Advanced optimization methods for power systems

P. Panciatici, M.C. Campi, S. Garatti, S.H. Low, D.K. Molzahn, A.X. Sun, L. Wehenkel

Abstract—**Power system planning and operation offers multitudinous opportunities for optimization methods. In practice, these problems are generally large-scale, non-linear, subject to uncertainties, and combine both continuous and discrete variables. In the recent years, a number of complementary theoretical advances in addressing such problems have been obtained in the field of applied mathematics. The paper introduces a selection of these advances in the fields of non-convex optimization, in mixed-integer programming, and in optimization under uncertainty. The practical relevance of these developments for power systems planning and operation are discussed, and the opportunities for combining them, together with high-performance computing and big data infrastructures, as well as novel machine learning and randomized algorithms, are highlighted.**

I. Needs from a grid operator perspective

The electrical grids and their management become more and more complex. This state of affairs has different causes that will not disappear in the near future.

In Europe, the first reason is the massive integration of renewable but generally intermittent generation in the system. Power flows in the grid are created by differences in the location of sinks and sources. With a significant amount of intermittent generation, the predictability of the sources (location and amount of power injections) decreases and affects the predictability of the flows. Furthermore, some of these new power plants could be small units (e.g. PV) connected to the distribution grid, changing the distribution grid into an active system. Moreover, Transmission System Operators (TSOs) have a poor observability of these power injections and have no control at all over them. Another factor is the inconsistency between the relatively short time to build new wind farms (2 or 3 years) and the time to go through all administrative procedures to build new lines (more than 5 years everywhere in Europe). In Europe, the best locations for wind farms are mostly along the coasts and offshore, while for photo-voltaic generation they are in the south of Europe. Since these locations do not generally match those of the large load

centers, a transmission network is required and this network will have to cope with the variability of the flows induced by the stochastic nature of the novel generation subsystems.

The second main reason is that it is more difficult than ever to build new overhead lines because of low public acceptance and "Not In My BackYard" (NIMBY) attitude. People are more and more afraid of hypothetical electromagnetic effects or just don't like to see big towers in the landscape and in particular in protected areas which are more and more numerous around Europe. It is very difficult to explain the need for new power lines to people who already have access to electricity at a reasonable price and with high reliability. An increase in the European Social Welfare with a positive feedback for the European economy and hopefully for all European citizens is a concept that is too theoretical compared to the negative local impact. Alternative solutions are technically complex, costly and need more time to be deployed.

The third reason is linked to the setup of electricity markets crossing the administrative and historical borders. Generators, retailers and consumers view the transmission system as a public resource to which they should have unlimited access. This approach has the desirable effect of pushing the system towards a maximization of the social welfare and an optimal utilization of the assets. However, this optimization is constrained by security considerations because wide-spread service interruptions spanning over long periods of time are unacceptable in our modern societies due to their huge economic and social costs. Since TSOs are responsible for maintaining the reliability of the electric power system, they must therefore define the operating limits that must be respected. As in any constrained optimization problem, the optimal solution towards which the market evolves tends to be limited by these security constraints. The stakeholders therefore perceive reliability management by the TSOs as constraining their activities and reducing the European Social Welfare rather than as enablers of this large physical market place, as it would be the case if the grid were a copper plate. The transparent definition and the precise assessment of the distance to these limits thus become more and more critical.

The last reason is that the ageing of grid assets needs increasing attention. A significant part of the European grids' assets are more than 50 years old. Asset management, and maintenance in systems that can't be stopped, are extremely challenging and need to be precisely anticipated when large numbers of assets are approaching simultaneously the end of their expected life times.

To maintain the security of the supply in this context, TSOs have to change the architecture of the system by considering the following technologies:

- Long distance HVAC underground cables with large reactive compensators.

P. Panciatici is with R&D Dept. of RTE, Versailles, France; email: patrick.panciatici@rte-france.com

M.C. Campi is with the Dept. of Information Engineering of the University of Brescia, Brescia, Italy; email: marco.campi@unibs.it

S. Garatti is with the Dept. of Electronics, Computer Science and Bioengineering of Politecnico di Milano, Milan, Italy; email: simone.garatti@polimi.it

S.H. Low is with the Engineering and Applied Science Division, California Institute of Technology, Pasadena, USA; email: slow@cms.caltech.edu

D.K. Molzahn is with the Dept. of EE&CS of the University of Michigan, Ann Arbor, USA; email: molzahn@umich.edu

A.X. Sun is with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, USA; email: andy.sun@isye.gatech.edu

L. Wehenkel is with the Dept. of EE&CS of the University of Liège (Institut Montefiore), Belgium; email: L.Wehenkel@ulg.ac.be

- HVDC underground cables in parallel with the AC grid with smart controls of AC/DC converters.
- And, ultimately, HVDC grids, first to connect efficiently offshore wind farms and then to provide cheaper interconnections between distant areas.

Meanwhile, TSOs will try to optimize the existing systems by adding more and more special devices such as Phase Shifting Transformers, Static VAr Compensators and advanced controls and protection schemes, taking also advantage of the flexibility provided by HVDC links embedded in AC grids. At the same time, demand response or dispersed storage could offer new ways to control the system, even if business models and costs are still questionable. But in any case, this flexibility will require a rethinking of historical operating practices where grid operators made the assumption that the load is an uncontrollable exogenous stochastic variable.

We have heard so often in conferences, seminars and workshops, that the power grid will soon be operated very near to its limits, so that this statement has become a cliché. This cliché is now a reality. To be more precise, it is no longer possible to respect the classical preventive $N - 1$ security standards during all hours in a year. The system is indeed no longer able to survive all single faults without post-fault actions, i.e. corrective controls. More and more corrective control strategies are hence elaborated and prepared to maintain the security of the system. The number of hours during which the system requires corrective actions to be secure is increasing, and that seems to be a natural trend associated with the massive integration of intermittent generation. More and more local or centralized Special Protection Schemes (SPS)/Remedial Actions Schemes(RAS) are deployed to implement automatically some of these corrective actions based on advanced measurement devices (Phasor Measurement Units, Dynamic Line Ratings, ...) and high bandwidth communication networks.

Grid operators have to manage an extremely complex decision making process in order to ensure the reliability and quality of supply at minimal cost over different time horizons. For the sake of clarity, while not aiming at being exhaustive, the following problems need to be dressed by the grid operators:

- Long term (10-20 years): planning stage
  - where to build new power lines? which technology? which capacity?
- Mid term (2-5 years):
  - installation of control devices: substation design, var/reactive support, PSTs, replacement of conductors, SPS/RAS design;
  - asset management and maintenance: which equipment to upgrade, to replace, to repair and when?
- Short term (monthly-weekly):
  - outage management, must-run generators, preparation of corrective actions, required margins.
- Real Time (two days ahead to real time):
  - interaction with energy markets: definition of grid capacities;
  - selection of substation's topology, settings of SPS/RAS, adjustment of generating units.

In all these contexts, the grid operators want to make "optimal" decisions over these different time horizons, even if some decision making processes are currently not formalized mathematically as optimization problems but are rather based solely on knowledge of experts. However, as complexity increases, decision support tools become mandatory to help these experts to make their decisions:

- For the long term planning, there is hyper uncertainty associated with the implementation of energy transition policies and long term market behavior (Priority to renewable energies, Demand Growth in context of efficiency promotion, Technology Costs: electrical batteries, Demand Response, EV, Distributed Generation, Carbon Tax, Fuel Costs) and Grid operators have to make robust decisions based on multi-scenario grid planning.
- In all these processes, the increasing level of uncertainty associated to wind and solar power must be taken into account, hence pushing towards the use of probabilistic methods.
- Operation nearest to the limits requires an accurate modeling of all pieces of equipment, of the corrective actions and of the dynamic behaviors, so as to allow an accurate enough assessment of security margins. Moreover, the active constraints could be related to voltage/reactive or stability issues and not only to thermal limits.

Grid operators must ensure an adequate consistency between these decision making processes. They are in fact multistage decision making processes considering all the different time horizons. At the planning stage, they have to consider the decisions which could be made in lower level problems: asset management and operation and the same between asset management and operation. The modeling of these lower level problems seems very challenging when these lower level problems become more complex. Approximations are required and relevant "proxies" must be found for this modeling.

In this paper we defend the idea that in order to address all these different questions, it is valuable to explicitly formulate them as optimization problems. Most of these problems, once stated, are hard to solve exactly. On the other hand significant progress has been collected in the recent years both in computational and in mathematical respects. The goal of the paper is to highlight the main avenues of progress in these respects and explain how they can be leveraged for improving power systems management.

The rest of the paper is organized as follows. In the next section, we present a taxonomy of optimization problems associated with the practical needs of power systems. In Section 3, we focus on the main recent progresses in optimization and their impacts on solving power system management problems. In Section 4, we discuss how these progresses could/should be combined with progresses in sister fields, such as massively parallel computation platforms and novel approaches developed in Computer Science for the exploitation of very big amounts of data.

II. TAXONOMY OF OPTIMIZATION PROBLEMS

The objective of this section is to dig into the different ways one can formulate and model optimization problems to

be addressed in power systems (with a focus on system-wide problems, as those addressed by TSOs).

A. Modeling the optimization problem from a formal viewpoint

Based on examples, we discuss the intrinsic nature of the different optimization problems, by distinguishing different possible formulations.

The general formulation can be summarized as a multistage decision making problem under uncertainty. But the formulation of this very general problem depends on the different time horizons and the type of decisions to make.

We can divide these decisions in three classes, illustrated here by an analogy with IT systems.

1) Decisions changing the structure of the system (developing the hardware)
2) Decisions changing policies or control/protection schemes (developing the software)
3) Decisions modifying the operating points of the system (selecting input data to run the software on the hardware)

We focus our discussion on the first problem which is the most challenging. This problem ideally requires the modeling of all the aspects of power systems: from possible long term energy policies to system operation using realistic modeling of the physical system and expectations of the grid users.

The decisions related to the structure of the system ("hardware") take long time to be implemented, they have to go through long permitting processes and need quite long construction times. They are investment decisions. The objective is to optimize the associated capital expenditures (capex) by comparing them to future operational expenditures (opex) saving. The time frame is varying from around ten years for the construction a new power line to one year for changing conductors on existing power lines. A part of the problem is to choose the relevant mixture of technological options: ac overhead power lines, ac underground cables, hvdc links, phase shifter transformers (PSTs), new conductors for existing power lines, new reactive compensation devices, ...

The problem can be formulated as stochastic dynamic programming problem. For long term expansion planning (as illustrated in Figure 1), scenario-based approaches seem the most attractive formulations in order to ensure some level of robustness as proposed in [1].

How to define reliability criteria and how to implement them, are key questions in these optimization problems. "Energy Not Served" or "Loss of Load" are generally used. An "artificial" monetization is performed and estimated costs are associated to these indexes. These large costs are simply added to operational expenditures. In stochastic formulations, generally only expected values are minimized without any cap on the maximum risk. This could be questionable and chance constraint programming or robust optimization could offer more relevant solutions. We could imagine that a generalization of Demand Response could change dramatically the definition of reliability and the foundations of power system design, pushing to less "hardware" and more "software" solutions as anticipated very optimistically in 1978 by F. C. Schweppe [2].
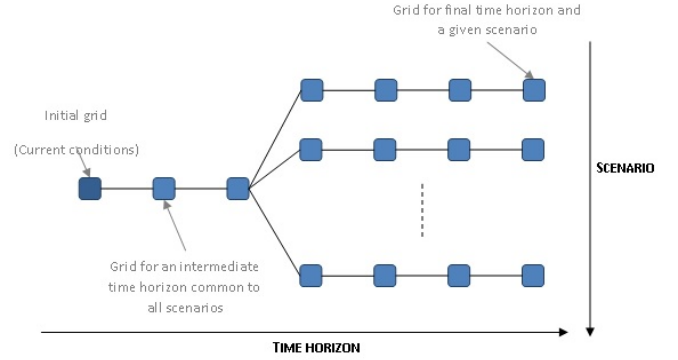

Fig. 1. Expansion Planning

A review of the current formulation and associated optimization problem is mandatory as proposed for example in the two on-going European projects: e-HIGHWAY2050 [3] and GARPUR [4]. In this global optimization problem, the sub problem on selection of relevant technological options leads to a combinatorial optimization very similar to a "knapsack" problem [1], increasing even more the complexity.

We can identify three different dimensions: spatial, temporal and stochastic. The spatial complexity is increasing: "more and more the electrical phenomena don't stop at administrative borders". We have to consider systems very extended (Pan-European Transmission System, Eastern or Western Interconnection in US, ...) and at the same time local active distribution grids. Time constants range from milliseconds to several years, leading to temporal complexity. Uncontrollable load and renewable energy sources implies to take into account more than ever stochastic behaviors. Considering spatial, temporal and stochastic complexity all together is still out of reach. Trade-offs must be made to take into account at most two of them in details at the same time and using approximation for third one.

