CrossMark

EU OR

A comparison of four approaches from stochastic programming for large-scale unit-commitment

Wim van Ackooij[1]

Abstract In energy management, the unit-commitment problem deals with computing the most cost-efficient production schedule that meets customer load, while satisfying the operational constraints of the units. When the problem is large scale and/or much modelling detail is required, decomposition approaches are vital for solving this problem. The recent strong increase in intermittent, relative unforeseeable production has brought forth the need of examining methods from stochastic programming. In this paper we investigate and compare four such methods: probabilistically constrained programming, robust optimization and 2-stage stochastic and robust programming, on several large-scale instances from practice. The results show that the robust optimization approach is computationally the least costly but difficult to parameterize and has the highest recourse cost. The probabilistically constrained approach is second as computational cost is concerned and improves significantly the recourse cost functions with respect to the robust optimization approach. The 2-stage optimization approaches do poorly in terms of robustness, because the recourse decisions can compensate for this. Their total computational cost is highest. This leads to the insight that 2-stage flexibility and robustness can be (practically) orthogonal concepts.

Keywords Probabilistically constrained programming · Robust optimization · Stochastic programming · Unit-commitment

Mathematics Subject Classification 49M37 · 65K05 · 90C15

✉ Wim van Ackooij
   wim.van.ackooij@gmail.com; wim.van-ackooij@edf.fr

[1] EDF R&D, OSIRIS, 1, avenue du Général de Gaulle, 92141 Clamart Cedex, France

🙆 Springer

1 Introduction

Unit-commitment is an optimization problem that deals with finding the most cost-effective production schedule while satisfying the operational constraints of the production units. These units are typically hydro valleys, i.e. series of reservoirs connected through pumps/turbines and thermal units. The obtained production schedule has to be sent to the grid operator one day in advance. Intra-daily changes to this schedule are allowed but need to be communicated to the grid operator at very specific moments in time. Any deviations between production and load will be highly penalized by the grid operator. This strong incentive ensures system safety when generation companies and the grid operator are legally separated units. When uncertainty is neglected, e.g. a forecast based approach is used, unit-commitment can already be challenging for two reasons: the system is large scale and much modelling detail is required for production schedules to be practically meaningful.

The recent increase in renewable generation has significantly increased overall uncertainty in the system. This fact combined with significant theoretical/algorithmic advances in stochastic methods has brought to the foreground stochastic programming approaches in unit-commitment. In this paper we propose to compare four of these approaches on several instances from the large-scale French system, namely an approach based on joint probabilistic constraints, a robust optimization approach and two-stage stochastic and robust approaches. Although, ideally one would like to consider uncertainty on customer load, renewable generation, inflows for hydro reservoirs and unit availability, we will focus on net load (i.e. customer load + renewable generation). We refer to Tahanan et al. (2015) for a recent survey on (stochastic) optimization approaches in unit-commitment. The survey also covers methods from robust optimization.

1.1 A structural viewpoint on (deterministic) unit-commitment

Unit-commitment problems are already challenging in a deterministic setting. The reason for this is that the units are coupled through constraints such as the offer-demand equilibrium constraint and are moreover subject to many complex technical constraints, specific to the their type (thermal, hydraulic, contracts). Frequently, computing an optimal production schedule of a unit, when seen as acting on a price signal, requires specific approaches for an efficient solution.

We consider m units; for any unit indexed by $i = 1, \ldots, m$, we denote its decision variables (including production) $x_i \in \mathbb{R}^{n_i}$, its production cost $f_i(x_i)$ and its specific production constraints $x_i \in X_i$. The decision variable is $x = (x_1, ..., x_m) \in \mathbb{R}^n$ where $\sum_{i=1}^{m} n_i = n$. Units are usually linked through the offer-demand equilibrium constraints that state, in a deterministic setting, that deviation between production and customer load has to remain small. These constraints have the typical form:

$$s^d \leq D - Ax \leq s^u, \tag{1}$$

where $s^d, s^u \in \mathbb{R}^T$ are operator chosen bounds, T is the number of time steps in the considered time horizon, $D \in \mathbb{R}^T$ is the customer load and A is the $T \times n$ matrix

summing the production in the decision vector $x = (x_1, ..., x_m) \in \mathbb{R}^n$. In this paper we will not explicitly consider network constraints, but under a DC-current assumption this would not perturb the linear structure of (1).

An abstract formulation of (deterministic) unit-commitment then has the following form:

$$\min_{x=(x_1,...,x_m)} \sum_{i=1}^{m} f_i(x_i),$$
$$\text{s.t.} \quad x_i \in X_i \subseteq \mathbb{R}^{n_i}, i = 1, ..., m$$
$$s^d \leq D - Ax \leq s^u.$$

Using aggregated objects $f : \mathbb{R}^n \to \mathbb{R}$ as $f(x) = \sum_{i=1}^{m} f_i(x_i)$,

$$X^1 := \prod_{i=1}^{m} X_i \quad \text{and} \quad X^2 := \left\{ y \in \mathbb{R}^n : s^d \leq D - Ax \leq s^u \right\},$$

we can write the above unit-commitment problem in a short manner as

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t.} \quad x \in X^1 \cap X^2. \tag{2}$$

There has been quite some undertaking in formulating (2) as a mixed integer linear program (of very large size). With the advent of strong commercial solvers for mixed integer programming, this has become an important approach (e.g. Carrión and Arroyo 2006; Morales-España et al. 2013a, b). However, the success of such an approach strongly depends on the amount of modelling detail (i.e. the description of the sets X_i) taken into account. We will make the assumption that decomposition approaches are the only possible tools to solve (2). This assumption is driven by the characteristics of the French system, both large scale and requiring substantial modelling details. This hydro-thermal system divides into hydro valleys, a set of connected reservoirs, pumps and turbines, and thermal units, both conventional and nuclear. The hydro valleys may for instance rely on joint probabilistic constraints for taking into account uncertainty on inflows (e.g. van Ackooij et al. 2014) or require a fine description of the power-to-discharge curve (e.g. Finardi and Da Silva 2006). The thermal units are also subject to many additional constraints beyond the model that is standard in the literature (e.g. Tahanan et al. 2015). A popular decomposition approach is Lagrangian decomposition of the coupling constraints, hidden in the set X^2 (see Dubost et al. 2005; Frangioni et al. 2011 and references therein). Maximizing the Lagrangian dual is of interest in its own, since the optimal dual vector is partially used as marginal prices (e.g. Zaourar and Malick 2013). Although, in general one cannot expect to obtain a primal feasible solution, well-established primal recovery heuristics exist and allow us to obtain very good solutions (see Wang et al. 1995; Beltran and Heredia 2002; Dubost et al. 2005; Frangioni et al. 2008; Sagastizábal 2012; Zhuang and Galiana 1988).

1.2 Approaches for stochastic unit-commitment

In order to present the approaches, let us assume that D is unknown in (1). The key modelling choice is how x relates to (partial) observation of D. If we have to decide on x prior to observing D, the inequality system (1) is meaningless. If no changes to x are allowed (within the model), it becomes meaningful to request that (1) holds in a sufficiently large set of situations. A second possibility is that we are allowed to make changes to x at a later stage in time after having observed partially D. This leads to models involving the so-called *recourse decisions*. Indeed, in practice, the solution of problem (2) defines a production schedule x (commitment decisions and power output). This schedule is sent to the grid operator before being activated and most importantly before observing uncertainty. If, in real time, the constraint (1) is violated, a new production schedule, redefining both commitment decisions and power output, can be sent to the grid operator at specific moments in time. The former change can be particularly useful if a later moment in the day might lack production. This implies that the recourse problem is exactly of the same structure as (2) and has the same complexity, but with a smaller time horizon. The modelling choices discussed here are but a few out of many (e.g. Kall and Mayer 2005).

In a first setting we do not model recourse decisions, but would like x to behave nonetheless in a controlled manner. This leads us to replace (1) with

$$\mathbb{P}[s^d \leq D - Ax \leq s^u] \geq p, \tag{3}$$

where $p \in (0, 1)$ is a user defined safety level and $\mathbb{P}$ a probability measure. The constraint (3) is a joint probabilistic constraint. This means that we wish that $s^d \leq D - Ax \leq s^u$ holds with high enough probability for all time steps simultaneously. We refer to van Ackooij (2014) for a thorough discussion of such a unit-commitment model. We also emphasize that this model differs from the individual chance-constrained approaches considered in Ding et al. (2010), Ozturk et al. (2004). Indeed the individual chance constraints on the offer-demand equilibrium with uncertainty on load have an equivalent linear formulation with an additional safety term and lead to insufficient robustness (e.g. van Ackooij et al. 2010, 2014).

This choice leads to the following probabilistically constrained unit-commitment problem:

$$\min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t.} \quad x \in X^1 \cap X^2_{\text{ccp}}, \tag{4}$$

where $X^2_{\text{ccp}} = \{x \in \mathbb{R}^n : \mathbb{P}[s^d \leq D - Ax \leq s^u] \geq p\}$.

A second related approach consists in replacing (1) with

$$s^d \leq D - Ax \leq s^u \ \forall D \in \mathcal{D}, \tag{5}$$

where $\mathcal{D} \subseteq \mathbb{R}^T$ is a user defined *uncertainty set*. This choice corresponds to a model coming from robust optimization (e.g. Ben-Tal et al. 2009).

This choice leads to the following robust unit-commitment problem:

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t.} \quad x \in X^1 \cap X^2_{\text{rob}}, \tag{6}$$

where $X^2_{\text{rob}} = \left\{ x \in \mathbb{R}^n : s^d \leq D - Ax \leq s^u \ \forall D \in \mathcal{D} \right\}$.

