Optimal Power Generation under Uncertainty via Stochastic Programming *

Darinka Dentcheva and Werner Römisch
Humboldt-Universität zu Berlin
Institut für Mathematik
Unter den Linden 6
10099 Berlin, Germany

Abstract

A power generation system comprising thermal and pumped-storage hydro plants is considered. Two kinds of models for the cost-optimal generation of electric power under uncertain load are introduced: (i) a dynamic model for the short-term operation and (ii) a power production planning model. In both cases, the presence of stochastic data in the optimization model leads to multi-stage and two-stage stochastic programs, respectively. Both stochastic programming problems involve a large number of mixed-integer (stochastic) decisions, but their constraints are loosely coupled across operating power units. This is used to design Lagrangian relaxation methods for both models, which lead to a decomposition into stochastic single unit subproblems. For the dynamic model a Lagrangian decomposition based algorithm is described in more detail. Special emphasis is put on a discussion of the duality gap, the efficient solution of the multi-stage single unit subproblems and on solving the dual problem by bundle methods for convex nondifferentiable optimization.

Keywords: hydro-thermal power system, uncertain load, stochastic programming, multi-stage, two-stage, mixed-integer, Lagrangian relaxation, bundle methods.

MSC 1991: 90C15, 90C90, 90C11

*This research is supported by the Schwerpunktprogramm "Echtzeit- Optimierung grosser Systeme" of the Deutsche Forschungsgemeinschaft

1 Introduction

The efficient operation and planning of electric power generation systems play an important role for electric utilities as well as the whole human activity. On the one hand, the efficient use of the available fuel for the production of electrical energy is of growing importance, both monetarily and because most of the primary energy sources, which today's energy supply is based on, are not renewable and have limited scope. Savings of a small percentage in the operation of a moderately large power system represent a significant reduction in operation cost as well as in the quantities of fuel consumed. On the other hand, in the future, the human community and, in particular, the power supply industry will be confronted with general economic and ecological conditions that are partly contradictory and aggravating. Some of these conditions are the rise in global energy demand, the scarcity of essential resources and the limits to the local and global environmental damage. Another contemporary challenge for the electric utility industry arises from the changes of market structures for electric power. There has been a world-wide movement towards deregulation of the electric utility industry and an opening of the market to nonutility participants. Moreover, there are plans to open the use of the transmission system in the European Community. All this has led and will further lead to a growth of the number and size of energy transactions. This development raises questions about the prices involved which are based on market actions rather than on costs as in traditional delivery contracts.

These issues have motivated a growing interest in applying mathematical modelling and optimization techniques for optimal system operation. Indeed, there is already a long tradition for applying mathematical programming methods and software to the solution of many relevant engineering problems (e. g. economic dispatch and unit commitment; see [66], [68] and the references therein). The recent substantial progress in many areas of mathematical optimization (e. g. in linear, mixed-integer, nonlinear, nondifferentiable and stochastic programming) opens the road to solving more and more involved models (e. g. [21]). Such complex and large optimization models arise, for instance, for the optimal operation of a hydro-thermal system when including additional aspects like data uncertainty, other regenerative sources of energy, the mid-term management of reservoirs, electricity trading etc. Models of this type are usually characterized by a combination of several difficulties like continuous as well as binary decision variables, very large dimension, nonlinearities (e. g. in hydro modelling, fuel costs, price structures in fuel as well as in electricity purchases) and the uncertainty of problem data (e. g. uncertainty of load forecasts, streamflows to reservoirs, pricing schemes, generator failures etc.).

The present paper aims, in particular, at applying a mathematical methodology, called *stochastic programming*, for handling uncertain data in optimization models. Stochastic programming is mostly concerned with problems that require a here-and-now decision on the basis of given probabilistic information on random quantities, but without making further observations. Possible formulations of stochastic programming models depend on when decisions must be taken relative to the realization

