [109], and include: *1*) maintaining a quality of transmission of electricity to the distribution grid 2) leading studies for the development of the electrical grid, that follows objectives of new renewable energy generators provided by the administration These missions emphasize the dual nature of RTE: it has to have strong regional links to ensure the effective electrical connection, while being responsible for the equilibrium of the electric system at the national and pan-European levels. As described in 1.1.4, the TSO has an emphasis on the "system" point of view, while also having a "network" approach. Challenges for the TSO include preservation of the generation/consumption equilibrium, and the dynamic maintenance and continuous monitoring of different types of reserve (namely the frequency containment, frequency restoration and replacement reserves).

In Europe, a vast majority of TSOs is represented by the association *European Network of Transmission System Operators for Electricity* (ENTSO-E) which brings together 42 TSOs from 35 European countries.

1.2.2 DSO

Distribution refers to a set of legal and contractual responsibilities in the process of electrical power delivery that is internationally encountered. A distributor is a role-concept in charge of this set of responsibilities, either found as independent entities (as in European countries) or integrated in entities having responsibilities over Transmission and/or supply as well (as in North America). They can be public or private, but distribution is generally a public service concession: the distributor therefore has a natural monopoly over a designated area for a fixed renewable time period. In France, Enedis is the leading distributor handling 95% of the distribution grid. The remaining 5% are concessions held by 154 small electric distributors, which account for 3.5 million people living in 2 800 municipalities evenly spread over the national territory. Dual point of view to the "system" one, the distributor traditionally has a "network" vision: the material aspects of the grid are more visible than in a systemic setting.

Differently to the TSOs, European DSOs have several representing associations including *European Distribution System Operators* (E.DSO) which brings together 41 DSOs from 24 countries, *GEODE* or *Eurelectric*.

1.2.3 Regulator

The Regulator's main mission is to participate in the good functioning of the electric market, to the benefit of the end users. In particular, it ensures the access to the electric

network is fair, and that the actors respect their contracts. It overlooks the coherence of exchanges on the market, the good implementation of European directives, the settling of transactions, and is responsible for the tariff method. For this latter responsibility, the regulator receives necessary financial data from RTE as well as from the distributors, and sets the calculation method that should reflect the national energy development policies. The output is a cost called TURPE that should favour virtuous grid management for the TSO and DSOs, be independent from the distance power plant to consuming point, from the location and the distributor. This entity embodies the bilateral link between the French State and the electrical grid.

1.2.4 Aggregator

The Aggregator is a new role in the electric system, whose creation is motivated by the requirement of balancing generation and consumption in a liberalized setting. Such an actor is the intermediary between grid users and the electricity market: it is in relation with a set of grid users, remunerates them for providing some power modulations that it can in turn sell on the market or over-the-counter to specific clients. Its main role relies on teleinformation gathering and sharing (including forecasts of consumption and/or production) in order to assess in real time the potential of power modulations. It has to be interfaced with 1. the grid to have information on its state; 2. its clients; 3. the market. It also has to have the capacity to send signals for power modulation, when committed to through its activated offers.

1.2.5 Grid user

A Grid User (GU) is defined as an actor connected to the grid, that can inject or withdraw energy. It is neither owned nor directly controlled by the DSO or the TSO. Connection to the grid is ensured by the corresponding operator, being either the TSO or DSO. The operator in turn, has to continuously maintain an optimised access to, and usage of, the network and its services. On the transmission level, a grid-user concept has been introduced through *Significant Grid-users* (SGU) established in the Grid Code. Such SGU include power generation or demand facilities deemed significant because of their impact on the transmission system in terms of the security of supply.

It is interesting to set such an abstract model of producers and consumers in the current electric grid context. Indeed, with interactions DSO-grid users evolving towards a contractualised model and the increased development of bi-directional information exchanges, an encompassing grid-user model where such actors can respond by increasing or decreasing their current access to the grid is fruitful. In our model, grid users all have contracts with the DSO, that uniquely define the possible interactions between the two actors. In the case of renewable energy production for example, two types of connection contracts exist in France. One is a contract that prevents any modulation of the injected power modulation from the DSO (which corresponds to Enedis' *Offre de raccordement de référence* [99]). The other one where power modulation is allowed and bounded (which models new contracts between deployed, *Offres de Raccordement Intelligent*). New grid users requesting to be connected to the grid can negotiate with the DSO the connection cost or time. Accepting an ORI can for instance lower the upfront cost, and provide power

modulation capabilities for the DSO during the exploitation time, and allow a quicker installation on the grid.

1.2.6 Market operator

Following the current European evolutions on the electrical system, the Market Operator is a new and upcoming actor whose role has gained much importance in the electrical balance. Electricity in Europe is now a commodity and the underlying of derivatives in multiple markets. Whereas common articles refer to "the" electricity market, one could distinguish different markets including a retail market (a consequence of the end of monopoly on the supply sector), a wholesale market (an electric power exchange, similar to a Spot commodities exchange, *e.g.* EPEX SPOT in France), and currently emerging ancillary services market (a codified market where TSOs can contractualise services to secure different types of reserves). Each market has a market operator, responsible for the settlement and proper compliance of actors to market rules.

1.3 A brief historical account of electricity and its development in France

In this section we take a slightly different point of view on electricity. Leaving Electricity as a "mystical" concept, not yet theorised, to the impressive reference [298], we start our account in the late 19th century, where electricity has already left laboratories. As such, history of electricity can roughly be separated into the following time periods: 1881-1918 where experiments were conducted to make appear the capacities of electricity, 1918-1946 where electricity was indeed regionally deployed, 1946 till nowadays with a nationalisation of the electric system, a massification of uses, and liberalisation.

Electricity and proofs of concept In 1881, it has its own International Exposition in Paris, raising a lot of optimistic ideas: a new technology is maturing, and one should find its applications. Being promoted by means of expositions straddling the two centuries, it became characterised as a "fairy" (the painting «La Fée Electricité», currently in the Musée d'Art Moderne de Paris, though unveiled much later in 1937, interestingly might have contributed to this lasting qualifier and explicitly bears in itself the motivations for the development of a national grid). Apart from lighting, the first major planned application of electricity lies in agriculture with, for example, demonstrations of electrical ploughing that inspired journalists, scientists, and farmers (see [49] for a comprehensive set of examples, and references therein). This is not merely a remark of the lesser importance, as alerted by Alain Beltran cited in [273]: "Étudier l'électricité en France sans aborder la question des campagnes et l'attitude des ruraux, c'est se condamner à parler d'une France hypothétique, future". Though electrification will require a minimum of density of population for industries to invest in the creation and management of grids, at its roots it is deeply related to agriculture. Bringing electricity to rural areas has accomplished multiple objectives, including strengthening nationalist feelings. A legal step is taken in 1906, when electricity distribution became a public service with law [103]; it also established that municipalities were the owner of the infrastructures of the distribution grid. Local (and rural) communitieds were encouraged to create larger associations,

partnerships, or joint municipalities, that would finance and overview electrification of their territories.

Regional developments of electrical grids From 1918, on the aftermath of World War I, and in particular until the breakout of World War II, electricity went from local tryouts to a rapid extension in (nearly) any parts of metropolitan France. The years 1919, when the French Parliament voted into law the legal basis for hydroelectricity [112], and 1923, when financial modalities were established [113], also are landmarks in the steps taken towards a rapid development. Led by entrepreneurs and private companies at the beginning of the 21st century, financed for more than 40% by the national State, electrification cities and villages grew from around 7 000 municipalities in 1918, to 31 000 in 1923 and 36 500 in 1937. The flexibility offered by the electric motor, the image of "modernity" borne by electricity in a period marked by Scientism, were strong arguments in favor of this source of power. Publicity had an important social role, as studied in [273]. From a higher perspective, the uncovered possibility to transport this power unleashed new groundbreaking works in urbanism: now, proximity to energy sources no longer was a crucial criteria when establishing a city (see [80], and in particular the "Cité Industrielle" by Tony Garnier described in it, an architect whose work has had a major influence of forthcoming urbanism and architecture in France).

Creation of EDF Before the breakout of World War II, as a result of this «private management of public services» that was dominant in France until the middle of the XXth century [120, p.55], there were 200 companies of production, over 100 in charge of transmission, and 1 150 for distribution. Efforts to interconnect these different meshes were speeded up in 1946 with the creation of *Electricité de France* (EDF) as a national company. Nowadays, 154 distributors (*Entreprises Locales de Distribution* - ELD) exist along with Enedis, the major Distributor; they are the result of the French electric development, have responsibilities over 5% of the national distribution grid and individually usually have less than 10 000 clients.

Production, Transmission, Distribution, and Supply were managed by EDF: this vertically integrated national company thus had the economic power to launch major industrial projects, as nuclear power development (industrialisation from the 60's to the 80's), large transmission network and extensive distribution grid, all while ensuring three fundamental principles:

- tariff equalisation: every private consumer pays the same tariff as comparable consumers i.e. clients that share similar consumption profiles and supply offer, independently of the distributor;
- tariff is independent of the distance to the electricity production plant.

Let us point out the relative scarcity of literature on the French development of electricity: as pointed out in [49], though it has been extensively covered from the end of the XIXth till the middle of the XXth (see for example [71] and [193] which provide a comprehensive work for this time period), an extensive survey study seems to be missing from 1946. The complexity of the subject, mixing a lot of different fields, has to be underlined. An interested reader could turn to EDF-related initiatives as Les Annales Historiques de l'Electricité (now defunct), and more recently Journal of Energy

History/Revue d'Histoire de l'Energie, two journals that provide solid works on electricity as a field of historical research largely (but not exclusively) related to the French case.

The European idea of the grid finds its roots in between the two world wars. Technological improvements made possible the exploitation of HV/EHV lines of 220, then 400kV, and significant ideas as maximalisation of the *economic mix* effectiveness (see [155]) initiated several initiatives to further develop a, still vague, European grid [175]. On another level, major historical events, including WWII, the Cold-War, economical growth of the 50's and a series of economical crises, a continuous set of diplomatic and inter-nations tractations well covered in [174] led to the interconnected European grid in 1995. The modification and impacts on distribution from this situation are the subject of the following section.

To conclude, our perception and imagination on electricity have recently drastically changed. As pointed out in [202], energy (and *a fortiori* electricity) cannot be looked at as the "electrical fairy".

1.4 RECENT EVOLUTIONS OF THE ELECTRICAL GRID

We here discuss about the main landmarks of the past 30 years on the distribution system. They can be classified into two categories: liberalisation and decarbonation. Both sets of objectives are transcribed, in France, into compatible schemes, strategies and plans. In a "top-down" vision, the leading national texts are the *Strategie Nationale bas-carbone*, an environmental roadmap for France to meet its international commitments. The strategic counterpart in the energy field is the *Programmation Pluriannuelle de l'Energie*, that is one of the three mainstays for the *Schéma décennal de développement du réseau*, the other ones being regional land use plans and the European network development ten-year plan.

The legal acting principles can roughly be summarised as the union of the following texts:

- the law founding the principles of electricity as a concession of public service [103]
- the European directive [235] that set up the electricity market in every European countries. It specified that actors on the electric network now had to be split into a Transmission Network Operator, a Distribution Network Operator, imposed the creation of a competent authority for setting litigation over contracts. It has been transposed into French law [110], that specifies that the DNO is EDF (now Enedis) and the "non nationalised distributors", creates the Regulator (the *Commission de Régulation de l'Energie*, CRE).
- Law [111] is currently fully in order, and can be seen as an entry point to the French electrical system legal basis. It implements a set of ordinances relative to the electrical grid, renewable energies.

A comprehensive summary of in action laws and decrees can be found in [97].

References for a thorough study of the legal requirements and objectives of Distribution can be found in the *Code de l'Énergie* [109]. We here summarise the necessary points for an understanding of the overview (see [109, Partie Législative, Livre III, Titre II, Chapitre II]).

Among all the missions defined in the Code, a Distributor is responsible for:

- defining and enforcing investments policies in order to allow new consumers and producers to be connected to grid;
- designing, constructing necessary facilities, while being the project manager;
- maintaining in a transparent manner a non discriminatory access to the grid;
- conducting studies on the impact of projects that are submitted to it, that are related to insertion of renewable energy, charging points for electric vehicles, urban or energy planning;
- operating and maintaining these networks;
- conducting the activities of energy metering.

More precisely, the exploitation comes with the obligation of monitoring at all time the flows of electric power, the efficiency, the security and safety of the distribution grid. The regulatory part of the Code defines what quality of electricity means, and we here expose the key elements for this definition. The Distributor is required to *1*) maintain tension on the distribution grid in a continuous way, and around values defined by the Minister in charge of Energy and *2*) ensure a minimum level of power failure.

The first item refers to the quality of power delivery, and upper as well as lower bounds on tension are thus provided. The second item refers to the continuity of delivery. Failure being an heterogeneous concept, without canonical measure, French distributors use two different criteria:

1. *Critère B*: measures the average duration of power failure for a LV consumer, per year (in minutes);
2. *END*, which stands for *Energie Non Distribuée*: measures the energy not supplied to a consumer on a yearly basis (in MWh). Not supplied energy is the quantity of electrical energy that would have been supplied, were there no failure.

For illustrative purposes, we consider a client on the HTA grid. Its contractual setpoint has to be within 5% range from the Nominal Tension; Enedis has to maintain the tension around this contractual tension with a 5% tolerance. Depending on the client zone (areas are ranged according to the number of inhabitants in a 1 to 4 scale), the maximum acceptable number of power failures is in-between 6 and 2 for major failures (more than 3 minutes, more acceptable failures in smaller municipalities), 30 to 2 for small failures (1 second to 3 minutes). From Enedis data [98], the average time of power failure per LV client is split into the planned shut-off (14.1 minutes) and unplanned failure (50.1 minutes).

For comparability purposes in Enedis' efforts to improve the quality of electric delivery, a global cost for ENS is computed with components from:

- the volume of energy not supplied, to which a cost is associated depending on the type of incident (€/kWh);
- the cost of power not delivered (€/kW);
- the number of clients not reconnected within a 5 day time-period after an exceptional weather event.

Interestingly, the majors changes on this particular Chapter that defines the missions of a DNO since 2011 are twofold. Firstly, its missions are enriched with an obligation to promote the development of renewable energy, and electric vehicles charging points, which are explicitly stated. Secondly, the CRE also sees its role grow: it is now more explicitly and often referred to, having to be called upon for counsel that, in turn, will be a prerequisite for issuing decrees.

One would easily expect regular changes in these texts, and this brief presentation should be taken as a rough overview subject to numerous incremental modifications. As seen with current European projects as TDX-ASSIST [286], an example of upcoming change can be an enriched exchange between the TSO and DSO, in particular when it comes to services that can be provided by the DSO. The following illustration highlights the necessity of improved exchanges TSO-DSO: liberalization of the electric network created markets of services like the Frequency Containment Reserve. As a consequence, producers, including the ones connected to the MV grid, can voluntarily participate to this market and make offers. In turn, the TSO can activate an offer on the market, made by a producer connected to the Distribution grid. Impacts on the Distribution grid are at the center of future studies (see [259, Section 4.10.4]).

On the liberalization

As well described in [126], liberalization of the electricity sector, previously considered to be a "natural monopoly", is unique on several levels. It has required more involvement from a regulator when compared to more classic markets (as commodities futures exchanges where oil is the underlying), which in turn had to deal with national breakdown of European rules. Regulatory changes are continuously made, having to adapt to the market maturing (where imperfections can arise) as well as to the evolution of actors on the market (new actors are introduced, and already active actors can have behavioural modifications, *e.g.* upcoming "citizen energy communities" as defined in the final 2019 European Clean Energy Package [83]). Difficulties of such an enterprise are easy to explicit: a common European market has to relate different technologies (a meaningful example from [126]: Norway main source of electricity is from hydro-power, while neighbouring Danemark produced around half of its electricity using wind power [86]), TSOs (currently 43 of them), physical interconnection points (420 at the European level, 50 of them being at the French borders), national regulations. The European vision is that of an economically optimal set point attainable once a complete market coupling has been attained.

Due to the intrinsic characteristics of electricity, namely its storing sill being scarce, the necessity of a continuous supply, and requirements of quality and balance, strict rules have to be followed in order for the setpoint coming from market decisions to be implementable on the grid. Nowadays, one could observe that market decisions do not often lead to network blackouts, thanks to robust design and planning of the electrical grid. This goes hand in hand with current regulations that do not require DSOs to provide particular services to other actors, apart from respecting its commitments to the TSO and possibly connected DSOs. In a forthcoming liberalized future, due to an increasing installation of distributed electricity generators, among which a majority will be connected to the MV grid, there is an emerging consensus that the probability of market output setpoints overshooting the boundaries of electrical networks, dating from

a *fit-and-forget* design period, is strictly positive.

A vertically integrated electrical company previously could conduct calculations on its grid relying on internal information exchanges, as *N-1* or short-circuit power calculations. In the unbundled setting, following on these two examples, external information exchanges are required, leading to the definition of standards of communication based on possible use-cases that can obviously evolve through time. Current Grid Codes, although originally emerging from the Transmission side, now include DSOs and in particular better define what the exchanges between the TSO and DSOs: for instance, a DSO has to provide forecasts of production on its MV grid to its connecting TSO. With information, exchanges of services could also be considered: this is the subject of on-going regulatory and technical works (see for example [131]).

1.5 ENEDIS' METHODOLOGY TO MEET ITS OBJECTIVES

Maintaining and developing the distribution grid is multiple time-scale process, than can be split into a strategic part and a tactical one. Mastering these contiguous time-scales is necessary as the distributor also has the objective of minimizing its costs. New developments are based on blueprints that span for 30 years, have long-lasting impacts (total cost of ownership of lines are computed on a 40 year exploitation time), and requires significant investments. They also have to take into account the future demographic dynamics, the evolving uses of electricity, the associated quantities of energy and power consumed, and how the tactical management can evolve and adapt to these new challenges. Strategic and tactical management are therefore naturally nested problems, the outputs of the former being the inputs of the latter.

1.5.1 Network Planning

Network planning refers to the set of means to anticipate future necessary network evolutions in order to maintain the distribution of electricity, achieving the same quality and maintaining the same level of security. It is a multiobjective problem: the output should comply to foreseen evolutions, while at the same time lower the environmental and monetary costs. Distribution networks have some specificities that emphasize the sensibility and need for tailored approach [93]:

- construction and reinforcement of infrastructures require large investments;
- new infrastructures usually have planned life cycle of 40 years: construction decision thus have long-lasting impacts, and there is little to no recourse;
- new developments can have long construction times to the hazardous nature of the environment, requiring special skills.

Let us also mention the influence of political choices, the integration of renewable energy being one example. This has been a fast-growing demand, that have had deep impacts on the historical network planning. Adaptation to future decisions is thus key to keep a continuously low cost of exploitation.

The challenges here are to provide robust decisions that will allow the continuous exploitation of the grid, at the lowest cost, for more than the next 30 years. More detailed

information on this subject can be found in [93], and technical information on Enedis' reports [258].

1.5.2 Operational Planning

In a chronological order, operational planning (OP) comes after network planning: although not explicitly defined, we consider it to span from a year till close to real time. The infrastructure is now considered to be mainly stable, without major changes allowed on the nature of network elements. What is variable within this scope is the state of existing elements on the grid: activations and levels are activations are modifiable. Operational planning is itself divided into long-term and short-term operations, each one with its own means of action and stakes. It has been introduced and properly defined in the Grid Code [215], which is a document that defines the operating rules and relations between the TSO and any type of user of the Transmission Network. As a consequence, it initially is a concept related to the TSO rather than the DSO. The Grid Code solely defines operational planning for the TSO in a sense that resembles what we would define as "long-term". Indeed, the OP phase for the TSO is from 8 weeks to 5 years in advance of real-time. We recall here its formal definition:

Definition 1 (Operational Planning from the Grid Code [215])**.** Planning through various timescales the matching of generation output with forecast National Electricity Transmission System Demand together with a reserve of generation to provide a margin, taking into account outages of certain Generating Units or Power Generating Modules, of parts of the National Electricity Transmission System and of parts of User Systems to which Power Stations and/or Customers are connected, carried out to achieve, so far as possible, the standards of security set out in The Company's Transmission Licence, each Relevant Transmission Licensee's Transmission Licence or Electricity Distribution Licence, as the case may be. •

Within the scope of our work, we have to define an appropriate definition for OP, from the DSO point of view.

Long-term Operational Planning

The objectives in this long-term process include negotiation and the set up of long-term contracts with producers, in order to plan construction or maintenance works on the networks. The time horizon still remains coarse, from a year or more, and the output should be a yearly planning that will ease the necessary adjustments of short-term planning. Decisions are based on climate variations and on a yearly basis, with visibility on necessary works on a big time scale.

Short-term Operational Planning

This is the subject of interest in this present work: short-term operational planning in the DSO's decision process from 1 month up to 30 minutes before real time. In this particular setting, the distributor has access to a specific set of information, and control variables. The inputs to short-term OP are the outputs of long-term OP and network planning. Only short-term construction and maintenance works are considered in the process. The

objective on the other hand remains essentially identical: the DSO has to define a tactic of minimal operational cost under operational constraints that allows users to safely access the grid, while respecting legal, contractual and thermal constraints.

We derive a formal definition for the DSO short-term operational planning:

Definition 2. Short-term Operational Planning is the coordination of means and human resources, respecting their activation limits, to ensure maximal network availability to users and meet legal, contractual and technical limits (including commitments to the TSO) at minimal cost. •

Whereas a traditional approach has been a temporal top-down vision, with unidirectional information being processed from the long-term to short-term, the latter has seen its importance grow with the evolution of electrical usages. The rationale lies in the occurrence of electrical constraints on some MV networks even under conditions deemed as "normal" i.e. without any infrastructure loss nor extreme weather event. As a consequence, "normal" network operations need to be redefined: information from the short-term should therefore be an input to longer-term decisions.

Equilibrium on the electrical grid

In France, the TSO is legally responsible for the electrical balance. In this equilibration process, RTE relies on Balance Responsible Parties, BRP, or entities that assume responsibility of local equilibria over a designated set of injection and consumption points. In case imbalances are observed at the settlement, RTE can ask for compensated adjustments: BRPs and the TSO operate financial exchanges. BRPs are market participants, and more explicitly do not need to be producers or suppliers in order to qualify as such. Their goal on the electrical market is to commercially balance the energy flows within their perimeter.

Means at the DSO's hand

Recent evolutions in the access to the grid and usages of electricity came with technological and regulatory changes concerning abilities and capabilities of the DSO. As a consequence, the DSO has at hand a set of internal and external levers that we discuss in this section.

Difference between flexibility and lever:

- *internal levers*: these are owned, activated and manoeuvred by the DSO. They do not constitute offers on a market. Their activations are limited by technical bounds.
- *external levers*: opposite to the previous class of levers, these are not owned by the DSO. Their activations are possible through a contract, either coming from a selected and paid offer on a market, or a mutually set up agreement. This is still a prospective field, as there is currently no market to which a DSO can participate in France, apart from local experiments.

We will refer to flexibilities as defined by Enedis:

Definition 3. A flexibility is a voluntary power modulation of one or more sites, either an upwards or downwards, injection or withdrawal, at a given time step for a given time period, as a reaction to an external signal requesting a service. Enedis refers to *local* flexibility as the knowledge of the flexibility localisation is necessary to ensure that its activation will ease an electrical constraint. •

Existing flexibilities Current flexibilities are numerous (see for instance [278])

1.6 CONCLUSION

The objective of this chapter is to provide a better hindsight into the French electrical grid and its evolutions. The DSO is bound to evolve on several aspects in order to fulfil its missions.

A REVIEW OF EXISTING OPF MODELS AND SOLVING ALGORITHMS

2

Short-term operational planning in a deregulated and liberalized environment, as well as in one requiring operation to be conducted at an economically optimal cost, is a competitive sector. Demand for an optimal management of an ever-growing and evermore complex electrical system has fostered a myriad of propositions since the difficult-to-find yet renowned work of Jacques Carpentier from 1962 [72][1]. In this chapter our ambition is to provide a general overview the Optimal Power Flow. To this end, we start by discussing the main existing modelling techniques used by researchers in the field of optimisation of electrical distribution systems. While some physical laws seem to be used in close to all of the existing models, some other modelling aspects (as variables choices) have led to numerous differing experiments. Modelling choices are the subject of section 2.1 along with some basic yet necessary definitions of physical quantities in an electrical grid. Building upon Section 2.1, we present in Sections 2.2 and 2.4 the OPF main structure and look into how this has been transposed into specifically tailored optimisation problems. The third section briefly presents the different objective functions that have been used in literature. Sections 2.6 and 2.7 are more specific than the previous ones. In the fifth section we look into the subset of literature that addresses the OPF from a DSO point of view. The sixth section is then dedicated to the OPF with uncertainties: on top of the multiple mathematical fields that are tailored to tackle such optimisation problems with uncertainties, a huge number of heuristics have been developed for these programs.

[1] We were not able to put our hands on this French article, although an overwhelming majority of posterior articles recognize it as the founder of the Optimal Power Flow as a mathematical object.

Notations in this Chapter

$\mathcal{N}, \mathcal{A}$	Sets of nodes and lines
u	Decision variables
x	Dependent or state variables
$\underline{a}, \overline{a}$	Lower and upper bounds of a given variable a
$f(u,x)$	Objective function
$g(u,x)$	Vector function of equality constraints
$h(u,x)$	Vector function of inequality constraints
$p_i^l(u,x), q_i^l(u,x)$	Vector function of load consumption of active and reactive power in the network
$p_i^g(u,x), q_i^g(u,x)$	Vector function of generator production of active and reactive power in the network
$p_i^l(u,x), q_i^l(u,x)$	Active and reactive power consumptions at bus $i \in \mathcal{N}$
$p_i^g(u,x), q_i^g(u,x)$	Active and reactive power injections at bus $i \in \mathcal{N}$
δ	Angle between bus voltages
θ	Angle between current and voltage
$\|V\|$	Voltage norm
δ	Voltage phasor
j	Complex number, $j^2 = -1$
z^*	Complex conjugate of z: if $z = z_R + jz_I$, then $z^* = z_R - jz_I$

Table 2.1: List of notations we use in this chapter.

2.1 Basic models for the OPF

In this first section, we start by presenting modelling aspects in the OPF. In order to provide a comprehensive section, we include definitions of some physical quantities related to the modelling of electrical distribution grids.

2.1.1 Preliminary elements for electrical grid modelling

While electrical grids can be modelled differently depending on the type of problem we are faced with, all models rely on graphs. An electrical grid is modelled by a set of nodes $\mathcal{N}$, connected by a set of lines $\mathcal{A}$. With more assumptions on the electrical grid we model, the graph can be further characterized:

- When dealing with transmission networks, the graph is connected and has cycles. It is referred to as a *meshed network*.

- When dealing with distribution networks, the graph is a tree (i.e. a connected graph without cycles).

More specifically for distribution networks, the modelling tree usually has one entry node called the *slack bus*. This latter node is the connection (the substation) between the distribution network at hand and the transmission one it is connected to. From an practical point of view, the slack bus of the model is placed either at the primary (the MV side of the substation) or the secondary (the LV side of the substation) with an observed preference for the secondary. Other nodes in the tree model either MV GUs, or connection points between the MV network and the LV one.

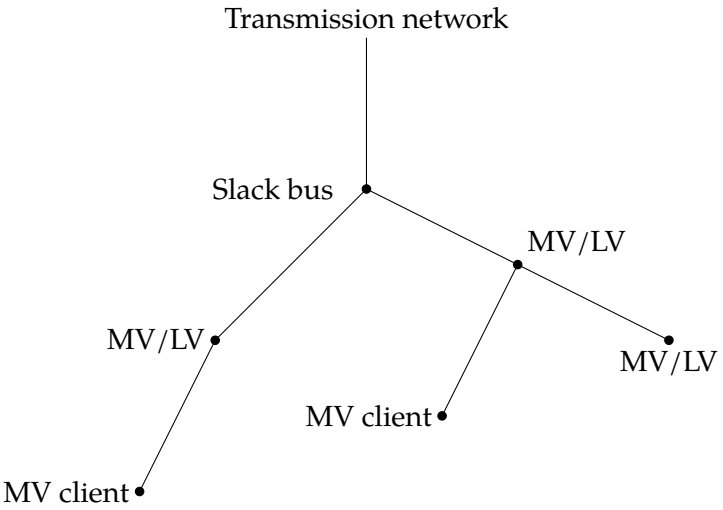


FIGURE 2.1: Illustration of a tree modelling an electrical grid.

The next step is to characterize how the electrical power flows behave in the electrical distribution network as the one of figure 2.1. These flows are necessary for the DSO to safely manage the grid. Modelling the power flows requires additional parameters held on lines and nodes of the tree depicted in figure 2.1, as well as the introduction of some variables, which we refer to as *state variables*.

CHARACTERISTIC PARAMETERS OF AN ELECTRICAL DISTRIBUTION GRID

Each electrical component has a specific reaction to electricity, i.e. the flow of electrons. In order to model this reaction and how an electrical component affects the flow of electrons, some specific parameters are necessary. Thanks to these parameters and to physical laws that are presented in section 2.1.4, a user is able to study the evolution of state variables, and thus appropriately monitor an electrical grid.

Within the scope of our work, electrical components (or Grid Elements (GEs)) are comprised of electrical lines (arcs in the tree of figure 2.1) and buses (nodes in the tree of figure 2.1). Most common models for electrical lines involve a complex parameter called the *electrical impedance* and noted z_{ij} for a line from node i to node j. The real part of the impedance is called the *resistance* noted r_{ij}, and the imaginary one is called the *reactance*, noted x_{ij}. Parameters are also affected to nodes, called *shunt impedances* (they are impedances from a node i to the ground), which are complex quantities too ad noted $z_i = r_i + jx_i$. We represent these parameters in an electrical diagram as follows:

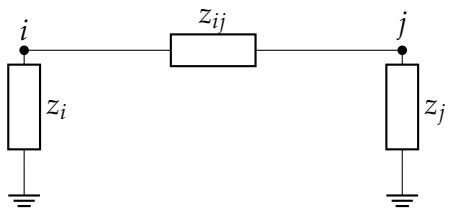


FIGURE 2.2: Illustration of electrical parameters.

The inverse of impedance is called the *admittance*, noted y_{ij} for a line connecting i to j. The following is therefore always true: $y_{ij} = \frac{1}{z_{ij}}$. The real part of the admittance is called the conductance, noted g_{ij}, while the imaginary part is the susceptance, noted b_{ij}: $y_{ij} = g_{ij} + jb_{ij}$.

2.1.2 STATE VARIABLES

Modelling the electrical grid amounts to defining the relations between *state* and *decision* variables. While the values given to the latter quantify the actions taken by the decision-maker, the former describe the state of the grid. Once state values are set the state of the grid is by definition unique and known. In this first subsection, we are interested in the different choices of state variables which are always continuous variables (this observation is backed in [117, Section 2.3] which provides an extensive overview of electrical grid modelling).

The cornerstone of state variables is the voltage. This latter variable is a nodal quantity that is closely related to the amount of electrical charges at a given point. In an Alternative Current context, the voltage is a complex value that requires at least two real variables to be described:

$$V = |V| e^{j\delta}, \tag{2.1}$$

where $|V|$ is the modulus and δ the phasor, or

$$V = V_{\text{Real}} + jV_{\text{Imag}}. \tag{2.2}$$

The first formulation equation (2.1) is called the "polar formulation", while the second is referred to as the "rectangular formulation". It is clear that both formulations have different advantages as they make explicit values that have different interpretations. In equation (2.1), the phasor is explicitly available and bounding this variable can be done linearly: $\delta \in [\underline{\delta}, \overline{\delta}]$ for instance. When using the rectangular form, the phasor is the following quantity: $\arctan(\frac{V_{\text{Imag}}}{V_{\text{Real}}})$, which requires for instance an additional proxy variable in order to include box constraints on this value. This is often found in literature where phasors are upper and lower-bounded for stability/synchronicity purposes. In more advanced modelling, phasor-shifters[2] can be modelled in which case using a polar form is evidently easier than using the rectangular one (see [227] for an introductory reference into modelling techniques of such advanced materials). As a

[2] Phasor-shifters are electrical components whose essential feature is to regulate the phasor of a given voltage in an electrical grid.

downside, the polar formulation also causes trigonometric functions to appear: the vast majority of works in literature uses real-valued functions (this is discussed in [268]). A reformulation with real variables while using the polar form immediately yields: $V_{Real} = |V|\cos(\delta)$, $|V|\, e^{j\delta}_{Imag} = |V|\sin(\delta)$, thus making trigonometric functions appear in the optimisation problem. On the other hand, polynomials of V_{Real} and V_{Imag} appear in place of cosinus and sinus functions when the rectangular form is used. This formulation makes it difficult to limit voltage magnitude as it leads to a non-linear constraint. This becomes visible as, when using the rectangular form, $|V| = \sqrt{V^2_{\text{Real}} + V^2_{\text{Imag}}}$.

The use of rectangular coordinates is first discussed in solving the Power Flow (PF) system of equations. As shown in [2, 156], where the authors solve a PF system with polar and rectangular formulations, convergence of their algorithm is significantly modified depending on the formulation choice. From a general point of view, it is unclear whether one can argue in favour of using the polar or the rectangular formulation: it seems that this choice is case-dependent.

A more recent work [268] advocates the use of complex variables. Complex Analysis, which involves holomorphic functions, has been recently successfully applied to solving OPFs ([177]) and PF systems ([284, 108]).

Apart from the voltage, other electrical quantities are of interest: first of all, the *current* which measure the quantity of electrons per time-unit through an electrical component (as a distribution line for instance). This quantity is noted I_{ij} for the current in the line that connects nodes i and j. The relation between the voltage and the current will be discussed in section 2.1.4.

Remark 2 (From current variables indexed by lines to current variables indexed by nodes)**.** Note that in some models, the current is a *nodal* variable, and is thus not indexed by a line but by a node. This merely is an equivalent reformulation, as described for instance in [277, Lemma 1]. Let us denote by $\widetilde{I}_i$ for a node i the nodal current values, while I_{ij} is the current indexed by lines. Let the grid be a directed graph, with a randomnly chosen orientation. Then, $\widetilde{I}_i = \sum_{j:\, j\to i} I_{ji} - \sum_{j:\, i\to j} I_{ij}$: there is a linear transformation from one set of variables to another. ▷

We then turn our attention to the electrical power. Similarly to the voltage, and the current, the electrical power is a complex value with the real part being the *active power* and the imaginary part being the *reactive power*.

Definition of the electrical power

The electrical power can seem ambiguous at first. On the one side, it designates a physical quantity that can characterize a steady-state for an electrical grid. On the other side, in the OPF context, it designates the level of power generation or consumption for a Grid User (GU). These two definitions, even if they both are electrical power quantities, are different: the former one is a state variable, while the latter is either data or a decision variable. To differentiate the two types of power, we use capital letters for the electrical power as a state variable, and minuscule letters for the electrical power as data or decision variable.

A first definition is as follows:

$$S_i = |V|_i \, e^{j\delta_i} I_i^*, \quad \forall i \in \mathcal{N}. \tag{2.3}$$

This is a nodal definition, and using equation (2.6) we can make the current variable disappear:

$$S_i = |V|_i \, e^{j\delta_i} \sum_{k \sim i} y_{ik}^* \, |V|_k \, e^{-j\delta_k}, \quad \forall i \in \mathcal{N}, \tag{2.4}$$

where $k \sim i := \{k \in \mathcal{N} \mid k \text{ and } i \text{ are connected}\}$[3]. This definition has been by far the most used one in literature. Apart from this nodal definition, there exists another one where the electrical power variables are held by lines. In this second version, the current and electrical power variables are indexed by lines, and the definition becomes:

$$S_{ij} = V_i I_{ij}^*, \quad \forall (ij) \in \mathcal{A}. \tag{2.5}$$

First introduced in [43, 44] where the authors develop a model called *Distflow*, it has recently been successfully used for convex relaxations in a model now called Branch-Flow. For instance, see [277, 198, 199] which discuss the equivalence of models with nodal versus line variables, [1, 53] present more advances on relaxations making use of line variables, [90] which is a case that takes uncertainties into consideration and [278] where the author adds integer variables.

Active and reactive power From a mathematical perspective, what is understood by "electrical power" can in fact turn out to be more complex than previously anticipated, and it is our belief that there are several significant operational aspects to understand in this concept. The *complex* power S always verifies $S = U\, I^*$, where U is the voltage and I the current (amount of electrical particles per time unit). Its magnitude, called the *apparent* power $|S|$, is measured in volt-ampere VA. We are interested in the alternative-current context, which is one operation mode for the electrical power, where electrical particles are set into motion one way and then the other periodically. Vector I is thus modelled by a complex time-dependent rotating vector, and according to Ohm's Law (which is linear) this also holds for U. As a consequence so is S, whose time-averaged value is affected by the complex linear factor of Ohm's Law (the impedance). More precisely, the real and imaginary parts of S respectively called the active (P) and reactive (Q) power, are impacted by the impedance: if it has a real value, then there is active power, whereas a purely complex impedance leads to only reactive power.

Taking a little more perspective, real power is the only component of electrical power to provide *work* (an energy) which is necessary and required from electricity as an energy vector. The reactive power on the other hand comes from the magnetic energy stored by grid elements that are not purely resistive (i.e. their impedance have a non-zero imaginary part). This type of power is thus inherent to each electrical grid, even if it does not contribute to any work.

[3] By convention, we here assume that i is connected to itself.

INSERT 3: Angles in power systems.
We have introduced δ as the phasor of the voltage, which is complex value. One then has: $\tan(\delta) = \frac{V_{\text{Imag}}}{V_{\text{Real}}}$. Another angle is often encountered in Power System studies, which is the phasor difference between the current and the voltage which we denote here as θ. Unfortunately these notations, although mostly encountered, are not unique. Angle θ is used to compute the *power factor* which is the proportion of real power out of apparent power: $P = |S| \cos(\theta)$. Angle δ is used to compute the amount of active power effectively transmitted from a sending end to a receiving end of a line.

2.1.3 DECISION VARIABLES

The second subset of variables is the decision variables. In the short-term operational planning, decision variables model the potential actions, called 'levers', available to solve the problem. As a consequence, there are as many decision variables as levers that are modelled in literature. Decision variables can be split between continuous and integer variables. The former ones have initially been mostly used in optimisation, due to the simpler resolution of a continuous program. Power modulations (i.e. the decision to modify the levels of power generation or consumption), especially in case of generation modulations, can be assumed to be continuous; this is the case in the canonical OPF equation (2.7). This is also true from a TSO point of view, where levels of production are modifiable. For load modulations, some models use continuous variables (see [125]) while others use integers (see [125]).

Integer variables can naturally arise in models of the electrical system. For instance, the following actions require integer variables to be modelled:

- Transformer taps: transformers are components used to modify the voltage between two electrical circuits (in alternative current). Some newer transformers can receive orders to change their set points. This decision is naturally a discrete choice, with typically around 20 possibilities. Depending on the number of transformers on the grid, it can be necessary to develop a tailored solving method to tackle these problems as in [278, 170]. We care to emphasize that other transformers, in particular in France, have a different way of functioning (see [60] for a detailed presentation of such transformers).

- Line switches: this models the state of lines in the network, and is a binary operation since a line is either connected or disconnected. This leads to very interesting and challenging OPF problems, as they are non-convex with integers programs that be of have large sizes. A technical model is presented in [278].

- Activation of capacitor banks: as part of the Flexible AC Transmission System (FACTS) framework, capacitor banks can be connected to, or disconnected from the grid. Modelling FACTS is didactically presented in [307].

- Start-up and standby of power plants. As extensively described in [10], some problems can combine decisions on the binary status of power plants with network state variables: this is called a *Unit Commitment*, and is a subproblem of the OPF as it includes PF constraints.

The list is well described in [117] and in the more recent survey [50].

2.1.4 Physical laws for the OPF and approximations

The main set of constraints are related to the physical model of the electrical grid. Following the vast literature not only of the OPF but of basic electrical studies, three main physical laws are necessary to describe the steady-state behaviour of an electrical circuit:

- The first one is Ohm's Law, which draws a linear relation between the current and a difference of voltage values. The linear coefficient characterizes a circuit element, and more specifically its opposition to current when a difference of voltage is applied across its ends.
- The second one is Kirchhoff's Current law which translates the nodal equilibrium between inputs, outputs, and internal production / demand.
- The last one is the definition of electrical power from voltage and current values.

Ohm's Law

This law defines that the relation between the current I through an electrical conductor (as a line) and the difference of voltage between the two extremities is linear.

$$I_{ij} = \frac{1}{z_{ij}}(|V|_j\, e^{j\delta_j} - |V|_i\, e^{j\delta_i}) = y_{ij}(|V|_j\, e^{j\delta_j} - |V|_i\, e^{j\delta_i}). \tag{2.6}$$

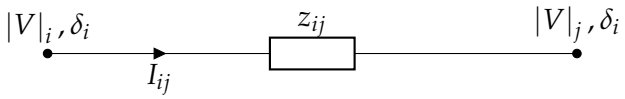


Figure 2.3: Illustration of Ohm's law.

The linear coefficient y_{ij} is called the admittance of line ij, and is the inverse of z_{ij} which is the impedance. The initial objective of most power systems analysis books is to build the admittance matrix (see for instance the first chapter of [139, 214]). Within the limits of validity of Ohm's law (see Section 2.1.4), the admittance matrix provides a complete description of the material characteristics of an electrical network.

Kirchhoff's law

The full name of this second law is Kirchhoff's circuit law, and is obtained as a flow balance at each node. In other words, it is a translation of the conservation of electrical charges: at each node of the electrical grid, the algebraic sum of intensities must be equal to 0.

For instance, in figure 2.4 Kirchhoff's law is written: $I_{15} + I_{25} = I_{53} + I_{54}$. This law also holds for current variables indexed by nodes, as already mentioned in remark 2.

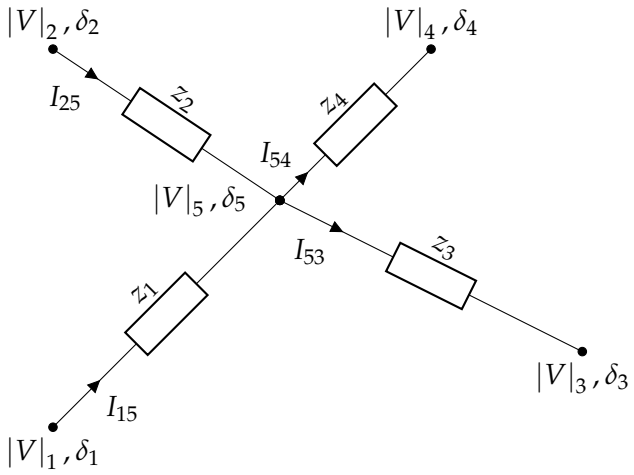


FIGURE 2.4: Illustration of Kirchhoff's law.

SPECIFICITIES OF THE DISTRIBUTION GRID

There are some particular rules that a distribution grid currently follows. These specificities are on the topology of the grid, its characteristic values and the level of production.

First of all, we recall that a distribution grid generally is exploited as a tree (i.e. a connected graph with no simple cycle). Lines can be oriented in literature for convenience (see DistFlow [43] or Branch Flow [198] for instance), but disregarding this possibility it is always assumed that there is a root node also referred to the *slack node*. The slack node is by convention set with $\delta = 0$, $|V| = 1$: this contributes to the definition of the *per unit* system. In practice, this node can be connected to either the primary or the secondary of the transformer HV/MV depending, for example, on whether the transformer is to be modelled or not.

An electrical network is characterized by its admittance or impedance matrices. Transmission networks have a reasonable approximation readily available: usually the reactance is significantly greater than the resistance. As a consequence, it is possible to neglect r in front of x in OPF approximations. Neglecting r is not possible for distribution lines: the rationale is that materials, lengths and diameters of transmission and distribution lines are different.

Finally, the last significant specificity of distribution grids when compared to transmission ones is that producers have lower levels of generation. As a consequence, distribution grids have historically been more "consuming" than "producing". This particular point is central in the electrical transition: some distribution grids can now experience reverse power flows for instance, due to the arrival of distributed renewable energy sources.

SOME ADDITIONAL ELEMENTS ON THE DISTINCTION BETWEEN ACTIVE AND REACTIVE POWER

Interestingly, due to recent technological advances (for instance in Power Electronics, which is "the study of electronic circuits intended to control the flow of electricity" as described in [172]) there is more "freedom" between P and Q values. Whereas we know that the proportion of real and reactive power is set by the impedance, new technologies make it possible to enlarge the set of admissible Q for a given P which previously was,

roughly speaking, a singleton. An interested reader can turn to [294], and in particular to Section 32.4.1.3 of this latter reference for a specific example of what Power Electronics can do.

This is significant as a first-order approximation leads to the following relations: P has a higher impact on δ and Q has a higher impact on $|V|$. The cost of modifying Q is also subject of debate as it is difficult to appropriately assess. In [47, 27], the authors discuss the difficulty of defining this cost of service to the grid.
All in all, one can easily perceive the interest of having a full AC model:

1. Taking into account Q is more realistic for the electrical model. As a matter of fact, in article [41] yet to be published, solutions to AC and DC OPF are compared. Interestingly, and as expected, under mild assumptions it turns out there is an empty intersection between the two solutions sets. This holds even after adding some improvements to the DC OPF. Considering the advances in tractability of AC-OPF, it leads the author to argue in favour of only considering DC approximation for more complicated OPF-related problems, such as mixed-integers or stochastic versions.

2. Reactive power, by its first order differentiated impact of $|V|$ when compared to its effect on δ, is a source of interesting local regulation. As a matter of fact, it is currently used on an operational basis [246], which is another argument that a mathematical model should indeed take Q into account.

Usual assumptions in models of electrical distribution systems, in a short-term operational planning setting

We have presented thus far in this section the main physical laws that are used to model PF in an OPF. These laws, as any physical model, are approximations. In this subsection, we highlight some elements which make these latter approximations acceptable, and thus allow the OPF to be a useful tool. The following assumptions can also be taken as sources of data uncertainty, which we do not consider in this work:

- Short-term operational planning is temporally macroscopic. In other words, as the time unit is about 30 minutes, physical events that might occur at a time scale close to an electrical event's typical time constant (roughly between 10^{-14} and 10^{-1} seconds) are discarded. The assumption that the distribution grid is a tree also holds because transient behaviours are usually disregarded. Topology changes in real life (usually) create temporary loops and cycles in the electrical grid, but we disregard such events and their corresponding impacts as their time-lasting order of magnitude is negligible with respect to our time scale.

- Similarly, this setting is spatially macroscopic. We are interested in a distribution grid, which implies that our spatial scale is well above the meter level. As a consequence we disregard any physical phenomena that can occur at smaller scales (e.g., microscopic behaviours of electrons are not considered at all).

- We consider all the grid elements to have a linear behaviour (as in Ohm's Law for instance), and we neglect the impact of weather (as temperature) on the linear coefficient.

- Full knowledge of all parameters in the grid is usually assumed.
- Loads and generators are modelled as p, q sources or consumers. Literature on PF offers more precise load and generator models. Usual improvements of load modelling include a relation between active power consumption and voltage at the connection point (see [157]). The impact of different load models on the OPF is studied in [163]. The authors of this latter work compare three different load models: it is shown in their test-cases that the p, q model provides almost always a different OPF set point to the one obtained using the two other models. Differences between the set points obtained from the p, q model and the ones obtained from a model that accounts for voltage sensibility can exceed 10% in some specific cases. This is a very significant value as usual MV grids should not exceed 5% increase/decrease from a contractualised set point. Our model choice remains a strongly motivated one as it is (1) widely used in literature, (2) a good trade-off between simplicity of use and quality of model, especially in a difficult mathematical problem.

Some approximations, or sources of uncertainties, can be addressed in the OPF and thus become a somewhat "generalized" OPF. For instance, using a better load model in a convex reformulation of an OPF is proposed in [266]. Transient Stability Constrained OPF is another example where identified limits are discussed and expanded, namely the fact the standard OPF setting is a steady-state one, thus *a priori* discarding voltage stability studies that can occur on a smaller time scale. In [304] the authors present an OPF with additional embedded constraints that account for transient constraints.

2.1.5 The canonical OPF

Building upon the first three subsections, we now turn to the most encountered OPF problem in literature. This formulation, which we refer to as the *canonical* OPF, is directly related to the formulations found in [72, 92]. Its objective is to minimise generation costs, constrained by Power Flows and bounding constraints. For any generator index k, let $C_k \colon \mathbb{R} \to \mathbb{R}_+$ be its associated cost function.

$$\min_{p^g, p^l, q^g, q^l, V, I} \quad \sum_{k \text{ Generator}} C_k(p_k^g) \tag{2.7a}$$

$$\text{s.t.} \quad p_i^g - p_i^l = \text{Real}(V_i I_i^*) \qquad \forall i \in \mathcal{N} \tag{2.7b}$$

$$q_i^g - q_i^l = \text{Imag}(V_i I_i^*) \qquad \forall i \in \mathcal{N} \tag{2.7c}$$

$$I = YV \tag{2.7d}$$

$$\underline{p_i^g} \leq p_i^g \leq \overline{p_i^g} \qquad \forall i \in \mathcal{N} \tag{2.7e}$$

$$\underline{q_i^g} \leq q_i^g \leq \overline{q_i^g} \qquad \forall i \in \mathcal{N} \tag{2.7f}$$

$$\underline{|V|_i} \leq |V|_i \leq \overline{|V|_i} \qquad \forall i \in \mathcal{N} \tag{2.7g}$$

$$\underline{\delta_i} \leq \delta_i \leq \overline{\delta_i} \qquad \forall i \in \mathcal{N} \tag{2.7h}$$

Bounding constraints (2.7e), (2.7f) are lower and upper bounds on the allowed levels of power generation and consumption. These bounds reflect the fact generators, for instance, cannot produce more electrical power than a given level determined right at the designing process of the power plants at hand. Conversely, some power plants (typically

thermal ones) cannot lower their production levels to a given threshold without having to shut-down. Bounding constraints (2.7g) and (2.7h) are related both to safety and legal regulations: firstly, too low/high voltages can lead to deteriorations as all connected equipments are designed to work at a specific voltage level. Secondly, these lower and upper bounds on voltage values are explicitly given in the legal texts that regulate the activities of the DSO. Respecting these boundaries is significant in the quality assessment of electrical distribution (see [173, Section 3.2]).

2.2 CONSTRAINTS IN THE OPF

In this section, we are interested in constraints always encountered in an OPF formulation called the *Power Flow* (PF) constraints. Conceptually, solving an OPF amounts to finding an optimal decision on power variables while ensuring the existence of state variables that comply to selected bounds. The main difficulty of an instance of the OPF as an optimisation problem lies in the constraints, referred to as the *Power Flow* constraints. These constraints link the information related to GUs, decision variables and state variables. Without approximation, they always involve non-convex functions. Power flow constraints can also be non-smooth, depending on which levers are modelled.

2.2.1 Power Flow constraints

The PF constraints model the equilibrium of power in a network. The most encountered formulation, called *bus injection*, only involves nodal variables and is written as follows:

$$\begin{aligned} p_i^g - p_i^l &= \text{Real}(V_i I_i^*) \qquad \forall i \in \mathcal{N}, \\ q_i^g - q_i^l &= \text{Imag}(V_i I_i^*) \qquad \forall i \in \mathcal{N}. \end{aligned}$$

When using equation (2.6), we rewrite the PF as follows:

$$\begin{aligned} p_i^g - p_i^l &= \text{Real}(V_i \sum_{k \sim i} V_k^* y_{ik}^*) \qquad \forall i \in \mathcal{N}, \\ q_i^g - q_i^l &= \text{Imag}(V_i \sum_{k \sim i} V_k^* y_{ik}^*) \qquad \forall i \in \mathcal{N}. \end{aligned} \tag{2.8}$$

Following power definitions that use line variables as in equation (2.5), the PFs are modified as follows:

$$s_j = \sum_{i:\, i \to j} (S_{ij} - z_{ij}|I_{ij}|^2) - \sum_{k:\, j \to k} S_{jk}, \quad \forall j \in \mathcal{N}. \tag{2.9}$$

This formulation has been developed from the 1980s (see [45, 43]) and is now often called the *Branch-Flow model*. Formulation (2.9) has in general been used in works that study relaxations (see Section 2.5.1), although in [277] it is clear that both formulations are equivalent and convex relaxations have been developed for both models.

2.2.2 Bounding constraints

Additional constraints are always considered in an OPF, which ensure safety in the electrical grid management. These constraints monitor state variables (voltage magnitudes

and phasors, current or power flows) as well as decision variables (box constraints on power consumption and generation). They usually enjoy a stronger structure than the PFs, as they are either linear or convex.

Starting with the linear constraints, voltage magnitudes and phasors, as well as power variables $|V|, \delta, p^g, p^l, q^g, q^l$ are lower and upper bounded in the vast majority of articles, independently of the levers considered. The only notable exceptions are for some particular relaxations (see [122] for instance): for these relaxations to hold with a zero-gap, some authors assume that there is no jointly binding box constraints at both end of a line for active and reactive power. One way to ensure this assumption is to drop some of the box constraints. As a consequence, the solution should be checked *a posteriori* to verify its physical feasibility.

Apart from these box constraints, thermal constraints are also common. The objective of these constraints is to limit the current or power flows in lines. From a physical point of view, there are at least two reasons for not allowing unlimited flows:

- safety and security devices are scaled for a given range of values,
- aerial lines expand when overheating, which increases their deflection and lowers their insulation.

Without relaxations, these constraints naturally result in line variables to model flows from a node i to a neighbouring node j. Upper bounds for these flows are either on apparent power magnitudes (usually the case for transmission networks) or current magnitudes (usually for distribution networks).

On the distribution grid, thermal constraints can be written as:

$$|I_{ij}| \leq \overline{I}_{ij}, \quad \forall (ij) \in \mathcal{A}. \tag{2.10}$$

However equation (2.10) is more often re-written as:

$$|I_{ij}|^2 \leq \overline{I}_{ij}^2, \quad \forall (ij) \in \mathcal{A}, \tag{2.11}$$

The equation (2.11) is now smooth. Without using the current variable, it is also possible to bound transit using apparent power flows. This new formulation is particularly applicable for transmission networks as indicated before. The transit constraint thus becomes:

$$|S_{ij}|^2 \leq \overline{S}_{ij}^2, \quad \forall (ij) \in \mathcal{A}, \tag{2.12}$$

where $|S_{ij}|^2 = P_{ij}^2 + Q_{ij}^2$. This latter constraint is smooth and convex. Moreover, thermal constraints usually are applied on power variables in models for transmission networks as noted in [117], which explains why these constraints are often used in literature. In [196], the authors discuss the relation between apparent power limits and current limits. This discussion is valuable as both types of limits coexist. Power limits are usually considered on HV networks, whereas current limits are considered on MV networks. As a consequence, one cannot discard one type of constraints, as both are useful for OPF studies, and the authors discuss the differences of one choice over the other. For instance, the feasible sets differ when one chooses one type of transit constraints in place of the other, which implies that for a given OPF one might get different results depending on this choice of constraints.

We care to point out another approach to thermal limits: in [133] no transit limit is applied on a Transmission network, but both active and reactive powers are upper and lower bounded. One may assume that these latter constraints, jointly with the assumption of a appropriately designed "robust" network, will lead to safe power flows.

Taking all constraints into consideration (voltage magnitudes and phasors, current or power flows, box constraints on power consumption and generation), the authors of [206] propose a method to "learn" the active sets at the optimal point of an OPF. This method has the interesting property of preserving the quality of the model, and maintaining relevant safety constraints, and giving less importance to constraints that are not likely to be binding at optimality.

A generic OPF does not take into account other constraints apart from the linear and thermal ones. There are nonetheless other subsets of problems from the "OPF family" in which additional constraints are considered. One example in the Security-constrained OPF (SCOPF), where the additional constraints account for post-contingency cases. The aim of a SCOPF is to provide a solution which remains feasible after a failure from a set of predetermined failures (e.g., the loss of a generator or of a transmission line). This particular subset of OPF problems has attracted a lot of researchers and has been identified as a key problem from the OPF family: in [66] where the authors discuss the recent advances in the OPF field, SCOPF is described as somewhat indistinguishable from the OPF as a whole since "most real-world OPF problem consider security constraints[4]". This explains why SCOPF has been at the center of works from the 1970's (see for instance [31]) to modern days (see [251, 32, 40] where the post-contingency constraints are usually treated in a *robust* way).

2.3 SOURCES OF UNCERTAINTIES

The electrical system is one of the largest human-made systems comprised of complex elements connected to one another, on top of being among the most in interaction with humans. As a consequence, sources of uncertainties are astonishingly numerous. Moreover, and somewhat unlike other large human-made systems, the subset of these sources that are "significant-enough" to be considered still is consistent. For instance, in the airline industry, generic crew pairing optimisation studies can (hopefully) safely disregard the possibility of a plane not reaching its destination. The outputs of events that are not expected usually include delays and loss of revenue for the company, possibly amplified by a domino effect, one delay leading to another one. On the other hand, in the electrical system, incidents resulting in the loss of some arcs in the network are more frequent and optimisation models taking such losses into consideration are of significant interest. The possible outputs of a material loss are also of greater concern due to the our dependence on the grid.

Limiting ourselves to short term operational planning[5], sources of uncertainty include:

[4] Security constraints are to be understood as post-contingency constraints in this context.

[5] Uncertainties evidently are hugely dependent on the time horizon. For instance, in the long term, political, social, technical decisions impact the development of the electrical grid, these aspects are not significant on a short term.

- Errors on characteristic parameters of grid elements. For instance, R and X values for a given line might be wrong or missing/unknown in an operational setting.
- Model errors: these errors are numerous, and encompass physical laws (e.g. Ohm's law as a linear function provides insight on the relation between values of voltage and intensity up to a precision generally deemed sufficient, in a system which conditions are appropriate[6]) as well as the mathematical model. Among model errors, ENEDIS financed a study in 2018 in the PhD work of J. Buire [60] on the uncertainty related to the On-Load Tap Changer. French networks often have On-Load Tap Changers that are control loop processes, following a selected voltage value for the secondary windings by automatically changing taps. The automatic rule, shifting step-by-step the selected taps, introduces hysteresis in the behaviour of the transformer, more importantly it is unknown which taps are currently being selected. The PhD thesis proposes a model for this uncertainty and its spread throughout the network. A key element from this work is that there are realistic cases where the uncertainties due to the selected taps are not negligible when related to forecasts uncertainties.
- Forecasts on production/consumption: it appears to be among the most studied source of uncertainties, and the increasing deployment of distributed renewable energy sources (DRES) as well as the growing current of electrical usages further stress the importance of studies on these sources of uncertainties.
- Numerical and coding errors: these we only tackle by repetitive verifications.

In this work we are interested in the forecasts errors/uncertainties. Our model for the loads and producers are fairly common on Medium voltage grids as they are treated as p, q grid elements. This obviously comes with associated model errors that we argue are beyond the scope of this work. A possible extension, where loads would be better modelled, could probably rely on [57] where the authors describe the "ZIP" load model, and useful data for the parametrization of this model. It is intended as a better voltage dependent description of the steady state of loads.

2.4 THE OBJECTIVE FUNCTION OF THE OPF

Compared to constraints, which in general are sources of non-convexity for the OPF, the objective function is usually not a source of numerical difficulties. The canonical OPF aims at minimising cost of power generation, and the vast majority of works studying numerical optimisation methods share this objective function: this is easily observable from, for instance, the extensive surveys [73, 117, 209].

Other classic objectives for the OPF include the minimisation of power losses ([263, 309, 167, 77]), improvement of voltage profiles ([78]), maximisation of profit when the operator acts on a market ([10]), maximisation of renewable DG penetration ([226, 68, 207]), or materials sizing ([43, 166, 140]) / allocation ([45, 119, 166, 297, 140]).

[6] There are many electrical devices than do not have a linear law between voltage and intensity, a "Tunnel diode" being one of them.

As described in [10], most of the encountered objective functions are linear, piecewise-linear convex functions, or polynomials (mostly quadratic convex); nonconvex functions are used in more specific cases which are discussed in this latter reference.

2.5 NUMERICAL METHODS TO SOLVE OPFS

From the original OPF definition at the beginning of the 1960s, more than 8.000 research publications have been proposed (see Figure 2.5 for the time distribution) specifically on the OPF subject. Although our Figure 2.5 solely relies on the Scopus database and consequently should not be regarded as an extensive analysis, well established operational research and energy related journals are included as ones affiliated to IEEE or IET. The trend is therefore significant: more than 2000 scientific documents have been published on the OPF for the past 3 years, more than 5500 for the past 10 years.

In this section we are interested in the numerical methods that have been used to solve OPFs. Although our interest lies mostly in mathematical methods, heuristics make up a large part of literature. We start by recalling some basic properties of the OPF, based on previously introduced material in this chapter. After a brief discussion on heuristics applied to the OPF, we then present the historical approaches and current trends.

An abstract OPF presentation is as follows:

$$\begin{aligned} \min_{x,u} \quad & f(u,x) \\ \text{s.t.} \quad & g(u,x) = 0 \\ & h(u,x) \leq 0 \\ & \underline{u} \leq u \leq \overline{u} \\ & \underline{x} \leq x \leq \overline{x} \end{aligned} \tag{2.13}$$

From Sections 2.2 and 2.4, it is known that g is a nonconvex function, and h is possibly nonconvex.

INSERT 4: Complexity of the AC-OPF.
In this formulation, u can be a mixed integer vector, x is a continuous variable, f and g are in general non-convex non-smooth functions. The canonical AC-OPF has been shown to be *strongly NP-hard* on general graphs, and *weakly NP-hard* on trees [51]. More specifically, the latter article presents a formal proof of the strong NP-hardness of AC-OPF by a reduction to an another NP-hard problem (the *3-SAT* problem).

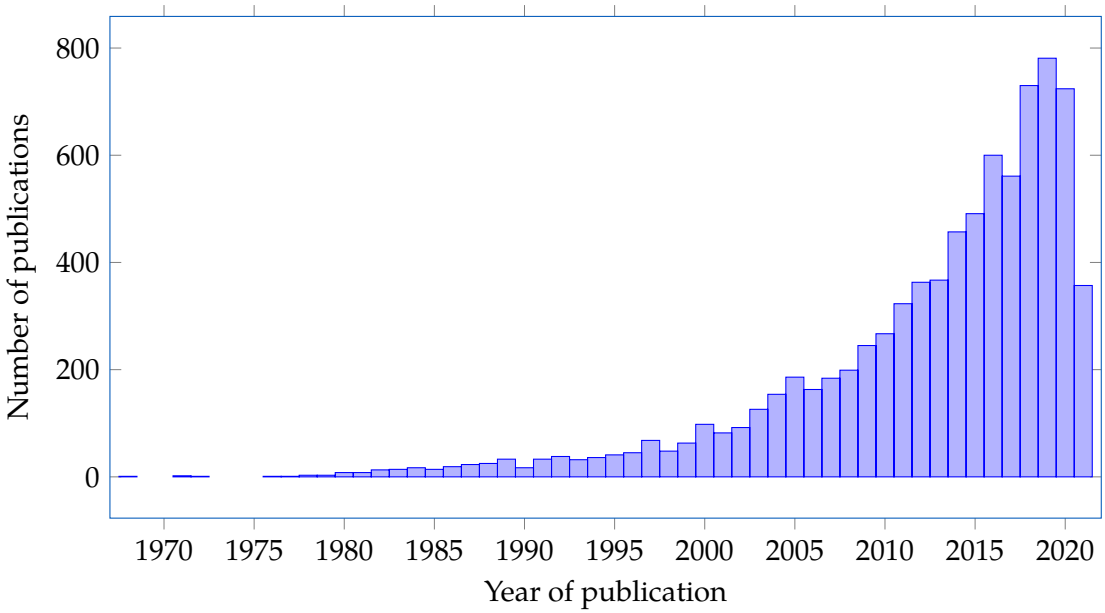


FIGURE 2.5: Number of publications with "Optimal Power Flow" in their titles, keywords or abstracts when searching on Scopus. Document types include mostly articles (4330), conference papers (3541), conference reviews (100), book chapters (83). Year 2021 is truncated: publications are counted up to the first of July 2021.

2.5.1 Overview of the state of the art on the OPF as an optimisation problem

The original OPF formulations were non-convex and the first numerical methods did not tackle the associated program without relaxations. In the impressive 1968 article [92] [7], the authors propose a decomposition scheme to retrieve a "good" feasible point for the original OPF. Using Newton's method to solve the PF at a given iterate, they then use the output of the first step to solve an unconstrained OPF with a penalisation of inequality constraints. If the stopping condition is not met, they update the control parameters estimate, and proceed to the next iterate. Interestingly, the latter article already emphasizes the scaling difficulties in numerical experiments for the OPF.

From this initial heuristic, the authors of [31] build a security-constrained version of the OPF. Their objective is to ensure the safe steady-state exploitation of the system after a contingency and associated automatic control devices activations, but before slower decisions (as decisions that require a human input). To that end, the authors rely on the methodology from [92] to iteratively solve an OPF. At each iteration, they penalise decisions that would lead to insecure grid state post-contingency. This can be related as an iterative two-step optimisation: at each iteration the first step sets values for decision variables, while the second step is on monitoring the grid state post-contingencies. If an insecure grid state is observed, a penalisation of the associated set of decision variables is applied.

J.L. Carpentier, who initiated the OPF also provides a still highly relevant review in [73]: from the canonical (full) OPF, he presents several approximations that are in use

[7] This article cites other works, claiming that they propose "numerical methods" to solve OPF related problems, but these works are (1) not easily found, (2) mostly in German or Russian.

30 years on. The challenges that are foreseen in this article are still faced by researchers today: the author identifies the ability of an algorithm that can provide a "feasible close to optimal" solution on a "minicomputer located at the energy control center" as a dominant research objective.

To that end, the OPF has been at the crossroads of evolutions in optimisation algorithms and modelling. Extensive and close to up-to-date surveys are available, and an interested reader should turn to the following one for more references and details:

- The authors of [117, 118] produced two milestones in 2012, that provide an extensive overview of deterministic optimisation techniques and heuristics applied to the OPF.
- In [10], a large and recent state-of-the-art of the Unit Commitment problem is provided.
- A more recent and mathematical review of OPF formulation can be found in [50].
- References [64, 66] are practical, more operational critical points of view of the OPF.

Optimisation algorithms for the continuous OPF

In this subsection, we are interested in optimisation methods applied to the continuous OPF. This excludes heuristics, even though it appears a significant amount of articles in literature do not separate heuristics from optimisation methods. Global trends in deterministic optimisation algorithms that have been applied to the OPF in its continuous formulation can broadly be reduced to:

- Gradient methods, and related algorithms. They have been among the first ones to be applied to the OPF in the 1960s, and still have been used recently (for instance [92, 187, 195]). Newton's methods, which use second order information, also have been used for the OPF [281]. Taking a broader view on these methods, Newtonian methods are extensively introduced in [159, 160].
- Sequential Quadratic Programming (SQP), and related algorithms. This class of algorithm has been developed from the 1970s, and is among the most successful for nonlinearly constrained optimisation problems (see [55], [186, Part III], and [56] for a thorough description of SQP).
- Interior Point Methods (IPMs): this class of methods applies to convex problems. A sequence of subproblems is iteratively solved, where each iterate is obtained by incorporating inequality constraints in the objective function using a "barrier" function. This latter function penalizes the violation of the inequalities, and usually is chosen as the logarithmic function. IPMs have proven to be numerically efficient for convex problems in terms of complexity and CPU times (see for instance [219], and more particularly [219, Section 4.3.6]).

Of these methods, the gradient ones are the first optimisation techniques to have been applied to an OPF problem. In [92, 274] for instance, the authors propose optimisation techniques using first order information to solve an OPF where the constraints are

penalised in the objective function. Numerous evolutions of gradient-based techniques have since been applied to the OPF: a list of reduced gradient, conjugate gradient, and generalized reduced gradient methods is extensively in the 2012 reference [117]. Gradient based methods are reliable, have been extensively studied since the 1960s, are easier to implement than other algorithms, and only require first order information on the functions. Convergence is also guaranteed, but is both slower and weaker (one cannot prove local optimality with first order information) than higher order methods which use more than first order information. Thanks to its robustness and reliability, this class of method is still used in recent works: for instance, in [195], the authors propose a Lagrange projected gradient for a generic (deterministic) OPF problem. This method relies in part on projected gradient method, in which a gradient step is followed by a projection onto the feasible set. The authors of [105] use a (sub-)gradient method to solve an OPF with automatic local voltage control. In [121], the authors propose to apply a gradient method to a logarithmically penalised OPF and are able to achieve significant speedups in CPU times (from 5 to over 70 times faster) while reaching objective values close to the ones obtained with other relaxation methods from the literature. Leaving aside the validity of the state variables at the solution point as the authors only present the objective values without discussing the exactness of their relaxed model, this work highlights that gradient methods are still relevant for the OPF.

In a different approach, sequential quadratic programming methods solve a non-convex problem by iteratively solving quadratic problems, called subproblems. Each subproblem is a quadratic approximation of the original problem around an iterate point x^k. Since the 1980s, SQP has been applied to the OPF (see [117, Section 4.5] for a broad overview of SQP applied to OPFs from the 1980s to early 2010), and is still developed nowadays for commercial OPF solvers. Sequential quadratic programming has recently been successfully applied to a Direct-Current OPF (DC-OPF) in [213]. In this latter work, the authors show that this specific type of OPF is particularly well suited for SQP methods as the DC-OPF, under "normal operative conditions", allow very accurate linearisations of the non-convex constraints; as a consequence, in their proposed model, only the objective function is quadratic while the constraints are linear. In a similar setting, the authors of [130] propose a tailored SQP approach while considering more precise load models (namely voltage depend models). In both latter works, the SQP algorithm is explicitly used for its numerical efficiency by the authors, in particular when applied to quadratic convex approximations of the OPF. For another specialised OPF, where stability of the solution to the OPF is controlled, the authors of [194] also rely on a SQP method. In majority of SQP applications to the OPF, the key element is the linear approximation of the non-convex constraints. It usually is done using Taylor approximation of the functions at the current iterate. More recently, reference [192] presents a more complex procedure to obtain local convex approximations of the non-convex constraints. It relies on the main axis transformation, a method that enables one to distinguish non-convex parts of a function from convex ones. The authors then linearise the non-convex parts, to obtain a convex approximation of the non-convex constraints.

Lastly, IPM methods which were initially developed for linear programs, have since been expanded to non-linear programs. They gained much interest due to their numerical efficiency when compared to the Simplex Method for linear programs, which at the time was the mostly used algorithm. Moreover, there exists an upper pseudo-polynomial bound on the worst-case running time (for this matter, see for instance [30]). In the

OPF community, IPM has been applied since its early beginning (as a replacement of the Simplex Method for linearised OPFs) and now is widely used in more advanced versions. Several of these improved IPM algorithms are applied and compared in [70]. Reference [74] extends this previous work by proposing an IPM algorithm for a security-constrained OPF. Both [117, Section 4.6] and [50] highlight that IPM algorithms for OPFs are currently widely used and improved in literature: as an example, one of the most renowned package for electrical network optimisation [310] relies on IPMs.

As already described, the difficulty of solving the OPF, combined with the high stakes residing in the development of an efficient solving method, explains the large number of works addressing this subject. The multiplicity of optimisation algorithms from the operational research field is combined with the multiplicity of models for the electrical network and its material side. In this overview of the main optimisation algorithms, we have left aside two main possible aspects for the general OPF: the use of integer variables, and optimisation algorithms that search for global optima (even though the original OPF is non-convex). For the first part, mixed-integer OPFs are an active and challenging optimisation problems. It has historically been tackled using heuristics (solving the relaxed problem and rounding to the closest discrete value, which can lead to infeasible solutions [67]). Optimisation techniques for the mixed-integer OPF rather rely on linearisation of the OPF, which is in turn embedded in a branch-and-bound tree (see for instance [67, 240], or [278] where a convexification in place of the linearisation is proposed). Another successful and tractable method for the mixed-integer OPF is ot be found in [271], where the authors iteratively solve continuous non-linear problems with a penalisation of variables that should be discrete but take non-integer values. For the second part, some authors indeed have worked on exact optimisation techniques, which we discuss in section 2.5.2.

Optimisation techniques constitute one side of the OPF, the other being the model itself. As the non-convexity is difficult to tackle, relaxations have an important role in solving the OPF. The following subsection is therefore dedicated to a review of existing relaxations.

Simplifications for the OPF

Due to the non-convexity of OPF problems and the difficulty of numerically solve an OPF, researchers have intensively worked on relaxations of these programs. The goal of relaxing a mathematical problem should always be twofold:

(1) obtain a relaxed problem which is easier to solve than the original one,

(2) be able to derive an upper bound on the optimality gap between the original problem and the relaxed one.

In other words, one should be able to control how "wrong" is the relaxed problem when compared to the original one, the best situation being when the optimality gap is zero.

Among the first relaxations is the DC approximation[8]. Assuming that:

- $X_{ij} >> R_{ij}$ for all lines (ij);
- $\sin(\delta_i - \delta_j) \approx \delta_i - \delta_j$ and $\cos(\delta_i - \delta_j) \approx 1$ for all lines (ij);

[8] We recall that DC stands for Direct Current.

then one can observe that the obtained model is linear: a review, and improvements on this relaxation can be found in [95]. This has shown success as this relaxation, although without general zero-gap property, still enjoys good quality solutions (see [231], where DC-OPF is applied in a market setting, where the objective is to localize congestion in order to activate closest levers). A more recent article [41] (still unpublished) proves that solutions to the DC-OPF cannot be feasible for the AC-OPF under (very) mild assumptions, namely: positive resistances and reactances, symmetric admittance matrix Y, strictly positive line losses and only p, q loads. One could argue in return that load models also have significant impact on feasible solution to the AC-OPF (recall [163]). Nevertheless, knowing that DC-OPF can be linear, can produce satisfactory results in some cases and that fast iterative methods to improve solutions *a posteriori* exist (detailed in [95]), it is still widely used today.

A *Second-Order Cone Programming* (SOCP) relaxation has been proposed in [162] without providing a zero-gap proof: it is stated that convergence to the "true load flow solution" is achieved in practice. But relaxations have gained significantly more attention from early 2010 with [179, 180] where the authors prove that a *Semi-Definite Programming* (SDP) relaxation of the OPF has a zero-duality gap. This immediately implies that, on some cases, one could efficiently solve the AC-OPF to global optimality with usual laptops. More generally, a set of conic relaxations have since been proposed. In [84] the authors propose a *Quadratic Convex* (QC) relaxation, while more recently [53] presented an improvement of both SDP and SOCP. Leveraging a recent reformulation of non-convex quadratic constraints by its explicit convex hull proposed in the mathematical literature, the authors prove their relaxation (nSDP) is a trade-off between SDP and SOCP. It has since been applied to other OPF problems in [52]. Generally speaking, as detailed in the latter reference, one has the following relations between the relaxations:

- SDP is numerically the slowest relaxation, while SOCP is the fastest. QC and nSDP both are in-between SDP and SOCP on the time of computing scale.
- SDP is the tightest relaxation, while SOCP is not as tight. Regarding the model strength, QC is not dominated nor dominates SDP, while nSDP is claimed to be "equivalent" to SDP.

Related relaxations have been studied for some specific network configurations (namely radial ones). For instance, in [208, 268] SDP relaxations are obtained thanks to the sparsity of the network, when the OPF is formulated as a system of polynomials (thus using rectangular coordinates, see section 2.1.2). An additional procedure applicable on polynomial optimisation problems (called "Lasserre hierarchy", see the references in [268] for a better insight of this concept) enables the authors to strengthen their relaxation. The relaxation that is obtained with a first-order relaxation in the Lasserre hierarchy is equivalent to the SDP one. More discussion on the Lasserre hierarchy and its relation with the OPF can be found in [203, Section V].

To conclude on the subject of relaxations, the goal of relaxations applied to the OPF usually is to obtain a continuous problem from a mixed-integer one, or a linear/convex problem from a non-convex one. Linear relaxations have long be used, as direct-current formulations of the OPF can lead to a linear problem. Convex relaxations have gained much interest when some authors proved there exists, in a particular setting, SOCP/SDP relaxations with zero-duality gaps.

Subclasses of the OPF and other related problems

We have thus far explored the main optimisation methods for the OPF, as well as the usual relaxations applied to this optimisation problem. We have thus distinguished the OPF on the resolution process, which is one way of analysing this class of programs. Another possibility is to study the different types of programs that belong to the OPF class while having different motivations. In this subsection, we overview the main subclasses of OPF which include the following ones:

- Security-constrained OPFs.
- Reactive-power OPFs.
- Unit commitment.
- Economic dispatch.

Security-constrained OPF Broadly speaking, an OPF is said to be "security-constrained" when it is designed to account for contingencies such as a power plant outage or the loss of a power line. The objective of a security-constrained OPF is consequently to provide an optimal solution that remains feasible after the occurrence of such a contingency. The set of possible contingencies is known/defined *a priori*. As a general rule, the solution of a security-constrained OPF will lead to a worse objective value when compared to that of the equivalent standard OPF (because the former is more constrained than the latter). Works in security-constrained models also enjoy a vast literature as they can be traced back from at least 1974 ([31]) and is still active nowadays ([64, 66], and more recently a "stochastic" version of the security constrained OPF [29]). Additional "security"-constraints can be interpreted as a way of taking some uncertainties into account: solving a security-constrained OPF is aiming at a decision of minimal cost, that is feasible for a known-set of possible contingencies U. This latter interpretation highlights that these models are related to "robust" optimisation.

Reactive-power OPF While reactive power has a "zero-cost", a sub-set of literature is dedicated to solving OPFs: the rationale is that in a first-order approximation, reactive power has a greater impact on voltage magnitude than on voltage phasor. As a consequence, grid levers that modify reactive power can be used as a means of mitigation of voltage magnitude if and when needed. This is also known as the Volt/Var control, and is of great importance due to its ability of providing efficient voltage control without network reinforcement (see [308, 209, 236] and references therein, which highlight the involvement of a IEEE taskforce on this subject). The Optimal Reactive Power Dispatch is the OPF program that aims at finding the optimal activation of reactive power levers on a grid. Mathematically speaking, this often amounts to solving non-convex mixed-integer programs (MINLP). This is highlighted in [52] where the authors propose a conic relaxation of the MINLP and a rounding procedure to obtain the integer variables. Similarly, in [90] the authors study conic relaxations of the OPF for the Reactive Power Optimisation with uncertainties. Reactive-power OPFs are also studied from a planning ([306]) and micro-grid ([134]) point of view. A key point of this subclass of problems is that they capture some important material, operational and upcoming aspects of the network.

Unit Commitment Unit commitment (UC) and Economic dispatch (ED, see the next paragraph) are two optimisation problems that are closely related. They are two sub-classes of decision analysis for operational planning in the electrical (transmission) network that naturally appear when one seeks to decompose this difficult problem into smaller, easier ones at the expense of obtaining a sub-optimal solution to the original problem. In the transmission network, one has to decide which plants to start-up, shut down, keep idle/active: these are known as the *commitment* decisions, which cannot be modified in a short-term horizon. This longer-term step has historically been the UC goal. Once the solution to this problem is known, the ED provides the decision on levels of production of active plants (the *production decisions*). This traditional distinction is not as clear nowadays, as one can see in the extensive reviews [280, 10]. One can now numerically tackle a UC with commitment and production decisions. Mathematically speaking, UC problems are non-convex mixed-integer programs over multiple time steps as there usually are some *ramp constraints* (see [114, 115]). It is not considered to be a "well-solved" problem in the deterministic case, and uncertain cases are still scarce (see [10, Section 4]).

Economic Dispatch The Economic Dispatch (ED) is intrinsically related to optimisation problems of energy systems, it is not always a subclass of the OPF strictly speaking. First of all, note that the ED is a term also encountered in other fields than optimisation related to electricity networks. The ED conceptually is the economically optimal allocation of levels of production to meet levels of demand. Some authors include PFs into this problem (see [10, Section 2.1] for a short discussion on the subject, [116, Section 3.1] which presents the security-constrained ED as having PFs, or [76] where PFs are linearized) in which case the obtained ED problems are indeed OPF problems. Other authors do not consider PFs (see [114, 299, 224]). Note that not taking PFs into consideration amounts to ignoring the electrical network. In that latter case, a simplified approach often encountered is the Merit Order methodology which amounts to a production price ranking matched to a consumption price ranking as shown in Figure 2.6.

In more complex cases, ED can lead to non-trivial optimisation problems even when the network is still ignored: generators can have non-convex efficiency curves, as shown in [24] for the hydro power plants case.

2.5.2 The OPF as a global optimisation problem

Even though relaxations and studies to provide bounds on the distance from the solution of a relaxation and the one of the associated non-relaxed problem appear to dominate recent scientific developments for the OPF, several researchers were interested in finding global optima of OPFs. Evidently, the authors in this field are interested in cases where relaxations with zero-gap fail, which occur in realistic cases: in [191] provides an explicit example over a 3-bus network where a recent semi-definite-programming relaxation fail to have a zero optimality-gap when a thermal constraint is considered. This is a drawback for this particular convex relaxation, as thermal constraints are important for both TSOs and DSOs. The authors of [191] build their counter-example by considering a large thermal constraint (where there is a zero optimality-gap), and a realistic but tighter one (where the gap is not zero). This is further developed in [136] where the authors directly work on open-source IEEE networks to set up cases with no zero-gap relaxation.

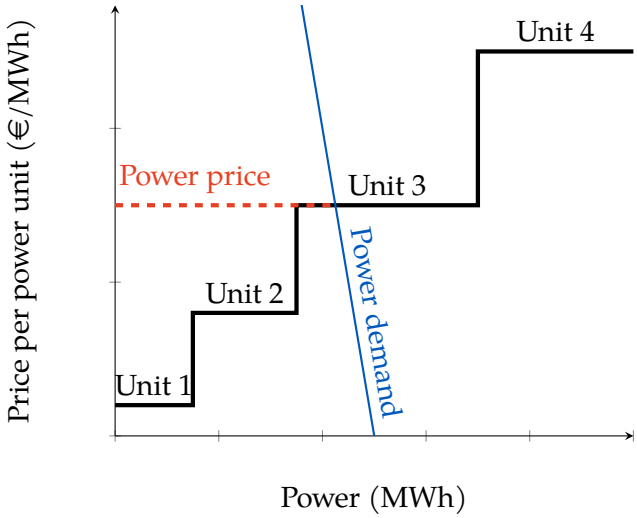


FIGURE 2.6: Merit order example graph. Production costs of the available power units (e.g. their marginal costs) are ranked, and the customer price as a function of power supplied is plotted. This latter plot is to be related to an elasticity of demand. In this simplified model, the intersection of the two plots provides the price of power in this market.

Moreover, the authors of [62] explicitly provide material that highlight several cases where semi-definite relaxations fail, along with their associated local minima. The last three references work on the key element of recent relaxations where zero-gap is possible: the ability of dropping a rank constraint, while still obtaining an exact solution to the original problem.

As is usual when seeking for a global optimum in the absence of convexity, works addressing the global optimisation of the OPF mostly rely on a *branch-and-bound* algorithm. This latter algorithm is time-consuming, but still provides "robust" results and has long been addressed in literature. This application to the OPF is discussed in [132, 133]. In these two works, the authors detail a branch-and-bound algorithm for three different OPF problems with (1) linear objective function, (2) quadratic objective function, (3) binary variables for on/off status of producers.

Using Complex Analysis tools has seen some success in energy related problems, starting with finding PF solutions: the author of [284] proposes a novel deterministic non-iterative holomorphic[9] resolution with a "perfect" convergence. In other words, except in pathological cases where the electrical system is beyond voltage collapse, it is more likely to converge to the right solution. This is an improvement from traditional iterative methods, where convergence problems can arise: it can either be very slow, or converge to a physically non-realistic but mathematically feasible solution to the PF. Building on this article, the PhD thesis [108] uses the same complex analysis to study the number and interpretation of all feasible solutions to a given PF system. Following the intuition of the author of [284], who predicted his novel methodology could be applied to the OPF, the authors of [177] propose an interesting methodology that tackles the (full) AC-OPF. Leveraging the strong convergence properties of the holomorphic resolution, they are able to address a deterministic mixed-integer OPF.

[9] Using complex-valued functions.

2.6 THE OPF FROM A DSO POINT OF VIEW

The OPF was traditionally formalised from a transmission system (and hence TSO in current terms) point of view, as at the beginning of the second half of 20th century the distribution actors obviously did not experience the same challenges as today. This might explain why specific DSO-oriented OPFs have long been significantly less studied than TSO-oriented ones. This section is dedicated to the specific challenges that a DSO faces, and works in literature that address them.

We identify the following elements as both (1) possible to model in an OPF, (2) interesting to model when taking a DSO point of view:

- Power injections are not directly controlled, but are levers that can be activated following constraints defined in pre-existing contracts.
- The same holds for loads on the grid: modulation is assumed to be possible only through contracts.
- Transit constraints should be on current flows (and not apparent power flows).
- Some of the characteristic parameters of the electrical grid, resistance R and reactance X of the lines, do not usually verify the following property $\frac{R}{X} << 1$. As a consequence, a DC-OPF is not well suited for the DSO, and the AC-OPF appears as a better problem to tackle.

These four elements are necessary to appear in an OPF, and are taken into consideration within our work, but one should note they are not sufficient for an ideal OPF for the DSO. One additional consideration we have not investigated is to consider the possibility of having an unbalanced electrical grid, whereas in our case we make the implicit assumption that the grid is balanced between the three electrical phases. Techniques for OPFs considering unbalanced networks can be found in [247] or in [85] where a low-voltage grid is considered (LV networks are natural use-cases for unbalanced OPF). Another important consideration in an ideal OPF for the DSO should be the ability to take integer variables into account, as several levers available to the DSO naturally require discrete variables to be modelled (as line switches for instance). As the main goal of this work lies in the uncertainties, we have chosen to discard this latter consideration.