A final model involves modelling explicitly the recourse decisions. More precisely, we consider an abstract random process affecting uncertainty on customer load and renewable generation denoted $\xi \in \mathbb{R}^k$. Observing this process at time step $\tau \in \{1, ..., T\}$ results in "observing" the net customer load $D(\xi) \in \mathbb{R}^T$. This load consists of $D(\xi)_1, ..., D(\xi)_\tau$, the actually observed net customer load of the previous time $t = 1, \ldots, \tau$, and $D(\xi)_{\tau+1}, ..., D(\xi)_T$, the current best forecast of net customer load after τ.

We introduce the appropriate modification of X^2 involving the change in D denoted as

$$X^2(\xi) := \left\{ y \in \mathbb{R}^n : s^d \leq D(\xi) - Ay \leq s^u \right\}$$

and the recourse cost function $c : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ as

$$c(x, \xi) := \begin{cases} \min_{y \in \mathbb{R}^n} f(y) & \\ \text{s.t.} \quad y \in X^1 \cap X^2(\xi) \ , \\ \quad\quad Px = Py, \end{cases} \tag{7}$$

where P is a $\ell \times n$ matrix having a single non-zero element for each line and column. The size of ℓ is the number of variables restricted to the first stage. The equation $Px = Py$ models the fact that the power output of each unit $i = 1, ..., m$ prior to τ is fixed and that the recourse decision y can only modify power output after τ. This equation $Px = Py$ can be understood as a non-anticipativity condition. We care to emphasize that y plays exactly the same role as x, and in particular it can contain binary variables.

The expected recourse cost function is then naturally defined as

$$v : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \quad\quad v(x) := \mathbb{E}\left(c(x, \xi)\right).$$

This leads to the following formulation of the two-stage unit-commitment problem:

$$\min_{x \in \mathbb{R}^n} f(x) + v(x)$$
$$\text{s.t.} \quad x \in X^1 \cap X^2. \tag{8}$$

The constraints of problems (4), (6) and (8) are the same as the initial unit-commitment problem (2). That is, we are keen on computing a first stage solution $x \in X^1$. Under the natural assumption that $X^2_{\text{rob}} \subseteq X^2$ and $X^2_{\text{ccp}} \subseteq X^2$, it is moreover immediate that any feasible solution to problems (4), (6) and (8) is feasible for (2).

Note that a set of popular approaches called adaptive robust optimization (or 2-stage robust optimization) can also be closely related to problem (8) above. Indeed this would lead to a problem of the type:

$$\min_{x \in \mathbb{R}^n} f(x) + \max_{D \in \mathcal{D}} c(x, D)$$
$$\text{s.t.} \quad x \in X^1 \cap X^2, \tag{9}$$

where $\mathcal{D}$ is a user defined uncertainty set. Here we have replaced the argument ξ by D to emphasize the "non-random" nature, but the meaning of $c(x, D)$ is immediately understood. Such a methodology is for instance considered in Bertsimas et al. (2013), Zugno and Conejo (2013). The frequent assumption is that binary variables are restricted to the first-stage variables. Then in principle a cutting plane methodology can be deployed as soon as $x \mapsto c(x, D)$ is convex for each D in the uncertainty set $\mathcal{D}$. The difficulty then boils down to being able to identify in an efficient way which of the family of mappings $x \mapsto c(x, D)$ is active at some given $\bar{x} \in X^1$. For a very explicit form of the set X^1 and mapping c, the authors in Zugno and Conejo (2013) suggest two formulations, either by using a bilinear program or a complementarity constrained problem. The latter problem is linearized with the help of binary variables. Alternative approaches considered in Ben-Salem (2011), Minoux (2014) excessively simplify the mapping c, i.e. problem (7) by making the assumption of simple recourse. Then computing the value of the map $x \mapsto \max_{D \in \mathcal{D}} c(x, D)$ and a subgradient at some $\bar{x} \in X^1$ can be done by using a dynamic programming method. In our setting we wish to keep implicit the description of the set X^1 and non-trivial the mapping c so that it would seem that these 2-stage robust approaches need significant work to be extended to this case. Still to have some idea of what such a method may give, we will consider problem (9) wherein $\mathcal{D}$ is a finite set consisting of the same scenarios as used for problem (8). The extension of the methodology for (8) presented below is then straightforward.

1.3 Organization of this document

This paper is organized as follows: Sect. 2 sketches the approaches, assumptions and employed algorithms. Section 3 contains the numerical results and setup of the datasets. The paper ends with conclusions and perspectives. Appendix 1 contains an illustrative example highlighting the differences and similarities between the suggested approaches.

2 Details on the approaches

In this section we provide global details on how we plan to solve problems (4), (6), (8) and (9). The suggested solution approaches involve classic ingredients from unit-commitment that are Lagrangian dualization, Lagrangian-based primal recovery heuristics and bundle methods (e.g. Tahanan et al. 2015; Sagastizábal 2012).

As explained in Sect. 1, X^1 can contain many binary variables, implicit constraints and joint probabilistic constraints. In this document, we do not suppose to know X^1 explicitly; we just make the following assumptions on problems (4), (6) or (8) and (9):

- Practical assumption 1: we can solve (approximatively) the (sub)problems defined as minimizing the sum of f_i and a linear term over X_i:

$$\min_{x_i \in X_i} f_i(x_i) + b_i^\mathsf{T} x_i.$$

- Practical assumption 2: Lagrangian-based primal recovery heuristics (e.g. Feltenmark and Kiwiel 2000; Takriti and Birge 2000; Borghetti et al. 2003) are readily available to build a primal feasible solution out of primal iterates and dual information.
- Theoretical assumption on X^1: each $X_i \subset \mathbb{R}^{n_i}$ is compact. The compactness of X^1 implies that its convex hull conv(X^1) is compact as well (by (Hiriart-Urruty and Lemaréchal 1996, III.1.4.3)). Thus, the sets $X^1 \cap X^2$ and conv $X^1 \cap X^2$ and conv($X^1 \cap X^2$) are compact as well.
- Theoretical assumption on f: each $f_i \colon \mathbb{R}^{n_i} \to \mathbb{R}$ is a closed convex function on $\mathbb{R}^{n_i}$, bounded from below. In view of the first practical assumption above, the f_i should be simple, as piece-wise linear or quadratic. Notice that we can assume without loss of generality that $f_i \geq 0$ upon replacing f_i by $f_i(x_i) - \underline{f}_i$, where $\underline{f}_i$ is the lower bound on f_i.
- Consistency assumption 1: Observe that problem (8) has implicit constraints coming from v, whose domain is

$$\mathcal{D}\mathrm{om}(v) := \{x : v(x) < +\infty\}$$
$$= \left\{x : \text{for all } \xi \in \varXi, \exists\, y \in X^1 \cap X^2(\xi) \text{ such that } Px = Py\right\}.$$

We assume that $\mathcal{D}\mathrm{om}(v)$ is nonempty, so that there exists a solution to (8). Note that following our specific form of the uncertainty set $\mathcal{D}$ in (9), we also have $\{x \in \mathbb{R}^n : \max_{D \in \mathcal{D}} c(x, D) < \infty\} = \mathcal{D}\mathrm{om}(v)$. Consequently there exists a solution to problem (9).
- Consistency assumption 2: In problem (4), the set $X^1 \cap X^2_{\mathrm{ccp}}$ is not empty so that problem (4) admits a solution.
- Consistency assumption 3: In problem (6), the set $X^1 \cap X^2_{\mathrm{rob}}$ is not empty so that problem (6) admits a solution.
- Comparative assumption 1: The uncertainty set $\mathcal{D}$ of (5) is set up in such a way that $\mathbb{P}[D \in \mathcal{D}] \geq p$. This implies $X^2_{\mathrm{rob}} \subseteq X^2_{\mathrm{ccp}}$.
- Comparative assumption 2: The random realizations used in problem (8) are drawn according to their joint distribution used in (3). Throughout this document we will assume that the latter follows a (regular) multivariate Gaussian distribution centred around the currently used deterministic load. This assumption is reasonably realistic (e.g. Bruhns 2005). Although we will not make the assumption in this paper that D follows a discrete distribution with a finite set of realizations, this assumption can be compatible with probabilistically constrained optimization.

Indeed efficient approaches from mixed integer programming have been designed for these situations (e.g. Lejeune 2012; Luedtke 2010, 2014). Note however that adapting the methods presented here to that situation will not be straightforward as the resulting set X^2_{ccp} will be non-convex.

If network constraints are integrated in model (1) under a DC-current assumption, then the multivariate Gaussian random variable D may be singular. In which case the continuous differentiability of the probability function $\mathbb{P}[s^d \leq D - z \leq s^u]$ is altered. We refer to Henrion and Möller (2012), van Ackooij et al. (2015) for more details and a thorough discussion of conditions under which differentiability holds or alternatively a (Clarke)-subgradient can be identified. This technicality does not influence the methods exposed in this paper, and they can carry through verbatim in the case of a singular Gaussian distribution (for D).

In theory, the different methods can define the same set of "first stage" constraints: when $\mathcal{D}\text{om}(v) = X^2_{\text{ccp}} = X^2_{\text{rob}}$. Obviously this is a somewhat utopic request. To begin with, the set X^2_{ccp} is defined by a joint probabilistic constraint, which is only known implicitly and does not admit an analytic representation. Several authors (e.g. Bremer et al. 2015) consider robust optimization as an approximate way to solve chance-constrained problems much like convex (safe) approximations Nemirovski and Shapiro (2006). Fine tuning the conservativeness has led to several works (e.g. Hong et al. 2011; Shan et al. 2014) highlighting the difficulty. The obtained formulations typically result in a feasible set much smaller than X^2_{ccp}. Based on the result Dyer and Frieze (1988) several authors argue that solving joint chance-constrained programming problems is NP-hard, since computing the volume (uniformly distributed random variables) of a polyhedron is NP-hard. We can deduce from this that establishing an explicit equivalent description of X^2_{ccp}, e.g. $X^2_{\text{ccp}} = X^2_{\text{rob}}$ is in general also NP-hard. Similarly, defining a recourse function v in such a way that $\mathcal{D}\text{om}(v) = X^2_{\text{ccp}}$ is also quite a challenging undertaking. Consequently the suggested approaches will differ and their comparison is non-trivial and meaningful.