of the random variables (e. g. at several stages in a dynamic model), the degree to which the constraint structure must be satisfied (e. g. with some probability), and the choice of the (stochastic) objective function (e. g. expected costs).
Stochastic programming approaches for tackling models in electric power generation under uncertainty have already found considerable attention (cf. chapters 24-26 in [17] for earlier works). We briefly mention here some of the recent and relevant works in this direction. A multi-stage stochastic optimization model for the optimal scheduling of a hydro-thermal generation system with uncertain inflows is developed in [51]. The authors present a solution strategy based on Benders decomposition and test results for a system comprising 39 hydroelectric plants, one aggregate thermal unit and a yearly planning period with monthly stages. In [31] a multi-stage stochastic program for scheduling hydroelectric generation under uncertainty is described and solved by an enhanced version of nested Benders decomposition. The paper also reports on the generation of monthly streamflow scenario trees and on model validation in the user's environment of the Pacific Gas & Electric Company. In [11] stochastic programming techniques based on Benders decomposition and importance sampling are applied to the facility expansion planning of electric power systems under uncertainty of the availability of generators and transmission lines, and on the demand. Schemes for the pricing of electric power, which is subject to demand and supply uncertainties, are designed and compared in [30] by means of a two-stage stochastic recourse model. The following papers deal with power scheduling under uncertain load. A two-stage stochastic program with simple recourse for the daily economic dispatch in a thermal power system is developed and solved in [8] under the assumption that the marginal distributions of the load are normal. In [24] and [25], this model is extended to power systems comprising thermal and pumped-storage hydro units and general load distributions. The extended model is solved by combining a smooth nonparametric estimation procedure for the marginal load distributions with standard nonlinear programming methods and it is validated by solving the daily economic dispatch problem of a system involving 24 thermal and 5 pumped-storage plants. Further extensions of the latter model by allowing for more general dynamics between decision and observation and for more appropriate recourse cost functions are discussed in [22] and [58]. These models do not yet include start-up and shut-down decisions into the optimization process. This is realized in [64], where a stochastic unit commitment problem for a thermal power system and a corresponding solution technique based on progressive hedging are developed. The progressive hedging methodology (cf. [57]) leads to a successive decomposition into scenario subproblems, which are deterministic unit commitment problems, and solved by Lagrangian relaxation and by an adapted subgradient method for dual maximization. In [65], the authors report on encouraging test runs for large real-life models.
The present paper aims at the development of two kinds of models for the cost-optimal scheduling of electric power in a hydro-thermal generation system under uncertain load: a dynamic stochastic recourse model for the short-term operation and a two-stage stochastic production planning model. Both models are further extensions of

the stochastic models described in [24], [22] and [58]. They represent mixed-integer stochastic optimization problems which are large-scale for moderately large power systems. The second aim of the present paper consists in designing Lagrangian decomposition procedures for the two models by exploiting the particular structure of coupling constraints.

The models arise from a cooperation with the electric utility VEAG Vereinigte Energiewerke AG, which supplies the Eastern part of Germany. The VEAG owned generation system (in 1995) consists of 25 (coal-fired or gas-burning) thermal units and 6 pumped-storage hydro plants. Its total capacity is about 13.150 megawatts (MW) including a hydrogeneration capacity of 1.700 MW; the systems peak load amounts to 8.620 MW (in 1995). Hence, optimal scheduling of the VEAG-system exhibits two special features: the simultaneous optimization of thermal and hydro capacity is indispensable and the model is more large-scale than ever when including stochasticity. This gives rise to the need of solution algorithms for large-scale stochastic optimization problems which allow for handling mixed-integer decisions.

Existing solution procedures for large-scale stochastic programs are mostly based on approximating the underlying probability distribution by a discrete measure having finite support and on utilizing decomposition techniques for solving the large-scale approximate (deterministic) programs. For an overview and a discussion of much of the work done in this direction we refer to [15], [17], [19], [32], [52], [67]. In addition, we mention some of the recent relevant papers on decomposition approaches in stochastic programming. Primal decomposition techniques are based on the L-shaped or Benders decomposition method ([62]), its nested extension for multi-stage models ([4], [23]), and on regularized decomposition ([60]). A second group of (sometimes called *dual* or *scenario*) decomposition methods relax nonanticipativity constraints by introducing Lagrangian terms. For instance, the progressive hedging algorithm ([57]) and the scenario decomposition methods in [45], [59] are based on introducing augmented Lagrangians. Another augmented Lagrangian method by relaxing the recourse constraints is developed in [12]. A third group of methods consists of algorithms that combine decomposition and sampling techniques in various ways. For instance, sampling techniques are used for the generation of cuts in stochastic decomposition methods ([27]), for the efficient calculation of multivariate expected values by importance sampling ([29]), and for reducing the large dimensionality via EVPI-sampling ([13]) within nested Benders decomposition. Methods of a fourth group combine decomposition schemes and iterated approximations via refinement strategies (cf. [19], [20] and chapt. 3.5 in [32]).

