PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Mines Paris-PSL

Gestion prévisionnelle optimisée sous incertitudes jointes
Operational planning under joint uncertainties

Soutenue par

Ksenia SYRTSEVA

Le 17 février 2025

Ecole doctorale n°84

Sciences et Technologies de l'Information et de la Communication

Spécialité

Contrôle, Optimisation, Prospective

Composition du jury :

Sourour, ELLOUMI Professeure, ENSTA Paris	*Présidente*
Line, ROALD Professor, University of Wisconsin	*Rapporteuse*
Claudia, SAGASTIZÁBAL Professor, University of Campinas	*Rapporteuse*
Hugo, MORAIS Senior Researchor, INESC-ID	*Examinateur*
Welington, DE OLIVEIRA Chargé de Recherche, Mines Paris-PSL	*Directeur de thèse*
Sophie, DEMASSEY Maître-Assistante, Mines Paris-PSL	*Directrice de thèse*

MINES PARIS | PSL

Abstract

The expansion of renewable energy sources (RES) leads to a growth of uncertainty in the power distribution network operation. The inherent variability and intermittency of RES present significant challenges to the efficient and reliable operation of power systems. To address these challenges, operational planning performed by distribution system operators should evolve, in particular, to allow the efficient utilization of different flexibility mechanisms (so-called levers), such as active power modulation and reactive power management. Decisions on lever activation are based on the resolution of an alternating current optimal power flow problem (AC-OPF). This thesis develops algorithms for handling two stochastic AC-OPF models. These optimization problems are simultaneously nonconvex, nonsmooth, and discrete. The thesis aims to grasp these complexities accurately, by addressing the AC power flow equations without relying on convexification and by handling interdependent uncertainties either through a joint probability constraint or via scenario decomposition to cope with the discrete levers.

More specifically, the first proposed methodology addresses a continuous version of the joint chance-constrained AC-OPF. A first contribution of this work is the design of a numerical procedure (oracle) that enables the representation of the probability constraint as a difference of two convex functions. This step is followed by applying a known Difference-of-Convex (DoC) bundle method to the resulting continuous optimization problem. A second contribution concerns a new bundle algorithm with stronger convergence guarantees under weaker assumptions. For the chance-constrained AC-OPF, this algorithm provides a critical (generalized KKT) point. The work builds upon the employed DoC bundle method and proposes a different master program and an original rule to update proximal parameter. The algorithm is capable of handling a broad class of nonsmooth and nonconvex optimization problems beyond the stochastic AC-OPF framework, provided the objective and constraint functions can be represented as differences of convex and weakly convex (CwC) functions. The practical performance of the algorithm is illustrated through numerical experiments on some nonconvex stochastic problems and is compared to the DoC bundle method for the chance-constrained AC-OPF in a 33-bus distribution network.

The second proposed methodology addresses operational planning rules for power modulation and curtailment, like priority and fairness, which result in logical and discrete formulations. The numerical results demonstrate the limitations of the bundle method for integrating integer variables. As an alternative, an optimization model is proposed that assigns a binary variable to each scenario and maximizes the number of satisfied scenarios within a limited budget. Applying penalization and block coordination allows separating those discrete considerations from the stochastic AC-OPF component, which is then decomposed into an individual deterministic AC-OPF for each scenario. Although it lacks theoretical convergence guarantees, the relevance of this approach is validated in practice.

Résumé

L'expansion des sources d'énergie renouvelable accroît le degré d'incertitude dans l'exploitation des réseaux de distribution d'électricité. La variabilité et l'intermittence inhérentes à ces énergies posent aussi d'importants défis aux gestionnaires de réseaux au niveau opérationnel. La gestion prévisionnelle doit ainsi évoluer pour intégrer des leviers de flexibilité, telles la modulation de puissance active et la gestion de puissance réactive. La décision relative à l'activation de ces leviers se traduit par un problème d'Optimal Power Flow. Cette thèse développe des algorithmes de résolution pour deux modèles stochastiques en courant alternatif (AC-OPF). Ces problèmes d'optimisation sont, à la fois, non-convexes, non-lisses et discrets. Cette thèse vise à appréhender ces complexités, sans recourir à la convexification des équations de flux de puissance, et en considérant l'interdépendance des incertitudes, via une contrainte probabiliste jointe ou une décomposition par scénarios dans le cas de leviers discrets.

Précisément, la première méthodologie proposée s'applique à une version continue de l'AC-OPF sous contrainte probabiliste jointe. Une contribution de ce travail porte sur la conception d'une procédure numérique (oracle) traitant la contrainte probabiliste comme la différence de deux fonctions convexes. L'oracle est alors associé à une méthode de faisceaux pour les problèmes DoC (différence de convexes). Une seconde contribution porte sur le développement d'un nouvel algorithme de faisceaux offrant des garanties de convergence plus fortes sous des hypothèses plus faibles. Il produit ainsi un point critique (satisfaisant des conditions KKT généralisées) de l'AC-OPF probabiliste. Basé sur la méthode DoC précédente, cet algorithme exploite un programme maître différent, ainsi qu'une règle originale de mise à jour du paramètre proximal. Il s'applique à la classe générale des problèmes d'optimisation non-convexes et non-lisses dont objectif et contraintes sont modélisables comme différence de fonctions convexes et faiblement convexes (CwC). L'évaluation empirique de l'algorithme est menée sur différents problèmes non-convexes et stochastiques. Ses performances pratiques sont comparées à celles de la méthode DoC sur un cas d'étude de l'AC-OPF probabiliste dans un réseau de distribution à 33 nœuds.

La seconde méthodologie proposée considère des règles discrètes en gestion prévisionnelle, telles que des règles de priorité et d'équité pour la modulation de puissance.

L'expérimentation montre les limites de la méthode des faisceaux pour intégrer des variables entières. Comme alternative, il est proposé un modèle d'optimisation attachant une variable binaire par scénario, et maximisant le nombre de scénarios réalisés dans un budget limité. La dualisation des contraintes couplantes et la coordination par blocs permettent de séparer les règles discrètes de l'AC-OPF stochastique, qui se décompose, à son tour, en AC-OPF déterministes individuels par scénario. Si la convergence théorique n'est plus garantie par cette séparation, la pertinence pratique de l'approche est illustrée numériquement.

Acknowledgements

As I write these words some time after my defense, I find myself looking back on this journey with deep gratitude and reflection. This thesis represents far more than an academic milestone — it has been an important chapter of my life, shaping me both as a scientist and as a person.

I am incredibly grateful to my supervisors, Welington de Oliveira and Sophie Demassey, who guided me through this journey with their intellectual rigor and unwavering support. Sophie, thank you for our countless discussions — about science, life, and everything in between. They gave me a true appreciation for research and the academic lifestyle. Welington, your infectious enthusiasm kept me motivated even during challenging moments. Thank you for your persistence and for believing in me. To both of you, I am profoundly grateful — for your patience, for the discussions I miss dearly, and for your faith in my work. Your mentorship has been invaluable.

I would like to express my sincere gratitude to the members of my thesis jury: Prof. Claudia Sagastizábal and Dr. Line Roald for their thorough review of my manuscript and deeply helpful feedback; Prof. Hugo Morais and Dr. Wim van Ackooij for their insightful contributions during the defense and research collaboration; and Prof. Sourour Elloumi for her valuable perspectives and support throughout this process. The engaging discussion during my defense and the generous encouragement I received about my work remain both an honor and a lasting source of motivation.

I am deeply thankful to my tutors at EDF R&D: Bhargav Swaminathan for initiating this thesis project, Aurel Garry for his careful review of my manuscript, and Nicolas Cussigh for his administrative support and personal encouragement. I also wish to thank EDF R&D for their financial support and for providing me with this exceptional professional opportunity.

My heartfelt thanks go to the entire CMA team: senior researchers, administrative staff, Alice and Amel, and Sébastien for technical support. A special thanks to my fellow PhD students and post-docs who made this time so memorable, and Maju for cheering up the entire lab. I also cherish the time I spent at EDF R&D with my colleagues from the Système and Osiris departments, sharing countless cozy coffee breaks and lunches.

My deepest gratitude extends to my former colleagues, professors, and teachers, without whom this journey would not have been possible. First and foremost, Riadh Zorgati and

Clémence Alasseur, my Master's internship supervisors, whose guidance played a pivotal role in shaping my professional trajectory. I am profoundly grateful to Dr. Aleksei Pirkovskii, my Bachelor's thesis supervisor, for his tireless dedication and mentorship — his functional analysis course and exceptional teaching unveiled the profound beauty of mathematics to me. Finally, I owe special thanks to Dr. Petr Sergeev and Lev Davydovich Altshuler, my extraordinary teachers at School 57, who instilled in us not only the joy of learning but also the inspiration to pursue a research path.

Above all, I recognize my greatest supporters: my dear friends; my parents, Alexander and Marina; my grandmother, Larisa; and my boyfriend, Ilia — for standing by me with endless encouragement and inspiration.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

This chapter discusses the paradigm shift for distribution system operators (DSOs) driven by the large-scale integration of renewable energy sources (RES) and advancements in grid technologies. We review regulatory and contractual incentives such as Feed-in Tariffs (FiT) and Smart Connection Points (SCP) related to RES adoption, along with the operational tools DSOs employ for effective grid management. Special emphasis is placed on short-term operational planning, particularly active power modulation and curtailment control levers — the core focus of this thesis.

The mathematical framework explores optimization methods for DSO operational planning, with a focus on stochastic optimization to address uncertainties in RES generation. The discussion highlights the shift from deterministic approaches to chance-constrained models, which ensure reliable grid operation under probabilistic constraints. These models lead to nonconvex and nonsmooth optimization problems, presenting significant theoretical and computational challenges.

The chapter then outlines the key contributions of this thesis, including theoretical and numerical results, followed by its organizational structure. It concludes with an overview of peer-reviewed publications and conference presentations derived from this research.

Ce chapitre aborde le changement de paradigme des gestionnaires de réseaux de distribution (GRD), induit par l'intégration massive des énergies renouvelables (ENR) et les avancées technologiques dans les réseaux électriques. Nous évoquons les dispositifs réglementaires et contractuels liés au déploiement des ENR, notamment les offres de raccordement de référence (ORR) et les offres de raccordement intelligentes (ORI), ainsi que les outils opérationnels déployés par les GRD pour la gestion du réseau. Un accent particulier est mis sur la gestion prévisionnelle à court terme, plus spécifiquement sur les leviers de modulation et de réduction de la puissance active, qui constituent l'axe central de cette thèse.

Les méthodes d'optimisation pour la gestion prévisionnelle sont ensuite explorées, avec un focus particulier sur l'optimisation stochastique pour tenir compte des incertitudes liées à la production des ENR. La discussion souligne la transition des approches déterministes vers des modèles à contraintes probabilistes, qui garantissent un fonctionnement fiable du réseau dans un cadre stochastique. Ces modèles conduisent à des problèmes d'optimisation non-convexes et non-lisses, posant d'importants défis théoriques et numériques.

Le chapitre présente ensuite les principales contributions de cette thèse, incluant les résultats théoriques et numériques, puis sa structure organisationnelle. Il se termine par un aperçu des publications scientifiques et des présentations en conférence issues de cette recherche.

1.1 Industrial context

The use of renewable energy sources (RES) is becoming increasingly important as the world seeks to transition to more sustainable and environmentally friendly energy systems. Many countries are setting new targets to achieve carbon neutrality. For example, in Europe, several programs and initiatives have been introduced, including *Green Deal*, *Fit to 55*, or *RePower EU*. These initiatives outline progressively ambitious goals, supported by concrete measures to promote, in particular, the development of renewable energies and the electrification of the transport sector [15]. A closer analysis of these initiatives highlights the key role of transmission and distribution power systems as strategic infrastructures in achieving those objectives.

Variable and intermittent by nature, RES such as solar and wind power pose significant challenges to the efficient and reliable operation of power systems. To address these challenges, system operators must adopt a more proactive approach to network planning and operation [39]. This shift primarily concerns distribution system operators (DSOs), as a traditional *fit-and-forget* approach has historically been used for distribution grid management. However, this concept would require tremendous investments in grid expansion to meet the demand of modern distribution systems. The new approach involves optimizing multi-annual investments in human resources, infrastructure, and technology, as well as adopting new solution methods and operating tools.

In this context, many European DSOs are implementing an integrated approach that aligns network planning, operational planning, and real-time operations over a continuous time horizon, enabling faster and more efficient integration of new distributed energy resources (DER). This approach is highlighted, for example, in the multi-annual network development plan of Enedis [36], the largest DSO in France. Typically, this methodology involves

characterizing major investments several years in advance, identifying and prioritizing medium-term investments, optimizing work schedules in the short term, and operating networks in an efficient and coordinated manner.

Although a DSO is responsible for the operation of the distribution network, its activities may be subject to unbundling rules. For instance, in Europe, DSOs are prohibited from owning or operating generating units. Consequently, their observability and level of control are limited, which in practice increases uncertainty in power generation and consumption and makes it more challenging for the DSO to effectively manage the grid. In this regard, the mechanisms used to ensure the network operation can be informally divided into two groups. The first group includes the techniques implemented to motivate the grid users, both power producers and consumers, to adapt their behavior in a specific way. As these incentives include different agents of the distribution system, they are often developed by governments and regulators, as part of broader political decisions. One such tariff-based incentive will be discussed in the next section. The second group consists of mechanisms, often referred to as "levers", that can be directly implemented and activated by DSOs for the network operational planning. The modeling and optimization of these mechanisms are the focus of this thesis. We will introduce the concept of levers, along with the DSO's objectives and main challenges in this context, later in the introduction.

1.1.1 Feed-in tariffs and smart connection points

Feed-in tariffs (FiT) for renewable generation were among the first incentive measures implemented by governments to encourage the development of RES [155]. According to the [40, 155], FiT tariffs can be defined as "the price per unit of electricity that a utility or supplier has to pay for renewable electricity from private generators". These tariffs allowed RES production units to inject all the energy they produced while receiving a fixed and economically attractive remuneration.

In this context, system operators had to prepare to accommodate the contractualized production capacity, which required significant investments in grid infrastructure to handle peak active power production from RES. Depending on the circumstances, this resulted in substantial costs for these producers, delays in commissioning, or even temporary limitations on the power injection during reinforcement works.

Nowadays, considering the level of maturity of the RES technologies — particularly wind and solar photovoltaic — FiTs are increasingly being replaced by alternative mechanisms [77]. Among these alternatives are smart and interruptible connection points (SCPs) and contracts, which have been proposed to mitigate the need for grid reinforcement [5] and are now being implemented across Europe [45]. Under such contracts, the DSO is permitted to temporarily

reduce the contractual consumption or generation capacity at a connection point of DERs. This approach increases the hosting capacity of existing systems, enabling faster integration of RES while reducing investment costs.

1.1.2 Operational planning

The operational planning for a DSO can be divided into long-term planning and short-term, in accordance with the objectives that need to be met. The former focuses on long-term goals such as planning construction and maintenance works, whereas the latter addresses operational constraints to ensure effective grid access for users while respecting legal, contractual, and technical requirements. Short-term operational planning typically spans a time scale ranging from 30 minutes to one month. Although these two aspects of planning are closely interconnected — long-term decisions influence short-term actions, and short-term decisions must align with long-term goals — our focus will be on short-term operational planning for the medium voltage (MV) network and its framework.

Short-term operational planning involves anticipating the network operational constraints (we will refer to these constraints as technical constraints or simply grid constraints) and addressing them using various control levers [91]. The first step, the detection of potential operational constraints, is based on forecasts of power production and consumption by grid users. Once constraints are identified, a decision is made regarding which control levers to activate from the available set to resolve the predicted issues. These options include, among others, active power modulation and curtailment, reactive power regulation, and network topology reconfiguration [121]. While this work focuses on modeling active power modulation and curtailment, the proposed methodology is not limited to these specific levers and can be adapted to incorporate models for other control actions.

The rules governing active power modulation and curtailment — essentially distinguished by the direction of power change — depend on the type of grid connection contract. Production facilities connected via FiT contracts are authorized to inject any time up to the maximum active power requested during the connection. Consequently, the DSO has no contractual authority to modulate or curtail power from FiT-connected users under normal operating conditions. Curtailment of such units is only permitted under critical operating conditions and must be financially compensated by the DSO. In contrast, SCP contracts guarantee each producer a minimum level of power injection into the grid. Within the bounds defined by these contracts, the DSO can perform power modulation at a relatively low cost. However, any curtailment beyond those limits is treated as a denial of service, similar to FiT arrangements. As for consumers, any curtailment of their power usage is treated as a forced

outage. For modeling purposes, this effectively places consumers in the same category as FiT grid users.

1.1.3 Operational planning methods

The decision on lever activation is taken in practice using operational planning tools specifically tailored to these needs. These tools have already been deployed and are continuously being improved to address emerging challenges, such as those highlighted above and described in [36]. These improvements include mathematical optimization-based approaches to solving problems, which represent a complete shift in the operational planning paradigm for DSOs. By integrating practical rules, use-case scenarios, and contractual obligations into mathematical models, this approach leads to a more systematic and efficient decision-making process. Ongoing efforts focus on deploying advanced methods capable of handling additional challenges and opportunities, such as uncertainty in DER production, interaction with external actors (e.g., flexibility providers and TSOs), and integrating these tools into IT and OT systems.

The employed optimization methods must handle power flow equations to ensure technical constraints are satisfied. Combined with the objective of minimizing costs associated with the operational solution, this results in an alternating current optimal power flow (AC-OPF) problem. AC-OPF is, in general, an optimization problem with continuous variables that is nonconvex and strongly NP-hard, as shown in various studies [12]. Several simplified formulations and solution methods have been proposed [43, 101], including classical optimization methods that rely mainly on convex relaxations or approximations [82, 48], nondeterministic search techniques (also known as heuristic, stochastic or random search methods) [44], and machine learning methods [57]. In an industrial optimization approach [122], nonconvex relationships are approximated by piecewise linear functions, while logical and discrete formulations representing operational planning rules are introduced. This results in a mixed-integer linear programming model (MILP) [122].

Traditional OPF models assume the system parameters are deterministic, which is hardly compatible with the integration of RES. Therefore, modeling uncertainties has become an important research topic in the field of power systems in the context of the energy transition. The next section explains the emerging challenges and provides an overview of approaches developed to address the resulting optimization models.

1.2 Mathematical context

1.2.1 OPF under uncertainties

Solution methodologies to address OPF models with uncertainties are diverse and include meta-heuristic methods, as well as robust and stochastic approaches. Meta-heuristic methods, in particular, have gained increasing attention in the field of power and energy systems within the last ten years [19], due to their ability to search for implementable decisions with relatively low computational cost. An overview of such methods can be found in [96]. However, the quality of the obtained decisions can be difficult to assess in practice, as no optimality certificate is provided from a theoretical point of view.

Robust optimization methods rely on constructing an uncertainty set and searching for a reliable solution for any scenario realization in that set. This approach is proposed for the distribution network management in [65, 116, 150] with different types of uncertain data. For instance, uncertainties are solely on renewable energy sources in [65], whereas [150] deals with stochastic load composition. The authors in [116] describe multi-period grid management applying a convex hull technique to define an uncertainty set. Solutions obtained with robust optimization methods are optimal for the worst-case scenario and thus tend to be conservative. In contrast, stochastic optimization methods assume that the probability distribution of uncertain variables is known [114, Chapter 1] and, in general, enable obtaining a less conservative decision.

One widespread approach in stochastic programming is chance-constrained optimization, where one searches for a decision that minimizes costs while satisfying a set of random constraints with a prescribed probability level (also known as security level) [114, 6]. Chance-constrained optimization models are intuitive and straightforward to explain. However, they are generally difficult to solve, because they often lack essential mathematical properties such as convexity and differentiability [126]. The use of chance constraints for energy management is discussed in detail in [142], where the authors show how to ensure feasibility to the greatest extent possible while aligning with the goals of the system operator. This approach is widely used for different energy-related problems, including hydro-reservoir and gas network management [141, 54, 52]. In this thesis, we focus on chance-constrained formulations for operational planning models in power grid management.

In the framework of OPF problems, most chance-constrained models rely on convex approximations or linearizations of power flow equations and deal with individual probability constraints, i.e., correlations between random variables are ignored. In [154] an OPF problem with load uncertainties is considered, where probability constraints are represented with individual bounds on state variables. The model equations are linearized at the random

vector expected value, establishing a monotonic relation between the output and a random input. This enables the treatment of probability constraints with respect to the distribution of random variables. An analytical reformulation of chance constraints for the direct current optimal power flow (DC-OPF) model with Gaussian distributions is proposed in [103]. This approach is generalized in [104] to accommodate a broader class of distributions with a given mean vector and covariance matrix. A partial linearization of AC power flow equations is applied in [102] to transform the individual chance constraints into deterministic ones at the forecasted operating point. This transformation assumes relatively small forecast errors in model uncertainties. Another approximation is introduced in [25], where chance constraints are used to enforce voltage regulation with prescribed probability. The authors address chance constraints with a convex relaxation closely related to the concept of conditional value at risk (CVaR) and exploit linearizations of power flow equations.

Reference [11] deals with minimizing an average generation cost over random renewable power injections, while controllable generators mitigate power fluctuations. Chance constraints are imposed separately for each type of operational limit and reformulated as second-order cone constraints. The model is thus reduced to a deterministic convex optimization problem, more precisely, a second-order cone program (SOCP), and resolved with a cutting-plane algorithm. A SOCP reformulation is also used for the case of individual chance constraints in [94, 10]. Paper [84] presents a robust modification of the approach given in [11], which addresses the uncertainty in parameters of probability distributions by restricting them to an uncertainty set.

1.2.2 Joint chance-constrained models

To address the strong correlation between the renewable generation profiles and loads, a joint chance-constrained formulation is used in [144, 100]. Despite recent advances in theory and numerical methods for chance-constrained problems, dealing with multivariate probability functions in optimization problems generally remains a challenging task. This amounts to computing numerically a multidimensional integral, a difficult task for even moderate dimensions [127]. The situation becomes even more demanding for assessing (sub)gradients. For this reason, a common practice is to estimate the probability function with scenario-based methods, such as Monte-Carlo simulations using a finite sample of scenarios [97]. In the latter work, the authors analyze the convergence properties of the sample average approximation when the sample number goes to infinity and show how to construct good approximate solutions.

To mitigate the complexity of joint chance constraints, a decomposition into a series of individual chance constraints is commonly used in the OPF literature. This approach is

applied in [8], where the individual chance constraints are then reformulated analytically. A key challenge of this decomposition is selecting appropriate security levels for the individual chance constraints. In [68] this issue is addressed by employing an iterative decomposition process that provides a less conservative solution. However, even with the optimal selection of individual security levels, this approach cannot guarantee an optimal solution to the original joint chance-constrained problem [18].

Another group of references addresses the joint chance constraint without decomposition. These approaches rely on two key steps: transforming the underlying AC-OPF into a convex model and converting the joint chance constraint into a deterministic form. The latter step often relies on sample-based approaches, which do not require particular assumptions on the distribution. For instance, a method based on empirical quantiles is applied to a DC-OPF model in [99] giving rise to a continuous nonlinear programming (NLP) model. Another scenario-based technique is applied in [146, 143], which guarantees satisfaction of the original constraint in probabilistic sense. This technique requires convexity of the underlying problem, and these works exploit semidefinite programming (SDP) relaxations for AC-OPF models. In [143] an alternative approach involving an analytical reformulation of the probability constraint is proposed for the case of a multivariate Gaussian distribution.

1.2.3 Chance-constrained OPF with discrete variables

Incorporating discrete variables and logical formulations significantly enhances the ability to model real-world decision-making processes. They are widely employed in operational planning problems such as goods distribution, production scheduling, and machine sequencing, where decisions are naturally binary (yes/no, on/off) or involve selecting among discrete options [95]. Meanwhile, such variables and formulations introduce a fundamentally different and inherently combinatorial challenge. In contrast to continuous optimization, where the primary concern is often how to find the optimum, discrete problems could, in principle, be solved by enumerating all possible configurations. However, as the problem size increases, this quickly becomes a computationally challenging task. As a result, the key question shifts to: how can these problems be solved efficiently? In addition to this change in perspective, theoretical framework and solution methodologies differ: standard constraint qualification conditions fail at isolated feasible points, and concepts like gradients are not well-defined in the discrete setting.

These factors introduce additional complexity to chance-constrained OPF models. The literature on such models remains relatively limited, and existing approaches often rely on linear approximations of power flow equations, along with simplifications of the joint

chance-constrained framework. As a result, an overall formulation is typically reduced to a mixed-integer linear approximation to ensure tractability.

In [3], the authors address discrete curtailment of renewable generation in the context of transmission grid operations. They propose a chance-constrained DC-OPF model with two joint chance constraints that separately incorporate the limits on power generation outputs and transmission line flows. Discrete curtailment decisions are modeled through discrete variables included in the deterministic part of the formulation. To handle the computational complexity of the chance constraints, the authors employ a robust approximation technique and a sample-based approach to construct suitable uncertainty sets. These methods lead to a MILP approximation of the original problem.

Another approach to the chance-constrained operational framework is proposed in [69]. The authors seek to identify minimum-cost operational planning decisions such that the system can be operated while satisfying a certain probabilistic security criterion. Adopting DC approximation of power flow equations, a deterministic MILP model representing the planning process for a finite number of scenarios is constructed. The employed two-step algorithm iteratively refines the scenario set to balance cost and reliability. In the first step, the scenario set is enlarged, and an associated operational planning cost is calculated by solving the MILP. In the second step, inappropriate scenarios are eliminated based on feasibility and cost considerations.

Papers [151, 115] deal with the two-stage framework, which additionally models either a proactive role of prosumers [151] or a preliminary step of contracting flexibilities in a day-ahead market by DSO [115]. The chance-constrained operational planning aspect remains integral to one of the stages. Both papers employ linearizations of AC power flow equations but address the chance constraints differently. In [151], individual chance constraints are reformulated as second-order cone constraints, which results in a mixed-integer second-order cone programming (MISOCP) problem. In [115], the authors adopt a scenario-based approach to a joint chance constraint and compare two reformulation techniques, Big-M method and McCormick relaxations, both leading to MILP models.

1.3 Contributions and organization of the thesis

This thesis focuses on joint-chance constrained AC-OPF models to address short-term operational planning for medium voltage (MV) distribution networks, specifically, power curtailment and modulation levers. From an optimization perspective, the resulting problems are simultaneously nonconvex, nonsmooth, and discrete, making them particularly challenging. Some level of simplification is unavoidable when tackling these problems.

The main paradigm of this thesis, which sets the proposed approaches apart from most existing methods discussed in the previous section, is the commitment to address these complexities without preliminary relaxations or approximations. Rather than simplifying the original problem upfront, which often results in an accumulation of errors from approximations, the underlying complexities are decomposed. Some approximations, when necessary, are introduced later in the process.

In Chapter 2, we consider a continuous version of the joint chance-constrained AC-OPF, addressing two complexities of the model, nonconvexity and nonsmoothness. This model, introduced in [67], incorporates a simplified version of power curtailment and modulation rules, avoiding logical variables and discrete formulations. The solution approach is structured into three steps, and our contribution includes the first of them. We propose a parallelizable numerical procedure (oracle) enabling decomposing the underlying formulation in probability constraint into a deterministic AC-OPF per scenario. Using this procedure and applying Monte-Carlo simulations, the chance constraint is approximated by a difference of two convex functions. This step contains all the approximations used in the method. A known algorithm, the DoC bundle method [129], originally designed to handle optimization problems with the objective function and constraints being differences of two convex functions (DoC), is then applied to the resulting model. However, since this model does not meet several favorable assumptions, the algorithm is not guaranteed to provide a critical point to the original model. Therefore, two challenges arise: finding a critical point to the resulting continuous model and incorporating discrete variables to represent operational planning rules.

To address the first challenge, we propose a new bundle algorithm in Chapter 3, with stronger convergence results under weaker assumptions. A detailed convergence analysis and a comparison with the DoC bundle method are also included. This algorithm applies to a broad class of nonsmooth and nonconvex optimization problems, provided the objective and constraint functions can be represented as differences of convex and weakly convex functions (CwC). As a weakly convex function can, in theory, be decomposed itself as a difference of two convex functions, the class of optimization problems remains unchanged. However, in practice, the CwC structure often arises more naturally in certain applications, whereas a DoC decomposition of the associated problems may not be available. This explains the practical interest of the new method, going beyond the chance-constrained AC-OPF framework. In contrast to the DoC bundle method, the new algorithm provides a critical (generalized KKT) point in the latter framework.

Next, in Chapter 4, we propose a model for operational planning rules that address priority and fairness in power curtailment, resulting in additional logical and discrete deterministic constraints. These rules are based on the notion of priority levels — a concept that generalizes

FiT and SCP contracts by explicitly defining the order in which power curtailment is applied to producers. The new model thus incorporates binary variables along with the joint chance constraint and the underlying AC-OPF, resulting in the three announced complexities. We begin Chapter 5 by testing the ability of the DoC bundle method to handle these additional constraints, which does not yield reliable numerical results. As an alternative approach, we assign a binary variable to each scenario and transition to a new model that maximizes the number of satisfied scenarios under a budget constraint, partly employing the methodology proposed in [2]. This transition allows for separating discrete formulations and breaking down the stochastic component into a deterministic AC-OPF per scenario, enabling the use of existing solvers and methods to address these components independently.

Finally, all the proposed methods for the joint chance constrained AC-OPF models are tested in a 33-bus distribution network. The practical performance of the new bundle method is illustrated on several nonconvex stochastic problems and compared to the DoC bundle method for the chance-constrained AC-OPF on the same network. The numerical results are presented in the corresponding chapters.

Since the proposed methodology provides a deterministic AC-OPF per scenario as one of the final outputs that are addressed with existing methods, we investigate an SDP relaxation of the resulting AC-OPF in Chapter 6. While the discussed methods can readily incorporate this relaxation, further considerations and numerical testing are required. Additionally, several perspectives for alternative approaches to the new model are discussed, along with the limitations of the presented approach and its potential application to other operational planning models.

Articles in peer-reviewed journals and presentations in conferences. The main content of Chapter 2 appeared in [123], and the content of Chapter 3 was published in [124]. The following presentations were delivered based on the material presented in this Thesis:

- "Scenario-Based Decomposition for Optimal Power Curtailment with Priority-Level Producers" by K. Syrtseva, W. de Oliveira, S. Demassey:
  - PGMO Days, *Palaiseau, France*, November 2024.
- "Joint Chance Constraint and Difference-of-Convex Approach for Operational Planning of Distribution Network" by K. Syrtseva, W. de Oliveira, S. Demassey:
  - Optimisation et Energie sous contraintes climatiques (journée ROD-RADIA), *Lyon, France*, November 2024 (*in French*).
  - Électricité de France, *Palaiseau, France*, November 2024 (*in French*).

  - Energy System Modelling for Transition to a net-Zero 2050 for EU via REPowerEU (work group Man0EUvRE), online, November 2024.

- "Difference-of-Convex Approach to Chance-Constrained OPF" by K. Syrtseva, W. de Oliveira, S. Demassey, H. Morais, P. Javal, and B. Swaminathan:

  - ROADEF, *Amiens, France*, March 2024 (*in French*).
  - Mines Paris, Centre de Mathématiques Appliquées, *Sophia Antipolis, France*, December 2023.
  - PGMO Days, *Palaiseau, France*, November 2023.
  - Électricité de France, *Palaiseau, France*, September 2023 (*in French*).

- "Minimizing the Difference-of-Convex and Weakly Convex functions via Bundle method" by K. Syrtseva, W. de Oliveira, S. Demassey, and W. van Ackooij:

  - SIAM Conference on Optimization, *Seattle, USA*, June 2023.
  - PGMO Days, *Palaiseau, France*, November 2022.
  - Électricité de France, *Palaiseau, France*, September 2022 (*in French*).

- "Optimization Approaches to Stochastic Problems":

  - Mines Paris, Centre de Mathématiques Appliquées, *Sophia Antipolis, France*, January 2024 (for students of the Master program of Advanced Studies in Energy Systems Optimization).

Chapter 2

Continuous Model and Difference-of-Convex Approach

In this chapter, we consider an operational planning problem that includes the activation of power modulation and power curtailment levers modeled as a joint chance-constrained AC-OPF. For now, we adhere to the framework of levers proposed in [67], where the prioritization between SCP and FiT users is addressed through differences in power modulation costs (without discrete and logical conditions). The solution method involves adopting a Difference-of-Convex (DoC) [1] approach, which enables tackling the resulting optimization problem without convexification or linearization of the core OPF equations. Furthermore, the approach yields a natural and parallelizable scenario decomposition. The method starts with a reformulation of the employed model as a DoC optimization problem, which is the main novelty of the present chapter. Next, a proximal bundle method algorithm developed in [129] is applied to solve it.

The main content of this chapter has appeared in [123] (K. Syrtseva, W. de Oliveira, S. Demassey, H. Morais, P. Javal, and B. Swaminathan (2023). "Difference-of-Convex approach to chance-constrained Optimal Power Flow modelling the DSO power modulation lever for distribution networks". *Sustainable Energy, Grids and Networks*).

Ce chapitre modélise un problème de gestion prévisionnelle intégrant l'activation de leviers de flexibilité, notamment la modulation et la réduction de la puissance active. Ce problème est modélisé comme un AC-OPF (Optimal Power Flow en courant alternatif) sous contrainte probabiliste jointe. Pour l'instant, nous adoptons le cadre des leviers de [67], où la priorisation entre les utilisateurs ORI et ORR repose sur des différences de

[1] The standard abbreviation for Difference-of-Convex is DC. To avoid confusion with "Direct Current" in OPF problems, we refer to the former as DoC.

coûts de modulation de puissance (sans conditions discrètes ou logiques). La méthode de résolution consiste à adopter une méthode de faisceaux pour les problèmes DoC (différence de convexes). Cette méthode permet de résoudre le problème d'optimisation sans recourir ni à une convexification ni à une linéarisation des équations de flux de puissance. De plus, elle induit naturellement une décomposition par scénarios parallélisable. Le modèle est d'abord reformulé sous la forme d'un problème d'optimisation DoC, ce qui constitue la contribution principale de ce chapitre. Puis, un algorithme de méthode des faisceaux, développé dans [129], est appliqué pour résoudre le problème d'optimisation obtenu.

Le contenu principal de ce chapitre a été publié dans [123] (K. Syrtseva, W. de Oliveira, S. Demassey, H. Morais, P. Javal, et B. Swaminathan (2023). "Difference-of-Convex approach to chance-constrained Optimal Power Flow modelling the DSO power modulation lever for distribution networks". Sustainable Energy, Grids and Networks).

2.1 Operational planning under uncertainties: framework and contributions

2.1.1 Optimal Power Flow (OPF) under uncertainties

In this chapter, we model an operational planning problem as a joint chance-constrained optimization problem. For a given parameter $\alpha \in [0,1]$ and a probability measure $\mathbb{P}$, the problem we investigate in this work can be synthetized as follows:

$$\min_{\text{Levers}} \quad \{ \text{ Levers activation cost } \} \tag{2.1.1a}$$

$$\text{s.t.} \quad \text{Activated levers satisfy contractual constraints} \tag{2.1.1b}$$

$$\mathbb{P}\left[\begin{array}{c}\text{Existence of a grid state within bounds}\\ \text{satisfying stochastic power-flow equations}\end{array}\right] \geq 1-\alpha. \tag{2.1.1c}$$

In this formulation, the decision variables are related to the activation of flexibility contracts (levers), which in our problem corresponds to DSO decisions on power modulation and power curtailment of distributed generation, and limitations on power supply to consumers (energy not supplied). Each grid user, who may be either a producer or a consumer, is characterized by its grid connection contract (FiT or SCP in our case). Depending on the contracts, levers activated by the DSO must satisfy specific deterministic constraints.

Stochastic equations are related to technical decision feasibility: given a DSO decision and a scenario realization on power generation and loads, grid operating conditions must

remain within technical limits. The literature varies in the types of these limits: these bounds may apply, for instance, to generation output, line flows, voltage magnitudes, transmission line capacities, or other parameters. We will precise the technical limits in Subsection 2.2.2. Similar framework is addressed, for instance, in [99, 146] and [143].

In practice, a decision on levers activation satisfying all the bounds and stochastic power flow equations for all the possible scenarios may not exist. Hence, it makes sense to search for a reliable, affordable decision that satisfies the stochastic constraints with a probability level of at least $1-\alpha$. Observe that, if we take $\alpha = 0$ and a solution of problem (2.1.1) exists, then such a solution is robust over all the possible scenarios. The decision maker can thus adjust her risk aversion by appropriately setting the parameter α: small α reflects high risk aversion.

In the proposed approach, the primary goal is to secure the required level $1-\alpha$, with $\alpha > 0$, for the system: we search for an implementable decision that is feasible with a probability of at least $1-\alpha$. As we will see shortly, the goal is obtained without any stringent assumption on the probability distribution or modeling simplifications such as linearization and convexification of core OPF equations. The deterministic case $\alpha = 0$ is not within the scope of the proposed methodology, which is applicable but not tailored for this case. In Section 2.2, we introduce a realistic mathematical model for (2.1.1) that is consequently nonconvex, nondifferentiable, and thus challenging.

We adhere to the principle to keep the model complexity, including AC-OPF and joint chance constraint, without preliminary simplifications. As discussed in Chapter 1, new mathematical approaches are required to face nonconvexity and nondifferentiability in problem (2.1.1). Based on the observation that requirement under the probability in (2.1.1c) can be written as a difference of two convex functions, we apply a Difference-of-Convex (DoC)[2] optimization approach to tackle the problem.

2.1.2 Chance-constrained optimization and the DoC approach

A function is called DoC if it is expressible as the difference of two convex functions. As already investigated in [63] and [129], probability functions can be approximated as accurately as one wishes by DoC functions. In Subsection 2.3.1, it is shown how to model the existence requirement in (2.1.1c) by a DoC function. Therefore, we end up with a composition of DoC functions that is itself DoC, and thus we can approximate probability

[2]The standard abbreviation for Difference-of-Convex is DC. To avoid confusion with "Direct Current" in OPF problems, we refer to the former as DoC.

constraint (2.1.1c) with a DoC constraint:

$$\mathbb{P}\left[\begin{array}{c}\text{Existence of a grid state within bounds}\\ \text{satisfying stochastic power-flow equations}\end{array}\right] \geq 1-\alpha$$
$$\approx \quad c_1(x) - c_2(x) \leq 0,$$

where $c_1, c_2 : \mathbb{R}^n \to \mathbb{R}$ are (nondifferentiable) convex functions. Given this approximation, we will show how to construct a DoC optimization model for problem (2.1.1) fitting into the following structure:

$$\min_{x \in X} f(x) \tag{2.1.2a}$$
$$\text{s.t.} \;\; c(x) := c_1(x) - c_2(x) \leq 0, \tag{2.1.2b}$$

where X is a nonempty bounded polyhedron contained in $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function as well. In the present work, we will refer to (2.1.2) as a DoC optimization problem.

DoC programming forms an important subfield of nonconvex programming, as it covers a large class of nonconvex optimization problems from real-life applications. At the same time, convex analysis apparatus enables to establish optimality conditions for DoC problems and design algorithms to solve them. These facts explain the increasing interest in this field, which started in the 80s [74]. Important facts about DoC programming such as optimality conditions and duality can be found in the tutorial paper [28]. The survey article [74] gives a large spectrum of examples and algorithm developments in this field. Currently, most used algorithms are based on iterative linearizations of components f_2 and c_2 [74], on penalization technique applied to the DoC constraint [73], and on improvement functions that combine constraint and objective in a single level [129]. The latter is a well-known and successful strategy in the nonsmooth optimization literature [110, 7, 90, 129]. In particular, a bundle method with improvement function is proposed in [129] for dealing with DoC-constrained DoC-problems. Due to its good numerical performance reported in [129], we choose this bundle algorithm – denoted by DoC bundle method – to tackle our DoC optimization model (2.1.2) of the chance-constrained OPF problem (2.1.1).

2.1.3 Contributions and organization of the chapter

Our main contribution is the design of an *oracle* (black-box) that enables finding a DoC decomposition of the constraint under probability sign in (2.1.1c), which imposes the OPF solution to be in the required bounds. This fills a critical gap, allowing the DoC approach from [129] to be applied to a joint chance-constrained AC-OPF. Once future random events are

approximated through a sample of N scenarios, our oracle makes accessing the probability constraint parallelizable, while the DoC bundle method efficiently handles the resulting optimization problem. The method's performance is demonstrated on a 33-bus distribution network in two use cases.

The remainder of this chapter is organized as follows. Section 2.2 is dedicated to the modeling of short-term operational planning on grid users power modulation under uncertainties as a chance-constrained OPF. A reformulation of the obtained problem as a DoC model is given in Section 2.3, followed by an explanation of DoC bundle method. In Section 2.4, numerical results for two use cases are provided, while Section 2.5 discusses the approach's advantages and limitations. Finally, Section 2.6 closes the chapter with some concluding remarks.

2.2 Chance-constrained OPF model

In what follows the chance-constrained OPF model is described by setting the decision variables, state variables and constraints. The employed notation is introduced along the way.

2.2.1 Decision variables and random vector

We denote by $\mathscr{N}$ the set of buses of the grid, and by $\mathscr{A}$ the set of lines. The set $\mathscr{A}$ consists of pairs (i,j) with $i,j \in \mathscr{N}$ such that there is a conductor between the nodes i and j. Among the end buses, there is one slack bus $sb \in \mathscr{N}$ and other buses with at most one connected grid user. Each grid user is either a producer or a consumer. The set of all grid users is denoted by $\tilde{\mathscr{N}} \subset \mathscr{N}$, identifying a grid user $i \in \tilde{\mathscr{N}}$ with a corresponding bus $i \in \mathscr{N}$. Let $\mathscr{G}$ and $\mathscr{L}$ be the subsets of producers and consumers, respectively. We wish to determine the active power p_i (positive for generation and negative for consumption) and reactive power q_i of each grid user $i \in \tilde{\mathscr{N}}$, based on the following decomposition:

$$p_i = p_i^{\phi}(\xi) - \boldsymbol{p}_i^{\gamma} - \boldsymbol{p}_i^{\beta}, \quad q_i = q_i^{\phi}(\xi) - \boldsymbol{q}_i^{\gamma} - \boldsymbol{q}_i^{\beta}. \tag{2.2.1}$$

For buses $i \in \mathscr{N} \setminus \left(\tilde{\mathscr{N}} \cup \{sb\}\right)$ with no grid users, excluding the slack bus, we set $p_i = q_i = 0$. The active and reactive power demand (production and consumption), $p_i^{\phi}(\xi)$ and $q_i^{\phi}(\xi)$, are the *uncertain parameters* of our model, given as scenario realizations of a random vector $\xi \in \Xi \subset \mathbb{R}^d$. In case of considered flexibility contracts, the decision bears on $\boldsymbol{p}_i^{\gamma} \in \mathbb{R}$, the active power modulation for an SCP grid user $i \in \mathscr{G}$ within the bounds of its guaranteed power, and on $\boldsymbol{p}_i^{\beta} \in \mathbb{R}$, which denotes, either the active power curtailment beyond the bounds of guaranteed power of an SCP producer or the active power curtailment for a FiT producer,

or the energy not supplied to a consumer $i \in \mathscr{L}$. We introduce also pair variables for reactive power modulation: $\boldsymbol{q}_i^{\gamma}$ for SCP grid users within the bounds of their guaranteed power; $\boldsymbol{q}_i^{\beta}$ for FiT and SCP grid users beyond the bounds of their guaranteed power, and for consumers. Consumers are modeled as if they have a FiT contract. In fact, each reactive power variable is a function of the paired active power variable, which will be defined in Subsection 2.2.3. This formulation was chosen for its generality and flexibility, allowing for straightforward adaptation to cases involving other DSO control levers, such as reactive power regulation (in the specific case of active power modulation and curtailment, the reactive power variables can be eliminated from the model). As the decision must be taken before a scenario realization is known, the aforementioned decision variables do not depend directly on ξ. In order to distinguish different types of decision variables, we denote by $\mathscr{N}_{SCP} \subset \mathscr{G}$ the set of indexes for $\boldsymbol{p}^{\gamma}$ and $\boldsymbol{q}^{\gamma}$, and by $\mathscr{N}_{FiT} \subset \mathscr{L}$ that for $\boldsymbol{p}^{\beta}$ and $\boldsymbol{q}^{\beta}$. To simplify the notation, if there is no need to distinguish user types, we will denote decision variables by

$$\mathbf{p} := (\{\boldsymbol{p}_i^{\gamma}\}_{i\in\mathscr{N}_{SCP}}, \{\boldsymbol{p}_i^{\beta}\}_{i\in\mathscr{N}_{FiT}}) \quad \in \quad \mathbb{R}^n$$

$$\mathbf{q} := (\{\boldsymbol{q}_i^{\gamma}\}_{i\in\mathscr{N}_{SCP}}, \{\boldsymbol{q}_i^{\beta}\}_{i\in\mathscr{N}_{FiT}}) \quad \in \quad \mathbb{R}^n$$

with $n = |\mathscr{N}_{SCP}| + |\mathscr{N}_{FiT}|$.

On the contrary, the state variables $|V_i|$ and δ_i representing voltage magnitude and angle at bus $i \in \mathscr{N} \setminus \{sb\}$ in polar formulation (the slack bus serves as a reference bus with $|V_{sb}| = 1$ pu and $\delta_{sb} = 0$), as well as p_{sb} and q_{sb} representing active and reactive power at slack bus, depend on the scenario realization ξ. Given ξ, as well as the active and reactive power p_i and q_i for all $i \in \tilde{\mathscr{N}}$, state variables result from power-flow analysis, using standard Gauss-Seidel, Newton-Raphson or similar methods [53].

Our problem thus has two main types of variables. First, the power modulation variables $\mathbf{p}_i$ and $\mathbf{q}_i$ are determined for all $i \in \mathscr{N}_{SCP} \cup \mathscr{N}_{FiT}$ before realization of the uncertain event. Then, after a scenario ξ is observed, which defines the power production and consumption, $p_i^{\phi}(\xi)$ and $q_i^{\phi}(\xi)$ for each grid user $i \in \tilde{\mathscr{N}}$, a grid-state defined by the values $|V_i|$ and δ_i, $i \in \mathscr{N}$, and p_{sb}, q_{sb}, has to be determined and checked against the operational conditions. In the next two subsections, we will detail how these two types of variables are combined in our chance-constrained OPF.

2.2.2 Power flow constraints under probability sign

The grid operating conditions are constrained by technical limits, which can be expressed as bounds on voltage variables and constraints on the power variables at the slack bus.

More specifically, we denote upper and lower bounds on voltage variables by $\overline{|V_i|}, \overline{\delta_i}$ and $\underline{|V_i|}, \underline{\delta_i}$, respectively. Constraints on active and reactive power at slack bus p_{sb} and q_{sb}, are represented by a set $\mathscr{F}_{sb}$ defined in Demand Connection Code [22] and depicted in Figure 2.1. In addition, other congestion constraints are formulated as thermal limits on the current transit: for each line (i, j) belonging to the set $\mathscr{A}$, we impose an upper limit $(I_{i,j}^{\max})^2$ on a quadratic form $l_{i,j}$ from the current (thermal constraints employed for the use cases are described in Subsection 2.4.3).

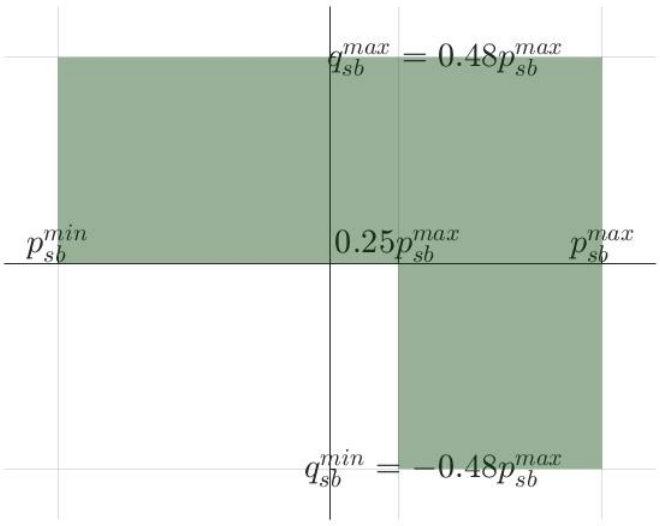


Fig. 2.1 Feasible *P*-*Q* diagram for slack-bus following Demand Connection Code.

All in all, we get the following set of constraints:

$$\underline{\delta_i} \leq \delta_i \leq \overline{\delta_i}, \quad \forall i \in \mathscr{N} \setminus \{sb\} \tag{2.2.2a}$$

$$\underline{|V_i|} \leq |V_i| \leq \overline{|V_i|}, \quad \forall i \in \mathscr{N} \setminus \{sb\} \tag{2.2.2b}$$

$$l_{i,j}(|V_i|, |V_j|, \delta_i, \delta_j) \leq (I_{i,j}^{\max})^2, \quad \forall (i,j) \in \mathscr{A} \tag{2.2.2c}$$

$$(p_{sb}, q_{sb}) \in \mathscr{F}_{sb}. \tag{2.2.2d}$$

In order to ease the model representation, let us define the following functions (here the sum is taken over all buses k connected to the bus i, including bus i):

$$\begin{aligned} &L_i^R(\mathbf{p},\delta,|V|,\xi) := \\ &\left(p_i^{\phi}(\xi) - \boldsymbol{p}_i^{\gamma} - \boldsymbol{p}_i^{\beta}\right) + \sum_{k\sim i} Y_{i,k}^R |V_i||V_k| \cos(\delta_i - \delta_k) \\ &\qquad\qquad + \sum_{k\sim i} Y_{i,k}^I |V_i||V_k| \sin(\delta_i - \delta_k) \\ &L_i^I(\mathbf{q},\delta,|V|,\xi) := \\ &\left(q_i^{\phi}(\xi) - \boldsymbol{q}_i^{\gamma} - \boldsymbol{q}_i^{\beta}\right) + \sum_{k\sim i} Y_{i,k}^R |V_i||V_k| \sin(\delta_i - \delta_k) \\ &\qquad\qquad - \sum_{k\sim i} Y_{i,k}^I |V_i||V_k| \cos(\delta_i - \delta_k). \end{aligned}$$

These functions define the stochastic power flow equations of our problem, where $Y_{i,k}^R$ and $Y_{i,k}^I$ represent the real and imaginary parts of the element (i,k) in the bus admittance matrix Y, respectively.

We can now mathematically state the requirements under the probability sign in problem (2.1.1) by defining the following random set

$$X(\xi) := \left\{ (\mathbf{p},\mathbf{q}) \;\middle|\; \begin{array}{c} \text{there exist } |V|,\delta,p_{sb},q_{sb} \text{ satisfying (2.2.2)}, \\ L_i^R(\mathbf{p},\delta,|V|,\xi) = 0 \text{ for all } i \in \mathscr{N}, \\ L_i^I(\mathbf{q},\delta,|V|,\xi) = 0 \text{ for all } i \in \mathscr{N} \end{array} \right\}. \tag{2.2.3}$$

This set is generally nonconvex and may contain holes and disconnected regions, as illustrated on IEEE test cases in [76].

Note that variables $|V|,\delta,p_{sb},q_{sb}$ depend directly on the random event ξ, whereas $(\mathbf{p},\mathbf{q})$ must be decided before realization of ξ. We are thus interested in finding a decision $(\mathbf{p},\mathbf{q})$ belonging to the random set $X(\xi)$ with probability $1-\alpha$, which is represented by a probability constraint in our optimization model: $\mathbb{P}\big[(\mathbf{p},\mathbf{q}) \in X(\xi)\big] \geq 1-\alpha$.

2.2.3 Chance-constrained OPF formulation

We continue by defining the deterministic constraints of the problem. We start with the conservative bound on generation and consumption

$$\underline{p}_i^{\phi} := \min_{\xi \in \Xi} p_i^{\phi}(\xi) \quad \text{for all } i \in \mathscr{G}$$
$$\bar{p}_i^{\phi} := \max_{\xi \in \Xi} p_i^{\phi}(\xi) \quad \text{for all } i \in \mathscr{L}.$$

When the support set of the random vector ξ is represented by finitely many Monte-Carlo scenarios, Ξ is a finite set, and thus computing $\underline{p}_i^{\phi}$, $\bar{p}_i^{\phi}$ is a straightforward task. This parameter enters in our optimization problem for bounding the decision variables on active power curtailment and modulation, $\boldsymbol{p}^{\beta}$ and $\boldsymbol{p}^{\gamma}$. The power curtailment must satisfy both contractual and technical bounds, for all realizations of the random vector ξ. Hence, we set the following constraints:

$$0 \leq \boldsymbol{p}_i^{\beta} \leq \underline{p}_i^{\phi} \text{ and } \boldsymbol{q}_i^{\beta} = \tan\phi_i \, \boldsymbol{p}_i^{\beta}, \; i \in \mathscr{N}_{FiT} \cap \mathscr{G}$$
$$\bar{p}_i^{\phi} \leq \boldsymbol{p}_i^{\beta} \leq 0 \text{ and } \boldsymbol{q}_i^{\beta} = \tan\phi_i \, \boldsymbol{p}_i^{\beta}, \; i \in \mathscr{N}_{FiT} \cap \mathscr{L},$$

where $\cos\phi_i$ is the given power factor associated with grid user i. Meanwhile, for $i \in \mathscr{N}_{SCP}$, the bounds on power modulation are determined contractually. They can be modeled as fractions of the installed power. Due to the data used in practice, we implemented them as a fraction of $\underline{p}_i^{\phi}$. All in all, the operational constraints are modeled as follows :

$$a_- \underline{p}_i^{\phi} \leq \boldsymbol{p}_i^{\gamma} \leq a_+ \underline{p}_i^{\phi} \quad \text{and} \quad \boldsymbol{q}_i^{\gamma} = \tan\phi_i \, \boldsymbol{p}_i^{\gamma},$$

with $a_- \in [-1, 0]$ and $a_+ \in [0, 1]$.

As for the objective function, we consider the case when $f(\mathbf{p}, \mathbf{q})$ is convex, Subsection 2.4.3. However, our approach remains valid as long as the objective function is DoC using the algorithm proposed in [129].

Based on the assumptions described in the previous paragraphs, it is now possible to present the chance-constrained model for the considered stochastic OPF problem:

$$\min_{\mathbf{p},\mathbf{q}} f(\mathbf{p},\mathbf{q}) \tag{2.2.4a}$$

$$\text{s.t. } 0 \le \boldsymbol{p}_i^{\beta} \le \underline{p}_i^{\phi} \quad \forall i \in \mathscr{N}_{FiT} \cap \mathscr{G} \tag{2.2.4b}$$

$$\bar{p}_i^{\phi} \le \boldsymbol{p}_i^{\beta} \le 0 \quad \forall i \in \mathscr{N}_{FiT} \cap \mathscr{L} \tag{2.2.4c}$$

$$\boldsymbol{q}_i^{\beta} = \tan\phi_i \, \boldsymbol{p}_i^{\beta}, \quad \forall i \in \mathscr{N}_{FiT} \tag{2.2.4d}$$

$$a_- \underline{p}_i^{\phi} \le \boldsymbol{p}_i^{\gamma} \le a_+ \underline{p}_i^{\phi}, \quad \forall i \in \mathscr{N}_{SCP} \tag{2.2.4e}$$

$$\boldsymbol{q}_i^{\gamma} = \tan\phi_i \, \boldsymbol{p}_i^{\gamma}, \quad \forall i \in \mathscr{N}_{SCP} \tag{2.2.4f}$$

$$\mathbb{P}\big[(\mathbf{p},\mathbf{q}) \in X(\xi)\big] \ge 1-\alpha. \tag{2.2.4g}$$

2.3 DoC reformulation and Bundle method

In this section, we first propose a DoC model for problem (2.2.4), and then we show how to solve it by the bundle method of [129]. The proposed model is built upon two DoC reformulations: one exact for the mathematical requirement $(\mathbf{p},\mathbf{q}) \in X(\xi)$ and another approximate for the probability measure $\mathbb{P}$, as detailed in Subsections 2.3.1 and 2.3.2. Subsection 2.3.3 recalls the main details of the chosen optimization algorithm for solving the underlying problem.

2.3.1 DoC reformulation of the condition $(\mathbf{p},\mathbf{q}) \in \mathbf{X}(\xi)$

Let $\xi \in \Xi$ be fixed. For notational convenience, we will denote the decision variables $(\mathbf{p},\mathbf{q})$ of our problem by $x \in \mathbb{R}^{2n}$. Our development starts by noting that the squared distance function to $X(\xi)$, i.e.,

$$d^2_{X(\xi)}(x) := \min_{z \in X(\xi)} \frac{1}{2}\|z-x\|^2,$$

yields the following useful relation:

$$x := (\mathbf{p},\mathbf{q}) \in X(\xi) \quad \Longleftrightarrow \quad d^2_{X(\xi)}(x) = 0, \quad x \in \mathbb{R}^{2n}.$$

Since the set $X(\xi)$ given in (2.2.3) is not convex, the squared distance function is nonconvex. The following development shows that it is a DoC function:

$$\begin{aligned} d^2_{X(\xi)}(x) &= \min_{z\in X(\xi)} \left\{ \frac{1}{2}\|x\|^2 - \langle z, x\rangle + \frac{1}{2}\|z\|^2 \right\} \\ &= \frac{1}{2}\|x\|^2 - \max_{z\in X(\xi)} \left\{ \langle z, x\rangle - \frac{1}{2}\|z\|^2 \right\} = g_1(x,\xi) - g_2(x,\xi). \end{aligned}$$

Indeed, for any arbitrary ξ fixed, the convexity of both functions

$$g_1(x,\xi) := \frac{1}{2}\|x\|^2 \tag{2.3.1a}$$

$$g_2(x,\xi) := \max_{z\in X(\xi)} \left\{ \langle z, x\rangle - \frac{1}{2}\|z\|^2 \right\} \tag{2.3.1b}$$

with respect to x directly results from the definition of convexity. Moreover, observe that $g_1 - g_2$ is non-negative. As a result, we have the following DoC reformulation for checking whether $(\mathbf{p},\mathbf{q})$ satisfies the constraints in (2.2.3):

$$x = (\mathbf{p},\mathbf{q}) \in X(\xi) \quad \Longleftrightarrow \quad g_1(x,\xi) - g_2(x,\xi) \le 0.$$

While g_1 is a simple quadratic function, $g_2(\cdot,\xi)$ is more involved: it is the optimal value of a nonconvex optimization problem and is thus generally nondifferentiable. Its subdifferential at a given point x is given by

$$\begin{aligned} \partial g_2(x,\xi) &:= \left\{ s_2(\xi) \in \mathbb{R}^{2n} :\ g_2(z,\xi) \ge g_2(x,\xi) + \langle s_2(\xi), z - x\rangle \ \ \forall\, z \in \mathbb{R}^{2n} \right\} \\ &= \arg\max_{z\in X(\xi)} \left\{ \langle z, x\rangle - \frac{1}{2}\|z\|^2 \right\} = \arg\min_{z\in X(\xi)} \frac{1}{2}\|z - x\|^2, \end{aligned}$$

which corresponds to the set of projections of x onto $X(\xi)$. If the set $X(\xi)$ were convex (and closed), then the above projection problem would have had a unique solution and, thus, g_2 would have been differentiable. However, even for nonconvex $X(\xi)$, differentiability at x can still occur if the projection of x onto $X(\xi)$ is unique. This condition corresponds to the squared distance function $d^2_{X(\xi)}$ having a unique minimizer at x.

Note that evaluating g_2 at a given point x and computing one of its subgradient (an element of the subdifferential) amounts to projecting x onto the set $X(\xi)$ given in (2.2.3). This task can be accomplished (at least approximately) by OPF tools because this projection problem is indeed an OPF with a quadratic objective function. A numerical procedure (oracle) for accessing the DoC function $g_1(x,\xi) - g_2(x,\xi)$ and its first-order information

is as follows. Here, we use the notation $x^k = (\mathbf{p}^k, \mathbf{q}^k)$ to highlight that this point is fixed, conceivably given by an algorithm at its iteration k.

Oracle 1. Black-box for the DoC function $g_1(x,\xi) - g_2(x,\xi)$.

1: Let $(\mathbf{p}^k, \mathbf{q}^k) \in \mathbb{R}^{2n}$ and event/scenario $\xi \in \Xi$ be given
2: **if** $(\mathbf{p}^k, \mathbf{q}^k)$ belongs to the set $X(\xi)$ in (2.2.3) **then**
3: Set $(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) \leftarrow (\mathbf{p}^k, \mathbf{q}^k)$
4: **else**
5: Let $(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ be a solution of the (scenario-dependent) OPF

$$\min_{(\mathbf{p},\mathbf{q}) \in X(\xi)} \frac{1}{2} \|(\mathbf{p},\mathbf{q}) - (\mathbf{p}^k, \mathbf{q}^k)\|^2. \tag{2.3.2}$$

6: **end if**
7: Set $g_1(\mathbf{p}^k, \mathbf{q}^k, \xi) \leftarrow \frac{1}{2}\|\mathbf{p}^k\|^2 + \frac{1}{2}\|\mathbf{q}^k\|^2$ and $s_1^k(\xi) \leftarrow \begin{pmatrix} \mathbf{p}^k \\ \mathbf{q}^k \end{pmatrix}$
8: Set $g_2(\mathbf{p}^k, \mathbf{q}^k, \xi) \leftarrow \langle \mathbf{p}^k, \tilde{\mathbf{p}} \rangle - \frac{1}{2}\|\tilde{\mathbf{p}}\|^2 + \langle \mathbf{q}^k, \tilde{\mathbf{q}} \rangle - \frac{1}{2}\|\tilde{\mathbf{q}}\|^2$ and $s_2^k(\xi) \leftarrow \begin{pmatrix} \tilde{\mathbf{p}} \\ \tilde{\mathbf{q}} \end{pmatrix}$
9: Return the first order information $(g_1(\mathbf{p}^k, \mathbf{q}^k, \xi), s_1^k(\xi))$ and $(g_2(\mathbf{p}^k, \mathbf{q}^k, \xi), s_2^k(\xi))$.