Literature is not yet abundant on DSO oriented short-term operational planning. For instance, reference [204] proposes a reformulation of the OPF based on the definition of a Hosting Capacity. Starting from the ground rules of a DSO (broadly speaking, it has to authorize a maximal activity from GU while ensuring exploitation constraints), the authors present a model to maximize the level of power injection. Interestingly, thermal constraints are on the current and not on power flows, which is closer to reality. The constraints are the full compliance with network's exploitation limits, the safety DSOs' interactions with local producers are regulated by contracts. Other works are rather interested in DSO-TSO communication and exchanges, and used to be in another community than the optimisation one (see for instance the European Project TDX-ASSIST, and reference [129] which discusses a framework for an efficient coordination TSO-DSO). Recent works on the other hand create links between these DSO-TSO exchanges and optimisation techniques: [267] discusses the available quantity of flexibilities at the

interface DSO-TSO, while [65], although adopting a TSO point of view, introduces the interaction mode between the DSOs and the TSO in an AC-OPF.

These latter works highlight the growing importance of the DSO as an actor in the optimisation community. We expect that future works will either include more and more DSOs in AC-OPF with a TSO point of view, or propose solving methodologies well fitted for the DSO particularities as taking into account uncertainties.

2.6.1 Comments on decision making in an uncertain setting

As described in [278, Section 4.4], an *ideal OPF* formulation for the Operational Planning should meet the following requirements:

- Computational tractability. This is in particular true for short-term operational planning, where the user can be required to compute a solution up to a few minutes before real-time.
- Meet optimality requirements. As described in its roles, the DSO is required to manage the grid at an optimal cost.
- Provide an accurate solution. Accuracy is the relation between the solution of a mathematical model and real-world values from the modelled system. It is evidently necessary that this relation be "close enough" to ensure that the model is indeed a tool for decision analysis. How "close enough" depends on the user.

These latter requirements are completed by additional considerations that appear more particularly in an industrial context. Summarized in figure 2.7, they include the analysis of the existing material *a priori*: for the identified problem, what is available to the one in charge of the decision analysis? More detailed, follow-up questions, include:

- How much time is made available to design a solving methodology? How much time can be allocated for one resolution? What solvers are available? This is the "Ressources" box in figure 2.7.
- What information is made available *a priori*? Is there any *expert* knowledge that can be useful? This is the "Information" box in figure 2.7.
- What is the objective of the end-user? This is the "User problem" box in figure 2.7.

Once these questions are answered, the resolution process is ready to be unfolded: it includes a model set-up, and a solving strategy for this latter model. The model is evidently dependent on all the previous questions of the Initial State; the solving strategy is dependent on the allocated resources.

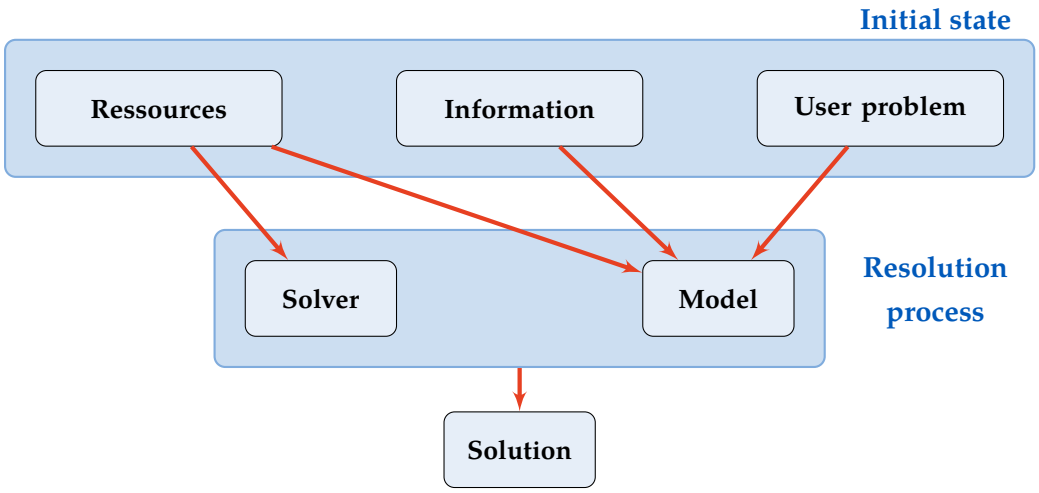


FIGURE 2.7: Main steps from the identification of a problem to solving it. Dependences are depicted by orange connecting lines.

Aside from highlighting some of the most sensible points of developing a decision analysis in an industrial context, figure 2.7 also enables us to point some elements where literature and industrial requirements are not aligned. On the one side, the "User problem" box is significantly developed in literature: solely looking at literature related to optimisation, numerous operational problems are addressed by a wide variety of methods. For instance, the authors of [117, 118, 116] provide an extensive review of problems related to the OPF and associated solving methodologies. On the other side, the "Information" box is more scarce: even if recent technological developments have enabled better data collection on grid topologies, consumption and generation levels, access to such information is limited due to proprietary rights. Some network topologies are both realistic and available in literature, most of which can be found in the dataset of MATPOWER [310] with mostly transmission networks. Two other openly available databases have been very recently made available. Firstly the authors of [178] propose a large collection of networks topologies and associated characteristic values. Interestingly, these networks closely resemble that of distribution grids. The downside is that only load-flow results are proposed, and not that of OPF; as a matter of fact, no lever is modelled which implies that no generation / load modulation costs are presented. Secondly, the work in [38] (yet to be published) is the result of a large collaboration of researchers from IEEE PES PGLib-OPF Task Force where OPF are conducted on 35 test cases. Results provided include the minimal costs, time of computation, and known optimality gaps. On the downside for a DSO point of view, this survey only deals with transmission networks: it should nevertheless be seriously considered for the assessment of OPF solving methodologies.

2.7 TAILORED METHODOLOGIES TO DEAL WITH AN OPF WITH UNCERTAINTIES

Modelling uncertainties and solving associated mathematical programs have obviously interested a large number of researchers: literature is vast on OPF with uncertainties.

In this section we present works that specifically address OPF with uncertainties. As described in part III, considering uncertainties in an optimisation problem is a holistic process. From the available information and expected decision analysis we classify these works into three different sets, depending on the type of methodology: 1. heuristic; 2. robust; 3. stochastic.

A major difficulty that arises from solving the OPF with uncertainties is ensuring that an *a priori* decision leads to a viable grid state as the link between this latter decision and the *a posteriori* grid state is a non-convex non-smooth one. This is discussed in the review of solving methodologies for the OPF under uncertainties [248].

2.7.1 Meta-heuristics for the OPF with uncertainties

We consider meta-heuristics to be methodologies that aim at finding a solution without certificate for optimality for a given resolution[10]. Such a certificate is to be understood in a *generalized* meaning, in particular in a non-convex setting: "global optimum", "local optimum" or "stationary point" are for example three different types of optimality certificates. Unlike usual optimisation techniques in mathematical programming, meta-heuristics are less concerned with associated a tailored optimality certificate to their end points. In other words, one cannot theoretically ensure any optimality-derived property of a point provided by a meta-heuristic. This obviously is a major concern: in complex optimisation problems, it is often difficult to humanly assess the quality of a solution. As a consequence, having a theoretical certificate that guarantees the quality of the obtained point is close to necessary. Adding that among the DSO's objectives is the pledge to manage the electrical grid at an optimal cost, the case for optimality certificates becomes even stronger. On the other side, meta-heuristics usually have the following advantages:

- It usually is more "intuitive" and easier to explain to non-specialists. As an illustration, one can observe the vast literature of meta-heuristics applied to OPF where techniques are often inspired by real-life phenomena: Genetic Algorithm as in [234], Particle-Swarm as in [3], Ant colony in [269], Bee Colony in [255], or Simulated Annealing in [272].
- As a consequence of the first point, it can relatively be easily modifiable and adjustable when applying the meta-heuristic to a new case.
- Meta-heuristics can be applied to a lot of mathematical problems, with a relative independence on their mathematical structure.
- As shown by literature, one can obtain satisfactory results using meta-heuristics.

While our work is not related to meta-heuristics, as these latter techniques are still very much used nowadays they should be mentioned. In [117] some meta-heuristics are presented and applied to the OPF with uncertainties, and [220] provides a more specific overview of meta-heuristics applied to this latter problem. Many of these meta-heuristics are either based on the iteration of a solution update (a succession of points

[10] Some methods we define as belonging to meta-heuristic ones can be proven to converge *in probability* to some optimal point: see for instance [262] for the proof of convergence in probability of some meta-heuristics to a global optimal, or [59] where the authors go further and study rates of convergence of the same class of meta-heuristics.

is calculated, and a point is updated only if an improvement criterion is met) or on a population-based search (many parallel processes are run at each generation, keeping only the best ones for the new generation). Drawbacks discussed in [220] include the difficulty to set parameters values, and the high computation intensity, to which we add the lack of theoretical certificate of quality of the solution. As further discussed in [79], the number of meta-heuristics applied in power and energy systems has been growing significantly within the last 10 years but gains are sometimes questionable. The authors of the latter reference recognize that meta-heuristics are in general difficult to compare and contrast in their current state, due to the "rush-to-heuristics" they highlight: a lot of energy is spent on proving the superiority of one meta-heuristic over another, while the concept of superiority requires a definition of "efficiency" which is yet to be clearly adopted.

Other methodologies are deterministic optimisation methods, which are algorithms with no random components that are applicable to programs with uncertainties.

2.7.2 Robust methodologies

A major subset of optimisation under uncertainties, Robust Optimisation (RO) has long been a thriving field for multiple reasons that span on several facets of a mathematical program. First of all, robust optimisation is a paradigm that provides a meaning to (1) an optimisation problem with uncertain parameters, (2) the solution to such a problem, which is not trivial. Secondly, on the numerical side, this paradigm can lead to more tractable programs when compared to other stochastic problems. And finally, robust models captures information on uncertainty as well as provides a solution than is often qualified as "conservative" which can suit large systems with human interactions and safety issues[11]. The core of RO is to provide a deterministic approximation of the uncertainty set.

While not at the core of our work which is rather centred on stochastic studies, we believe that RO and related framework will be thriving in the field of optimisation under uncertainties, and in particular applied to the OPF. First tractable methods and application of RO to OPF can be dated back to the early 2010s (see [69, 161, 217]). As usual in OPF with uncertainties, in these initial works the sources of uncertainties are mostly distributed renewable generators, followed by loads. In [69], the authors present what can be interpreted as a RO approach by developing an algorithm that select "most constrained scenarios" for the OPF where the levels of power injections are uncertain and provide a solution for this finite set of scenarios. An iterative process then update the solution in order to comply to all selected scenarios. The author of [161] then propose a more formalized model, closer to a canonical RO model. In this work, uncertainties solely are on the renewable energy sources, and it is assumed the operator has *here-and-now* and *wait-and-see* variables (respectively non-adjustable and adjustable after observing uncertainties). This is interesting, as adaptative automatic protocols can then be modelled. The model is based on a linearization of the physical laws in the wait-and-see step.

[11] The preface [48] is a classic introduction to RO.

2.7.3 Stochastic methodologies

Another main type of approach to optimisation under uncertainties is Stochastic Optimisation (SO). It has experienced a growing interest in the last decade due to new numerical codes and algorithms to overcome the numerical tractability obstacles of probabilistic models. A stochastic model intrinsically differ from a robust one because of the guarantees obtained in a solution to these models. While a solution to the former model enjoys a probabilistic guarantee (*the constraints are verified with a user-given probability* as in a chance-constrained model), the a solution to the latter has a robust guarantee (*the constraints are enforced for all uncertainty realisations deductible from the model*).

Stochastic models arise quite naturally when modelling uncertainties in an OPF, through the use of chance-constraints. The "natural use" of chance-constraints is discussed for instance in [23]: such constraints guarantee feasibility "as much as possible", which can appropriately translate the wish of an operator. This is not only true from the OPF, but also for other energy related problems (see [25, 22, 24, 280, 137, 135] for applications to various energy-related problems, including hydro-reservoir management and gas network problems).

First attempts at solving a stochastic OPF relied on linear or convex approximations. The authors of [305] appear to be among the first to look into tractable numerical treatment of a chance-constrained OPF. The chance-constraints are disjoint (see table 5.1 for a definition of this concept), and the probabilistic constraints are the boundary ones (upper and lower bounds on state variables). With these simplifications, the authors proposes a "two-layer approach": on the one side, an oracle that provides functional and partial derivatives of the involved functions; on the other side, a non-linear solver (a SQP algorithm was used), which takes the previous information as input and outputs the next iterate. The drawbacks of this initial approach include the fact the chance-constraints are not jointly considered, only the boundary constraints hold the uncertainties, and the sources of uncertainties are limited. In [290], the authors present a chance-constrained OPF that accounts for some predetermined contingencies. Their methodology relies on a convex approximation of the PFs equations, and a substitution of chance-constraints by a finite number of "hard-constraints" (see references [20-22] within this latter work). This numerical technique for chance-constraints is tractable, but has the disadvantage of not being scalable (a large number of scenario is desirable for accurate results, which implies that a large number of hard constraints have to be added). More recently, in [250] the authors propose a chance-constrained AC-OPF with local and partial linearisations. It is assumed that the forecast errors are small. This is again a tractable proposition, with individual chance-constraints on every boundary constraints. Power flow constraints are "locally" enforced: rather than having probabilistic power flow constraints, only a set of deterministic equations for the forecasted operating point is included in the model. On the one side, this model is fast to solve even on networks with over 100 nodes, but on the other hand it is not yet a fully and jointly chance-constrained one and the probabilistic constraints are not accurately computed. In [295], a work based on the PhD of the first author, the authors propose a (tractable) polynomial approximation of the chance-constrained OPF, relying on SDPs. Their model interestingly enforces (deterministic) constraints on the mean value of the random vector, as well as probabilistic constraints on every realisation of the uncertainty. Denoting the decision variables as x, the state variables as y_x, the random vector as ω with a zero mean value, the proposed

chance-constrained OPF is written as:

$$\begin{aligned}
\min_{x,y_x,y(\omega)} \quad & c(x,y_x) \\
\text{s.t.} \quad & f_i^0(x,y_x) = 0,\ i = 1,\cdots,m, && (2.14)\\
& g_j^0(x,y_x) \geq 0,\ j = 1,\cdots,k, && (2.15)\\
& \mathbb{P}\left[f_i(x,y_x,y(\omega)) = 0,\ i = 1,\cdots,m\right] \geq 1-\epsilon_1, && (2.16)\\
& \mathbb{P}\left[g_j(x,y_x,y(\omega)) \geq 0\right] \geq 1-\epsilon_2,\ j = 1,\cdots,k, && (2.17)
\end{aligned}$$

where $f_i^0(x,y_x) = f_i(x,y_x,0)$, $g_j^0(x,y_x) = g_j(x,y_x,0)$. One can observe that this model is mixed with deterministic constraints (equations (2.14) and (2.15)) and probabilistic ones (equations (2.16) and (2.17)). Moreover probabilistic constraint (2.16) is a joint one, while (2.17) is a disjoint probability constraint. The solving methodology relies on inner and outer approximations of the feasible sets defined by the chance-constraints.

Building on this review of literature, and in particular referring to [270], to our knowledge there is little work in the optimisation field on integrating uncertainties in short operational planning from a active-DSO point of view. Though this still is a prospective work, as a vast majority of DSOs cannot be yet considered as "active" (see for a discussion on “active network management” reference [278]), the list of new levers and of uncertainties due to an increased access to the grid is already vast.

2.8 CONCLUSION

This Chapter aims at providing a broad overview of the OPF and its numerical process. Literature is vast and an exhaustive description is a bottomless pit. We have highlighted some key-elements of the setting up the OPF as an optimisation problem. On the modelling side a large part of literature is dedicated to variables choices, and available relaxations. Significant advances were made possible thanks to convex relaxations and their associated optimality gap studies, which prove in some cases that this latter gap is equal to zero. Recent works also highlight the interest of using rectangular variables, due to new developments in polynomial optimisation. The OPF becomes significantly harder when considering uncertainties. It rapidly becomes clear that the different sources of uncertainties are numerous, and that an OPF that accounts for uncertainties will always include a choice of which ones to include, and which ones to discard. Modelling uncertainties is then about giving a meaning to an optimisation problem with uncertain vectors. This process is somewhat a multi-objective one: one looks for a model with an adequate level of uncertainty modelling according to needs of the user, while still maintaining tractability at the solving step. This last remark can explain the reason why, contrary to deterministic OPF where there is a widespread generic formulation, there are many different formulations of OPFs with uncertainties. It is particularly acute in stochastic programming applied to the OPF, and more specifically to OPFs with chance-constraints: probabilistic constraints often differ from one reference to another, some propose fully individual chance-constraints while other present a mix of individual and joint chance-constraints. The following elements are key for the remaining parts of our work: (1) there are not yet (fully) jointly chance-constrained OPFs, (2) this latter problem is of interest for the DSO, (3) new tractable numerical methods can be necessary to solve such a problem, that has yet to be proposed.

Part II

Difference-of-convex Programming

INTRODUCTION

One of the take-away of Part I is that when facing an OPF with uncertainties, one faces a nonsmooth nonconvex optimisation problem. Relaxations and approximations have usually been used to obtain tractable numerical methods. A recent work on the OPF with uncertainties without relaxations of the (difficult) PF constraints found in [295] highlights the numerical difficulties of this type of approach. Recent progresses in the area of nonconvex and nonsmooth optimization could provide better numerical tractability for these problems: Difference-of-Convex (DoC)[12] programming has proven to be an efficient structure for nonconvex optimisation problems. This vast concept encompasses basically every continuous optimisation studies that do not fold into convex (albeit linear) nor smooth fields. We recall that in case an optimisation problem is indeed convex or smooth, tailored approaches that take advantage from these properties will (in general) always outperform one that does not use them, counter-examples having to be considered as peculiar cases.

DoC programming has been active since at least the 1970's, with initial works on duality for nonconvex functions by John Francis Toland [282], and a pioneering introduction specific to DoC programming, with initial properties and challenges[13] by Jean-Baptiste Hiriart-Urruty is found in [152]. A numerical breakthrough was provided by an optimisation algorithm introduced in 1985 called DCA (Difference-of-Convex Algorithm) (see [183] which provides an extensive review of works on this matter). It still is a backbone for DoC programming, along with a somewhat similar DoC-tailored algorithm called CCP (Convex-Concave Procedure) (see [303, 197]). Most of the algorithms in DoC programs were initially developed for unconstrained/convexly constrained DoC programs. Extensions using penalization techniques have been proposed, with the difficulty of setting the penalization parameter to a "good" value. In order to tackle DoC-constrained DoC programs, we decide to consider bundle methods, which are very successful in nonsmooth optimisation, and more particularly to one bundle method that is tailored for convexly-constrained DoC programs [228]. Following another inspiring work on bundle methods [22], we propose an algorithm for DoC-constrained DoC programs.

The first Chapter of this Part, Chapter 3, is dedicated to introductory elements of DoC programming. Based on these preliminaries, we then propose our algorithm in Chapter 4. The DoC structure is interesting for its scope: every convex function is a trivial DoC function, and every linear combination of convex functions is a DoC one. As such, the set of DoC functions is the smallest vector space that includes the set of convex functions (see for instance [152]). This first example highlights the generality of DoC functions, which still maintains a strong structure: DoC programming is somewhat close to manipulations of convex functions. The intrinsic link between DoC and convex functions is fruitful as the latter setting is well understood, studied, and enjoys strong properties starting

[12] DoC is often referred to as DC in the mathematical field. In order to avoid confusion with DC (Direct-Current) from the electrical field, we use DoC in this work.

[13] Interestingly, some of the listed challenges are still open and seemingly difficult questions to answer, as on finding what could constitute a "good" (or the "best") DoC decomposition of a function.

with the regularity of functions at hand. The ingenious setting of a general structure built using particular elements with strong properties is a thriving dive into non-convex optimisation.

RÉSUMÉ EN FRANÇAIS

L'une des conclusions de la Partie I est que lorsqu'on fait face à un OPF avec des incertitudes, on est confronté à un problème d'optimisation non-convexe non lisse. Des relaxations et des approximations ont généralement été utilisées pour obtenir des méthodes numériques abordables. Un travail récent sur l'OPF avec incertitudes sans relaxations des (difficiles) contraintes PF présenté dans [295] souligne les difficultés numériques de ce type d'approche. Les progrès récents dans le domaine de l'optimisation non convexe et non lisse pourraient fournir une meilleure tractabilité numérique pour ces problèmes. L'optimisation de problèmes de différences de convexes (DoC)[14] s'est avérée être une structure efficace pour les problèmes d'optimisation non convexes. L'ensemble des problèmes DoC est réputé vaste, notamment car tout problème d'optimisation continue peut être approximé aussi proche que souhaité par un problème DoC. Nous rappelons que si un problème d'optimisation est effectivement convexe ou lisse, les approches adaptées qui tirent parti de ces propriétés seront (en général) toujours plus performantes que celles qui ne les utilisent pas, les contre-exemples devant être considérés comme des cas particuliers.

La programmation DoC est apparue de façon pratique depuis au moins les années 1970, avec les premiers travaux sur la dualité pour les fonctions non convexes de John Francis Toland [282], ainsi que [152] où l'auteur présente une introduction à la programmation DoC, les propriétés initiales et les défis à relever[15]. Une première avancée numérique a été réalisée grâce un algorithme d'optimisation introduit en 1985 appelé DCA (Difference-of-Convex Algorithm) (voir [183] qui fournit une revue extensive des travaux sur ce sujet). Il constitue toujours une pièce maîtresse de la programmation DoC. Un algorithme quelque peu similaire adapté au DoC appelé CCP (Convex-Concave Procedure) (see [303, 197]) proposé plus tardivement est aussi important. La plupart des algorithmes des programmes DoC ont été initialement développés pour des programmes DoC non contraints/convexes. Des extensions utilisant des techniques de pénalisation ont été proposées, avec la difficulté de fixer le paramètre de pénalisation à une "bonne" valeur. Afin de s'attaquer aux programmes DoC contraints, nous décidons de considérer les méthodes de faisceaux, qui ont sont réputées parmi les plus efficaces dans l'optimisation non lisse. Plus particulièrement, nous nous inspirerons d'une méthode de faisceaux qui est adaptée aux programmes DoC avec contraintes convexes [228]. En développant ce dernier travail avec un autre travail inspirant sur les méthodes de faisceaux [22], nous proposons un algorithme pour les programmes DoC avec contraintes DoC.

Le premier chapitre de cette partie, Chapter 3, est consacré aux éléments introductifs de la programmation DoC. Sur la base de ces éléments préliminaires, nous proposons ensuite notre algorithme dans le Chapter 4. La structure DoC est intéressante en particulier pour sa versatilité : par exemple, toute fonction convexe est une fonction DoC triviale, et

[14] DoC est souvent appelée DC dans le domaine mathématique. Afin d'éviter toute confusion avec le DC (Direct-Current) du domaine électrique, nous utilisons DoC dans ce travail.

[15] De façon intéressante, certains des défis énumérés sont encore ouverts et sont toujours considérés difficiles, comme pour la recherche de ce qui pourrait constituer une "bonne" (ou la "meilleure") décomposition DoC d'une fonction.

toute combinaison linéaire de fonctions convexes est une fonction DoC. Ainsi, l'ensemble des fonctions DoC est le plus petit espace vectoriel qui inclut l'ensemble des fonctions convexes (voir par exemple [152]). Ce premier exemple met en évidence la généralité des fonctions DoC, qui conservent néanmoins une structure forte : les méthodes de la programmation DoC sont largement basées sur celles développées de longue date dans l'optimisation convexe. Le lien intrinsèque entre les fonctions DoC et les fonctions convexes est fructueux car ce dernier cadre est bien compris, étudié, et bénéficie de propriétés fortes à commencer par la régularité des fonctions en question.

PRELIMINARIES AND CONCEPTS ON THE DIFFERENCE-OF-CONVEX STRUCTURE

3

This Chapter is divided as follows: the first section provides a broad overview of applications of DoC programming. Secondly, we recall some properties and definitions that are necessary to develop an optimisation algorithm for DoC functions. The main content of this Chapter has appeared in [9].

3.1 INITIAL DEFINITIONS AND CONSIDERATIONS

In this section, we recall some necessary definitions from the field of optimisation. Starting from an abstract presentation of an optimisation problem, we briefly provide an overview of some of the strong properties that have been used in optimisation. This overview highlights the existing continuity between the structure we study in our work (the *difference-of-convex* one) with the already very well-studied ones in literature. We also discuss the notion of *stationarity*, which is both central to the field of optimisation and not trivial to define. This section is divided into two main parts: in the first one we present necessary definitions on optimisation from a general point of view, some important structures (convex and differentiable functions). Discussing the limits encountered in the discussion this first part motivates our second part where we introduce our structure of choice, and its main associated properties.

3.1.1 A GENERAL OVERVIEW ON AN OPTIMISATION PROBLEM AND SOME ASSOCIATED CONCEPTS

We consider a function f that is defined on an open set $\Omega \subseteq \mathbb{R}^n$ and is valued in $\mathbb{R}$. In the field of mathematical programming, one studies optimisation problems that can generally be formulated as follows:

$$\min_{x \in X} \quad f(x), \tag{3.1}$$

where X is a given set which we suppose is included in Ω for simplicity. A *feasible* point x is one that verifies $x \in X$. A *solution* to a problem is a more complex concept than a feasible point, and requires additional precisions in order to be characterized in practice. First of all, from a theoretical point of view, one can directly expose the following definitions:

- A *global* solution x^g is a point that verifies: $x^g \in X$ and $f(x^g) \leq f(y)$, $\forall y \in X$.
- A *local* solution is a point x^l that verifies: $x^l \in X$, $\exists \epsilon > 0$ such that $\left\|x - x^l\right\| \leq \epsilon$ and $x \in X \implies f(x^l) \leq f(x)$.

We recall that global solutions, which are in optimisation the most interesting ones, always are local solutions, but the reverse does not hold. From a practical point of view,

one can easily observe a significant limit to these definitions: in a general setting, in order to check if a given point verifies one of these definitions, one requires an infinite quantity of information which is not acceptable for numerical applications. This limit is a strong motivation for more characterization of optimality.

First of all, one way to "lower" this obstacle is to add assumptions and work in a setting with a convex structure. Convexity has long been a significant property for optimisation. We recall the following basic definitions:

- A function $f : \Omega \mapsto \mathbb{R}$ is said to be convex if for any $x_1, x_2 \in \mathbb{R}^n$ and t such that $0 \leq t \leq 1$, $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$.
- A set $X \in \mathbb{R}^n$ is said to be convex if for any $x_1, x_2 \in X$ and t such that $0 \leq t \leq 1$, $tx_1 + (1-t)x_2 \in X$. This mathematical property means that the segment from x_1 to x_2 entirely belongs to X.
- A mathematical program, as problem (3.1), is said to be convex if both f and X are convex.

Assuming that f and X are both convex, the following implication is true:

$$\bar{x} \text{ is a local minimiser of } f \implies \bar{x} \text{ is a global minimiser of } f.$$

Thanks to convexity, the search for global optimum is reduced to a search for a local optimum[1]. On the other hand, the search for local optimality still requires an infinite amount of (local) information.

We now turn our attention more particularly to local optimality for continuous functions in an unconstrained setting, i.e. $X = \mathbb{R}^n$. Differentiability is significant for checking local optimality and designing algorithms aimed at finding local optima thanks to the following properties:

Theorem 1 (Theorem 1.1.3 [150])**.** Suppose f is a differentiable function.

- First-order necessary condition: if $\bar{x}$ is a local minimum, then

$$\nabla f(\bar{x}) = 0.$$

Suppose now f is twice differentiable.

- Second-order necessary condition: if $\bar{x}$ is a local minimum, then

$$\langle h, \nabla^2 f(\bar{x}) h \rangle \geq 0 \text{ for all } h \in \mathbb{R}^n. \tag{3.2}$$

- Second-order sufficient condition: if $\bar{x}$ satisfies equation (3.2) together with

$$\langle h, \nabla^2 f(\bar{x}) h \rangle > 0 \text{ for all } h \in \mathbb{R}^n \backslash \{0\}, \tag{3.3}$$

then $\bar{x}$ is a local minimum.

◀

[1] In a general continuous setting, note that without convexity it is not possible to verify global optimality within reasonable CPU time. In a general, discrete setting of low cardinality, one can hope to complete an exhaustive evaluation of all possibilities for instance.

In other words, equation (3.2) requires $\nabla^2 f(\bar{x})$ to be positive semi-definite, while equation (3.3) requires $\nabla^2 f(\bar{x})$ to be positive definite. As discussed in [150, p. 49], checking the second-order conditions can be difficult and a numerically expensive task especially when the dimension n becomes large. Points that verify the first-order condition are thus of particular practical importance:

Definition 4 (Critical and stationary points in an differentiable and unconstrained setting)**.** For a differentiable f, a point $x \in \mathbb{R}^n$ that verifies $\nabla f(x) = 0$ is called *critical* or *stationary*. •

Remark 3. Note that in Definition 4 both terms "critical" and "stationary" are synonyms, but that is only true in the unconstrained differentiable setting. ▷

A general optimisation framework for an unconstrained differentiable problem can thus be one that looks for points $\bar{x}$ that verify $\|\nabla f(\bar{x})\| \leq \epsilon$, where $\epsilon > 0$ is a user-given precision. More details, and numerous improvements on optimisation algorithms for this type of problems are found in [150].

There are more interesting properties to convex functions, in particular on their regularity and the existence of elements that reflect some of their variational information.

Regularity of convex functions

We start by recalling a well-known property on convex functions that is discussed in [82]:

Proposition 1 (Proposition 2.2.6 from [82])**.** Let f be a convex function, bounded above on a neighbourhood of some point of $\mathbb{R}^n$. Then, for any $x \in \mathbb{R}^n$, f is Lipschitz near x. ◀

Local Lipschitz continuity of a convex function f provides another strong result on the regularity of f:

Theorem 2 (Section 9.J [253], Rademacher's Theorem)**.** Let $\Omega \in \mathbb{R}^n$ be open, and let $f\colon \Omega \to \mathbb{R}$ be locally Lipschitz continuous. Let D be the subset of Ω consisting of the points where f is differentiable. Then $\Omega \backslash D$ is a set of Lebesgue measure zero. ◀

Rademacher's theorem states that convex functions are differentiable almost everywhere in the interior of their domain. Both Proposition 1 and theorem 2 help to understand that, at "most" points, convex functions are in-between continuous and differentiable ones. As shown in Example 1, non-differentiability evidently occurs for convex functions, sometimes at important points (as global optima). This is a reminder that, despite the regularity properties of convex functions, variational studies of such functions belong in general to that of non-differentiable functions.

Example 1. Let $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$. Then f is convex and non-differentiable at 0 (which is the global minimum of f). ▶

Directional derivatives and subdifferentials

Some additional material on the variational behaviour of convex functions can be derived from their regularity properties. These variational information are evidently of importance for the minimisation of convex functions. Let $f\colon \Omega \to \mathbb{R}$ be a convex function.

We call *directional derivative* of f at $x \in \Omega$ in the direction $d \in \mathbb{R}^n$ the following quantity:

$$f'(x;d) = \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t}. \tag{3.4}$$

Directional derivatives exist for convex functions thanks to their local Lipschitz property (see [82, Proposition 2.2.7] on this matter). It coincides with the Gâteaux derivative and can be represented by $f'(x;d) = \max_{g \in \partial f(x)} \langle g, d \rangle$, where $\partial f(x)$ is the *subdifferential* of f at point x:

$$\partial f(x) := \{ s \in \mathbb{R}^n : f(y) \geq f(x) + \langle s, y - x \rangle \quad \forall\, y \in \mathbb{R}^n \}.$$

The *approximate subdifferential* is defined, for $\epsilon \geq 0$, by

$$\partial_\epsilon f(x) := \{ s \in \mathbb{R}^n : f(y) \geq f(x) + \langle s, y - x \rangle - \epsilon \quad \forall\, y \in \mathbb{R}^n \}.$$

The following known results about subdifferentials will be of particular interest in our analysis. First, the mapping ∂f is locally bounded in the interior of $\mathcal{D}\mathrm{om}(f) := \{x \in \mathbb{R}^n : f(x) < \infty\}$ [150, Prop. VI.6.2.2]. Thus the image $\partial f(X)$ of the bounded set X is bounded in $\mathbb{R}^n$ (because $X \subset \Omega \subset \mathcal{D}\mathrm{om}(f_i)$ by hypothesis). Second, let $(g_i)_{i \in I}$ be a family of convex functions indexed by a finite index set I. The subdifferential of the pointwise maximum of finitely many convex functions $g(x) := \max_{i=1,\dots,m} g_i(x)$ at a point $x \in \bigcap_{i=1}^m \mathcal{D}\mathrm{om}(g_i)$ is the convex hull of the subdifferentials at x of the active functions [150, Corollary VI 4.3.2]:

$$\partial g(x) = \mathrm{Conv} \left\{ \bigcup_j \partial g_j(x) \,|\, j \in \{ i \in I : g_i(x) = g(x) \} \right\}. \tag{3.5}$$

Since convex functions are locally Lipschitz continuous, the directional derivatives of f are well defined, as well as the Clarke's directional derivatives and subdifferential at x:

$$f^{\mathrm{C}}(x;d) := \max_{s \in \partial^{\mathrm{C}} f(x)} \langle s, d \rangle,$$

with

$$\partial^{\mathrm{C}} f(x) := \ \mathtt{Conv} \left\{ \lim_{\ell \to \infty} \nabla f(x_\ell), x_\ell \to x, f \text{ differentiable at } x_\ell \right\} \tag{3.6}$$

the Clarke subdifferential. Moreover, convex functions with non-empty domains are *Clarke regular*, which means that the Clarke subdifferential coincides with the convex one: $\partial f(x) = \partial^{\mathrm{C}} f(x)$ for all x in Ω.

Now, as seen in part I, convexity and differentiability are not to be expected in an OPF with uncertainties: what can we say in a more general setting, one where neither convexity nor differentiability holds? In order to tackle this question, we investigate s subsects work a set of functions that is larger than convex one, which is set of difference-of-convex functions.

3.1.2 The difference-of-convex structure

We turn to the set of difference-of-convex functions. Several ingredients are necessary for our future developments: first, a formal definition is required. We then highlight the core properties of DoC functions that can be used in optimisation algorithms. We also further discuss how "optimality" is characterized for DoC functions.

First of all, we formally define DoC functions:

Definition 5. Let Ω be a convex set on $\mathbb{R}^n$. A function f is *DoC on* Ω if there exists two functions g, h convex on Ω such that $f = g - h$. Functions g and h are called the components of f. •

Far from a purely mathematical gimmick, this structure has now long been introduced and studied, and in fact naturally arises as a generalisation of the set of convex functions. A natural implication of this latter remark is that some optimality concepts should be carefully redefined in a DoC setting. Fortunately, the DoC structure is inherently linked to the well-studied convex one, as components of DoC functions are convex.

Regularity and variational properties of DoC functions

We start by presenting regularity and variational properties of DoC functions. Since DoC functions are locally Lipschitz continuous (because their components are so), the directional derivatives of $f = g - h$ are well defined and:

$$f'(x;d) = g'(x;d) - h'(x;d) \quad \forall x \in \Omega, d \in \mathbb{R}^n$$

The Clarke's directional derivatives and subdifferential at x of f are also well defined:

$$f^{\mathrm{C}}(x;d) := \max_{s \in \partial^{\mathrm{C}} f(x)} \langle s, d \rangle ,$$

with $\partial^{\mathrm{C}} f(x)$ as defined for convex functions in (3.6). DoC functions are in general non-regular in the sense of Clarke, i.e., the inclusion $\partial^{\mathrm{C}} f(x) \subset \partial g(x) - \partial h(x)$ always holds for $x \in \Omega$ but the equality may not (see the basic calculus rules in [82, Section 2.3] for instance).

Some characteristics of the set of DoC functions

As already mentioned, the set of DoC functions is the smallest vector space that includes the set of convex ones (this is discussed in [152]). Evidently, every convex or concave function is a DoC one. Being DoC is a global property: for a function f, being *locally DoC* on an open or closed set $D \subset \mathbb{R}^n$ amounts to being DoC on D ([285, Proposition 4.3]). From [285, Proposition 4.2] it is known that every twice-differentiable function is DoC on any compact convex set included within its definition domain. More generally speaking, [253, Theorem 10.33] states that every functions that is lower-$\mathcal{C}^2$ on an open set of its definition set is DoC. When considering a function q defined on Ω and valued in $\mathbb{R}$, q is lower-$\mathcal{C}^2$ if for all $x \in \Omega$ there is a local representation of q of the form $q(x) = \max_{t \in T} q_t(x)$, where the index set T is a compact space with $q_t(x)$ and all their partial derivatives through order 2 depend continuously not just on x but on (t, x). Recalling that every continuous function can be approximated by twice-differentiable functions (this is the Stone-Weierstrass theorem), we easily deduce that any continuous function can be approximated as precisely as desired by a DoC function.

Whereas convex functions have to be carefully manipulated in order to maintain the convexity property (e.g., recall that a multiplication by a negative scalar does not preserve convexity, neither does taking the pointwise minimum), there are numerous operations that maintain the DoC property. Being a vector space, the addition of two DoC functions as well as the multiplication by any scalar still create a DoC function. Moreover, the following operation also preserve the DoC property:

Proposition 2 (Proposition 4.4 from [285]). Let $\Omega_1 \subset \mathbb{R}^n$ be convex, open or closed, $\Omega_2 \subset \mathbb{R}^n$ be convex and open. If $f_1 \colon \Omega_1 \to \Omega_2$, $f_2 \colon \Omega_2 \to \mathbb{R}$ are DoC, then $f_1 \circ f_2 \colon \Omega_1 \to \mathbb{R}$ is DoC. ◀

All in all, and as discussed in [229, Section 2.1], the set of DoC functions is closed under all operations usually considered in the field of optimisation.

Optimality in a DoC setting

Important properties and optimality conditions of DoC optimization problems can be found in [58] some of which we recall here:

Theorem 3 (Theorem for Global Optimality [58]). Suppose that $\bar{x}$ satisfies $g(\bar{x}) < \infty$. Then $\bar{x}$ is a global minimizer of $f = g - h$ if and only if

$$\partial_\epsilon h(\bar{x}) \subseteq \partial_\epsilon g(\bar{x}), \text{ for all } \epsilon \geq 0.$$

◀

Theorem 4 (Theorem for Local Optimality [58]). Suppose that $\bar{x}$ satisfies $g(\bar{x}) < \infty$. Then $\bar{x}$ is a local minimizer of $f = g - h$ if and only if for some $\delta > 0$

$$\partial_\epsilon h(\bar{x}) \subseteq \partial_\epsilon g(\bar{x}), \text{ for all } 0 \leq \epsilon \leq \delta.$$

◀

Both Theorems 3 and 4 can be impossible to apply in practical situations as they require the full knowledge of the approximate subdifferentials of g and h. As a matter of fact, industrial applications can involve models where the operator does not have an explicit knowledge of a function, but relies on an *oracle*: the operator can, for instance, request a functional value and a subgradient at a given point x. On the other hand, in case x is a point where g is non-differentiable, the characterisation of $\partial g(x)$ would require an infinite number of calls to the oracle with each call returning a different subgradient: Theorems 3 and 4 are not applicable in such a situation. This example motivates the definition of weaker optimality concepts for practical applications, which is the subject of section 3.3.

Difficulties with DoC functions

The universality of DoC functions is desirable as it provides modelling tools to operators. As expected, it comes with some difficulties which we here discuss. The first one is somewhat of a drawback from the convex setting: as previously mentioned, one can approximate any continuous function as precisely as desired with DoC functions. A direct consequence, is that the uniform limit of a sequence of DoC functions is not guaranteed to be a DoC function itself. Secondly, one should not forget that DoC functions are in general nonsmooth nonconvex functions. As a consequence, there is little to no hope to find global optima. Moreover, tailored algorithms that tackle optimisation problems with DoC functions are likely to be numerically intensive. Finally, and most importantly, there are several difficulties related to DoC formulations. On the one hand, one can immediately see that there is an infinite number of acceptable DoC formulations. If $f = g - h$ is a DoC function, then for any convex function φ, $(g + \varphi) - (h + \varphi)$ is also a

valid DoC decomposition for f. Choosing which decomposition to choose is still an open question. In [152] for instance, the author proposes to set a normalisation by choosing a DoC decomposition where the second component h verifies $\min_{x\in\Omega} h(x) = 0$. Apart from this normalisation, there is no general rule for deciding whether a DoC formulation is "good" or not. Before defining what is a good DoC decomposition, finding an explicit decomposition already is a non-trivial task. One technique for Lipschitz continuous functions is presented in [229, Proposition 1]; Let f be a Lipschitz continuous function, with modulus $L > 0$: then $f = (f - \frac{L}{2}\|x\|^2) - \frac{L}{2}\|x\|^2$ is a valid DoC decomposition[2]. In another recent work [28] the authors propose a DoC decomposition framework for polynomials. Apart from these works, it often is difficult to obtain explicit DoC formulations which is a limit to practical DoC programming.

3.2 PROBLEM STATEMENT

We are concerned in this section with nonsmooth and nonconvex optimization problems of the form

$$\min_{x\in X^c} f(x), \quad \text{with} \quad f(x) := f_1(x) - f_2(x) \quad \text{and} \quad X^c := \{x \in X : c_1(x) - c_2(x) \leq 0\} \tag{3.7}$$

where X is a nonempty bounded polyhedron contained in an open set $\Omega \subset \mathbb{R}^n$, and the functions f_1, f_2, c_1, $c_2 : \Omega \to \mathbb{R}$ are all convex but possibly nonsmooth. In this setting, there is no loss of generality in considering the scalar constraint function $c(x) := c_1(x) - c_2(x)$: if the problem possesses m DoC constraints $c_{1,i}(x) - c_{2,i}(x) \leq 0$, $i = 1, \ldots, m$, then we can represent these constraints as $\max_{i=1,\ldots,m}\{c_{1,i}(x) - c_{2,i}(x)\} \leq 0$, which is equivalent to the scalar function of (3.7) with DoC decomposition $c_1(x) := \max_{i=1,\ldots,m}[c_{1,i}(x) + \sum_{j\neq i}^m c_{2,j}(x)]$ and $c_2(x) := \sum_{i=1}^m c_{2,i}(x)$; see for instance [285, Prop. 4.1(ii)]. Note that the latter is a smooth function provided that all $c_{2,j}$, $j = 1, \ldots, m$, are smooth.

Industrial applications fitting (3.7) include bilevel optimization problems originating from energy management [8], reformulation of mixed-binary problems [285], chance-constrained energy management problems [228], gas network management under uncertainty [135], and others [183, 285]. The motivation of this work is chance-constrained optimization, that is, optimization problems having a probability constraint of the form

$$\mathbb{P}\left[x \in M(\xi)\right] \geq p\,. \tag{3.8}$$

In this notation, $M(\xi) \subset \mathbb{R}^n$ is a random set depending on the random vector ξ (in a given sample space Ξ) following a known probability distribution, and $p \in (0,1)$ is a given confidence level. Constraints of this type are encountered in many engineering problems involving uncertain data. Fields of application include water management, telecommunications, electricity network expansion, hydro reservoir management, etc. We refer the interested reader to [245] for classical references on optimization problems with probability constraints and to [20, 14] for new developments and other applications. It is well known that the probability functions may fail to be convex even in the simpler setting where $M(\xi)$ is convex for almost surely all $\xi \in \Xi$. Moreover, evaluating the

[2] Note that one still has to know the Lipschitz modulus of the function.

probability function for a given point involves in general computing a multidimensional integral. This is a difficult task when the random vector has large dimension [20]. In this work we investigate smooth and nonsmooth DoC approximations of a class of probability functions. The goal is to construct a workable DoC function $c_1(x) - c_2(x)$ such that, for all $x \in \Omega$,

$$c_1(x) - c_2(x) \leq 0 \quad \approx \quad \mathbb{P}\left[x \in M(\xi)\right] \geq p\,.$$

In doing that, the chance-constrained problem can be represented in the generic form (3.7).

We care to mention that the interest in investigating DoC problems as in (3.7) goes far beyond chance-constrained programming. DoC optimization problems, that form a relevant discipline of nonconvex programming, have been receiving significant attention from researchers and practitioners. DoC programs are, in the general situation, NP-hard, and therefore cannot be solved up to global optimality in reasonable CPU times.

We refer to [285, Part II] for a comprehensive presentation of several algorithms for global optimization of DoC problems. Beyond their practical interest, local solution methods play an important role in global optimization, since algorithms of the latter class typically employ local methods to find stationary/critical points (see definition (3.11) below) that in turn feed a certain search strategy for global solutions.

Among the local-solution methods, the most employed numerical technique for DoC programming is the DoC Algorithm (DCA) and its variants reviewed in [184, 239, 182, 183]. This class of algorithms computes trial points by solving a sequence of convex subproblems obtained by replacing the second-component functions f_2 and c_2 with their first-order linearizations computed at the current point. Another closely related algorithm is the proximal linearized method studied in [233]. Recent developments on algorithms for DoC programming have been made by considering line-searches [35], inertial-force [230], (approximate) DoC decompositions[8], and bundle methods [123, 164, 228].

Most of the DoC techniques found in the literature deal with the particular case of convexly-constrained DoC programs. Having at least one DoC constraint increases the problem's complexity significantly, both theoretically and practically. For instance, linearizing the second-component function c_2 at an infeasible point can lead to an infeasible convex subproblem (see [182]).

Another main obstacle consists in verifying whether or not a given point is stationary/critical when (3.7) is nonsmooth. The recent works [233, 8] have shed some light on these issues. For instance, [8, 7] proposes a Slater-type constraint qualification (CQ) weaker than the extended Mangasarian-Fromowitz CQ or the one found in [233]. If problem (3.7) satisfies a CQ, then a workable verification of stationarity is possible (see [8, § 2.3], [233], and the brief overview in Section 4.2.3 below). In a different approach, [275] uses a penalty strategy to lever DoC constraints to the objective function. Exact penalization in DoC programming can be found in [183] and references therein.

In this work, we deal with nonsmooth DoC-constrained DoC programs by resorting to neither penalization techniques nor linearization of the constraints. Our algorithmic development is built upon the so-called *improvement functions* (see definition (4.1) below) and bundle methods [22]. Bundle methods are among the most efficient algorithms for solving nonsmooth convex optimization problems. This class of methods constitutes a very active area of research in the nonsmooth optimization community. Extensions of proximal bundle algorithms to nonconvex programs have been investigated by different

authors in [143, 222, 34]. Unconstrained DoC bundle methods are investigated in [123], convexly-constrained ones in [228], and the DoC-constrained setting is considered in [211, 8]. The paper [211] deals with nonsmooth multiobjective DoC optimization with the help of an improvement function, and in [8] the DoC constraint is handled by linearizing the second-component function c_2 at every iteration of the method. For that strategy to be well defined, it is necessary to start the algorithm with a feasible point so that the sequence of points generated by the algorithm is feasible for (3.7). This is not required in [211] nor in this work, whose sequence of trial points may be infeasible. Under the assumptions that f_2 and c_2 are the pointwise maxima of finitely many differentiable functions, we equip our approach with an escaping procedure capable of computing a B(ouligand)-stationary point (see definition (3.9)) of problem (3.7).

This chapter is organized as follows: Section 3.3 recalls some important tools of convex analysis and optimality conditions of (3.7). In Section 4.1 we present our proximal bundle algorithm for computing critical points of (3.7). Convergence analysis of the algorithm is given in Section 4.2 and the escaping procedure is described in Section 4.3. Section 5.3.1 presents a DoC model for a class of chance-constrained programs and Section 4.4 provides computational results on two deterministic and four chance-constrained problems from the literature.

3.3 SOLUTIONS OF GENERAL DOC PROGRAMS

This section recalls what is expected from "solving" a general DoC program. We rely on [233] and [8] and define B(ouligand)-stationarity and criticality for problem (3.7). We extend the initial definitions presented in section 3.1.2.

Stationary and critical points. Local optimality can generally not be achieved in nonconvex nonsmooth optimization, but weaker conditions can be expected, at different degrees, depending on the complexity of the problem. We define such conditions for the DoC-constrained DoC problem (3.7) following [233]. A feasible point $\bar{x} \in X^c = \{x \in X : c_1(x) - c_2(x) \leq 0\}$ is called a *B-stationary point* of (3.7) if the directional derivative of f is nonnegative in all directions d in the *Bouligand tangent cone* $\mathcal{T}_{X^c}(\bar{x})$ of X^c at $\bar{x}$:

$$f_1'(\bar{x}; d) \geq f_2'(\bar{x}; d) \quad \forall\, d \in \mathcal{T}_{X^c}(\bar{x})\,. \tag{3.9}$$

We recall that $d \in \mathcal{T}_{X^c}(\bar{x})$ if there exists a sequence of vectors $\{x^k\} \subset X^c$ converging to $\bar{x}$ and a sequence of positive scalars $\tau_k \to 0$ such that $d = \lim_{k\to\infty}(x^k - \bar{x})/\tau_k$. B-stationarity is a necessary condition for local optimality, but it is not sufficient [233]. The following proposition provides a workable description of the tangent cone $\mathcal{T}_{X^c}(\bar{x})$ under a constraint qualification.

Proposition 3. (Proposition 2.1 from [7]). Let $\bar{x} \in X^c$ be such that $c_1(\bar{x}) = c_2(\bar{x})$. Suppose that the following Slater constraint qualification (CQ) holds: there exists $\bar{d} \in \mathcal{T}_X(\bar{x})$ such that $c_1'(\bar{x}; \bar{d}) < c_2'(\bar{x}; \bar{d})$. Then

$$\mathcal{T}_{X^c}(\bar{x}) = \{d \in \mathcal{T}_X(\bar{x}) : c_1'(\bar{x}; d) \leq c_2'(\bar{x}; d)\}.$$

◀

The CQ of Proposition 3 is weaker than both the *extended Mangasarian-Fromowitz constraint qualification* (EMFCQ) and the one given in [233] (see [8, §2]). The CQ given in [233] is tailored to a particular setting of (3.7) where the second-component functions f_2 and c_2 are the pointwise maximum of finitely many continuously differentiable functions. The EMFCQ conditions holds at a point $x \in X$, such that $c(x) = c_1(x) - c_2(x) = 0$, if there exists $\bar{d} \in \mathcal{T}_X(x)$ such that the Clarke directional derivative of c at x in direction $\bar{d}$ is negative, i.e., $c^{\mathrm{C}}(\bar{x}; \bar{d}) < 0$, with $c^{\mathrm{C}}(\bar{x}; \bar{d}) := \max_{s \in \partial^{\mathrm{C}} c(\bar{x})} \langle s, d \rangle$.

Under the assumptions of Proposition 3, the B-stationary condition (3.9) becomes

$$f_1'(\bar{x}; d) \geq f_2'(\bar{x}; d) \quad \forall\, d \in \mathcal{T}_X(\bar{x}) \;\text{ s.t. }\; c_1'(\bar{x}; d) \leq c_2'(\bar{x}; d)\,,$$

which is, according to [8], equivalent to $\bar{x}$ being a solution of

$$\begin{cases} \min\limits_{x \in X} & f_1(x) - [f_2(\bar{x}) + \langle s_{f_2}, x - \bar{x} \rangle] & \forall\, s_{f_2} \in \partial f_2(\bar{x}) \\ \text{s.t.} & c_1(x) - [c_2(\bar{x}) + \langle s_{c_2}, x - \bar{x} \rangle] \leq 0 & \forall\, s_{c_2} \in \partial c_2(\bar{x}). \end{cases} \tag{3.10}$$

This last definition unveils the difficulty of checking whether a point $\bar{x} \in X^c$ is B-stationary in the general nonsmooth context, as we may not know the whole subdifferentials $\partial f_2(\bar{x})$ and $\partial c_2(\bar{x})$. In practice, we compute a *critical point*, i.e., a vector $\bar{x} \in X^c$ satisfying the assumptions of Proposition 3 and the following weaker condition

$$\begin{cases} \min\limits_{x \in X} & f_1(x) - [f_2(\bar{x}) + \langle s_{f_2}, x - \bar{x} \rangle] & \text{for some arbitrary } s_{f_2} \in \partial f_2(\bar{x}) \\ \text{s.t.} & c_1(x) - [c_2(\bar{x}) + \langle s_{c_2}, x - \bar{x} \rangle] \leq 0 & \text{and } s_{c_2} \in \partial c_2(\bar{x}). \end{cases} \tag{3.11}$$

Under additional assumptions. Note that the notions of criticality and B-stationarity coincide under the assumption that both f_2 and c_2 are respectively continuously differentiable at $\bar{x} \in X^c$, since $\partial f_2(\bar{x}) = \{\nabla f_2(\bar{x})\}$ and $\partial c_2(\bar{x}) = \{\nabla c_2(\bar{x})\}$ in this case. Furthermore, we care to mention that B-stationarity becomes numerically verifiable under the assumptions that the second-component functions f_2 and c_2 are the pointwise maximum of finitely many differentiable functions [233]:

$$f_2(x) := \max_{i=1,\ldots,m_f} \psi_i(x) \quad \text{and} \quad c_2(x) := \max_{i=1,\ldots,m_c} \varphi_i(x)\,. \tag{3.12}$$

It follows from (3.5) that the subdifferential of f_2 (respectively c_2) at any given point $x \in X$ is the convex hull of gradients of functions ψ_i (respectively φ_j) that are active:

$$\forall\, x \in X\,,\;\; \partial f_2(x) = \;\texttt{Conv}(\{\nabla \psi_i(x)\}_{i \in A^f(x)})\,,\;\; \partial c_2(x) = \;\texttt{Conv}(\{\nabla \varphi_j(x)\}_{j \in A^c(x)})\,, \;\text{ with} \tag{3.13a}$$

$$A^f(x) := \{1 \leq i \leq m_f : \; f_2(x) = \psi_i(x)\} \quad \text{and} \quad A^c(x) := \{1 \leq j \leq m_c : \; c_2(x) = \varphi_j(x)\}\,. \tag{3.13b}$$

Under assumption (3.12), verifying the B-stationarity condition (3.10) of a given point $\bar{x}$ feasible for (3.7) amounts to checking whether $\bar{x}$ solves

$$\begin{cases} \min\limits_{x \in X} & f_1(x) - [\psi_i(\bar{x}) + \langle \nabla \psi_i(\bar{x}), x - \bar{x} \rangle] \\ \text{s.t.} & c_1(x) - [\phi_j(\bar{x}) + \langle \nabla \phi_j(\bar{x}), x - \bar{x} \rangle] \leq 0, \end{cases} \tag{3.14}$$

for all $i \in A^f(\bar{x})$ and all $j \in A^c(\bar{x})$ (see [233]). This is the key fact behind the escaping procedure of Section 4.3.

Relation between critical and KKT points of DoC-constrained DoC programs, [8, Remark 2.4]. Consider the convex subproblem (3.11). If $\bar{x}$ satisfies the assumptions of Proposition 3, then there exists a Lagrange multiplier $\bar{\lambda} \geq 0$ such that $(\bar{x}, \bar{\lambda})$ satisfies the KKT system of (3.11):

$$\begin{cases} 0 \in \partial f_1(\bar{x}) - s_{f_2} + \bar{\lambda}(\partial c_1(\bar{x}) - s_{c_2}) + N_X(\bar{x}) \\ c_1(\bar{x}) - c_2(\bar{x}) \leq 0,\ \bar{\lambda}[c_1(\bar{x}) - c_2(\bar{x})] = 0,\ \bar{\lambda} \geq 0,\ \bar{x} \in X, \end{cases}$$

where $N_X(\bar{x})$ is the normal cone of the convex set X at $\bar{x}$. Furthermore, suppose that (i) either f_1 or f_2 is continuously differentiable, and (ii) either c_1 or c_2 is continuously differentiable. These assumptions ensure that $\partial^c f = \partial f_1 - \partial f_2$ and $\partial^c c = \partial c_1 - \partial c_2$ according to [82, Prop. 2.3.3 and Corol. 2]. As a result, the above system becomes the KKT system of problem (3.7):

$$\begin{cases} 0 \in \partial^c[f_1(\bar{x}) - f_2(\bar{x})] + \bar{\lambda}\,\partial^c[c_1(\bar{x}) - c_2(\bar{x})] + N_X(\bar{x}) \\ c_1(\bar{x}) - c_2(\bar{x}) \leq 0,\ \bar{\lambda}[c_1(\bar{x}) - c_2(\bar{x})] = 0,\ \bar{\lambda} \geq 0,\ \bar{x} \in X. \end{cases} \tag{3.15}$$

This result shows that, under some constraint qualification and smoothness conditions, a critical vector is a KKT point of (3.7).

In the absence of the DoC constraint. For convexly-constrained problems, the B-stationarity and criticality conditions become simpler. When the DoC constraint $c(x) = c_1(x) - c_2(x) \leq 0$ in problem (3.7) is absent, it is immediately seen from the optimality conditions of problems (3.10) and (3.11), that B-stationarity reduces to *d(irectional)-stationarity*:

$$\partial f_2(\bar{x}) \subset \partial f_1(\bar{x}) + N_X(\bar{x}) = \partial f_1(\bar{x}) + \partial \mathbb{1}_X(\bar{x})\,, \tag{3.16}$$

where $\mathbb{1}_X$ is the indicator function of the set X and the last equality holds for $\bar{x} \in X$, and criticality is nothing else than:

$$\emptyset \neq \partial f_2(\bar{x}) \cap [\partial f_1(\bar{x}) + N_X(\bar{x})]\,. \tag{3.17}$$

An intermediate condition between (3.16) and (3.17) is Clarke stationarity [233]:

$$0 \in \partial^c f(\bar{x}) + N_X(\bar{x})\,.$$

3.4 CONCLUSION

This chapter proposes an overview basic elements of the DoC structure. These elements are necessary for the development of a tailored algorithm to solve problems similar to equation (3.7). The DoC structure is an attractive extension to the convex one, that comes with some additional costs. On the one hand the set of DoC functions appears to be a large one: every continuous function is either DoC, or can be approximated as closely as desired by DoC functions. Moreover, the set of DoC functions is closed under all usual operations used in the optimisation field. On the other hand, for a given function finding a DoC decomposition is a difficult task that its not always possible to overcome. Discrimination between DoC decompositions should also be regarded as an important field, as described in [28]: the behaviour of their DoC algorithm is significantly modified when selecting another DoC decomposition for a single DoC function. The next chapter is dedicated to the presentation of a novel bundle algorithm for DoC constrained DoC functions.

A PROXIMAL BUNDLE METHOD FOR DoC PROBLEMS

4

Based on chapter 3, we are now in a position to present our algorithm for the optimization of DoC-constrained DoC problems. The gist of this algorithm is the combination of the proximal bundle method for convexly constrained DoC program (see [228]) with the improvement function that is defined in equation (4.1). The convergence study is a challenging step, and requires some involved and technical developments.

This chapter is organised as follows: we start in the first section by presenting our algorithm, before developing a proper and tailored convergence analysis in the second section.

4.1 NONLINEARLY CONSTRAINED DoC BUNDLE METHOD

In this section, we present a proximal bundle method addressing the DoC-constrained DoC programs such as (3.7). Instead of penalization techniques [183] or constraint linearization [233, 8], our approach employs, for a suitable target $\tau = (\tau_f, \tau_c) \in \mathbb{R}^2$, an *improvement function* $H_\tau : \Omega \to \mathbb{R}$ defined by

$$\begin{aligned} H_\tau(x) &:= \max\{f(x) - \tau_f,\, c(x) - \tau_c\} \\ &= \max\{f_1(x) - f_2(x) - \tau_f,\, c_1(x) - c_2(x) - \tau_c\} \\ &= \underbrace{\max\{f_1(x) + c_2(x) - \tau_f,\, f_2(x) + c_1(x) - \tau_c\}}_{F_\tau(x)} - \underbrace{[f_2(x) + c_2(x)]}_{G(x)}, \end{aligned} \tag{4.1}$$

which is a DoC function itself. For a given parameter τ, we consider the convexly-constrained DoC program

$$\min_{x \in X} \; H_\tau(x). \tag{4.2}$$

Note that if $\tau = (f^*, 0)$ with f^* the global (local) value of problem (3.7), then the global (local) value of (4.2) is zero and its global (local) solutions globally (locally) solve (3.7). We refer the interested reader to [34, Lemma 5.1] for the mathematical relations between (3.7) and (4.2) in a more general context (without the DoC structure). Although transforming the DoC-constrained DoC program (3.7) into the convexly-constrained DoC program (4.2) is an appealing strategy, certain difficulties arise in this approach:

- The optimal value of the original problem (3.7) is usually unknown. Therefore, the target parameter τ must be estimated iteratively. As a downside, the optimal value of (4.2) for an arbitrary τ is unknown as well.
- When dealing with problem (4.2), DoC algorithms provide us with a critical point. How do the critical points of (4.2) and (3.7) relate?

We partially anticipate the answer to the latter question by mentioning that a critical point $\bar{x}$ of (4.2) may not be critical, even not feasible, for (3.7) (see Theorem 6 below). We

will come to this matter at the end of Section 4.2 and Section 4.3, where we will require the second-component functions to satisfy additional assumptions. But first, we will present our algorithm in the general setting where all functions in (3.7) are nonsmooth, and prove that it yields a critical point $\bar{x}$ of (4.2) for a certain parameter $\bar{\tau}$.

4.1.1 Proximal DoC bundle method with improvement functions

We start by mentioning that bundle methods employing improvement functions have been investigated in the convex setting in [22] for the proximal variant, and in [20] for the level variant. In a more general nonconvex framework, the improvement function is called *progress function* in [34], where a proximal bundle method is proposed for a class of optimal control problems. The DoC setting has recently been investigated in [211] with a focus on multi-objective optimization and the computation of Pareto critical/stationary points. Our approach is similar to [211] but the algorithm described below is simpler to present, analyze and implement. Furthermore, we employ the slightly different improvement function (4.1) where the target value τ_c can be different from zero. For these reasons, our approach requires a convergence analysis of its own. One major difference lies in the procedure from [164] which is enforced in the algorithm of [211] to escape critical points of poor quality and lead to KKT points of (3.7) under mild assumptions. While this procedure can also be employed by our algorithm, we present in Section 4.3 an alternative to provide better candidate solutions as B-stationary points.

We highlight that for a fixed τ, problem (4.2) could be solved by the algorithm presented in [230]. Indeed, a natural idea is to generate a sequence of iterates $\{x^k\} \subset X$ by successively setting targets τ^k and defining the trial point x^{k+1} as a critical point of (4.2) with $\tau = \tau^k$. However, this strategy can be too time-consuming depending on the nature of the four convex functions composing $H_\tau(x)$. Instead of computing a critical point of (4.2) with τ fixed, an alternative is to address (4.2) iteratively by updating τ^k and x^k simultaneously, at each iteration $k \in \mathbb{N}$, as follows:

$$\tau_f^k := f(x^k) + \rho \max\{c(x^k), 0\} \quad \text{and} \quad \tau_c^k := \sigma \max\{c(x^k), 0\} \tag{4.3a}$$

$$s_{f_2}^k \in \partial f_2(x^k) \quad \text{and} \quad s_{c_2}^k \in \partial c_2(x^k) \quad (\text{i.e., } s_{f_2}^k + s_{c_2}^k \in \partial G(x^k)) \tag{4.3b}$$

$$x^{k+1} := \arg\min_{x \in X} \; F_{\tau^k}(x) - \langle s_{f_2}^k + s_{c_2}^k, x \rangle + \frac{\mu}{2} \left\| x - x^k \right\|^2, \tag{4.3c}$$

with $\mu > 0$, $\rho \geq 0$, $\sigma \in [0, 1)$ and $x^0 \in \Omega$ given. This scheme follows the general lines of proximal linearized algorithms for convexly-constrained DoC programming as proposed e.g. in [233], or [42]. However, it differs from these methods in so much that here the first component F_{τ^k} changes along the iterates every time τ^k is updated.

Again, solving the nonsmooth convex optimization problem (4.3c) at each iteration might not be trivial depending on the functions of (3.7). To overcome this drawback, we follow the lead of bundle methods and replace the convex function F_τ at iteration $k \in \mathbb{N}$ with a model $\mathcal{M}_\tau^k$ approximating F_τ from below. Our approach requires four oracles (black-boxes), one for each component function f_i or c_i, $i = 1, 2$, providing the function value $f_i(x^k)$ or $c_i(x^k)$ and an arbitrary subgradient $s_{f_i}^k \in \partial f_i(x^k)$ or $s_{c_i}^k \in \partial c_i(x^k)$ at the given point $x^k \in X$. At iteration $k \in \mathbb{N}$, let $\mathcal{B}^k \subset \{0 \ldots, k\}$ be an index set defining the

following individual cutting-plane models for the convex functions f_i, c_i, $i = 1, 2$:

$$\begin{aligned} \check{f}_i^k(x) &:= \max_{j\in\mathcal{B}^k}\{f_i(x^j) + \langle s_{f_i}^j, x - x^j\rangle\} \quad \leq \quad f_i(x) \quad \forall\, x \in \Omega \text{ and } i = 1,2, \\ \check{c}_i^k(x) &:= \max_{j\in\mathcal{B}^k}\{c_i(x^j) + \langle s_{c_i}^j, x - x^j\rangle\} \quad \leq \quad c_i(x) \quad \forall\, x \in \Omega \text{ and } i = 1,2. \end{aligned}$$

Our model $\mathcal{M}_\tau^k$ for F_τ is defined as

$$\mathcal{M}_\tau^k(x) := \max\left\{\check{f}_1^k(x) + \check{c}_2^k(x) - \tau_f,\ \check{f}_2^k(x) + \check{c}_1^k(x) - \tau_c\right\} \quad \leq \quad F_\tau(x) \quad \forall\, x \in \Omega. \tag{4.4}$$

Notice that $\mathcal{M}_\tau^k(x^j) = F_\tau(x^j)$ for all $j \in \mathcal{B}^k$. When replacing F_τ with $\mathcal{M}_\tau^k$ in the subproblem (4.3c), the solution of the resulting optimization problem does not necessarily imply descent of the improvement function. For this reason, bundle algorithms yield descent by keeping a stability center fixed and improving the model $\mathcal{M}_\tau^k$ until a better trial point is obtained. When this happens, we say that the algorithm performs a serious step. At a given iteration $k \in \mathbb{N}$, we denote by $\ell \in \mathbb{N}$ the number of serious steps performed, and by $k(\ell) \in \{\ell, \ldots, k\}$ the iteration of the last serious step: $x^{k(\ell)}$ is thus the current stability center. To simplify the notation and because τ is only updated at serious steps according to (4.3a), we identify $\tau^{k(\ell)}$ with τ^ℓ:

$$\tau_f^\ell := f(x^{k(\ell)}) + \rho\max\{c(x^{k(\ell)}), 0\} \quad \text{and} \quad \tau_c^\ell := \sigma\max\{c(x^{k(\ell)}), 0\}. \tag{4.5}$$

Hence, at iteration $k \in \mathbb{N}$, we compute the next trial point x^{k+1} by solving the following subproblem instead of (4.3c):

$$x^{k+1} := \arg\min_{x\in X}\ \mathcal{M}_{\tau_\ell}^k(x) - \langle s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}, x\rangle + \frac{\mu^k}{2}\left\|x - x^{k(\ell)}\right\|^2. \tag{4.6}$$

The optimality condition for the above problem yields (e.g., [230, Prop. 1])

$$x^{k+1} = x^{k(\ell)} - \frac{1}{\mu^k}[p^{k+1} + s_X^{k+1} - (s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)})], \quad \text{with} \quad \begin{cases} p^{k+1} \in \partial\mathcal{M}_{\tau^\ell}^k(x^{k+1}) \\ s_X^{k+1} \in N_X(x^{k+1}), \end{cases} \tag{4.7}$$

which follows from the property that the vector of zeros in $\mathbb{R}^n$ must belong to the subdifferential of the convex objective function of (4.6) at its solution x^{k+1}. Notice that (4.6) can be rewritten as a quadratic program (QP) by adding the extra vector $r \in \mathbb{R}^5$ of variables:

$$\begin{cases} \min\limits_{x,r} & r_5 - \langle s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}, x\rangle + \frac{\mu^k}{2}\left\|x - x^{k(\ell)}\right\|^2 \\ \text{s.t.} & f_1(x^j) + \langle s_{f_1}^j, x - x^j\rangle \leq r_1 \quad \forall j \in \mathcal{B}^k \\ & f_2(x^j) + \langle s_{f_2}^j, x - x^j\rangle \leq r_2 \quad \forall j \in \mathcal{B}^k \\ & c_1(x^j) + \langle s_{c_1}^j, x - x^j\rangle \leq r_3 \quad \forall j \in \mathcal{B}^k \\ & c_2(x^j) + \langle s_{c_2}^j, x - x^j\rangle \leq r_4 \quad \forall j \in \mathcal{B}^k \\ & r_1 + r_4 - \tau_f^\ell \leq r_5,\ r_2 + r_3 - \tau_c^\ell \leq r_5 \\ & x \in X,\ r \in \mathbb{R}^5. \end{cases} \tag{4.8}$$

As usual in bundle methods, after a serious step, we may keep only the last oracle information in the bundle. Otherwise, we also keep all the linearizations $\overline{\mathcal{B}}^k \subset \mathcal{B}^k$ that

are active in the QP subproblem (4.8), i.e.,

$$\overline{\mathcal{B}}^k := \left\{ j \in \mathcal{B}^k : \sum_{i=1}^{4} \alpha_i^j > 0 \right\}, \tag{4.9}$$

where $\alpha^j \in \mathbb{R}^4$, $j \in \mathcal{B}^k$ denote the Lagrange multipliers associated with the first four constraints of (4.8). Note that this requires four times more storagethan standard convex bundle methods.For simplicity, we assume synchronous bundle management in the sense that the quartet of linearizations issued at any previous iteration j is either entirely kept or removed from the bundle. However, except for the information related to the new iterate x^{k+1} and the last stability center $x^{k(\ell)}$, there is no theoretical reason for the other quartets to be managed synchronously: asynchronous bundle management is possible. We have opted for the less involved notation of the synchronous case.

Our algorithm employs the following descent test, where $\mu_{\min} > 0$ is a lower bound on the prox-parameters μ^k for all $k \in \mathbb{N}$ and $\kappa \in (0,1)$: we define a serious step if the inequality

$$H_{\tau^\ell}(x^{k+1}) \leq H_{\tau^\ell}(x^{k(\ell)}) - \kappa \frac{\mu_{\min}}{2} \left\| x^{k+1} - x^{k(\ell)} \right\|^2 \tag{4.10}$$

holds, and declare a null step otherwise. After a serious step, we increment the counter ℓ and update $k(\ell) := k+1$ and τ^ℓ according to (4.3a) with $k = k(\ell)$. The following lemma shows that a serious step either improves the objective function or decreases infeasibility in (3.7).

Lemma 1. Let $x^{k(\ell)} \in X\ (\subset \Omega)$ be the current stability center and τ^ℓ as in (4.5) with $\rho \geq 0$ and $\sigma \in [0,1)$ be given. Then $H_{\tau^\ell}(x^{k(\ell)}) \geq 0$ holds true. Furthermore, if inequality (4.10) holds, then either

i) $f(x^{k+1}) \leq f(x^{k(\ell)}) - \kappa \frac{\mu_{\min}}{2} \left\| x^{k+1} - x^{k(\ell)} \right\|^2$ and $c(x^{k+1}) \leq 0$ when $c(x^{k(\ell)}) \leq 0$; or

ii) $c(x^{k+1}) \leq c(x^{k(\ell)}) - \kappa \frac{\mu_{\min}}{2} \left\| x^{k+1} - x^{k(\ell)} \right\|^2$ when $c(x^{k(\ell)}) > 0$.

◀

Proof. In the first case, if $c(x^{k(\ell)}) \leq 0$, then $\tau_f^\ell = f(x^{k(\ell)})$ and $\tau_c^\ell = 0$ according to (4.5). Definition (4.1) of H_{τ^ℓ} states that

$$\begin{aligned} H_{\tau^\ell}(x^{k(\ell)}) &= \max\{f(x^{k(\ell)}) - f(x^{k(\ell)}), c(x^{k(\ell)})\} = 0 \\ H_{\tau^\ell}(x^{k+1}) &= \max\{f(x^{k+1}) - f(x^{k(\ell)}), c(x^{k+1})\}, \end{aligned} \tag{4.11}$$

and (4.10) gives $\max\{f(x^{k+1}) - f(x^{k(\ell)}), c(x^{k+1})\} \leq 0 - \kappa \frac{\mu_{\min}}{2} \left\| x^{k+1} - x^{k(\ell)} \right\|^2$, which implies i).

Now suppose that $c(x^{k(\ell)}) > 0$, i.e., $x^{k(\ell)}$ is an infeasible point for (3.7). In this case,

$$\begin{aligned} H_{\tau^\ell}(x^{k+1}) &= \max\{f(x^{k+1}) - f(x^{k(\ell)}) - \rho c(x^{k(\ell)}), c(x^{k+1}) - \sigma c(x^{k(\ell)})\}, \\ H_{\tau^\ell}(x^{k(\ell)}) &= \max\{-\rho\, c(x^{k(\ell)}), (1-\sigma)\, c(x^{k(\ell)})\} = (1-\sigma) c(x^{k(\ell)}) > 0, \end{aligned} \tag{4.12}$$

where the last equality is due to the assumptions $\rho \geq 0$ and $1-\sigma > 0$. Therefore, (4.10) reads as

$$\max\{f(x^{k+1}) - f(x^{k(\ell)}) - \rho c(x^{k(\ell)}), c(x^{k+1}) - \sigma c(x^{k(\ell)})\} \leq (1-\sigma)c(x^{k(\ell)}) - \kappa \frac{\mu_{\min}}{2} \left\| x^{k+1} - x^{k(\ell)} \right\|^2,$$

which implies ii). Finally, equations (4.11) and (4.12) show that $H_{\tau^\ell}(x^{k(\ell)}) \geq 0$. □

This result shows that the descent test can either be chosen as (4.10), which is used in Section 4.4, or based on the inequalities of Lemma 1. The rationale of serious iterates is to ensure sufficient decrease on one component function of H_{τ^ℓ} while maintaining feasibility for (3.7) once reached.

Algorithm 1 provides a pseudocode of our proximal bundle method algorithm for the DoC-constrained DoC program (3.7). This algorithm solves the strongly convex QP (4.8) at every iteration instead of (4.3c). We focus in the next section on the analysis of Algorithm 1 and skip proving the convergence of (4.3) as this is is essentially a byproduct of the analysis given below, where several results can be either simplified or eliminated.

Algorithm 1 Proximal Bundle Method for DoC-constrained DoC programs - `PBMDC` [2]

Step 0: Initialization. Let $x^0 \in X$, $\kappa \in (0,1)$, $0 < \mu_{\min} \leq \mu^0 \leq \mu_{\max} < \infty$, $\rho \geq 0$, $\sigma \in [0,1)$ and $\delta_{\text{Tol}} \geq 0$ be given. Compute $(f_i(x^0), s^0_{f_i} \in \partial f_i(x^0))$ and $(c_i(x^0), s^0_{c_i} \in \partial c_i(x^0))$ for $i = 1,2$, and $\tau^0 = (\tau^0_f, \tau^0_c)$ as in (4.5). Define $k := \ell := k(\ell) = 0$ and $\mathcal{B}^0 := \{0\}$.

Step 1: Trial point. Compute x^{k+1} by solving the QP (4.8).

Step 2: Stopping test. If $\left\| x^{k+1} - x^{k(\ell)} \right\| \leq \delta_{\text{Tol}}$, then stop and return $x^{k(\ell)}$.

Step 3: Oracles call. Compute $(f_i(x^{k+1}), s^{k+1}_{f_i} \in \partial f_i(x^{k+1}))$ and $(c_i(x^{k+1}), s^{k+1}_{c_i} \in \partial c_i(x^{k+1}))$ for $i = 1,2$.

Step 4: Descent test. If (4.10) holds, then declare a *serious step*:
define $\ell := \ell + 1$, $k(\ell) := k+1$ and τ^ℓ as in (4.5); choose $\mathcal{B}^{k+1} \subset \{0, \ldots, k+1\}$ with $\{k+1\} \in \mathcal{B}^{k+1}$ and $\mu^{k+1} \in [\mu_{\min}, \mu^k]$.
Else, declare a *null step*: choose $\mathcal{B}^{k+1} \subset \{0, \ldots, k+1\}$ with $\overline{\mathcal{B}}^k \cup \{k+1, k(\ell)\} \subset \mathcal{B}^{k+1}$ ($\overline{\mathcal{B}}^k$ as in (4.9)) and $\mu^{k+1} \in [\mu^k, \mu_{\max}]$.

Step 5: Loop Set $k := k+1$ and go back to Step 1.

4.2 CONVERGENCE ANALYSIS

Before linking to the original problem (3.7), we show in this section that, with $\delta_{\text{Tol}} = 0$, Algorithm 1 yields a sequence in which any cluster point $\bar{x}$ is critical for

$$\min_{x \in X} H_{\bar{\tau}}(x), \quad \text{with} \quad \begin{cases} H_{\bar{\tau}}(x) & = F_{\bar{\tau}}(x) - G(x) \\ \bar{\tau}_f & = f(\bar{x}) + \rho \max\{c(\bar{x}), 0\}, \quad \rho \geq 0 \\ \bar{\tau}_c & = \sigma \max\{c(\bar{x}), 0\}, \quad \sigma \in [0,1). \end{cases} \tag{4.13}$$

Since $\{x^{k(\ell)}\}_\ell$ is contained in the compact set X, the sequence of stability centers has at least one cluster point in X.

As is standard in the study of convergence of proximal bundle methods, we will split our analysis into two exclusive cases: the sequence $\{x^{k(\ell)}\}_\ell$ is either finite or infinite. This amounts to having a finite or infinite number of serious steps. In the former case, the analysis follows from the standard theory of convex bundle methods: the function F_{τ^ℓ} becomes fixed after finitely many steps and the *convex* subproblem solved by the bundle algorithm is (with ℓ fixed)

$$\min_{x\in X} F_{\tau^\ell}(x) - \langle s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}, x\rangle.$$

The case of infinitely many serious steps is more involved and requires elements of "DoC analysis". The results of [228] (for the convexly-constrained case) do not apply in our setting due to the fact that τ in (4.2) changes along the serious steps (c.f. (4.5)). Throughout this section we assume that $\delta_{\text{Tol}} = 0$.

4.2.1 Finitely many serious steps

Assume that the algorithm performs a finite number of serious steps (and a finite or infinite number of null steps), then the following result shows that our assertion holds with $\bar{x}$ being the last stability center, and (4.13) being the subproblem (4.2) corresponding to the last serious iteration.

Proposition 4. Let $\delta_{\text{Tol}} = 0$, $\bar{\ell} \in \mathbb{N}$ be the number of serious steps performed by Algorithm 1, and $\bar{x} := x^{k(\bar{\ell})}$ be the last generated stability center. Then $\bar{x}$ is a critical point of problem (4.13), and the sequence of iterates $\{x^k\}_k$ generated by Algorithm 1 is either finite (the algorithm stops), or it converges to $\bar{x}$. ◀

Proof. At every iteration $k \geq k(\bar{\ell})$, the target $\bar{\tau} := \tau^{\bar{\ell}}$ is fixed as defined in (4.13) with $\bar{x} := x^{k(\bar{\ell})}$. Then Algorithm 1 behaves as a convex bundle algorithm applied to the convex function $x \mapsto F_{\bar{\tau}}(x) - \langle s_{f_2}^{k(\bar{\ell})} + s_{c_2}^{k(\bar{\ell})}, x\rangle$, and the convergence analysis applies as well: if the algorithm stops then
$\bar{x} := x^{k(\bar{\ell})}$ is a critical point of (4.13) according to [228, Lemma 2]; otherwise, according to [228, Proposition 3], the sequence of iterates generated after the last serious step converges to the last stability center $\bar{x}$ which is a critical point of (4.13). Notice that Algorithm 1 ensures that the active linearizations are kept in the bundle and that $\{\mu^k\}_{k\geq k(\bar{\ell})}$ is a nondecreasing sequence contained in $[\mu_{\min}, \mu_{\max}]$, two important assumptions in [228]. □

4.2.2 Infinitely many serious steps

Throughout this section, we will use the notation $i(\ell) = k(\ell+1) - 1$ for $\ell \in \mathbb{N}$ to refer to the iteration of Algorithm 1 yielding the $(\ell+1)^{th}$ serious step. At such an iteration $i(\ell)$, the quadratic subproblem (4.6) reads as

$$x^{k(\ell+1)} = \arg\min_{x\in X} \mathcal{M}_{\tau^\ell}^{i(\ell)}(x) - \langle s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}, x\rangle + \frac{\mu^{i(\ell)}}{2}\left\|x - x^{k(\ell)}\right\|^2.$$

The following result, which mirrors [228, Lemma 3], shows that $p^{k(\ell+1)}$ given in (4.7) is an approximate subgradient of the function F_{τ^ℓ} at the point $x^{k(\ell+1)}$.