Most of these numerical methods cannot be applied directly to stochastic programs involving integrality constraints. Methods for solving (mixed-) integer stochastic programs are rather rare. We refer to [61] for a brief overview of some recent approaches to stochastic integer programming. Moreover, let us mention a recently developed stochastic branch and bound method ([47]) and a dual decomposition method based on relaxing the scenario constraints and on (deterministic) branch and bound techniques ([9]), which also applies to mixed-integer situations.

Our paper is organized as follows. We introduce and discuss the two stochastic power scheduling models in Section 2. In Section 3 we briefly recall the Lagrangian relaxation approach and review some recent progress in solving the nondifferentiable duals. In the remaining two sections we develop Lagrangian decomposition methods for the dynamic recourse as well as for the two-stage stochastic model by relaxing coupling constraints. The dualization argument and the duality gap, the separability structure and the solution of the stochastic single unit subproblems are discussed in more detail for the dynamic model.

2 Models

2.1 Modelling a Hydro-Thermal System

We consider a power generation system comprising (coal-fired and gas-burning) thermal units, pumped-storage hydro plants and interchange contracts between interconnected utilities. We will develop and describe a mathematical model for a power system of this kind which has its origin in the earlier papers [24], [25]. The models allow for the simultaneous scheduling of all units and contracts over a certain time horizon.

Let T denote the number of time intervals obtained by discretizing the operation horizon. This discretization may be chosen uniformly (e. g. hourly or half-hourly) or non-uniformly. Let I and J denote the number of thermal and pumped-storage hydro units in the system. Delivery contracts are regarded as particular thermal units, but may have cost functions that are essentially different (e. g. nonconvex) from typical thermal costs. The decision variables in the model correspond to the outputs of each unit, i. e., the electric power generated or consumed by each unit of the system. These decision variables are denoted by

$$\begin{array}{llll} u_i^t \;, & p_i^t \;, & i=1,\ldots,I \;, & t=1,\ldots,T, \\ s_j^t \;, & w_j^t \;, & j=1,\ldots,J \;, & t=1,\ldots,T, \end{array}$$

where $u_i^t \in \{0,1\}$ and p_i^t are the on/off decisions and the production levels of the thermal unit i during the time period t. Correspondingly, s_j^t, w_j^t are the generation and pumping levels of the pumped-storage plant j during the period t, respectively. Thus, $u_i^t = 0$ and $u_i^t = 1$ mean that unit i is off-line and on-line during period t, respectively. Further, by ℓ_j^t we denote the storage volume in the upper reservoir of plant j at the end of interval t. All variables mentioned above have finite upper and lower bounds representing unit capacity limits and reservoir capacities of the generation system:

$$\begin{array}{rcl} p_{it}^{\min} u_i^t & \leq & p_i^t \leq p_{it}^{\max} u_i^t \;,\; u_i^t \in \{0,1\},\; i=1,\ldots,I,\; t=1,\ldots,T, \\ 0 \leq s_j^t & \leq & s_{jt}^{\max}, 0 \leq w_j^t \leq w_{jt}^{\max}, 0 \leq \ell_j^t \leq \ell_{jt}^{\max}, j=1,\ldots,J,\; t=1,\ldots,T. \end{array} \tag{2.1}$$

The constants $p_{it}^{\min}, p_{it}^{\max}, s_{jt}^{\max}, w_{jt}^{\max}$, and $\ell_{jt}^{\max}$ denote the minimal/maximal outputs of the units and the maximal storage volumes in the upper reservoirs during period

t, respectively. The dynamics of the storage volume, which is measured in electrical energy, is modelled by the equations:

$$\begin{array}{rcll} \ell_j^t & = & \ell_j^{t-1} - s_j^t + \eta_j w_j^t & , \quad t = 1, ..., T, \\ \ell_j^0 & = & \ell_j^{\text{in}}, \ell_j^T = \ell_j^{\text{end}} & , \quad j = 1, ..., J. \end{array} \qquad (2.2\ )$$