2.1 Probabilistically constrained unit-commitment

Our first suggestion involves reducing the dimension of the set X^2_{ccp}, which is currently a subset of $\mathbb{R}^n$. In order to do so, we define $\tilde{X}^2_{\text{ccp}} = \left\{ z \in \mathbb{R}^T : \mathbb{P}[s^d \leq D - z \leq s^u] \geq p \right\}$. In this formulation $z \in \mathbb{R}^T$ directly represents the total amount of generated power for each time step. We henceforth reformulate problem (4) as

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \ & f(x) \\
\text{s.t.} \quad & x \in X^1, z \in \tilde{X}^2_{\text{ccp}}, \\
& Ax = z.
\end{aligned}
\tag{10}
$$

The coupling linear constraint $Ax = z$ can be dualized, and the corresponding dual problem is

$$
\max_{\lambda \in \mathbb{R}^T} \Theta_{\text{ccp}}(\lambda),
\tag{11}
$$

where

$$\Theta_{\text{ccp}}(\lambda) = \min_{x \in X^1, z \in \tilde{X}^2_{\text{ccp}}} f(x) + \langle \lambda, Ax - z \rangle. \tag{12}$$

Solving the optimization problem appearing in (12) is not computationally expensive as it decomposes over the m production units due to the special structure of X^1 and function f. One also needs to solve an additional *equilibrium problem*. The mapping $\Theta_{\text{ccp}} : \mathbb{R}^T \to \mathbb{R}$ is concave. For any given $\lambda \in \mathbb{R}^T$, let (x^*, z^*) be the optimal solutions for the optimization problem in (12). Then, we have $Ax^* - z^* \in \partial \Theta_{\text{ccp}}(\lambda)$. Solving problem (11) can thus be carried out by a bundle method (e.g. Bonnans et al. 2006 and references therein). Solving the m subproblems inexactly can immediately be incorporated in this framework by considering recent variants of bundle methods (e.g. de Oliveira et al. 2014 and references therein).

The *equilibrium problem* has the following form:

$$\begin{aligned} \min_{z \in \mathbb{R}^T} \quad & \lambda^\mathsf{T} z \\ s.t. \quad & \mathbb{P}[s^d \leq D - z \leq s^u] \geq p, \\ & z \in Z, \end{aligned} \tag{13}$$

where Z is a polyhedral set containing additional constraints on the total amount of generated power (e.g. bounds). Under the assumption that D follows a multivariate Gaussian random variable, the set $\tilde{X}^2_{\text{ccp}}$ is convex (Prékopa 1995, Theorem 4.2.4). Convexity can also be assured under many other distributions (e.g. Prékopa 2003; Dentcheva 2009). Still the resulting problem (13) is a non-linear convex optimization problem for which specific algorithms have been developed, such as supporting hyperplane methods (e.g. Prékopa 2003) or bundle methods (van Ackooij and Sagastizábal 2014; van Ackooij and de Oliveira 2014). For further details on the probabilistically constrained unit-commitment problem we refer to van Ackooij (2014).

2.1.1 Primal recovery

At convergence of the bundle method solving (11), we do not only dispose of a sequence of iterations (x_ℓ, z_ℓ) produced while evaluating (12) but also of the so-called pseudo-schedule. The latter is an appropriate convex combination of $\{(x_\ell, z_\ell)\}_{\ell \geq 1}$. We will denote this solution by $(\hat{x}, \hat{z})$ and remark that it is the optimal solution of the problem bi-dual to (10). In particular $\hat{z} \in \tilde{X}^2_{\text{ccp}}$ since the latter set is convex. We suggest to interpret $\hat{z}$ as the "load" that needs to be generated and employ the (deterministic) Lagrangian-based primal recovery heuristics where $\hat{z}$ plays the role of the deterministic load.

In the situation wherein D follows a discrete distribution, or wherein X^2_{ccp} is non-convex, we only have $\hat{z} \in \text{conv}(\tilde{X}^2_{\text{ccp}})$. Consequently, for primal recovery, it is needed to identify an appropriate element of the set $\tilde{X}^2_{\text{ccp}}$. For instance by projecting $\hat{z}$ on $\tilde{X}^2_{\text{ccp}}$, i.e. by solving:

$$\min_{z \in Z \subseteq \mathbb{R}^T} \|z - \hat{z}\|^2$$

$$s.t. \quad \mathbb{P}[s^d \leq D - z \leq s^u] \geq p$$

and then using the resulting solution as a substitute for "load" to be produced. Although this will ultimately produce a feasible solution x, the effect on the "optimality" of such a simple procedure is unclear.

2.2 Robust unit-commitment

Quite similarly as for the probabilistic unit-commitment problem, we suggest to operate a dimension reduction first. Hence, we introduce $\tilde{X}^2_{\mathrm{rob}} = \{z \in \mathbb{R}^T : s^d \leq D - z \leq s^u \ \forall D \in \mathcal{D}\}$. We reformulate problem (6) as

$$\min_{x \in \mathbb{R}^n} \ f(x)$$
$$\text{s.t.} \quad x \in X^1, z \in \tilde{X}^2_{\mathrm{rob}},$$
$$Ax = z. \tag{14}$$

The coupling linear constraint $Ax = z$ can be dualized and the corresponding dual problem is

$$\max_{\lambda \in \mathbb{R}^T} \Theta_{\mathrm{rob}}(\lambda), \tag{15}$$

where

$$\Theta_{\mathrm{rob}}(\lambda) = \min_{x \in X^1, z \in \tilde{X}^2_{\mathrm{rob}}} f(x) + \langle \lambda, Ax - z \rangle. \tag{16}$$

2.2.1 Structure and setup of the equilibrium problem

The aspect that still needs to be commented is the solution and setup of the *equilibrium problem*. It has the following form:

$$\min_{z \in \mathbb{R}^T} \ \lambda^{\mathsf{T}} z$$
$$s.t. \quad s^d \leq D - z \leq s^u, \ \forall D \in \mathcal{D}$$
$$z \in Z, \tag{17}$$

where Z is a polyhedral set containing additional constraints on the total amount of generated power (e.g. bounds).

Under the assumption that $D \sim \mathcal{N}(\mu, \Sigma)$ follows a multivariate Gaussian distribution, we can provide an interesting form for the set $\mathcal{D}$. For a centred multivariate Gaussian random variable $\xi \in \mathbb{R}^T$ having covariance matrix Σ, it is well known that its lines of iso-probability are of the form $y^{\mathsf{T}} \Sigma^{-1} y \leq \beta$. It is also readily seen that $\xi^{\mathsf{T}} \Sigma^{-1} \xi \sim \chi^2(T)$, where $\chi^2(T)$ is a 1-dimensional χ-squared distribution with T degrees of freedom. Hence, if β is defined as the p-th quantile of a $\chi^2(T)$ distribution, we have $\mathbb{P}[\xi \in \mathcal{D}] \geq p$, where $\mathcal{D} = \{y \in \mathbb{R}^T : (y - \mu)^{\mathsf{T}} \Sigma^{-1} (y - \mu) \leq \beta\}$. A computation gives $\beta = T + \Phi^{-1}(p)\sqrt{2T}$, where Φ^{-1} is the inverse of the standard normal distribution function (see also van Ackooij et al. 2014). In order to understand the subsequent reasoning, it is important to recall that

$$\min_{x \in \mathbb{R}^n} f(x),$$
$$s.t. \quad Ax \leq b \; \forall b \in \mathcal{D},$$

for a given function $f : \mathbb{R}^n \to \mathbb{R}$, is equivalent with

$$\min_{x \in \mathbb{R}^n} f(x),$$
$$s.t. \quad Ax \leq \underline{b},$$

where $\underline{b} \in \mathbb{R}^m$ is defined as

$$\underline{b}_i := \min_{b \in \mathcal{D}} b_i.$$

The specific form of problem (17) therefore implies that we need to compute $\underline{z}, \overline{z} \in \mathbb{R}^T$ by solving T minimization and maximization problems over the uncertainty set. However, in practice it turns out that $\mathbb{P}[\xi \in [\underline{z}, \overline{z}]] >> p$, when β is defined as above. It is therefore important to integrate a threshold, $\beta_{\mathrm{rob}} \in (0, 1]$ and define

$$\beta = \beta_{\mathrm{rob}}(T + \Phi^{-1}(p)\sqrt{2T}). \tag{18}$$

Problem (17) still appears to have an infinite set of constraints but can actually be reduced to a linear programming problem. For this we need the following auxiliary result:

Lemma 1 *Let $\gamma \in \mathbb{R}^n$ be given and Q be a positive definite $n \times n$ matrix, $\alpha > 0$ a scalar. Then the optimization problem*

$$\min_{x \in \mathbb{R}^n} \gamma^{\mathsf{T}} x$$
$$s.t. \quad \frac{1}{2} x^{\mathsf{T}} Q x \leq \alpha$$

admits the optimal solution $x^* = -\frac{Q^{-1}\gamma}{\lambda^*}$, where $\lambda^* = \sqrt{\frac{1}{2} \frac{\gamma^{\mathsf{T}} Q^{-1} \gamma}{\alpha}}$ is the optimal dual solution and Q^{-1} denotes the inverse of matrix Q.