Lemma 2. Let $p^{k(\ell+1)} \in \mathcal{M}_{\tau^\ell}^{i(\ell)}(x^{k(\ell+1)})$. There exists a constant $L > 0$ such that

$$p^{k(\ell+1)} \in \partial_{e_{i(\ell)}} F_{\tau^\ell}(x^{k(\ell)}), \text{ with } e_{i(\ell)} = L \left\| x^{k(\ell)} - x^{k(\ell+1)} \right\|.$$

◀

Proof. Definition (4.4) of the model and the fact that $k(\ell)$ is in the bundle $\mathcal{B}^{k(\ell)}$ give the following chain of inequalities:

$$\begin{aligned}
\mathcal{M}_{\tau^\ell}^{i(\ell)}(x^{k(\ell+1)}) &\geq \max\Big\{ f_1(x^{k(\ell)}) + c_2(x^{k(\ell)}) + \langle s_{f_1}^{k(\ell)} + s_{c_2}^{k(\ell)}, x^{k(\ell+1)} - x^{k(\ell)} \rangle - \tau_f^\ell, \\
&\qquad\qquad f_2(x^{k(\ell)}) + c_1(x^{k(\ell)}) + \langle s_{f_2}^{k(\ell)} + s_{c_1}^{k(\ell)}, x^{k(\ell+1)} - x^{k(\ell)} \rangle - \tau_c^\ell \Big\} \\
&\geq \max\Big\{ f_1(x^{k(\ell)}) + c_2(x^{k(\ell)}) - \tau_f^\ell, f_2(x^{k(\ell)}) + c_1(x^{k(\ell)}) - \tau_c^\ell \Big\} + \\
&\qquad \min\Big\{ \langle s_{f_1}^{k(\ell)} + s_{c_2}^{k(\ell)}, x^{k(\ell+1)} - x^{k(\ell)} \rangle, \langle s_{f_2}^{k(\ell)} + s_{c_1}^{k(\ell)}, x^{k(\ell+1)} - x^{k(\ell)} \rangle \Big\}.
\end{aligned} \tag{4.14}$$

Once again, using the Cauchy-Schwarz inequality and the boundedness of $\partial f_i(X)$ and $\partial c_i(X)$, $i = 1, 2$, there exist $L_1, L_2 > 0$ such that, for any $\ell \in \mathbb{N}$:

$$\begin{aligned}
\langle s_{f_1}^{k(\ell)} + s_{c_2}^{k(\ell)}, x^{k(\ell+1)} - x^{k(\ell)} \rangle &\geq -\| s_{f_1}^{k(\ell)} + s_{c_2}^{k(\ell)} \| \| x^{k(\ell)} - x^{k(\ell+1)} \| \geq -L_1 \| x^{k(\ell)} - x^{k(\ell+1)} \| \\
\langle s_{f_2}^{k(\ell)} + s_{c_1}^{k(\ell)}, x^{k(\ell+1)} - x^{k(\ell)} \rangle &\geq -\| s_{f_2}^{k(\ell)} + s_{c_1}^{k(\ell)} \| \| x^{k(\ell)} - x^{k(\ell+1)} \| \geq -L_2 \| x^{k(\ell)} - x^{k(\ell+1)} \|,
\end{aligned}$$

By taking $L := \max\{L_1, L_2\}$ and combining the two previous inequalities with (4.14) and the definition (4.1) of $F_{\tau^\ell}(x^{k(\ell)})$, we get

$$\mathcal{M}_{\tau^\ell}^{i(\ell)}(x^{k(\ell+1)}) \geq F_{\tau^\ell}(x^{k(\ell)}) - L \left\| x^{k(\ell)} - x^{k(\ell+1)} \right\|.$$

Since the model approximates F_{τ^ℓ} from below, the above inequality yields, for any $x \in \Omega$, $\ell \in \mathbb{N}$,

$$\begin{aligned}
F_{\tau^\ell}(x) &\geq \mathcal{M}_{\tau^\ell}^{i(\ell)}(x) \\
&\geq \mathcal{M}_{\tau^\ell}^{i(\ell)}(x^{k(\ell+1)}) + \langle p^{k(\ell+1)}, x - x^{k(\ell)} \rangle \\
&\geq F_{\tau^\ell}(x^{k(\ell)}) + \langle p^{k(\ell+1)}, x - x^{k(\ell)} \rangle - L \left\| x^{k(\ell)} - x^{k(\ell+1)} \right\|,
\end{aligned}$$

i.e. $p^{k(\ell+1)} \in \partial_{e_{i(\ell)}} F_{\tau^\ell}(x^{k(\ell)})$, with $e_{i(\ell)} = L \left\| x^{k(\ell)} - x^{k(\ell+1)} \right\|$. □

Remark 4. In the proof of Lemma 2 we do not use the fact that $x^{k(\ell+1)}$ satisfies the descent test (4.10). Hence, the stated results also hold for any arbitrary iteration $k(\ell) \leq k < k(\ell+1)$:

$$p^{k+1} \in \partial_{L\|x^{k+1} - x^{k(\ell)}\|} F_{\tau^\ell}(x^{k(\ell)}).$$

▷

Furthermore, taking $s_X^{k+1} \in N_X(x^{k(\ell)})$, one can derive from Lemma 2 the following relation:

$$p^{k+1} + s_X^{k+1} \in \partial_{L\|x^{k+1} - x^{k(\ell)}\|} F_{\tau^\ell}(x^{k(\ell)}) + N_X(x^{k(\ell)}).$$

Proposition 5. Assume that $X \neq \emptyset$ is a bounded polyhedron, and that Algorithm 1 performs infinitely many serious steps, then any cluster point $\bar{x} \in X$ of the sequence $\{x^{k(\ell)}\}_\ell$ is a critical point of problem (4.13). ◀

Proof. We first prove that

$$\lim_{\ell \to \infty} \left\| x^{k(\ell+1)} - x^{k(\ell)} \right\| = 0 \,. \tag{4.15}$$

To this end, we need to analyze the two cases of Lemma 1. In case i), Algorithm 1 produces a feasible point for (3.7) after finitely many serious steps and all the subsequent points are feasible. Let $x^{k(\ell_1)}$ be the first feasible serious iterate. Then, Lemma 1 i) at iteration $k = i(\ell)$ yields

$$f(x^{k(\ell+1)}) \leq f(x^{k(\ell)}) - \kappa \tfrac{\mu_{\min}}{2} \left\| x^{k(\ell+1)} - x^{k(\ell)} \right\|^2 \text{ and } c(x^{k(\ell+1)}) \leq 0 \text{ for all } \ell \geq \ell_1.$$

The telescopic sum of the first inequality above yields

$$\begin{aligned} \sum_{\ell=\ell_1}^{\infty} \|x^{k(\ell+1)} - x^{k(\ell)}\|^2 &\leq \frac{2}{\kappa \mu_{\min}} \sum_{\ell=\ell_1}^{\infty} \left(f\left(x^{k(\ell)}\right) - f\left(x^{k(\ell+1)}\right) \right) \\ &\leq \frac{2}{\kappa \mu_{\min}} \left(f\left(x^{k(\ell_1)}\right) - \lim_{\ell \to \infty} f\left(x^{k(\ell+1)}\right) \right). \end{aligned}$$

Since f is finite-valued and continuous over the bounded set X, the right-hand side of the above inequality is finite. Hence, (4.15) holds. Assume now that the sequence $\{x^{k(\ell)}\}_\ell$ is infeasible for (3.7). Lemma 1 ii) at iteration $k = i(\ell)$ yields

$$0 < c(x^{k(\ell+1)}) \leq c(x^{k(\ell)}) - \kappa \tfrac{\mu_{\min}}{2} \left\| x^{k(\ell+1)} - x^{k(\ell)} \right\|^2 \text{ for all } \ell.$$

Once again, by using the telescopic sum we get (4.15). In order to show that $\bar{x}$ is a critical point of problem (4.13), following definition (3.17) we need to prove: $[\partial F_{\bar{\tau}}(\bar{x}) + N_X(\bar{x})] \cap \partial G(\bar{x}) \neq \{\emptyset\}$. To this end, we study the asymptotic behavior of the three sequences $\{x^{k(\ell)}\}_{\ell \in \mathbb{N}}$, $\{p^{k(\ell+1)}\}_{\ell \in \mathbb{N}}$, and $\{s_X^{k(\ell+1)}\}_{\ell \in \mathbb{N}}$ that verifies by definition $s_X^{k(\ell+1)} \in N_X(x^{k(\ell+1)})$ for any ℓ. Firstly, as $\{x^{k(\ell)}\}_{\ell \in \mathbb{N}} \subset X$, this sequence is bounded. Secondly, $\{p^{k(\ell+1)}\}_{\ell \in \mathbb{N}}$ is also bounded as Lemma 2 ensures that $p^{k(\ell+1)} \in \partial_{e_{i(\ell)}} F_{\tau^\ell}(x^{k(\ell)})$ with $e_{i(\ell)} = L \left\| x^{k(\ell+1)} - x^{k(\ell)} \right\|$ for a constant $L > 0$ and for any $\ell \in \mathbb{N}$. Lastly, rewriting equality (4.7) provides

$$s_X^{k+1} = \mu^k (x^{k(\ell)} - x^{k+1}) - p^{k+1} + (s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}),$$

which proves that $\{s_X^{k(\ell+1)}\}_{\ell \in \mathbb{N}}$ also is a bounded sequence as $\{s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}\}_{\ell \in \mathbb{N}} \subset \partial G(x^{k(\ell)})$, $\{x^{k(\ell)}\}_{\ell \in \mathbb{N}} \subset X \subset \Omega$, and $\{\mu^k\}_{k \in \mathbb{N}}$ is bounded by definition. Note also that $\lim_{\ell \to \infty} e_{i(\ell)} = 0$ because $\lim_{\ell \to \infty} \left\| x^{k(\ell+1)} - x^{k(\ell)} \right\| = 0$. Thus, there exist subsets $\mathcal{L}'' \subset \mathcal{L}' \subset \mathcal{L} \subset \{0, 1, \ldots\}$ such that:

- $\{x^{k(\ell)}\}_{\ell \in \mathcal{L}}$ converges to a vector $\bar{x} \in X$, and
- $\{p^{k(\ell+1)}\}_{\ell \in \mathcal{L}'}$ converges to a point $\bar{p} \in \partial F_{\bar{\tau}}(\bar{x})$ as a result of the outer semi-continuity of the subdifferential of convex functions (see [151, Prop. 4.1.1] for more details, and note that $\bar{\tau} = \lim_{\ell \in \mathcal{L}'} \tau^\ell$ by definition (4.5)), and

- $\{s_X^{k(\ell+1)}\}_{\ell\in\mathcal{L}''}$ converges to a point $\bar{s} \in N_X(\bar{x})$ (using the definition of normal cone and taking the limit in the underlying inequality).

All in all, $\bar{p} + \bar{s} \in \partial F_{\bar{\tau}}(\bar{x}) + N_X(\bar{x})$. We then show that

$$\bar{p} + \bar{s} \;=\; \lim_{\ell\in\mathcal{L}''} (s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}) \quad (\text{ i.e., } \; \bar{p} + \bar{s} \in \partial G(\bar{x}) \;),$$

which concludes the proof because $(s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}) \in \partial G(x^{k(\ell)})$ for all ℓ. This latter result follows directly from equation (4.7) and condition $\mu^k \le \mu_{\max} < \infty$:

$$\begin{aligned}
\bar{p} + \bar{s} = \lim_{\ell\in\mathcal{L}''} [p^{k(\ell+1)} + s_X^{k(\ell+1)}] &= \lim_{\ell\in\mathcal{L}''} \left[(s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}) + \mu^{i(\ell)} \left(x^{k(\ell)} - x^{k(\ell+1)} \right) \right] \\
&= \lim_{\ell\in\mathcal{L}''} (s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}) + \lim_{\ell\in\mathcal{L}''} \mu^{i(\ell)} \left(x^{k(\ell)} - x^{k(\ell+1)} \right) \\
&= \lim_{\ell\in\mathcal{L}''} (s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}) \quad (\in \partial G(\bar{x}) \;),
\end{aligned}$$

where the third equality holds because $\{s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}\}_{\ell\in\mathbb{N}}$ is a bounded sequence, and the last equality is due to (4.15) and boundedness of $\{\mu^k\}$. □

4.2.3 Convergence analysis: main results

Convergence analysis of Algorithm 1 is summarized in the following two theorems. The first one ensures that a critical point of the reformulated problem with improvement function is (asymptotically) computed by our algorithm. The second theorem establishes the link with the computed point and criticality for the original problem (3.7).

Theorem 5. Consider Algorithm 1 and suppose that $X \neq \emptyset$ is a bounded polyhedron. If the stopping-test tolerance satisfies $\delta_{\text{Tol}} = 0$, then any cluster point $\bar{x}$ of the sequence of stability centers $\{x^{k(\ell)}\}_\ell$ generated by the algorithm is critical for the problem (4.13). (If $\{x^{k(\ell)}\}_\ell$ is finite, $\bar{x} = x^{k(\ell)}$ is the last stability center.)

Moreover, if the stopping-test tolerance $\delta_{\text{Tol}} > 0$, then the algorithm stops after finitely many steps with an approximate critical point $\bar{x} = x^{k(\ell)}$, namely

$$[\partial_{L\delta_{\text{Tol}}} F_{\bar{\tau}}(\bar{x}) + N_X(\bar{x})] \cap [\partial G(\bar{x}) + B(0; \mu_{\max}\, \delta_{\text{Tol}})] \neq \{\emptyset\},$$

where $L > 0$ is the constant in Lemma 2 and $B(0; \mu_{\max}\, \delta_{\text{Tol}})$ is the closed ball in $\mathbb{R}^n$ with center at zero and with radius $\mu_{\max}\, \delta_{\text{Tol}}$. ◀

Proof. The analysis of the case with $\delta_{\text{Tol}} = 0$ results from Proposition 5 if infinitely many serious steps are generated, and from Proposition 4 otherwise.

Proposition 4 ensures that $\lim_{k\to\infty} \left\| x^{k+1} - x^{k(\ell)} \right\| = 0$ if $x^{k(\ell)}$ is the last stability center, and Proposition 5 yields $\lim_{\ell\to\infty} \left\| x^{k(\ell+1)} - x^{k(\ell)} \right\| = 0$ otherwise (see (4.15)). In both cases, Algorithm 1 stops after a finite number of steps provided that $\delta_{\text{Tol}} > 0$. Suppose $\delta_{\text{Tol}} > 0$, let $x^{k(\ell)}$ be the last stability center computed by the algorithm, and k its last iteration. Then $\left\| x^{k+1} - x^{k(\ell)} \right\| \le \delta_{\text{Tol}}$. Lemma 2 and Remark 4 assert that $p^{k+1} + s_X^{k+1} \in \partial_{L\delta_{\text{Tol}}} F_{\tau^\ell}(x^{k(\ell)}) + N_X(x^{k(\ell)})$. Furthermore, Eq. (4.7) yield

$$\left\| p^{k+1} + s_X^{k+1} - (s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}) \right\| = \mu^k \left\| x^{k+1} - x^{k(\ell)} \right\| \le \mu_{\max}\, \delta_{\text{Tol}},$$

with $s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)} \in \partial G(x^{k(\ell)})$. These properties show that $\bar{x} = x^{k(\ell)}$ is, as stated, an approximate critical point of problem (4.13). □

Remark 5. From Theorem 5, one could derive another formulation to characterize the approximate criticality of $x^{k(\ell)} = \bar{x} \in X$ that resembles formulation (3.17). Under the assumptions of Theorem 5, assume that $G : \mathbb{R}^n \mapsto \mathbb{R} \cup \{\infty\}$ is lower-semicontinuous and that $\mathcal{D}\mathrm{om}(G)$ is bounded: there exists $\Lambda > 0$ such that $\|x - \bar{x}\| \leq \Lambda$ for all $x \in \mathcal{D}\mathrm{om}(G)$. Let $s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)} \in \partial G(\bar{x})$, and $x \in \mathcal{D}\mathrm{om}(G)$ arbitrarily chosen; invoking the convexity of G and equality (4.7) yields:

$$\begin{aligned} G(x) &\geq G(\bar{x}) + \langle s_{f_2}^{k(\ell)} + s_{c_2}^{k(\ell)}, x - \bar{x} \rangle \\ &= G(\bar{x}) + \langle \mu^k (x^{k+1} - \bar{x}) + p^{k+1} + s_X^{k+1}, x - \bar{x} \rangle \\ &= G(\bar{x}) + \langle p^{k+1} + s_X^{k+1}, x - \bar{x} \rangle - \langle \mu^k (\bar{x} - x^{k+1}), x - \bar{x} \rangle \\ &\geq G(\bar{x}) + \langle p^{k+1} + s_X^{k+1}, x - \bar{x} \rangle - \mu_{\max} \delta_{\mathrm{Tol}} \Lambda \,. \end{aligned}$$

As this last inequality trivially holds for $x \notin \mathcal{D}\mathrm{om}(G)$, we conclude that $p^{k+1} + s_X^{k+1} \in \partial_{\mu_{\max}\delta_{\mathrm{Tol}}\Lambda} G(\bar{x})$. Since $p^{k+1} + s_X^{k+1} \in \partial_{L\delta_{\mathrm{Tol}}} F_{\bar{\tau}}(\bar{x}) + N_X(\bar{x})$ (see the proof of Theorem 5), this shows that the approximate critical point $\bar{x}$ of problem (4.13) verifies

$$[\partial_{L\delta_{\mathrm{Tol}}} F_{\bar{\tau}}(\bar{x}) + N_X(\bar{x})] \cap \partial_{\mu_{\max}\delta_{\mathrm{Tol}}\Lambda} G(\bar{x}) \neq \{\emptyset\}.$$

▷

Theorem 5 provides the convergence analysis of Algorithm 1. The next theorem establishes the link between (4.13) and (3.7). The stated results are closely related to the optimality conditions discussed in [34, Lemma 5.1], where a different setting is considered: no DoC structure is assumed (but f is twice continuously differentiable). For the sake of completeness, we include the mathematical proof tailored to our DoC setting. We start by presenting a simple lemma that will be useful for the next Theorem: it helps compute the subdifferential of $F_\tau(x) = \max\{f_1(x) + c_2(x) - \tau_f,\, f_2(x) + c_1(x) - \tau_c\}$.

Lemma 3. Let $\bar{x} \in X$ and $\bar{\tau} = (\bar{\tau}_f, \bar{\tau}_c)$ as in (4.13) be given.

i) If $c(\bar{x}) > 0$, then $f_1(\bar{x}) + c_2(\bar{x}) - \bar{\tau}_f < f_2(\bar{x}) + c_1(\bar{x}) - \bar{\tau}_c$.

ii) If $c(\bar{x}) < 0$, then $f_1(\bar{x}) + c_2(\bar{x}) - \bar{\tau}_f > f_2(\bar{x}) + c_1(\bar{x}) - \bar{\tau}_c$.

iii) If $c(\bar{x}) = 0$, then $f_1(\bar{x}) + c_2(\bar{x}) - \bar{\tau}_f = f_2(\bar{x}) + c_1(\bar{x}) - \bar{\tau}_c$.

◀

Proof. The definition of $\bar{\tau} = \big(f(\bar{x}) + \rho \max\{c(\bar{x}), 0\}, \sigma \max\{c(\bar{x}), 0\}\big)$ with $\rho \geq 0$ and $\sigma \in [0, 1)$ gives:

$$\begin{aligned} f_1(\bar{x}) + c_2(\bar{x}) - \bar{\tau}_f - (f_2(\bar{x}) + c_1(\bar{x}) - \bar{\tau}_c) &= (\sigma - \rho) \max\{c(\bar{x}), 0\} - c(\bar{x}) \\ &= \begin{cases} (\sigma - 1 - \rho) c(\bar{x}) < 0 & \text{if } c(\bar{x}) > 0 \\ -c(\bar{x}) \geq 0 & \text{if } c(\bar{x}) \leq 0. \end{cases} \end{aligned}$$

□

We now present the main convergence Theorem.

Theorem 6. Under the assumptions of Theorem 5 and $\delta_{\text{Tol}} = 0$, let $\bar{x} \in X$ be the point (asymptotically) computed by Algorithm 1. Furthermore, assume that f_2 and c_2 are continuously differentiable at $\bar{x}$. Then $\bar{x}$ is a d-stationary point for (4.13) and the following statements hold:

i) If $c(\bar{x}) > 0$, then $\bar{x}$ is a d-stationary point of the DoC problem $\min_{x\in X}\ c_1(x) - c_2(x)$.

ii) If $c(\bar{x}) < 0$, the pair $(\bar{x}, 0)$ satisfies the KKT system (3.15) of the DoC-constrained DoC program (3.7). Furthermore, $\bar{x}$ is a d-stationary point of the DoC problem $\min_{x\in X}\ f_1(x) - f_2(x)$.

iii) If $c(\bar{x}) = 0$ and problem (3.7) satisfies the extended Mangasarian-Fromowitz constraint qualification (EMFCQ), then there exists a scalar $\bar{\lambda} > 0$ such that the pair $(\bar{x}, \bar{\lambda})$ satisfies the KKT system (3.15) of the DoC-constrained DoC program (3.7). Furthermore, $\bar{x}$ is a B-stationary point of (3.7).

◀

Proof. According to Theorem 5, $\bar{x}$ is critical for (4.13). By definition, $\bar{x}$ solves the convex subproblem $\min_{x\in X}\ F_{\bar{\tau}}(x) - \langle \bar{s}_{f_2} + \bar{s}_{c_2}, x\rangle$ for the pair of vectors $\bar{s}_{f_2} = \nabla f_2(\bar{x})$ and $\bar{s}_{c_2} = \nabla c_2(\bar{x})$, then $\bar{x}$ is indeed a d-stationary point of (4.13) and, according to (3.16),

$$\nabla G(\bar{x}) = \nabla f_2(\bar{x}) + \nabla c_2(\bar{x}) \in \partial F_{\bar{\tau}}(\bar{x}) + N_X(\bar{x}).$$

If $c(\bar{x}) > 0$, Lemma 3(i) (in the Appendix A), implies that $F_{\bar{\tau}}(\bar{x}) = f_2(\bar{x}) + c_1(\bar{x}) - \bar{\tau}_c$ $(> f_1(\bar{x}) + c_2(\bar{x}) - \bar{\tau}_f)$ and, therefore, $\partial F_{\bar{\tau}}(\bar{x}) = \nabla f_2(\bar{x}) + \partial c_1(\bar{x})$. By replacing this identity in the equality above we get $\nabla c_2(\bar{x}) \in \partial c_1(\bar{x}) + N_X(\bar{x})$. This proves item i).

If $c(\bar{x}) \leq 0$, the subdifferential of $F_{\bar{\tau}}(x)$ satisfies (thanks to (3.5))

$$\partial F_{\bar{\tau}}(x) \subset \bigcup_{\lambda\in[0,1]} \lambda[\partial f_1(x) + \partial c_2(x)] + (1-\lambda)[\partial f_2(x) + \partial c_1(x)].$$

As a consequence of $\nabla G(\bar{x}) \in \partial F_{\bar{\tau}}(x) + N_X(\bar{x})$, there exists $\tilde{\lambda} \in [0,1]$ such that:

$$\nabla f_2(\bar{x}) + \nabla c_2(\bar{x}) \in \tilde{\lambda}[\partial f_1(\bar{x}) + \nabla c_2(\bar{x})] + (1-\tilde{\lambda})[\nabla f_2(\bar{x}) + \partial c_1(\bar{x})] + N_X(\bar{x})\,,$$

or equivalently,

$$0 \in \tilde{\lambda}[\partial f_1(\bar{x}) - \nabla f_2(\bar{x})] + (1-\tilde{\lambda})[\partial c_1(\bar{x}) - \nabla c_2(\bar{x})] + N_X(\bar{x})\,. \tag{4.16}$$

By using the Clarke's sum rule, note that (4.16) is equivalent to

$$0 \in \tilde{\lambda}\partial^{\text{C}}[f_1(\bar{x}) - f_2(\bar{x})] + (1-\tilde{\lambda})\partial^{\text{C}}[c_1(\bar{x}) - c_2(\bar{x})] + N_X(\bar{x})\,.$$

If $c(\bar{x}) < 0$, Lemma 3(ii) implies that $F_{\bar{\tau}}(\bar{x}) = f_1(\bar{x}) + c_2(\bar{x}) - \bar{\tau}_f$ and, therefore, $\tilde{\lambda} = 1$ above. This shows that $\nabla f_2(\bar{x}) \in \partial f_1(\bar{x}) + N_X(\bar{x})$, proving thus item ii) (take $\bar{\lambda} = 0$ in (3.15) to conclude that $(\bar{x}, \bar{\lambda})$ satisfies the KKT system of problem (3.7)).

We now show that if $c(\bar{x}) = 0$ and the EMFCQ holds, then $\tilde{\lambda} > 0$. If $c(\bar{x}) = 0$ and should $\tilde{\lambda} = 0$ be possible, then there would exist an element $s \in N_X(\bar{x})$ such that $0 \in \partial c_1(\bar{x}) - \nabla c_2(\bar{x}) + s$ (c.f. (4.16)). However, since EMFCQ holds for problem (3.7) at

$\bar{x}$, there exists $d \in T_X(\bar{x})$, such that $\langle s_1 - \nabla c_2(\bar{x}), d\rangle < 0$ for all $s_1 \in \partial c_1(\bar{x})$. Consequently: $0 = \langle 0, d\rangle = \langle s_1 - \nabla c_2(\bar{x}) + s, d\rangle = \langle s_1 - \nabla c_2(\bar{x}), d\rangle + \langle s, d\rangle \leq \langle s_1 - \nabla c_2(\bar{x}), d\rangle < 0$, a contradiction. Hence, $\tilde{\lambda} > 0$. By setting $\bar{\lambda} := \frac{(1-\tilde{\lambda})}{\tilde{\lambda}} > 0$ we conclude that the pair $(\bar{x}, \bar{\lambda})$ satisfies the KKT system (3.15). As the EMFCQ implies the CQ of Proposition 3, then (3.9) holds (see [8, § 2.4]). □

In Theorem 6, the smoothness assumption on f_2 and c_2 at $\bar{x}$ is crucial to obtain the inclusion (4.16) and, therefore, ensure that the computed vector is a KKT point of (3.7) under a constraint qualification (CQ). In other words, this inclusion makes possible the characterization of d-stationarity at $\bar{x}$ in (4.13). In the next section we discuss how to equip Algorithm 1 with escaping procedures to provide a point satisfying KKT or, even stronger, B-stationarity without the smoothness assumption. To that end, we rely on [164] and [233].

Before finalizing this section, we put the convergence results of Theorems 5 and 6 in perspective with what is known in the literature.

4.2.4 Different assumptions and results: a few comments on the related literature

The work [211] deals with multiobjective DoC optimization problems, and similarly uses a bundle algorithm with an improvement function. When tailored to the setting of a single objective and using an additional escaping procedure, the algorithm proposed in [211] is shown to compute a point $\bar{x}$ that, if feasible, satisfies the KKT system (3.15) provided that the following assumptions are met: a CQ holds at this point, and the subdifferentials of $\partial F_\tau(x)$ and $\partial G(x)$ are polytopes for all $x \in \mathbb{R}^n$. Differently, the analysis of Theorem 6 also covers infeasibility as well as establishes a direct link between the KKT system of the original problem (3.15) and the obtained criticality for the reformulated DoC problem using the improvement function (4.2). Necessary assumptions include that a CQ holds and continuous differentiability of f_2 and c_2 at $\bar{x}$, but exclude the assumption on ∂F_τ used in [211].

As in [8, 211], our analysis concerns any arbitrary cluster point of the sequence of stability centers $\{x^{k(\ell)}\}_\ell$. However, we have not established that such a sequence is indeed convergent. Two natural questions arise: (i) Is the sequence of stability centers convergent? (ii) If yes, what is its rate of convergence? These are non-trivial questions that deserve consideration. Unfortunately, this work answers neither (i) nor (ii). The reason is that additional and non-trivial assumptions are required not only about the involved functions f and c but also about the structure of the model approximating the improvement function. Nevertheless, we care to mention that (i) and (ii) have already been addressed in other publications in different contexts [222, 296]. A common assumption in these references is the requirement that the functions at hand satisfy the Kurdyka-Łojasiewicz (KŁ) property/inequality. Roughly speaking, the KŁ property says that one can bound the subgradients of a function from below by a reparametrization of its function values; see [222, § 2.1] for the mathematical definition.

The reference [222] deals with nonsmooth, nonconvex unconstrained optimization problems under the assumption that the objective function satisfies the KŁ property. The work does not assume any DoC structure and presents convergence results of a bundle method employing more sophisticated models requiring, among other hypotheses,

Clarke subgradients of the objective function. The latter assumption is not required by our method, which works with subgradients of the DoC component functions. If the so-called first-order standard model verifies the (strong) KŁ property [222, Theorem 3.2], then the sequence of stability centers converges to a (Clarke) stationary point. The rate of convergence is not assessed.

An analysis on the rate of convergence of proximal-like methods for nonsmooth, nonconvex unconstrained optimization under the KŁ property can be found in [37]. Concerning the unconstrained DoC setting, the work [296] studies a DCA-like algorithm and proves convergence of the generated sequence to a critical point under KŁ and the additional assumption that the underlying functions are *good-DC* (roughly speaking, the functions are Clarke regular). Furthermore, rates of convergence are established. Likewise, but with a double proximal algorithm for unconstrained DoC programs, the paper [42] establishes convergence of the generated sequence and assesses its rate of convergence under the KŁ property. Reference [33] also establishes similar results for a Generalized Proximal Point Algorithm, applied to a somewhat larger nonconvex nonsmooth setting that encompasses the structure of unconstrained DoC problems (namely only the second component is assumed to be convex).

Rates of convergence in the DoC-constrained setting are investigated in the recent technical report [301], where another DCA-like method is proposed. To the best of our knowledge, this is the first attempt to analyze rate of convergence of an algorithm for DoC-constrained DoC problems satisfying the KŁ property. In contrast to our work that employs a model approximating the improvement function, the DoC component functions f_1 and c_1 are handled by the method of [301]. This is an issue when dealing with general problems fitting the structure of (3.7).

The study of the KŁ property in a bundle algorithm applied to DoC problems is left for future research. We point out a main difficulty in this endeavor: when considering the DoC-constrained setting, the resulting convexly-constrained DoC problem (4.2) changes at every serious step. This fact precludes the application of the analysis developed in the references [33, 222, 37, 296, 42, 301] discussed above.

4.3 KKT AND B-STATIONARY POINTS UNDER MILDER ASSUMPTIONS: ESCAPING PROCEDURES

The smoothness assumption in Theorem 6 essentially ensures that the vector $\bar{x}$ delivered by Algorithm 1 is a d-stationary point and, therefore, a Clarke-stationary point of (4.13). A result similar to Theorem 6 (iii) (but with no details on cases (i) or (ii)) is given in [211], pointing out that any Clarke-stationary point of (4.13) (if feasible and if a CQ holds) is a KKT point of (3.7). To compute a Clarke-stationary point of (4.13) without assuming smoothness on f_2 and c_2 at $\bar{x}$, the authors of [211] equip their algorithm with the escaping procedure proposed in [164]. Such a procedure requires the assumption that the subdifferentials $\partial F_\tau(x)$ and $\partial G(x)$ are polytopes to either prove that the given vector $\bar{x}$ is an approximate Clarke-stationary point of (4.13) (and therefore an approximated KKT point of (3.7)) or provide a better candidate to solve (3.7). The same escaping procedure of [164] and [211] can be employed by our algorithm in a straightforward manner:

- let $\bar{x} = x^k$ be the point delivered by Algorithm 1. Assume that $\bar{x}$ is feasible to the original problem (3.7);
- perform the escaping procedure of [164, 211] on $\bar{x}$; see [211, Algorithm 1] for details;
- if $\bar{x}$ is a proven approximate Clarke-stationary point of (4.13), then stop and return $\bar{x}$;
- if $\bar{x}$ is not approximately Clarke-stationary, then the escaping procedure delivers a better point, say x^{k+1}, that should be used to continue the iterative process of Algorithm 1.

Under the assumptions that $\partial F_\tau(x)$ and $\partial G(x)$ are polytopes for all $x \in \mathbb{R}^n$, Algorithm 1 implementing the above steps terminates after finitely many steps provided $\delta_{\text{Tol}} > 0$: to see that, just combine Lemma 3 in [211] with Theorem 5 above. If a CQ holds, then the delivered vector is a KKT point of the original problem (3.7). A stronger result (B-stationarity) is possible under the assumption that both f_2 and c_2 are the pointwise maximum of finitely many differentiable convex functions. This is the subject of the next subsection, where we rely on the ideas investigated in [233].

4.3.1 Ensuring B-stationarity

Suppose that functions f_2 and c_2 have the structure given in (3.12). For any given point $x \in X$, their subdiffentials are defined in (3.13). As discussed in Section 3.3, verifying the B-stationarity condition (3.10) of a given point $\bar{x}$ feasible for (3.7) amounts to checking whether $\bar{x}$ solves (3.14) for all $i \in A^f(\bar{x})$ and all $j \in A^c(\bar{x})$. Since $\bar{x}$ is by assumption feasible for the original problem (3.7), then $\bar{x}$ is also feasible for the convex problem (3.14). To ensure that (3.14) has a unique solution, we can w.l.o.g. add the quadratic term $\frac{1}{2}\|x - \bar{x}\|^2$ to its objective function. The escaping procedure for Algorithm 1 we propose is as follows.

Algorithm 2 Escaping procedure

Hypothesis: f_2 and c_2 satisfying (3.12) and x^k a feasible point of (3.7) satisfying the stopping test of Algorithm 1.

1: Let $\delta_{\text{Tol}} > 0$, $\mu > 0$ be given, set $\bar{x} := x^k$, $A^f(\bar{x})$ and $A^c(\bar{x})$ as in (3.13)
2: **for** $j \in A^c(\bar{x})$ **do**
3: **for** $i \in A^f(\bar{x})$ **do**
4: Let $z^{i,j}$ be the solution of the strongly convex subproblem

$$\begin{cases} \min\limits_{x \in X} & f_1(x) - [\psi_i(\bar{x}) + \langle \nabla \psi_i(\bar{x}), x - \bar{x} \rangle] + \frac{\mu}{2}\|x - \bar{x}\|^2 \\ \text{s.t.} & c_1(x) - [\phi_j(\bar{x}) + \langle \nabla \phi_j(\bar{x}), x - \bar{x} \rangle] \leq 0 \end{cases}$$

5: **if** $\left\|z^{i,j} - \bar{x}\right\| > \delta_{\text{Tol}}$ **then**
6: Set $x^{k(\ell+1)} := z^{i,j}$, $\ell := \ell + 1$ and continue Algorithm 1 from Step 1
7: Stop both this pseudo-algorithm and Algorithm 1: $\bar{x}$ is an approximate B-stationary point of problem (3.7)

Proposition 6. Let $z^{i,j}$ be as in Pseudo-algorithm 2 for any $i \in A^f(\bar{x})$ and $j \in A^c(\bar{x})$.

i) $z^{i,j}$ is feasible for (3.7) and it satisfies $f(z^{i,j}) \leq f(\bar{x}) - \frac{\mu}{2}\left\|z^{i,j} - \bar{x}\right\|^2$.

ii) If $\left\|z^{i,j} - \bar{x}\right\| \leq \delta_{\text{Tol}}$ for all $i \in A^f(\bar{x})$ and $j \in A^c(\bar{x})$, then $\bar{x}$ satisfies (3.10) within the tolerance $\delta_{\text{Tol}} > 0$. If, in addition, problem (3.7) satisfies EMFCQ, then $\bar{x}$ is an approximate *B*-stationary point of the original problem (3.7).

◀

Proof. i) By definition, $z^{i,j}$ satisfies the constraint in subproblem (3.14). Then, the definition of c_2 in (3.12) and the convexity of ϕ_j yield

$$c_1(z^{i,j}) - c_2(z^{i,j}) \leq c_1(z^{i,j}) - \phi_j(z^{i,j}) \leq c_1(z^{i,j}) - [\phi_j(\bar{x}) + \langle \nabla\phi_j, z^{i,j} - \bar{x}\rangle] \leq 0\,,$$

showing that $z^{i,j}$ is feasible to problem (3.7). We show in the same way, by definition of f_2 in (3.12), by convexity of ψ_i and by definition of $z^{i,j}$, that $z^{i,j}$ is a better candidate than $\bar{x}$:

$$\begin{aligned} f(z^{i,j}) + \frac{\mu}{2}\left\|z^{i,j} - \bar{x}\right\|^2 &= f_1(z^{i,j}) - \psi_i(z^{i,j}) + \frac{\mu}{2}\left\|z^{i,j} - \bar{x}\right\|^2 \\ &\leq f_1(z^{i,j}) - [\psi_i(\bar{x}) + \langle \nabla\psi_i(\bar{x}), z^{i,j} - \bar{x}\rangle] + \frac{\mu}{2}\left\|z^{i,j} - \bar{x}\right\|^2 \\ &\leq f_1(\bar{x}) - \psi_i(\bar{x}) = f_1(\bar{x}) - f_2(\bar{x}) = f(\bar{x})\,. \end{aligned}$$

ii) If $\left\|z^{i,j} - \bar{x}\right\| \leq \delta_{\text{Tol}}$ for all $i \in A^f(\bar{x})$ and $j \in A^c(\bar{x})$, then the point $\bar{x}$ delivered by Algorithm 1 approximately solves (3.14) for every $i \in A^f(\bar{x})$ and $j \in A^c(\bar{x})$. Thanks to (3.13), we then conclude that $\bar{x}$ approximately solves (3.10); thus it is a *B*-stationary point of (3.7) under the CQ [8, subsection 2.2].

□

We care to mention that the subproblem of Pseudo-algorithm 2 does not need to be solved up to optimality when $\bar{x}$ is not among its solutions: it suffices to compute a feasible point $z^{i,j}$ providing a better functional value than $f(\bar{x})$. To that end, we can for example employ the convex bundle algorithm of [22]: by starting its iterative process with the feasible point $\bar{x}$ then all the generated iterates will remain feasible. Therefore, as soon as an iterate of better quality than $\bar{x}$ is found we can terminate the convex algorithm, the escape procedure, and continue with Algorithm 1 to obtain a new and necessarily better critical point of (4.13).

Theorem 7. Suppose, in addition to the assumptions of Theorem 5, that (3.12) holds. If $\delta_{\text{Tol}} > 0$, then Algorithm 1 with Pseudo-algorithm 2 stops after finitely many steps with an approximate *B*-stationary point of (3.7) provided a CQ holds. ◀

Proof. Since $\delta_{\text{Tol}} > 0$, every run of Algorithm 1 terminates after finitely many steps. This is also the case for Pseudo-algorithm 2 that requires solving at most $m_f \cdot m_c$ convex subproblems (c.f. (3.12)). Let $\{\bar{x}^\iota\}$ be the subsequence of points at which Algorithm 1 calls the escape procedure of Pseudo-algorithm 2. For a fixed $\bar{x}^\iota$, if $\left\|z^{i,j} - \bar{x}^\iota\right\| > \delta_{\text{Tol}}$ in the escaping procedure then Proposition 6 ensures that

$$f(z^{i,j}) \leq f(\bar{x}^\iota) - \frac{\mu}{2}\left\|z^{i,j} - \bar{x}^\iota\right\|^2 < f(\bar{x}^\iota) - \frac{\mu}{2}\delta_{\text{Tol}}^2. \tag{4.17}$$

Algorithm 1 receives $x^{k(\ell+1)} := z^{i,j}$ and continues its iterative process until stopping with another (approximated) critical point $\bar{x}^{\iota+1} \neq \bar{x}^{\iota}$ (note, however, that $\bar{x}^{\iota+1}$ can be equal to $z^{i,j}$). Since $x^{k(\ell+1)}$ is feasible for problem (3.7), Lemma 1(i) ensures that $\bar{x}^{\iota+1}$ is feasible as well and, moreover, $f(\bar{x}^{\iota+1}) \leq f(z^{i,j})$. By combining this inequality with (4.17) we get

$$f(\bar{x}^{\iota+1}) < f(\bar{x}^{\iota}) - \frac{\mu}{2}\delta_{\text{Tol}}^2, \quad \text{for all } \iota = 1, 2, \ldots .$$

Since f is bounded from below, because f is continuous on compact X, then the above inequality ensures that the loop between Algorithm 1 and Pseudo-Algorithm 2 is finite. Proposition 6 (ii) then concludes the proof. □

With a proper management, in Algorithm 1, of the tolerance $\delta_{\text{Tol}} > 0$ such that $\delta_{\text{Tol}} \to 0$, the escaping procedure could be able to asymptotically compute an exact B-stationary point of (3.7). To that end, we could rely on Theorem 7 with $\delta_{\text{Tol}} \to 0$ by following the general idea of [233, § 5.1] as in [8, Corollary 3.4] with γ therein replaced by δ_{Tol}. But we shall not go into theoretical analysis of such modifications, as they will cause some technical complications, while the conceptual ideas are clear from our simpler presentation of the escaping algorithm.

4.4 NUMERICAL EXPERIMENTS

Before dealing with chance-constrained problems, we illustrate the numerical performance of Algorithm 1 on a bi-dimensional example and the challenging circle packing problem, which have simple geometric interpretations. The goal is to show that although we start the optimization process with a infeasible point, the algorithm quickly computes a feasible point and the subsequent iterates remain feasible (c.f. Lemma 1). All the experiments in this section have been carried out using Matlab R2017a in a computer with Intel(R) Core(TM), i3-3110M CPU 2.40, 4G (RAM), under Windows 10, 64Bits. The choice of the parameters in Algorithm 1 (named `PBMDC`^{2} for short) is as follows:

> `PBMDC`^{2}- Algorithm 1 with $\mu_{\min} = 10^{-6}$, $\mu_{\max} = 10^5$, $\mu^0 = 1$, $\rho = \sigma = 0.5$ and $\delta_{\text{Tol}} = 10^{-6}$. Whenever the size of the bundle is greater or equal to 100 we keep in the bundle only the active constraints according to (4.9).

4.4.1 AN ILLUSTRATIVE EXAMPLE

In order to plot the objective function and feasible set, let us consider the following bi-dimensional problem

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^2} & \|x\|^2 + \sum_{i=1}^2 x_i - \|x\|_1 \\ \text{s.t.} & \min\limits_{j=1,2,3} \left\{ \left\| x - p^j \right\|^2 - r_j^2 \right\} \leq 0 \\ & -2 \leq x_i \leq 2,\ i = 1, 2, \end{array} \tag{4.18}$$

with $p^1 = (-1; 0)^\top$, $p^2 = (0; 0)^\top$, $p^3 = (0.5; -1)^\top$, and $r = (0.5, 0.3, 0.5)^\top$. Problem (4.18) fits the general formulation (3.7) with

$$f_1(x) = \|x\|^2 + \sum_{i=1}^{2} x_i, \qquad f_2(x) = \|x\|_1 ,$$

$$c_1(x) = \sum_{j=1}^{3} \left[\left\| x - p^j \right\|^2 - r_j^2 \right], \quad \text{and} \quad c_2(x) = \max_{j=1,2,3} \sum_{l \neq j} \left\{ \left\| x - p^l \right\|^2 - r_l^2 \right\}.$$

The objective function $f = f_1 - f_2$ is plotted in Figure 4.1(a), and its level curves are visible in Figures 4.1(b)-(f). Observe that the constraint $c_1(x) - c_2(x) \leq 0$ defines a disconnected feasible set, which is the union of three blue circles in Figures 4.1(b)-(f). The global solution of the problem is $\bar{x}^1 = (-1, -0.5)^\top$ (with $f(\bar{x}^1) = -1.75$), and the local solutions are $\bar{x}^2 = (-0.2121, -0.2121)^\top$ (with $f(\bar{x}^2) = -0.7585$) and $\bar{x}^3 = (0, -1)^\top$ (with $f(\bar{x}^3) = -1$). As functions f_2 and c_2 are nonsmooth, the problem has critical points that fail to be d-stationary. This is the case, for instance, of points $\bar{x}^4 = (-0.3, 0)$, $\bar{x}^5 = (0, 0)^\top$ and $\bar{x}^6 = (0, -0.3)$. However, since both f_2 and c_2 are the pointwise maximum of finitely many smooth functions, we can apply the escaping procedure of Pseudo-Algorithm 2 to ensure that Algorithm 1 always computes a d-stationary point.

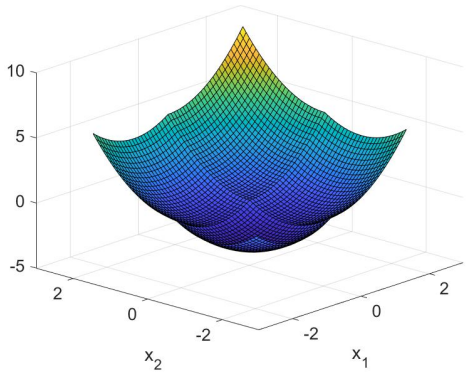
We first run Algorithm 1 without the escaping procedure on problem (4.18). Figure 4.1(b) (respectively 4.1(c)) presents one hundred sequences of points $\{x^k\}$ generated by Algorithm 1 with $\mu^0 = 10$ (respectively $\mu^0 = 1$) starting from one hundred initial points uniformly generated in the square $-2 \leq x_i \leq 2$, $i = 1, 2$. In this example, when the initial prox-parameter μ^0 is large, all the (one hundred) sequences converge to a local solution (and therefore d-stationary point): as we can see from Figure 4.1(b), all the sequences are well stabilized, converging to three limit points, and do not get trapped by critical points of poor quality. However, with $\mu = 1$ the sequences oscillate (see Figure 4.1(c)) and six of the hundred sequences terminate at either $(0, -0.3)^\top$ or $(-0.3, 0)^\top$, which are critical but not d-stationary for problem (4.18).

To facilitate visualization, we consider in Figures 4.1(e) and (f) a single initialization that leads to $\bar{x} = (0, -0.3)^\top$. Let us focus on this sequence. We can see that once the algorithm computes the first feasible point, the remainder of the sequence follows a direction of descent until it stops at $\bar{x}$. The reason why the algorithm terminates at $\bar{x}$ is the subgradient $s_{f_2} = (0, -1)^\top$ of $f_2(\bar{x})$ provided by the oracle (note that $\partial f_2(\bar{x}) = \{(0, \alpha) : \alpha \in [-1, 1]\}$ but c_2 is smooth at $\bar{x}$). With this subgradient, the objective function $f_1(x) - [f_2(\bar{x}) + \langle s_{f_2}, x - \bar{x} \rangle]$ in (3.11) is simply $\|x\|^2 + x_1 + 2x_2$. After some calculations, we can see that $\bar{x}$ solves (3.11) (confirming thus that $\bar{x}$ is critical). However, if we pick $s_{f_2} = (0, 1)^\top \in \partial f_2(\bar{x})$, the function $f_1(x) - [f_2(\bar{x}) + \langle s_{f_2}, x - \bar{x} \rangle]$ becomes $\|x\|^2 + x_1 - 0.6$ and $\bar{x}$ does not minimize it over the feasible set of (3.11). This shows that $\bar{x}$ does not solve (3.10) (i.e., $\bar{x}$ is not d-stationary).

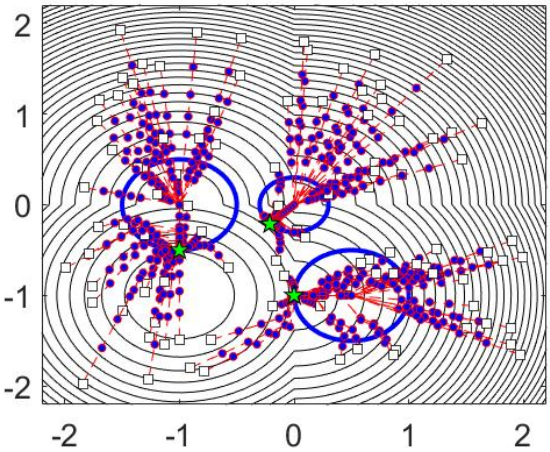
When we equip our method with the Pseudo-Algorithm 2, the algorithm is able to escape from these critical points in all the cases; see Figures 4.1(d) and 4.1(f), compared respectively to Figures 4.1(c) and 4.1(e) which have the same setting but no escaping procedure. The reason is that Pseudo-Algorithm 2 checks all the generators of $\partial f_2(\bar{x})$; c.f. Theorem 7.

4.4.2 The circle packing problem

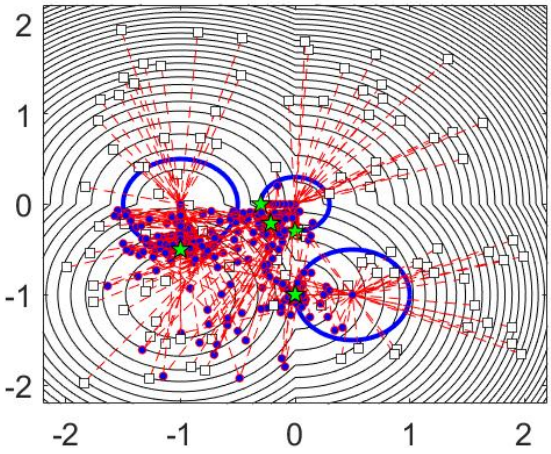
The circle packing problem is to find the largest percentage of a polygon that can be covered by a given number of equal circles. Mathematically, it consists of positioning the circles and maximizing their radius while keeping the intersection of all circles empty. We consider m circles and take the polygon to be the unit square in $\mathbb{R}^2$. Let $p_i \in \mathbb{R}^2$ be



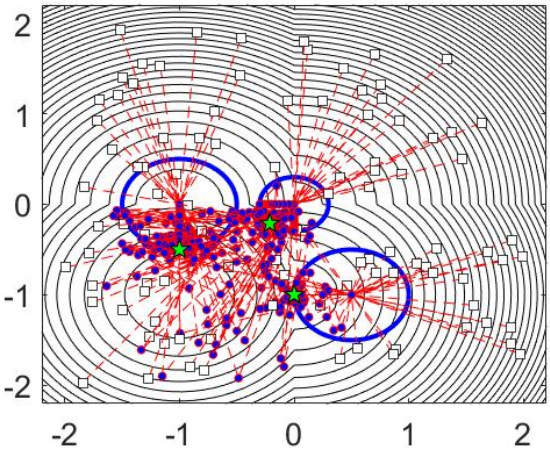
(a) Graph of f.



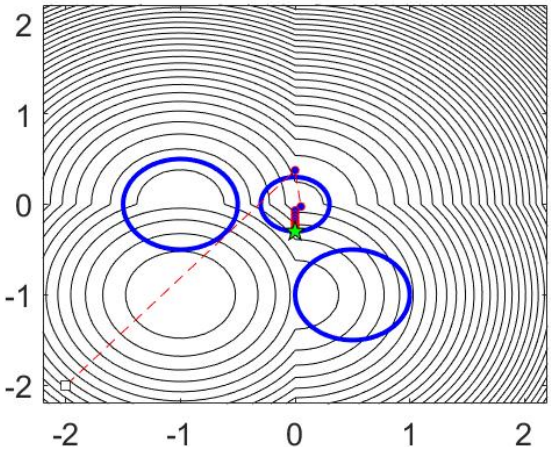
(b) 100 trajectories generated by Algorithm 1 with $\mu^0 = 10$ and without escaping procedure.



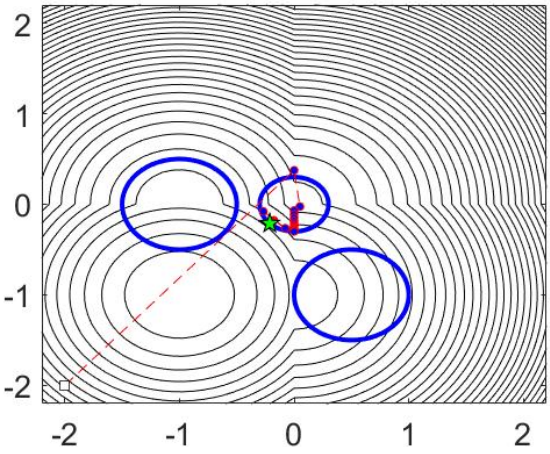
(c) 100 trajectories generated by Algorithm 1 with $\mu^0 = 1$ and without escaping procedure.



(d) 100 trajectories generated by Algorithm 1 with $\mu^0 = 1$ and with escaping procedure.



(e) Without escaping procedure.



(f) With escaping procedure.

FIGURE 4.1: Illustration of the DoC problem (4.18). The initial points, randomly generated, are marked by the white little squares, x^k by blue dots, and the computed critical points are represented by the green stars. With $\mu^0 = 10$, all the 100 sequences converge to one of the local (or global) solutions without the need of applying the escaping procedure. However, for $\mu^0 = 1$, some of these sequences terminate at critical points that are not d-stationary (subfigures (c) and (e)). When the escaping procedure is applied, then all the sequences converge to one of the local (or global) solutions (Subfigures (d) and (f)).

the position of the i^{th} center and r be the radius of the circles. The problem reads as

$$\begin{cases} \min\limits_{p\in\mathbb{R}^{2m}, r\in\mathbb{R}} & -r \\ \text{s.t.} & \|p_i - p_j\| \geq 2r \qquad i = 1,\dots,m-1, \quad j = i+1,\dots,m \\ & \mathbb{1}r \leq p_i \leq \mathbb{1}(1-r) \quad i = 1,\dots,m, \end{cases} \tag{4.19}$$

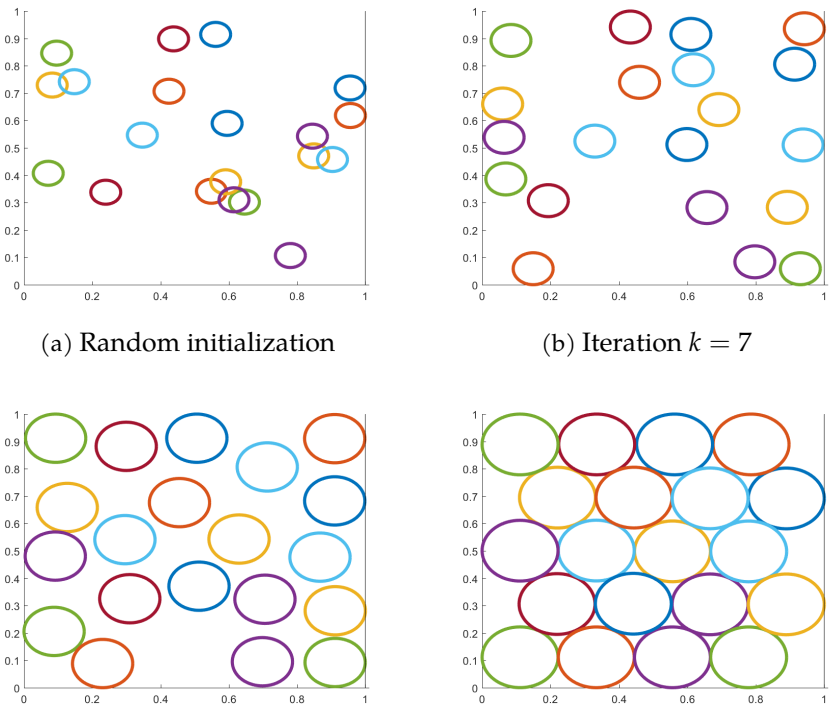
where $\mathbb{1}$ is the vector of ones in $\mathbb{R}^2$. This problem fits (3.7) with $x = (p, r)$,

$$f_1(x) = -r,\ f_2(x) \equiv 0,\ c_1(x) = 2r + \max_{i,j} \left\{ \sum_{(k,s)\neq(i,j)} \|p_k - p_s\| \right\},$$

and

$$c_2(x) = \sum_{(i,j)} \|p_i - p_j\|.$$

Although c_2 is a nonsmooth function, it is differentiable at any feasible point whenever the radius is strictly positive. Figure 4.2 presents the results obtained from Algorithm 1 for the configuration of $m = 20$ circles starting from a random initialization. The figure shows



(a) Random initialization

(b) Iteration $k = 7$

(c) Iteration $k = 60$

(d) Iteration $k = 617$

FIGURE 4.2: Packing 20 circles. PBMDC^2 computed a global solution yielding $r = 0.11138$ in 47 seconds.

that PBMDC^2 finds a feasible point for the circle packing problem in few iterations. From that iteration on, all the iterates produced by the method remain feasible which illustrates Lemma 1 (ii). For this instance, our solver PBMDC^2 was successful in computing a global solution from a random initialization in 47 seconds (and in 13 seconds for $m = 10$). The

global solution for several number of circles can be found on the specialized web page `www.packomania.com`.

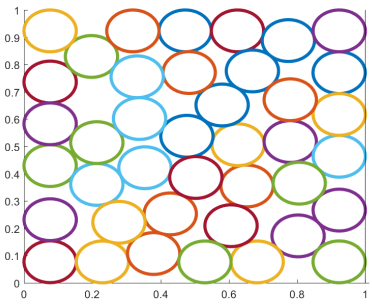


FIGURE 4.3: Packing 40 circles. PBMDC^2 computed a solution with an optimality gap of 2.8% in 160 seconds.

Figure 4.3 presents a feasible packing of $m = 40$ circles computed by PBMDC^2. In this setting, PBMDC^2 failed in computing a global solution, however the obtained radius is within an error of 2.8% of the (global) optimal value.

4.4.3 CHANCE-CONSTRAINED PROBLEMS

We now assess the numerical performance of solver PBMDC^2 on 4 sets of optimization problems with probability constraints: PLANTOY, PNORM, PROBUST, and GAS. As proposed in Section 5.3, we approximated the probability constraint $\mathbb{P}[\mathtt{lb} \leq c(x,\xi) \leq \mathtt{ub}]$ in these problems with the DoC constraint $c_1(x) - c_2(x) \leq 0$ and c_1 and c_2 given in (5.14). For these experiments, we estimated the expectation in (5.14) with a sample of $N = 10^4$ randomly generated scenarios and set $t = 0.0025$. The "moniker" PBMDC2S refers to the application of our solver to the smooth variant of the chance-constrained problem: the max-operations were replaced by their smooth versions. Because of the extra step to smooth the "max", the variant PBMDC2S requires more time than PBMDC^2 to evaluate the constraint function.

We compared our approach with two solvers exploiting the DoC structure of the problems:

SCA is our implementation in Matlab of the approach given in [153], denoted by *Sequential Convex Approximations*. The iterate x^{k+1} of this method is obtained by solving the convex subproblem

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} & f_1(x) - [f_2(x^k) + \langle s^k_{f_2}, x - x^k\rangle] \\ \text{s.t.} & c_1(x) - [c_2(x^k) + \langle s^k_{c_2}, x - x^k\rangle] \leq 0 \\ & x \in X\,. \end{cases}$$

To this end, we apply the convex bundle method of [22] (with exact oracles instead). We care to mention that the original SCA of [153] deals with convex objectives only. Its extension to DoC objectives boils down to the classic DCA algorithm for DoC-constrained DoC problems [183].

`PCCP` is our implementation in Matlab of Algorithm 3.1 in [197], called *Penalty Convex-Concave Procedure*. The iterate x^{k+1} of `PCCP` is given by solving the penalized convex subproblem

$$\begin{cases} \min\limits_{(x,r)\in\mathbb{R}^{n+1}} & f_1(x) - [f_2(x^k) + \langle s^k_{f_2}, x - x^k\rangle] + \tau^k r \\ \text{s.t.} & c_1(x) - [c_2(x^k) + \langle s^k_{c_2}, x - x^k\rangle] \leq r \\ & x \in X,\, r \geq 0\,, \end{cases}$$

using the same convex bundle method employed by `SCA`. Here, $\tau^k > 0$ is a penalization parameter initialized as $\tau^0 = 1$ and updated by the rule $\tau^{k+1} \leftarrow \min\{1.5\tau^k, 10^4\}$ whenever the slack variable r^{k+1} in the solution of the above problem satisfies $r^{k+1} > \delta_{\text{Tol}} := 10^{-6}$. If $r^{k+1} \leq \delta_{\text{Tol}}$ then $\tau^{k+1} \leftarrow \tau^k$.

`BD` is a specialized approach for chance-constrained programs proposed in [26]. Binary variables are introduced for each scenario ξ^j. The resulting integer nonlinear program is then successively relaxed as continuous, regularized, then solved iteratively for each regularization parameter value, by Benders decomposition, as a two-stage problem after deporting the scenario variables into a slave linear program. Differently from the previous solvers, this approach does not exploit the DoC structure of the problem. In our tests, we used the Matlab code made available by the authors at `staff.utia.cas.cz/adam/research.html` with the default parameters.

We employed the same stopping test $\left\|x^{k+1} - x^k\right\| \leq \sqrt{n}\,10^{-6}$ for the bundle method in all 4 solvers `PBMDC`[2], `PBMDC2S`, `SCA`, and `PCCP`.

The numeric results on the 4 benchmark sets are provided in Table 4.1. Each problem is defined by the dimension n, the confidence level p and a set-specific parameter Π described hereafter. Column f^* reports the optimal value of the problem, when known, and columns $f(\bar{x})$ the function value found by each solver within a CPU time limit of 900s (8h for GAS). The column "Probability" represent the actual probability value $\mathbb{P}\left[\texttt{lb} \leq c(x,\xi) \leq \texttt{ub}\right]$ at $\bar{x}$, as computed by the Matlab statistical function `mvncdf`.

For the largest benchmark sets, we also depict the performance profiles of the solvers as defined in [91] for two metrics: objective values and runtimes. Given a metric g (the smaller, the better), scores $g(s,i) \geq 0$ for each solver $s \in S$ on each instance $i \in I$, and $g^*(i)$ the best known score on instance $i \in I$ (which is set to $\min_{s\in S} g(s,i)$ by default), then the performance profile of solver $s \in S$ is defined as the distribution curve $C_s(\tau) = card(\{i \in I : g(s,i) \leq \tau g^*(i)\})$, for ratio $\tau \geq 1$. For a quick comparison, the higher is the curve, the better is the solver. Note that in the case of the minimization of a negative function $f < 0$, we consider the objective value metric defined by $g(s,i) = -1/f(\bar{x})$ where $\bar{x}$ is the point computed by solver s on instance i, instead of $g(s,i) = f(\bar{x})$ when $f \geq 0$.

PLANTOY: AN ACADEMIC CHANCE-CONSTRAINED PLANNING PROBLEM

We first consider an example from [228] of a small management problem consisting of planing two fictitious refineries for producing two types of fuel to meet a demand that is deterministic in the first month of planning, but uncertain in the second one. The goal is to decide, at the minimal cost, the amount of processed oil by the two refineries (x_1

and x_2 for the first month, and x_5 and x_6 for the second one), storing oil (x_3 and x_7), and importation (x_4 and x_8). The storage and importation decisions must ensure that the second-month demand is satisfied with probability of at least p. The fictitious planning problem PlanToy reads as

$$\begin{cases} \min\limits_{x \geq 0} & 2x_1 + 3x_2 + 0.5x_3 + 12x_4 + 2x_5 + 3x_6 + 0.5x_7 + 12.5x_8 \\ \text{s.t.} & 2x_1 + 6x_2 = 190 \\ & 3x_1 + 2.8x_2 = 168 \\ & x_1 + x_2 + x_3 - x_4 = 60 \\ & -x_3 + x_5 + x_6 + x_7 - x_8 = 47.16 \\ & \mathbb{P}[c(x, \xi) \leq 0] \geq p \\ & x_3 \leq 10\,, x_7 \leq 10\,, \end{cases} \tag{4.20}$$

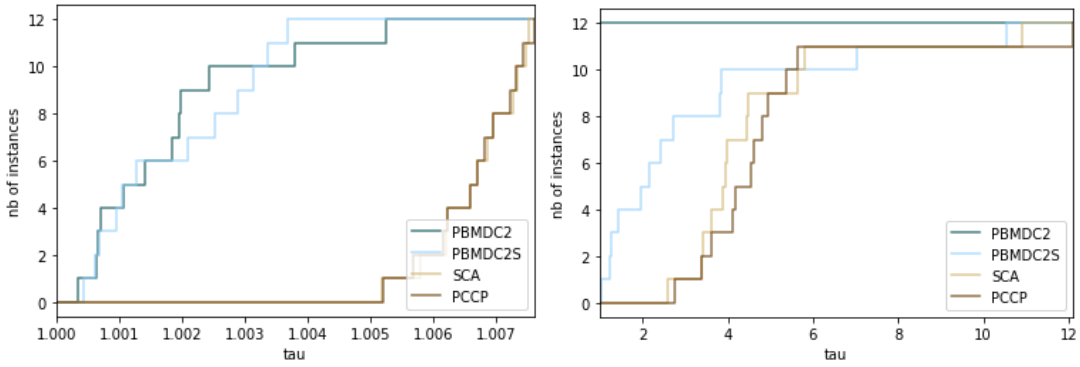
where $c(x, \xi) = \max\{\xi_1 - 2x_5 - 6x_6,\ \xi_2 - 3x_5 - 2.8x_6\}$, and $\xi = (\xi_1, \xi_2)$ is a random vector (of fuel demand) following a normal distribution with mean $\mathbb{E}[\xi] = (193, 178)$ and covariance matrix

$$C = \begin{pmatrix} 9 & \mathtt{Cov}(\xi_1, \xi_2) \\ \mathtt{Cov}(\xi_2, \xi_1) & 10.24 \end{pmatrix}.$$

We created 12 instances of PlanToy by varying the confidence level:

$$p \in \{0.8,\ 0.85,\ 0.9,\ 0.95\},$$

and the covariance coefficient $\Pi = \mathtt{Cov}(\xi_1, \xi_2) \in \{-4.8,\ 0,\ 4.8\}$. The optimal value of each instance is known (see [228] for details). As starting points for the solvers, we have considered the solution of the simpler individual chance-constrained program, as described in [228].



(a) objective $f(\bar{x})$ vs. optimum f^*, mean opt = 428

(b) CPU times, mean best = 31s

FIGURE 4.4: Performance profiles of 4 solvers on 12 PlanToy instances. For example, (a) reads: *PBMDC^2 solves 9 instances out 12 within an error inferior to* 0.2%; (b) reads: *PCCP was more than 4 times slower than the quickest solver (here PBMDC^2) on every instances but 3.*

Since the function $c(x, \xi)$ is nonsmooth, our solver PBMDC^2 is not ensured to compute a KKT point of the problem. However, as one can see in Figure 4.4a, PBMDC^2 was successful

in computing approximated global solutions for all these instances within errors less than 0.6% (0.2% in average). The smoothed variant PBMDC2S has a maximum error inferior to 0.4% (0.2% in average), and the SCA and PCPP also provided good results with errors inferior to 0.8% (0.6% in average) but with higher CPU times (see Figure 4.4b): they were respectively 3x, 5x and 5x slower than PBMDC^2 in average. See Table 4.1 below for more details.

Norm optimization problem with chance-constraint

We now consider the chance-constrained problem from [153, § 5.1]:

$$\begin{cases} \max\limits_{x \in \mathbb{R}^n_+} & \|x\|_1 \\ \text{s.t} & \mathbb{P}[\|\xi\, x\|_2 \leq 10] \quad \geq p\,. \end{cases} \tag{4.21}$$

which, using the non-negativity constraints, can be written as a minimization problem of function $x \mapsto -\sum_{j=1}^n x_j$. This problem fits (5.11) with:

$$c_2(x, \xi) = \max_{i=1,\ldots,10} \{\sum_{j=1}^n \xi_{ij}^2 x_j^2 - 100\},$$

and

$$c_1(x, \xi) \equiv 0,$$

where ξ_{ij} ($i = 1, \ldots, 10$ and $j = 1, \ldots, n$) are the random variables. As discussed in [153], when ξ_{ij} are independent and identically distributed standard normal random variables the global solution and optimal value of (4.21) are, respectively,

$$x_i^* = \frac{10}{\sqrt{F^{-1}_{\chi^2_n}(p^{1/10})}}, \quad i = 1, \ldots, n, \quad \text{and} \quad f^* = \frac{10\,n}{\sqrt{F^{-1}_{\chi^2_n}(p^{1/10})}},$$

where $F^{-1}_{\chi^2_n}$ denotes the inverse distribution function of a Chi-square distribution with n degrees of freedom. When the scenarios are correlated the optimal value is unknown.

We tested a set of 8 instances, that we denote Pnorm, with dimension $n = 30$, by considering 4 confidence levels $p \in \{0.8, 0.85, 0.9, 0.95\}$ and either independent $\Pi = I$ or correlated $\Pi = C$ scenarios. Results are reported in Table 4.1 and depicted in Figure 4.5.

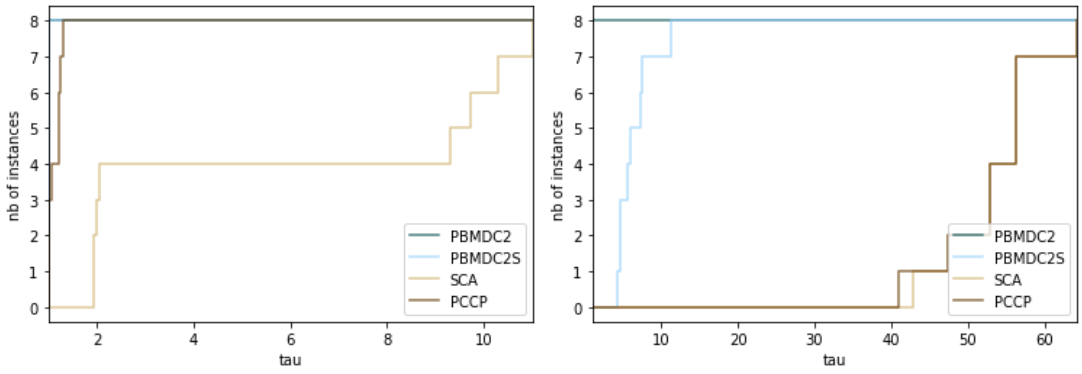
The solvers PBMDC^2 and PBMDC2S succeeded in computing approximately global solutions for the 4 independent cases within errors of 2.5%. They also obtained the best results on the 4 correlated cases. The solver PCCP performed well too, but reached the maximum CPU time of 900 seconds in 7 instances out of 8. In comparison, the average CPU times of PBMDC^2 and PBMDC2S on the 8 instances were respectively 17 and 109 seconds.

Probust: A hybrid chance-constrained/robust model

Consider the problem of minimizing a linear function subject to linear constraints

$$Ax \leq \xi\,, \tag{4.22}$$

where both ξ and A are subject to uncertainty. As pointed out in [5], this setting is of interest for instance in energy management, where x represents an energy production



(a) objective $-1/f(\bar{x})$, mean best = 0.08 (b) CPU times, mean best = 17s

FIGURE 4.5: Performance profiles of 4 solvers on the 8 PNORM instances.

schedule, and (4.22) means that one wishes to produce sufficient energy in all but the most extreme and implausible scenarios. Knowledge of the distribution of ξ (energy demand) is available, since its characterization has received considerable attention, while A is related to the underlying physics of generation plants and/or to the behavior of other generation companies, and much less information is available.

Let $A = [a_i]_{i\in I}$; following [5] we will assume that the uncertainty about the coefficients matrix can be expressed in the form $a_i(u) = \bar{a}_i + P_i u$, where $\bar{a}_i \in \mathbb{R}^n$, P_i is an $n \times n_i$ matrix, and the *uncertainty set* $u \in \mathcal{U}_i = \{u \in \mathbb{R}^{n_i} : \|u\| \leq \kappa_i\}$ is the ball of radius κ_i in the ℓ_2 norm. For the sake of notation we define $\mathcal{U} = [\mathcal{U}_i]_{i\in I}$, and we write $A(u)$ for $u \in \mathcal{U}$ to mean $[a_i(u_i)]_{i\in I}$, where $u_i \in \mathcal{U}_i$. On the other hand, $\xi \in \mathbb{R}^m$ is a random variable with known distribution, in our setting represented by $N = 10^4$ scenarios equally probable. We can then express our requirement under the form of the *probabilistic-robust* (probust) constraint:

$$\mathbb{P}\left[\, A(u)x \leq \xi \quad \forall u \in \mathcal{U} \,\right] \geq p\,. \tag{4.23}$$

These are joint-probust constraints, which in general represents an additional source of difficulty due to its robust structure, as they have to hold jointly for all $u \in \mathcal{U}$. For this particular structure, it was shown in [5] that (4.23) reduces to:

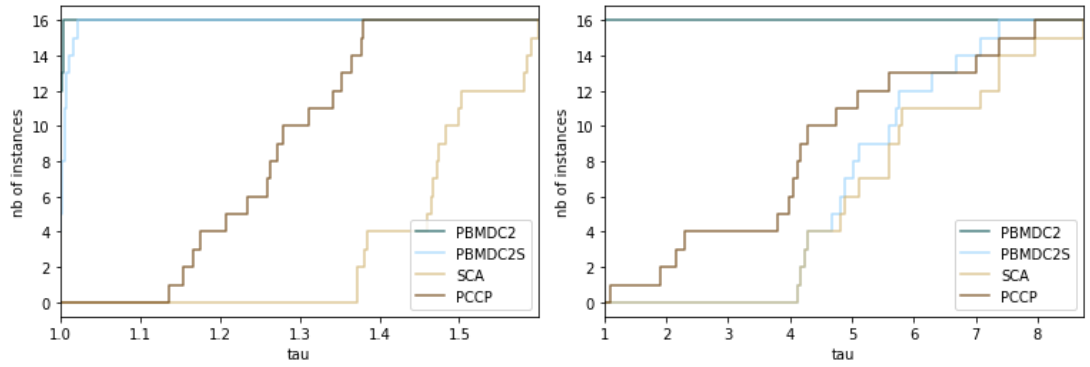
$$\mathbb{P}\left[\bar{a}_i^\mathsf{T} x + \kappa_i \left\|P_i^\mathsf{T} x\right\| \leq \xi_i \quad i \in I\right] \geq p,$$

which readily falls in the setting (5.11) with $c_2(x,\xi) = \max_{i\in I}\{\bar{a}_i^\mathsf{T} x + \kappa_i \left\|P_i^\mathsf{T} x\right\| - \xi_i\}$ and $c_1(x,\xi) \equiv 0$. The problem we are willing to solve is thus

$$\min_{x\in\mathbb{R}^n} \langle c, x\rangle \quad \text{s.t.} \quad \mathbb{P}\left[c_2(x,\xi) \leq 0\right] \geq p\,.$$

We generated 16 instances of PROBUST as described in [5], with $n = 30$, $\Pi = |I| \in \{10, 15, 20, 30\}$ and $p \in \{0.8, 0.85, 0.9, 0.95\}$.

Once again, `PBMDC` [2] was systematically the best and the quickest of the 4 solvers on these instances with an average CPU time of 164 seconds. `PBMDC2S` gave similar results but it was always more than 4 times slower and reached the maximum CPU time of 900 seconds on 10 of the 16 instances.



(a) objective $-1/f(\bar{x})$, mean best = 0.086 (b) CPU times, mean best = 164s

FIGURE 4.6: Performance profiles of 4 solvers on the 16 PROBUST instances.

CHANCE-CONSTRAINED GAS NETWORK PROBLEM

In this subsection we deal with a gas network problem from [137], whose instances are the ones considered in [26]. In the network there are n exit nodes $V = \{1, \dots, n\}$ with random exit loads $\xi = (\xi_1, \dots, \xi_n)$; one injection point corresponding to the root node (node 0); a set of pipes (directed edges) E and coefficient of the pressure drop in Φ_e, for $e \in E$ with $e = (i, j) \subset V \times V$. In a general setting, there are lower and upper pressure bounds $p_i^{\min}$ and $p_i^{\max}$ to be decided. If the minimal capacities $p_i^{\min}$ are considered fixed, then the problem reads as

$$\min_{p^{\max} \geq p^{\min}} \langle c, p^{\max} \rangle \quad \text{s.t.} \quad \mathbb{P}\left[(p_0^{\max})^2 \geq v_0(\xi), \dots, (p_n^{\max})^2 \geq v_n(\xi)\right] \geq p\,,$$

where $p^{\max}$ is the vector of decision variables; $v_0(\xi) = \max_{i=0,\dots,n}\{(p_i^{\min})^2 + h_i(\xi)\}$, $v_i(\xi) = v_0(\xi) - h_i(\xi)$ for $i = 1, \dots, n$; and $h_i(\xi)$ is a function of the random vector, pressure drop Φ_e, nodes and edges of the network [26, Eq. (21)]. Note that

$$\mathbb{P}\left[(p_0^{\max})^2 \geq v_0(\xi), \dots, (p_n^{\max})^2 \geq v_n(\xi)\right] = \mathbb{P}\left[\max_{i=0,\dots,n}\{v_i(\xi) - (p_i^{\max})^2\} \leq 0\right],$$

and thus:

$$\mathbb{P}\left[(p_0^{\max})^2 \geq v_0(\xi), \dots, (p_n^{\max})^2 \geq v_n(\xi)\right] =$$
$$\mathbb{P}\left[\max_{i=0,\dots,n}\{(p_i^{\max})^2 - v_i(\xi)\} - \sum_{i=0}^{n}[(p_i^{\max})^2 - v_i(\xi)]\right],$$

i.e., the function under the probability has an explicit DoC decomposition. In what follows we consider the two networks of [26]: the smaller network has $|V| = 4$ nodes, while the larger has $|V| = 12$. As in [26], we set $p = 0.85$ and employed $N = 10^4$ scenarios of $v(\xi)$ available at `staff.utia.cas.cz/adam/research.html` to approximate the probability function by (5.13). The GAS entries of Table 4.1 reports the results obtained with our two solvers `PBMDC`^{2} and `PBMDC2S` based on the DoC approximations of the probability constraints, and the results obtained when running the Matlab code of `BD` on the same machine.

Solver `BD` did not solve the larger instance in 8 hours of processing. The function value reported in [26, Table 3] is 3145.47. The solvers `PBMDC`^{2} and `PBMDC2S` were significantly

faster than BD for these instances of the chance-constrained gas network problem, and yet the estimated optimal values are very close to the ones computed by BD; see Table 4.1.

Table 4.1: Numerical results. The probability function (5.13) was estimated with $N = 10^4$ scenarios. CPU time limit is set to 900s for PlanToy, Norm and Probust problems, and 8 hours for Gas problem.

	Problem data				$f(\bar{x})$				Probability				CPU (s)				Iterations			
	Π	n	p	f^*	PBMDC[2]	PBMDC2S	SCA	PCCP	PBMDC[2]	PBMDC2S	SCA	PCCP	PBMDC[2]	PBMDC2S	SCA	PCCP	PBMDC[2]	PBMDC2S	SCA	PCCP
PlanToy	-4.8	8	0.95	434.81	435.67	435.91	437.33	437.28	0.948	0.949	0.948	0.947	22	84	87	106	165	92	32	32
	-4.8	8	0.90	429.30	430.09	430.88	432.24	432.22	0.899	0.900	0.900	0.900	33	90	119	138	236	99	42	43
	-4.8	8	0.85	425.61	426.21	426.50	428.70	428.68	0.850	0.850	0.852	0.851	39	50	134	141	279	55	48	48
	-4.8	8	0.80	422.69	422.99	422.87	425.85	425.83	0.801	0.800	0.803	0.803	55	57	143	151	403	62	50	51
	0	8	0.95	434.79	436.44	436.15	437.48	437.46	0.949	0.950	0.949	0.949	21	81	94	104	158	83	31	32
	0	8	0.90	429.25	431.50	429.51	432.23	432.23	0.903	0.903	0.902	0.901	23	50	130	130	176	49	43	42
	0	8	0.85	425.53	425.81	425.97	428.65	428.64	0.854	0.852	0.853	0.853	40	57	136	135	301	59	46	46
	0	8	0.80	422.58	422.72	423.12	425.75	425.79	0.802	0.805	0.804	0.804	35	85	156	159	258	92	55	52
	4.8	8	0.95	434.57	434.85	434.87	436.82	436.83	0.950	0.951	0.948	0.948	11	116	120	133	86	124	41	39
	4.8	8	0.90	428.94	429.98	430.18	431.81	431.81	0.903	0.903	0.904	0.904	22	155	128	118	164	168	45	41
	4.8	8	0.85	425.14	425.97	426.57	427.94	427.94	0.854	0.854	0.854	0.854	36	44	140	148	273	47	49	50
	4.8	8	0.80	422.13	422.58	422.53	424.75	424.75	0.806	0.805	0.801	0.801	37	72	147	171	258	75	53	56
Norm	I	30	0.95	-40.984	-40.869	-40.871	-3.719	-40.286	0.950	0.950	1.000	0.966	21	153	900	862	110	122	16	137
	I	30	0.90	-42.133	-41.835	-42.009	-4.100	-41.625	0.900	0.900	1.000	0.920	17	191	900	900	90	142	16	147
	I	30	0.85	-42.904	-42.834	-42.834	-4.412	-42.138	0.850	0.850	1.000	0.894	16	121	900	900	86	91	16	112
	I	30	0.80	-43.514	-43.423	-43.423	-4.672	-40.658	0.800	0.800	1.000	0.955	17	96	900	900	88	91	16	46
	C	30	0.95	-	-6.568	-6.567	-3.392	-5.355	0.950	0.950	1.000	0.993	16	75	900	900	82	75	17	146
	C	30	0.90	-	-6.844	-6.843	-3.509	-5.633	0.900	0.900	1.000	0.987	19	80	900	900	98	81	20	102
	C	30	0.85	-	-7.247	-7.248	-3.628	-5.812	0.850	0.850	1.000	0.982	16	75	900	900	82	73	23	91
	C	30	0.80	-	-7.773	-7.780	-3.752	-5.966	0.800	0.800	1.000	0.977	14	83	900	900	72	81	26	87
Probust	10	30	0.95	-	-11.470	-11.508	-7.197	-8.429	0.950	0.950	0.992	0.947	103	648	900	417	686	674	20	24
	10	30	0.90	-	-11.902	-11.882	-7.527	-9.857	0.900	0.900	0.979	0.923	127	726	900	890	822	758	20	145
	10	30	0.85	-	-12.282	-12.324	-7.779	-9.757	0.850	0.850	0.975	0.890	113	799	900	900	741	828	20	66

10	30	0.80	-	-12.635	-12.680	-7.971	-9.201	0.800	0.800	0.967	0.912	122	817	900	900	805	828	20	28
15	30	0.95	-	-13.075	-13.005	-8.723	-9.669	0.950	0.950	0.988	0.941	184	900	900	351	1179	693	19	25
15	30	0.90	-	-13.236	-13.085	-8.978	-11.345	0.900	0.907	0.971	0.918	210	900	900	799	1313	690	19	145
15	30	0.85	-	-13.438	-13.220	-9.160	-11.426	0.850	0.861	0.962	0.875	212	900	900	842	1351	689	19	98
15	30	0.80	-	-13.614	-13.322	-9.325	-10.710	0.800	0.803	0.955	0.892	216	900	900	900	1399	687	19	36
20	30	0.95	-	-10.362	-10.300	-6.894	-7.510	0.950	0.950	0.987	0.953	122	900	900	281	762	690	19	16
20	30	0.90	-	-10.552	-10.484	-7.114	-9.288	0.900	0.905	0.975	0.922	176	900	900	895	1102	694	19	144
20	30	0.85	-	-10.691	-10.628	-7.264	-9.260	0.850	0.853	0.967	0.878	187	900	900	888	1179	701	19	90
20	30	0.80	-	-10.849	-10.770	-7.402	-8.608	0.800	0.808	0.963	0.896	218	900	900	900	1369	699	19	39
30	30	0.95	-	-10.737	-10.733	-7.820	-8.001	0.950	0.950	0.979	0.952	161	753	900	176	980	732	18	18
30	30	0.90	-	-10.898	-10.877	-7.940	-8.306	0.900	0.901	0.962	0.931	156	900	900	338	946	886	18	39
30	30	0.85	-	-11.076	-11.091	-8.035	-8.667	0.850	0.850	0.951	0.877	155	777	900	664	943	762	18	58
30	30	0.80	-	-11.244	-11.215	-8.116	-9.115	0.800	0.803	0.940	0.840	161	900	900	900	976	901	18	67

Gas	Π	n	p	f^*	PBMDC^2	PBMDC2S	BD	PBMDC^2	PBMDC2S	BD	PBMDC^2	PBMDC2S	BD	PBMDC^2	PBMDC2S	BD
		4	0.85	-	739.12	738.60	738.05	0.850	0.850	0.850	299	351	2422	369	220	646
		12	0.85	-	3144.81	3142.42	3132.27	0.850	0.850	0.847	1066	1492	8h	720	637	-

4.5 CONCLUSION

We have presented in this chapter a novel convergent algorithm for DoC constrained DoC problems (also called *generalized DoC problems* in some woks). Our algorithm, based on the improvement function and existing algorithms, has the property of converging to d-stationay or B-stationary points, which are the sharpest optimality concepts in DoC programming. Numerical experiments show that they indeed are efficient for academic problems. This is a first necessary step in order to tackle the OPF under uncertainties.

The next chapter will develop a missing part for our attempt to tackle an OPF under uncertainties, namely the modelling of uncertainties. Based on our introductory elements on DoC programming, where the universality of DoC functions is presented, there are hopes that probabilistic functions can be DoC/be approximated by DoC functions.

Part III

Chance-constraints

INTRODUCTION

Before proceeding to the application of PBMDC^2 to an electricity distribution problem, some exertion is still needed on the *uncertainties* aspect of our work which has not yet been rigorously addressed. Broadly speaking, uncertainties have to be understood as a cause of breach between decisions *now* and determined consequences *later*. "Taking into account uncertainties" in a decision-making process has become a somewhat ubiquitous concept: thousands of research articles (literally) have both words "optimisation" and "uncertainty" in their titles. This ubiquity comes from two concurring observations:

1. real-life problems are invariably subject to uncertain data (independently of the application domain, this is inherent to implementation being prone to errors, or measurement tools and techniques),

2. it is possible that a "small" change in the uncertain data leads to significant sub-optimality/non applicability of a previously taken decision.

As a consequence, a decision-maker is naturally dragged into the subject of uncertainties. Actual implementations can be on the simpler side (e.g. computing a sequence of deterministic problems to determine a notion of *sensitivity* of the solution depending on input data) or on a more detailed one in particular when uncertainties follow a known probability distribution (e.g. stochastic programming applied to energy or natural resources management are good examples of applied domains where one can encounter more refined approaches of uncertainties). This already vast variety of techniques that fold under "optimisation that account for uncertainties" can grow even bigger depending on what qualifies as an "optimisation technique". To clarify on such techniques, we emphasize we are interested in optimisation methodologies that, for a given optimization problem, provide a point and an associated *certificate of quality*. Such a certificate stems from convergence analyses, and takes the form of *global/local optimality*, or weaker definitions of *criticality/stationarity* as detailed in Part II. Following this definition, *heuristic* methods (such as simulated annealing, genetic algorithms, tabu searches) are out of the scope of this work.

Now from this optimisation standpoint, we are faced with the following type of problem for which we ought to give a meaning:

$$\begin{aligned} \min_{x \in X} \quad & f(x, \xi) \\ & c(x, \xi) \leq 0 \end{aligned} \tag{4.24}$$

with X an abstract appropriate set, functions f, c respectively are real-valued and valued in a real vector space, and ξ is a parameter. Assuming that ξ is deterministic, (4.24) is a deterministic problem and the subsequent steps are to classify it along the usual criteria *linear/non-linear*, *convex/nonconvex*, *continuous/integers*, *smooth/nonsmooth*. Now, assuming that ξ is a random vector, (4.24) is meaningless as it is, and frameworks for optimisation under uncertainties are precisely aimed at providing a meaning. As pointed out, a first technique could be to approximate the random variable by a single deterministic value,

say by $E[\xi]$. The initial problem (4.24) thus becomes:

$$\begin{aligned} \min_{x \in X} \quad & f(x, E[\xi]) \\ & c(x, E[\xi]) \leq 0 \end{aligned} \tag{4.25}$$

This model (sometimes referred to as the "deterministic counterpart") is not satisfactory in general, mainly due to the fact it does not account for the variance/extremal values of the random variables which makes its solutions not robust, i.e., prone to sub-optimality gaps in case the uncertainty realisation is not equal to its mean - see for instance the example in [289, Section 1.1.3].

One can broadly divide optimisation problems with uncertainties according to which step, of the *decision* or the *observation* of uncertainty, comes first. In case the latter comes first, i.e. the decision can be postponed to after the observation of uncertainty, solving such a problem amounts to solving a set of deterministic optimisation. The question at hand is to find:

$$\begin{aligned} \widehat{f}(\xi) = \min_{x \in X} \quad & f(x, \xi) \\ & c(x, \xi) \leq 0. \end{aligned} \tag{4.26}$$

The forthcoming challenges lie in the study of the optimal value $\widehat{f}$ as a function of the uncertainty parameter ξ. This setting is referred to *sensitivity analysis*, and is one way of taking uncertainties into account as described in [48].

When the decision, or at least one of the decisions, has to be taken prior to observing uncertainty, there are two usual methodologies, each of them with different assumptions and complementary objectives. We here provide broad descriptions of them, and refer to [48, 252] for more details. The shared key elements between these methodologies is that they provide a setting where we have tools to properly define *feasibility* and *optimality* properties for given decision vectors.

- The first methodology is Robust Optimisation: here knowledge of the uncertainty set $\Xi \ni \xi$ is assumed, but more precise information on the "randomness" of ξ is not necessary. As described in the thorough and didactic reference [48], in its simplest form this paradigm associates an uncertain optimisation problem with a *robust counterpart*. This association is usually done using $\min - \max$ operations over the set of uncertainties Ξ aiming at obtaining a computationally tractable, and this time meaningful, problem. From a modelling/interpretation standpoint, the core principle of this paradigm is *safety*: this decision criteria then makes it possible to give a meaning to the uncertainties and how to deal with them. Often viewed as "the worst-case scenario approach", this optimisation framework provides complementary modelling and solving methods to the stochastic paradigm. A very general mathematical formulation of a robust counterpart in this framework could be the following:

$$\min_{x,t \in X \times \mathbb{R}} \{t : f(x, \xi) \leq t,\ c(x, \xi) \leq 0\ \forall \xi \in \Xi\} \tag{4.27}$$

  Solving an optimisation problem within the Robust Optimisation framework can be broadly summarized, at a "first order approximation", as providing to the user a solution to (4.27).

- The second one is Stochastic Optimisation: here the main assumption (which is not necessary for the other approaches) is the *a priori* knowledge of the "type of randomness". This amounts to know/define probability laws for the random variables, which then makes it possible to develop mathematical programming models. Chance-constraints constitute as subfield of stochastic optimisation, just as the expectation violation penalty approach where the expected value of violation is minimized. Recent works are looking into "bridging" the gap between a robust and a stochastic approach: a first example is the *hybrid robust/chance-constrained* approach of [5]. A second example can be found in [101, 100] where the authors propose a framework allowing only partial knowledge of the class of probability distribution. This second example, which is often referred to as *ambiguous extension* to a stochastic setting wherein the probabilistic model is itself unknown, is further developed in the didactic book [238].