The appropriate modeling of uncertainties is also a key factor to find realistic optimal solutions. The spatial and temporal correlations between these uncertainties must be taken into account not to be too optimistic or too pessimistic. This pushes towards probabilistic methods and risk based approaches. When the probabilistic properties of the uncertainties are only partially known, generalized semi-infinite programming seems an appealing method proposing to find robust solutions when the uncertainties live in a defined domain:

$$\min_{x \in X} f(x)$$
$$\text{subject to: } \forall \delta \in \Delta : g(x, \delta) \le 0,$$

where x are the decision variables and δ the uncertainties.

The objective function is related to satisfaction of the grid users (consumers and suppliers): maximization of social welfare. We need to estimate the expectations and the behaviors of the grid users. For long term decisions, it seems reasonable

---

[1]the knapsack problem: given a set of items, each with a mass and a value, determine those to include in a collection so that the total weight is less than or equal to a given limit and the total value is as large as possible.

to simulate a "perfect market" leading to a global rational behavior minimizing the costs. For more short term decisions, it could be important to simulate the actual behavior of the market players and the imperfect market design. These estimations could be formulated as optimization problems based on game theory finding Nash equilibrium[2].

In practice, we could have to formulate multi-objective optimization problems which are generally transformed in a single optimization problem using a weighted sum of the different objective functions

$$\min(w_1.f_1(x) + w_2.f_2(x) + ... + w_n.f_n(x)).$$

(For example, we want to minimize the production costs and the amount of emitted CO_2). Finding the associated weighting factors could be difficult and questionable. A more rational approach should be to formulate a true multi-objective function

$$\min(f_1(x), f_2(x), ..., f_n(x)).$$

But for a nontrivial multi-objective optimization problem, there does not exist a single solution that simultaneously optimizes each objective. In that case, the objective functions are said to be conflicting, and there exists a (possibly infinite number of) Pareto optimal solutions. A solution is called nondominated, Pareto optimal, Pareto efficient or noninferior, if none of the objective functions can be improved in value without degrading some of the other objective values. This leads to complex optimization problems which could be solved using meta-heuristics methods.

We could see that power system management could lead to a large diversity of optimization problems. The proper formulation of each problem has to be well thought out before searching for computational solutions.

B. Modeling the physics of the power system

The objective of this section is to analyze the physics and technological constraints arising from the power system, and explain their implications in terms of the nature of the above formulated optimization problems.

The quality of physical modeling of power systems used in optimization problems is essential in order to make "optimal" decisions. Solving optimization problems with a high accuracy based on not realistic enough modeling is useless. We need to find the right balance between realism and complexity. The usage of static and deterministic modeling using a linearization of the associated mathematical formulations should be questioned in the new context presented in the introduction.

A significant number of controls in electrical grids are discrete: switch on/off of breakers, switch on/off capacitor or reactor banks, tap changers on transformers, generating units producing with non zero minimal active power when they are started. These controls become integer variables in optimization problems and their treatments require a special

---

[2]Nash equilibrium is a solution concept of a non-cooperative game involving two or more players, in which each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing only their own strategy

attention; with a naive relaxation (round off strategy) is not always possible even to find feasible solutions.

Some controls and protection schemes implemented in local or centralized SPS/RAS are not event-based but measurement-based. They acts conditionally when a measurement or a set of measurements don't fulfill a given rule (for example, when a measurement value is beyond a given limit). This kind of behavior must taken into account in optimization problem. This leads to conditional corrective actions and the modeling of this type of hybrid system (continuous and discrete) requires binary variable [5].

Some active constraints in power systems could be more and more related to stability and the system dynamic behavior. The ultimate solution should be use a DAE-constrained optimization formulation but a reasonable first step could be to use "proxies" to hide this complexity. The idea is to learn using Monte-Carlo simulation, rules which ensure that the study case has a very low probability to be prone to stability issues [6] and to introduce these rules in static optimization problems.

III. RECENT DEVELOPMENTS IN THE FIELD OF OPTIMIZATION

In this section we highlight a set of novel results obtained in the last 10 years, that should be leveraged to better address the optimization problems encountered by TSOs.

A. Convexification

In this subsection, we introduce two recently explored directions aiming at developing algorithms which can compute, under some relatively general conditions, global optima for OPF problems and provide optimality bounds / certificates of optimality / certificates of infeasibility.

At the core of our problem, we have a nonconvex program to solve (AC power flow) even without considering discrete variables or discrete decisions. FERC sponsored a number of studies on ACOPF [7]. The motivation extracted from these studies is: *"The AC Optimal Power Flow (ACOPF) is at the heart of Independent System Operator (ISO) power markets and vertically integrated utility dispatch. ACOPF simultaneously optimizes real and reactive power. An approximated form of the ACOPF is solved in some form annually for system planning, daily for day-ahead commitment markets, and even every 5 minutes for real-time market balancing. The ACOPF was first formulated in 1962 by Carpentier. With advances in computing power and solution algorithms, we can model more constraints and remove unnecessary limits and approximations that were previously required to find a good solution in reasonable time. Today, 50 years after the problem was formulated, we still do not have a fast, robust solution technique for the ACOPF. Finding a good solution technique for the ACOPF could potentially save tens of billions of dollars annually."*

The main conclusion is: *"The iterative linear approximation (ILIV-ACOPF) solves faster and is more robust than most other approaches examined. Parameter tuning can improve performance. With binary variables, for example, as in the unit commitment and optimal transmission switching problems,*

linear approximations can be solved faster than nonlinear models."

This is indeed the state of the art but a large amount of dedicated tunings and heuristics is still required to solve each specific practical problem. The main drawback is that it is impossible to ensure that an AC feasible solution could be obtained using an iterative linear method. Moreover, when this iterative linear method is able to find an AC feasible and optimal solution, we believe that advanced nonlinear solvers are also able to find it in a quite efficient way. The comparison between solvers is always a very difficult task, a specific team has always a better knowledge of one solver and has a good understanding of all possible tunings, this is not generally the case for all the other solvers under comparison The convexification of ACOPF is the most promising generic method to avoid most of these tunings and heuristics. In the following sections, we present two recent promising convexification methods.

Notations. Let $\mathbb{C}$ denote the set of complex numbers, $\mathbb{R}$ the set of real numbers, $\mathbb{N}$ the set of non-negative integers, and $\mathbf{i} := \sqrt{-1}$. For $a \in \mathbb{C}$, Re a and Im a denote the real and imaginary parts of a respectively. For $a, b \in \mathbb{C}$, $a \leq b$ means Re $a \leq$ Re b and Im $a \leq$ Im b. A variable without a subscript denotes a vector with appropriate components, e.g. $s := (s_j, j = 0, \ldots, n)$, $S := (S_{jk}, (j, k) \in E)$. For vectors x, y, $x \leq y$ denotes componentwise inequality. The transpose of a matrix A is denoted as A^T, and the Hermitian (complex conjugate) transpose is denoted by A^H. A matrix A is Hermitian if $A = A^H$. A is positive semidefinite (or psd), denoted by $A \succeq 0$, if A is Hermitian *and* $x^H A x \geq 0$ for all $x \in \mathbb{C}^n$. Let $\mathbb{S}^n$ be the set of all $n \times n$ Hermitian matrices.

1) Semidefinite programming primer: ACOPF can be formulated as a quadratic constrained quadratic program (QCQP); see (22) below and the paper [8] in the invited session of PSCC2014 for more details. Convex relaxation of quadratic programs has been applied to many engineering problems [9] (see also [10] for some applications in power systems). There is a rich theory, mature algorithms, and extensive empirical experiences; see, e.g., [11], [12], [9]. We now summarize QCQP and its semidefinite relaxation. We formulate these problems in the complex domain for notational simplicity.

QCQP is the following optimization problem:

$$\min_{x \in \mathbb{C}^n} \quad x^H C_0 x \tag{1a}$$
$$\text{subject to} \quad x^H C_l x \leq b_l, \quad l = 1, \ldots, L, \tag{1b}$$

where the optimization variables are $x \in \mathbb{C}^n$, and for $l = 0, \ldots, L$, b_l are given real numbers and C_l are given Hermitian matrices in $\mathbb{S}^n$ (so that $x^H C_l x$ are real). If C_l, $l = 0, \ldots, L$, are positive semidefinite then (1) is a convex QCQP. Otherwise it is generally nonconvex, as OPF problems are.

Any psd rank-1 matrix X has a spectral decomposition $X = xx^H$, unique up to a rotation. Using $x^H C_l x = \text{tr } C_l xx^H =: \text{tr } C_l X$ we can rewrite a QCQP as the following equivalent problem where the optimization is over Hermitian matrices:

$$\min_{X \in \mathbb{S}^n} \quad \text{tr } C_0 X \tag{2a}$$
$$\text{subject to} \quad \text{tr } C_l X \leq b_l, \quad l = 1, \ldots, L \tag{2b}$$
$$X \succeq 0, \quad \text{rank } X = 1 \tag{2c}$$

The key observation is that, while the objective function and the constraints in (1) are quadratic in x they are linear in X in (2a)–(2b). The constraint $X \succeq 0$ in (2c) is convex ($\mathbb{S}^n_+$ is a convex cone). The rank constraint in (2c) is the only nonconvex constraint. Removing the rank constraint results in a semidefinite program (SDP):

$$\min_{X \in \mathbb{S}^n} \quad \text{tr } C_0 X \tag{3a}$$
$$\text{subject to} \quad \text{tr } C_l X \leq b_l, \quad l = 1, \ldots, L \tag{3b}$$
$$X \succeq 0. \tag{3c}$$

SDP is a convex program and can be efficiently computed. We call (3) an *SDP relaxation* of QCQP (1) because the feasible set of (2) is a subset of the feasible set of SDP (3). A strategy for solving QCQP (1) is to solve SDP (3) for an optimal X^{opt} and check its rank. If rank $X^{\text{opt}} = 1$ then X^{opt} is optimal for (2) as well and an optimal solution x^{opt} of QCQP (1) can be recovered from X^{opt} through spectral decomposition $X^{\text{opt}} = x^{\text{opt}}(x^{\text{opt}})^H$. If rank $X^{\text{opt}} > 1$ then, in general, no feasible solution of QCQP can be directly obtained from X^{opt} but the optimal objective value of SDP provides a lower bound on that of QCQP. If the SDP (3) is infeasible, then it is a certificate that the original QCQP (1) is infeasible.

2) Moment relaxation primer: Although the semidefinite relaxation in Section III-A1 globally solves many QCQP problems, the rank condition fails to be satisfied (i.e., rank $(X^{\text{opt}}) > 1$) for some practical power system optimization problems. Thus, the semidefinite relaxation does not yield the globally optimal solution to all problems of interest. Recognizing that QCQP is a special case of polynomial optimization problems (i.e., optimization problems with a polynomial objective function and polynomial constraints), tools from polynomial optimization theory can be exploited to solve a broader class of QCQP problems.

In particular, by exploiting the fact that polynomial optimization problems are themselves special cases of generalized moment problems, the Lasserre hierarchy [13], [14] can be applied to solve QCQP problems. The Lasserre hierarchy is composed of "moment" relaxations that take the form of SDPs. Increasing the relaxation order in the Lasserre hierarchy adds constraints that are redundant in the original QCQP but tighten the relaxation's feasible space, thus enabling global solution of a broader class of problems at the computational cost of larger SDPs.[3]

We next describe the moment relaxations resulting from the Lasserre hierarchy. Consider the generic polynomial optimiza-

---

[3]The dual form of the moment relaxations are sum-of-squares programs. The relaxation order corresponds to maximum degree of the sum-of-squares polynomials. See [13], [14] for details.

tion problem

$$\min_{x \in \mathbb{R}^n} \quad f_0(x) \tag{4a}$$

$$\text{subject to} \quad f_l(x) \geq 0, \quad l = 1, \ldots, L, \tag{4b}$$

where $f_l(x)$ for $l = 0, \ldots, L$ are polynomial functions of the decision variables x. The QCQP in (1) is a special case of (4) where all polynomials $f_l(x)$ are quadratic and a complex x is decomposed into real and imaginary parts. We use the notation x^α to represent a monomial in $f_l(x)$, where $\alpha \in \mathbb{N}^n$ is a vector denoting the exponents for each variable in x (i.e., $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$). Each polynomial is then written as

$$f_l(x) = \sum_{\alpha \in \mathbb{N}^n} c_{l,\alpha} x^\alpha, \tag{5}$$

where $c_{l,\alpha}$ is the scalar coefficient associated with each monomial x^α in $f_l(x)$.

The moment relaxations in the Lasserre hierarchy are derived using the fact that every polynomial optimization problem can be written as a generalized moment problem:

$$\min_{\mu \in \mathscr{M}(\mathbb{K})_+} \quad \int_{\mathbb{K}} f_0 \, d\mu \tag{6a}$$

$$\text{subject to} \quad \int_{\mathbb{K}} d\mu = 1, \tag{6b}$$

where $\mathbb{K}$ is the feasible space defined by the constraints (4b) and $\mathscr{M}(\mathbb{K})_+$ is the space of finite Borel measures μ on $\mathbb{K}$. (See Theorem 1.1 and its corresponding proof in [14].) The formulation (6) represents the polynomial optimization problem (5) as a linear (and therefore convex), infinite-dimensional optimization problem.