Proof Let $\lambda \geq 0$ denote the dual multiplier belonging to the convex constraint $\frac{1}{2} x^{\mathsf{T}} Q x \leq \alpha$. Writing down the KKT conditions gives

$$\gamma + \lambda Q x = 0,$$
$$\frac{1}{2} x^{\mathsf{T}} Q x \leq \alpha,$$
$$\lambda \geq 0,$$
$$\lambda \left(\frac{1}{2} x^{\mathsf{T}} Q x - \alpha \right) = 0.$$

The form of the asserted solution x^* follows from the first inequality. Substituting this form in the last inequality provides the value for λ^*, which is readily seen to be non-negative. Substituting x^* in the second equation allows us to deduce primal

feasibility of x^* so that the couple (x^*, λ^*) is indeed a solution to the KKT conditions. Since the constraint $\frac{1}{2}x^\top Q x \leq \alpha$ satisfies the slater condition for $x = 0$, the pair is also optimal.   □

It is now immediate that $z \leq D$ for all $D \in \mathcal{D}$ if and only if $z_i \leq e_i^\top D$ for all $D \in \mathcal{D}$, where e_i is the ith unit vector in $\mathbb{R}^T$. Hence, by solving T optimization problems of the form given in Lemma 1, one can compute $\underline{z} \in \mathbb{R}^T$ such that $\underline{z} \leq D$ for all $D \in \mathcal{D}$. These T problems have γ equal to $e_1, ... e_T$, respectively, $Q = \Sigma^{-1}$ and $\underline{z}_i = \mu_i + x_i^{*,i}$, where $x^{*,i}$ is the optimal solution of the ith problem. Likewise one can compute an upper bound which is $\bar{z}_i = \mu_i - x_i^{*,i}$: it suffices to replace e_i by $-e_i$ in the objective function.

Problem (17) then becomes

$$\begin{aligned}
\min_{z \in \mathbb{R}^T} \quad & \lambda^\top z \\
s.t. \quad & z \leq \underline{z} - s^d \\
& -z \leq s^u - \bar{z} \\
& z \in Z.
\end{aligned} \tag{19}$$

We care to emphasize the importance of setting β_{rob} appropriately. A too high value will make (19) infeasible.

After having solved (15), primal recovery can be conducted as for the probabilistically constrained unit-commitment problem.

2.3 2-Stage unit-commitment

Solving (8) requires 2 steps which can be identified with the 2 decision-making stages. The first step involves replacing v by a cutting plane model $\check{v}_k$ defined as

$$\check{v}_k(x) := \max_{i=1,...,k-1} \{ \langle \bar{g}^i, x - x^i \rangle + \bar{v}^i \} \leq v(x), \tag{20}$$

for values $\bar{v}_i$ and vectors $\bar{g}_i \in \mathbb{R}^n$. The second step deals with improving the cutting plane model $\check{v}_k$. Both steps rely on Lagrangian dualization. The Lagrangian dual of problem (7) with respect to the coupling constraints in $X^2(\xi)$ and linking constraints $Px = Py$ allows us to derive a cutting plane for the mapping v at any given x. To this end let λ_1 be the dual multiplier associated with the constraints $Px = Py$ and λ_2 the multiplier associated with the constraints in the set $X^2(\xi)$. Then, for a given set of realizations $\xi_1, ..., \xi_S$ with associated probabilities $p_1, ..., p_S$, we have

$$\bar{g}^i := P^\top \left(\sum_{j=1}^S p_j \, \lambda_1(x^i, \xi_j) \right) \quad \text{and} \quad \bar{v}^i := \sum_{j=1}^S p_j \theta_{x^i, \xi_j} \big(\lambda_1(x^i, \xi_j), \lambda_2(x^i, \xi_j) \big) \tag{21}$$

for the dual variables $\big(\lambda_1(x^i, \xi_j), \lambda_2(x^i, \xi_j) \big)$. Here $\theta_{x,\xi}$ denotes the Lagrangian dual to problem (7). Note that, we decide to aggregate cuts as above to simplify pre-

sentation. As usual in stochastic programming (see e.g. Ruszczyński 2003), these cuts could also be combined in other ways; see for example the so-called multi-cut of Birge and Louveaux (1988), Birge and Louveaux (1997), which is a disaggregate cutting plane model for v. Other variants consist of partially aggregating cuts as explained in Xiong and Jirutitijaroen (2011). A second feature worth mentioning is that maximizing the Lagrangian dual $\theta_{x,\xi}$ can be hot-started since $X^2(\xi)$ only constrains right-hand side uncertainty and $Px = Py$ is linear. Consequently the solution of the subproblems of this dual are independent of (x, ξ) and one can recycle a previous cutting plane model for $\theta_{x,\xi}$ when (x, ξ) changes to (x', ξ') with no additional cost. When dealing with problem (9), the only modification needed is to figure out which scenario ξ^* is active and redefine $\bar{g}^i$, $\bar{v}^i$ in (21) by setting $\bar{g}^i = P^\mathsf{T}\lambda_1(x^i, \xi^*)$, $\bar{v}^i = \theta_{x^i,\xi^*}(\lambda_1(x^i, \xi^*), \lambda_2(x^i, \xi^*))$. When the uncertainty set $\mathcal{D}$ consists of a finite, not very large, collection of scenarios, one readily identifies the active scenario ξ^*.

The approximated first-stage problem

$$
\begin{aligned}
\min\, & f(x) + \check{v}_k(x), \\
& x \in X^1 \cap X^2,
\end{aligned}
\tag{22}
$$

can be reformulated as

$$
\begin{aligned}
\min_{(x,v)}\, & f(x) + v, \\
\text{s.t.}\quad & (\bar{g}^i)^\mathsf{T}x + \bar{v}^i \leq v, \quad i = 1, \ldots, k-1 \\
& x \in X^1 \cap X^2.
\end{aligned}
\tag{23}
$$

Once again we suggest to consider the Lagrangian dual of problem (23) involving multipliers for the constraints in the set X^2 and $k - 1$ multipliers belonging to the cuts in the model $\check{v}_k(x)$.

The algorithm now behaves as follows. First we maximize the Lagrangian dual to problem (22). This should be followed by a Lagrangian-based primal recovery heuristic for recovering a feasible solution x^k to problem (22). We define the observed duality gap $\Delta_G^k := f(x^k) + \check{v}_k(x^k) - \Theta_k(\mu^k, v^k)$. Here Θ_k is the dual function to problem (22) with dual multipliers μ^k (to the cutting plane constraints) and v^k (to the constraints in X^2). The solution x^k is then processed by a step enriching the model for v, i.e. solving the S dual problems to (7). Once this step is completed, we define the approximation error $\Delta_A^k := \check{v}_{k+1}(x^k) - \check{v}_k(x^k)$. Whenever Δ_A^k is small enough, we stop the algorithm. Note that, under usual assumptions in Bender's type decomposition, we can establish that the final iterate x^k is a $\Delta_A^k + \Delta_G^k$-optimal solution to problem (8) (problem (9), respectively) wherein v is replaced with an appropriately convexified version of it (not the convex hull). This convexified mapping is automatically computed by having moved to the Lagrangian dual. For full details on the approach, we refer the reader to van Ackooij et al. (2014).

3 Numerical experiments

The approaches given in Sect. 2 do not rely on the specific form of the set X^1. They, however, all require at some stage, a primal recovery step. One of the general purpose
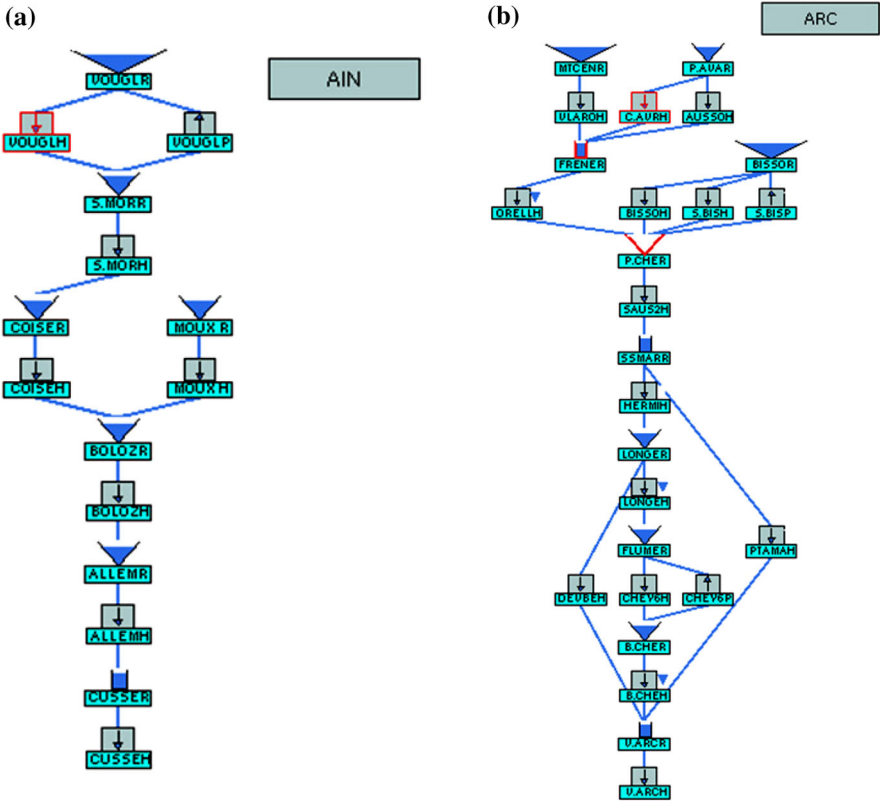
(a)

(b)

Fig. 1 An example of two hydro valleys. **a** Ain valley. **b** Arc valley

heuristics is the one given in Takriti and Birge (2000) that does not rely on the specific form of X^1. In general, the primal recovery heuristics do exploit specific knowledge of the set X^1. For this reason we will restrict our model to standard choices in the unit-commitment literature, i.e. a convex model for hydro valleys (e.g. van Ackooij et al. 2014) and a standard non-convex model for thermal units (e.g. Frangioni et al. 2011). The data are taken from actual datasets from the French system. We will consider a total of 96 half hourly time steps. The total number of variables describing the set X^1 is around 50000 continuous variables, 27000 binary variables and uses around 815000 constraints. The 2-stage approach will consider 50 scenarios (see Figure 2 for an example), and both stages deal with the set of technically feasible production schedules X^1. This means that part of the commitment decisions can be changed in the second stage. Figure 1 shows two typical hydro valleys.