Generally speaking, Stochastic optimisation can be qualified as "less-conservative" than the Robust framework; one should keep an acute mind and critical opinion on framework comparisons. The paradigm selection is a multi-dimensional decision: it is dependent on the available data, prior knowledge of the problem, available solvers / solving techniques, question to answer, own position on uncertainties, type of constraints and type of guarantee expected (is the guarantee expected to be robust or probabilistic?).

We have chosen to look into the stochastic setting, and started more particularly to study CC (Chance-Constraint). We will discuss the rationale and interest of such a modelling tool, before moving on technicalities. In this second phase, we will discuss the eventual DoC structure of chance-constraints, and provide DoC approximations. We will conclude with a more in-depth analysis of these constraints to derive a characterization of their subdifferentials in a more general case than what is encountered in literature.

RÉSUMÉ EN FRANÇAIS

Avant de passer à l'application de PBMDC[2] à un problème de distribution d'électricité, il est nécessaire de travailler sur l'aspect *incertain* de notre travail, lequel n'a pas encore été abordé de manière rigoureuse. De manière générale, les incertitudes doivent être comprises comme une rupture entre les décisions *maintenant* (*now*) et les conséquences déterminées *plus tard* (*then*). La "prise en compte des incertitudes" dans un processus décisionnel est devenue un concept quasi-omniprésent : il existe (littéralement) des milliers d'articles de recherche dont le titre comprend les mots "optimisation" et "incertitude" dans leur titre. Cette quasi-omniprésence provient notamment des deux observations suivantes :

1. les problèmes de la vie réelle sont invariablement soumis à des données incertaines (indépendamment du domaine d'application, ceci est inhérent à la mise en œuvre sujette à des erreurs, ou aux outils et techniques de mesure),

2. il est possible qu'un "petit" changement dans les données incertaines conduise à une sous-optimalité significative/non applicabilité d'une décision prise précédemment.

En conséquence, le sujet des incertitudes s'impose naturellement à un décideur ou une décideuse. Les implémentations réelles peuvent être plus simples (par exemple, le traitement des incertitudes peut se résumer au calcul d'une séquence de problèmes déterministes pour déterminer une notion de *sensibilité* de la solution en fonction des données d'entrée) ou plus détaillées, en particulier lorsque les incertitudes suivent une distribution de probabilité connue (par exemple, la programmation stochastique appliquée à la gestion de l'énergie ou des ressources naturelles est un bon exemple de domaines appliqués où l'on peut rencontrer des approches plus raffinées des incertitudes). Ce vaste ensemble d'optimisation "prenant en compte les incertitudes" s'accroît encore dépendemment de ce que l'on qualifie de "technique d'optimisation". Pour clarifier ce dernier point, nous soulignons que nous nous intéressons aux méthodologies d'optimisation qui, pour un problème d'optimisation donné, fournissent un point et un *certificat d'optimalité* associé. Un tel certificat provient d'analyses de convergence, et prend la forme de *optimalité globale/optimalité locale*, ou des définitions plus faibles de *criticalité/stationnarité* comme détaillé dans la partie II. Suivant cette définition, les méthodes *heuristiques* (telles que le recuit simulé, les algorithmes génétiques, la recherche tabou) ne sont pas considérées dans ce travail.

D'un point de vue générique, nous sommes confrontés au type de problème suivant pour lequel nous devrions donner un sens :

$$\begin{aligned} \min_{x \in X} \quad & f(x, \xi) \\ & c(x, \xi) \leq 0 \end{aligned} \tag{4.28}$$

avec X un ensemble approprié, les fonctions f, c sont à valeur réelle et évaluées dans un espace vectoriel réel, et ξ est un paramètre. En supposant que ξ est déterministe, (4.28) est

un problème déterministe. Pour le résoudre, il s'agit premièrement de classifier le problème selon les critères habituels suivants : *linéaire/non-linéaire*, *convexe/non-convexe*, *continu/intégral*, *lissé/non-lissé*. A présent, en supposant que ξ est un vecteur aléatoire, (4.28) n'a pas de sens en l'état : il n'est pas possible d'interpréter ce problème tel quel. Les cadres d'optimisation sous incertitudes visent précisément à fournir un sens à (4.28). Comme souligné, une première technique pourrait consister à approximer la variable aléatoire par une seule valeur déterministe, disons par exemple $E[\xi]$. Le problème initial (4.28) devient alors :

$$\begin{aligned} \min_{x \in X} \quad & f(x, E[\xi]) \\ & c(x, E[\xi]) \leq 0 \end{aligned} \tag{4.29}$$

Ce modèle (parfois appelé "contrepartie déterministe" ou *deterministic counterpart*) n'est pas satisfaisant en générale, principalement parce qu'il ne tient pas compte de la variance ni des valeurs extrêmes des variables aléatoires. Ceci rend ses solutions non robustes, c'est-à-dire sujettes à des écarts de sous-optimalité dans le cas où la réalisation de l'incertitude n'est pas égale à sa moyenne - voir par exemple l'exemple dans [289, Section 1.1.3].

On peut globalement diviser les problèmes d'optimisation avec incertitudes selon quelle étape vient en premier, à savoir la *décision* ou l'*observation* de l'incertitude. Dans le cas où cette dernière vient en premier, i.e. la décision peut être reportée après l'observation de l'incertitude, la résolution d'un tel problème revient à résoudre un ensemble de problèmes d'optimisations déterministes. Il s'agit alors de trouver :

$$\begin{aligned} \widehat{f}(\xi) = \min_{x \in X} \quad & f(x, \xi) \\ & c(x, \xi) \leq 0, \end{aligned} \tag{4.30}$$

où ξ prend la valeur de l'observation. Les défis dans ce cas sont dans l'étude de la valeur optimale $\widehat{f}$ en fonction du paramètre d'incertitude ξ. Ce paramètre est appelé *analyse de sensibilité*, et constitue une façon de prendre en compte les incertitudes comme décrit dans [48].

Lorsque la décision, ou au moins l'une des décisions, doit être prise avant l'observation de l'incertitude, il existe deux méthodologies principales, chacune d'entre elles comportant des hypothèses différentes et des objectifs complémentaires. Nous en donnons ici une description générale, et nous renvoyons à [48, 252] pour plus de détails. L'élément clé de ces méthodologies est qu'elles fournissent un cadre dans lequel nous disposons d'outils pour définir correctement les propriétés de *faisabilité* et d'*optimalité* pour des décisions données.

- La première méthodologie est l'optimisation robuste : ici, nous supposons connaître l'ensemble $\Xi \ni \xi$, sans avoir besoin d'informations plus précises sur le caractère "aléatoire" de ξ. Comme décrit de façon approfondie dans la référence [48], dans sa forme la plus simple ce paradigme associe un problème d'optimisation incertain à un problème *robuste*. Cette association est généralement réalisée à l'aide d'opérations $\min - \max$ sur l'ensemble des incertitudes Ξ dans le but d'obtenir un problème résolvable en temps acceptable. Souvent considéré comme "l'approche par le pire des scénarios", ce cadre d'optimisation fournit des méthodes de modélisation et de résolution complémentaires au paradigme stochastique. Une formulation

mathématique très générale d'une contrepartie robuste dans ce cadre pourrait être la suivante :

$$\min_{x,t \in X \times \mathbb{R}} \{t : f(x,\xi) \leq t, (x,\xi) \leq 0 \; \forall \xi \in \Xi\} \tag{4.31}$$

La résolution d'un problème d'optimisation dans le cadre de l'optimisation robuste peut être résumée de manière générale, à une "approximation de premier ordre", comme fournissant à l'utilisateur une solution à (4.31).

- La seconde est l'optimisation stochastique : ici, la principale hypothèse (qui n'est pas nécessaire pour les autres approches) est la connaissance *a priori* du "type d'aléa". Cela revient à connaître/définir des lois de probabilité pour les variables aléatoires, ce qui permet ensuite de développer des méthodes de programmation mathématique. L'optimisation avec contraintes probabilité (*chance-constraint programming*) constitue un sous-domaine de l'optimisation stochastique. Des travaux récents cherchent à "combler" le fossé entre une approche robuste et une approche stochastique : un premier exemple est l'approche *hybride robust/chance constraint* de [5]. Un deuxième exemple peut être trouvé dans [101, 100] où les auteurs proposent un cadre d'optimisation où seule une connaissance partielle de la classe de distribution de probabilité est nécessaire. Ce deuxième exemple est développé dans le livre didactique [238].

D'une manière générale, l'optimisation stochastique peut être qualifiée de "moins conservatrice" que le cadre robuste ; il convient de maintenir une opinion critique sur les comparaisons de ces deux paradigmes. Le choix du paradigme est une décision multidimensionnelle : il dépend des données disponibles, de la connaissance préalable du problème, des solveurs / techniques de résolution disponibles, de la question à laquelle il faut répondre, de sa connaissance des incertitudes à traiter, du type de contraintes et du type de garantie attendue (la garantie doit-elle être robuste ou probabiliste ?).

Nous avons choisi de nous intéresser au cadre stochastique, et plus particulièrement aux CC (Chance-Constraint). Nous discuterons de l'intérêt d'un tel outil de modélisation, avant de passer aux aspects techniques. Dans cette deuxième phase, nous discuterons de l'éventuelle structure DoC des contraintes de chance, et fournirons des approximations DoC. Nous conclurons par une analyse variationnelle de ces contraintes en étudiant leurs sous-différentiels dans un cas plus général que celui rencontré dans la littérature. Cette dernière partie est d'une portée dépassant l'application visée par ce travail.

A CONVENIENT BUT CHALLENGING MODELLING TOOL

5

5.1 NOTATIONS, ASSUMPTIONS AND TERMINOLOGY

We start by recalling some basic terminology as an attempt to maintain clarity in technicalities. We are interested in functions of the form $\mathbb{P}\left[c(x,\xi) \leq 0\right]$ which are to be defined in the following sections. Studying continuous probabilities unveiled several concepts shared but not similarly defined between some authors, or others referred to by different names.

PRELIMINARY TOOLS AND RECALLS ON PROBABILITY-RELATED MATERIAL

In order to formally define random variables, we appropriately equip ourselves with a probability space $(\Xi, \mathfrak{A}, \mathbb{P})$. Ξ is the sample space, $\mathfrak{A}$ a σ-algebra defined on Ξ and $\mathbb{P}$ a probability measure. The combination of a sample space and a σ-algebra of subsets of the sample space canonically defines a measurable space.

We recall that measures, here denoted μ, are maps defined on $(\Xi, \mathfrak{A})$ with values in $\mathbb{R}_+ \cup \{\infty\}$. They are *$\sigma$-additive*[1] (for any disjoint sets $(A_i)_{i\in[1,n]} \in \mathfrak{A}^n$, $\mu(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n \mu(A_k)$) and indeed *positive* (i.e. $\mu(A) \geq 0$, $\forall A \in \mathfrak{A}$). A measure is said to be a probability measure if moreover it verifies a *norming* assumption, namely $\mu(\Xi) = 1$. A function α between two measurable spaces $(\Xi, \mathfrak{A})$ and $(\Xi_2, \mathfrak{A}_2)$ is measurable (with respect to $\mathfrak{A}, \mathfrak{A}_2$) if: $A \in \mathfrak{A}_2 \implies \alpha^{-1}(A) \in \mathfrak{A}$.

A random vector[2] ξ is any measurable mapping defined from $(\Xi, \mathfrak{A}, \mathbb{P})$ to a measurable space $\Xi_2, \mathfrak{A}_2$. The function F defined by $F(x) = \mathbb{P}\left[\xi \leq x\right]$ is called the (cumulative) probability distribution function of ξ[3]. F can be "abusively" overloaded as F_ξ to emphasize to which random variable is related the distribution function. As a matter of fact F, as every probability distribution function by definition, is non-decreasing, right-continuous, defined on the real line and verifies $F(-\infty) = 0$, as well as the norming condition $F(\infty) = 1$. If $A \in \mathfrak{A}_2$, then $\{\xi \in A\} = \xi^{-1}A \in \mathfrak{A}$, and the following chain of equalities helps us define the *law* μ_ξ of the random variable ξ: $\mathbb{P}\left[\xi \in A\right] = \mathbb{P}\left[\xi^{-1}A\right] = (\mathbb{P} \circ \xi^{-1})A = \mu_\xi(A)$, for all $A \in \mathfrak{A}_2$. One can verify that μ_ξ is a probability measure on $(\Xi_2, \mathfrak{A}_2)$. For clarity within this terminology, in the unidimensional case the *law* and the *distribution function* of a random variable ξ are related to each other as follows: $F_\xi(x) = \mu_\xi\{(-\infty, x]\}$. From now on, for simplicity reasons, we drop the formal use of curly brackets and leave the distinction between law and distribution function to the context. A thorough and complete introduction to this formalism is to be found in [107]. Moving on other notions related to a random vector of size m, the mean of ξ is denoted $\mathfrak{m} = \mathbb{E}[\xi] = (\mathbb{E}[\xi_i])_{i\in[1,m]}$ with

1 σ-additivity is also called "complete additivity" in literature, which is synonymous.

2 ξ is usually referred to as "random variable" when $\Xi_2 = \mathbb{R}$, the term "vector" being used if $\Xi_2 = \mathbb{R}^m$ and $m > 1$.

3 The adjective *cumulative* is sometimes omitted by some authors.

by definition $\mathbb{E}[\xi_i] = \int_\Xi \xi_i(\omega) d\mathbb{P}(\omega)$, $\forall i \in [1, m]$ while its covariance matrix is denoted $\Sigma = (\mathbb{E}[(\xi_i - \mathfrak{m}_i)(\xi_j - \mathfrak{m}_j)])_{(i,j)\in[1,m]^2}$.

In an attempt to keep our arguments crisp, we here leave aside a special case that could be deemed "pathological" in our context: we assume our random variables have their expectations properly defined (i.e. the defining integrals do exist). We recall that when $F_{\tilde{\xi}}$ is absolutely continuous with respect to a chosen measure μ (e.g. the Lebesgue measure) if there exists a density function f such that $F(A) = \mathbb{P}[\mathbf{X} \in A] = \int_A f(x)\mu\{dx\}$ for all set A. On top of this definition, we recall the Radon-Nikodym theorem [107, p. 140] that is concerned with the existence of this density function:

Theorem 8 (Radon-Nikodym theorem)**.** The probability distribution $F_{\tilde{\xi}}$ is absolutely continuous with respect to μ if, and only if, $F_{\tilde{\xi}}(A) = 0$ whenever $\mu(A) = 0$, $\forall A \in \mathfrak{A}$. ◀

Additional elements on terminology

We now recall two different types of structural properties of random variables, that often constitute salient points of works in literature.

- Firstly, studies in the stochastic field can evidently be categorized according to which class is assumed for the random variables. A class defines additional properties verified by its elements, which in general give access to reformulation, or more assumptions on the probability distribution function. For instance, in our work we are interested in the class of elliptical symmetric random variables which appears to be promising (see definition 10). It provides us with a useful reformulation of this random vector into an expression of elements with stronger properties (namely a radial unidimensional probability distribution and a uniform distribution on a sphere). Other subclasses include notably the multivariate Gaussian random variable, already well studied and widespread. An influential and general class of probability distribution functions is the quasi-concave one: a complete introduction to this class can be found in [245] by the author who first introduced it.

- Secondly, another significant structural attribute in literature is on the *non-degenerate* and *singular* properties of distributions. Formally speaking, we recall the definition of *degeneracy*:

  Definition 6 (Degeneracy, see Chapter III Sections 5-6 of [107])**.** The probability distribution F is degenerate if it is concentrated on a manifold of lower dimension than the dimension of the manifold in which it is considered. •

  At the end of Chapter III, Section 5 of this reference it is proved that if the covariance matrix is positive definite, then the probability distribution function F is non-degenerate.

  Among practical consequences of non-degeneracy is (i) the random variable does not reduce to a constant (i.e. no singleton is assigned probability 1, which notably excludes the Dirac distribution), (ii) associated to these distribution functions, are density functions which explicit form can be found for instance in [245, Section 4.4] for the Gaussian distribution. This last consequence is to be related with the introduction to this chapter, and namely the Radon-Nikodym theorem. We also provide a formal definition for singular probability distributions:

Definition 7 (Singularity, see Definition V.3.3a of [107])**.** The probability distribution F is singular with respect to a measure (e.g. the Lebesgue measure) μ if it is concentrated on a set R such that $\mu(R) = 0$. •

It appears in literature that the different concepts (i) singular covariance matrix, (ii) singular probability distribution, (iii) degenerate probability distribution are interchangeable (while not explicitly defined, this is implicitly understood for instance in [146, Section 2]). To clarify these notions, we settle on an unifying proper definition from [107, Chapter III Sections 6], [147, 146]:

Definition 8. Let η be a m_2-dimensional random vector, $A \in \mathcal{M}_{m_1,m_2}(\mathbb{R})$, and $b \in \mathbb{R}^{m_1}$ such that $\xi = A\eta + b$. The element ξ is consequently a m_1-dimensional vector, which distribution is said to be degenerate in m_1 dimension if $m_1 > m_2$. If $m_1 \leq m_2$, it is non-degenerate provided that $\text{rank}(A) = m_1$. •

Again, singularity has to be understood as a special case that contradicts Theorem 8.

Remark 6. Some articles in literature interestingly discuss why singular random variables actually stand out, especially those dealing with networks where stochastic demands at some nodes are observed, as gas networks. In such problems, one "small" random non-degenerate vector ξ of size m can model uncertainties (typically m can be the number of nodes), while the constraint set can make appear the value $\eta = A\xi + b$ with a given matrix $A \in \mathcal{M}_{w,m}(\mathbb{R})$ and vector $b \in \mathbb{R}^w$. The new element η is a random variable of its own, and if A is of rank strictly lower than m (which happens *de facto* if $k < m$) then it is a singular random variable (see the introductory parts of [147, 6] for instance). It is thus clear that singularity naturally arises in stochastic problems. ▷

Notations used throughout this part

Here are some additional notation:

- In our case, the sample space Ξ_2 (in which our random variables denoted ξ are valued) is assumed to be $\mathbb{R}^m$ and the σ-algebra $\mathfrak{A}_2$ is the one of Borel sets in $\mathbb{R}^m$.
- The decision variable, usually denoted x, is in $\mathbb{R}^n$.
- We will refer to $c : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^d$ as the *operand* (the function on which the probabilistic operation is applied).
- The safety parameter is a user-given parameter p which defines a level of acceptable risk and verifies $\mathrm{p} \in [0, 1]$.
- Throughout the chapter, i is an index and $\mathcal{I}$ is a given finite index set.

5.2 PRELIMINARY ELEMENTS ON PROBABILISTIC FUNCTIONS

The first chapter outlines the interest of CC from a modelling point of view. Let φ be defined as follows:

$$\begin{aligned} \varphi \colon \mathbb{R}^n &\to [0,1] \\ x &\mapsto \mathbb{P}\left[c(x,\xi) \leq 0\right], \end{aligned} \tag{5.1}$$

where c is continuous map from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^d$, and ξ is a multivariate random vector in $\mathbb{R}^m$. The function φ is qualified as *probabilistic*. An optimisation problem is *chance-constrained* when it involves this function φ and takes as an input a user-given safety parameter denoted p. Such a constraint can consequently be written as:

$$\varphi(x) \geq p. \tag{5.2}$$

The interpretation of a decision vector $x \in \mathbb{R}^n$ verifying (5.2) is as follows: the probability of observing $c(x,\xi) \leq 0$ is greater that $p \in [0,1]$. In practise, p is often chosen close to one of its boundaries: in an electrical distribution context, one can for instance request that power transit on lines be lower that a given upper-bound with a probability set to 0.95 or 0.99. Let us briefly discuss of the special cases where $\mathrm{p} = 1$ or $\mathrm{p} = 0$: in the latter, constraint (5.2) has no impact on the feasible set of the optimisation problem at hand, and should be excluded from the optimisation model. In the former case, $\varphi(x) \geq 1$ becomes $\varphi(x) = 1$ by definition and in fact states that the user requests that a constraint be verified independently of the realisation of the random variable. It becomes visible that only the support of the random variable is significant, and the probability distribution function of ξ is irrelevant: this matches the robust framework rather than the chance-constrained one. As a consequence, we can assume $p \in (0,1)$ without loss of generality.

We also introduce the feasible set as follows:

$$M_f := \{x \in \mathbb{R}^n \mid \varphi(x) \geq \mathrm{p}\}. \tag{5.3}$$

Note that, following some authors as R. Henrion, this set M_f is parametrized by p; one could therefore rather use the notation $M_f(\mathrm{p})$ which we omit for simplicity.
Interestingly, another set worth studying (which sometimes shares the same notation in literature) is the following:

$$M(x) := \{z \in \mathbb{R}^m \mid c(x,z) \leq 0\}. \tag{5.4}$$

This set then allows us to write the chance-constraint (5.2) as $\mathbb{P}\left[\xi \in M(x)\right] \geq \mathrm{p}$. This reformulation, taking a "set-theoretic" point of view, will turn out to be particularly fruitful for the variational studies of the upcoming developments (see in particular Chapter 6).

Following the definition of the mapping c, we introduce the set of functions $(c_i)_{i\in\mathcal{I}}$ as c's coordinate-functions. A chance-constraint can be qualified as *individual* or *joint* depending on the localisation of the quantifier $\forall$:

Joint chance constraint	$\mathbb{P}\left[c_i(x,\xi) \leq 0,\ \forall i \in \mathcal{I}\right] \geq \mathrm{p}.$
Individual chance constraint	$\mathbb{P}\left[c_i(x,\xi) \leq 0\right] \geq \mathrm{p},\ \forall i \in \mathcal{I}.$

Table 5.1: Different types of chance-constraints.

5.2.1 Brief historical overview

As briefly discussed in the Introduction to this Part, uncertainties are among the most trending topics in optimisation, but a clear description on how we consider them is yet to be given. Recalling that our upcoming interest lies in *a priori* decision analysis for the electrical distribution domain, we look for decisions to be made prior to observing uncertainty while the output (i.e. the combination of our decision and the uncertainty) becomes known only *a posteriori*.

To that end, CC is a popular modelling tool albeit a challenging mathematical object to study. This is easily seen in the wide variety of active research domains, energy being among the most intensive one (see [25, 291, 22]). Usual energy subfields involve unit-commitments or economic dispatch problems references [241], electrical network planning [300, 171], wind energy management [36], water dam management [23], gas networks [137, 135, 144], while [94] studies the application of CC to energy-related markets. More generally, studies on chance-constrained network design find numerous applications as in telecommunication or power transmission [264]. Mathematical finance also is a very active application to chance-constraints, as these class of constraints relate well to the concept of Value-at-Risk (VaR). A brief standard search of applications of chance-constraints immediately yields numerous other fields [4] and we suggest that the interested reader turn to [127] for a recent monograph on chance-constrained programming.

Taking a step back, CC are usually dated back to a 1958 article [75], which is a work on individual chance-constraints for an oil management/scheduling problem where the inputs and sales of oil at each time step are assumed to be random with a know distribution function. The main contribution was the introduction of chance-constraints, while the resolution method was a tailored sequence of approximations. Propositions of chance-constrained models grew more numerous with, for example, [165] where the authors present an interesting alternative objective function to the expected value of random variables, or [205] which introduces joint chance constraints (though random variables are assumed to be independent). Key elements in these pioneering works included the reformulation (possibly with approximations, as first order Taylor's expansions of involved functions) of chance-constrained programs into non-linear but convex problems. Major contributions are then credited to Andras Prékopa, which are mostly included in the large reference [245]. Among groundbreaking works compiled in that book, let us point out some advances pushed by its author as (i) the study of joint chance-constraints with dependency (see [243]), (ii) development of *efficient points* and associated resolution methods (see [88, 188, 12]), (iii) convexity and duality properties of chance-constrained problems (see [89, 87]). A brief notwithstanding thorough overview

[4] As an illustration of the variety of domain, here is a list of applications easily encountered in literature: water quality management, air traffic flow management, blood supply chain optimisation, optimisation of chemicals processes, job scheduling, autonomous vehicle control, irrigation planning

of the subject including a discussion of critics, difficulties, theoretical cornerstones and algorithmic key-points can be found in the recent article [13].

We will briefly describe the usual structures of chance-constraints that are found in literature and leave the interested reader to [237, Chapters I-II] and the classic reference [244].

INSERT 5: A debate on chance-constrained programming framework as a mean to provide decision analysis under uncertainty.
On a historical side, chance-constrained programming interestingly led to clashes, even extending on somewhat more philosophical grounds. R. Blau presented what appeared to a "dilemma" using the chance-constraint framework in [54]: in his example it appeared that the expected value of perfect information was lower than the expected value of sample information, and moreover both are negative. This counter-intuitive result led to arguments during the following decades on the usefulness of chance-constraints as a framework that could make efficient use of available information to the decision maker: numerous suggestions were made to change the chance-constrained setting to "stochastic programming with recourse", "Linear Programming under uncertainty" or others. Central to this issue was a modelling question: is chance-constrained programming adequate to provide decision analysis in an uncertain environment? As a consequence, arguments are focussed on "expected value of information", and confront chance-constrained programming to another decision theory, namely the Bayesian Decision Theory. Following works, notably led by A. Charnes, R. Jagannathan on the one hand, A. Hogan, I. LaValle on the other hand, argued on this expected value of information, how chance-constraint programming behaves when information quality improves, and how solving a chance-constrained program could yield equivalent solution sets when compared to solving a well-chosen "bayesian fashion" problem (cited from [181]). Reference [216] provides a rather broad overview of this long debate. Nowadays there still are debates on decision theories, for instance between the stochastic and robust ones, but their relevance is not questioned and complementarity is well recognized. It is clear that these theories are tools that provide limited results, which will turn out to be wrong with probability one, and selecting one rather than the other is mostly motivated by (i) the *a priori* available information, (ii) the available resources to solve a problem, (iii) the rules expected to be followed by the output.

5.2.2 Some theoretical properties of probabilistic functions

The popularity of probabilistic constraints brought some flourishing studies on their variational and structural properties. We will summarize in this subsection the elements related to these two subjects. Far from being exhaustive, we will refer to some works that detail the technicalities while remaining on the superficial aspects. Let us stress out that without any additional assumptions, the probabilistic function φ as defined in (5.1) and its associated feasible set M_f defined in (5.3) enjoy no particular properties. We here

introduce the definition of *separable* probabilistic functions which has a significant impact of the derived properties of these functions.

Definition 9 (Separable function)**.** A function $c\colon \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^d$ is *separable* if there exists functions $\widetilde{c}$ and $\widehat{c}$ such that

$$c(x,\xi) = \widetilde{c}(x) - \widehat{c}(\xi), \quad \forall x \in \mathbb{R}^n,\ \xi \in \mathbb{R}^m. \tag{5.5}$$

Usual definition also implicitly define $\widehat{c}$ as a "nice" function, namely the identity or a linear function. •

Remark 7. Definition 9 merely states that we can separate the decision variable from the random one. It is to be related to the structural aspects of chance-constrained programming studied in the introductory work by R. Henrion [145]. This separation property is interesting: let us consider φ as defined in (5.1), and $c(x,\xi) = \xi - \widetilde{c}(x)$ for some appropriate function $\widetilde{c}$. Using the distribution function $F_\xi(a) := \mathbb{P}[\xi \leq a]$ for our random variable ξ, we may write (modulo a sign function) $\varphi(x) = F_\xi(\widetilde{c}(x))$. We thus have more tools at hand to derive some properties of the probabilistic function. ▷

Remark 8. Individual chance-constraints with usual random distributions can actually be written as deterministic problems:

$$\mathbb{P}[c_i(x) - \xi_i \leq 0] \geq \mathrm{p} \iff c_i(x) \geq \text{p-quantile of } \xi_i. \tag{5.6}$$

One might wonder what is lost when considering individual chance-constraints opposed to joint ones: this is answered in [288] where the authors compare the solutions to a water management problem with uncertainties modelled by two-sided separable chance-constraints. It turns out treating the probabilistic individually allows for simpler numerical methods but the solution is deemed less "robust". In other words, over their set of 100 generated scenarios, they looked at how many of them are actually violated at (at least) one time step: as expected, the individual chance-constraints proposed a solution that violates 68 scenarios, compared to 10 scenarios in the joint case. ▷

These properties are particularly dependent on the separability assumption. The rationale lies in the reformulation allowed when separability is verified presented in remark 7. Properties of the operand function c as well as the class of the random variable ξ are significant when studying the properties of φ.

General properties

Under the (reasonable) assumption that the family $(c_i)_{i\in\mathcal{I}}$ only contains upper-semi-continuous functions and that $\mathcal{I}$ is finite, φ is upper-semicontinuous and M_f as defined in (5.3) is closed. Continuity of the probabilistic function is not evident, and only ensured with additional assumptions. In the particular case where:

- every functions c_i is continuous,
- and $\mathbb{P}[c_i(x,\xi) = 0] = 0,\ \forall x \in \mathbb{R}^n,\ \forall i \in \mathcal{I}$,

then φ is continuous. Continuity of φ, and another somewhat related result that also ensures continuity of φ are to be found in [13, Section 2].

Example 2 (Pathological examples where continuity fails)**.** To make continuity fail in the separable case, one should produce a function where there exist a $i \in \mathcal{I}$ that does not verify $\mathbb{P}\left[c_i(x,\xi) = 0\right] = 0$, $\forall x \in \mathbb{R}^n$. A "classic" way to set up such a pathological case (see [13, Example 2.1]) is to include a deterministic coordinate-function in the operand. Let, for instance, ξ be an elliptical symmetric random variable, c be defined on $\mathbb{R} \times \mathbb{R}$ with $c(x,\xi) = x$. As such, we obtain:

$$\varphi(x) = \mathbb{P}\left[c(x,\xi) \leq 0\right] = \begin{cases} 0 & \text{if } x > 0 \\ 1 & \text{if } x \leq 0. \end{cases} \tag{5.7}$$

Observe that we have a trivial function c, linear in both arguments and separable.

Discontinuity in the nonseparable case seems to be found in more diverse cases when compared to the separable case, for which we do not know any other "simple" wit to produce the desired result. Let us define c as $c(x,\xi) = \max(\xi, x)$, and set ξ to be a Gaussian random variable, say $\xi \sim \mathcal{N}(0,1)$. We then obtain:

$$\varphi(x) = \begin{cases} 0.5 & \text{if } x \leq 0, \\ 0 & \text{if } x > 0. \end{cases} \tag{5.8}$$

φ, again, clearly is discontinuous. ▶

A take-away on continuity of probabilistic functions should be that this property can be expected in practise.

Variational properties

As expected, more assumptions are needed to study variational properties of probabilistic functions. "First-order" information is important for probabilistic functions when one wants to rely on the vast and efficient literature of non-linear programming when solving a chance-constrained optimisation problem.

Broadly speaking, one could classify the historical studies on differentiability into two somewhat different approaches. The first one makes little to no use of specific properties of ξ (as its class), and rather proposes results on differentiation that consequently are more general in terms of random variables. These methods are generally based on multiple integral of density functions, one example of which can be found [201]. Within this framework, which main results are summarized in [13, Section 2.2], it appears that the numerical implementation necessary to obtain partial derivatives of the probabilistic function, let alone verify that all required assumptions are met, is not generally accessible (see [245, Section 6.6.4], and more recent works [14, 13]). Moreover it appears that compactness of the set $M(x)$ (see (5.4)) is often assumed, which turns out to be rather restrictive for random variables without compact support (see [15, p. 4] where the author provide a broad list of references relying on this assumption). A trivial example of non-differentiability even in a favourable case (the operand is continuously differentiable) is provided in [15, Proposition 2.2]: the authors there prove that from continuity property (which is obtained in relatively general cases) to differentiability of probabilistic functions, simply adding the differentiability of the operand function is not sufficient and compactness of $M(x)$ turns out to be a key assumption. They also

paved the way for a set of special conditions (namely growth conditions imposed on either partial derivatives of the operand as in [25, 15, 14, 19, 142], or on some density function appearing in the decomposition of the random variable as will be discussed in Chapter 6, that is a summary of [17]) on probabilistic functions that can provide (general) differentiability information, while not requiring compactness of $M(x)$. Within our artificial distinction, this falls into the second type of approach, which investigates a another side of the trade-off between few assumptions on the random variables, and implementation practicality.

Variational studies of probabilistic functions now consequently become a "case-by-case" process, with significant impact factors being the class of random variable, the properties verified by the operand function and its structure (separable or not, generally). The objective here is to leverage properties verified by more specific operands or random variables. In order to derive some (sub-)differential properties, in particular in the sense of Clarke, one has to verify some stronger regularity properties than mere continuity. Local Lipschitz continuity is one of these stronger regularity proprieties. We recall that the famed Theorem of Rademacher states that Lipschitzian continuity of a function f implies differentiability almost everywhere on an open subset of the space where f is defined (see [82, Theorem 2.5.1]).

A. Prékopa proved smoothness of the non-degenerate probability distribution function of multivariate Gaussian random vectors in [245, Section 6.6.4]. As a consequence, when c is separable (see Definition 9), this property holds for equation (5.1) and the same reference provides a practical formulation of partial derivatives of the probability distribution function. It turns out these partial derivatives are dependent on functional evaluation of other Gaussian distribution functions in lower dimensions, which is computationally interesting - this is described as a "reduction". In [147], for random variables with quasi-concave probability distribution function[5], two major results are proposed. Firstly, the distribution function of quasi-concave random vectors is proved to be locally Lipschitzian if and only if none of the component ξ_i, $i \in [1, m]$ has zero variance (see [147, Theorem 2.2]). When applied to the following form:

$$\varphi(x) = \mathbb{P}\left[A\xi - \widetilde{c}(x) \leq 0\right], \text{ with } \widetilde{c} \text{ locally Lipschitzian,} \tag{5.9}$$

this implies that φ is itself locally Lipschitzian provided that none of the rows of A is in $\operatorname{Ker}\Sigma$ (see [147, Corollary 2.2]). The second important result builds upon this Lipschitz continuity, and proves smoothness of the probability distribution function even in the singular case, except on a set of zero-measure. When in (5.9), A turns out to be a function of x and ξ is multivariate Gaussian, differentiability results are proved in [11] which also includes a detailed numerical scheme to follow for computational results. Interestingly, the computation of functional values of φ and of its gradient $\nabla\varphi$ is boiled down to evaluation of Gaussian distribution functions, thus paving the way for efficient implementation using existing code. In [146], the authors study a case similar to (5.9) where $\widetilde{c} = Id$, for a (possibly singular, this time) multivariate Gaussian random vector ξ. Provided that the rows of A are linearly independent, φ is continuously differentiable: the main take-away of [146] is the (now proved) ability to have a generalized-linear

[5] This set of random vectors is relatively wide and includes the multivariate Gaussian singular or nondegenerate, uniform distributions over compact, convex subsets of $\mathbb{R}^m$,...a more comprehensive list is found in this article.

$\widehat{c}$ functions in a separable operant, which was only allowed for non-degenerate linear transformation in [245]. Again, the authors propose a gradient formulation that solely relies on functional values of the original probabilistic function and density values of one-dimensional Gaussian distributions. This is often referred to as a "reduction", and this particular one is valid everywhere except on a set of zero measure. An extension that accounts for these sets is proposed in [6]: for a multivariate (possibly singular) Gaussian vector, assuming $\widetilde{c}$ is differentiable, the probabilistic function (5.9) is locally Lipschitzian (which was already known) and an explicit formula for Clarke subdifferentials is readily available everywhere. Although Gaussian and Gaussian-like distributions receives most of the attention, it is worth noting that in a separable setting, [138] provides an efficient numerical methods for the computation of functional values and partial derivatives for Dirichlet distributions.

In the more general setting:

$$\varphi(x) = \mathbb{P}\left[c(x, \xi) \leq 0\right], \tag{5.10}$$

the authors of [15] are able to prove strong local properties. At a given decision point $\bar{x}$, assuming that $c : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuously differentiable, convex relatively to its second argument and verifies a growth condition at $\bar{x}$, and taking ξ as a standard multivariate Gaussian vector, φ is proved to be continuously differentiable on a neighbourhood of $\bar{x}$. An explicit formulation of its gradient is provided, as well as easy to verify conditions that ensure that $c(\bar{x}, 0) < 0$ (see Proposition 3.11 of this reference). Finally, these results are expanded to "Gaussian like" distributions (that include χ^2 and log-normal distributions) multivariate Student's distributions in [15, Section 4]. A direct extension of this latter work, in [19] is studied second-order differentiability in the same setting, with an involved but explicit formula for the Hessian of φ. Building upon this chain of works, in reference [14] the authors characterize the Clarke subdifferential of a probabilistic function by explicitly defining an including set. Aside from assumptions now becoming classic in literature (continuous differentiability of c, convexity relatively to the second argument, a certain growth condition for each coordinate function of c, and a "Slater" constraint qualification that imposes strict validity of $c(\bar{x}, \mathfrak{m}) < 0$ at the point of interest $\bar{x}$), [14, Theorem 3.6] provides this characterization as well as an explicit gradient formula. The assumption of continuous differentiability of the operand c is later weakened to Lipschitzian continuity of c in [142], where the authors investigate a setting similar to [14]. Still assuming continuous differentiability of c, in [4] the authors weaken the Gaussian assumption to extend the results of [14] to a broader class of random variables, namely the elliptical symmetric ones (the other assumptions being identical). Leaving aside the "finite" setting, it is worth noting that recent works are interested in probabilistic functions of the form $\mathbb{P}\left[c_t(\xi, x) \leq 0,\ \forall t \in T\right]$ with T being a given index set, possibly arbitrary as [21]. This latter work embraces a wide setting, wherein ξ belongs the class of elliptically symmetrically distributed random variables and c is again continuously differentiable, convex relatively to its second argument. Characterizations of several subdifferentials (Clarke, and more general ones) which provides a strong theoretical framework for (relatively to literature) nonrestrictive assumptions. On a theoretical basis, it can be seen as an natural extension of several pioneering works, as [14, 4, 142].

Cases where convexity is proved

When studying convexity, the separability assumption (see definition 9) also plays a key role: one could divide convexity-related studies between results on the separable case, and the (yet more scarce) results on nonseparable one.

Similarly to continuity and differentiability properties, convexity has long been studied. In fact, in [243] studied convexity/concavity of probability distribution functions for multivariate non-degenerate Gaussian variables. In particular, for the bivariate case, quasi-concavity of the joint probability distribution function with correlation coefficient r, i.e. $F_{\xi_1,\xi_2}(x_1, x_2, r)$, is proved for $r \geq 0,\ x_1 \geq 0,\ x_2 \geq 0$ and concavity of $F_{\xi_1,\xi_2}(x_1, x_2, r)$ as a function of either x_1 or x_2 is proved if $-1 < r \leq 0$. As a consequence, convexity results for the set $\{(x_1, x_2),\ |\ F_{\xi_1,\xi_2}(x_1, x_2, r) \geq \mathtt{p}\}$ follow, with lower bounds provided on $\mathtt{p}$ for this to hold.

In a more general setting, as convexity turns out to be verified only in (very) specific settings, some "neighbouring" properties were introduced which are equally desirable. First of all, log-concavity has been proposed by A. Prékopa [242]. In this article, the author investigates the logarithmically concave probability distribution function, which turns out to be verified for several classic random variables. Taking an opposite point of view, the rationale of log-concavity, when in a separable setting, is as follows: knowing that $\mathbb{P}\left[\widetilde{c}(x) \geq \xi\right] \geq \mathtt{p} \iff F_\xi(\widetilde{c}(x)) \geq \mathtt{p}$, is there a function α that could yield $F_\xi(\widetilde{c}(x)) \geq \mathtt{p} \iff -\alpha \circ F_\xi(\widetilde{c}(x)) \leq -\alpha(\mathtt{p})$ such that $-\alpha \circ F_\xi$ is convex ? Benefits from the existence of such an α for a given random variable ξ would be significant: one could use elements of convex optimisation, and have stronger optimality guarantees on the solution point of an appropriately designed algorithm. One can immediately see that α should either be increasing and lead to $\alpha \circ F_\xi$ concave, or decreasing and lead to $\alpha \circ F_\xi$ convex. Setting $\alpha = \log$ is interesting, as it is relatively easy to verify most of classic distributions happen to have a concave composition $\alpha \circ F_\xi$: in [242, Section 5], the multivariate Gaussian, Wishart, Beta, Dirichlet are proved to be log-concave (more log-concave distributions are to be found in [39]).

As a result from the pioneering work of A. Prékopa, it is now well-known that the feasible set M_f (recall (5.3)) is convex for any $\mathtt{p} \in [0, 1]$, provided that ξ is such that:

- it has a density f_ξ, which is log-concave,
- the operand is separable, of the form $c(x, \xi) = A\xi - \widetilde{c}(x)$, and $\widetilde{c}$ is concave.

Convexity of the feasible set in a nonseparable setting can be found in [148], where it is dependent on the value of $\mathtt{p}$; moreover, compacity of M_f is also proved, evidently with a condition on $\mathtt{p}$.

Convexity-related property of the probabilistic function itself is studied in the detailed reference [245]. In Theorem 10.2.1 of this reference, provided that all coordinate functions of the operand are quasi-concave as functions of both variables, that ξ has a log-concave density, then φ is proved to be logconcave in $\mathbb{R}^n$. Other closely related results are presented in Chapter 10 of this reference. A brief historical account on convexity is found in the first Chapter of [237]. while [18] proposes a thorough literature review on convexity studies of probabilistic functions and/or feasible sets. For an elliptical symmetrical

random vector[6], reasonable assumptions on the operand (namely lower-semicontinuity and convexity with respect to its first variable), the feasible set M_f is convex for p "large enough". As previously discussed, the results presented in [18] on convexity is obtained by assuming specific, not particularly restrictive, properties hold for both the class of random variable and the operand.

Moreover, the recent work [18] provides an overview of cases where convexity of the feasible set holds. It can be seen as an extension of works on quasi-concavity and log-concavity, as the authors propose a concept of α-concavity which is a generalized concavity concept that includes the two previously stated generalized concavity properties. This new concept interestingly defines tight assumptions on the properties of the random variable that ensure convexity of the feasible set of the chance-constraint. Remaining and future challenges are then discussed. For instance, current results prove convexity of M_f provided that p is above an explicit threshold: is there "some convexity" to obtain even if p is lower than the threshold?

5.2.3 Numerical methods to deal with probabilistic constraints

The development of complex models using chance-constraints has motivated thrilling works, and paved the way for exciting research both on the algorithmic/numeric side and the theoretical one. Although we will taking a nonlinear programming point of view, i.e. studying the variational properties of probabilistic functions, we briefly point out the existing numerical methods. First of all, let us stress that even computing a functional value can be difficult: computing a value $\varphi(x)$ (with φ as in (5.1)) often requires a multidimensional integration. In general, no analytical formulation of φ is available, so this task is a case-by-case matter. A brief overview of numerical method to compute the value of probabilistic functions can be found in [124, Section 1.2].

Here is a subset of existing numerical methods in literature: a detailed presentation of these is beyond the scope of this work, and we refer to a handful of selected references for more details on each of them.

Sample-average approximation

The rationale in such an approach is to approximate the expectation of a random variable by the mean of a user-generated independent samples. A classic reference to Sample-average approximation is [232]: in this work the authors replace the distribution at hand in the probabilistic function by a user-generated distribution comprised of a random sample. Proposition 2.2 of this article proves that as the random sample size increases, the sets of optimal solutions and the optimal values of the SAA-problem converge to the sets of optimal solutions and the optimal values of the original chance-constrained problem. The necessary assumptions seem to be reasonable, namely that the objective and function under the probability measure sign have some continuity properties, and a type of constraint qualification holds (see Assumption (A) therein).

[6] More precisely, results are proven when ξ belongs to a subset of this class of random vector. Efficient numerical methods to very the belonging to this subset are provided, so that the results are readily implementable.

Scenario approximation

This approach is also *sample based*, as is the sample-average one, but differs in that the ambition is not to approximate the expected value of random variables. As described in the didactic and classic article [63], after having generated N independent samples, the scenario-approximated chance-constrained problem is set up by replacing the probabilistic constraints by a set of N constraints, where each one is obtained by setting the random variables to the associated scenario values. This approach is more general than the probabilistic setting and was, in fact, initially applied to *semi-infinite convex programming* (i.e. one has finite number of variables, but an infinite number of constraints). Note that a lower-bound on the number of scenarios N is provided in [63], depending on the safety parameter p.

Boolean approaches

This class of methodologies is relatively recent, first introduced in 2012 and extended by the same author and his co-authors in [189]. In this work, a reformulation framework is provided: a chance-constrained problem is casted as MINLP. Two reformulations are presented, and a key element is that the number of binary variables introduced is not increasing in the number of scenarios considered. The reformulated problem at hand is still a difficult one to solve as it is non-convex quadratic with integer variables. The authors nevertheless are able to provide numerical results on real-life network-related problems with up to 10.000 scenarios (and interestingly only a 25% increase of CPU time between 5.000 and 10.000 scenarios). It appears that this approach is successful when the size of the random vector, m, is "relatively" small (in practise: less than a few dozens). When m increases, the number of integer variables increases and this approach becomes numerically ineffective.

Mixing with robust optimisation

Seeking to combine the computational tractability of robust optimisation while maintaining a less conservative standpoint, the authors of [302] provide a list of robust approximations of a chance-constrained problem where the uncertainty sets include only a subset of all possible outcomes. In other words, some extreme scenarios are left out, which number is determined by the safety parameter p of the original probabilistic constraint.

Convex approximation

Convexity in chance-constrained programming is a large subdomain: in [245] is recalled a result on log-concave distribution. Aside from the detailed reference [245], a summary on convexity properties of probabilistic constraints is to be found in [237, Section 1.1].

Convex property is not sufficient to ensure an efficient solving technique, one also requires a capable computational tool for the probability at hand: convex approximations are thus useful in this matter. In [218] the authors derive a class of convex, explicit, and tractable approximations of the probabilistic constraint, deemed as "conservative". They provide an approximation of $\mathbb{P}\left[c(x,\xi) \leq 0\right] \geq \mathrm{p}$, with possibly non-separability of c but still some structural assumptions, by a constraint of the form $d(x,t) \leq 0$ with

t a parameter and d a convex function defined on an appropriate set. In this class of approximations, a duple x, u verifying the latter implies that x verifies the former (this implication is the reason why the convex approximations are qualified as conservative). Their method is interesting in that it is analytical (i.e. not scenario based) and expands the allowed structural assumptions on functions and distributions at hand, which previously were more on restrictive.

5.3 DIFFERENCE-OF-CONVEX APPROXIMATIONS OF CHANCE-CONSTRAINTS

In this section, we are concerned with chance-constrained optimization problems of the form

$$\begin{cases} \min\limits_{x \in X} & f_1(x) - f_2(x) \\ \text{s.t.} & \mathbb{P}\left[\mathtt{lb} \leq c_2(x,\xi) - c_1(x,\xi) \leq \mathtt{ub}\right] \geq p\,, \end{cases} \tag{5.11}$$

where c_2, $c_1 : \mathbb{R}^n \times \Xi \to \mathbb{R}$ are convex functions w.r.t argument x. The bounds $\mathtt{lb} \in \mathbb{R} \cup \{-\infty\}$ and $\mathtt{ub} \in \mathbb{R}$ are given parameters, and $\xi \in \Xi$ is a random vector following a known probability distribution. Parameter $p \in (0,1)$ is a confidence level: the constraint $\mathtt{lb} \leq c_2(x,\xi) - c_1(x,\xi) \leq \mathtt{ub}$ must be satisfied with probability at least p. Note that the constraint of (5.11) fits (3.8) with $M(\xi) := \{x \in \mathbb{R}^n : \mathtt{lb} \leq c_2(x,\xi) - c_1(x,\xi) \leq \mathtt{ub}\}$.

5.3.1 DoC FORMULATION VIA APPROXIMATION OF THE CHARACTERISTIC FUNCTION

In this section, we consider a DoC approximation of the probability function. We follow the lead of [153] and extend the approach therein to handle intervals $[\mathtt{lb}, \mathtt{ub}]$ under the probability function. Different from [153], one of our DoC approximations is smooth provided that c_2 and c_1 are smooth functions with respect to x. Suppose the random vector ξ follows a continuous probability distribution, and assume that a model/software is available for generating independent and identically distributed scenarios of ξ. Let $c(x,\xi) := c_2(x,\xi) - c_1(x,\xi)$. The probability function $\mathbb{P}\left[\mathtt{lb} \leq c(x,\xi) \leq \mathtt{ub}\right]$ can be alternatively written as

$$\mathbb{E}[\mathbb{1}_{[\mathtt{lb},\mathtt{ub}]}(c(x,\xi))], \quad \text{with} \quad \mathbb{1}_{[\mathtt{lb},\mathtt{ub}]}(z) = \begin{cases} 1 & \text{if } z \in [\mathtt{lb}, \mathtt{ub}] \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbb{E}[\cdot]$ is the expected value operator of the random vector ξ w.r.t. the probability measure $\mathbb{P}$. The discontinuous characteristic function $\mathbb{1}_{[\mathtt{lb},\mathtt{ub}]}(\cdot)$ can be approximated by

the following continuous one, with $t > 0$ a given small parameter:

$$\zeta^t(z) := \begin{cases} 0 & \text{if } z \leq \mathtt{lb} - t \\ \frac{z+t-\mathtt{lb}}{t} & \text{if } \mathtt{lb} - t \leq z \leq \mathtt{lb} \\ 1 & \text{if } \mathtt{lb} < z \leq \mathtt{ub} \\ \frac{-z+t+\mathtt{ub}}{t} & \text{if } \mathtt{ub} < z \leq \mathtt{ub} + t \\ 0 & \text{if } \mathtt{ub} + t < z. \end{cases} \tag{5.12}$$

The above is nothing other than a DoC function: for $z = c(x,\xi)$ we get

$$\zeta^t(c(x,\xi)) = \frac{\max\{c(x,\xi)+t, \mathtt{lb}\}}{t} + \frac{\max\{c(x,\xi)-t, \mathtt{ub}\}}{t} - \left[\frac{\max\{c(x,\xi), \mathtt{lb}\}}{t} + \frac{\max\{c(x,\xi), \mathtt{ub}\}}{t}\right].$$

Note that the four terms above are themselves DoC functions:

$$\begin{aligned} \max\{c(x,\xi)+t, \mathtt{lb}\} &= \max\{c_2(x,\xi)+t, c_1(x,\xi)+\mathtt{lb}\} - c_1(x,\xi) \\ -\max\{c(x,\xi), \mathtt{lb}\} &= -\max\{c_2(x,\xi), c_1(x,\xi)+\mathtt{lb}\} + c_1(x,\xi) \\ \max\{c(x,\xi)-t, \mathtt{ub}\} &= \max\{c_2(x,\xi)-t, c_1(x,\xi)+\mathtt{ub}\} - c_1(x,\xi) \\ -\max\{c(x,\xi), \mathtt{ub}\} &= -\max\{c_2(x,\xi), c_1(x,\xi)+\mathtt{ub}\} + c_1(x,\xi). \end{aligned}$$

By summing these four equalities and dividing by t we get

$$\begin{aligned} \zeta^t(c(x,\xi)) = \quad & \frac{\max\{c_2(x,\xi)+t, c_1(x,\xi)+\mathtt{lb}\}}{t} + \frac{\max\{c_2(x,\xi)-t, c_1(x,\xi)+\mathtt{ub}\}}{t} \\ & - \left[\frac{\max\{c_2(x,\xi), \quad c_1(x,\xi)+\mathtt{lb}\}}{t} + \frac{\max\{c_2(x,\xi), \quad c_1(x,\xi)+\mathtt{ub}\}}{t}\right]. \end{aligned}$$

The above is a nonsmooth DoC approximation of the function $\mathbb{1}_{[\mathtt{lb},\mathtt{ub}]}(c(x,\xi))$. A smooth one ζ_s^t is readily available by replacing $\max\{a,b\}$ with a smooth and convex function, e.g.,

$$s^t(a,b) := \max\{a\,u_1 + b\,u_2 - \frac{t}{2}d(u) \quad | \quad u \in \mathbb{R}^2_+,\ u_1 + u_2 = 1\},$$

with $d : \mathbb{R}^2 \to \mathbb{R}_+$ a differentiable function. Classical choices for d are $d(u) = \sum_i u_i \ln(u_i)$ and $d(u) = \|u\|^2$. In this first alternative, the smooth function $s^t(a,b)$ has a closed form (but may suffer from numerical issues due to scaling). In the second, efficient algorithms for projecting a vector onto a simplex are available in the literature. Given s^t, provided c_2 and c_1 are smooth functions, we can approximate $\zeta^t(c(x,\xi))$ with the smooth DoC function $\zeta_s^t(c(x,\xi)) = \frac{1}{t}(\psi_1^t(x,\xi) - \psi_2^t(x,\xi))$ of components:

$$\begin{aligned} \psi_1^t(x,\xi) &:= s^t(c_2(x,\xi)+t, c_1(x,\xi)+\mathtt{lb}) + s^t(c_2(x,\xi)-t, c_1(x,\xi)+\mathtt{ub}), \\ \psi_2^t(x,\xi) &:= s^t(c_2(x,\xi), \quad c_1(x,\xi)+\mathtt{lb}) + s^t(c_2(x,\xi), \quad c_1(x,\xi)+\mathtt{ub}). \end{aligned}$$

Note that assessing the gradient of $\zeta_s^t(c(x,\xi))$ amounts to computing the gradient of $s^t(a,b)$. We present the gradient expression for the case $a = c_2(x,\xi)+t$ and $b = c_1(x,\xi)+$

$\texttt{lb}$ (the gradients for the other cases follow analogously). Under the assumption that c_2 and c_1 are smooth as functions of x, $d(\cdot) = \|\cdot\|^2$, and letting $u^\star$ be the projection of $\frac{1}{t}[c_2(x,\xi)+t, c_1(x,\xi)+\texttt{lb}]$ onto the simplex $\{u \in \mathbb{R}^2_+, u_1+u_2=1\}$, the gradient g^t of s^t at $(c_2(x,\xi), c_1(x,\xi))$ is $g^t := [\nabla_x c_2(x,\xi), \nabla_x c_1(x,\xi)]^\top u^\star$. This is the formula considered in our numerical experiments.

Figure 5.1 illustrates the smooth DoC function ζ_s^t for two different values of t. Note that the smaller is t, the better is the approximation of $\mathbb{1}_{[\texttt{lb},\,\texttt{ub}]}(\cdot)$.

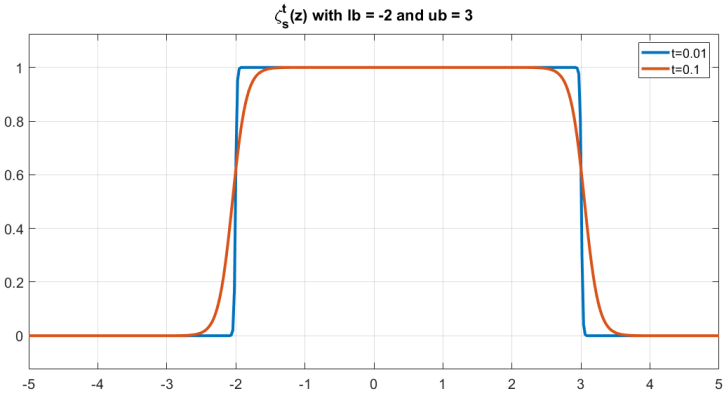


FIGURE 5.1: Illustration of the smooth DoC function approximating $\mathbb{1}_{[-2,3]}(z)$.

Given this DoC decomposition, the probability function in (5.11) can be approximated by

$$\begin{aligned}\mathbb{P}\left[\texttt{lb} \le c(x,\xi) \le \texttt{ub}\right] &= \mathbb{E}[\mathbb{1}_{[\texttt{lb},\,\texttt{ub}]}(c(x,\xi))] \\ &\approx \mathbb{E}[\zeta_s^t(c(x,\xi))] \\ &= \mathbb{E}[\psi_1^t(x,\xi)]/t - \mathbb{E}[\psi_2^t(x,\xi)]/t.\end{aligned} \tag{5.13}$$

The expectation $\mathbb{E}[\cdot]$ can be efficiently approximated via Monte-Carlo simulation by considering a fixed sample of scenarios randomly generated according to the distribution of ξ. In our numerical experiments we randomly generate a sample of N scenarios $\{\xi^1, \ldots, \xi^N\}$ and estimate the convex functions $\mathbb{E}[\psi_1^t(x,\xi)]$ and $\mathbb{E}[\psi_2^t(x,\xi)]$ by their sample average approximations (SAA). Once the sample of scenarios is available, the estimations are easily computed. This allows us to employ large samples ($N \ge 10\,000$) and, due to the *Law of Large Numbers*, to get reliable estimators for the expected values. Accordingly, problem (5.11) fits formulation (3.7) with

$$c_1(x) := \frac{1}{N}\sum_{i=1}^{N}[\psi_2^t(x,\xi^j)] + t\,p \quad \text{and} \quad c_2(x) := \frac{1}{N}\sum_{i=1}^{N}[\psi_1^t(x,\xi^j)]\,. \tag{5.14}$$

We point out that if $c(x,\xi)$ is smooth w.r.t x, then c_1 and c_2 are smooth convex functions. If in addition f_2 is smooth, then Algorithm 1 applied to our reformulation of the chance-constrained problem (5.11) computes a B-stationary point (because (5.11) easily satisfies a CQ if $p < 1$).

5.4 DIFFERENCE-OF-CONVEX PROPERTY OF A SUBCLASS OF PROBABILISTIC FUNCTIONS

For this section, we assume that $\varphi(x) = \mathbb{P}\left[c_1(x,\xi) - c_2(x,\xi) \leq 0\right]$ with ξ a random variable, c_1, c_2 are convex functions of there first component defined on $\mathbb{R}^n \times \mathbb{R}^m$ with values in $\mathbb{R}^d$.

Proving the DoC property of φ is not trivial, and indeed this is not true in general: in Example 2 we clearly have DoC operands for which the associated φ is discontinuous, and consequently not DoC. We present here the original result showing that φ is in fact DoC provided that a list of assumptions is verified. Before presenting this result, some preliminary tools are necessary.

5.4.1 PRELIMINARY ELEMENTS FOR THE DOC PROPERTY OF CCS

We start by recalling that φ is properly defined because $c = c_1 - c_2$ is a DoC function, it is thus locally-Lipschitz and consequently a *Baire* function[7]. Then by [107, Theorem IV.4 p.117], for a given $x \in \mathbb{R}^n$ $c(x,\xi)$ is a random vector thus measurable in our probability space, which implies the correct definition of $\mathbb{P}\left[c(x,\xi) \leq 0\right]$ for all x.

We then define a structure for the random variable ξ, and follow the lead of recent literature by assuming that it is elliptically symmetrically distributed. This class of random variables is an extension of the multivariate Gaussian distributions. For an extensive study on spherical and elliptical symmetric random variables, we refer to [104].

Definition 10 (Spherical and Elliptical symmetric distributions)**.** We say that a random vector $\xi \in \mathbb{R}^m$ has a spherically symmetric distribution if for every orthogonal matrix $\Gamma \in \mathcal{O}_m(\mathbb{R})$:

$$\Gamma\xi = \xi.$$

We say that the random vector $\xi \in \mathbb{R}^m$ is elliptically symmetrically distributed with mean $\mathfrak{m}$, covariance-like matrix Σ if:

$$\xi = \mathfrak{m} + A'y, \quad y \text{ is spherically distributed.} \tag{5.15}$$

•

Remark 9. In definition 10, the sign "=" is to be understood as equality in terms of distributions: let x, y be two random vectors, then "$x = y$" exactly means both vectors have the same distribution. ▷

Remark 10. In Definition 10, Σ is not the exact covariance matrix of ξ. As detailed in [104, Theorem 2.17], there is a scalar factor between Σ and $\mathrm{Cov}(\xi)$. ▷

An immediate and precious property that will be useful is the following:

Lemma 4 (see page 32 of [104] or, for example, p.6 of [18])**.** Let $\xi \in \mathbb{R}^m$ be elliptically symmetrically distributed, and $\mathrm{rank}(\Sigma) = d$. Then ξ has the following representation:

$$\xi = \mathfrak{m} + \mathcal{R}L\zeta, \tag{5.16}$$

[7] Basically speaking, it belongs to the smallest class of functions that contains all continuous functions: see [107, p.106].

with:

- $L \in \mathcal{M}_{d,m}(\mathbb{R})$ such that $\Sigma = LL^{\mathsf{T}}$.
- $\mathcal{R}$ a univariate random variable with support in $\mathbb{R}_+$ (which density, when it exists, is explicitly known and to be found in [104, Theorem 2.9]).
- ζ is a random vector with a uniform distribution over the Euclidean d-dimensional unit sphere $\mathbb{S}^{d-1} := \{z \in \mathbb{R}^d : \sum_{i=1}^d z_i^2 = 1\}$.
- ζ and $\mathcal{R}$ are independent random variables.

This representation is called the *radial-spherical representation*. ◀

Note that this representation enables us to "decompose" a relatively general type of random variable into elements with stronger properties. This leads to a key argument for future developments which is the new formulation of the probabilistic function $\mathbb{P}[\xi \in M]$ (see Insert 6 for an associated geometric interpretation).

Reformulation of a chance-constraint in the elliptical case

Going back to $\varphi(x)$ and using the set $M(x)$ previously introduced, we obtain: $\varphi(x) = \mathbb{P}[\xi \in M(x)]$, with $M(x) = \{z \in \mathbb{R}^m \mid c_1(x,z) - c_2(x,z) \leq 0\}$. It is a Lebesgue measurable set in $\mathbb{R}^m$. From Lemma 4, we can rewrite φ at a given $x \in \mathbb{R}^n$ as:

$$\varphi(x) = \int_{v \in \mathbb{S}^{k-1}} \mu_{\mathcal{R}}\left(\{r \geq 0 \ : \ \mathfrak{m} + rLv \cap M(x) \neq 0\}\right) d\mu_\zeta.$$

This is obtained directly from definitions. Under some additional assumptions on $x \mapsto M(x)$, an enriched formulation is available in literature, which we recall here for clarity:

Theorem 9 ([18, Theorem 2.1])**.** Let $\xi \in \mathbb{R}^m$ be an elliptical random vector with mean $\mathfrak{m}$. Let $x \in \mathbb{R}^n$ be such that $c_1(x,\mathfrak{m}) - c_2(x,\mathfrak{m}) \leq 0$. Assume furthermore that:

$$c_1(x, \mathfrak{m} + \lambda(z - \mathfrak{m})) - c_2(x, \mathfrak{m} + \lambda(z - \mathfrak{m})) \leq 0, \quad \forall \lambda \in [0,1] \tag{5.17}$$

holds for all z such that $c_1(x,z) - c_2(x,z) \leq 0$. Then the probability function φ satisfies

$$\varphi(x) = \int_{v \in \mathbb{S}^{m-1}} F_{\mathcal{R}}(\rho(x,v)) d\mu_\zeta(v) \tag{5.18}$$

where the mapping $\rho\colon \mathbb{R}^n \times \mathbb{S}^{m-1} \to [-\infty, \infty]$, called ray function, is defined by:

$$\rho(x,v) := \begin{cases} \sup_{t \geq 0} & t \\ \text{s.t.} & c_1(x, \mathfrak{m} + tLv) - c_2(x, \mathfrak{m} + tLv) \leq 0. \end{cases} \tag{5.19}$$

◀

Condition (5.17) is referred to a "star-shaped" set: it states that from $\mathfrak{m} \in \mathbb{R}^m$, for every $z \in \mathbb{R}^m$ every segment $[\mathfrak{m}, z]$ should be feasible for the constraint $c_1(x,\cdot) - c_2(x,\cdot) \leq 0$. One can immediately observe that every convex set is star-shaped, and the reverse does not hold (see Example).

Example 3. We here provide a graphical example, for a "star-shaped-nonconvex" valued mapping $x \mapsto M(x)$. We consider $\bar{x} \in \mathbb{R}^n$ to be such that: $c_1(\bar{x}, \mathfrak{m}) - c_2(\bar{x}, \mathfrak{m}) < 0$. By continuity, we can find a neighbourhood of $\bar{x}$ on which this strict inequality is verified, and we select two points in it denoted x_1, x_2.

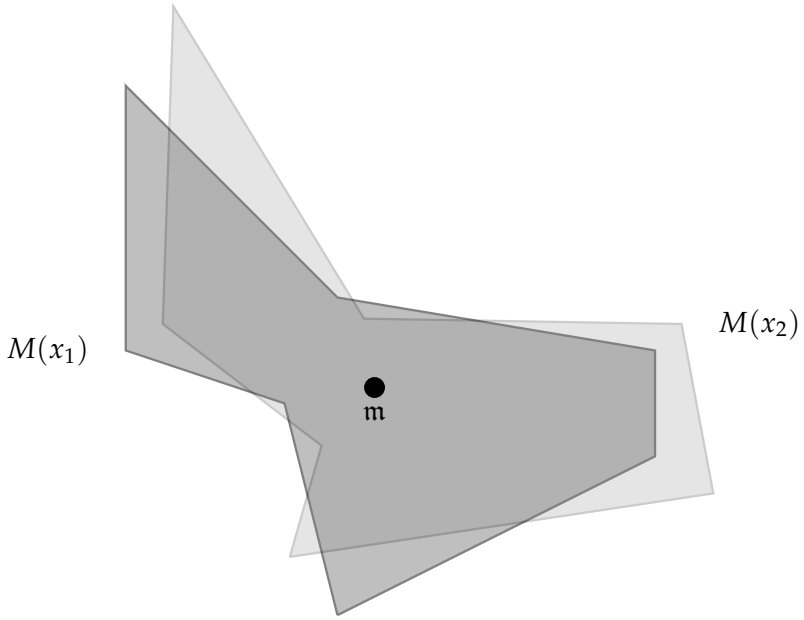


FIGURE 5.2: Example of star shaped non convex mapping, in the neighbourhood of a point where $c(x, \xi) < 0$.

▶

INSERT 6: Geometric interpretation of the radial-spherical representation. From Lemma 4, we can derive the following somewhat enlightening geometric interpretation. Let us consider the following function:

$$\mathbb{P}\left[\xi \in M\right] = \int_{v \in \mathbb{S}^{m-1}} \mu_{\mathcal{R}}\left(\{r \geq 0 \ : \ \mathfrak{m} + rLv \cap M \neq 0\}\right) d\mu_{\zeta}$$

with M a polyhedron in $\mathbb{R}^m$. Now, broadly speaking, the computation of this probability value amounts to summing every intersections between the set M and a segment from $\mathfrak{m}$, in direction v and of length $r > 0$ [a]. In a two dimensional space, here is a visualisation with an arbitrary set M which is here presented as bounded for convenience:

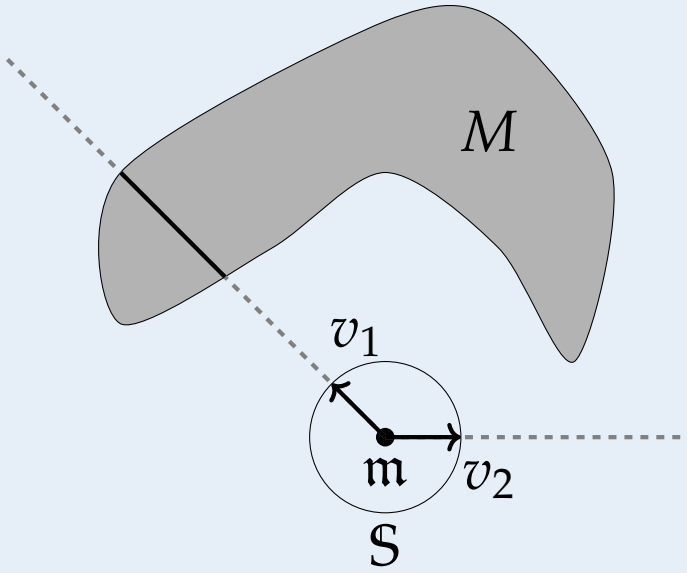


FIGURE 5.3: Geometric visualisation of the spherical-radial decomposition.

In this figure, the circle $\mathbb{S}$ is the sphere of unit radius, v_1 and v_2 are two "aiming directions" belonging to $\mathbb{S}$, and the dark line in M is the set which measure contributes to the computation of the aforementioned probability value. In this example M is depicted as bounded, which is not required.

[a] This is a superficial and inexact description as, for instance, it does not account for matrix L. Nevertheless we still believe that this leap proposes an easy to understand geometric interpretation that can help to grasp more intuitive feeling on the importance and significance of the radial-spherical representation

Remark 11. This class of distribution is not in fact of any particular limitations: as described in [176, Section 2], the multivariate-Gaussian, -Student-t, -logistic, -exponential power families are all included in the symmetric elliptic one. ▷

5.4.2 Chance-constraints with DoC functions

With this preliminary elements in hand, we now turn our attention to the CC of interest (see equation (5.2)). The following theorem is a direct combination of theorem 9 and DoC preserving operations.

Theorem 10 (DoC property of CC functions)**.** Let ξ be an elliptically symmetrically distributed random variable which mean is denoted $\mathfrak{m}$ and φ be defined as follows:

$$\varphi(x) = \mathbb{P}\left[c_1(x,\xi) - c_2(x,\xi) \leq 0\right]. \tag{5.20}$$

Let us moreover assume:

1. c_1 is convex and c_2 concave in their respective second argument;
2. $F_{\mathcal{R}}$ is in $\mathcal{C}^2(\mathbb{R}_+)$;
3. $x \to \rho(x,v)$ is DoC for all $v \in \mathbb{S}^{m-1}$;
4. x is such that $c_1(x,\mathfrak{m}) - c_2(x,\mathfrak{m}) \leq 0$.

Then $x \to \varphi(x)$ is a DoC function. ◀

Proof. The proof is largely based on the formulation obtained in Theorem 9 and DoC preserving operations that are described in [285, Corollary 4.2] and [229]. Note that the star-shape requirement of 5.17 is directly obtained from item 1. With all of our assumptions, we have on the one hand:

$$\varphi(x) = \int_{v\in\mathbb{S}^{m-1}} F_{\mathcal{R}}\left(\rho(x,v)\right) d\mu_\zeta(v).$$

And on the other hand we know that $x \to F_{\mathcal{R}}(\rho(x,v))$ is DoC for all $v \in \mathbb{S}^{m-1}$ by [285, Corollary 4.2]. As a consequence, φ being the integral over a bounded set of a DoC function, it is DoC. □

Comments on the assumptions of Theorem 10

We start by stressing that assumption 4 from theorem 10 should not be considered as a limitation: it merely states that our decisions x should be feasible for the mean vector $\mathfrak{m}$. This is the case when considering a CC as in (5.2) with a safety parameter $\mathfrak{p}$ that verifies $\mathfrak{p} > \frac{1}{2}$.

Assumption 1 on the respective convexity and concavity of c_1, c_2 as functions of their second arguments ensures that reformulation (5.18), which is true as soon as $M(x)$ is star-shaped, is verified. Evidently, this assumption is sufficient: one could replace it by assuming $M(x)$ is start-shaped for every x of a selected set.

Assumptions 2 and 3 of Theorem 10 can appear to be particularly restrictive: in particular, asking for $F_{\mathcal{R}}$ to be $\mathcal{C}^2$ is not trivial. A slightly less restrictive alternative is to replace the assumption of $F_{\mathcal{R}}$ being $\mathcal{C}^2$ by the following one: $F_{\mathcal{R}}$ is DoC. This is merely thanks to [285, Proposition 4.4] which, similarly to the Corollary used in the proof, states that every composition of DoC function also is DoC provided that some easily verified assumptions are met (for instance, the domains and image sets of the functions have to be convex).

Finally, we care to emphasize that Theorem 10 has indirectly proven that φ is continuous (which is to be expected, as it has been proven it is a DoC function). To verify this claim, we need to invoke [13, Theorem 2.2]:

Theorem 11 (Thm 2.2 from [13])**.** Let $c\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ be a continuous mapping and assume that the following regularity condition holds for all $x \in \mathbb{R}^n, j = 1, \dots, k$:

$$\mathbb{P}\left[c_j(x, \xi) = 0\right] = 0, \tag{5.21}$$

then the probability function $\varphi(x) = \mathbb{P}\left[c(x, \xi) \leq 0\right]$ is continuous at any $x \in \mathbb{R}^n$. ◀

As discussed in in [13, p. 590], for the regularity condition (5.21) to hold, one way is to ensure the following two conditions:

- ξ has a density with respect to the Lebesgue measure;
- $\{z \in \mathbb{R}^m\colon c(x, z) = 0\}$ is a Lebesgue null set. This last condition holds whenever c is convex in its second argument, and admits a Slater point.

One can observe that the firs item is verified in our case. For the second item, as we assume that $\mathfrak{m} \in \text{ int } (M(x))$, we have: $c_1(x, \mathfrak{m}) - c_2(x, \mathfrak{m}) < 0$. In other words, $\mathfrak{m}$ is a Slater point. From the assumptions of Theorem 10, it is clear that function $c_1 - c_2$ is convex in its second argument: we conclude that the regularity (5.21) condition is verified. As a consequence, we have verified that φ is continuous, as expected.

5.5 CONCLUSION

We have recalled in this Chapter some elements of probability theory in order to properly introduce chance-constraints. Interestingly, chance-constraints are DoC functions although we do not have an explicit formulation; there exist a DoC approximation of chance-constraints, which we use in Part IV where we introduce our jointly chance-constrained OPF. We care to emphasize that this methodology is only one possibility: the following Chapter is yet another proof that other methodologies are possible.

MORE IN-DEPTH ANALYSIS OF VARIATIONAL PROPERTIES OF CC

6

This chapter is based on the work [17] that has been accepted for publication. We here provide an overview of this work, while leaving large parts of the technicalities to the original work: our motivation is to show that they are other works to deal with jointly-chance-constrained OPF and we leave numerical results for future works.

Until now, we have presented two practical implementable approximations DoC approximations of CC, and provided a more theoretical result on the (exact) DoC property of these functions. Based on a dense and remarkably continuous corpus, we move on to present some variational properties on the a particular case of the probability function φ. This work has been started by Wim van Ackooij and Pedro Pérez-Aros, and builds on some of their previous works as well as others by András Prékopa [241, 245], René Henrion [146], Jérôme Malick [106] and their co-authors. Following this stream of works, the probabilistic function φ is studied from a variational point of view. Among the practical interests of these theoretical work is to apply these in an approximate projected subgradient method, an example of which is provided in Appendix C.

Assumption 1. In this section, we will work on the following special case of the probability function φ:

$$\begin{aligned} \varphi\colon \mathbb{R}^n &\to [0,1] \\ x &\mapsto \mathbb{P}\left[\xi \in \bigcup_{j=1}^{\ell} M_j(x)\right]. \end{aligned} \tag{6.1}$$

where $M_j(x) = \{z \in \mathbb{R}^m \colon A_j(x)z \leq b_j(x)\}$, A_j, b_j respectively being a continuous matrix- and vector-valued maps. ξ is still assumed to be an elliptical symmetric random variable, and $\ell < \infty$.

We will investigate the (sub)differential/generalized differential properties of (6.1). As a matter of fact, this assumption is relatively nonrestricting. Observe that in Assumption 1 we impose some simplifications for this difficult problem. First of all, we are in a separable case: we believe this assumption to be reasonable as it is commonly made in literature. Secondly, we impose the including set to be defined by a finite-union of polyhedra. The interest of this particular case can be found in examples example 4.

Example 4. Let us consider the following situation $g\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ where g admits the following representation

$$g(x,z) = g_1(x,z) - g_2(x,z), \tag{6.2}$$

where g_1 and g_2 verify:

$$g_1(x,z) = \max_{i=1,\dots,\ell_1} \left\{ a_1^i(x) + \left\langle s_1^i(x), z \right\rangle \right\}, \tag{6.3}$$

and

$$g_2(x,z) = \max_{j=1,\dots,\ell_2} \left\{ a_2^j(x) + \left\langle s_2^j(x), z \right\rangle \right\}, \tag{6.4}$$

for appropriate scalar and vector valued functions $x \mapsto a_k^i(x), x \mapsto s_k^i(x), i = 1,\dots,\ell_k$, $k = 1,2$. Note that these functions clearly are DoC as function of their second arguments.

Now the original constraint $g(x,z) \leq 0$ can be re-written as:

$$\min_{j=1,\dots,\ell_2} \max_{i=1,\dots,\ell_1} a_1^i(x) + \left\langle s_1^i(x), z \right\rangle - a_2^j(x) - \left\langle s_2^j(x), z \right\rangle \leq 0, \tag{6.5}$$

or when introducing, for $j = 1,\dots,\ell_2$, the set-valued maps $M_j : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ defined as:

$$M_j(x) := \left\{ z \in \mathbb{R}^m \ : (s_1^i(x) - s_2^j(x))^\mathsf{T} z \leq a_2^j(x) - a_1^i(x), i = 1,\dots,\ell_1 \right\} \tag{6.6}$$

it becomes clear that $\{z \in \mathbb{R}^m \ : g(x,z) \leq 0\} = \bigcup_{j=1}^{\ell_2} M_j(x)$: a finite union of polyhedra.

It thus holds that

$$\mathbb{P}\left[g(x,\xi) \leq 0\right] := \mathbb{P}\left[\xi \in \bigcup_{j=1}^{\ell_2} M_j(x)\right],$$

which is exactly the structure we will investigate in this work. ▶

We also require the following definition for this chapter.

Definition 11 (First-order polynomial growth condition)**.** Let $\xi \in \mathbb{R}^m$ be elliptically symmetrically distributed with associated radial density function $f_\mathcal{R} : \mathbb{R}_+ \to \mathbb{R}_+$. Then we say that ξ or $f_\mathcal{R}$ is compatible with the first order polynomial growth condition if and only if:

$$\lim_{r\to\infty} f_\mathcal{R}(r) r^2 = 0. \tag{6.7}$$

•

Remark 12. The table 1 in [4] provides a large family of radial distributions compatible with this first order polynomial growth condition. When ξ is multivariate Gaussian random, the condition is satisfied. It also holds for multi-variate Student random vectors ξ, whenever the related degrees of freedom ν is larger than 1. ▷

Definition 11 turns out to be an additional assumption for our results in this Chapter. It can be related to another assumption that is found in literature, where some works assume their random variables have compact (thus bounded) support (see [15, p. 4]). This last assumption is rather restrictive: the case where the set $M(x) = \{z \in \mathbb{R}^m : g(x,z) \leq 0\}$ is unbounded is discarded under this assumption. Using Definition 11, it is possible to drop the boundedness assumption on $M(x)$ at a given $x \in \mathbb{R}^n$. For an elliptically symmetrically distributed random variable ξ, compatibility with the first order polynomial growth condition makes it possible to control the growth of the first partial derivative of g with respect to x: this latter control is found under the name *polynomial growth condition* (note that *compatibility* is dropped) in [4] for instance. The key elements to relate both definitions lie in the polynomial structure in x of the operand of equation (6.1) and [4, Lemma 6].

The importance of the first-order growth condition is highlighted in [15, Definition 3.6], and discussed in [4, Section 3]. An example that highlights the necessity of the

first order polynomial growth condition is provided in [15, Proposition 2.2]. First-order growth condition turns out to be a necessary assumption throughout this Chapter.

The main question to investigate is "how does φ behave locally?", and to characterize these variations we will look at its Clarke subdifferential $\partial^C\varphi(x)$. We recall that in a general setting, φ is not DoC let alone convex, nor is it yet locally Lipschitz. Our approach is based on space partition, and rigorous analysis of the behaviour of the function at hand on each partition. From these analyses we are able to:

- prove that φ is locally-Lipschitz, and consequently define its Clarke subdifferential $\partial^C\varphi$;
- propose an inclusion of its subdifferential in an explicitly defined set;
- present practical results for computational-friendly implementation.

6.1 REPRESENTATION OF THE PROBABILITY FUNCTION WITH THE SPHERICAL-RADIAL DECOMPOSITION

We recall that by Lemma 4, for any Lebesgue measurable set $M \subseteq \mathbb{R}^m$ the following identity holds:

$$\mathbb{P}\left[\xi \in M\right] = \int_{v\in\mathbb{S}^{m-1}} \mu_{\mathcal{R}}\left(\{r \geq 0 : \mathfrak{m} + rLv \cap M \neq \emptyset\}\right) d\mu_\zeta(v), \tag{6.8}$$

with $\mu_{\mathcal{R}}$ and μ_ζ are the laws of $\mathcal{R}$ and ζ, respectively. Therefore, in particular for any $j = 1, \ldots, \ell$, it follows that

$$\mathbb{P}\left[\xi \in M_j(x)\right] = \int_{v\in\mathbb{S}^{m-1}} \mu_{\mathcal{R}}\left(\{r \geq 0 : r(A_j(x)Lv) \leq (b_j(x) - A_j(x)\mathfrak{m})\}\right) d\mu_\zeta(v), \tag{6.9}$$

and consequently (6.1) can be written in the following form:

$$\varphi(x) = \int_{v\in\mathbb{S}^{m-1}} \mu_{\mathcal{R}}\left(\{r \geq 0 : \exists j = 1, \ldots, \ell \ : r(A_j(x)Lv) \leq (b_j(x) - A_j(x)\mathfrak{m})\}\right) d\mu_\zeta(v). \tag{6.10}$$

There is an immediate keen advantage to representation (6.10) when it comes to numerical evaluation. As observed (e.g., [15, eq. (1.5)]), representation (6.10) allows for a reduction in sample variance when compared directly with sampling from the "native" representation (6.1).

6.2 THE SINGLE POLYHEDRON CASE

In this submitted work, we first look into the case where $\ell = 1$, i.e. a single polyhedron is at hand, before moving on multiple polyhedra. Recall that a polyhedron is an intersection of half spaces, and having $\ell > 0$ amounts to studying unions of intersections.
The probabilistic function equation (6.1) becomes in this case:

$$g(x, z) = A(x)z - b(x),$$

and ξ elliptical symmetric. From preliminary elements of Chapter chapter 5, we know the following subcases are already treated in literature:

- the study of φ' subdifferential around a point $\bar{x}$ that verifies $g(\bar{x}, \mathfrak{m}) < 0$.
- The subcase with constant A and b.
- The subcase with $A(\bar{x})$ of full-rank

As a consequence, our interest lies in the subcase where $g(x, \mathfrak{m}) < 0$ does not hold, which means that the mean point $\mathfrak{m}$ is outside of the feasible set.

6.2.1 Set partitioning and intuition on irregularities

Let us consider the following setting, with a single polyhedron $M(x)$ at hand:

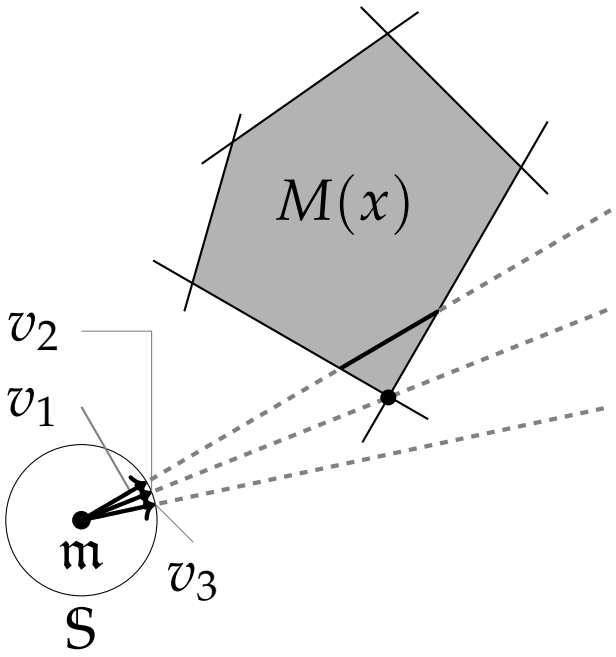


Figure 6.1: Geometric example of the computation of φ as an integral.

When looking at a single row $a_i(\bar{x})$ of $A(\bar{x})$ that defines polyhedron $M(\bar{x})$, several different cases arise:

- the mean vector $\mathfrak{m}$ can be on one side or the other of the half space defined by $a_i(x)$. It cannot, following our assumption that $g(\bar{x}, \mathfrak{m}) < 0$ does not hold, be in the polyhedron defined by the intersection of all half-space $a_i(\bar{x})$.
- a half-ray v of $\mathbb{S}$ can either aim to "exit" the half-space, or oppositely, to "enter" the half space

By continuity, we can define a neighbourhood U of $\bar{x}$ on which $a_i(x)^\mathsf{T}\mathfrak{m} > b_i(x)$ holds true. Following reformulation (6.9) and the geometric interpretation, we study the intersections of $\mathfrak{m} + rLv$ and $M(x)$ on U:

$$\{r \geq 0 \ : g_i(x, \mathfrak{m} + rLv) \leq 0\} = \left[\frac{b_i(x) - a_i(x)^\mathsf{T}\mathfrak{m}}{a_i(x)^\mathsf{T}Lv}, \infty\right), \tag{6.11}$$

for all $v \in \mathbb{S}$ for which $a_i(x)^\mathsf{T}Lv < 0$. For all $v \in \mathbb{S}$, with $a_i(x)^\mathsf{T}Lv \geq 0$, it holds that $\{r \geq 0 \ : g_i(x, \mathfrak{m} + rLv) \leq 0\} = \emptyset$. Now should we be given a neighbourhood U of $\bar{x}$ on which, for $x \in U$, $a_i(x)^\mathsf{T}\mathfrak{m} < b_i(x)$ holds true, then

$$\{r \geq 0 \ : g_i(x, \mathfrak{m} + rLv) \leq 0\} = \left[0, \frac{b_i(x) - a_i(x)^\mathsf{T}\mathfrak{m}}{a_i(x)^\mathsf{T}Lv}\right], \tag{6.12}$$

for all $v \in \mathbb{S}$ for which $a_i(x)^\mathsf{T}Lv > 0$. Finally, for all $v \in \mathbb{S}$, with $a_i(x)^\mathsf{T}Lv \leq 0$, it holds that $\{r \geq 0 \ : g_i(x, \mathfrak{m} + rLv) \leq 0\} = [0, \infty)$.