We next define a linear functional $L_y(f)$ which converts a polynomial argument $f(x) = \sum_{\alpha \in \mathbb{N}^n} c_{l,\alpha} x^\alpha$ to a linear combination of scalars y indexed by α:

$$L_y(f) = \sum_{\alpha \in \mathbb{N}^n} c_{l,\alpha} y_\alpha. \tag{7}$$

For matrix arguments $F(x)$ (i.e., entries of the matrix F are polynomials), $L_y(F)$ applies the linear functional to each element of F.

We can then use (7) to reformulate (6) as

$$\min_y \quad L_y(f_0) \quad \left(= \sum_{\alpha \in \mathbb{N}^n} c_{0,\alpha} y_\alpha \right) \tag{8a}$$

$$\text{s.t.} \quad y_0 = 1 \tag{8b}$$

$$y_\alpha = \int_{\mathbb{K}} x^\alpha d\mu, \quad \alpha \in \mathbb{N}^n, \text{ for some } \mu \in \mathscr{M}(\mathbb{K})_+ \tag{8c}$$

See Chapter 4 of [14]. Note that constraint (8b), which corresponds to (6b), enforces the fact that $x^0 = 1$.

With this reformulation, the optimization problem is described solely with the moments y_α of μ rather than μ itself. We form a *moment relaxation* by rewriting (8) entirely in terms of y_α (i.e., eliminating explicit dependence on μ) and adding

constraints that enforce necessary conditions for y_α satisfying (8c).

The *order* of the moment relaxation determines the stringency of the necessary conditions used in place of (8c). For the order-γ moment relaxation, define the vector χ_γ containing all monomials of the decision variables x up to order-γ:

$$\chi_\gamma = \begin{bmatrix} 1 & x_1 & \ldots & x_n & x_1^2 & x_1 x_2 & \ldots & x_n^2 & x_1^3 & x_1^2 x_2 & \ldots & x_n^\gamma \end{bmatrix}^\mathsf{T}. \tag{9}$$

The necessary conditions for y_α to satisfy (8c) take the form of positive semidefinite matrix constraints on a single *moment* matrix and on *localizing* matrices corresponding to each polynomial constraint in (4b). The order-γ moment relaxation has a moment matrix formed by applying the linear functional L_y to the outer product $\chi_\gamma \chi_\gamma^\mathsf{T}$:

$$L_y\left(\chi_\gamma \chi_\gamma^\mathsf{T}\right). \tag{10}$$

The localizing matrices result for application of the linear functional L_y to the polynomial matrices formed by multiplying the scalar polynomial constraints $f_l(x)$ by the outer product $\chi_{\gamma-\beta_l} \chi_{\gamma-\beta_l}^\mathsf{T}$:

$$L_y\left(f_l(x) \chi_{\gamma-\beta_l} \chi_{\gamma-\beta_l}^\mathsf{T}\right), \qquad l = 1, \ldots, L, \tag{11}$$

where the polynomial $f_l(x)$ in the lth constraint of (4b) has degree $2\beta_l$ or $2\beta_l - 1$. (Thus, for QCQP, $\beta_l = 2$, $l = 1, \ldots, L$.)

The order-γ moment relaxation is then

$$\min_y \quad L_y(f_0(x)) \tag{12a}$$

$$\text{s.t.} \quad y_0 = 1 \tag{12b}$$

$$L_y\left(\chi_\gamma \chi_\gamma^\mathsf{T}\right) \succeq 0 \tag{12c}$$

$$L_y\left(f_l(x) \chi_{\gamma-\beta_l} \chi_{\gamma-\beta_l}^\mathsf{T}\right) \succeq 0, \, l = 1, \ldots, L. \tag{12d}$$

In the same way as the semidefinite relaxation in Section III-A1, satisfaction of a rank condition is sufficient for exactness of the moment relaxation:

$$\text{rank}\left(L_y\left(\chi_1 \chi_1^\mathsf{T}\right)\right) = 1. \tag{13}$$

When the rank condition (13) is satisfied, the globally optimal decision variables are obtained from an spectral decomposition of the matrix $L_y\left(\chi_1 \chi_1^\mathsf{T}\right)$. Define λ as the non-zero eigenvalue of the matrix $L_y\left(\chi_1 \chi_1^\mathsf{T}\right)$ with corresponding unit-length eigenvector η. The globally optimal solution to the polynomial optimization problem is

$$x^{\text{opt}} = \sqrt{\lambda}\, \eta. \tag{14}$$

Since positive semidefinite moment and localizing matrices are only *necessary* for satisfaction of (8c), rather than necessary and sufficient, the moment relaxation may not be exact. That is, the moment relaxation may only provide a lower bound on the optimal objective value rather than the globally optimal decision variables x^{opt}. When the rank condition (13) is not satisfied, either 1) there are multiple global solutions, necessitating the more general conditions for exactness of the moment relaxations that are described in [14], which include the ability to extract multiple global solutions, or

2) the moment relaxation is not exact and only provides a lower bound on the optimal objective value. In the latter case, increasing the relaxation order can improve the lower bound on the optimal objective value and potentially yield a global solution. Solutions to the moment relaxations converge to the global optima of a polynomial optimization problem with increasing relaxation order [14].

Applying the first-order moment relaxation to a QCQP problem results in a moment matrix equivalent to X in (3). The localizing "matrices" for the first-order moment relaxation are in fact scalars which impose equivalent constraints as (3b).[4] Thus, the first-order relaxation is equivalent to the semidefinite relaxation described in Section III-A1.

The upper-left block of the moment and localizing matrices for a moment relaxation contain the moment and localizing matrices for the lower-order relaxations. Thus, positive semidefinite constraints on the order-γ moment and localizing matrices ensure that the $(\gamma - k)$-order moment relaxations' constraints are also satisfied for $k = 1, \ldots, \gamma - 1$. Therefore, the higher-order moment relaxations generalize the lower-order moment relaxations (and also generalize the semidefinite relaxation in Section III-A1).

As an illustrative example, consider the polynomial optimization problem

$$\min_{x} \quad x_1^2 + x_1 x_2 + x_2^2 \tag{15a}$$
$$\text{subject to} \quad x_1 x_2 - x_2^2 \geq 1 \tag{15b}$$
$$x_1 + x_2 \geq 0 \tag{15c}$$

The first-order moment relaxation for this problem is

$$\min_{y} \quad y_{20} + y_{11} + y_{02} \tag{16a}$$
$$\text{subject to} \quad y_{00} = 1 \tag{16b}$$
$$y_{11} - y_{02} - 1y_{00} \geq 0 \tag{16c}$$
$$y_{10} + y_{01} \geq 0 \tag{16d}$$
$$\begin{bmatrix} y_{00} & y_{10} & y_{01} \\ y_{10} & y_{20} & y_{11} \\ y_{01} & y_{11} & y_{02} \end{bmatrix} \succeq 0 \tag{16e}$$

where (16b) corresponds to (12b); (16c) and (16d) are the constraints on the localizing matrices (which are, in fact, scalars for the first-order relaxation) corresponding to (15b) and (15c), respectively; and (16e) is the moment matrix constraint.

The solution to (16) has a moment matrix with rank 2. Since the rank condition (13) is not satisfied, the first-order relaxation only provides a lower bound on the globally optimal objective value rather than the globally optimal decision variables. (The lower bound is 2.130.)

For this example, solving the second-order moment relaxation yields the globally optimal solution. The second-order

---

[4]For the first-order moment relaxation, the lth localizing matrix is $L_y \left(f_l \left(x \right) \chi_0 \chi_0^\mathsf{T} \right) = L_y \left(f_l \left(x \right) \cdot 1 \cdot 1^\mathsf{T} \right) = L_y \left(f_l \left(x \right) \right)$ (i.e., a scalar that imposes an equivalent constraint as (3b)).

moment relaxation for this example problem is

$$\min_{y} \quad y_{20} + y_{11} + y_{02} \tag{17a}$$
$$\text{s.t.} \quad y_{00} = 1 \tag{17b}$$

$$\left[\begin{array}{c|cc} y_{11}{-}y_{02}{-}y_{00} & y_{21}{-}y_{12}{-}y_{10} & y_{12}{-}y_{03}{-}y_{01} \\ y_{21}{-}y_{12}{-}y_{10} & y_{31}{-}y_{22}{-}y_{20} & y_{22}{-}y_{13}{-}y_{11} \\ y_{12}{-}y_{03}{-}y_{01} & y_{22}{-}y_{13}{-}y_{11} & y_{13}{-}y_{04}{-}y_{02} \end{array}\right] \succeq 0 \tag{17c}$$

$$\left[\begin{array}{c|cc} y_{10} + y_{01} & y_{20} + y_{11} & y_{11} + y_{02} \\ y_{20} + y_{11} & y_{30} + y_{21} & y_{21} + y_{12} \\ y_{11} + y_{02} & y_{21} + y_{12} & y_{12} + y_{03} \end{array}\right] \succeq 0 \tag{17d}$$

$$\left[\begin{array}{ccc|ccc} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{array}\right] \succeq 0 \tag{17e}$$

where (17c) and (17d) are the localizing matrix constraints corresponding to (15b) and (15c), respectively, and (17e) is the moment matrix constraint. Note that the upper-left block of the second-order moment and localizing matrices in (17) are equivalent to the first-order matrices in (16), which demonstrates that the moment relaxations build on themselves.

Since the solution to (17) satisfies the rank condition, we can obtain the globally optimal solution to (15): $x = [2.076 \quad 0.760]^\mathsf{T}$. Observe that the optimal objective value of 6.464 is indeed greater than the lower bound of 2.130 provided by the first-order relaxation.

3) Application to the OPF problem: Solving OPF through convex relaxation offers several *advantages*. First, while the linear DC OPF approximation is useful in a wide variety of applications, it is not always applicable, e.g., when power loss is not negligible, voltage magnitudes can fluctuate significantly, and reactive power needs to be optimized to stabilize voltages. Second, a solution of DC OPF may not be feasible. In this case, an operator may tighten some constraints in DC OPF and solve again. This may not only reduce efficiency but also relies on heuristics that are hard to scale to larger systems or faster control in the future. Third, when they converge, most nonlinear algorithms compute a local optimum, usually without assurance on the quality of the solution. In contrast, a convex relaxation provides for the first time the ability to check if a solution is globally optimal. If it is not, the solution provides a lower bound on the minimum cost and hence a bound on how far any feasible solution is from optimality. Moreover, unlike approximations, if a relaxed problem is infeasible, it is a certificate that the original OPF is infeasible.

Convex relaxations can also be used to verify the optimality of a feasible solution obtained by other methods (e.g. interior point methods, sequential linear or quadratic programming): if such a solution attains the same optimal value of the SDP relaxation of OPF, then it is globally optimal.

Further, if a candidate solution from any solver that provides both primal and dual solution values satisfies the Karush-Kuhn-Tucker (KKT) conditions for a relaxation, then the candidate solution is, in fact, the global optimum. (Note that these conditions are *sufficient* but not *necessary* for global

optimality of the candidate solution.) Since evaluating the KKT conditions for a semidefinite program only requires linear algebra operations (specifically, an inner product and a Cholesky decomposition) rather than solving a semidefinite program, testing for global optimality can be computationally advantageous. In [15], this approach is applied using the semidefinite relaxation described in Section III-A1. As an example of the computational benefits of this approach, solving the 2736-bus Polish system with MATPOWER [16] and verifying global optimality with the semidefinite relaxation's KKT conditions only requires 1.6% of the time required to actually solve the semidefinite relaxation directly.

Overall, research in convex relaxations lends itself to practical algorithms which are applicable to large-scale problems. For radial networks, sufficient conditions are known that guarantee exact semidefinite relaxations. Numerical experience with more general mesh networks demonstrates the relaxations' ability to find globally optimal solutions.

a) Semidefinite relaxation of the OPF problem: To the best of our knowledge, solving OPF through convex relaxation is first proposed in [17] as an SOCP for radial (tree) networks and in [18] as an SDP for general networks in a bus injection model. It is first proposed in [19], [20] as an SOCP for radial networks in the branch flow model of [21]. While these convex relaxations have been illustrated numerically in [17] and [18], whether or when they will turn out to be exact is first studied in [22]. Exploiting graph sparsity to simplify the SDP relaxation of OPF is first proposed in [23], [24] and analyzed in [25], [26]. See a comprehensive tutorial in [27], [28] (also [8]).