3.1 Description of the primal recovery heuristics

In our numerical experiments, we use a heuristic inspired by Borghetti et al. (2003). The heuristic uses information returned by the bundle algorithm maximizing (11),

(15) or the dual to (22). More precisely, denoting by p the number of iterations of this algorithm and x^j a primal iterate obtained at iteration $j \in \{1, \ldots, p\}$, the heuristic uses the following quantities:

1. the dual simplicial multipliers α of the last quadratic program that the bundle method solved.
2. the so-called pseudo-schedule $\hat{x}$ defined as $\sum_{j=1}^{p} \alpha_j x^j$, (see Daniildis and Lemaréchal (2005); Dubost et al. (2005));
3. the pseudo-costs $(\hat{c}_1, ..., \hat{c}_m)$ defined as $\hat{c}_i = \sum_{j=1}^{p} \alpha_j c_i^j$ where c_i^j is the pure production cost (i.e. not involving any Lagrange multiplier valued generation) of subproblem i at iteration $j = 1, ..., p$;
4. the pseudo-commitment decisions defined as $\hat{u}_i^j = \sum_{j=1}^{p} \alpha_j u_i^j$, where $u_i^j \in \{0, 1\}^T$ are the commitment decisions of each thermal plant for each iteration $j = 1, ..., p$.

Another common ingredient is the resolution of an economic dispatch problem: for a fixed set of commitment decisions, we let a continuous optimization problem adjust production levels in order to generate a solution in X^2. We begin by remarking that the pseudo-schedule is a technically feasible solution as hydro valleys are concerned (since these subproblems have convex feasible sets). Also the pseudo-schedule is directly related to offer-demand equilibrium constraints through the bundle stopping criterium. We therefore keep the pseudo-schedule as hydro valleys are concerned and remove their generation from the load D in order to form $\tilde{D}$. The set of generations assets is such that the obtained net load is always strictly positive. The heuristic is therefore mostly concerned with thermal plants.

We begin with an initial guess for the commitment decisions called $\tilde{u}$, for instance one of the commitment decisions encountered during the optimization of problems (11), (15) or the dual to (22). We now build a time-independent priority list. This list is related to sorting the pseudo-costs divided by total generated pseudo-power in increasing order. A lower value indicates a unit with higher priority (best cost to power ratio). Note that this idea is directly inspired from the heuristic presented in Borghetti et al. (2003) wherein a time-dependent priority list is set up, by dividing the pseudo-commitment decisions by the above pseudo-cost over pseudo-power ratio. A higher value indicates a unit more likely to be started. We have experimented both heuristics and found little difference. Hence, we restrict our presentation to the simpler method.

Starting from our initial commitment guess $\tilde{u}$, we first begin by computing the generation envelope, i.e. the minimum and maximum power the plants can generate over the whole time horizon at these commitment decisions. We now move from the first time step to the last one, if $\tilde{D}$ is in the generation envelope, nothing more needs to be done. If generation is insufficient, we check if we can start the highest priority unit (if not done so already), we continue in this manner until generation covers load. If generation is in excess, we try to decommit the lowest priority unit (if not already off) and continue in this manner until the minimum generation is below load. The hence-generated commitment decision is post-processed with an economic dispatch in order to finely adjust generation to actual load. In this manner, we post-process any of the generated commitment decisions u^j, $j = 1, ..., p$ in order to retain the best one.

3.2 Computational results

We have tested the four approaches of Sect. 2 on a total of 6 datasets covering roughly the year. Table 1 provides an identifier to each dataset:

We now provide for each method information related to the performance, gaps and validation criteria. We also provide information on CPU time that should be taken as an indication only. Various algorithmic improvements such as parallelization have not been used, but could obviously be so, both while processing the subproblems and in the algorithmic interior of the methods.

In Table 2, we report the results found when using the method of Sect. 2.2 on the 6 instances. We have empirically set $\beta_{\mathrm{rob}} = 0.025$ in order to avoid being to demanding as robustness is concerned. Nonetheless, one of the data sets is found to be infeasible. This shows the difficulty in appropriately setting this parameter. We have also evaluated the value of the probability appearing in (3) as an indication of found feasibility (by using Genz' code Genz 1992; Genz and Bretz 2009). Finally the number of dual iterations is provided as well. The values $\frac{1}{2}v^\top \Sigma^{-1} v$ for $v = s^d + Ax - \mathbb{E}(D)$ and $v = s^u + Ax - \mathbb{E}(D)$ have also been evaluated but not found to be meaningfully interpretable. Notice that in principle these values correspond to v belonging to the ellipsoidal uncertainty set for D, which in practice they do not, i.e. we have obtained conservative solutions. In order to see the effect of the parameter β_{rob}, we have varied it slightly from 0.03 to 0.05. The results are given in Tables 3 and 4. We can observe that moving the parameter from 0.03 to 0.031 results in an additional infeasible instance.

Table 1 Instance label and corresponding date, the datasets contain 96 half hourly time steps

Instance	Date	# Thermal units	# Hydro valleys
1	15/01/2013	106	41
2	20/03/2013	106	32
3	13/05/2013	104	35
4	22/08/2013	104	34
5	25/10/2013	104	37
6	10/12/2013	104	40

Table 2 Characteristics of the solutions with the robust optimization approach of Sect. 2.2

Instance	Dual value	Primal value	Gap %	$\mathbb{P}$	# It.	CPU (s)
1	$2.77733e7$	$2.78751e7$	0.36	0.89	46	85
2	$1.10651e7$	$1.10991e7$	0.31	0.91	67	386
3	$6.20137e6$	$6.22951e6$	0.45	0.89	47	215
4	–	–	–	–	–	–
5	$4.03614e6$	$4.06488e6$	0.71	0.87	55	186
6	$1.84763e7$	$1.87301e7$	1.37	0.89	117	197

Here β_{rob} was set to 0.025

Table 3 Characteristics of the solutions with the robust optimization approach of Sect. 2.2

Instance	Dual value	Primal value	Gap %	$\mathbb{P}$	# It.	CPU (s)
1	2.78600e7	2.79594e7	0.36	0.94	48	118
2	1.10913e7	1.11251e7	0.30	0.92	71	469
3	6.22056e6	6.24777e6	0.44	0.91	33	206
4	–	–	–	–	–	–
5	–	–	–	–	–	–
6	1.85176e7	1.86566e7	0.75	0.91	123	287

Here β_{rob} was set to 0.03

Table 4 Characteristics of the solutions with the robust optimization approach of Sect. 2.2

Instance	Dual value	Primal value	Gap %	$\mathbb{P}$	# It.	CPU (s)
1	2.78768e7	2.79762e7	0.36	0.94	45	110
2	1.10965e7	1.11321e7	0.32	0.92	72	488
3	–	–	–	–	–	–
4	–	–	–	–	–	–
5	–	–	–	–	–	–
6	1.85260e7	1.87150e7	1.02	0.92	162	363

Here β_{rob} was set to 0.031

Table 5 Characteristics of the solutions with the probabilistically constrained optimization approach of Sect. 2.1

Instance	Dual value	Primal value	Gap %	$\mathbb{P}$	# It.	CPU (s)	Dual-1 %
1	2.78891e7	2.79614e7	0.26	0.82	144	5813	2.76102e7
2	1.10811e7	1.11138e7	0.30	0.88	53	15503	1.09703e7
3	6.26619e6	6.30719e6	0.65	0.84	179	12309	6.20350e6
4	3.22850e6	3.26396e6	1.10	0.82	185	26012	3.19622e6
5	4.07367e6	4.11380e6	0.99	0.82	198	20940	4.03290e6
6	1.85051e7	1.88178e7	1.69	0.85	171	11510	1.83200e6

Here p was set to 0.8

When the parameter is set to 0.05 all instances were found to be infeasible. The results also show that the dual/primal values can only be precise within the given oracle error (here at 0.05 %).

Table 5 reports the results found when using the method of Sect. 2.1. Even though the probability level was set to 0.8, and this is indeed the feasibility level obtained as the "equilibrium" pseudo-solution is concerned, primal recovery slightly increases the actually found probability level. The strongly increased CPU times with respect to those reported in Table 2 can be attributed to the solution of the equilibrium sub-problem: a joint probabilistically constraint problem with random vector in dimension 96. Nonetheless, due to hot-starting of the supporting hyperplane method, only early
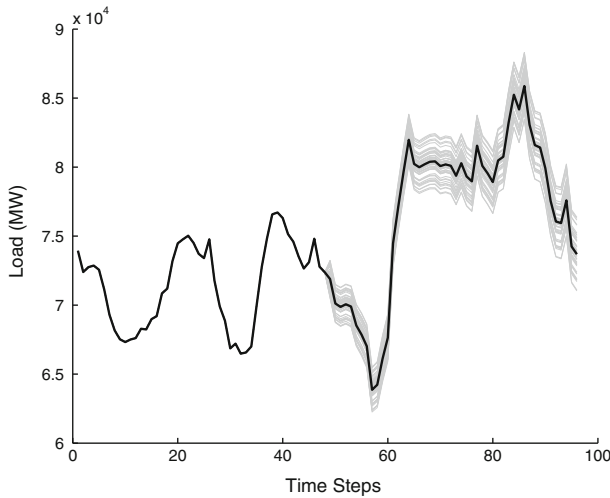
Fig. 2 An illustration of 50 generated scenarios for the 2-stage approach

dual iterations are costly and require several iterations of this method. Algorithmic improvements such as on-demand accuracy (de Oliveira and Sagastizábal 2014) and bundle methods (van Ackooij and de Oliveira 2014; van Ackooij and Sagastizábal 2014) specially designed for probabilistically constrained optimization problems may provide further algorithmic improvements bringing significantly down this algorithmic burden. An issue worth commenting is the fact that the dual bounds provided in Table 5 are not below those found in Table 2 as one would expect. The reason for this is two-fold:

1. the original uncertainty set $\mathcal{D}$ for D is set up in such a way as to satisfy $\mathbb{P}[D \in \mathcal{D}] \geq p$. Consequently any feasible solution of the robust problem is feasible for the probabilistically constrained model. However, we shrunk $\mathcal{D}$ in order to be less demanding. In practice, $\mathbb{P}[\underline{z} \leq D \leq \overline{z}] \approx 0.64$, where $\underline{z}, \overline{z}$ are as in (19). This implies that several dual iterations may provide a "static solution", i.e. solution to problem (19) infeasible for the probabilistic constraint. This effect is illustrated in Tables 3 and 4.