All those facts are immediately checked through standard inequality simplifications.

Let us lay down formally our trial point of interest:

Assumption 2. Let $\bar{x} \in \mathbb{R}^n$ be given such that $a_i(\bar{x})^\mathsf{T}\mathfrak{m} \neq b_i(\bar{x})$ for all $i = 1, \ldots, p$.

Let us thus distinguish and separate the index set $\{1, \ldots, p\} = I^+ \cup I^-$ as well as fix the neighbourhood U of $\bar{x}$ with $a_i(x)^\mathsf{T}\mathfrak{m} < b_i(x)$ for all $i \in I^+$ and $x \in U$; $a_i(x)^\mathsf{T}\mathfrak{m} > b_i(x)$ for all $i \in I^-$ and $x \in U$. As already argued, we will assume the existence of $i = 1, \ldots, p$ such that $a_i(x)^\mathsf{T}\mathfrak{m} > b_i(x)$ holds on an appropriate neighbourhood U of the point of interest $\bar{x}$. This condition translates as I^- being not empty. It is still however possible for I^+ to be empty. Observe that we rule out that $\bar{x}$ is such that $a_i(\bar{x})^\mathsf{T}\mathfrak{m} = b_i(\bar{x})$ for some $i = 1, \ldots, p$. This condition, combined with definition of U, that requires the whole neighbourhood to verify strict inequalities, ensures the local stability with respect to x of our index sets I^- and I^+.

Following the geometric interpretation, one could notice that I^- corresponds to the set of hyperplane indices that (strictly) do not include mean vector $\mathfrak{m}$ for $x \in U$; symmetrically, I^+ corresponds to hyperplane indices that strictly include $\mathfrak{m}$. Additionally, one could see that the intersection of the closed half-line emanating from $\mathfrak{m}$ and of direction v and the hyperplane defined by row i of A and b is intrinsically related to the sign of $a_i(x)^\mathsf{T}Lv$.

These observations will allow us to partition the set $U \times \mathbb{S}$ usefully, with the help of the following sets:

Definition 12. Let $\bar{x} \in \mathbb{R}^n$ be given as by Assumption 2 and let U be the neighbourhood

of $\bar{x}$ on which the index sets I^-, I^+ are stable. Define the following subsets of $U \times \mathbb{S}$:

$$B^- = \left\{(x,v) \in U \times \mathbb{S} \ : a_i(x)^\mathsf{T} L v < 0, \forall i \in I^-\right\}, \tag{6.13a}$$

$$\mathcal{O}^- = \left\{(x,v) \in U \times \mathbb{S} \ : \ a_i(x)^\mathsf{T} L v \leq 0, \forall i \in I^- \ \wedge \ \exists i \in I^- \mid a_i(x)^\mathsf{T} L v = 0\right\} \tag{6.13b}$$

$$A^- = \left\{(x,v) \in U \times \mathbb{S} \ : \exists i \in I^- \mid a_i(x)^\mathsf{T} L v > 0\right\}, \tag{6.13c}$$

$$B^+ = \left\{(x,v) \in B^- \ : a_i(x)^\mathsf{T} L v < 0, \forall i \in I^+\right\} \tag{6.13d}$$

$$\mathcal{O}^+ = \left\{(x,v) \in B^- \ : \ a_i(x)^\mathsf{T} L v \leq 0, \forall i \in I^+ \ \wedge \ \exists i \in I^+ \mid a_i(x)^\mathsf{T} L v = 0\right\} \tag{6.13e}$$

$$A^+ = \left\{(x,v) \in B^- \ : \exists i \in I^+ \mid a_i(x)^\mathsf{T} L v > 0\right\} \tag{6.13f}$$

•

The intuition behind these set definitions is partly inspired by previous works in literature (see for instance [4]). The rationale is to be able to distinguish sets where the candidate mapping for behind an estimation of φ's subdifferential changes behaviour.

One can immediately derive the following observations:

- A^+, B^-, B^+ are open sets;
- $\mathcal{O}^-, \mathcal{O}^+$ are closed sets;

The following lemma shows the structure of Definition 12 and backs up Figure 6.2.

Lemma 5. With notation as in Definition 12, it holds that the sets $B^+, \mathcal{O}^+, A^+$ form a partition of B^- (i.e., are mutually disjoint and their union makes up the whole set). Furthermore the sets $B^+, \mathcal{O}^+, A^+, \mathcal{O}^-, A^-$ form a partition of $U \times \mathbb{S}$. ◀

For any $i = 1, \ldots, p$, we can now define $\mathfrak{r}_i \colon B^- \to \mathbb{R} \cup \{-\infty, \infty\}$ as

$$\mathfrak{r}_i(x,v) = \frac{b_i(x) - a_i(x)^\mathsf{T} \mathfrak{m}}{a_i(x)^\mathsf{T} L v}, \tag{6.14}$$

where division "by zero" is interpreted as leading to $\pm\infty$. Our first observation is that for $i \in I^-$, and $(x,v) \in B^-$, $\mathfrak{r}_i(x,v) \in (0,\infty)$ always. Furthermore, since B^- is open, $\mathfrak{r}_i$ is evidently continuously differentiable on this set. For any $i \in I^+$ and any $(x,v) \in A^+ \cup B^+$, we observe that $\mathfrak{r}_i(x,v)$ is finite valued. Moreover $\mathfrak{r}_i$ is continuously differentiable on $A^+ \cup B^+$, the latter being an open set.

With the help of these mappings, we now define $r_1 \colon B^- \to \mathbb{R}_+$ as

$$r_1(x,v) = \max_{i \in I^-} \mathfrak{r}_i(x,v) = \max_{i \in I^-} \frac{b_i(x) - a_i(x)^\mathsf{T} \mathfrak{m}}{a_i(x)^\mathsf{T} L v} > 0. \tag{6.15}$$

Since r_1 is the maximum of continuously differentiable maps on the open set B^-, it is locally Lipschitz there. Figure 6.1 shows how at a specific point (x, v_1), $r_1(x, v_1) = \mathfrak{r}_2(x, v_1)$, $\mathfrak{r}_1(x, v_1)$ being strictly smaller at this (x, v_1).

We can also define $r_2 \colon B^- \to \mathbb{R}_+ \cup \{\infty\}$ as follows[1]:

$$r_2(x,v) = \begin{cases} \sup_t & t \\ \text{s.t.} & t a_i(x)^\mathsf{T} L v \leq b_i(x) - a_i(x)^\mathsf{T} \mathfrak{m}, i \in I^+ \end{cases}$$

[1] One could observe the ressemblance with the ray-function as defined in (5.19)

It now follows that $\mathcal{D}\mathrm{om}(r_2) = A^+$ and with $\mathcal{I}(x,v) = \{i \in I^+ \ : a_i(x)^\mathsf{T} L v > 0\}$ (being locally stable), it follows that $r_2(x,v) = \min_{i\in\mathcal{I}(x,v)} \mathfrak{r}_i(x,v) = \min_{i\in\mathcal{I}(x,v)} \frac{b_i(x)-a_i(x)^\mathsf{T}\mathfrak{m}}{a_i(x)^\mathsf{T} L v} > 0$ and this mapping is once more locally Lipschitzian on its domain (being open). This follows from [14, Lemma 3.1], when considering a reduced inequality system to the index set I^+.

From a practical point of view, one could have the following interpretations:

- at a given $(x,v) \in B^-$, $r_1(x,v)$ is the largest distance $r \geq 0$ on the half-line from $\mathfrak{m}$ in direction v before reaching the intersection of all hyperplanes defined by rows of I^-.
- At a given $(x,v) \in A^+$, $r_2(x,v)$ is the smallest distance $r \geq 0$ on the half-line from $\mathfrak{m}$ in direction v before leaving any hyperplane defined by rows of I^+.

Now it is readily visible in (6.10) that the value of the integrand depends on the order between $r_1(x,v)$ and $r_2(x,v)$. Depicted in figure 6.1 is the particular case where $r_1(x,v) < r_2(x,v)$ but obviously this does not hold in general. We may therefore partition further the set A^+ as follows:

$$A^{+-} = \{(x,v) \in A^+ \ : r_1(x,v) > r_2(x,v)\} \tag{6.16a}$$
$$A^{+0} = \{(x,v) \in A^+ \ : r_1(x,v) = r_2(x,v)\} \tag{6.16b}$$
$$A^{++} = \{(x,v) \in A^+ \ : r_1(x,v) < r_2(x,v)\} \tag{6.16c}$$
$$\tag{6.16d}$$

and it follows, from the previous observations, that A^{+-}, A^{++} are both open (since r_1, r_2 are continuous), and A^{+0} is closed. Based on this partition, we will now provide an explicit representation of the integrand occurring in the spherical radial representation (6.9) that will be called e.

Figure 6.2 provides a visualization of these index sets.

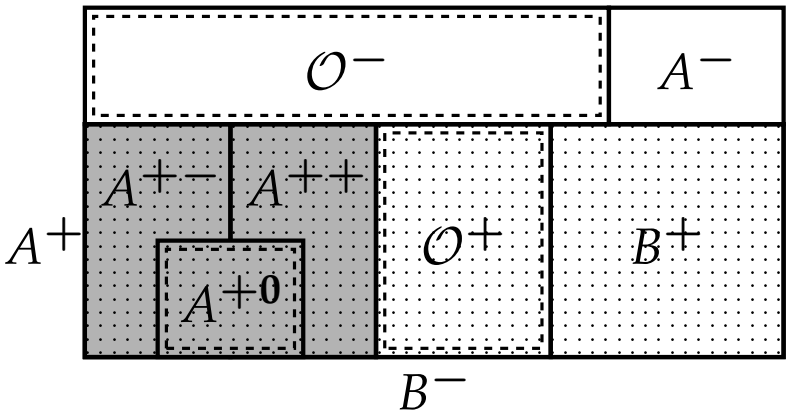


FIGURE 6.2: Illustration of the partition of $U \times \mathbb{S}$. The dotted set is B^-, while A^+ is the darker set. Sets with dashed inner linings are closed, the other ones being open.

6.2.2 Studying and representing the inner radial probability function e

We now define the mapping $e\colon U \times \mathbb{S} \to [0,1]$ as

$$e(x,v) = \mu_{\mathcal{R}}\left(\{r \geq 0 : r(A(x)Lv) \leq (b(x) - A(x)\mathfrak{m})\}\right), \tag{6.17}$$

and recall that $\phi(x) = \int_{v\in\mathbb{S}} e(x,v)d\mu_\zeta(v)$ as a result of (6.9).

Due to the preparatory material above, we can now provide a full partition-wise description of e:

Lemma 6. The mapping $e\colon U \times \mathbb{S} \to [0,1]$ specified in (6.17) is also given by:

$$e(x,v) = \begin{cases} 0 & \text{if} \quad (x,v) \in A^- \cup \mathcal{O}^- \cup A^{+-} \cup A^{+0} \\ 1 - F_{\mathcal{R}}(r_1(x,v)) & \text{if} \quad (x,v) \in B^+ \cup \mathcal{O}^+ \\ F_{\mathcal{R}}(r_2(x,v)) - F_{\mathcal{R}}(r_1(x,v)) & \text{if} \quad (x,v) \in A^{++} \end{cases}. \tag{6.18}$$

◀

This is where the set partition is necessary: one could have the naïve interpretation that it would be desirable to discard the cases where $r_1(x,v) \geq r_2(x,v)$. In fact, when this holds, there is no contribution of half-ray v to the integral.

Now we have to properly define e on the entire set $U \times \mathbb{S}$ and prove its regularity. We will need to study limits at the boundaries of every subset, and use a *gluing* Lemma in the following section that enables us to *unify* every local behaviour into a correctly defined mapping.

6.2.3 Useful partition-wise limits

The main focus of the lemmas in this section is to characterize the asymptotic behaviour of the sub-differential of e over certain subsets of our partition. These lemmas will be useful to go from an initial subset-wise study to the unified result over $U \times \mathbb{S}$. In particular, asymptotic studies are necessary along sequences from within an open subset of our partition to a neighbouring closed subset. Difficulties arise from the fact mapping r_1 and r_2 can, in principle, reach arbitrarily large values: we will need to carefully study the behaviour of "their" sub-differentials under such limits. This will be the topic of the current section. First, we will concentrate on r_2, for which the result follows relatively analogously to earlier established results.

Lemma 7. Let $f_{\mathcal{R}}$ be compatible with a 1st order polynomial growth condition (see Definition 11).

- Let $(x_k, v_k) \in A^+$ be a sequence such that $(x_k, v_k) \to (\bar{x}, \bar{v}) \in \mathcal{O}^+$, then it holds that

  i) $r_2(x_k, v_k) \to \infty$

  ii)
  $$\lim_{k\to\infty} f_{\mathcal{R}}(r_2(x_k, v_k))\partial_x^{\mathsf{C}} r_2(x_k, v_k) = \{0\},$$
  where the last limit is to be understood in the Painlevé-Kuratowski sense.

- Now let $(x_k, v_k) \in B^-$ be a sequence such that $(x_k, v_k) \to (\bar{x}, \bar{v}) \in \mathcal{O}^-$, then it holds that
  - $r_1(x_k, v_k) \to \infty$
  - $$\lim_{k\to\infty} f_{\mathcal{R}}(r_1(x_k, v_k))\partial_x^{\mathsf{C}} r_1(x_k, v_k) = \{0\},$$
    where the last limit is to be understood in the Painlevé-Kuratowski sense.
- Finally There can not exist any sequence $(x_k, v_k) \in A^{+0} \cup A^{+-}$ with limit $(\bar{x}, \bar{v}) \in \mathcal{O}^+$ (likewise with B^+).

◀

This lemma is useful to define the behaviour of e, and more precisely its subdifferential, at the boundaries of some defined subsets. As such, we can slightly update our Figure 6.2.

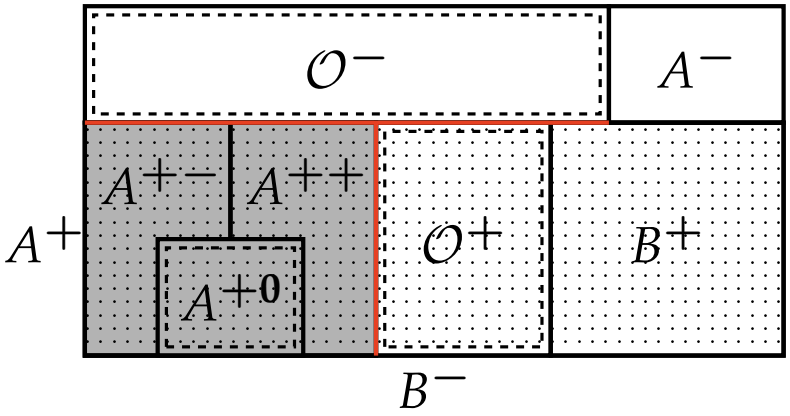


FIGURE 6.3: Illustration of the partition of $U \times \mathbb{S}$. Orange borders are those for which we have asymptotic information on r_1 and r_2. Observe that the sets are properly placed, in the sense that no sequence can originate from $A^{+-} \cup A^{+0}$ and reach $\mathcal{O}^+$ nor B^+. From the set definition we also already know that there cannot exist sequence originating from A^- converging to A^+.

6.2.4 On the partial subgradients of e

Now that some of the preparatory material has been laid out, we turn our attention to the identification of the sub-differential of e. First, on part of the set $U \times \mathbb{S}$, the partial sub-differential of e with respect to x is easily identified. Let us recall that local Lipschitz property ensures the existence and non-emptyness of Clarks's subdifferentials. For brevity, we omit the proofs that in general are made of extensive use of Clarke's rules on differentiation in [82], as the chain rule of Theorem 2.3.9(ii). This is the cornerstone of the following Lemmas.

Lemma 8. The mapping e, given in (6.17), is locally Lipschitzian on the open set $A^- \cup B^+ \cup A^{++} \cup A^{+-}$ and has partial Clarke derivative, w.r.t. x satisfying:

$$\partial_x^{\mathrm{C}} e(x,v) \subseteq \begin{cases} \{0\} & \text{if} \quad (x,v) \in A^{+-} \cup A^- \\ -f_{\mathcal{R}}(r_1(x,v))\partial_x^{\mathrm{C}} r_1(x,v) & \text{if} \quad (x,v) \in B^+ \\ f_{\mathcal{R}}(r_2(x,v))\partial_x^{\mathrm{C}} r_2(x,v) - f_{\mathcal{R}}(r_1(x,v))\partial_x^{\mathrm{C}} r_1(x,v) & \text{if} \quad (x,v) \in A^{++} \end{cases} \tag{6.19}$$

◀

As a final observation, we note that the generalized derivatives of r_1 and r_2, as finite extrema are also immediately identified, and we will not make them fully explicit.

Our next idea is to identify a set-valued mapping which is a good candidate for "being" $\partial_x^{\mathrm{C}} e$. Observe however that for the time being we have not yet established that e is locally Lipschitz on the whole $U \times \mathbb{S}$. We therefore define the map $e_x : U \times \mathbb{S} \rightrightarrows \mathbb{R}^n$ as follows

$$e_x(x,v) = \begin{cases} \{0\} & \text{if} \quad (x,v) \in A^{+-} \cup \mathcal{O}^- \cup A^- \\ -f_{\mathcal{R}}(r_1(x,v))\partial_x^{\mathrm{C}} r_1(x,v) & \text{if} \quad (x,v) \in \mathcal{O}^+ \cup B^+ \\ \mathrm{Co}(\{0\} \cup f_{\mathcal{R}}(r_2(x,v))\partial_x^{\mathrm{C}} r_2(x,v) - & \\ \quad f_{\mathcal{R}}(r_1(x,v))\partial_x^{\mathrm{C}} r_1(x,v)) & \text{if} \quad (x,v) \in A^{+0} \\ f_{\mathcal{R}}(r_2(x,v))\partial_x^{\mathrm{C}} r_2(x,v) - & \\ \quad f_{\mathcal{R}}(r_1(x,v))\partial_x^{\mathrm{C}} r_1(x,v) & \text{if} \quad (x,v) \in A^{++} \end{cases} \tag{6.20}$$

and observe that this map coincides with the upper-estimation found in (8) over the open sets on which e was shown to be locally Lipschitz (recall Lemma 8). Our first endeavour will be to establish that e_x is outer semi-continuous (o.s.c.). To this end, we require the following result:

Lemma 9. Let $f_{\mathcal{R}}$ be compatible with a 1st order polynomial growth condition. Along any sequence $(x_k, v_k) \in A^{++} \cup \mathcal{O}^+ \cup B^+$ converging to $(\bar{x}, \bar{v}) \in \mathcal{O}^-$ it holds that $\lim_{k\to\infty} e_x(x_k, v_k) = \{0\}$, where the last limit is to be understood in the Painlevé-Kuratowski sense. ◀

Lemma 10. Let $f_{\mathcal{R}}$ be compatible with a 1st order polynomial growth condition. Along any sequence $(x_k, v_k) \in A^{++} \cup A^{+-}$ converging to $(\bar{x}, \bar{v}) \in A^{+0}$ it holds that

$$\limsup_{k\to\infty} e_x(x_k, v_k) \subseteq e_x(\bar{x}, \bar{v})$$

. ◀

We can now gather all are findings into the following proposition:

Proposition 7. Let $f_{\mathcal{R}}$ be compatible with a 1st order polynomial growth condition. Then the mapping e_x defined in (6.20) is outer semi-continuous on $U \times \mathbb{S}$ and locally bounded. ◀

6.2.5 Gluing pieces together: the main result

In order to show that e is locally Lipschitzian on $U \times \mathbb{S}$, we require a technical result that allows us to bridge, the closed sets that interspace A^-, B^-, A^{++} and A^{+-}. We thus need

to glue together in a way, the various open sets.

Lemma 11 (Gluing lemma)**.** Let $U \subseteq \mathbb{R}^n$ be an open set, U be partitioned into U' open and $\mathcal{O}$ (closed). Assume given the following objects:

i) $f\colon U \to [0,1]$ a continuous mapping that is locally Lipschitzian on U', constant on $\mathcal{O}$;

ii) $\partial f\colon U \to \mathbb{R}^n$ a set-valued mapping such that $\partial f = \partial^{\mathrm{c}} f$ on U' that is moreover outer semi-continuous (o.s.c.) and locally bounded.

Then f is locally Lipschitzian on U and $\partial^{\mathrm{c}} f \subseteq \mathrm{Co}\partial f$ if ∂f contains 0 on $\mathcal{O}$ and f is extremal on $\mathcal{O}$ (i.e., takes the values 0 or 1). ◀

We can now put to use all these results to establish that the mapping e itself, is locally Lipschitz, on the whole set $U \times \mathbb{S}$.

Proposition 8. Let $f_{\mathcal{R}}$ be compatible with a 1st order polynomial growth condition. The mapping $e\colon U \times \mathbb{S} \to [0,1]$ defined in (6.18) is locally Lipschitzian. Moreover, for any $(x,v) \in U \times \mathbb{S}$, the following inclusion holds $\partial_x^{\mathrm{c}} e(x,v) \subseteq e_x(x,v)$, where e_x is as in (6.20). ◀

We can now gather all previously established material and show that the probability function ϕ itself is locally Lipschitzian. We can also obtain an outer estimate of its Clarke sub-differential:

Theorem 12. Let $A\colon \mathbb{R}^n \to \mathbb{R}^{p\times m}$ be a continuously differentiable matrix valued map, and let $b\colon \mathbb{R}^n \to \mathbb{R}^p$ be continuously differentiable. Let the random vector $\xi \in \mathbb{R}^m$ be elliptically symmetric with mean $\mathfrak{m}$, covariance-like matrix Σ and generator θ. Let ξ be compatible with the 1st order polynomial growth condition.

Let $g\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ be defined as $g(x,z) = A(x)z - b(x)$. Let $\bar{x}$ and the neighbourhood U of $\bar{x}$ be such that $g_i(x,\mathfrak{m}) \neq 0$ for all $i = 1,\ldots,p$ and all $x \in U$. Then, the probability function $\phi\colon \mathbb{R}^n \to [0,1]$ defined as

$$\phi(x) = \mathbb{P}\left[A(x)\xi \leq b(x)\right], \tag{6.21}$$

is locally Lipschitz on U and has Clarke sub-differential satisfying:

$$\partial^{\mathrm{c}}\phi(x) \subseteq \int_{v\in\mathbb{S}} e_x(x,v)d\mu_\zeta(v), \tag{6.22}$$

where μ_ζ is the uniform measure over the sphere $\mathbb{S} = \left\{z \in \mathbb{R}^m \ : \|z\|^2 = 1\right\}$, and e_x is as in (6.20). ◀

6.3 THE GENERAL CASE: UNION OF POLYHEDRA

We can now turn our attention to the case of probability functions given in the form of (6.1), and drop the assumption $\ell = 1$ we had in section section 6.2. Let us first justify the interchange of sub-differentiation and integration, prior to giving formulæ for the sub-differential of the probability function.

Theorem 13. Consider the probability function (6.1), wherein the random vector $\xi \in \mathbb{R}^m$ is elliptically symmetric with mean $\mathfrak{m}$, covariance-like matrix Σ and generator θ. For each $j = 1, \ldots, \ell$, let $A_j \colon \mathbb{R}^n \to \mathbb{R}^{p_j \times m}$ be a continuously differentiable matrix valued map, and let $b_j \colon \mathbb{R}^n \to \mathbb{R}^{p_j}$ be continuously differentiable. Assume moreover that ξ is compatible with the first order polynomial growth condition (see Definition 11). Let $\bar{x} \in \mathbb{R}^n$ be given along with a neighbourhood of $\bar{x}$ such that $a_{ij}(x)^{\mathsf{T}}\mathfrak{m} - b_{ij}(x) \neq 0$ for all $i = 1, \ldots, p_j$, $j = 1, \ldots, \ell$ and all $x \in U$ holds true. Here a_{ij} denotes the ith row of matrix A_j. Then φ is (locally) Lipschitz on U and

$$\partial^{\mathrm{c}}\varphi(x) \subseteq \int_{v \in \mathbb{S}} \partial_x^{\mathrm{c}} e(x, v) d\mu_\zeta(v), \tag{6.23}$$

where μ_ζ is the uniform measure over the sphere $\mathbb{S} = \left\{ z \in \mathbb{R}^m \ : \|z\|^2 = 1 \right\}$, and $e \colon U \times \mathbb{S} \to [0, 1]$ is also locally Lipschitz and given by:

$$e(x, v) = \mu_{\mathcal{R}}\left(\left\{ r \geq 0 \ : \mathfrak{m} + rLv \in \bigcup_{j=1}^{\ell} M_j(x) \right\} \right). \tag{6.24}$$

◀

Remark 13 (On subdifferential inclusion)**.** If we make the additional assumption that the mapping e is regular at x (see [82, Definition 2.4.10]), then the inclusion (6.23) is in fact an equality at x (see [82, Theorem 2.7.2]). ▷

The issue with the expanded form of e, through the inclusion-exclusion formula, is that it does not help produce an efficient estimate of $\partial_x^{\mathrm{c}} e(x, v)$. We will illustrate this in Example 5 below. First however we will introduce notation that will help us provide a more elegant representation of e and $\partial_x^{\mathrm{c}} e(x, v)$. To this end, let us introduce the set

$$\mathcal{O}_{ij} = \left\{ (x, v) \in U \times \mathbb{S} \ : a_{ij}(x)^{\mathsf{T}} L v = 0 \right\}, \tag{6.25}$$

for $i = 1, \ldots, p_j$ and $j = 1, \ldots, \ell$ and $\mathcal{O} = \bigcup_{j=1}^{\ell} \cup_{i=1}^{p_j} \mathcal{O}_{ij}$ as well as $\mathcal{Z} = U \times \mathbb{S} \setminus \mathcal{O}$. Then we can observe that the sets $\mathcal{O}_{ij}$ are all of zero measure (the latter fact follows from [18, Lemma 2.2]. Consequently $\mathcal{O}$ is also of zero measure (in $\lambda \otimes \mu_\zeta$).

Next for each $j = 1, \ldots, \ell$ we define the following sets:

$$\begin{aligned}
I_j^- &= \left\{ i = 1, \ldots, p_j \ : a_{ij}(x)^{\mathsf{T}}\mathfrak{m} > b_{ij}(x), \forall x \in U \right\} \\
I_j^+ &= \left\{ i = 1, \ldots, p_j \ : a_{ij}(x)^{\mathsf{T}}\mathfrak{m} < b_{ij}(x), \forall x \in U \right\} \\
B_j^- &= \left\{ (x, v) \in \mathcal{Z} \ : a_{ij}(x)^{\mathsf{T}} L v < 0, \forall i \in I_j^- \right\} \\
A_j^- &= \left\{ (x, v) \in \mathcal{Z} \ : \exists i \in I_j^- \ : a_{ij}(x)^{\mathsf{T}} L v > 0 \right\},
\end{aligned}$$

where implicitly it is assumed that U is sufficiently small, for the index sets I_j^- and I_j^+ to form a partition of $\{1, \ldots, p_j\}$. Observe that this can be achieved as a result of continuity and the standing assumption at $\bar{x}$.

Next we may observe that $B_j^- \cup A_j^- = \mathcal{Z}$, and that both B_j^- and A_j^- are open sets. Moreover, should $I_j^- = \emptyset$, then simply $B_j^- = \mathcal{Z}$. Moreover define for all $i = 1, \ldots, p_j$,

$j = 1, \dots, \ell$, the map $\mathfrak{r}_{ij} \colon \mathcal{Z} \to \mathbb{R}$ as

$$\mathfrak{r}_{ij}(x, v) = \frac{b_{ij}(x) - a_{ij}(x)^{\mathsf{T}}\mathfrak{m}}{a_{ij}(x)^{\mathsf{T}} L v} \tag{6.26}$$

and observe that this map is well defined and continuously differentiable on $\mathcal{Z}$. Now for each $j = 1, \dots, \ell$, we define $r_1^j \colon B_j^- \to \mathbb{R}_+$ as

$$r_1^j(x, v) = \begin{cases} 0 & \text{if} \quad I_j^- = \emptyset. \\ \max_{i \in I_j^-} \mathfrak{r}_{ij}(x, v) & \text{otherwise} \end{cases}$$

and we define $r_2^j \colon B_j^- \to \mathbb{R}_+ \cup \{\infty\}$ as

$$r_2^j(x, v) = \begin{cases} \sup_t & t \\ \text{s.t.} & t a_{ij}(x)^{\mathsf{T}} L v \leq b_{ij}(x) - a_{ij}(x)^{\mathsf{T}}\mathfrak{m}, i \in I_j^+ \end{cases}.$$

Now at any $(x, v) \in \mathcal{Z}$, let us define the index set $F(x, v) = \left\{ j = 1, \dots, \ell \ : (x, v) \in B_j^- \right\}$. Then, we can observe that F is locally stable since B_j^- is open and that

$$e(x, v) = \mu_{\mathcal{R}} \left(\bigcup_{j \in F(x,v)} [r_1^j(x, v), \max \left\{ r_2^j(x, v), r_1^j(x, v) \right\}] \right), \tag{6.27}$$

since $\left\{ r \geq 0 \ : \mathfrak{m} + rLv \in M_j(x) \right\} = \emptyset$ if $(x, v) \in A_j^-$. In order to further simplify the formula, let us introduce for any $(x, v) \in \mathcal{Z}$, the index set $F'(x, v)$, defined as follows:

$$F'(x, v) = \left\{ j \in F(x, v) \ : r_2^j(x, v) \geq r_1^j(x, v) \right\}.$$

Hence, formula (6.27) becomes:

$$e(x, v) = \mu_{\mathcal{R}} \left(\bigcup_{j \in F'(x,v)} [r_1^j(x, v), r_2^j(x, v)] \right). \tag{6.28}$$

We have already hinted on the potential issue, in representing economically $\partial_x^{\mathsf{c}} e(x, v)$ when employing the inclusion-exclusion formula. The problem is fully illustrated in the next example:

Example 5. We consider a trial point $x \in \mathbb{R}^n$ and two polyhedra $M_1(x)$ and $M_2(x)$, as well as a direction v at which we have $r_1^1(x, v) < r_1^2(x, v)$, $r_2^1(x, v) < r_2^2(x, v)$ and $r_1^2(x, v) < r_2^1(x, v)$.

The inclusion-exclusion based formula from Theorem 13 yields $e(x, v) = F_{\mathcal{R}}(r_2^2(x, v)) - F_{\mathcal{R}}(r_1^2(x, v)) + F_{\mathcal{R}}(r_2^1(x, v)) - F_{\mathcal{R}}(r_1^1(x, v)) - (F_{\mathcal{R}}(r_2^1(x, v)) - F_{\mathcal{R}}(r_1^2(x, v)))$, which in turn provides a non-economic representation of its sub-differential:

$$\begin{aligned} \partial_x^{\mathsf{c}} e(x, v) \subseteq [& f_{\mathcal{R}}(r_2^2(x, v)) \partial_x^{\mathsf{c}} r_2^2(x, v) - f_{\mathcal{R}}(r_1^2(x, v)) \partial_x^{\mathsf{c}} r_1^2(x, v) + f_{\mathcal{R}}(r_2^1(x, v)) \partial_x^{\mathsf{c}} r_2^1(x, v) - \\ & f_{\mathcal{R}}(r_1^1(x, v)) \partial_x^{\mathsf{c}} r_1^1(x, v) - (f_{\mathcal{R}}(r_2^1(x, v)) \partial_x^{\mathsf{c}} r_2^1(x, v) - f_{\mathcal{R}}(r_1^2(x, v)) \partial_x^{\mathsf{c}} r_1^2(x, v))]. \end{aligned}$$

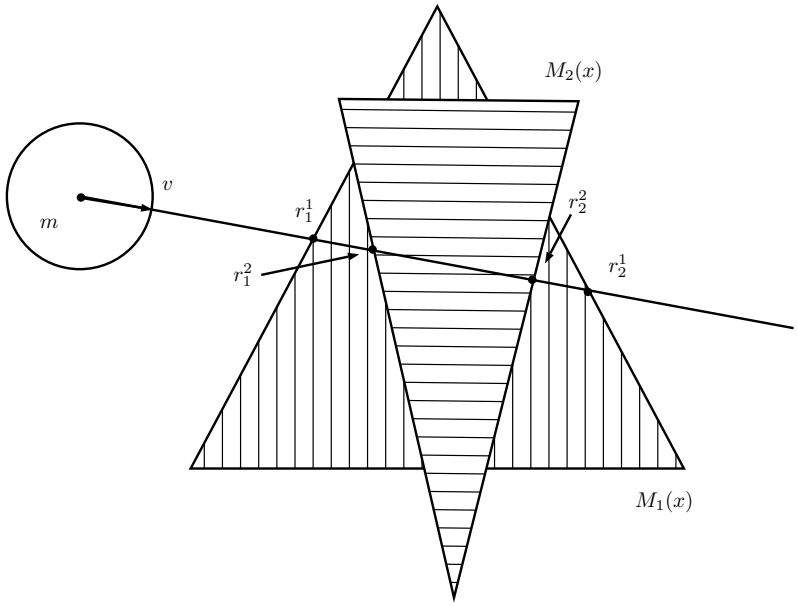


FIGURE 6.4: Example where formula from Theorem 13 fails to provide an economic representation of its sub-differential.

One can of course observe that in most cases, (6.28), through potentially tedious computations, will allow us to identify swiftly the most economic representation of e. One would simply have to work out, what the union of intervals amounts to and, unless some local degeneracy occurs, this will immediately provide the appropriate estimate of $\partial_x^{\mathrm{c}} e(x,v)$. Degeneracy will occur as soon as any two interval bounds match and locally "cross". The subsequent result provides a unified formula, rendering this interval calculus unnecessary in the above situation. We will also illustrate, that typically, an efficient estimate of the sub-differential is obtained. Let us just briefly mention, prior to stating the result, that $F_{\mathcal{R}}$ is increasing in increasing arguments. Therefore, minimum/maximum operations may be pulled through $F_{\mathcal{R}}$ at will.

Proposition 9. At any $(x,v) \in \mathcal{Z}$, the following identity holds true:

$$e(x,v) = \max\left\{F_{\mathcal{R}}(\max_{j\in F'(x,v)} r_2^j(x,v)) - F_{\mathcal{R}}(\min_{j\in F'(x,v)} r_1^j(x,v)), 0\right\} - \sum_{S,S'\subseteq F'(x,v), S\cap S'=\emptyset, S\cup S'=F'(x,v), S,S'\neq\emptyset} \max\left\{F_{\mathcal{R}}(\min_{i\in S'} r_1^i(x,v)) - F_{\mathcal{R}}(\max_{i\in S} r_2^i(x,v)), 0\right\}, \tag{6.29}$$

where $F_{\mathcal{R}}(\infty) = 1$ is to be understood. ◀

Proof. The proof will be performed by induction. To this end, let us fix an arbitrary $(\bar{x},\bar{v}) \in \mathcal{Z}$, and first of all observe that on an appropriate neighbourhood of $(\bar{x},\bar{v})$, $F(\bar{x},\bar{v})$ remains unchanged. We may thus, without loss of generality assume that $F(\bar{x},\bar{v}) = \{1,\dots,\ell\}$.

When $\ell = 1$, the stipulated formula is immediately seen to be identical to the formula of (6.18).

Let us thus assume that the formula is valid for a given $\ell' \geq 1$, we will then show that the formula is valid for $\ell'+1$. First of all, we may assume that for all $j = 1,\dots,\ell'+1$, it

holds that $r_2^j(\bar{x},\bar{v}) \geq r_1^j(\bar{x},\bar{v})$. Indeed, should this not be the case, i.e., should there exist $j = 1,\dots,\ell'+1$ for which $r_2^j(\bar{x},\bar{v}) < r_1^j(\bar{x},\bar{v})$ holds true, then as a result of continuity, this continues to hold true on an appropriate neighbourhood of $(\bar{x},\bar{v})$, therefore, for (x,v), near $(\bar{x},\bar{v})$, $j \notin F'(x,v)$. Re-arranging indices if needed, we may assume $j = \ell'+1$ and hence locally,

$$e(x,v) = \mu_{\mathcal{R}}(\left\{ r \geq 0 \ : \mathfrak{m} + rLv \in \bigcup_{j=1}^{\ell'} M_j(x) \right\}).$$

We may thus use the induction assumption to obtain the formula for the case ℓ'. Since $\ell'+1 \notin F(x,v)$ for (x,v) near $(\bar{x},\bar{v})$, the formulae (6.29) with ℓ' and $\ell'+1$ do indeed coincide.

From now on we will thus assume that $F'(x,v) = \{1,\dots,\ell'\}$.

Next, let us assume the existence of $k = 1,\dots,\ell'$ such that $r_1^k(\bar{x},\bar{v}) \leq r_1^{\ell'+1}(\bar{x},\bar{v}) \leq r_2^{\ell'+1}(\bar{x},\bar{v}) \leq r_2^k(\bar{x},\bar{v})$. We may employ the induction assumption to derive the formula (6.29), since indeed (6.28) holds true also when eliminating the $\ell'+1$th term. It is moreover clear that:

$$\begin{aligned}
\max_{j\in F'(x,v)} F_{\mathcal{R}}(r_2^j(x,v)) &= \max_{j\in F'(x,v)\cup\{\ell'+1\}} F_{\mathcal{R}}(r_2^j(x,v)) \\
\min_{j\in F'(x,v)} F_{\mathcal{R}}(r_1^j(x,v)) &= \min_{j\in F'(x,v)\cup\{\ell'+1\}} F_{\mathcal{R}}(r_1^j(x,v)).
\end{aligned}$$

Let us now pick an arbitrary but fixed partition S', S of $F'(x,v)\cup\{\ell'+1\}$, then, an exhaustive case distinction yields the identity:

$$\begin{aligned}
&\max\left[F_{\mathcal{R}}\left(\min_{i\in S'} r_1^i(\bar{x},\bar{v})\right) - F_{\mathcal{R}}\left(\max_{i\in S} r_2^i(\bar{x},\bar{v})\right), 0\right] = \\
&= \begin{cases}
\max\left[F_{\mathcal{R}}\left(\min\limits_{i\in S'\setminus\{\ell'+1\}} r_1^i(\bar{x},\bar{v})\right) - F_{\mathcal{R}}\left(\max\limits_{i\in S} r_2^i(\bar{x},\bar{v})\right), 0\right] & \text{if } k, \ell'+1 \in S' \\
\max\left[F_{\mathcal{R}}\left(\min\limits_{i\in S'} r_1^i(\bar{x},\bar{v})\right) - F_{\mathcal{R}}\left(\max\limits_{i\in S\setminus\{\ell'+1\}} r_2^i(\bar{x},\bar{v})\right), 0\right] & \text{if } k, \ell'+1 \in S \\
0 & \text{if } k \in S \text{ and } \ell'+1 \in S' \\
0 & \text{if } \ell'+1 \in S \text{ and } k \in S'
\end{cases}
\end{aligned}$$

and thus consequently, the sum over all partitions in (6.29) is unaltered when moving from partitions of $F'(\bar{x},\bar{v})$ to all partitions of $F'(\bar{x},\bar{v})\cup\{\ell'+1\}$. We have thus shown the formula to extend to $\ell'+1$ in this particular situation.

For the remainder we may, upon reordering the indices if needed, assume that $r_1^{\ell'+1}(\bar{x},\bar{v}) \leq r_1^i(\bar{x},\bar{v})$ for all $i \in F'(\bar{x},\bar{v})$. Now two further cases appear

- The situation wherein $r_2^{\ell'+1}(\bar{x},\bar{v}) < r_1^i(\bar{x},\bar{v})$ for all $i \in F'(\bar{x},\bar{v})$
- Or the existence of $i \in F'(\bar{x},\bar{v})$ such that $r_1^i(\bar{x},\bar{v}) \leq r_2^{\ell'+1}(\bar{x},\bar{v})$.

The first case corresponds to adding a disjoint interval to the union of intervals. It thus immediately follows that:

$$e(\bar{x},\bar{v}) = F_{\mathcal{R}}(r_2^{\ell'+1}(\bar{x},\bar{v})) - F_{\mathcal{R}}(r_1^{\ell'+1}(\bar{x},\bar{v}))+ \max\left\{\max_{j\in F'(x,v)} F_{\mathcal{R}}(r_2^j(x,v)) - \min_{j\in F'(x,v)} F_{\mathcal{R}}(r_1^j(x,v)),0\right\} - \sum_{S,S'\subseteq F'(x,v),S\cap S'=\emptyset,S\cup S'=F'(x,v),S,S'\neq\emptyset} \max\left\{F_{\mathcal{R}}(\min_{i\in S'} r_1^i(x,v)) - F_{\mathcal{R}}(\max_{i\in S} r_2^i(x,v)),0\right\},$$

upon employing the indication hypothesis for the ℓ' remaining intervals. Observe now that on $F'(\bar{x},\bar{v})$, $\max_{j\in F'(x,v)} F_{\mathcal{R}}(r_2^j(x,v)) - \min_{j\in F'(x,v)} F_{\mathcal{R}}(r_1^j(x,v)) \geq 0$, so that the first max operation can be removed. We can now identify:

$$\max_{j\in F'(\bar{x},\bar{v})\cup\{\ell'+1\}} F_{\mathcal{R}}(r_2^j(\bar{x},\bar{v})) = \max_{j\in F'(\bar{x},\bar{v})} F_{\mathcal{R}}(r_2^j(\bar{x},\bar{v}))$$

$$F_{\mathcal{R}}(r_1^{\ell'+1}(\bar{x},\bar{v})) = \min_{j\in F'(\bar{x},\bar{v})\cup\{\ell'+1\}} F_{\mathcal{R}}(r_1^j(\bar{x},\bar{v}))$$

$$F_{\mathcal{R}}(r_2^{\ell'+1}(\bar{x},\bar{v})) - \min_{j\in F'(\bar{x},\bar{v})} F_{\mathcal{R}}(r_1^j(\bar{x},\bar{v})) = -\max\left\{F_{\mathcal{R}}(\min_{j\in S'} r_1^j(\bar{x},\bar{v})) - F_{\mathcal{R}}(\max_{j\in S} r_2^j(\bar{x},\bar{v})),0\right\},$$

with $S' = F'(\bar{x},\bar{v})$, $S = \{\ell'+1\}$. Now let S, S' be an arbitrary partition of $F'(\bar{x},\bar{v}) \cup \{\ell'+1\}$. Then as soon as $\ell'+1 \in S'$, it follows that: $\max\left\{F_{\mathcal{R}}(\min_{j\in S'} r_1^j(\bar{x},\bar{v})) - F_{\mathcal{R}}(\max_{j\in S} r_2^j(\bar{x},\bar{v})),0\right\} = 0$. If $|S| \geq 2$ and $\ell'+1 \in S$, then, clearly:

$$\max\left\{F_{\mathcal{R}}(\min_{j\in S'} r_1^j(\bar{x},\bar{v})) - F_{\mathcal{R}}(\max_{j\in S} r_2^j(\bar{x},\bar{v})),0\right\} = \max\left\{F_{\mathcal{R}}(\min_{j\in S'} r_1^j(\bar{x},\bar{v})) - F_{\mathcal{R}}(\max_{j\in S\setminus\{\ell'+1\}} r_2^j(\bar{x},\bar{v})),0\right\}$$

Combining these derivations with the already derived formula for $e(\bar{x},\bar{v})$, we thus establish the validity of (6.29) for $\ell'+1$.

In the last situation, we may thus assume the existence of some i for which $r_1^i(\bar{x},\bar{v}) \leq r_2^{\ell'+1}(\bar{x},\bar{v})$. We may assume w.l.o.g. that $i = \ell'$. Then setting $\bar{r}_2^{\ell'} = \max\left\{r_2^{\ell'}(\bar{x},\bar{v}), r_2^{\ell'+1}(\bar{x},\bar{v})\right\}$ and $\bar{r}_1^{\ell'} = \min\left\{r_1^{\ell'}(\bar{x},\bar{v}), r_1^{\ell'+1}(\bar{x},\bar{v})\right\}$, and $\bar{r}_{1,2}^i(\bar{x},\bar{v}) = r_{1,2}^i(\bar{x},\bar{v})$ for $i < \ell'$, we may employ the induction assumption to derive (6.29) holds for $e(\bar{x},\bar{v})$ but involving $\bar{r}_{1,2}^i$. It is however clear that:

$$\max_{j\in F'(\bar{x},\bar{v})\cup\{\ell'+1\}} F_{\mathcal{R}}(r_2^j(\bar{x},\bar{v})) = \max_{j\in F'(\bar{x},\bar{v})} F_{\mathcal{R}}(\bar{r}_2^j(\bar{x},\bar{v}))$$

$$\min_{j\in F'(\bar{x},\bar{v})\cup\{\ell'+1\}} F_{\mathcal{R}}(r_1^j(\bar{x},\bar{v})) = \min_{j\in F'(\bar{x},\bar{v})} F_{\mathcal{R}}(\bar{r}_1^j(\bar{x},\bar{v})).$$

Once again, let S, S' be an arbitrary partition of $F'(\bar{x},\bar{v}) \cup \{\ell'+1\}$. Now should $\ell' \in S'$, $\ell'+1 \in S$ or vice-versa, then it follows that: $\max\left\{F_{\mathcal{R}}(\min_{j\in S'}(r_1^j(\bar{x},\bar{v}))) - F_{\mathcal{R}}(\max_{j\in S}(r_2^j(\bar{x},\bar{v}))),0\right\} = 0$. For the latter term to con-

tribute it must thus hold that $\ell', \ell'+1 \in S'$ or $\ell', \ell'+1 \in S'$, and thus by construction:

$$F_{\mathcal{R}}(\min_{j\in S'}(r_1^j(\bar{x},\bar{v}))) = F_{\mathcal{R}}(\min_{j\in S'\setminus\{\ell'+1\})}(\bar{r}_1^j(\bar{x},\bar{v})))$$
$$F_{\mathcal{R}}(\max_{j\in S}(r_2^j(\bar{x},\bar{v}))) = F_{\mathcal{R}}(\max_{j\in S\setminus\{\ell'+1\}}(\bar{r}_2^j(\bar{x},\bar{v}))),$$

so that upon combining these derivations, the formula (6.29) is shown for $\ell'+1$.

The proof is thus completed, by induction over ℓ'. □

6.4 CONCLUSION

This Chapter provides a better theoretical insight into local behaviours of probabilistic functions in a less restrictive setting relatively to what is usually encountered in literature. Although Theorem (6.23) provides "only" an inclusion, this still is valuable "first-order" information on φ. A practical method to verify whether the inclusion turns out to be an equality is readily available. Numerical experiments using for instance algorithm C is to be found in the original article [17]. This work has to be understood in the line of previous works we have detailed in Chapter 6: building upon the fruitful spherical-radial representation of elliptical random variables, we are able to work out a reformulation of probabilistic functions with interesting geometric interpretation. This, in turn, enables us to expose more information on such function, that can then be used for numerical algorithms. As one can observe, it requires quite a technical background in order to achieve incremental improvements, but the perspectives are still vast. As a matter of fact, modelling tools with enriched probabilistic functions (such as probust constraints for instance, see [16]) are fast, on-going developments made possible thanks to elliptical distributions. First-order information for these "new" functions is thus a promising field.

Part IV

Applications to the Optimal Power Flow

INTRODUCTION

We recall that before this last Part, we have set up the following elements:

1. the OPF is a tool available to the DSO to provide optimised decisions in the short-term operational planning process. Part I detailed the objective of this optimisation tool for the short-term operational planning from a DSO point of view.

2. From the abundant literature on the generic OPF, it is clear an instance of this class of problems with uncertainties is, without relaxation, a non-convex nonsmooth optimisation problem. Part II provides an optimisation algorithm tailored to solve a large class of nonsmooth nonconvex problems, namely Difference-of-Convex (DoC) problems.

3. The uncertainties are at the core of many engineering problems but still are significant features of complexification in a model. In Part III we present a short study into chance-constraints, which constitute one possibility to model uncertainties which involves probabilistic functions. Apart from a variational study of these functions, which we believe is interesting as it provides a bridge between chance-constrained problems and (sub-)gradient algorithms, we present the DoC property of chance-constraints and a DoC approximation.

In the light of these elements, we are now ready to turn to the OPF. The first chapter of this part presents our model of an instance of the OPF with uncertainties, using a DoC formulation. To the best of our knowledge, DoC programming has not yet been applied to the OPF class of problems. This is to be expected for the deterministic case, as casting a smooth OPF problem (for instance, without integer variables) as a DoC problem is in general not a "good" idea. Using an "off-the-shelf" solver for smooth programs will ultimately be a (significantly) better option, especially in terms of computational effort. Our DoC method does not leverage the smoothness assumption, and a Sequential Quadratic Programming method will probably be more efficient numerically. We cannot yet infer on the efficiency of our DoC method in the integer case (see the Discussion for some insights on this subject). It nevertheless appears to be promisingly applicable to a Chance-Constrained OPF, as defined in the following pages. From a broader perspective, we will propose two different models that account for uncertainties. In order to test these models we also present a realistic data creation process based on Enedis Open Data.
This fourth part is structured as follows: we first present preliminary elements for the DoC formulation of functions that appear in an OPF. This enables us to propose our DoCOPF, building upon the model initiated in Part I. In Appendix D, we briefly describe our data creation scheme. This plodding work, crucial to optimisation albeit sometimes somewhat slightly overlooked, aimed at providing a tool based on Enedis Open Data that is able to produce exhaustive open-source test data for our algorithms. Moreover, heuristics to automatically define forecasts of production and consumption on the grid that would lead to electrical constraints have been implemented and tested. Lastly, in Chapter 8 we present our results obtained from different DoC algorithms.

RÉSUMÉ EN FRANÇAIS

Nous rappelons qu'avant cette dernière partie, nous avons présenté les éléments suivants :

1. l'OPF est un outil à la disposition du DSO pour fournir des décisions optimisées dans le processus de planification opérationnelle à court terme. La partie I a détaillé l'objectif de cet outil d'optimisation de la planification opérationnelle à court terme du point de vue du DSO.

2. D'après l'abondante littérature sur l'OPF, il est clair qu'une instance de cette classe de problèmes avec incertitudes est, sans relaxation, un problème d'optimisation non convexe et non lisse. La partie II fournit un algorithme d'optimisation adapté à la résolution d'une grande classe de problèmes non convexes non lisses, à savoir les problèmes de différence de convexes (DoC).

3. Les incertitudes sont au cœur de nombreux problèmes d'ingénierie mais restent des sources importantes de complexification d'un modèle. Dans la partie III est présentée une brève étude des contraintes de probabilité, qui constituent une possibilité de modélisation des incertitudes impliquant des fonctions probabilistes. Outre une étude variationnelle de ces fonctions, qui nous semble intéressante car elle permet de relier les problèmes de contraintes de probabilité et les algorithmes de (sous-)gradient, nous présentons la propriété DoC des contraintes aléatoires et une approximation DoC.

A l'aide de ces éléments, nous sommes maintenant prêts à aborder l'OPF. Le premier chapitre de cette partie présente notre modèle d'une instance d'OPF avec incertitudes, en utilisant une formulation DoC. À notre connaissance, la programmation DoC n'a pas encore été appliquée à la classe de problèmes OPF. Ceci paraît attendu pour le cas déterministe, car transformer un problème OPF lisse (par exemple, sans variables entières) en un problème DoC n'est en général pas une "bonne" idée. L'utilisation d'un solveur "standard" pour les programmes lisses sera très probablement une meilleure option, notamment en termes de temps de calcul. Notre méthode DoC n'exploite pas l'hypothèse lisse, et une méthode de programmation quadratique séquentielle sera probablement plus efficace numériquement. Nous ne pouvons pas encore inférer sur l'efficacité de notre méthode DoC dans le cas des nombres entiers (voir le chapitre de Discussion pour quelques idées à ce sujet). Elle semble néanmoins s'appliquer de manière prometteuse à un OPF avec contrainte de probabilité, tel que nous allons le définir. Dans une perspective plus large, nous proposerons deux modèles différents qui tiennent compte des incertitudes. Afin de tester ces modèles, nous présentons également un processus réaliste de création de données basé sur Enedis Open Data, bien qu'il n'ait pas pu être totalement utilisé. Cette quatrième partie est structurée comme suit : nous présentons d'abord des éléments préliminaires pour la formulation de la DoC des fonctions qui apparaissent dans un OPF. Cela nous permet de proposer notre DoCOPF, en nous appuyant sur le modèle initié dans la partie I. Dans la partie D, nous décrivons brièvement notre schéma de création de données. Ce travail laborieux, crucial pour l'optimisation bien que parfois un peu négligé, a pour but de fournir un outil basé sur

Enedis Open Data qui est capable de produire des données de test réalistes ne nécessitant que des données libres pour nos algorithmes. De plus, des heuristiques permettant de définir automatiquement les prévisions de production et de consommation sur le réseau qui conduiraient à des contraintes électriques ont été implémentées et testées. Enfin, dans le chapitre 8 les résultats obtenus par différents algorithmes de DoC sur un cas test produit.

PRELIMINARY ELEMENTS FOR THE OPF AND DOC-OPF

7

7.1 INTRODUCTION

From part I, we know that the OPF is a non-convex, possibly non-smooth optimisation problem, for which a wide variety of (re)formulations and approximations are proposed in literature. A generic OPF is, broadly speaking, an optimisation problem that takes into account physical laws of electricity, namely the definition of electrical power, Ohm's and Kirchhoff's laws. Usual variables include power generation/consumption (we use the word *interaction* as an abstract word that can either be generation or consumption, depending on the actor), voltage norm, voltage phasor, and current. In literature, each proposition of an OPF formulation consists of a subset of these variables, the associated physical laws of electricity, and case-dependent constraints (e.g. specific constraints on GUs' interactions) as well as an objective function.

A generic deterministic OPF formulation that relies on a bus-injection formulation[1] is as follows:

$$
\begin{aligned}
\min_{\delta,p,q,|V|} \quad & f(\delta,p,q,|V|) \\
\text{s.t.} \quad & I = Y\,V && \text{Ohm's Law} && (7.1a)\\
& p_i^g - p_i^l = \mathrm{Re}(V_i\,I_i^\star) && \forall i \in \mathcal{N} && \text{Real power eq.} && (7.1b)\\
& q_i^g - q_i^l = \mathrm{Img}(V_i\,I_i^\star) && \forall i \in \mathcal{N} && \text{Reactive power eq.} && (7.1c)\\
& \ell(|V_i|,|V_j|,\delta_i,\delta_j) \leq (I_{i,j}^{\max})^2 && \forall (i,j) \in \mathcal{A} && \text{Transit limit} && (7.1d)\\
& A\,([\delta,p,q,|V|) \leq b && && \text{Linear constraints,}
\end{aligned}
$$

where f is a given objective function, ℓ is an appropriate quadratic function, A and b respectively an adequate matrix and vector.

One can immediately observe that formulation (7.1) is not well suited for our case. First of all, it is set up under the assumptions that the operator has full control over the power interactions. This obviously is not the case from a DSO point of view, as discussed in chapter 1. Secondly, we need to properly define and include uncertainties in it. Making DoC decompositions explicit is not a trivial step either.

While working towards a model for the OPF with uncertainties, the importance of the distinction between control and state variables becomes more significant than in the deterministic setting. As it turns out, in a deterministic setting the aim is to find a single set of control variables for which the associated set of state variables[2] respects the constraints defined by the user. In a stochastic setting, the objective is to provide a single set of control variables for which, broadly speaking in "as many scenarios as desired",

1 This formulation is, at the time of writing, the most widely found in literature.

2 By "associated set of state variables" we designate the state variables that have a real-life meaning. We assume that this set of variables, if it exists, is unique (see [295, Assumption 1]). It is possible that it does not exist when the electrical grid is managed outside of its "usual bounds".

there exists a set of admissible state variables. Although this explanation is not exact as we are dealing with continuous random variables and not a finite set of scenarios, it advantageously provides an hindsight into the necessary distinction of control variables from state ones in a stochastic setting: it is clear that having a single set of state variables in such a setting is meaningless. One solution, that has been studied in literature (see for instance [295]), is to have one set of state variables per generated scenario. Without any additional technique, this can rapidly lead to large, untractable optimisation problems. Available techniques to cope with this scalability problem is to use admissible envelopes of the feasible sets defined by the probabilistic constraints. This approximation leads to tractable, but intense computation. Our approach is different, as we do not introduce one set of state variables per scenario. We introduce a two-step approach, where the first step is an optimisation on control variables and the second step is an oracle that involves state variables and provides information on the uncertain operators of our model.

This Chapter is organized as follows: we start by detailing how our DSO point of view changes model (7.1), as well as the model of contracts that DSO has with GUs. We then provide insights on the DoC formulation of the functions at hand in our problem. We finally conclude by presenting the our formulations of the OPF with uncertainties, and discuss the differences between these proposed formulations.

7.1.1 Notations and preliminary remarks

Table 7.1: Relevant notations for this chapter

f, f_1, f_2	Objective function, in a DoC context $f = f_1 - f_2$ is to be understood	functions
c, c_1, c_2	Constraint function, in a DoC context $c = c_1 - c_2$ is to be understood	functions
$\mathcal{N}, \mathcal{A}$	List of grid' nodes and lines	parameters
$\lvert\mathcal{N}\rvert, \lvert\mathcal{A}\rvert, \lvert GU\rvert$	Cardinalities of $\mathcal{N}$, $\mathcal{A}$, and number of GU connected to the grid	parameters
p_u^ϕ, q_u^ϕ	Active and reactive power injection or consumption of GU u	parameters
ξ	Random vector in $\mathbb{R}^{\lvert GU\rvert}$	parameter
Overlined variable	Upper bound of the variable	parameter
Underlined variable	Lower bound of the variable	parameter
c_u^ν, c_u^γ	Marginal cost on variable p_u^ν and p_u^γ respectively	parameter
N	Number of generated outcomes of a random vector ξ	parameter
n, m, k	Number of variables, size of the random vector ξ and size of the constraint vector	parameters
p_u^ν, q_u^ν	DSO decision of SNS on active/reactive power for GU u	variables (control)
p_i^γ, q_j^γ	DSO decision on power modulation within contractualized rules for GU u	variables (control)
$\lvert V\rvert, \delta$	voltage norm and phasor	variables (state)
I	current, either indexed by a node or a line	variable (state)
S_{ij}	Apparent power on line $(i,j) \in \mathcal{A}$	variable (state)

Remark 14. Cost parameters can have more indices when the cost function is polynomial, in which case $c_{u,k}^\nu$ is the coefficient of monomial $(p_u^\nu)^k$. The same holds for c_u^γ. ▷

We add for clarity the following immediate properties:

- n is the sum of the number of all variables i.e. $p_u^\gamma, q_u^\gamma, p_u^\nu, q_u^\nu, \delta$ and $|V|$.
- Function c is valued in $\mathbb{R}^k$.

When necessary, we introduce scalar parameters or matrices which, unless specified otherwise, are respectively denoted by a and A. These two notations can also have indices. For clarity, we will mostly present our models for a single time step, which allows us to drop a t index in our notations. This should not be regarded as a limitation, as inter-temporal constraints could be included at the (only) cost of implementation work.

7.2 THE OPF FROM A DSO POINT OF VIEW

We here recall our high-level model of the grid and how the DSO interacts with other actors in the following Entity-Relationship model, which is detailed in part I:

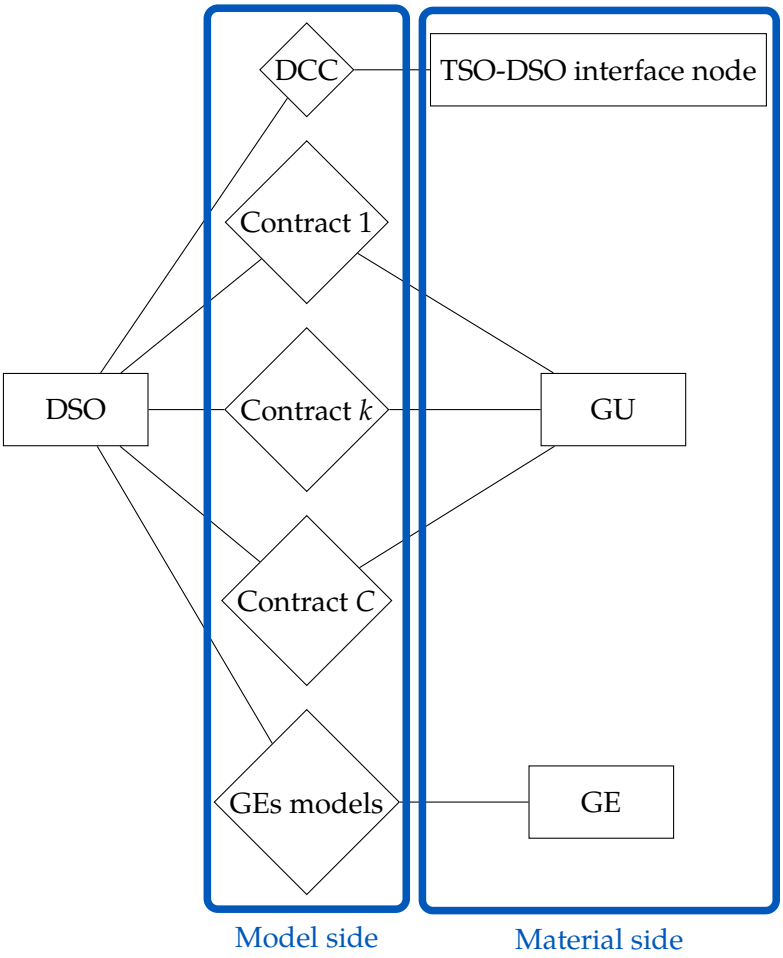


FIGURE 7.1: Abstract overview of the OPF model from a DSO point of view. For illustrative purposes, we consider a set of C different contracts, k being a given index within this set. We recall that GU stands for Grid User (an actor connected to the grid), while GE stands for Grid Element (a technical element connected to the grid, as a power line).

In Figure 7.1, DCC stands for Demand Connection Code which is a set of regulations on the flows of power at the interface between the DSO and TSO.

A *decision* for the DSO is a set of recommendations of actions that follows all the rules defined by the interactions elements of Figure 7.1. As the output of the ST-OP process, a decision is reached after the following generic steps:

1. Retrieving of information on the operational status of GEs. After this step, we expect the DSO to have a full knowledge of the input topology consisting of all active GEs in the grid. The DSO also has full knowledge of the allowed flexibilities on GUs.

2. Obtain forecasts of GUs interactions with the grid, which amounts for GUs to

communicate to the DSO their expected levels of consumption and production for a given set of time steps.

3. With this information, evaluate the forecasted operational state of the grid.
4. If electrical constraints are foreseen, then derive a decision on which flexibilities to activate motivated by the (inherently opposed) objectives of minimisation of potential failures of GEs and maximisation of the access to the grid for GUs.

Using the last point of this enumeration, it is evident that a DSO *decision* translates into a modulation of some forecasted levels of GUs interactions with the grid. Further detailing DSO' decisions, from a modelling point of view we distinguish two different types:

- either a decision that falls within the setting of a service from the GU for the benefit of the DSO;
- or a decision that falls outside from this service: we refer to this decision as "service not supplied" (SNS).

Decision of the first type Decisions of the first type follow constraints individually decided and negotiated upon beforehand between a GU and DSO. For instance in our case, these negotiated constraints are defined in "contracts" or *offres* within Enedis' nomenclature. Within the framework of traditional electrical connections, the DSO does not negotiate flexibility in the access of GUs to the grid: broadly speaking, when in a usual operational setting, the DSO cannot modulate the interaction between the grid and the GUs. The DSO can propose an electrical connection at a lower cost for a GU in exchange of some flexibility in the service provided by DSO. From a DSO perspective, this lowers the risk of faults, material failure or life endangering situations. From a GU perspective, this lowers the financial costs which can be significant when creating a new connection, and lowers the investment risks.

Decision of the second type The second type of decision is in our case an approximation of real-life decisions: when in hazardous situations (following significant meteorological events for example), the DSO may be compelled to temporally limit its service for some GUs. Un-contractualised load shedding is an example of such degradation of electrical quality, and can be included in an OPF as control variables. One could be tempted to model this latter type of load shedding variables with binary/integer variables, as in real life they are closer to an *on*/*off* decision than a continuous one. From our previous choices of not considering integer variables, we relax the integrality constraint on this second type of decision and thus have continuous variables. Evidently, considering integer variables would bring in some challenging mathematical considerations within this DoC optimisation problem requiring significantly more studies that are beyond the scope of this first DoC approach of an uncertain OPF. Moreover having continuous SNS variables has an interesting interpretation. Bearing in mind that SNS variables are the ones of *last resort*, we expect that their values should hold the information of "missing" flexibility in a given situation. An implication of this design is that the marginal cost of SNS should always be higher than maximal unitary cost of decisions of the the first type.

As a consequence, to further explicit the DSO decision on a GU's forecast, the left hand-sides of equations (7.1b) and (7.1c) are reformulated as follows:

$$\begin{aligned} p_u^g &= p_u^{\phi,g} + p_u^{\gamma,g} - p_u^{\nu,g} \\ p_u^l &= p_u^{\phi,l} + p_u^{\gamma,l} - p_u^{\nu,l} \\ q_u^g &= q_u^{\phi,g} + q_u^{\gamma,g} - q_u^{\nu,g} \\ q_u^l &= q_u^{\phi,l} + q_u^{\gamma,l} - q_u^{\nu,l}, \end{aligned} \tag{7.2}$$

where the exponents ϕ, γ, ν respectively designate the forecast parameter, the first-type decision variable and the second-type decision variable.

We can rewrite equations equations (7.1b) and (7.1c) as:

$$\sum_{j\ GU\sim i} p_j^{\phi,g} - p_j^{\phi,l} + (p_j^{\gamma,g} - p_j^{\gamma,l}) - (p_j^{\nu,g} - p_j^{\nu,l}) = \mathrm{Re}(V_i\ I_i^\star) \qquad \forall i \in \mathcal{V} \tag{7.3}$$

$$\sum_{j\ GU\sim i} q_j^{\phi,g} - q_j^{\phi,l} + (q_j^{\gamma,g} - q_j^{\gamma,l}) - (q_j^{\nu,g} - q_j^{\nu,l}) = \mathrm{Img}(V_i\ I_i^\star) \qquad \forall i \in \mathcal{V}, \tag{7.4}$$

7.3 FLEXIBILITIES IN OPERATIONAL PLANNING WITHIN OUR FRAMEWORK

In this section we provide an overview of the interactions described in figure 7.1. After some general comments on the GU model, we describe the different contracts at hand. For these contracts, we follow Enedis' white paper on flexibilities [99]. We focus on three different types of contracts between the DSO and GUs, that are referred to by their numbers: contract 1, 2 and 3.

7.3.1 GU MODEL

All GUs, either loads or DRES, share the same model framework that includes:

- The type of relation with the DSO, which in turn will define its contract,
- Its technical bounds: the lower bounds $\underline{p_u}, \underline{q_u}$ and upper ones $\overline{p_u}, \overline{q_u}$,
- Its forecasts of injection/consumption of real power and, when applicable, reactive power.

For clarity, when necessary we include for our presentation superscripts g and l to explicitly specify the role (power producer or consumer respectively) of the given GU. Since we do not consider batteries as a potential source of flexibility, the following relations:

$$\begin{aligned} p^{\phi,g} \geq 0,\ p^{\phi,l} \geq 0 \\ q^{\phi,g} \geq 0,\ q^{\phi,l} \geq 0, \end{aligned}$$

hold true.

We also consider the following to always hold within the scope of our application:

$$p_u^\nu \geq 0, \quad p_u = p_u^\phi + p_u^\gamma - p_u^\nu \geq 0,$$

for all GU except at the slack node which is modelled as node with a single GU that can be a consumer or a producer. These latter equations imply that the SNS variables are always positive (one cannot shed a negative amount of load for instance), and specify the decomposition of power variables: real power is the sum of forecasted power and of the contractualised modulation of power, minus the SNS value.

7.3.2 Contract type 1

This is the *Offre de raccordement de référence* which used to be the only one available for new producers. A GU connected to the grid through this contract, provided that the grid is in its usual exploitation scheme, can inject/consume its desired level of power. In other words, the DSO has no contractualized lever with this GU, and we will consider that the only possible mitigation of injection or consumption is denial of service.

The injection/consumption of power for a GU u verifies the following relations:

$$p_u = p_u^\phi - p_u^\nu \tag{7.5}$$

$$0 \leq p_u^\nu \leq p_u^\phi \tag{7.6}$$

The cost of the SNS variable p_j^ν is 10.000 €.MWh^{-1}.

7.3.3 Contract type 2

The first *Offre de Raccordement Intelligent* (ORI, see [99] for a public description of this contract) that has been experimented by Enedis ensures to a GU a guaranteed level of power injection to the grid, whose production/consumption is greater than this guaranteed power can be modulated by the DSO if necessary. We let this lower bound verify:

$$\underline{p_u^{\gamma,g}} = a_1 p_u^{\texttt{installed}}, \qquad \forall u \; GU \text{ generator},$$

$$\underline{p_u^{\gamma,l}} = a_2 p_u^{\texttt{subscribed}}, \qquad \forall u \; GU \text{ consumer},$$

with $a_1, a_2 \in \mathbb{R}_+$.

From a practical point of view, due to our data and case creation that is described in appendix D, we have implemented a slightly different rule:

$$\underline{p_u^{\gamma,g}} = a_1 p_u^{\phi,g}, \qquad \forall u \; GU \text{ generator}, \tag{7.7}$$

$$\underline{p_u^{\gamma,l}} = a_2 p_u^{\phi,l}, \qquad \forall u \; GU \text{ consumer}, \tag{7.8}$$

still with $a_1, a_2 \in \mathbb{R}_+$. What this new rule implements amounts to have "lower" lower bounds, and thus more flexibility in our optimisation problem. It is possible that at a given time step for a given generator i, $p_u^\phi < a_1 p_u^{\texttt{installed}}$. Consequently, this generator would behave as a contract of type 1 from the DSO point of view, and provide no flexibility. As part of this work aims at exploring flexibilities, this modification appears to be a reasonable workaround to make sure flexibilities are implemented, while maintaining the structure of this contract's rules.

In the objective function, it is assumed that p^γ is of zero cost (and implemented as a close-to-zero cost, i.e. $c^\gamma = 0.001$).

7.3.4 Contract type 3

The second ORI[3] does not implement a minimal injection threshold, but rather provides to the DSO a given, contractualized yearly level of energy curtailment. One could see contract 2 as a *contract on power* as opposed to contract 3 which is one on *energy*. Conceptually, from an optimisation point of view the cost of lever activation within this contract could be modelled as having a marginal cost dependent on the level of energy curtailed up to the current time of activation. This implies that for GU under this contract, no constraints are added to the optimisation problem: the associated p_u^γ variable is solely constrained by the usual box-constraints $\underline{p_u^\gamma}$ and $\overline{p_u^\gamma}$ (with in our case $\underline{p_u^\gamma} = 0$ and $\overline{p_u^\gamma} = p_u^\phi$):

$$\underline{p_u^\gamma} \leq p_u^\gamma \leq \overline{p_u^\gamma}, \quad \forall u \; GU \tag{7.9}$$

On the other hand, the objective function is not linear any more as it is the case for contract 2. In a multitemporal setting, this is modelled as:

$$c_{u,t}^\gamma = \text{cost}(\{p_{u,\tau}^\gamma, q_{u,\tau}^\gamma\}_{\tau \in [u,t-1]}). \tag{7.10}$$

Within our work' scope, for a single time step we decide to take another stance and consider that the cost for p^γ within this contract is a quadratic function:

$$\text{cost}(x) = c_{i,2}^\gamma x^2 + c_{i,1}^\gamma x + c_{i,0}^\gamma, \tag{7.11}$$

with $(c_{i,2}^\gamma, c_{i,1}^\gamma, c_{i,0}^\gamma) \in \mathbb{R}_+ \times \mathbb{R}^2$.

Remark 15 (Acceptable class of objective function)**.** Thanks to DoC programming and `PBMDC`[2], any DoC is in fact eligible as our objective function. Turning to literature on the OPF, most used objective functions are polynomials of degree 2 which obviously are DoC (and possibly non-convex). The rationale behind this class of functions is simply that cost of plant production is satisfyingly modelled by the square of the energy produced. ▷

7.3.5 Slack bus contract

At the slack bus level, which is the interface between HV and MV grids, a special contract is needed. As is usual, this node is treated as a GU which is at the same time a load and a generator while having no forecasts. Constraints at this interface are legally set by two official codes (see [60] for more detailed information): the Demand Connection Code (DCC) and System Operations. DCC sets a feasible region in a P, Q space, while SO defines the temporality and contents the DSO has to share with its associated TSO.

It is clearly seen in figure 7.2 that the feasible set is not convex.

[3] Again, see [99] for more information on this contract.

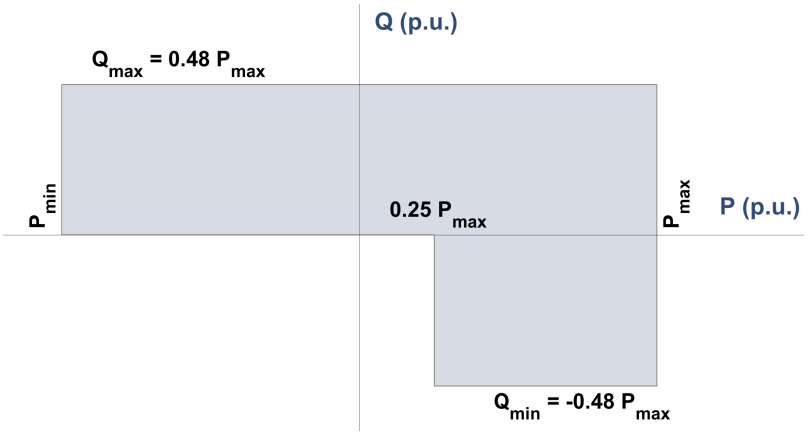


FIGURE 7.2: Feasible P-Q diagram following DCC rules.

INSERT 7: A DoC approximation of the DCC constraint.
We care to mention that the non-convex set of figure 7.2 can be approximated as closely as desired using DoC functions. This approximation is therefore a DoC set as defined in [285, Section 4.6], and is depicted in the following figure:

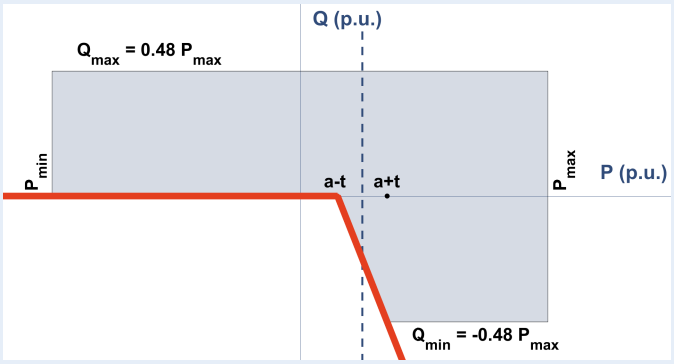


FIGURE 7.3: A DoC approximation for the non-convex set of figure 7.2 with $a = 0.25P_{max}$ and $t > 0$.

Setting the feasible set of figure 7.3 as $\mathcal{S}$, a DoC description of $\mathcal{S}$ amounts to finding the equation of the orange line. Therefore one has:

$$\mathcal{S} = \{(P,Q) \in \mathbb{R}^2 \mid P_{\min} \leq P \leq P_{max},\ Q_{\min} \leq Q \leq Q_{max},$$
$$Q \geq \min(0, \frac{-0.24P_{max}}{t}P + \frac{0.24(a-t)P_{max}}{t})\}. \quad (7.12)$$

In our setting, as (i) a DoC version of the DCC constraint adds a computational burden and (ii) such a constraint is not our prime interest, we have decided to stick to a simpler convex relaxation, consisting of the convex-hull of the original DCC set. Its P-Q diagram is as follows:

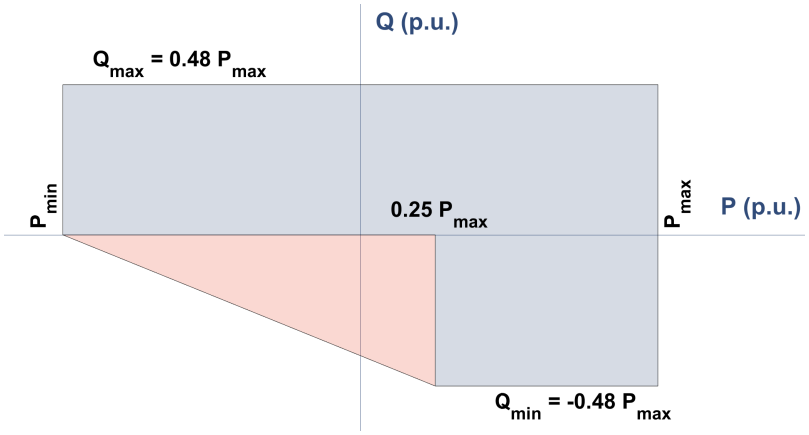


FIGURE 7.4: A "lazy" convex approximation of the P-Q diagram following DCC rules, consisting of the union of the blue and orange shapes.

As such, the feasible set at the slack bus is:

$$\mathcal{F}_{\text{slack}} := \{(P,Q) \in \mathbb{R}^2,\ P_{\min} \leq P \leq P_{\max},\ Q_{\min} \leq Q \leq Q_{\max},\ \frac{-0.48P_{\max}}{-P_{\min}+0.25P_{\max}}P + \frac{0.48P_{\min}\ P_{\max}}{-P_{\min}+0.25P_{\max}} \leq Q\}. \quad (7.13)$$

7.4 A FIRST DOC FORMULATION FOR THE OPF

In this section, we present a first deterministic DoC reformulation of our model, before discussing a natural stochastic extension of this first model. We will see that the result of this work turns out to be difficult to interpret for the original short-term operational planning objective. An analysis of these difficulties provides enlightening motivation for a second DoC model, presented in section 7.5.

7.4.1 THE DETERMINISTIC DOCOPF

First of all, as is visible in the problem (7.1), in order to derive a DoC formulation for the OPF, one has to make explicit the DoC form of constraints equations (7.1b) and (7.1c). This turns out to be the most challenging part of the DoC formulation in a deterministic setting, as constraint equation (7.1d) often turns out to be convex, and the boundary constraints are linear.

An evident obstacle is the necessity to transform an equality into two inequalities. Constraints equations (7.1b) and (7.1c) are thus rewritten as:

$$|p_i^g - p_i^l - \text{Real}(V_i\ I_i^\star)| \leq \epsilon, \quad (7.14)$$

$$|q_i^g - q_i^l - \text{Img}(V_i\ I_i^\star)| \leq \epsilon, \quad (7.15)$$

where $\epsilon \geq 0$ is a fixed user-selected parameter. In our case, we used values between 10^{-3} and 10^{-5}, with a clear trade-off between numerical efficiency and interpretability/usefulness of the solution point when varying ϵ.

Changing these equality constraints that translate a power equilibrium can rightfully appear to be problematic. In the deterministic setting, we explore if this approximation

dependent on ϵ can appropriately hold. "Holding" in this case obviously is dependent on the user's definition of what is a "good" approximation. In any optimisation process, different types of numerical approximations/roundings that sequentially impact one another imply that one should humbly study the outputs of optimisation. As our computations are done within a selected unit parameters setting, quality of a numerical solutions has to account for it. In our case, the potential base parameter is approximately $V_{\texttt{base}} = 20.10^3$V: when one have two different values for a voltage, say $|V|_1, |V|_2$, with $||V|_1 - |V|_2| \leq a$ with $a \in \mathbb{R}_+$, then the absolute spread between these two values is $a \times 20.10^3$. We assume that two different absolute norms of voltage values with a difference of order of magnitude 10 Volts can reasonably be accepted as *similar enough* in order to consider them numerically equal. Back in the per-unit system, 10 Volts are equal to 5.10^{-4} per unit, which justifies our selection for values of ϵ. Now, as in our model we impose loose constraints on voltage phasors, this discussion is not relevant for these latter variables.

Let us start by defining the following functions, in order to ease the coming model presentations:

$$\begin{aligned} c_i^{\mathtt{R}}(\delta, p, q, |V|) = & \left(p_i^{\phi,l} + p_i^{\gamma,l} - p_i^{\nu,l} - \left(p_i^{\phi,g} + p_i^{\gamma,g} - p_i^{\nu,g}\right)\right) + \\ & \sum_{k\sim i} Y_{i,k}^R |V|_i |V|_k \cos(\delta_i - \delta_k) + Y_{i,k}^I |V|_i |V|_k \sin(\delta_i - \delta_k) \\ = & \; c_{i,1}^{\mathtt{R}}(\delta, p, q, |V|) - c_{i,2}^{\mathtt{R}}(\delta, p, q, |V|), \end{aligned} \tag{7.16a}$$

$$\begin{aligned} c_i^{\mathtt{I}}(\delta, p, q, |V|) = & \left(q_i^{\phi,l} + q_i^{\gamma,l} - q_i^{\nu,l} - \left(q_i^{\phi,g} + q_i^{\gamma,g} - q_i^{\nu,g}\right)\right) + \\ & \sum_{k\sim i} Y_{i,k}^R |V|_i |V|_k \sin(\delta_i - \delta_k) - Y_{i,k}^I |V|_i |V|_k \cos(\delta_i - \delta_k) \\ = & \; c_{i,1}^{\mathtt{I}}(\delta, p, q, |V|) - c_{i,2}^{\mathtt{I}}(\delta, p, q, |V|), \end{aligned} \tag{7.16b}$$

where functions $c_{i,1}^{\mathtt{R}}, c_{i,2}^{\mathtt{R}}, c_{i,1}^{\mathtt{I}}, c_{i,2}^{\mathtt{I}}$ are convex functions in all their components. We expose the technical description of our methodology to obtain these functions in appendix A.

A direct DoC reformulation of our OPF is then:

$$\begin{aligned} \min_{\delta,p,q,|V|} \quad & f(\delta, p, q, |V|) \\ \text{s.t.} \quad & c_{i,1}^{\mathtt{R}}(\delta, p, q, |V|) - c_{i,2}^{\mathtt{R}}(\delta, p, q, |V|) - \epsilon \leq 0 && (7.17a) \\ & c_{i,2}^{\mathtt{R}}(\delta, p, q, |V|) - c_{i,1}^{\mathtt{R}}(\delta, p, q, |V|) - \epsilon \leq 0 && (7.17b) \\ & c_{i,1}^{\mathtt{I}}(\delta, p, q, |V|) - c_{i,2}^{\mathtt{I}}(\delta, p, q, |V|) - \epsilon \leq 0 && (7.17c) \\ & c_{i,2}^{\mathtt{I}}(\delta, p, q, |V|) - c_{i,1}^{\mathtt{I}}(\delta, p, q, |V|) - \epsilon \leq 0 && (7.17d) \\ & \ell(|V_i|, |V_j|, \delta_i, \delta_j) \leq (I_{i,j}^{\max})^2 \quad \forall (i,j) \in \mathcal{A} && (7.17e) \\ & A\,([\delta, p, q, |V|) \leq b, \end{aligned}$$

where we recall $\epsilon \in \mathbb{R}_+$ is a user-given approximation parameter.