Step 2 amounts to solving a system of power flow equations and verifying if the computed solution $(\delta, |V|, p_{sb}, q_{sb})$ satisfies the bounds specified in (2.2.2). This task can be accomplished relatively quickly using standard Gauss-Seidel, Newton-Raphson or similar methods [53]. If the power flow solution is operational, then $(\mathbf{p}^k, \mathbf{q}^k)$ automatically solves the OPF of Step 5: there is no need for calling an OPF solver, but only a (simpler) power flow algorithm. However, if the computed solution of the power flow equations does not satisfy (2.2.2), then we say that $(\mathbf{p}^k, \mathbf{q}^k)$ is infeasible for the future event ξ. In this case, an OPF solver must be applied to compute a point $(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ that is feasible for the scenario ξ and as close as possible to $(\mathbf{p}^k, \mathbf{q}^k)$. The most time-consuming task in Oracle 1 is thus Step 5, which is not necessarily executed for all considered scenarios. Such a step thus amounts to answering the following question: given that $(\mathbf{p}^k, \mathbf{q}^k)$ is infeasible for scenario ξ,

what is the smallest necessary perturbation on $(\mathbf{p}^k, \mathbf{q}^k)$ to render it feasible?

The answer is $(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) - (\mathbf{p}^k, \mathbf{q}^k)$, which is nothing but the opposite direction to the subgradient $s_1^k(\xi) - s_2^k(\xi)$. This fact explicitly reveals the practical role of the subgradients computed by Oracle 1: it guides the optimization process to seek for a better candidate solution (at the next iteration). However, caution is necessary: in this analysis, the oracle only sees the given individual scenario ξ. It is thus necessary to account for all the scenarios and the probability

measure in a higher-level oracle (that, for the same point $(\mathbf{p}^k, \mathbf{q}^k)$, can call Oracle 1 in parallel over the scenarios). That is the goal in the following subsection.

2.3.2 DoC reformulation of probability constraint

This subsection aims at presenting a DoC approximation for the probability constraint (2.2.4g). From the previous subsection we have

$$\mathbb{P}[(\mathbf{p},\mathbf{q}) \in X(\xi)] \quad \equiv \quad \mathbb{P}[g_1(\mathbf{p},\mathbf{q},\xi) - g_2(\mathbf{p},\mathbf{q},\xi) \leq 0],$$

with g_1, g_2 two convex functions given in (2.3.1). Hence,

$$\mathbb{P}[(\mathbf{p},\mathbf{q}) \in X(\xi)] \geq 1-\alpha$$

is equivalent to

$$\mathbb{P}[g_1(\mathbf{p},\mathbf{q},\xi) - g_2(\mathbf{p},\mathbf{q},\xi) > 0] \leq \alpha.$$

Next, we follow the lead of [129] to approximate the probability measure by a DoC function. To this end, let $v(\xi) = g_1(\mathbf{p},\mathbf{q},\xi) - g_2(\mathbf{p},\mathbf{q},\xi)$ be the random variable of interest, $\mathbb{E}[\cdot]$ the expected value operator w.r.t. $\mathbb{P}$, and let $\mathbf{1}_{(0,\infty)}(\cdot)$ denote the *characteristic function* of the segment $(0,\infty)$, that equals to 1 if $v > 0$, and 0 if $v \leq 0$. Recall the following useful equivalence:

$$\mathbb{P}[v(\xi) > 0] \;=\; \mathbb{E}[\mathbf{1}_{(0,\infty)}(v(\xi))].$$

The main source of difficulties is that $\mathbf{1}_{(0,\infty)}(\cdot)$ is not convex and, even worse, it is discontinuous at 0. As in [63] and [129], we now approximate the characteristic function by a DoC one. Given a small parameter $t > 0$, the discontinuous characteristic function can be approximated by the continuous one

$$\zeta^t(v) := \begin{cases} 0, & \text{if } v \leq 0 \\ \frac{v}{t}, & \text{if } 0 < v \leq t \\ 1, & \text{if } t < v. \end{cases} \tag{2.3.3}$$

Observe that $\lim_{t\downarrow 0} \zeta^t(v) = \mathbf{1}_{(0,\infty)}(v)$ and $\zeta^t(v)$ has the following DoC decomposition:

$$\zeta^t(v) = \max\left\{\frac{v}{t}, 0\right\} - \max\left\{0, \frac{v-t}{t}\right\}.$$

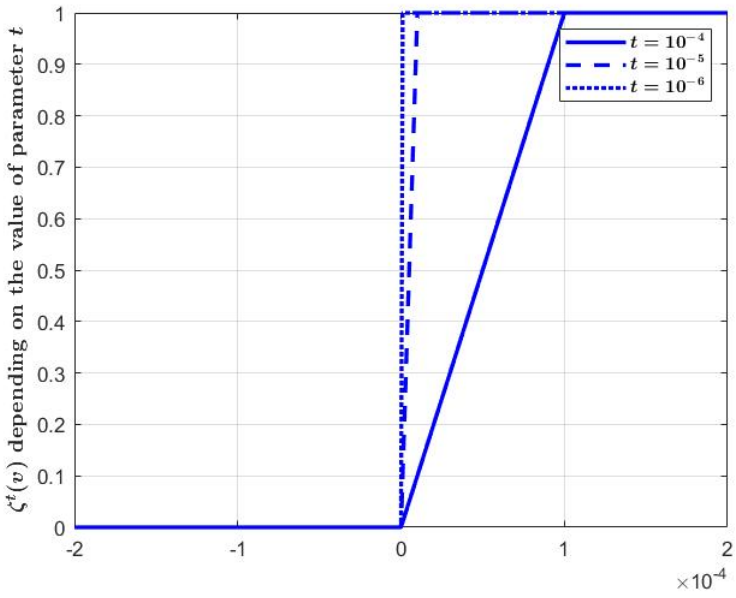


Fig. 2.2 Function $\zeta^t(v)$ for $t = 10^{-4}$, $t = 10^{-5}$ and $t = 10^{-6}$.

Hence, the following expected value

$$\mathbb{E}[\zeta^t(g_1(\mathbf{p},\mathbf{q},\xi) - g_2(\mathbf{p},\mathbf{q},\xi))]$$

is an approximation of the probability

$$\mathbb{P}[g_1(\mathbf{p},\mathbf{q},\xi) - g_2(\mathbf{p},\mathbf{q},\xi) > 0].$$

Such approximation is as good as one wishes: the smaller is the parameter $t > 0$, the closer $\zeta^t(\cdot)$ is to $\mathbf{1}_{(0,\infty)}(\cdot)$. Furthermore, the composition of DoC functions under the expectation is itself a DoC function:

$$\zeta^t(g_1(\mathbf{p},\mathbf{q},\xi) - g_2(\mathbf{p},\mathbf{q},\xi)) = \frac{\max\{g_1(\mathbf{p},\mathbf{q},\xi), g_2(\mathbf{p},\mathbf{q},\xi)\}}{t} - \frac{\max\{g_1(\mathbf{p},\mathbf{q},\xi), g_2(\mathbf{p},\mathbf{q},\xi) + t\}}{t} + 1.$$

It is well-known that 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. As usual in the stochastic programming literature, in our numerical experiments we randomly generate a sample of N scenarios $\Xi := \{\xi^1, \ldots \xi^N\}$ and estimate

the convex functions

$$\mathbb{E}\Big[\frac{\max\{g_1(\mathbf{p},\mathbf{q},\xi),g_2(\mathbf{p},\mathbf{q},\xi)\}}{t}\Big]+1 \quad \text{and} \quad \mathbb{E}\Big[\frac{\max\{g_1(\mathbf{p},\mathbf{q},\xi),g_2(\mathbf{p},\mathbf{q},\xi)+t\}}{t}\Big]. \tag{2.3.4}$$

by their sample average approximations. The justification of such an approach is well documented in the literature (e.g., [97]), and specialized to the DoC setting in [63] and [67, Subsection 7.7.2]. Hence, we can approximate the probability constraint (2.2.4g) with the following DoC constraint

$$c_1(\mathbf{p},\mathbf{q})-c_2(\mathbf{p},\mathbf{q})\leq 0,$$

where

$$\begin{cases} c_1(\mathbf{p},\mathbf{q}) := \dfrac{1}{N}\displaystyle\sum_{j=1}^{N}\max\{g_1(\mathbf{p},\mathbf{q},\xi^j),g_2(\mathbf{p},\mathbf{q},\xi^j)\}+t(1-\alpha), \\ c_2(\mathbf{p},\mathbf{q}) := \dfrac{1}{N}\displaystyle\sum_{j=1}^{N}\max\{g_1(\mathbf{p},\mathbf{q},\xi^j),g_2(\mathbf{p},\mathbf{q},\xi^j)+t\} \end{cases} \tag{2.3.5}$$

are convex functions. As a result, we have our DoC optimization model for the chance-constrained problem (2.2.4):

$$\begin{cases} \min\limits_{(\mathbf{p},\mathbf{q})\in X} & f(\mathbf{p},\mathbf{q}) \\ \text{s.t.} & c_1(\mathbf{p},\mathbf{q})-c_2(\mathbf{p},\mathbf{q})\leq 0, \end{cases} \tag{2.3.6a}$$

where

$$X := \{(\mathbf{p},\mathbf{q})\in\mathbb{R}^{2n}:\ (2.2.4b)-(2.2.4f)\} \tag{2.3.6b}$$

is a polyhedral set. As discussed in Subsection 2.1.2, this optimization problem fits the structure of (2.1.2), and we can apply the DoC bundle method of [129] to tackle it.

The case of a discrete probability distribution. Consider a probability distribution with N outcomes. Then, this DoC reformulation is a relaxation of the chance constraint at a given point.

Indeed, the inequality $\mathbf{1}_{(0,\infty)}(v)\geq\zeta^t(v)$, which is satisfied for all v and $t>0$, leads to

$$\mathbb{E}[\mathbf{1}_{(0,\infty)}(v(\xi))]\geq\mathbb{E}[\zeta^t(v(\xi))].$$

Due to the equivalence

$$\mathbb{P}[x \in X(\xi)] \geq 1-\alpha \quad \Longleftrightarrow \quad \mathbb{E}[\mathbf{1}_{(0,\infty)}(v(\xi))] \leq \alpha, \tag{2.3.7}$$

we conclude that $\mathbb{P}[x \in X(\xi)] \geq 1-\alpha$ implies $\mathbb{E}[\zeta^t(v(\xi))] \leq \alpha$.

Applying (2.3.5) for N scenarios, we obtain the following equivalence:

$$\mathbb{E}[\zeta^t(g_1(\mathbf{p},\mathbf{q},\xi) - g_2(\mathbf{p},\mathbf{q},\xi))] \leq \alpha \quad \Longleftrightarrow \quad c_1(\mathbf{p},\mathbf{q}) - c_2(\mathbf{p},\mathbf{q}) \leq 0,$$

which means that the DoC reformulation is a relaxation of the chance constraint.

2.3.3 DoC Bundle method

We start by noting that oracles for the functions in problem (2.3.6) are readily available. Indeed, f is a convex function and thus simple. Furthermore, an oracle for c_1 and c_2, providing their values and first-order information, is readily implementable thanks to Oracle 1 and assumption that we have finitely many scenarios to represent the future random events: given $x := (\mathbf{p},\mathbf{q}) \in X$, an oracle provides $(c_1(x), s_1 \in \partial c_1(x))$ and $(c_2(x), s_2 \in \partial c_2(x))$. Although such an oracle can be run in parallel over N scenarios, it is not a straightforward one: it requires calling Oracle 1 N times for the same given $x = (\mathbf{p},\mathbf{q})$. Accordingly, N deterministic power flow problems (Step 2 of Oracle 1) and $n_x \leq N$ deterministic OPFs (Step 5 of Oracle 1) must be solved to compute the function values and a pair of subgradients for c_1 and c_2 via (2.3.5). We highlight that the number n_x of deterministic OPFs to be solved to access the probability constraint depends on the quality (in terms of feasibility) of the point $x = (\mathbf{p},\mathbf{q})$. This is an intuition from constraint (2.2.4g): at a feasible point x of problem (2.2.4), we have that $n_x \leq \alpha N$.

Given these oracles, we can go further and briefly present the solution methodology, which has many numerical advantages: it does not require a feasible initial point, nor penalty parameters, and numerical experience suggests that the approach is likely to escape bad-quality critical points. The interested reader is referred to [129] for a detailed presentation of DoC bundle method, as well as its mathematical properties. In that paper, the algorithm is given for a general case of DoC objective function, while the version presented below is adapted for convex objective. We start with the *improvement function* definition, an essential tool for presenting the algorithm. For a given parameter $\hat{\tau} \in \mathbb{R}^2$, the improvement function is

defined as

$$H_{\hat{\tau}}(x) = \max\{f(x) - \hat{\tau}_f, c(x) - \hat{\tau}_c\}, \quad \text{for } c(x) = c_1(x) - c_2(x). \tag{2.3.8}$$

It allows us to define the following convexly-constrained DoC problem

$$\min_{x \in X} H_{\hat{\tau}}(x), \tag{2.3.9}$$

which has the notable property that if $\hat{\tau} = (f^*, 0)$, with f^* the optimal value of (2.3.6), then any solution to problem (2.3.9) solves (2.3.6). However, as f^* is unknown, the parameter $\hat{\tau}$ must be iteratively updated as follows: given $\hat{x} \in X$, a candidate point to solve (2.3.6) produced by the algorithm, we set

$$\hat{\tau} := \big(\hat{\tau}_f, \hat{\tau}_c\big) = \big(f(\hat{x}) + \rho \max\{c(\hat{x}), 0\},\ \sigma \max\{c(\hat{x}), 0\}\big), \tag{2.3.10}$$

where $\rho > 0$ and $\sigma \in [0, 1)$ can be freely chosen. With this rule, the theory presented in [129] ensures that the following algorithm computes a point satisfying necessary optimality conditions to problem (2.3.9).

We highlight that Algorithm 1 is a simplified version of the one presented in [129]. For instance, the prox-parameter $\mu > 0$ can be updated iteratively, the number of constraints in the quadratic (master) program can be kept bounded, and the objective function can be DoC. According to Theorems 1 and 2 from [129], convergence analysis can be summarized as follows.

Theorem 2.3.1. *If the stopping-test tolerance is set to zero (*$\texttt{Tol} = 0$*), then any cluster point of the sequence of stability centers* $\hat{x}$ *generated by Algorithm 1 satisfies certain necessary optimality conditions (criticality) for problem* (2.3.9)*. Moreover, if* $\texttt{Tol} > 0$*, then Algorithm 1 stops after finitely many steps, and the last stability center is an approximate critical point of* (2.3.9)*. Under the additional assumption that* c_2 *is continuously differentiable at* $\hat{x}$*, the same statements hold for problem* (2.1.2)*.*

In summary, the necessary optimality conditions ensured by Algorithm 1 are extensions of the well-known Karush-Kuhn-Tucker conditions in mathematical optimization of differentiable functions. We refer the interested reader to [129, Section 4.4] and Chapter 3 for a dedicated discussion. However, the latter statement of Theorem 2.3.1 cannot be applied in our case, as $g_2(\cdot, \xi)$ is nondifferentiable for any scenario ξ, and consequently, so is c_2. We address this challenge in Chapter 3 by proposing an improvement of the method developed in [129] that yields stronger optimality conditions under weaker assumptions.

Algorithm 1 DoC bundle method

Given $x^0 \in X$, $\alpha \in [0,1]$, choose $\mu > 0$ and $\kappa \in (0,1)$. Set $\hat{x} \leftarrow x^0$ and $\hat{\tau}$ as in (2.3.10)
Compute $(f(x^0), s_f^0 \in \partial f(x^0))$ and $(c_i(x^0), s_i^0 \in \partial c_i(x^0))$, $i = 1,2$

for $k = 0,1,2\ldots$ **do**

▷ Trial point

Let x^{k+1} be the x-part solution of the quadratic program

$$\begin{cases} \min\limits_{x\in X, r\in\mathbb{R}^4} & r_4 - \langle s_2^k, x\rangle & + & \frac{\mu}{2}\|x-\hat{x}\|^2 \\ \text{s.t.} & f(x^j) + \langle s_f^j, x - x^j\rangle & \leq r_1, & j = 0,\ldots,k \\ & c_1(x^j) + \langle s_1^j, x - x^j\rangle & \leq r_2, & j = 0,\ldots,k \\ & c_2(x^j) + \langle s_2^j, x - x^j\rangle & \leq r_3, & j = 0,\ldots,k \\ & r_1 + r_3 - \hat{\tau}_f & \leq r_4 & \\ & r_2 - \hat{\tau}_c & \leq r_4 & \end{cases}$$

▷ Stopping test

if $\|x^{k+1} - \hat{x}\| \leq$ `Tol` **then**
    Stop and return $\hat{x}$
end if

▷ Oracles call

Compute $(f(x^{k+1}), s_f^{k+1} \in \partial f(x^{k+1}))$ and $(c_i(x^{k+1}), s_i^{k+1} \in \partial c_i(x^{k+1}))$, $i = 1,2$

▷ Descent test

if $H_{\hat{\tau}}(x^{k+1}) \leq H_{\hat{\tau}}(\hat{x}) - \kappa\frac{\mu}{2}\|x^{k+1} - \hat{x}\|^2$ **then**

▷ Stability center update

    Set $\hat{x} \leftarrow x^{k+1}$ and update $\hat{\tau}$ as in (2.3.10).
end if

end for

2.4 Results

2.4.1 Network

The current research study utilizes a medium voltage distribution network, inspired by [9, 117], consisting of 33 buses, accommodating 31 loads and with a total peak consumption of approximately 16.6 MW. The network incorporates three distributed generation (DG) units, comprising two biomass plants (one with SCP contract connected in bus 12 with an installed capacity of 6.97 MW and another with FiT contract connected in bus 29 with an installed capacity of 3.86 MW), and one wind farm with a FiT contract connected in bus 32 with an installed capacity of 0.47 MW. The network is connected to the high-voltage

network in bus 1 which is considered the slack bus in the present study. The consumption and generation values were defined to have the network operating near its technical limits creating the need for flexibility activation. This is an extreme case, with a much higher number of constraints compared to a realistic situation. The schematic representation of the network is illustrated in Figure 2.3.

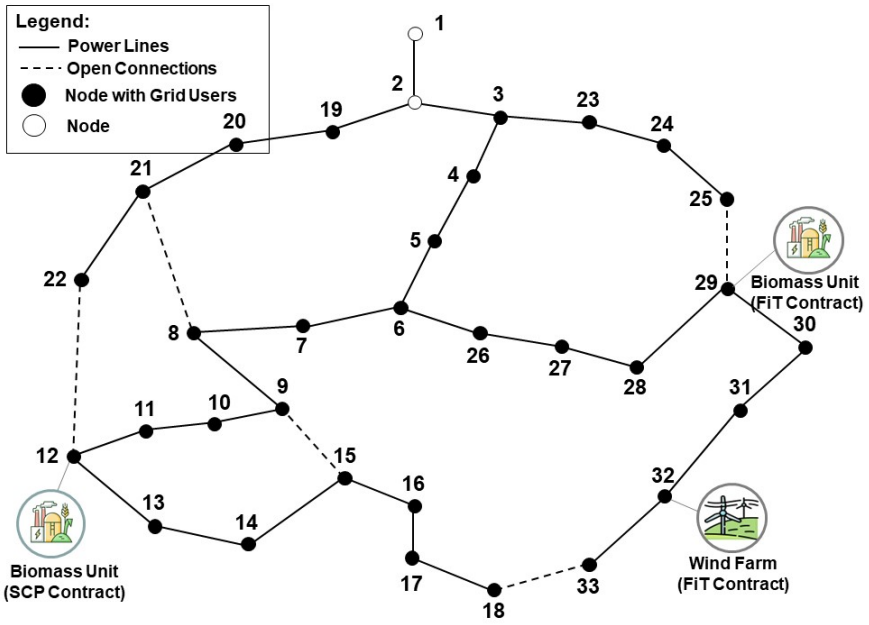


Fig. 2.3 Medium voltage distribution network, 32 buses.

2.4.2 Scenario generation

In order to test the proposed DoC approach on the test network described in Subsection 2.4.1, we use Enedis Open Data [37] on July 27, 2020, to construct load and generation profiles for N scenarios (unless otherwise specified, $N = 1000$). Three types of grid users are considered: biomass generation, wind generation and consumption (data based on small and medium enterprises). For each grid user $i \in \mathcal{N}$, we denote the provided data by $(\tilde{p}_i^{\phi}, \tilde{q}_i^{\phi})$. Next, we attribute an individual variance σ_i^2: 0.0248 pu for biomass generation type, 0.01 pu for wind generation type, and 4 different values in the range between $6.01 \cdot 10^{-5}$ pu and 0.01 pu for consumers. Consider covariance matrix Σ with $\Sigma_{i,j} = r_{i,j}\sigma_i \cdot \sigma_j$, where $r_{i,j} = 1$ if the types of users i and j coincide, and 0 otherwise; and a normalization matrix A with $A_{i,j} = 0.5$ if the user $i = j$ is a generator, $A_{i,j} = 5000$ if the user $i = j$ is a consumer, and 0 otherwise. Applying the procedure described in [61], we compute the nearest symmetric positive semidefinite matrix $\tilde{\Sigma}$ to $A\sigma$. Finally, we generate vectors p^{ϕ} and q^{ϕ} following

multivariate Gaussian distributions with the means $\tilde{p}^{\phi}$ and $\tilde{q}^{\phi}$, respectively, and covariance $\tilde{\Sigma}$.

2.4.3 Parameters of DoC approach

In what follows $P_{base} = 1$ MW, $V_{base} = 12.66$ kV and $I_{base} = 78.99$A.

The upper and lower bounds in (2.2.2b) are defined in the network design manual of a French DSO [33, Annex 1.3] specifically related to connection contracts for grid users, and represent $\pm 5\%$ of nominal voltage for MV network. The upper and lower bounds on δ in (2.2.2a) are set to $\pm\frac{\pi}{2}$. For all $(i,j) \in \mathscr{A}$, thermal constraints (2.2.2c) are reduced to $|I_{i,j}|^2 \leq (I_{i,j}^{\max})^2$. In order to simplify computations of Oracle 1, we convexify the set $\mathscr{F}_{sb}$ in (2.2.2d) in the following form:

$$\mathscr{S} = \left\{ (p_{sb}, q_{sb}) \in \mathbb{R}^2 : \quad p_{sb}^{\min} \leq p_{sb} \leq p_{sb}^{\max}, \quad q_{sb}^{\min} \leq q_{sb} \leq q_{sb}^{\max}, \right.$$
$$\left. q_{sb} \geq \frac{-0.48 p_{sb}^{\max}}{-p_{sb}^{\min} + 0.25 p_{sb}^{\max}} p_{sb} + \frac{0.48 p_{sb}^{\max} p_{sb}^{\min}}{-p_{sb}^{\min} + 0.25 p_{sb}^{\max}} \right\}.$$

In our model, the objective function is convex and has the following structure (all the coefficients are non-negative):

$$f(\mathbf{p}, \mathbf{q}) = f_1(\mathbf{p}) + f_2(\mathbf{p}), \tag{2.4.1a}$$

where

$$f_1(\mathbf{p}) = \sum_{i \in \mathscr{N}_{SCP}} C_{SCP}^i |\boldsymbol{p}_i^{\gamma}| + \sum_{i \in \mathscr{N}_{FiT} \cap \mathscr{G}} C_{FiT,g}^i |\boldsymbol{p}_i^{\beta}| + \sum_{i \in \mathscr{L}} C_{FiT,l}^i |\boldsymbol{p}_i^{\beta}| \tag{2.4.1b}$$

and

$$f_2(\mathbf{p}) = \sum_{i \in \mathscr{N}_{SCP}} c_{SCP}^i (\boldsymbol{p}_i^{\gamma})^2 + \sum_{i \in \mathscr{N}_{FiT} \cap \mathscr{G}} c_{FiT,g}^i (\boldsymbol{p}_i^{\beta})^2 + \sum_{i \in \mathscr{L}} c_{FiT,l}^i (\boldsymbol{p}_i^{\beta})^2. \tag{2.4.1c}$$

The values of coefficients in (2.4.1b) and (2.4.1c) are given in Table 2.1. Our motivation for this choice is as follows. First, the component $f_1(\mathbf{p})$ reflects the cost of lever activation (since coefficients in Table 2.1 are set as dimensionless quantities, this cost is expressed in pu). As the power modulation of a FiT producer is more expensive compared to that of an SCP grid user, the inequality [3] $C_{FiT,g}^i \gg C_{SCP}^j$ should be respected for $i, j \in \mathscr{G}$. Moreover, as the limitation on the power supply to a consumer is more expensive than the power curtailment of a producer, the inequality $C_{FiT,l}^i \gg C_{FiT,g}^j$ should be satisfied for $i, j \in \mathscr{N}_{FiT}$. Thus, the

[3] The relation "$a \gg b$" means that a is much greater than b.

following relation $C^i_{FiT,l} \gg C^j_{FiT,g} \gg C^k_{SCP}$ is respected in (2.4.1b). Further, as a penalty for the energy not supplied is defined by a regulator, we set $C^i_{FiT,l} = C^j_{FiT,l}$ for all consumers $i, j \in \mathscr{L}$. Since the contracts may differ depending on generation technology, we fix different coefficients for biomass generation and wind generation. Meanwhile, the quadratic terms are introduced as a penalty that encourages fairness in power modulation among the same types of users, as the minimum in (2.4.1c) is attained at a point where activation of levers is equal among grid users with equal coefficients. Thus, we set $c^i_{FiT,l} = c^j_{FiT,l}$ for $i, j \in \mathscr{N}_{FiT}$, i.e. we impose the coefficients to be equal for FiT consumers. We also assume that the quadratic coefficients are equal among FiT producers of the same generation type.

Table 2.1 Coefficients in the objective function (2.4.1) for each grid user (GU).

GU	**Contract**	**Type**	**Coeff. in $f_1(\mathbf{p})$**	**Coeff. in $f_2(\mathbf{p})$**
12	SCP	Biomass generation	$4.2 \cdot 10^{-5}$	0
12, 29	FiT	Biomass generation	$4.2 \cdot 10^{-3}$	0.01
32	FiT	Wind generation	0.02	0.1
Others	FiT	Consumption	1	1

Unless otherwise specified, the approximation parameter t from (2.3.3) is set to 10^{-5}. The choice of parameters in DoC bundle method [129] is as follows: $\rho = 10^7$, $\sigma = 0.5$, $\kappa = 0.9$, $\mu_{\min} = 10^{-6}$, $\mu_{\max} = 10^6$ and $\mu^0 = 10^2$. The stopping test parameter `Tol` is set to 10^{-7} unless otherwise specified. To test the performance of the presented DoC approach for the chance-constrained problem (2.2.4), we set 11 values of the safety parameters $1-\alpha$ ranging from 0.75 to 1 with a step size of 0.025 (the case $\alpha = 0$ is tested as an extreme one, the algorithm is not designed for deterministic framework). The algorithm is initialized with a zero vector that corresponds to the initial state of the grid, i.e. without levers activation.

A tailored rule is implemented for updating the prox-parameter μ^k: it is decreased after two consecutive updates of the stability center, and increased if the stability center remains unchanged for more than five consecutive iterations. This approach is motivated by the following considerations. Choosing a large prox-parameter tends to yield solutions close to the previous stability center, which can help satisfy the stopping criterion more rapidly. However, if μ^k increases too quickly and reaches $\mu^{\max}$ prematurely, the algorithm may lose flexibility. To avoid this, the prox-parameter is decreased when the stability center is updated too frequently. This adaptive update rule balances two dynamics.

We consider two cases: one where only voltage constraints are detected (buses 9, 10, 11, 15, 16 and 17) in the initial state of the grid, and another with an additional congestion constraint. For the second case, we set an upper limit on current for the line connecting

buses 2 and 19. The constraint value is chosen in such a way that it is violated for 533 scenarios in the initial state of the grid.

Numerical experiments were performed using MATLAB R2020a on a personal computer with the following characteristics: Windows 10 Professional, 32 Go, Intel i7-10850H (6 cores). Gurobi 9.5.1 was employed for solving the convex quadratic program in Algorithm 1, and the Parallel Computing Toolbox was used to parallelize the calls to Oracle 1. The Newton-Raphson method was applied to solve the power flow equations in Step 2.

2.4.4 Case 1: Voltage constraints

For the default values of parameters $N = 1\,000$, $\texttt{Tol} = 10^{-7}$ and $t = 10^{-5}$, and each value of the safety parameter, the algorithm manages to find a feasible critical point with an average execution time of 1665 seconds ranging from 1059 ($1-\alpha = 0.875$) to 2279 ($1-\alpha = 0.775$) seconds. For several values of the safety parameter $1-\alpha$, Figure 2.4 illustrates the number $n_{x^k} \leq N$ of different deterministic OPF problems solved at every iteration k (see Oracle 1, Step 5). Recall that an OPF solver is called in Oracle 1 only if, for given scenario ξ and current point $x^k = (\mathbf{p}^k, \mathbf{q}^k)$, no solution $(\delta, |V|, p_{sb}, q_{sb})$ of a system of power flow equations satisfies the bounds in (2.2.2). Note that n_{x^k} approaches αN as the iteration number k increases. This is as expected, as we have already argued in Subsection 2.3.3. Note, however, that n_{x^k} cannot coincide with αN due to our approximation (2.3.3) of the characteristic function.

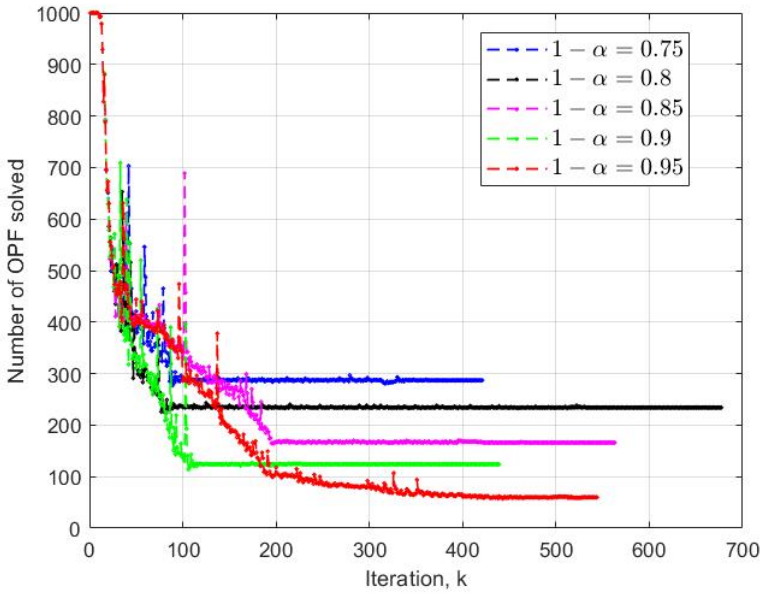


Fig. 2.4 Number of deterministic (scenario-based) OPF problem solved per iteration for different values of the safety parameter $1-\alpha$.

Figure 2.5 shows that increasing the stopping test parameter `Tol` or reducing the number N of scenarios improves the execution time. The average improvement among all values of the safety parameter if increasing the stopping test tolerance from 10^{-7} to 10^{-5} is 34.6%. As Figure 2.6 shows, power modulation cost represented by (2.4.1b) in the objective function, is not very sensitive to the change of parameter `Tol` in the range $[10^{-7}, 10^{-5}]$. The maximum relative difference between the cost for the reference case with $\texttt{Tol} = 10^{-7}, N = 1000$, and the case with $\texttt{Tol} = 10^{-5}$ ($N = 1000$) among all values of the safety parameter is 3.8%. The analogous difference for the reference case and the case with $\texttt{Tol} = 10^{-6}$ ($N = 1000$) is 2.7%. Figure 2.6 also shows that, on the contrary, a decrease in a number of scenarios N has an important impact on the power modulation cost, therefore on the solution of the problem as well. It confirms an interest in the proposed approach enabling to treat large number of scenarios: even if considering fewer scenarios is more efficient in terms of execution time, the obtained solution is less robust with respect to future events.

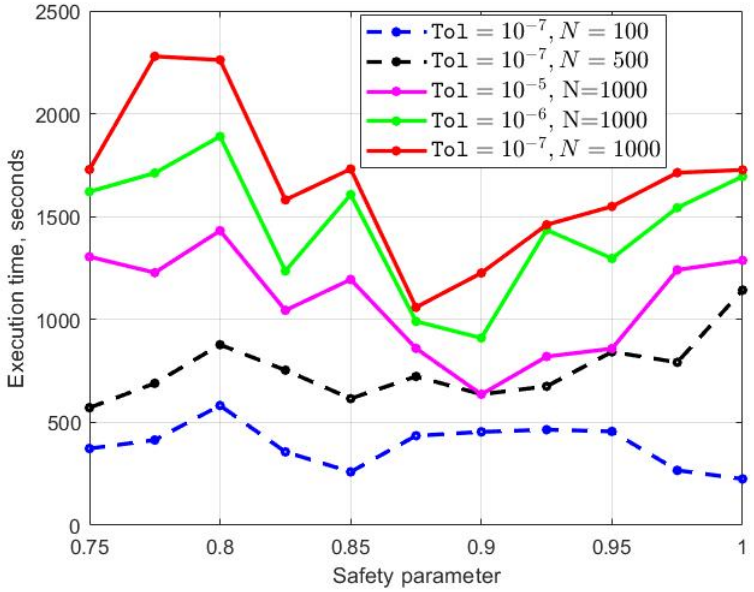


Fig. 2.5 Execution time for different number of scenarios N and stopping test parameter `Tol`.

To check the validation of the chance constraint, we compare the targeted value of the safety parameter $1-\alpha$ with the observed value defined as the ratio of scenarios satisfying the power flow equations. The tests have been performed for three values of parameter t that participates in the DoC approximation of the probability constraint (2.3.3), $t = 10^{-4}, 10^{-5}$ and 10^{-6}, Figure 2.7. The values of parameters N and `Tol` are set by default. In the initial state of the grid, without lever activation, the ratio of scenarios satisfying the power flow equations is equal to 0.545. Figure 2.7 shows that it gets closer to the targeted safety

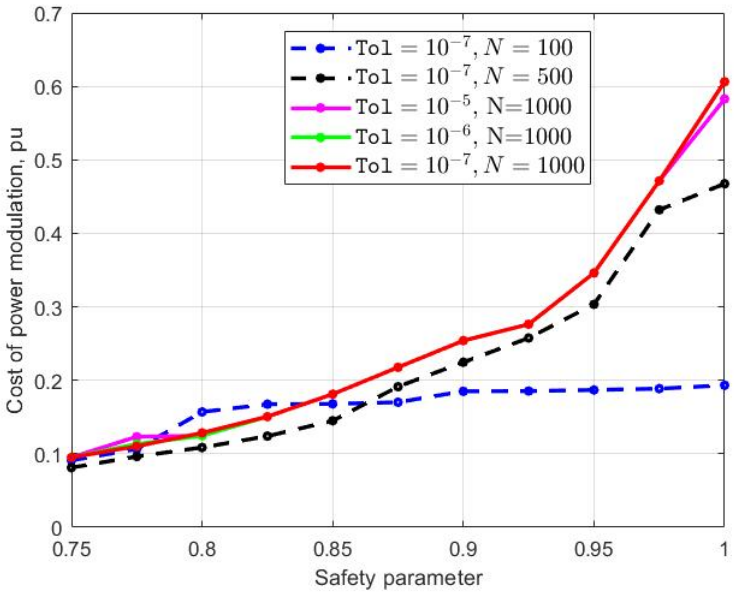


Fig. 2.6 Comparison of power modulation cost for different number of scenarios N and stopping test parameter `Tol`.

parameter after optimization. We observe that the gap between the latter and the ratio of scenarios satisfying the power flow equations tends to decrease with an increase in safety parameter $1-\alpha$ for all parameter values t. However, the gap remains always positive due to the approximation of the probability function used for DoC formulation (see Subsection 3.2). Moreover, Figure 2.7 reveals that, for a fixed value of the safety parameter $1-\alpha$, the ratio of scenarios satisfying the power flow equations is always higher for a smaller parameter t. This illustrates, as expected, an increase in accuracy as t goes to zero. Thus, for $1-\alpha = 0.9$ the required ratio constitutes 0.84, 0.876 and 0.893 for $t = 10^{-4}$, $t = 10^{-5}$ and $t = 10^{-6}$, respectively. This accuracy can be improved by considering a larger number of scenarios: for $N = 10\,000$ the analogous figures are 0.8427, 0.8822, and 0.8941 for $t = 10^{-4}$, $t = 10^{-5}$ and $t = 10^{-6}$, respectively. Power modulation cost is not very sensitive to the change of parameter t. The maximum relative difference between the cost for the reference case with $t = 10^{-5}$ and the case with $t = 10^{-4}$ among all values of the safety parameter is 3.9%. The analogous difference between the reference case and the case with $t = 10^{-6}$ is 3.4%.

The values of all parameters are set by default in what follows ($N = 1000$, $\texttt{Tol} = 10^{-7}$, $t = 10^{-5}$). As the targeted safety parameter increases, the volume of active power modulation grows linearly up to $1-\alpha = 0.925$ and accelerates afterward, as illustrated in Figure 2.8. We observe that the same tendency is valid for the volume supplied by FiT grid users, whose share in total power curtailment remains within $69-77\%$. Meanwhile, an SCP grid user

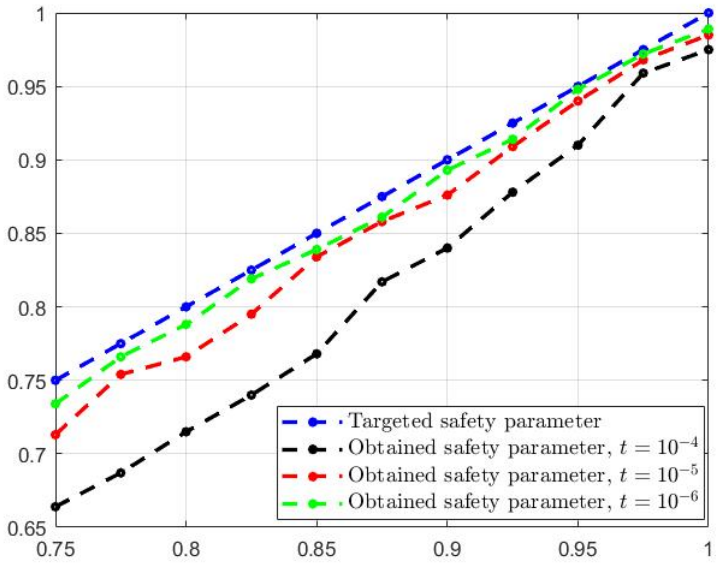


Fig. 2.7 Comparison of targeted safety parameter $1-\alpha$ with the obtained one for $t = 10^{-4}$, $t = 10^{-5}$ and $t = 10^{-6}$.

supplies the remaining part of the volume within the bounds of her guaranteed power for all values of the safety parameter. This situation illustrates that the use case is indeed extreme, as FiT users need to be actively engaged even at the lowest safety level $1-\alpha = 0.75$. Power modulation cost, represented by (2.4.1b) in the objective function, follows the same upward trend as power modulation volume. It is consistent with the price formation, which is linear in the volume. Therefore, a slight decrease in safety parameters implies a reduction in the total cost of power modulation. This result aligns with our expectations from probabilistic modeling of power flow constraints.

Computing maximum and average amplitudes of constraint violations over all scenarios where a constraint violation is detected, we compare them with an average volume of power modulation supplied by one grid user. Without lever activation, the maximum amplitude of constraint violation is 0.0339 pu, whereas the average is 0.0057 pu. Figure 2.9 reveals a notable decrease in the maximum amplitude of constraint violation while increasing the system reliability up to 0.0102 pu for $\alpha = 0$ and a significant growth in the average volume of power modulation. This confirms that the higher the volume of levers activated, the less amplitude of constraint violation is. To ensure the system reliability with $1-\alpha \geq 0.9$, the average power modulation volume exceeds the maximum amplitude of constraint violation. The gap between the two figures drastically increases as $1-\alpha$ approaches 1 due to the stochastic character of our problem: if we did not have to deal with uncertainties and the

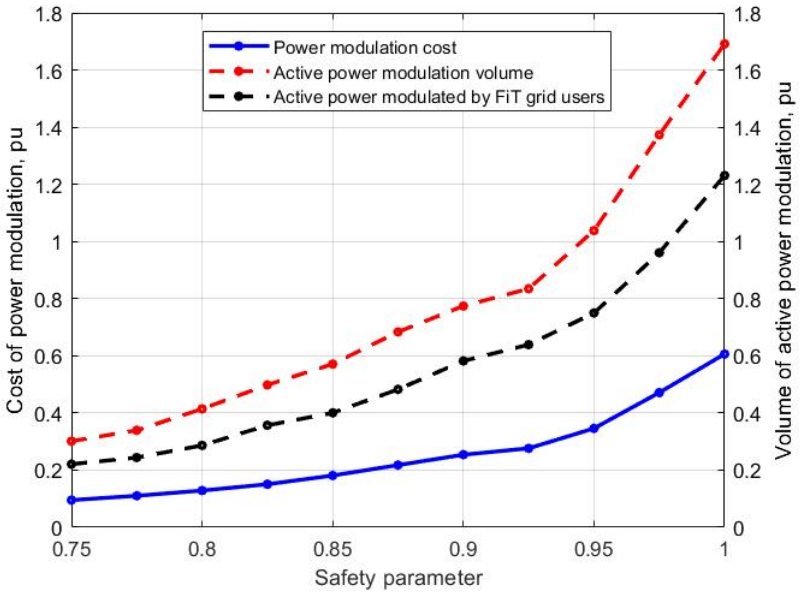


Fig. 2.8 Cost and volume of power modulation.

scenario realization was known, covering the maximum amplitude of constraint violation would have been sufficient.

Meanwhile, the average amplitude of constraint violation fluctuates near the value 0.0045 pu with a slight tendency to decrease. These upward and downward trends in the amplitude of constraint violation lead to the following conclusions. More reliable but costly solutions cover risky scenarios that allow a higher amplitude of constraint violation. Although this tendency is not directly implied by the chance-constrained formulation, it is expected for our problem, as resolving one voltage constraint often helps mitigate others due to their interdependence. The fact that it does not apply to the curve of average amplitude highlights that the chance-constrained formulation is not made to reduce the latter, since it imposes no specific requirements on the amplitude of constraint violations.

2.4.5 Case 2: Voltage and congestion constraints

The algorithm finds a critical point with an average execution time of 1947 seconds ranging from 1353 ($1-\alpha = 0.85$) to 2850 ($1-\alpha = 0.825$) seconds. However, for the deterministic case $\alpha = 0$, the algorithm does not manage to find a feasible solution possibly because none exists. The ratio of scenarios satisfying the power flow equations is equal to 0.266 for the initial state of the grid. As Figure 2.10 shows, this ratio approaches the targeted value of the safety parameter once lever activation is optimized. However, for $1-\alpha > 0.95$, the

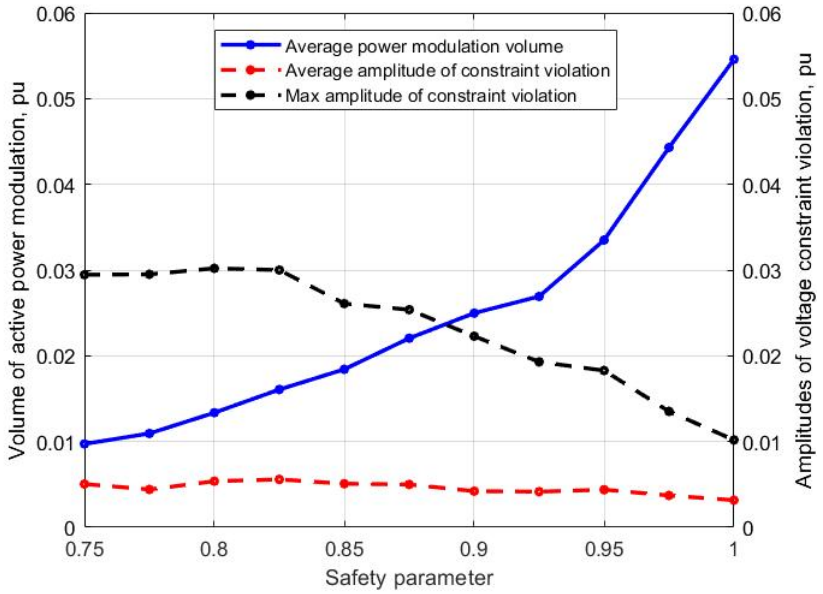


Fig. 2.9 Volume of power modulation and amplitudes of constraint violation.

difference between the targeted and obtained value of the safety parameter increases. This can be due to the fact that a robust solution may not exist. The difference between the targeted and obtained value of the safety parameter represents 4.6% of the targeted value for $\alpha = 0$, whereas, initially, it was only 3.4% for $1 - \alpha = 0.75$. The gap for $\alpha = 0$ is significant due to the infeasibility of the obtained solution.

Similarly to the case without congestion constraints, power modulation cost repeats the growth dynamics of power modulation volume, Figure 2.11. The part supplied by FiT grid users constitutes 79% to 87% of the total volume, and the remaining part is due to an SCP grid user within the bounds of her guaranteed power. However, compared to the previous case, the growth acceleration of power modulation cost and volume is less pronounced as the safety parameter goes to 1. Moreover, these values become almost stable at $1 - \alpha = 0.975$ and $\alpha = 0$, which is consistent with the fact that the corresponding number of covered scenarios is very close (953 for $1 - \alpha = 0.975$ and 954 for $\alpha = 0$). Thus, except for the deterministic case with $\alpha = 0$ when the obtained solution is not feasible, a small decrease in the safety parameter allows us to reduce the total cost of lever activation.

As in the previous case, we compute the average and maximum amplitudes of voltage constraint violation. These values are calculated over all scenarios where a voltage constraint violation is detected (scenarios where only a congestion constraint is violated, are not included). Without lever activation, they remain the same as for the case without congestion constraints. Meanwhile, maximum and average amplitudes of congestion constraint viola-

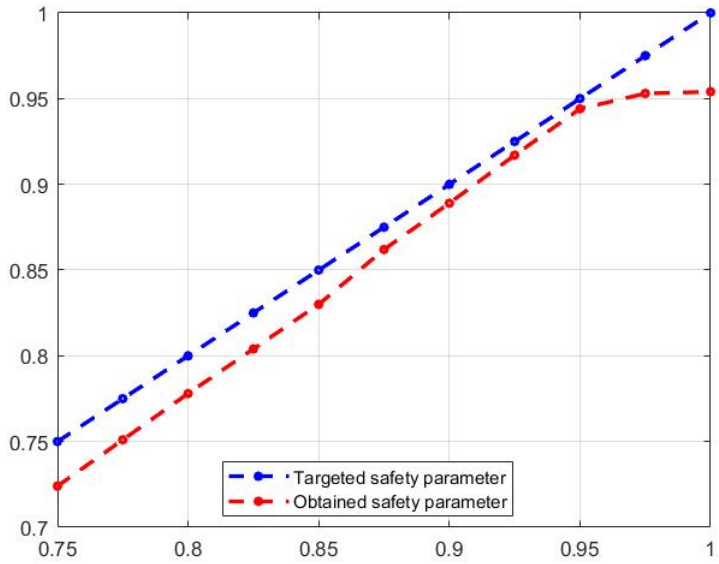


Fig. 2.10 Comparison of targeted safety parameter $1-\alpha$ with the obtained one.

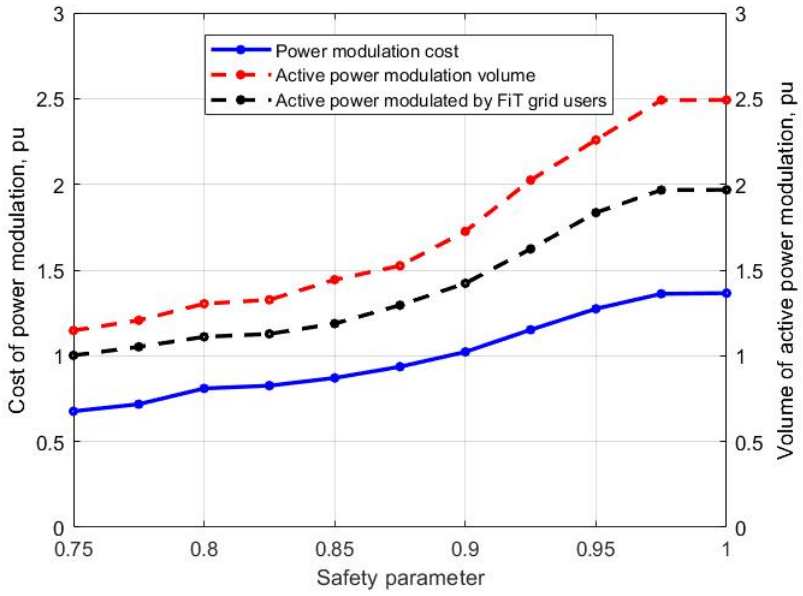


Fig. 2.11 Cost and volume of power modulation.

tions calculated over all scenarios where a congestion constraint violation is detected, are 0.3851 pu and 1.5155 pu, respectively. As the order of magnitude of congestion constraint violation is greater than that of voltage constraint violation, the corresponding values are plotted separately, Figure 2.12 and Figure 2.13. All in all, they illustrate the same trends

as in the previous case, namely, a steady decrease in the maximum amplitude of constraint violation, both for voltage and congestion constraints. At the same time, Figure 2.12 reveals a significant growth in the average volume of power modulation while increasing the system reliability. Thus, the conclusion that an increase in the volume of lever activation reduces the maximum amplitude of constraint violation remains valid.

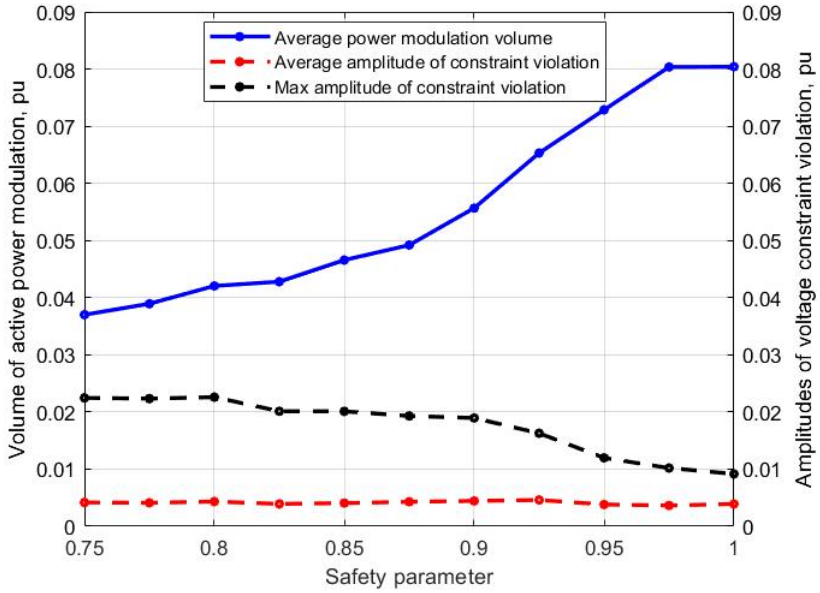


Fig. 2.12 Volume of power modulation and amplitudes of constraint violation (excluding congestion constraint violation).

Comparing curves of average power modulation volume in Figure 2.9 and Figure 2.12, we observe that it is higher for the case with congestion constraint. Moreover, the average power modulation volume always exceeds the maximum amplitude of voltage constraint violation in the latter case. This is due to the stochastic character of our problem, but also to the additional amplitude of congestion constraint violation that should be covered by power modulation. Meanwhile, the curve of maximum power modulation volume is higher in the first case. It can be explained by the curtailment of grid users downstream of the congestion constraint, which weakens voltage constraints at other buses of the grid and reduces the amplitude of their violation for scenarios not covered. Furthermore, the ratio of scenarios where congestion constraint is violated among all not covered scenarios remains within $47-59\%$ and tends to slightly increase as the safety parameter goes to 1. In other words, more reliable but costly solutions are more prone to cover scenarios without congestion constraints.

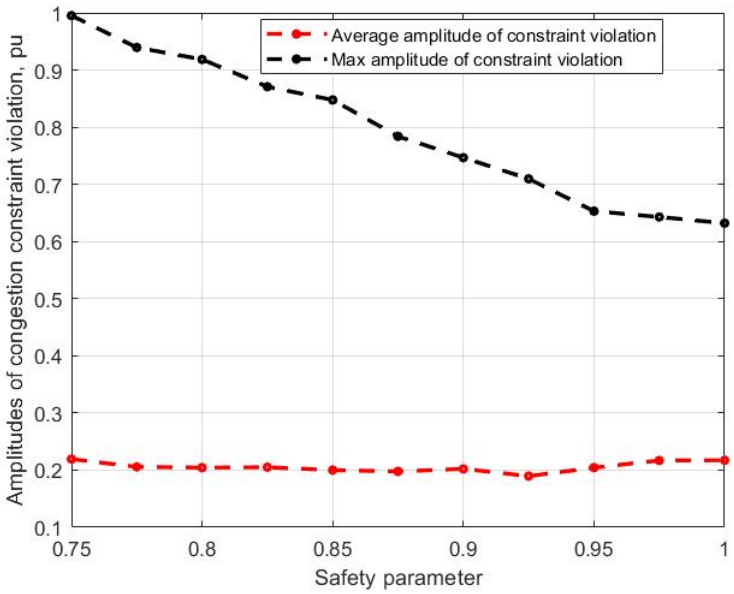


Fig. 2.13 Amplitudes of congestion constraint violation.

Comparing the activation of levers for grid users downstream the line connecting buses 2 and 19 (congestion constraint in Case 2) in two considered use cases, we observe that there is no power curtailment of those grid users for all values of the safety parameter $1-\alpha$, $\alpha > 0$ in Case 1. Meanwhile, they are curtailed in Case 2 for all values of the safety parameter $1-\alpha$, $\alpha > 0$. As the voltage constraints are the same in both use cases, we conclude that their curtailment resolves only the congestion constraint. Moreover, we also observe that the power curtailment is not equal among those users, although all of them are consumers. Consequently, the soft constraint imposed by quadratic terms (2.4.1c) in the objective function, is not sufficient to force an equal power curtailment inside the same group of grid users. This is due to the nonconvex nature of the optimization model.

2.5 Discussion

The proposed methodology enables the integration of uncertainties in distribution network operational planning problems modeled as a joint chance-constrained OPF. One main advantage of this approach is its versatility: no particular assumption on the system or type of uncertainties is necessary. For instance, no matter the (random) system one has in mind, as long as the objective function is DoC, the random set $X(\xi)$ in (2.2.3) can be mathematically defined, and a solver able to handle the scenario-based optimization problem (2.3.2) is available, the presented methodology applies. In other words, given a DoC objective function

and any system represented by $X(\xi)$, all one needs to make the approach functional is an implementable Oracle 1.

Thanks to Oracle 1, this approach naturally decomposes along scenarios and allows for parallel computations. Therefore, the decision variables that depend explicitly on the random vector can be handled in different processors. As a result, the algorithm's QP master program is independent of the number N of scenarios used to estimate the probability function. However, the algorithm's computational burden depends on N as Oracle 1 needs to be called N times per iteration.

The above observation is related to the approach's main limitation: Oracle 1 is assumed to be exact, i.e., a globally optimal solution to (2.3.2) must be computed whenever Step 5 of Oracle 1 is accessed. This requirement is strong as it requires projecting a point onto the nonconvex set $X(\xi)$. One way to overcome this difficulty is to consider convex approximations of $X(\xi)$, a path we have not explored in this chapter (we have studied an SDP relaxation of OPF (2.3.2) with modified constraints on voltage angles in Subsection 6.1.1). Instead, we have taken the problem of projecting onto $X(\xi)$ as it is: a difficult nonconvex OPF problem. However, instead of employing a global optimization solver, we have applied a local solver based on the interior-point algorithm. In other words, we have run Algorithm 1 with a potentially inexact oracle. As such, the convergence analysis provided in [129] does not apply. We mention in passing that certain DoC (bundle) methods can handle inexact oracles provided their errors asymptotically vanish. This is the case of Algorithm 1 in [28], whose general presentation encompasses several methods. However, that paper addresses only the convexly-constrained setting, a simpler setting than the one considered here. Extending that analysis (on inexact oracles) to our more general DoC-constrained setting is an interesting direction for future research.

A further limitation of this method is the requirement that the deterministic set X in (2.3.6) must be convex (note that X should not be confused with $X(\xi)$, which is generally nonconvex). This excludes the situation where the vector of decision variable x (equal to $(\mathbf{p},\mathbf{q})$ in our application) must satisfy binary constraints. However, a binary constraint $z \in \{0,\, 1\}$ can be reformulated as a combination of convex and DoC constraints

$$z \in \{0,1\} \quad \equiv \quad z \in [0,1] \text{ and } z - z^2 \leq 0.$$

Thus, binary constraints can, in principle, be handled by our approach, as the DoC proximal bundle algorithm can address multiple DoC constraints. We have applied this reformulation trick to the new model of power curtailment described in Chapter 4.2, where binary variables are used to model operational planning rules. However, this technique did not result in good numerical performance in practice. For more details, see Section 5.1. As a result, we have

abandoned the DoC approach in the context of the chance-constrained problem with discrete variables. Chapter 5 outlines the alternative path we chose for this case.

Another aspect related to the additional assumption on the differentiability of the second convex component c_2 in Theorem 2.3.1, which is not satisfied in our case, encouraged us to explore how to strengthen the DoC bundle method and eliminate this condition. In addition, this study led us to the algorithm generalization, so that it no longer relies on an explicit DoC decomposition of the involved functions. Chapter 3 focuses on the generalization of DoC bundle method and its convergence analysis, while Section 3.8 compares the performance of the two algorithms on the first use case presented in this chapter. Although the new algorithm provides a more cost-effective solution, achieving criticality leads to an increase in execution time.

2.6 Conclusion

This chapter presents a formulation of the distribution network operational planning problem as a joint chance-constrained AC-OPF, modeled as a Difference-of-Convex-constrained optimization problem. In this formulation the decision bears on the activation of flexibility (power modulation) from SCP and FiT grid users. This approach takes into account the correlation between RES generation profiles and load profiles, which represent uncertainties of the model via a joint probability constraint approximated with scenarios. The solution methodology consists of model reformulation as a DoC optimization problem with the use of a parallelizable special numerical procedure (oracle), and the subsequent use of a DoC proximal bundle method.

The case study on a 33-bus network accommodating 31 loads, among which three DG units (one with SCP contract and two with FiT contracts), has been analyzed. Numerical experiments, considering 1 000 load and generalization profile scenarios, evidence the approach's advantages and limitations in addressing joint chance-constrained OPF problems without linearization and convexification of power flow equations: decisions that are robust against future events are computed at the expense of more computational burden. Further, running times can be kept at acceptable levels depending on the available computational resources (number of available processors), the efficiency of the employed deterministic OPF solver, and the number of scenarios used to approximate the probability constraint. The approach demonstrates flexibility concerning the balance between model accuracy and computational efficiency while keeping solutions' robustness close to prescribed values.

Chapter 3

Convex-Weakly-Convex Algorithm

This chapter is dedicated to developing the Convex-Weakly-Convex (CwC) bundle method. This direction was initially chosen to strengthen the convergence properties of the DoC bundle method discussed in Chapter 2. As noted in Subsection 2.3.3, the latter does not guarantee providing a critical point for the DoC model (2.1.2) of the chance-constrained AC-OPF. The new algorithm addresses this limitation effectively. Furthermore, the new method applies to a broad class of nonsmooth and nonconvex optimization problems beyond this framework, with the objective and constraints being the difference of convex and weakly convex functions. The practical performance of the algorithm is illustrated by numerical experiments on some nonconvex stochastic problems.

The main content of this chapter has appeared in [124] (K. Syrtseva, W. de Oliveira, S. Demassey, and W. van Ackooij (2024). "Minimizing the difference of convex and weakly convex functions via bundle method". *Pacific Journal of Optimization*).