Here, ℓ_j^{in} and ℓ_j^{end} denote the initial and final volumes in the upper reservoir, respectively, and η_j is the cycle efficiency of plant j. The cycle efficiency is defined as the quotient of the generation and of the pumping load that correspond to the same volume of water. The equalities (2.2) show, in particular, that there occure no in- or outflows in the upper reservoirs and, hence, that the pumped storage plants of the system operate with a constant amount of water. Together with the upper and lower bounds for ℓ_j^t the equations (2.2) mean that certain reservoir constraints have to be maintained for all pumped-storage plants during the whole time horizon.
Further single-unit constraints are minimum up- and down-times and possible must-on/off constraints for each thermal unit. Minimum up- and down-time constraints are imposed to prevent the thermal stress und high maintenance costs due to excessive unit cycling. Denoting by τ_i the minimum down-time of unit i, the corresponding constraints are described by the inequalities:

$$u_i^{t-1} - u_i^t \leq 1 - u_i^\tau \ , \ \tau = t+1, ..., \min\{t + \tau_i - 1, T\}, \ t = 1, ..., T. \qquad (2.3\ )$$

Analogous constraints can be formulated describing minimum-up times. Note that further single-unit constraints could be added, such as generator fuel limit constraints or air quality constraints in the form of limits on emissions from fossil-fired units.
The next constraints are coupling across the units: the loading and reserve constraints. The first constraints are essential for the operation of the power system and mean that the sum of the output powers is greater than or equal to the load demand in each time period. Denoting by d_t the load demand during period t, the loading constraints are described by the inequalities:

$$\sum_{i=1}^{I} p_i^t + \sum_{j=1}^{J} (s_j^t - w_j^t) \geq d^t, \ t = 1, ..., T. \qquad (2.4\ )$$

In order to compensate unexpected events within a specified short time period, a spinning reserve, describing the total amount of generation available from all units synchronized on the system minus the present load, is prescribed. For instance, such events are sudden load increases and the outage of one or more units. Beyond spinning reserve various classes of off-line reserves may be involved. These include gas-turbine units and pumped-storage hydro plants that can quickly be brought on-line and up to full capacity. Hence, the spinning reserve constraints concern the synchronized thermal units and are given by the following inequalities:

$$\sum_{i=1}^{I} (p_{it}^{\max} u_i^t - p_i^t) \geq r^t, \ t = 1, ..., T, \qquad (2.5\ )$$

where $r^t > 0$ is a specified spinning reserve in period t.
The objective function is given by the total costs for operating the thermal units. These costs consist of the sum of the costs of each individual unit over the whole time horizon, i. e.,

$$\sum_{i=1}^{I}\sum_{t=1}^{T}\left[FC_{it}(p_i^t,\, u_i^t) + SC_{it}\left(u_i(t)\right)\right], \tag{2.6 }$$

where FC_{it} is the fuel cost function and SC_{it} are the start-up costs for the operation of the thermal unit i during period t. We make the natural assumption that $FC_{it}(0,\, 0) = 0$ and that $FC_{it}(\cdot, 1)$ is strictly monotonically increasing. Often fuel cost functions are piecewise linear-quadratic and convex, i. e., they are functions of the form

$$FC_{it}(p,\, u) = \max_{\ell=1,\ldots,L} f_{i\ell}(p) + u\ c_i, \tag{2.7 }$$

where $f_{i\ell}$ are linear or convex quadratic functions having the property $\max_{\ell=1,\ldots,L} f_{i\ell}(0) = 0$ and c_i is a fixed cost term. Non-convex set-ups for fuel costs are also possible and of particular importance for modelling costs in delivery contracts including discounts. Typical cost functions of this kind are general piecewise linear functions. Note that such functions can be modeled using binary variables for selecting the correct line segment for a given value of p (see e. g. [46]).
The start-up costs $SC_{it}\left(u_i(t)\right)$, where $u_i(t) = (u_i^1, ..., u_i^t)$, can vary from a maximum cold-start value to a much smaller value when the unit i is still relatively close to the operating temperature. A simple description for start-up costs is given by

$$SC_{it}\left(u_i(t)\right) = C_i^f \max\left\{u_i^t - u_i^{t-1}, 0\right\},\ t = 2, \ldots, T,$$

where C_i^f are fixed costs. This description has the advantage that it can be expressed in linear terms. On the other hand, it does not reflect that the costs depend on the cooling time. Alternatively, a more involved start-up cost function, which is time-dependent, is given by