7.4.2 An associated stochastic extension

We now invoke (i) the previous discussion on the uncertainties section 2.3, (ii) and the deterministic model of section 7.4.1. From the former, we recall that we consider the

power forecasts values to be uncertain, and to be associated to probability distribution functions of the elliptical class. To explicit this dependency, we now write the following:

$$\begin{array}{ccc} \text{Deterministic} & & \text{Uncertain} \\ p_u^{\phi} & \longrightarrow & p_u^{\phi}(\xi) \\ q_u^{\phi} & \longrightarrow & q_u^{\phi}(\xi). \end{array}$$

From the latter, a "canonical" jointly chance-constrained extension of the deterministic model would lead us to:

$$\begin{aligned} \min_{\delta,p,q,|V|} \quad & f(\delta,p,q,|V|) \\ \text{s.t.} \quad & \mathbb{P}\begin{bmatrix} c_{i,1}^{\mathrm{R}}(\delta,p,q,|V|,\xi) - c_{i,2}^{\mathrm{R}}(\delta,p,q,|V|,\xi) - \epsilon & \leq 0 \\ c_{i,2}^{\mathrm{R}}(\delta,p,q,|V|,\xi) - c_{i,1}^{\mathrm{R}}(\delta,p,q,|V|,\xi) - \epsilon & \leq 0 \\ c_{i,1}^{\mathrm{I}}(\delta,p,q,|V|,\xi) - c_{i,2}^{\mathrm{I}}(\delta,p,q,|V|,\xi) - \epsilon & \leq 0 \\ c_{i,2}^{\mathrm{I}}(\delta,p,q,|V|,\xi) - c_{i,1}^{\mathrm{I}}(\delta,p,q,|V|,\xi) - \epsilon & \leq 0 \\ -p_i^{\phi,l}(\xi) - p_i^{\gamma,l} + p_i^{\nu,l} & \leq 0 \\ -p_i^{\phi,g}(\xi) - p_i^{\gamma,g} + p_i^{\nu,g} & \leq 0 \\ -q_i^{\phi,l}(\xi) - p_i^{\gamma,l} + q_i^{\nu,l} & \leq 0 \\ -q_i^{\phi,g}(\xi) - p_i^{\gamma,g} + q_i^{\nu,g} & \leq 0 \end{bmatrix} \geq \mathtt{p} \qquad (7.18a) \\ & \ell(|V_i|,|V_j|,\delta_i,\delta_j) \leq (I_{i,j}^{\max})^2 \quad \forall (i,j) \in \mathcal{A} \qquad (7.18b) \\ & A\,([\delta,p,q,|V|) \leq b. \end{aligned}$$

Remark 16. For an easier presentation, we have overloaded the notations c_i^{R} and c_i^{I} initially defined for the deterministic case (7.16), now also defined for the stochastic case (7.19). The definition of these functions is then context-dependent, the latter being easily identifiable with the presence of ξ parameter. We then explicitly have:

$$\begin{aligned} c_i^{\mathrm{R}}(\delta,p,q,|V|,\xi) = & \left(p_i^{\phi,l}(\xi) + p_i^{\gamma,l} - p_i^{\nu,l} - \left(p_i^{\phi,g}(\xi) + p_i^{\gamma,g} - p_i^{\nu,g}\right)\right) + \qquad (7.19a) \\ & \sum_{k\sim i} Y_{i,k}^{R}\,|V|_i\,|V|_k \cos(\delta_i-\delta_k) + Y_{i,k}^{I}\,|V|_i\,|V|_k \sin(\delta_i-\delta_k) \\ = & \; c_{i,1}^{\mathrm{R}}(\delta,p,q,|V|,\xi) - c_{i,2}^{\mathrm{R}}(\delta,p,q,|V|,\xi), \\ c_i^{\mathrm{I}}(\delta,p,q,|V|,\xi) = & \left(q_i^{\phi,l}(\xi) + q_i^{\gamma,l} - q_i^{\nu,l} - \left(q_i^{\phi,g}(\xi) + q_i^{\gamma,g} - q_i^{\nu,g}\right)\right) + \qquad (7.19b) \\ & \sum_{k\sim i} Y_{i,k}^{R}\,|V|_i\,|V|_k \sin(\delta_i-\delta_k) - Y_{i,k}^{I}\,|V|_i\,|V|_k \cos(\delta_i-\delta_k) \\ = & \; c_{i,1}^{\mathrm{I}}(\delta,p,q,|V|,\xi) - c_{i,2}^{\mathrm{I}}(\delta,p,q,|V|,\xi). \end{aligned}$$

▷

One can observe that uncertain forecasts parameters naturally introduce uncertainty in our non-convex, DoC equations (7.14) and (7.15). Problem (7.18) obviously is chance-constrained, with DoC functions under the probability operator, as well as a possibly DoC objective function. Other constraints are convex, and possibly linear. Further studying equations (7.18) and (7.19), it is immediately apparent that we have separable

functions at hand (see Definition 9). Thanks to the separability and the continuously differentiability of the functions under the probabilistic operator, we know from chapter 5 that constraint (7.18a) is (locally) continuously differentiable, the neighbourhoods where this differentiability result holds being those where a Slater condition is verified.

Now that we have discussed the expected regularity conditions of this problem, let us move on two follow-up questions:

1. How can we solve such a problem?
2. How can we interpret a solution of this problem?

The solving methodology

The first question is rapidly answered when one recalls PBMDC^2 algorithm and section 5.3. From the latter section, we obtain a DoC approximation of constraint equation (7.18a) thus leading us with problem equation (7.20) at hand. Then, as PBMDC^2 is an algorithm that can tackle DoC constrained DoC problems.

Let us define $c^{\mathbb{P}}_{1,\tau,\epsilon}, c^{\mathbb{P}}_{2,\tau,\epsilon}$ as the two DoC components of the DoC approximation of equation (7.18a). As detailed in section 5.3, this approximation is parametrized and we denote this parameter by τ. Moreover, these approximation are dependent on ϵ (recall (7.18a) for instance). Formally speaking, $c^{\mathbb{P}}_{1,\tau,\epsilon}, c^{\mathbb{P}}_{2,\tau,\epsilon}$ are functions defined from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}$, convex in all their arguments.

We can explicit the DoC constrained DoC optimisation problem:

$$
\begin{aligned}
\min_{\delta,p,q,|V|} \quad & f(\delta, p, q, |V|) \\
\text{s.t.} \quad & c^{\mathbb{P}}_{1,\tau,\epsilon}(\delta, p, q, |V|, \xi) - c^{\mathbb{P}}_{2,\tau,\epsilon}(\delta, p, q, |V|, \xi) \geq \mathrm{p} && \text{(7.20a)} \\
& \ell(|V_i|, |V_j|, \delta_i, \delta_j) \leq (I^{\max}_{i,j})^2 \quad \forall (i,j) \in \mathcal{A} && \text{(7.20b)} \\
& A\,([\delta, p, q, |V|) \leq b.
\end{aligned}
$$

Proposition for the interpretation of a solution to equation (7.20)

Let us denote by $x^\star = (\delta^\star, p^\star, q^\star, |V|^\star)$ a solution to the optimisation problem equation (7.20). From an operational point of view, the interpretation is not trivial and this difficulty exposes the significant limit of this first model.

Strictly speaking, when our problem equation (7.20) has a non-empty feasible set to which belong $x^\star$, then $x^\star$ is a point that minimizes our cost function while being feasible with a probability greater or equal to p. An immediate question is what are $\delta^\star$ and $|V|^\star$?

To answer this question, we first rewrite equation (7.18a) as follows:

$$\mathbb{P}\left[\xi \in M_\epsilon(x^\star)\right] \geq \mathrm{p},$$

for an appropriately defined set $M_\epsilon(x^\star)$ (whose explicit definition is easily retrieved from equation (7.18a)). The random vector ξ holds the information on possible future outcomes, while $M_\epsilon(x^\star)$ is the set of outcomes that, combined with decisions $x^\star$, lead to an "approximately" feasible electrical state. The approximation here has to be understood as "up to a translation of ϵ".

The main difficulty in this model, that possibly was already visible, now becomes clear: it assumes that a single grid electrical state, defined by $|V|^\star, \delta^\star$, can be ϵ-feasible with a probability greater than p. This assumption in turn requires that:

(i) the variances of all coordinates of the random vector ξ are "small enough", or

(ii) the behaviour of the equilibrium constraint vector equation (7.16) around some of its zeros is calm "enough"[4].

The first point is problematic as it limits the operational usefulness; the second point is numerically not verified. This latter point is related to interesting questions on the localisation of solutions to OPF or PF, which is the subject of future sections. Significant associated questions include: (i) how many solutions does a PF system have? (ii) what is the sensibility of PF solutions relatively to data values?

Conclusion on this first model

The first model is a chance-constrained problem that we can cast as a DoC-constrained DoC problem, and consequently solve. It appears as a naïve stochastic extension to the OPF, well fitted to our previously developed material. Nevertheless, a solution to this DoC problem should not be considered as a decision-making tool to the ST-OP process as described in part I. A mathematical model aiming at being a consistent decision-making tool should clearly include the *two-step* property of the underlying problem. As a matter of fact, one looks for values for control variables such that there exists feasible associated state variables with probability greater or equal to p.

The next section is dedicated to our second DoC model which is a two-step program. It is built on the limits made explicit in our first attempt.

7.5 A second DoC formulation for the stochastic OPF

We now move on our second proposition for a satisfactory DoCOPF. The main drawback of the naïve approach is not tackling the two inherently different roles of involved variables. On a more operational point of view, we will distinguish *state* variables from *control* ones. The former ones are variables a DSO does not explicitly have control over but requires their values to be in given sets - one could also refer to them as the *controlled* variables. The latter ones are the levers the DSO has control over. Constraints in our model will therefore enforce the relation between *controllable* variables and *controlled* ones.

Based on this variable separation, we propose a two-step approach which is an iterative/looping process. The first step is about control values selection, thus decision over the control variables. The second step is about the impact on *state* variables, given the values selected for the control variables. The loop closes as information from the second step is passed to the first step.

[4] Calmness, as described in [253, Chapter 8 Section F], is an intuitive concept that applies at a given point x. A function f is said to be calm if the differences between $f(x)$ and $f(x')$ are bounded for every x' in a neighbourhood of x. In that matter, it differs from Lipschitz continuity that is a uniform property.

In our case, state variables are comprised of the voltage norm $|V|$ and the phasor δ, whereas control variables are DSO decisions on active power p. Reactive power q will be considered as a state variable or as a control one, depending on the model.

We first of all define:

$$\mathcal{R}(p,q,\xi) = \min_{\delta,|V|} \quad \sum_{i=1}^{|\mathcal{N}|} (c_{i,1}^{\mathrm{R}}(\delta,p,q,|V|,\xi) - c_{i,2}^{\mathrm{R}}(\delta,p,q,|V|,\xi))^2 + (c_{i,1}^{\mathrm{I}}(\delta,p,q,|V|,\xi) - c_{i,2}^{\mathrm{I}}(\delta,p,q,|V|,\xi))^2 \tag{7.21a}$$

$$\text{s.t.} \quad \underline{\delta}_i \leq \delta_i \leq \overline{\delta}_i \quad \forall i \in \mathcal{N} \tag{7.21b}$$

$$\underline{|V|}_i \leq |V|_i \leq \overline{|V|}_i \quad \forall i \in \mathcal{N}. \tag{7.21c}$$

as well as:

$$\mathcal{R}_{\delta,|V|}(p,q,\xi) = \sum_{i=1}^{|\mathcal{N}|} (c_{i,1}^{\mathrm{R}}(\delta,p,q,|V|,\xi) - c_{i,2}^{\mathrm{R}}(\delta,p,q,|V|,\xi))^2 + (c_{i,1}^{\mathrm{I}}(\delta,p,q,|V|,\xi) - c_{i,2}^{\mathrm{I}}(\delta,p,q,|V|,\xi))^2. \tag{7.22}$$

Note that $\mathcal{R}_{\delta,|V|}$ is the objective function in problem (7.21). We have thus defined an infinite family of functions $(\mathcal{R}_{\delta,|V|})_{\delta,|V|}$. Each element of this family is a function defined on $\mathbb{R}^n \times \mathbb{R}^m$ with values in $\mathbb{R}_+$. From equation (7.16) it is clear that by virtue of being a sum of compositions of $\mathcal{C}^2$ functions $c_{i,1}^{\mathrm{R}}, c_{i,2}^{\mathrm{R}}, c_{i,1}^{\mathrm{I}}, c_{i,2}^{\mathrm{I}}$ with the square function, each element $\mathcal{R}_{\delta,|V|}$ is a $\mathcal{C}^2$ function as well.

Remark 17. Problem (7.21) is a redefinition of a classical Power-Flow, this time cast as an optimisation problem with additional constraints. When the constraints are neglected, this method has already been investigated in the past, as in [292, 274] and more recently discussed in [108, 61] (although neither implementation is discussed nor numerical experiments are reported in the latter two works). The existence of very efficient load-flow solvers as, for instance, those based on Newton-Raphson technique might limit the interest of writing the load-flow as an optimisation problem. In fact, it is unsure whether there is any computational gain by doing so: [292] reports improvements in computation time, while [274, 108] rather discard this method. Numerical instabilities were also reported in the early works of the 60's-70's, although no evidence was given. The idea of casting the PF as an optimisation problem stemmed from the apparition of efficient gradient-based algorithms. It can be found in these works referred to as the "minimisation of the squared power mismatches". ▷

INSERT 8: Interpretation of function $\mathcal{R}$.
The values of function $\mathcal{R}$ can be interpreted as follows: for a given (first-step) decision, it provides the minimal violation of PFs (if any). In case this value is not equal to zero, there is little hope one can find a physical interpretation to the optimal point obtained in the computation of the functional value of $\mathcal{R}$. The only conclusion in this particular case is that no point respecting PFs equations as well as constraints on voltage and current transit was found.
In case the functional value of $\mathcal{R}$ is zero, then there exists a point satisfying the PF equations and respecting the constraints on voltage and current transits.

Lemma 12 (Regularity and structure of $\mathcal{R}$)**.** $\mathcal{R}$ is a DoC function. ◀

Proof. For clarity purposes, and in order to match some definitions and theorems from our reference [253], we introduce the following functions:

$$\begin{aligned}\mathcal{S}_{\delta,|V|}(p,q,\xi) = &-\sum_{i=1}^{|\mathcal{N}|}(c^{\mathrm{R}}_{i,1}(\delta,p,q,|V|,\xi)-c^{\mathrm{R}}_{i,2}(\delta,p,q,|V|,\xi))^2+\\ &(c^{\mathrm{I}}_{i,1}(\delta,p,q,|V|,\xi)-c^{\mathrm{I}}_{i,2}(\delta,p,q,|V|,\xi))^2\\ = &-\quad \mathcal{R}_{\delta,|V|}(p,q,\xi),\end{aligned}$$

and:

$$\begin{aligned}\mathcal{S}(p,q,\xi) \quad = \quad &\max_{\delta,|V|} \quad \mathcal{S}_{\delta,|V|}(p,q,\xi) && \text{(7.24a)}\\ &\text{s.t.} \quad \underline{\delta}_i \le \delta_i \le \overline{\delta}_i \quad \forall i \in \mathcal{N} && \text{(7.24b)}\\ &\qquad \underline{|V|}_i \le |V|_i \le \overline{|V|}_i \quad \forall i \in \mathcal{N}. && \text{(7.24c)}\end{aligned}$$

As a consequence, the following inequality holds everywhere: $\mathcal{S}(p,q,\xi) = -\mathcal{R}(p,q,\xi)$. We are now in position to properly invoke results from [253].

A direct application of [253, Definition 10.29], and observing that every function of the family $(\mathcal{S}_{\delta,|V|})_{\delta,|V|}$ is a $\mathcal{C}^2$ function on $\mathbb{R}^n \times \mathbb{R}^m$, proves that $\mathcal{S}$ belongs to the set of Lower-$\mathcal{C}^2$ functions on $\mathbb{R}^n \times \mathbb{R}^m$. As a consequence of well-known DoC properties (see for instance the discussion on the universality of DoC functions in [229, Section 2.1]), $\mathcal{S}$ is a DoC function. It follows that $\mathcal{R}$ is DoC. □

Here is a list of immediate considerations from the definition of problem (7.21):

- If $\mathcal{R}(p,q,\xi) = 0$, then there exists some values for $\delta, |V|$ that describes a feasible electrical state of a network studied with the problem.
- If $\mathcal{R}(p,q,\xi) > 0$, then some further considerations are necessary and section 7.5.5 is dedicated to this matter.
- Problem (7.21) is non-convex, smooth, deterministic for a fixed outcome of ξ, and can consequently be efficiently solved by numerous existing solvers (namely Sequential Quadratic Programming ones).

Before presenting different models, we need to make sure that an explicit DoC form for the non-convex and non-smooth function $\mathcal{R}$ is available. First of all, as $\mathcal{S}$ is lower-$\mathcal{C}^2$, we know that there exists a fixed parameter $a \in \mathbb{R}_+$ such that $\mathcal{S} = \mathcal{S} + a\|\cdot\|^2 - a\|\cdot\|^2$ is a valid DoC formulation. Then, as $\mathcal{S} = -\mathcal{R}$, we have $\mathcal{R} = a\|\cdot\|^2 - (-\mathcal{S} + a\|\cdot\|^2)$. We also claim that another one is close-to-readily available, thanks to the work conducted in the first model section 7.4. This latter DoC formulation of $\mathcal{R}$ is obtained from the generic DoC decomposition of $x \mapsto h^2(x)$ where h is a DoC function for which a decomposition is explicitly available. We recall that this decomposition can be found in [285, Proposition 4.12], and holds provided that the components functions of h are positive-valued (this is always true modulo a translation of the component functions).

7.5.1 First-order information on $\mathcal{R}$

We here recall some well-known results in variational analysis:

Theorem 14 (Theorem 10.31 in [253])**.** Let us consider a lower-$\mathcal{C}^1$ function f, that is, $f(x) = \max_{t\in T} f_t(x)$ with T a compact set and $(f_t)_{t\in T}$ a family of $\mathcal{C}^1$ functions. Let $T(x) = \arg\max_{t\in T} f_t(x)$.
Then, one has:

$$\partial f(\bar{x}) = \mathrm{Co}\{\nabla f_t(\bar{x}) \mid t \in T(\bar{x})\}. \tag{7.25}$$

◀

As a consequence of Theorem 14, when considering an outcome ξ_l of random vector ξ we can exhibit an element of $\mathcal{S}(p, q, \xi_l)$:

$$\nabla \mathcal{S}_{\delta^\star, |V|^\star}(p, q, \xi_l) \in \partial \mathcal{S}(p, q, \xi_l), \tag{7.26}$$

where $\delta^\star, |V|^\star$ are solutions to Problem (7.24). With the Lipschitz property of $\mathcal{R}$ and $\mathcal{S}$, and recalling [82, Proposition 2.3.1] we can conclude on the first-order information on $\mathcal{R}$ with:

$$\nabla \mathcal{S}_{\delta^\star, |V|^\star}(p, q, \xi_l) \in \partial \mathcal{S}(p, q, \xi_l) = -\partial \mathcal{R}(p, q, \xi_l) \implies \nabla \mathcal{R}_{\delta^\star, |V|^\star}(p, q, \xi_l) \in \partial \mathcal{R}(p, q, \xi_l).$$

Remark 18. In order to have a correct oracle for PBMDC^2 and ensure a convergence, "first-order information" is necessary. For this information, it is sufficient to provide an element from the subdifferentials of each DoC function at hand, which is reasonable to expect. This is due to the relatively mild assumptions on the oracle used in this algorithm. There exists other bundle methods that make use of oracles with stronger assumptions (see for instance [223]), and their convergence results are evidently at least as good as those of PBMDC^2.

A correct oracle also means that it requires exact values of $\delta^\star, |V|^\star$: the importance of having a good solving methodology for the second-step is again explicited. ▷

7.5.2 Presentation of different types of models within the second approach

The following models make use of the load-flow cast as an optimisation problem. Within our two-step framework, all models can basically be classified as a risk minimisation

under a cost constraint, or a cost minimisation under a risk constraint. In each of these two types, one can evidently use either the expected value operator $\mathbb{E}[\cdot]$ or the probabilistic one $\mathbb{P}[\cdot]$. The justifications for this affirmation is that

(i) by virtue of the linearity of $\mathbb{E}$, the composition of this operator with a DoC function is a DoC function[5];

(ii) we know from Chapter 5 that the composition by $\mathbb{P}[\cdot]$ can be approximated by a DoC function, and the first item concludes our justification.

Following Part III, we constrain the random vector $\xi \in \mathbb{R}^{|GU|}$ to have an elliptic symmetric distribution.

The first variant model is a joint-chance constrained optimisation problem (JCCP). For this model, the user defines a parameter $\mathrm{p} > 0$ which defines the minimal satisfaction parameter to enforce.

$$\begin{aligned} \min_{p,q} \quad & f(p,q) \\ \text{s.t.} \quad & \mathbb{P}_{\xi}\left[\mathcal{R}(p,q,\xi) \leq 0\right] \geq \mathrm{p} \\ & \textit{Constraints (7.5) to (7.9) and (7.13)}, \end{aligned} \tag{Const-$\mathbb{P}$}$$

The second variant is closely related to (Const-$\mathbb{P}$), and is the maximization of the probabilistic constraint of problem (Const-$\mathbb{P}$) under a cost constraint. In that setting, the user provides a parameter $\overline{f} > 0$ that models the maximal allowed cost for the operation. Such a problem is referred to as "Max-P" (see [25, Section 5.1.2.3]):

$$\begin{aligned} \max_{p,q} \quad & \mathbb{P}_{\xi}\left[\mathcal{R}(p,q,\xi) \leq 0\right] \\ \text{s.t.} \quad & f(p,q) \leq \overline{f} \\ & \textit{Constraints (7.5) to (7.9) and (7.13)}, \end{aligned} \tag{Max-$\mathbb{P}$}$$

Remark 19. "Max-P" also is of interest for solving the JCCP problem (Const-$\mathbb{P}$): it can be used to compute an initial point. ▷

Similarly to Problems Const-$\mathbb{P}$ and Max-$\mathbb{P}$, we derive the following two models:

- An expected-value minimisation problem:

$$\begin{aligned} \min_{p,q} \quad & E_{\xi}\left[\mathcal{R}(p,q,\xi)\right] \\ \text{s.t.} \quad & f(p,q) \leq \overline{f} \\ & \textit{Constraints (7.5) to (7.9) and (7.13)}, \end{aligned} \tag{Min-$\mathbb{E}$}$$

  with $\overline{f} > 0$ the maximum acceptable cost.

- A cost minimisation under uncertain operational constraints:

$$\begin{aligned} \min_{p,q} \quad & f(p,q) \\ \text{s.t.} \quad & E_{\xi}\left[\mathcal{R}(p,q,\xi)\right] \leq \epsilon \\ & \textit{Constraints (7.5) to (7.9) and (7.13)}, \end{aligned} \tag{Const-$\mathbb{E}$}$$

  with $\epsilon \geq 0$ being an abstract parameter defining the maximum acceptable risk.

[5] A proper proof is to be found in [285, Proposition 4.4]

The two parameters $\overline{f} > 0$ and $\epsilon > 0$ are user-given, with simple interpretations: the first one is a max-cost allowance, while the second one is a mask-risk allowance.

Remark 20. As opposed to the first two models, Problems Min-$\mathbb{E}$ and Const-$\mathbb{E}$ are not joint chance-constrained problems. They are DoC optimisation models that account for uncertainty in a stochastic setting. They also have a particular interest of their own from a modelisation point of view, as their description and the interpretation of their respective solutions' are simple. As a matter of fact, from the implementation and numerical point of view the expected value operator amounts to a count of scenarios that either show no electrical constraints, or on the opposite forecast some. ▷

INSERT 9: Distinction between a two-step and a bilevel/two-stage program. Let us emphasize that a two-step algorithm is different from a two-stage or bilevel one, the former being a special case of the latter. Following [46] which provides a thorough introduction to bilevel programming, we define a general bilevel program as:

$$\begin{aligned} \min_{x \in X} \quad & F_b(x, y) \\ \text{s.t.} \quad & G_b(x, y) \leq 0 \\ & \min_{y \in Y} \quad f_b(x, y) \\ & \text{s.t.} \quad g_b(x, y) \leq 0 \end{aligned}$$

where $F_b, f_b : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $G_b : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ and $g_b : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$ are assumed to be continuous and twice differentiable for simplicity. It is essential to notice the constraint is itself a minimisation problem over set of variables that is not independent from the one of the master problem. In our case these two sets are disjoint, and this turns out to be a simpler model. Abstractly our model is rather of the following form:

$$\begin{aligned} \min_{x \in X} \quad & F(x) \\ \text{s.t.} \quad & \mathbb{P}_\xi \left[G(x, \xi) := \min_{y \in Y} g(x, y, \xi) \leq 0 \right] \geq \mathrm{p}, \end{aligned}$$

or

$$\begin{aligned} \min_{x \in X} \quad & F(x) \\ \text{s.t.} \quad & E_\xi \left[G(x, \xi) := \min_{y \in Y} g(x, y, \xi) \right] \leq \overline{E}. \end{aligned}$$

where $F : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$, $g : \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^m \to \mathbb{R}^p$, and $\overline{E}$ is a maximal acceptable risk parameter. It becomes clear that, contrary to bilevel programming which essence is having an optimisation problem as a constraint, in our case the optimisation solely is in the *evaluation* of the constraint.

7.5.3 General structure of the two-step approach

With more insight, we here propose a modular overview of the two-step approach. This modularity is interesting as several of these blocks are indeed challenging problems of their own. A more efficient block can evidently improve the global performance of the algorithm.

The First-step problem

The Master Problem, which we can also refer to as the *first step*, is a DoC constrained DoC optimisation problem. This is the case for Problems Const-$\mathbb{E}$ and Const-$\mathbb{P}$, whereas Problems Min-$\mathbb{E}$ and Max-$\mathbb{P}$ can be DoC optimisation problem with "simpler" constraints (linear or convex).

This first step problem is on control variables: state variables are relegated to the second step. This is an optimisation problem with only power variables and constraints at hand. Observe that first-step variables are the only ones with a contribution to the objective function. A consequence of this observation is that when this function is a cost function, all variables which activations imply a cost contribution should appear in the first-step.

The Second-step problems

Once a decision p^k, q^k is reached in the Master, the second step uses this decision to compute the values $\mathcal{R}(p^k, q^k, \xi)$. As it turns out, the evaluation of this value amounts to a scenarii decomposition which bears the nice property of being *embarrassingly parallel*. For each and every scenario, there is a load-flow cast as an optimisation problem to solve. The set of outputs are gathered and processed differently depending on which operator is selected (either $E[\cdot]$ or $\mathbb{P}[\cdot]$).

This second step only deals with state variables, and integrates the associated constraints. The objective is to have the "simplest" second step possible, as per iteration of PBMDC^2 it is required to solve this step as many times as there are outcomes of ξ considered by the user.

Remark 21. We care to emphasize the distinctions between these second-step problems and the sequence of subproblems in PBMDC^2. The former problems are smooth optimisation problems to solve at each oracle call, whereas the latter are the linearly constrained quadratic optimisation problems solved at each iteration of PBMDC^2 to compute the next iterate. ▷

INSERT 10: Taking a step back on the two steps.
From classic load-flow/power-flow models and methods, usual elements/actors of an electrical grid are modelled as one of the three following node:

- $p, |V|$ node: such a voltage controlled node is typical of a large producer;
- p, q node: this is the most common node on a distribution grid. It is well suited for loads or small generators;
- $|V|, \delta$ node: the slack node, it is the interface between the distribution and transport grids.

Apart from the slack bus, we only consider in our case p, q nodes. It is therefore no coincidence our variables are separated in such a way, *i.e.* power variables in the first-step and voltage-related ones in the second step. As a matter of fact, within the scope of our framework, our objective is to find modifications of elements/users' fixed values, constrained by contracts' bounds while ensuring the variable values (to be understood as the "*non-fixed*" ones) do not violate any constraints. Having p, q nodes, our first step variables are indeed p, q and the other variables ($|V|, \delta$) are effectively in the second step. If we were to consider $p, |V|$ nodes, we would (probably) include these variables into the first step, while considering the associated variables q, δ into the second step.

Broad overview

We here provide an overview of the solving methodology within the framework of our second type of models using `PBMDC`^{2}. It abstractly presents all the required steps, each one having to be implemented in a tailored fashion depending on the model that is at hand.

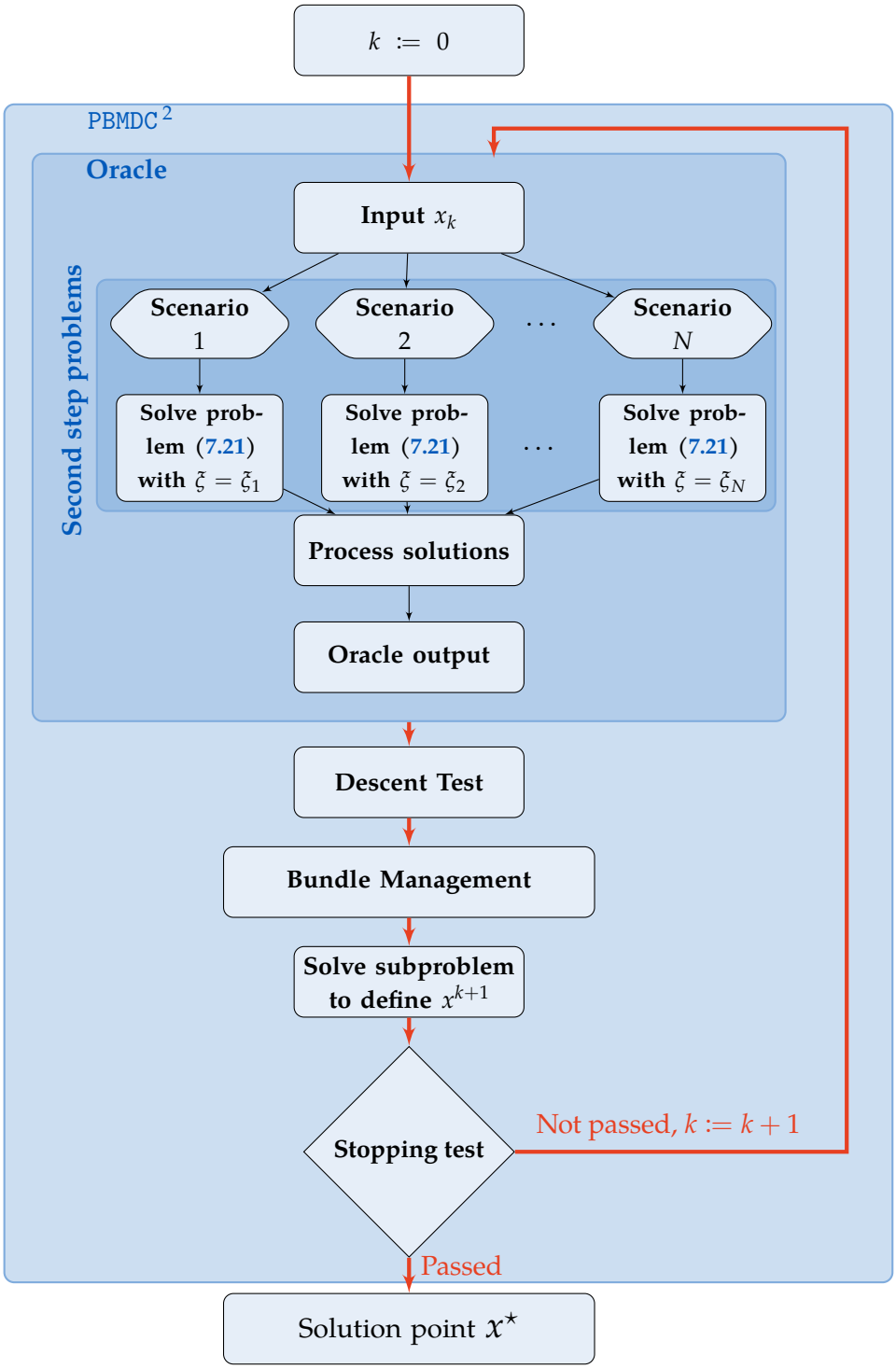


FIGURE 7.5: A broad overview of the Two-step approach within PBMDC 2.

In figure 7.5, step "Process solutions" is the most differing one between probabilistic models and expectation models. In the former case, we use the previously presented elements of section 5.3. With these elements, we are able to process the outputs of all resolutions for each generated outcome of ξ and provide the functional value and a subgradient of the DoC approximation of $\mathcal{R}$. In the latter case, the "Process solutions" amounts to computing the average value of all the outputs from each generated outcome

of ξ, weighted by the associated probabilities of these outcomes.

7.5.4 Possible extensions to these models

We here discuss three different aspects related to a better modelling of the ST-OP. The first one is related to the consideration of constraints on current transit also called thermal constraints. The second aspect is related to local voltage regulation, while the third is on the OLTC (On Load Tap Changer) at the slack node.

Thermal constraints

An interesting extension to these models is to add operational constraints on current transit, as previously considered in the first model equation (7.20b). This in an important gain on the operational side of the model, and a significant numerical challenge as it complexifies the second step. Let us point out that this constraint does not change the regularity properties of $\mathcal{R}$ (see Lemma 12). As a consequence, we can rewrite $\mathcal{R}$ as follows:

$$\mathcal{R}(p,q,\xi) = \min_{\delta,|V|} \quad \textit{Equation}(7.21a) \tag{7.27a}$$

$$\text{s.t.} \quad \textit{Constraints (7.21b) and (7.21c)} \tag{7.27b}$$

$$\ell(|V_i|,|V_j|,\delta_i,\delta_j) \leq (I_{i,j}^{\max})^2 \quad \forall (i,j) \in \mathcal{A}, \tag{7.27c}$$

Now, as discussed on the second step, when aiming to have the fastest problems to solve per each scenario, a possibility is to move some constraints from the second step to the first one. As such, transit limits on current could be modelled as constraints on power variables in the first step.

Local voltage regulation

Local voltage regulation belongs to the thriving studies of distributed control mechanisms, leveraging new technologies of power electronics. This methodology has the advantage of not relying on a large centralized system, and some special cases do not even require communications. We are interested here in a voltage-based reactive power control, which is a local voltage control. As found in [105, Definition 1], a local voltage control is the assignment of reactive power interaction of a selected GU connected to the grid computed according to the measured voltage value at its connection point. In other words, when a GU has the required technological elements to provides this local regulation, it can modify its reactive injection q, which in turn has an influence on $|V|$ at its connection point. This is an already operational technology, and has been implemented as it is often assumed that, on a "first order approximation", the reactive power has more impact on $|V|$ than on δ, the converse being true for the active power p.

When a GU u has the technological elements to provide this service, the associated first-step variables q_u^γ become *null* within our model: the DSO cannot control them any more. The reactive variables of this GU become a second-step variable. The associated constraints on these variables can take several forms:

- A user can for instance consider the local regulation law from [246]. It is a parametrized, piecewise linear local regulation function of type Q as a function of V. This has not been implemented in our models, but should be seen as a direct and valuable extension.
- We have opted for an easier implementation, which is to consider a continuous variable $\underline{q} \leq q \leq \bar{q}$. This model amounts to considering that power electronics are capable of delivering any kind of service at no cost (recall that for now reactive power is not monetized).

Slack node voltage set point

Yet another possible enrichment of our models is the modification of the voltage set point at the slack node's transformer. With the modernisation of these transformers, it becomes possible to modify the set point in a closed interval. In this case, $|V|_{\text{slack}}$ becomes a variable where it is a parameter set to the fixed set point value in Problems (Const-$\mathbb{P}$ - Min-$\mathbb{E}$).

Two possibilities to consider this variable are available:

- Either consider it to be a control variable, thus leading to a modification of the first step. In this step, $|V|_{\text{slack}}$ becomes a continuous, lower and upper bounded variable.
- Another interesting possibility is somewhat similar to our proposition for the Local voltage regulation. In this second model, we let $|V|_{\text{slack}}$ be a continuous, lower and upper bounded variable in the second step problem without modifying the first step. This models the ability to modify, close to real time, the voltage set-point at the slack node at no cost. There are several drawbacks to this model:
  - As described in [278], the "no-cost" assumption is not realistic at the moment. It turns out that the activation of modifying the set point at the slack node deteriorates the transformer, and this should be accounted for.
  - By passing the variable $|V|_{\text{slack}}$ to the second step, there is no possibility to control the number of its activations within a given set of time periods. On an operational setting, the operator cannot activate this lever as many times as desired: if more than 2 activations are required over a 24hours period, this is considered to be above usual bounds.

  We believe nevertheless this slight modification to be interesting from a possibly "prospective" point of view. Modifying the set point value at the entry node is an important lever for the future of electrical grids that is still to be fully deployed and is a still improving technology: for instance, newer OLTCs can be expected to operate 600.000 times before requiring maintenance[6].

7.5.5 Analysis of the second step problem (7.21)

The aim of this section is to provide insights on the solving a of PF, cast as an optimisation problem. As first described in Remark 17, several works have studied this method.

[6] This converts to over 34 years, considering one operation every 30 minutes.

Historically the ambition was to benefit from rapidly developing numerical methods for smooth optimisation problems. In our case, the ambition is to be able to provide valid cuts for the DoC function $\mathcal{R}$. The main obstacle is to ensure that our solution to problem (7.21) is a global optimum. This is an open and difficult question, to which some insights suggest that we indeed can expect global optimality. If we fail to reach global optimality when computing the value of $\mathcal{R}$ are a given iteration, then we do not have an exact oracle, which is a necessary condition for our algorithm PBMDC^2.

In order to reduce numerical difficulties, we start by proposing the following scheme when entering the second step:

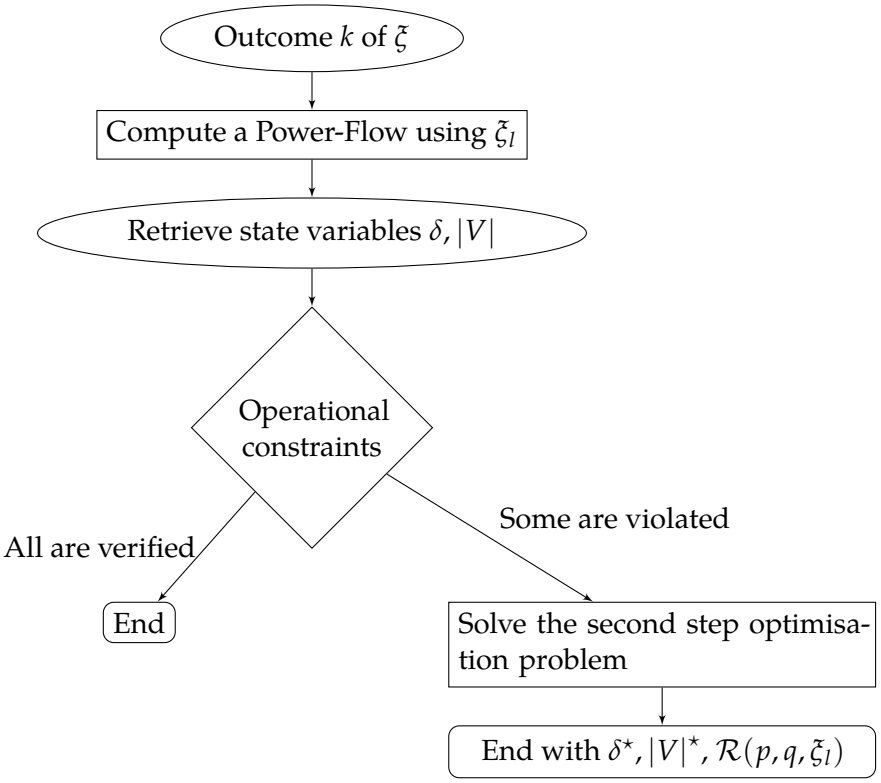


FIGURE 7.6: The second-step overview.

The rationale is that when no lever activation is required, we already know the functional value and a subgradient of $\mathcal{R}$. For a given outcome ξ_l of random vector ξ, when all operational constraints are verified by the state variables computed in the PF, then $\mathcal{R}(p, q, \xi_l) = 0$. Moreover, a subgradient of $\mathcal{R}$ at (p, q) is a zero vector of appropriate size.

A smooth optimisation problem needs to be solved only for outcomes where operational constraints are violated. We thus lower the numerical burden, as solving a PF system is less demanding: we rely for this matter on existing efficient code, namely the one initially proposed by Joe H. Chow and improved by Graham Rogers, that is included within their MATLAB toolbox [81].

Several obstacles to an exact oracle for PBMDC^2 can arise in the second-step problem (7.21) and equation (7.27), or other second-step described in Section 7.5.4.

1. The Power-flow computation can provide an undesirable solution point.

2. The optimisation problem (7.21) being non-convex, the solution can be a local minimum and not a global one.

For the first point, we look into a slightly related question, which is the number of solutions that a PF system admits: as expected, there are several mathematical solutions to such a system. As proved in works [156, 169] and more recently [203, 210], it is clear that the solution set to a PF system is not a singleton. The latter work even explicitly provides solutions to a PF system (in a lossless network, thus not realistic but simpler for computations). Another very interesting work on this matter is [108]: in a $|\mathcal{N}|$-bus network, there are at most $2^{|\mathcal{N}|-1}$ solutions to the associated PF system. One of these solutions is called the high-voltage, while the others are the low-voltage ones. The former is "the" operational solution which is the one that should be observed in real-life (leaving aside modelling errors), and the latter ones are non-desirable because they only have a mathematical meaning and no physical one. The distance of low-voltage points from the high-voltage points is an indicator of the electrical stability of the electrical grid. As described in the works cited in this section (see for instance [108, p.9]), when the system is not ill-conditioned, it is usually observed that numerical methods to solve a PF system converge to the desired high-voltage point. A similar assumption is made in [295], namely that in the usual operational conditions (i.e. with usual operational constraints on $\delta, |V|$), the PF system has a unique solution when it exists. This assumption could enable us to discard the first possible source of error: assuming our Power-flow code is able to find a solution that verifies the operational constraints, it is unique. If it cannot find this solution, we assume there is no Power-flow solution within the operational constraints, and move on to the second-step optimisation problem.

For the second point, by definition and assumptions, there is no Power-flow solution that respects the operational constraints. As a consequence, the objective at hand is not about finding a feasible PF solution anymore. The second-step optimisation problems we have described thus far belong to a class of constrained Power-flow problem, which is a hybrid PF-OPF problem as it is a true PF problem written as an optimisation problem[7]. The immediate properties of these problems include smoothness and a non-convex but DoC structure. One could then eventually derive an algorithm capable of computing a global optimum. But the gain on optimality would be at the expense of computational time. Still aiming to have a fast solving second step, we rely on the practical efficiency of the SQP solver that we assume is able to provide a global optimum to our second-step optimisation problems.

We conclude with the following two assumptions:

Assumption 3. Our PF code is capable of finding the high-voltage solution point to a PF system. If it verifies every operational constraints, it is unique.

Assumption 4. When solving the optimisation problem of the second-step is necessary, our SQP solver is able to provide a global solution.

[7] It is not an OPF problem as generic OPFs have power variables whereas our second-step is an optimisation where power values are parameters.

7.5.6 General characterization of our approach

When analysing our two step approach, one might wonder how our methodology relates to a "scenario decomposition" method. Although we do compute values per scenarios, this should not be considered as a classic scenario decomposition method. In our case, we do require scenarios solely to compute values of the probabilistic functions. To clarify, by setting $\eta = \mathcal{R}(p, q, \xi)$ which is a random vector itself, each scenario is a realisation of η. We immediately see that more scenarios we use, the better are these values. The physical interpretation, or trying to associate to these scenarios a real-life interpretation, should be regarded as secondary. They indeed are generated for the probabilistic approximations and not their physical meaning. As such, observing the number of scenarios that have electrical constraints along the iterates of `PBMDC`^{2} is not a good criteria *a priori*.

The only chance-constrained problem among our four propositions is problem (Const-$\mathbb{P}$). Problem (Max-$\mathbb{P}$) is often referred as a "max-P" problem: it also is a probabilistic problem. On the other hand, problem (Min-$\mathbb{E}$) is closer to a canonical two-stage stochastic program, as described in [265]. The only difference lies in that the constraints in the canonical two-stage stochastic program from [265] are linear, whereas there are convex polynomial in problem (Min-$\mathbb{E}$). Out last model, problem (Const-$\mathbb{E}$), is a stochastic problem which is simply known as an expected value constrained problem in [293].

7.5.7 Conclusion on the second DoC formulation for the stochastic OPF

This new class of models seems more promising than the first one: the distinction between control and state variables is explicit. A solution to a problem of this class should consequently be considered as a better assistance to the DSO decision making in the ST-OP process. Two different types of models are presented: probabilistic ones (with the $\mathbb{P}[\cdot]$ operator), and expectations ones (with the $\mathbb{E}[\cdot]$ operator). The solving methodologies are similar for both, the difference only residing in the step "Process solutions" of figure 7.5.

In this section, some numerical burdens have been withheld: the explicit DoC formulations of the involved functions is not immediately obtained. Some obstacles are discussed in Section 7.5.5 among which Assumption (4) can appear to be restrictive. This restriction is discussed in chapter 8, and promising research is likely to lift this particular obstacle.

7.6 A discussion on our two-steps models

Comparing mathematical models is a sensible subject as the underlying question often is "which model is best to use?", which answers should always be taken with care. In a decision making context, a mathematical model and its resolution aim at providing useful information to a question asked by a decision-maker. Just as a mathematical model in mathematical programming has an optimality and a feasibility criteria that can characterize what the "best" point is, the "best" model can only be determined if an optimality and a feasibility criteria are readily available. As a consequence, a fair and honest answer to which model to use is possible if this latter question comes with:

1. The question the decision-maker is facing. An example can be: what decision should the DSO take on a given electrical grid in order to have a 0.95 probability of experiencing no electrical constraints according to our current knowledge of the uncertainties?

2. The constraints related to the decision-maker and his/her environment. For example, the DSO has a time/material constraint: no more than a given amount of time is allowed.

From this observation, it is evident that all our four models do not provide answers to a single question, and as such they should not be frontally compared. More precisely:

- Problems Const-$\mathbb{P}$ and Const-$\mathbb{E}$ aim at minimizing a cost function while having a low measure of risk;

- Problems Max-$\mathbb{P}$ and Min-$\mathbb{E}$ aim at minimizing a measure of risk of failure at a maximal known and given cost.

The objective of the decision-maker and the *a priori* available information (do we have an information on the maximal acceptable cost or on the maximal acceptable risk?) are key to decide which model should be used.

Comparisons can be conducted on the use of the operator $\mathbb{E}$ or $\mathbb{P}$ which differently measure the risk of exploiting the grid outside of its usual operating bounds. This amounts to basic probabilistic considerations: we here only provide some key points of this comparison. The expected value operator applied to a multivariate random vector does not include any covariance information. Recalling that the expected value of a multivariate random vector is by definition the vector of the expected values of each coordinate, it is clear that no information on a possible "joint" effect cannot be included in our models using the operator $\mathbb{E}[\cdot]$. As a consequence, a probabilistic model is arguably more conservative than one with the expected value operator, as in the former case more information on the distribution is taken into consideration.

Remark 22. We care to add that if cost and risk were to be minimized in a single model, this falls into a multi-objective context. Interestingly, the recent and major work [212] from the DoC literature proposes a "double bundle" method in a multi-objective context with DoC functions. For a DoC-constrained DoC problem, reformulated thanks to an improvement function, this work proposes a bundle-based solving methodology to achieve a Clarke stationary solution point[8]. The authors discuss the link between this achievable Clarke stationarity and notions of Pareto optimality which are key in multi-objective optimisation. The bridge between DoC programming and multi-objective optimisation is thus complete. ▷

Finally, we care to emphasize two key "ingredients" that enable us to set up these models and have reasonable expectations on the numerical efficiency of `PBMDC` [2] when it is applied on them:

- The availability of a procedure to compute functional values and subgradients of $\mathbb{P}_{\tilde{\xi}}\left[\mathcal{R}(p,q,\xi) \leq 0\right]$ and $E_{\tilde{\xi}}\left[\mathcal{R}(p,q,\xi) \leq 0\right]$, which is found in section 5.3.

[8] This is similar to an unconstrained problem solved by `PBMDC` [2]

- The DoC framework which is the cornerstone of the communication between the two steps. Thanks to the DoC structure of the function $\mathcal{R}$, and of the composition of operator $\mathbb{P}$ or $\mathbb{E}$ with $\mathcal{R}$, information of the second step can be used as input for the first step. Other known cases where information of the second step could be brought back to the first usually require convexity, which is significantly stronger than DoC property required in our case.

7.7 CONVERGENCE OF THE *SAMPLE AVERAGE APPROXIMATION* APPLIED TO OUR PROBLEM

While previous sections develop and discuss our two-step approach which 1) is interesting from a physical point of view and 2) can fit in a DoC setting, little has yet been studied on the numerical treatment of the evaluation of uncertainties. We will rely on a *Sample Average Approximation* (SAA), an extensively used numerical method to obtain tractable approximation of chance-constraints. The motivation for this choice rather than, say, Genz' code [128], is the relative simplicity of this method.

Let us define:

$$\begin{aligned} \varphi_{\mathtt{lb}}^{\mathtt{ub}} \colon \mathbb{R}^n &\to [0,1] \\ x &\mapsto \mathbb{P}\left[\mathtt{lb} \leq c_1(x,\xi) - c_2(x,\xi) \leq \mathtt{ub}\right], \end{aligned} \tag{7.28}$$

as well as:

$$\mathbb{1}_{[\mathtt{lb},\,\mathtt{ub}]}(z) = \begin{cases} 1 & \text{if } z \in [\mathtt{lb},\,\mathtt{ub}] \\ 0 & \text{otherwise.} \end{cases} \tag{7.29}$$

We also recall that:

$$\varphi_{\mathtt{lb}}^{\mathtt{ub}}(x) = \mathbb{E}[\mathbb{1}_{[\mathtt{lb},\,\mathtt{ub}]}(c_1(x,\xi) - c_2(x,\xi))]. \tag{7.30}$$

The SAA method requires a set of independent Monte Carlo sample of vector ξ, that we denote $\{\xi_j\}_{j\in[1,N]}$, N being the sample size. The main idea of this method is to replace the integral computation of equation (7.30), which is necessary to compute $\varphi_{\mathtt{lb}}^{\mathtt{ub}}(x)$, by the following finite sum:

$$\widehat{\varphi}_{\mathtt{lb}}^{\mathtt{ub}}(x) \coloneqq \frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{[\mathtt{lb},\,\mathtt{ub}]}(x).$$

Now, recalling that our initial problem is:

$$\begin{aligned} \min_{x\in X} \quad & f_1(x) - f_2(x) \\ \text{s.t.} \quad & \varphi_{\mathtt{lb}}^{\mathtt{ub}}(x) \geq \mathtt{p}, \end{aligned} \tag{$\mathcal{P}_{\mathtt{p}}$}$$

the SAA problem is defined as follows:

$$\begin{aligned} \min_{x\in X} \quad & f_1(x) - f_2(x) \\ \text{s.t.} \quad & \widehat{\varphi}_{\mathtt{lb}}^{\mathtt{ub}}(x) \geq \mathtt{p}. \end{aligned} \tag{$\mathcal{P}_{\mathtt{p}}^{N}$}$$

SAA ensures in some particular cases that this is indeed an approximation as the approximation error goes to zero when the sample size grows. Our objective here is to verify that this property holds in our case where we have DoC functions.

Let z^* and z^N be the optimal solutions of problems $\mathcal{P}_\mathtt{p}$ and $\mathcal{P}_\mathtt{p}^N$ respectively. The authors of [200] study the relations between z^* and z^N, as well as possible inclusion of the feasible set of problem $(\mathcal{P}_\mathtt{p}^N)$ into the one of problem $(\mathcal{P}_\mathtt{p})$.

In this section, we start by presenting results on the probability of obtaining a lower bound to the optimal value from solving the SAA approximation, as well as cases where the set of feasible points to the SAA problem is included in the set of feasible points to the original problem.

7.7.1 Initial properties of the SAA problem and its optimal value

Using [200, Theorem 3] and provided that problem $(\mathcal{P}_\mathtt{p})$ has an optimal solution, the optimal solution to $\mathcal{P}_{\mathtt{p}-\gamma}^N$ will be a lower bound to the optimal solution of problem $(\mathcal{P}_\mathtt{p})$ for all $\gamma \in (0,1)$, with a probability converging to one exponentially fast with the sample size N.

An additional relation between z^* and z^N is readily available in literature. Let us define $\delta \in (0,1)$ as the desired confidence level for $z_{\mathtt{p}-\gamma}^N$ to be a lower bound to $z_\mathtt{p}^*$. If the sample size N verifies:

$$N \geq \frac{1}{2\gamma^2} \log\left(\frac{1}{1-\delta}\right),$$

then by [200, Theorem 3], we obtain $z_{\mathtt{p}-\gamma}^N \leq z_\mathtt{p}^*$ with a probability of at least δ.

In other words, the authors of [200] are thus able to derive a relation between the sample size and the probability that $z_{\mathtt{p}-\gamma}^N$ is a lower estimate of $z_\mathtt{p}^*$, for all $\gamma \in (0,1)$. They have also studied the conditions under which a feasible solution to $\mathcal{P}_{\mathtt{p}-\gamma}^N$ turns out to be feasible for problem $(\mathcal{P}_\mathtt{p})$. For clarity, let us define the following notations:

- $X_\mathtt{p} := \{x \in X \colon \mathbb{P}\left[\mathtt{lb} \leq c_1(x,\xi) - c_2(x,\xi) \leq \mathtt{ub}\right] \geq \mathtt{p}\}$;
- $X_\mathtt{p}^N := \{x \in X \colon \dfrac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{[\mathtt{lb},\mathtt{ub}]}(c_2(x,\xi^i) - c_1(x,\xi^i)) \geq \mathtt{p}\}$.

Within the scope of our work, where our application (the *Optimal Power Flow*) has a random right-hand side, we will apply the results presented in [200, Section 2.2.2]. In more details, our interest lies in a case where the following three assumptions are verified:

Assumption 5. There exist functions $\widetilde{c_1}, \widetilde{c_2}, g$ such that $c_1(x,\xi) - c_2(x,\xi) = \widetilde{c_1}(x) - \widetilde{c_2}(x) - g(\xi)$.

Assumption (5) states we assume our model has a random right-hand side.

Assumption 6. The feasible set $X_\mathtt{p}$ verifies

$$X_\mathtt{p} \subseteq \{x \in X \colon l \leq \widetilde{c_1}(x) - \widetilde{c_2}(x) \leq u\} =: X(l,u),$$

for some u and l in $\mathbb{R}^m$.

Assumption 7. The cumulative distribution function F of the random vector ξ is Lipschitz continuous with constant L.

Theorem 15 (Theorem 9 in [200])**.** Assume that Assumptions (5 - 7) are verified. Let $\gamma \in [0, 1-\mathrm{p})$, $\beta \in (0, 1-\mathrm{p}-\gamma)$ and $D = \max\{u_j - l_j,\ j \in [1,m]\}$. Let us also define:

$$X_{\mathrm{p}}^N(l,u) = \left\{ x \in X(l,u) \colon \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{[\mathtt{lb},\mathtt{ub}]}(c_2(x,\xi^i) - c_1(x,\xi^i)) \geq \mathrm{p} \right\}.$$

Then:

$$\mathbb{P}\left[X_{\mathrm{p}}^N(l,u) \subseteq X_{\mathrm{p}}\right] \geq 1 - \left\lceil \frac{DL}{\beta} \right\rceil^m \exp\{-2N(1-\mathrm{p}-\gamma-\beta)^2\}.$$

◀

In other words, one can now have an estimation of the sample size necessary to reach a confidence level on the inclusion of the feasible set $X_{\mathrm{p}}^N(l,u)$ into the feasible set X_{p} of our original problem: let $\delta \in (0,1)$ be this desired confidence level. Then, if N verifies:

$$N \geq \frac{1}{2(1-\mathrm{p}-\gamma-\beta)^2}\log\left(\frac{1}{1-\delta}\right) + \frac{m}{2(1-\mathrm{p}-\gamma-\beta)^2}\log\left(\left\lceil \frac{DL}{\beta} \right\rceil\right),$$

the probability of having $X_{\mathrm{p}}^N(l,u) \subseteq X_{\mathrm{p}}$ wille reach at least the desired confidence level δ.

Assumption (7) shall not be seen as too restrictive, as for instance all random vector ξ with quasi-concave distributions have Lipschitz-continuous cumulative distribution functions if and only if none of the components ξ_i has zero variance (see [147, Theorem 2.2]). The latter reference also highlights that the class of quasi-concave multivariate distributions is very large, and provides some examples of such distributions.

7.7.2 Convergence of the SAA solution to the original value

Thanks to [232], results from the last section can be improved. In this latter work, the authors also study a joint chance constrained problem:

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & \mathbb{P}\left[G(x,\xi) \leq 0\right] \geq \mathrm{p}, \end{aligned} \tag{7.31}$$

where f is a continuous real-valued function and G is a given constraint function from $\mathbb{R}^n \times \Xi$ to $\mathbb{R}^m$ without any other assumption. It is clear that our problem $\mathcal{P}_{\mathrm{p}}$ fits into this setting.

Let us add some introductory material presented in [232]:

- For two sets $A, B \subset \mathbb{R}^n$, the *deviation* of set A from set B is defined as $\mathbb{D}(A,B) := \sum_{x \in A} \mathrm{dist}(x,B)$. One can immediately observe that if $\mathbb{D}(A,B) = 0$, then $A \subseteq B$.
- S^N and S are defined as the set of optimal points for problems $\mathcal{P}_{\mathrm{p}}^N$ and $\mathcal{P}_{\mathrm{p}}$ respectively.

We here recall an important assumption used in [232]:

Assumption 8 (Assumption (A) in [232]). There is an optimal solution $\bar{x}$ of the true problem $\mathcal{P}_{\mathrm{p}}$ such that for any $\epsilon > 0$ there is $x \in X$ with $\|x - \bar{x}\| \leq \epsilon$ and $\mathbb{P}\left[\mathtt{lb} \leq c_1(x,\xi) - c_2(x,\xi) \leq \mathtt{ub}\right] \geq \mathrm{p}$.

With this assumption, the following theorem is proposed in [232]:

Theorem 16 (Proposition 2 in [232]). Assume that X is compact, f is continuous, $G(\cdot,\xi)$ is continuous for almost every $\xi \in \Xi$, $G(x,\cdot)$ is measurable for every $x \in \mathbb{R}^n$, and Assumption 8 holds.

Then $z^N \to z^*$ and $\mathbb{D}(S^N, S) \to 0$ with probability 1 as N goes to infinity. ◀

Theorem 16 merely states that as the sample size goes to infinity, provided that there exists an optimal point $\bar{x}$ for the true problem $\mathcal{P}_{\mathrm{p}}$ that is an accumulation point, then the optimal value and the set of optimal solutions for the SAA problem converge to their true counterparts.

Remark 23. We have directly applied results to our DoC setting with a random right-hand side. These direct applications prove that the SAA method is theoretically useful as the approximated solutions that it provides converge to solutions of the true problem. ▷

Note that we will be using SAA on the following problem:

$$\begin{aligned} \min_{x \in X} \quad & f_1(x) - f_2(x) \\ \text{s.t.} \quad & \frac{1}{N} \sum_{j=1}^{N} \zeta^t(c_1(x,\xi^j) - c_2(x,\xi^j)) \geq \mathrm{p} \end{aligned} \tag{7.32}$$

where ζ^t is as defined in equation (5.12). Due to this additional approximation, another step is needed in order to ensure the proper convergence of the SAA method. To complete this step, one can turn to [153, Theorem 1] where the authors prove that the approximation ζ^t of the indicator function is asymptotically exact as t positively goes to zero. This proves the equivalence of Problems 7.32 and $\mathcal{P}_{\mathrm{p}}^N$ (as defined by the authors of [153]).

7.8 CONCLUSION

In this chapter we have presented and discussed two classes of models for a DSO oriented OPF problem. Although the first model is not satisfactory, it has led to the second type of models that shows promise. This second type of model is a two-step approach: the first step is an optimisation over control variables, while the second one is solely on state variables. This decomposition into two steps is made available thanks to the DoC structure of our problem. Once this framework is set-up, our four models naturally arise through the use of different operators ($\mathbb{P}$ and $\mathbb{E}$). After part II where we have developed an algorithm for DoC constrained DoC programs, and part III where we have presented a DoC formulation of chance-constraints (see section 5.3), building these models is the last step before our first numerical experiments on the OPF.

NUMERICAL APPLICATIONS

8

This chapter concludes our efforts to provide a numerical tool for decision making for the DSO in a ST-OP context. It assembles two main facets of our work:

- a model proposition, that is built on our initial understanding of the operational ambitions (see Part I), explicitly presented in Chapter 7.
- a solving methodology that relies on several mathematical elements: DoC programming and a new generic algorithm for a large class of problems (see Chapter 4), as well as a better understanding on chance-constraints in Chapter 5.

The main objective of this Chapter is to provide a "proof-of-concept" of our methodology to tackle an OPF under uncertainties. This (first) attempt should evidently be improved, and several possible extensions.

Our ambition is to test and compare different existing DoC algorithms for our models. For these mixed comparisons of models and algorithms, it is necessary to define discriminating criteria; this is addressed in the first section of this present Chapter.

As is common in non-convex, nonsmooth optimisation, results are significantly influenced by parametrisations and initialisations. We are interested in these parameters and their possible combinations. Due to the number of parameters (there are at least 10 different parameters in `PBMDC`[2] for instance), expecting an omniscient and extensive overview of their combinations is unrealistic. In the second section of this chapter we discuss which parameters we have selected. The third section is dedicated to the numerical strictly speaking. All numerical results are produced using MATLAB2019b, Gurobi 9.1.1, Windows 10 Professional, 16Go and an Intel i7-6820 processor (4 cores).

8.1 DEFINITION OF OUR CRITERIA

We here present our criteria that we then use to determine the quality of the resolutions of a model by an algorithm with a selected combination of parameters. Time of computing (CPU time), and number of iterations of the DoC algorithms are reported. We also include the objective and constraints values at the initial point (x^0) as well as at the last serious step ($\bar{x}$). These values respectively denote $f(x^0)$, $c(x^0)$, $f(\bar{x})$ and $c(\bar{x})$. We also have included the number of scenarios for which decisions x^0 and $\bar{x}$ lead to no electrical constraints. These number of "covered" scenarios are denoted N_{SC}^0 and $\overline{N_{SC}}$ respectively.

The notion of "covered" scenarios can be misleading in our context, as our objective is not strictly speaking to increase $\overline{N_{SC}}$ to a level deemed sufficient by the operator. We recall that we use scenarios for the numerical evaluations of probabilistic functions, which in turn appear in our optimisation models. Algorithms then optimise values of these latter functions, which can in some cases happen to be related to a number of covered scenarios and not in others. We care to emphasize that the analogy between the number covered scenarios and values of our probabilistic functions is false in general, and should be studied with attention.

Moving slightly into more details, the reasons why there are differences between the number of covered scenarios and the values of probabilistic functions are distinct depending on the probabilistic operator at hand: $\mathbb{E}$ or $\mathbb{P}$.

In the former case, one can easily conceive the existence of these differences: see Example 6 for a simple explanation. In the latter case, as we do make use of an approximation of the probabilistic operator $\mathbb{P}$ (with a user given precision parameter) the number of covered scenarios can evidently be different from the computed value of the approximation of the probability. Even when a significant number of scenarios is used for the numerical computation, a loosely chosen precision parameter for our approximation will undoubtedly lead to poor analogy between the number of covered scenarios and the probability values.

Example 6 (Highlights on the misleading aspect of the relation between the number of covered scenarios and the expected value of power imbalances.)**.** Let us consider a set of 7 scenarios, two decisions x_1, x_2 and their associated imbalance values (sum of squared distances to 0 in the power equilibrium constraints equations (7.1b) and (7.1c) for instance) reported in Table 8.1.

Scenario	1	2	3	4	5	6	7	Expected Value
x_1	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}
x_2	0	0	0	4.10^{-4}	0	0	0	3e-4

Table 8.1: Example of power imbalances for 7 scenarios for two decisions x_1 and x_2.

One can immediately observe that decision x_1 has a lower expected value although no scenario is covered, while decision x_2 covers 6 scenarios out of the 7 but has an expected value three times greater than the one related to x_1. ▶

8.2 CHOICE OF SIGNIFICANT PARAMETERS

In Table 8.2 we recall the list of parameters that can be modified. From *a priori* experiments, and in order to limit the combinatorial possibilities we want to investigate, we claim that our algorithms are significantly less sensible to some parameters than others.

We start by the parameters which values seem to have a low impact in our numerical experiments. Different values of κ, μ^0 and the maximal bundle size, once this value is above a threshold, seem to have little impact on the resolution of our problems. As a consequence, we set $\kappa = 0.5$, $\mu^0 = 1$ and the maximal bundle size to 500.

Remark 24 (About the maximal bundle size parameter)**.** As a reminder, the maximal bundle size parameter retains the amount of information we allow to gather along `PBMDC` and `PBMDC`^{2}. While the number of linearisations in the bundle is less than this value, we allow for the addition of new linearisations at each iteration. Once (i) the bundle size is above its maximal value and (ii) a serious step is made we can in fact solely keep the information of the last iteration. Another possibility, which we use in our numerical experiments, is to keep all active linearisations and discard the other ones. Note that this rule implies the first one: the linearisation from the last iteration obviously is active (provided that the sub-problem is exactly solved).

Parameter	Constraint	Comment
ρ	$\rho \geq 0$	PBMDC^2
σ	$\sigma \in [0,1)$	PBMDC^2
κ	$\kappa \in (0,1)$	PBMDC^2
δ_{Tol}	$\delta_{\text{Tol}} \geq 0$	PBMDC^2 and PBMDC
$\mu_{\max}, \mu_{\max}$	$\mu_{\max}, \mu_{\min} \geq 0,\ \mu_{\min} \leq \mu_{\max}$	PBMDC^2 and PBMDC
μ^0	$\mu^0 \in [\mu_{\min}, \mu_{\max}]$	PBMDC^2 and PBMDC
μ update rule	Defined in PBMDC^2	PBMDC^2 and PBMDC
Maximal bundle size	Strictly positive value	PBMDC^2 and PBMDC
p	p $\in (0,1)$	For models with operator $\mathbb{P}$
rhs	rhs ≥ 0	For models with operator $\mathbb{E}$
Initialisation	Initial point must verify linear bounds	For all models
λ_c	$\lambda_c > 0$	Scaling factor for the constraints
DoC decompositions		For all models

Table 8.2: List of parameters to set.

Once this parameter value is greater, in order of magnitude, than 100, then we have experienced that it has few impacts on the resolutions. Let us also add that as a general rule, the largest the bundle the longer is the computation time: each subproblem has a higher cost to solve. Nevertheless, in our case, at each iteration the step with the highest computation cost is by far the oracle evaluation. Solving the subproblem has a comparatively very low computation time, which implies that we can allow large bundles: we do not gain any significant computing time by reducing the maximal bundle size. ▷

The remaining parameters have more significant impact on the outcomes. On the algorithmic side:

- δ_{Tol} impacts the solution for obvious reasons, and has been set to 10^{-7} for all our experiments.
- ρ is critical for the rate of reaching a feasible point: in general, the larger it is, the faster PBMDC^2 reaches feasibility. This is easily understandable as, by design of the improvement function we use, ρ provides weight to infeasibility. With a large weight, the optimality criterion is lessened relatively to the feasibility criterion until feasibility is reached. Once reached at a step k, ρ has no impact from $k+1$ until the end of PBMDC^2.
- $\mu_{\max}, \mu_{\min}$ are upper and lower bounds on the weight given to the proximal term in the subproblems of PBMDC^2. With a large weight, computing the next iterate will more likely provide a solution close to the last serious iterate (also called "stability center" in literature). Conversely, a small weight reduces this attraction to the last serious iterate. In order to reach the stopping condition rapidly, and avoid long computation tails, one should aim at large $\mu_{\max}$. We set $\mu_{\max} = 10^5$, and $\mu_{\min} = \dfrac{1}{\mu_{\max}} = 10^{-5}$, keeping in mind Gurobi's advice on admissible numerical

scales (i.e. avoid large differences in order of magnitudes, and for right hand sides of the linear constraints avoid orders of magnitude greater than 10^4).

- Our update rule on μ is based on the behaviour the observed of PBMDC^2: in general, more null steps are achieved than serious ones. As a rule defined for PBMDC^2, when a null step is performed μ can be updated to an greater value, while it can decrease solely after a serious one. In other words, we observe more opportunities to increase μ than decrease its value. As a consequence, we implemented a rule that increases μ after every 5 consecutive null steps, and decreases this parameter after 3 consecutive serious steps. Requiring several consecutive steps of the same type is an attempt to prevent μ from converging to its bounds "too quickly", which can be observed in practise, and let the algorithm gather more information in the bundle.

- The right hand side parameters `p` and `rhs` are part of the user-provided information, and reflect the user's requirements: they naturally have a huge impact on the solution.

- As is usual in nonsmooth nonconvex optimisation, initialisation of a problem impacts its solution. An initialisation in the neighbourhood of a critical point can lead the algorithm to stop at this point, which can be unobserved when initialising at another point. This expectation holds in our case.

- Finally, we recall that in general DoC decompositions have an impact on the solutions reached. As we do not know how to discriminate *a priori* a DoC decomposition from another, this also appeared to be a significant parameter to control.

8.3 OBSERVATIONS ON THE INITIALISATION

Aside from the particular structure that our problem possesses, our setting introduces a singular property: by design, there always exists a non-empty feasible set and we can easily conceive a heuristic to reach feasibility. This claim is easily ascertainable: thanks to our variables of service not delivered p^v, q^v, it is theoretically possible to set all interactions with the grid to zero which is a trivial feasible point. Building upon this observation, we can derive a heuristic to find a feasible point. To this end, we define the *sign matrix* A^σ as the diagonal matrix which diagonal elements have values in $\{-1, 1\}$:

$$A^\sigma_{i,i} = \begin{cases} 1 & \text{if } x_i \text{ is a decision on service not delivered} \\ -1 & \text{otherwise.} \end{cases} \tag{8.1}$$

This matrix A^σ is such that for a given decision x_1, setting $x_2 = x_1 + \alpha A^\sigma$, $\alpha > 0$ leads to a decision x_2 with lower interactions of the GUs with the electrical grid and consequently less forecasted electrical constraints, if any. Algorithm 3 is a simple heuristic to find a feasible point for instance.

Finding a feasible point is obviously theoretically interesting: as usually noted in works on nonconvex optimisation, finding a "good" initial point can have a significant advantage over a "bad" one (see for instance the short discussion in the general introduction of [283], which is focussed on generalized version of DCA). In our case, we could

Algorithm 3 A simple heuristic to find a point within the feasible set

Step 0: Initialization. Define a step length $\tau > 0$, set $k = 0$ and A^σ the sign matrix as defined in equation (8.1). Let x^0 be a zero vector of the appropriate size.

Step 1: Stopping test. Compute a load-flow accounting for decision x^k. If no electrical constraint is observed, then stop and return x^k.

Step 2: Incrementation. Set $x^{k+1} = x^k + \tau A^\sigma$.

Step 3: Loop Set $k := k + 1$ and go back to Step 1.

leverage PBMDC^2 's property of maintaining feasibility while decreasing the value of the objective function: nevertheless, this has not been useful in our case, and leave this interesting property for future developments.

8.4 NUMERICAL RESULTS FOR OUR MODELS OF PROBABILISTIC OPF

We looked into different randomly generated cases on a 33 nodes topology. The GUs selection and the scenario generation used to approximate either the expected value function or the probabilistic one are all initialized to a single fixed seed, arbitrarily fixed to 10[1]. The following tables report results for our four different models, and present combinations of different parametrisations/choices.

First of all, we compare the following DoC algorithms:

- our PBMDC^2;
- algorithm PBMDC from [228], where the DoC constraints are penalized.

Remark 25. We have decided to discard both DCA and the convex-concave procedure (CCP, from [197]) algorithms, as it turned out to be inefficient: at each iteration, it is required to solve a DoC problem which is too time-consuming. Trying to develop a tailored approach using relying on either DCA or the convex-concave procedure is probably still possible, but requires some more involved work and possibly smooth reformulations in the subproblems: we believe that it falls beyond the scope of our work. ▷

We have considered the following input parameters:

- Initialisation: as is usual in nonconvex optimisation, using different initial points can lead to different outputs. As a consequence, we consider 3 types of initialisation, referred to as "Zero", "Rand" and "Mean". The first two types are direct affectations: "Zero" is the zero vector of appropriate size (i.e. the DSO does not take any decision), while "Rand" is a random vector taken in-between the upper-bound and the lower-bound. Lastly, "Mean" is the point obtained from solving a deterministic OPF where the vector ξ is replaced by its mean value. Note that the initial point, in this case, can be unfeasible which is visible from the numerical results.

[1] As such, we claim that our cases cannot be deemed *a priori* "favourable" for our algorithm.

- Right-hand-side `rhs`: this parameter is a user-given one, and has different interpretations depending on the model in use. For problems Min-$\mathbb{E}$ and Max-$\mathbb{P}$, `rhs` is the allocated maximal cost. For problem (Const-$\mathbb{E}$), `rhs` is the maximal expected value of the recourse function $\mathcal{R}$. For problem (Const-$\mathbb{P}$), `rhs` is the minimal probability of observing no electrical constraints.

In order to study the behaviour of our algorithms, we consider the following criteria:

- the initial values of the objective function and the DoC constraint function;
- the final values of the objective function and the DoC constraint function;
- the number of "covered" scenarios at the solution point;
- the number of steps in `PBMDC` and `PBMDC`^{2};
- the time of computing for `PBMDC` and `PBMDC`^{2};

8.4.1 Numerical results for problem (Min-$\mathbb{E}$)

Recall that problem (Min-$\mathbb{E}$) aims at minimising the expected value of the recourse function $\mathcal{R}$, whose interpretation is found in Insert 8.

Table 8.3: Numerical results for problem (Min-IE). The number of scenarios generated is $N = 10^3$. CPU time limit is set to 7200s. The number N_{SC}^0 is the number of covered scenarios with the initial point, $\overline{N_{SC}}$ being the number of covered scenarios with the solution point.

Init	rhs	$f(x^0)$	$c(x^0)-$rhs	$f(\bar{x})$		$c(\bar{x})-$rhs		$\overline{N_{SC}}$		CPU (s)		Iterations	
				PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC
Zero	10^{-3}	$3.0347.10^{-2}$	-10^{-3}	$2.3921.10^{-4}$	$2.5627.10^{-4}$	$-7.5503.10^{-9}$	$-8.617.10^{-5}$	948	982	3848	2511	141	127
	10^{-4}	$3.0347.10^{-2}$	-10^{-4}	$2.5845.10^{-4}$	$2.6724.10^{-4}$	$-2.2545.10^{-7}$	$1.000.10^{-7}$	827	872	2511	820	1151	114
	10^{-5}	$3.0347.10^{-2}$	-10^{-5}	$3.4800.10^{-3}$	$1.8575.10^{-3}$	$-1.379.10^{-11}$	$1.009.10^{-9}$	545	643	2725	1827	2593	1446
Rand	10^{-3}	$5.4871.10^{-1}$	$1.9224.10^{1}$	$2.3995.10^{-4}$	$2.6234.10^{-4}$	$-6.0173.10^{-9}$	$6.7548.10^{-5}$	943	977	2359	1897	148	127
	10^{-4}	$5.4871.10^{-1}$	$1.9225.10^{1}$	$2.5913.10^{-4}$	$2.6792.10^{-4}$	-2.4010^{-7}	$5.5963.10^{-6}$	818	872	4088	3899	533	415
	10^{-5}	$5.4871.10^{-1}$	$1.9225.10^{1}$	$3.4912.10^{-3}$	$1.9554.10^{-3}$	-9.2561^{-8}	$2.3695.10^{-6}$	460	660	4560	4207	1423	1418
Mean	10^{-3}	$1.5696.10^{-3}$	0	$2.3930.10^{-4}$	$2.3958.10^{-4}$	$-7.5503.10^{-9}$	$7.2540.10^{-4}$	927	945	1603	1597	211	224
	10^{-4}	$2.0140.10^{-3}$	0	$2.5846.10^{-4}$	$2.9958.10^{-4}$	$-2.2545.10^{-7}$	$5.8456.10^{-5}$	788	877	2513	1988	514	551
	10^{-5}	$1.1144.10^{-2}$	0	$3.4801.10^{-3}$	$1.9958.10^{-3}$	$-5.1589.10^{-11}$	$6.7189.10^{-7}$	560	588	3554	2651	956	1053

In this model, it is interesting to see that feasible points are nearly always reached, with PBMDC and PBMDC^2 providing different solution points. When studying the impact of parametrization, we obtain the following graphs:

- figure 8.1 is a plot of the final objective values against the maximal allowed cost;

- figure 8.2 is a plot of the number of scenarios covered at the solution point against the maximal allowed cost.

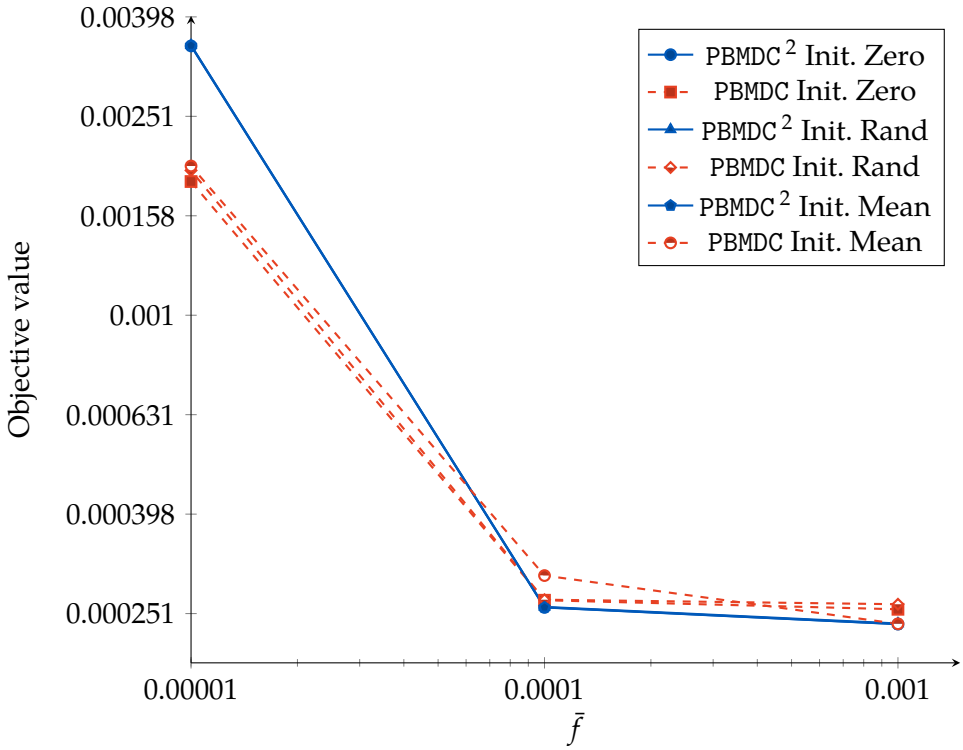


FIGURE 8.1: Final values of $\mathbb{E}[\mathcal{R}(\bar{x}, \xi)]$ as a function of $\bar{f}$. Some values for PBMDC^2 are not visible on the graph due to the very small relative differences between the values.

It is visible that the expected value of function $\mathcal{R}$ increases as the allowed cost is lowered. Moreover, it appears that PBMDC is able to find better objective values when compared to PBMDC^2, while not ensuring feasibility. This should be interpreted as a behavioural difference between an algorithm using the improvement function (as PBMDC^2), and a penalization algorithm (as PBMDC). The improvement function has the interesting property of iteratively selecting points that lower the constraint violation, until the DoC constraints are verified. Once verified, future iterates remain feasible. This is not the case of a penalization algorithm, which appears to look for "slightly" infeasible points. It does not enjoy the property of maintaining feasibility once reached.

Among our four models, and although parametrization has remained a difficult task for all of them, we believe this model to be the most robust from a numerical point of view. Another advantage of this model is that, in case the cost function is a convex function, problem (Min-$\mathbb{E}$) is a convexly-constrained DoC program. In other words, there would be no difference in applying PBMDC^2 or PBMDC as they are equivalent on DoC problems without DoC constraints.

On the downside, it is difficult to *a priori* interpret the expected values we minimize. Truly, it is clear that minimizing this expected value minimizes the expected of value of the risk of observing the grid out of its bounds for a given decision. Nevertheless, analysing a numerical value of this expected value seems impractical for the operator: for instance, its magnitude can be dependent on the size of the problem for instance. As

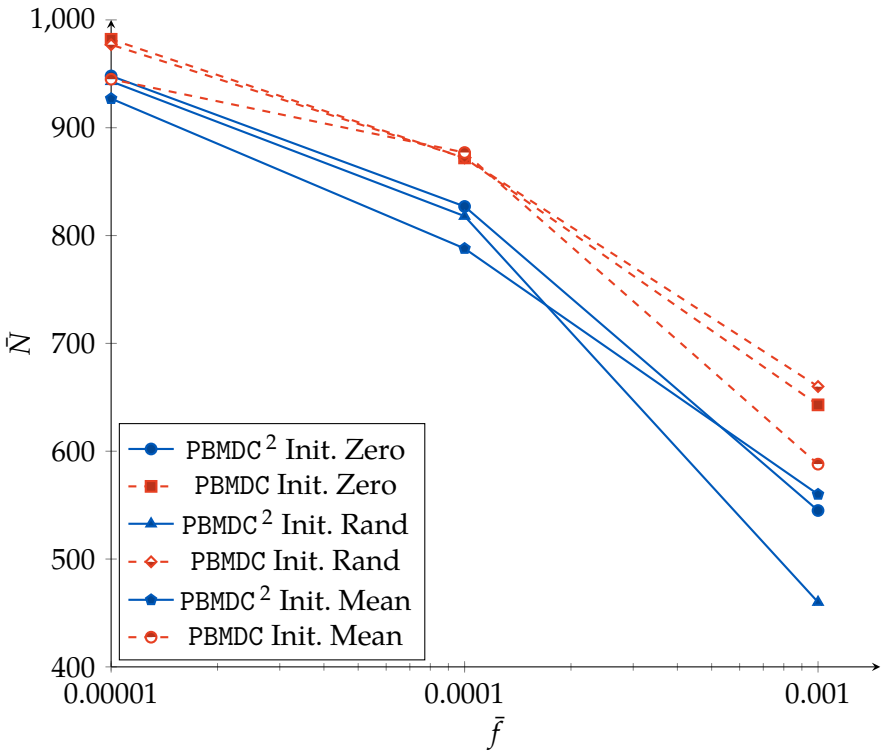


FIGURE 8.2: Number of scenario covered at the end point ($\overline{N_{SC}}$) as a function of $\bar{f}$.

a result, we believe that this model is interesting for preliminary studies on a grid, its behaviour and sensibility to input data, and didactic purposes rather than for industrial decisions for the DSO.

All in all, this first model should be regarded as an interesting, didactic study of an OPF with uncertainties. It highlights the differences in behaviours of PBMDC and PBMDC^2 regarding the feasibility of the obtained solutions. While the former provides solutions that are close to be feasible, the latter manages to reach feasibility and improve the objective function while maintaining feasibility.

8.4.2 NUMERICAL RESULTS FOR PROBLEM (CONST-$\mathbb{E}$)

In this second model there are DoC functions in the objective and the constraints. It aims at minimizing the cost function, with an upper-bound on the expected value of the second-step function $\mathcal{R}$. In other words, the goal is to reach a minimal cost with a bound on a measure of risk.

Table 8.4: Numerical results for problem (Const-E). The number of scenarios generated is $N = 10^3$. CPU time limit is set to 7200s. The number N_{SC}^0 is the number of covered scenarios with the initial point, $\overline{N_{SC}}$ being the number of covered scenarios with the solution point.

Init	rhs	$f(x^0)$	$c(x^0)-$rhs	$f(\bar{x})$		$c(\bar{x})-$rhs		$\overline{N_{SC}}$		CPU (s)		Iterations	
				PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC
Zero	10^{-2}	0	$3.0347.10^{-4}$	$2.9484.10^{-5}$	$3.1742.10^{-5}$	$-1.7270.10^{-4}$	$-8.6170.10^{-4}$	751	751	3258	3547	2411	1869
	10^{-3}	0	$3.0347.10^{-4}$	$8.1372.10^{-3}$	$5.0891.10^{-3}$	$-9.6644.10^{-7}$	$3.1878.10^{-6}$	874	852	3606	820	2960	1151
	10^{-4}	0	$3.0347.10^{-4}$	$9.1319.10^{-1}$	1.1754	$6.8227.10^{-6}$	$3.3559.10^{-4}$	990	995	4221	3014	1727	665
Rand	10^{-2}	$1.9225.10^{1}$	$2.3995.10^{-4}$	$3.2598.10^{-3}$	$3.0014.10^{-4}$	$1.2544.10^{-6}$	$6.2147.10^{-6}$	724	699	3544	3480	2746	2578
	10^{-3}	$1.9225.10^{1}$	$2.3995.10^{-4}$	$9.2156.10^{-2}$	$2.1441.10^{-2}$	$5.2245.10^{-6}$	$4.1120.10^{-6}$	855	855	3142	2877	2541	1352
	10^{-4}	$1.9225.10^{1}$	$2.3995.10^{-4}$	$9.4945.10^{-1}$	2.8458	$-4.2521.10^{-11}$	$2.2455.10^{-5}$	990	987	4752	3857	2033	1257
Mean	10^{-2}	$5.5697.10^{-5}$	0	$3.5247.10^{-5}$	$3.9475.10^{-4}$	$-4.9624.10^{-6}$	$-9.8874.10^{-4}$	745	760	3614	3101	2122	2035
	10^{-3}	$8.2700.10^{-4}$	0	$1.1144.10^{-2}$	$9.4174.10^{-3}$	$-2.2241.10^{-4}$	$1.3477.10^{-5}$	880	845	2877	2413	1874	699
	10^{-4}	$2.4770.10^{-2}$	0	1.2456	2.8451	$-9.9775.10^{-7}$	$1.1247.10^{-5}$	994	925	3510	2747	3569	1661

Similarly to the previous subsection, we plot the following graphs:

- figure 8.3 is a plot of the final objective values against the maximal allowed risk;
- figure 8.4 is a plot of the number of scenarios covered at the solution point against the maximal allowed risk.

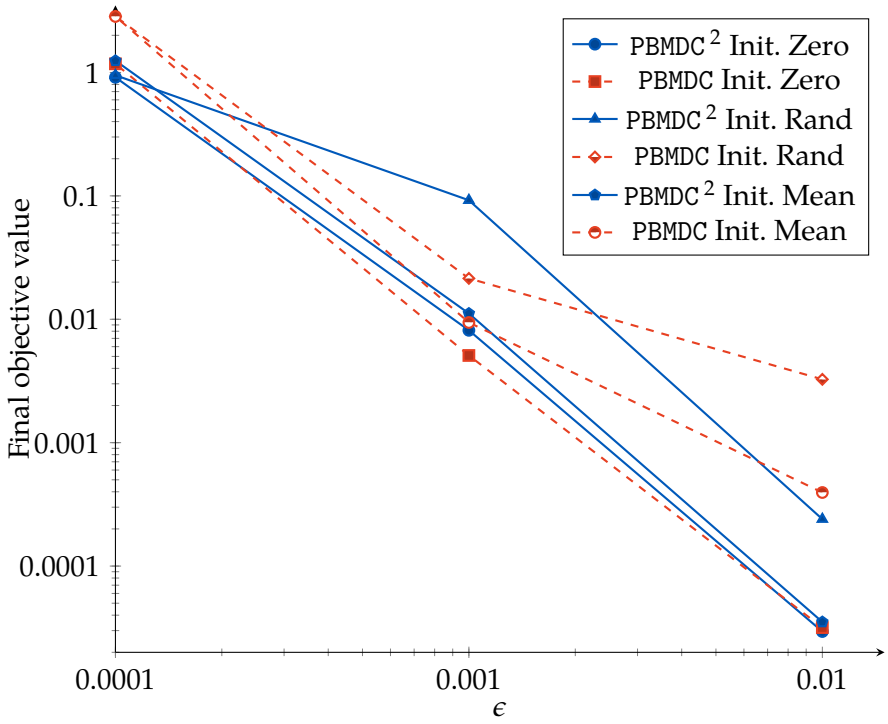


FIGURE 8.3: Values of $\mathbb{E}[\mathcal{R}(\bar{x}, \xi)]$ as a function of ϵ.