Ce chapitre présente le développement d'une méthode de faisceaux dédiée aux problèmes d'optimisation dont l'objectif et les contraintes sont modélisables comme différence de fonctions convexes et faiblement convexes (CwC). Initialement, cette approche a été conçue pour renforcer les propriétés de convergence de la méthode de faisceaux DoC présentée dans le Chapitre 2. Comme évoqué dans la Sous-section 2.3.3, cette dernière ne garantit pas l'obtention d'un point critique pour le modèle DoC (2.1.2) *de l'AC-OPF sous contrainte probabiliste jointe. Le nouvel algorithme permet de surmonter les limitations mentionnées. De plus, il s'applique à une large classe de problèmes d'optimisation non-lisses et non-convexes au-delà de ce cadre, où la fonction objectif et les contraintes sont CwC. La performance pratique de l'algorithme est illustrée par des expérimentations numériques sur des problèmes stochastiques non-convexes.*

Le contenu principal de ce chapitre a été publié dans [124] (K. Syrtseva, W. de Oliveira, S. Demassey, et W. van Ackooij, (2024). "Minimizing the difference of convex and weakly convex functions via bundle method". Pacific Journal of Optimization.)

3.1 Motivation and main contributions

In this chapter, we present a bundle method for nonsmooth and nonconvex optimization problems of the form

$$\min_{x\in X} f(x) \quad \text{s.t.} \quad c(x) \leq 0, \tag{3.1.1a}$$

where X is a nonempty bounded polyhedron contained in an open convex set $\mathscr{O} \subset \mathbb{R}^n$, and functions $f : \mathscr{O} \to \mathbb{R}$ and $c : \mathscr{O} \to \mathbb{R}$ are decomposable as the difference of *convex* and (locally) *weakly convex* functions. More specifically, we assume that the following convex-weakly convex (CwC) decompositions are available:

$$f(x) = f_1(x) - f_2(x) \quad \text{and} \quad c(x) = c_1(x) - c_2(x), \tag{3.1.1b}$$

with $f_1, c_1 : \mathscr{O} \to \mathbb{R}$ convex and $f_2, c_2 : \mathscr{O} \to \mathbb{R}$ *weakly convex functions on some neighbourhood of each* $x \in \mathscr{O}$. We adopt the more general definition of weakly convex functions (see Definition 3.2.2 below) given in [145, Def. 4.2] so that we can exploit the equivalence between the families of locally weakly convex and Lower-C^2 functions [113, Thm. 1.3, Cor. 1.3] to highlight the breadth of our approach. In particular, we have in mind the following settings for f_2 (as well as for c_2):

i) $f_2(x) = \phi(x)$ is a (possibly nonsmooth) convex function;

ii) $f_2(x) = -h(x)$ with h having Lipschitz continuous gradient;

iii) $f_2(x) = \phi(x) - h(x)$, with ϕ and h as given above;

iv) $f_2(x)$ is the optimal value of $\max_{t\in T} F(t,x)$, with T a (possibly nonconvex) compact set and F of class C^2;

v) $f_2(x) = \phi(G(x))$, with $\phi : \mathbb{R}^m \to \mathbb{R}$ convex and Lipschitz and $G : \mathbb{R}^n \to \mathbb{R}^m$ a smooth mapping with Lipschitz Jacobian.

Analogous settings for c_2, and their combinations with the ones for f_2, are covered by our analysis (see Section 3.2 below for details).

Weakly convex functions enjoy favorable properties in so much as that they can be recast as DoC functions [56]. Hence, problem (3.1.1) can, in theory, be recast as a DoC-constrained DoC program, a setting that proves *practical if explicit DoC decompositions are available*;

see for instance [75, 70, 118, 90, 49, 129, 128, 120] and references therein. However, if no DoC decomposition is known for f or c, the DoC machinery is unsuitable, and the methods proposed in these references are not applicable. This is already the case for the more straightforward items ii) and iii) above if the underlying Lipschitz constant is unknown and no upper bound is readily available[1] The situation becomes even more complicated for items iv) and v): in general, there are no formulae, rules, or practical insights to obtain a DoC decomposition for f_2 in these cases (see Example 3.1.1 below for a particular case of iv). A strategy to handle problem (3.1.1) via DoC programming algorithms is to replace functions f_i and c_i ($i = 1,2$) with $f_i(x) + \frac{\mu}{2}\|x\|^2$ and $c_i(x) + \frac{\mu}{2}\|x\|^2$ for a large parameter $\mu > 0$ estimating upper bounds on the unknown weakly-convex moduli μ_f and μ_c of f_2 and c_2 (see Proposition 3.2.4), hoping that $f_2(x) + \frac{\mu}{2}\|x\|^2$ and $c_2(x) + \frac{\mu}{2}\|x\|^2$ are convex on X. As, in general, there is no reliable way to assert the convexity of these latter functions, DoC programming algorithms applied in this context must be understood as heuristics. Remarkably, the work [140] exploits such a strategy by combining a dynamic rule to update μ with a nonconvexity test to achieve convergence, but only in a probabilistic sense. Differently, for a class of nonconvex two-stage stochastic problems, the authors of [79] exploit an implicitly convex-concave structure of the objective function and propose an algorithm based on the so-called partial Moreau envelope that disregards DoC decompositions at the price of non-negligible computational costs.

In contrast to the above references, we investigate a bundle method approach for tackling (3.1.1), which neither requires explicit DoC decompositions of the involved functions (in particular, bounds on the weakly-convex moduli μ_f and μ_c need not be known), nor relies on (often costly) Moreau envelopes. For the method to work, it suffices to dispose of a difference of *convex and weakly convex* (CwC) decomposition of the involved functions, as in (3.1.1b). Compared to DoC, the latter structure appears more naturally in applications (see [67, § 7.5]) and, at the time of writing the article [124], has yet to be exploited to design optimality conditions and numerical algorithms. This chapter aims to fill this gap. Currently, research on the problems involving CwC functions is ongoing. For example, the behavior of Difference-of-Convex algorithm (DCA) for unconstrained optimization is studied in [109], while the authors of [27] apply a progressive decoupling algorithm (PDA) for minimizing CwC functions over a linear subspace. Meanwhile, the paper [112] focuses on the generalization of proximal methods on a broader class of nonsmooth and nonconvex problems.

Our approach broadens and enhances the method proposed in [129] for dealing with DoC-constrained DoC-problems, and its particular case given in Subsection 2.3.3, in two

[1] It is worth noting that in many practical problems, mainly those from data science, an upper bound on such a constant can often be computed, although it may be quite large.

ways. First, the availability of DoC decompositions is no longer needed, which makes our approach applicable to a larger scope of problems. Second, it is ensured to compute critical points for the original problem without any additional assumption on the second components: f_2 and c_2 need not be continuously differentiable as assumed in [129, Thm. 2]. In addition, it has a lower cost per iteration (the master subproblem has fewer constraints than the one of [129]). Similarly to [129], our approach builds upon a problem reformulation via improvement function, a well-known and successful strategy in the nonsmooth optimization literature [110, 7, 90]. However, due to the above modifications, the convergence analysis of our extension of the method proposed in [129] must be done anew. Furthermore, a new criticality definition for the reformulated problem links directly with (necessary) optimality conditions for the original problem (3.1.1), which makes it a major ingredient for these enhancements. Such a criticality concept is introduced and analyzed in Section 3.3 below, where we also extend the alternative characterization of Bouligand stationarity given in [98] to our CwC setting. Before that, we motivate this work with the following example that presents a class of problems (of great practical appeal) where the CwC decomposition arises upon applying a well-known interior-penalty strategy.

Example 3.1.1 (Nonconvex two-stage programming)**.** Let $\Xi := \{\xi^1, \dots \xi^S\}$ be a set of scenarios and $\pi_s > 0$ the probability of occurrence of event ξ^s, $s = 1, \dots, S$. Consider the following nonconvex two-stage program

$$\begin{cases} \min\limits_{x \in X} & f_1(x) + \sum\limits_{s=1}^{S} \pi_s Q(x; \xi^s) \\ \text{s.t.} & c_1(x) - c_2(x) \leq 0 \end{cases} \quad \text{with} \quad Q(x;\xi) := \begin{cases} \min\limits_{y \in Y} & q(x,y;\xi) \\ \text{s.t.} & \psi_i(x,y;\xi) \leq 0,\ i = 1, \dots, m. \end{cases} \tag{3.1.2}$$

Assume that:

- $f_1, c_1, c_2 : \mathbb{R}^n \to \mathbb{R}$ are convex (possibly nonsmooth) functions;
- $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^{n_2}$ are two (non-empty) convex and bounded polytopes;
- $q, \psi_i : \mathbb{R}^n \times \mathbb{R}^{n_2} \times \Xi \to \mathbb{R}$, $i = 1, \dots, m$, possess the following characteristics: $q(\cdot, \cdot, \xi)$ and $\psi_i(\cdot, \cdot, \xi)$ are twice-continuously differentiable for every $\xi \in \Xi$ fixed and, moreover, $q(x, \cdot, \xi)$ and $\psi_i(x, \cdot, \xi)$ are convex for every x and ξ fixed;
- the constraints in the subproblem $Q(x;\xi)$ satisfy the Slater condition: for every $x \in X$ and $\xi \in \Xi$, there exists $y^\circ(x;\xi) \in Y$ such $\psi_i(x, y^\circ(x;\xi), \xi) < 0$, $i = 1, \dots, m$.

As presented in Chapter 2, the DoC constraint $c_1(x) - c_2(x) \leq 0$ above is particularly useful in this stochastic programming setting to model chance constraints.

Under the above assumptions, evaluating the recourse function $Q(x;\xi)$ amounts to solving a well-defined convex optimization problem on variable y. Although this essential property is present, the recourse function itself fails to be convex on variable x (but $Q(\cdot;\xi)$ is continuous as a result of [13, Prop. 4.4]). Furthermore, without further assumptions, computing a (generalized) subgradient of $Q(\cdot;\xi)$ at x as well as asserting additional properties about this function are challenging tasks. This could for instance be done if the constraints satisfy a further Aubin or Lipschitz like property upon exploiting [92, Chapter 4]. Still though, most likely, at best we would be dealing with subdifferentials inclusions - and concrete algorithms to handle such general "marginal functions" would be unavailable.

A possible manner to curtail these difficulties is to approximate the recourse function with a more tractable one. As explained in [14], with the help of the log-barrier penalty function and a penalization parameter $\varepsilon > 0$, we may approximate $Q(x;\xi)$ with

$$Q^{\varepsilon}(x;\xi) := \min_{y\in Y} \; q(x,y;\xi) - \frac{1}{\varepsilon}\sum_{i=1}^{m} \log(-\psi_i(x,y;\xi)). \tag{3.1.3}$$

Given the above assumptions, it is well known that $Q^{\varepsilon}(x;\xi) \downarrow Q(x;\xi)$ as $\varepsilon \downarrow 0$ (e.g., [14, § 2.2] and [13, p. 266]), and thus the model

$$\begin{cases} \min\limits_{x\in X} & f_1(x) - f_2(x) \\ \text{s.t.} & c_1(x) - c_2(x) \leq 0 \end{cases} \quad \text{with} \quad f_2(x) := \sum_{s=1}^{S} \pi_s[-Q^{\varepsilon}(x;\xi^s)]$$

is an accurate approximation of (3.1.2) when $\varepsilon > 0$ is small enough. Furthermore, as $-Q^{\varepsilon}(x;\xi) = \max_{y\in Y} \; \frac{1}{\varepsilon}\sum_{i=1}^{m} \log(-\psi_i(x,y;\xi)) - q(x,y;\xi)$ is a weakly convex function (c.f. item iv) above), this model fits the structure (3.1.1). We highlight that $Q^{\varepsilon}(x;\xi)$ is generally a nonsmooth (nonconvex) function; hence, the above problem is challenging. To our knowledge, no practical and mathematically sound optimization algorithm could tackle this class of problems before this work. For instance, [88] requires f_2 to be smooth, [14] introduces an additional Tikhonov regularization term to (3.1.3) to ensure smoothness of $Q^{\varepsilon}(x;\xi)$, and [79] assumes $q(\cdot,\cdot,\xi)$ and $\psi_i(\cdot,\cdot,\xi)$ to be concave-convex functions. In all these references, function c_2 is absent. Being nonsmooth, we mention in passing that a (generalized) subgradient of f_2 at x can be computed and seen to be $\sum_{s=1}^{S} \pi_s g(y(x;\xi^s))$, where $g(\cdot) := \nabla_x[\frac{1}{\varepsilon}\sum_{i=1}^{m} \log(-\psi_i(x,\cdot;\xi^s)) - q(x,\cdot;\xi^s)]$ is the gradient w.r.t. x of the objective function of (3.1.3) multiplied by -1, and $y(x;\xi)$ is an arbitrary optimal solution of (3.1.3) (see Proposition 3.2.1 and [105, Thm. 7.3]). □

Our approach is still applicable in a more general case, where the probability vector π in the above example is a function (of class C^2) of the first-stage variable x, i.e., $\pi_s(x)$,

$s = 1, \ldots, S$. Hence, the class of optimization problems considered in this chapter includes the challenging family of stochastic programming recourse models with decision-dependent uncertainty (e.g. [59] and [79]).

The remainder of this chapter is organized as follows. Section 3.2 recalls essential definitions, key elements, and well-known concepts from variational analysis. Necessary optimality conditions for problem (3.1.1) are presented in Section 3.3 as well as the problem reformulation via an improvement function. Once the link between the reformulated and the original problem is established in the same section, Section 3.4 focuses on an improvement-function-based bundle method for problem (3.1.1). Section 3.5 presents the method's convergence analysis to critical points, while Section 3.7 illustrates the practical performance of our approach on some nonconvex stochastic optimization problems and a compressed sensing problem. Finally, it is compared to the DoC bundle method from Chapter 2 in Section 3.8.

Notation. The following notation is employed throughout the chapter. For a real number a, we denote by $[a]_+$ the value $\max\{a, 0\}$. For any points $x, y \in \mathbb{R}^n$, $\langle x, y\rangle$ stands for the Euclidean inner product, and $\|\cdot\|$ for the associated norm, i.e., $\|x\| = \sqrt{\langle x, x\rangle}$. For a convex set X, $N_X(x)$ stands for its normal cone at the point x, i.e., the set $\{y : \langle y, z - x\rangle \leqslant 0 \text{ for all } z \in X\}$ if $x \in X$ and the empty set otherwise. The Bouligand tangent cone to a (possibly nonconvex) set $W \subset \mathbb{R}^n$ at a point $w \in W$ is the set $\mathscr{T}_W(w)$ of all tangent directions in the following sense: $d \in \mathscr{T}_W(w)$ if there exist a sequence of vectors $\{w^k\} \subset W$ converging to w and a sequence of positive scalars $t_k \to 0$ such that $d = \lim_{k\to\infty}(w^k - w)/t_k$. The indicator function of $X \subset \mathbb{R}^n$ is defined as $i_X(x) = 0$ if $x \in X$ and $i_X(x) = +\infty$ otherwise. The convex hull of a set X is $\mathtt{conv}X$ and the relative interior is denoted by $\mathtt{ri}X$. The domain of a function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is represented by $\mathscr{D}\mathrm{om}(\varphi) = \{x \in \mathbb{R}^n : \varphi(x) < +\infty\}$. Notation $\mathscr{O}$ stands for an open convex set of the Euclidean space $\mathbb{R}^n$ and, given the definitions of f and c, we have that $\mathscr{O} \subset \mathscr{D}\mathrm{om}(f)$ and $\mathscr{O} \subset \mathscr{D}\mathrm{om}(c)$. The component functions of f and c are f_1, f_2, and c_1, c_2 respectively: f_1 and c_1 are convex, whereas f_2 and c_2 are weakly convex on some neighbourhood of every $x \in \mathscr{O}$. Finally, f^* stands for the Legendre-Fenchel transform of a function $f : \mathbb{R}^n \to (-\infty, +\infty]$.

3.2 Definition and prerequisites

This section starts by recalling the concept of (generalized) directional derivatives and subdifferentials. Basic subdifferential calculus is summarized in Proposition 3.2.1 below, followed by the definitions of weakly convex and lower-C^2 functions. The section closes

with Proposition 3.2.4 asserting that the definition of (locally) weakly convex function can be globally extended to the whole convex and compact set X.

A function $\mathsf{f} : \mathscr{O} \to \mathbb{R}$ is said to be *locally Lipschitz continuous* if for each $x' \in \mathscr{O}$ there is a neighbourhood $V_{x'} \subset \mathscr{O}$ of x' such that, for some $L_{x'} \geq 0$,

$$|\mathsf{f}(x) - \mathsf{f}(y)| \leq L_{x'} \|x - y\| \quad \forall\, x, y \in V_{x'}.$$

The function f is said to be *Lipschitz continuous* on $\mathscr{O}$ if $L_{x'} = L$ can be taken independent of $x' \in \mathscr{O}$, and $V_{x'}$ in the above inequality is replaced with $\mathscr{O}$.

Directional derivatives and subdifferentials. Let $\phi : \mathscr{O} \to \mathbb{R}$ be a convex function. Then ϕ is locally Lipschitz continuous and, for each $x \in \mathscr{O}$, the directional derivative

$$\phi'(x; d) := \lim_{\tau \downarrow 0} \frac{\phi(x + \tau d) - \phi(x)}{\tau}$$

exists (and is finite) in every direction $d \in \mathbb{R}^n$ [93, Prop. 2.81 and Cor. 2.82]. Such a derivative can be represented by $\phi'(x; d) = \max_{s \in \partial\phi(x)} \langle s, d \rangle$, where $\partial\phi(x)$ is the *subdifferential* of ϕ at point x:

$$\partial\phi(x) := \{s \in \mathbb{R}^n : \phi(y) \geq \phi(x) + \langle s, y - x \rangle \quad \forall y \in \mathbb{R}^n\}. \tag{3.2.1}$$

The elements of $\partial\phi(x)$ are referred to as the *subgradients* of ϕ at x. The *approximate subdifferential* is defined, for $\varepsilon \geq 0$, by

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

Let $\mathsf{f} : \mathscr{O} \to \mathbb{R}$ be a locally Lipschitz continuous function. Then the generalized directional derivative defined by

$$\mathsf{f}^\circ(x; d) := \limsup_{x' \to x,\ \tau \downarrow 0} \frac{\mathsf{f}(x' + \tau d) - \mathsf{f}(x')}{\tau}$$

is finite for all $x \in \mathscr{O}$ in every direction $d \in \mathbb{R}^n$ [20, Prop. 2.1.1(a)]. Such a mathematical concept permits us to define the *Clarke subdifferential* of f at $x \in \mathscr{O}$,

$$\partial^{\mathsf{C}}\mathsf{f}(x) := \{g \in \mathbb{R}^n : \langle g, d \rangle \leq \mathsf{f}^\circ(x; d) \ \text{ for all } \ d \in \mathbb{R}^n\}, \tag{3.2.2}$$

which is a nonempty, convex, and compact subset of $\mathbb{R}^n$ [20, Prop. 2.1.2(a)] satisfying $\mathsf{f}^\circ(x;d) = \max_{g\in\partial^{\mathsf{C}}\mathsf{f}(x)}\langle g,d\rangle$. The elements of $\partial^{\mathsf{C}}\mathsf{f}(x)$ are referred to as *generalized (or Clarke) subgradients*, as they are the usual subgradients, i.e., $\partial^{\mathsf{C}}\mathsf{f} = \partial\mathsf{f}$, when f is convex [20, Prop. 2.2.7]. Furthermore, when f is continuously differentiable, $\partial^{\mathsf{C}}\mathsf{f}(x)$ reduces to the singleton $\{\nabla\mathsf{f}(x)\}$. An alternative representation, in finite dimensions, of $\partial^{\mathsf{C}}\mathsf{f}(x)$ is (see [20, Thm. 2.5.1])

$$\partial^{\mathsf{C}}\mathsf{f}(x) := \mathtt{conv}\left\{\lim_{\iota\to\infty}\nabla\mathsf{f}(x_\iota), x_\iota\to x, \mathsf{f}\text{ differentiable at }x_\iota\right\}.$$

A fundamental result, often evoked in this chapter, is the following one [20, Prop. 2.1.2]: the mapping $\partial^{\mathsf{C}}\mathsf{f}$ is locally bounded in the interior of $\mathscr{D}\mathrm{om}(\mathsf{f}) := \{x\in\mathbb{R}^n : \mathsf{f}(x)<\infty\}$. As a result, the image $\partial^{\mathsf{C}}\mathsf{f}(X)$ of every bounded set $X\subset\mathscr{O}$ $(\subset\mathscr{D}\mathrm{om}(\mathsf{f}))$ is bounded in $\mathbb{R}^n$. Useful calculus rules of subdifferentials are listed in Proposition 3.2.1 below and rely on the concept of regularity.

A locally Lipschitz continuous function $\mathsf{f}:\mathscr{O}\to\mathbb{R}$ is *subdifferentially regular* (or simply regular) at $x\in\mathscr{O}$ if for every $d\in\mathbb{R}^n$ the ordinary directional derivative at x exists and coincides with the generalized one:

$$\mathsf{f}'(x;d) = \mathsf{f}^\circ(x;d) \quad \forall\, d\in\mathbb{R}^n.$$

It holds that smooth functions, as well as convex ones, are regular at every point in the interior of their domains. Moreover, a finite linear combination (by non-negative scalars) of regular functions at x is regular [20, Prop. 2.3.6].

Proposition 3.2.1. *Let $\mathsf{f}_t:\mathscr{O}\to\mathbb{R}$, $t=1,2,\ldots,m$, be locally Lipschitz functions and $x\in\mathscr{O}$ an arbitrary point. Then*

i) $\partial^{\mathsf{C}}[\sum_{t=1}^m a_t\mathsf{f}_t](x)\subset\sum_{t=1}^m a_t\partial^{\mathsf{C}}\mathsf{f}_t(x)$ *for all $a\in\mathbb{R}^m$, and equality holds if*

- *all but one of f_t are smooth [20, Prop. 2.3.3 and Cor. 2];*
- *or if every f_t is regular at x and $a\in\mathbb{R}^m_+$ [20, Cor. 3];*

ii) $\partial^{\mathsf{C}}\mathsf{f}(x)\subset\mathtt{conv}\{\partial^{\mathsf{C}}\mathsf{f}_t(x) : t\in I(x)\}$, *for* $\mathsf{f}(x)=\max\limits_{t=1,\ldots,m}\mathsf{f}_t(x)$ *and* $I(x):=\arg\max\limits_{t=1,\ldots,m}\mathsf{f}_t(x)$, *and equality holds and f is regular if every f_t is regular at x, [20, Prop. 2.3.12].*

The last item can be strengthened when more structure is present, such as in the case of weakly convex functions (see Eq. (3.2.3) below).

Weakly convex functions: definition and main properties.

Definition 3.2.2 (Def. 4.2 [145])**.** A function $\mathsf{f} : \mathscr{O} \to \mathbb{R}$ is said to be (locally) *weakly convex* on $\mathscr{O}$ if, on some neighbourhood $V_{x'} \subset \mathscr{O}$ of each $x' \in \mathscr{O}$, there exists $\mu_{x'} \geq 0$ such that, for all $\mu \geq \mu_{x'}$

$$\phi(x) := \mathsf{f}(x) + \frac{\mu}{2}\|x\|^2 \quad \text{is finite and convex on } V_{x'}.$$

Furthermore, f is said to be *weakly convex in the global sense* on $\mathscr{O}$ if the above property holds for $V_{x'} = \mathscr{O}$ and $\mu_{x'} = \bar{\mu} \geq 0$ regardless of $x' \in \mathscr{O}$. □

Clearly, a convex function on $\mathscr{O}$ is weakly convex in the global sense: it suffices to take $\mu_{x'} = 0$ and $V_{x'} = \mathscr{O}$ for all $x' \in \mathscr{O}$. When f is a smooth function with a Lipschitz continuous gradient, then f is weakly convex in the global sense with $\mu = L$ the Lipschitz constant of $\nabla\mathsf{f}$ [28, Prop. 1]. Moreover, it follows from [31, Lemma 4.2] that the family of composite functions given in item v) of Section 3.1 is also weakly convex in the global sense.

Definition 3.2.2 implies that weakly convex functions are *locally DoC*: the decomposition $\mathsf{f}(x) = \phi(x) - \frac{\mu}{2}\|x\|^2$ holds on some neighbourhood of every $x' \in \mathscr{O}$. As a result, [107, Thm. 10.33] ensures that the class of weakly convex functions coincides with that of Lower-C^2 functions; see also [113, Thm. 1.3, Cor. 1.3].

Definition 3.2.3 (Def. 10.29 [107])**.** (LC^2 functions). A function $\mathsf{f} : \mathscr{O} \to \mathbb{R}$ is said to be *Lower-C^2* or LC^2 on $\mathscr{O}$ if, on some neighbourhood $V_{x'} \subset \mathscr{O}$ of each $x' \in \mathscr{O}$, there is a representation

$$\mathsf{f}(x) = \max_{t \in T} \mathsf{f}_t(x).$$

in which the functions f_t are of differentiability class C^2 on $V_{x'}$ and the index set T is a compact space such that $\mathsf{f}_t(x)$, $\nabla\mathsf{f}_t(x)$, and $\nabla^2\mathsf{f}_t(x)$ depend continuously not just on $x \in V_{x'}$ but jointly on $(t,x) \in T \times V_{x'}$. □

In particular, if f is given by $\mathsf{f}(x) = \max\{\mathsf{f}_1(x), \ldots, \mathsf{f}_m(x)\}$ and all functions $\mathsf{f}_1, ..., \mathsf{f}_m$ are of class C^2, then f is Lower-C^2/weakly convex. Furthermore, the functions of item iv) are also weakly convex, since they are Lower-C^2 by definition.

An important property of LC^2/weakly convex functions is regularity [106, Thm. 1]: for every $x \in \mathscr{O}$, the equality $\mathsf{f}'(x;d) = \mathsf{f}^\circ(x;d)$ holds in every direction $d \in \mathbb{R}^n$. Theorem 7.3 in [105] gives the following characterization of the Clarke subdifferential of f at $x \in \mathscr{O}$: for $I(x) = \arg\max_{t \in T} \mathsf{f}_t(x)$,

$$\partial^{\mathsf{C}}\mathsf{f}(x) = \mathtt{conv}\,\{\nabla_x \mathsf{f}_t(x) \,:\, t \in I(x)\} \quad \text{for all } x \in \mathscr{O}. \tag{3.2.3}$$

Furthermore, the concept of the approximate subdifferential can be generalized to the class of weakly convex functions; for details, see [139]. When constrained to a compact convex

set $X \neq \emptyset$, we can say more about weakly convex functions. Indeed, the local property in Definition 3.2.2 globally extends to the whole X, and we have the following result.

Proposition 3.2.4. *Let $\mathsf{f} : \mathscr{O} \to \mathbb{R}^n$ be a weakly convex function, and $X \subset \mathscr{O}$ a compact and convex set. Then there exist a real number $\mu_{\mathsf{f}} \geq 0$ and an open convex set $\mathscr{O}'$ satisfying $X \subset \mathscr{O}' \subset \mathscr{O}$ such that, for all $\mu \geq \mu_{\mathsf{f}}$:*

i) the function $\phi(x) := \mathsf{f}(x) + \frac{\mu}{2}\|x\|^2$ is convex on $\mathscr{O}'$ and $\partial\phi(x) = \partial^{\mathsf{C}}\mathsf{f}(x) + \mu x$ for all $x \in \mathscr{O}'$;

ii) for all $s_f \in \partial^{\mathsf{C}} f(x)$ with $x \in \mathscr{O}'$, the following inequality holds

$$\mathsf{f}(y) \geq \mathsf{f}(x) + \langle s_f, y - x\rangle - \frac{\mu}{2}\|y - x\|^2 \quad \forall y \in X. \tag{3.2.4}$$

Proof. Since $\mathsf{f} : \mathscr{O} \to \mathbb{R}^n$ is weakly convex, it follows by definition that, relative to some neighbourhood $V_{x'}$ of each point $x' \in \mathscr{O}$, there exist $\mu_{x'} > 0$ such that for all $\mu \geq \mu_{x'}$ the function $\phi(x) = \mathsf{f}(x) + \frac{\mu}{2}\|x\|^2$ is finite and convex on $V_{x'}$. In such a representation, there is no loss of generality in assuming that $V_{x'} \subset \mathscr{O}$ (if necessary we can define a new/smaller neighbourhood as $V_{x'} \cap \mathscr{O}$ for which the above conclusion obviously stands). By considering all the points in X, let $V := \{V_{x'} : x' \in X\}$ be the collection of all such neighbourhoods. Then, by construction, V is an open cover of the compact set X and, by definition of compactness, it has a finite open subcover, i.e., there exists finitely many points $\{x'_1, \ldots, x'_m\} \subset X$ such that $\mathscr{O}' := \cup_{i=1}^m V_{x'_i} \supset X$, and by construction $\mathscr{O}'$ is an open subset of $\mathscr{O}$. The first part of item i) thus follows by taking $\mu_{\mathsf{f}} := \max_{i=1,\ldots,m} \mu_{x'_i} < \infty$. By writing $\mathsf{f}(x) = \phi(x) - \frac{\mu}{2}\|x\|^2$ and recalling Proposition 3.2.1 i) we get $\partial^{\mathsf{C}}\mathsf{f}(x) = \partial\phi(x) - \mu x$ for all $x \in \mathscr{O}'$. This concludes item i).

To show item ii), let us now define $\tilde{\phi}(x) = \mathsf{f}(x) + \frac{\mu}{2}\|x\|^2 + i_{\mathscr{O}'}(x)$, an extended real-valued convex function: $\tilde{\phi} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. Note that for each $x \in \mathscr{O}'$, there exists a neighbourhood $V_x \subset \mathscr{O}'$ such that $\tilde{\phi}(x') = \phi(x')$ for all $x' \in V_x$. This fact permits us to conclude that $\partial\tilde{\phi}(x) = \partial\phi(x)$ for all $x \in \mathscr{O}'$. It thus follows from item i) that, for every $x \in \mathscr{O}'$ and every $s \in \partial\tilde{\phi}(x)$, there exists $s_f \in \partial^{\mathsf{C}} f(x)$ such that $s = s_f + \mu x$ and the subgradient inequality reads as

$$\tilde{\phi}(y) \geq \tilde{\phi}(x) + \langle s_f + \mu x, y - x\rangle \quad \forall y \in \mathbb{R}^n,$$

i.e., $\mathsf{f}(y) + \frac{\mu}{2}\|y\|^2 + i_{\mathscr{O}'}(y) \geq \mathsf{f}(x) + \frac{\mu}{2}\|x\|^2 + i_{\mathscr{O}'}(x) + \langle s_f + \mu x, y - x\rangle$ for all $y \in \mathbb{R}^n$. The latter simplifies to

$$\mathsf{f}(y) + i_{\mathscr{O}'}(y) \geq \mathsf{f}(x) + \langle s_f, y - x\rangle - \frac{\mu}{2}\|y - x\|^2 \quad \forall y \in \mathbb{R}^n.$$

By restricting y to the set X and recalling that $s_f = s - \mu x \in \partial^{\mathsf{C}} \mathsf{f}(x)$ is an arbitrary subgradient (because no restriction was imposed to $s \in \partial\tilde{\phi}(x)$), the above inequality becomes (3.2.4). □

Concerning the setting of this chapter where X is compact, the appealing DoC decomposition $\mathsf{f}(x) = \phi(x) - \frac{\mu_{\mathsf{f}}}{2}\|x\|^2$ is, unfortunately, unavailable: the threshold μ_{f} in Proposition 3.2.4 is in general unknown. This fact precludes the application of DoC techniques to optimization problems featuring general Lower-C^2/weakly convex functions. Interested readers are referred to [140] for a strategy that uses approximated DoC decompositions based on item i) of Proposition 3.2.4.

3.3 Necessary optimality conditions and problem reformulation

Let $f, c : \mathscr{O} \to \mathbb{R}$ be given by (3.1.1b). We highlight that some properties of their components can be transferred to these functions. (To ease the presentation, let us focus only on $f(x) = f_1(x) - f_2(x)$, as the same conclusions hold for c.) For instance, f is locally Lipschitz continuous because f_1 and f_2 are so. Furthermore, as f_1 is convex and f_2 is (locally) weakly convex, they are both directional differentiable and these properties extend to f as well: for every $x \in \mathscr{O}$, the directional derivative of f is finite in every direction $d \in \mathbb{R}^n$ as result of the following relation:

$$f_1'(x;d) - f_2'(x;d) = \lim_{\tau \downarrow 0} \left[\frac{f_1(x+\tau d) - f_1(x)}{\tau} - \frac{f_2(x+\tau d) - f_2(x)}{\tau} \right] = f'(x;d).$$

However, the important regularity condition of both f_1 and f_2 does not extend to f as a mere fact that the latter is not a linear combination with non-negative coefficients of the two former functions (see Proposition 3.2.1.i)). Hence, we cannot expect to have equality in the following inclusion

$$\partial^{\mathsf{C}} f(x) \subset \partial f_1(x) - \partial^{\mathsf{C}} f_2(x)$$

unless one of the component functions is smooth at x. Such an inclusion impacts stationary concepts as we will now discuss. Let us first consider the convexly-constrained problem

$$\min_{x \in X} f(x), \quad \text{with} \quad f(x) = f_1(x) - f_2(x). \tag{3.3.1}$$

A point $\bar{x} \in X$ is said to be directional (d)-stationary for this problem if $f'(\bar{x};d) \geq 0$ for all $d \in \mathscr{T}_X(\bar{x})$. The following result generalizes [98, Prop. 5], where a specific case of problem (3.3.1) with $f_2(x) = \max\{\psi_1(x), \ldots, \psi_m(x)\}$ and convex $\psi_1, \ldots, \psi_m$ is considered.

Proposition 3.3.1. *A point $\bar{x} \in X$ is d-stationary of problem* (3.3.1) *if, and only if,*

$$\bar{x} \in \operatorname*{arg\,min}_{x \in X} f_1(x) - [f_2(\bar{x}) + \langle s_{f_2}, x - \bar{x} \rangle] \quad \forall\, s_{f_2} \in \partial^{\mathtt{C}} f_2(\bar{x}).$$

Proof. Observe that $\mathscr{T}_X(\bar{x}) = \mathtt{cl}\,\{d \in \mathbb{R}^n : d = t(x - \bar{x}),\ x \in X, t \in \mathbb{R}_+\}$ due to convexity of X. Therefore, the definition of d-stationarity can be equivalently written as $f'(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in X$. Recall that f_2 is (locally) weakly convex and thus regular, which implies that $f_2'(\bar{x}; x - \bar{x}) = \max_{s_{f_2} \in \partial^{\mathtt{C}} f_2(\bar{x})} \langle s_{f_2}, x - \bar{x} \rangle$. Hence,

$$\begin{aligned}
& f'(\bar{x}; x - \bar{x}) \geq 0 && \forall\, x \in X \\
\Leftrightarrow\ & f_1'(\bar{x}; x - \bar{x}) - f_2'(\bar{x}; x - \bar{x}) \geq 0 && \forall\, x \in X \\
\Leftrightarrow\ & f_1'(\bar{x}; x - \bar{x}) - \langle s_{f_2}, x - \bar{x} \rangle \geq 0 && \forall\, s_{f_2} \in \partial^{\mathtt{C}} f_2(\bar{x}),\ \forall\, x \in X \\
\Leftrightarrow\ & \bar{x} \in \operatorname*{arg\,min}_{x \in X} f_1(x) - \langle s_{f_2}, x - \bar{x} \rangle && \forall\, s_{f_2} \in \partial^{\mathtt{C}} f_2(\bar{x}).
\end{aligned}$$

□

A point $\bar{x} \in X$ is said to be Clarke-stationary to problem (3.3.1) if

$$0 \in \partial^{\mathtt{C}} f(\bar{x}) + N_X(\bar{x}). \tag{3.3.2}$$

Furthermore, by following the lead of DoC programming (see for instance [28, §3.1]), $\bar{x} \in X$ is said to be a critical point if

$$0 \in \partial f_1(\bar{x}) - \partial^{\mathtt{C}} f_2(\bar{x}) + N_X(\bar{x}). \tag{3.3.3}$$

It is not difficult to see that this inclusion means that

$$\bar{x} \in \operatorname*{arg\,min}_{x \in X} f_1(x) - [f_2(\bar{x}) + \langle s_{f_2}, x - \bar{x} \rangle] \quad \text{for some} \quad s_{f_2} \in \partial^{\mathtt{C}} f_2(\bar{x}).$$

Note that the concept of criticality is weaker than that of Clarke-stationarity, which in turn is weaker than d-stationarity (because $f'(\cdot;d) \leq f^{\circ}(\cdot;d)$ for all $d \in \mathbb{R}^n$). However, criticality and Clarke-stationarity coincide when at least one component function is smooth (in which case f is regular). Furthermore, we can see from Proposition 3.3.1 and the above alternative

characterization of criticality that the three concepts coincide when f_2 is continuously differentiable at $\bar{x}$.

For the more general problem (3.1.1), $\bar{x} \in X$ is said to be a Bouligand (B)-stationary point of (3.1.1) if $f'(\bar{x};d) \geq 0$ for all $d \in \mathscr{T}_{X^c}(\bar{x})$, with X^c the feasible set of (3.1.1). If the considered point strictly satisfies the nonconvex constraint, i.e. $c(\bar{x}) < 0$, then B-stationarity condition is equivalent to d-stationarity condition for problem (3.3.1), as $\mathscr{T}_{X^c}(\bar{x}) = \mathscr{T}_X(\bar{x})$. Analogously, B-stationarity boils down to d-stationarity if constraint $c(x) \leq 0$ is absent. Necessary and sufficient conditions for B-stationarity are given in [98, Prop. 4] for the case of DoC-constrained DoC problems. The next result deals with a more general case: we assume that c_2 is convex, while f_2 remains a weakly-convex function.

Proposition 3.3.2. *In addition to our assumptions on problem* (3.1.1)*, let $c_2 : \mathscr{O} \to \mathbb{R}$ be a convex function and $\bar{x} \in X^c := \{x \in X : c(x) \leq 0\}$ such that $c(\bar{x}) = 0$. Moreover, assume that the following constraint qualification (CQ) holds*

$$\mathtt{cl}\left\{d \in \mathscr{T}_X(\bar{x}) : c'(\bar{x};d) < 0\right\} = \left\{d \in \mathscr{T}_X(\bar{x}) : c'(\bar{x};d) \leq 0\right\}. \tag{3.3.4}$$

Then, $\bar{x}$ is a B-stationary point of problem (3.1.1) *if and only if $\bar{x}$ solves the convex problems*

$$\left\{ \begin{array}{ll} \min\limits_{x \in X} & f_1(x) - [f_2(\bar{x}) + \langle s_{f_2}, x - \bar{x}\rangle] \\ s.t. & c_1(x) - [c_2(\bar{x}) + \langle s_{c_2}, x - \bar{x}\rangle] \leq 0 \end{array} \right\} \quad \forall\, s_{f_2} \in \partial^{\mathtt{C}} f_2(\bar{x}),\ \forall\, s_{c_2} \in \partial c_2(\bar{x}). \tag{3.3.5}$$

Proof. Denote $\bar{Y}(\bar{x}) = \{x \in X : c_1(x) \leq c_2(\bar{x}) + c_2'(\bar{x};x-\bar{x})\}$. As the CQ (3.3.4) holds, Proposition 2.1 of [128] ensures that

$$\mathscr{T}_{X^c}(\bar{x}) = \mathscr{T}_{\bar{Y}(\bar{x})}(\bar{x}) = \mathtt{cl}\left\{d \in \mathbb{R}^n : d = t(x-\bar{x}), x \in \bar{Y}(\bar{x}), t \in \mathbb{R}_+\right\}.$$

Thus, the B-stationary definition is equivalent to

$$\begin{array}{ll} f'(\bar{x};x-\bar{x}) \geq 0 & \forall\, x \in \bar{Y}(\bar{x}) \\ \Leftrightarrow f_1'(\bar{x};x-\bar{x}) \geq \langle s_{f_2}, x-\bar{x}\rangle & \forall\, s_{f_2} \in \partial^{\mathtt{C}} f_2(\bar{x}),\ \forall\, x \in \bar{Y}(\bar{x}). \end{array} \tag{3.3.6}$$

The stated result follows upon establishing the equivalence between (3.3.6) and (3.3.5).

[(3.3.6) $\Rightarrow$ (3.3.5)]. Suppose (3.3.6) holds and let $s_{c_2} \in \partial c_2(\bar{x})$ be arbitrary. As

$$Y(s_{c_2}) := \{x \in X : c_1(x) \leq c_2(\bar{x}) + \langle s_{c_2}, x-\bar{x}\rangle\} \subset \bar{Y}(\bar{x})$$

due to convexity of c_2, we conclude that $f_1'(\bar{x};x-\bar{x}) \geq \langle s_{f_2}, x-\bar{x}\rangle$ for all $s_{f_2} \in \partial^{\mathsf{C}} f_2(\bar{x})$ and all $x \in Y(s_{c_2})$. Convexity of the latter set implies that $\bar{x}$ minimizes $f_1(x) - \langle s_{f_2}, x-\bar{x}\rangle$ over $Y(s_{c_2})$ for all $s_{f_2} \in \partial^{\mathsf{C}} f_2(\bar{x})$. Thus, condition (3.3.5) holds because $s_{c_2} \in \partial c_2(\bar{x})$ was taken arbitrarily.

[(3.3.5) $\Rightarrow$ (3.3.6)]. To show the reverse implication, we proceed with a proof by contrapositive. Suppose that there exist $s'_{f_2} \in \partial^{\mathsf{C}} f_2(\bar{x})$ and $x' \in \bar{Y}(\bar{x})$ such that $f_1'(\bar{x};x'-\bar{x}) < \langle s'_{f_2}, x'-\bar{x}\rangle$ (and hence $x' \neq \bar{x}$), i.e., (3.3.6) does not hold. Let $s'_{c_2} \in \partial c_2(\bar{x})$ be such that $c_2'(\bar{x};x'-\bar{x}) = \langle s'_{c_2}, x'-\bar{x}\rangle$. Therefore, x' is feasible for the convex problem

$$\begin{aligned} &\min_{x\in X} f_1(x) - [f_2(\bar{x}) + \langle s'_{f_2}, x-\bar{x}\rangle] \\ &\text{s.t. } c_1(x) - [c_2(\bar{x}) + \langle s'_{c_2}, x-\bar{x}\rangle] \leq 0. \end{aligned}$$

Together with our assumption $f_1'(\bar{x};x'-\bar{x}) < \langle s'_{f_2}, x'-\bar{x}\rangle$, we have that $d = x'-\bar{x}$ is a feasible descent direction for the above problem, and thus $\bar{x}$ cannot be one of its solution. Hence, $\bar{x}$ does not satisfy (3.3.5). The proof is thus complete. □

Note that convexity of c_2 plays an important role in the above proposition. Indeed, if c_2 is nonconvex, then the set $\{x \in X : c_1(x) \leq c_2(\bar{x}) + \langle s_{c_2}, x-\bar{x}\rangle\}$ is not necessarily a subset of X^c and when solving the linearized subproblem (3.3.5) we may get a point that is infeasible for the original problem (3.1.1).

Example 3.3.3. Let $f_1 = x$, $f_2 = 0$, $c_1 = 0$, $c_2 = \frac{x^3}{3}$ and $X = [-2,2]$. We are not in the framework of Proposition 3.3.2, since c_2 is not convex on $[-2,2]$, but weakly convex (with modulus $\mu = 4$). At $\bar{x} = 0$, which globally solves (3.1.1), the convex problem (3.3.5) becomes $\min_{x\in[-2,2]} x$ (because we have dropped the trivial constraint $0 \leq 0$), and thus does not provide a feasible point for the original problem. □

However, if the modulus μ is known for the weakly convex function c_2, adding a quadratic term in the constraint of the convex problem (3.3.5) makes the corresponding set feasible for the original problem. Moreover, Proposition 3.3.2 is generalized in case of weakly convex c_2.

Corollary 3.3.4. *Let $c_2 : \mathscr{O} \to \mathbb{R}$ be a weakly convex function and $\mu_c \geq 0$ be a real number from Proposition 3.2.4 corresponding to c_2. Moreover, assume that the CQ* (3.3.4) *holds. Then, $\bar{x}$ is a B-stationary point of problem* (3.1.1) *if and only if, for any given $\mu \geq \mu_c$, $\bar{x}$ solves*

the convex problems

$$\left\{ \begin{array}{ll} \min\limits_{x\in X} & f_1(x)-[f_2(\bar{x})+\langle s_{f_2},x-\bar{x}\rangle] \\ s.t. & c_1(x)-[c_2(\bar{x})+\langle s_{c_2},x-\bar{x}\rangle]+\frac{\mu}{2}\|x-\bar{x}\|^2\leq 0 \end{array} \right\} \quad \forall s_{f_2}\in\partial^{\mathtt{C}} f_2(\bar{x}),\ \forall s_{c_2}\in\partial^{\mathtt{C}} c_2(\bar{x}). \tag{3.3.7}$$

Proof. Consider the convex functions $\tilde{c}_1(x)=c_1(x)+\frac{\mu}{2}\|x\|^2$ and $\tilde{c}_2(x)=c_2(x)+\frac{\mu}{2}\|x\|^2$ with $\mu\geq\mu_c$. The result follows from Proposition 3.3.2 by using instead the DoC decomposition $c(x)=\tilde{c}_1(x)-\tilde{c}_2(x)$ and by noting that, for an arbitrary $\tilde{s}_{c_2}\in\partial\tilde{c}_2(\bar{x})$, we obtain

$$\tilde{c}_1(x)-[\tilde{c}_2(\bar{x})+\langle\tilde{s}_{c_2},x-\bar{x}\rangle]=c_1(x)-[c_2(\bar{x})+\langle s_{c_2},x-\bar{x}\rangle]+\frac{\mu}{2}\|x-\bar{x}\|^2$$

with $s_{c_2}=\tilde{s}_{c_2}-\mu\bar{x}$, $s_{c_2}\in\partial^{\mathtt{C}} c_2(\bar{x})$. □

Except for some particular cases, checking B-stationarity numerically is out of reach. Therefore, weaker stationarity concepts need to come into play: $\bar{x}\in X$ is said to be Clarke-stationary for (3.1.1) if there exists a Lagrange multiplier $\bar{\lambda}$ such that

$$\left\{ \begin{array}{l} 0\in\partial^{\mathtt{C}} f(\bar{x})+\bar{\lambda}\,\partial^{\mathtt{C}} c(\bar{x})+N_X(\bar{x}) \\ c(\bar{x})\leq 0,\ \bar{\lambda}\,c(\bar{x})=0,\ \bar{\lambda}\geq 0,\ \bar{x}\in X. \end{array} \right. \tag{3.3.8}$$

Analogously, $\bar{x}$ is a critical point of (3.1.1) if there exists a Lagrange multiplier $\bar{\lambda}$ such that

$$\left\{ \begin{array}{l} 0\in\partial f_1(\bar{x})-\partial^{\mathtt{C}} f_2(\bar{x})+\bar{\lambda}\,[\partial c_1(\bar{x})-\partial^{\mathtt{C}} c_2(\bar{x})]+N_X(\bar{x}) \\ c(\bar{x})\leq 0,\ \bar{\lambda}\,c(\bar{x})=0,\ \bar{\lambda}\geq 0,\ \bar{x}\in X. \end{array} \right. \tag{3.3.9}$$

Observe that if f_1 or f_2 and c_1 or c_2 are smooth, then criticality boils down to Clarke stationarity. Next, we revisit the proximal bundle method of [129] and extend it to the more general setting of problem (3.1.1). To this end, the method must be modified, and its convergence analysis must be done anew.

3.3.1 Problem reformulation via improvement function

Nonsmooth and noncovex constraints in optimization problems are in general numerically treated via exact penalization [73, 81, 118, 66], linearization of certain components [98, 140], and improvement functions [110, 7, 129]. The latter has a recognized good practical performance, does not require the additional assumptions normally assumed in exact penalization methods, and employs parameters that are simple to set. For these reasons, we handle

problem (3.1.1) via the *improvement function* $H : \mathscr{O} \times \mathscr{O} \to \mathbb{R}$ given by

$$H(x;y) = \max\left\{f(x) - \tau_f(y),\ c(x) - \tau_c(y)\right\}, \tag{3.3.10a}$$

$$\text{with} \quad \tau_f(y) = f(y) + \rho[c(y)]_+ \quad \text{and} \quad \tau_c(y) = \sigma[c(y)]_+, \quad \text{for } \rho \geq 0 \text{ and } \sigma \in [0,1). \tag{3.3.10b}$$

Observe that if $\bar{x}$ is a global solution of (3.1.1), then $H(x;\bar{x}) \geq 0$ for all $x \in X$ and $H(\bar{x};\bar{x}) = 0$.

Improvement functions (also known as progress functions) have been considered within bundle methods in [110, 138, 127] for convex problems, in [7] for a class of (nonconvex) optimal control problems, and in [90, 129] for DoC-constrained DoC programs. In what follows we exploit some relevant mathematical properties of (3.3.10) and its link to the original problem (3.1.1). To this end, we need to consider necessary conditions for a point $\bar{x}$ to be a local solution of the reformulated problem

$$\min_{x \in X} H(x;\bar{x})\,. \tag{3.3.11}$$

As the second argument of H is fixed, it follows from (3.3.2) that $\bar{x} \in X$ is a Clarke-stationary point of (3.3.11) if

$$0 \in \partial_1^{\mathtt{C}} H(\bar{x};\bar{x}) + N_X(\bar{x}), \tag{3.3.12}$$

where $\partial_1^{\mathtt{C}} H$ stands for the generalized subdifferential of H with respect to the first argument. Proposition 3.2.1 ii) yields

$$\partial_1^{\mathtt{C}} H(\bar{x};\bar{x}) \subset \begin{cases} \partial^{\mathtt{C}} c(\bar{x}) & \text{if } f(\bar{x}) - \tau_f(\bar{x}) < c(\bar{x}) - \tau_c(\bar{x}) \\ \mathtt{conv}\left\{\partial^{\mathtt{C}} f(\bar{x}), \partial^{\mathtt{C}} c(\bar{x})\right\} & \text{if } f(\bar{x}) - \tau_f(\bar{x}) = c(\bar{x}) - \tau_c(\bar{x}) \\ \partial^{\mathtt{C}} f(\bar{x}) & \text{if } f(\bar{x}) - \tau_f(\bar{x}) > c(\bar{x}) - \tau_c(\bar{x}). \end{cases}$$

Since we do not work with generalized subgradients of either f or c, but only with subgradients of the functions yielding their CwC decompositions (3.1.1b), we must consider a weaker stationary definition: we say that $\bar{x} \in X$ is a critical point of the composite problem (3.3.11) with CwC decompositions (3.1.1b) if

$$0 \in N_X(\bar{x}) + \begin{cases} \partial c_1(\bar{x}) - \partial^{\mathtt{C}} c_2(\bar{x}) & \text{if } f(\bar{x}) - \tau_f(\bar{x}) < c(\bar{x}) - \tau_c(\bar{x}) \\ \mathtt{conv}\left\{\partial f_1(\bar{x}) - \partial^{\mathtt{C}} f_2(\bar{x}), \partial c_1(\bar{x}) - \partial^{\mathtt{C}} c_2(\bar{x})\right\} & \text{if } f(\bar{x}) - \tau_f(\bar{x}) = c(\bar{x}) - \tau_c(\bar{x}) \\ \partial f_1(\bar{x}) - \partial^{\mathtt{C}} f_2(\bar{x}) & \text{if } f(\bar{x}) - \tau_f(\bar{x}) > c(\bar{x}) - \tau_c(\bar{x}). \end{cases} \tag{3.3.13}$$

Note that if both f and c are regular, then the above condition coincides with that of Clarke stationarity: recall (3.3.2), Proposition 3.2.1 i), and observe that the set defined by the expressions in the curly brackets above is nothing but $\partial_1^{\mathtt{C}} H(\bar{x};\bar{x})$. The following result,

inspired by both [7, Lemma 5.1] that deals with the (stronger) Clarke stationarity and [129, Thm. 2] that works with the (weaker) criticality definition from DoC programming, links condition (3.3.13) with criticality of the original problem.

Theorem 3.3.5. *Let $\bar{x} \in X$ be a point satisfying condition* (3.3.13)*. Then, the following hold:*

i) If $c(\bar{x}) > 0$, then $\bar{x}$ is a critical point (in the sense of (3.3.3)*) of the optimization problem*

$$\min_{x \in X} \; c_1(x) - c_2(x). \tag{3.3.14}$$

ii) If $c(\bar{x}) = 0$ and $\bar{x}$ is not a critical point of (3.3.14)*, then $\bar{x}$ satisfies* (3.3.9) *for some $\bar{\lambda} > 0$.*

iii) If $c(\bar{x}) < 0$, then $\bar{x}$ satisfies (3.3.9) *with $\bar{\lambda} = 0$.*

Proof. With the τ function defined in (3.3.10b), note that

$$f(\bar{x}) - \tau_f(\bar{x}) - [c(\bar{x}) - \tau_c(\bar{x})] = \begin{cases} -[\rho + (1-\sigma)]c(\bar{x}) < 0 & \text{if } c(\bar{x}) > 0, \\ 0 & \text{if } c(\bar{x}) = 0, \\ -c(\bar{x}) > 0 & \text{if } c(\bar{x}) < 0. \end{cases}$$

Hence,

$$\begin{aligned} c(\bar{x}) > 0 \quad &\Leftrightarrow \quad f(\bar{x}) - \tau_f(\bar{x}) < c(\bar{x}) - \tau_c(\bar{x}), \\ c(\bar{x}) = 0 \quad &\Leftrightarrow \quad f(\bar{x}) - \tau_f(\bar{x}) = c(\bar{x}) - \tau_c(\bar{x}), \\ c(\bar{x}) < 0 \quad &\Leftrightarrow \quad f(\bar{x}) - \tau_f(\bar{x}) > c(\bar{x}) - \tau_c(\bar{x}), \end{aligned}$$

and items i) and iii) follow directly from (3.3.13). To show item ii), recall that $c(\bar{x}) = 0$ and condition (3.3.13) ensures the existence of $\lambda \in [0,1]$ such that

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

By assumption, $\bar{x}$ is not a critical point of (3.3.14). Then $0 \notin \partial c_1(\bar{x}) - \partial^{\mathsf{C}} c_2(\bar{x}) + N_X(\bar{x})$, implying that λ above must be strictly positive. Dividing the displayed inclusion by $\lambda > 0$ we obtain the criticality condition (3.3.9) with $\bar{\lambda} = (1-\lambda)/\lambda > 0$. □

At item ii) above, the assumption that $\bar{x}$ is not a critical point of (3.3.14) can be seen as a constraint qualification, which turns out to be more restrictive than (3.3.4). Indeed, the latter excludes d-stationary points of (3.3.14), but not necessarily critical ones. The following example gives a critical point $\bar{x}$ of (3.3.14) that satisfies (3.3.4) but not the criticality condition (3.3.9) for the nonlinearly-constrained problem (3.1.1).

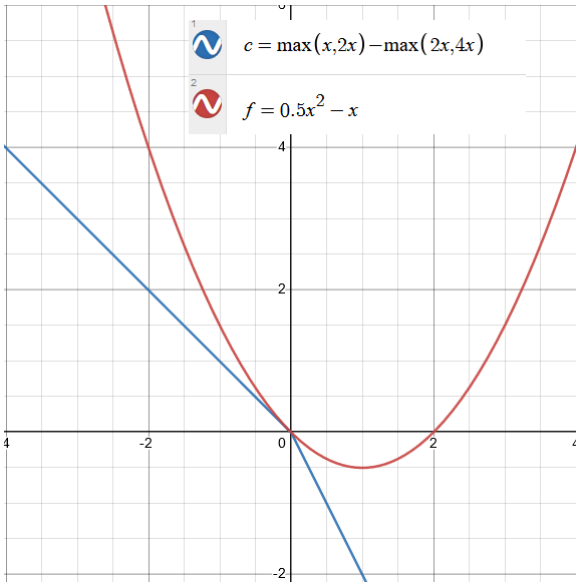


Fig. 3.1 Function $f(x) = \frac{1}{2}x^2 - x$ in red and $c(x) = \max\{x, 2x\} - \max\{2x, 4x\}$ in blue.

Example 3.3.6. Take $c_1(x) = \max\{x, 2x\}$, $c_2(x) = \max\{2x, 4x\}$, $X = [-2, 2]$ and $\bar{x} = 0$. Then $\mathscr{T}_X(\bar{x}) = \mathbb{R}$, $N_X(\bar{x}) = \{0\}$, and $\bar{x}$ is a critical point of (3.3.14) because $0 \in \partial c_1(\bar{x}) - \partial c_2(\bar{x}) = [1, 2] - [2, 4] = [-3, 0]$. Furthermore, note that

$$c_1'(\bar{x}; d) = \max_{s \in [1,2]} sd = \begin{cases} 2d & \text{if } d \geq 0 \\ d & \text{if } d \leq 0 \end{cases} \quad \text{and} \quad c_2'(\bar{x}; d) = \max_{s \in [2,4]} sd = \begin{cases} 4d & \text{if } d \geq 0 \\ 2d & \text{if } d \leq 0, \end{cases}$$

thus $c'(\bar{x}; d) = \min\{-d, -2d\}$. We conclude that $\{d \in \mathscr{T}_X(\bar{x}) : c'(\bar{x}; d) < 0\} = \mathbb{R}_+$, whereas $\{d \in \mathscr{T}_X(\bar{x}) : c'(\bar{x}; d) \leq 0\} = \mathbb{R}_+ \cup \{0\}$, showing that $\bar{x} = 0$ satisfies the CQ (3.3.4). However, if we take $f_1(x) = 0$ and $f_2(x) = -\frac{1}{2}x^2 + x$, the following system does not have a solution:

$$\begin{cases} 0 & \in \partial f_1(0) - \partial^{\mathrm{C}} f_2(0) + \bar{\lambda}[\partial c_1(0) - \partial c_2(0)] \\ \bar{\lambda} & \geq 0 \end{cases} \equiv$$

$$\begin{cases} 0 & \in -1 + \bar{\lambda}[-3, 0] \\ \bar{\lambda} & \geq 0 \end{cases} \equiv \begin{cases} 0 & \in [-3\bar{\lambda} - 1, -1] \\ \bar{\lambda} & \geq 0, \end{cases}$$

i.e., $\bar{x}$ does not satisfy (3.3.9). Figure 3.1 illustrates the objective and constraint function in this example: it is clear that $\bar{x}$ is indeed a global maximizer of $f(x)$ under the constraints $x \in X$ and $c(x) \leq 0$.

□

This example shows that, at item ii) of Theorem 3.3.5, we cannot replace the assumption that $\bar{x}$ is not a critical point of (3.3.14) with the CQ (3.3.4).

3.3.2 The DoC setting

In the DoC setting, functions f_2 and c_2 are convex and the improvement function (3.3.10) is DoC. Indeed, for $\bar{x}$ fixed, we can write

$$H(x;\bar{x}) = F(x;\bar{x}) - G(x), \quad \text{with}$$
$$\begin{cases} F(x;\bar{x}) & = \max\left\{f_1(x) + c_2(x) - \tau_f(\bar{x}),\, f_2(x) + c_1(x) - \tau_c(\bar{x})\right\}, \\ G(x) & = f_2(x) + c_2(x). \end{cases} \tag{3.3.15}$$

Since F and G are convex functions, the criticality condition (3.3.3) for (3.3.11) (under this DoC decomposition) reads as

$$0 \in \partial_1 F(\bar{x};\bar{x}) - \partial G(\bar{x}) + N_X(\bar{x}), \tag{3.3.16}$$

where $\partial_1 F$ stands for the subdifferential of F with respect to the first argument. It turns out that our new condition (3.3.13) is stronger than (3.3.16), used in [129].

Lemma 3.3.7. *In addition to our basic assumptions on problem* (3.1.1)*, suppose that f_2 and c_2 are convex. Then the necessary optimality condition* (3.3.13) *implies* (3.3.16)*.*

Proof. Let $\bar{x} \in X$ be a point satisfying (3.3.13). Let us first observe that since f_2, c_2 are convex and thus regular, we have $\partial G(\bar{x}) = \partial f_2(\bar{x}) + \partial c_2(\bar{x})$. A similar observation can be made concerning the computation for F. Our analysis splits into three possible cases.

a) $f(\bar{x}) - \tau_f(\bar{x}) < c(\bar{x}) - \tau_c(\bar{x})$. It follows from (3.3.13) that

$$0 \in N_X(\bar{x}) + \partial c_1(\bar{x}) - \partial c_2(\bar{x}) \subset N_X(\bar{x}) + \partial f_2(\bar{x}) + \partial c_1(\bar{x}) - [\partial f_2(\bar{x}) + \partial c_2(\bar{x})].$$

We claim that this inclusion implies (3.3.16). To see that, observe that the above inequality implies $f_1(\bar{x}) + c_2(\bar{x}) - \tau_f(\bar{x}) < f_2(\bar{x}) + c_1(\bar{x}) - \tau_c(\bar{x})$, which in turn gives $\partial_1 F(\bar{x};\bar{x}) = \partial f_2(\bar{x}) + \partial c_1(\bar{x})$. Therefore, the right-hand side of the above inclusion is (3.3.16).

b) $f(\bar{x}) - \tau_f(\bar{x}) = c(\bar{x}) - \tau_c(\bar{x})$. It follows from (3.3.13) that there exists $\lambda \in [0,1]$ such that

$$\begin{aligned} 0 &\in N_X(\bar{x}) + \lambda[\partial f_1(\bar{x}) - \partial f_2(\bar{x})] + (1-\lambda)[\partial c_1(\bar{x}) - \partial c_2(\bar{x})] \\ &= N_X(\bar{x}) + \lambda[\partial f_1(\bar{x}) + \partial c_2(\bar{x})] - \lambda\partial f_2(\bar{x}) + (1-\lambda)\partial c_1(\bar{x}) - \partial c_2(\bar{x}) \\ &\subset N_X(\bar{x}) + \lambda[\partial f_1(\bar{x}) + \partial c_2(\bar{x})] + (1-\lambda)[\partial f_2(\bar{x}) + \partial c_1(\bar{x})] - [\partial f_2(\bar{x}) + \partial c_2(\bar{x})] \\ &\subset N_X(\bar{x}) + \partial_1 F(\bar{x};\bar{x}) - \partial G(\bar{x}). \end{aligned}$$

c) $f(\bar{x}) - \tau_f(\bar{x}) > c(\bar{x}) - \tau_c(\bar{x})$. Again, (3.3.13) gives

$$0 \in N_X(\bar{x}) + \partial f_1(\bar{x}) - \partial f_2(\bar{x}) \subset N_X(\bar{x}) + \partial f_1(\bar{x}) + \partial c_2(\bar{x}) - [\partial f_2(\bar{x}) + \partial c_2(\bar{x})].$$

The proof is complete because in this case $\partial_1 F(\bar{x};\bar{x}) = \partial f_1(\bar{x}) + \partial c_2(\bar{x})$ due to the fact that $f_1(\bar{x}) + c_2(\bar{x}) - \tau_f(\bar{x}) > f_2(\bar{x}) + c_1(\bar{x}) - \tau_c(\bar{x})$.