$$SC_{it}\left(u_i(t)\right) = \left(C_i^f + C_i^c\left(1 - \exp\left(-(t - t_{s_i})/\alpha_i\right)\right)\right) \max\left\{u_i^t - u_i^{t-1},\, 0\right\},$$

where C_i^f are again fixed costs, C_i^c cold-start costs, α_i the thermal time constant for the unit i and $t - t_{s_i}$ the down-time of unit i until period t, i. e.,
$s_i = \max\left\{s \in I\!N : u_i^{t-j} = u_i^{t-1},\, j = 2, \ldots, s\right\}$.
Altogether, minimizing the objective function (2.6) subject to the constraints (2.1)-(2.5) leads to a cost-optimal schedule for all units of the power system during the specified time horizon. It is worth mentioning that a cost-optimal schedule has the following two interesting properties, which are both a consequence of the strict monotonicity of the fuel costs. If a schedule $(u,\, p,\, s,\, w)$ is optimal, then the loading constraints (2.4) are typically satisfied with equality and we have $s_j^t w_j^t = 0$ for all $j = 1, \ldots, J,\ t = 1, \ldots, T$, i. e., generation and pumping do not occur simultaneously (see [26]).

The minimization problem (2.1)-(2.6) represents a mixed-integer program with (possibly) nonlinear objective, linear constraints, and IT binary and $(I+2J)T$ continuous variables, respectively. For a typical configuration of the VEAG owned generation system with $I = 22$ (thermal), $J = 6$ (hydro) and $T = 192$ (i. e., 8 days with hourly discretization), this amounts to 4224 binary and 6528 continuous variables.