Consider a power network modeled by a connected undirected graph $G(N^+, E)$ where $N^+ := \{0\} \cup N$, $N := \{1, 2, \ldots, n\}$, and $E \subseteq N^+ \times N^+$. Each node in N^+ represents a bus and each edge in E represents a transmission or distribution line or transformer. For each edge $(i, j) \in E$ let $y_{ij} \in \mathbb{C}$ be its admittance. A bus $j \in N^+$ can have a generator, a load, both or neither. Let y_j denote a shunt element at bus j. Let V_j be the complex voltage at bus $j \in N^+$ and $|V_j|$ denote its magnitude. Bus 0 is the *slack bus*. Its voltage is fixed and we assume without loss of generality that $V_0 = 1\angle 0°$ per unit (pu). Let s_j be the net complex power injection (generation minus load) at bus $j \in N^+$.

The state of the network can be represented by the complex voltages vector $V \in \mathbb{C}^{n+1}$.[5] It satisfies the Kirchhoff's laws (power balance):

$$s_j = \sum_{k:j\sim k} y_{jk}^H V_j(V_j^H - V_k^H) + y_j^H |V_j|^2 \quad (18)$$

The power injections are constrained: $\underline{s}_j \leq s_j \leq \overline{s}_j$, $j \in N^+$, where $\underline{s}_j$ and $\overline{s}_j$ are given bounds on the injections at buses j. We can eliminate the variables s_j from the OPF formulation by using (18) to translate power injection constraints into quadratic constraints on V:

$$\underline{s}_j \leq \sum_{k:(j,k)\in E} y_{jk}^H V_j(V_j^H - V_k^H) + y_j^H |V_j|^2 \leq \overline{s}_j, \quad j \in N^+ \quad (19)$$

---

[5]For convenience we include V_0 in the vector variable $V := (V_j, j \in N^+)$ with the understanding that $V_0 := 1\angle 0°$ is fixed.

In addition, all voltage magnitudes must satisfy:

$$\underline{v}_j \leq |V_j|^2 \leq \overline{v}_j, \quad j \in N^+ \quad (20)$$

where $\underline{v}_j$ and $\overline{v}_j$ are given lower and upper bounds on the squared voltage magnitudes. Other constraints, such as line limits, can also be written as quadratic constraints on V, e.g., a limit on a line current may take the form $|y_{jk}^H (V_j - V_k)|^2 \leq \overline{\ell}_{jk}$. For notational simplicity, we restrict our discussion to constraints (19) and (20).

The OPF problem can be defined in terms of the complex voltage vector V. Suppose the cost function takes the form $V^H C V$ for some psd cost matrix C (e.g., the cost of real power at each generator bus or line loss over the network). Then the problem of interest is:

OPF:

$$\min_V C(V) \quad \text{subject to} \quad V \text{ satisfies } (19), (20). \quad (21)$$

Since (19) is quadratic, the feasible set is generally a nonconvex set. To see this, we transform (21) into the standard form QCQP (1), using the derivation in [29].

Let Y denote the $(n+1) \times (n+1)$ admittance matrix. Let I_j be the net injection current from bus j to the rest of the network. Then the current vector I and the voltage vector V are related by the Ohm's law $I = YV$. The power flow equations (18) are equivalent to:

$$s_j = V_j I_j^H = (e_j^H V)(I^H e_j), \quad j \in N^+$$

where e_j is the $(n+1)$-dimensional vector with 1 in the jth entry and 0 elsewhere. Hence, since $I = YV$, we have

$$s_j = \text{tr} \left(e_j^H V V^H Y^H e_j \right) = \text{tr} \left(Y^H e_j e_j^H \right) V V^H = V^H Y_j^H V$$

where $Y_j := e_j e_j^H Y$ is an $(n+1) \times (n+1)$ matrix with its jth row equal to the jth row of the admittance matrix Y and all other rows equal to the zero vector. Y_j is in general not Hermitian so that $V^H Y_j^H V$ is in general a complex number. Its real and imaginary parts can be expressed in terms of the Hermitian and skew Hermitian components of Y_j^H defined as:

$$\Phi_j := \frac{1}{2} \left(Y_j^H + Y_j \right) \text{ and } \Psi_j := \frac{1}{2\mathbf{i}} \left(Y_j^H - Y_j \right)$$

Then

$$\text{Re } s_j = V^H \Phi_j V \quad \text{and} \quad \text{Im } s_j = V^H \Psi_j V$$

Let their upper and lower bounds be denoted by

$$\underline{p}_j := \text{Re } \underline{s}_j \quad \text{and} \quad \overline{p}_j := \text{Re } \overline{s}_j$$
$$\underline{q}_j := \text{Re } \underline{s}_j \quad \text{and} \quad \overline{q}_j := \text{Re } \overline{s}_j$$

Let $J_j := e_j e_j^H$ denote the Hermitian matrix with a single 1 in the (j, j)th entry and 0 everywhere else. Then OPF (21) can be written as a standard form QCQP:

$$\min_{V \in \mathbb{C}^{n+1}} V^H C V \quad (22a)$$

$$\text{subject to} \quad V^H \Phi_j V \leq \overline{p}_j, \quad V^H(-\Phi_j)V \leq -\underline{p}_j \quad (22b)$$

$$V^H \Psi_j V \leq \overline{q}_j, \quad V^H(-\Psi_j)V \leq -\underline{q}_j \quad (22c)$$

$$V^H J_j V \leq \overline{v}_j, \quad V^H(-J_j)V \leq -\underline{v}_j, \quad (22d)$$

where $j \in N^+$ in (22), and the matrices Φ_j, Ψ_j, J_j are Hermitian. It is clear that the equivalent OPF (22) is a QCQP. Moreover, since Φ_j, Ψ_j are generally not psd, OPF is a nonconvex problem and NP-hard to solve in general.

As explained in Section III-A1, SDP relaxation can be applied to OPF and it takes the form

$$\min_{W \in \mathbb{S}^{n+1}} \quad \text{tr } CW$$

$$\text{subject to} \quad \text{tr } (\Phi_j W) \leq \overline{p}_j, \ \text{tr } ((-\Phi_j)W) \leq -\underline{p}_j,$$
$$\text{tr } (\Psi_j W) \leq \overline{q}_j, \ \text{tr } ((-\Psi_j)W) \leq -\underline{q}_j,$$
$$\text{tr } (J_j W) \leq \overline{v}_j, \ \text{tr } ((-J_j)W) \leq -\underline{v}_j.$$

If an SDP solution W^{opt} is of rank 1, then an optimal solution V^{opt} of OPF (22) can be recovered from W^{opt}. Otherwise, the optimal value of SDP provides a lower bound on the optimal cost of OPF.

b) Moment relaxations of the OPF problem: There are a variety of approaches for handling cases where the semidefinite relaxation fails to yield a global solution to OPF (i.e., $\text{rank}\,(X^{\text{opt}}) > 1$). These approaches include heuristics that attempt to find "nearby" local solutions (that may often be, in fact, the global solution) [30], [31] and the use of a branch-and-bound method to eliminate the gap between the lower bound from a relaxation and the upper bound from a nonlinear programming solver [32]. We next discusses an alternative to these approaches that generalizes the semidefinite relaxation using the Lasserre hierarchy of moment relaxations.

By specifying the voltage decision variables in rectangular coordinates (i.e., decomposing the complex voltage variable V in (21) such that the OPF problem has real decision variables V_d and V_q with $V = V_d + \mathbf{i}V_q$), the OPF problem takes the form of a polynomial optimization problem of the form described in (4). That is, the objective and constraints are polynomial functions of the voltage components. This formulation enables global solution of OPF problems using tools from polynomial optimization theory. In particular, the Lasserre hierarchy [13], [14] described in Section III-A2 can be applied to the OPF problem.

The order-γ moment relaxation of the OPF problem is developed using the vector χ_γ from (9) containing all monomials of the voltage components $V_{d1}, V_{d2}, \ldots, V_{dn}, V_{q1}, V_{q2}, \ldots, V_{qn}$ up to degree γ. (To maintain notational consistency with Section III-A2, the slack bus is labeled as bus 1 in the remainder of this section.)

Next construct the moment and localizing matrices. The moment matrix $L_y\left(\chi_\gamma\chi_\gamma^\intercal\right)$ from (10) contains terms corresponding to all monomials of the voltage components with up to degree 2γ. With all constraint polynomials having degree 2, the localizing matrices $L_y\left(f_l(x)\chi_{\gamma-1}\chi_{\gamma-1}^\intercal\right)$ from (11) are composed of polynomials with maximum degree 2γ. (The monomials in the matrix $\chi_{\gamma-1}\chi_{\gamma-1}^\intercal$ have maximum degree $2\gamma - 2$. Multiplication by the second-degree polynomials $f_l(x)$ results in a matrix with up to 2γ-degree polynomials.)

To eliminate the degree-of-freedom corresponding to the choice of the reference angle, include the angle reference constraint $V_{q1} = 0$. (Alternatively, V_{q1} can be eliminated from the optimization problem.) Also add constraint (12b) to enforce

$x^0 = 1$. Applying the linear functional L_y to the quadratic cost function in (21) results in a linear objective function. The semidefinite program corresponding to this objective function with these constraints and positive semidefinite constraints on the moment and localizing matrices is the order-γ moment relaxation of the OPF problem.

As a small illustrative example, consider the second-order relaxation of a two-bus OPF problem. (For notational conciseness, the reference angle constraint $V_{q1} = 0$ is used to eliminate V_{q1}.) This relaxation is developed using

$$\chi_1 = \begin{bmatrix} 1 \ V_{d1} \ V_{d2} \ V_{q2} \end{bmatrix}^\intercal \tag{23a}$$

$$\chi_2 = \begin{bmatrix} 1 \ V_{d1} \ V_{d2} \ V_{q2} \ V_{d1}^2 \ V_{d1}V_{d2} \ V_{d1}V_{q2} \ V_{d2}^2 \ V_{d2}V_{q2} \ V_{q2}^2 \end{bmatrix}^\intercal . \tag{23b}$$

The second-order moment matrix for a two-bus system is

$$L_y\left(\chi_2\chi_2^\intercal\right) = \left[\begin{array}{cccc|cccccc} y_{000} & y_{100} & y_{010} & y_{001} & y_{200} & y_{110} & y_{101} & y_{020} & y_{011} & y_{002} \\ y_{100} & y_{200} & y_{110} & y_{101} & y_{300} & y_{210} & y_{201} & y_{120} & y_{111} & y_{102} \\ y_{010} & y_{110} & y_{020} & y_{011} & y_{210} & y_{120} & y_{111} & y_{030} & y_{021} & y_{012} \\ y_{001} & y_{101} & y_{011} & y_{002} & y_{201} & y_{111} & y_{102} & y_{021} & y_{012} & y_{003} \\ \hline y_{200} & y_{300} & y_{210} & y_{201} & y_{400} & y_{310} & y_{301} & y_{220} & y_{211} & y_{202} \\ y_{110} & y_{210} & y_{120} & y_{111} & y_{310} & y_{220} & y_{211} & y_{130} & y_{121} & y_{112} \\ y_{101} & y_{201} & y_{111} & y_{102} & y_{301} & y_{211} & y_{202} & y_{121} & y_{112} & y_{103} \\ y_{020} & y_{120} & y_{030} & y_{021} & y_{220} & y_{130} & y_{121} & y_{040} & y_{031} & y_{022} \\ y_{011} & y_{111} & y_{021} & y_{012} & y_{211} & y_{121} & y_{112} & y_{031} & y_{022} & y_{013} \\ y_{002} & y_{102} & y_{012} & y_{003} & y_{202} & y_{112} & y_{103} & y_{022} & y_{013} & y_{004} \end{array}\right] \tag{24}$$

where the subscripts indicate the exponents α in the monomials x^α for $x = \begin{bmatrix} V_{d1} & V_{d2} & V_{q2} \end{bmatrix}^\intercal$. Observe that the upper-left block of the moment matrix in (24) is equivalent to the moment matrix from the first-order relaxation $L_y\left(\chi_1\chi_1^\intercal\right)$.

Consider the polynomial constraint $V_{d2}^2 + V_{q2}^2 - (0.9)^2 \geq 0$, which enforces a minimum voltage magnitude at bus 2 of 0.9 per unit. The localizing matrix corresponding to this constraint is shown in eqn. (25).

The (1,1) element of the localizing matrices are the scalar constraints from the first-order relaxation, again demonstrating that the higher-order moment relaxations build on the lower-order relaxations.

The first-order moment relaxation in the Lasserre hierarchy is equivalent to the SDP relaxation. In the same way as the SDP relaxation, the first-order relaxation in the Lasserre hierarchy is sufficient for globally solving many practical OPF problems [22], [25]. However, there exist OPF problems for which the first-order relaxation yields only a lower bound on the optimal objective value rather than the globally optimal decision variables [33], [34].

Although there is an upper bound on the relaxation order necessary for obtaining a global solution, this bound is too large for practical application [14]. If large relaxation orders were necessary for practical OPF problems, the moment relaxations would be computationally intractable. Fortunately, numerical experience demonstrates that the second-order relaxation globally solves most small OPF problems for which the first-order relaxation fails, with a third-order relaxation being necessary for only small parameter ranges of some problems [35], [36], [37].