□

Remark 3.3.8. In the DoC setting, the three concepts of criticality (3.3.12), (3.3.13), and (3.3.16) are equivalent when f_2 and c_2 are continuously differentiable at $\bar{x}$. Indeed, in this case f and c are regular at $\bar{x}$ and (3.3.12) coincides with (3.3.13) (regardless of convexity of f_2 and c_2). Theorem 2 in [129] ensures that, under these assumptions, (3.3.16) is equivalent to (3.3.12). □

The following example shows that (3.3.16) does not necessarily imply (3.3.13) in the nondifferentiable DoC case.

Example 3.3.9. Let $X = [-1,1]$, $f_1 = 2x$, $f_2 = |x|$, $c_1 = 4x$ and $c_2 = 2|x|$. At $\bar{x} = 0$, $f(\bar{x}) = c(\bar{x}) = 0$ and thus, $\tau_f(\bar{x}) = \tau_c(\bar{x}) = 0$ due to (3.3.10b). Furthermore, we have that

$$\partial f_1(0) = \{2\}, \quad \partial f_2(0) = [-1,1], \quad \partial c_1(0) = \{4\}, \text{ and } \partial c_2(0) = [-2,2].$$

As a result, $\partial f_1(0) - \partial f_2(0) = [1,3]$, $\partial c_1(0) - \partial c_2(0) = [2,6]$, $\partial f_1(0) + \partial c_2(0) = [0,4]$, and $\partial f_2(0) + \partial c_1(0) = [3,5]$. As in $N_X(0) = \{0\}$, we conclude that

$$0 \notin [1,6] = \texttt{conv}\{\partial f_1(0) - \partial f_2(0), \partial c_1(0) - \partial c_2(0)\} + N_X(0),$$

whereas

$$\texttt{conv}\{\partial f_1(0) + \partial c_2(0), \partial c_1(0) + \partial f_2(0)\} + N_X(0) = [0,5]$$

and $\partial f_2(0) + \partial c_2(0) = [-3,3]$, showing that (3.3.16) is satisfied but not (3.3.13). □

The paper [129] proposes a bundle method for DoC-constrained DoC programs employing the DoC decomposition $H = F - G$ above. Once a critical point satisfying (3.3.16) is computed, the link with criticality of the original problem is adequate if f_2 and c_2 are continuously differentiable at $\bar{x}$. In the next section we modify that method to compute a point satisfying the stronger criticality condition (3.3.13). As a result, the link with criticality of the original problem is nicely established by Theorem 3.3.5 without any additional assumption. In fact, f_2 and c_2 need not even be convex, but weakly convex on some neighbourhood of each $x \in \mathscr{O}$. We, therefore, strengthen the analysis provided in [129] even though significantly

fewer assumptions are required: [129] works in the DoC configuration, whereas here, we deal with the more general CwC structure. These improvements, together with the optimality conditions presented above, feature the main contributions of this chapter.

3.4 Proximal bundle method with improvement function

This section extends the proximal bundle method of [129] for computing a critical point of problem (3.1.1). The main tool in our analysis is the improvement function H given in (3.3.10). In the DoC setting, the algorithm of [129] works with the explicit DoC decomposition (3.3.2) of H and computes a point $\bar{x} \in X$ satisfying the classic criticality condition in DoC programming (3.3.16). In this section we do not decompose H and consider the milder assumption that f_2 and c_2 are weakly convex and target the stronger criticality condition (3.3.13).

3.4.1 The method's main ingredients: model, subproblem, and descent test

The algorithm requires four oracles (black-boxes) providing, for every given $x \in X$, $i \in \{1,2\}$, the function values $f_i(x)$, $c_i(x)$, arbitrary subgradients $s_{f_1} \in \partial f_1(x)$, $s_{c_1} \in \partial c_1(x)$ (c.f., (3.2.1)) and arbitrary generalized subgradients $s_{f_2} \in \partial^{\mathtt{C}} f_2(x)$, $s_{c_2} \in \partial^{\mathtt{C}} c_2(x)$ (c.f. (3.2.2)). We do not impose any assumption on these (generalized) subgradients, as they are assumed to be computed by (external) oracles that do not accept any intervention from the algorithm. (This is particularly useful in industrial applications where companies do not want or cannot share information on the underlying functions with optimizers.)

At iteration $k \in \mathbb{N}$, given a trial point $x^k \in X$, we construct a linearization of every component (here $s^k_{f_i}$, $s^k_{c_i}$, $i \in \{1,2\}$, denote the respective - generalized - subgradients at x^k):

$$\bar{f}^k_i(x) := f_i(x^k) + \langle s^k_{f_i}, x - x^k \rangle \quad (i = 1,2) \tag{3.4.1a}$$

$$\bar{c}^k_i(x) := c_i(x^k) + \langle s^k_{c_i}, x - x^k \rangle \quad (i = 1,2). \tag{3.4.1b}$$

By convexity of f_1 and c_1, we have the following inequalities

$$\bar{f}^k_1(x) \le f_1(x) \quad \text{and} \quad \bar{c}^k_1(x) \le c_1(x) \quad \text{for all } x \in \mathbb{R}^n. \tag{3.4.2}$$

Since X is compact and components f_2 and c_2 are assumed to be only weakly convex on some neighbourhood of each $x \in \mathscr{O}$, we have weaker inequalities for these functions. Let

$\mathscr{O}' \subset O$ be an open convex set and μ_f, μ_c real numbers ensured by Proposition 3.2.4. As $x^k \in X \subset O'$, the following inequalities are due to Proposition 3.2.4, item ii)

$$\bar{f}_2^k(x) \leq f_2(x) + \tfrac{\bar{\mu}}{2}\|x - x^k\|^2 \quad \text{and} \quad \bar{c}_2^k(x) \leq c_2(x) + \tfrac{\bar{\mu}}{2}\|x - x^k\|^2 \quad \text{for all } x \in X, \quad (3.4.3)$$

where $\bar{\mu} := \max\{\mu_f, \mu_c\} > 0$. Observe that the threshold $\bar{\mu}$ is in general unknown, and the inequalities in (3.4.3) are only supposed to hold for x in X, in contrast with the (subgradient) inequalities in (3.4.2).

Let $\mathscr{B}_f^k$ and $\mathscr{B}_c^k$ be two index sets gathering the *bundle* of information (function values and subgradients) given by the oracles. In general, $\mathscr{B}_f^k, \mathscr{B}_c^k \subset \{0, \ldots, k\}$ but other possibilities exist making it possible to design a limited-memory method (see Remark 3.5.5 below). These index sets are useful to define the following individual cutting-plane models for the convex functions f_1 and c_1:

$$\begin{aligned} \check{f}_1^k(x) := \max_{j \in \mathscr{B}_f^k} \bar{f}_1^j(x) \quad &\leq \quad f_1(x) \quad \text{for all } x \in \mathbb{R}^n \\ \check{c}_1^k(x) := \max_{j \in \mathscr{B}_c^k} \bar{c}_1^j(x) \quad &\leq \quad c_1(x) \quad \text{for all } x \in \mathbb{R}^n. \end{aligned}$$

Furthermore, let $\ell_k \in \{0, \ldots, k\}$ be the iteration index of the best candidate solution (*stability center*, in the parlance of bundle methods) among the trial points $\{x^0, \ldots, x^k\}$: whenever a better candidate solution x^{k+1} is computed by the algorithm, at a so-called *serious step*, such a point becomes the new stability center and the counter ℓ is increased by one: for $\kappa \in (0, \frac{1}{2})$, we declare a *serious step* and let $\ell_{k+1} := k+1$ if $x^{k+1} \neq x^{\ell_k}$ and the inequality

$$H(x^{k+1}; x^{\ell_k}) \leq H(x^{\ell_k}; x^{\ell_k}) - \frac{\kappa}{2}\|x^{k+1} - x^{\ell_k}\|^2 \quad (3.4.5)$$

holds, and declare a *null step* and let $\ell_{k+1} := \ell_k$ otherwise. Since the descent test is independent of the model, the following result from [129] also holds in our framework.

Lemma 3.4.1 (Lemma 1 in [129])**.** *Let $x^{\ell_k} \in X$ be the stability center at iteration k. Then $H(x^{\ell_k}; x^{\ell_k}) \geq 0$ and if inequality* (3.4.5) *holds, we have that either*

i) $f(x^{k+1}) \leq f(x^{\ell_k}) - \frac{\kappa}{2}\|x^{k+1} - x^{\ell_k}\|^2$ *and* $c(x^{k+1}) \leq 0$ *when* $c(x^{\ell_k}) \leq 0$*; or*

ii) $c(x^{k+1}) \leq c(x^{\ell_k}) - \frac{\kappa}{2}\|x^{k+1} - x^{\ell_k}\|^2$ *when* $c(x^{\ell_k}) > 0$.

The rationale of serious iterates is to ensure sufficient decrease on one component function of $H(\cdot; x^{\ell_k})$ while maintaining feasibility for (3.1.1) once reached. Having all these ingredients at our disposal, we can now define our convex model for the improvement function (3.3.10)

at iteration k :

$$\check{\mathsf{H}}^k(x;x^{\ell_k}) = \max\left\{\check{f}_1^k(x) - \bar{f}_2^{\ell_k}(x) - \tau_f(x^{\ell_k}),\ \check{c}_1^k(x) - \bar{c}_2^{\ell_k}(x) - \tau_c(x^{\ell_k})\right\}. \tag{3.4.6}$$

(Even in the particular setting where f_2 and c_2 are convex functions, this model differs from the one employed in [129] and is crucial to obtain convergence results stronger than the ones in that paper.) Given a prox-parameter $\mu^k > 0$ estimating the threshold $\bar{\mu}$ in (3.4.3), the next iterate is the solution of the following strict convex subproblem

$$x^{k+1} = \operatorname*{arg\,min}_{x\in X}\ \check{\mathsf{H}}^k(x;x^{\ell_k}) + \frac{\mu^k}{2}\|x - x^{\ell_k}\|^2, \tag{3.4.7}$$

which can be transformed into a QP (provided X is a polyhedron) by adding an extra variable $r \in \mathbb{R}$

$$\left\{\begin{array}{lll} \min\limits_{x,r} & r + \frac{\mu^k}{2}\|x - x^{\ell_k}\|^2 & \\ \text{s.t.} & \bar{f}_1^j(x) - \bar{f}_2^{\ell_k}(x) - r & \le \tau_f(x^{\ell_k}) \quad \forall\, j \in \mathscr{B}_f^k \\ & \bar{c}_1^j(x) - \bar{c}_2^{\ell_k}(x) - r & \le \tau_c(x^{\ell_k}) \quad \forall\, j \in \mathscr{B}_c^k \\ & x \in X,\ \ r \in \mathbb{R}. & \end{array}\right. \tag{3.4.8}$$

The optimality condition for (3.4.7) gives

$$x^{k+1} = x^{\ell_k} - \frac{1}{\mu^k}[p^{k+1} + s_X^{k+1}], \quad \text{with} \quad \left\{\begin{array}{l} p^{k+1} \in \partial_1 \check{\mathsf{H}}^k(x^{k+1};x^{\ell_k}) \\ s_X^{k+1} \in N_X(x^{k+1}). \end{array}\right. \tag{3.4.9}$$

As usual in bundle methods, we may remove from the model the inactive linearizations to keep (3.4.8) small. To this end, we denote by $\bar{\mathscr{B}}_f^k \subset \mathscr{B}_f^k$ and $\bar{\mathscr{B}}_c^k \subset \mathscr{B}_c^k$ the index set of active linearizations in the QP subproblem (3.4.8), i.e.,

$$\bar{\mathscr{B}}_f^k := \left\{j \in \mathscr{B}_f^k :\ \alpha_f^j > 0\right\} \quad \text{and} \quad \bar{\mathscr{B}}_c^k := \left\{j \in \mathscr{B}_c^k :\ \alpha_c^j > 0\right\} \tag{3.4.10}$$

where $\alpha_f^j \ge 0,\ j \in \mathscr{B}_f^k$, denote the Lagrange multipliers associated with the first set of constraints and $\alpha_c^j \ge 0,\ j \in \mathscr{B}_c^k$, the ones associated with the second family of constraints. We mention in passing that the index sets $\mathscr{B}_f^k$ and $\mathscr{B}_c^k$ can be kept bounded at the price of including artificial (aggregate) linearizations. We postpone this discussion to Remark 3.5.5, right after the analysis of null steps (the only place in the convergence analysis where bundle management plays a role.)

We can now present the following proximal bundle method algorithm for CwC-constrained CwC programs (3.1.1), which modifies [129, Alg. 1] in two ways. First, the convex model

(3.4.6) for the improvement function is distinct. On the one hand, it is a key element to obtain the stronger criticality condition (3.3.13), and on the other hand, it leads to a simpler/smaller strongly convex QP (3.4.8) (more details can be found in Subsection 3.4.2 below). Second, Algorithm 2 employs an ad-hoc rule to update the proximal parameter μ^k so that no pre-estimation of the underlying weakly-convex moduli is needed. The proposed rule employs the following value

$$\nu^k := 2\max\left\{\frac{\bar{f}_2^{\ell_k}(x^{k+1}) - f_2(x^{k+1})}{\|x^{k+1}-x^{\ell_k}\|^2}, \frac{\bar{c}_2^{\ell_k}(x^{k+1}) - c_2(x^{k+1})}{\|x^{k+1}-x^{\ell_k}\|^2}, 0\right\}. \tag{3.4.11}$$

(For numerical performance it can be preferable to replace the last term (equal to zero) in (3.4.11) by a small machine epsilon.)

Algorithm 2 Proximal Bundle Method for `CwC`-constrained `CwC` programs - `CwC-PBM`

Step 0 (Initialization) Let $x^0 \in X$, $\kappa \in (0, \frac{1}{2})$, $\kappa \leq \mu^0$, $\rho \geq 0$, $\sigma \in [0,1)$, and $\texttt{Tol} \geq 0$ be given. Call the oracles to compute $f_i(x^0)$, $c_i(x^0)$, and (generalized) subgradients $s_{f_i}^0$, $s_{c_i}^0$, $i = 1,2$. Define $k := \ell_k = 0$ and $\mathscr{B}_f^0 = \mathscr{B}_c^0 := \{0\}$.

Step 1 (Trial point) Compute x^{k+1} by solving the QP (3.4.8).

Step 2 (Stopping test) If $\|x^{k+1} - x^{\ell_k}\| \leq \texttt{Tol}$, then stop and return x^{ℓ_k}.

Step 3 (Oracles call) Compute $f_i(x^{k+1})$, $c_i(x^{k+1})$, and subgradients $s_{f_i}^{k+1}$, $s_{c_i}^{k+1}$, $i = 1,2$.

Step 4 (Descent test)

(a) If (3.4.5) holds, then declare a *serious step*: define $\ell_{k+1} := k+1$, choose $\mathscr{B}_f^{k+1}, \mathscr{B}_c^{k+1} \subset \{0,\ldots,k+1\}$ with $\{k+1\} \in \mathscr{B}_f^{k+1} \cap \mathscr{B}_c^{k+1}$ and arbitrarily select $\mu^{k+1} \in (0, \mu^k]$.

(b) Else, declare a *null step*: define $\ell_{k+1} := \ell_k$ and choose $\mathscr{B}_f^{k+1}, \mathscr{B}_c^{k+1} \subset \{0,\ldots,k+1\}$ with $\bar{\mathscr{B}}_f^k \cup \{k+1, \ell_k\} \subset \mathscr{B}_f^{k+1}$ and $\bar{\mathscr{B}}_c^k \cup \{k+1, \ell_k\} \subset \mathscr{B}_c^{k+1}$ ($\bar{\mathscr{B}}_f^k$ and $\bar{\mathscr{B}}_c^k$ as in (3.4.10)).
Compute ν^k by (3.4.11). If $\nu^k \geq \mu^k - 2\kappa$, set $\mu^{k+1} = \nu^k + 1$; otherwise $\mu^{k+1} = \mu^k$.

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

A drawback of the rule for updating the prox-parameter is that μ^k only increases after a null step when the inequality $\nu^k \geq \mu^k - 2\kappa$ is verified. As a result, μ^k may never increase: this is, for instance, the case when f_2 and c_2 are convex (thus $\nu^k = 0$ for all k). The motivation for this rule is to eventually keep the prox-parameter fixed if the algorithm performs an infinite sequence of null steps after a last serious step (see Lemma 3.5.1). This is a condition necessary to prove Proposition 3.5.4 below. We care to mention that increasing μ^k after a

null step is a simple strategy that pays off in practice: it helps the algorithm to either stop or produce a new serious step, and thus accelerate the numerical performance.

3.4.2 The DoC setting: a comparison with the earlier bundle method for DoC programs

In the DoC setting, both functions f_2 and c_2 are convex and the improvement function (3.3.10) is DoC. The DoC decomposition given in (3.3.2), with $\bar{x}$ replaced with x^{ℓ_k}, was exploited in the bundle method of [129] through the following model for the improvement function $H(\cdot; x^{\ell_k})$ (see Eq. (18) therein):

$$\max\left\{\check{f}_1^k(x)+\check{c}_2^k(x)-\tau_f(x^{\ell_k}),\ \check{f}_2^k(x)+\check{c}_1^k(x)-\tau_c(x^{\ell_k})\right\}-[\bar{f}_2^{\ell_k}(x)+\bar{c}_2^{\ell_k}(x)]. \tag{3.4.12}$$

Differently from our model (3.4.6), the above gathers also cutting-planes for f_2 and c_2 and, although gathering more information, only the weaker criticality condition (3.3.16) is ensured by the method of [129]. Hence, the proposed model (3.4.6) is more advantageous than (3.4.12) from both practical and theoretical point of view:

- the quadratic program (QP) issued by our model has only half of the linearizations, and is thus simpler to solve;
- convexity of f_2 and c_2 are not required in (3.4.6) in contrast to (3.4.12);
- both models (3.4.6) and (3.4.12) are iteratively updated to ensure that every cluster point $\bar{x} \in X$ of the sequence of stability centers satisfies a criticality condition. To show that such a point is also critical for (the DoC counterpart of) (3.1.1), [129, Thm. 2] requires both f_2 and c_2 to be continuously differentiable at $\bar{x}$. As we will show in Theorem 3.5.9 below, neither convexity nor differentiability of f_2 and c_2 are required to establish that $\bar{x}$ issued by Algorithm 2 is also critical for (3.1.1) in the sense of (3.3.9). Thus, Algorithm 2 strengthens the results of [129] even though significantly fewer assumptions are required.

Although the apparently small changes concerning [129, Alg. 1], the convergence analysis in that paper cannot be reused here. The reason is that the analysis in [129] strongly depends on the DoC decomposition of the employed model for the improvement function. That reasoning is no longer valid for our new model, even if f_2 and c_2 were convex. Furthermore, our more general setting requires extra steps to cope with the weakly convex functions.

3.5 Convergence Analysis

The goal of this section is to show that every cluster point $\bar{x}$ of the sequence $\{x^{\ell_k}\}_k \subset X$ generated by Algorithm 2 satisfies the necessary optimality condition (3.3.13). To this end, we first observe that the sequence of prox-parameters issued by Algorithm 2 is bounded.

Lemma 3.5.1. *The value $\mu_{\max} := \sup_{k\in\mathbb{N}} \mu^k$ is finite. Furthermore, if the algorithm produces an infinite sequence of null steps after a last serious step, then the prox-parameter becomes eventually constant.*

Proof. Let $\bar{\mu} := \max\{\mu_{f_2}, \mu_{c_2}, \mu^0\} > 0$ be given, where μ_{f_2} and μ_{c_2} are as in Proposition 3.2.4 for the weakly convex functions f_2 and c_2, and μ^0 is the parameter given to the algorithm at initialization. Then, by taking $y := x^{k+1}$ and $x := x^{\ell_k}$ in (3.2.4) it follows that

$$2\frac{\bar{f}_2^{\ell_k}(x^{k+1}) - f_2(x^{k+1})}{\|x^{k+1} - x^{\ell_k}\|^2} \leq \bar{\mu}, \quad \text{and} \quad 2\frac{\bar{c}_2^{\ell_k}(x^{k+1}) - c_2(x^{k+1})}{\|x^{k+1} - x^{\ell_k}\|^2} \leq \bar{\mu} \quad \text{for all } k \text{ with } x^{k+1} \neq x^{\ell_k}.$$

As a result, $v^k \leq \bar{\mu}$ for all k. Note that the prox-parameter is only increased after a null step such that $v^k \geq \mu^k - 2\kappa$. In this case, the rule employed in Step 4 of the algorithm sets $\mu^{k+1} = v^k + 1$, which gives $\mu^{k+1} = v^k + 1 \leq \bar{\mu} + 1$. Since the algorithm does not increase the prox-parameter after a serious step or null step such that $v^k < \mu^k - 2\kappa$, we conclude that $\mu_{\max} := \sup_{k\in\mathbb{N}} \mu^k \leq \bar{\mu} + 1$ is finite. Finally, note that the prox-parameter is sharply increased after a null step such that $v^k \geq \mu^k - 2\kappa$: $\mu^{k+1} = v^k + 1 \geq \mu^k - 2\kappa + 1 > \mu^k + \delta$ because $\kappa \in (0, \frac{1}{2})$, with $\delta = \frac{1}{2} - \kappa > 0$. As a result, if the algorithm produces an infinite sequence of null steps after a last serious step, then the inequality $v^k < \mu^k - 2\kappa$ will be satisfied for all k large enough and the prox-parameter will become constant (otherwise μ^k would increase indefinitely, which contradicts that $\mu_{\max}$ is finite). □

We now define the following function $\bar{\mathsf{H}} : \mathscr{O} \times \mathscr{O} \to \mathbb{R}$, which is of key importance in our analysis:

$$\bar{\mathsf{H}}(x;y) := \max\left\{f_1(x) - [f_2(y) + \langle s_{f_2}, x-y\rangle] - \tau_f(y),\ c_1(x) - [c_2(y) + \langle s_{c_2}, x-y\rangle] - \tau_c(y)\right\}, \tag{3.5.1}$$

with $s_{f_2} \in \partial^{\mathsf{C}} f_2(y)$ and $s_{c_2} \in \partial^{\mathsf{C}} c_2(y)$. As these subgradients are not specified, the above definition is ambiguous. However, when y is a point previously computed by the algorithm, say $y = x^j$ for $j \leq k$, then $s_{f_2}^j \in \partial^{\mathsf{C}} f_2(x^j)$ and $s_{c_2}^j \in \partial^{\mathsf{C}} c_2(x^j)$ are the subgradients provided by the oracles and ambiguity disappears:

$$\bar{\mathsf{H}}(x;x^j) := \max\left\{f_1(x) - \bar{f}_2^j(x) - \tau_f(x^j),\ c_1(x) - \bar{c}_2^j(x) - \tau_c(x^j)\right\}.$$

It follows from convexity of f_1 and c_1 that, for every $y \in \mathscr{O}$ fixed, the function $\bar{\mathsf{H}}(\cdot;y)$ is convex and satisfies $\bar{\mathsf{H}}(\cdot;x^{\ell_k}) \geq \check{\mathsf{H}}^k(\cdot;x^{\ell_k})$ for all k. Furthermore, as $\ell_k \in \mathscr{B}_f^k \cap \mathscr{B}_c^k$ for all k, we have that $\check{f}_i^k(x^{\ell_k}) = f_i(x^{\ell_k})$, $\check{c}_i^k(x^{\ell_k}) = c_i(x^{\ell_k})$, $i = 1,2$, and thus

$$\bar{\mathsf{H}}(x^{\ell_k};x^{\ell_k}) = \check{\mathsf{H}}^k(x^{\ell_k};x^{\ell_k}) = H(x^{\ell_k};x^{\ell_k}). \tag{3.5.2}$$

The following lemma is of particular interest in the remainder of this chapter.

Lemma 3.5.2. *Suppose that $\bar{x}$ minimizes $\bar{\mathsf{H}}(\cdot;\bar{x})$ over X. Then, $\bar{x}$ satisfies the necessary optimality condition* (3.3.13).

Proof. Convexity of $\bar{\mathsf{H}}(\cdot;\bar{x})$ in the first argument and assumption on $\bar{x} \in X$ imply that $0 \in \partial_1 \bar{\mathsf{H}}(\bar{x};\bar{x}) + N_X(\bar{x})$. The result follows by noting that, for some pair of generalized subgradients $\bar{s}_{f_2} \in \partial^{\mathsf{C}} f_2(\bar{x})$ and $\bar{s}_{c_2} \in \partial^{\mathsf{C}} c_2(\bar{x})$, the following set

$$\partial_1 \bar{\mathsf{H}}(\bar{x};\bar{x}) = \begin{cases} \partial c_1(\bar{x}) - \bar{s}_{c_2} & \text{if } f(\bar{x}) - \tau_f(\bar{x}) < c(\bar{x}) - \tau_c(\bar{x}) \\ \mathtt{conv}\left\{\partial f_1(\bar{x}) - \bar{s}_{f_2},\, \partial c_1(\bar{x}) - \bar{s}_{c_2}\right\} & \text{if } f(\bar{x}) - \tau_f(\bar{x}) = c(\bar{x}) - \tau_c(\bar{x}) \\ \partial f_1(\bar{x}) - \bar{s}_{f_2} & \text{if } f(\bar{x}) - \tau_f(\bar{x}) > c(\bar{x}) - \tau_c(\bar{x}) \end{cases}$$

is contained in the one defined by the curly brackets in (3.3.13). □

We begin the convergence analysis for the case $\mathtt{Tol} = 0$ with the remark that the sequence of stability centers $\{x^{\ell_k}\}_k$ has at least one cluster point, since it is contained in the compact set X. We split the analysis into three cases: the algorithm performs only finitely many steps; the algorithm performs infinitely many steps and the sequence $\{x^{\ell_k}\}_k$ is either finite or infinite.

Proposition 3.5.3 (Finitely many iterations)**.** *Assume that Algorithm 2 stops at iteration k with $\mathtt{Tol} = 0$. Then, the last stability center $\bar{x} := x^{\ell_k} = x^{k+1}$ satisfies condition* (3.3.13).

Proof. It follows from the model's definition (3.4.6) and (3.5.1) that $\bar{\mathsf{H}}(x;x^{\ell_k}) \geq \check{\mathsf{H}}^k(x;x^{\ell_k})$ for all $x \in \mathscr{O}$. Hence, as $x^{\ell_k} \in X$ we have that

$$\begin{aligned} \bar{\mathsf{H}}(x^{\ell_k};x^{\ell_k}) &\geq \min_{x\in X}\ \bar{\mathsf{H}}(x;x^{\ell_k}) + \frac{\mu_k}{2}\|x - x^{\ell_k}\|^2 \\ &\geq \min_{x\in X}\ \check{\mathsf{H}}^k(x;x^{\ell_k}) + \frac{\mu_k}{2}\|x - x^{\ell_k}\|^2 \\ &= \check{\mathsf{H}}^k(x^{k+1};x^{\ell_k}) + \frac{\mu_k}{2}\|x^{k+1} - x^{\ell_k}\|^2 \\ &= \check{\mathsf{H}}^k(x^{\ell_k};x^{\ell_k}) = \bar{\mathsf{H}}(x^{\ell_k};x^{\ell_k}), \end{aligned}$$

where the first equality is due to (3.4.7), the second one follows by the fact that $x^{k+1} = x^{\ell_k}$ since the algorithm stops at iteration k with $\texttt{Tol} = 0$, and the last one is due to (3.5.2). Hence, x^{ℓ_k} minimizes $\bar{\mathsf{H}}(\cdot; x^{\ell_k}) + \frac{\mu_k}{2}\|\cdot - x^{\ell_k}\|^2$ over X and the quadratic term vanishes in the corresponding optimality condition: $0 \in \partial_1 \bar{\mathsf{H}}(\bar{x}; x^{\ell_k}) + N_X(\bar{x})$ and the stated result follows from Lemma 3.5.2. □

If the algorithm performs finitely many serious steps and infinite number of null steps, the following result shows that the last stability center satisfies (3.3.13).

Proposition 3.5.4 (Finitely many serious steps)**.** *Suppose that Algorithm 2 with* $\texttt{Tol} = 0$ *does not stop but produces only finitely many serious steps. Then the last stability center* $\bar{x}$ *satisfies the condition* (3.3.13)*, and* $\lim_{k\to\infty} x^k = \bar{x}$.

Proof. Let $\ell \in \mathbb{N}$ denote the last serious iteration, then $\bar{x} = x^\ell$ and note that, for all subsequent (null) iterations $k > \ell$, $\ell_k = \ell$ and the linearizations $\bar{f}_2^\ell$ and $\bar{c}_2^\ell$ are fixed in the model $\check{\mathsf{H}}^k(\cdot;\bar{x})$, which is in this case a cutting-plane model for the convex function $\bar{\mathsf{H}}(\cdot;\bar{x})$. Here we take $\tau_f^\ell = \tau_f(x^\ell)$, $\tau_c^\ell = \tau_c(x^\ell)$, and function $\bar{\mathsf{H}}(\cdot;\bar{x})$ defined with the fixed linearizations $\bar{f}_2^\ell$ and $\bar{c}_2^\ell$, i.e.,

$$\bar{\mathsf{H}}(\cdot; x^\ell) := \max\left\{ f_1(\cdot) - \bar{f}_2^\ell(\cdot) - \tau_f^\ell,\ c_1(\cdot) - \bar{c}_2^\ell(\cdot) - \tau_c^\ell \right\}.$$

We highlight that the updating rule for μ^k in Algorithm 2 ensures that the sequence $\{\mu^k\}_{k>\ell}$ is non-decreasing and becomes constant at a certain value $\mu' \in (0, \mu_{\max}]$ after finitely many steps $k' > \ell$, as a consequence of Lemma 3.5.1. More precisely, the updating rule at Step 4(b) of Algorithm 2 ensures that

$$\mu^k = \mu' \quad \text{and} \quad v^k + 2\kappa < \mu' \quad \text{for all } k > k'. \tag{3.5.3}$$

Hence, from iteration k' on, Algorithm 2 becomes a cutting-plane procedure to compute the unique solution $\tilde{x}$ of

$$\min_{x\in X}\ \bar{\mathsf{H}}(x;\bar{x}) + \frac{\mu'}{2}\|x - \bar{x}\|^2. \tag{3.5.4}$$

As the algorithm keeps all the active linearizations in the bundles (Step 4(b)), standard arguments from the convex bundle methods' theory (see [24, Prop. 4.3]) ensure that

$$\lim_{k\to\infty} x^k = \tilde{x} \quad \text{and} \quad \lim_{k\to\infty} [\check{\mathsf{H}}^k(x^{k+1};\bar{x}) - \bar{\mathsf{H}}(x^{k+1};\bar{x})] = 0.$$

(The last inequality implies that the convex model asymptotically coincides with the function at the limit point.) We claim that $\tilde{x} = \bar{x}$. To show that, let us assume the opposite, i.e., $\tilde{x} \neq \bar{x}$, and arrive to a contradiction. In this case, for some $\delta > 0$, we may find and index

k_1 such that $\|x^{k+1}-\bar{x}\|^2 > \delta$ for all $k \geq k_1$. We may furthermore find an index k_2 such that $\check{\mathsf{H}}^k(x^{k+1};\bar{x}) - \bar{\mathsf{H}}(x^{k+1};\bar{x}) \geq -\frac{\kappa}{2}\delta$ for all $k \geq k_2$ as the left-hand side vanishes. Therefore, for $k'' \geq \max\{k_1, k_2, k'\}$, we have

$$\check{\mathsf{H}}^k(x^{k+1};\bar{x}) - \bar{\mathsf{H}}(x^{k+1};\bar{x}) \geq -\frac{\kappa}{2}\|x^{k+1}-\bar{x}\|^2 \neq 0 \quad \text{for all } k > k''.$$

The following chain of inequalities holds at every iteration $k > k''$:

$$\begin{aligned}
\check{\mathsf{H}}^k(\bar{x};\bar{x}) &\geq \check{\mathsf{H}}^k(x^{k+1};\bar{x}) + \tfrac{\mu'}{2}\|x^{k+1}-\bar{x}\|^2 && \text{(by (3.4.7) and (3.5.3))}\\
&= [\check{\mathsf{H}}^k(x^{k+1};\bar{x}) - \bar{\mathsf{H}}(x^{k+1};\bar{x})] + \bar{\mathsf{H}}(x^{k+1};\bar{x}) + \tfrac{\mu'}{2}\|x^{k+1}-\bar{x}\|^2\\
&\geq -\tfrac{\kappa}{2}\|x^{k+1}-\bar{x}\|^2 + \bar{\mathsf{H}}(x^{k+1};\bar{x}) + \tfrac{\mu'}{2}\|x^{k+1}-\bar{x}\|^2\\
&\geq -\tfrac{\kappa}{2}\|x^{k+1}-\bar{x}\|^2 + \max\left\{\begin{array}{l} f_1(x^{k+1}) - \bar{f}_2^\ell(x^{k+1}) - \tau_f^\ell + \frac{\mu'}{2}\|x^{k+1}-\bar{x}\|^2 \\ c_1(x^{k+1}) - \bar{c}_2^\ell(x^{k+1}) - \tau_c^\ell + \frac{\mu'}{2}\|x^{k+1}-\bar{x}\|^2 \end{array}\right\} && \text{(by (3.5.1))}\\
&> -\tfrac{\kappa}{2}\|x^{k+1}-\bar{x}\|^2 + \max\left\{\begin{array}{l} f_1(x^{k+1}) - \bar{f}_2^\ell(x^{k+1}) - \tau_f^\ell + \frac{\nu^k+2\kappa}{2}\|x^{k+1}-\bar{x}\|^2 \\ c_1(x^{k+1}) - \bar{c}_2^\ell(x^{k+1}) - \tau_c^\ell + \frac{\nu^k+2\kappa}{2}\|x^{k+1}-\bar{x}\|^2 \end{array}\right\} && \text{(by (3.5.3))}\\
&\geq -\tfrac{\kappa}{2}\|x^{k+1}-\bar{x}\|^2 + \max\left\{\begin{array}{l} f_1(x^{k+1}) - f_2(x^{k+1}) - \tau_f^\ell + \frac{2\kappa}{2}\|x^{k+1}-\bar{x}\|^2 \\ c_1(x^{k+1}) - c_2(x^{k+1}) - \tau_c^\ell + \frac{2\kappa}{2}\|x^{k+1}-\bar{x}\|^2 \end{array}\right\} && \text{(by (3.4.11))}\\
&= H(x^{k+1};\bar{x}) + \tfrac{\kappa}{2}\|x^{k+1}-\bar{x}\|^2. && \text{(by (3.3.10))}
\end{aligned}$$

As $\bar{x} = x^\ell$ and $\check{\mathsf{H}}^k(x^\ell;x^\ell) = H(x^\ell;x^\ell)$ due to (3.5.2), we have shown that the descent test (3.4.5) is satisfied at $x^{k+1} \neq x^\ell$:

$$H(x^{k+1};x^\ell) \leq H(x^\ell;x^\ell) - \frac{\kappa}{2}\|x^{k+1}-x^\ell\|^2,$$

contradicting thus the assumption that only null steps are performed for $k > \ell$. Hence, $\tilde{x} = \bar{x}$ and the last stability center solves (3.5.4). This allows us to conclude (thanks to convexity of $\bar{\mathsf{H}}(\cdot;\bar{x})$) that $\bar{x} = x^\ell$ solves $\min_{x\in X}\bar{\mathsf{H}}(x;x^\ell)$. Lemma 3.5.2 then concludes the proof. □

Remark 3.5.5 (Bundle compression)**.** It is worth mentioning that the index sets $\mathscr{B}_f^k$ and $\mathscr{B}_c^k$ gathering the information bundle can be kept bounded; each one having at most $M_{\max}$ indices, for a chosen integer $M_{\max} \geq 3$. Indeed, it suffices to keep in the bundles the linearizations issued by the stability center x^{ℓ_k}, the new trial point x^{k+1} and the so-called aggregate linearization as in [24, Eq. 4.5]. When transcribed to our setting, the aggregate

linearizations for f_1 and c_1 read as

$$\check{f}_1^{a_f^k}(x) := \check{f}_1^k(x^{k+1}) + \langle p_f^k, x - x^{k+1}\rangle \leq f_1(x) \quad \forall x \in \mathbb{R}^n$$

$$\check{c}_1^{a_c^k}(x) := \check{c}_1^k(x^{k+1}) + \langle p_c^k, x - x^{k+1}\rangle \leq c_1(x) \quad \forall x \in \mathbb{R}^n,$$

with $p_f^k := \sum_{j\in\mathscr{B}_f^k} \alpha_f^j s_{f_1}^j$, $p_c^k := \sum_{j\in\mathscr{B}_c^k} \alpha_c^j s_{c_1}^j$ and multipliers α_f, α_c as in (3.4.10). We claim that the following economical rule for managing $\mathscr{B}_f^k$ and $\mathscr{B}_c^k$ (in Step 4 of Algorithm 2) is enough to ensure convergence:

Serious step: set $\mathscr{B}_f^{k+1} = \{k+1\}$ and $\mathscr{B}_c^{k+1} = \{k+1\}$;

Null step: set $\mathscr{B}_f^{k+1} = \{k+1, \ell_k, a_f^k\}$ and $\mathscr{B}_c^{k+1} = \{k+1, \ell_k, a_c^k\}$.

Indeed, Proposition 3.5.4 is still valid if the algorithm employs the above economical rule for updating the bundles: the key Proposition 4.3 from [24] still applies and thus the displayed equations right after (3.5.4) hold. As it can be noted in the sequel, no bundle management restriction (besides the requirement that $k+1 \in \mathscr{B}_f^{k+1} \cap \mathscr{B}_c^{k+1}$) is required after a serious steps. □

We consider now the case of infinitely many serious steps. To this end, we need the following auxiliary result.

Lemma 3.5.6. *There exist constants $L, M > 0$ such that, for all $k \in \mathbb{N}$, the three following conditions hold for $p^{k+1} \in \partial_1 \check{\mathsf{H}}^k(x^{k+1}; x^{\ell_k})$, $s_X^{k+1} \in N_X(x^{k+1})$, and $e^{k+1} = L||x^{k+1} - x^{\ell_k}||$:*

$$||p^{k+1} + s_X^{k+1}|| \leq \mu_{\max} ||x^{k+1} - x^{\ell_k}|| \leq M, \tag{3.5.5a}$$

$$p^{k+1} + s_X^{k+1} \in \partial_{e^{k+1}} \left[\bar{\mathsf{H}}(x^{\ell_k}; x^{\ell_k}) + i_X(x^{\ell_k})\right], \tag{3.5.5b}$$

$$p^{k+1} \in \partial_{e^{k+1}} \bar{\mathsf{H}}(x^{\ell_k}; x^{\ell_k}). \tag{3.5.5c}$$

Proof. As $\mu^k \in (0, \mu_{\max}]$ (c.f. Lemma 3.5.1), expression (3.4.9) yields the first inequality in (3.5.5a). Recall that the iterates x^{k+1} and x^{ℓ_k} are contained in the bounded set X for all k. The second inequality in (3.5.5a) then follows. Convexity of the function $\check{\mathsf{H}}^k + i_X$ and (3.4.9) gives that, for all $x \in \mathbb{R}^n$,

$$\begin{aligned}
\check{\mathsf{H}}^k(x; x^{\ell_k}) + i_X(x) &\geq \check{\mathsf{H}}^k(x^{k+1}; x^{\ell_k}) + \langle p^{k+1} + s_X^{k+1}, x - x^{k+1}\rangle \\
&\geq \check{\mathsf{H}}^k(x^{k+1}; x^{\ell_k}) + \langle p^{k+1} + s_X^{k+1}, x - x^{\ell_k}\rangle + \langle p^{k+1} + s_X^{k+1}, x^{\ell_k} - x^{k+1}\rangle \\
&\geq \check{\mathsf{H}}^k(x^{k+1}; x^{\ell_k}) + \langle p^{k+1} + s_X^{k+1}, x - x^{\ell_k}\rangle - M||x^{\ell_k} - x^{k+1}||,
\end{aligned} \tag{3.5.6}$$

where the last inequality is due to (3.5.5a) and Cauchy-Schwarz inequality. Definition (3.4.6) of $\check{\mathsf{H}}^k(\cdot;x^{\ell_k})$ as well as the fact that $\ell_k \in \mathscr{B}_f^k \cap \mathscr{B}_c^k$ give the following chain of inequalities:

$$\begin{aligned}
\check{\mathsf{H}}^k(x^{k+1};x^{\ell_k}) &\geq \max\left\{\bar{f}_1^{\ell_k}(x^{k+1}) - \bar{f}_2^{\ell_k}(x^{k+1}) - \tau_f(x^{\ell_k}), \bar{c}_1^{\ell_k}(x^{k+1}) - \bar{c}_2^{\ell_k}(x^{k+1}) - \tau_c(x^{\ell_k})\right\} \\
&= \max\Big\{ f_1(x^{\ell_k}) - f_2(x^{\ell_k}) + \langle s_{f_1}^{\ell_k} - s_{f_2}^{\ell_k}, x^{k+1} - x^{\ell_k}\rangle - \tau_f(x^{\ell_k}), \\
&\quad c_1(x^{\ell_k}) - c_2(x^{\ell_k}) + \langle s_{c_1}^{\ell_k} - s_{c_2}^{\ell_k}, x^{k+1} - x^{\ell_k}\rangle - \tau_c(x^{\ell_k})\Big\} \\
&\geq \max\left\{ f_1(x^{\ell_k}) - f_2(x^{\ell_k}) - \tau_f(x^{\ell_k}),\, c_1(x^{\ell_k}) - c_2(x^{\ell_k}) - \tau_c(x^{\ell_k})\right\} \\
&\quad + \min\left\{ \langle s_{f_1}^{\ell_k} - s_{f_2}^{\ell_k}, x^{k+1} - x^{\ell_k}\rangle,\, \langle s_{c_1}^{\ell_k} - s_{c_2}^{\ell_k}, x^{k+1} - x^{\ell_k}\rangle \right\}.
\end{aligned}$$

Since $X \subset \mathscr{O}$ is compact, we have that $\partial f_1(X)$, $\partial c_1(X)$, $\partial^{\mathsf{C}} f_2(X)$, and $\partial^{\mathsf{C}} c_2(X)$ are bounded sets (see Section 3.2). Hence, there exist $K_f > 0$ and $K_c > 0$ such that, $\|s_{f_1}^{\ell_k} - s_{f_2}^{\ell_k}\| \leq K_f$ and $\|s_{c_1}^{\ell_k} - s_{c_2}^{\ell_k}\| \leq K_c$ for all k. Applying the Cauchy-Schwarz inequality to the inequalities above and recalling that

$$\max\left\{ f_1(x^{\ell_k}) - f_2(x^{\ell_k}) - \tau_f(x^{\ell_k}),\, c_1(x^{\ell_k}) - c_2(x^{\ell_k}) - \tau_c(x^{\ell_k})\right\} = H(x^{\ell_k};x^{\ell_k}) = \bar{\mathsf{H}}(x^{\ell_k};x^{\ell_k})$$

by definition, we get

$$\check{\mathsf{H}}^k(x^{k+1};x^{\ell_k}) \geq \bar{\mathsf{H}}(x^{\ell_k};x^{\ell_k}) - L_0\,||x^{k+1} - x^{\ell_k}||, \quad \text{with } L_0 = \max\{K_f, K_c\}. \tag{3.5.7}$$

Recall that $\bar{\mathsf{H}}(x;x^{\ell_k}) \geq \check{\mathsf{H}}^k(x;x^{\ell_k})$ for all $x \in \mathbb{R}^n$ and combine (3.5.6) with (3.5.7) to obtain

$$\begin{aligned}
\bar{\mathsf{H}}(x;x^{\ell_k}) + i_X(x) &\geq \check{\mathsf{H}}^k(x;x^{\ell_k}) + i_X(x) \\
&\geq \check{\mathsf{H}}^k(x^{k+1};x^{\ell_k}) + \langle p^{k+1} + s_X^{k+1}, x - x^{\ell_k}\rangle - M||x^{\ell_k} - x^{k+1}|| \\
&\geq \bar{\mathsf{H}}(x^{\ell_k};x^{\ell_k}) - (L_0 + M)||x^{k+1} - x^{\ell_k}|| + \langle p^{k+1} + s_X^{k+1}, x - x^{\ell_k}\rangle.
\end{aligned}$$

We have thus shown (3.5.5b) with $L = M + L_0$. To prove the last inclusion (3.5.5c), observe that this chain of inequalities remains true if the term $i_X(x)$ is excluded together with corresponding subdifferential s_X^{k+1}: for all $x \in \mathbb{R}^n$,

$$\begin{aligned}
\bar{\mathsf{H}}(x;x^{\ell_k}) &\geq \check{\mathsf{H}}^k(x;x^{\ell_k}) \geq \check{\mathsf{H}}^k(x^{k+1};x^{\ell_k}) + \langle p^{k+1}, x - x^{\ell_k}\rangle - M||x^{\ell_k} - x^{k+1}|| \\
&\geq \bar{\mathsf{H}}(x^{\ell_k};x^{\ell_k}) + \langle p^{k+1}, x - x^{\ell_k}\rangle - (L_0 + M)||x^{k+1} - x^{\ell_k}||.
\end{aligned}$$

□

Lemma 3.5.6 can be applied to any iteration $k+1$ between the serious step ℓ_k and ℓ_{k+1} (ℓ_{k+1} included). Indeed, equation (3.4.9) used to show (3.5.5a) holds true. Other arguments used in the proof remain valid for any iteration between ℓ_k and ℓ_{k+1}, as from the point of view of the algorithm, the only change is the bundle information, which is not explicitly used in the proof.

Observe that $\check{\mathsf{H}}^k(\cdot;x^{\ell_k})$ given in (3.4.6) is the pointwise maximum of finitely many affine functions. Hence, its subdifferential is the convex hull of the "active" linearization slopes, i.e., Proposition 3.2.1 ii) asserts that

$$\partial_1\check{\mathsf{H}}^k(x^{k+1};x^{\ell_k}) := \mathtt{conv}\left\{\left\{s^j_{f_1}-s^{\ell_k}_{f_2}\right\}_{j\in\bar{\mathscr{B}}^k_f},\left\{s^j_{c_1}-s^{\ell_k}_{c_2}\right\}_{j\in\bar{\mathscr{B}}^k_c}\right\}, \tag{3.5.8}$$

with $\bar{\mathscr{B}}^k_f$ and $\bar{\mathscr{B}}^k_c$ given in (3.4.10). Since $X\subset\mathscr{O}$ is compact, we have that $\partial f_1(X)$, $\partial c_1(X)$, $\partial^{\mathsf{C}} f_2(X)$, and $\partial^{\mathsf{C}} c_2(X)$ are bounded sets (see Section 3.2). Thus, (3.4.9) and (3.5.8) certificate that the sequence of model's subgradients $\left\{p^k\right\}$ is bounded. To prove the next proposition we will use the latter property, along with the following characteristic of subdifferentials for a sequence of convex functions that converges pointwise.

Lemma 3.5.7. *Let $\varphi:\mathbb{R}^n\to\mathbb{R}$ be a convex function, and $\{\varphi_\ell\}_{\ell\in\mathbb{N}}$ a sequence of convex functions $\varphi_\ell:\mathbb{R}^n\to\mathbb{R}$ converging pointwise to φ, i.e., $\lim_{\ell\to\infty}\varphi_\ell(x)=\varphi(x)$ for every given point x. Furthermore, let $\{x^\ell\}\subset\mathbb{R}^n$ be such that $\lim_{\ell\to\infty}x^\ell=\bar{x}$ and $\{\varepsilon^\ell\}\subset\mathbb{R}_+$ satisfy $\lim_{\ell\to\infty}\varepsilon^\ell=0$. If $g^\ell\in\partial_{\varepsilon^\ell}\varphi_\ell(x^\ell)$ for all ℓ and $\lim_{\ell\to\infty}g^\ell=\bar{g}$, then $\bar{g}\in\partial\varphi(\bar{x})$.*

Proof. First, let us prove that $\liminf_\ell\varphi_\ell(x^\ell)\geq\varphi(\bar{x})$. Since $\mathtt{dom}(\varphi_\ell)=\mathtt{dom}(\varphi)=\mathbb{R}^n$, it follows from [111, Cor. 2C] that the pointwise convergence of $\{\varphi_\ell\}_{\ell\in\mathbb{N}}$ is equivalent to epi-convergence, which in turn is equivalent (see [111, Eq. (4.2)]) to epi-convergence of $\{\varphi^*_\ell\}_{\ell\in\mathbb{N}}$, the sequence of conjugate functions to φ_ℓ. Hence, it follows that $\lim_\ell\varphi^*_\ell(x)=\varphi^*(x)$ for every given $x\in\mathbb{R}^n$. Now consider the following development:

$$\varphi_\ell(x^\ell)=(\varphi^*_\ell)^*(x^\ell)=\sup_{y\in\mathbb{R}^n}\left[\langle y,x^\ell\rangle-\varphi^*_\ell(y)\right]\geq\langle y,x^\ell\rangle-\varphi^*_\ell(y)\ \ \forall y\in\mathbb{R}^n.$$

Accordingly, $\liminf_\ell\varphi_\ell(x^\ell)\geq\liminf_\ell[\langle y,x^\ell\rangle-\varphi^*_\ell(y)]=\langle y,\bar{x}\rangle-\varphi^*(y)$ for all $y\in\mathbb{R}^n$, showing that

$$\liminf_\ell\varphi_\ell(x^\ell)\geq\sup_y\left[\langle y,\bar{x}\rangle-\varphi^*(y)\right]=(\varphi^*)^*(\bar{x})=\varphi(\bar{x}).$$

Recall that $g^\ell \in \partial_{\varepsilon^\ell}\varphi_\ell(x^\ell)$. Then, $\varphi_\ell(x) \geq \varphi_\ell(x^\ell) + \langle g^\ell, x - x^\ell\rangle - \varepsilon^\ell$ for all $x \in \mathbb{R}^n$. By taking the limit when ℓ goes to infinity we get

$$\begin{aligned}\varphi(x) = \lim_\ell \varphi_\ell(x) = \liminf_\ell \varphi_\ell(x) &\geq \liminf_\ell [\varphi_\ell(x^\ell) + \langle g^\ell, x - x^\ell\rangle - \varepsilon^\ell] \\ &\geq \liminf_\ell \varphi_\ell(x^\ell) + \liminf_\ell \langle g^\ell, x - x^\ell\rangle - \limsup_\ell \varepsilon^\ell \\ &\geq \varphi(\bar{x}) + \langle \bar{g}, x - \bar{x}\rangle,\end{aligned}$$

showing that $\bar{g} \in \partial\varphi(\bar{x})$. □

Proposition 3.5.8 (Infinitely many serious steps)**.** *Assume that the algorithm performs infinitely many serious steps. Then, any cluster point $\bar{x} \in X$ of the sequence $\{x^{\ell_k}\}_k$ satisfies the necessary optimality condition* (3.3.13).

Proof. We first show that

$$\lim_{k\to\infty} \|x^{\ell_{k+1}} - x^{\ell_k}\| = 0. \tag{3.5.9}$$

To this end, we must analyze the two cases of Lemma 3.4.1. In case i), Algorithm 2 produces a feasible point for (3.1.1) after finitely many serious steps and all subsequent points are feasible. Let $x^{\ell_{k_1}}$ be the first feasible serious iterate. Then, Lemma 3.4.1 i) yields

$$f(x^{\ell_{k+1}}) \leq f(x^{\ell_k}) - \tfrac{\kappa}{2}\|x^{\ell_{k+1}} - x^{\ell_k}\|^2 \text{ and } c(x^{\ell_{k+1}}) \leq 0 \text{ for all } k \geq \ell_{k_1}.$$

The telescopic sum of the first inequality above yields

$$\sum_{k=k_1}^{\infty} \|x^{\ell_{k+1}} - x^{\ell_k}\|^2 \leq \frac{2}{\kappa}\sum_{k=k_1}^{\infty}\left(f\left(x^{\ell_k}\right) - f\left(x^{\ell_{k+1}}\right)\right) \leq \frac{2}{\kappa}\left(f\left(x^{\ell_{k_1}}\right) - \lim_{k\to\infty} f\left(x^{\ell_{k+1}}\right)\right).$$

Since f is finite-valued and continuous over the bounded set X, the right-hand side of the above inequality is finite. Hence, (3.5.9) holds. Assume now that the sequence $\{x^{\ell_k}\}$ is infeasible for (3.1.1). Lemma 3.4.1 ii) yields

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

Once again, by using the telescopic sum we get (3.5.9).

As $X \subset \mathscr{O}$ is compact, with $\mathscr{O}$ an open set contained in the domains of component functions, and the generalized subdifferential is locally compact, we conclude that for $s_{f_2}^{\ell_k} \in \partial^{\mathsf{C}} f_2(x^{\ell_k})$ and $s_{c_2}^{\ell_k} \in \partial^{\mathsf{C}} c_2(x^{\ell_k})$

$$\left\{x^{\ell_k}\right\}, \quad \left\{s_{f_2}^{\ell_k}\right\} \quad \text{and} \quad \left\{s_{c_2}^{\ell_k}\right\} \quad \text{are bounded sequences.}$$

By taking subsequences, we can define an index set $\mathscr{K} \subset \{0,1,2,\ldots\}$ such that

$$\lim_{\mathscr{K}\ni k\to\infty} x^{\ell_k} = \bar{x} \in X, \quad \lim_{\mathscr{K}\ni k\to\infty} s_{f_2}^{\ell_k} = \bar{s}_{f_2} \in \partial^{\mathtt{C}} f_2(\bar{x}) \quad \text{and} \quad \lim_{\mathscr{K}\ni k\to\infty} s_{c_2}^{\ell_k} = \bar{s}_{c_2} \in \partial^{\mathtt{C}} c_2(\bar{x}),$$

where the two last limits are due to the fact that the generalized subdifferential is outer-semicontinuous [20, Prop. 2.1.5(b)]. Let us now define $\varphi_k(\cdot) = \bar{\mathsf{H}}(\cdot; x^{\ell_k})$ and $\varphi(\cdot) = \bar{\mathsf{H}}(\cdot;\bar{x})$, with the latter defined in (3.5.1) with $y = \bar{x}$ and the pair of generalized subgradients $(\bar{s}_{f_2}, \bar{s}_{c_2})$ above. For every $x \in \mathbb{R}^n$ fixed, the above limits imply

$$\lim_{\mathscr{K}\ni k\to\infty} \bar{\mathsf{H}}(x; x^{\ell_k}) = \bar{\mathsf{H}}(x;\bar{x}),$$

i.e., $\{\varphi_k\}_{k\in\mathscr{K}}$ converges pointwise to φ. Applying Lemma 3.5.6 for $k+1 = \ell_{k+1}$, we obtain $p^{\ell_{k+1}} \in \partial_{e^{\ell_{k+1}}} \varphi_\ell(x^{\ell_k})$ (Eq. (3.5.5c)). Furthermore, as the sequence $\{p^{\ell_{k+1}}\}$ is bounded (see the paragraph right before this proposition), we may take another subsequence indexed by $\mathscr{K}' \subset \mathscr{K}$ so that $\lim_{\mathscr{K}'\ni\ell\to\infty} p^{\ell_{k+1}} = \bar{p} \in \mathbb{R}^n$ and $\lim_{\mathscr{K}'\ni\ell\to\infty} e^{\ell_{k+1}} = 0$ in view of (3.5.9) and definition of e^{k+1} given in Lemma 3.5.6. With these conditions at hand, Lemma 3.5.7 ensures that $\bar{p} \in \partial\varphi(\bar{x})$, i.e., $\bar{p} \in \partial_1 \bar{\mathsf{H}}(\bar{x};\bar{x})$. Next, observe that $\{s_X^{\ell_{k+1}}\}$ is a bounded sequence as the inequality

$$||p^{\ell_{k+1}} + s_X^{\ell_{k+1}}|| \leq \mu_{\max} ||x^{\ell_{k+1}} - x^{\ell_k}|| \tag{3.5.10}$$

holds due to Lemma 3.5.6 (with $\mu_{\max}$ finite due to Lemma 3.5.1). By definition of the convex normal cone, it follows that there exists a suitable subsequence of $\{s_X^{\ell_{k+1}}\}_{k\in\mathscr{K}''}$, with $\mathscr{K}'' \subset \mathscr{K}'$ converging to a cluster point $\bar{s} \in N_X(\bar{x}) = \partial i_X(\bar{x})$. Hence, since X is polyhedral and $\mathtt{ri}(\mathscr{D}\mathrm{om}(\bar{\mathsf{H}}(\cdot;\bar{x}))) = \mathscr{O} \neq \emptyset$,

$$\bar{p} + \bar{s} \in \partial_1 \bar{\mathsf{H}}(\bar{x};\bar{x}) + \partial i_X(\bar{x}) = \partial_1 \big[\bar{\mathsf{H}}(\bar{x};\bar{x}) + i_X(\bar{x})\big]. \tag{3.5.11}$$

Finally, inequality (3.5.10) combined with (3.5.9) yield $\bar{p} + \bar{s} = 0$. Hence, $0 \in \partial_1 \big[\bar{\mathsf{H}}(\bar{x};\bar{x}) + i_X(\bar{x})\big]$, showing that $\bar{x}$ minimizes $\bar{\mathsf{H}}(\cdot;\bar{x})$ over X. Lemma 3.5.2 thus concludes the proof. □

The following theorem sums up the algorithm's convergence analysis.

Theorem 3.5.9 (Convergence analysis)**.** *Let $X \neq \emptyset$ be a bounded polyhedron contained in the open set $\mathscr{O}$, $f_1, c_1 : \mathscr{O} \to \mathbb{R}$ convex, and $f_2, c_2 : \mathscr{O} \to \mathbb{R}$ weakly convex functions on some neighbourhood of each $x \in \mathscr{O}$. If in Algorithm 2 the stopping test tolerance* `Tol` *is set to zero, then any cluster point $\bar{x}$ of the sequence of stability centers $\{x^{\ell_k}\}$ satisfies the necessary optimality condition* (3.3.13)*.*

Furthermore, concerning the original problem (3.1.1)*, the following holds:*

i) If $c(\bar{x}) > 0$ (which cannot happen if x^0 is feasible), then $\bar{x}$ is a critical point of

$$\min_{x \in X} c_1(x) - c_2(x).$$

ii) If $c(\bar{x}) = 0$ and $\bar{x}$ is not a critical point of $\min_{x \in X} c_1(x) - c_2(x)$, then $\bar{x}$ satisfies the criticality condition (3.3.9) *with $\bar{\lambda} > 0$.*

iii) If $c(\bar{x}) < 0$, then $\bar{x}$ satisfies the criticality condition (3.3.9) *with $\bar{\lambda} = 0$.*

If $\texttt{Tol} > 0$, then the algorithm stops after finitely many steps $k \in \mathbb{N}$ with an approximate critical point x^{ℓ_k} of (3.3.13).

Proof. For the case $\texttt{Tol} = 0$, condition (3.3.13) follows directly from Proposition 3.5.3 if $\{x^{\ell_k}\}$ is finite, from Proposition 3.5.4 if the the algorithm produces only finitely many serious steps followed by an infinite sequence of null steps, and from Proposition 3.5.8 if infinitely many serious steps are produced.

Furthermore, the connection with the necessary optimality condition for the original problem (3.1.1) is established by Theorem 3.3.5.