2. the subproblems during the dual iterations are solved inexactly. The thermal subproblems up to 0.5 % gap and the probabilistically constrained subproblem up to 1 % gap. Consequently the dual values are only precise up to 1 % as well. This is shown in the last column of Table 5. This imprecision on the dual values also has an effect on the primal recovery.

We provide the results for the full 2-stage stochastic approach explained in Sect. 2.3 in Table 6. One can observe that a total of roughly 2500 oracle calls (subproblem) resolutions are needed for obtaining an optimality threshold of around 0.05 %. This is quite reasonable, when one recalls that we are solving a problem of the size of 510 full large-scale unit-commitment problems. If each unit-commitment problem would require between 50 and 100 dual iterations, this would require a total of 25,500–51,000 oracle calls, about 10 times more. The primal recovery heuristic provides

Table 6 Characteristics of the solutions with the 2-stage optimization approach of Sect. 2.3 using 50 scenarios

Instance	Primal value	Duality gap %	Stop gap %	$\mathbb{P}$	# It.	# Or. (1st).	# Or. (2nd).	CPU (s)
1	2.76685e7	2.68	0.01	0.16	22	968	2637	26763
2	1.08665e7	3.43	0.05	0.15	11	589	1241	32337
3	6.01728e6	1.20	0.01	0.15	11	441	993	16046
4	3.02334e6	2.84	0.05	0.15	16	656	1961	37380
5	3.84513e6	2.70	0.05	0.16	13	713	1201	16669
6	1.85152e7	3.30	0.02	0.16	16	1035	2392	28430

Stopping criteria $\delta_{\mathtt{tol}} = 0.05$ %

Table 7 Probability level obtained when producing exactly the average load

Instance	$\mathbb{P}$
1	0.3295
2	0.3249
3	0.3667
4	0.3242
5	0.3157
6	0.3364

solutions with gaps of around 2 %. The obtained probability levels are lower than what one would obtain when producing a schedule satisfying exactly the average load. In that case one would obtain the probability levels given in Table 7. This can be explained by the availability of additional recourse actions which allow us to re-adjust the production levels after having observed partially the load. Due to the existence of temporal dependencies, this observation provides useful information about the future. The obtained costs for the reference schedule (i.e. x) are also slightly lower than those for the probabilistically constrained and robust optimization solution, which again can be explained by the availability of recourse decisions, which are not present in the former models.

The results for the full 2-stage robust optimization approach are given in Table 8. One can observe that the computational burden is rather similar to the 2-stage stochastic approach, as can be expected. Indeed the employed algorithm differs only in the subgradient and function value computed during the model enrichment step explained in Sect. 2.3. Rather surprisingly, the obtained solutions are even less robust than those that one would obtain when using a 2-stage stochastic optimization approach. The obtained costs related to the first-stage decision x are also above those obtained with the probabilistically constrained solution.

Finally, Tables 9 and 10 provide the in- and out-sample recourse cost evaluated on each of the obtained solutions. The in-sample evaluation is related to the actually seen second-stage objective function. We have not presented the evaluation of the 2-stage robust solution in the expected recourse cost function as this solution did worse than

Table 8 Characteristics of the solutions with the 2-stage robust approach (9) using an uncertainty set consisting of 50 scenarios

Instance	Primal value	Duality gap %	Stop gap %	$\mathbb{P}$	# It.	# Or. (1st).	# Or. (2nd).	CPU (s)
1	$2.79846e7$	0.98	0.01	0.085	6	230	692	6176
2	$1.12717e7$	1.78	0.05	0.085	12	588	1305	32160
3	$6.19538e6$	1.32	0.05	0.084	7	329	682	12934
4	$3.25917e6$	10.6	0.01	0.084	12	498	1709	36424
5	$4.13115e6$	1.94	0.05	0.083	16	627	1489	23774
6	$1.91428e7$	2.16	0.01	0.084	8	417	1221	16043

Stopping criteria $\delta_{\texttt{tol}} = 0.05$ %

Table 9 Recourse costs obtained with each approach

Instance	Expected recourse cost			Worst-case recourse cost		
	2-stage stochastic	ccp	rob	2-stage robust	ccp	rob
1	$2.71700e7$	$2.91192e7$	$4.37984e7$	$2.85564e7$	$3.01098e7$	$4.49801e7$
2	$1.08317e7$	$1.49400e7$	$1.54511e7$	$1.14763e7$	$1.53260e7$	$1.58225e7$
3	$6.00571e6$	$6.22994e6$	$6.29292e6$	$6.35631e6$	$6.47817e6$	$6.47230e6$
4	$3.01836e6$	$3.41765e6$	–	$3.37643e6$	$3.54545e6$	–
5	$3.82574e6$	$4.27767e6$	$4.26898e6$	$4.29860e6$	$4.45282e6$	$4.44699e6$
6	$1.81980e7$	$3.72008e7$	$3.95870e7$	$1.92693e7$	$3.82183e7$	$4.04473e7$

This table involves the actually used recourse function
We report both the worst-case and expected recourse cost

that obtained from the 2-stage stochastic approach. The similar symmetric observation was also made. Note that Tables 9 and 10 show that the 2-stage solutions provides the lower recourse cost (expected or worst case), which obviously makes sense since this is (partially) part of the optimization criterium for this approach. Both the probabilistically constrained and robust optimization solution provide a higher recourse cost. However, with a significant advantage for the probabilistically constrained solution which seems to preserve more flexibility than the robust optimization approach. We have also observed that if one computes the 2-stage solutions with a too small sample size, the out-sample costs become very high. The 2-stage stochastic solution was found to be more sensitive. For the same "insufficient" optimization sample size (here 40 samples), the 2-stage robust optimization solution still obtained correct out-sample values (both expected and worst case), whereas the stochastic solution performed worst than all other methods.

3.3 On scenario selection

From a Monte-Carlo viewpoint, considering only 50 scenarios may seem inappropriate. Yet, this corresponds to the size of interest in practice. Météo France provides

Table 10 Recourse costs obtained with each approach

Instance	Expected recourse cost				Worst-case recourse cost			
	2-stage stochastic	2-stage robust	ccp	rob	2-stage stochastic	2-stage robust	ccp	rob
1	2.78307e7	2.77605e7	3.13328e7	4.53014e7	2.90045e7	2.89349e7	3.44846e7	4.67034e7
2	1.12408e7	1.12422e7	1.62822e7	1.72722e7	1.16143e7	1.16148e7	1.79119e7	1.96120e7
3	6.18286e6	6.18833e6	6.49266e6	6.54154e6	6.43975e6	6.43354e6	6.91366e6	7.01332e6
4	3.24865e6	3.24936e6	3.59475e6	–	3.44964e6	3.44965e6	3.91913e6	–
5	4.11590e6	4.11673e6	4.42744e6	4.49746e6	4.39576e6	4.39379e6	4.82023e6	4.94401e6
6	1.89607e7	1.89757e7	3.86212e7	4.08404e7	1.96978e7	1.96175e7	3.95812e7	4.12288e7

This table involves an a posteriori evaluation with 500 scenarios
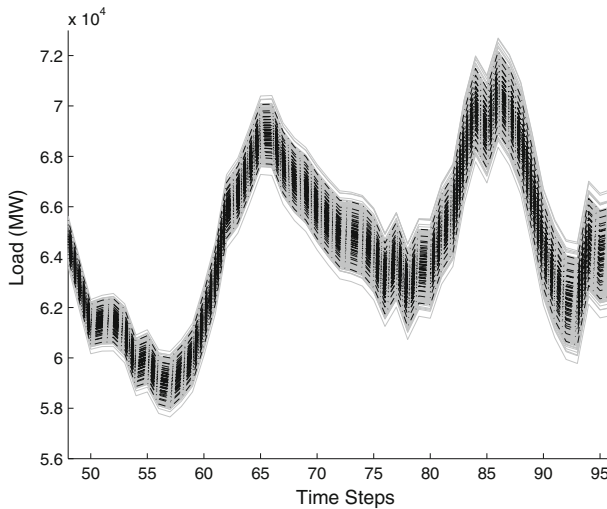We report both the worst-case and expected recourse cost

Fig. 3 An illustration of 50 quantized scenarios (*black-dashed lines*) representative of 50,000 (only 5000 shown, *grey lines*)

EDF with a set of 51 scenarios of temperature that can be used to derive 51 likely scenarios of load. Being able to exploit this size in number of scenarios would therefore already be of interest. Secondly, both the first stage and second stage are full large-scale unit-commitment problems. This also limits the total number of scenarios quite below what is common in Monte-Carlo "computations". In order to capture the essential of the distribution, we suggest to employ methods from quantization Pagès and Printems (2003). We have therefore generated a total of 50,000 load scenarios following the underlying distribution and employed a variant of Lloyd's algorithm Lloyd (1982), Gersho and Gray (1992) implemented in Matlab's VQDtool to obtain a selection of 50 representative scenarios. Figure 3 shows how the obtained scenarios overlay with the original set.