One particularly illustrative example is the nine-bus system in [38]. MATPOWER [16] with the default interior point solver

$$L_y \left(\left(V_{d2}^2 + V_{q2}^2 - 0.81 \right) \chi_1 \chi_1^\mathsf{T} \right) =$$

$$\left[\begin{array}{c|ccc} y_{020} + y_{002} - 0.81y_{000} & y_{120} + y_{102} - 0.81y_{100} & y_{030} + y_{012} - 0.81y_{010} & y_{021} + y_{003} - 0.81y_{001} \\ y_{120} + y_{102} - 0.81y_{100} & y_{220} + y_{202} - 0.81y_{200} & y_{130} + y_{112} - 0.81y_{110} & y_{121} + y_{103} - 0.81y_{101} \\ y_{030} + y_{012} - 0.81y_{010} & y_{130} + y_{112} - 0.81y_{110} & y_{040} + y_{022} - 0.81y_{020} & y_{031} + y_{013} - 0.81y_{011} \\ y_{021} + y_{003} - 0.81y_{001} & y_{121} + y_{103} - 0.81y_{101} & y_{031} + y_{013} - 0.81y_{011} & y_{022} + y_{004} - 0.81y_{002} \end{array} \right]. \tag{25}$$

and default solver options either fails to converge or converges to one of three local optima (with objective values that are 10.0%, 37.5%, and 38.1% greater than the global optimum) for typical initializations. The first-order relaxation yields a lower bound that is 11% less than the global optimum. Thus, the first-order relaxation and a typical interior point method both perform poorly for this problem while a second-order relaxation finds the global solution.

While low-order relaxations solve many small OPF problems, there are computational challenges associated with large OPF problems. The size of the SDP matrices grows combinatorially with the relaxation order.[6] Only OPF problems with up to approximately ten buses are computationally tractable for direct applications of the second-order relaxation. Fortunately, the sparsity of typical OPF problems can be exploited in the moment-based relaxations using a matrix completion decomposition [39], [37], [40]. This approach decomposes the positive semidefinite constraint on the moment matrix into constraints on many smaller matrices. With direct application of this approach, the first-order moment-based relaxation is feasible for problems with thousands of buses [41], [25], [37]. Direct application of the matrix completion decomposition to the second-order relaxation results in computational intractability for problems with more than approximately forty buses [37], [40]. Solving higher-order moment relaxations for larger OPF problems can be accomplished by exploiting the fact that the first-order relaxation is sufficient for most regions of typical large OPF problems. An iterative technique that selectively applies the higher-order constraints to specific buses enables global solution of OPF problems with up to 300 buses [40].

There are many directions for future research related to the moment relaxations. For instance, one direction is extension to larger OPF problems, potentially using distributed optimization techniques. Expanding on research for existing convex relaxations [28], [34], another direction is the development of sufficient conditions for which the moment relaxations are guaranteed to yield globally optimal solutions and characterization of OPF problems for which moment relaxations fail.

B. Integer variables

The treatment of integer variables in optimization problems is a difficult task but some integer variables appear naturally in the modeling of power systems (See Section II.B).

For linear systems, very efficient solvers based on branch and bound methods exist (CPLEX, Xpress, Gurobi, ...), They

---

[6]The size of the moment matrix in the order-γ relaxation of an n-bus system is $\frac{(2n+\gamma)!}{(2n)!\gamma!}$.

are able to deal with very large systems (millions of constraints, several thousands of variables). They have generally very impressive presolvers capable to reduce drastically the number of constraints taken into account in the core algorithm. Some solvers can take advantage of multicore architectures and implement parallel evaluations over the combinatorial tree. The efficiency of these parallel methods is very problem depend and it must be tuned. The solvers offer hot start capabilities which can speedup a sequence of optimization problems (for example optimization of successive power system states encountered in real-time).

The integer variables could be useful not only for the modeling of physical components but also to implement complex behaviors and decisions making processes.

For example, for simulating automatic devices or operating rules acting when the system is in a specific state (conditional corrective actions) that must be taken into account in the power system optimization, we propose below a simple illustrative example, the modeling of the control of Phase Shifter Transformers (PST) in optimization problems.

We want to find the initial active powers of n_g generations (P) and initial phase shift ϕ_0 of a transformer phase shifter which minimize the deviation from a given generation schedule (P^0) and which ensure that the security remains inside security boundaries after each single contingency belonging to a finite set of contingencies (S_c) while simulating the effect of the local control of a single PST. Using a DC approximation, θ is the vector of voltage angles of the base case, θ_k the vector of voltage angles in the post contingency (k) and θ_k^c is the vector of voltage angles in the post contingency (k) and action of the PST. ϕ_k is the phase shift after contingency (k). ϕ_k is a free variable after the contingency (k) if a local quantity $g_{pst}(\theta_k, \phi_0)$ is greater than a given limit L_{pst}

$$\min \sum_{g=1}^{n_g} |p_g - p_g^0|$$

$$\begin{aligned} \text{subject to:} \quad & F(P, \theta, \phi_0) = 0 \\ & C(\theta, \phi_0) \le L_0 \\ & P_{min} \le P \le P_{max} \\ & -\phi_m \le \phi_0 \le \phi_m \\ & \text{for each contingency: } k \in S_c \\ & F_k(P, \theta_k, \phi_0) = 0 \\ & C_k(\theta_k, \phi_0) \le L_1 \\ & \text{if } (g_{pst}(\theta_k, \phi_0) \le L_{pst}) \text{ then } \{\phi_k = \phi_0\} \\ & -\phi_m \le \phi_k \le \phi_m \\ & F_k^c(P, \theta_k^c, \phi_k) = 0 \\ & C_k^c(\theta_k^c, \phi_k) \le L_2 \end{aligned}$$

where F, F_k, C, C_k and g_{pst} are linear functions.

The conditional action of PST can be modeled using a integer variable.

$$\delta.L_{pst} \leq g_{pst}(\theta_k, \phi_0) \leq \delta.M_g + L_{pst}$$
$$-\delta.M_\phi + \phi_0 \leq \phi_k \leq \delta.M_\phi + \phi_0$$

where δ is a binary variable M_g and M_ϕ are two big positive constants;

Indeed:
when $\delta = 0$:
$$0 \leq g_{pst}(\theta_k, \phi_0) \leq L_{pst}$$
$$\phi_k = \phi_0$$

when $\delta = 1$:
$$L_{pst} \leq g_{pst}(\theta_k, \phi_0) \leq M_g + L_{pst}$$
$$-M_\phi + \phi_0 \leq \phi_k \leq M_\phi + \phi_0$$

The problem becomes a MILP:

$$\min \sum_{g=1}^{n_g} |p_g - p_g^0|$$

subject to:
$$F(P, \theta, \phi_0) = 0$$
$$C(\theta, \phi_0) \leq L_0$$
$$P_{min} \leq P \leq P_{max}$$
$$-\phi_m \leq \phi_0 \leq \phi_m$$
for each contingency: $k \in S_c$
$$F_k(P, \theta_k, \phi_0) = 0$$
$$C_k(\theta_k, \phi_0) \leq L_1$$
$$\delta.L_{pst} \leq g_{pst}(\theta_k, \phi_0) \leq \delta.M_g + L_{pst}$$
$$-\delta.M_\phi + \phi_0 \leq \phi_k \leq \delta.M_\phi + \phi_0$$
$$-\phi_m \leq \phi_k \leq \phi_m$$
$$F_k^c(P, \theta_k^c, \phi_k) = 0$$
$$C_k^c(\theta_k^c, \phi_k) \leq L_2.$$

This requires to find the two big positive constants M_g and M_ϕ. To avoid possible numerical problems, it is recommended to select these big constants as small as possible. For example, in this case: $M_\phi = 2.\phi_m + \epsilon$ with ϵ: a small positive constant

This is only an illustrative example of what could be done for the modeling of complex behaviors in optimization problems. A case by case analysis is required for each specific problem but a powerful general framework is proposed in [5].

Another important example is the formulation of the optimal expansion planning problem. We want to find the power lines to add in an existing grid in order to have a feasible state and in order to optimize a given objective. In this problem, using constant incremental values for the power lines' parameters (incremental reactance ΔX and incremental capacity ΔC), the power flow between two buses could be expressed as the product between the difference of phase angles and the sum of incremental admittance.

$$P_{sr} = K_v.\frac{(\theta_s - \theta_r)}{X_0} + \sum_{k=1}^{n} \delta_k K_v.\frac{(\theta_s - \theta_r)}{\Delta X}$$
$$-C_0 - \sum_{k=1}^{n} \delta_k \Delta C \leq P_{sr} \leq C_0 + \sum_{k=1}^{n} \delta_k \Delta C$$

where δ_k are binary variables
$$\delta_{k+1} \leq \delta_k.$$

These equations are non linear; the product between two decision variables appears in the formulation $\delta_k(\theta_s - \theta_r)$ but it is possible to change it in a linear problem using the method proposed in [42].

$$p_k = K_v.\delta_k \frac{(\theta_s - \theta_r)}{\Delta X}$$
$$-\delta_k \Delta C \leq p_k \leq \delta_k \Delta C$$

where δ_k are binary variables.

These equations are equivalent to:

$$-M.(1 - \delta_k) \leq p_k - K_v.\frac{(\theta_s - \theta_r)}{\Delta X} \leq M.(1 - \delta_k)$$
$$-\delta_k \Delta C \leq p_k \leq \delta_k \Delta C,$$

where M is a big positive constant and δ_k are binary variables.

Indeed:

when $\delta_k = 1$:
$$p_k = K_v.\frac{(\theta_s - \theta_r)}{\Delta X}$$
$$-\Delta C \leq p_k \leq \Delta C$$

when $\delta_k = 0$:
$$-M \leq p_k - K_v.\frac{(\theta_s - \theta_r)}{\Delta X} \leq M$$
$$p_k = 0$$

$$\theta \in [-\pi, \pi]$$
$$M > K_v.\frac{2\pi}{\Delta X}.$$

Mixed Integer Non Linear Programming (MINLP) remains very challenging; the paper [43] gives a general diagnostic on practical problems. The branch and bound method remains attractive only if the NLP relaxation sub problems are not too expensive to solve, or if only few of them need to be solved. The existing MINLP solvers may run in trouble on non convex problems because the solution of the NLP relaxed sub problem cannot provide a lower bound on the solution of the original problem. Moreover, the very large CPU time spent diagnosing infeasible nodes for non convex problems may result in the failure of branch and bound solvers.

For practical applications, two heuristic approaches are currently used: sequential MILPs [7] or round-off methods, both of them must be tuned for each specific problem; even finding a feasible solution could be difficult for a new problem. An interesting alternative approach is to use mathematical programming with equilibrium (or complementarity) constraints (MPEC) [44]. This method ensures the feasibility of the solution. The mixed integer nonlinear programming (MINLP) problems can be seen as belonging to this more general class of MPEC optimization problems.

$$\min_{x \in D_x} J(x)$$
subject to: $c(x) \geq 0$
$$x_1 \perp x_2 = 0,$$
where $x = (x_0, x_1, x_2)$ and $\perp$ is the complementarity operator which requires that either a component $x_{1i} = 0$ or the corresponding component $x_{2i} = 0$.

The motivation is to use this complementarity operator for binary variables: if x is a binary variable, $x \in \{0, 1\} \iff x \perp (x - 1) = 0$

One attractive way of solving the above problem is to replace $\perp$ by a set of nonlinear inequalities such as:

$$\min_{x \in D_x} J(x)$$
$$\text{subject to: } c(x) \geq 0$$
$$x_1 \geq 0, x_2 \geq 0, X_1 x_2 \leq 0$$
$$\text{where } X_1 = diag(x_1).$$

Unfortunately, it has been shown that this equivalent problem violates the MFCQ (Mangasarian and Fromovitz Constraint Qualifications) condition at any feasible point. The domain $x_1 \geq 0, x_2 \geq 0, X_1 x2 \leq 0$ has no interior. Formulations compatible with Interior Point Methods have been proposed. To remedy the nonexistence of the central path, the idea is either to perturb or to penalize the complementarity constraint [45].

The First scheme is based on a "relaxation": a standard primal-dual method is applied to the following constraint: $x_1 \geq 0, x_2 \geq 0, X_1 x2 \leq \tau$.

The parameter τ is controlled in conjunction with the barrier parameters of the Interior Point Method. It can be shown that near a strongly stationary solution, the multipliers remain bounded and the central path exists.

The second scheme proposes a "penalization": an alternative to relaxation is to introduce a penalty for the complementarity constraint by adding $\rho x_1^T x_2$ to the objective. The resulting penalised NLP problem satisfies MFCQ and is well behaved.

Near a strongly stationary point, a sufficient large (but finite) penalty parameter exists and any IPM method converges to this point.

The two schemes can be shown to be equivalent in the sense that for every relaxation τ there exists a penalty parameter ρ such as both approaches give the same solution.