Proposition 3.5.4 ensures that $\lim_{k \to \infty} ||x^{k+1} - x^{\ell}|| = 0$ if x^{ℓ} is the last stability center. Otherwise, $\lim_{k \to \infty} ||x^{\ell_{k+1}} - x^{\ell_k}|| = 0$, as shown in the proof of Proposition 3.5.8. Thus, Algorithm 2 stops after finitely many steps provided $\texttt{Tol} > 0$. □

3.6 Simplified algorithm for the case without nonlinear constraints

This section describes how Algorithm 2 can be simplified to deal with the simpler convexly-constrained problem

$$\min_{x \in X} f(x), \quad \text{with} \quad f(x) = f_1(x) - f_2(x). \tag{3.6.1}$$

In this case, the problem's model (3.4.6) reduces to $\check{\mathsf{H}}^k(x; x^{\ell_k}) = \check{f}_1^k(x) - \bar{f}_2^{\ell_k}(x)$, and the descent test (3.4.5) becomes $f(x^{k+1}) \leq f(x^{\ell_k}) - \frac{\kappa}{2}\|x - x^{\ell_k}\|^2$. Hence, Algorithm 2 boils down to the following plainer scheme.

Convergence analysis for Algorithm 3 follows from that of Algorithm 2 upon several simplifications. Instead of doing this exercise, we simply state the following result.

Theorem 3.6.1. *Consider problem* (3.6.1) *with $X \neq \emptyset$ a bounded polyhedron contained in the open set $\mathscr{O}$, $f_1 : \mathscr{O} \to \mathbb{R}$ convex, and $f_2 : \mathscr{O} \to \mathbb{R}$ weakly convex on some neighbourhood*

Algorithm 3 Proximal Bundle Method for Convexly-Constrained `CwC` programs

Step 0 (Initialization) Let $x^0 \in X$, $\kappa \in (0, \frac{1}{2})$, $\kappa \leq \mu^0$, and $\texttt{Tol} \geq 0$ be given.
Call the oracles to compute $f_i(x^0)$ and (generalized) subgradients $s_{f_i}^0$, $i = 1,2$.
Define $k := \ell_k = 0$ and $\mathscr{B}_f^0 := \{0\}$.

Step 1 (Trial point) Compute x^{k+1} the (x-part) solution of the QP

$$\begin{cases} \min\limits_{x,r} & r + \frac{\mu^k}{2}\|x - x^{\ell_k}\|^2 \\ \text{s.t.} & \bar{f}_1^j(x) - \bar{f}_2^{\ell_k}(x) - r \leq 0 \quad \forall\, j \in \mathscr{B}_f^k \\ & x \in X, \;\; r \in \mathbb{R}. \end{cases}$$

Step 2 (Stopping test) If $\|x^{k+1} - x^{\ell_k}\| \leq \texttt{Tol}$, then stop and return x^{ℓ_k}.

Step 3 (Oracles call) Compute $f_i(x^{k+1})$, and subgradients $s_{f_i}^{k+1}$, $i = 1,2$.

Step 4 (Descent test)

(a) If $f(x^{k+1}) \leq f(x^{\ell_k}) - \frac{\kappa}{2}\|x - x^{\ell_k}\|^2$ holds, then declare a *serious step*: define $\ell_{k+1} := k+1$, choose $\mathscr{B}_f^{k+1} \subset \{0,\ldots,k+1\}$ with $\{k+1\} \in \mathscr{B}_f^{k+1}$ and arbitrarily select $\mu^{k+1} \in (0, \mu^k]$.

(b) Else, declare a *null step*: define $\ell_{k+1} := \ell_k$ and choose $\mathscr{B}_f^{k+1} \subset \{0,\ldots,k+1\}$ with $\bar{\mathscr{B}}_f^k \cup \{k+1, \ell_k\} \subset \mathscr{B}_f^{k+1}$ ($\bar{\mathscr{B}}_f^k$ as in (3.4.10)).
Compute $v^k := 2\max\left\{\frac{\bar{f}_2^{\ell_k}(x^{k+1}) - f_2(x^{k+1})}{\|x^{k+1} - x^{\ell_k}\|^2},\, 0\right\}$. If $v^k \geq \mu^k - 2\kappa$, set $\mu^{k+1} = v^k + 1$; otherwise $\mu^{k+1} = \mu^k$.

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

of each $x \in \mathscr{O}$. If in Algorithm 3 the stopping test tolerance `Tol` is set to zero, then any cluster point $\bar{x}$ of the sequence of stability centers $\{x^{\ell_k}\}$ satisfies the necessary optimality condition (3.3.3).

If `Tol` > 0, then the algorithm stops after finitely many steps $k \in \mathbb{N}$ with an approximate critical point x^{ℓ_k} of (3.3.3).

To have an intuition of why the above theorem is valid, the reader may think of adding a dummy convex nonlinear convex function $c(x) = c_1(x) - 0$ to (3.6.1) and rely on the results from Sections 3.4 and 3.5. Indeed, by selecting a constant $M > 0$ large enough and function c such that $c(x) \leq -M < 0$ for all $x \in X$, we can see that Algorithm 2 applied to (3.6.1) with the additional and superfluous constraint $c(x) \leq 0$ boils down to Algorithm 3. Furthermore, in this artificial setting, the above convergence result follows directly from Theorem 3.5.9, item iii).

3.7 Illustrative numerical examples

This section aims to illustrate our approach to solving some challenging test problems. Here we have two goals: show that it provides good-quality approximate critical points (examples of Subsection 3.7.2 - Subsection 3.7.4, Section 3.8) and is able to solve problems that, to the best of our knowledge, could not be resolved with other solvers (Subsection 3.7.1). We consider four nonconvex stochastic optimization problems and one coming from signal processing. Notice that the stochastic problems (example of Subsection 3.7.1 - Subsection 3.7.3) do not have explicit DoC decompositions, and thus DoC programming algorithms are not directly applicable. Another numerical example of a chance-constrained energy management problem can be found in [137].

Numerical experiments were performed using MATLAB R2020a and Gurobi 9.5.1 (for solving the master QP subproblem (3.4.8) in Algorithm 2) on a personal computer with the following characteristics: Windows 10 Professional, 32 Go, Intel i7-10850H (6 cores). Our implementation enables applying Algorithm 3 for simpler problems without non-linear constraints, specifically, to the problem considered in Subsection 3.7.2.

Unless otherwise specified, the choice of the parameters in Algorithm 2 is as follows: $\rho = \sigma = \frac{1}{2}$, $\kappa = 0.3$, $\mu^0 = 10^2$ and $\texttt{Tol} = 10^{-4}$.

3.7.1 Highly nonconvex chance-constrained problem

In this section we investigate the following optimization problem (having weakly-convex constraint):

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\top x \\ \text{s.t.} \quad & \mathbb{P}\left[\frac{1}{2}\xi^\top Q_j(x)\xi + q_j(x)^\top \xi + d_j(x) \leq 0,\ j = 1,...,k\right] \geq p \\ & \underline{x} \leq x \leq \bar{x}. \end{aligned} \tag{3.7.1}$$

We first note that as a result of [136] and the upfollowing concrete data, that the probability function is continuously differentiable. Furthermore the underlying feasible set is compact and so we are in case ii) of the introduction: $c_1(x) = p$ and $-c_2$ indicating the probability function. The underlying data is not convex in the parameter replaced by the random vector. As a result, the underlying feasible set is not expected to be convex. Concretely we will

consider the following data, for $k = 2$:

$$Q_1(x) = \begin{bmatrix} 3(x_1-1) & -x_2 \\ -x_2 & 3(x_1-1) \end{bmatrix}, \quad Q_2(x) = \begin{bmatrix} -2x_2 & x_1-1 \\ x_1-1 & -2x_2 \end{bmatrix}$$

as well as

$$q_1(x) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}(x_1-1), \quad q_2(x) = \begin{bmatrix} 1 \\ 4 \end{bmatrix} x_2$$

$d_1(x) = -2, d_2(x) = -2$. We have also picked $p = 0.7$ together with $c = (-1,-1)$, $\underline{x} = (-2,-2)$, $\bar{x} = (2,2)$. The random vector is taken to be multivariate Gaussian with mean vector 0 and covariance matrix

$$\Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

This optimization problem is quite challenging. First it can be observed that an alternative sample average approximation along the lines of [85] would be a MILP. It is thus tempting to first try to solve the resulting optimization problem with such a formulation. We have done so for the following sample sizes $\{100, 1\,000, 5\,000, 10\,000, 50\,000\}$, using CPLEX 12.10. The resulting computation times are $0.5, 4, 22, 100, > 8\,000$ seconds respectively. The last computation was aborted still showcasing a 25.9% gap after more than 2 hours of computation. Unfortunately, none of the obtained solutions turn out to be feasible, quite to the contrary: the typically obtained final probability value is roughly 0.04 being far from the required 0.7. We have also performed a run of a sampled problem with 10 000 random realizations, but with a significantly higher probability level of 0.9. In this case, after roughly one hour of computation, the resulting solution being at a MIPgap of 3.3 %, is still highly infeasible having only a probability value of 0.02.

Furthermore, we have tested IPOPT solver for the problem resolution: tests have been performed for the six initial points listed in Table 3.1 and default tolerance 10^{-4}. After at most 26 seconds of computation, IPOPT halted with highly infeasible points with a probability constraint value equal to 0. The difficulty of generating feasible points might come from the form of probability distribution as the probability level sharply raises from zero (blue, Figure 3.2) to nearly 1 (yellow, Figure 3.2) when approaching the feasible area from most directions, which causes the loss of gradient information in a large zone of zero probability. But of course this gradient information is not exploited by the MILP formulation at all.

In contrast, as Table 3.1 shows, the `CwC-PBM` algorithm manages to improve the probability level if the starting point is infeasible, and to improve the objective function value while satisfying probability constraint for a feasible initial point. Moreover, for one of the

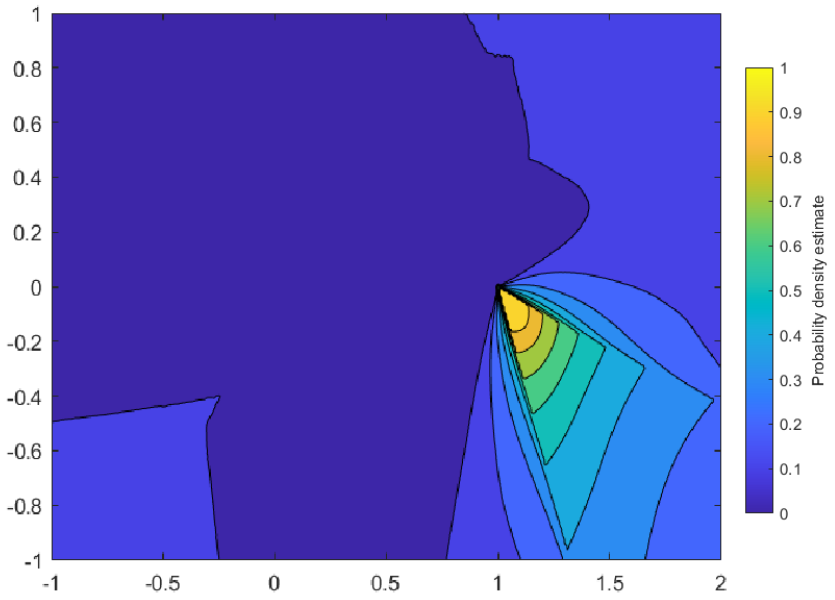


Fig. 3.2 Probability distribution associated with the chance constraint in (3.7.1)

tested starting points we have managed to generate the near (globally) optimal solution $(1.2400; -0.1126)$. Since the problem is indeed very difficult, a precise internal sampling scheme for the probability function is required. This amounts to the number of samples used to compute a formula of the type (3.7.4), which subsequently leads to design of the oracle for c_2 component. We can prematurely end further sampling, very much like the implementation of Genz' code [50], by checking if sampling variance - in fact the confidence interval bounds - are sufficiently small. Unlike Genz' code a crude antithetic Monte-Carlo scheme has been used for sampling, thus leaving much room for significant improvement in terms of speed and precision, by using for instance QMC as in [58].

Table 3.1 Results obtained with the `CwC-PBM` algorithm for problem (3.7.1) depending on initialization.

Initial point	Time (s)	Iterations	Initial probability level	Final probability level	Objective value
$(1.2, -0.1)^\top$	47.6	17	0.80	0.70	-1.1097
$(1.5, -0.4)^\top$	78.4	28	0.43	0.70	-1.1346
$(1.6, -0.3)^\top$	36.9	14	0.42	0.70	-1.0815
$(1.7, -0.4)^\top$	59.8	20	0.36	0.70	-1.1443
$(1.1, -0.4)^\top$	21.3	10	0.43	0.43	-0.7001
$(1, -0.8)^\top$	10.8	9	0.18	0.18	-0.2001

3.7.2 Investment like problems

We will see how the structure of Example 3.1.1 can appear in practice. Here we will follow the general discussion in [135]. We are interested in the situation wherein we dispose of a set of different technologies $i = 1, ..., m$ capable of generating electricity. Each technology comes with a specific and detailed set of constraints P_i, cost function c_i attributing to $p_i \in \mathbb{R}^T$ the cost of generation. Altogether, the various technologies are meant to ensure the satisfaction of a given customer load $d \in \mathbb{R}^T$. We are interested in finding the optimal mix. Thus for $i = 1, ..., m$, we are given $\theta_i \in \mathbb{N}$, the number of "units" of a given type we would like to invest in. The vector θ comes with an investment cost $F(\theta)$. In a deterministic setting this would amount to solving

$$\begin{aligned} \min_{\theta \in \Theta, p_i^j \in P_i} \quad & F(\theta) + \sum_{i=1}^{m} \sum_{j=1}^{\theta_i} c_i(p_i^j) \\ \text{s.t.} \quad & \sum_{i=1}^{m} \sum_{j=1}^{\theta_i} p_i^j \geq d. \end{aligned}$$

Now should for each i, the mappings c_i as well as the feasible sets P_i be convex, then it must be so that the averaged solution: $p_i^* = \frac{1}{\theta_i}\sum_{j=1}^{\theta_i}(p_i^j)^*$, in which each power plant of technology i produces exactly this amount is also optimal. This follows from using convexity of P_i showing feasibility of p_i^* and through using Jensen's inequality for c_i. This is also exactly what would happen if we would solve the subproblems of the Lagrangian dual (w.r.t. the load constraint) for a given fixed investment vector. The convexifying effect of the Lagrangian is well known, e.g., [78, 133] and thus for this dual setting convexity of c_i or P_i would not be essential. Either way, as a result we may thus assume that each power plant of the same technology produces the same amount. This would thus lead to the simpler problem (less variables):

$$\begin{aligned} \min_{\theta\in\Theta, p_i\in P_i} \quad & F(\theta) + \sum_{i=1}^{m} \theta_i c_i(p_i) \\ \text{s.t.} \quad & \sum_{i=1}^{m} \theta_i p_i \geq d. \end{aligned}$$

We will investigate a two-stage stochastic version of the last problem, wherein d is uncertain. We thus define:

$$Q(\theta,\xi) = \min_{p_i\in P_i} \sum_{i=1}^{m} \theta_i c_i(p_i) \text{ s.t. } \sum_{i=1}^{m} \theta_i p_i \geq d. \tag{3.7.2}$$

and consider the optimization problem

$$\min_{\theta\in\Theta} F(\theta) + \mathbb{E}[Q(\theta,\xi)], \tag{3.7.3}$$

where for the sake of simplicity we will assume θ to be allowed to take continuous values (Θ is a polytope). We will also assume that the feasible set P_i is convex, although one could investigate problem (3.7.3) without this assumption - for instance by arguing through Lagrangian duality.

Let us now look at a concrete instance. We will consider a time horizon of $t = 1,...,T$ time steps where each time step is considered to be Δt hours long. The problem disposes of m types of technology, having the following characteristics. Each technology type has a maximum power output level $p_i^{\mathtt{mx}}$, proportional cost c_i and gradient condition g_i. Additionally, each unit is assumed to dispose of a carbon emission rate e_i, and the system subject to a carbon cost f. Concretely this means that the proportional cost gets updated through the formula $c_i \leftarrow c_i + fe_i$.

The system is moreover endowed with a given customer load that we will assume to be multivariate Gaussian with a given mean and positive definite Covariance matrix. We refer to

[125, § 5] for the description of P_i (polyhedral). Furthermore for the various technologies we will assume that $F(\theta) = F^\top \theta$. The purpose of our experiment is to showcase how concretely the new algorithm can process specifically structured problems such as these.

Following the description of Example 3.1.1, we need to compute $Q^\varepsilon(\theta, \xi)$ at each given θ. The latter involves the solution of a convex optimization problem, wherein ψ_t is given by the t-th component of $d - \sum_{i=1}^m \theta_i p_i$. As a result of the logarithm, the objective function defining Q^ε is convex in y. We will therefore use a cutting-plane approach to internally compute Q^ε, as well as it's gradient. The inner optimization is initialized from the optimal solution y_0 of the inner optimization problem of $Q(x, \xi)$. The latter can be computed by solving a linear program. This will give us the oracle for f_2 (in the notation of Subsection 3.4.1).

Table 3.2 provides the concrete data.

Table 3.2 Data for the stylized investment problem.

	1	2	3	4	5	6	7	8	9	10
$p^{\mathtt{mx}}$ (MW)	900	900	900	300	300	200	200	200	100	10000
g (MW/h)	100	100	100	30	30	20	20	70	70	5000
c (€/MWh)	30	35	37	45	47	60	100	110	150	10000
$F^{\mathtt{inv}}$ (€)	493151	493151	493151	41096	41096	32877	32877	32877	21918	0
e (t/MW)	0	0	0	1	1	0.5	0.5	0.5	1.1	0

We can observe that the last unit described in the previous table is an imbalance unit - a computational trick to ensure that one can always meet the load, in this case even despite a potentially completely unbalanced set of invested assets. In terms of constraints on investment, we do not allow investment in this last unit, the capacity will remain at 1. The cost of investment was set up using typical values of investment cost per kW, upon rescaling to match T and while accounting for life span of the various technologies. The data of the case is stylized and the general purpose of the study is more a demonstration of the capabilities of the algorithm rather than an attempt to provide practical insights into investment regarding the electrical system.

We have performed tests with the `CwC-PBM` algorithm and IPOPT solver, both applied to approximation with Q^ε, $\varepsilon = 10^{-2}$, for $f = 0$ (zero carbon cost) and $f = 100$. The latter case enables to see the potential impact of such a penalized setting for emitting technologies. The methods provide comparable results in terms of objective value with the average relative difference of 0.98% (both for default tolerance 10^{-4}), see Table 3.3. (Both algorithms exploit the problem's decomposable structure and employ the same oracles.) The average execution time (among 6 considered cases) is 2579 seconds for the `CwC-PBM` algorithm,

and constitutes to 2241 seconds for IPOPT (including function evaluation). For initial states $v_2 = (1\,1\,1\,1\,1\,0\,1\,0\,1\,1)^\top$, the resulting total amount of installed capacity is less for the case with $f = 100$ (around 3500 MW for both methods) compared to the case with zero carbon cost (more than 3900 MW for both methods). While the latter remains close to the nominal investment vector, the methods update the set of installed capacities quite significantly, by shifting essentially all generation from carbon emitting technologies to technologies $1-3$ not emitting CO2 at all for the case with $f = 100$. This can be explained since the original setting was slightly overcapacitated - and as a result of the introduction of the "fictive" imbalance unit - infeasibility is no longer an issue. The maximum load over the considered scenarios being roughly 3500 as well. A similar situation occurs for the initial state $v_3 = (1\,1\,1\,1\,1\,1\,1\,1\,1\,1)^\top$ (roughly, 4000 MW for $f = 100$ against 3700 MW for zero carbon cost). For $v_1 = (1\,1\,1\,1\,1\,0\,0\,0\,0\,1)^\top$, the solution found by IPOPT corresponds to 3700 MW of installed capacity for $f = 100$, compared to approximately 3500 MW suggested by the `CwC-PBM` algorithm. This numerical experiment thus clearly shows that the `CwC-PBM` algorithm provides meaningful good quality solutions for this type of problems.

Table 3.3 Results obtained with `CwC-PBM` algorithm and IPOPT solver for (3.7.3) for carbon cost $f = 0$ and 100. Average execution time is 2579 seconds for the `CwC-PBM` algorithm and 2241 seconds for IPOPT.

		`CwC-PBM` algorithm		IPOPT solver	
Initial state	CO2 cost	Iterations	Objective value	Iterations	Objective value
$v_1 = (1\,1\,1\,1\,1\,0\,0\,0\,0\,1)^\top$	0	49	5 781 983	16	5 864 083
	100	33	5 886 797	10	6 001 847
$v_2 = (1\,1\,1\,1\,1\,0\,1\,0\,1\,1)^\top$	0	29	5 965 697	9	5 865 690
	100	31	5 938 787	15	5 920 723
$v_3 = (1\,1\,1\,1\,1\,1\,1\,1\,1\,1)^\top$	0	16	5 985 021	6	5 970 187
	100	32	5 945 049	11	5 930 397

Setting the prox-parameter μ^0 equal to its value at the first serious step μ^{l_1} predictably speeds up the `CwC-PBM` algorithm's performance. For initialization at v_2 and v_3, the prox-parameter μ^{l_1} is of the order of 10^6, while it is equal to its default value of 10^2 for initialization at v_1. The number of iterations decreases significantly for the former cases: down to 15 and 24 in the case of v_2, and 9 and 28 in the case of v_3, with an average improvement

of 27% in execution time. When altering the stopping test tolerance `Tol` within the set $\{10^{-4}, 10^{-6}, 10^{-8}, 10^{-12}\}$ (with fixed $\mu^0 = \mu^{l_1}$), we observe that the objective value does not change for all considered initial states and carbon costs. To progress to the next tolerance level within the specified set, the algorithm undergoes between 1 and 4 iterations, resulting in duration ranging from 44 to 456 seconds depending on the case. Moreover, the objective value shows minimal variation with alterations in the parameter κ. Varying κ within the set $\{0.001, 0.01, 0.03, 0.1, 0.3\}$, the objective value remains within the limits of $\pm 0.01\%$ (tests have been performed with the initial state v_2). The algorithm is thus robust with respect to the choice of `Tol` and κ and can be accelerated with the selection of an appropriate prox-parameter μ.

3.7.3 Decision dependent probability constraints in two stage problems

In this section we consider the following stochastic problem having different random vectors:

$$\begin{cases} \min\limits_{x\in X} & f_1(x) + \sum\limits_{s=1}^{S} \pi_s Q(x;\xi^s) \\ \text{s.t.} & \mathbb{P}[A_1 x + b_1 \geq \omega_1] \geq p_1 \end{cases} \quad \text{with } Q(x;\xi) := \begin{cases} \min\limits_{y\in Y} & q(x,y;\xi) \\ \text{s.t.} & \mathbb{P}_x[A_2(\xi)y + b_2(\xi) \geq \omega_2(x)] \geq p_2. \end{cases}$$

In this problem, $\xi \in \Xi := \{\xi^1, \ldots, \xi^S\}$, $\omega_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$, and $\omega_2(x) \sim \mathcal{N}(\mu_2(x), \Sigma_2(x))$. The latter random vector depends on the first-stage decisions. We assume that the covariance matrices Σ_1 and $\Sigma_2(x)$ are positive definite for all $x \in X$. As a result, the probability functions are twice-differentiable [60, 134]. Furthermore, as the multivariate Gaussian distribution is log-concave, we get that $c_1(x) = \log(p_1) - \log(\mathbb{P}[A_1 x + b_1 \geq \omega_1])$ is a convex function and so is the objective of the penalized subproblem

$$Q^\varepsilon(x;\xi) = \min_{y\in Y} \; q(x,y;\xi) - \tfrac{1}{\varepsilon} \log\Big(\mathbb{P}_x[A_2(\xi)y + b_2(\xi) \geq \omega_2(x)] - p_2\Big).$$

We are thus in the setting of Example 3.1.1 with $f_2(x) = \sum_{s=1}^{S} \pi_s[-Q^\varepsilon(x;\xi^s)]$. We can observe that the just given optimization problem is a version of two-stage stochastic program having unhedgeable, or post-decision random realizations.

Now in order to compute the gradient of both of the involved probability functions, we can rely on two different formulæ for the gradients. The mapping c_1 is continuously differentiable and its gradient can be evaluated by employing the results shown in [132]. The second stage probability function is also differentiable under a mild regularity condition, its gradient can be evaluated using the formulæ from [131, Thm. 5.1]. Indeed with $L_2(x)$ the

matrix resulting from the Cholesky decomposition $\Sigma_2(x) = L_2(x)L_2(x)^\top$, we may write

$$c_2^i(x,y) = \mathbb{P}[-A_2(\xi^i)y - b_2(\xi^i) + \mu_2(x) + L_2(x)\omega_2 \leq 0],$$

where $\omega_2 \in \mathbb{R}^m$, $\omega_2 \sim \mathcal{N}(0,I)$. Hence we can observe that:

$$\nabla c_2^i(x,y) = \int\limits_{v\in\mathbb{S}^{m-1}: J^*(v)\neq\emptyset, |J^{**}(v)|=1} -\frac{\chi(\hat{\rho}(v))}{(L_2(x))_{j(v)}v}\left(\hat{\rho}(v)\nabla(L_2(x))_{j(v),.}v) + \nabla\mu_2(x), -(A_2)^\top_{j(v),:}\right) d\mu_\zeta(v) \tag{3.7.4}$$

with

$$\begin{aligned} J^*(v) &= \left\{ j = 1,...,r \ : (L_2(x))_j v > 0 \right\} \\ \hat{\rho}(v) &= \min_{j\in J^*(v)} \frac{A_2 y + b_2 - \mu_2(x)}{(L_2(x))_j v} \\ J^{**}(v) &= \left\{ j \in J^*(v) \ : \hat{\rho}(v) = \frac{A_2 y + b_2 - \mu_2(x)}{(L_2(x))_j v} \right\}. \end{aligned}$$

and $j(v)$ being the unique element of $J^{**}(v)$. In this case since L_2 has linearly independent rows - which is the case since Σ_2 is positive definite - the aforementioned regularity condition (R2CQ) holds true. In fact (R2CQ) holds true locally and as a consequence it is indeed so that both c_1 and c_2 are twice continuously differentiable. This was already clear for c_1 upon using well known classic arguments.

Let us now consider the following concrete example of a problem of this kind. We are interested in a situation considering a manufacturer capable of producing two different products. The first-stage decision variables of the problem consist of setting prices for the products and an advertisement levels. The price will be assumed to be in relation to the average second-stage demand for the given product. We will use the following rule $\mu_2(x) = (\bar{\mu}_1/x_1, \bar{\mu}_2/x_2)$, with x_1, x_2 being the price levels for product 1 and 2 respectively. Advertisement is assumed to have a beneficial effect on the variance of the demand, but simultaneous advertisement for both products will be counterproductive. In other words:

$$\Sigma_2(x) := \begin{bmatrix} (0.1\bar{\mu}_1/x_1x_3)^2 & -0.4(0.01\bar{\mu}_1/x_1x_3\bar{\mu}_2/x_2x_4) \\ -0.4(0.01\bar{\mu}_1/x_1x_3\bar{\mu}_2/x_2x_4) & (0.1\bar{\mu}_2/x_2x_4)^2 \end{bmatrix}.$$

The second stage decision y involves the production of the goods. The production process of the goods is subject to some possible unreliability as the amount of actually produced goods are concerned. The matrix A_2 is thus a diagonal matrix, where the first entry is a uniform random variable over the interval $[0.9, 1]$ - on average only 95% of the commissioned

products actually get manufactured. The second diagonal entry is uniform over the interval $[0.8,1]$ - the process of production here is more unreliable. However producing with the more unreliable process is slightly cheaper. Any products that are manufactured but not sold, will incur a penalty. The second stage cost function is thus given by

$$q(x,y,\xi) = (2-x_1)y_1 + (1-x_2)y_2 + 12\mathbb{E}[\max\{y_1 - (\omega_2)_1, 0\}] + 12\mathbb{E}[\max\{y_2 - (\omega_2)_2, 0\}].$$

The last two terms correspond to the penalization of produced, but not sold goods. It turns out that the latter expectations can be computed "analytically" as they are related to the computation of an expectation of a truncated Gaussian random variable. Therefore, we can observe that the following identity holds true:

$$\begin{aligned}\mathbb{E}[\max\{y_1 - (\omega_2)_1, 0\}] = \Phi((y_1 - (\mu_2(x))_1)x_1/(0.1\bar{\mu}_1 x_3))(y_1 - (\mu_2(x))_1) \\ - \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}((y_1-(\mu_2(x))_1)x_1/(0.1\bar{\mu}_1 x_3))^2} 0.1\bar{\mu}_1/x_1 x_3.\end{aligned}$$

The second formula is of course immediately deduced as it is analogous. Both products require a different setup of the factory, so $y_1 + y_2 \leq 10$. Furthermore, the first-stage cost function is related to the cost of advertisement $f_1(x) = q_1 x_3^2 + q_2 x_4^2$. Furthermore all first-stage variables are bounded.

The implementation of this example requires first the implementation of the formulae for the gradient of the probability function. Here we can exploit the earlier given formula immediately. It can be observed (see the more extensive discussion in [126]) that the probability value itself can be computed with exactly the same cost. Subsequently the algorithm scheme is very similar to the one of the investment problem. In particular, combining the computations for probability function value and subgradient (3.7.4) with the reasoning of the previous example, we will obtain the oracle for f_2 component (in the notation of Subsection 3.4.1), while the oracle for f_1 is straightforward from the formula of the advertisement cost.

Therefore, we have also run this case with the `CwC-PBM` algorithm and found an approximate critical solution after a total of 12 iterations (1600 seconds on personal laptop). The found solution is $x = (3.37, 3.21, 0.096, 0.784)$, showing that there is interest in balancing the prices, i.e., not taking maximal prices, while also investing in advertisement. We have done the same test with IPOPT solver. The computation was aborted after 50000 seconds with the resulting infeasible point $x = (9.997, 9.983, 0.0004, 0.167)$ slowly approaching the bound $(10,10,0,0)$.

This example thus shows that the new algorithm allows us to consider settings beyond classic convexity, even when dealing with probability functions - in this case with decision dependent random vectors.

3.7.4 Compressed sensing problem

In this section we focus on the problem of compressed sensing considered in [152]:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \|x\|_1 - \|x\| \\ \text{s.t.} \quad & \|Ax-b\|^2_{LL_2,\gamma} \leq \delta, \end{aligned} \tag{3.7.6}$$

where $A \in \mathbb{R}^{q\times n}$ is a full row rank matrix and $b \in \mathbb{R}^q$. For given $\gamma > 0$, Lorentzian norm $\|\cdot\|_{LL_2,\gamma}$ of a vector $y \in \mathbb{R}^q$ is defined as

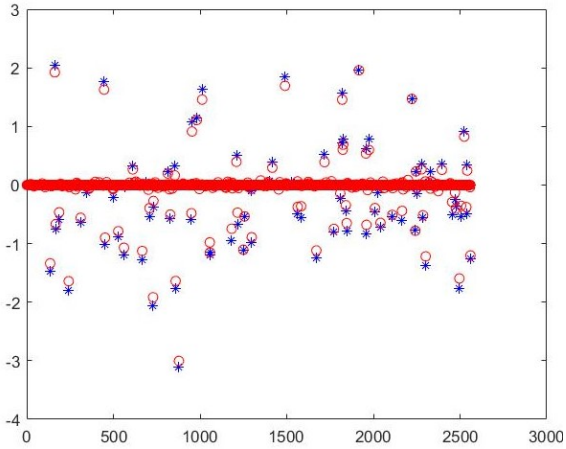
$$\|y\|^2_{LL_2,\gamma} = \sum_{i=1}^{q} \log\left(1+\frac{y_i^2}{\gamma^2}\right).$$

As discussed in [152], the problem (3.7.6) is DoC with twice continuously differentiable constraint, whose Lipschitz constant of the gradient is known. This allows us to construct an oracle for the constraint component. For the objective function oracle, the sign function is chosen as a subgradient of the component $f_1(x) = \|x\|_1$.

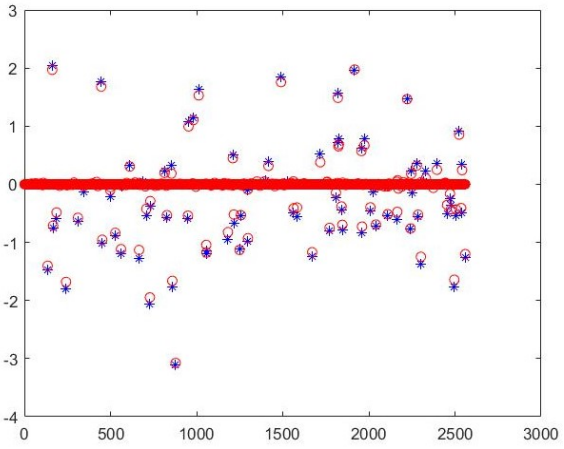
As in [152], we have generated $A \in \mathbb{R}^{q\times n}$ with normally distributed random entries normalizing it so that each column has a unit norm. To set the original point, we have chosen a subset of size $s_0 = [\frac{q}{9}]$ among basis vectors and generated a s_0-sparse vector x_{orig} with i.i.d. normally distributed random entries. We have taken $b = Ax_{orig} + 0.01\eta$, each η_i having a standard Cauchy distribution, and $\delta = 1.1\|0.01\eta\|_{LL_2,\gamma}$ with $\gamma = 0.02$.

We have performed tests with the `CwC-PBM` algorithm and SCP$_{ls}$ algorithm coded based on Algorithm 2.1 of the paper [152] for $q = 720\,i$, $n = 2560\,i$, with $i = 1$ (Figure 3.3) and $i = 5$ (Figure 3.4). Both algorithms were initialized at $x_0 = A^+b$ with matrix A^+ denoting the Moore-Penrose pseudoinverse of A. The gap between solution provided by the SCP$_{ls}$ algorithm (with tolerance 10^{-4}) and `CwC-PBM` algorithm is 4.45% and 2.71% for 2400 and 3000 iterations, respectively, for the case $i = 1$. It constitutes 2.2% and 1.16% for 2400 and 3000 iterations, respectively, for the case $i = 5$, Table 3.4. However, the execution time of the `CwC-PBM` algorithm is higher compared to SCP$_{ls}$, which was designed for a more specific framework (constraint functions have Lipschitz continuous gradient, and the objective function is decomposed as a sum of a smooth function and DoC function), Table 3.4.

We have also run the CwC-PBM algorithm with initialization at zero vector. It manages to obtain a feasible solution after 50 iterations (for both $i = 1$ and 5), as well as to recover an optimal solution within tolerance 3.2×10^{-4} for $i = 1$ (10 000 iterations) and 1.2×10^{-3} for $i = 5$ (3 000 iterations) with execution time of 1 347 seconds and 2 717 seconds, respectively. The SCP$_{ls}$ algorithm cannot be applied in this case, as a feasible initial point is required.



CwC-PBM algorithm.



SCP$_{ls}$ algorithm.

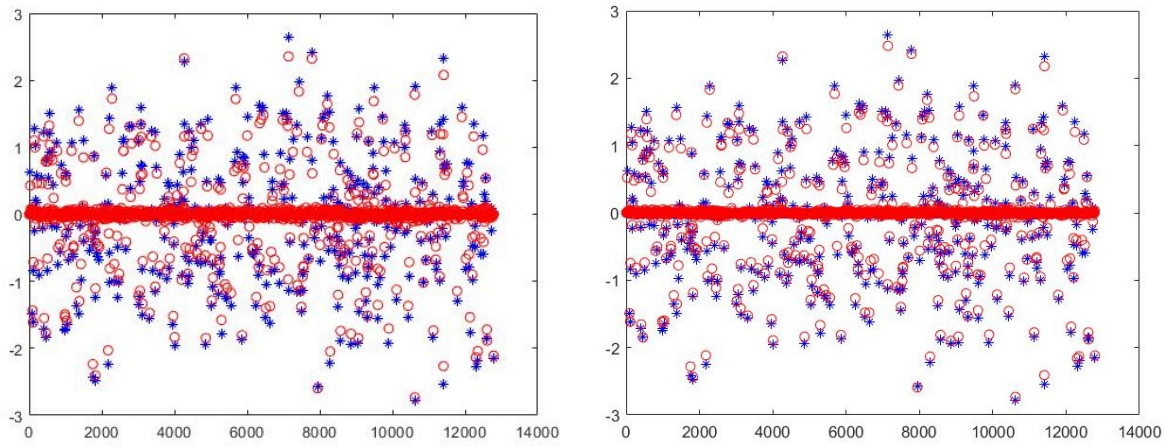
Fig. 3.3 Computed solutions (marked by circle) of (3.7.6) and x_{orig} (marked by asterisk) for $i = 1$

Table 3.4 Results obtained with CwC-PBM and SCP$_{ls}$ algorithms for (3.7.6) with the initial point $x_0 = A^+ b$.

		CwC-PBM algorithm		SCP$_{ls}$ algorithm	
	Iterations	Time (s)	Objective value	Time (s)	Objective value
i=1	2 400	324.27	54.16	16.02	51.85
	3 000	592.92	53.26		
i=5	2 400	2 396.65	289.52	133.37	283.27
	3 000	3 180.52	286.56		

3.8 Chance-Constrained Optimal Power Flow

This section is dedicated to the operational planning problem of distribution power grid under uncertainties related to the probabilistic nature of nodal generation and consumption considered in Chapter 2. It is formulated as a chance-constrained AC-OPF where the objective



`CwC-PBM` algorithm. SCP$_{ls}$ algorithm.

Fig. 3.4 Computed solutions (marked by circle) of (3.7.6) and x_{orig} (marked by asterisk) for $i = 5$

function represents the operational planning cost, the deterministic constraints are convex and reflect contractual engagements related to the operation of the grid, and the joint probability constraint incorporates the stochastic nature of the model. More precisely, it ensures that for an operational decision, grid operating conditions remain within technical limits with a given probability $1-\alpha$ (security level). The content of this section has been published in [124].

First, the chance-constrained OPF is reformulated as a DoC-constrained DoC problem (2.3.6) (with convex objective function, i.e. $f_2 = 0$). This reformulation is made with the use of Oracle 1 and the sample average approximation, which provides oracles for the DoC constraint as given by formula (2.3.5). The resulting model is then solved using two algorithms: the DoC bundle method (`PBMDC`^{2}) from [129] and the `CwC-PBM` algorithm.

In order to ensure that `PBMDC`^{2} provides a critical point of (2.3.6) with `Tol` $= 0$, the function c_2 needs to be differentiable as stated in [129, Thm. 2] and Theorem 2.3.1. This requirement does not hold for our problem. According to Theorem 3.5.9, this requirement is not necessary for the `CwC-PBM` algorithm. If `Tol` > 0, the `PBMDC`^{2} algorithm provides an approximate critical point of (3.3.16), while the `CwC-PBM` algorithm provides an approximate critical point of (3.3.13). Lemma 3.3.7 states that the latter condition (3.3.13) is stronger than (3.3.16) in the case when c_2 is not differentiable.

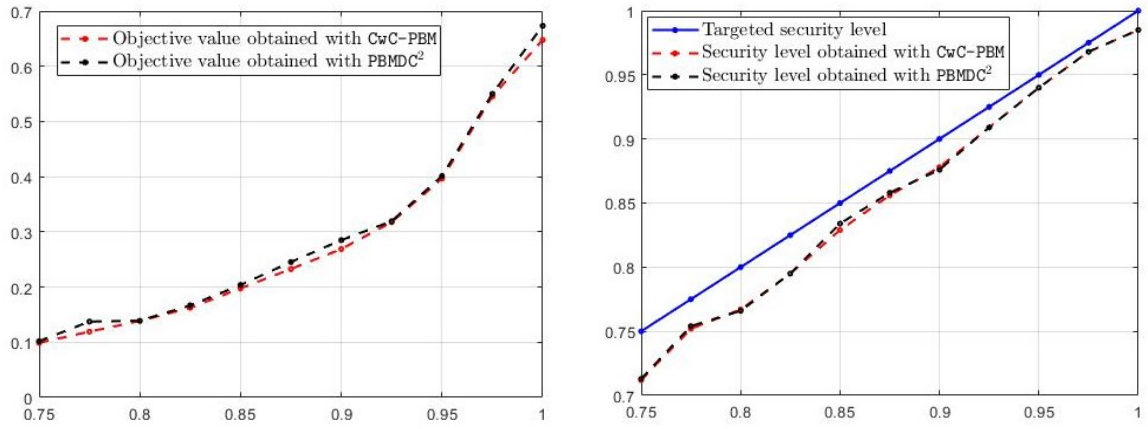
We consider the use case from Subsection 2.4.4. The parameters ρ, σ, κ and `Tol` for both methods are set to: $\rho = 10^7$, $\sigma = 0.5$, $\kappa = 0.3$ and `Tol` $= 10^{-5}$. The parameter μ^0 is chosen to be 80 for the `CwC-PBM` algorithm, while $\mu^0 = 10^2$, $\mu_{\min} = 10^{-6}$ and $\mu_{\max} = 10^6$ for `PBMDC`^{2}. We set 11 values of the security level $1-\alpha$ ranging from 0.75 to 1 with a step

size of 0.025. The initial vector $(\mathbf{p}^0, \mathbf{q}^0)$ is equal to zero, which corresponds to the case with no power modulation.

The `CwC-PBM` algorithm provides better solutions in terms of the objective value for all values of security level $1-\alpha$, Figure 3.5 (a). The maximal improvement of 13.33% is obtained for $1-\alpha = 0.775$ and constitutes 3.56% on average. As explained in Subsection 2.4.4, the obtained security level is less than $1-\alpha$ (Figure 3.5 (b)) due to the DoC approximation of the probability constraint. The relative maximal difference between the targeted and obtained security level is 5.07% for the `CwC-PBM` algorithm compared to 4.93% for the `PBMDC`^{2} algorithm (both at $1-\alpha = 0.75$). Nevertheless, solutions supplied by both algorithms are feasible for the DoC model (2.3.6). This means that more cost-effective solutions provided by the `CwC-PBM` algorithm are still feasible, and the difference in obtained security levels is related to the DoC approximation of the chance-constrained problem.

The average execution time is less for the `PBMDC`^{2} algorithm: 1 340 seconds compared to 4 232 for `CwC-PBM`. However, the objective value corresponding to `PBMDC`^{2}'s solution is attained after 2 387 seconds on average. The latter difference can be explained by the fact that we have not exploited the DoC structure in the `CwC-PBM` algorithm.

This use case thus shows that the new algorithm is applicable to real-life industrial problems and capable of generating approximate critical points in the DoC-constrained framework, without assumptions on the differentiability of DoC components. However, achieving stronger criticality comes at the cost of an increase in execution time.



(a) Objective value (b) Security level

Fig. 3.5 Objective value and security level obtained with `CwC-PBM` and `PBMDC`^{2} algorithms.

3.9 Conclusion

In this chapter, we have considered nonsmooth and nonconvex optimization problems where the objective function and nonlinear constraint are represented as the difference of convex and weakly convex functions (CwC). Our work studies various stationary conditions and a bundle method approach enabling to compute critical (generalized KKT) points. The latter broadens and enhances the algorithm developed in [129] for the case of DoC-constrained DoC-problems, and likewise relies on problem reformulation via an improvement function. To the best of our knowledge, the proposed method is the first one that directly exploits the CwC-structure of the involved functions and does not require additional assumptions or transformations as, for instance, explicit DoC decompositions or Moreau envelopes. We have illustrated the method performance with a few stochastic problems, including two-stage and chance-constrained problems, and a compressed sensing problem with nonlinear constraint. Preliminary results are meaningful and show that the algorithm can tackle settings beyond the classic DoC setting. For the chance-constrained optimal power flow modeling power modulation and curtailment levers, the method provides a more cost-effective solution. However, achieving stronger criticality leads to an increase in execution time.

Chapter 4

Integrating Priority and Fairness

The French Energy Code establishes the main principles and the guidelines for consumers and producers access to the grid [46]. In particular, it introduces alternative grid connection offers for energy producers [45]. Although the rules of power modulation are not described explicitly, the spirit of this law is also reflected in informative documents provided by DSOs, such as in [35].

Two concepts that emerge from the spirit of these documents are priority levels and fairness in power modulation. The first concept, priority levels, generalizes the SCP and FiT mechanisms: the higher the modulation priority, the lower the cost of power modulation associated with that level, as it reflects power that is more readily accessible according to the grid connection contract. The second concept, fairness, ensures that all producers assigned the same priority level equally share the responsibility in resolving a grid constraint to which they contribute. These principles are not standardized in industry practice yet, and as a result, there is no universally accepted definition or model for them.

The model in Chapter 2, which does not employ binary variables, handles priority levels through differences in power modulation costs. This formulation effectively prioritizes the modulation of SCP power over the curtailment of FiT power for the same producer. However, for different users, this approach is not universally applicable: if curtailing one user is more effective in resolving the constraint than curtailing another, it may be more cost-efficient to curtail that user's FiT power, even when SCP power from another user is still available. Furthermore, as mentioned in Subsection 2.4.5, power curtailment is not distributed equally among users of the same type, despite the quadratic penalty (2.4.1c) intended to promote fairness in power modulation.

In this chapter, we bring the model closer to real-life requirements with the use of binary variables and additional logical and discrete constraints. We begin by defining the concepts of priority levels, followed by the priority and fairness principles in power modulation. Next,

we develop an optimization model that accurately integrates these considerations. Finally, we study some symmetry properties of the resulting model.

The main content of this chapter was presented at the PGMO Days 2024 conference, held in Palaiseau, France, in November 2024.

Le Code de l'énergie établit les principes fondamentaux et les lignes directrices régissant l'accès des consommateurs et des producteurs au réseau public d'électricité [46]. Il prévoit notamment des offres de raccordement alternatives pour les installations de production d'énergies renouvelables [45]. Bien que les règles de modulation de puissance ne soient pas explicitement définies par la loi, ses principes directeurs apparaissent également dans les documents techniques des gestionnaires de réseau de distribution (GRD), en particulier dans [35].

Deux concepts émergent de ces documents : les niveaux de priorité et l'équité dans la modulation de puissance. Le premier concept, les niveaux de priorité, généralise les mécanismes ORI et ORR : plus la priorité de modulation est élevée, plus le coût de modulation associé à ce niveau est faible, car il correspond à une puissance plus accessible selon le contrat de raccordement. Le second concept, l'équité, garantit que tous les producteurs d'un même niveau de priorité partagent équitablement la responsabilité de résoudre toute contrainte réseau à laquelle ils contribuent. Ces principes ne sont pas encore standardisés dans la pratique industrielle et, par conséquent, il n'existe pas de définition ou de modèle universellement accepté.

Le modèle présenté au Chapitre 2, qui n'utilise pas de variables binaires, gère les niveaux de priorité par des différences de coûts de modulation de puissance. Cette formulation privilégie effectivement la modulation de puissance ORI par rapport à la réduction de puissance ORR pour le même producteur. Cependant, pour des utilisateurs différents, cette approche n'est pas systématiquement applicable : si la réduction de puissance d'un utilisateur résout plus efficacement la contrainte réseau que celle d'un autre, la solution optimale peut conduire à réduire la puissance ORR de cet utilisateur, même en présence de puissance ORI disponible chez d'autres. Comme souligné dans la Sous-section 2.4.5, les réductions de puissance ne sont pas équitablement réparties entre utilisateurs de même type, malgré la pénalité quadratique (2.4.1c) *visant à garantir l'équité de modulation.*

Ce chapitre rapproche le modèle des exigences réelles par l'intégration de variables binaires et de contraintes logiques et discrètes. Le concept de niveaux de priorité est d'abord présenté, suivi des principes de priorité et d'équité pour la modulation de puissance. Ces considérations sont ensuite intégrées au modèle d'optimisation. Enfin, certaines propriétés de symétrie du modèle sont étudiées.

Le contenu principal de ce chapitre a été présenté aux PGMO Days 2024 (Palaiseau, France, novembre 2024).

4.1 Main principles and rules

4.1.1 Priority levels

A set of priority levels is assigned to each producer, which determines the order in which power modulation is applied. For the same priority level, all producers to whom it is assigned and who are contributing to the same constraint of the grid, must be equitably involved in its resolution. Let the priority levels be denoted by $1,\dots,M$, from the highest to the lowest, and let $\mathscr{N}_m$ represent the set of producers for whom priority level m is available according to the rules.

Different rules for assigning levels can be considered here. For instance, if a producer has a priority level m, they also have all priority levels k where $k \geq m$. This corresponds to the case where $\mathscr{N}_m \subset \mathscr{N}_k$ if $m \leq k$. Another possibility is to assume that higher priorities are assigned to larger producers. Curtailment limits for each level are defined as a fraction of the producer's installed or guaranteed power. Therefore, if a larger producer has a higher priority than a smaller one, a greater amount of power can be curtailed at a lower cost. As a result, assigning higher priority to larger producers (one or two levels to each) could present a cost-effective strategy for a DSO. In this chapter and Chapter 5, we do not assume any specific rule. For simplicity, the reader may consider the first case, where $\mathscr{N}_m \subset \mathscr{N}_k$ if $m \leq k$.

Consumer curtailment is only initiated when no additional power can be curtailed from producers to resolve a grid constraint, meaning that consumers are assigned the lowest priority.[1] In our model, we assume that consumer curtailment can be started simultaneously, i.e. they share a single priority level. As a result, we do not introduce any specific notation for the consumer priority level. However, this assumption is not essential to the methodology we apply in Chapter 5. The model can easily be generalized to the cases where consumers have different priority levels or where different groups of consumers follow specific curtailment rules.

4.1.2 Critical nodes and impacted elements

In what follows, we consider the same grid model as described in Chapter 2. In particular, we assume that among the end buses, there is one slack bus and other buses with at most

[1]This rule for consumer curtailment is designed for the case where high-voltage constraint resolution is prioritized, as curtailing producers and consumers has opposite effects on grid constraints.

one connected grid user. Each grid user is either a producer or a consumer. Potential grid constraints include voltage constraints and thermal constraints on current transit (congestion constraints). We maintain the notation from Chapter 2, while introducing new notation to describe the principles of priority and fairness.

Denote by C the set of grid constraints, including both voltage and congestion constraints, which are active for at least one scenario. For each element $J \in C$, we define a set $C(J)$ of *critical nodes* that includes producers and consumers influencing the constraint J. The set $C(J)$ can be described explicitly as follows:

- if J denotes a congestion constraint, the set $C(J)$ includes all producers and consumers downstream of the constraint: the distribution grid is operated in a radial topology, and according to Kirchhoff's current law, only the power modulation of downstream users affects J;
- if J denotes a voltage constraint, the set $C(J)$ includes all producers and consumers downstream of the same feeder (primary line extending from a substation) to which J belongs.

For a more sophisticated model, it would be useful to further distinguish between types of constraints (e.g. high-voltage or low-voltage constraints) and types of users whose curtailment contributes to resolving these constraints (producers or consumers, respectively) to define the sets of critical nodes. However, these details are beyond the scope of our model, as they do not change the type of the resulting optimization problem.

Next, we introduce the notion of *impacted elements* for each producer and consumer $i \in \mathcal{N}$. Denote by $C^{-1}(i)$ the set of all voltage constraints associated with the same feeder as the bus i, and congestion constraints upstream of bus i. The sets $C(J)$ and $C^{-1}(i)$ are related by the following property: the bus i belongs to the set $C(J)$ if and only if J belongs to the set $C^{-1}(i)$.

We assume for simplicity that there is only one feeder. Therefore, if J denotes a voltage constraint, $C(J)$ refers to all producers and consumers in the grid. Meanwhile, the set $C^{-1}(i)$ includes all voltage constraints of the grid, as well as congestion constraints upstream of bus i. The sets of critical nodes and impacted elements are fixed and depend only on the grid's topology, as we do not consider any changes in the latter.

As an example, consider a small grid shown in Figure 4.1, with grid users connected at the buses $2-6$ (each bus hosting either a producer or a consumer). Suppose there is a voltage constraint, J_1, at bus 3, and a congestion constraint, J_2, at the line connecting buses 2 and 4. In this case, the set C consists of two elements: J_1 and J_2. The set of critical nodes for the voltage constraint, $C(J_1)$, includes 5 grid users at buses $2-6$, while the set of critical

nodes for the congestion constraint, $C(J_2)$, includes 3 grid users at buses $4-6$. For grid users at buses $4-6$, the sets of impacted elements include both the voltage and congestion constraints. The sets of impacted elements for grid users at buses 2 and 3 include only the voltage constraint.

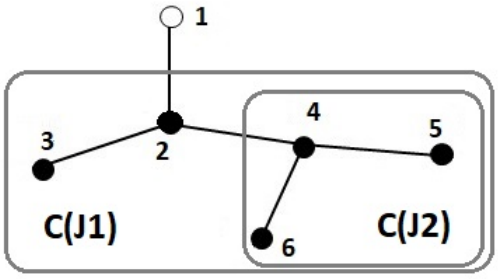


Fig. 4.1 Example for a small grid.

In the use cases considered in Subsection 2.4.4, the only constraints are the voltage constraints at nodes 9, 10, 11, 15, 16, and 17. The set of critical nodes for each of these voltage constraints includes all 31 buses, which accommodate 3 producers and 28 consumers. The set of impacted elements for each producer and consumer includes all the voltage constraints, specifically those at nodes 9, 10, 11, 15, 16, and 17. In the use cases in Subsection 2.4.5, there is an additional congestion constraint on the line connecting nodes 2 and 19. Consequently, the set of critical nodes for this congestion constraint includes buses 19, 20, 21, and 22 (see Figure 2.3), while the set of critical nodes for each voltage constraint remains the same as in the previous case. The sets of impacted elements for buses 19, 20, 21, and 22 include the congestion constraint and all voltage constraints. For other producers and consumers, the sets of impacted elements remain as in the previous case, including only the voltage constraints.

We say that grid constraints J_1 and J_2 are equivalent if sets of critical nodes $C(J_1)$ and $C(J_2)$ coincide, i.e. contain the same set of buses, and $J_1 \leq J_2$ if $C(J_2)$ contains all buses from $C(J_1)$. The relation $\leq$ is a partial order relation on the set C of grid constraints. The greatest element of the partially ordered set $(C, \leq)$ corresponds to any voltage constraint if one exists. We will use this consideration in Subsection 4.2.1, while exploring the symmetry properties of the model.

4.1.3 Modeling the priority levels

For a producer $i \in \mathscr{G}$, we introduce decision variables $\boldsymbol{p}_i^m$, $\boldsymbol{q}_i^m$ for the active and reactive power modulated at level m, respectively. Let $\bar{p}_i^m \geq 0$ denote the upper bound on active

power curtailment at level m for producer i. For simplicity, we assume the lower bound is 0, meaning we consider power curtailment with non-negative $\boldsymbol{p}_i^m \geq 0$. For a consumer $i \in \mathscr{L}$, we introduce decision variables $\boldsymbol{p}_i^{\mathscr{L}}$ and $\boldsymbol{q}_i^{\mathscr{L}}$ for the active and reactive power curtailment, respectively, along with the lower bound on active power curtailment $\underline{p}_i^{\mathscr{L}} \leq 0$. Note that, in contrast to Chapter 2, we separate consumers from the priority levels in the notation, in particular $\mathscr{N}_m \subset \mathscr{G}$.

With this notation, the resulting active and reactive power for a scenario ξ at bus i, where a user is connected, is given by

$$\begin{aligned} p_i &= p_i^{\phi}(\xi) - \Big(\sum_{m=1}^{M} \boldsymbol{p}_i^m\Big), & q_i &= q_i^{\phi}(\xi) - \Big(\sum_{m=1}^{M} \boldsymbol{q}_i^m\Big), & &\text{if } i \in \mathscr{G} \\ p_i &= p_i^{\phi}(\xi) - \boldsymbol{p}_i^{\mathscr{L}}, & q_i &= q_i^{\phi}(\xi) - \boldsymbol{q}_i^{\mathscr{L}}, & &\text{if } i \in \mathscr{L}. \end{aligned}$$

To simplify the notation, if there is no need to distinguish priority levels and user types, we will denote decision variables on power modulation by

$$\mathbf{p} = \big(\boldsymbol{p}_i^m, \boldsymbol{p}_j^{\mathscr{L}}, \quad i \in \mathscr{N}_m, m = 1, \ldots, M, \quad j \in \mathscr{L}\big) \tag{4.1.1a}$$

$$\mathbf{q} = \big(\boldsymbol{q}_i^m, \boldsymbol{q}_j^{\mathscr{L}}, \quad i \in \mathscr{N}_m, m = 1, \ldots, M, \quad j \in \mathscr{L}\big). \tag{4.1.1b}$$

For a grid constraint $J \in C$ and level m, let $T_{J,m} \in \{0,1\}$ be a binary variable that indicates whether curtailment of producers up to the m-th priority level is authorized for activation to resolve the grid constraint J (with $T_{J,m} = 1$ if authorized, and $T_{J,m} = 0$ otherwise). Additionally, we introduce binary variables $T_{J,\mathscr{L}} \in \{0,1\}$ to indicate whether curtailment of consumers is authorized to resolve the grid constraint J. For each consumer $i \in \mathscr{L}$, we associate a binary variable $T_i^{\mathscr{L}} \in \{0,1\}$, which indicates whether curtailment for this specific consumer is authorized for activation. To simplify the notation, if there is no need to distinguish the types of the introduced binary variables, we will denote them by

$$T = \big(T_{J,m}, T_{J,\mathscr{L}}, T_i^{\mathscr{L}}, \quad J \in C, m = 1, \ldots, M, \quad i \in \mathscr{L}\big) \in \{0,1\}^{|T|}.$$

4.1.4 Priority and fairness principles

We base our model on the following priority principle for producer power modulation: for a given constraint J, curtailment at critical nodes of priority level $m+1$ cannot be activated unless the available active power at critical nodes from all previous levels k, with $k \leq m$ has

already been curtailed to its maximum. We model this principle as follows:

$$T_{J,m+1} \leq T_{J,m}, \qquad J \in C, \quad m = 1,\dots,M-1 \tag{4.1.2a}$$

$$T_{J,m+1} \leq \frac{\boldsymbol{p}_i^m}{\bar{p}_i^m}, \qquad J \in C,\ i \in C(J) \cap \mathscr{N}_m, \quad m = 1,\dots,M-1. \tag{4.1.2b}$$

The first constraint states that the activation of level $m+1$ is permitted only if all previous levels have been authorized for activation. The second constraint explicitly models the requirement that active power at the previous level must be curtailed to its maximum. By induction, this requirement extends to all previous levels.

The same principle applies to consumers: for a given constraint J, consumer curtailment at critical nodes cannot be activated unless the available active power from all producers at critical nodes has been curtailed to its maximum. We model this principle in a similar way:

$$T_{J,\mathscr{L}} \leq T_{J,M}, \qquad J \in C \tag{4.1.3a}$$

$$T_{J,\mathscr{L}} \leq \frac{\boldsymbol{p}_i^M}{\bar{p}_i^M}, \qquad J \in C,\ i \in C(J) \cap \mathscr{N}_M. \tag{4.1.3b}$$

For a consumer i, individual curtailment is permitted only if there is a constraint $J \in C^{-1}(i)$ where curtailment at the consumer level is authorized (i.e. $T_{J,\mathscr{L}} = 1$). We introduce the following constraints:

$$T_i^{\mathscr{L}} \leq \sum_{J \in C^{-1}(i)} T_{J,\mathscr{L}}, \qquad i \in \mathscr{L}. \tag{4.1.4}$$

There is no unique strategy for modeling the fairness principle [4]. Although a profound research and comparison of fairness models are beyond the scope of this work, we explored several approaches, including penalizing the maximum gap between the ratios of achieved and available curtailment for participating generation unit as proposed in [122]. Finally, we decided to integrate fairness using hard constraints, introducing a proportion of active power that should be uniformly curtailed to resolve each grid constraint.

For a grid constraint $J \in C$ and level m, let $\rho_{J,m} \in [0,1]$ denote the proportion of active power curtailed at level m for all producers connected to critical nodes $C(J)$. We introduce the notation:

$$\rho = \big(\rho_{J,m}, \quad J \in C, m = 1,\dots,M\big) \in [0,1]^{|\rho|}.$$

This proportion $\rho_{J,m}$ is equal to zero, if curtailment at the m-th priority level is not authorized for grid constraint J. This requirement is modeled as follows:

$$\rho_{J,m} \leq T_{J,m}, \qquad J \in C, \quad m = 1,\dots,M. \tag{4.1.5}$$

For a producer $i \in C(J)$, the resulting curtailed power consists of the active power curtailed to resolve each constraint $J \in C^{-1}(i)$:

$$\boldsymbol{p}_i^m = \sum_{J \in C^{-1}(i)} \rho_{J,m} \cdot \bar{p}_i^m, \qquad i \in \mathscr{N}_m, \quad m = 1,\dots,M. \tag{4.1.6}$$

In this equation $\sum_{J \in C^{-1}(i)} \rho_{J,m} \leq 1$, which is implicitly incorporated into the model by imposing an upper bound on $\boldsymbol{p}_i^m$.

An additional factor based on the sensitivity matrix can be introduced to each term in the sum to adjust the influence of the curtailed power on the constraint J. This adjustment term does not change the type of the resulting optimization problem, we will thus assume that it is equal to 1.

The variables $\rho_{J,m}$ explicitly capture the fairness of modulation for each constraint J and level m, which is the primary motivation behind our modeling choice. However, the combination of constraints (4.1.2b) and (4.1.6) imposes very restrictive conditions on the activation of the next level of producer curtailment $T_{J,m+1}$. Similarly, constraints (4.1.3b) and (4.1.6) impose strict limitations on the activation of consumer curtailment $T_{J,\mathscr{L}}$. As a result, the model yields feasible but potentially more expensive solutions compared to a version where these constraints (or some of them) are relaxed.