Using this set of 50 quantized scenarios, we have solved problems (8) and (9) to obtain alternative (with respect to 50 randomly drawn scenarios) 2-stage solutions. The computational results have been found to differ little from the results reported in Tables 6 and 8 except for the a priori robustness obtained with the 2-stage stochastic solution that now aligns more with the level obtained from the 2-stage robust solution (i.e. levels of around 0.08 are observed). We have performed an in- and out-sample evaluation of the respective recourse functions, the results of which are presented in Table 11. It is very interesting to see that the in- and out-sample values almost perfectly match. This observation holds both for the worst-case and expected recourse cost functions. This contrasts when comparing the results presented in Tables 9 and 10 where the out-sample costs are seen to be significantly higher than the in-sample costs. This argues in favour of a scenario preprocessing step in the form of quantization to keep the overall number of scenarios low, while still capturing many features of the underlying distribution. When comparing the "out-sample" recourse values for the probabilistically constrained and robust optimization solution presented in Table

Table 11 Recourse costs obtained with quantized scenario selection by the 2-stage approaches

Instance	Expected recourse cost				Worst-case recourse cost			
	In-sample		Out-sample		In-sample		Out-sample	
	Stochastic	Robust	Stochastic	Robust	Stochastic	Robust	Stochastic	Robust
1	$2.77885e7$	$2.78242e7$	$2.78200e7$	$2.78452e7$	$2.91021e7$	$2.91294e7$	$2.90006e7$	$2.90201e7$
2	$1.13300e7$	$1.13310e7$	$1.12759e7$	$1.12779e7$	$1.17699e7$	$1.17493e7$	$1.16607e7$	$1.16594e7$
3	$6.26034e6$	$6.26408e6$	$6.17362e6$	$6.17906e6$	$6.58332e6$	$6.58354e6$	$6.43046e6$	$6.43549e6$
4	$3.21644e6$	$3.21423e6$	$3.20531e6$	$3.20505e6$	$3.43253e6$	$3.42720e6$	$3.40612e6$	$3.40632e6$
5	$4.12491e6$	$4.12518e6$	$4.11600e6$	$4.11591e6$	$4.44640e6$	$4.44592e6$	$4.39937e6$	$4.39914e6$
6	$1.90848e7$	$1.90467e7$	$1.89863e7$	$1.89577e7$	$1.99591e7$	$1.98533e7$	$1.96942e7$	$1.95975e7$

We report both the worst-case and expected recourse cost for an in-sample (using the 50 quantized scenarios) and out-sample (using 500 randomly drawn samples)

10 with the values presented in Table 11, one observes that the earlier drawn conclusions regarding the classification are not altered by the quantization procedure (which only influences the 2-stage methods as they are the only methods that are scenario dependent).

4 Conclusions and perspectives

We have investigated four different approaches for taking uncertainty into account in the unit-commitment problem: robust optimization, probabilistically constrained optimization and 2-stage stochastic and robust approaches.

These approaches can be classed as follows when looking at computation times: robust optimization, probabilistically constrained optimization and 2-stage methods. However, it is important to highlight that the robust optimization approach is difficult to set up. Indeed, we have empirically fine tuned the uncertainty set in order to have a reasonable solution, yet this parametrization did not work on one of the datasets.

When examining the a priori robustness the approaches can be classified as follows: 2-stage methods, probabilistically constrained optimization and robust optimization. Even though the uncertainty set for robust optimization was tuned in such a way as to make these solutions comparable with the ones resulting from the probabilistically constrained method, in practice we have observed significant over-robustness. The 2-stage solution provides even less a priori robustness than the solution meeting up with average load. This shows that the potential for recourse is very important. Indeed it allows one to relax the constraints (on robustness) because we observe uncertainty in a later stage. If we take the recourse cost function as a measure of flexibility, the methods are classed as 2-stage method, probabilistically constrained optimization and robust optimization. These conclusions lead to the insight that 2-stage flexibility and robustness can be (practically) orthogonal concepts.

We can therefore conclude that the probabilistically constrained solution allows us to preserve an equilibrium between a priori robustness, computation times and flexibility. Obviously the 2-stage approach also requires further investigation. One could for instance add a request of a priori robustness in the second stage. Algorithmic improvements also need to be investigated in order to smooth up the resolution of these approaches. Early tests with a regularized supporting hyperplane method suggest that the resolution times of the probabilistically constrained approach can be reduced on average by 33% and sometimes even 66 %. Likewise in the 2-stage approach first-stage regularization should bring down the number of "first stage" oracle calls and the number of total iterations. We also suggest considering further improvements to the 2-stage approach, on the one hand, by integrating other risk-related criteria, on the other hand, by adding additional recourse stages.

Appendix: Didactic toy example

In this appendix we present a small example that may help the reader visualize the difference and similarities between the suggested models.

The set of assets consists of two thermal plants, subject to power limits and ramping rates. There are only two stages and time steps in the problem and only the load in the second time step, i.e. second stage, is uncertain (see Table 12 for further data).

The deterministic load is equal to $d = (0.75, 0.75)$, and the second-stage load varies uniformly in the interval $(0.25, 1.25)$. The bounds in (1) are $s^d = (0.0, -0.1)$, $s^u = (0.0, 0.5)$. In what follows $p^1 = (p_1^1, p_2^1)$ will denote the production level of the first unit and $p^2 = (p_1^2, p_2^2)$ that of the second.

The deterministic approach

One can readily observe that the optimal solution is to use unit 1 only and produce $(0.75, 0.5)$ for a total cost of 13.75. Note that the minimal stable generation of 0.5 prevents this unit from decreasing its production in the second stage further. Unit 2 remains offline.

The 2-stage approach

The situation is now rather different and partially depends on the possibility for us to switch off unit 1 in the second time step, when d_2 has been observed. If this can be done, unit 1 can still be online in the first time step. Then a feasible strategy is one wherein unit 1 produces 0.75 in the first time step and

1. if $d_2 < 0.4$, shutdown unit 1 and use unit 2 to generate 0.25 MW.
2. if $d_2 \in [0.4, 1.0]$, unit 1 produces at MSG in the second time step
3. if $d_2 \in (1.0, 1.25]$, unit 1 produces $d_2 - 0.5$ in the second time step.

One can readily compute that this solution leads to an expected cost of 15.3063. A second feasible solution is to only use unit 2. In this situation unit 2 produces 0.75 in the first time step and

Table 12 Characteristics of the toy example

Feature	Unit 1	Unit 2
Cost/MW	3	5
Minimal stable generation (MSG) (MW)	0.5	0.25
Full load (MW)	1.25	1.25
Ramp rates (MW/h)	1	1
Startup cost	10	10

For simplicity hourly time steps are assumed

1. if $d_2 < 0.75$, unit 2 generates 0.25 MW in the second time step.
2. if $d_2 \in [0.75, 1.25]$, unit 2 produces $d_2 - 0.5$ in the second time step.

The expected cost is 15.6250, making this a suboptimal solution.

If unit 1 has to remain online in the second stage, the first strategy described above leads to an infeasible second-stage problem for several random realizations. Consequently only the second strategy remains feasible and hence optimal.

The probabilistically constrained approach

In this situation the probability constraint (3) becomes $\mathbb{P}[d_2 - 0.5 \leq p_2^1 + p_2^2 \leq d_2 + 0.1] \geq p$. For a given probability level of say $p = 0.55$, we can observe that the probability request can be alternatively cast as the following linear constraint $0.3 \leq p_2^1 + p_2^2 \leq 0.8$. In this particular situation, the probabilistically constrained model is more demanding than the deterministic model, which requests $0.25 \leq p_2^1 + p_2^2 \leq 0.85$, but the difference is not seen on the optimal solution due to the MSG restriction on the first unit. The optimal solution is therefore to use unit 1 to generate $p^1 = (0.75, 0.5)$ and keep unit 2 offline.

The robust constrained approach

An uncertainty set $\mathcal{D}$ consistent with the above probabilistically constrained model would be $\mathcal{D} = \left\{ d \in R^2 : d_1 = 0.75, d_2 \in [\bar{d}, \bar{d} + 0.55] \right\}$. With such an uncertainty set, the constraint (5) related to the second time step reads $\bar{d} + 0.05 \leq p_2^1 + p_2^2 \leq \bar{d} + 0.1$.

This leads to the following set of solutions:

1. For $\bar{d} \in [0.25, 0.4)$, use unit 2 to produce 0.75 in the first time step and $\bar{d} + 0.05$ in the second time step. The cost of this solution lies in the interval $[15.25, 16)$.
2. For $\bar{d} \in [0.4, 0.45]$, use unit 1 to produce 0.75 in the first time step and 0.5 in the second time step. The cost of this solution is 13.75.
3. For $\bar{d} \in (0.45, 0.7]$, use unit 1 to produce 0.75 in the first time step and $\bar{d} + 0.05$ in the second time step. The total cost lies in the interval $[13.75, 14.50]$.

The 2-stage robust approach

Here the situation depends only on the given uncertainty set $\mathcal{D}$ and not on the flexibility of unit 1 as in the stochastic 2-stage approach.

Indeed, assuming that the first unit can be switched off in the second stage, then as soon as there exists $d \in \mathcal{D}$ with $d_2 < 0.4$, the worst-case cost of strategy 1 exhibited in the 2-stage stochastic approach is highly costly as it involves two starting costs (it is 23.5). Consequently only the second solution involving only unit 2 is retained. If unit 1 cannot be switched off, the solution strategy involving unit 1 is not even feasible.