Another completely different alternative is to use a function for the modeling of complementary constraints (called NRS function: smooth natural residual). The desired property is that an NRS function vanishes at (a,b) if and only if a and b are non negative and their product ab is non positive. For example the following function:

$$\phi(a, b) = \frac{1}{2}(a + b - \sqrt{(a - b)^2 + \frac{ab}{\sigma}})$$
$$\text{for } \sigma > \frac{1}{2}, \text{ the property holds.}$$

The use of NRS function is appealing because it appears to allow standard large-scale NLP solvers to take into account MPEC.

The MPEC method seems to be useful to manage very large size problems when no other method could provide results as shown in the European Project PEGASE [46]. This idea is intensively used in the on going European Project iTesla to build realistic base cases in Monte Carlo simulations; a paper [47] describing this work, is presented in PSCC 2014. Of course as any non global MINLP solver, only sub-optimal solutions are expected from MPEC solvers for non-convex MINLP problems.

C. Chance constrained optimization

The basic idea of chance constrained optimization (CCO) is to solve a problem by minimising some function of the decision variables, while ensuring that the probability of some some constraint violations is small enough. This problem setting is useful in the context of many decision making problems under uncertainty.

1) The nature of CCO: A chance constrained optimization problem may be formulated as follows:

choose x^* to minimize $f(x)$ subject to $x \in \mathcal{X}, \delta \in \Delta$ and
$$\mathbb{P}\{\delta : x \in \mathcal{X}_\delta\} \geq 1 - \epsilon.$$

In this framework, δ is denoting an exogenous (vector of) variable(s) modeled in a stochastic way, and influencing the set of constraints of the problem to be solved, while x is a vector of decision variables to be chosen by the decision-maker. The latter wishes to choose $x^* \in \mathcal{X}$ so as to avoid the $\mathcal{X}_\delta$ constraint violation, with a probability ϵ-close to 1.

In the context of power systems management, we could apply this framework in order to choose preventive controls, while bounding the risk of insecure operation over the next time period, linked to contingencies and operating condition changes modelled as exogenous stochastic processes.

Notice that if the set Δ of possible exogenous random events is finite (i.e. $\Delta = \{\delta_1, \ldots, \delta_K\}$), then the chance constraint may be written as $\sum_{i=1}^{K} \mathbb{P}\{\delta_i\} 1_{x \in \mathcal{X}_{\delta_i}} \geq 1 - \epsilon$. While this is in general a non-convex constraint, in the finite case the subset of δ_i-values such that $x \in X_{\delta_i}$ could be modeled by integer variables, thus reducing the chance constrained optimization problem to a Mixed-Integer-Programming problem. However, when the set of uncertainties Δ is not finite, then the problem can in general not be reduced to a finite-dimensional optimization problem.

2) The scenario approach to CCO: When the function $f(\cdot)$ to be minimized is convex, and so are the constraints $\mathcal{X}$ and $\{\mathcal{X}_\delta, \forall \delta \in \Delta\}$, chance constrained optimization can be addressed by a recently developed and extremely powerful technique, known under the name of scenario approach, [48], [49], [50], [51]. To explain this technique, suppose that N samples $\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}$ independent and identically distributed according to the probability measure $\mathbb{P}$ are available. The idea behind the scenario approach of [48], [49], [50], [51] is to substitute the vast multitude of constraints in the infinite initial domain Δ with these N constraints only, and to find the optimal solution that satisfies *all* of these N constraints. This boils down to the program

$$\text{SP}_N : \quad \min_{x \in \mathcal{X}} c^T x$$
$$\text{s.t. } x \in \mathcal{X}_{\delta^{(i)}}, \ i \in \{1, \ldots, N\}.$$

Note that linearity of the cost function $c^T x$ comes to no loss of generality within a convex set-up. In fact, should the cost be a convex function $f(x)$, one could reformulate the problem of minimizing $f(x)$ with the constraints $\mathcal{X}_{\delta^{(i)}}$ as the minimization of an additional variable y with the constraints $\mathcal{X}_{\delta^{(i)}} \cap \{y \geq f(x)\}$ (epigraphic reformulation). Contrary to the original chance constrained formulation, SP_N is a standard

convex *finite* (i.e. with a finite number of constraints) optimization problem and, consequently, a solution can be found at low computational cost via available solvers.

Though totally disregarding all constraints but N of them may appear naive, the scenario approach stands on a very solid mathematical footing. Precisely, in [50] it has been shown that if

$$N \geq \frac{2}{\epsilon} \left(\ln \frac{1}{\beta} + d \right)$$

(d is the number of optimization variables), then, with probability no smaller than $1 - \beta$, the solution x_N^* of SP_N satisfies all constrains in Δ but at most an ϵ-fraction, i.e. $\mathbb{P}\{x_N^* \notin \mathcal{X}_\delta\} \leq \epsilon$, that is, the obtained solution is chance constrained feasible with high confidence $1 - \beta$.

On the other hand, the solution provided by SP_N can be conservative in that the achieved optimal cost is away from the optimal cost of the initial chance constrained problem. To improve the performance, one can a posteriori remove k constraints among the initial set of N constraints. In this way, the solution improves, and a solid theory permits one to still guarantee chance constrained feasibility. More precisely, one can show, [52], [53], that the solution obtained after removing k constraints satisfies

$$\mathbb{P}\{x_N^* \notin \mathcal{X}_\delta\} \leq \frac{k}{N} + O\left(\frac{\sqrt{k \ln k}}{N} \right).$$

$O\left(\frac{\sqrt{k \ln k}}{N} \right)$ has been used here for short, but it has a precise expression valid for any finite k and N, see [53]. Interestingly, this expression shows that $\mathbb{P}\{x_N^* \notin \mathcal{X}_\delta\}$ rapidly approaches the empirical chance constrained violation $\frac{k}{N}$ as N increases, so that the approach bears very little conservatism.

3) Applications of CCO to power systems: In the context of power systems, chance constrained optimization offers a rich framework for formulating decision making problems under uncertainty, when it is desired to avoid extreme situations, e.g. situations where the power system operation is not feasible without significant degradation of the quality of service.

In particular, in operation planning and real-time operation, one needs to take into account many uncertainties about load and generation patterns, combined with contingencies and possible failures of corrective and/or emergency controls. In such cases, rather than (as it is carried out in classical deterministic SCOPF formulations [54]) choosing a set of decisions that given a best guess of future operating conditions will ensure feasibility of operation for all possible contingencies (assuming a finite set of contingencies), or, alternatively, by ensuring that even for the worst-case scenario operation is still feasible given any contingency [55], one may prefer to model the problem as avoiding with high enough probability a set of undesirable operating conditions, which naturally leads to a chance constrained optimal power flow problem. We refer the reader to the paper [56], which provides a nice bibliography of such formulations, and discusses the interest of the chance-constrained OPF with respect to the classical security constrained OPF.

To our best knowledge, in the context of power systems management the first body of work leveraging the above described scenario approach to chance constrained optimization has been proposed by the group of ETH Zürich. In particular, reference [57] proposes to use this approach in the context of day-ahead operation planing while taking into account uncertainties about the next day load and renewable generation, so that the proposed schedule covers with high probability all next day operation conditions. On the other hand, the recent work reported in [58] focuses on real-time operation and more specifically on the tradeoff between preventive and corrective control, while taking into account probabilities of contingencies and also the probabilities of different failure modes of corrective controls. The authors propose to frame real-time security control as a chance constrained optimization problem, where the goal is to minimise expected costs (those related to preventive controls, combined with the expected costs of corrective controls and of service interruptions) while upper-bounding the probability of large service interruptions. The scenario based approach can be applied to this problem fruitfully as soon as the exhaustive treatment of all possible combinations of contingencies and control failure modes becomes intractable, and/or if uncertainties about the demand and renewable generation need to be taken into accout, e.g. when working in look-ahead mode over a few hours of real-time operation.

To summarise, using the scenario based approach to chance constrained optimization essentially opens the way to a much broader applicability of the chance constrained optimization paradigm, since the scenario based approach is completely agnostic with respect to the nature of the constraints and of the probabilistic properties of the problem, provided that convexity assumptions are valid. In this latter respect, the OPF "convexification" approaches discussed earlier in this paper should play also a very significant role in the future to further leverage the chance constrained optimization framework to many other power system management problems, such as operation planning, asset management and long-term system expansion problems. A first paper proposing such a combination of SDP relaxations with the scenario based chance-constrained optimization, is provided in [59], for the context of day-ahead operation planning.

D. Robust optimization

Another appealing approach to deal with uncertainty is "robust optimization" (RO), which essentially targets at finding the decision variables that would maximize the minimum possible objective function value under uncertainties. Indeed, as shown in the previous section "chance constraint optimization" may be technically difficult and requires the modeling of the probabilistic behavior of uncertainties. An advantage of RO is that operators are quite familiar with worst-case approaches and they are hence more prone to accept RO formulations when uncertainty can not be represented in a credible probabilistic way. One of the central questions regarding the RO framework is its conservativeness against high-impact-low-probability events. As will be shown in this section, choosing

proper optimization structure (e.g. static formulation versus two-stage and multi-stage formulations) and designing more flexible uncertainty modeling techniques are key to controlling the conservatism of the robust formulations.

RO could be used to solve different type of problems from Security Assessment to Unit Commitment. For example, the paper [60] proposes an application to Security Asssesment. This paper deals with day-ahead security management with respect to a postulated set of contingencies, while taking into account uncertainties about the next day generation/load scenario. The paper is based on a general method proposed in [61] with in this case, a specific bilevel linear optimization with discrete decisions variables in the follower problem.

In the following section, we present more in details an application to robust unit commitment.

1) Two-stage robust unit commitment models: The two-stage adaptive robust unit commitment (UC) model can be written in the following compact form

$$\min_{\boldsymbol{x} \in X} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \max_{\boldsymbol{d} \in \mathcal{D}} \min_{\boldsymbol{y} \in Y(\boldsymbol{x}, \boldsymbol{d})} \boldsymbol{b}^\top \boldsymbol{y} \right\}. \tag{26}$$

This formulation and solution methods have been first proposed and discussed in a series of papers, see e.g. [62], [63], [64], [65], [66]. It provides a fundamental robust optimization model in power systems operation. Here, the first-stage decision is the unit commitment decision $\boldsymbol{x}$ in the feasible region X, and the second-stage decision is the dispatch decision $\boldsymbol{y}$, which adapts to the realization of uncertainty $\boldsymbol{d}$ in the uncertainty set $\mathcal{D}$ and satisfies various operational constraints in $Y(\boldsymbol{x}, \boldsymbol{d})$ parametrized by the first-stage commitment decision $\boldsymbol{x}$ and $\boldsymbol{d}$. Usually, the uncertainty $\boldsymbol{d}$ models load uncertainty and generation uncertainty in variable resources such as wind and solar power.

Another important source of uncertainty is generation and transmission contingencies. One of the first applications of robust optimization to UC problems concerns contingency-constrained UC problems with $N - k$ generation security criterion (see [67]), where the traditional $N - k$ condition is modeled by the so-called budgeted uncertainty set [68]. The work [67] is further extended to include transmission contingencies and post-contingency redispatch in [69]. The proposed model has the two-stage structure as described in (26), except here $\boldsymbol{d}$ is the uncertainty of generation and transmission outages.

Several extensions of the fundamental model (26) have been proposed. For example, regret minimization objective is introduced as a means to decrease the conservativeness of the robust UC solution [70], where the regret of a UC solution is defined as the maximum difference between the minimum total cost by adjusting the dispatch decision for the fixed UC solution and the total cost in the perfect-information case where the UC and dispatch solutions can be jointly optimized after knowing the uncertainty. A hybrid approach of combining the worst-case cost of a robust UC model with the average cost of a stochastic UC model is proposed in [71].

The two-stage robust model (26) is also applied to the AGC control problem. The first-stage decision $\boldsymbol{x}$ stands for the operating point for the nominal load level, then the second-stage decision is the AGC dispatch signal responding to the uncertain load realization, see [72]. An alternative approach using affine decision policy for the AGC dispatch signal is discussed in [73]. A multiperiod adaptive robust model is proposed for look-ahead ED under significant wind [74], where the first-stage decision is the dispatch decision in the current time period, and the dispatch decisions for future look-ahead periods are the second-stage decision. A rolling-horizon simulation platform is developed for real-time dispatch integrated with statistical modules for updating uncertainty sets.