4.2 New model with priority and fairness rules

In this section, we adjust and complement model (2.2.4) with additional constraints to incorporate priority and fairness principles discussed above.

Constraints (2.2.4b) - (2.2.4f) of model (2.2.4) are modified in the new framework as follows:

$$0 \leq \boldsymbol{p}_i^m \leq \bar{p}_i^m, \qquad i \in \mathscr{N}_m, \quad m = 1,\dots,M \tag{4.2.1a}$$

$$\boldsymbol{q}_i^m = \tan\phi_i \, \boldsymbol{p}_i^m, \qquad i \in \mathscr{N}_m, \quad m = 1,\dots,M \tag{4.2.1b}$$

$$\underline{p}_i^{\mathscr{L}} \leq \boldsymbol{p}_i^{\mathscr{L}} \leq 0, \qquad i \in \mathscr{L} \tag{4.2.1c}$$

$$\boldsymbol{q}_i^{\mathscr{L}} = \tan\phi_i \, \boldsymbol{p}_i^{\mathscr{L}}, \qquad i \in \mathscr{L}. \tag{4.2.1d}$$

We add additional constraints that state that the curtailed power of consumers is zero if their curtailment is not individually authorized:

$$|\boldsymbol{p}_i^{\mathscr{L}}| \leq T_i^{\mathscr{L}} \cdot |\underline{p}_i^{\mathscr{L}}|, \qquad i \in \mathscr{L}. \tag{4.2.2}$$

Clearly, constraints (4.2.2) and (4.2.1c) can be expressed more succinctly as follows

$$T_i^{\mathscr{L}} \cdot \underline{p}_i^{\mathscr{L}} \leq \boldsymbol{p}_i^{\mathscr{L}} \leq 0, \qquad i \in \mathscr{L}.$$

We maintain the stochastic aspect as in model (2.2.4). To formulate it accurately, we specify the power flow equations for our framework (here the sum is taken over all buses k connected to the bus i, including bus i):

$$L_i^R(\mathbf{p}, \delta, |V|, \xi) := \tag{4.2.3}$$
$$\begin{cases} p_i^{\phi}(\xi) - \left(\sum_{m=1}^{M} \boldsymbol{p}_i^m\right) + \sum_{k\sim i} Y_{i,k}^R |V_i||V_k| \cos(\delta_i - \delta_k) + \sum_{k\sim i} Y_{i,k}^I |V_i||V_k| \sin(\delta_i - \delta_k), & \text{if } i \in \mathscr{G} \\ p_i^{\phi}(\xi) - \boldsymbol{p}_i^{\mathscr{L}} + \sum_{k\sim i} Y_{i,k}^R |V_i||V_k| \cos(\delta_i - \delta_k) + \sum_{k\sim i} Y_{i,k}^I |V_i||V_k| \sin(\delta_i - \delta_k), & \text{if } i \in \mathscr{L}, \end{cases}$$
$$L_i^I(\mathbf{q}, \delta, |V|, \xi) :=$$
$$\begin{cases} q_i^{\phi}(\xi) - \left(\sum_{m=1}^{M} \boldsymbol{q}_i^m\right) + \sum_{k\sim i} Y_{i,k}^R |V_i||V_k| \sin(\delta_i - \delta_k) - \sum_{k\sim i} Y_{i,k}^I |V_i||V_k| \cos(\delta_i - \delta_k), & \text{if } i \in \mathscr{G} \\ q_i^{\phi}(\xi) - \boldsymbol{q}_i^{\mathscr{L}} + \sum_{k\sim i} Y_{i,k}^R |V_i||V_k| \sin(\delta_i - \delta_k) - \sum_{k\sim i} Y_{i,k}^I |V_i||V_k| \cos(\delta_i - \delta_k), & \text{if } i \in \mathscr{L}. \end{cases}$$

The set of constraints on the grid state and the definition of $X(\xi)$ remain the same as in Chapter 2, with the power flow equations specified by (4.2.3):

$$\underline{\delta_i} \leq \delta_i \leq \overline{\delta_i}, \qquad i \in \mathscr{N} \setminus \{sb\} \tag{4.2.4a}$$
$$\underline{|V_i|} \leq |V_i| \leq \overline{|V_i|}, \qquad i \in \mathscr{N} \setminus \{sb\} \tag{4.2.4b}$$
$$l_{i,j}(|V_i|, |V_j|, \delta_i, \delta_j) \leq (I_{i,j}^{\max})^2, \qquad (i,j) \in \mathscr{A} \tag{4.2.4c}$$
$$(p_{sb}, q_{sb}) \in \mathscr{F}_{sb}, \tag{4.2.4d}$$

$$X(\xi) := \left\{ (\mathbf{p}, \mathbf{q}) \;\middle|\; \begin{array}{c} \text{there exist } |V|, \delta, p_{sb}, q_{sb} \text{ satisfying (4.2.4)}, \\ L_i^R(\mathbf{p}, \delta, |V|, \xi) = 0 \text{ for all } i \in \mathscr{N}, \\ L_i^I(\mathbf{q}, \delta, |V|, \xi) = 0 \text{ for all } i \in \mathscr{N} \end{array} \right\}, \tag{4.2.5}$$

as well as the chance constraint:

$$\mathbb{P}\big[(\mathbf{p}, \mathbf{q}) \in X(\xi)\big] \geq 1 - \alpha.$$

The objective function represents the term (2.4.1b) using the new notation. Given coefficients $c_m^i \geq 0$, $m = 1, \ldots, M$, $i \in \mathscr{G}$, and $c_{\mathscr{L}} \leq 0$ (since $\boldsymbol{p}^{\mathscr{L}}$ is negative), we minimize

$$f(\mathbf{p}) := \sum_{m=1}^{M} \sum_{i \in \mathscr{N}_m} c_m^i \boldsymbol{p}_i^m + \sum_{i \in \mathscr{L}} c_{\mathscr{L}} \boldsymbol{p}_i^{\mathscr{L}}.$$

We then incorporate constraints (4.1.2) - (4.1.6), resulting in the following optimization problem:

$$\min_{\mathbf{p},\mathbf{q},T,\rho} \sum_{m=1}^{M} \sum_{i \in \mathscr{N}_m} c_m^i \boldsymbol{p}_i^m + \sum_{i \in \mathscr{L}} c_{\mathscr{L}} \boldsymbol{p}_i^{\mathscr{L}} \tag{4.2.6a}$$

$$\text{s.t. } T \in \{0,1\}^{|T|}, \quad \rho \in [0,1]^{|\rho|} \tag{4.2.6b}$$

$$0 \leq \boldsymbol{p}_i^m \leq \bar{p}_i^m, \qquad i \in \mathscr{N}_m, \quad m = 1, \ldots, M \tag{4.2.6c}$$

$$\boldsymbol{q}_i^m = \tan\phi_i \, \boldsymbol{p}_i^m, \qquad i \in \mathscr{N}_m, \quad m = 1, \ldots, M \tag{4.2.6d}$$

$$\underline{p}_i^{\mathscr{L}} \leq \boldsymbol{p}_i^{\mathscr{L}} \leq 0, \qquad i \in \mathscr{L} \tag{4.2.6e}$$

$$\boldsymbol{q}_i^{\mathscr{L}} = \tan\phi_i \, \boldsymbol{p}_i^{\mathscr{L}}, \qquad i \in \mathscr{L} \tag{4.2.6f}$$

$$|\boldsymbol{p}_i^{\mathscr{L}}| \leq T_i^{\mathscr{L}} \cdot |\underline{p}_i^{\mathscr{L}}|, \qquad i \in \mathscr{L} \tag{4.2.6g}$$

$$T_{J,m+1} \leq T_{J,m}, \qquad J \in C, \quad m = 1, \ldots, M-1 \tag{4.2.6h}$$

$$T_{J,m+1} \leq \frac{\boldsymbol{p}_i^m}{\bar{p}_i^m}, \qquad J \in C, \; i \in C(J) \cap \mathscr{N}_m, \quad m = 1, \ldots, M-1 \tag{4.2.6i}$$

$$T_{J,\mathscr{L}} \leq T_{J,M}, \qquad J \in C \tag{4.2.6j}$$

$$T_{J,\mathscr{L}} \leq \frac{\boldsymbol{p}_i^M}{\bar{p}_i^M}, \qquad J \in C, \; i \in C(J) \cap \mathscr{N}_M \tag{4.2.6k}$$

$$T_i^{\mathscr{L}} \leq \sum_{J \in C^{-1}(i)} T_{J,\mathscr{L}}, \qquad i \in \mathscr{L} \tag{4.2.6l}$$

$$\rho_{J,m} \leq T_{J,m}, \qquad J \in C, \quad m = 1, \ldots, M \tag{4.2.6m}$$

$$\boldsymbol{p}_i^m = \sum_{J \in C^{-1}(i)} \rho_{J,m} \cdot \bar{p}_i^m, \qquad i \in \mathscr{N}_m, \quad m = 1, \ldots, M \tag{4.2.6n}$$

$$\mathbb{P}\big[(\mathbf{p},\mathbf{q}) \in X(\xi)\big] \geq 1 - \alpha. \tag{4.2.6o}$$

4.2.1 Symmetry properties

In this subsection, we explore symmetries in this model and identify valid inequalities to help reduce the solution space, by forbidding some redundant binary combinations.

Proposition 4.2.1. *The symmetry-breaking constraints below are valid and non-redundant inequalities for model* (4.2.6)*:*

$$T_{\tilde{J},m} + T_{J,m+1} - 1 \leq T_{\tilde{J},m+1} \qquad \tilde{J} \leq J, \quad J,\tilde{J} \in C \quad m = 1,\dots,M-1 \quad \text{(4.2.7a)}$$

$$T_{\tilde{J},M} + T_{J,\mathscr{L}} - 1 \leq T_{\tilde{J},\mathscr{L}} \qquad \tilde{J} \leq J, \quad J,\tilde{J} \in C. \quad \text{(4.2.7b)}$$

Proof. Consider a feasible solution (p,q,T',ρ) of (4.2.6) with $T'_{\tilde{J},m} = T'_{J,m+1} = 1$ and $T'_{\tilde{J},m+1} = 0$ for some $J,\tilde{J} \in C$ with $\tilde{J} \leq J$ and $m \in \{1,\dots,M-1\}$. These conditions do not violate any constraint in (4.2.6). Moreover, $C(\tilde{J}) \subseteq C(J)$ and $T'_{J,m+1} = 1$ with (4.2.6i) imply $\frac{\boldsymbol{p}_i^m}{\bar{p}_i^m} = 1$ for all $i \in C(\tilde{J}) \cap \mathscr{N}_m$, and $T'_{\tilde{J},m+1} = 0$ with (4.2.6m) imply $\rho_{\tilde{J},m+1} = 0$. Let define vector T identical to T' except for one term $T_{\tilde{J},m+1} = 1$, then the solution (p,q,T,ρ) satisfies $1 = T_{\tilde{J},m+1} \leq T_{\tilde{J},m} = 1$ in (4.2.6h) and $1 = T_{\tilde{J},m+1} \leq \frac{\boldsymbol{p}_i^m}{\bar{p}_i^m} = 1$ for all $i \in C(\tilde{J}) \cap \mathscr{N}_m$ in (4.2.6i). Therefore, this solution is feasible for (4.2.6), it has the same cost as (p,q,T',ρ), and it satisfies (4.2.7a) while the solution (p,q,T',ρ) does not. The proof is similar for (4.2.7b): we can enforce $T_{\tilde{J},\mathscr{L}} = 1$ whenever $T_{\tilde{J},M} = T_{J,\mathscr{L}} = 1$ in any solution of (4.2.6). □

We will enforce constraints (4.2.7) to model (4.2.6) in Chapter 5, while studying different approaches to resolve it. As an alternative, such conditions could be enforced as dominance rules to remove redundant branches in a Branch-and-bound method, instead as explicit constraints in the model: when setting a binary variable $T_{J,m+1}$ (or $T_{J,\mathscr{L}}$) to 1, we can fix $T_{\tilde{J},m+1} = 1$ (or $T_{\tilde{J},\mathscr{L}}$) for all dominated constraint $\tilde{J}$ such that $\tilde{J} \leq J$, thus branching decisions may focus on the maximal elements in each chain of the partially ordered set of critical nodes. We will discuss this approach further in Chapter 6.

Now, we consider the case where all grid constraints belong to the same equivalent class, in particular when there are only voltage constraints. The next proposition shows that, in this case, the voltage constraints can be aggregated into one, reducing the size of the model accordingly.

Proposition 4.2.2. *In case $C(J) = \mathscr{N}$ for all $J \in C$ and $C^{-1}(i) = C$ for all $i \in \mathscr{N}$, then constraints (4.2.6h-4.2.6j), and (4.2.6m-4.2.6n) can be replaced in model* (4.2.6) *by:*

$$T_{m+1} \leq T_m, \qquad m = 1,\dots,M-1$$

$$T_{m+1} \leq \rho_m, \qquad m = 1,\dots,M-1$$

$$T_{J,\mathscr{L}} \leq T_M, \qquad J \in C$$

$$\rho_m \leq T_m, \qquad m = 1,\dots,M$$

$$\boldsymbol{p}_i^m = \rho_m \cdot \bar{p}_i^m, \qquad i \in \mathscr{N}_m, \quad m = 1,\dots,M.$$

on variables $T_m \in \{0,1\}$ *and* $\rho_m \geq 0$ *for* $m = 1,\ldots,M$ *in place of* $(T_{J,m}, \rho_{J,m})_{J\in C}$.

Proof. We show that to any feasible solution (p,q,T',ρ) of (4.2.6) corresponds a feasible solution (p,q,T,ρ) of same cost such that $T_{\tilde{J},m} = T_{J,m}$ for all $J,\tilde{J} \in C$ and $m \in \{1,\ldots,M-1\}$. This is obvious if $T' \equiv 0$, so let $\tilde{m} \geq 1$ be the highest level such that $T'_{\tilde{J},\tilde{m}} = 1$ for some $\tilde{J} \in C$. Define for all m, $\rho_m := \sum_{J\in C} \rho_{J,m}$, then $\rho_m = \frac{\boldsymbol{p}_i^m}{\bar{p}_i^m} \leq 1$ for $i \in \mathcal{N}_m$ according to (4.2.6n) and (4.2.6c). By definition, and given (4.2.6m), $T'_{J,m} = \rho_m = 0$ for all $J \in C$ and $m > \tilde{m}$. According to (4.2.6h), (4.2.6i), and (4.2.6n), $T'_{\tilde{J},m} = \rho_{m-1} = 1$ for all $m \leq \tilde{m}$. Thus, the desired feasible solution (p,q,T,ρ) is obtained by defining vector T identical to T' except possibly for terms $T_{J,m} = 1$ for all $J \in C$ and $m \leq \tilde{m}$ and it satisfies (4.2.8) for $T_m := T_{\tilde{J},m}$. Hence, we can enforce constraints $\rho_m \leq T_m$ for $m \in \{1,\ldots,M\}$ together with $T_{J,m} = T_m$, for all $J,\tilde{J} \in C$ and $m \in \{1,\ldots,M-1\}$ in model (4.2.6), or, alternatively, identify all components $T_{J,m}$ to only one T_m leading to reformulation (4.2.8). □

4.3 Conclusion

Inspired by the core principles and guidelines of the French Energy Code, we define the concept of priority levels and formulate the corresponding priority and fairness principles in power modulation. We extend the model from Chapter 2 by incorporating deterministic logical and discrete constraints that reflect these principles. Finally, we explore several properties of the resulting model to reduce the number of constraints and restrict the solution space. In the next chapter, we will focus on optimization approaches for solving the resulting model.

Chapter 5

Optimization approaches to the model with priority and fairness rules

In this chapter, we explore several approaches to solving model (4.2.6) with discrete constraints. In Section 5.1 we consider different integrations of binary variables into bundle methods and related theoretical challenges. Then, we test the proposed approaches with the DoC bundle method.

In Section 5.2, we consider an alternative approach that consists of adding an additional binary variable for each scenario. The resulting optimization problem features two types of nonconvexity: one related to binary variables and the other arising from alternating current power flow equations. By employing a scenario decomposition method, we can separate these issues and break down the stochastic component into a deterministic AC-OPF problem per scenario. This approach results in a mixed-integer linear problem (MILP) or mixed-integer quadratic problem (MIQP) that captures priority rules, and parallelizable AC-OPFs that ensure technical feasibility.

The main content of this chapter was presented at the PGMO Days 2024 conference, held in Palaiseau, France, in November 2024.

Ce chapitre explore différentes méthodes de résolution du modèle (4.2.6) *avec des variables binaires et contraintes discrètes. La Section 5.1 présente différentes stratégies d'intégration de variables binaires au sein des méthodes de faisceaux, tout en abordant les défis théoriques associés. Les approches d'intégration proposées sont testées avec la méthode de faisceaux DoC.*

La Section 5.2 propose une approche alternative consistant à ajouter une variable binaire supplémentaire par scénario. Le problème d'optimisation résultant présente deux types de non-convexité : l'une liée aux variables binaires et l'autre provenant des équations de flux de

puissance en courant alternatif. Grâce à une méthode de décomposition par scénarios, ces complexités peuvent être séparées, permettant de décomposer la composante stochastique en un problème AC-OPF déterministe par scénario. Cette approche conduit à un problème d'optimisation linéaire (MILP) ou quadratique (MIQP) en nombres entiers reflétant les règles de priorité, ainsi qu'à des problèmes AC-OPF parallélisables assurant la faisabilité technique.

Le contenu principal de ce chapitre a été présenté aux PGMO Days 2024 (Palaiseau, France, novembre 2024).

5.1 Integrating binary variables into a bundle method

We begin this section by noting that the binary variables $T \in \{0,1\}^{|T|}$ in model (4.2.6) appear only in the deterministic constraints, specifically in constraints (4.2.6g) - (4.2.6m), while the chance constraint (4.2.6o) does not involve binary variables. Therefore, the DoC reformulation described in Subsections 2.3.1 and 2.3.2 remains applicable. Meanwhile, the continuous constraints (4.2.6c) - (4.2.6f) are inherited from model (2.2.4), and the continuous constraint (4.2.6n) models fairness in power curtailment. In what follows, we will refer to the version of model (4.2.6) that includes only the continuous constraints (4.2.6c) - (4.2.6f), (4.2.6n), and the chance constraint (4.2.6o) as the *continuous version* of the model. The DoC bundle method can handle the continuous version of model (4.2.6), while all the challenges arise from the deterministic discrete constraints. We consider three different ways to incorporate them.

5.1.1 Defining three approaches

The first approach consists of integrating the discrete constraints directly into the master program of a bundle method applied to the continuous version of model (4.2.6). In this case, there is no theoretical guarantee of convergence or solution quality. The challenges related to the binary variables $T \in \{0,1\}^{|T|}$ are handled by a solver (with an integrated Branch-and-bound method) in the numerical experiments.

If there is no need to distinguish between types of binary variables $T \in \{0,1\}^{|T|}$, we will enumerate them as T_t. Relaxing the binary requirements, i.e. considering variables T_t as continuous ones, we introduce the additional constraints enforcing them to be equal to zero or one:

$$T_t(1-T_t) \leq 0, \quad 0 \leq T_t \leq 1, \qquad\qquad t \in \{1,\dots,|T|\}.$$

Since variables T_t are non-negative, these constraints are equivalent to

$$h_1(T) - h_2(T) \leq 0, \quad 0 \leq T_t \leq 1, \qquad t \in \{1, \ldots, |T|\} \tag{5.1.1}$$

with

$$h_1(T) = \sum_{t=1}^{|T|} T_t \quad \text{and} \quad h_2(T) = \sum_{t=1}^{|T|} T_t^2.$$

The two remaining approaches are based on replacing the requirement $T \in \{0,1\}^{|T|}$ by the nonconvex constraints (5.1.1) in model (4.2.6). There are two options: integrating the DoC constraint $h_1(T) - h_2(T) \leq 0$ either as a soft constraint, by adding a penalization term to the objective function, or as a hard constraint. In the latter case, the resulting model will include two DoC constraints, which can be combined into a single DoC constraint by taking their maximum. We will discuss the convergence results applicable in this case for the DoC and CwC bundle methods.

According to Theorem 3.5.9, any solution provided by the CwC bundle method (with $\texttt{Tol} = 0$) satisfies the criticality condition (3.3.13) related to the problem of minimizing the improvement function (3.3.11). If the binary requirement $T \in \{0,1\}^{|T|}$ is satisfied (that we want to obtain), then T is a critical point for $\min_{T \in [0,1]^{|T|}} h_1(T) - h_2(T)$. In this case, according to the second case (ii) of Theorem 3.5.9, the stronger criticality condition (3.3.9) does not necessarily hold.

For the DoC bundle method, a similar situation occurs. According to [129, Thm. 1] and Theorem 2.3.1, any solution provided by the DoC bundle method (with $\texttt{Tol} = 0$) satisfies the criticality condition (3.3.16) related to the problem of minimizing the corresponding improvement function. To obtain a critical point to the original DoC model, c_2 must be continuously differentiable ([129, Thm. 2] and Theorem 2.3.1). Since we take the maximum between the DoC constraint approximating the chance constraint and the DoC constraint $h_1(T) - h_2(T) \leq 0$, this requirement may be satisfied at the solution point if only the latter constraint is active. In this case, $T \in \{0,1\}^{|T|}$ holds, which corresponds to the case (iii) of Theorem 2 in [129]. The point (iii) provides a certain necessary qualification condition, the extended Mangasarian–Fromovitz constraint qualification (EMFCQ), for the stronger criticality.

The EMFCQ condition holds at $\bar{x} \in X$ with $c(\bar{x}) = 0$ if there exists a direction $d \in \mathscr{T}_X(\bar{x})$ such that the corresponding generalized directional derivative is negative: $c^\circ(\bar{x}; d) = \max_{s \in \partial^{\mathrm{c}} c(\bar{x})} \langle s, d \rangle < 0$. In our case, $c(x) := h_1(T) - h_2(T)$ is smooth, and thus regular: $(h_1 - h_2)^\circ(T; d) = h_1'(T; d) - h_2'(T; d)$. For $T_t \in [0,1]$, the tangent cone is $\mathbb{R}_+ \cup \{0\}$ at

$T_t = 0$, and $\mathbb{R}_- \cup \{0\}$ at $T_t = 1$. Therefore, $h_1'(\bar{T};d) - h_2'(\bar{T};d) \geq 0$ at $\bar{T} \in \{0,1\}^{|T|}$ for all d from the tangent cone. Consequently, the EMFCQ condition does not hold.

As the qualification conditions are not satisfied for either bundle method at binary values of $T \in \{0,1\}^{|T|}$, and convergence to a critical point of the DoC model is not guaranteed, we apply the DoC bundle method. It is worth noting that in the case of soft constraint integration, there is only one DoC constraint approximating the chance constraint, and therefore the convergence analysis from Section 3.8 applies.

First approach: integrating binary variables into the master program. Integrating the discrete constraints into the master program in Step 4 of Algorithm 1 leads to the following program at iteration k:

$$\min_{\mathbf{p},\mathbf{q},T,\rho,r\in\mathbb{R}^4} \quad r_4 - \left\langle s_2^k, \begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix}\right\rangle + \quad \frac{\mu}{2}\left\| \begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} - \begin{pmatrix}\hat{\mathbf{p}}\\ \hat{\mathbf{q}}\end{pmatrix}\right\|^2 \tag{5.1.2}$$

s.t. $T \in \{0,1\}^{|T|}, \quad \rho \in [0,1]^{|\rho|}$

$$0 \le \boldsymbol{p}_i^m \le \bar{p}_i^m, \qquad i \in \mathscr{N}_m, \quad m = 1,\dots,M$$

$$\boldsymbol{q}_i^m = \tan\phi_i \, \boldsymbol{p}_i^m, \qquad i \in \mathscr{N}_m, \quad m = 1,\dots,M$$

$$\underline{p}_i^{\mathscr{L}} \le \boldsymbol{p}_i^{\mathscr{L}} \le 0, \qquad i \in \mathscr{L}$$

$$\boldsymbol{q}_i^{\mathscr{L}} = \tan\phi_i \, \boldsymbol{p}_i^{\mathscr{L}}, \qquad i \in \mathscr{L}$$

$$|\boldsymbol{p}_i^{\mathscr{L}}| \le T_i^{\mathscr{L}} \cdot |\underline{p}_i^{\mathscr{L}}|, \qquad i \in \mathscr{L}$$

$$T_{J,m+1} \le T_{J,m}, \qquad J \in C, \quad m = 1,\dots,M-1$$

$$T_{J,m+1} \le \frac{\boldsymbol{p}_i^m}{\bar{p}_i^m}, \qquad J \in C,\ i \in C(J) \cap \mathscr{N}_m, \quad m = 1,\dots,M-1$$

$$T_{J,\mathscr{L}} \le T_{J,M}, \qquad J \in C$$

$$T_{J,\mathscr{L}} \le \frac{\boldsymbol{p}_i^M}{\bar{p}_i^M}, \qquad J \in C,\ i \in C(J) \cap \mathscr{N}_M$$

$$T_i^{\mathscr{L}} \le \sum_{J \in C^{-1}(i)} T_{J,\mathscr{L}}, \qquad i \in \mathscr{L}$$

$$\rho_{J,m} \le T_{J,m}, \qquad J \in C, \quad m = 1,\dots,M$$

$$\boldsymbol{p}_i^m = \sum_{\tilde{J} \in C^{-1}(i)} \rho_{\tilde{J},m} * \bar{p}_i^m, \qquad i \in \mathscr{N}_m, \quad m = 1,\dots,M$$

$$f(\mathbf{p}^l) + \langle s_f^l, \mathbf{p} - \mathbf{p}^l \rangle \le r_1, \qquad l = 0,\dots,k$$

$$c_1\left(\begin{pmatrix}\mathbf{p}^l\\ \mathbf{q}^l\end{pmatrix}\right) + \left\langle s_1^l, \begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} - \begin{pmatrix}\mathbf{p}^l\\ \mathbf{q}^l\end{pmatrix}\right\rangle \le r_2, \qquad l = 0,\dots,k$$

$$c_2\left(\begin{pmatrix}\mathbf{p}^l\\ \mathbf{q}^l\end{pmatrix}\right) + \left\langle s_2^l, \begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} - \begin{pmatrix}\mathbf{p}^l\\ \mathbf{q}^l\end{pmatrix}\right\rangle \le r_3, \qquad l = 0,\dots,k$$

$$r_1 + r_3 - \hat{\tau}_f \le r_4$$

$$r_2 - \hat{\tau}_c \le r_4.$$

Second and third approaches: integrating an additional DoC constraint. To simplify the notation, let us denote the continuous variables as follows:

$$x := \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix}, \qquad x_\rho := \begin{pmatrix} x \\ \rho \end{pmatrix}.$$

A feasible set defined by the deterministic constraints (4.2.6c) - (4.2.6n) and $T \in [0,1]^{|T|}$, is convex. We denote this set as $X_\rho \times B$.

In the case of soft integration of the DoC constraint, we add a penalization term $\omega\big(h_1(T) - h_2(T)\big)$ with $\omega \geq 0$ to the objective function, while retaining the lower and upper bounds on T_t, $t \in \{1,\ldots,|T|\}$, as hard constraints.

Note that the proportion variables ρ are present only in the convex deterministic constraints, and thus they have no impact on oracles. The new oracles are defined by the following formulas:

$$f_1\begin{pmatrix} x_\rho \\ T \end{pmatrix} := f(x) + \omega h_1(T), \qquad f_2\begin{pmatrix} x_\rho \\ T \end{pmatrix} := \omega h_2(T),$$
$$c_1\begin{pmatrix} x_\rho \\ T \end{pmatrix} := c_1(x), \qquad c_2\begin{pmatrix} x_\rho \\ T \end{pmatrix} := c_2(x),$$

while the oracles for subgradients s_{c_1} and s_{c_2} remain unchanged (as in Section 2.3). Since the component f_2 is nonzero, the master program solved at Step 4 of Algorithm 1 must be modified. According to the theory presented in [129], it becomes:

$$
\begin{aligned}
\min_{(x_\rho,T)\in(X_\rho\times B),\,T\in[0,1]^{|T|},\,r\in\mathbb{R}^5} \quad & r_5 - \left\langle s^k_{f_2} + s^k_{c_2}, \begin{pmatrix} x_\rho \\ T \end{pmatrix} \right\rangle + \frac{\mu}{2}\left\| \begin{pmatrix} x_\rho \\ T \end{pmatrix} - \begin{pmatrix} \hat{x}_\rho \\ \hat{T} \end{pmatrix} \right\|^2 \\
\text{s.t. } & f_1 \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} + \left\langle s^l_{f_1}, \begin{pmatrix} x_\rho \\ T \end{pmatrix} - \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} \right\rangle \le r_1, \qquad l = 0,\dots,k \\
& f_2 \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} + \left\langle s^l_{f_2}, \begin{pmatrix} x_\rho \\ T \end{pmatrix} - \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} \right\rangle \le r_2, \qquad l = 0,\dots,k \\
& c_1 \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} + \left\langle s^l_{c_1}, \begin{pmatrix} x_\rho \\ T \end{pmatrix} - \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} \right\rangle \le r_3, \qquad l = 0,\dots,k \\
& c_2 \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} + \left\langle s^l_{c_2}, \begin{pmatrix} x_\rho \\ T \end{pmatrix} - \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} \right\rangle \le r_4, \qquad l = 0,\dots,k \\
& r_1 + r_4 - \hat{\tau}_f \le r_5 \\
& r_2 + r_3 - \hat{\tau}_c \le r_5.
\end{aligned}
$$

In the case of hard integration of constraints (5.1.1), we deal with two DoC constraints: $c_1(x) - c_2(x) \le 0$, approximating the chance constraint, and $h_1(T) - h_2(T) \le 0$. Since these constraints are not scaled, we introduce a parameter $\tilde{\omega} \ge 0$ to equalize their values. Representing them as a single DoC constraint $\max\{c_1(x) - c_2(x),\, \tilde{\omega}\big(h_1(T) - h_2(T)\big)\} \le 0$, we equivalently obtain:

$$
\max\{c_1(x) + \tilde{\omega}\, h_2(T),\, c_2(x) + \tilde{\omega}\, h_1(T)\} - \big((c_2(x) + \tilde{\omega}\, h_2(T)\big) \le 0.
$$

Similar to the previous case, the variables ρ appear only in convex deterministic constraints and have no impact on oracles. We define the new oracles by the following formulas:

$$
\begin{aligned}
c_1 \begin{pmatrix} x_\rho \\ T \end{pmatrix} &:= \max\{c_1(x) + \tilde{\omega}\, h_2(T),\, c_2(x) + \tilde{\omega}\, h_1(T)\}, \\
c_2 \begin{pmatrix} x_\rho \\ T \end{pmatrix} &:= c_2(x) + \tilde{\omega}\, h_2(T), \\
f \begin{pmatrix} x_\rho \\ T \end{pmatrix} &:= f(x),
\end{aligned}
$$

which also lead to the oracles for subgradients s_f, s_{c_1} and s_{c_2}. As the objective function is convex, the master program solved at Step 4 of Algorithm 1 remains unchanged:

$$\begin{aligned}
\min_{(x_\rho,T)\in(X_\rho\times B),T\in[0,1]^{|T|},r\in\mathbb{R}^4} \quad & r_4 - \left\langle s^k_{c_2}, \begin{pmatrix} x_\rho \\ T \end{pmatrix} \right\rangle \quad + \frac{\mu}{2}\left\| \begin{pmatrix} x_\rho \\ T \end{pmatrix} - \begin{pmatrix} \hat{x}_\rho \\ \hat{T} \end{pmatrix} \right\|^2 \\
\text{s.t.} \quad & f\begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} + \left\langle s^l_f, \begin{pmatrix} x_\rho \\ T \end{pmatrix} - \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} \right\rangle \le r_1, && l=0,\dots,k \\
& c_1\begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} + \left\langle s^l_{c_1}, \begin{pmatrix} x_\rho \\ T \end{pmatrix} - \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} \right\rangle \le r_2, && l=0,\dots,k \\
& c_2\begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} + \left\langle s^l_{c_2}, \begin{pmatrix} x_\rho \\ T \end{pmatrix} - \begin{pmatrix} x^l_\rho \\ T^l \end{pmatrix} \right\rangle \le r_3, && l=0,\dots,k \\
& r_1 + r_3 - \hat{\tau}_f \le r_4 \\
& r_2 - \hat{\tau}_c \le r_4.
\end{aligned}$$

5.1.2 Numerical results for integration into the DoC bundle method

We have constructed use cases based on the use case considered in Subsection 2.4.4 with the same network configuration, load and generation profiles, and voltage constraints. Due to the fact that only voltage constraints are present, the number of constraints in model (4.2.6) can be reduced, as discussed in Subsection 4.2.1. In this section, we have extended the set of priority levels for producers and adapted the coefficients in the objective function, which results in three different cases.

The first use case corresponds to a single priority level for producers, i.e. $M = 1$ and $\mathcal{N}_1$ includes all three producers connected in buses 12, 29 and 32. In this case, the priority constraints (4.2.6h) and (4.2.6i) are not involved.

For the second use case, we consider two priority levels ($M = 2$) with $\mathcal{N}_1$ including the producer connected in bus 12, and $\mathcal{N}_2$ including all three producers at buses 12, 29 and 32. This assignment of priority levels corresponds to that in Subsection 2.4.4.

For the third use case, we consider three priority levels ($M = 3$) with $\mathcal{N}_1$ including the producer connected in bus 12, $\mathcal{N}_2$ including the producers at buses 12 and 32, and $\mathcal{N}_3$ including all three producers at buses 12, 29 and 32.

The coefficients in the objective function (4.2.6a) are provided in Table 5.1 with a distinction between different types of generation.

First approach: integrating binary variables into the master program. Recall that the first approach consists of modifying solely the master programs (5.1.2) at each iteration k

Table 5.1 Coefficients in the objective function (4.2.6a) for each grid user (GU).

GU	Priority level	Type	Coefficient in (14a)
12	m=1 m=2 m=3	Biomass generation	$4.2 \cdot 10^{-5}$ $4.2 \cdot 10^{-4}$ $4.2 \cdot 10^{-3}$
29	m=1	Biomass generation	$4.2 \cdot 10^{-3}$
32	m=1 m=2	Wind generation	$2 \cdot 10^{-3}$ $2 \cdot 10^{-2}$
Others	-	Consumption	1

of Algorithm 1, with a solver handling the binary variables. The remaining task is to set an initial starting point $(\mathbf{p}^0, \mathbf{q}^0, T^0, \rho^0)$, which is a crucial step, as the approach is sensitive to this choice. However, we currently do not have reliable heuristics for selecting good starting points.

We initially ran the algorithm with the starting point corresponding to zero power curtailment, and set $T^0 = \mathbf{1}$ to allow the activation of all priority levels and curtailment of consumers. After 1 500 iterations, the algorithm did not perform any serious step in all three use cases. Then, we decided to begin with a solution of the continuous version of model (4.2.6), allowing the solver to search for a nearby actual solution. Specifically, we solve the continuous version in the 1st phase and denote its solution, provided by Algorithm 1, as $(\bar{\mathbf{p}}, \bar{\mathbf{q}}, \bar{\rho})$. In the 2nd phase, we set $\mathbf{p}^0 = \bar{\mathbf{p}}$, $\mathbf{q}^0 = \bar{\mathbf{q}}$ and $\rho^0 = \bar{\rho}$. This choice was based on the assumption that the solution of the continuous version is sufficiently close to a good solution for the actual (discrete) problem.

We have performed tests for the first and the second use cases, with $N = 1\,000$ scenarios and the safety parameter ranging from $1 - \alpha = 0.75$ to 1 with a step size of 0.025. In the 1st phase, for the continuous version of model (4.2.6), the execution time varies from 376 to 1 618 seconds, with an average of 1 048 seconds in the first use case, and from 1 536 to 2 241 seconds, with an average of 1 838 seconds in the second use case. Initialized at zero power curtailment values, the DoC bundle method manages to find feasible solutions of the continuous version with safety parameter values up to 0.9 except for $1 - \alpha = 0.875$. Figure 5.1 illustrates a drastic decrease in the ratio of scenarios satisfying the power flow equations (the obtained safety parameter) for the targeted ratios $1 - \alpha = 0.925$, 0.95 and 0.975, when infeasible solutions have been provided by the approach. Meanwhile, the curtailment of producers is activated at maximum for all values of the parameter $1 - \alpha$. Consequently, constraints (4.2.6j) and (4.2.6k) allow curtailment of consumers.

In the 2nd phase, the starting point is set to $(\mathbf{p}^0, \mathbf{q}^0, \rho^0) = (\bar{\mathbf{p}}, \bar{\mathbf{q}}, \bar{\rho})$, and T^0 is set to $\mathbf{1}$. It provides feasible solutions for all safety parameter values, with the execution time ranging

from 156 to 1 396 seconds, and an average of 1 046 seconds. As Figure 5.1 shows, the ratio of scenarios satisfying the power flow equations always remains higher for the 2nd phase (the mixed-integer version) than for the 1st phase (continuous version). The power curtailment cost is also higher: 1.8002 pu on average, against 0.2391 pu in the 1st phase, Figure 5.2. It is worth mentioning that for $1-\alpha=1$, the ratio of scenarios satisfying the power flow equations is 0.997. The corresponding power curtailment cost drastically increases up to 3.0942 pu, Figure 5.2.

The power curtailment is activated at maximum for all producers, while only several consumers are affected: three consumers if $1-\alpha=0.775$, 0.8 and 0.825, 13 consumers if $1-\alpha=1$, and four consumers for other security levels. In the 1st phase, the number of activated consumers varies from 6 to 25, with an average of 12.4. Therefore, the 2nd phase of the algorithm significantly decreases the number of consumers participating in power curtailment. However, the average volume of power curtailed per user increases by 15 times (an average value among all security levels). This explains the gap in power curtailment costs for the 1st phase and 2nd phase, Figure 5.2.

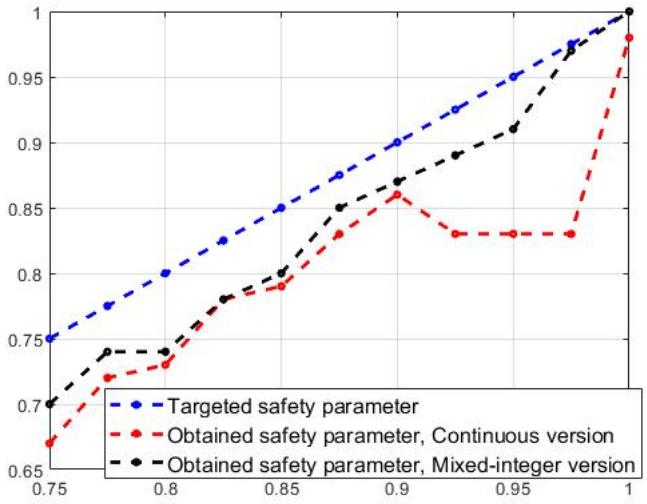


Fig. 5.1 Comparison of targeted safety parameter $1-\alpha$ with the obtained one for the first approach, Continuous (1st phase) and Mixed-integer (2nd phase) versions.

For the second use case, we obtain less promising results. For all values of the safety parameter except for $1-\alpha=0.95$ and 1, the algorithm provides a feasible solution in the 1st phase. However, in the 2nd phase (T^0 set to $\mathbf{1}$), it does not succeed in finding a feasible point to model (4.2.6). The algorithm fails to meet the targeted security level, consistently stabilizing at 0.545, which corresponds to the initial security level (without power curtailment).

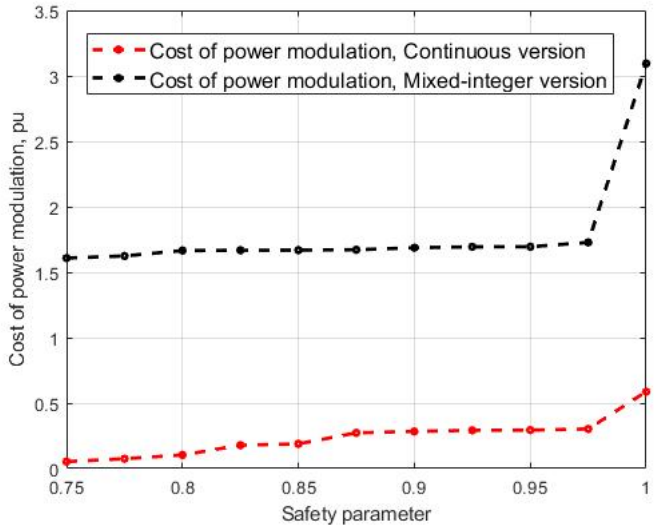


Fig. 5.2 Cost of power curtailment obtained with the first approach, Continuous (1st phase) and Mixed-integer (2nd phase) versions.

This outcome can be explained as follows. In contrast to the first use case, the power curtailment of producers is not activated at maximum in the 1st phase. This volume of power curtailment is insufficient to satisfy the required security level. However, during the 2nd phase, the algorithm does not deviate significantly from $(\bar{\mathbf{p}}, \bar{\mathbf{q}}, \bar{\rho})$ and thus fails to find another critical point capable of enabling consumer curtailment. The issue cannot be resolved by setting T^0 differently; in fact, doing so only makes it more restrictive for the algorithm. We have conducted tests with several initializations of T^0, and the resulting values remained unchanged.

The tests conducted with $N = 100$ scenarios yield similar results, that is the 2nd phase provides feasible solutions for all security levels in the first use case, and guarantees security level of only 0.52 for the second use case. The execution time is lower for the 1st phase: it varies from 412 to 1 118 seconds, with an average of 748 seconds in the first use case, and from 536 to 1 641 seconds, with an average of 1 107 seconds in the second use case. For the 2nd phase, it is comparable with the execution time of $N = 1\,000$ scenarios. At the same time, the average difference in objective function values is 4.9% for the 1st phase and 3.5% for the 2nd phase among all values of security level in the first use case. For the second use case, the 1st phase shows an average difference of 5.2% in values of the objective function.

The results described above demonstrate the limited capacity of the methodology in handling binary variables. In fact, the key success factor in the first use case is the solution in the 1st phase (continuous version) which enabled the activation of consumers' power curtailment with respect to the corresponding binary and discrete constraints. However, as it

does not hold for the 1st phase in the second use case, the results in the 2nd phase are not satisfactory. Therefore, we conclude that the described approach is not universally applicable and demonstrates good performance only for specific use cases.

Second and third approaches: integrating an additional DoC constraint. To test the second approach with soft constraint integration, we have conducted numerical experiments for three use cases, the safety parameter $1-\alpha$ equal to 0.9, and $N = 1\,000$ scenarios. First, we set the penalization term ω to 0, which corresponds to the continuous relaxation of binary variables. The algorithm identifies feasible approximate critical points for all three use cases with the corresponding execution time of 1 615, 1 416 and 1 478 seconds. In each case, the power curtailment of producers is activated to its maximum at the priority level $m = 1$. In the second and third use cases, the power curtailment of producers is uniformly activated at 12.4% from the upper bound at the remaining priority levels. The consumers' power curtailment is uniformly activated at 15.2% in the first use case, while it is activated only for 12 consumers in the second use case, and for 10 consumers in the third use case, each at a non-maximum level. We have also tried the first approach (with the integration of binary variables into the master program (5.1.2) at each iteration k of Algorithm 1) starting from the obtained solutions, but the algorithm fails to perform serious steps after 1 500 iterations.

For the second run, we set the penalization parameter ω to 100 and initialize the algorithm at the resulting point of the first run with $\omega = 0$. The idea behind this approach resembles the one of the first approach: by initializing the algorithm at a feasible solution of the continuous problem, we search for nearby binary values of variables T_t, $t \in \{0,\dots,|T|\}$. Instead of requesting a solver to provide binary values of T_t, we guide the algorithm through a penalization term.

In the three use cases, the algorithm delivers feasible solutions after 284, 1 639 and 4 916 seconds, respectively. However, not all resulting values of variables T_t, $t \in \{0,\dots,|T|\}$, are binary. In contrast to the first run with $\omega = 0$, the power curtailment of producers at priority level $m = 1$ is no longer activated to its maximum in the first use case (activated at approximately 15.8%) and second use case (activated at approximately 18.8%), and the values of corresponding variables $T_{J,1}$ are not binary. Furthermore, the number of consumers who participate in power curtailment for all three use cases, reduces up to 9, 4 and 6 consumers, respectively.

The algorithm thus decreases the resulting value of the penalization term by switching off the power curtailment of consumers. This effect is beneficial and aligns with our expectations. However, we are still seeking to satisfy the priority for producer curtailment. Setting ω to 1 000 does not help improve the situation — the algorithm fails to perform serious steps (only

null-steps up to 1 500 iterations) for the first use case, and thus the resulting point remains unchanged. This behavior may be explained by the fact that the penalization term becomes too dominant.

We attempted another strategy involving a progressive increase of the penalization parameter ω by 1%. Starting with initial values of $\omega = 10^{-3}$, 10^{-6} and 10^{-9}, we have performed simulations for the first use case with $N = 10\,000$. The latter choice was motivated by the intention to better approximate the probability constraint. The goal of this approach was to delicately guide the algorithm to preserve the binary values of variables T_t, $t \in \{0,\dots,|T|\}$, obtained after the first run (with $\omega = 0$), and gradually push other values towards binaries. However, the algorithm immediately switched to a solution close to that obtained with $\omega = 100$ (while maintaining a comparable execution time): power curtailment of producers is activated at 13.4% from the upper bound, while consumers' power curtailment is activated for 9 consumers. With a gradual increase of the penalization parameter ω, the values of power curtailment remain within 0.05%. We conclude that solutions provided by the second approach with soft constraint integration do not satisfy the priority rules.

The third approach with hard constraint integration, is much less intuitive, as the scaling parameter $\tilde{\omega}$ does not have a direct interpretation. We have conducted tests for the first use case, with $N = 1\,000$ scenarios and the parameter $\tilde{\omega}$ varying within the set $\{0.1, 1, 10\}$. After an average of 1 477 seconds, the algorithm produces unfeasible solutions. The resulting producers' power curtailment ranges between 14.6% and 16.2% from the upper bound, while the number of activated consumers is 24, 26 and 22, respectively. As in the previous case, solutions provided by the third approach with hard constraint integration do not satisfy the priority rules.

5.1.3 Discussion

In this section, we proposed three approaches for integrating binary variables into a bundle method. The first approach involves directly incorporating discrete constraints into the master programs solved at each iteration of the algorithm, and does not provide any convergence guarantee. The other two approaches add an additional DoC constraint to enforce binary outcomes for the relaxed binary variables, either as a penalization term or as a hard constraint. In this case, the convergence theory applies. However, for the hard constraint integration, convergence to a critical point of the original problem is not guaranteed by either the DoC or CwC bundle methods, due to non-satisfaction of constraint qualification conditions.

We have tested these integrations with the DoC bundle method. None of the three approaches succeeded in finding a feasible solution that fully satisfies the fairness and priority rules. For the first one, the potential issue is a lack of information about binary variables,

as there are no subgradients to indicate a direction for the bundle method. The solution remains around the initial point. For the second approach with soft constraint integration, the penalization term is not sufficient to enforce binary outcomes for the relaxed binary variables, although the algorithm tends to decrease the number of non-binary values. Indeed, as the bundle method provides only a critical point rather than a global solution, soft constraint penalization does not ensure binary satisfaction. The CwC bundle method can potentially provide a better solution with stronger criticality guarantees for this approach. However, we anticipate the same challenge. For the third approach with hard constraint integration, the algorithm finds an unfeasible critical point.

This opens a direction for future research: finding nonconvex constraints enforcing binary outcomes for the relaxed binary variables with an appropriate CwC decomposition, such that the qualification condition for the CwC bundle method is satisfied. In this case, a stronger convergence is ensured for hard constraint integration, which could lead to better numerical results.

5.2 Scenario decomposition approach

In the case of finite discrete probability distributions, one of the approaches used in the literature to deal with stochastic models is a scenario decomposition approach. For two-stage stochastic problems, it was applied in [1] for the case of binary first-stage decisions. In [17] authors consider multi-stage stochastic problems with integrality requirements. The methods rely on introducing copies of the first-stage (here-and-now) variables along with additional non-anticipativity constraints ensuring these copies take the same value. This results in an extended mathematical program with a block structure. Any block-decomposition method allows separating the model into scenario-dependent subproblems.

A column generation approach in [30] and a generalized Benders decomposition approach in [130] have been employed for chance-constrained models. Both methods rely on dualizing the condition under the probability sign and thus require the convexity, even linearity, of the underlying formulations. This requirement is omitted in [2], where the linear non-anticipativity constraints are dualized. In the absence of a linear dual for AC-OPF and to avoid relaxing the nonconvex power flow equations, we adhere to the latter methodology.

We begin in Subsection 5.2.1, by introducing the extended model, with a binary variable per scenario, indicating whether the decision $(\mathbf{p},\mathbf{q})$ belongs to the set $X(\xi)$ for a given scenario ξ. In Subsection 5.2.2, we transition to a model that maximizes the number of satisfied scenarios following the first step of the heuristic scheme given in [2][Section 6]. In Subsections 5.2.3 and 5.2.4, we evaluate different Lagrangian methods: with or without a

regularization term. In Subsection 5.2.5, we numerically test the Lagrangian methods employing different dual updates within the maximization of Lagrangian functions. Depending on the algorithm, we solve one MILP or MIQP, and a deterministic AC-OPF problem per each scenario, at each iteration.

5.2.1 Binary variable per scenario

For a scenario ξ^j, let $z^j \in \{0,1\}$ be a binary variable associated with this scenario. Consider the following set of upper and lower bounds on state variables for given $\underline{M}_\delta$, $\overline{M}_\delta$, $\underline{M}_V$, $\overline{M}_V$, $\underline{M}_{sb}$, $\overline{M}_{sb}$, $\overline{M'}_{sb}$, and $\overline{M}_{curr}$:

$$\underline{M}_\delta(1-z^j)+\underline{\delta_i} \le \delta_i^j \le \overline{\delta_i}+\overline{M}_\delta(1-z^j), \qquad i \in \mathscr{N}\setminus\{sb\} \quad (5.2.1a)$$

$$\underline{M}_V(1-z^j)+\underline{|V_i|} \le |V_i|^j \le \overline{|V_i|}+\overline{M}_V(1-z^j), \qquad i \in \mathscr{N}\setminus\{sb\} \quad (5.2.1b)$$

$$\underline{M}_{sb}(1-z^j)+p_{sb}^{\min} \le p_{sb}^j \le p_{sb}^{\max}+\overline{M}_{sb}(1-z^j), \quad (5.2.1c)$$

$$\underline{M}_{sb}(1-z^j)+q_{sb}^{\min} \le q_{sb}^j \le q_{sb}^{\max}+\overline{M}_{sb}(1-z^j), \quad (5.2.1d)$$

$$\frac{-0.48p_{sb}^{\max}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}}p_{sb}^j+\frac{0.48p_{sb}^{\max}p_{sb}^{\min}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}}-q_{sb}^j \le \overline{M'}_{sb}(1-z^j), \quad (5.2.1e)$$

$$l_{i,k}(|V_i|,|V_k|,\delta_i,\delta_k) \le (I_{i,k}^{\max})^2+\overline{M}_{curr}(1-z^j), \qquad (i,k) \in \mathscr{A} \quad (5.2.1f)$$

and the upper and lower bounds on power curtailment $(\mathbf{p},\mathbf{q})$:

$$0 \le \sum_{m=1}^{M} \boldsymbol{p}_i^m \le \bar{p}_i^{\mathscr{G}}, \qquad i \in \mathscr{G} \quad (5.2.2a)$$

$$\underline{p}_i^{\mathscr{L}} \le \boldsymbol{p}_i^{\mathscr{L}} \le 0, \qquad i \in \mathscr{L} \quad (5.2.2b)$$

$$0 \le \sum_{m=1}^{M} \boldsymbol{q}_i^m \le \bar{q}_i^{\mathscr{G}}, \qquad i \in \mathscr{G} \quad (5.2.2c)$$

$$\underline{q}_i^{\mathscr{L}} \le \boldsymbol{q}_i^{\mathscr{L}} \le 0. \qquad i \in \mathscr{L} \quad (5.2.2d)$$

Using these bounds on power curtailment and corresponding state variables, we define the following set:

$$LF(\xi^j,z^j) := \left\{ (\mathbf{p},\mathbf{q}) \;\middle|\; \begin{array}{c} \text{satisfy (5.2.2)} \\ \text{there exist } |V|,\delta,p_{sb},q_{sb} \text{ satisfying (5.2.1)}, \\ L_i^R(\mathbf{p},\delta,|V|,\xi^j)=0 \text{ for all } i \in \mathscr{N}, \\ L_i^I(\mathbf{q},\delta,|V|,\xi^j)=0 \text{ for all } i \in \mathscr{N} \end{array} \right\},$$

where (according to (4.2.3))

$$L_i^R(\mathbf{p},\delta,|V|,\xi) :=$$

$$\begin{cases} p_i^{\phi}(\xi) - \left(\sum_{m=1}^{M} \boldsymbol{p}_i^m\right) + \sum_{k\sim i} Y_{i,k}^R |V_i||V_k|\cos(\delta_i-\delta_k) + \sum_{k\sim i} Y_{i,k}^I |V_i||V_k|\sin(\delta_i-\delta_k), & \text{if } i\in\mathscr{G} \\ p_i^{\phi}(\xi) - \boldsymbol{p}_i^{\mathscr{L}} + \sum_{k\sim i} Y_{i,k}^R |V_i||V_k|\cos(\delta_i-\delta_k) + \sum_{k\sim i} Y_{i,k}^I |V_i||V_k|\sin(\delta_i-\delta_k), & \text{if } i\in\mathscr{L}, \end{cases}$$

$$L_i^I(\mathbf{q},\delta,|V|,\xi) :=$$

$$\begin{cases} q_i^{\phi}(\xi) - \left(\sum_{m=1}^{M} \boldsymbol{q}_i^m\right) + \sum_{k\sim i} Y_{i,k}^R |V_i||V_k|\sin(\delta_i-\delta_k) - \sum_{k\sim i} Y_{i,k}^I |V_i||V_k|\cos(\delta_i-\delta_k), & \text{if } i\in\mathscr{G} \\ q_i^{\phi}(\xi) - \boldsymbol{q}_i^{\mathscr{L}} + \sum_{k\sim i} Y_{i,k}^R |V_i||V_k|\sin(\delta_i-\delta_k) - \sum_{k\sim i} Y_{i,k}^I |V_i||V_k|\cos(\delta_i-\delta_k), & \text{if } i\in\mathscr{L}. \end{cases}$$

The set $LF(\xi^j,1)$ coincides with $X(\xi^j)$ given by (4.2.5).

Even though the solution set to the system of power flow equations is not a singleton, there is only one physically meaningful solution within the region where the system is operated. The assumption on solution uniqueness is common in the literature for modeling purposes (see, for instance, [149, 41, 67]). Moreover, as discussed in [41, Section 2.1], [67, Section 7.5.5], when the system is not ill-conditioned (it is obviously the case for our framework, as otherwise, stabilization mechanisms would need to be activated and included in the model), numerical methods typically provide an operational solution (physically meaningful within the region where the system is operated) to a power flow system. In what follows, we will thus focus on the operational solution. The Big-M bounds $\underline{M}_{\delta}, \overline{M}_{\delta}, \underline{M}_{V}, \overline{M}_{V}, \underline{M}_{sb}, \overline{M}_{sb}, \overline{M'}_{sb}$, and $\overline{M}_{curr}$ are sufficiently large, set in such a way that for any decision on power curtailment (within the lower and upper bounds) when $z^j = 0$, the operational solution satisfies (5.2.1). An estimation for $\overline{M}_V$ can be found in [34], which specifies technical and engineering regulations on grid parameters.

We replace the chance constraint by the condition $\frac{1}{N}\left(\sum_{j=1}^{N} z^j\right) \geq 1-\alpha$. Since $z^j = 1$ implies $(\mathbf{p},\mathbf{q}) \in X(\xi^j)$, the latter condition implies

$$\frac{1}{N}\Big(\sum_{j=1}^{N} \mathbf{1}_{(\mathbf{p},\mathbf{q})\in X(\xi^j)}\Big) \geq 1-\alpha.$$

For a finite number N of scenarios, this sum is equivalent to $\mathbb{E}[\mathbf{1}_{(\mathbf{p},\mathbf{q})\in X(\xi)}]$, which corresponds to $\mathbb{P}[(\mathbf{p},\mathbf{q}) \in X(\xi)]$. Therefore, the described substitution of the probability constraint corresponds to an inner approximation of the latter.

To simplify the notation, we denote a discrete feasible set defined by the constraints (4.2.6c) - (4.2.6n) as $X_\rho \times \bar{B}$. The problem (4.2.6) transforms into

$$\min_{\mathbf{p},\mathbf{q},T,\rho,|V|,\delta,p_{sb},q_{sb},z} \quad f(\mathbf{p}) \tag{5.2.3a}$$

$$\text{s.t. } (\mathbf{p},\mathbf{q},T,\rho) \in X_\rho \times \bar{B}, \tag{5.2.3b}$$

$$(\mathbf{p},\mathbf{q}) \in LF(\xi^j, z^j), \qquad j = 1,\dots,N \tag{5.2.3c}$$

$$\frac{1}{N}\Big(\sum_{j=1}^{N} z^j\Big) \geq 1-\alpha. \tag{5.2.3d}$$

5.2.2 Maximizing probability

The resulting problem (5.2.3) is a nonconvex MINLP (mixed-integer nonlinear program) due to the binary variables in (5.2.3b) and (5.2.3d), as well as the nonconvex constraints (5.2.3c) representing an AC-OPF problem for each scenario. Our goal is to separate these issues by employing a scenario decomposition approach and to handle each component individually. We start by reformulating this problem as a probability maximization problem and show their link using a bisection procedure.

For each scenario ξ^j, we duplicate variables $(\mathbf{p},\mathbf{q})$ in (5.2.3c) denoting them by $(\tilde{\mathbf{p}}^j,\tilde{\mathbf{q}}^j)$. In what follows, we again denote $x = (\mathbf{p},\mathbf{q})$, and $\tilde{x}^j = (\tilde{\mathbf{p}}^j,\tilde{\mathbf{q}}^j)$. Additional linear non-anticipativity constraints $A(x-\tilde{x}^j) = 0$ for $j = 1,\dots,N$ ensure that

$$\sum_{m=1}^{M} \boldsymbol{p}_i^m = \sum_{m=1}^{M} \tilde{\mathbf{p}}_i^{m,j}, \qquad \text{if } i \in \mathscr{G},$$

$$\mathbf{p}_i^{\mathscr{L}} = \tilde{\mathbf{p}}_i^{\mathscr{L},j}, \qquad \text{if } i \in \mathscr{L}.$$

Finally, for a given value $y \geq 0$, we add a budget constraint $f(x) \leq y$. The problem (5.2.3) becomes

$$\min_{x,\tilde{x},\rho,T,|V|,\delta,p_{sb},q_{sb},z} \quad \sum_{j=1}^{N}(1-z^j) \tag{5.2.4a}$$

$$\text{s.t. } (x,\rho,T) \in X_\rho \times \bar{B}, \tag{5.2.4b}$$

$$\tilde{x}^j \in LF(\xi^j, z^j), \qquad j = 1,\dots,N \tag{5.2.4c}$$

$$A(x-\tilde{x}^j) = 0, \qquad j = 1,\dots,N \tag{5.2.4d}$$

$$f(x) \leq y. \tag{5.2.4e}$$

Let $\bar{v} = (x, \tilde{x}, \rho, T, |V|, \delta, p_{sb}, q_{sb}, z)$ be a global solution of problem (5.2.4) for a given y. If the corresponding value of the objective function is lower than αN, then the constraint (5.2.3d) is satisfied at $\bar{v}$. In this case, the given value of y is an upper bound for the optimal objective value of problem (5.2.3). Conversely, if the objective function value exceeds αN, then the constraint (5.2.3d) is not satisfied at $\bar{v}$. Since $\bar{v}$ is a global minimum for (5.2.4), this constraint is not satisfied at any other point. Therefore, the given y is a lower bound on the optimal objective value of problem (5.2.3).

If $\bar{v}$ is a feasible point of problem (5.2.4) and the corresponding objective function value is lower than αN, the observation about the value of y being an upper bound remains valid. However, in the case of a lower bound, the assumption that $\bar{v}$ is a global solution is crucial. Solving problem (5.2.4) up to global optimality is challenging, as it still involves binary variables along with AC-OPF constraints for all scenarios.

Assume that there exists a relaxation, which is easier to solve. In this case, the observation for lower bound remains valid: if $\bar{v} = (x, \tilde{x}, \rho, T, |V|, \delta, p_{sb}, q_{sb}, z)$ is a global solution of a relaxed problem and the corresponding value of $\sum_{j=1}^{N}(1 - z^j)$ is greater than αN, then the given value of y is a lower bound on the optimal objective value of problem (5.2.3). This does not hold for an upper bound: if the value of $\sum_{j=1}^{N}(1 - z^j)$ is lower than αN in the relaxed problem, it may still exceed αN in problem (5.2.4).