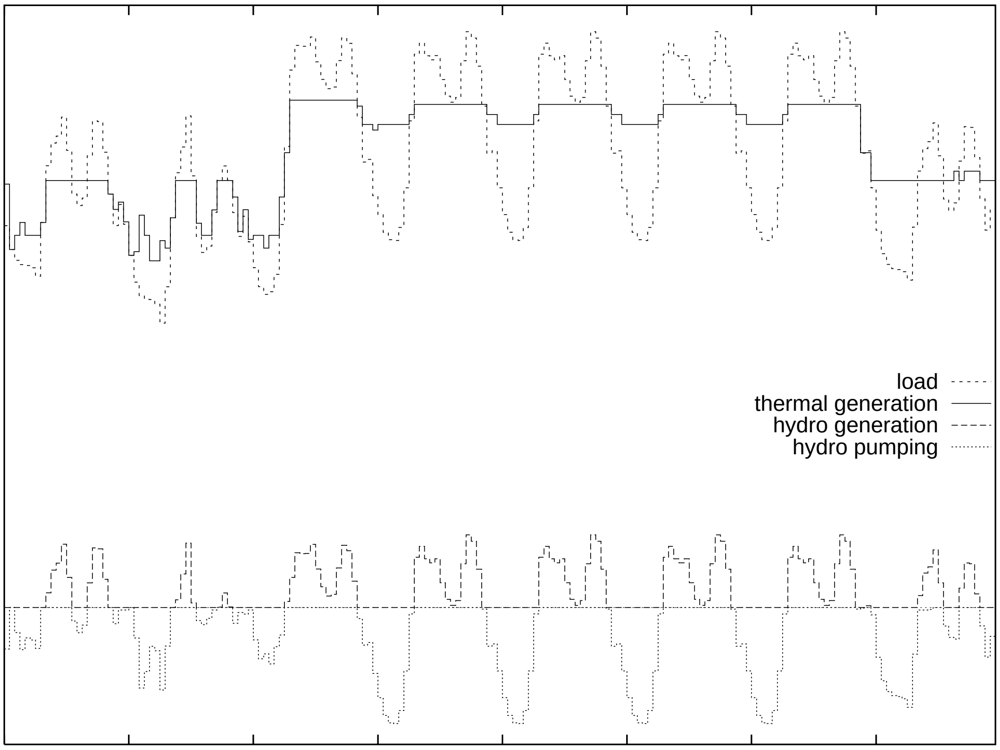

Fig. 1: load curve and hydro-thermal schedule

For this park of the power system and for a peak load week, Figure 1 shows a typical load curve and a corresponding cost-optimal hydro-thermal schedule. Note that the mixed-integer program is solved by the methods described in [14], which Figure 1 is taken from. The load curve in Figure 1 shows two types of cycles: In general, the load is higher during the morning and the early evening (peak), with a small valley during the early afternoon, and it is lower during the night. In addition, the consumption of electric power exhibits a weekly cycle, because the load is lower over weekend days than weekdays. The efficient operation of pumped-storage hydro plants exploits these two cycles. They are designed to save fuel costs by serving the peak load with hydro-energy and then pumping to refill the reservoir during off-peak periods, i. e., during the nights and weekends. The hydro schedule in Figure 1 reflects this typical operation of pumped-storage plants. They may, in fact, be operated on a daily or weekly cycle. Figure 1 records a schedule when operating on a weekly cycle. The remaining load, i. e., the difference between the original system load and the hydro schedule, shows a much more uniform structure than the original load. This portion of the load is covered by the total thermal output. Among the thermal plants of the system, the base-load units are loaded nearly 100% of the time horizon and the "cycling" units are loaded for periods depending on their costs and the shape of the load pattern.

So far we have tacitly assumed that the electrical load is deterministic over the whole time horizon. In electric utilities, schedulers forecast the electrical load for each time period of the day or week in advance. For this purpose they make use of historical load data (e. g. of the same week from previous years), of their personal experience and of statistical methods (e. g. time series or regression analysis). But, clearly, the actual load demand may deviate from the predicted load at any time period for various reasons. Usually electric utilities record the actual system load and save the data over several years. These statistical data provide a basis for the development of stochastic models for the load process and the optimization of power scheduling.

Next we decribe two stochastic models for the optimal scheduling of electric power which differ mainly in the quality of available information on the load stochasticity. The first one represents a model for the optimal on-line or short-term operation of a power system, where future consequences of actual scheduling decisions as well as the future load uncertainty are taken into account. In this model we assume that the load is completely known (i. e., deterministic) at the beginning of the time horizon and that the load uncertainty increases with the growing number of time periods. Secondly, a model for short- or mid-term power production planning is developed. The essential difference to the first model is that the quality of available information on the load uncertainty does not depend on time. It aims at determining (optimal) power production schedules for a future planning period (e. g. next week or month). The second model represents a two-stage stochastic program, whereas the first one is a dynamic (multi-stage) stochastic optimization problem. Both models involve mixed-integer decisions in all stages.

2.2 Dynamic Recourse Model

We assume that the load $\{d^t : t = 1, \ldots, T\}$ forms a (discrete-time) stochastic process on some probability space $(\Omega, \mathcal{A}, \mu)$, that the information on the load is complete for $t = 1$, and that the uncertainty increases with growing t. Let $\{\mathcal{A}_t : t = 1, \ldots, T\}$ be the filtration generated by the load process, where $\mathcal{A}_t$ is the μ-completed σ-field defined by the random vector $(d^1, \ldots, d^t)$. Hence, we have $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \ldots \subseteq \mathcal{A}_t \subseteq \ldots \subseteq \mathcal{A}_T \subseteq \mathcal{A}$ and $\mathcal{A}_1$ is the μ-completion of $\{\emptyset, \Omega\}$. The sequence of scheduling decisions $\{(u^t, p^t, s^t, w^t) : t = 1, \ldots, T\}$ also forms a stochastic process on $(\Omega, \mathcal{A}, \mu)$, which is assumed to be *adapted* to the filtration of σ-fields, i.e., *nonanticipative*. The latter condition means that the decision (u^t, p^t, s^t, w^t) depends only on the data history $(d^1, \ldots, d^t)$ or, equivalently, that (u^t, p^t, s^t, w^t) is $\mathcal{A}_t$-measurable. We mention that this condition is often formulated in terms of a closed linear subspace that is determined by the conditional expectations with respect to the σ-fields $\mathcal{A}_t$ ([55], [12]). Since all decision variables are uniformly bounded, we may restrict our attention to decisions (u, p, s, w) belonging to $L^\infty(\Omega, \mathcal{A}, \mu; \mathbb{R}^m)$, where $m := 2(I + J)T$. Then the nonanticipativity condition can be formulated equivalently as

$$x = (u,\, p,\, s,\, w) \in \mathop{\times}_{t=1}^{T} L^{\infty}\left(\Omega,\, \mathcal{A}_t,\, \mu; I\!R^{m_t}\right), \tag{2.8 }$$