However, as long as any $d \in \mathcal{D}$ has $d_2 \geq 0.4$ the first strategy is optimal, independently of whether or not unit 1 can be switched off or not.

A remark on comparing 2-stage and mono-stage approaches

Following the framework laid out in (8) both 2-stage approaches compute an optimal strategy that depends on various problem characteristics, but also a first-stage vision of the future (called x in (7)). For, the strategy involving unit 1 this vision would be $p^1 = (0.75, 0.5)$, $p^2 = (0, 0)$. The strategy involving unit 2 would provide a first-stage vision $p^1 = (0, 0)$, $p^2 = (0.75, 0.25)$. Note that this first-stage vision can be compared with the solutions obtained from the mono-stage approaches and vice versa.

References

Beltran C, Heredia FJ (2002) Unit commitment by augmented lagrangian relaxation: testing two decomposition approaches. J Optim Theory Appl 112(2):295–314

Ben-Salem S (2011) Gestion Robuste de la production électrique à horizon court-terme. PhD thesis, Ecole Centrale Paris, Mars

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) Robust optimization. Princeton University Press, Princeton

Bertsimas D, Litvinov E, Sun XA, Zhao J, Zheng T (2013) Adaptive robust optimization for the security constrained unit commitment problem. IEEE Trans Power Syst 28(1):52–63

Birge JR, Louveaux F (1988) A multicut algorithm for two-stage stochastic linear programs. Eur J Oper Res 34(3):384–392

Birge JR, Louveaux F (1997) Introduction to stochastic programming. Springer, New York

Bonnans JF, Gilbert JC, Lemaréchal C, Sagastizábal C (2006) Numerical optimization: theoretical and practical aspects, 2nd edn. Springer, Berlin

Borghetti A, Frangioni A, Lacalandra F, Nucci CA (2003) Lagrangian heuristics based on disaggregated bundle methods for hydrothermal unit commitment. IEEE Trans Power Syst 18:313–323

Bremer I, Henrion R, Möller A (2015) Probabilistic constraints via SQP solver: application to a renewable energy management problem. Comput Manag Sci 12:435–459

Bruhns A, Deurveilher G, Roy JS (2005) A non-linear regression model for mid-term load forecasting and improvements in seasonality. PSCC 2005 Luik

Carrión M, Arroyo JM (2006) A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem. IEEE Trans Power Syst 21(3):1371–1378

Daniildis A, Lemaréchal C (2005) On a primal-proximal heuristic in discrete optimization. Math Program Ser A 104:105–128

de Oliveira W, Sagastizábal C (2014) Level bundle methods for oracles with on demand accuracy. Optim Methods Softw 29(6):1180–1209

de Oliveira W, Sagastizábal C, Lemaréchal C (2014) Convex proximal bundle methods in depth: a unified analysis for inexact oracles. Math Program Ser B 148:241–277

Dentcheva D (2009) Optimisation models with probabilistic constraints. In: Shapiro A, Dentcheva D, Ruszczynski A (eds) Lectures on stochastic programming. Modeling and theory, MPS-SIAM series on optimization, vol 9. SIAM and MPS, Philadelphia

Ding X, Lee W-J, Jianxue W, Liu L (2010) Studies on stochastic unit commitment formulation with flexible generating units. Electr Power Syst Res 80:130–141

Dubost L, Gonzalez R, Lemaréchal C (2005) A primal-proximal heuristic applied to french unitcommitment problem. Math Program 104(1):129–151

Dyer ME, Frieze AM (1988) On the complexity of computing the volume of a polyhedron. SIAM J Comput 17(5):967–974

Feltenmark S, Kiwiel KC (2000) Dual applications of proximal bundle methods, including lagrangian relaxation of nonconvex problems. SIAM J Optim 10(3):697–721

Finardi EC, Da Silva EL (2006) Solving the hydro unit commitment problem via dual decomposition and sequential quadratic programming. IEEE Trans Power Syst 21(2):835–844

Frangioni A, Gentile C, Lacalandra F (2008) Solving unit commitment problems with general ramp constraints. Int J Electr Power Energy Syst 30:316–326

Frangioni A, Gentile C, Lacalandra F (2011) Sequential Lagrangian-MILP approaches for unit commitment problems. Int J Electr Power Energy Syst 33:585–593

Genz A (1992) Numerical computation of multivariate normal probabilities. J Comput Graph Stat 1:141–149

Genz A, Bretz F (2009) Computation of multivariate normal and t probabilities. In: Number 195 in Lecture Notes in Statistics. Springer, Dordrecht

Gersho A, Gray RM (1992) Vector Quantization and Signal Compression. In: Number 159 in The Springer International Series in Engineering and Computer Science. Springer, US

Henrion R, Möller A (2012) A gradient formula for linear chance constraints under Gaussian distribution. Math Oper Res 37:475–488

Hiriart-Urruty JB, Lemaréchal C (1996) Convex Analysis and Minimization Algorithms I. In: Number 305 in Grundlehren der mathematischen Wissenschaften, 2nd edn. Springer

Hong LJ, Yang Y, Zhang L (2011) Sequential convex approximations to joint chance constrained programed: a monte carlo approach. Oper Res 3(59):617–630

Kall P, Mayer J (2005) Stochastic linear programming: models, theory and computation. International series in operations research and management science, 1st edn. Springer, New York

Lejeune MA (2012) Pattern-based modeling and solution of probabilistically constrained optimization problems. Oper Res 60(6):1356–1372

Lloyd S (1982) Least squares quantization in pcm. IEEE Trans Inf Theory 28(2):129–137

Luedtke J (2010) An integer programming and decomposition approach to general chance-constrained mathematical programs. In: Eisenbrand F, Shepherd FB (eds) Integer programming and combinatorial optimization, lecture notes in computer science, vol 6080. Springer, Berlin, pp 271–284

Luedtke J (2014) A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. Math Program 146(1–2):219–244

Minoux M (2014) Two-stage robust optimization, state-space representable uncertainty and applications. RAIRO-Oper Res 48:455–475

Morales-España G, Latorre JM, Ramos A (2013) Tight and compact MILP formulation for the thermal unit commitment problem. IEEE Trans Power Syst 28(4):4897–4908

Morales-España G, Latorre JM, Ramos A (2013) Tight and compact MILP formulation of start-up and shut-down ramping in unit commitment. IEEE Trans Power Syst 28(2):1288–1296

Nemirovski A, Shapiro A (2006) Convex approximations of chance constrained programs. SIAM J Optim 17(4):969–996

Ozturk UA, Mazumdar M, Norman BA (2004) A solution to the stochastic unit commitment problem using chance constrained programming. IEEE Trans Power Syst 19(3):1589–1598

Pagès G, Printems J (2003) Optimal quadratic quantization for numerics: the gaussian case. Monte Carlo Methods Appl 9(2):135–166

Prékopa A (1995) Stochastic programming. Kluwer, Dordrecht

Prékopa A (2003) Probabilistic programming. In: Ruszczyński A, Shapiro A (eds) Stochastic programming. Handbooks in operations research and management science, vol 10. Elsevier, Amsterdam

Ruszczyński A (2003) Decomposition methods. In: Ruszczyński A, Shapiro A (eds) Stochastic programming. Handbooks in operations research and management science, vol 10. Elsevier, Amsterdam

Sagastizábal C (2012) Divide to conquer: decomposition methods for energy optimization. Math Program 134(1):187–222

Shan F, Zhang LW, Xiao X (2014) A smoothing function approach to joint chance constrained programs. J Optim Theory Appl 163:181–199

Tahanan M, van Ackooij W, Frangioni A, Lacalandra F (2015) Large-scale unit commitment under uncertainty: a literature survey. 4OR 13(2):115–171

Takriti S, Birge JR (2000) Using integer programming to refine lagrangian-based unit commitment solutions. IEEE Trans Power Syst 15(1):151–156

van Ackooij W (2014) Decomposition approaches for block-structured chance-constrained programs with application to hydro-thermal unit commitment. Math Methods Oper Res 80(3):227–253

van Ackooij W, de Oliveira W (2014) Level bundle methods for constrained convex optimization with various oracles. Comput Optim Appl 57(3):555–597

van Ackooij W, Henrion R, Möller A, Zorgati R (2010) On probabilistic constraints induced by rectangular sets and multivariate normal distributions. Math Methods Oper Res 71(3):535–549

van Ackooij W, Henrion R, Möller A, Zorgati R (2014) Joint chance constrained programming for hydro reservoir management. Optim Eng 15:509–531

van Ackooij W, Malick J (2014) Decomposition algorithm for large-scale two-stage unit-commitment. Draft submitted, http://www.optimization-online.org/DB_HTML/2015/04/4873.html, pages 1–26, 2014

van Ackooij W, Minoux M (2015) A characterization of the subdifferential of singular Gaussian distribution functions. Set valued and variational analysis. pages 1–19 (to appear in)

van Ackooij W, Sagastizábal C (2014) Constrained bundle methods for upper inexact oracles with application to joint chance constrained energy problems. SIAM J Optim 24(2):733–765

Wang SJ, Shahidehpour M, Kirschen DS, Mokhtari S, Irisarri GD (1995) Short-term generation scheduling with transmission and environmental constraints using an augmented lagrangian relaxation. IEEE Trans Power Syst 10(3):1294–1301

Xiong P, Jirutitijaroen P (2011) Stochastic unit commitment using multi-cut decomposition algorithm with partial aggregation. In IEEE Power and Energy Society General Meeting

Zaourar S, Malick J (2013) Prices stabilization for inexact unit-commitment problems. Math Methods Oper Res 78(3):341–359

Zhuang F, Galiana FD (1988) Towards a more rigorous and practical unit commitment by lagrangian relaxation. IEEE Trans Power Syst 3(2):763–773

Zugno M, Conejo AJ (2013) A robust optimization approach to energy and reserve dispatch in electricity markets. Technical report, University of Denmark