To find an upper bound on the optimal objective value of problem (5.2.3), it is sufficient to obtain a feasible point for either (5.2.4) or (5.2.3). One approach is to fix the binary variable values (for example, using the values from the solution obtained for the lower bound). Then, apply the DoC bundle method to solve the chance-constrained problem (4.2.6) with fixed binary variables. As we observed in numerical experiments in Section 2.4, the obtained security level, which is exactly $\frac{1}{N}\left(\sum_{j=1}^{N} z^j\right)$, is lower than targeted level. Consequently, we need to set a higher value for the safety parameter (which can be approximately estimated based on numerical results from the same section). Since any feasible point provides an upper bound, we can terminate the DoC bundle method earlier.

The discussion above can be summarized in the following scheme of bisection procedure for determining values of y:

1. **Initialization.** Let $\hat{y} = y_0$. (By default, set an upper bound on maximal cost of power curtailment)

2. **Solve a relaxation.** For the given $\hat{y}$, solve a relaxation of problem (5.2.4), and denote its solution by $\bar{v}$.

3. **Lower bound.** If $\sum_{j=1}^{N}(1 - z^j) \geq \alpha N$, then set LB $\leftarrow \hat{y}$.

4. **Upper bound.** Fix values of binary variables as in $\bar{v}$. Find a feasible point of the resulting continuous problem and denote the corresponding cost value by $\hat{f}$. Set $\text{UB} \leftarrow \hat{f}$.

5. **Update $\hat{y}$.** (Default update: $\hat{y} \leftarrow \frac{1}{2}\big(\text{LB} + \text{UB}\big)$.) Return to Step 2.

In the next sections, we will closely consider Step 2, which consists of solving a relaxation of problem (5.2.4). Specifically, we explore the dualization of non-anticipativity constraints in (5.2.4) leading to a scenario decomposition. Remaining consistent with the logic of keeping the AC-OPF formulation, we do not use any relaxation of the latter. However, the method can be easily adapted to subsequently incorporate convex relaxations of AC-OPFs, such as Quadratic Convex relaxation [21], Second-Order Cone relaxation [64], and Semidefinite Programming relaxation [71]. Conversely, directly relaxing AC-OPFs without dualizing the non-anticipativity constraints does not allow the decomposition per scenario and thus remains an intractable problem in the case of a large number of scenarios.

5.2.3 Scenario decomposition with the Lagrangian relaxation

By weak duality, for any values of λ_j, $j = 1,\dots,N$, the solution of the Lagrangian relaxation

$$\begin{aligned} \min_{x,\tilde{x},\rho,T,|V|,\delta,p_{sb},q_{sb},z} \quad & \sum_{j=1}^{N} (1-z^j) + \lambda_j^T A(x-\tilde{x}^j) \\ \text{s.t. } & \text{(5.2.4b)}, \\ & \text{(5.2.4c)}, \\ & \text{(5.2.4e)}, \end{aligned} \tag{5.2.5}$$

provides a lower bound on (5.2.4). Moreover, it is decomposable into a MILP problem

$$\begin{aligned} \min_{x,\rho,T} \quad & (\sum_{j=1}^{N} \lambda_j)^T Ax \\ \text{s.t. } & \text{(5.2.4b)}, \\ & \text{(5.2.4e)}, \end{aligned} \tag{5.2.6}$$

and $j = 1,\dots,N$ problems

$$\begin{aligned} \min_{\tilde{x}^j,|V|^j,\delta^j,p^j_{sb},q^j_{sb},z^j} \quad & (1-z^j) - \lambda_j^T (A\tilde{x}^j) \\ \text{s.t. } & \tilde{x}^j \in LF(\xi^j, z^j). \end{aligned} \tag{5.2.7}$$

For each scenario ξ^j, the problem (5.2.7) splits into two cases: one with $z^j = 1$ and another with $z^j = 0$. The resulting optimization problems are deterministic AC-OPFs, which differ only in the bounds specified in (5.2.1) (the latter are tight for $z^j = 1$ and defined by Big-M method for $z^j = 0$):

$$\begin{array}{ll} \text{Case } z^j = 1: & \text{Case } z^j = 0: \\ \min\limits_{\tilde{x}^j, |V|^j, \delta^j, p^j_{sb}, q^j_{sb}} -\lambda_j^T (A\tilde{x}^j) & \min\limits_{\tilde{x}^j, |V|^j, \delta^j, p^j_{sb}, q^j_{sb}} 1 - \lambda_j^T (A\tilde{x}^j) \\ \text{s.t.} \quad \tilde{x}^j \in LF(\xi^j, 1). & \text{s.t.} \quad \tilde{x}^j \in LF(\xi^j, 0). \end{array}$$

We will show that it is sufficient to solve only one deterministic AC-OPF per scenario and the sign of λ_j is known. Assume that there exists $\tilde{i} \in \mathscr{G}$ such that component of λ_j corresponding to the total active power curtailment of user $\tilde{i}$ is strictly positive, i.e. $\lambda_j^{\tilde{i}} > 0$ (or, if $\tilde{i} \in \mathscr{L}$, then $\lambda_j^{\tilde{i}} < 0$). For this component, $-\lambda_j^{\tilde{i}} \left(\sum_{m=1}^M \mathbf{p}_{\tilde{i}}^{m,j}\right) \leq 0$ (or, $-\lambda_j^{\tilde{i}} \mathbf{p}_{\tilde{i}}^{\mathscr{L},j} \leq 0$).

Recall that the condition $\tilde{x}^j \in LF(\xi^j, 0)$ is always satisfied for $z^j = 0$ (for $\tilde{x}^j$ within the bounds (5.2.2)). The minimum of problem (5.2.7) with $z^j = 0$ is thus attained at the upper bound specified by (5.2.2): $\sum_{m=1}^M \mathbf{p}_{\tilde{i}}^{m,j} = \bar{p}_{\tilde{i}}^{\mathscr{G}}$ (or, lower bound if $\tilde{i} \in \mathscr{L}$: $\mathbf{p}_{\tilde{i}}^{\mathscr{L},j} = \underline{p}_{\tilde{i}}^{\mathscr{L}}$).

Let x be a feasible point of problem (5.2.6). We will reduce the modulus of its $\tilde{i}$-th component while preserving feasibility. Consider a point $\hat{x}$ with components $\hat{\mathbf{p}}_i = \mathbf{p}_i$, $i \neq \tilde{i}$ and

$$\hat{\mathbf{p}}_{\tilde{i}} = \begin{cases} 0, & \text{if } \tilde{i} \in \mathscr{L} \\ (\mathbf{p}_{\tilde{i}}^1, \ldots, \mathbf{p}_{\tilde{i}}^{m-1}, 0), & \text{if } \tilde{i} \in \mathscr{G} \text{ and} \\ & \text{maximal activated curtailment level at } x_{\tilde{i}} \text{ is } m. \end{cases}$$

The point $\hat{x}$ is indeed feasible, as decreasing power curtailment at the highest activated level (including consumers) respects deterministic constraints. Meanwhile, the value of the objective function is lower at $\hat{x}$ compared to x. Thus, any solution $\hat{x}$ of problem (5.2.6) satisfies $\sum_{m=1}^M \hat{\mathbf{p}}_{\tilde{i}}^m < \bar{p}_{\tilde{i}}^{\mathscr{G}}$ if $\tilde{i} \in \mathscr{G}$. Moreover, there exists a positive constant $C > 0$ such that $\bar{p}_{\tilde{i}}^{\mathscr{G}} - \left(\sum_{m=1}^M \hat{\mathbf{p}}_{\tilde{i}}^m\right) \geq C$ by construction. An analogous bound holds for consumers.

The update rule for λ_j^{k+1} (dual ascent method) is $\lambda_j^{k+1} = \lambda_j^k + \beta_k A(x^{k+1} - \tilde{x}^{j,k+1})$. The coefficients $\beta_k > 0$ should not be too small, ensuring that $\sum_k \beta_k = +\infty$. Consequently, $|\lambda_j^{\tilde{i},k+1} - \lambda_j^{\tilde{i},k}| \geq \tilde{C}\beta_k$, and after finite number of steps $\lambda_j^{\tilde{i}}$ will change its sign. Therefore, we set initially $\lambda_j^{\tilde{i}} \leq 0$ for $\tilde{i} \in \mathscr{G}$, and $\lambda_j^{\tilde{i}} \geq 0$ for $\tilde{i} \in \mathscr{L}$.

Consequently, the solution to problem (5.2.7) with $z^j = 0$ is zero. This aligns with the intuition that power curtailment is unnecessary in unfavorable scenarios, as it increases power curtailment cost if activated. Therefore, $z^j = 1$ if $-\lambda_j^T (A\tilde{x}^j) \leq 1$. More precisely, it is

sufficient to solve (5.2.7) with $z^j = 1$ and verify whether the objective value is lower than 1. If so, we conclude that solution of (5.2.7) is $z^j = 1$. Otherwise, the solution is $z^j = 0$ and $\tilde{x}^j = 0$. Algorithm 4 summarizes the described scenario decomposition approach. It can be accelerated by terminating the resolution of problem (5.2.8) as soon as $-\lambda_j^T(A\tilde{x}^j) \leq 1$.

5.2.4 Defining the here-and-now decision

Augmented Lagrangian. The main challenge for us in scenario decomposition approach is gluing together $\tilde{x}^j$ for $j = 1,\dots,N$, and x, as there is no theoretical guarantee that the Stopping Criterion $\|A(x^{k+1} - \tilde{x}^{j,k+1})\| \leq \varepsilon$ will be met. To enforce the satisfaction of the non-anticipativity constraints, we strengthen the penalization of their violation with quadratic penalization terms. This results in considering an augmented Lagrangian function:

$$\min_{x,\tilde{x},\rho,T,|V|,\delta,p_{sb},q_{sb},z} \sum_{j=1}^{N} (1 - z^j) + \lambda_j^T A(x - \tilde{x}^j) + r_j \|A(x - \tilde{x}^j)\|^2 \tag{5.2.10}$$

$$\text{s.t. (5.2.4b)},$$
$$\text{(5.2.4c)},$$
$$\text{(5.2.4e)},$$

where $r_j > 0$ for $j = 1,\dots,N$. Due to the quadratic term, which is equal to

$$\sum_{j=1}^{N} r_j \|A(x - \tilde{x}^j)\|^2 = \sum_{j=1}^{N} r_j \big(\|Ax\|^2 - 2\langle Ax, A\tilde{x}^j\rangle + \|A\tilde{x}^j\|^2\big),$$

problem (5.2.10) is no longer separable.

One of the approaches widely used when individual subproblems are simpler to solve than the original problem is the block coordinate method. It relies on optimizing one block of variables while keeping other blocks fixed, and iterating between blocks. For the biconvex framework, convergence of this method is shown in [51]. Another approach, the alternating direction method of multipliers (ADMM) was proposed in [42] and further studied in detail in [16]. This method combines the ideas from dual decomposition and augmented Lagrangian methods and relies on updating primal and dual variables alternately, promoting both decomposition and coordination. While there are no convergence guarantees due to the presence of binary variables and nonconvexity in applying these methods for our case, we adopt the scheme of the latter algorithm to address problem (5.2.10).

At iteration k, we obtain the following problems (following block coordinate scheme):

$$\min_{\tilde{x}^j, |V|^j, \delta^j, p^j_{sb}, q^j_{sb}, z^j} r_j \|A\tilde{x}^j\|^2 - \langle 2r_j A x^k + \lambda_j, A\tilde{x}^j \rangle + (1 - z^j) \tag{5.2.11}$$

$$\text{s.t. } \tilde{x}^j \in LF(\xi^j, z^j)$$

for $j = 1, \ldots, N$, and

$$\min_{x, \rho, T} \Big(\sum_{j=1}^{N} \lambda_j\Big)^T Ax + \Big(\sum_{j=1}^{N} r_j\Big) \|Ax\|^2 - \Big\langle Ax, \sum_{j=1}^{N} 2r_j A\tilde{x}^{j,k+1} \Big\rangle \tag{5.2.12}$$

$$\text{s.t. (5.2.4b),}$$

$$\text{(5.2.4e).}$$

For the case $z^j = 0$, problem (5.2.11) becomes:

$$\min_{\tilde{x}^j, |V|^j, \delta^j, p^j_{sb}, q^j_{sb}, z^j} r_j \|A\tilde{x}^j\|^2 - \langle 2r_j A x^k + \lambda_j, A\tilde{x}^j \rangle + 1 \tag{5.2.13a}$$

$$\text{s.t. } \tilde{x}^j \in LF(\xi^j, 0). \tag{5.2.13b}$$

Since the Big-M bounds are set so that the constraint $\tilde{x}^j \in LF(\xi^j, 0)$ is always satisfied (for $\tilde{x}^j$ within the bounds (5.2.2)), the latter problem can be solved independently for each component i of $(A\tilde{x}^j)_i$. The minimum for component can be reached either at bounds (5.2.2), which are defined specifically for $A\tilde{x}^j$, or at $(A\tilde{x}^j)_i = (Ax^{k-1} + \frac{\lambda_j}{2r_j})_i$. Therefore, the values of $A\tilde{x}^j$ are determined component-wise based on the comparison of the corresponding objective values at identified points. As the constraints defining priority rules are not present in (5.2.13), any allocation of the curtailed power among priority levels m for producers is feasible. Consequently, only the AC-OPF with tight bounds

$$\min_{\tilde{x}^j, |V|^j, \delta^j, p^j_{sb}, q^j_{sb}, z^j} r_j \|A\tilde{x}^j\|^2 - \langle 2r_j A x^k + \lambda_j, A\tilde{x}^j \rangle \tag{5.2.14a}$$

$$\text{s.t. } \tilde{x}^j \in LF(\xi^j, 1) \tag{5.2.14b}$$

needs to be solved at each iteration and per each scenario. The value of $z^{j,k+1}$ is identified based on the comparison of the objective values for AC-OPF (5.2.14) and (5.2.13).

The scheme of Algorithm 4 remains almost unchanged: problem (5.2.8) is replaced by (5.2.14), and problem (5.2.9) is replaced by (5.2.12). The update for λ_j at Step 16 is given

by

$$\lambda_j^{k+1} = \lambda_j^k + 2r_j A(x^{k+1} - \tilde{x}^{j,k+1}).$$

The complete procedure is summarized in Algorithm 5.

Proximal Lagrangian. For problem (5.2.4), the Proximal Lagrangian is defined as

$$\mathscr{L}_{\text{prox}}(x, \tilde{x}, z, \lambda, r) = \sum_{j=1}^{N} (1 - z^j) + \lambda_j^T A(x - \tilde{x}^j) + r\|A(x - \tilde{x}^j)\|^2,$$

where $r \geq 0$ also becomes a dual variable. A notable property of the Proximal Lagrangian is closing the duality gap, even for nonconvex mixed-integer problems under mild assumptions ([23, Corollary 1]). Similar to the problem (5.2.10), the presence of the quadratic term means that the Lagrangian subproblems are no longer separable. The primal-dual bundle method (Algorithm 2 in [23]) allows the latter to be solved inexactly, with known uniformly bounded errors. This method yields an approximate solution, with an error estimate based on these bounds.

In our case, we apply the block coordinate scheme to find approximate solutions to the Lagrangian subproblems. With the dual variables λ and r fixed, the problems at iteration k correspond exactly to (5.2.11) and (5.2.12) (r_j are the same for all scenarios $j = 1, \ldots, N$). However, we are unable to estimate the errors for the obtained solutions. Therefore, we cannot expect any guarantees on the quality of the resulting solution provided by the primal-dual bundle method.

5.2.5 Numerical results for scenario decomposition methods

We have conducted numerical experiments with the scenario decomposition approach for three Lagrangian methods: Lagrangian relaxation, augmented Lagrangian, and Proximal Lagrangian. So far, we have focused on the first use case from Subsection 5.1.2, where only voltage grid constraints are present, and all three producers connected at buses 12, 29, and 32, have a single priority level ($M = 1$).

In this case, the priority and fairness rules reduce to two requirements: first, a simultaneous and proportional curtailment of producers due to the constraint (4.2.6n); second, consumer curtailment begins only after all producers are curtailed to their maximum due to the constraints (4.2.6j) and (4.2.6k).

The tests have been performed for $N = 1\,000$ scenarios. The Stopping Criterion, which measures the distance between the solutions of MILP or MIQP and the stochastic part of

the model at iteration k, denoted by $\|A(x^{k+1} - \tilde{x}^{j,k+1})\| \leq \varepsilon$ in Algorithm 4, is calculated as follows:

$$\max_{j=1,\ldots,N} \|A(x^{k+1} - \tilde{x}^{j,k+1})\|_1.$$

We will refer to this value as the solution distance at iteration k.

The ratio of satisfied scenarios at iteration k can be calculated in two different ways: based on the solution x^{k+1} of MILP or MIQP by counting the number of satisfied scenarios with power flow equations, or directly from the values of binary variables $z^{j,k+1}$ related to the solutions of deterministic AC-OPFs. If the solution distance is large, these two ratios are different, and we therefore distinguish between them. We define the ratio of satisfied scenarios for the solution x^{k+1} as the obtained safety level and refer to $\sum_{j=1}^{N}(1 - z^{j,k+1})$ as the number of unsatisfied scenarios.

Based on the results provided by the first approach with the integration of binary variables into the master problem of the DoC bundle method, we estimate the power curtailment cost $\hat{y}$ in the budget constraint $f(x) \leq \hat{y}$. More precisely, we refer to the results of the 2nd phase and $N = 1\,000$ scenarios analyzed in Subsection 5.1.2 for three values of the targeted safety parameter $1 - \alpha = 0.75$, 0.975, and 1. The corresponding ratios of scenarios satisfying the power flow equations (the obtained safety parameter) are 0.709, 0.97, and 0.997 with the corresponding power curtailment costs (value of the objective function) 1.6079 pu, 1.7076 pu, and 3.0942 pu. We thus consider three use cases with the latter values of $\hat{y}$ rounded.

Lagrangian relaxation. We apply the scenario decomposition approach with the Lagrangian relaxation, as described in Subsection 5.2.3, with $\hat{y} = 1.71$ pu and the dual ascent subgradient method for updating λ given by Step 16 of Algorithm 4. Several initial values of λ^0 are tested, setting them equal for all scenario $j = 1,\ldots,N$, and for all producers and consumers: $\lambda_j^{i,0} \in \{-1,-10,-100\}$ for $i \in \mathscr{G}$ and $\lambda_j^{i,0} \in \{1,10,100\}$ for $i \in \mathscr{L}$. The primal starting points are not explicitly involved in Algorithm 4. However, as deterministic AC-OPFs (5.2.8) are addressed in practice with a local solver (as in Chapter 2), the choice of starting points at each iteration k influences the provided solutions $\tilde{x}^{j,k+1}$. We will discuss this limitation later in the section. For all initial dual values, the solution distance stabilizes after first iterations and the ratio of satisfied scenarios (safety level) remains zero after 200 iterations.

Since the Lagrangian relaxation (5.2.5) is convex with respect to the dual variable, it can be addressed with a convex bundle method. The oracles, which rely on the resolution of MILP (5.2.9) and AC-OPFs (5.2.8), are inexact due to the local solutions obtained for the

latter problems. After 62 iterations for $\hat{y} = 1.61$ pu, 33 iterations for $\hat{y} = 1.71$ pu, and 48 iterations for $\hat{y} = 3.095$ pu, the bundle method provides a dual value for (5.2.5), terminating based on the stopping criterion with a tolerance of 10^{-6}. However, the corresponding (primal) solution distances remain significant for all three use cases, 6.621, 11.191, and 1.109, respectively, and the safety levels are zero. At the same time, the number of unsatisfied scenarios is zero for $\hat{y} = 1.71$ pu and $\hat{y} = 3.095$ pu, and 152 scenarios for $\hat{y} = 1.61$ pu.

Significant solution distances observed for these two approaches can be explained by the lack of interdependence between MILP (5.2.9) and AC-OPF problems (5.2.8) (decomposition is too strong): they are linked only by the dual variables, but it turns out to be insufficient to enforce convergence between x and $\tilde{x}^j$. For the first approach with the subgradient update, we tried to increase the communication between these problems by setting the solution of MILP at iteration k as a starting point for the local solver (with an interior-point method) addressing AC-OPFs at iteration $k+1$. In this case, we observe the dual variable iterating between two values, which implies the same behavior for solutions of MILP and AC-OPFs, but does not ensure convergence in the solution distances.

Augmented Lagrangian. For the augmented Lagrangian dual method described in Subsection 5.2.4, we test its capability to recover a point satisfying the priority and fairness rules. The algorithm fails to improve the safety level starting from zero; we also initialize it at $x^0 = (\bar{\mathbf{p}}, \bar{\mathbf{q}})$ corresponding to the solution provided by the 1st phase (continuous version) of the first approach integrating binary variables into the DoC bundle method for a corresponding safety level. In the latter case, the algorithm serves as an alternative to the 2nd phase of the first approach.

The method is sensitive to the choice of the penalization parameter. For $r_j = 1$ and $r_j = 10$, $j = 1,\dots,N$, the algorithm behaves similarly to the Lagrangian relaxation. We observe that the solutions of MIQP and AC-OPFs iterate between several values, and the solution distance does not decrease. Setting a higher value of the penalization parameter $r_j = 10\,000$ for $j = 1,\dots,N$ enforces a decrease in the solution distance.

However, another challenge occurs: after several iterations (depending on the use case and a starting point for the local solver addressing AC-OPFs), the number of unsatisfied scenarios reaches N, and the solution distance stops decreasing. It means that solutions provided by the local solver for AC-OPFs with tight bounds (5.2.14) are more expensive in terms of the objective function compared to the solutions of AC-OPFs with Big-M bounds (5.2.13) for all scenarios. Recall that the initial safety level (corresponding to zero power curtailment) is 0.545. This implies that for at least 545 scenarios the objective value of the AC-OPF with tight bounds (5.2.14a) should be lower compared to the AC-OPF with

Big-M bounds (5.2.13a). Consequently, this issue is related to the limitations of the local solver.

The issue persists when other algorithms for the local solver, such as Sequential Quadratic Programming (SQP), are used. Running the solver from multiple initial points (the tests have been performed with 100 random initial points) helps to improve the convergence of the solution distance and provides a higher safety level than the initial one (0.545). However, the latter option is highly time-consuming. Therefore, we apply an intermediate time-efficient strategy: when solving an AC-OPF (5.2.14), we first check additional points corresponding to zero power curtailment (as in Step 4 in Algorithm 4) and the extreme point of the parabola (5.2.13a). If one of the corresponding grid states satisfies the tight bounds (i.e. one of these points belongs to $LF(\xi^j, 1)$) and the corresponding value of the objective function (5.2.14a) is lower compared to the objective value for the AC-OPF with Big-M bounds (5.2.13a), then this scenario is satisfied ($z^{j,k+1} = 1$, without solving (5.2.14)).

This brute-force approach significantly reduces the number of unsatisfied scenarios, at least for a sufficient number of iterations: the Stopping Criterion in terms of the solution distances is not satisfied (except for $\hat{y} = 3.095$ pu), but we terminate when the safety level is sufficiently high. This termination is based on the practical observation that the highest safety level is typically achieved at the iteration just before the number of unsatisfied scenarios reaches N. Once this happens, the algorithm is unable to return to significantly lower numbers of unsatisfied scenarios. Informally speaking, it loses the information that it could have obtained from AC-OPFs with tight bounds, and thus cannot transfer it to the solutions of MIQP.

The corresponding solution distances at the final iteration are as follows: 0.152 for $\hat{y} = 1.61$ pu, 1.2134 for $\hat{y} = 1.71$ pu, and 10^{-14} for $\hat{y} = 3.095$ pu. The obtained safety levels are relatively high: 0.662, 0.911, and 0.998, respectively. In the latter case, the safety level is higher than the ratio of scenarios satisfying the power flow equations (the obtained safety parameter) provided with the 2nd run of the first approach (integrating binary variables into the DoC bundle method). Meanwhile, the execution time is lower for all three use cases: 172 against 664 seconds ($\hat{y} = 1.61$ pu), 630 against 1 396 seconds ($\hat{y} = 1.71$ pu), and 348 against 977 seconds ($\hat{y} = 3.095$ pu).

Proximal Lagrangian. We apply the Proximal Lagrangian method discussed in Section 5.2.4 to the same use cases as the augmented Lagrangian. We also set $x^0 = (\bar{\mathbf{p}}, \bar{\mathbf{q}})$ corresponding to the solution provided by the 1st phase of the first approach integrating binary variables into the DoC bundle method. The dual variables are initialized at $(\lambda^0, r^0) = (-1, 10000)$.

Table 5.2 Numerical results for three considered scenario decomposition methods.

Method	Budget constraint, pu	Average time per iteration, AC-OPFs, s	Average time per iteration, MILP/MIQP, s	Safety level	Final solution distance
LagRel	1.61	0.48	0.25	0	6.621
	1.71	0.52	0.24	0	11.19
	3.1	0.51	0.31	0	1.109
AugLag	1.61	61.2	0.47	0.662	0.152
	1.71	81.8	0.66	0.911	1.213
	3.1	56.57	0.51	0.998	10^{-14}
ProxLag	1.61	25.75	0.61	0.654	0.704
	1.71	3.13	0.55	0.875	0.007
	3.1	24.48	0.57	0.982	0.302

Recall that solving MIQP (5.2.12) and AC-OPFs (5.2.14) corresponds now to the oracle phase. We apply the same modification to address AC-OPFs (solving power flow equations for additional points) with the local solver. Moreover, we implement the termination condition based on the practical observation (i.e. terminating just before the number of unsatisfied scenarios reaches N). These modifications enable significantly improve the numerical results: without them, the algorithm fails to provide a non-zero safety level.

The primal-dual bundle method terminates based on the stopping criterion with a tolerance of 10^{-6}. The safety levels corresponding to the primal solutions are 0.654 for $\hat{y} = 1.61$ pu, 0.875 for $\hat{y} = 1.71$ pu, and 0.982 for $\hat{y} = 3.095$ pu. Compared to the augmented Lagrangian method, these values are lower, as well as the average time per an oracle call compared to the average time per iteration in the previous method. However, the primal-dual algorithm requires a higher number of iterations.

The numerical results for the discussed methods are summarized in Table 5.2.

5.2.6 Discussion

In this section, we considered three methods for scenario decomposition relying on the Lagrangian relaxation, and an augmented and proximal Lagrangian functions. Directly decomposing the problem in the first case, and applying a block coordination in the latter

two cases, we obtain one MILP or MIQP, and one AC-OPF per scenario, at each iteration. Assuming that each subproblem can be effectively handled by known methods, the main challenge lies in gluing together the solutions of the decomposed components.

Numerical experiments demonstrate that the methods based on augmented and proximal Lagrangian functions provide relatively good lower-bound estimations for the considered use cases, with moderate execution times. However, these methods lack convergence guarantees, and are sensitive to parameter choices and initializations. Therefore, they can be considered as heuristics, and are efficient when combined with a bundle method (continuous version) providing a convenient initial point (warm start). Future testing should focus on applying these methods to other use cases with different priority levels and additional congestion grid constraints.

Furthermore, the quality of local solutions to deterministic AC-OPFs significantly influences performance of the method. The next step in this direction is to employ convex approximations of AC-OPFs (possibly, with recovery procedure) and testing global solvers to address original AC-OPFs with higher computational resources.

5.3 Conclusion

In this chapter, we considered several approaches to address the chance-constrained AC-OPF model with discrete constraints representing the priority and fairness rules. The first group of approaches relies on incorporating binary variables into the bundle methods, directly into the master program, or using additional DoC constraints with simultaneous relaxation of binary variables. While the convergence results are applicable in the latter case, the binary outcomes are not provided in practice by the DoC bundle method due to the infeasibility of the obtained solutions. Finding another nonconvex constraint compatible with the qualification conditions of the CwC bundle methods and enforcing binary outcomes for binary variables could improve the numerical performance.

An alternative approach addresses the model maximizing the number of satisfied scenarios under a budget constraint, representing a discrete counterpart of the probability maximization model. Applying different Lagrangian methods, with or without block coordination, it can be decomposed into subproblems that are solved independently. Numerical results validate numerical performance of the method on three use cases provided warm start initialization. Further numerical experiments should address broader range of use cases with integration of more accurate AC-OPF solvers.

Algorithm 4 Scenario decomposition (Lower bound loop): Lagrangian relaxation

Set $\hat{y} \geq 0$, and λ_j^0 for $j = 1, \ldots, N$. Choose $\varepsilon > 0$, $K_{\max} > 0$ and a sequence $(\beta_k)_k$.
for $k = 0, 1, 2, \ldots$ **do**
  for $j = 1, \ldots, N$ **do**
    if $0 \in LF(\xi^j, 1)$ **then**
      $\tilde{x}^{j,k+1} \leftarrow 0$ and $z^{j,k+1} \leftarrow 1$
    else
      Solve

$$\min_{\tilde{x}^j, |V|^j, \delta^j, p_{sb}^j, q_{sb}^j} \; -(\lambda_j^k)^T (A\tilde{x}^j) \tag{5.2.8a}$$

$$\text{s.t.} \quad \tilde{x}^j \in LF(\xi^j, 1), \tag{5.2.8b}$$

      if $-(\lambda_j^k)^T (A\tilde{x}^j) \leq 1$ **then**
        $\tilde{x}^{j,k+1} \leftarrow \tilde{x}^j$ and $z^{j,k+1} \leftarrow 1$
      else
        $\tilde{x}^{j,k+1} \leftarrow 0$ and $z^{j,k+1} \leftarrow 0$
      end if
    end if
  end for
  Solve

$$\min_{x, \rho, T} \; (\sum_{j=1}^{N} \lambda_j^k)^T Ax \tag{5.2.9a}$$

$$\text{s.t. } (x, \rho, T) \in X_\rho \times \bar{B}, \tag{5.2.9b}$$

$$f(x) \leq \hat{y}, \tag{5.2.9c}$$

  Set $x^{k+1} \leftarrow x$ and $\lambda_j^{k+1} \leftarrow \lambda_j^k + \beta_k A(x^{k+1} - \tilde{x}^{j,k+1})$
  if one of the Stopping Criteria holds **then**
    return $(x^{k+1}, \tilde{x}^{j,k+1}, z^{j,k+1})$
  end if
end for

▷ Stopping Criteria

*Stopping Criteria: $\|A(x^{k+1} - \tilde{x}^{j,k+1})\| \leq \varepsilon$ for $j = 1, \ldots, N$, or $k = K_{\max}$.

Algorithm 5 Scenario decomposition (Lower bound loop): Augmented Lagrangian

Set $\hat{y} \geq 0$, $r_j > 0$ and λ_j^0 for $j = 1, \ldots, N$. Choose $\varepsilon > 0$, $K_{\max} > 0$ and x^0.
for $k = 0, 1, 2, \ldots$ **do**
  for $j = 1, \ldots, N$ **do**
    Solve

$$\min_{\tilde{x}_M^j, |V|^j, \delta^j, p_{sb}^j, q_{sb}^j, z^j} r_j \|A\tilde{x}_M^j\|^2 - \langle 2r_j A x^k + \lambda_j^k, A\tilde{x}_M^j \rangle + 1$$
$$\text{s.t. } \tilde{x}_M^j \in LF(\xi^j, 0),$$

▷ Check the extreme points

    Solve

$$\min_{\tilde{x}^j, |V|^j, \delta^j, p_{sb}^j, q_{sb}^j, z^j} r_j \|A\tilde{x}^j\|^2 - \langle 2r_j A x^k + \lambda_j^k, A\tilde{x}^j \rangle$$
$$\text{s.t. } \tilde{x}^j \in LF(\xi^j, 1),$$

    if $r_j \|A\tilde{x}^j\|^2 - \langle 2r_j A x^k + \lambda_j^k, A\tilde{x}^j \rangle \leq r_j \|A\tilde{x}_M^j\|^2 - \langle 2r_j A x^k + \lambda_j^k, A\tilde{x}_M^j \rangle + 1$ **then**
      $\tilde{x}^{j,k+1} \leftarrow \tilde{x}^j$ and $z^{j,k+1} \leftarrow 1$
    else
      $\tilde{x}^{j,k+1} \leftarrow \tilde{x}_M^j$ and $z^{j,k+1} \leftarrow 0$
    end if
  end for
  Solve

$$\min_{x, \rho, T} \Big(\sum_{j=1}^{N} \lambda_j^k\Big)^T Ax + \Big(\sum_{j=1}^{N} r_j\Big) \|Ax\|^2 - \Big\langle Ax, \sum_{j=1}^{N} 2r_j A\tilde{x}^{j,k+1} \Big\rangle$$
$$\text{s.t. } (x, \rho, T) \in X_\rho \times \bar{B},$$
$$f(x) \leq \hat{y},$$

  Set $x^{k+1} \leftarrow x$ and $\lambda_j^{k+1} \leftarrow \lambda_j^k + 2r_j A(x^{k+1} - \tilde{x}^{j,k+1})$
  if one of the Stopping Criteria holds **then**
    return $(x^{k+1}, \tilde{x}^{j,k+1}, z^{j,k+1})$
  end if
end for

▷ Stopping Criteria

*Stopping Criteria: $\|A(x^{k+1} - \tilde{x}^{j,k+1})\| \leq \varepsilon$ for $j = 1, \ldots, N$, or $k = K_{\max}$.

Chapter 6

Perspective and future works

In this chapter, we discuss alternative approaches for solving model (4.2.6), as well as the potential for applying the methods considered in previous chapters to other operational planning models.

We begin by investigating the potential applicability of the Branch-and-bound method, with the primary challenge being the determination of an accurate relaxation of the chance constraint. As an initial step in this direction, we consider an SDP relaxation of the deterministic AC-OPFs introduced in Chapter 2 and Chapter 4.

Next, we discuss the interest in applying an Alternating Direction Method of Multipliers (ADMM) to a spatial decomposition of the chance-constrained AC-OPF, highlighting the main challenge of handling the joint chance constraint. Another promising direction for future research is reducing the sample size of considered scenarios to enhance computational efficiency.

Finally, we review the limitations of model generalizations within the operational planning framework, while preserving the applicability of DoC and CwC approaches to the chance-constrained formulation. We also discuss the potential for transitioning to the probability maximization model and applying the scenario decomposition approach.

Ce chapitre explore des approches alternatives pour résoudre le modèle (4.2.6)*, ainsi que la possibilité d'appliquer les méthodes présentées dans les chapitres précédents à d'autres modèles de gestion prévisionnelle.*

Nous analysons d'abord l'applicabilité potentielle de la méthode Branch-and-Bound, dont le défi principal consiste à relaxer la contrainte probabiliste. Comme première étape dans cette direction, nous considérons une relaxation SDP (semi-définie positive) des problèmes AC-OPF déterministes introduits dans le Chapitre 2 et le Chapitre 4.

Nous abordons ensuite l'intérêt d'appliquer une méthode ADMM (Alternating Direction Method of Multipliers) à une décomposition spatiale du problème AC-OPF sous contrainte probabiliste, en soulignant la difficulté principale que représente le traitement de la contrainte probabiliste jointe. Une autre piste prometteuse pour de futures recherches concerne la réduction de la taille de l'échantillon de scénarios afin d'améliorer l'efficacité numérique.

Enfin, nous évaluons les limites des extensions possibles du modèle dans le contexte de la gestion prévisionnelle, tout en maintenant l'applicabilité des approches DoC et CwC. Le potentiel de transition vers un modèle de maximisation de probabilité et d'implémentation de la décomposition par scénarios est également discuté.

6.1 Branch-and-bound method

The symmetry properties described in Proposition 4.2.1 and Proposition 4.2.2 for the constraints representing the priority and fairness principles enable removing redundant branches from the search tree when applying a Branch-and-bound method to model (4.2.6).

To eliminate branches that do not contain an optimal solution, it is essential to effectively compute a lower bound on the optimal value of (4.2.6) for a given candidate solution (with fixed binary values). Therefore, a convex relaxation of the chance constraint in (4.2.6) is required. Conservative approximations of the chance constraint, such as the conditional value-at-risk (CVaR) approximation [62], widely used in risk-averse framework (see, for instance, [29]), are not suitable for this goal, as they provide an upper bound. For a discrete probability distribution, a convex relaxation of the chance-constrained is derived from the DoC reformulation given in Subsection 2.3.2 (which is a relaxation in this case) by linearizing the second convex component. However, while the DoC reformulation is defined by a closed formula at a given point, an explicit global definition of the feasible set is still required.

Providing a global convex relaxation for a joint chance-constrained AC-OPF is a challenging problem that, to the best of our knowledge, has not yet been solved in a general framework. Although a probability constraint does not necessarily preserve the convexity of the underlying function, convex relaxations of AC-OPFs and linearizations of power flow equations are commonly used approaches in the literature. In [103, 83], individual chance constraints are analytically reformulated along with linearizations of underlying power flow equations, while SOCP reformulations are used in [55]. A scenario-based method is applied to a joint chance-constrained optimization problem in [87] and joint chance-constrained AC-OPF in [146]. In these works, the chance constraint is replaced by a certain number of deterministic constraints which ensure satisfaction of the original constraint in probabilistic

sense. The underlying deterministic problem must be convex, and an SDP relaxation of AC-OPF is thus used in the latter paper.

This motivates us to take the first step by selecting a suitable relaxation for a deterministic AC-OPF. For radial distribution networks, AC-OPF relaxations are extensively studied and well-understood. In particular, SDP relaxations, obtained by transforming the nonconvex power flow constraints into convex constraints on a rank-one positive semidefinite matrix and then relaxing the rank-one constraint. These relaxations have been analyzed from both algebraic [71] and geometric [72] perspectives. The exactness of SDP relaxations for radial networks was shown in [71, 72, 47, 153] under additional assumptions.

However, the deterministic AC-OPF (2.3.2) does not fit within the considered framework of AC-OPFs, where the exactness of SDP relaxations is established. Specifically, the upper and lower bounds on voltage angles, constraints (2.2.2a), cannot be expressed in terms of the positive semidefinite matrix, as we will discuss in the next subsection. We modify AC-OPF (2.3.2) by incorporating the constraints on the differences between voltage angles at adjacent buses. We then apply methodology proposed in [71] to the resulting problem, which differs from the framework considered in the paper, to explore the applicability of the method to our case and identify the main challenges.

6.1.1 SDP Relaxation

In this subsection, we consider a modified version of AC-OPF (2.3.2) at a given point (p^k, q^k) and scenario ξ. First, we denote by $(\mathbf{p}_i, \mathbf{q}_i)$ the resulting power modulated at the node $i \in \mathscr{N} \setminus \{sb\}$ to harmonize the notation in Chapter 2 and Chapter 4.

An SDP relaxation is based on the representation of problem variables in terms of the product of two voltage variables, which is used to construct the positive semidefinite matrix. This step is not possible for initial constraints on voltage angles (2.2.2a). Hence, we replace the upper and lower bounds on δ_i, $i \in \mathscr{N} \setminus \{sb\}$ with the constraints on the differences between voltage angles at adjacent buses $\delta_i - \delta_j$ for $(i, j) \in \mathscr{A}$:

$$\underline{\delta_{ij}} \leq \delta_i - \delta_j \leq \overline{\delta_{ij}}, \quad (i, j) \in \mathscr{A}. \tag{6.1.1}$$

If the bounds on the voltage angles are sufficiently close for the adjacent buses in (2.2.2a), constraints (6.1.1) represent a relaxation of the former constraints.

Following [71], we use the complex notation for voltage, current, and the admittance matrix. We denote the conjugate transpose of the operator A as A^*.

For simplicity, we assume that the bounds in (6.1.1) lie within the range $(-\frac{\pi}{2}, \frac{\pi}{2})$. Under this assumption, the constraints on the differences between voltage angles can equivalently

be expressed as:

$$\tan(\underline{\delta_{ij}})\,\mathrm{Re}(V_iV_j^*) \le \mathrm{Im}(V_iV_j^*) \le \tan(\overline{\delta_{ij}})\,\mathrm{Re}(V_iV_j^*), \quad (i,j) \in \mathscr{A}.$$

Therefore, we consider the following problem:

$$\min_{\alpha,\beta,V} \quad \sum_{i\in\mathscr{N}\setminus\{sb\}} \alpha_i + \beta_i \tag{6.1.2a}$$

$$\underline{|V_i|}^2 \le |V_i|^2 \le \overline{|V_i|}^2, \qquad i \in \mathscr{N} \tag{6.1.2b}$$

$$p_{sb}^{\min} \le \mathrm{Re}(V_{sb}I_{sb}^*) \le p_{sb}^{\max}, \tag{6.1.2c}$$

$$q_{sb}^{\min} \le \mathrm{Im}(V_{sb}I_{sb}^*) \le q_{sb}^{\max}, \tag{6.1.2d}$$

$$\mathrm{Im}(V_{sb}I_{sb}^*) \ge \frac{-0.48p_{sb}^{\max}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}}\mathrm{Re}(V_{sb}I_{sb}^*) + \frac{0.48p_{sb}^{\max}p_{sb}^{\min}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}}, \tag{6.1.2e}$$

$$|V_i - V_j|^2 \le (V_{i,j}^{\max})^2, \qquad (i,j) \in \mathscr{A} \tag{6.1.2f}$$

$$\tan(\underline{\delta_{ij}})\,\mathrm{Re}(V_iV_j^*) \le \mathrm{Im}(V_iV_j^*) \le \tan(\overline{\delta_{ij}})\,\mathrm{Re}(V_iV_j^*), \qquad (i,j) \in \mathscr{A} \tag{6.1.2g}$$

$$\begin{pmatrix} (p_i^k)^2 - 2p_i^k(p_i^\phi(\xi) - \mathrm{Re}(V_iI_i^*)) - \alpha_i & p_i^\phi(\xi) - \mathrm{Re}(V_iI_i^*) \\ p_i^\phi(\xi) - \mathrm{Re}(V_iI_i^*) & -1 \end{pmatrix} \preceq 0, \quad i \in \mathscr{N}\setminus\{sb\} \tag{6.1.2h}$$

$$\begin{pmatrix} (q_i^k)^2 - 2q_i^k(q_i^\phi(\xi) - \mathrm{Im}(V_iI_i^*)) - \beta_i & q_i^\phi(\xi) - \mathrm{Im}(V_iI_i^*) \\ q_i^\phi(\xi) - \mathrm{Im}(V_iI_i^*) & -1 \end{pmatrix} \preceq 0, \quad i \in \mathscr{N}\setminus\{sb\}. \tag{6.1.2i}$$

Constraints (6.1.2b) represent constraints on voltage magnitude in model (2.3.2), and an additional constraint on voltage magnitude at the slack bus with $\underline{|V_{sb}|}^2 = \overline{|V_{sb}|}^2 = 1$. Constraints (6.1.2f) represent congestion constraints in model (2.3.2) with $V_{i,j}^{\max} = I_{i,j}^{\max}/|Y_{i,j}|$ for all $(i,j) \in \mathscr{A}$. Observe that minimizing $\sum_{i\in(\mathscr{N}\setminus\{sb\})}(\mathbf{p}_i - p_i^k)^2 + (\mathbf{q}_i - q_i^k)^2$ is equivalent to minimizing $\sum_{i\in(\mathscr{N}\setminus\{sb\})}\alpha_i + \beta_i$ under the constraints

$$(\mathbf{p}_i - p_i^k)^2 \le \alpha_i, \quad i \in \mathscr{N}\setminus\{sb\}$$
$$(\mathbf{q}_i - q_i^k)^2 \le \beta_i, \quad i \in \mathscr{N}\setminus\{sb\}.$$

Constraints (6.1.2h) and (6.1.2i) are obtained using Schur complement [71]. Using the equation $I = YV$, the current variables can be eliminated from problem (6.1.2).

Compared to the problem studied in [71], our case includes additional constraints (6.1.2c)-(6.1.2g). As a result, the proof must be adapted, specifically regarding the formulation of the

Lagrangian, the structure of the matrix $A(x,r)$ and the demonstration of the existence of an interior point for the dual problem.

Dual problem. Denote the corresponding Lagrange multipliers for (6.1.2b) - (6.1.2g) as $\mu^{\min}, \mu^{\max} \in \mathbb{R}_+^{|\mathscr{N}|}$, $\lambda_{sb}^{\min}, \lambda_{sb}^{\max} \in \mathbb{R}_+$, $\overline{\lambda}_{sb}^{\min}, \overline{\lambda}_{sb}^{\max} \in \mathbb{R}_+$, $\hat{\lambda}_{sb} \in \mathbb{R}_+$, $\rho \in \mathbb{R}_+^{|\mathscr{A}|}$, $\overline{\mu}^{\min}, \overline{\mu}^{\max} \in \mathbb{R}_+^{|\mathscr{A}|}$. In accordance with the generalized Lagrangian theory, the multiplier associated with the inequalities (6.1.2h) and (6.1.2i) are symmetric positive semidefinite matrices:

$$\begin{pmatrix} r_{i0} & r_{i1} \\ r_{i1} & r_{i2} \end{pmatrix} \succeq 0, \quad \begin{pmatrix} \tilde{r}_{i0} & \tilde{r}_{i1} \\ \tilde{r}_{i1} & \tilde{r}_{i2} \end{pmatrix} \succeq 0, \qquad i \in \mathscr{N} \setminus \{sb\},$$

and the inner product is defined as the trace of the operators.

Let $n = |\mathscr{N}|$, and let $e_1, \ldots, e_n$ be the standard vector basis in $\mathbb{R}^n$. For $i \in \mathscr{N}$, let $Y_i = e_i e_i^* Y$ and

$$\mathbf{Y}_i = \frac{1}{2} \begin{bmatrix} \mathrm{Re}\{Y_i + Y_i^T\} & \mathrm{Im}\{Y_i^T - Y_i\} \\ \mathrm{Im}\{Y_i - Y_i^T\} & \mathrm{Re}\{Y_i + Y_i^T\} \end{bmatrix},$$

$$\overline{\mathbf{Y}}_i = -\frac{1}{2} \begin{bmatrix} \mathrm{Im}\{Y_i + Y_i^T\} & \mathrm{Re}\{Y_i - Y_i^T\} \\ \mathrm{Re}\{Y_i^T - Y_i\} & \mathrm{Im}\{Y_i + Y_i^T\} \end{bmatrix},$$

and define $M_i \in \mathbb{R}^{2n \times 2n}$ a matrix with zero entries, except for the elements (i,i) and $(n+i, n+i)$ that are equal to 1. For $(i,j) \in \mathscr{A}$, let $P_{ij} \in \mathbb{R}^{2n \times 2n}$ be

$$P_{ij} = \begin{bmatrix} e_i e_i^* - e_i e_j^* - e_j e_i^* + e_j e_j^* & 0 \\ 0 & e_i e_i^* - e_i e_j^* - e_j e_i^* + e_j e_j^* \end{bmatrix}.$$

For $(i,j) \in \mathscr{A}$, we introduce as well $\mathscr{M}_{ij}, \overline{\mathscr{M}}_{ij} \in \mathbb{R}^{2n \times 2n}$ given by

$$\mathscr{M}_{ij} = \begin{bmatrix} \frac{e_i e_j^* + e_j e_i^*}{2} & 0 \\ 0 & -\frac{e_i e_j^* + e_j e_i^*}{2} \end{bmatrix}, \quad \overline{\mathscr{M}}_{ij} = \begin{bmatrix} 0 & \frac{e_i e_j^* + e_j e_i^*}{2} \\ \frac{e_i e_j^* + e_j e_i^*}{2} & 0 \end{bmatrix}.$$

Denoting $U = \begin{bmatrix}\mathrm{Re}\{V\}^T & \mathrm{Im}\{V\}^T\end{bmatrix}^T$, we get the following chains of equations

$$\mathrm{Re}\{V_i I_i^*\} = \mathrm{Re}\{V^* Y_i V\} = U^T \mathbf{Y}_i U = \mathrm{trace}\{\mathbf{Y}_i U U^T\}, \qquad i \in \mathscr{N}, \tag{6.1.3a}$$

$$\mathrm{Im}\{V_i I_i^*\} = \mathrm{Im}\{V^* Y_i V\} = U^T \overline{\mathbf{Y}}_i U = \mathrm{trace}\{\overline{\mathbf{Y}}_i U U^T\} \qquad i \in \mathscr{N}. \tag{6.1.3b}$$

Furthermore,

$$|V_i|^2 = U^T M_i U = \mathrm{trace}\{M_i U U^T\}, \qquad i \in \mathscr{N}, \tag{6.1.4a}$$

$$|V_i - V_j|^2 = U^T P_{ij} U = \mathrm{trace}\{P_{ij} U U^T\}, \qquad (i,j) \in \mathscr{A}, \tag{6.1.4b}$$

$$\mathrm{Re}\{V_i V_j^*\} = U^T \mathscr{M}_{ij} U = \mathrm{trace}\{\mathscr{M}_{ij} U U^T\}, \qquad (i,j) \in \mathscr{A}, \tag{6.1.4c}$$

$$\mathrm{Im}\{V_i V_j^*\} = U^T \overline{\mathscr{M}}_{ij} U = \mathrm{trace}\{\overline{\mathscr{M}}_{ij} U U^T\}, \qquad (i,j) \in \mathscr{A}. \tag{6.1.4d}$$

The Lagrangian corresponding to problem (6.1.2) is given by

$$
\begin{aligned}
&\sum_{i\in\mathscr{N}\setminus\{sb\}}\alpha_i+\sum_{i\in\mathscr{N}\setminus\{sb\}}\beta_i+\sum_{i\in\mathscr{N}}\mu_i^{\min}\big(\underline{|V_i|}^2-|V_i|^2\big)-\mu_i^{\max}\big(\overline{|V_i|}^2-|V_i|^2\big)\\
&+\lambda_{sb}^{\min}\big(p_{sb}^{\min}-\mathrm{Re}(V_{sb}I_{sb}^*)\big)-\lambda_{sb}^{\max}\big(p_{sb}^{\max}-\mathrm{Re}(V_{sb}I_{sb}^*)\big)\\
&+\overline{\lambda}_{sb}^{\min}\big(q_{sb}^{\min}-\mathrm{Im}(V_{sb}I_{sb}^*)\big)-\overline{\lambda}_{sb}^{\max}\big(q_{sb}^{\max}-\mathrm{Im}(V_{sb}I_{sb}^*)\big)\\
&+\hat{\lambda}_{sb}\Big(\frac{-0.48p_{sb}^{\max}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}}\mathrm{Re}(V_{sb}I_{sb}^*)+\frac{0.48p_{sb}^{\max}p_{sb}^{\min}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}}-\mathrm{Im}(V_{sb}I_{sb}^*)\Big)\\
&+\sum_{(i,j)\in\mathscr{A}}\rho_{ij}\big(|V_i-V_j|^2-(V_{i,j}^{\max})^2\big)\\
&+\sum_{(i,j)\in\mathscr{A}}\overline{\mu}_{ij}^{\min}\big(\tan(\underline{\delta_{ij}})\,\mathrm{Re}(V_iV_j^*)-\mathrm{Im}(V_iV_j^*)\big)-\sum_{(i,j)\in\mathscr{A}}\overline{\mu}_{ij}^{\max}\big(\tan(\overline{\delta_{ij}})\,\mathrm{Re}(V_iV_j^*)-\mathrm{Im}(V_iV_j^*)\big)\\
&+\sum_{i\in\mathscr{N}\setminus\{sb\}}r_{i0}\left((p_i^k)^2-2p_i^k\left(p_i^{\phi}(\xi)-\mathrm{Re}(V_iI_i^*)\right)-\alpha_i\right)+2r_{i1}\left(p_i^{\phi}(\xi)-\mathrm{Re}(V_iI_i^*)\right)-r_{i2}\\
&+\sum_{i\in\mathscr{N}\setminus\{sb\}}\tilde{r}_{i0}\left((q_i^k)^2-2q_i^k\left(q_i^{\phi}(\xi)-\mathrm{Im}(V_iI_i^*)-\beta_i\right)+2\tilde{r}_{i1}\left(q_i^{\phi}(\xi)-\mathrm{Im}(V_iI_i^*)\right)-\tilde{r}_{i2}\right.\\
&=\sum_{i\in\mathscr{N}\setminus\{sb\}}(1-r_{i0})\,\alpha_i+\sum_{i\in\mathscr{N}\setminus\{sb\}}(1-\tilde{r}_{i0})\beta_i+h(x,r)+\sum_{i\in\mathscr{N}}(\mu_i^{\max}-\mu_i^{\min})|V_i|^2\\
&+(\lambda_{sb}^{\max}-\lambda_{sb}^{\min})\mathrm{Re}(V_{sb}I_{sb}^*)+(\overline{\lambda}_{sb}^{\max}-\overline{\lambda}_{sb}^{\min})\mathrm{Im}(V_{sb}I_{sb}^*)\\
&+\hat{\lambda}_{sb}\Big(\frac{-0.48p_{sb}^{\max}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}}\mathrm{Re}(V_{sb}I_{sb}^*)-\mathrm{Im}(V_{sb}I_{sb}^*)\Big)+\sum_{(i,j)\in\mathscr{A}}\rho_{ij}|V_i-V_j|^2\\
&+\sum_{(i,j)\in\mathscr{A}}\big(\overline{\mu}_{ij}^{\min}\tan(\underline{\delta_{ij}})-\overline{\mu}_{ij}^{\max}\tan(\overline{\delta_{ij}})\big)\,\mathrm{Re}(V_iV_j^*)-\sum_{(i,j)\in\mathscr{A}}\big(\overline{\mu}_{ij}^{\min}-\overline{\mu}_{ij}^{\max}\big)\,\mathrm{Im}(V_iV_j^*)\\
&+\sum_{i\in\mathscr{N}\setminus\{sb\}}2\big(p_i^k\cdot r_{i0}-r_{i1}\big)\,\mathrm{Re}(V_iI_i^*)+\sum_{i\in\mathscr{N}\setminus\{sb\}}2\big(q_i^k\cdot\tilde{r}_{i0}-\tilde{r}_{i1}\big)\,\mathrm{Im}(V_iI_i^*)\\
&=\sum_{i\in\mathscr{N}\setminus\{sb\}}(1-r_{i0})\,\alpha_i+\sum_{i\in\mathscr{N}\setminus\{sb\}}(1-\tilde{r}_{i0})\,\beta_i+h(x,r)+\mathrm{trace}\big(A(x,r)\,UU^T\big),
\end{aligned}
\tag{6.1.5}
$$

where

$$
\begin{aligned}
x&=(\mu^{\min},\mu^{\max},\lambda_{sb}^{\min},\lambda_{sb}^{\max},\overline{\lambda}_{sb}^{\min},\overline{\lambda}_{sb}^{\max},\hat{\lambda}_{sb},\rho,\overline{\mu}^{\min},\overline{\mu}^{\max}),\\
r&=(r_{i0},r_{i1},r_{i2},\tilde{r}_{i0},\tilde{r}_{i1},\tilde{r}_{i2})_{i\in\mathscr{N}\setminus\{sb\}},
\end{aligned}
$$

and

$$h(x,r) = \sum_{i\in\mathscr{N}} \left(\mu_i^{\min}|\underline{V_i}|^2 - \mu_i^{\max}\overline{|V_i|}^2\right) + \lambda_{sb}^{\min} p_{sb}^{\min} - \lambda_{sb}^{\max} p_{sb}^{\max} + \overline{\lambda}_{sb}^{\min} q_{sb}^{\min} - \overline{\lambda}_{sb}^{\max} q_{sb}^{\max}$$
$$+\hat{\lambda}_{sb}\frac{0.48 p_{sb}^{\max} p_{sb}^{\min}}{-p_{sb}^{\min} + 0.25 p_{sb}^{\max}} - \sum_{(i,j)\in\mathscr{A}} \rho_{ij}(V_{i,j}^{\max})^2$$
$$+ \sum_{i\in\mathscr{N}\setminus\{sb\}} \left(r_{i0}\cdot(p_i^k)^2 + \tilde{r}_{i0}\cdot(q_i^k)^2 + 2(r_{i1} - p_i^k\cdot r_{i0})p_i^\phi(\xi) + 2(\tilde{r}_{i1} - q_i^k\cdot\tilde{r}_{i0})q_i^\phi(\xi) - r_{i2} - \tilde{r}_{i2}\right),$$

$$A(x,r) = \sum_{i\in\mathscr{N}} (\mu_i^{\max} - \mu_i^{\min}) M_i + \sum_{i\in\mathscr{N}\setminus\{sb\}} 2(p_i^k\cdot r_{i0} - r_{i1})\,\mathbf{Y}_i + 2(q_i^k\cdot\tilde{r}_{i0} - \tilde{r}_{i1})\,\overline{\mathbf{Y}}_i$$
$$+ \sum_{(i,j)\in\mathscr{A}} \rho_{ij}P_{ij} + \left(\overline{\mu}_{ij}^{\min}\tan(\underline{\delta_{ij}}) - \overline{\mu}_{ij}^{\max}\tan(\overline{\delta_{ij}})\right)\mathscr{M}_{ij} + \left(\overline{\mu}_{ij}^{\min} - \overline{\mu}_{ij}^{\max}\right)\overline{\mathscr{M}}_{ij}$$
$$+\left(\lambda_{sb}^{\max} - \lambda_{sb}^{\min} + \hat{\lambda}_{sb}\frac{-0.48 p_{sb}^{\max}}{-p_{sb}^{\min} + 0.25 p_{sb}^{\max}}\right)\mathbf{Y}_{sb} + \left(\overline{\lambda}_{sb}^{\min} - \overline{\lambda}_{sb}^{\max} - \hat{\lambda}_{sb}\right)\overline{\mathbf{Y}}_{sb}.$$

The dual problem for (6.1.2) is obtained by minimizing the Lagrangian over U, α and β, and then maximizing it over the Lagrange multipliers x and r.

- The minimum of the term $(1 - r_{i0})\,\alpha_i$ over α_i is $-\infty$, unless $r_{i0} = 1$, in which case the minimum is zero. The same argument holds for the term $(1 - \tilde{r}_{i0})\,\beta_i$. Therefore, we set $r_{i0} = 1$ and $\tilde{r}_{i0} = 1$ for $i \in \mathscr{N}\setminus\{sb\}$ and denote $\hat{r} = (r_{i1}, r_{i2}, \tilde{r}_{i1}, \tilde{r}_{i2})_{i\in\mathscr{N}\setminus\{sb\}}$.
- The minimum of the term $\mathrm{trace}\left(A(x,r)UU^T\right)$ over U is $-\infty$, unless $A(x,r) \succeq 0$, in which case the minimum is zero.

The dual problem to (6.1.2) is thus defined as follows:

$$\max_{x\geq 0,\,\hat{r}} h(x,\hat{r}) \tag{6.1.6a}$$
$$\begin{pmatrix} 1 & r_{i1} \\ r_{i1} & r_{i2} \end{pmatrix} \succeq 0, \quad \begin{pmatrix} 1 & \tilde{r}_{i1} \\ \tilde{r}_{i1} & \tilde{r}_{i2} \end{pmatrix} \succeq 0, \quad i \in \mathscr{N}\setminus\{sb\} \tag{6.1.6b}$$
$$A(x,\hat{r}) \succeq 0. \tag{6.1.6c}$$

SDP Relaxation. Substituting $W = UU^T$ in (6.1.2) and using the equations (6.1.3-6.1.4), we obtain an SDP relaxation of problem (6.1.2):

$$\min_{\alpha,\beta,W} \quad \sum_{i\in\mathscr{N}\setminus\{sb\}} \alpha_i + \beta_i \tag{6.1.7a}$$

$$\underline{|V_i|}^2 \le \mathrm{trace}\{M_i W\} \le \overline{|V_i|}^2, \qquad i \in \mathscr{N} \tag{6.1.7b}$$

$$p_{sb}^{\min} \le \mathrm{trace}\{\mathbf{Y}_{sb} W\} \le p_{sb}^{\max} \tag{6.1.7c}$$

$$q_{sb}^{\min} \le \mathrm{trace}\{\overline{\mathbf{Y}}_{sb} W\} \le q_{sb}^{\max}, \tag{6.1.7d}$$

$$\mathrm{trace}\{\overline{\mathbf{Y}}_{sb} W\} \ge \frac{-0.48 p_{sb}^{\max}}{-p_{sb}^{\min} + 0.25 p_{sb}^{\max}} \mathrm{trace}\{\mathbf{Y}_{sb} W\} + \frac{0.48 p_{sb}^{\max} p_{sb}^{\min}}{-p_{sb}^{\min} + 0.25 p_{sb}^{\max}}, \tag{6.1.7e}$$

$$\mathrm{trace}\{P_{ij} W\} \le (V_{i,j}^{\max})^2, \qquad (i,j) \in \mathscr{A} \tag{6.1.7f}$$

$$\tan(\underline{\delta_{ij}}) \mathrm{trace}\{\mathscr{M}_{ij} W\} \le \mathrm{trace}\{\overline{\mathscr{M}}_{ij} W\} \le \tan(\overline{\delta_{ij}}) \mathrm{trace}\{\mathscr{M}_{ij} W\}, \qquad (i,j) \in \mathscr{A} \tag{6.1.7g}$$

$$\begin{pmatrix} (p_i^k)^2 - 2p_i^k (p_i^\phi(\xi) - \mathrm{trace}\{\mathbf{Y}_i W\}) - \alpha_i & p_i^\phi(\xi) - \mathrm{trace}\{\mathbf{Y}_i W\} \\ p_i^\phi(\xi) - \mathrm{trace}\{\mathbf{Y}_i W\} & -1 \end{pmatrix} \preceq 0, \qquad i \in \mathscr{N}\setminus\{sb\} \tag{6.1.7h}$$

$$\begin{pmatrix} (q_i^k)^2 - 2q_i^k (q_i^\phi(\xi) - \mathrm{trace}\{\overline{\mathbf{Y}}_i W\}) - \beta_i & q_i^\phi(\xi) - \mathrm{trace}\{\overline{\mathbf{Y}}_i W\} \\ q_i^\phi(\xi) - \mathrm{trace}\{\overline{\mathbf{Y}}_i W\} & -1 \end{pmatrix} \preceq 0, \qquad i \in \mathscr{N}\setminus\{sb\} \tag{6.1.7i}$$

$$W \succeq 0. \tag{6.1.7j}$$

By adding an additional rank constraint, we obtain the fourth optimization problem required for the proof:

$$\begin{aligned} &\min_{\alpha,\beta,W} \quad \sum_{i\in\mathscr{N}\setminus\{sb\}} \alpha_i + \beta_i \\ &(6.1.7b) - (6.1.7j), \\ &\mathrm{rank}(W) = 1. \end{aligned} \tag{6.1.8}$$

Problem equivalence. The exactness of the SDP relaxation is established under additional assumptions by proving the following statements:

(i) The original problem (6.1.2) and the rank-one SDP model (6.1.8) are equivalent,

(ii) Strong duality holds between the SDP relaxation (6.1.7) and the dual problem (6.1.6),

(iii) The SDP relaxation (6.1.7) and the rank-one SDP model (6.1.8) are equivalent .