2) New Uncertainty Modeling Techniques: Besides the classic budget uncertainty set used in most of the robust UC models, some new uncertainty modeling techniques are proposed. For example, instead of using one uncertainty set $\mathcal{D}$ as in (26), multiple nested uncertainty sets $\mathcal{D}_k$ can be used for the second-stage problem as $\sum_k \rho_k \max_{\boldsymbol{d} \in \mathcal{D}_k} \min_{\boldsymbol{y} \in \Omega_k(\boldsymbol{x}, \boldsymbol{d})} \boldsymbol{b}_k^\top \boldsymbol{y}$ (see [75]). The positive weights ρ_k reflect the decision maker's risk choice. Some generalizations of uncertainty sets are discussed in [76]. In particular, some simple interval based models are proposed for modeling spatial and temporal correlations, non-polyhedral uncertainty sets such as ellipsoids can be used to model correlation, and using multiple uncertainty sets to model specific regions of uncertain data is also discussed, which reduces conservativeness through refining uncertainty sets, having a similar spirit as in [75]. A different type of uncertainty sets, termed dynamic uncertainty sets, is proposed in [74], which explictly model the dynamic relations between uncertainty in different times and space. A concrete example is proposed for modeling wind speed and power uncertainty of adjacent wind farms over multiple periods. A data-driven approach that fuses time-series models and uncertainty set modeling is implemented [74].

3) Multistage robust UC model: All of the above models are two-stage robust optimization formulations. A general multistage adaptive robust UC model is proposed in [77], where the dispatch decision depends on the sequential revelation of uncertainty, thus respecting the non-anticipativity constraints for real-time operations. The concept of simplified affine policies is introduced and efficient algorithms are developed. Extensive computational experiments on small, medium, and large-scale systems show significant improvement over existing deterministic and two-stage robust UC models in both cost reduction and reliability improvement.

IV. POSSIBLE SYNERGIES WITH SISTER FIELDS

On top of the above presented significant developments specific to the field of optimization, and their declination in the context of the core optimization problems ubiquitous in the context of power systems planning, maintenance, and operation activities, the present section aims at highlighting the potential for additional progress raised by ongoing developments in Information Technology (IT) (large scale computing and data management infrastructures, Internet of Things, broadband communications...) and by parallel scientific developments in the field of Computer Science (CS) (Machine Learning

and Randomized Algorithms). While these sister fields may provide many different opportunities in the context of power systems, we focus on those that are directly linked with the question of advanced optimization in power systems addressed in this paper.

A. Opportunities/needs from progress in Information Technology (IT)

Recent technological developments make High Performance Computing (HPC) as well as distributed Big Data (BD) management become low cost solutions of great value in many areas. The question is "How to take the best advantage of these novel technologies in the context of power system's optimization practice?". In this section we suggest some opportunities and needs in this respect.

1) High Performance Computing (HPC) opportunities: In the context of very large-scale optimization of power system problems, HPC offers still a lot of opportunities to yield significant practical progress. Indeed, if we analyse the availability (at constant costs) of computing infrastructures and compare it to those actually exploited in power system optimization problems, we observe that the gap between available means and actually used ones is increasing over the last years. This means that further work should as well focus on how to exploit these HPC environments to solve more efficiently many of the optimization problems encountered in power system practice.

There are basically two ways of taking advantage of HPC in this context: i.) by developing parallel/distributed versions of the generic solvers used (MILP, IP, SDP, . . .); ii) by mapping the power system optimization applications in a clever and thought out way on the available low cost HPC architectures.

Both directions are of interest, but given the pressure on power systems application scale up and the larger number of opportunities to do this, we believe that the leveraging of HPC infrastructures should be also directed by the nature of power system applications, rather than solely rely on the expected progress in generic optimization solver implementations on HPC infrastructures.

Notably, many Monte-Carlo methods, e.g. used in power systems reliability assessment, trivially lend themselves to decompositions into a large number of independent computational tasks, each one corresponding to a sub-sample. On the other hand, many of the optimization problems used for power system planning and operation can be decomposed naturally into a large set of much smaller optimization problems formulated over partial objectives and coupled only by a reduced number of equality constraints, while the overall objective function is decomposable into a (possibly weighted) sum of the partial objectives. In this case, the so-called Progressive Hedging Algorithm (PHA) [1], initially proposed for scenario-tree based multi-stage stochastic programming, can be applied in order to exploit multiple processors, where at each iteration a given processor solves one of the subproblems modified by penalizing its partial objective by the discrepancy of its decision variables to the coupling constraints' satisfaction. The resulting algorithm offers convergence guarantees under the assumption that the concerned optimization problem is convex,

and even if the partial problems are not solved perfectly at each iteration. In particular, the classical security constrained OPF, as well as various scenario-based stochastic versions of this problem, may be reformulated in order to possibly take advantage of this decomposition strategy when a HPC infrastructure is available. For example, the SP_N problem of section III-C2, may be rewritten as follows:

$$SP_N: \quad \min \frac{1}{N} \sum_{i=1}^{N} c^T x_{\delta(i)}$$
$$\text{s.t. } x_{\delta(i)} \in \mathcal{X} \cap \mathcal{X}_{\delta(i)}, \ i \in \{1, \ldots, N\},$$
$$\text{s.t. } x_{\delta(i)} = \frac{1}{N} \sum_{i=1}^{n} x_{\delta(i)}, \ i \in \{1, \ldots, N\},$$

where the last set of constraints are the coupling constraints among the solutions of the N subproblems, each one of which would be written as:

$$SP_N(i): \quad \min c^T x_{\delta(i)}$$
$$\text{s.t. } x_{\delta(i)} \in \mathcal{X} \cap \mathcal{X}_{\delta(i)}.$$

This formulation lends itself directly to the application of the PHA (we refer the reader to [78] for an explanation of the use of PHA to optimization problems having a similar structure).

On the other hand, when it comes to the use of HPC infrastructures for the resolution of non-convex optimization problems, further complications arise. In this respect, we refer the interested reader to the PSCC2014 paper [79], which addresses the problem of leveraging HPC for the efficient resolution of non-convex optimization problems arising in the context of energy systems.

Still, no software packages are available to solve such problems in a seamless way by exploiting available HPC infrastructures. Further work is thus required to make this happen as soon as possible.

2) Big Data (BD) exploitation needs: When it comes to real-world applications, it is notable that: i) optimization problem solving may take advantage of a large body of existing data (input data, scenarios for validation and tuning of algorithms); ii) the recurrent applications of optimization solvers generate as well huge amounts of output data that may be exploited by other applications (see also Section 4.2).

It is therefore of interest to be attentive about the progresses in the context of BD management and exploitation methods, in order to well target the added value of novel optimization workflows in the context of power systems planning, maintenance and operation.

While the power systems field could undoubtedly benefit from the rapid growth of measurement, data-collection, and data-processing technologies, some important kinds of data will at the same time remain scarce, such as for example data about component failures, or data about system performance in the context of large deviations from the steady-state operation (short-term and medium-term dynamics). Hence, some of the mathematical models used in the context of optimization will continue to suffer from a *Lack of Big Data*. Therefore, one of the main questions to be addressed is to develop methods able to translate the growing amounts of data not only into "optimal" parameter settings to be used in optimization, but also to quantify the uncertainties on these parameters given the inevitable lack of some of the relevant data. Essentially,

this means that all optimization problems should be formulated in such a way so as to make their solution robust to inaccuracies in their parameter settings, given the data exploitation technology that is applied to the available amounts of data.

Modeling uncertainties about models used in optimization (sic) allows in principle to assess suboptimalities induced by these uncertainties, and hence should allow one to justify investments in data collection, sharing, and processing. To do this, one will have to formulate "meta-optimization problems", of even larger sizes than those discussed in the present paper (see [80]).

B. Opportunities raised by progress in Computer Science (CS)

Recent progresses in theoretical computer science of interest in the context of the present paper concern mainly two directions: i) the development of methodologies to construct randomized algorithms of low complexity and with probabilistic guarantees of performance; ii) the huge progress in machine learning allowing these methods to be applied both to very large scale problems and also to problems with structured input and output spaces.

1) Randomized Algorithms (RA): It is well known that the Monte-Carlo method is a randomized algorithm for computing an integral over a high dimensional space, and an algorithm that does not suffer from the curse of dimensionality of alternative deterministic discretization techniques. The Monte-Carlo method essentially trades deterministic guarantees and corresponding intractable computational complexities of high-dimensional deterministic discretization-and-integration schemes for an extremely simple randomized sampling-and-averaging algorithm, the latter however presenting "only" probabilistic guarantees of accuracy scaled by the complexity (i.e. the variance) of the function that has to be integrated [81]. The generic Monte-Carlo method has been evolved during the last 50 years into a sophisticated set of modified sampling and aggregation techniques that make the approach very efficient in many applications by exploiting knowledge about the structure of the practical problem to be addressed. Furthermore, deterministic low-discrepancy sampling schemes have been developed in this context, so as to further reduce the variance of the Monte-Carlo methods when the problem structure is well understood.

More broadly, the study of randomized algorithms and their de-randomization has become one of the major avenues for progress in algorithmics, and specially in the study of the distinction between truly difficult (but often artificial) computational problems and those that only appear as difficult but can be in practice tackled efficiently at least in a probably-approximately-correct way [82]. For the sake of space, we will not dig further in this direction, but we clearly indicate that the realm of randomized algorithms is a major direction for future progress, specially in the context of large scale power system optimization problems. As a matter of fact, the scenario-based approach to chance constrained optimization presented in section III-C is an instance of this generic approach.

2) Machine Learning (ML): Machine Learning is the modern word for denominating Statistics. Statistics, as well as Machine Learning, aims at drawing conclusions from observational data in the form of mathematical models. While the statistics community has been educated mostly by the mathematical theory of probability, machine learning researchers have as well (and mostly) been influenced by computer science and algorithmics. It also is important to notice that progress in machine learning has been driven in the last twenty years by a combination of progresses in data collection (such as in the field of bioinformatics, in environmental sciences, and social networks) yielding novel needs for problem solutions, and by the leveraging of optimization solvers (at the heart of most machine learning algorithms, in combination with randomization techniques) to the very large scale problems raised by machine learning. In this context, the possibility to reformulate machine learning problems in a flexible way has been a major source of progress.

From the point of view of power systems optimization, machine learning offers the possibility to exploit both input and output data-streams of optimization solvers in order to build simplified models (we have called them "proxies", in section 2.2) of the corresponding input-output relation [6], [80]. Such machine-learnt proxies extracted from a given context (say short-term power system operation) can then be exploited in other contexts (say mid-term operation planning) by being "plugged" into the optimization problems to be solved in these other contexts. To make this possible, the machine-learnt models not only need to be accurate (which is the main target of machine learning research), but they need as well to be expressed in a mathematical form that is exploitable within the upstream optimization contexts, which should be the target of power systems research.

V. Summary

Power system planning and operation raises many important decision making problems, which can generally be stated as large-scale, non-linear, mixed-integer continuous, non-convex, stochastic and/or robust optimization problems.

In the last years, many progresses have been made in the theory and implementation of optimization algorithms, driven by research in applied mathematics and by multitudinous opportunities of application. The combination of these novel ideas to improve the state-of-the-art of power systems optimization is an important direction of future work.

On the other hand, low cost information technology (HPC and Big Data) as well as progresses in machine learning and randomized algorithms offer other enabling approaches to apply optimization techniques in power systems.

We suggest that the research community should further focus on the proper formulation of power system optimization problems with the help of power system experts, and develop more intensively fruitful collaborations with researchers in applied mathematics and computer science to determine the most effective solution strategies for these problems.

At the same time, we think that more systematic investments in a more effective use of modern information technologies, especially in the context of high-performance computing and massive data exploitation should be made by the power systems industry.

REFERENCES

[1] R.T. Rockafellar, R.J.-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16(1):119-147, 1991.

[2] F. C. Schweppe. Power systems "2000": hierarchical control strategies. *Spectrum, IEEE*, Volume:15, Issue: 7, July 1978

[3] e-HighWay2050: Modular Development Plan of the Pan-European Transmission System 2050. *http://www.e-highway2050.eu/*

[4] GARPUR: Generally Accepted Reliability Principle with Uncertainty modelling and through probabilistic Risk assessment. *http://www.garpur-project.eu/*

[5] A. Bemborad and M. Morari. Control of systems integrating logic, dynamics, and constraints. *Automatica*, 35:407-427, 1999.

[6] L. Wehenkel, M. Pavella, E. Euxibie, B. Heilbronn. Decision tree based transient stability method - a case study. *IEEE Trans. on Power Syst.*, 9(1):459-469, 1994.

[7] http://www.ferc.gov/industries/electric/indus-act/market-planning/opf-papers.asp#p8

[8] L. Gan and S.H. Low. Convexification of AC optimal power flow. In *Proc. of the 18th Power Systems Computation Conference (PSCC)*, Wroclaw, Poland, 2014.

[9] Z. Luo, W. Ma, A.M.C. So, Y. Ye, and S. Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, 27(3):20-34, 2010.

[10] G.B. Giannakis, V. Kekatos, N. Gatsis, S.-J. Kim, H. Zhu, and B. Wollenberg. Monitoring and optimization for power grids: A signal processing perspective. *IEEE Signal Processing Magazine*, 30(5):107-128, 2013.

[11] S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[12] H. Wolkowicz, R. Saigal, and L. Vandenberghe. *Handbook of semidefinite programming: theory, algorithms, and applications*, volume 27. Springer Netherlands, 2000.