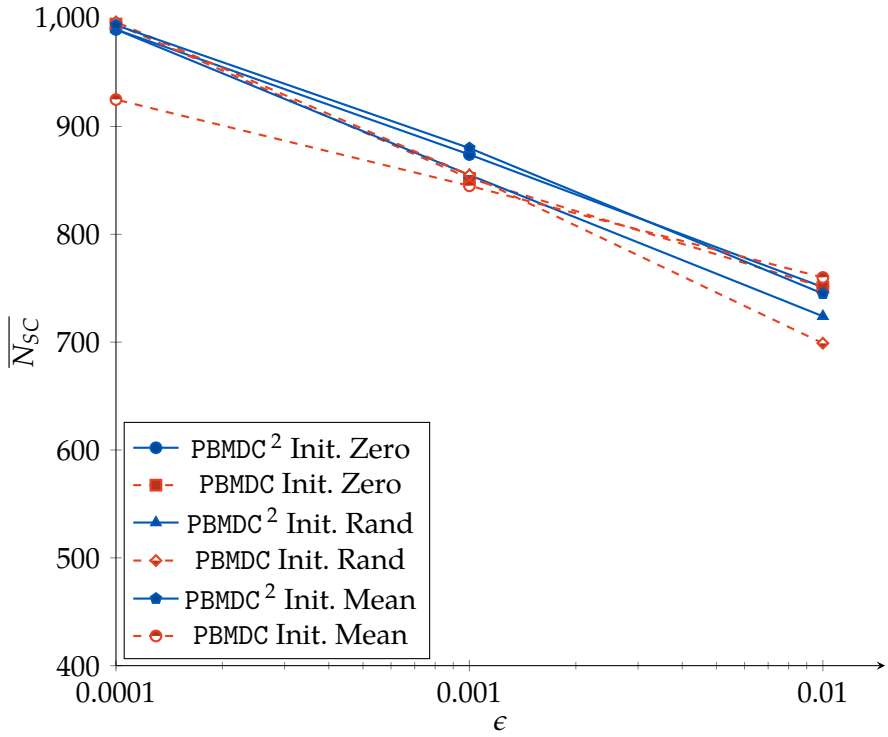


FIGURE 8.4: Number of scenario covered at the end point ($\overline{N_{SC}}$) as a function of ϵ.

Again, as visible in figure 8.3, when the allowed risk is increased, the cost is lowered which is expected. In figure 8.4, we present the plot that relates our "measure of risk" (the upper-bound on the expected value of the second-step function $\mathcal{R}$) to the number

of scenarios where no electrical constraints are observed. First of all, we recall that this notion of "covered scenarios" should be taken with care: this concept should only be regarded as an inexact proxy to measure the quality of a solution. Having noted this remark, figure 8.4 emphasizes the difficulty of analysing a risk using the expected value operator, in particular in this model. In fact, problem (Const-$\mathbb{E}$) requires that the operator provides an upper-bound on the expected value of the second-step function. This value is an absolute one, which implies that the operator is required to know if the acceptable risk is to be set at 10^{-2} or 10^{-4}. In general, it is not possible to *a priori* discriminate between these different choices without an extensive experience. We are only *a posteriori* able to say that setting the upper-bound to 10^{-4} can provide solution where close to all the scenarios are covered, while setting the upper-bound to 10^{-2} can provide solution where between 700 and 800 of the 1000 scenarios are covered.

In this model, it is visible that `PBMDC`^{2} has a more interesting behaviour than `PBMDC`: the end points are nearly all feasible, while `PBMDC` stops mainly at non-feasible points (although close to being feasible). One explanation is that problem (Const-$\mathbb{E}$) is somewhat better matched with `PBMDC`^{2} than with `PBMDC`. Differently to problem (Min-$\mathbb{E}$) which could be in a special case written as a convexly-constrained DoC program, problem (Const-$\mathbb{E}$) inherently is a DoC constrained DoC program.

8.4.3 Numerical results for problem (Max-ℙ)

This subsection and the next one are dedicated to models with the probabilistic operator ℙ. We recall that problem (Max-ℙ) is a maximisation problem: the objective is to maximise the probability of observing the grid within its bounds, while ensuring that the cost is upper bounded.

Table 8.5: Numerical results for problem (Max-ℙ). The number of scenarios generated is $N = 10^3$. CPU time limit is set to 7200s. N_{SC}^0 is the number of covered scenarios with the initial point, $\overline{N_{SC}}$ being the number of covered scenarios with the solution point. Recall here that this is a maximisation problem.

Init	rhs	$f(x^0)$	$c(x^0)-$rhs	$f(\bar{x})$		$c(\bar{x})-$rhs		$\overline{N_{SC}}$		CPU (s)		Iterations	
				PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC
Zeros	10^{-3}	0.445	-10^{-3}	0.990	0.993	$-3.1421.10^{-6}$	$1.4114.10^{-4}$	990	992	1422	1341	167	222
	10^{-4}	0.445	-10^{-4}	0.845	0.855	$-6.1895.10^{-6}$	$1.5917.10^{-4}$	845	855	1894	1445	268	551
	10^{-5}	0.445	-10^{-5}	0.640	0.689	$-1.0114.10^{-8}$	$3.7541.10^{-6}$	640	742	2941	1925	1446	2593
Rand	10^{-3}	0.910	$1.9215.10^{1}$	0.745	0.802	$4.6652.10^{-4}$	$6.4721.10^{-4}$	745	802	1945	1554	1342	1024
	10^{-4}	0.910	$1.9224.10^{1}$	0.701	0.710	$3.1014.10^{-6}$	$6.4721.10^{-4}$	701	710	1844	1719	985	714
	10^{-5}	0.910	$1.9225.10^{1}$	0.560	0.560	$7.9178.10^{-8}$	$6.9824.10^{-8}$	560	560	1123	1463	1056	564
Mean	10^{-3}	0.851	0	0.990	0.990	$-3.1420.10^{-6}$	$-3.1200.10^{-6}$	990	990	856	841	189	144
	10^{-4}	0.760	0	0.845	0.850	$2.5478.10^{-4}$	$9.1247.10^{-5}$	845	990	1422	1014	714	213
	10^{-5}	0.464	0	0.651	0.684	$3.4100.10^{-5}$	$7.6600.10^{-6}$	651	684	2521	1498	698	532

We plot in figure 8.5 the objective value as a function of upper-bound on the allowed cost. As a general trend, it is visible as expected that the greater is the allowed cost, the higher is the probability of observing the grid within its bounds. In more details, this

model highlights the importance of initialisation: the random initialisation provides significantly better end points ($\sim$ 20% improvement). Interestingly, this model is the only one to have this particularity: in other models, initialising with the "mean value", or even with the "zero vector", provides end points that are at least as good as end points obtained from the random initialisation.

Further studying the impact of the random initialisation, one can observe that this initial vector more often than not verifies the two following items:

- It leads to a probability of the second-step function being lower or equal than 0 close to 1. The rationale lies in the two following remarks: (1) non-zero decision in our model often leads to better electrical states, (2) as zero is a bound on close to all the decisions the DSO can take, when generating the random initial vector we often end up with a non-zero decision.
- It leads to a decision with a high cost, and this is explained similarly to the previous point.

These two remarks are visible in table 8.5 (see columns $f(x^0)$ and $c(x^0)-\texttt{rhs}$). They imply that with the random initialisation, the algorithms start at a "highly" infeasible point with an objective function close to 1 (which is the upper bound on its value).

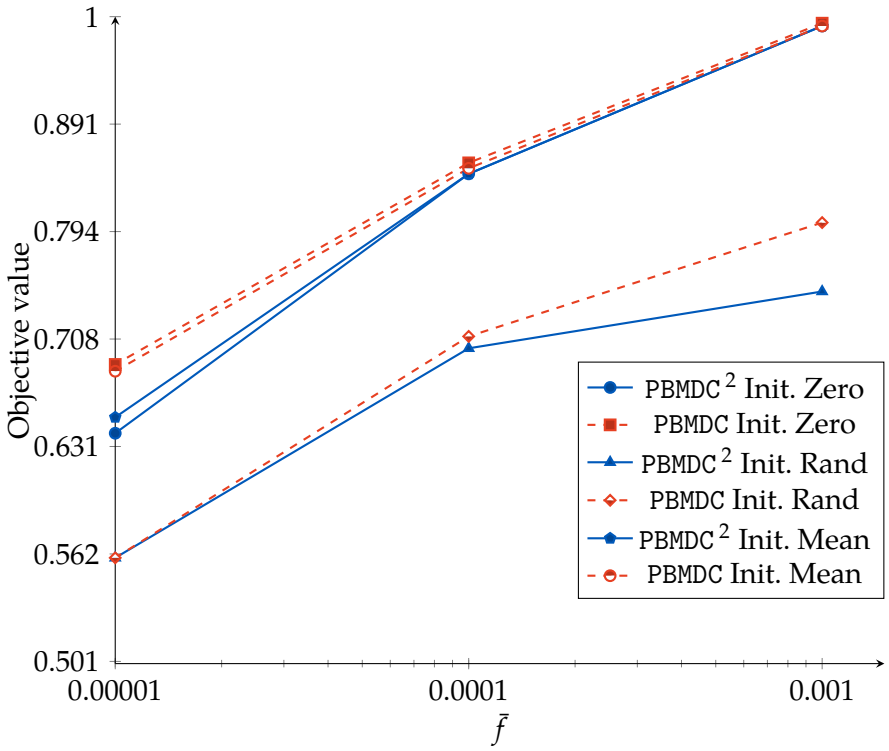


FIGURE 8.5: Objective values at the end point as a function of $\bar{f}$.

8.4.4 Numerical results for problem (Const-$\mathbb{P}$)

This last model is a chance-constrained OPF. The aim is to minimise a cost function, with a constraint on a probabilistic function. In other words, the goal is to provide the decision with the lowest cost that ensure the grid state is respected in probability.

Table 8.6: Numerical results for problem (Const-$\mathbb{P}$). The number of scenarios generated is $N = 10^3$. CPU time limit is set to 7200s. N_{SC}^0 is the number of covered scenarios with the initial point, $\overline{N_{SC}}$ being the number of covered scenarios with the solution point.

Init	rhs	$f(x^0)$	$c(x^0)$	$f(\bar{x})$		$c(\bar{x})$		$\overline{N_{SC}}$		CPU (s)		Iterations	
				PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC	PBMDC2	PBMDC
Zero	0.95	0	0.445	$6.9244.10^{-2}$	$1.1475.10^{-1}$	0.950	0.931	950	930	1952	1757	1744	1256
	0.90	0	0.445	$5.4631.10^{-3}$	$5.8744.10^{-3}$	0.914	0.861	914	861	1394	820	1568	1151
	0.80	0	0.445	$3.4800.10^{-3}$	$2.1410.10^{-3}$	0.833	0.758	833	756	1827	2725	1446	2593
Rand	0.95	$1.9225.10^{1}$	0.910	$7.5510.10^{-2}$	$7.4951.10^{-2}$	0.990	0.941	988	940	1655	1420	1452	866
	0.90	$1.9225.10^{1}$	0.910	$6.2540.10^{-3}$	$6.1540.10^{-3}$	0.924	0.885	902	891	1445	1225	1952	1223
	0.80	$1.9225.10^{1}$	0.910	$4.9225.10^{-3}$	$4.9225.10^{-3}$	0.842	0.791	804	895	2582	1749	2257	1977
Mean	0.95	1.10^{-3}	0.851	$4.1142.10^{-2}$	$9.5814.10^{-2}$	0.950	0.912	950	912	1655	1258	965	827
	0.90	1.10^{-4}	0.760	$5.4010.10^{-3}$	$5.1600.10^{-3}$	0.905	0.865	905	865	2245	1985	855	789
	0.80	1.10^{-5}	0.464	$2.9182.10^{-3}$	$2.4258.10^{-3}$	0.811	0.755	810	755	3266	2561	468	744

In figs. 8.6 and 8.7 we plot the final costs and final probability against the upper-bound on the chance-constraint (denoted p).

Figure 8.7 highlights that PBMDC is not able to provide a feasible end point: the chance-constraint is always violated. On the other hand, PBMDC^2 provides end points that respect the probabilistic constraints, even providing over-conservative solutions (when the

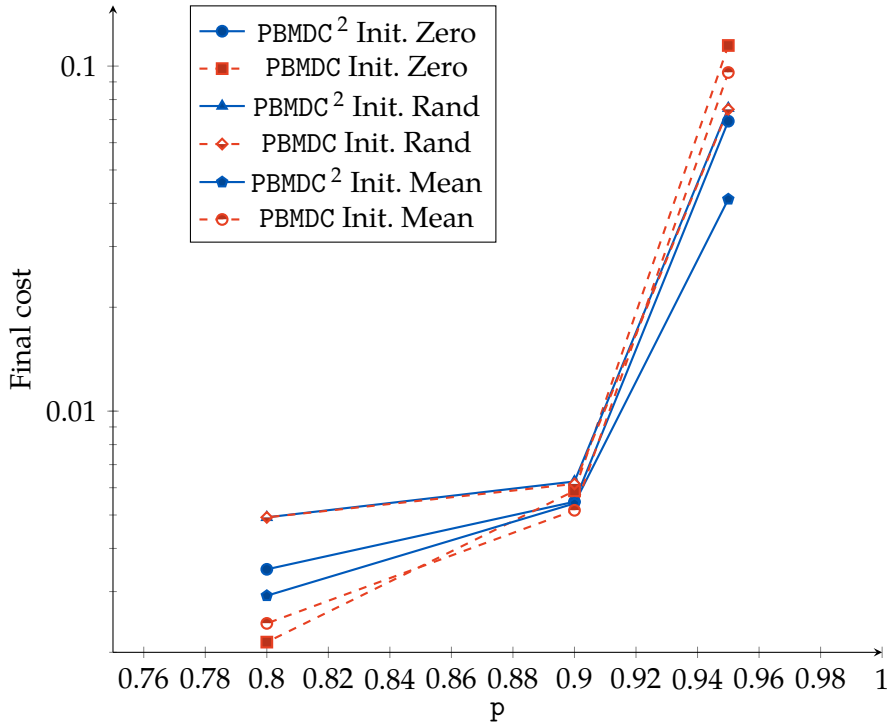


FIGURE 8.6: Final costs as a function of p.

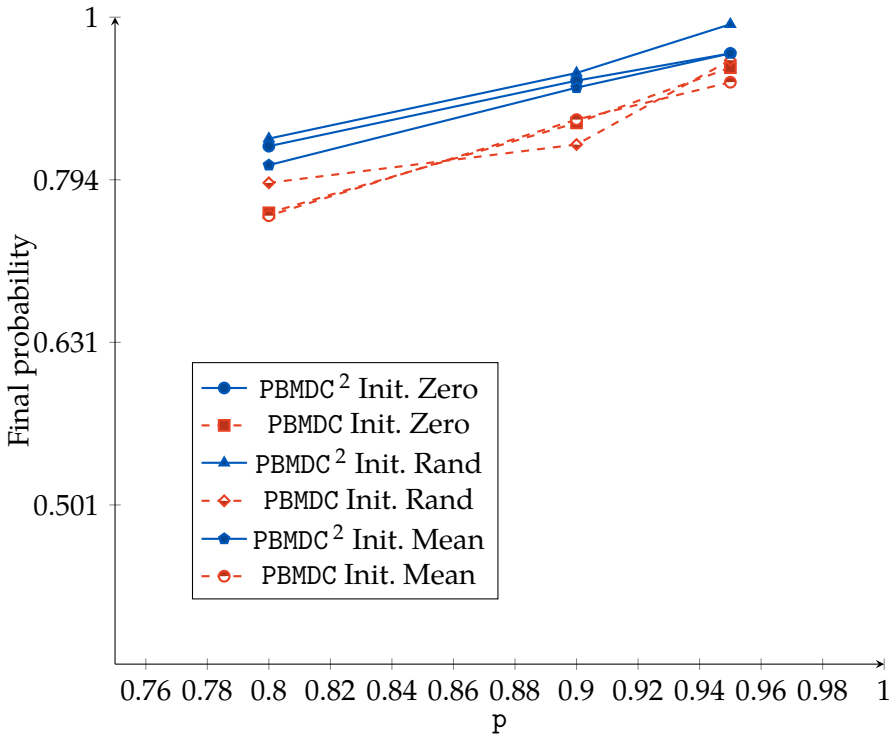


FIGURE 8.7: Final probability values as a function of p.

constraint is strictly verified). Similarly to problem (Const-$\mathbb{E}$), problem (Const-$\mathbb{P}$) is better matched with PBMDC^2 than with PBMDC probably because it inherently is a DoC constrained DoC program.

8.5 DISCUSSION ON OUR RESULTS

Immediate observations from our first numerical results can readily be made, and are often aligned with possibly *a priori* analysis. We discuss in this section our four models and compare the two algorithms we used for this test. We start by commenting the differences between the models. We then discuss general and important aspects of the resolutions we present. Finally, although the two algorithms PBMDC and PBMDC^2 share similarities in that they both make use of bundles, rely on similar oracles and have somewhat resembling convergence properties, we found out some significant distinctions that further justify the advantage of PBMDC^2.

8.5.1 COMPARISONS OF OUR MODELS BASED ON THE NUMERICAL RESULTS

We start by recalling that comparisons between models should primarily be conducted by observing their capacity of providing a useful answer to a given question. The objective of our work is to provide to the DSO a decision analysis tool for the short-term operational planning. The decision is required to be:

- cost-optimal, due to the status of the DSO;
- explicitly control a risk measure.

Following this observation and reminders, we here gather our comments on our models.

First of all, it is clear that not all of our models match these two requirements. More precisely, problems Min-$\mathbb{E}$ and Max-$\mathbb{P}$ do not look for cost optimal solutions. On the other hand, problems Const-$\mathbb{E}$ and Const-$\mathbb{P}$ do look for cost-optimal solutions. Secondly, the measurement of risk is different between problems Min-$\mathbb{E}$ and Const-$\mathbb{E}$ and problems Max-$\mathbb{P}$ and Const-$\mathbb{P}$: while the former models use the operator $\mathbb{E}$, the latter ones use the operator $\mathbb{P}$. The operator $\mathbb{P}$ enables us to have a better understanding of the risk. A first, immediate justification for this fact is simply because it is humanly easier to manipulate probabilities than abstract values whose order of magnitudes are modified from one grid to another.

As a consequence, we claim that problem (Const-$\mathbb{P}$) is the best model with respect to our two main requirements discussed at the beginning of this section. Now we care to highlight the main positive aspects of the other models:

- Problems Min-$\mathbb{E}$ and Const-$\mathbb{E}$ are easier to implement than problems Max-$\mathbb{P}$ and Const-$\mathbb{P}$.
- If *a priori* to the optimisation step the operator only has a cost information (for instance the operator only knows the maximum allowed cost), then problems Min-$\mathbb{E}$ and Max-$\mathbb{P}$ are well suited.
- Problems Const-$\mathbb{E}$ and Const-$\mathbb{P}$ are better suited when the operator has knowledge on the maximal allowed risk.

8.5.2 Analysis of the time spent in PBMDC^2

The computing time in a DSO context is important: a proper answer should provide a solution point well under thirty minutes and ideally under ten minutes. The rationale is that this solution has to be studied by an operator, and other optimisation can possibly be relaunched before real-time. A first observation is that in general, our current version of the algorithm cannot meet this requirement as a run time can be greater than two hours when parameters are not properly set. We believe that the choice of parameters for PBMDC and PBMDC^2 have a significant impact on its efficiency when different input data sets (i.e. grid parameters and forecasts) are provided. This postulate is expected from non-convex optimisation insights, and verified in our case. We highlight that we have little knowledge on *a priori* definition of a "good" choice of parameters.

Taking a step back, the time of computing can be split into:

1. the time spent in the oracle;
2. the time spent in the subproblem of PBMDC^2.

We would like to highlight that the time required for steps with lower complexity (such as for bundle management or descent test) are not considered here. These steps have linear complexity, and the time spent in them is significantly lower than the time spent in the two previously identified steps.

Solving the second-step (see section 7.5.5 for a description of the second step) turns out to be very efficient using readily available solver. We recall that the second step is comprised of a number of small subproblems (one per generated scenario). We also recall that these subproblems are (linearly constrained) quadric programs whose size is bounded thanks to the bundle management step of PBMDC^2. On the other hand, calls to the oracle are more costly. To highlight this last result, we consider our 33 nodes case and use 1000 scenarios; we then call the oracle on 100 randomly selected points. In order to observe decision points with low as well as high probabilities of leading to electrical constraints, the random selection is as follows:

$$x_{\text{rand}} = \texttt{lb} + (ub - lb)\zeta, \tag{8.2}$$

with ζ a Gaussian random variable of mean value 0 and variance 0.1.

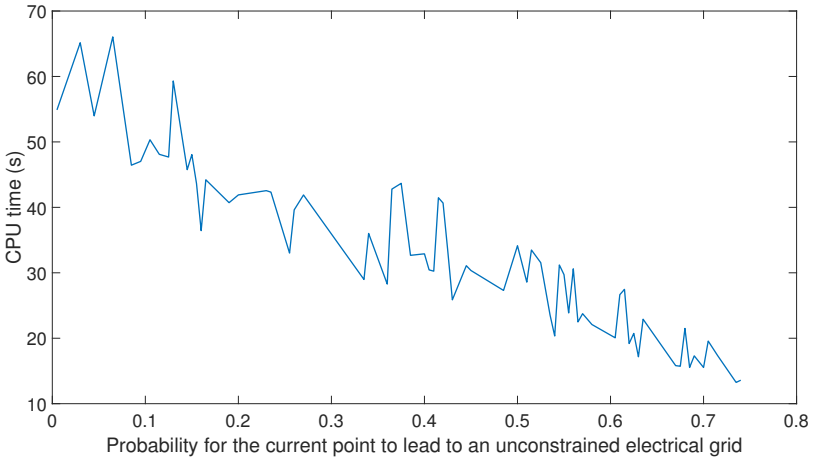


Figure 8.8: Computing time for 100 calls to the oracle on random points.

To create section 8.5.2, we have computed 100 points using a combination of random generation and the same incrementation idea found in algorithm 3. Our goal is to obtain points with different probabilities of leading to an unconstrained electrical grid. For these 100 points, the time spent in the second step[2] is reported as a function of their respective probability of leading to an unconstrained state. Section 8.5.2 clearly shows that the more a point can lead to an unconstrained state in our setting, the less computing time is required in the oracle. This is evident when looking on section 8.5.2: if, for a given decision and a given generated scenario, no electrical constraint is encountered by the PF solver, then for this scenario no optimisation is necessary and the algorithm moves on to the next scenario. On the other hand, when a combination of decision and scenario leads to a constrained state, an optimisation is required and the solving process requires more time.

From this observation, one could therefore hope that with a "good" initialisation, the overall solving time could be improved. This hope, although reasonable at first sight, has to be mitigated: if x^k has low computation time because a small number of scenarios require an additional optimisation step, we cannot expect this property to hold for x^{k+1} in general (in particular if x^{k+1} is a null step).

8.5.3 Comparisons of the different algorithms

We already know that (generalized) `DCA` and `CCP` are easier to implement when compared to `PBMDC` and `PBMDC`^{2}, but prove to be inefficient in our context. The main advantage of `CCP` over `DCA` is that it does not require a feasible point for the DoC constraints for the initialisation. Neither algorithm uses a model for the first component of the DoC functions, nor do they rely on bundle methods. A direct conclusion on these algorithms is that they are easier to implement, but can possibly have longer steps (as they do not make use of approximations) or even be practically unusable (due to too long steps).

At first sight, one might wonder if `PBMDC`^{2} is worth its development and convergence studies, as `PBMDC` applied to a simple penalised version of our program provides interesting results. There are two answers to this question. First of all, thanks to properties of the improvement function we recall that `PBMDC`^{2} has the property of maintaining feasibility once it is reached, which is not the case when applying `PBMDC` to the penalised version. This property is interesting in a context with a computing time constraint for instance. It is possible to parametrise our algorithm to emphasize the importance of reaching feasibility (this is done by applying a large value for ρ) so that the feasible set is reached within a short time period, and the operator can decide to stop the algorithm after a given additional time while being certain that the last point found by the algorithm is feasible.

Secondly, the penalisation parameter of the `PBMDC` algorithm is not trivial to set. On this latter observation, we recall that [185] explores exact penalty for DoC programming but in the particular case where the constraints are defined by concave functions (thus a sub-case of DoC programming). In our case, with an OPF at hands, our DoC functions have both convex and concave components.

Looking more closely, one can observe that `PBMDC` usually has a relatively lower CPU time when compared to `PBMDC`^{2}. We recall that these differences in the computing time

[2] Recall that the time spent in the second step is the time spent in the oracle.

are to be considered taking into consideration the end points are different from one algorithm to the other: we thus compare times to obtain a numerical solutions and not times to converge to a single point. Time of computing can be lower in PBMDC due to the more easily satisfied descent test in the algorithm in particular when a feasible point has been reached. To highlight this fact, let us consider the following problem:

$$\begin{aligned} \min_{x\in X} \quad & f(x) \\ & c(x) \leq 0, \end{aligned} \tag{8.3}$$

where f and c are given functions. When using PBMDC we reformulate problem (8.3) into the parametrized problem (8.4):

$$\min_{x\in X} \quad f(x) + \gamma \max(c(x), 0)^2, \tag{8.4}$$

with $\gamma > 0$. We consider a fixed penalization, with a large γ. When using PBMDC^2, we introduce the improvement function $H_\tau(x) = \max\{f(x) - \tau_f, c(x) - \tau_c\}$ with τ_f, τ_c as defined in Chapter 4, to reformulate problem (8.3) into the parametrized problem (8.5):

$$\min_{x\in X} \quad H_\tau(x). \tag{8.5}$$

Let us consider a setting where the last serious point is denoted $x^{k(\ell)}$, the current iterate is k and verifies $c(x^{k(\ell)}) \leq 0$, i.e., the last serious point is feasible.

Now let us consider the set of points that are admissible to become the next serious iterate $x^{k(\ell+1)}$. When optimising problem (8.4) with PBMDC, this set denoted S_1 is equal to:

$$\{x \in X \quad | \quad f(x) + \gamma \max(c(x), 0)^2 \leq f(x^{k(\ell)}) - \kappa \frac{\mu_{\min}}{2} \left\| x - x^{k(\ell)} \right\|^2 \} =: S_1. \tag{8.6}$$

When optimising problem (8.5) with PBMDC^2, this set is denoted S_2 is equal to:

$$\{x \in X \quad | \quad H_{\tau^\ell}(x) \leq H_{\tau^\ell}(x^{k(\ell)}) - \kappa \frac{\mu_{\min}}{2} \left\| x - x^{k(\ell)} \right\|^2 \} =: S_2. \tag{8.7}$$

One can then invoke Lemma 1 to conclude that $S_2 \subseteq S_1$. In other words, the set of points that can be selected as the next serious step for PBMDC includes the one for PBMDC^2.

8.5.4 Impact of some parameter values

The choice for numerical values for parameters at hand turn out to be crucial for convergence and the solution's quality. Several factors contribute to this importance:

(i) this is a nonsmooth nonconvex optimisation, and this class of problem is known for this significant reliance on parameter tuning.

(ii) Although the proof of convergence in bundle-like methods do no rely on the existence of parameter values that are difficult to obtain *a priori* as in penalisation methods, in practise one has to set "appropriately-chosen" parameters. This choice is based on experience.

(iii) The OPF problem requires a high level of precision: this is also true in our modelisation.

A parametrisation difficulty appears when the objective and constraint functions have values of different order of magnitude. In this case, we applied some fixed scale parameters $\alpha_f, \alpha_c \in \mathbb{R}_+^2$ for objective and constraint functions as follows:

$$\begin{aligned} f(x) &\leftarrow \alpha_f f(x), \\ f(x) &\leftarrow \alpha_c f(x). \end{aligned} \tag{8.8}$$

Parameter ρ as defined in chapter 4 for `PBMDC`^{2} also has a significant impact. In our case, DoC constraints are in practise close to zeros (positive values that are lower than 1), whereas the objective function has values which order of magnitude is greater than 10^3. Setting ρ proportionally to the value $\dfrac{\text{order of magnitude } f}{\text{order of magnitude } c}$ turns out to be efficient in practise.

CONCLUSION

9

New challenges have appeared from the long-term and profound transformations implied by the energy transition and the evolving role of the DSO. Among them is the increasing impact of uncertainties on the operational planning, which proves to be a significant obstacle both from a mathematical point of view, as new optimisation tools are required for this challenge, and from the operational field, as new levers and a good understanding of the uncertainties are necessary. For a better integration of the uncertainties in the mathematical tools the OPF is often identified as a major mathematical tool to address the operational planning step. Literature on the OPF is thus large, as numerous works address the necessary improvement of this tool. For the last 60 years, more efficient modelling choices have led to relaxations with interesting properties in some cases; more electrical elements can now be included in an electrical grid model; new models, as security-constrained OPFs, have been proposed which are well suited for optimisation of energy systems. New algorithms from the optimisation field have been applied to the OPF, with significant numerical improvements: from gradient methods applied to a linear OPF, there are now solvers for the full AC-OPF. With the increasing development of chance-constraints since the early 2010, probabilistic OPFs have gained more and more attention. In parallel, bundle methods have been proposed for non-convex problems, and a DoC formulation for chance-constraints has been presented. The combination of these last three recent developments motivated us to look into particular subset of these challenges: the OPF problem from the DSO, taking uncertainties into account and modelling them using chance-constraints.

As described before, several obstacles had to be overcome in this work:

- The program at hand was a nonsmooth nonconvex one.
- Usual network and PF approximations available to a TSO do not hold in the case of the Medium Voltage (MV) perimeter considered in the work.
- Taking a DSO point of view implies that we do not possess the means of production/consumption, but can interact with them with *a priori* set contracts.
- Forecasts data, as well as a mean of modelling uncertainties linked to these forecasts, are necessary.
- Joint chance-constraint was an initial goal of our work, which usually leads to untractable problems.

In this small subset of the OPF studies, even leaving aside integer numbers, we still immediately obtain a difficult mathematical program and a list of obstacles.

9.1 ACHIEVEMENTS

In a setting that is as much time constrained as it is replete of possibilities, our chosen course of action has placed an emphasis on the mathematical side of the chance-constrained OPF. As a consequence, while the initial motivation and future developments

are focused on the OPF, the technical core of this thesis is DoC programming. This choice is backed by the following facts:

(1) DoC programming is a suitable structure for nonconvex nonsmooth programs, which is the case of the OPF,

(2) Chance-constraints can be approximated by DoC functions.

The first group of achievements of this thesis is thus on DoC programming. We propose a new bundle method for DoC constrained DoC programs, which is a large set of optimisation problems. Convex problems are in this set, and every continuous optimisation problem can be approximated by a DoC one up to a user-given precision. This bundle method extends the current state-of-the-art in nonsmooth nonconvex optimisation. In other words, we only require functional values and subgradients at given points of our functions' domains, which is a fair requirement, and prove that the convergence point is either a *critical* or a *stationary* point depending on the initial assumptions. Note that stationarity is, in DoC programming, the strongest optimality certificate one can hope for in general (see [233]). This algorithm is assessed on usual problems from the literature, where numerical results prove its efficiency.

We then propose a DoC formulation of the Chance-constrained OPF. The latter problem is re-written as a two-step problem, which emphasizes the inherent differences of state and decision variables.

9.2 FOLLOWING STEPS

Numerous questions and new possibilities arise from this work, the answers to which will require additional research. These questions / possibilities are classified into "Mathematical methods", "OPF models" and "data".

9.2.1 MATHEMATICAL METHODS

We discuss in this subsection possible future developments from our work, from a mathematical point of view. The possibilities we present do not all require the same amount of work: some appear to us as direct interesting extensions, while others have been open questions for decades now.

Our solution methodology relies on our proposed bundle algorithm for DoC programs. This algorithm relies on an exact, deterministic oracle, which can be a strong hypothesis (see assumptions 3 and 4). It turns out, in our case, that this hypothesis amounts to assume we always are in a position to find the global optimum of a nonconvex, smooth program. Although this assumption has been encountered in literature, we believe it to be too strong and possibly unnecessary. Following [22] where the authors propose a bundle method with inexact oracles for convex programs, developing a new bundle method for DoC constrained DoC programs seems achievable. The main difficulty seems to be the fact in our case the model function for the original DoC function is not convex, yet convexity of the model is a key assumption in the developments of [22]. Note that inexactitudes of the oracle in our case always follow the following rule: in case the oracle provides an inexact value, this latter value will always be greater than the optimal value.

Another major question, which is still open (see the theoretical considerations of [152, 149]), is the impact of the DoC decomposition on the results. For a deterministic OPF, we have performed trials with different DoC decompositions of the functions at hand. Thanks to [221], and adaptations presented in [285, Chapter 4], we have two different DoC decompositions. The models built during the iterations of `PBMDC`^{2} differ from one decomposition to another, and it is unclear if there exists criteria to discriminate one decomposition from the other. For the stochastic OPF, we only use one decomposition as it is not yet clear if the second one built similarly to the second one of the deterministic OPF is valid.

Thirdly, a difficulty, more or less shared between all nonconvex algorithms, is finding an appropriate parametrization. There usually are no fixed and *a priori* rules for this parametrization, and it requires experience and numerous trials to encounter a set of parameters that suits our current program. Developing rules of parametrization (or lowering the importance of parameters' values in the behaviour of the algorithm) should be considered as an important future work. It should also be clear that without fine parametrization rules, no industrial code tackling nonconvex programs is and will be independent of a designated experienced operator capable of conducting try-outs of parameters.

On our numerical treatment of chance-constraints, it appears that there exists numerous relaxations of this type of constraints. We believe there are multiple, yet laborious, work leads on leveraging these relaxations. For example, one could use Bonferroni approximations of the chance-constraints, as proposed in [76, Section 2.2]: these types of inequalities would enable us to leverage the particular structure of our uncertainties which are in fact separable. We used separability to speed up computation time, while Bonferroni inequalities are an interesting way to approximate joint chance-constraint by individual ones, and thus lower the computational burden. Moreover, considering chance-constrained programs from a variational analysis point of view, a direct future work could be to use `PBMDC`^{2} and apply the subgradient formula of Chapter 6 to our chance-constrained OPF. More precisely, it would be interesting to implement an algorithm as algorithm 3, where the functional and subgradients values of our probabilistic functions are computed as described in Part III. This approach does not fall under DoC programming methods, but rather is "purely" variational approach which could be directly compared to our DoC one.

Finally, we have omitted integer numbers from our models as we are unsure on how our solving methodology could cope with integrality constraints. As a general rule, when adding integer numbers one could either solve a branching tree where each node would be a continuous DoC constrained DoC program or include integer numbers in the DoC constrained DoC program. The latter choice can rapidly appear unfit to one accustomed to DoC programming when studying the usual condition of criticality or stationarity (see for instance [228]).The former choice on the other hand appears too time-consuming to be useful in real-life, letting aside the possible necessity of fine-tuning the parameters at each node of the branching tree.

9.2.2 OPF MODELS AND POSSIBLE IMPROVEMENTS

As discussed in the previous subsection, a major limitation to our developments on the OPF is the lack of integrality constraints. Considering such constraints should be

at the core of the future work on optimization for the OPF under uncertainties. As a matter of fact, integer variables naturally arise in already operational levers (choice of flexibility offers for instance). Although deterministic OPF with integer variables are still significant challenges in large scale networks, in smaller networks ($\sim$100 nodes) OPF with uncertainties and integer variables seem to be the current "go-to".

Nonetheless, letting aside integer variables, we believe our model is of interest for uncertain continuous OPF. Due to the DoC structure, we are able to decompose the problem in a two-step program. The first step relates to the DSO decisions, while the second step computes the probability of observing a network whose state variables are in-between their bounds. As such, this should be interpreted as a framework, where every user can modify or improve a part. In particular, following the successful works on relaxations for the OPF in literature, we believe that the second step could be improved. For instance, we are inspired by the convex relaxations of the OPF, which are exact under some hypothesis: proposing an exact, convex relaxation for our second step optimisation problems would be a significant improvement for our framework.

In a different approach, recent optimisation works could bring a significant improvement to our developments. First of all, from the point of view of methodologies, machine learning techniques appear to provide fast and robust methods for the OPF. For instance, in [177] the authors report a 12x increase in speed on medium sized networks (200 nodes) while the number of cases where their algorithm fails in converging to a solution with an interesting physical meaning is 40% lower than when using a basic mixed-integer solver for an OPF (the one used in MATPOWER [310]). This is interesting as our second step problems are somewhat related to the OPF, and we do require high quality solutions (for our exact oracle hypothesis) and low computing time (for practical usefulness). Secondly, from a material point of view, we care to point out to [168, 254] which describe GPU-accelerated LF and OPF. The benefit of this material incremental improvement within our framework would be a faster function values and subgradients evaluations. It would therefore improve the second-step from a computing time point of view, which is the most challenging step of our proposed algorithms.

9.2.3 Data considerations

While Enedis proposes good quality data including consumption and production profiles for different types of GUs, in stochastic programming it is assumed one has a good knowledge of the variables with uncertainties. The variance of different profiles are absent in Enedis Open Data, which are necessary for our OPF. Moreover, covariances also are missing, which also are necessary when considering joint chance-constraints. Without these two moments, we cannot quantify the importance of uncertainties in our problem, let alone the impact of these uncertainties in a real-world problem. A better assessment and quantification of the different sources of uncertainties should be necessary for real-world applications.

For instance, the author of [60] studies a case where the most significant source of uncertainties is not on the forecasts of GUs, but on the taps of the HV/MV transformer. Other sources of uncertainties include the possible unforecasted loss of infrastructure or means of production (this particular case is studied in the *Security constrained* OPF), or the unknown characteristic values of electrical elements (resistance and reactance values could be modelled as uncertain parameters).

DETAILS ON THE DOC DECOMPOSITIONS OF THE DETERMINISTIC OPF

A

In this Appendix, we are interested in the DoC decomposition of the deterministic OPF. The interest of this development is twofold:

(1) It is a good introduction into practical DoC decompositions of functions;

(2) We consequently prove that the deterministic continuous OPF is a DoC problem by explicitly showing this peculiar structure.

We are able to propose two different DoC decompositions which is also interesting as there is no general knowledge of which DoC decomposition is better: this is still an open question in the optimisation field. As a consequence, having two different decompositions readily available could contribute to research on discrimination of DoC decompositions.

This Appendix is organised as follows: we start with necessary elements for both DoC decompositions, before presenting the first decomposition and the second.

a.1 FIRST NECESSARY DOC DECOMPOSITIONS FOR THE DOCOPF

First of all, it is clear the a continuous deterministic OPF as equation (7.1) is a DoC problem: the theoretical justification is that all involved functions are $\mathcal{C}^2$ everywhere. The objective of this section is to provide an explicit DoC formulation.

We immediately tackle the linear components of nonconvex functions as in the following example:

Example 7. Let $c_1, c_2 : \mathbb{R}^n \to \mathbb{R}$ be two convex functions, and $f : \mathbb{R}^n \to \mathbb{R}$ be a linear function. For any $\lambda \in \mathbb{R}$, we define $c_1^\lambda = c_1 + \lambda f$ and $c_2^\lambda = c_2 + (\lambda - 1)f$ which are both convex functions. Then $c_1^\lambda - c_2^\lambda$ is a valid DoC decomposition of $x \mapsto c_1(x) - c_2(x)$. ▶

Having dealt with the linear parts of functions, we now turn to the more interesting nonconvex parts of the functions at hands. Developing equations (7.3) and (7.4) yields:

$$p_i^g - p_i^l = \sum_{\mathcal{N} \ni k \sim i} y_{i,k}^R \, |V|_i \, |V|_k \cos(\delta_i - \delta_k) + y_{i,k}^I \, |V|_i \, |V|_k \sin(\delta_i - \delta_k) \tag{A.1}$$

$$q_i^g - q_i^l = \sum_{\mathcal{N} \ni k \sim i} y_{i,k}^R \, |V|_i \, |V|_k \sin(\delta_i - \delta_k) - y_{i,k}^I \, |V|_i \, |V|_k \cos(\delta_i - \delta_k). \tag{A.2}$$

It becomes clear that the two key elements to decompose into a DoC formulation are the following:

$$\begin{array}{rccl} \Psi : & \mathbb{R}^4 & \to & \mathbb{R} \\ & (x, y, z, h) & \mapsto & xy \cos(z - h), \end{array} \tag{A.3}$$

$$\begin{array}{rccl} \Phi: & \mathbb{R}^4 & \to & \mathbb{R} \\ & (x,y,z,h) & \mapsto & xy\sin(z-h), \end{array} \tag{A.4}$$

the last step to find the full DoC decomposition being a linear combination of the components. As a consequence, the following sections will aim at providing DoC formulations for these two functions, as functions of four variables. We emphasize that the clear link between formulations equations (A.3) and (A.4) and our OPF problem is:

$$x \leftrightarrow |V|_i\,,\ y \leftrightarrow |V|_k\,,\ z \leftrightarrow \delta_i,\ h \leftrightarrow \delta_k.$$

Again, Ψ and Φ are immediately seen to be DoC functions as they are lower-$\mathcal{C}^2$. We recall that there are an infinite number of valid DoC decompositions for a given DoC function: we here propose two of these. Moreover, even though the DoC property is a global one, the validity of a DoC decomposition can be local as we can see in example 8.

Example 8. Let us consider $f: \mathbb{R} \ni x \mapsto x^3$; f is a nonconvex DoC function, and a possible decomposition over $\mathbb{R}$ is

$$f = \underbrace{\frac{13\,x^4}{16} + \frac{x^3}{2} + \frac{5\,x^2}{8}}_{\text{convex}} - \underbrace{(\frac{13\,x^4}{16} - \frac{x^3}{2} + \frac{5\,x^2}{8})}_{\text{convex}}.$$

Interestingly, another DoC decomposition this time over $\mathbb{R}_+$ is f itself as it is convex on this set, or yet again $f = \frac{1}{2}\left((x^2+x)^2 - (x^4+x^2)\right)$ is a valid DoC decomposition for $x > \frac{-3+\sqrt{3}}{6}$. This simple example outlines that the decomposition is a local concept, as opposed to the global property of being DoC. ▶

We now present two DoC decompositions, both of them using different techniques. The motivation lies in the fact there no clear way to discriminate a DoC decomposition from another *a priori*. In accordance with Example 8, we will present DoC formulations that are valid on a set that includes all values that are possible to obtain in an OPF problem. To that matter, we define $\mathcal{S} := [-\pi,\pi]^2 \times [0,2]^2$ which includes all attainable values of the variable vector $[\delta_1, \delta_2, |V|_1\,, |V|_2]$.

A.1.1 DoC decomposition A

Let f be a lower-$\mathcal{C}^2$ function. Among its properties is the existence of a parameter $\rho > 0$ such that:

$$f(x) = \underbrace{f(x) + \frac{\rho}{2}\|x\|^2}_{\text{convex}} - \underbrace{\frac{\rho}{2}\|x\|^2}_{\text{convex}}. \tag{A.5}$$

We will refer to ρ as the *convexification parameter*. In the particular case where f is a $\mathcal{C}^2$ function, a rule to select this parameter ρ is readily available: it is sufficient to set $\rho = \max(0, -\lambda_{\min})$, where $\lambda_{\min}$ is a uniform lower bound on the minimal eigen values of the hessian matrix of f. Similarly, in case f has a Lipschitz continuous gradient with modulus $L > 0$ then setting $\rho = L$ in equation (A.5) provides a valid DoC decomposition ([229, Proposition 1]).

Recalling that in our case we are interested in a DoC decomposition within a set that verifies $x, y \in [0;2]$ and $z, h \in [-\pi;\pi]$, the following decomposition is valid (see Appendix A for more details):

$$\Psi = \overbrace{xy\cos(z-h) + 4(|x|^2+|y|^2+|z|^2+|h|^2)}^{\text{convex}} - \overbrace{4(|x|^2+|y|^2+|z|^2+|h|^2)}^{\text{convex}},$$

$$\Phi = \overbrace{xy\sin(z-h) + 4(|x|^2+|y|^2+|z|^2+|h|^2)}^{\text{convex}} - \overbrace{4(|x|^2+|y|^2+|z|^2+|h|^2)}^{\text{convex}}.$$

A.1.2 DoC decomposition B

Leveraging the proposed algorithm of [221] and using DoC calculus rules one can encounter in the introduction of [229], we propose another DoC decomposition. The explicit formulation is can be found in Appendix A.

The advantage of this decomposition is that it does not rely on any parameter that is difficult to validate: an operator solely has to compute the Lipschitz modulus of the gradient of $(z,h) \mapsto \cos(z-h)$, which is easier done than for a function of four variables. Moreover, as depicted in Appendix B, the two different DoC decompositions are of different numerical interests.

A.1.3 DoC decomposition of the OPF

From subsections appendices A.1.1 and A.1.2 we have obtained decompositions $\Psi = \Psi_1 - \Psi_2$ and $\Phi = \Phi_1 - \Phi_2$. As visible in equations equations (A.1) and (A.2), the last step to obtain a DoC formulation for these two constraints is to compute a linear combinations of $\Psi_1, \Psi_2, \Phi_1, \Phi_2$.

Remark 26. Evidently for $y < 0$, one has: $y(\Psi_1 - \Psi_2) = \Psi_1' - \Psi_2'$ with $\Psi_1' = |y|\,\Psi_2$, $\Psi_2' = |y|\,\Psi_1$. ▷

Once this step is completed, we are now in possession of the following formulation:

$$\begin{aligned} c_i^{\mathrm{R}}(\delta,p,q,|V|) &= p_i^l - p_i^g + \sum_{\mathcal{N}\ni k\sim i} y_{i,k}^R\,|V|_i\,|V|_k\cos(\delta_i-\delta_k) + y_{i,k}^I\,|V|_i\,|V|_k\sin(\delta_i-\delta_k) \\ &= c_{i,1}^{\mathrm{R}}(\delta,p,q,|V|) - c_{i,2}^{\mathrm{R}}(\delta,p,q,|V|), \end{aligned}$$

$$\begin{aligned} c_i^{\mathrm{I}}(\delta,p,q,|V|) &= q_i^l - q_i^g + \sum_{\mathcal{N}\ni k\sim i} y_{i,k}^R\,|V|_i\,|V|_k\sin(\delta_i-\delta_k) - y_{i,k}^I\,|V|_i\,|V|_k\cos(\delta_i-\delta_k) \\ &= c_{i,1}^{\mathrm{I}}(\delta,p,q,|V|) - c_{i,2}^{\mathrm{I}}(\delta,p,q,|V|), \end{aligned}$$

where, by construction, functions $c_{i,1}^{\mathrm{R}}, c_{i,2}^{\mathrm{R}}, c_{i,1}^{\mathrm{I}}, c_{i,2}^{\mathrm{I}}$ are convex. Equations equations (A.1) and (A.2) now become:

$$\begin{aligned} c_{i,1}^{\mathrm{R}}(\delta,p,q,|V|) - c_{i,2}^{\mathrm{R}}(\delta,p,q,|V|) &= 0, \qquad && \forall i\in\mathcal{N} \\ c_{i,1}^{\mathrm{I}}(\delta,p,q,|V|) - c_{i,2}^{\mathrm{I}}(\delta,p,q,|V|) &= 0, && \forall i\in\mathcal{N} \end{aligned}$$

which we modify into inequalities equations (A.6) and (A.7) for a specified $\epsilon > 0$ to fit our DoC programming setting:

$$\left|c_{i,1}^{\mathrm{R}}(\delta,p,q,|V|) - c_{i,2}^{\mathrm{R}}(\delta,p,q,|V|)\right| \le \epsilon, \qquad \forall i\in\mathcal{N} \qquad \text{(A.6)}$$

$$\left|c_{i,2}^{\mathrm{I}}(\delta,p,q,|V|) - c_{i,2}^{\mathrm{I}}(\delta,p,q,|V|)\right| \le \epsilon, \qquad \forall i\in\mathcal{N} \qquad \text{(A.7)}$$

INSERT 11: Note on inequalities equations (A.6) and (A.7) and ϵ.
The aforementioned inequalities are, strictly speaking, relaxations of equations (A.1) and (A.2) with a user-given parameter $\epsilon > 0$. In practice, we set $\epsilon = 10^{-5}$

For clarity, before proposing our first DoC formulation of the deterministic OPF, we recall a classic DoC programming identity:

Lemma 13. [285, Proposition 4.1] Let $\mathcal{I}$ be a finite index set, and let $(a_{1,i} - a_{2,i})_{i\in\mathcal{I}}$ be a finite set of DoC functions defined on an abstract set A of $\mathbb{R}^n$. Then $\max_{i\in\mathcal{I}}[a_{1,i} - a_{2,i}]$ is DoC, with a readily available DoC formulation as follows:

$$\max_{i\in\mathcal{I}}[a_{1,i}(x) - a_{2,i}(x)] = \max_{i\in\mathcal{I}}\left[a_{1,i}(x) + \sum_{\mathcal{I}\ni j\neq i} a_{2,j}(x)\right] - \sum_{i\in\mathcal{I}} a_{2,i}(x) \tag{A.8}$$

◀

As a consequence of Lemma 13, we can re-write constraints equations (A.6) and (A.7) that are enforced for every node as:

$$\overbrace{\max_{i\in\mathcal{I}}\left[c^{\mathrm{R}}_{i,1}(\delta,p,q,|V|) + \sum_{\mathcal{I}\ni j\neq i} c^{\mathrm{R}}_{j,2}(\delta,p,q,|V|)\right]}^{c^{\mathrm{R}}_1(\delta,p,q,|V|)} - \overbrace{\sum_{i\in\mathcal{I}} c^{\mathrm{R}}_{i,2}(\delta,p,q,|V|)}^{c^{\mathrm{R}}_2(\delta,p,q,|V|)} - \epsilon \leq 0$$

$$\underbrace{\max_{i\in\mathcal{I}}\left[c^{\mathrm{I}}_{i,1}(\delta,p,q,|V|) + \sum_{\mathcal{I}\ni j\neq i} c^{\mathrm{I}}_{j,2}(\delta,p,q,|V|)\right]}_{c^{\mathrm{I}}_1(\delta,p,q,|V|)} - \underbrace{\sum_{i\in\mathcal{I}} c^{\mathrm{I}}_{i,2}(\delta,p,q,|V|)}_{c^{\mathrm{I}}_2(\delta,p,q,|V|)} - \epsilon \leq 0$$

We can finally explicitly propose the DoC formulation of our deterministic OPF:

$$\begin{aligned}
\min_{\delta,p,q,|V|} \quad & f(\delta,p,q,|V|) \\
\text{s.t.} \quad & A\,(\delta,p,q,|V|) && \leq b && \text{Linear constraints,} \\
& A_{eq}\,(\delta,p,q,|V|) && = 0 && \text{Linear equality constraints} \\
& c^{\mathrm{R}}_1(\delta,p,q,|V|) - c^{\mathrm{R}}_2(\delta,p,q,|V|) && \leq \epsilon && \text{Real power} && \text{(A.9a)} \\
& c^{\mathrm{I}}_1(\delta,p,q,|V|) - c^{\mathrm{I}}_2(\delta,p,q,|V|) && \leq \epsilon. && \text{Reactive power} && \text{(A.9b)}
\end{aligned}$$

We here provide more details to rigorously prove we indeed have two correct DoC decompositions. As a reminder, we are interested in DoC decompositions of the following functions:

$$\Psi(x,y,z,h) = xy\cos(z-h), \qquad \Phi(x,y,z,h) = xy\sin(z-h).$$

In our OPF context, it is sufficient to look into DoC decompositions that are valid on $[0,2]^2 \times [-\pi,\pi]^2$.

a.2 ON DECOMPOSITION A

Our first proposition relies on finding an upper-bound on the Lipschitz moduli of $\nabla\Psi$ and $\nabla\Phi$. When the functions at hand are twice continuously differentiable, these moduli are in fact uniform upper-bounds on the largest eigenvalue of the respective hessian matrices:

$$\nabla^2\Psi(x,y,z,h) \preceq L_\Psi \mathrm{id}, \qquad \nabla^2\Phi(x,y,z,h) \preceq L_\Phi \mathrm{id}.$$

From this point on, we computed eigenvalues of the 4×4 hessian matrices, and found an upper bound to the largest of these. This process has been conducted using MATLAB Symbolic Toolbox for verification. In this context, an eigenvalue is a function from $[0,2]^2 \times [-\pi,\pi]^2$ to $\mathbb{R}$. The largest value attained by one of the eigenvalue functions is 8, for both Ψ and Φ. We also have a very easy routine to numerically search for errors in a DoC decomposition, which can be found in Annex B.

a.3 ON DECOMPOSITION B

The drawback of the first decomposition is that it relies on a chosen parameter, which is difficult to verify either "by hand" or by using a CAS system. To overcome this drawback, one can observe that Ψ and Φ are two product functions, which elements we can modify to ensure they are non-negative.

An interested reader can find in [221] DoC decomposition methods for the particular case of polynomials. We adapted the proposed procedures to our case where we have cos and sin functions. The key elements can in fact be found in [285, Chapter 4]:

Proposition 10. [285, Proposition 4.12] For any two convex functions $f_1, f_2 \colon \mathbb{R}^n_+ \mapsto \mathbb{R}_+$ a DoC representation of their product is the following:

$$f_1(x)f_2(x) = \frac{1}{2}\left[f_1(x) + f_2(x)\right]^2 - \frac{1}{2}\left[f_1^2(x) + f_2^2(x)\right]$$

◀

Similar formulations are presented in [221], which also rely on the hypothesis of non-negativity:

Lemma 14. [221, Section 4] For g_1, g_2, h_1, h_2 convex polynomials that are sum-of-squares (*i.e.* a sum of monomials that have even exponents), the following formulation is a difference of sum-of-squares polynomials:

$$(g_1 - h_1)(g_2 - h_2) = \frac{1}{2}\left((g_1 + g_2)^2 + (h_1 + h_2)^2 - (g_1 + h_2)^2 - (g_2 + h_1)^2\right)$$

◀

We can observe that a cornerstone element of Lemma 14 is the non-negativity property of sums-of-squares. Let us recall two well-known convexity preserving operations: for convex functions f and g, $f + g$ is convex, and if f also is non-negative then f^k is convex, for $k \geq 1$. As a consequence, we can observe that Lemma 14 also holds for non-negative convex functions which are not necessarily polynomials.

Turning our attention back to Ψ and Φ, we start by making explicit a structure that will allow the application of Proposition 10 and Lemma 14:

$$\Psi(x,y,z,h) = xy(\cos(z-h)+|z|^2+|h|^2- \quad \Phi(x,y,z,h) = xy(\sin(z-h)+|z|^2+|h|^2+1-$$
$$|z|^2-|h|^2), \qquad |z|^2-|h|^2-1).$$

Now, applying twice the property of Proposition 10:

$$\begin{aligned}\Psi(x,y,z,h) =& \frac{1}{2}\left[(x+y)^2-\left(x^2+y^2\right)\right]\left[\cos(z-h)+z^2+h^2-z^2-h^2\right]\\ =& \frac{1}{4}\left[\left((x+y)^2+\cos(z-h)+z^2+h^2\right)^2+\left(x^2+y^2+z^2+h^2\right)^2-\right.\\ & \left.\left(x^2+y^2+\cos(z-h)+z^2+h^2\right)^2-\left((x+y)^2+z^2+h^2\right)^2\right]\\ =& \Psi_1(x,y,z,h)-\Psi_2(x,y,z,h),\end{aligned}$$

with:

$$\Psi_1(x,y,z,h) = \frac{\left(\cos(h-z)+(x+y)^2+h^2+z^2\right)^2}{4}+\left(h^2+x^2+y^2+z^2\right)^2,$$
$$\Psi_2(x,y,z,h) = \frac{\left(\cos(h-z)+h^2+x^2+y^2+z^2\right)^2}{4}+\left((x+y)^2+h^2+z^2\right)^2.$$

Similarly for Φ:

$$\begin{aligned}\Phi(x,y,z,h) =& \frac{1}{2}\left[(x+y)^2-\left(x^2+y^2\right)\right]\left[\sin(z-h)+z^2+h^2+1-z^2-h^2-1\right]\\ =& \frac{1}{4}\left[\left((x+y)^2+\sin(z-h)+z^2+h^2+1\right)^2+\left(x^2+y^2+z^2+h^2+1\right)^2-\right.\\ & \left.\left(x^2+y^2+\sin(z-h)+z^2+h^2+1\right)^2-\left((x+y)^2+z^2+h^2+1\right)^2\right]\\ =& \Phi_1(x,y,z,h)-\Phi_2(x,y,z,h),\end{aligned}$$

with:

$$\Phi_1(x,y,z,h) = \frac{\left(x^2+y^2+h^2+z^2+1\right)^2}{4}+\frac{\left((x+y)^2+\sin(z-h)+h^2+z^2+1\right)^2}{4}$$
$$\Phi_2(x,y,z,h) = \frac{\left((x+y)^2+h^2+z^2+1\right)^2}{4}+\frac{\left(x^2+y^2+\sin(z-h)+h^2+z^2+1\right)^2}{4}$$

It is readily seen that the functions $\Psi_1,\Psi_2,\Phi_1,\Phi_2$ are sums and squares of convex non-negative functions, which ensures that this is a valid DoC decomposition on a set that includes $[0,2]^2\times[-\pi,\pi]^2$.

A SIMPLE ROUTINE FOR NUMERICAL VERIFICATION OF A DoC DECOMPOSITION

B

Let f be a DoC function, and let us consider its associated DoC decomposition $f = f_1 - f_2$, with f_1, f_2 convex. Let us also assume all functions are defined on a set $\mathcal{S} \in \mathbb{R}^n$, have real values, and the DoC decomposition is valid on $\mathcal{S}$. There are two immediate properties one can check:

- functional values validity. This is a direct verification that for a given $x \in \mathcal{S}$, $f(x) = f_1(x) - f_2(x)$.
- Convexity verification of f_1 and f_2.

It turns out that convexity verification is not an easy task when the explicit formulation is a difficult one, making a direct convexity check using usual convexity-preserving rules inapplicable. The following numerical method is inspired by bundle methods: one can build a cutting-plane model for both DoC components, and check that this model is indeed below the function. In a convex setting, provided that the oracles are correct, if a functional value of a DoC component is above the model then there is the function at hand is not convex.

Algorithm 4 A cutting-plane procedure for numerical verification of convexity

Let $f = f_1 - f_2$ be a DoC decomposition to test with their associated oracles $x \mapsto (f_i(x), g_i(x))$, $g_i(x) \in \partial f_i(x)$, $i = 1, 2$. Select two sets S_1, S_2 of randomly generated points in $\mathcal{S}$. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two empty cutting plane models.
procedure CUTTING-PLANE MODEL GENERATION
    Set up $\mathcal{M}_1$ and $\mathcal{M}_2$ using points of S_1.
procedure NUMERICAL CONVEXITY VERIFICATION
    for $s_k \in \mathcal{S}_2$ **do**
        Call oracles at s_k.
        if $\mathcal{M}_i(s_k) > f_i(s_k)$, $i = 1, 2$ **then**
            STOP with error.
        else
            $\text{list}_i(k) = f_i(s_k) - \mathcal{M}_i(s_k)$, $i = 1, 2$.

Using this algorithm, and aiming at providing more insight into the DoC decompositions of functions at hands in our application (see equations (A.3) and (A.4)), we provide plots of distances between functions' and their associated models' at each point of $\mathcal{S}_2$. For these plots, the convexification parameter in decomposition A is set to 8, the models are set up on 10^5 randomly generated points of $\mathcal{S}$, and distances are computed on 10^5 randomly generated points of $\mathcal{S}$.

Remark 27. One should bear in mind that in this experiment, points are randomly generated: in a bundle methods, points at which the oracles are called to enrich the

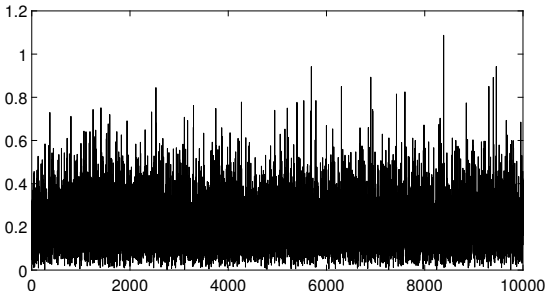


FIGURE B.1: Model distances for Ψ_1, DoC decomposition A

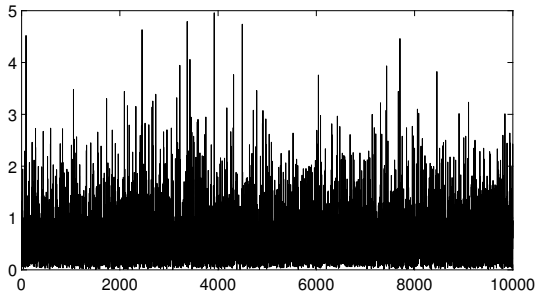


FIGURE B.2: Model distances for Ψ_1, DoC decomposition B

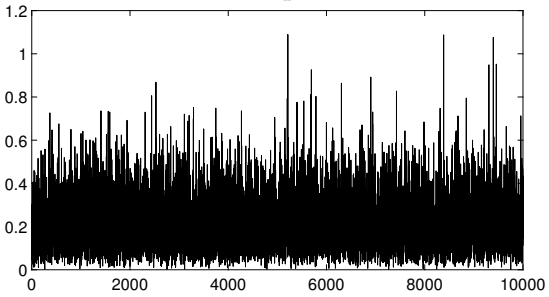


FIGURE B.3: Model distances for Ψ_2, DoC decomposition A

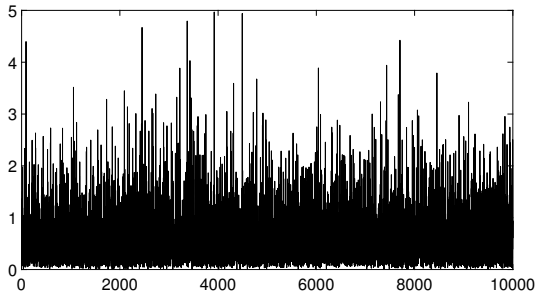


FIGURE B.4: Model distances for Ψ_2, DoC decomposition B

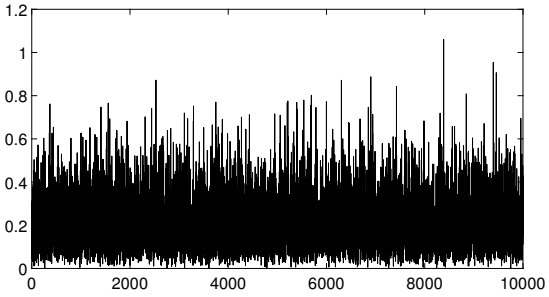


FIGURE B.5: Model distances for Φ_1, DoC decomposition A

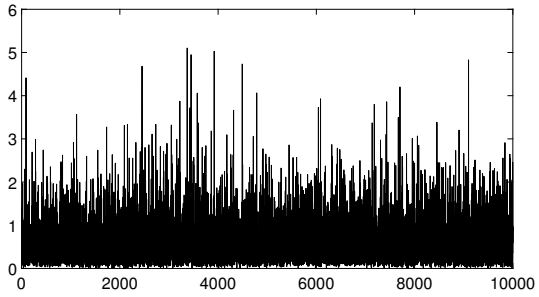


FIGURE B.6: Model distances for Φ_1, DoC decomposition B

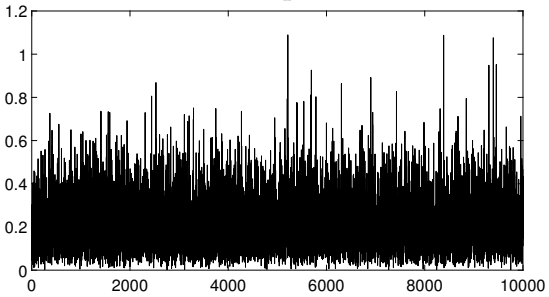


FIGURE B.7: Model distances for Φ_2, DoC decomposition A

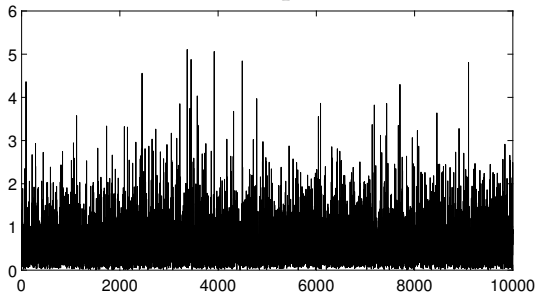


FIGURE B.8: Model distances for Φ_2, DoC decomposition B

cutting plane model are not randomly generated. The rationale is that these computed distances only matter at "interesting" points, namely critical/stationary points in our case. ▷

It is readily visible that decomposition A seems to be, within the scope of this experiment, a better approximation of the original function. As can be expected, it is visible in figure B.9 that different values of the convexification parameter in this decomposition have significant impacts on the output of this experiment. Another drawback, this time from a numerical precision point of view, is that a higher convexification parameter implies greater magnitudes of the DoC components and their associated (sub-)gradients' norms. Although no rigorous study has been conducted in this work, we experienced cases where components' values were of magnitude 10^9 while their difference (*i.e.* the original functional value) was of magnitude 10^{-5}: having two drastically different order of magnitude in the same optimisation program can lead to numerical difficulties. We care to emphasize this is entirely due to our methods of finding the DoC decompositions, which were the only ones at our disposal at the time of this work.

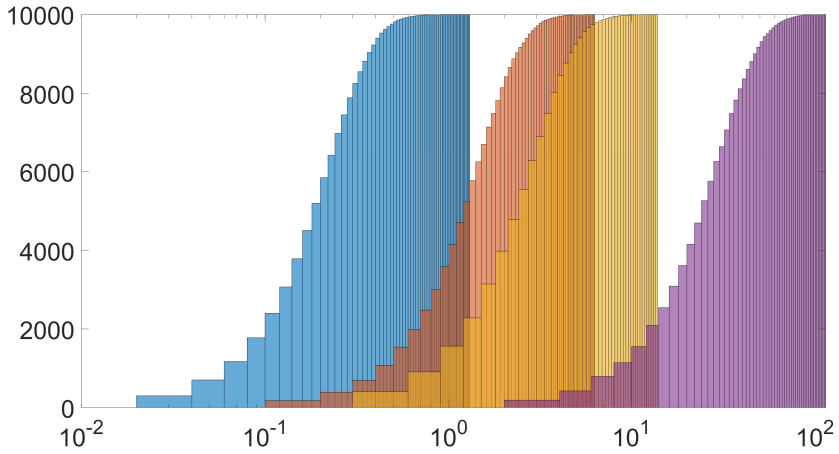


FIGURE B.9: Cumulative distances for Ψ_1 for 4 different convexification parameters ρ in decomposition A. Blue bars are obtained with $\rho = 8$, orange with $\rho = 50$, yellow with $\rho = 100$ and purple with $\rho = 10^3$. All four experiments share the same $\mathcal{S}_2$, comprised of 10^4 randomly chosen points.

AN APPROXIMATE PROJECTED GRADIENT ALGORITHM

C

We present in this Appendix a generic algorithm, similar to the one used in [144, Section 6] for solving the following problem:

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & \varphi(x) \geq p, \end{aligned}$$

where f is assumed to be affine, and φ is as defined in (3.8). We also assume that a point $x_0 \in X' := \{x \in X \ : \mathbb{P}\left[g(x, \xi) \leq 0\right] \geq p\}$ is known (and therefore this set is not empty). Let us define $\tau_{X'}^{x_0} : X \mapsto [0, 1]$ as follows:

$$\tau_{X'}^{x_0}(x) = \left\{ \begin{array}{ll} \min_{t \in [0,1]} & t \\ \text{s.t.} & \mathbb{P}\left[g((1-t)x + tx_0, \xi) \leq 0\right] \geq p \end{array} \right.$$

and let us define $P_{X'}^{x_0} \colon X \to X'$ as $P_{X'}^{x_0}(x) = (1 - \tau_{X'}^{x_0}(x))x + \tau_{X'}^{x_0}(x)x_0$. Since X' is a closed set, $\tau_{X'}^{x_0}$ and thus $P_{X'}^{x_0}$ are well defined. Since, it is not a proper projection, this algorithm is sometimes referred to as an approximate projected gradient algorithm. Convergence of such an algorithm is studied for example in [141]. Obtaining a numerical value for $\tau_{X'}^{x_0}$ can be done using a bisection method, having at each iteration k to compute $\mathbb{P}\left[g((1-t_k)x + t_k x_0, \xi) \leq 0\right]$ and stopping when it reaches p up to a given tolerance.

The formulæ (6.23) can be exploited to compute $\nabla\varphi(x_k)$ at a given trial point (upon verifying the condition of [17, Theorem 4.2]).

Algorithm 5 An approximate projected gradient algorithm

Step 0: Initialization. Let $k = 0$, $x_0 \in X'$, and Tol > 0 a given tolerance.

Step 1: Oracle call. Compute $f(x_k)$, $s_k^f = \nabla f(x_k)$, and $s_k^g = \nabla\varphi(x_k)$.
If $f(x_k) \leq f^{\text{best}}$, then $x^{\text{best}} \leftarrow x_k$ and $f^{\text{best}} \leftarrow f(x_k)$.

Step 2: Descent step. If $\left\|s_k^g\right\| = 0$ then define $d_k = -s_k^f$, else define $d_k = -s_k^f + \dfrac{\left\langle s_k^f, s_k^g \right\rangle}{\left\|s_k^g\right\|^2} s_k^g$.

Step 3: Line Search. Perform a line search to find θ such that $x'_{k+1} := x_k + \theta d_k \in X$ and achieves descent.

Step 4: Projection step. $x_{k+1} \leftarrow P_{X'}^{x_0}(x'_{k+1})$.

Step 5: Stopping test. If $\|x_{k+1} - x_k\| \leq$ Tol return x^{best} and f^{best} and terminate the algorithm. Else proceed to step 6.

Step 6: Loop. Set $k := k + 1$ and go back to Step 1.

NUMERICAL INPUTS FROM ENEDIS OPEN DATA

D

In this appendix, we summarize the processed data we used for all our test-cases. As discussed in Chapter 7, data is both (i) necessary at each stage (parametrisation and resolution) of the optimisation, (ii) difficult to obtain due to either technical considerations (on the material side, an electrical grid is expensive to exploit and maintain, let alone install sensors so there is no such operational omniscience) or privacy/intellectual property limitations.

As part of a larger initiative to develop a full multi-temporal approach for a DoC formulation of the OPF, we propose a data structure that is in part inspired by MATPOWER work [310]. Our motivation to develop a model of our own stems from the following ambitions:

- have a clear data separation between the ones relevant to the grid, the GU and the DSO;
- be able to generate copyright-free forecasts from open source data (specifically from [98]);
- make available an easier stream of process from data creation to the optimisation problem and vice-versa.

INSERT 12: Note on MATPOWER format.
On the one hand, MATPOWER format has an integrated load/generator model in its data structure and makes easily available grid parameters (topology, resistance and reactance values). This data structure is perfectly suited and efficient for PF and OPF, with the drawback of being "closer" to the mathematical model and "further" from GUs' models. On the other hand, Enedis develops since 2015 an open dataset that provides data detailed up to a legally privacy-preserving level. In order to benefit from this (huge) dataset, we aimed at creating a bridge between the detailed GUs models and a mathematical model. Keeping in mind that grid parameters as its topology, resistance and reactance values are easier to obtain when compared to injection/consumption forecasts, this motivates the definition of a new global model, while keeping it relatively close to MATPOWER's. On another matter, non-decomposable multi-temporal OPF is not a native functionality of MATPOWER's, which we have implemented in the spirit of our ST-OP definition from the DSO point of view.

We propose a data format for optimisation on electrical networks which is broadly described in figure D.1.

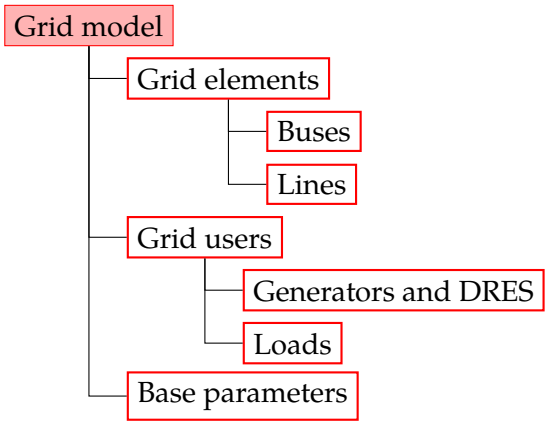


FIGURE D.1: High level description of our grid model.

We highlight and briefly discuss some key elements of data from a DSO point of view, which is comprised of:

- a network topology;
- a list of GUs;
- an affectation of GUs to nodes;
- a generation or consumption forecast for each GU.

Network topologies and Grid Elements Network topologies can be easily encountered in literature: see for instance [310, 178, 287, 38] which present electrical networks. Although a majority of them are transmission networks, some are distribution grids. In [60] a real and large distribution network is used.

Grid Users Differently to network topologies, grid users are more protected and as a consequence less encountered in literature. Thanks to Enedis Open Data (see [98]), we now have access to aggregated grid user data at different levels in France (mainly national, regional or by departments). Due to the long and unified history of the French electrical grid, data is somewhat robustly classified: each geographical scopes share the same classes for instance. One can find two technical documents [102, 260], which complement each other to describe Enedis' data collection. The first one provides an overview of the Open Data datasets, while the second one gives details on the classification of profiles. The key elements are the following:

- GUs are broadly categorized according to three of their characteristics: electric energy metering, power subscription/installation size and activity.
- Each category is given an aggregate profile. The aggregation is done following privacy-preserving rules, which can lead to inconsistent data. This is especially the case when a category has a time-dependent cardinality that is relatively close to thresholds defined in these privacy-preserving rules.
- For a given geographical area, one has access to the total number of producers/consumers, as well as number of actors in each category defined in the first point.

Taking these inputs into consideration, one can derive probabilities for random draws of grid users.

Affectation of GUs to nodes We randomly classify nodes into three categories:

1. a "load-only" node;
2. a constrained mixed node: loads and producers can be connected to this node, provided that the sum of subscribed powers (loads) is lower than the sum of installed powers (producers);
3. a fully mixed node: loads and producers can be connected to this node, without any limits on their proportions.

Note that the affectation is an iterative process. For each GU, we randomly select a node to which it will be affected, provided that the following rules are met:

- the node category allows the connection of the GU to the selected node;
- the power of this GU, added to the possibly already installed power, does not exceed the nominal power of the MV/LV transformer.

Forecast generation From Enedis Open Data one has access to aggregated consumption and production levels every 30 minutes of a given time horizon. We add marginal modifications to make appear electrical constraints in the forecasts. These modifications make sure that a given GU never has a production/consumption forecast that exceeds its installation/subscription level.

INSERT 13: Note on existing open data for electrical networks.
Since 2015, Enedis publishes and maintains datasets relative to energy (mainly electricity) matters. Now is included a vast number of data types, that ranges from electricity balance within different geographical areas, to electricity quality and various grid development indices. Although some data types are not detailed at a lower scale than the national/regional one (e.g. due to the limited number of electrical vehicles charging points, data on this matter are available only at a regional level, which lessens the interest of their use), as the grid and its use expand it is reasonable to expect a continuous improvement of data availability.
Having access to data without infringing any intellectual property law still is challenge, for the academia as well as for some industrial projects: for instance, data on consumption are technically retrieved by the DSO but are the property of the consumer. The consumer can then authorize their transfer to the supplier, who can operate on them. Aggregate data are made openly available by the DSO as a mean to further encourage the expansion of the energy transition.
The open-data initiative is an on-going process, and we care to point out existing research on this matter. Differently from our assumptions and approach, it appears that active open-source research is aimed at the creation of GEs: this is to be related to research on electrical planning. As a matter of fact, broadly speaking both these fields are aimed at developing tools to have access to plausible electrical topology and GEs' characteristics, either at present time for the former, or in the future for the latter. One example is [287] where research group provides transmission networks and recently added a combined transmission/distribution network. The main idea of this research group is to derive network topologies and GEs' characteristics from existing open-source datasets (which include for instance: density of population, geographical data, highly aggregated energy consumptions).
While this first reference does not yet provide useful data for the distribution level, two newly added references share a similar goal to ours. The first one is [178] (yet to be published) is an interesting gathering of topologies, R, X values for realistic radial distribution networks. A second dataset is [38] is a presentation of an update to PGLib-OPF, a dataset with "reasonable" values although it is not as explicitly aimed at the distribution level when compared to the latter reference.
The objective of our data generation module is to leverage this real data, process them in order to fill in the missing values in our OPF.

A "fully-implemented" electrical grid is a data set comprised of all necessary data in order to compute solutions to our models from Chapter 7. In this section we briefly detail the main steps for our generic data creation, and the rules that are enforced throughout this process.

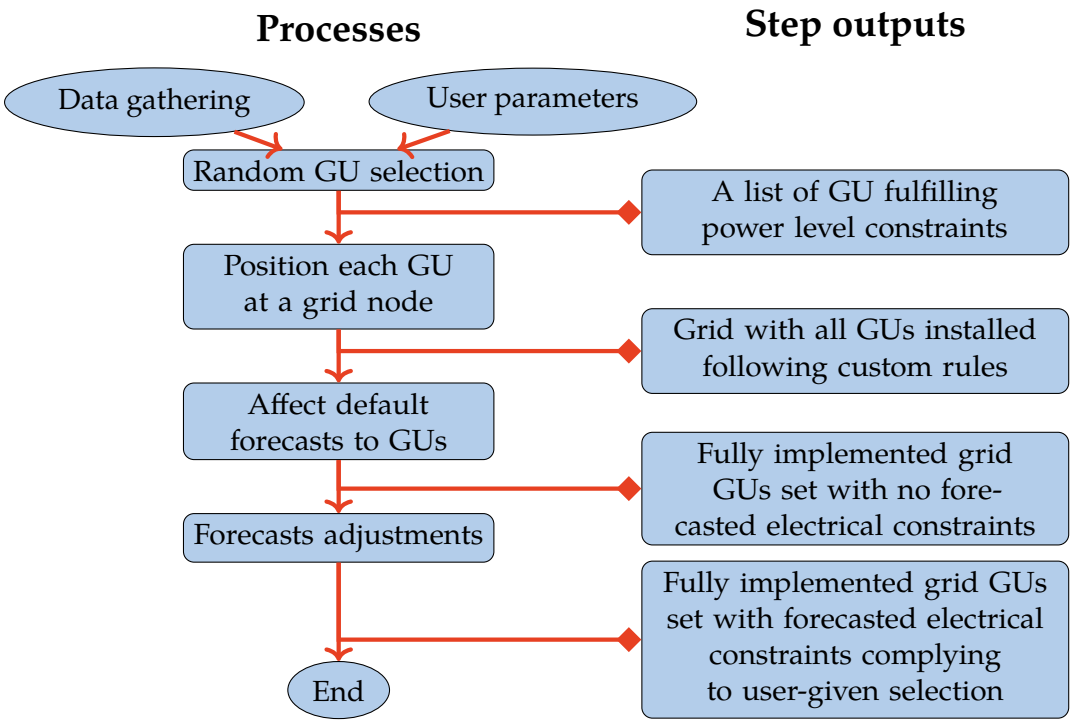


FIGURE D.2: Broad overview of the different steps in our data creation process.

As depicted in Figure D.2, we rely on pre-processed data from Enedis Open Data and user-given parameters. The latter are the dimensions of transformers on the grid, and a rule to compute the desired amount of installed and subscribed power connected to this network. The selection of GUs that will meet this aggregate goal of installed and subscribed power is computed according to random draws from statistical occurrences observed in Enedis Open Data.

USER GIVEN PARAMETERS

- matrix Y: grid topology and line characteristics;
- an installed consumption power target on the grid (MVA);

BIBLIOGRAPHY

[1] S. Y. Abdelouadoud et al. "Optimal power flow of a distribution system based on increasingly tight cutting planes added to a second order cone relaxation". In: *International Journal of Electrical Power and Energy Systems* 69 (2015), pp. 9–17. DOI: 10.1016/j.ijepes.2014.12.084. URL: http://dx.doi.org/10.1016/j.ijepes.2014.12.084.

[2] S. Abe et al. "Load Flow Convergence in the Vicinity of a Voltage Stability Limit". In: *IEEE Transactions on Power Apparatus and Systems* PAS-97.6 (Nov. 1978), pp. 1983–1993. DOI: 10.1109/TPAS.1978.354702. URL: http://ieeexplore.ieee.org/document/4181651/.

[3] M.A. Abido. "Optimal power flow using particle swarm optimization". In: *International Journal of Electrical Power & Energy Systems* 24.7 (Oct. 2002), pp. 563–571. DOI: 10.1016/S0142-0615(01)00067-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S0142061501000679.

[4] W. van Ackooij, I. Aleksovska, and M. Munoz-Zuniga. "(Sub-)Differentiability of Probability Functions with Elliptical Distributions". In: *Set-Valued and Variational Analysis* 26.4 (Dec. 2018), pp. 887–910. DOI: 10.1007/s11228-017-0454-3. URL: http://link.springer.com/10.1007/s11228-017-0454-3.

[5] W. van Ackooij, A. Frangioni, and W. de Oliveira. "Inexact stabilized Benders' decomposition approaches with application to chance-constrained problems with finite support". In: *Computational Optimization and Applications* 65.3 (2016), pp. 637–669. DOI: 10.1007/s10589-016-9851-z.

[6] W. van Ackooij and M. Minoux. "A Characterization of the Subdifferential of Singular Gaussian Distribution Functions". In: *Set-Valued and Variational Analysis* 23.3 (Sept. 2015), pp. 465–483. DOI: 10.1007/s11228-015-0317-8. URL: http://link.springer.com/10.1007/s11228-015-0317-8.

[7] W. van Ackooij and W. de Oliveira. "Addendum to the paper "Non-smooth DC-constrained optimization: constraint qualification and minimizing methodologies"". In: *Optimization Methods and Software* 34.4 (July 2019), pp. 890–920. DOI: 10.1080/10556788.2019.1595619. URL: https://www.researchgate.net/publication/348182693%7B%5C_%7DAddendum%7B%5C_%7DIDCApdf.

[8] W. van Ackooij and W. de Oliveira. "Non-smooth DC-constrained optimization: constraint qualification and minimizing methodologies". In: *Optimization Methods and Software* 34.4 (July 2019), pp. 890–920. DOI: 10.1080/10556788.2019.1595619. URL: https://www.tandfonline.com/doi/full/10.1080/10556788.2019.1595619.

[9] W. van Ackooij et al. "A bundle method for nonsmooth DC programming with application to chance-constrained problems". In: *Computational Optimization and Applications* 78.2 (Mar. 2021), pp. 451–490. DOI: 10.1007/s10589-020-00241-8. URL: http://link.springer.com/10.1007/s10589-020-00241-8.

[10] W. van Ackooij et al. "Large-scale unit commitment under uncertainty: an updated literature survey". In: *Annals of Operations Research* 271.1 (2018), pp. 11–85. DOI: 10.1007/s10479-018-3003-z. URL: https://doi.org/10.1007/s10479-018-3003-z.

[11] W. van Ackooij et al. "On joint probabilistic constraints with Gaussian coefficient matrix". In: *Operations Research Letters* 39.2 (Mar. 2011), pp. 99–102. DOI: 10.1016/j.orl.2011.01.005. URL: http://dx.doi.org/10.1016/j.orl.2011.01.005%20https://linkinghub.elsevier.com/retrieve/pii/S016763771100006X.

[12] W. van Ackooij et al. "Probabilistic optimization via approximate p-efficient points and bundle methods". In: *Computers & Operations Research* 77 (Jan. 2017), pp. 177–193. DOI: 10.1016/j.cor.2016.08.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0305054816301903.

[13] Wim van Ackooij. "A Discussion of Probability Functions and Constraints from a Variational Perspective". In: *Set-Valued and Variational Analysis* 28.4 (Dec. 2020), pp. 585–609. DOI: 10.1007/s11228-020-00552-2. URL: http://link.springer.com/10.1007/s11228-020-00552-2.

[14] Wim van Ackooij and R Henrion. "(Sub-)Gradient Formulae for Probability Functions of Random Inequality Systems under Gaussian Distribution". In: *SIAM/ASA Journal on Uncertainty Quantification* 5.1 (Jan. 2017), pp. 63–87. DOI: 10.1137/16M1061308. URL: https://epubs.siam.org/doi/10.1137/16M1061308.

[15] Wim van Ackooij and René Henrion. "Gradient Formulae for Nonlinear Probabilistic Constraints with Gaussian and Gaussian-Like Distributions". In: *SIAM Journal on Optimization* 24.4 (2014), pp. 1864–1889. DOI: 10.1137/130922689. URL: http://epubs.siam.org/doi/10.1137/130922689.

[16] Wim van Ackooij, René Henrion, and Pedro Pérez-Aros. "Generalized gradients for probabilistic/robust (probust) constraints". In: *Optimization* 69.7-8 (Aug. 2020), pp. 1451–1479. DOI: 10.1080/02331934.2019.1576670. URL: https://www.tandfonline.com/doi/full/10.1080/02331934.2019.1576670.