Furthermore, if these statements hold, a zero duality gap between the original problem (6.1.2) and the dual problem (6.1.6) is guaranteed, although proving this directly is challenging.

The equivalence (i) is shown based on the assumption of feasibility of (6.1.2) and its non-triviality, more precisely that $V = 0$ is not a feasible point of problem (6.1.2). In our case, $V = 0$ is not a feasible point of the deterministic AC-OPF due to the lower bound constraint on the voltage magnitude. Let V be a non-zero feasible vector of (6.1.2), then the matrix $W = UU^T$ with $U = \begin{bmatrix}\mathrm{Re}\{V\}^T & \mathrm{Im}\{V\}^T\end{bmatrix}^T$ has rank at most 1. On the other hand, any semidefinite matrix W with rank at most 1 can be decomposed as $W = UU^T$. This change of variables is a bijection (up to the sign of U). Moreover, constraints of (6.1.2) and (6.1.8) are mapped to each other by construction, and hence the optimization problems coincide.

The stong duality in (ii) follows from two observations. First, problem (6.1.7) is a dual problem for (6.1.6), where the Lagrangian is given by (6.1.5) with $W = UU^T$ and W is a Lagrangian multiplier for the constraint (6.1.6c). Second, strong duality holds by Slater condition provided the existence of an interior point of (6.1.6). While the proofs of the equivalence (i) and weak duality in (ii) remain the same as in [71], the existence of an interior point for (6.1.6) has to be shown anew due to a more complex structure of $A(x,\hat{r})$.

Proposition 6.1.1. *Optimization problem* (6.1.6) *has an interior point.*

Proof. Let $r_{i1} = p_i^k$, $\tilde{r}_{i1} = q_i^k$, and $r_{i2} = (p_i^k)^2 + 1$, $\tilde{r}_{i2} = (q_i^k)^2 + 1$, which implies the existence of an interior point in (6.1.6b). We set $\overline{\mu}_{ij}^{\min} = \overline{\mu}_{ij}^{\max} = 1$, and $\hat{\lambda}_{sb} = 1$, $\overline{\lambda}_{sb}^{\max} = 1$, $\overline{\lambda}_{sb}^{\min} = 2$. Moreover, let $\lambda_{sb}^{\min} = \left|\frac{-0.48p_{sb}^{\max}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}}\right| + 1$, and

$$\lambda_{sb}^{\max} = \begin{cases} 1, & \text{if} \quad \frac{-0.48p_{sb}^{\max}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}} \geq 0 \\ -2\frac{-0.48p_{sb}^{\max}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}} + 1, & \text{if} \quad \frac{-0.48p_{sb}^{\max}}{-p_{sb}^{\min}+0.25p_{sb}^{\max}} < 0. \end{cases}$$

For this choice of values, we obtain

$$A(x,\hat{r}) = \sum_{i\in\mathcal{N}} (\mu_i^{\max} - \mu_i^{\min})M_i + \sum_{(i,j)\in\mathcal{A}} \rho_{ij}P_{ij} + \left(\overline{\mu}_{ij}^{\min}\tan(\underline{\delta_{ij}}) - \overline{\mu}_{ij}^{\max}\tan(\overline{\delta_{ij}})\right)\mathscr{M}_{ij}.$$

For each $(i,j) \in \mathscr{A}$, the term $\rho_{ij}P_{ij} + \big(\tan(\underline{\delta_{ij}}) - \tan(\overline{\delta_{ij}})\big)\mathscr{M}_{ij}$ is a symmetric matrix

$$\begin{bmatrix} \rho_{ij}(e_ie_i^*+e_je_j^*) - \big(\rho_{ij}+\frac{a_{ij}}{2}\big)(e_ie_j^*+e_je_i^*) & 0 \\ 0 & \rho_{ij}(e_ie_i^*+e_je_j^*) - \big(\rho_{ij}-\frac{a_{ij}}{2}\big)(e_ie_j^*+e_je_i^*) \end{bmatrix}, \tag{6.1.9}$$

with $a_{ij} = \tan(\overline{\delta_{ij}}) - \tan(\underline{\delta_{ij}}) > 0$. Choose the same positive $\overline{\rho} := \rho_{ij}$ for all $(i,j) \in \mathscr{A}$, and denote by d_i the degree of a bus $i \in \mathscr{N}$. Taking the sum over $(i,j) \in \mathscr{A}$, we obtain a symmetric matrix $\tilde{S}$ with diagonal elements (i,i) and $(n+i,n+i)$ equal to $d_i \cdot \overline{\rho}$, $i = 1,\dots,n$.

Let $\overline{a} = \max a_{ij}$, $m = \frac{\overline{a}}{2}\max d_i + 1$, and $\mu_i^{\max} = m+1$, $\mu_i^{\min} = 1$. For this choice of values, the first sum in the formula of $A(x,\hat{r})$ is equal to $m \cdot \mathrm{Id}$ by definition of M_i, and $A(x,\hat{r})$ is equal to $S = m \cdot \mathrm{Id} + \tilde{S}$. It follows from (6.1.9) that $\sum_{j\neq i}|S_{ij}| \leq (\overline{\rho} + \frac{\overline{a}}{2})d_i < m + \overline{\rho}\cdot d_i = S_{ii}$. Hence, S is a strictly diagonally dominant matrix. We claim that every eigenvalue of S is positive, and thus $S \succ 0$.

Let λ be an eigenvalue of S with a corresponding eigenvector $x = (x_i)_{1\leq i\leq 2n}$. Find $\hat{i}$ such that $x_{\hat{i}} \geq x_j$ for all j. We obtain the following chain of inequalities

$$\sum_j S_{\hat{i}j}x_j = \lambda x_{\hat{i}} \quad \Leftrightarrow \quad |\lambda - S_{\hat{i}\hat{i}}| = \Big|\sum_{j\neq\hat{i}} \frac{S_{\hat{i}j}x_j}{x_{\hat{i}}}\Big| \leq \sum_{j\neq\hat{i}}|S_{\hat{i}j}| < S_{\hat{i}\hat{i}} \quad \Rightarrow \quad \lambda > 0.$$

□

The equivalence (iii) holds if the optimal solution of (6.1.7) has rank 1. As shown in [71], a sufficient condition for this equivalence to hold is the existence of an optimal solution $(x^*,\hat{r}^*)$ of (6.1.6) such that the matrix $A(x^*,\hat{r}^*)$ has a zero eigenvalue of multiplicity 2. The proof does not rely on a specific form of $A(x,\hat{r})$ and can, therefore, be directly applied to our case.

Verifying exactness of the relaxation. To check the exactness of the SDP relaxation of (6.1.2) in practice, we have three options. The first option is to solve the dual problem (6.1.6) and check whether its solution is feasible for problem (6.1.2) using the recovery procedure described in [71]. If this is the case, the SDP relaxation is exact due to strong duality in (ii).

The second option is to solve problem (6.1.7). If its solution has rank 1, the equivalence (i) implies the exactness of the SDP relaxation. Finally, the last option is to check the existence of a zero eigenvalue of multiplicity 2 for the matrix $A(x^*,\hat{r}^*)$ at the optimal solution $(x^*,\hat{r}^*)$ of (6.1.6).

All these options require solving convex semidefinite programs. However, in the second and third cases, determining the solution rank or the existence of a zero eigenvalue with

multiplicity 2 is challenging numerically. Since this involves verifying the satisfaction of equality constraints, the results are highly sensitive to computational errors. A numerical method to assess these conditions, based on the calculation of the minimum eigenvalue ratio, has been implemented in [89, 143].

6.2 Spatial decomposition of the chance-constrained OPF

The scenario decomposition approach applied in Section 5.2 to the chance-constrained AC-OPF enables isolating the binary variables describing priority and fairness principles from nonconvex power flow equations. To apply this approach, we introduced additional binary variables associated with each scenario and reformulated the chance constraint as a scenario-coupling linear constraint that we dualized. We also envisaged another decomposition for the chance-constrained model.

Motivated by a partitioning of the network which naturally arises from the definition of critical nodes and thus is determined by the topology of congestion constraints, we look in the direction of a spatial decomposition. In a deterministic framework, ADMM is a widely used decentralized algorithm for solving optimization problems with coupling constraints. In particular, it has been applied to convex relaxations [26] and approximations of AC-OPF, for instance, to the DC approximation [147], where theoretical convergence has been established. In [119, 86, 38], ADMM and its modifications have been applied to AC-OPF based on the separation of power flow equations along the nodes.

For nonconvex problems, there are no theoretical convergence guarantees for ADMM in the general case. However, the convergence of some ADMM schemes has been shown under additional assumptions. These assumptions often include the ability to solve non-convex subproblems to global optimality [148], or the availability of a high-quality starting point [119], ensuring that a stationary subproblem solution provides a reduced objective value compared to the previous iteration. The first assumption does not hold in our case, and the second assumption also may not hold in practice. For the scenario decomposition approach, the availability of good-quality starting points for several scenarios did allow us to increase the overall probability, while resolving deterministic AC-OPFs. Nevertheless, for a large number of scenarios, we observed that the value of the objective function in (5.2.8) and (5.2.14) fails to decrease or decreases insufficiently, and Algorithm 4 classifies this solution as too expensive at Step 8, which implies a decrease in overall probability. With the inclusion of binary variables describing the priority and fairness principles, we do not expect any theoretical convergence for ADMM.

Leaving aside binary variables, we decide to explore the potential of applying ADMM to a spatial decomposition of the chance-constrained AC-OPF. In the framework of individual chance constraints, the method has been applied, for example, to the energy management problem in [80] using preliminary analytical reformulations. However, to the best of our knowledge, spatial decomposition has not been explored yet within the joint chance-constrained framework.

This raises the following question: How can the joint nature of the chance constraint be reconciled with the distributed nature of ADMM? If we retain a chance constraint for each subproblem, that is, consider independent chance-constrained AC-OPF on each component of the network, the information from the initial joint probability constraint will be lost. Is it possible to find a reasonable relaxation of the joint chance constraint for the given partition of the network, enabling the application of ADMM on a spatial decomposition of AC-OPF?

To address this issue, we may consider the conditional probability of solution existence within each component, given that the corresponding power flows are correctly glued together at the intersection nodes. This introduces a new challenge: state variables are implicit, derived from active and reactive power, making it impossible to directly impose consensus constraints on them at the intersections. Consequently, the development of this approach would require a reformulation of deterministic AC-OPF for our specific case.

6.3 Scenario reduction method

The computational efficiency of the methods considered in this work, for both continuous and mixed-integer models, depends on the sample size N. For the DoC and CwC methods, Oracle 1 must be called N times per iteration to estimate the probability function, whereas the scenario decomposition approach results in N deterministic AC-OPFs. However, not all scenarios are computationally equivalent; some require solving only the system of power flow equations and verifying that the bounds on state variables are satisfied, instead of resolving an AC-OPF (as described in Step 2 of Oracle 1 and Step 4 of Algorithm 4). We propose several directions for future research on the sample properties.

The first idea is to integrate sample preprocessing to reduce the number of scenarios considered at each iteration. This can be approached in two ways: excluding scenarios that are expensive to satisfy a priori or representing scenarios with a smaller sample. For the probability maximization model, the first approach relies on an initial cost estimation of the power curtailment required to satisfy each scenario independently. If, for a given scenario, the estimated power curtailment calculated without considering the priority and fairness rules exceeds the budget constraint, this scenario can be excluded from the subsequent

iterations. For the chance-constrained model with a given security level $1-\alpha$, the first approach involves identifying scenarios that are guaranteed to fall within the αN unsatisfied scenarios. However, identifying such scenarios during a preliminary step can be challenging. This procedure can be implemented as a stepwise elimination of scenarios that do not satisfy the chance constraint after several serious iterations of the DoC or CwC algorithms. The theoretical properties of this procedure remain to be studied.

The second idea, representing scenarios with a smaller sample, relies on clustering techniques. This unavoidably results in a less accurate representation of uncertainty, and therefore the choice of a particular method should be analyzed on a case-by-case basis. For the chance-constrained problem, the authors of [2, 108] propose dynamic scenario grouping models that provide lower and upper bounds for the initial problem. By iteratively improving these bounds using scenario refinement and merging procedures, the solution of the initial problem can be recovered [108]. At each iteration, a chance-constrained model needs to be solved, and thus this method can potentially be integrated into the DoC and CwC bundle methods.

6.4 Model generalizations

In this section, we discuss the potential of applying the DoC and CwC approaches from Chapter 2 and Section 3.8, as well as the scenario decomposition approach to other operational planning models, considering the example of the reactive power control.

In the chance-constrained model (2.2.4), the decision bears only on the active power modulation, as reactive power variables are defined as functions from the active ones. Making reactive power a free variable enables modeling reactive power regulation as a control lever. In this case, the DoC and CwC approaches remain applicable, as long as the deterministic constraints are convex and the probability constraint is unchanged in (2.2.4).

The integration of a regulation system capable of dynamically adjusting reactive power consumption and production based on measured voltage is proposed in [32]. Specifically, the ratio of reactive to active power in a production facility is modeled as a piecewise linear function of the deviation in magnitude of the voltage from the nominal voltage. Modeling this system as a chance-constrained AC-OPF requires considering the reactive power as a state variable, as the decision no longer directly bears on its regulation. Consequently, the power flow equations and the definition of the random set $X(\xi)$ need to be adapted to integrate the dependence of the reactive power from the voltage magnitude deviation. As discussed in Chapter 2, the methodology for representing the probability constraint as a DoC function is applicable once the set $X(\xi)$ is mathematically defined and a solver capable of

handling the associated optimization problem is available. Therefore, after modifying the equations that define the set $X(\xi)$, the DoC and CwC approaches can be directly applied.

Combining the direct reactive power regulation and the reactive power adjustment depending on the voltage magnitude deviation for different groups of producers is also possible. If these groups are fixed (e.g., producers have signed contracts specifying the type of reactive power regulation), then it is sufficient to modify the power flow equations and the definition of the set $X(\xi)$, incorporating the dependence of reactive power on voltage magnitude deviation for one of the groups.

At the same time, the dynamic choice of these groups requires the integration of binary variables in the chance-constrained AC-OPF model. Since this choice determines whether the reactive power is a state variable for a given producer and thus influences the form of the power flow equations, the associate binary variables must be integrated into the probability constraint. Consequently, the methods for integrating binary variables into the DoC model considered in Section 5.1, cannot be applied, as they address only binary variables within the deterministic part of the model. Transitioning to a probability maximization model (following Subsection 5.2.2) and applying the scenario decomposition approach is justified in this case. However, separating binary variables associated with the choice of the group from the nonconvexity related to the power flow equations is not possible, and the complexity of the problem is thus embedded in the technical constraints. More precisely, in this case, technical constraints include the binary variables defining the form of the power flow equations (dependent on the choice of the group) and the scenario satisfaction along with the nonconvexity related to the power flow equations. This results in solving multiple deterministic AC-OPFs per scenario, one for each combination of binary values. Consequently, the practical interest of this method depends on its numerical performance for a specific use case.

Generalizing this discussion on reactive power regulation, we highlight the following principles. First, as long as the deterministic constraints in the chance-constrained model (2.2.4) remain convex, the DoC and CwC approaches are applicable. Moreover, the system of equations defining the set $X(\xi)$ can be quite general, with the only requirement being the availability of a solver capable of handling the associated optimization problem. This implies that the DoC and CwC approaches are not specific to distribution grid topology and can also be applied to operational planning models for power transmission grids. Further, integrating deterministic mixed-integer convex constraints places the problem within the framework discussed in Chapter 5. In this case, the modified methods are applicable, but their performance depends on the specific use case. Another option is transitioning to the probability maximization problem and applying the scenario decomposition approach. For

the case involving binary variables integrated into probability constraints, only the latter approach is applicable. This case requires further investigation.

6.5 Conclusion

In this chapter, we discussed potential directions for future research, including investigating alternative methods for solving the chance-constrained AC-OPF model with discrete formulations and enhancing computational efficiency by reducing sample size. As a preliminary step toward the first direction, we studied the SDP relaxation of deterministic AC-OPF and the conditions implying the exactness of this relaxation. Furthermore, we discussed generalization limits for operational planning models, emphasizing the applicability of the approaches studied in this work.

Conclusion

In this thesis, we investigated short-term operational planning problems modeled as joint chance-constrained AC-OPFs. In these formulations, the decisions bear on the activation of power modulation and curtailment levers for FiT and SCP producers and consumers, or their generalized counterparts represented by priority levels. First, a continuous joint chance-constrained AC-OPF model with simplified operational planning rules was considered. To enhance the realism of the model, logical and discrete formulations were then introduced to represent priority and fairness principles in power curtailment. The resulting models were addressed by decomposing the underlying complexities without relying on preliminary simplifications, such as decomposing the joint probability constraint or approximating the AC-OPF.

We began by considering the continuous version of the model and applied the DoC approach to resolve it. This approach consists of decomposing the joint probability constraint as a difference of two convex functions and employing a known DoC bundle method. To accomplish the first step, we proposed a parallelizable numerical procedure (oracle) enabling the decomposition of the underlying formulation in probability constraint into a deterministic AC-OPF per scenario. Combined with Monte-Carlo simulations, this procedure leads to a required decomposition. However, the resulting model does not satisfy several assumptions, and the DoC bundle method is not guaranteed to provide a critical point to the original model.

To address this challenge, we proposed a new bundle method with stronger convergence guarantees under weaker assumptions. For the joint chance-constrained AC-OPF, this algorithm provides a critical (generalized KKT) point. Numerical experiments in a 33-bus distribution network showed that it provides a more cost-effective solution compared to the DoC bundle method. However, achieving stronger criticality leads to an increase in execution time. Furthermore, the new method is capable of handling a broad class of nonsmooth and nonconvex optimization problems beyond this framework, provided that the objective and constraint functions can be represented as differences of convex and weakly convex functions. We studied various stationary conditions and provided a detailed convergence analysis for the new algorithm. Its practical performance was illustrated on several nonconvex stochastic problems and a compressed-sensing problem with nonlinear constraints.

For the joint chance-constrained AC-OPF with logical and discrete formulations, we first tested the capability of the bundle methods to address this model. Numerical experiments with the DoC bundle method revealed limitations in handling binary variables. We also considered an alternative approach, which involves introducing a binary variable for each scenario and reformulating the problem as a maximization of the number of satisfied scenarios. We evaluated different Lagrangian methods, with or without a regularization term, enabling the separation of the discrete and stochastic parts of the probability maximization model and decomposing the latter into scenario-dependent deterministic AC-OPF subproblems. Although it lacks theoretical convergence guarantees, the relevance of this approach was demonstrated in practice for several use cases in a 33-bus distribution network. Both approaches, based on the bundle methods and scenario decomposition, have the potential to address the identified limitations, either by improving theoretical foundations or practical performance.

We also discussed the potential for applying the proposed methods to other operational planning models and outlined directions for future research to develop alternative approaches for joint chance-constrained problems with discrete variables. In this context, we considered applying ADMM for a spatial decomposition of the model, and scenario reduction techniques. Additionally, we studied the SDP relaxation of the underlying deterministic AC-OPF as a preliminary step toward implementing a Branch-and-Bound method.

Two articles derived from this Thesis have been published in international journals: the main contents of Chapter 2 and Chapter 3 have appeared in [123] and [124], respectively.

References

[1] S. Ahmed. A scenario decomposition algorithm for 0–1 stochastic programs. *Operations Research Letters*, 41(6):565–569, 2013. doi:10.1016/j.orl.2013.07.009.

[2] S. Ahmed, J. Luedtke, Y. Song, and W. Xie. Nonanticipative duality, relaxations, and formulations for chance-constrained stochastic programs. *Mathematical Programming*, 162:51–81, 2017. doi:10.1007/s10107-016-1029-z.

[3] K.-M. Aigner, J.-P. Clarner, F. Liers, and A. Martin. Robust approximation of chance constrained DC optimal power flow under decision-dependent uncertainty. *European Journal of Operational Research*, 301(1):318–333, 2022. doi:10.1016/j.ejor.2021.10.051.

[4] S. Alyami. Fairness and usability analysis in renewable power curtailment: A microgrid network study using bankruptcy rules. *Renewable Energy*, 227:Article 120424, 2024. doi:10.1016/j.renene.2024.120424.

[5] K. L. Anaya and M. G. Pollitt. Going smarter in the connection of distributed generation. *Energy Policy*, 105:608–617, 2017. doi:10.1016/j.enpol.2017.01.036.

[6] R. Ansaripour, H. Barati, and A. Ghasemi. A chance-constrained optimization framework for transmission congestion management and frequency regulation in the presence of wind farms and energy storage systems. *Electric Power Systems Research*, 213:Article 108712, 2022. doi:10.1016/j.epsr.2022.108712.

[7] P. Apkarian, D. Noll, and A. Rondepierre. Mixed H_2/H_∞ Control via Nonsmooth Optimization. *SIAM J. Control Optim.*, 47(3):1516–1546, 2008. doi:10.1137/070685026.

[8] K. Baker and A. Bernstein. Joint Chance Constraints in AC Optimal Power Flow: Improving Bounds Through Learning. *IEEE Transactions on Smart Grid*, 10(6):6376–6385, 2019. doi:10.1109/TSG.2019.2903767.

[9] M. Baran and F. Wu. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Transactions on Power Delivery*, 4(2):1401–1407, 1989. doi:10.1109/61.25627.

[10] R. Bauer, T. Mühlpfordt, N. Ludwig, and V. Hagenmeyer. Analytical uncertainty propagation for multi-period stochastic optimal power flow. *Sustainable Energy, Grids and Networks*, 33:Article 100969, 2023. doi:10.1016/j.segan.2022.100969.

[11] D. Bienstock, M. Chertkov, and S. Harnett. Chance-constrained optimal power flow: Risk-aware network control under uncertainty. *SIAM Review*, 56(3):461–495, 2014. doi:10.1137/130910312.

[12] D. Bienstock and A. Verma. Strong NP-hardness of AC power flows feasibility. *Oper. Res. Lett.*, 47(6):494–501, 2019. doi:10.1016/j.orl.2019.08.009.

[13] J. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization problems*. Springer Ser. Oper. Res. Financ. Eng. . Springer - New York, 1st edition, 2000. doi:10.1007/978-1-4612-1394-9.

[14] P. Borges, C. Sagastizábal, and M. Solodov. A regularized smoothing method for fully parameterized convex problems with applications to convex and nonconvex two-stage stochastic programming. *Math. Program.*, 189(1):117–149, 2021. doi:10.1007/s10107-020-01582-2.

[15] R. Bouckaert. From the green deal to repower EU, 2023. Accessed: 2023-05-13. URL: https://feps-europe.eu/publication/from-the-green-deal-to-repower-eu/.

[16] S. P. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3:1–122, 2011. doi:10.1561/2200000016.

[17] C. C. Carøe and R. Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24(1):37–45, 1999. doi:10.1016/S0167-6377(98)00050-9.

[18] W. Chen, M. Sim, J. Sun, and C.-P. Teo. From CVaR to Uncertainty Set: Implications in Joint Chance-Constrained Optimization. *Operations Research*, 58(2):470–485, 2010. doi:10.1287/opre.1090.0712.

[19] G. Chicco and A. Mazza. Metaheuristic optimization of power and energy systems: Underlying principles and main issues of the 'rush to heuristics'. *Energies*, 13(19), 2020. Article 5097. doi:10.3390/en13195097.

[20] F. Clarke. *Optimization and Nonsmooth Analysis*. Classics Appl. Math. SIAM, 1990. doi:10.1137/1.9781611971309.

[21] C. Coffrin, H. L. Hijazi, and P. Van Hentenryck. The QC Relaxation: A theoretical and computational study on optimal power flow. *IEEE Transactions on Power Systems*, 31(4):3008–3018, 2016. doi:10.1109/TPWRS.2015.2463111.

[22] Commission Regulation (EU). Network code on demand connection DC. *Official Journal*, 18.8.2016:10–54, 2016. URL: http://data.europa.eu/eli/reg/2016/1388/oj.

[23] M. Cordova, W. de Oliveira, and C. Sagastizábal. Revisiting augmented lagrangian duals. *Math. Program.*, 196:235 – 277, 2022. doi:10.1007/s10107-021-01703-5.

[24] R. Correa and C. Lemaréchal. Convergence of some algorithms for convex minimization. *Math. Program.*, 62(2):261–275, 1993. doi:10.1007/BF01585170.

[25] E. Dall'Anese, K. Baker, and T. Summers. Chance-constrained AC optimal power flow for distribution systems with renewables. *IEEE Transactions on Power Systems*, 32(5):3427–3438, 2017. doi:10.1109/TPWRS.2017.2656080.

[26] E. Dall'Anese, S. V. Dhople, B. B. Johnson, and G. B. Giannakis. Decentralized optimal dispatch of photovoltaic inverters in residential distribution systems. *IEEE Transactions on Energy Conversion*, 29(4):957–967, 2014. doi:10.1109/TEC.2014.2357997.

[27] W. de Oliveira and J. C. Souza. A progressive decoupling algorithm for minimizing the difference of convex and weakly convex functions over a linear subspace, 2024. To appear in Journal of Optimization Theory and Applications. URL: https://optimization-online.org/?p=26953.

[28] W. de Oliveira. The ABC of DC programming. *Set-Valued Var. Anal.*, 28(4):679–706, 2020. doi:10.1007/s11228-020-00566-w.

[29] W. de Oliveira. Risk-averse stochastic programming and distributionally robust optimization via operator splitting. *Set-Valued and Variational Analysis*, 29:861–891, 2021. doi:10.1007/s11228-021-00600-5.

[30] D. Dentcheva, A. Prékopa, and A. Ruszczynski. Concavity and efficient points of discrete distributions in probabilistic programming. *Mathematical Programming*, 89:55–77, 2000. doi:10.1007/PL00011393.

[31] D. Drusvyatskiy and C. Paquette. Efficiency of minimizing compositions of convex functions and smooth maps. *Math. Program.*, 178(1):503–558, 2019. doi:10.1007/s10107-018-1311-3.

[32] Enedis. 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 (*in French*), 2017. URL: https://www.enedis.fr/media/2061/download.

[33] Enedis. Principes d'étude et de développement du réseau pour le raccordement des clients consommateurs et producteurs BT (*in French*), 2019. URL: https://www.enedis.fr/media/2168/download.

[34] Enedis. Description et étude des protections de découplage pour le raccordement des installations de production raccordées au réseau public de distribution (*in French*), 2022. URL: https://www.enedis.fr/media/2162/download.

[35] Enedis. Procédure de traitement des demandes de raccordement d'une Installation de Production en BT de puissance supérieure à 36 kVA et en HTA, au Réseau Public de Distribution géré par Enedis (*in French*), 2022. URL: https://www.enedis.fr/media/2173/download.

[36] Enedis. Plan de développement de réseau (*in French*), 2023. URL: https://www.enedis.fr/sites/default/files/documents/pdf/plan-de-developpement-de-reseau-document-preliminaire-2023.pdf.

[37] Enedis Open Data, Accessed: 2021-11-29. URL: https://data.enedis.fr/explore/dataset/coefficients-des-profils/table/?disjunctive.sous_profil&refine.horodate=2020.

[38] T. Erseghe. Distributed Optimal Power Flow Using ADMM. *IEEE Transactions on Power Systems*, 29(5):2370–2380, 2014. doi:10.1109/TPWRS.2014.2306495.

[39] S. Fattaheian-Dehkordi, M. Tavakkoli, A. Abbaspour, M. Fotuhi-Firuzabad, and M. Lehtonen. An incentive-based mechanism to alleviate active power congestion in a multi-agent distribution system. *IEEE Transactions on Smart Grid*, 12(3):1978–1988, 2020. doi:10.1109/TSG.2020.3037560.

[40] Federal Republic of Germany. Renewable Energy Sources Act (EEG 2017), 2017. URL: www.bmwk.de/Redaktion/EN/Downloads/renewable-energy-sources-act-2017.html.

[41] Y. Feng. *Solving for the low-voltage/large-angle power-flow solutions by using the holomorphic embedding method*. PhD thesis, Arizona State University, 2015. URL: https://hdl.handle.net/2286/R.I.34817.

[42] M. Fortin and R. Glowinski. Augmented lagrangian methods : applications to the numerical solution of boundary-value problems. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 65(12):622–622, 1985. Translation from French. doi:10.1002/zamm.19850651211.

[43] S. Frank, I. Steponavice, and S. Rebennack. Optimal power flow: a bibliographic survey I. *Energy Systems*, 3:221–258, 2012. doi:10.1007/S12667-012-0056-Y.

[44] S. Frank, I. Steponavice, and S. Rebennack. Optimal power flow: a bibliographic survey II. non-deterministic and hybrid methods. *Energy Systems*, 3:259–289, 2012. doi:10.1007/S12667-012-0057-X.

[45] French Republic. Arrêté du 12 juillet 2021 d'application de l'article D. 342-23 du Code de l'énergie (*in French*). *Journal Officiel*, 0162, 2021. URL: https://www.legifrance.gouv.fr/eli/arrete/2021/7/12/TRER2116996A/jo/texte.

[46] French Republic. Article L111-91 du Code de l'énergie (*in French*). *Journal Officiel*, 0060, 2023. URL: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000047302072.

[47] L. Gan, N. Li, U. Topcu, and S. H. Low. Exact convex relaxation of optimal power flow in radial networks. *IEEE Transactions on Automatic Control*, 60(1):72–87, 2015. doi:10.1109/TAC.2014.2332712.

[48] L. Gan and S. H. Low. Convex relaxations and linear approximation for optimal power flow in multiphase radial networks. In *2014 Power Systems Computation Conference*, pages 1–9. IEEE, 2014. doi:10.1109/PSCC.2014.7038399.

[49] M. Gaudioso, G. Giallombardo, G. Miglionico, and A. M. Bagirov. Minimizing nonsmooth DC functions via successive DC piecewise-affine approximations. *J. Global Optim.*, 71(1):37–55, 2018. doi:10.1007/s10898-017-0568-z.

[50] A. Genz. Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist.*, 1(2):141–149, 1992. doi:10.2307/1390838.

[51] J. Gorski, F. Pfeuffer, and K. Klamroth. Biconvex sets and optimization with biconvex functions: a survey and extensions. *Mathematical Methods of Operations Research*, 66:373–407, 2007. doi:10.1007/s00186-007-0161-1.

[52] C. Gotzes, H. Heitsch, R. Henrion, and R. Schultz. On the quantification of nomination feasibility in stationary gas networks with random load. *Mathematical Methods of Operations Research*, 84:427 – 457, 2016. doi:10.1007/s00186-016-0564-y.

[53] J. Grainger and W. Stevenson. *Power System Analysis*. McGraw-Hill Education, 1994.

[54] T. Grandón, H. Heitsch, and R. Henrion. A joint model of probabilistic/robust constraints for gas transport management in stationary networks. *Computational Management Science*, 14:1–18, 2017. doi:10.1007/s10287-017-0284-7.

[55] L. Halilbašić, P. Pinson, and S. Chatzivasileiadis. Convex Relaxations and Approximations of Chance-Constrained AC-OPF Problems. *IEEE Transactions on Power Systems*, 34(2):1459–1470, 2019. doi:10.1109/TPWRS.2018.2874072.

[56] P. Hartman. On functions representable as a difference of convex functions. *Pacific J. Math.*, 9(3):167–198, 1959.

[57] F. Hasan, A. Kargarian, and A. Mohammadi. A survey on applications of machine learning for optimal power flow. In *2020 IEEE Texas Power and Energy Conference (TPEC)*, pages 1–6. IEEE, 2020. doi:10.1109/TPEC48276.2020.9042547.

[58] H. Heitsch. On probabilistic capacity maximization in a stationary gas network. *Optimization*, 69(3):575–604, 2020. doi:10.1080/02331934.2019.1625353.

[59] L. Hellemo, P. I. Barton, and A. Tomasgard. Decision-dependent probabilities in stochastic programs with recourse. *Comput. Manag. Sci.*, 15(3):369–395, 2018. doi:10.1007/s10287-018-0330-0.

[60] R. Henrion and A. Möller. A gradient formula for linear chance constraints under Gaussian distribution. *Math. Oper. Res.*, 37:475–488, 2012. doi:10.1287/moor.1120.0544.

[61] N. Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and Its Applications*, 103:103–118., 1988. doi:10.1016/0024-3795(88)90223-6.

[62] L. J. Hong, Z. Hu, and L. Zhang. Conditional Value-at-Risk Approximation to Value-at-Risk Constrained Programs: A Remedy via Monte Carlo. *INFORMS Journal on Computing*, 26(2):385–400, 2014. doi:10.1287/ijoc.2013.0572.

[63] L. Hong, Y. Yang, and L. Zhang. Sequential convex approximations to joint chance constrained programed: A Monte Carlo approach. *Operations Research*, 3(59):617–630, 2011. doi:10.1287/opre.1100.0910.

[64] R. Jabr. Radial distribution load flow using conic programming. *IEEE Transactions on Power Systems*, 21(3):1458–1459, 2006. doi:10.1109/TPWRS.2006.879234.

[65] R. A. Jabr. Adjustable robust OPF with renewable energy sources. *IEEE Transactions on Power Systems*, 28(4):4742–4751, 2013. doi:10.1109/TPWRS.2013.2275013.

[66] F. Jara-Moroni, J.-S. Pang, and A. Wächter. A study of the difference-of-convex approach for solving linear programs with complementarity constraints. *Math. Program.*, 169(1):221–254, 2018. doi:10.1007/s10107-017-1208-6.

[67] P. Javal. *Integrating uncertainties in short-term operational planning*. PhD thesis, Mines Paris - PSL, 2021. URL: https://pastel.archives-ouvertes.fr/tel-03693993v2/document.

[68] M. Jia, G. Hug, and C. Shen. Iterative Decomposition of Joint Chance Constraints in OPF. *IEEE Transactions on Power Systems*, 36(5):4836–4839, 2021. doi:10.1109/TPWRS.2021.3072541.

[69] E. Karangelos and L. Wehenkel. Towards leveraging discrete grid flexibility in chance-constrained power system operation planning. *Electric Power Systems Research*, 188:Article 106571, 2020. doi:10.1016/j.epsr.2020.106571.

[70] W. Khalaf, A. Astorino, P. d'Alessandro, and M. Gaudioso. A DC optimization-based clustering technique for edge detection. *Optim. Lett.*, 11(3):627–640, 2017. doi:10.1007/s11590-016-1031-7.

[71] J. Lavaei and S. Low. Zero duality gap in optimal power flow problem. *IEEE Transactions on Power Systems*, 27(1):92–107, 2012. doi:10.1109/TPWRS.2011.2160974.

[72] J. Lavaei, D. Tse, and B. Zhang. Geometry of power flows and optimization in distribution networks. *IEEE Transactions on Power Systems*, 29(2):572–583, 2014. doi:10.1109/TPWRS.2013.2282086.

[73] H. A. Le Thi, H. V. Ngai, and P. D. Tao. DC programming and DCA for general DC programs. In T. Van Do, H. A. Le Thi, and N. T. Nguyen, editors, *Advanced Computational Methods for Knowledge Engineering: Proceedings of the 2nd International Conference on Computer Science, Applied Mathematics and Applications (ICCSAMA 2014)*, volume 282, pages 15–35. Springer, Cham, 2014. doi:10.1007/978-3-319-06569-4_2.

[74] H. A. Le Thi and T. Pham Dinh. DC programming and DCA: thirty years of developments. *Mathematical Programming*, 169(1):5–68, 2018. doi:10.1007/s10107-018-1235-y.

[75] H. A. Le Thi, T. Pham Dinh, and H. V. Ngai. Exact penalty and error bounds in DC programming. *J. Global Optim.*, 52(3):509–535, 2012. doi:10.1007/s10898-011-9765-3.

[76] D. Lee, H. D. Nguyen, K. Dvijotham, and K. Turitsyn. Convex restriction of power flow feasibility sets. *IEEE Transactions on Control of Network Systems*, 6(3):1235–1245, 2019. doi:10.1109/TCNS.2019.2930896.

[77] M. D. Leiren and I. Reimer. Germany: From feed-in tariffs to greater competition. In E. L. Boasson, M. D. Leiren, and J. Wettestad, editors, *Comparative Renewables Policy*, pages 75–102. Routledge, London, 2020.

[78] C. Lemaréchal and A. Renaud. A geometric study of duality gaps, with applications. *Math. Program.*, 90:399–427, 2001. doi:10.1007/PL00011429.

[79] H. Li and Y. Cui. A decomposition algorithm for two-stage stochastic programs with nonconvex recourse functions. *SIAM Journal on Optimization*, 34(1):306–335, 2024. doi:10.1137/22M1488533.

[80] W. Li, Y. Liu, H. Liang, Y. Man, and F. Li. Distributed tracking-ADMM approach for chance-constrained energy management with stochastic wind power in smart grid. *CSEE Journal of Power and Energy Systems*, pages 1–11, 2021. (Early Access). doi:10.17775/CSEEJPES.2021.00540.

[81] T. Lipp and S. Boyd. Variations and extension of the convex–concave procedure. *Optim. Eng.*, 17(2):263–287, 2016. doi:10.1007/s11081-015-9294-x.

[82] S. H. Low. Convex relaxation of optimal power flow—part I: Formulations and equivalence. *IEEE Transactions on Control of Network Systems*, 1(1):15–27, 2014. doi:10.1109/TCNS.2014.2309732.

[83] M. Lubin, Y. Dvorkin, and L. Roald. Chance constraints for improving the security of ac optimal power flow. *IEEE Transactions on Power Systems*, 34(3):1908–1917, 2019. doi:10.1109/TPWRS.2018.2890732.

[84] M. Lubin, Y. Dvorkin, and S. Backhaus. A robust approach to chance constrained optimal power flow with renewable generation. *IEEE Transactions on Power Systems*, 31(5):3840–3849, 2016. doi:10.1109/TPWRS.2015.2499753.

[85] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM J. Optim.*, 19:674–699, 2008. doi:10.1137/070702928.

[86] T. W. K. Mak, M. Chatzos, M. Tanneau, and P. V. Hentenryck. Learning Regionally Decentralized AC Optimal Power Flows With ADMM. *IEEE Transactions on Smart Grid*, 14(6):4863–4876, 2023. doi:10.1109/TSG.2023.3251292.

[87] K. Margellos, P. Goulart, and J. Lygeros. On the road between robust optimization and the scenario approach for chance constrained optimization problems. *IEEE Transactions on Automatic Control*, 59(8):2258–2263, 2014. doi:10.1109/TAC.2014.2303232.

[88] H. Mine and M. Fukushima. A minimization method for the sum of a convex function and a continuously differentiable function. *J. Optim. Theory Appl.*, 33(1):9–23, 1981. doi:10.1007/BF00935173.

[89] D. K. Molzahn, J. T. Holzer, B. C. Lesieutre, and C. L. DeMarco. Implementation of a large-scale optimal power flow solver based on semidefinite programming. *IEEE Transactions on Power Systems*, 28(4):3987–3998, 2013. doi:10.1109/TPWRS.2013.2258044.

[90] O. Montonen and K. Joki. Bundle-based descent method for nonsmooth multiobjective DC optimization with inequality constraints. *J. Global Optim.*, 72(3):403–429, 2018. doi:10.1007/s10898-018-0651-0.

[91] H. Morais, C. Paris, O. Carré, M. Carlier, and B. Bouzigon. Levers optimization in short-term operational planning for real distribution systems. In *CIRED 2019 Conference*, 2019. URL: http://dx.doi.org/10.34890/281.

[92] B. S. Mordukhovich. *Variational Analysis and Applications*. Springer Monogr. Math. Springer, Cham, 2018. doi:10.1007/978-3-319-92775-6.

[93] B. S. Mordukhovich and N. M. Nam. *An Easy Path to Convex Analysis and Applications*. Synth. Lect. Math. Stat. Springer Cham, 2014. doi:10.1007/978-3-031-02406-1.

[94] T. Mühlpfordt, T. Faulwasser, and V. Hagenmeyer. A generalized framework for chance-constrained optimal power flow. *Sustainable Energy, Grids and Networks*, 16:231–242, 2018. doi:10.1016/j.segan.2018.08.002.

[95] G. Nemhauser and L. Wolsey. *Integer and Combinatorial Optimization*. John Wiley & Sons, 1999. doi:10.1002/9781118627372.

[96] M. Niu, C. Wan, and Z. Xu. A review on applications of heuristic optimization algorithms for optimal power flow in modern power systems. *Journal of Modern Power Systems and Clean Energy*, 2:289–297, 2014. doi:10.1007/s40565-014-0089-4.

[97] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications*, 142(2):399–416, 2009. doi:10.1007/s10957-009-9523-6.

[98] J. S. Pang, M. Razaviyayn, and A. Alvarado. Computing B-stationary points of nonsmooth DC programs. *Math. Oper. Res.*, 42(1):95–118, 2017. doi:10.1287/moor.2016.0795.

[99] A. Peña-Ordieres, D. K. Molzahn, L. A. Roald, and A. Wächter. DC Optimal Power Flow With Joint Chance Constraints. *IEEE Transactions on Power Systems*, 36(1):147–158, 2021. doi:10.1109/TPWRS.2020.3004023.

[100] B. Qi, K. N. Hasan, and J. V. Milanović. Identification of critical parameters affecting voltage and angular stability considering load-renewable generation correlations. *IEEE Transactions on Power Systems*, 34(4):2859–2869, 2019. doi:10.1109/TPWRS.2019.2891840.

[101] Z. Qiu, G. Deconinck, and R. Belmans. A literature survey of optimal power flow problems in the electricity market context. In *2009 IEEE/PES Power Systems Conference and Exposition*, pages 1–6. IEEE, 2009. doi:10.1109/PSCE.2009.4840099.

[102] L. Roald and G. Andersson. Chance-constrained AC optimal power flow: Reformulations and efficient algorithms. *IEEE Transactions on Power Systems*, 33(3):2906–2918, 2018. doi:10.1109/TPWRS.2017.2745410.

[103] L. Roald, F. Oldewurtel, T. Krause, and G. Andersson. Analytical reformulation of security constrained optimal power flow with probabilistic constraints. In *2013 IEEE Grenoble Conference*, pages 1–6. IEEE, 2013. doi:10.1109/PTC.2013.6652224.

[104] L. Roald, F. Oldewurtel, B. V. Parys, and G. Andersson. Security constrained optimal power flow with distributionally robust chance constraints, 2015. URL: https://arxiv.org/abs/1508.06061.

[105] R. T. Rockafellar. Generalized subgradients in mathematical programming. In *Mathematical Programming The State of the Art: Bonn 1982*, pages 368–390, Berlin, 1983. Springer. doi:10.1007/978-3-642-68874-4_15.

[106] R. Rockafellar. Favorable Classes of Lipschitz Continuous Functions in Subgradient Optimization. In *Progress in Nondifferentiable Optimization*, IIASA Collaborative Proceedings Series, International Institute of Applied Systems Analysis, Laxenburg, Austria, pages 125–144, 1982.

[107] R. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren Math. Wiss.* Springer Verlag Berlin, 3rd edition, 2009. doi:10.1007/978-3-642-02431-3.

[108] M. Roland, A. Forel, and T. Vidal. Adaptive Partitioning for Chance-Constrained Problems with Finite Support, 2024. URL: https://arxiv.org/abs/2312.13180, arXiv:2312.13180.

[109] T. Rotaru, P. Patrinos, and F. Glineur. Improved convergence rates for the difference-of-convex algorithm, 2024. URL: https://arxiv.org/abs/2403.16864.

[110] C. Sagastizábal and M. Solodov. An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter. *SIAM J. Optim.*, 16(1):146–169, 2005. doi:10.1137/040603875.

[111] G. Salinetti and R. J. Wets. On the relations between two types of convergence for convex functions. *J. Math. Anal. Appl.*, 60(1):211–226, 1977. doi:10.1016/0022-247X(77)90060-9.

[112] G. M. Sempere, W. de Oliveira, and J. O. Royset. A proximal-type method for nonsmooth and nonconvex constrained minimization problems. *Journal of Optimization Theory and Applications*, 204:Article 54, 2025. doi:10.1007/s10957-024-02597-x.

[113] A. Shapiro and Y. Yomdin. On functions representable as a difference of two convex functions, and necessary conditions in a constrainted optimization. Technical report, Ben-Gurion University of the Negev, 1981. URL: https://sites.gatech.edu/alexander-shapiro/publications.

[114] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. MOS-SIAM Series on Optimization. SIAM, Philadelphia, 2009. doi:10.1137/1.9781611976595.

[115] T. Soares and R. J. Bessa. Proactive management of distribution grids with chance-constrained linearized AC OPF. *International Journal of Electrical Power & Energy Systems*, 109:332–342, 2019. doi:10.1016/j.ijepes.2019.02.002.

[116] T. Soares, R. J. Bessa, P. Pinson, and H. Morais. Active Distribution Grid Management Based on Robust AC Optimal Power Flow. *IEEE Transactions on Smart Grid*, 9(6):6229–6241, 2018. doi:10.1109/TSG.2017.2707065.

[117] T. Sousa, H. Morais, Z. Vale, P. Faria, and J. Soares. Intelligent energy resource management considering vehicle-to-grid: A simulated annealing approach. *IEEE Transactions on Smart Grid*, 3(1):535–542, 2012. doi:10.1109/TSG.2011.2165303.

[118] A. S. Strekalovsky and I. M. Minarchenko. A local search method for optimisation problem with d.c. inequality constraints. *Appl. Math. Model.* , 1(58):229–244, 2018. doi:10.1016/j.apm.2017.07.031.

[119] K. Sun and X. A. Sun. A Two-Level ADMM Algorithm for AC OPF With Global Convergence Guarantees. *IEEE Transactions on Power Systems*, 36(6):5271–5281, 2021. doi:10.1109/TPWRS.2021.3073116.

[120] K. Sun and X. A. Sun. Algorithms for Difference-of-Convex Programs Based on Difference-of-Moreau-Envelopes Smoothing. *INFORMS J. Optim.*, 5(4):321–339, 2023. doi:10.1287/ijoo.2022.0087.

[121] B. Swaminathan. *Operational Planning of Active Distribution Networks - Convex Relaxation under Uncertainty*. PhD thesis, Université Grenoble Alpes, 2017. URL: https://theses.hal.science/tel-01690509.

[122] B. Swaminathan, A. Matthieu, and B. Benoit. OptimGP - Industrial Optimisation for Enedis' Short-term Operational Planning. In *CIRED 2021 - The 26th International Conference and Exhibition on Electricity Distribution*, pages 1161–1165, 2021. doi:10.1049/icp.2021.1956.

[123] K. Syrtseva, W. de Oliveira, S. Demassey, H. Morais, P. Javal, and B. Swaminathan. Difference-of-Convex approach to chance-constrained Optimal Power Flow modelling the DSO power modulation lever for distribution networks. *Sustainable Energy, Grids and Networks*, 36:Article 101168, 2023. doi:10.1016/j.segan.2023.101168.

[124] K. Syrtseva, W. de Oliveira, S. Demassey, and W. van Ackooij. Minimizing the difference of convex and weakly convex functions via bundle method. *Pacific Journal of Optimization*, 20(4):699–741, 2024. doi:10.61208/pjo-2024-004.

[125] W. van Ackooij. Decomposition approaches for block-structured chance-constrained programs with application to hydro-thermal unit commitment. *Math. Methods Oper. Res.*, 80(3):227–253, 2014. doi:10.1007/s00186-014-0478-5.

[126] W. van Ackooij. A discussion of probability functions and constraints from a variational perspective. *Set-Valued Var. Anal.*, 28(4):585–609, 2020. doi:10.1007/s11228-020-00552-2.

[127] W. van Ackooij and W. de Oliveira. Level bundle methods for constrained convex optimization with various oracles. *Comput. Optim. Appl.*, 57(3):555–597, 2014. doi:10.1007/s10589-013-9610-3.

[128] W. van Ackooij and W. de Oliveira. Addendum to the paper 'nonsmooth DC-constrained optimization: constraint qualification and minimizing methodologies'. *Optim. Methods Softw.*, 37(6):2241–2250, 2022. doi:10.1080/10556788.2022.2063861.

[129] W. van Ackooij, S. Demassey, P. Javal, H. Morais, W. de Oliveira, and B. Swaminathan. A bundle method for nonsmooth DC programming with application to chance-constrained problems. *Comput. Optim. Appl.*, 78(2):451–490, 2021. doi:10.1007/s10589-020-00241-8.

[130] W. van Ackooij, A. Frangioni, and W. de Oliveira. Inexact stabilized Benders' decomposition approaches: with application to chance-constrained problems with finite support. *Comput. Optim. Appl.*, 65(3):637–669, 2016. doi:10.1007/s10589-016-9851-z.

[131] W. van Ackooij and R. Henrion. (Sub-) Gradient formulae for probability functions of random inequality systems under Gaussian distribution. *SIAM/ASA J. Uncertain. Quantif.* , 5(1):63–87, 2017. doi:10.1137/16M1061308.

[132] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati. On probabilistic constraints induced by rectangular sets and multivariate normal distributions. *Math. Methods Oper. Res.*, 71(3):535–549, 2010. doi:10.1007/s00186-010-0316-3.

[133] W. van Ackooij and J. Malick. Decomposition algorithm for large-scale two-stage unit-commitment. *Ann. Oper. Res.*, 238(1):587–613, 2016. doi:10.1007/s10479-015-2029-8.

[134] W. van Ackooij and J. Malick. Second-order differentiability of probability functions. *Optim. Lett.*, 11(1):179–194, 2017. doi:10.1007/s11590-016-1015-7.

[135] W. van Ackooij and N. Oudjane. On supply and network investment in power systems. *4OR*, 22:465–481, 2024. doi:10.1007/s10288-024-00566-8.

[136] W. van Ackooij and P. Pérez-Aros. Gradient formulae for probability functions depending on a heterogenous family of constraints. *Open J. Math. Optim.*, 2:1–29, 2021. doi:10.5802/ojmo.9.

[137] W. van Ackooij, P. Pérez-Aros, and C. Soto. Probability functions generated by set-valued mappings: A study of first order information. *Set-Valued and Variational Analysis*, 32(6), 2024. doi:10.1007/s11228-024-00709-3.

[138] W. van Ackooij and C. Sagastizábal. Constrained bundle methods for upper inexact oracles with application to joint chance constrained energy problems. *SIAM J. Optim.*, 24(2):733–765, 2014. doi:10.1137/120903099.

[139] W. van Ackooij, F. Atenas, and C. Sagastizábal. Weak convexity and approximate subdifferentials. *Journal of Optimization Theory and Applications*, 203:1686–1709, 2024. doi:10.1007/s10957-024-02551-x.

[140] W. van Ackooij and W. de Oliveira. Nonsmooth and nonconvex optimization via approximate difference-of-convex decompositions. *J. Optim. Theory Appl.*, 182(1):49–80, 2019. doi:10.1007/s10957-019-01500-3.

[141] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati. Joint chance constrained programming for hydro reservoir management. *Optimization and Engineering*, 15:509–531, 2014. doi:10.1007/s11081-013-9236-4.

[142] W. van Ackooij, R. Zorgati, R. Henrion, and A. Möller. Chance constrained programming and its applications to energy management. In I. Dritsas, editor, *Stochastic Optimization*, chapter 13. IntechOpen, 2011. doi:10.5772/15438.

[143] A. Venzke, L. Halilbasic, U. Markovic, G. Hug, and S. Chatzivasileiadis. Convex Relaxations of Chance Constrained AC Optimal Power Flow. In *2018 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2018. doi:10.1109/PESGM.2018.8586626.

[144] R. A. Verzijlbergh, L. J. De Vries, and Z. Lukszo. Renewable Energy Sources and Responsive Demand. Do We Need Congestion Management in the Distribution Grid? *IEEE Transactions on Power Systems*, 29(5):2119–2128, 2014. doi:10.1109/TPWRS.2014.2300941.

[145] J.-P. Vial. Strong and weak convexity of sets and functions. *Math. Oper. Res.*, 8(2):231–259, 1983. doi:10.1287/moor.8.2.231.

[146] M. Vrakopoulou, M. Katsampani, K. Margellos, J. Lygeros, and G. Andersson. Probabilistic security-constrained AC optimal power flow. In *2013 IEEE Grenoble Conference*, pages 1–6. IEEE, 2013. doi:10.1109/PTC.2013.6652374.

[147] Y. Wang, L. Wu, and S. Wang. A Fully-Decentralized Consensus-Based ADMM Approach for DC-OPF With Demand Response. *IEEE Transactions on Smart Grid*, 8(6):2637–2647, 2017. doi:10.1109/TSG.2016.2532467.

[148] Y. Wang, W. Yin, and J. Zeng. Global Convergence of ADMM in Nonconvex Nonsmooth Optimization. *Journal of Scientific Computing*, 78:29 – 63, 2015. doi:10.1007/s10915-018-0757-z.

[149] T. Weisser, L. A. Roald, and S. Misra. Chance-constrained optimization for nonlinear network flow problems, 2018. URL: https://arxiv.org/abs/1803.02696, arXiv:1803.02696.

[150] Y. Xu, J. Ma, Z. Y. Dong, and D. J. Hill. Robust transient stability-constrained optimal power flow with uncertain dynamic loads. *IEEE Transactions on Smart Grid*, 8(4):1911–1921, 2017. doi:10.1109/TSG.2015.2510447.

[151] Y. Yi and G. Verbič. Fair operating envelopes under uncertainty using chance constrained optimal power flow. *Electric Power Systems Research*, 213:Article 108465, 2022. doi:10.1016/j.epsr.2022.108465.

[152] P. Yu, T. K. Pong, and Z. Lu. Convergence rate analysis of a sequential convex programming method with line search for a class of constrained difference-of-convex optimization problems. *SIAM J. Optim.*, 31(3):2024–2054, 2021. doi:10.1137/20M1314057.

[153] B. Zhang, A. Y. Lam, A. D. Domínguez-García, and D. Tse. An optimal and distributed method for voltage regulation in power distribution systems. *IEEE Transactions on Power Systems*, 30(4):1714–1726, 2015. doi:10.1109/TPWRS.2014.2347281.

[154] H. Zhang and P. Li. Chance constrained programming for optimal power flow under uncertainty. *IEEE Transactions on Power Systems*, 26(4):2417–2424, 2011. doi:10.1109/TPWRS.2011.2154367.

[155] G. Zhao, P. Zhou, and W. Wen. Feed-in tariffs, knowledge stocks and renewable energy technology innovation: The role of local government intervention. *Energy Policy*, 156:Article 112453, 2021. doi:10.1016/j.enpol.2021.112453.

RÉSUMÉ

L'expansion des sources d'énergie renouvelable accroît le degré d'incertitude dans l'exploitation des réseaux de distribution d'électricité. La variabilité et l'intermittence inhérentes à ces énergies posent aussi d'importants défis aux gestionnaires de réseaux au niveau opérationnel. La gestion prévisionnelle doit ainsi évoluer pour intégrer des leviers de flexibilité, telles la modulation de puissance active et la gestion de puissance réactive. La décision relative à l'activation de ces leviers se traduit par un problème d'Optimal Power Flow. Cette thèse développe des algorithmes de résolution pour deux modèles stochastiques en courant alternatif (AC-OPF). Ces problèmes d'optimisation sont, à la fois, non-convexes, non-lisses et discrets. Cette thèse vise à appréhender ces complexités, sans recourir à la convexification des équations de flux de puissance, et en considérant l'interdépendance des incertitudes, via une contrainte probabiliste jointe ou une décomposition par scénarios dans le cas de leviers discrets.

Précisément, la première méthodologie proposée s'applique à une version continue de l'AC-OPF sous contrainte probabiliste jointe. Une contribution de ce travail porte sur la conception d'une procédure numérique (oracle) traitant la contrainte probabiliste comme la différence de deux fonctions convexes. L'oracle est alors associé à une méthode de faisceaux pour les problèmes DoC (différence de convexes). Une seconde contribution porte sur le développement d'un nouvel algorithme de faisceaux offrant des garanties de convergence plus fortes sous des hypothèses plus faibles. Il produit ainsi un point critique (satisfaisant des conditions KKT généralisées) de l'AC-OPF probabiliste. Basé sur la méthode DoC précédente, cet algorithme exploite un programme maître différent, ainsi qu'une règle originale de mise à jour du paramètre proximal. Il s'applique à la classe générale des problèmes d'optimisation non-convexes et non-lisses dont objectif et contraintes sont modélisables comme différence de fonctions convexes et faiblement convexes (CwC). L'évaluation empirique de l'algorithme est menée sur différents problèmes non-convexes et stochastiques. Ses performances pratiques sont comparées à celles de la méthode DoC sur un cas d'étude de l'AC-OPF probabiliste dans un réseau de distribution à 33 nœuds.

La seconde méthodologie proposée considère des règles discrètes en gestion prévisionnelle, telles que des règles de priorité et d'équité pour la modulation de puissance.

L'expérimentation montre les limites de la méthode des faisceaux pour intégrer des variables entières. Comme alternative, il est proposé un modèle d'optimisation attachant une variable binaire par scénario, et maximisant le nombre de scénarios réalisés dans un budget limité. La dualisation des contraintes couplantes et la coordination par blocs permettent de séparer les règles discrètes de l'AC-OPF stochastique, qui se décompose, à son tour, en AC-OPF déterministes individuels par scénario. Si la convergence théorique n'est plus garantie par cette séparation, la pertinence pratique de l'approche est illustrée numériquement.

MOTS CLÉS

Gestion opérationnelle, OPF sous incertitudes jointes, OPF discret
Optimisation stochastique, Différence de convexes, Faiblement convexe

ABSTRACT

The expansion of renewable energy sources (RES) leads to the growth of uncertainty in the power distribution network operation. The inherent variability and intermittency of RES present significant challenges to the efficient and reliable operation of power systems. To address these challenges, operational planning performed by distribution system operators should evolve, in particular, to allow the efficient utilization of different flexibility levers, such as active power modulation and reactive power management. Decisions on lever activation are based on the resolution of an alternating current optimal power flow problem (AC-OPF). This thesis develops algorithms for handling two stochastic AC-OPF models. These optimization problems are simultaneously nonconvex, nonsmooth, and discrete. The thesis aims to grasp these complexities accurately, by addressing the AC power flow equations without relying on convexification and by handling interdependent uncertainties either through a joint probability constraint or via scenario decomposition to cope with the discrete levers.

More specifically, the first proposed methodology addresses a continuous version of the joint chance-constrained AC-OPF. A first contribution of this work is the design of a numerical procedure (oracle) that enables the representation of the probability constraint as a difference of two convex functions. This step is followed by applying a known Difference-of-Convex (DoC) bundle method to the resulting continuous optimization problem. A second contribution concerns a new bundle algorithm with stronger convergence guarantees under weaker assumptions. For the chance-constrained AC-OPF, this algorithm provides a critical (generalized KKT) point. The work builds upon the employed DoC bundle and proposes a different master program and an original rule to update proximal parameter. The algorithm is capable of handling a broad class of nonsmooth and nonconvex optimization problems beyond the stochastic AC-OPF framework, provided the objective and constraint functions can be represented as differences of convex and weakly convex (CwC) functions. The practical performance of the algorithm is illustrated through numerical experiments on some nonconvex stochastic problems and is compared to the DoC bundle method for the chance-constrained AC-OPF in a 33-bus distribution network.

The second proposed methodology addresses operational planning rules for power modulation and curtailment, like priority and fairness, which result in logical and discrete formulations. The numerical results demonstrate the limitations of the bundle method for integrating integer variables. As an alternative, an optimization model is proposed that assigns a binary variable to each scenario and maximizes the number of satisfied scenarios within a limited budget. Applying penalization and block coordination allows separating those discrete considerations from the stochastic AC-OPF component, which is then decomposed into an individual deterministic AC-OPF for each scenario. Although it lacks theoretical convergence guarantees, the relevance of this approach is validated in practice.

KEYWORDS

Operational planning, Chance-constrained OPF, Discrete OPF, Stochastic optimization, Difference-of-convex, Weakly convex