[13] J.B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optimiz.*, 11(3):796-817, 2001.

[14] J.B. Lasserre. *Moments, Positive Polynomials and their Applications*. Imperial College Press, vol. 1, 2010.

[15] D.K. Molzahn, B.C. Lesieutre, and C.L. DeMarco. A Sufficient condition for global optimality of solutions to the optimal power flow problem. *IEEE Trans. on Power Syst.*, 29(2):978-979, 2014.

[16] R. Zimmerman, C. Murillo-Sánchez, and R. Thomas. MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans. Power Syst.*, 99:1-8, 2011.

[17] R.A. Jabr. Radial distribution load flow using conic programming. *IEEE Trans. on Power Syst.*, 21(3):1458-1459, 2006.

[18] X. Bai, H. Wei, K. Fujisawa, and Y. Wang. Semidefinite programming for optimal power flow problems. *Int'l J. of Electrical Power & Energy Systems*, 30(6-7):383-392, 2008.

[19] M. Farivar, C.R. Clarke, S.H. Low, and K.M. Chandy. Inverter var control for distribution systems with renewables. In *Proceedings of IEEE SmartGridComm Conference*, Brussels, Belgium, 2011.

[20] M. Farivar and S.H. Low. Branch flow model: relaxations and convexification (parts I, II). *IEEE Trans. Power Syst.*, 28(3):2554-2572, 2013.

[21] M.E. Baran and F.F. Wu. Optimal capacitor placement on radial distribution systems. *IEEE Trans. Power Delivery*, 4(1):725-734, 1989.

[22] J. Lavaei and S.H. Low. Zero duality gap in optimal power flow problem. *IEEE Trans. Power Syst.*, 27(1):92-107, 2012.

[23] X. Bai and H. Wei. A semidefinite programming method with graph partitioning technique for optimal power flow problems. *Int'l J. of Electrical Power & Energy Systems*, 33(7):1309–1314, 2011.

[24] R.A. Jabr. Exploiting sparsity in SDP relaxations of the OPF problem. *IEEE Trans. Power Syst.*, 27(2):1138-1139, 2012.

[25] D. Molzahn, J. Holzer, B. Lesieutre, and C. DeMarco. Implementation of a large-scale optimal power flow solver based on semidefinite programming. *IEEE Trans. Power Syst.*, 28(4):3987-3998, 2013.

[26] S. Bose, S.H. Low, T. Teeraratkul, and B. Hassibi. Equivalent relaxations of optimal power flow. *IEEE Trans. Automatic Control*, 2014.

[27] S.H. Low. Convex relaxation of optimal power flow, I: formulations and relaxations. *IEEE Trans. on Control of Network Systems*, 1(1):15-27, 2014.

[28] S.H. Low. Convex relaxation of optimal power flow, II: exactness. *IEEE Trans. on Control of Network Systems*, 1(2):177-189, 2014.

[29] S. Bose, D. Gayme, K.M. Chandy, and S.H. Low. Quadratically constrained quadratic programs on acyclic graphs with application to power flow. arXiv:1203.5599v1, 2012.

[30] R. Louca, P. Seiler, and E. Bitar. A rank minimization algorithm to enhance semidefinite relaxations of optimal power flow. In *51st Annu. Allerton Conf. Commun., Control, and Comput.*, 2013.

[31] R. Madani, S. Sojoudi, and J. Lavaei. Convex Relaxation for optimal Power Flow Problem: Mesh Networks. To appear in *IEEE Trans. on Power Syst.*, 2014. Preprint available: http://www.ee.columbia.edu/ lavaei/Penalized_SDP_2013.pdf

[32] A. Gopalakrishnan, A.U. Raghunathan, D. Nikovski, and L.T. Biegler. Global optimization of optimal power flow using a branch & bound algorithm. In *50th Annu. Allerton Conf. Commun., Control, and Comput.*, 2012.

[33] B. Lesieutre, D. Molzahn, A. Borden, and C. DeMarco. Examining the L-limits of the application of semidefinite programming to power flow problems. In *49th Annu. Allerton Conf. Commun., Control, and Comput.*, pp. 1492-1499, 2011.

[34] D. Molzahn, B. Lesieutre, and C. DeMarco. Investigation of non-zero duality gap solutions to a semidefinite relaxation of the optimal power flow problem. In *47th Hawaii Int. Conf. Syst. Sci. (HICSS)*, 2014.

[35] D.K. Molzahn and I.A. Hiskens. Moment-based relaxation of the optimal power flow problem. To appear in *18th Power Syst. Comput. Conf. (PSCC)*, 2014.

[36] C. Josz, J. Maeght, P. Panciatici, and J.C. Gilbert. Application of the moment-SOS approach to global optimization of the OPF problem. To appear in *IEEE Trans. Power Syst.*, Preprint available: http://arxiv.org/abs/1311.6370.

[37] B. Ghaddar, J. Marecek, and M. Mevissen. Optimal power flow as a polynomial optimization problem. IBM Research Ireland, Tech. Rep., 2014, http://arxiv.org/abs/1404.3626.

[38] W. Bukhsh, A. Grothey, K. McKinnon, and P. Trodden. Local solutions of the optimal power flow problem. *IEEE Trans. Power Syst.*, 28(4):4780-4788, 2013.

[39] H. Waki, S. Kim, M. Kojima, and M. Muramatsu. Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. *SIAM J. Optimiz.*, 17(1):218-242, 2006.

[40] D.K. Molzahn and I.A. Hiskens. Sparsity-exploiting moment-based relaxations of the optimal power flow problem. University of Michigan, Tech. Rep., April 2014, Submitted for publication, Preprint available: http://arxiv.org/abs/1404.5071.

[41] R. Jabr. Exploiting sparsity in SDP relaxations of the OPF problem. *IEEE Trans. Power Syst.*, 27(2):1138-1139, 2012.

[42] L. Bahiense, G.C. Oliveira, M. Pereira, S. Granville. A mixed integer disjunctive model for transmission network expansion. *IEEE Trans. on Power Syst.*, 16(3), 2001.

[43] S. Leyffer, J. Linderoth, J. Luedtke, A. Miller, T. Munson. Applications and algorithms for mixed integer non linear programming. *SciDAC Journal of Physics*, 2009.

[44] S. Leyffer, J. Linderoth. Mixed integer nonlinear programming (MINLP). In *Informs Annual Meeting*, San Francisco, 2005

[45] M. Anitescu. On solving mathematical programs with complementarity constraints as nonlinear programs. 2000. [Online]. Available: http://www.mcs.anl.gov/papers/P864.pdf

[46] L. Platbrood, S. Fliscounakis, F. Capitanescu, P. Panciatici, C. Merckx, M.O. Vazquez. Development of prototype software for steady state optimization of the ETN. PEGASE consortium, Tech. Rep., 2011. [Online]. Available: http://www.fp7-pegase.com/pdf/D3.2 proto.pdf

[47] M. Ruiz, J. Maeght, A. Marie, P. Panciatici, A. Renaud. A progressive method to solve large-scale AC Optimal Power Flow with discrete aspects and control of the feasibility. In *Proceedings of PSCC*, 2014.

[48] G. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25-46, 2005.

[49] G. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Trans. on Automatic Control*, 51(5):742-753, 2006.

[50] M.C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211-1230, 2008.

[51] M.C. Campi, S. Garatti, and M. Prandini. The scenario approach for systems and control design. *Annual Reviews in Control*, 33(2):149-157, 2009.

[52] M.C. Campi and S. Garatti. A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality. *Journal of Optimization Theory and Applications*, 148(2):257-280, 2011.

[53] S. Garatti and M.C. Campi. Modulating robustness in control design: principles and algorithms. *IEEE Control Systems Magazine*, 33(2):36-51, 2013.

[54] F. Capitanescu, J. Martinez Ramos, P. Panciatici, D. Kirschen, A. Marano Marcolini, L. Platbrood, and L. Wehenkel. State-of-the-art, challenges, and future trends in security constrained optimal power flow. *Electric Power Systems Research*, 81(8):1731-1741, 2011.

[55] F. Capitanescu, S. Fliscounakis, P. Panciatici, and L. Wehenkel. Day-ahead security assessment under uncertainty relying on the combination of preventive and corrective controls to face worst-case scenarios. In *Proceedings of PSCC*, 2011.

[56] C. Hamon, M. Perninge, and L. Söder. The value of using chance-constrained optimal power flows for generation re-dispatch under uncertainty with detailed security constraints. In *Proceedings of the 5th IEEE PES Asia-Pacific Power and Energy Engineering Conference*, 2013.

[57] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson. Probabilistic guarantees for the N-1 security of systems with wind power generation. In *Proceedings of PMAPS*, 2012.

[58] E. Karangelos, P. Panciatici, and L. Wehenkel. Whither probabilistic security management for real-time operation of power systems? In *Proceedings of IREP Symposium-Bulk Power Systems Dynamics and Control-IX*, 2013.

[59] M. Vrakopoulou, M. Katsampani, K. Margellos, J. Lygeros, and. G. Andersson. Probabilistic security-constrained ac optimal power flow. In *Proceedings of IEEE PowerTech Grenoble*, 2013.

[60] S. Fliscounakis, P. Panciatici, F. Capitanescu, L. Wehenkel, Contingency Ranking With Respect to Overloads in Very Large Power Systems Taking Into Account Uncertainty, Preventive, and Corrective Actions. *IEEE Trans. on Power Syst.* (28)4:4909-4917, 2013.

[61] A. Mitsos, Global solution of nonlinear mixed-integer bilevel. *J. Glob. Optim.* (47)4:557-582, 2010.

[62] R. Jiang, J. Wang, and Y. Guan. Robust unit commitment with wind power and pumped storage hydro. *IEEE Trans. on Power Syst.*, 27(2):800-810, 2012.

[63] L. Zhao and B. Zeng. Robust unit commitment problem with demand response and wind energy. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–8. IEEE, 2012.

[64] D. Bertsimas, E. Litvinov, X.A. Sun, J. Zhao, and T. Zheng. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Trans. on Power Syst.*, 28(1):52-63, 2013.

[65] B. Zeng and L. Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457-461, 2013.

[66] R. Jiang, M. Zhang, G. Li, and Y. Guan. Two-stage network constrained robust unit commitment problem. *European Journal of Operational Research*, 234(3):751-762, 2014.

[67] A. Street, F. Oliveira, and J.M. Arroyo. Contingency-constrained unit commitment with $n - k$ security criterion: A robust optimization approach. *IEEE Trans. on Power Syst.*, 26(3):1581-1590, 2011.

[68] D. Bertsimas, D.B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464-501, 2011.

[69] Q. Wang, J.P. Watson, and Y. Guan. Two-stage robust optimization for $n - k$ contingency-constrained unit commitment. *IEEE IEEE Trans. on Power Syst.*, 2013.

[70] R. Jiang, M. Zhang, and Y. Guan. Two-stage minimax regret robust unit commitment. *IEEE Trans. on Power Syst.*, 28(3):2271-2282, 2013.

[71] C. Zhao and Y. Guan. Unified stochastic and robust unit commitment. *IEEE Trans. on Power Syst.*, 2013.

[72] T. Zheng, J. Zhao, E. Litvinov, and F. Zhao. Robust optimization and its application to power system operation. In *CIGRE*, 2012.

[73] R. Jabr. Adjustable robust OPF with renewable energy sources. *IEEE Trans. on Power Syst.*, 28(4):4742-4751, 2013.

[74] A. Lorca and A.X. Sun. Adaptive robust optimization and dynamic uncertainty sets for multi-period economic dispatch with significant wind. *IEEE Trans. on Power Syst.*, 2014. under review.

[75] Y. An and B. Zeng. Exploring the modeling capacity of two-stage robust optimization: Variants of robust unit commitment model. *IEEE Trans. on Power Syst.*, 2014. to appear.

[76] Y. Guan and J. Wang. Uncertainty sets for robust unit commitment. *IEEE Trans. on Power Syst.*, 29(3):1439-1440, 2014.

[77] A. Lorca, A.X. Sun, E. Litvinov, and T. Zheng. Multistage adaptive robust optimization for electric power systems operation. 2014. Submitted.

[78] B. Defourny and L. Wehenkel. Large margin classification with the progressive hedging algorithm. In *Proc. of the Second NIPS Workshop on Optimization for Machine Learning*, Whistler, Canada, 2009.

[79] N. Chiang, C.G. Petra, and V.M. Zavala. Structured nonconvex optimization of large-scale energy systems using PIPS-NLP. In *Proc. of the 18th Power Systems Computation Conference (PSCC)*, Wroclaw, Poland, 2014.

[80] L. Wehenkel. Emergency control and its strategies. In *Proc. of the 13th Power Systems Computation Conference (PSCC)*, Trondheim, Norway, 1999.

[81] R.Y. Rubinstein. *Simulation and the Monte-Carlo Method.* J. Wiley, 1981.

[82] B. Chazelle. *The Discrepancy Method: Randomness and Complexity.* Cambridge University Press, 2000.