[17] Wim van Ackooij, Paul Javal, and Pedro Pérez-Aros. "Derivatives of Probability Functions: Unions of Polyhedra and Elliptical Distributions". In: *Set-Valued and Variational Analysis* (July 2021). DOI: 10.1007/s11228-021-00598-w. URL: https://link.springer.com/10.1007/s11228-021-00598-w.

[18] Wim van Ackooij and Jérôme Malick. "Eventual convexity of probability constraints with elliptical distributions". In: *Mathematical Programming* 175.1-2 (May 2019), pp. 1–27. DOI: 10.1007/s10107-018-1230-3. URL: http://link.springer.com/10.1007/s10107-018-1230-3.

[19] Wim van Ackooij and Jérôme Malick. "Second-order differentiability of probability functions". In: *Optimization Letters* 11.1 (2017), pp. 179–194. DOI: 10.1007/s11590-016-1015-7. URL: http://link.springer.com/10.1007/s11590-016-1015-7.

[20] Wim van Ackooij and Welington de Oliveira. "Level bundle methods for constrained convex optimization with various oracles". In: *Computational Optimization and Applications* 57.3 (Apr. 2014), pp. 555–597. DOI: 10.1007/s10589-013-9610-3. URL: http://link.springer.com/10.1007/s10589-013-9610-3.

[21] Wim van Ackooij and Pedro Pérez-Aros. "Generalized Differentiation of Probability Functions Acting on an Infinite System of Constraints". In: *SIAM Journal on Optimization* 29.3 (Jan. 2019), pp. 2179–2210. DOI: 10.1137/18M1181262. URL: https://epubs.siam.org/doi/10.1137/18M1181262.

[22] Wim van Ackooij and Claudia Sagastizábal. "Constrained Bundle Methods for Upper Inexact Oracles with Application to Joint Chance Constrained Energy Problems". In: *SIAM Journal on Optimization* 24.2 (Jan. 2014), pp. 733–765. DOI: 10.1137/120903099. URL: http://epubs.siam.org/doi/10.1137/120903099.

[23] Wim van Ackooij et al. "Chance Constrained Programming and Its Applications to Energy Management". In: *Stochastic Optimization - Seeing the Optimal for the Uncertain*. June. 2011. DOI: 10.5772/15438. URL: http://www.intechopen.com/books/stochastic-optimization-seeing-the-optimal-for-the-uncertain/chance-constrained-programming-and-its-applications-to-energy-management.

[24] Wim van Ackooij et al. "Joint chance constrained programming for hydro reservoir management". In: *Optimization and Engineering* 15.2 (2014), pp. 509–531. DOI: 10.1007/s11081-013-9236-4.

[25] Wim Van Ackooij. "Chance Constrained Programming: with applications in Energy Management". PhD thesis. Ecole Centrale Paris, 2013. DOI: 10.5772/15438. URL: https://tel.archives-ouvertes.fr/tel-00978519.

[26] Lukáš Adam et al. "Solving joint chance constrained problems using regularization and Benders' decomposition". In: *Annals of Operations Research* 292.2 (2020), pp. 683–709. DOI: 10.1007/s10479-018-3091-9. URL: http://link.springer.com/10.1007/s10479-018-3091-9%20https://doi.org/10.1007/s10479-018-3091-9.

[27] A. Águila Téllez et al. "Optimal reactive power compensation in electrical distribution systems with distributed resources. Review". In: *Heliyon* 4.8 (2018). DOI: 10.1016/j.heliyon.2018.e00746.

[28] Amir Ali Ahmadi and Georgina Hall. "DC decomposition of nonconvex polynomials with algebraic techniques". In: *Mathematical Programming* 169.1 (2018), pp. 69–94. DOI: 10.1007/s10107-017-1144-5. arXiv: 1510.01518.

[29] Mohammad Iman Alizadeh, Muhammad Usman, and Florin Capitanescu. "Envisioning security control in renewable dominated power systems through stochastic multi-period AC security constrained optimal power flow". In: (2021). arXiv: 2104.15038. URL: http://arxiv.org/abs/2104.15038.

[30] Xavier Allamigeon et al. "Log-Barrier Interior Point Methods Are Not Strongly Polynomial". In: *SIAM Journal on Applied Algebra and Geometry* 2.1 (2018), pp. 140–178. DOI: 10.1137/17M1142132. arXiv: 1708.01544.

[31] O. Alsac and B Stott. "Optimal Load Flow with Steady-State Security". In: *IEEE Transactions on Power Apparatus and Systems* PAS-93.3 (May 1974), pp. 745–751. DOI: 10.1109/TPAS.1974.293972. URL: http://ieeexplore.ieee.org/document/4075418/.

[32] Kyungsung An, Kyung Bin Song, and Kyeon Hur. "Incorporating charging/discharging strategy of electric vehicles into security-constrained optimal power flow to support high renewable penetration". In: *Energies* 10.5 (2017). DOI: 10.3390/en10050729.

[33] Nguyen Thai An and Nguyen Mau Nam. "Convergence analysis of a proximal point algorithm for minimizing differences of functions". In: *Optimization* 66.1 (2017), pp. 129–147. DOI: 10.1080/02331934.2016.1253694. arXiv: 1504.08079.

[34] Pierre Apkarian, Dominikus Noll, and Aude Rondepierre. "Mixed H_2/H_\infty Control via Nonsmooth Optimization". In: *SIAM Journal on Control and Optimization* 47.3 (Jan. 2008), pp. 1516–1546. DOI: 10.1137/070685026. URL: http://epubs.siam.org/doi/10.1137/070685026.

[35] Francisco J. Aragón Artacho, Ronan M. T. Fleming, and Phan T. Vuong. "Accelerating the DC algorithm for smooth functions". In: *Mathematical Programming* 169.1 (May 2018), pp. 95–118. DOI: 10.1007/s10107-017-1180-1. URL: http://link.springer.com/10.1007/s10107-017-1180-1.

[36] T Arnold. "A mixed-integer stochastic nonlinear optimization problem with joint probabilistic constraints". In: (2013), pp. 1–14. DOI: http://dx.doi.org/10.18452/8435.

[37] Hédy Attouch et al. "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-lojasiewicz inequality". In: *Mathematics of Operations Research* 35.2 (2010), pp. 438–457. DOI: 10.1287/moor.1100.0449. arXiv: 0801.1780.

[38] Sogol Babaeinejadsarookolaee et al. "The Power Grid Library for Benchmarking AC Optimal Power Flow Algorithms". In: *arXiv* (Aug. 2019), pp. 1–17. arXiv: 1908.02788. URL: http://arxiv.org/abs/1908.02788.

[39] Mark Bagnoli and Ted Bergstrom. "Log-concave probability and its applications". In: *Economic Theory* 26.2 (2005), pp. 445–469. DOI: 10.1007/s00199-004-0514-4. URL: http://link.springer.com/10.1007/s00199-004-0514-4.

[40] Shahab Bahrami and Vincent W.S. Wong. "Security-Constrained Unit Commitment for AC-DC Grids with Generation and Load Uncertainty". In: *IEEE Transactions on Power Systems* 33.3 (2018), pp. 2717–2732. DOI: 10.1109/TPWRS.2017.2749303.

[41] Kyri Baker. "Solutions of DC OPF are Never AC Feasible". In: (Nov. 2019), pp. 1–4. arXiv: 1912.00319. URL: http://arxiv.org/abs/1912.00319.

[42] Sebastian Banert and Radu Ioan Boţ. "A general double-proximal gradient algorithm for d.c. programming". In: *Mathematical Programming* 178.1-2 (Nov. 2019), pp. 301–326. DOI: 10.1007/s10107-018-1292-2. arXiv: /arxiv.org/abs/1610.06538 [https:]. URL: http://link.springer.com/10.1007/s10107-018-1292-2.

[43] M. Baran and F.F. Wu. "Optimal sizing of capacitors placed on a radial distribution system". In: *IEEE Transactions on Power Delivery* 4.1 (1989), pp. 735–743. DOI: 10.1109/61.19266. arXiv: arXiv:1011.1669v3. URL: http://ieeexplore.ieee.org/document/19266/.

[44] M.E. Baran and F.F. Wu. "Optimal capacitor placement on radial distribution systems". In: *IEEE Transactions on Power Delivery* 4.1 (1989), pp. 725–734. DOI: 10.1109/61.19265. URL: http://ieeexplore.ieee.org/document/19265/.

[45] Mesut E. Baran and Felix F Wu. "Network Reconfiguration in Distribution Systems for Loss Reduction and Load Balancing". In: *IEEE Power Engineering Review* 9.4 (1989), pp. 101–102. DOI: 10.1109/MPER.1989.4310642. URL: http://ieeexplore.ieee.org/document/4310642/.

[46] Jonathan F Bard. *Practical Bilevel Optimization*. Vol. 30. Nonconvex Optimization and Its Applications. Boston, MA: Springer US, 1998, xii, 473 p. DOI: 10.1007/978-1-4757-2836-1. URL: http://link.springer.com/10.1007/978-1-4757-2836-1.

[47] Julián Barquín et al. "On the Cost of the Reactive Power Generation and Voltage Support Service". In: *Bulk Power System Dynamics and Control IV-Restructuring* (1998).

[48] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Vol. 9. 2. Princeton: Princeton University Press, 2009, pp. 357–390. DOI: 10.1515/9781400831050. URL: http://epubs.siam.org/doi/10.1137/09075696X%20https://www.degruyter.com/view/books/9781400831050/9781400831050/9781400831050.xml.

[49] Arnaud Berthonnet. "L'électrification rurale". In: *Histoire & Sociétés Rurales* 19.1 (2003), p. 193. DOI: 10.3917/hsr.019.0193. URL: http://www.cairn.info/revue-histoire-et-societes-rurales-2003-1-page-193.htm?ref=doi.

[50] Dan Bienstock et al. "Mathematical programming formulations for the alternating current optimal power flow problem". In: *4OR* 18.3 (Sept. 2020), pp. 249–292. DOI: 10.1007/s10288-020-00455-w. arXiv: 2007.05334. URL: http://link.springer.com/10.1007/s10288-020-00455-w.

[51] Daniel Bienstock and Abhinav Verma. "Strong NP-hardness of AC power flows feasibility". In: *Operations Research Letters* 47.6 (Nov. 2019), pp. 494–501. DOI: 10.1016/j.orl.2019.08.009. arXiv: 1512.07315. URL: http://arxiv.org/abs/1512.07315%20https://linkinghub.elsevier.com/retrieve/pii/S016763771930247 0%20https://doi.org/10.1016/j.orl.2019.08.009.

[52] Christian Bingane, Miguel F. Anjos, and Sébastien Le Digabel. "Tight-and-Cheap Conic Relaxation for the Optimal Reactive Power Dispatch Problem". In: *IEEE Transactions on Power Systems* 34.6 (2019), pp. 4684–4693. DOI: 10.1109/TPWRS.2019.2912889. arXiv: 1810.03040.

[53] Christian Bingane, Miguel F. Anjos, and Sebastien Le Digabel. "Tight-and-cheap conic relaxation for the AC optimal power flow problem". In: *IEEE Transactions on Power Systems* 33.6 (2018), pp. 7181–7188. DOI: 10.1109/TPWRS.2018.2848965.

[54] Roger A. Blau. "Stochastic Programming and Decision Analysis: An Apparent Dilemma". In: *Management Science* 21.3 (Nov. 1974), pp. 271–276. DOI: 10.1287/mnsc.21.3.271. URL: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.21.3.271.

[55] Paul T. Boggs and Jon W. Tolle. "Sequential Quadratic Programming". In: *Acta Numerica* 4 (Jan. 1995), pp. 1–51. DOI: 10.1017/S0962492900002518. URL: https://www.cambridge.org/core/product/identifier/S0962492900002518/type/journal%7B%5C_%7Darticle.

[56] Paul T. Boggs and Jon W. Tolle. "Sequential quadratic programming for large-scale nonlinear optimization". In: *Journal of Computational and Applied Mathematics* 124.1-2 (2000), pp. 123–137. DOI: 10.1016/S0377-0427(00)00429-5.

[57] Abdullah Bokhari et al. "Experimental determination of the ZIP coefficients for modern residential, commercial, and industrial loads". In: *IEEE Transactions on Power Delivery* 29.3 (2014), pp. 1372–1381. DOI: 10.1109/TPWRD.2013.2285096.

[58] Immanuel M. Bomze and Claude Lemaréchal. "Necessary conditions for local optimality in difference-of-convex programming". In: *Journal of Convex Analysis* 17.2 (2010), pp. 673–680.

[59] Clément Bouttier and Ioana Gavra. "Convergence rate of a simulated annealing algorithm with noisy observations". In: *Journal of Machine Learning Research* 20 (2019), pp. 1–45. arXiv: 1703.00329.

[60] Jérôme Buire. "Intégration des incertitudes liées aux prévisions de consommation et de production à la gestion prévisionnelle d'un réseau de distribution". PhD thesis. Centrale Lille, 2018.

[61] Waqquas A. Bukhsh. *On Solving the Load Flow Problem as an Optimization Problem*. Tech. rep. University of Strathclyde, pp. 1–3.

[62] Waqquas A. Bukhsh et al. "Local Solutions of the Optimal Power Flow Problem". In: *IEEE Transactions on Power Systems* 28.4 (Nov. 2013), pp. 4780–4788. DOI: 10.1109/TPWRS.2013.2274577. URL: http://ieeexplore.ieee.org/document/6581918/.

[63] Marco C. Campi, Simone Garatti, and Maria Prandini. "The scenario approach for systems and control design". In: *Annual Reviews in Control* 33.2 (Dec. 2009), pp. 149–157. DOI: 10.1016/j.arcontrol.2009.07.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1367578809000479.

[64] F. Capitanescu et al. "State-of-the-art, challenges, and future trends in security constrained optimal power flow". In: *Electric Power Systems Research* 81.8 (Aug. 2011), pp. 1731–1741. DOI: 10.1016/j.epsr.2011.04.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0378779611000885.

[65] Florin Capitanescu. "AC OPF-Based methodology for exploiting flexibility provision at TSO/DSO interface via oltc-controlled demand reduction". In: *20th Power Systems Computation Conference, PSCC 2018* (2018), pp. 1–6. DOI: 10.23919/PSCC.2018.8442917.

[66] Florin Capitanescu. "Critical review of recent advances and further developments needed in AC optimal power flow". In: *Electric Power Systems Research* 136 (2016), pp. 57–68. DOI: 10.1016/j.epsr.2016.02.008. URL: http://dx.doi.org/10.1016/j.epsr.2016.02.008.

[67] Florin Capitanescu and Louis Wehenkel. "Sensitivity-Based Approaches for Handling Discrete Variables in Optimal Power Flow Computations". In: 25.4 (2010), pp. 1780–1789. DOI: 10.1109/TPWRS.2010.2044426.

[68] Florin Capitanescu et al. "Assessing the Potential of Network Reconfiguration to Improve Distributed Generation Hosting Capacity in Active Distribution Systems". In: *IEEE Transactions on Power Systems* 30.1 (Jan. 2015), pp. 346–356. DOI: 10.1109/TPWRS.2014.2320895. URL: http://ieeexplore.ieee.org/document/6818426/.

[69] Florin Capitanescu et al. "Cautious Operation Planning Under Uncertainties". In: *IEEE Transactions on Power Systems* 27.4 (Nov. 2012), pp. 1859–1869. DOI: 10.1109/TPWRS.2012.2188309. URL: http://ieeexplore.ieee.org/document/6175976/.

[70] Florin Capitanescu et al. "Interior-point based algorithms for the solution of optimal power flow problems". In: *Electric Power Systems Research* 77.5-6 (2007), pp. 508–517. DOI: 10.1016/j.epsr.2006.05.003.

[71] François Caron and Fabienne Cardot. *Histoire de l'électricité en France - 1881-1918*. Fayard. Paris, 1991, p. 1000.

[72] J. Carpentier. "Contribution à l'Étude du Dispatching Économique". In: *Bulletin de la Société Française des Électriciens*. 8e III.32 (1962), pp. 431–447.

[73] J.L. Carpentier. "Optimal Power Flows: Uses, Methods and Developments". In: *IFAC Proceedings Volumes* 18.7 (1985), pp. 11–21. DOI: 10.1016/S1474-6670(17)60410-5. URL: https://www.sciencedirect.com/science/article/pii/S1474667017604105.

[74] L. Casacio, C. Lyra, and A. R.L. Oliveira. "Interior point methods for power flow optimization with security constraints". In: *International Transactions in Operational Research* 26.1 (2019), pp. 364–378. DOI: 10.1111/itor.12279.

[75] A. Charnes, W. W. Cooper, and G. H. Symonds. "Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil". In: *Management Science* 4.3 (1958), pp. 235–263. DOI: 10.1287/mnsc.4.3.235.

[76] Jianqiang Cheng et al. "Chance-constrained economic dispatch with renewable energy and storage". In: *Computational Optimization and Applications* 70.2 (2018), pp. 479–502. DOI: 10.1007/s10589-018-0006-2. URL: https://doi.org/10.1007/s10589-018-0006-2.

[77] Michael Chertkov, Deepjyoti Deka, and Yury Dvorkin. "Optimal Ensemble Control of Loads in Distribution Grids with Network Constraints". In: *2018 Power Systems Computation Conference (PSCC)*. IEEE, June 2018, pp. 1–7. DOI: 10.23919/PSCC.2018.8442447. URL: https://ieeexplore.ieee.org/document/8442447/.

[78] Michael Chertkov et al. "Voltage Collapse and ODE Approach to Power Flows: Analysis of a Feeder Line with Static Disorder in Consumption/Production". In: (2018), pp. 1–8. arXiv: 1106.5003. URL: http://arxiv.org/abs/1106.5003.

[79] Gianfranco Chicco and Andrea Mazza. "Metaheuristic optimization of power and energy systems: Underlying principles and main issues of the 'rush to heuristics'". In: *Energies* 13.19 (2020). DOI: 10.3390/en13195097.

[80] Françoise Choay. *L'urbanisme, utopies et réalités - une anthologie*. Seuil. Paris, 1965, p. 448.

[81] Joe H. Chow. *PST Manual 3.0*. Tech. rep. 905. 2008, pp. 1–123. URL: https://www.ecse.rpi.edu/%7B~%7Dchowj/.

[82] Frank H. Clarke. *Optimization and Nonsmooth Analysis*. Society for Industrial and Applied Mathematics, Jan. 1990, pp. xii + 306. DOI: 10.1137/1.9781611971309. URL: http://epubs.siam.org/doi/book/10.1137/1.9781611971309.

[83] *Clean energy for all Europeans*. Tech. rep. 2. European Commission, 2019, p. 3. DOI: 10.2833/9937. URL: https://ec.europa.eu/energy/studies/study-quality-electricity-market-data-transmission-system-operators-electricity-supply%7B%5C_%7Den.

[84] Carleton Coffrin, Hassan L. Hijazi, and Pascal Van Hentenryck. "The QC Relaxation: A Theoretical and Computational Study on Optimal Power Flow". In: *IEEE Transactions on Power Systems* 31.4 (2016), pp. 3008–3018. DOI: 10.1109/TPWRS.2015.2463111. arXiv: 1502.07847.

[85] H. M. Costa et al. "A multi-temporal optimal power flow for managing storage and demand flexibility in LV networks". In: *2017 IEEE Manchester PowerTech, Powertech 2017* 645963 (2017), pp. 1–6. DOI: 10.1109/PTC.2017.7981166.

[86] Danish Energy Agency. *Energy Statistics 2018*. Tech. rep. Danish Energy Agency, 2018. URL: http://www.ens.dk.

[87] Darinka Dentcheva, Bogumila Lai, and Andrzej Ruszczyński. "Dual methods for probabilistic optimization problems". In: *Mathematical Methods of Operations Research* 60.2 (2004), pp. 331–346. DOI: 10.1007/s001860400371.

[88] Darinka Dentcheva, Bogumila Lai, and Andrzej Ruszczyński. *Efficient Point Methods for Probabilistic Optimization Problems*. Ed. by Giuseppe Calafiore and Fabrizio Dabbene. London: Springer London, 2006, pp. 1–17. DOI: 10.1007/b138725. URL: http://link.springer.com/10.1007/b138725.

[89] Darinka Dentcheva, András Prékopa, and Andrzej Ruszczyński. "Concavity and efficient points of discrete distributions in probabilistic programming". In: *Mathematical Programming, Series B* 89.1 (2000), pp. 55–77. DOI: 10.1007/PL00011393.

[90] Tao Ding et al. "A two-stage robust reactive power optimization considering uncertain wind power integration in active distribution networks". In: *IEEE Transactions on Sustainable Energy* 7.1 (2016), pp. 301–311. DOI: 10.1109/TSTE.2015.2494587.

[91] Elizabeth D. Dolan and Jorge J. Moré. "Benchmarking optimization software with performance profiles". In: *Mathematical Programming* 91.2 (Jan. 2002), pp. 201–213. DOI: 10.1007/s101070100263. URL: http://link.springer.com/10.1007/s101070100263.

[92] Hermann Dommel and William Tinney. "Optimal Power Flow Solutions". In: *IEEE Transactions on Power Apparatus and Systems* PAS-87.10 (Oct. 1968), pp. 1866–1876. DOI: 10.1109/TPAS.1968.292150. URL: http://ieeexplore.ieee.org/document/4073461/.

[93] Héloïse Dutrieux. "Méthodes pour la planification pluriannuelle des réseaux de distribution. Application à l'analyse technico-économique des solutions d'intégration des énergies renouvelables intermittentes." PhD thesis. Ecole Centrale de Lille, 2015, p. 332. URL: https://tel.archives-ouvertes.fr/tel-01491053.

[94] Yury Dvorkin. "A Chance-Constrained Stochastic Electricity Market". In: *IEEE Transactions on Power Systems* 35.4 (2020), pp. 2993–3003. DOI: 10.1109/TPWRS.2019.2961231. arXiv: 1906.06963.

[95] Brent Eldridge, Richard O'Neill, and Anya Castillo. "An Improved Method for the DCOPF With Losses". In: *IEEE Transactions on Power Systems* 33.4 (July 2018), pp. 3779–3788. DOI: 10.1109/TPWRS.2017.2776081. URL: https://ieeexplore.ieee.org/document/8119823/.

[96] Enedis. *Chiffres clés 2019*. Tech. rep. Enedis, 2019.

[97] Enedis. *Description physique du Réseau Public de Distribution*. Tech. rep. Enedis, 2017.

[98] Enedis. *Enedis Open Data*. 2020. URL: https://www.enedis.fr/open-data.

[99] Enedis. *Lancement des Offres de Raccordement Intelligentes*. Tech. rep. 2018.

[100] Emre Erdogan. "Ambiguous chance constrained programs: Algorithms and applications". PhD thesis. Columbia University, 2007, p. 113. URL: https://search.proquest.com/docview/304859203?accountid=188395.

[101] E. Erdoğan and G. Iyengar. "Ambiguous chance constrained problems and robust optimization". In: *Mathematical Programming* 107.1-2 (June 2006), pp. 37–61. DOI: 10.1007/s10107-005-0678-0. URL: http://link.springer.com/10.1007/s10107-005-0678-0.

[102] La Fabrique. *Aide en ligne - Dictionnaire de données Agrégats segmentés de consommation et de production électrique au pas demi-horaire*. Tech. rep. ENEDIS, 2019. URL: https://www.enedis.fr/agregats-segmentes-de-consommation-et-production-electriques-au-pas-12-h.

[103] Armand Fallieres, Barthou Louis, and Clemenceau Georges. *Loi du 15 juin 1906 sur les distributions d'énergie*. Paris, 1906. URL: https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=B2EE0B1B98E2E94EB8D05213464F0937.tplgfr38s%7B%5C_%7D2?idSectionTA=LEGISCTA000006100018%7B%5C&%7DcidTexte=JORFTEXT000000508752%7B%5C&%7DdateTexte=20151231.

[104] Kai-Tai Fang, Samuel Kotz, and Kai Wang Ng. *Symmetric Multivariate and Related Distributions*. Vol. 2. 19. Boston, MA: Springer US, 1990. DOI: 10.1007/978-1-4899-2937-2. URL: http://link.springer.com/10.1007/978-1-4899-2937-2.

[105] Masoud Farivar, Xinyang Zho, and Lijun Che. "Local voltage control in distribution systems: An incremental control algorithm". In: *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, Nov. 2015, pp. 732–737. DOI: 10.1109/SmartGridComm.2015.7436388. URL: http://ieeexplore.ieee.org/document/7436388/.

[106] M. H. Farshbaf-Shaker, R. Henrion, and D. Hömberg. "Properties of Chance Constraints in Infinite Dimensions with an Application to PDE Constrained Optimization". In: *Set-Valued and Variational Analysis* 26.4 (Dec. 2018), pp. 821–841. DOI: 10.1007/s11228-017-0452-5. URL: http://link.springer.com/10.1007/s11228-017-0452-5.

[107] William Feller. *An Introduction to Probability Theory and Its Applications, Vol 2*. Second edi. John Wiley and Sons, Inc, 1970, p. 669.

[108] Yang Feng. "Solving for the Low-Voltage/Large-Angle Power-Flow Solutions by Using the Holomorphic Embedding Method". PhD thesis. Arizona State University, 2015.

[109] Parlement Français. *Code de l'Energie*. 2020. URL: https://www.legifrance.gouv.fr/affichCode.do;jsessionid=79684B0C2EB3963219B8BABEAAFE6019.tplgfr42s%7B%5C_%7D2?idSectionTA=LEGISCTA000031748525%7B%5C&%7DcidTexte=LEGITEXT000023983208%7B%5C&%7DdateTexte=20200721.

[110] Parlement Français. *Loi 2000-108 du 10 février 2000 relative à la modernisation et au développement du service public de l'électricité*. Paris, 2000. URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000750321.

[111] Parlement Français. *Loi 2013-619 du 16 juillet 2013 portant diverses dispositions d'adaptation au droit de l'Union européenne dans le domaine du développement durable*. 2013. URL: https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=79684B0C2EB3963219B8BABEAAFE6019.tplgfr42s%7B%5C_%7D2?cidTexte=JORFTEXT000027713399%7B%5C&%7DcategorieLien=id%7B%5C#%7DJORFSCTA000027713402.

[112] Parlement Français. *Loi du 16 octobre 1919 relative à l'utilisation de l'énergie hydraulique*. Paris.

[113] Parlement Français. *Loi du 2 août 1923 facilitant par des avances de l'Etat la distribution de l'énergie électrique dans les campagnes*. Rambouillet, 1923. URL: https://energie-en-lumiere.fr/loi-du-2-aout-1923-facilitant-par-des-avances-de-letat-la-distribution-de-lenergie-electrique-dans-les-campagnes/.

[114] Antonio Frangioni and Claudio Gentile. "Solving Nonlinear Single-Unit Commitment Problems with Ramping Constraints". In: *Operations Research* 54.4 (2006), pp. 767–775. DOI: 10.1287/opre.1060.0309. URL: http://pubsonline.informs.org/doi/abs/10.1287/opre.1060.0309.

[115] Antonio Frangioni, Claudio Gentile, and Fabrizio Lacalandra. "Solving unit commitment problems with general ramp constraints". In: *International Journal of Electrical Power and Energy Systems* 30.5 (2008), pp. 316–326. DOI: 10.1016/j.ijepes.2007.10.003.

[116] Stephen Frank and Steffen Rebennack. "An introduction to optimal power flow: Theory, formulation, and examples". In: *IIE Transactions (Institute of Industrial Engineers)* 48.12 (2016), pp. 1172–1197. DOI: 10.1080/0740817X.2016.1189626.

[117] Stephen Frank, Ingrida Steponavice, and Steffen Rebennack. "Optimal power flow: a bibliographic survey I". In: *Energy Systems* 3.3 (Sept. 2012), pp. 221–258. DOI: 10.1007/s12667-012-0056-y. URL: http://link.springer.com/10.1007/s12667-012-0056-y.

[118] Stephen Frank, Ingrida Steponavice, and Steffen Rebennack. "Optimal power flow: a bibliographic survey II". In: *Energy Systems* 3.3 (Sept. 2012), pp. 259–289. DOI: 10.1007/s12667-012-0057-x. URL: http://link.springer.com/10.1007/s12667-012-0056-y%20http://link.springer.com/10.1007/s12667-012-0057-x.

[119] Stephen M. Frank and Steffen Rebennack. "Optimal design of mixed AC-DC distribution systems for commercial buildings: A nonconvex generalized benders decomposition approach". In: *European Journal of Operational Research* 242.3 (2015), pp. 710–729. DOI: 10.1016/j.ejor.2014.10.008.

[120] Pauline Gabillet. "Les entreprises locales de distribution à Grenoble et Metz, des outils de gouvernement énergétique urbain partiellement appropriés". PhD thesis. Université Paris-Est, 2015.

[121] Lingwen Gan and Steven H. Low. "An online gradient algorithm for optimal power flow on radial networks". In: *IEEE Journal on Selected Areas in Communications* 34.3 (2016), pp. 625–638. DOI: 10.1109/JSAC.2016.2525598.

[122] Lingwen Gan et al. "Exact Convex Relaxation of Optimal Power Flow in Radial Networks". In: *IEEE Transactions on Automatic Control* 60.1 (2015), pp. 72–87. DOI: 10.1109/TAC.2014.2332712. arXiv: 1311.7170.

[123] Manlio Gaudioso et al. "Minimizing nonsmooth DC functions via successive DC piecewise-affine approximations". In: *Journal of Global Optimization* 71.1 (May 2018), pp. 37–55. DOI: 10.1007/s10898-017-0568-z. URL: http://link.springer.com/10.1007/s10898-017-0568-z.

[124] Abebe Geletu et al. "Monotony analysis and sparse-grid integration for nonlinear chance constrained process optimization". In: *Engineering Optimization* 43.10 (Oct. 2011), pp. 1019–1041. DOI: 10.1080/0305215X.2010.532552. URL: http://www.tandfonline.com/doi/abs/10.1080/0305215X.2010.532552.

[125] Quentin Gemine, Damien Ernst, and Bertrand Cornélusse. "Active network management for electrical distribution systems: problem formulation, benchmark, and approximate solution". In: *Optimization and Engineering* 18.3 (Sept. 2017), pp. 587–629. DOI: 10.1007/s11081-016-9339-9. arXiv: 1405.2806. URL: http://link.springer.com/10.1007/s11081-016-9339-9.

[126] Busra Gencer, Erik Reimer Larsen, and Ann van Ackere. "Understanding the coevolution of electricity markets and regulation". In: *Energy Policy* 143.April 2018 (Aug. 2020), p. 111585. DOI: 10.1016/j.enpol.2020.111585. URL: https://doi.org/10.1016/j.enpol.2020.111585%20https://linkinghub.elsevier.com/retrieve/pii/S0301421520303256.

[127] Xinbo Geng and Le Xie. "Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization". In: *Annual Reviews in Control* 47 (2019), pp. 341–363. DOI: 10.1016/j.arcontrol.2019.05.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S1367578819300306.

[128] Alan Genz. "Numerical Computation of Multivariate Normal Probabilities". In: *Journal of Computational and Graphical Statistics* 1.2 (June 1992), p. 141. DOI: 10.2307/1390838. URL: https://www.jstor.org/stable/1390838?origin=crossref.

[129] Helena Gerard, Enrique Israel Rivero Puente, and Daan Six. "Coordination between transmission and distribution system operators in the electricity sector: A conceptual framework". In: *Utilities Policy* 50 (Feb. 2018), pp. 40–48. DOI: 10.1016/j.jup.2017.09.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957178717301285.

[130] Walter Gil-Gonzalez et al. "A Convex OPF Approximation for DC Networks Considering Voltage-Dependent Load Models". In: *2020 Ieee Andescon, Andescon 2020* (2020). DOI: 10.1109/ANDESCON50619.2020.9272042.

[131] Prettico Giuseppe, Marinopoulos Antonios, and Vitiello Silvia. *Distribution System Operator Observatory 2020*. Tech. rep. 2021. DOI: 10.2760/311966. URL: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/distribution-system-operator-observatory-2020.

[132] Hadrien Godard. "Résolution exacte du problème de l'optimisation des flux de puissance". PhD thesis. Conservatoire national des arts et metiers - CNAM, 2019. URL: https://tel.archives-ouvertes.fr/tel-02463203v2.

[133] Hadrien Godard et al. "Novel Approach Towards Global Optimality of Optimal Power Flow Using Quadratic Convex Optimization". In: *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, Apr. 2019, pp. 1227–1232. DOI: 10.1109/CoDIT.2019.8820584. URL: https://ieeexplore.ieee.org/document/8820584/.

[134] Sajjad Golshannavaz. "Cooperation of electric vehicle and energy storage in reactive power compensation: An optimal home energy management system considering PV presence". In: *Sustainable Cities and Society* 39 (May 2018), pp. 317–325. DOI: 10.1016/j.scs.2018.02.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S2210670717316098.

[135] T. González Grandón, H. Heitsch, and R. Henrion. "A joint model of probabilistic/robust constraints for gas transport management in stationary networks". In: *Computational Management Science* 14.3 (July 2017), pp. 443–460. DOI: 10.1007/s10287-017-0284-7. URL: http://link.springer.com/10.1007/s10287-017-0284-7.

[136] Ajit Gopalakrishnan et al. "Global optimization of Optimal Power Flow using a branch & bound algorithm". In: *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, Oct. 2012, pp. 609–616. DOI: 10.1109/Allerton.2012.6483274. URL: http://ieeexplore.ieee.org/document/6483274/.

[137] Claudia Gotzes et al. "On the quantification of nomination feasibility in stationary gas networks with random load". In: *Mathematical Methods of Operations Research* 84.2 (2016), pp. 427–457. DOI: 10.1007/s00186-016-0564-y.

[138] Ashraf A. Gouda and Tamás Szántai. "On numerical calculation of probabilities according to Dirichlet distribution". In: *Annals of Operations Research* 177.1 (June 2010), pp. 185–200. DOI: 10.1007/s10479-009-0601-9. URL: http://link.springer.com/10.1007/s10479-009-0601-9.

[139] John Grainger and William Stevenson. *Power System Analysis*. 1994.

[140] Etta Grover-Silva, Robin Girard, and George Kariniotakis. "Optimal sizing and placement of distribution grid connected battery systems through an SOCP optimal power flow algorithm". In: *Applied Energy* 219.August 2017 (2018), pp. 385–393. DOI: 10.1016/j.apenergy.2017.09.008. URL: https://doi.org/10.1016/j.apenergy.2017.09.008.

[141] William W. Hager, Dzung T. Phan, and Jiajie Zhu. "Projection algorithms for nonconvex minimization with application to sparse principal component analysis". In: *Journal of Global Optimization* 65.4 (2016), pp. 657–676. DOI: 10.1007/s10898-016-0402-z. arXiv: 1404.4132.

[142] Abderrahim Hantoute, René Henrion, and Pedro Pérez-Aros. "Subdifferential characterization of probability functions under Gaussian distribution". In: *Mathematical Programming* 174.1-2 (2019), pp. 167–194. DOI: 10.1007/s10107-018-1237-9. URL: http://link.springer.com/10.1007/s10107-018-1237-9.

[143] W. Hare, C. Sagastizábal, and M. Solodov. "A proximal bundle method for nonsmooth nonconvex functions with inexact information". In: *Computational Optimization and Applications* 63.1 (2016), pp. 1–28. DOI: 10.1007/s10589-015-9762-4.

[144] Holger Heitsch. "On probabilistic capacity maximization in a stationary gas network". In: *Optimization* 69.3 (Mar. 2020), pp. 575–604. DOI: 10.1080/02331934.2019.1625353. URL: https://www.tandfonline.com/doi/full/10.1080/02331934.2019.1625353.

[145] René Henrion. *Introduction to Chance Constrained Programming*. Tech. rep. 2004. URL: http://www.wias-berlin.de/people/henrion/ccp.ps.

[146] René Henrion and Andris Möller. "A gradient formula for linear chance constraints under gaussian distribution". In: *Mathematics of Operations Research* 37.3 (2012), pp. 475–488. DOI: 10.1287/moor.1120.0544.

[147] René Henrion and Werner Römisch. "Lipschitz and differentiability properties of quasi-concave and singular normal distribution functions". In: *Annals of Operations Research* 177.1 (June 2010), pp. 115–125. DOI: 10.1007/s10479-009-0598-0. URL: http://link.springer.com/10.1007/s10479-009-0598-0.

[148] René Henrion and Cyrille Strugarek. "Convexity of chance constraints with independent random variables". In: *Computational Optimization and Applications* 41.2 (Nov. 2008), pp. 263–276. DOI: 10.1007/s10589-007-9105-1. URL: http://link.springer.com/10.1007/s10589-007-9105-1.

[149] J.-B. Hiriart-Urruty. "Generalized Differentiability / Duality and Optimization for Problems Dealing with Differences of Convex Functions". In: *Journal of Chemical Information and Modeling*. Ed. by Jacob Ponstein. Vol. 256. Lecture Notes in Economics and Mathematical Systems 9. Berlin, Heidelberg: Springer Berlin Heidelberg, 1985, pp. 37–70. DOI: 10.1007/978-3-642-45610-7. arXiv: arXiv:1011.1669v3. URL: http://link.springer.com/10.1007/978-3-642-45610-7%20http://link.springer.com/10.1007/978-3-642-45610-7%7B%5C_%7D3.

[150] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Vol. 305. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, p. 432. DOI: 10.1007/978-3-662-02796-7. URL: http://link.springer.com/10.1007/978-3-662-02796-7.

[151] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms II*. Vol. 306. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993, p. 432. DOI: 10.1007/978-3-662-06409-2. URL: http://link.springer.com/10.1007/978-3-662-06409-2.

[152] Jean-Baptiste Hiriart-Urruty. *Generalized Differentiability, Duality and Optimization for problems dealing with Differences of Convex Functions*. Tech. rep. Université Paul Sabatier(Toulouse III), 1984. URL: https://www.carma.newcastle.edu.au/resources/jon/dc-hu.pdf.

[153] L. Jeff Hong, Yi Yang, and Liwei Zhang. "Sequential Convex Approximations to Joint Chance Constrained Programs: A Monte Carlo Approach". In: *Operations Research* 59.3 (June 2011), pp. 617–630. DOI: 10.1287/opre.1100.0910. URL: http://pubsonline.informs.org/doi/abs/10.1287/opre.1100.0910.

[154] L. Jeff Hong, Yi Yang, and Liwei Zhang. "Sequential Convex Approximations to Joint Chance Constrained Programs: A Monte Carlo Approach". In: *Operations Research* 59.3 (June 2011), pp. 617–630. DOI: 10.1287/opre.1100.0910. URL: http://pubsonline.informs.org/doi/abs/10.1287/opre.1100.0910.

[155] Thomas P. Hughes. *Networks of Power: Electrification in Western Society, 1880–1930*. Baltimore: Johns Hopkins University Press, 1983.

[156] Kenji Iba et al. "A Method for Finding A Pair of Multiple Load Flow Solutions in Bulk Power Systems". In: *IEEE Transactions on Power Systems* 5.2 (1990), pp. 582–591. DOI: 10.1109/59.54569.

[157] IEEE Task Force on Load Representation for Dynamic Performance. "Standard load models for power flow and dynamic performance simulation". In: *IEEE Transactions on Power Systems* 10.3 (1995), pp. 1302–1313. DOI: 10.1109/59.466523. URL: http://ieeexplore.ieee.org/document/466523/.

[158] INELFE. *Electricity Interconnection France-Spain web-page*. URL: https://www.inelfe.eu/en.

[159] A. F. Izmailov and M. V. Solodov. "Newton-Type Methods: A Broader View". In: *Journal of Optimization Theory and Applications* 164.2 (Feb. 2015), pp. 577–620. DOI: 10.1007/s10957-014-0580-0. URL: http://link.springer.com/10.1007/s10957-014-0580-0.

[160] Alexey F Izmailov and Mikhail V Solodov. *Newton-Type Methods for Optimization and Variational Problems*. Springer Series in Operations Research and Financial Engineering. Cham: Springer International Publishing, 2014. DOI: 10.1007/978-3-319-04247-3. URL: http://link.springer.com/10.1007/978-3-319-04247-3.

[161] Rabih A. Jabr. "Adjustable Robust OPF With Renewable Energy Sources". In: *IEEE Transactions on Power Systems* 28.4 (Nov. 2013), pp. 4742–4751. DOI: 10.1109/TPWRS.2013.2275013. URL: http://ieeexplore.ieee.org/document/6575173/.

[162] Rabih A. Jabr. "Radial distribution load flow using conic programming". In: *IEEE Transactions on Power Systems* 21.3 (2006), pp. 1458–1459. DOI: 10.1109/TPWRS.2006.879234.

[163] Marko Jereminov et al. "Impact of Load Models on Power Flow Optimization". In: *2019 IEEE Power & Energy Society General Meeting (PESGM)*. Vol. 2019-Augus. IEEE, Aug. 2019, pp. 1–5. DOI: 10.1109/PESGM40551.2019.8974085. URL: https://ieeexplore.ieee.org/document/8974085/.

[164] Kaisa Joki et al. "Double Bundle Method for finding Clarke Stationary Points in Nonsmooth DC Programming". In: *SIAM Journal on Optimization* 28.2 (Jan. 2018), pp. 1892–1919. DOI: 10.1137/16M1115733. URL: https://epubs.siam.org/doi/10.1137/16M1115733.

[165] Shinji Kataoka. "A Stochastic Programming Model". In: *Econometrica* 31.1/2 (Jan. 1963), p. 181. DOI: 10.2307/1910956. URL: https://www.jstor.org/stable/1910956?origin=crossref.

[166] M. Kefayat, A. Lashkar Ara, and S. A. Nabavi Niaki. "A hybrid of ant colony optimization and artificial bee colony algorithm for probabilistic optimal placement and sizing of distributed energy resources". In: *Energy Conversion and Management* 92 (2015), pp. 149–161. DOI: 10.1016/j.enconman.2014.12.037. URL: http://dx.doi.org/10.1016/j.enconman.2014.12.037.

[167] Vassilis Kekatos et al. "Stochastic Reactive Power Management in Microgrids with Renewables". In: *IEEE Transactions on Power Systems* 30.6 (2015), pp. 3386–3395. DOI: 10.1109/TPWRS.2014.2369452. arXiv: 1409.6758.

[168] Siddhartha Kumar Khaitan. "A survey of high-performance computing approaches in power systems". In: *2016 IEEE Power and Energy Society General Meeting (PESGM)*. Vol. 2016-Novem. IEEE, July 2016, pp. 1–5. DOI: 10.1109/PESGM.2016.7741984. URL: http://ieeexplore.ieee.org/document/7741984/.

[169] A. Klos and J. Wojcicka. "Physical aspects of the nonuniqueness of load flow solutions". In: *International Journal of Electrical Power and Energy Systems* 13.5 (1991), pp. 268–276. DOI: 10.1016/0142-0615(91)90050-6.

[170] Magnus Korpås et al. "Flexibility of controllable power transformers for managing wind uncertainty using robust adjustable linearised optimal power flow". In: *IET Renewable Power Generation* 13.2 (2018), pp. 262–272. DOI: 10.1049/iet-rpg.2018.5136.

[171] Nikolaos Koutsoukis and Pavlos Georgilakis. "A Chance-Constrained Multistage Planning Method for Active Distribution Networks". In: *Energies* 12.21 (Oct. 2019), p. 4154. DOI: 10.3390/en12214154. URL: https://www.mdpi.com/1996-1073/12/21/4154.

[172] Philip T. Krein. "Introduction". In: *Power Electronics Handbook*. Elsevier, 2011, pp. 1–14. DOI: 10.1016/B978-0-12-382036-5.00001-X. URL: https://linkinghub.elsevier.com/retrieve/pii/B978012382036500001X.

[173] Commission de régulation de L'énergie. *Délibération de la Commission de régulation de l'énergie du 17 novembre 2016 portant décision sur les tarifs d'utilisation des réseaux publics d'électricité dans les domaines de tension HTA et BT*. Tech. rep. 2016.

[174] Vincent Lagendijk. *Electrifying Europe - The power of Europe in the construction*. 2008. 2008. DOI: https://doi.org/10.6100/IR638264.

[175] Vincent Lagendijk. "Histoire de l'idée d'un système européen de l'électricité : projet, progrès, persistance". In: *Annales historiques de l'électricité* 6.1 (2008), p. 57. DOI: 10.3917/ahe.006.0057. URL: https://www.cairn.info/revue-annales-historiques-de-l-electricite-2008-1-page-57.htm%20http://www.cairn.info/revue-annales-historiques-de-l-electricite-2008-1-page-57.htm.

[176] Zinoviy M. Landsman and Emiliano A. Valdez. "Tail Conditional Expectations for Elliptical Distributions". In: *North American Actuarial Journal* 7.4 (Oct. 2003), pp. 55–71. DOI: 10.1080/10920277.2003.10596118. URL: http://www.tandfonline.com/doi/abs/10.1080/10920277.2003.10596118.

[177] Henning Lange et al. "Learning to Solve AC Optimal Power Flow by Differentiating through Holomorphic Embeddings". In: (Dec. 2020), pp. 1–10. arXiv: 2012.09622. URL: http://arxiv.org/abs/2012.09622.

[178] Y. Latreche et al. "Comprehensive Review of Radial Distribution Test Systems". 2020. DOI: 10.36227/techrxiv.12578648.v1.

[179] Javad Lavaei. "Zero duality gap for classical OPF problem convexifies fundamental nonlinear power problems". In: *American Control Conferenc* (2011), pp. 4566–4573. DOI: 10.1109/ACC.2011.5991104.

[180] Javad Lavaei and Steven H. Low. "Zero Duality Gap in Optimal Power Flow Problem". In: *IEEE Transactions on Power Systems* 27.1 (Feb. 2012), pp. 92–107. DOI: 10.1109/TPWRS.2011.2160974. URL: http://ieeexplore.ieee.org/document/5971792/.

[181] Irving H Lavalle. "On the 'Bayesability' of chance-constrained programming problems". In: *Operations Research Letters* 4.6 (Apr. 1986), pp. 281–283. DOI: 10.1016/0167-6377(86)90029-5. URL: https://linkinghub.elsevier.com/retrieve/pii/0167637786900295.

[182] Hoai An Le Thi, Van Ngai Huynh, and Tao Pham Dinh. "DC Programming and DCA for General DC Programs". In: 2014, pp. 15–35. DOI: 10.1007/978-3-319-06569-4_2. URL: http://link.springer.com/10.1007/978-3-319-06569-4%7B%5C_%7D2.

[183] Hoai An Le Thi and Tao Pham Dinh. "DC programming and DCA: thirty years of developments". In: *Mathematical Programming* 169.1 (2018), pp. 5–68. DOI: 10.1007/s10107-018-1235-y. URL: https://doi.org/10.1007/s10107-018-1235-y.

[184] Hoai An Le Thi, Tao Pham Dinh, and Huynh Van Ngai. "Exact penalty and error bounds in DC programming". In: *Journal of Global Optimization* 52.3 (Mar. 2012), pp. 509–535. DOI: 10.1007/s10898-011-9765-3. URL: http://link.springer.com/10.1007/s10898-011-9765-3.

[185] Hoai An Le Thi, Tao Pham Dinh, and Huynh Van Ngai. "Exact penalty and error bounds in DC programming". In: *Journal of Global Optimization* 52.3 (Mar. 2012), pp. 509–535. DOI: 10.1007/s10898-011-9765-3. URL: http://link.springer.com/10.1007/s10898-011-9765-3.

[186] Jon Lee and Sven Leyffer, eds. *Mixed Integer Nonlinear Programming*. Vol. 154. The IMA Volumes in Mathematics and its Applications. New York, NY: Springer New York, 2012. DOI: 10.1007/978-1-4614-1927-3. URL: http://link.springer.com/10.1007/978-1-4614-1927-3.

[187] C. Lehmkoster. "Security constrained optimal power flow for an economical operation of FACTS-devices in liberalized energy markets". In: *IEEE Transactions on Power Delivery* 17.2 (Apr. 2002), pp. 603–608. DOI: 10.1109/61.997946. URL: http://ieeexplore.ieee.org/document/997946/.

[188] Miguel Lejeune and Nilay Noyan. "Mathematical programming approaches for generating p-efficient points". In: *European Journal of Operational Research* 207.2 (2010), pp. 590–600. DOI: 10.1016/j.ejor.2010.05.025. URL: http://dx.doi.org/10.1016/j.ejor.2010.05.025.

[189] Miguel A. Lejeune and François Margot. "Solving Chance-Constrained Optimization Problems with Stochastic Quadratic Inequalities". In: *Operations Research* 64.4 (Aug. 2016), pp. 939–957. DOI: 10.1287/opre.2016.1493. URL: http://pubsonline.informs.org/doi/10.1287/opre.2016.1493.

[190] *Les services de Communications Electroniques en France*. Tech. rep. ARCEP, 2020.

[191] Bernard C. Lesieutre et al. "Examining the limits of the application of semidefinite programming to power flow problems". In: *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, Sept. 2011, pp. 1492–1499. DOI: 10.1109/Allerton.2011.6120344. URL: http://ieeexplore.ieee.org/document/6120344/.

[192] Thomas Leveringhaus et al. "Modelling of Sequential Optimal Power Flow by Piecewise Linear Convexificated Quadratic Approximations". In: *2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings* 201804270000112 (2019), pp. 87–92. DOI: 10.1109/POWERCON.2018.8601613.

[193] Maurice Levy-Leboyer and Henri Morsel. *Histoire de l'électricité en France - 1919-1946*. Fayard. Paris, 1994, p. 1438.

[194] Peijie Li et al. "An SQP method combined with gradient sampling for small-signal stability constrained OPF". In: *IEEE Transactions on Power Systems* 32.3 (2017), pp. 2372–2381. DOI: 10.1109/TPWRS.2016.2598266.

[195] Shin Yeu Lin and Jyun Fu Chen. "Distributed optimal power flow for smart grid transmission system with renewable energy sources". In: *Energy* 56 (2013), pp. 184–192. DOI: 10.1016/j.energy.2013.04.011. URL: http://dx.doi.org/10.1016/j.energy.2013.04.011.

[196] Paula A. Lipka et al. "Constructing transmission line current constraints for the IEEE and polish systems". In: *Energy Systems* 8.1 (Feb. 2017), pp. 199–216. DOI: 10.1007/s12667-016-0194-8. URL: http://link.springer.com/10.1007/s12667-016-0194-8.

[197] Thomas Lipp and Stephen Boyd. "Variations and extension of the convex–concave procedure". In: *Optimization and Engineering* 17.2 (June 2016), pp. 263–287. DOI: 10.1007/s11081-015-9294-x. URL: http://link.springer.com/10.1007/s11081-015-9294-x.

[198] Steven H. Low. "Convex Relaxation of Optimal Power Flow—Part I: Formulations and Equivalence". In: *IEEE Transactions on Control of Network Systems* 1.1 (Mar. 2014), pp. 15–27. DOI: 10.1109/TCNS.2014.2309732. URL: http://ieeexplore.ieee.org/document/6756976/.

[199] Steven H. Low. "Convex Relaxation of Optimal Power Flow—Part II: Exactness". In: *IEEE Transactions on Control of Network Systems* 1.2 (June 2014), pp. 177–189. DOI: 10.1109/TCNS.2014.2323634. URL: http://ieeexplore.ieee.org/document/6815671/.

[200] James Luedtke and Shabbir Ahmed. "A Sample Approximation Approach for Optimization with Probabilistic Constraints". In: *SIAM Journal on Optimization* 19.2 (Jan. 2008), pp. 674–699. DOI: 10.1137/070702928. URL: http://epubs.siam.org/doi/10.1137/070702928.

[201] K. Marti. "Differentiation formulas for probability functions: The transformation method". In: *Mathematical Programming, Series B* 75.2 (1996), pp. 201–220. DOI: 10.1007/BF02592152.

[202] Geneviève Massard-Guilbaud. "From the history of sources and sectors to the history of systems and transitions: how the history of energy has been written in France and beyond". In: *Journal of Energy History/Revue d'Histoire de l'Energie* 1 (2018). URL: http://www.energyhistory.eu/fr/dossier/history-sources-and-sectors-history-systems-and-transitions-how-history-energy-has-been.

[203] Dhagash Mehta, Daniel K. Molzahn, and Konstantin Turitsyn. "Recent advances in computational methods for the power flow equations". In: *Proceedings of the American Control Conference* 2016-July (2016), pp. 1753–1765. DOI: 10.1109/ACC.2016.7525170. arXiv: 1510.00073.

[204] D. Menniti et al. "A DSO-oriented mathematical model for dispersed generation management on MV networks". In: *IEEE Power and Energy Society General Meeting* (2012), pp. 1–8. DOI: 10.1109/PESGM.2012.6344600.

[205] Bruce L. Miller and Harvey M. Wagner. "Chance Constrained Programming with Joint Constraints". In: *Operations Research* 13.6 (1965), pp. 930–945. DOI: 10.1287/opre.13.6.930.

[206] Sidhant Misra, Line Roald, and Yeesian Ng. "Learning for Constrained Optimization: Identifying Optimal Active Constraint Sets". In: (2018), pp. 1–16. arXiv: 1802.09639. URL: http://arxiv.org/abs/1802.09639.

[207] Erfan Mohagheghi et al. "A Survey of Real-Time Optimal Power Flow". In: *Energies* 11.11 (2018), p. 3142. DOI: 10.3390/en11113142. URL: http://www.mdpi.com/1996-1073/11/11/3142.

[208] Daniel K. Molzahn and Ian A. Hiskens. "Sparsity-Exploiting Moment-Based Relaxations of the Optimal Power Flow Problem". In: *IEEE Transactions on Power Systems* 30.6 (2015), pp. 3168–3180. DOI: 10.1109/TPWRS.2014.2372478. arXiv: arXiv:1312.1992v2.

[209] Daniel K. Molzahn et al. "A Survey of Distributed Optimization and Control Algorithms for Electric Power Systems". In: *IEEE Transactions on Smart Grid* 8.6 (Nov. 2017), pp. 2941–2962. DOI: 10.1109/TSG.2017.2720471. arXiv: 1604.04683. URL: http://ieeexplore.ieee.org/document/7990560/.

[210] Daniel K. Molzahn et al. "Investigating the Maximum Number of Real Solutions to the Power Flow Equations: Analysis of Lossless Four-Bus Systems". In: (2016), pp. 2–7. arXiv: 1603.05908. URL: http://arxiv.org/abs/1603.05908.

[211] Outi Montonen and Kaisa Joki. "Bundle-based descent method for nonsmooth multiobjective DC optimization with inequality constraints". In: *Journal of Global Optimization* 72.3 (2018), pp. 403–429. DOI: 10.1007/s10898-018-0651-0. URL: https://doi.org/10.1007/s10898-018-0651-0.

[212] Outi Montonen and Kaisa Joki. "Multiobjective Double Bundle Method for DC Optimization". In: *Numerical Nonsmooth Optimization*. Cham: Springer International Publishing, 2020, pp. 481–497. DOI: 10.1007/978-3-030-34910-3_14. URL: http://link.springer.com/10.1007/978-3-030-34910-3%7B%5C_%7D14.

[213] Oscar Danilo Montoya, Walter Gil-González, and Alejandro Garces. "Sequential quadratic programming models for solving the OPF problem in DC grids". In: *Electric Power Systems Research* 169.July 2018 (2019), pp. 18–23. DOI: 10.1016/j.epsr.2018.12.008. URL: https://doi.org/10.1016/j.epsr.2018.12.008.

[214] I. J. Nagrath, D. P. Kothari, and R. C. Desai. *Modern Power System Analysis*. Vol. 12. 1. 2008, pp. 96–96. DOI: 10.1109/tsmc.1982.4308783.

[215] National Grid Electricity System Operator Limited. *The Grid code*. Tech. rep. 5 - Revision 38. 2019, pp. 1–998.

[216] Robert F Nau. "Note—Blau's Dilemma Revisited". In: *Management Science* 33.10 (Oct. 1987), pp. 1232–1237. DOI: 10.1287/mnsc.33.10.1232. URL: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.33.10.1232.

[217] Morteza Nazari-Heris and Behnam Mohammadi-Ivatloo. "Application of Robust Optimization Method to Power System Problems". In: *Classical and Recent Aspects of Power System Optimization*. Elsevier, 2018, pp. 19–32. DOI: 10.1016/B978-0-12-812441-3.00002-1. URL: http://dx.doi.org/10.1016/B978-0-12-812441-3.00002-1%20https://linkinghub.elsevier.com/retrieve/pii/B9780128124413000021.

[218] Arkadi Nemirovski and Alexander Shapiro. "Convex Approximations of Chance Constrained Programs". In: *SIAM Journal on Optimization* 17.4 (Jan. 2007), pp. 969–996. DOI: 10.1137/050622328. URL: http://epubs.siam.org/doi/10.1137/050622328.

[219] Yurii Nesterov. *Introductory Lectures on Convex Optimization*. Vol. 87. Applied Optimization 2. Boston, MA: Springer US, 2004, pp. 54–67. DOI: 10.1007/978-1-4419-8853-9. URL: http://repositorio.unan.edu.ni/2986/1/5624.pdf%20http://link.springer.com/10.1007/978-1-4419-8853-9.

[220] Ming NIU, Can WAN, and Zhao XU. "A review on applications of heuristic optimization algorithms for optimal power flow in modern power systems". In: *Journal of Modern Power Systems and Clean Energy* 2.4 (Dec. 2014), pp. 289–297. DOI: 10.1007/s40565-014-0089-4. URL: http://link.springer.com/10.1007/s40565-014-0089-4.

[221] Yi-shuai Niu. "On Difference-of-SOS and Difference-of-Convex-SOS Decompositions for Polynomials". In: March (2018). arXiv: arXiv:1803.09900v2.

[222] Dominikus Noll. "Convergence of Non-smooth Descent Methods Using the Kurdyka–Łojasiewicz Inequality". In: *Journal of Optimization Theory and Applications* 160.2 (Feb. 2014), pp. 553–572. DOI: 10.1007/s10957-013-0391-8. URL: http://link.springer.com/10.1007/s10957-013-0391-8.

[223] Dominikus Noll. "Cutting plane oracles for non-smooth trust-regions". In: (2019), pp. 1–36. arXiv: 1905.06797. URL: http://arxiv.org/abs/1905.06797.

[224] Hussam Nosair and Francois Bouffard. "Economic dispatch under uncertainty: The probabilistic envelopes approach". In: *IEEE Transactions on Power Systems* 32.3 (2017), pp. 1701–1710. DOI: 10.1109/TPWRS.2016.2602942.

[225] *Observatoire: Les marché de détail de l'électricité et du gaz naturel*. Tech. rep. Commission de régulation de l'énergie, 2020.

[226] L.F. Ochoa, C.J. Dent, and G.P. Harrison. "Maximisation of intermittent distributed generation in active networks". In: *CIRED Seminar 2008: SmartGrids for Distribution*. IEE, 2008, pp. 6–6. DOI: 10.1049/ic:20080411. URL: https://digital-library.theiet.org/content/conferences/10.1049/ic%7B%5C_%7D20080411.

[227] Tomasz Okon and Kazimierz Wilkosz. "Phase shifter models for steady state analysis". In: *2016 17th International Scientific Conference on Electric Power Engineering (EPE)*. IEEE, May 2016, pp. 1–6. DOI: 10.1109/EPE.2016.7521831. URL: http://ieeexplore.ieee.org/document/7521831/.

[228] Welington de Oliveira. "Proximal bundle methods for nonsmooth DC programming". In: *Journal of Global Optimization* 75.2 (Oct. 2019), pp. 523–563. DOI: 10.1007/s10898-019-00755-4. URL: http://link.springer.com/10.1007/s10898-019-00755-4.

[229] Welington de Oliveira. "The ABC of DC Programming". In: *Set-Valued and Variational Analysis* 28.4 (Dec. 2020), pp. 679–706. DOI: 10.1007/s11228-020-00566-w. URL: http://link.springer.com/10.1007/s11228-020-00566-w.

[230] Welington De Oliveira and Michel P Tcheou. "An inertial algorithm for DC programming". In: (2018).

[231] T.J. Overbye, Xu Cheng, and Yan Sun. "A comparison of the AC and DC power flow models for LMP calculations". In: *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*. Vol. 37. C. IEEE, 2004, 9 pp. DOI: 10.1109/HICSS.2004.1265164. URL: http://ieeexplore.ieee.org/document/1265164/.

[232] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro. "Sample Average Approximation Method for Chance Constrained Programming: Theory and Applications". In: *Journal of Optimization Theory and Applications* 142.2 (Aug. 2009), pp. 399–416. DOI: 10.1007/s10957-009-9523-6. URL: http://link.springer.com/10.1007/s10957-009-9523-6.

[233] Jong Shi Pang, Meisam Razaviyayn, and Alberth Alvarado. "Computing B-stationary points of nonsmooth DC programs". In: *Mathematics of Operations Research* 42.1 (2017), pp. 95–118. DOI: 10.1287/moor.2016.0795. arXiv: 1511.01796.

[234] S. R. Paranjothi and K. Anburaja. "Optimal Power Flow Using Refined Genetic Algorithm". In: *Electric Power Components and Systems* 30.10 (Oct. 2002), pp. 1055–1063. DOI: 10.1080/15325000290085343. URL: http://www.tandfonline.com/doi/abs/10.1080/15325000290085343.

[235] European Parliament. *Directive 96/92/EC of the European Parliament and of the Council of 19 December 1996 concerning common rules for the internal market in electricity*. 1996. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31996L0092%7B%5C&%7Dfrom=FR.

[236] Qiuyu Peng and Steven H. Low. "Distributed Optimal Power Flow Algorithm for Radial Networks, I: Balanced Single Phase Case". In: *IEEE Transactions on Smart Grid* 9.1 (Jan. 2018), pp. 111–121. DOI: 10.1109/TSG.2016.2546305. URL: http://ieeexplore.ieee.org/document/7440858/.

[237] Shen Peng. "Chance constrained problem and its applications". PhD thesis. Université Paris Saclay, 2019.

[238] Georg Ch. Pflug and Alois Pichler. *Multistage Stochastic Optimization*. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, 2014. DOI: 10.1007/978-3-319-08843-3. URL: http://link.springer.com/10.1007/978-3-319-08843-3.

[239] Tao Pham Dinh and Hoai An Le Thi. "Recent Advances in DC Programming and DCA". In: 2014, pp. 1–37. DOI: 10.1007/978-3-642-54455-2_1. URL: http://link.springer.com/10.1007/978-3-642-54455-2%7B%5C_%7D1.

[240] Ricardo B N M Pinheiro, Guilherme G Lage, and R M Geraldo. "A primal-dual integrated nonlinear rescaling approach applied to the optimal reactive dispatch problem". In: *European Journal of Operational Research* 276.3 (2019), pp. 1137–1153. DOI: 10.1016/j.ejor.2019.01.060. URL: https://doi.org/10.1016/j.ejor.2019.01.060.

[241] A. Prékopa and T. Szántai. "Flood control reservoir system design using stochastic programming". In: *Math Program Study*. Vol. 9. 9. 1978, pp. 138–151. DOI: 10.1007/BFb0120831. URL: http://link.springer.com/10.1007/BFb0120831.

[242] András Prékopa. "Logarithmic concave measures with application to stochastic programming". In: *Acta Scientiarum Mathematicarum* 32 (1971), pp. 301–316.

[243] András Prékopa. "On Probabilistic Constrained Programming". In: *Princeton Symposium on Mathematical Programming* (1970), pp. 113–138. DOI: 10.2307/j.ctt13x0wct.10. URL: http://rutcor.rutgers.edu/%7B~%7Dprekopa/prob.pdf.

[244] András Prékopa. "Probabilistic Programming". In: *Handbooks in Operations Research and Management Science*. 2003, pp. 267–351. DOI: 10.1016/S0927-0507(03)10005-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S0927050703100059.

[245] András Prékopa. *Stochastic Programming*. Dordrecht: Springer Netherlands, 1995. DOI: 10.1007/978-94-017-3087-7. URL: http://link.springer.com/10.1007/978-94-017-3087-7.

[246] *Principe et conditions de mise en oeuvre d'une régulation locale de puissance réactive pour les Installations de Production raccordées au Réseau Public de Distribution HTA*. Tech. rep. Enedis Direction Technique. URL: https://www.enedis.fr/sites/default/files/Enedis-NOI-RES%7B%5C_%7D60E.pdf.

[247] Bharath Rao, Friederich Kupzog, and Martin Kozek. "Three-Phase Unbalanced Optimal Power Flow Using Holomorphic Embedding Load Flow Method". In: *Sustainability* 11.6 (Mar. 2019), p. 1774. DOI: 10.3390/su11061774. URL: https://www.mdpi.com/2071-1050/11/6/1774.

[248] S. Surender Reddy, Vuddanti Sandeep, and Chan Mook Jung. "Review of stochastic optimization methods for smart grid". In: *Frontiers in Energy* 11.2 (2017), pp. 197–209. DOI: 10.1007/s11708-017-0457-7.

[249] Johannes Reichl, Michael Schmidthaler, and Stefan Schmidinger. *Blackout cost*. URL: http://www.blackout-simulator.com/.

[250] Line Roald and Goran Andersson. "Chance-Constrained AC Optimal Power Flow: Reformulations and Efficient Algorithms". In: *IEEE Transactions on Power Systems* 33.3 (May 2018), pp. 2906–2918. DOI: 10.1109/TPWRS.2017.2745410. arXiv: arXiv:1706.03241v1. URL: http://ieeexplore.ieee.org/document/8017474/.

[251] Line Roald et al. "Analytical reformulation of security constrained optimal power flow with probabilistic constraints". In: *2013 IEEE Grenoble Conference*. IEEE, June 2013, pp. 1–6. DOI: 10.1109/PTC.2013.6652224. URL: http://ieeexplore.ieee.org/document/6652224/.

[252] R. Tyrrell Rockafellar. *Optimization Under Uncertainty*. Tech. rep. Washington: University of Washington, 2001.

[253] R. Tyrrell Rockafellar and Roger J. B. Wets. *Variational Analysis*. Vol. 317. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer Berlin Heidelberg, Apr. 1998, pp. XII, 736. DOI: 10.1007/978-3-642-02431-3. URL: http://link.springer.com/10.1007/978-3-642-02431-3.

[254] Diego F. Rodriguez et al. "Low-Cost Analysis of Load Flow Computing Using Embedded Computer Empowered by GPU". In: *2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, Feb. 2021, pp. 1–5. DOI: 10.1109/ISGT49243.2021.9372280. URL: https://ieeexplore.ieee.org/document/9372280/.

[255] Ranjit Roy and H.T. Jadhav. "Optimal power flow solution of power system incorporating stochastic wind power using Gbest guided artificial bee colony algorithm". In: *International Journal of Electrical Power & Energy Systems* 64 (Jan. 2015), pp. 562–578. DOI: 10.1016/j.ijepes.2014.07.010. URL: http://dx.doi.org/10.1016/j.ijepes.2014.07.010%20https://linkinghub.elsevier.com/retrieve/pii/S0142061514004426.

[256] Johannes O. Royset. "Approximations of semicontinuous functions with applications to stochastic optimization and statistical estimation". In: *Mathematical Programming* 184.1-2 (Nov. 2020), pp. 289–318. DOI: 10.1007/s10107-019-01413-z. arXiv: 1709.06730. URL: https://doi.org/10.1007/s10107-019-01413-z%20http://link.springer.com/10.1007/s10107-019-01413-z.

[257] RTE. *Bilan Electrique 2018*. Tech. rep. RTE, 2018, pp. 1–175.

[258] RTE. *Documentation Technique de Référence*. Tech. rep. 2014, pp. 1–17. URL: https://www.enedis.fr/sites/default/files/Enedis-NOI-RES%7B%5C_%7D71E.pdf.

[259] RTE. *Règles relatives à la Programmation, au Mécanisme d'Ajustement et au Recouvrement des charges d'ajustement*. Tech. rep. RTE, 2020, pp. 1–107. URL: http://clients.rte-france.com/htm/fr/offre/telecharge/MARE%7B%5C_%7DSection%7B%5C_%7D1%7B%5C_%7D20200601.pdf.

[260] RTE. *Relations entre le GRD et le Responsable d'Equilibre Dispositions applicables pour l'estimation des Courbes de Charges*. Tech. rep. RTE, 2015, pp. 1–44.

[261] RTE. *Schéma décennal de développement du réseau*. Tech. rep. RTE, 2019.

[262] Gunter Rudolph. "Convergence analysis of canonical genetic algorithms". In: *IEEE Transactions on Neural Networks* 5.1 (1994), pp. 96–101. DOI: 10.1109/72.265964. URL: http://ieeexplore.ieee.org/document/265964/.

[263] Albert M. Sasson and G. J. Fisher. "Nonlinear Programming Solutions for Load-Flow, Minimum-Loss, and Economic Dispatching Problems". In: *IEEE Transactions on Power Apparatus and Systems* PAS-88.4 (1969), pp. 399–409. DOI: 10.1109/TPAS.1969.292460.

[264] Kathryn M. Schumacher et al. "Algorithm to solve a chance-constrained network capacity design problem with stochastic demands and finite support". In: *Naval Research Logistics (NRL)* 63.3 (Apr. 2016), pp. 236–246. DOI: 10.1002/nav.21685. URL: http://doi.wiley.com/10.1002/nav.21685.

[265] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming*. 2009. DOI: 10.1137/1.9780898718751.

[266] Zhangliang Shen et al. "Representing ZIP loads in convex relaxations of optimal power flow problems". In: *International Journal of Electrical Power & Energy Systems* 110.September 2018 (Sept. 2019), pp. 372–385. DOI: 10.1016/j.ijepes.2019.03.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S0142061518330151.

[267] Joao Silva et al. "Estimating the Active and Reactive Power Flexibility Area at the TSO-DSO Interface". In: *IEEE Transactions on Power Systems* 33.5 (2018), pp. 4741–4750. DOI: 10.1109/TPWRS.2018.2805765.

[268] Julie Sliwak et al. "A Julia Module for Polynomial Optimization with Complex Variables applied to Optimal Power Flow". In: *2019 IEEE Milan PowerTech*. IEEE, June 2019, pp. 1–6. DOI: 10.1109/PTC.2019.8810960. arXiv: 1904.08894. URL: https://ieeexplore.ieee.org/document/8810960/.

[269] J. Soares et al. "Ant Colony Search algorithm for the optimal power flow problem". In: *2011 IEEE Power and Energy Society General Meeting*. IEEE, July 2011, pp. 1–8. DOI: 10.1109/PES.2011.6039840. URL: https://ieeexplore.ieee.org/document/6039840/.

[270] Tiago Soares et al. "Active Distribution Grid Management Based on Robust AC Optimal Power Flow". In: *IEEE Transactions on Smart Grid* 9.6 (Nov. 2018), pp. 6229–6241. DOI: 10.1109/TSG.2017.2707065. URL: https://ieeexplore.ieee.org/document/7932508/.

[271] Edilaine Martins Soler, Vanusa Alves de Sousa, and Geraldo R.M. da Costa. "A modified Primal–Dual Logarithmic-Barrier Method for solving the Optimal Power Flow problem with discrete and continuous control variables". In: *European Journal of Operational Research* 222.3 (Nov. 2012), pp. 616–622. DOI: 10.1016/j.ejor.2012.05.021. URL: http://dx.doi.org/10.1016/j.ejor.2012.05.021%20https://linkinghub.elsevier.com/retrieve/pii/S0377221712003700.

[272] T. Sousa et al. "Simulated Annealing metaheuristic to solve the optimal power flow". In: *2011 IEEE Power and Energy Society General Meeting*. IEEE, July 2011, pp. 1–8. DOI: 10.1109/PES.2011.6039543. URL: https://ieeexplore.ieee.org/document/6039543/.

[273] Laure Soutenet-Martin. "Les politiques de promotion commerciale urbaine des sociétés de production-distribution d'électricité en France dans l'entre-deux-guerres". In: *Annales historiques de l'électricité* 7.1 (2009), p. 41. DOI: 10.3917/ahe.007.0041.

[274] Brian Stott. "Review of Load-Flow Calculation Methods". In: *Proceedings of the IEEE* 62.7 (1974), pp. 916–929. DOI: 10.1109/PROC.1974.9544.

[275] Alexander S. Strekalovsky and Ilya M. Minarchenko. "A local search method for optimization problem with d.c. inequality constraints". In: *Applied Mathematical Modelling* 58 (2018), pp. 229–244. DOI: 10.1016/j.apm.2017.07.031. URL: https://doi.org/10.1016/j.apm.2017.07.031.

[276] *Study on the quality of electricity market data of transmission system operators, electricity supply disruptions, and their impact on the European electricity markets - Final Report*. Tech. rep. March. European Commission, 2018. URL: https://ec.europa.eu/energy/studies/study-quality-electricity-market-data-transmission-system-operators-electricity-supply%7B%5C_%7Den.

[277] Bose Subhonmesh, Steven H. Low, and K. Mani Chandy. "Equivalence of branch flow and bus injection models". In: *50th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2012*. 2012, pp. 1893–1899. DOI: 10.1109/Allerton.2012.6483453.

[278] Bhargav Prasanna Swaminathan. "Gestion prévisionnelle des réseaux actifs de distribution". PhD thesis. Grenoble INP, 2017.

[279] *Synthèse de l'étude publiée par l'UIE: Patrimoine eau potable, assainissement collectif et eaux pluviales en France*. Tech. rep. Union Nationale des Industries et Entreprises de l'Eau, 2016.

[280] Milad Tahanan et al. "Large-scale Unit Commitment under uncertainty". In: *4OR* 13.2 (2015), pp. 115–171. DOI: 10.1007/s10288-014-0279-y.

[281] Yujie Tang, Krishnamurthy Dvijotham, and Steven Low. "Real-Time Optimal Power Flow". In: *IEEE Transactions on Smart Grid* 8.6 (2017), pp. 2963–2973. DOI: 10.1109/TSG.2017.2704922.

[282] J.F Toland. "Duality in nonconvex optimization". In: *Journal of Mathematical Analysis and Applications* 66.2 (Nov. 1978), pp. 399–415. DOI: 10.1016/0022-247X(78)90243-3. URL: https://linkinghub.elsevier.com/retrieve/pii/0022247X78902433.

[283] Bach Tran. "Advanced Difference of Convex functions Algorithms for some topics of Machine Learning with Big Data". PhD thesis. Université de Lorraine, 2020.

[284] Antonio Trias. "The Holomorphic Embedding Load Flow method". In: *IEEE Power and Energy Society General Meeting* July (2012), pp. 1–8. DOI: 10.1109/PESGM.2012.6344759.

[285] Hoang Tuy. *Convex Analysis and Global Optimization*. Vol. 110. Springer Optimization and Its Applications. Cham: Springer International Publishing, 2016, pp. XII, 340. DOI: 10.1007/978-3-319-31484-6. URL: http://link.springer.com/10.1007/978-3-319-31484-6.

[286] European Union. *TDX-ASSIST - Coordination of Transmission and Distribution data eXchanges for renewables integration in the European marketplace through Advanced, Scalable and Secure ICT Systems and Tools*. Tech. rep. Eurpean Union's Horizon 2020 research and innovation Program. URL: http://www.tdx-assist.eu/index.php.

[287] Texas AM University. *Texas A&M University Electric Grid Datasets*. URL: https://electricgrids.engr.tamu.edu/.

[288] Wim Van Ackooij et al. "On probabilistic constraints induced by rectangular sets and multivariate normal distributions". In: *Mathematical Methods of Operations Research* 71.3 (June 2010), pp. 535–549. DOI: 10.1007/s00186-010-0316-3. URL: http://link.springer.com/10.1007/s00186-010-0316-3.

[289] Kiriakos Vlahos, Peter Kall, and Stein W. Wallace. "Stochastic Programming." In: *The Journal of the Operational Research Society* 46.9 (1995), p. 1155. DOI: 10.2307/2584504. URL: https://www.jstor.org/stable/2584504?origin=crossref.

[290] Maria Vrakopoulou et al. "Probabilistic security-constrained AC optimal power flow". In: *2013 IEEE Grenoble Conference PowerTech, POWERTECH 2013* (2013), pp. 1–6. DOI: 10.1109/PTC.2013.6652374.

[291] Stein W. Wallace and Stein-Erik Fleten. "Stochastic Programming Models in Energy". In: *Handbooks in Operations Research and Management Science*. Vol. 10. C. 2003, pp. 637–677. DOI: 10.1016/S0927-0507(03)10010-2. URL: https://linkinghub.elsevier.com/retrieve/pii/S0927050703100102.

[292] Yehuda Wallach. "Gradient Methods for Load-Flow Problems". In: *IEEE Transactions on Power Apparatus and Systems* PAS-87.5 (1968), pp. 1314–1318. DOI: 10.1109/TPAS.1968.292223.

[293] Wei Wang and Shabbir Ahmed. "Sample average approximation of expected value constrained stochastic programs". In: *Operations Research Letters* 36.5 (2008), pp. 515–519. DOI: 10.1016/j.orl.2008.05.003.

[294] E. H. Watanabe et al. "Flexible AC transmission systems". In: *Power Electronics Handbook* (2011), pp. 851–877. DOI: 10.1016/B978-0-12-382036-5.00032-X.

[295] Tillmann Weisser, Line Roald, and Sidhant Misra. "Chance-Constrained Optimization for Non-Linear Network Flow Problems". In: (Mar. 2018), pp. 1–11. arXiv: 1803.02696. URL: http://arxiv.org/abs/1803.02696.

[296] Bo Wen, Xiaojun Chen, and Ting Kei Pong. "A proximal difference-of-convex algorithm with extrapolation". In: *Computational Optimization and Applications* 69.2 (Mar. 2018), pp. 297–324. DOI: 10.1007/s10589-017-9954-1. arXiv: 1612.06265. URL: http://link.springer.com/10.1007/s10589-017-9954-1.

[297] Shuli Wen et al. "Economic allocation for energy storage system considering wind power distribution". In: *IEEE Transactions on Power Systems* 30.2 (2015), pp. 644–652. DOI: 10.1109/TPWRS.2014.2337936.

[298] Edmund Taylor Whittaker. *A History of the Theories of, Aether and Electricity from the Age of Descartes to the Close of the Nineteenth Century*. London: Longmans, Green & Co, 1911, p. 502. DOI: 10.1126/science.34.871.316. URL: http://www.biodiversitylibrary.org/bibliography/19630.

[299] X. Xia and A. M. Elaiw. "Optimal dynamic economic dispatch of generation: A review". In: *Electric Power Systems Research* 80.8 (2010), pp. 975–986. DOI: 10.1016/j.epsr.2009.12.012. URL: http://dx.doi.org/10.1016/j.epsr.2009.12.012.

[300] H. Yu et al. "A Chance Constrained Transmission Network Expansion Planning Method With Consideration of Load and Wind Farm Uncertainties". In: *IEEE Transactions on Power Systems* 24.3 (Aug. 2009), pp. 1568–1576. DOI: 10.1109/TPWRS.2009.2021202. URL: http://ieeexplore.ieee.org/document/4909013/.

[301] Peiran Yu, Ting Kei Pong, and Zhaosong Lu. "Convergence rate analysis of a sequential convex programming method with line search for a class of constrained difference-of-convex optimization problems". In: (2020), pp. 1–28. arXiv: 2001.06998. URL: http://arxiv.org/abs/2001.06998.

[302] Yuan Yuan, Zukui Li, and Biao Huang. "Robust optimization approximation for joint chance constrained optimization problem". In: *Journal of Global Optimization* 67.4 (Apr. 2017), pp. 805–827. DOI: 10.1007/s10898-016-0438-0. URL: http://link.springer.com/10.1007/s10898-016-0438-0.

[303] A L Yuille and Anand Rangarajan. "The Concave-Convex Procedure". In: *Neural Computation* 15.4 (Apr. 2003), pp. 915–936. DOI: 10.1162/08997660360581958. URL: https://direct.mit.edu/neco/article/15/4/915-936/6726.

[304] R. Zarate-Minano et al. "Securing Transient Stability Using Time-Domain Simulations Within an Optimal Power Flow". In: *IEEE Transactions on Power Systems* 25.1 (Feb. 2010), pp. 243–253. DOI: 10.1109/TPWRS.2009.2030369. URL: http://ieeexplore.ieee.org/document/5288558/.

[305] Hui Zhang and Pu Li. "Chance Constrained Programming for Optimal Power Flow Under Uncertainty". In: *IEEE Transactions on Power Systems* 26.4 (Nov. 2011), pp. 2417–2424. DOI: 10.1109/TPWRS.2011.2154367. URL: http://ieeexplore.ieee.org/document/5871715/.

[306] Wenjuan Zhang, Fangxing Li, and Leon M. Tolbert. "Review of reactive power planning: Objectives, constraints, and algorithms". In: *IEEE Transactions on Power Systems* 22.4 (2007), pp. 2177–2186. DOI: 10.1109/TPWRS.2007.907452.

[307] Xiao-Ping Zhang, Christian Rehtanz, and Bikash Pal. "Modeling of FACTS-Devices in Optimal Power Flow Analysis". In: *Power Systems*. Vol. 11. 2012, pp. 113–156. DOI: 10.1007/978-3-642-28241-6_4. URL: http://link.springer.com/10.1007/978-3-642-28241-6%7B%5C_%7D4.

[308] Xinyang Zhou et al. "Reverse and Forward Engineering of Local Voltage Control in Distribution Networks". In: (Jan. 2018). arXiv: 1801.02015. URL: http://arxiv.org/abs/1801.02015.

[309] Y. Zhu and K. Tomsovic. "Optimal distribution power flow for systems with distributed energy resources". In: *International Journal of Electrical Power and Energy Systems* 29.3 (2007), pp. 260–267. DOI: 10.1016/j.ijepes.2006.07.007.

[310] Ray D Zimmerman and Carlos E Murillo-Sanchez. *Matpower User's Manual*. Version 7. 2019. DOI: 10.5281/zenodo.3251119. URL: https://matpower.org/docs/MATPOWER-manual-7.0.%7B%5C%%7D0Apdf.

RÉSUMÉ

L'arrivée massive de systèmes de production d'électricité renouvelable et la libéralisation du secteur de l'énergie ont un impact durable sur la gestion des réseaux électriques. En particulier, le réseau de distribution électrique est devenu un lieu d'interactions entre acteurs de marchés et son gestionnaire unique: l'opérateur du système de distribution (DSO). Parmi les enjeux majeurs du DSO il y a la gestion prévisionnelle court-terme: actionner une série de leviers pour permettre l'exploitation sûre du réseau suivant les prévisions d'activité à l'horizon de quelques heures; sa décision reposant sur le modèle mathématique d'Optimal Power Flow (OPF). Les sources d'incertitudes sur les prévisions se multiplient avec le nombre d'acteurs du réseau, et les outils de gestion doivent évoluer en conséquence.
Cette thèse porte sur l'intégration des incertitudes liées à la production et la consommation électrique par des contraintes de probabilités dans l'OPF. Le problème d'optimisation ainsi étendu est non-convexe et non-lisse, mais présente une structure Différence-de-Convexes (DoC). La classe des fonctions DoC généralise les classes de fonctions convexes et concaves, et inclut en particulier des approximations de précision arbitraire de toute fonction continue, tout en conservant des propriétés de régularité fortes qu'il s'agit d'exploiter dans une solution générique de programmation mathématique.
Une première contribution porte sur le développement et l'étude de convergence d'un algorithme de faisceaux pour les programmes mathématiques avec objectif et contraintes DoC. Une formulation DoC de contraintes de probabilités est ensuite présentée puis appliquée à l'OPF probabiliste. Une caractérisation de l'information du premier ordre des fonctions probabilistes est réalisée, basée sur une analyse des variations locales. Cette dernière étude des fonctions probabilistes rappelle que résoudre un OPF probabiliste en exploitant la structure DoC n'est pas l'unique choix de résolution. Quatre formulations explicites d'OPFs probabilistes, dont la structure DoC est démontrée, sont ainsi présentées. La performance de l'algorithme, l'impact de la paramétrisation et l'intérêt respectif des différents modèles sont validés numériquement sur un réseau électrique réaliste de 33 nœuds. Outre des temps de résolution raisonnables, cette méthodologie originale se distingue par sa capacité à rendre immédiatement accessible et contrôlable la viabilité électrique des décisions du DSO.

MOTS CLÉS

Différence-de-convexes, Contraintes de probabilité, Optimisation de flux de puissance électrique

ABSTRACT

The increasing integration of renewable energy sources has a long-lasting impact on the electrical grid, and the liberalisation of the energy sector has significantly changed its regulatory environment. In particular, the distribution network has become an area of interactions of competitive actors, while being managed by a single actor: the distribution system operator (DSO). Among the DSO' challenges is the short-term operational planning: the selection and activation of levers to ensure the safe exploitation of the grid, taking into account the forecasts of grid users' activities. Decisions in this context are based on the mathematical model of the Optimal Power Flow (OPF). Sources of uncertainties on these latter forecasts are growing due to the increasing number of actors on the grid.
The focus of this thesis is on the integration of uncertainties on power production and consumption in the OPF, using chance-constraints. The resulting probabilistic OPF model is a non-convex non-smooth optimization problem with a Difference-of-Convex (DoC) structure. The class of DoC functions is large enough to include convex, concave, and approximations of arbitrarily precision of every continuous function, while offering strong regularity properties that one can leverage to derive a generic optimisation algorithm.
A first contribution of this work is the development of an original bundle algorithm for the class of DoC constrained DoC problems. Chance-constraints are proved to be DoC, and a DoC approximation of chance-constraints is proposed before being applied to the probabilist OPF. A characterization of the first-order information of probabilist functions is presented, based on a variational study of these latter functions. This characterization highlights the variety of choices when it comes to solving chance-constrained programs. Four explicit formulations of probabilist OPFs are then proposed, and their DoC structure is proved. The algorithm's performance, the impact of parametrisation on its behaviour and the interest of each model are numerically validated on a 33 nodes network. Besides the reasonable computing times, this methodology is particularly relevant as, differently to other works in literature, the electrical viability and validity of a solution are directly accessible.

KEYWORDS

Difference-of-convex, Chance-constrained optimisation, Optimal Power Flow