where $m_t := 2(I + J)$, and the (stochastic) optimization problem consists in minimizing the expected cost (cf. (2.6))

$$F(x) = I\!E\left\{\sum_{i=1}^{I}\sum_{t=1}^{T}\left[FC_{it}\left(p_i^t,\, u_i^t\right) + SC_{it}\left(u_i(t)\right)\right]\right\} \tag{2.9 }$$

over all decisions $(u,\, p,\, s,\, w)$ satisfying the nonanticipativity constraint (2.8) and μ-almost surely the constraints (2.1)-(2.5). Among the constraints (2.1)-(2.5), (2.2) and (2.3) reflect the dynamics of the model and (2.4), (2.5) are coupling across units. Altogether, the stochastic program involves $2(I + J)T$ stochastic decision variables and, hence, an enormous number of stochastic scheduling decisions for real-life power generation systems. It is a discrete time dynamic or multi-stage recourse problem, where the "stages" do not necessarily refer to time periods, but correspond to steps in the decision process where observations of the uncertain environment (i. e. the load) take place. The number K of stages of the dynamic model thus corresponds to the (maximal) number of time steps $t_1 = 1 < t_2 < \ldots < t_k < \ldots < t_{K+1} = T$ such that we have the strict inclusion $\mathcal{A}_{t_k} \subset \mathcal{A}_{t_{k+1}},\ k = 1, \ldots, K-1$, for the σ-fields belonging to the filtration.
For the numerical solution of the dynamic recourse model we now assume that a *discrete* multivariate probability distribution of the stochastic load vector $d = \left(d^1, \ldots, d^T\right)$, whose finite support consists of the atoms or *scenarios* $d_n = \left(d_n^1, \ldots, d_n^T\right)$, with the probabilities $\pi_n = \mu\left(d = d_n\right),\ n = 1, \ldots, N$, is given. Let $n_k,\ k = 1, \ldots, K$, denote the number of atoms corresponding to the σ-field $\mathcal{A}_{t_k}$. Then we have $n_1 = 1 < n_2 < \ldots < n_k < \ldots < n_K = N$ and the following scenario constraints at each stage $k \in \{1, \ldots, K\}$:

$$d_n^{t_k} = d_{\hat{n}}^{t_k} \text{ implies } d_n^t = d_{\hat{n}}^t, \text{ for all } t = 1, \ldots, t_k. \tag{2.10 }$$

Hence, the information on the load can be represented in the form of a *scenario tree*. Each path from the root to a leaf of the tree corresponds to one scenario; each branching node corresponds to a (decision) stage. Figure 2 shows an example of a load scenario tree over a weekly time horizon, where observations of the load are made every day, leading to one additional daily scenario.

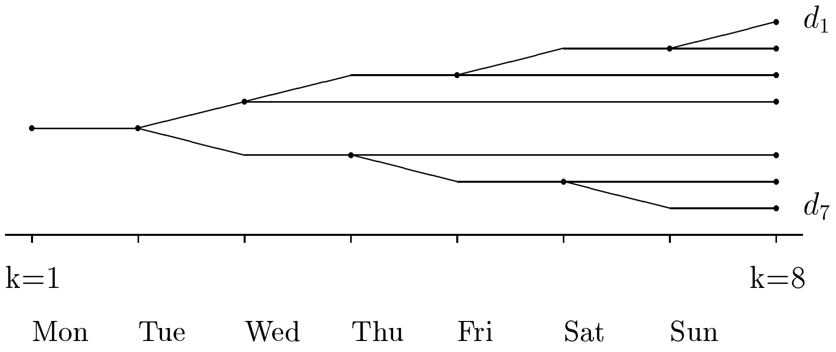

Fig. 2: Load scenario tree

The scenario information may have various origins. It can be obtained as an approximation of the multivariate load distribution, based on sampling from empirical data or on scenarios provided by experienced schedulers. We do not go into detail here, but refer to [16] (and the references therein) for a discussion of various approaches to the generation of scenarios that reflect the structure of the model as well as the information available on the underlying probability distribution. We also refer to [64] where several strategies for generating load scenarios (e. g. handling forecast uncertainty) are discussed. Although the primary aim of generating a scenario tree is to obtain a reasonable approximation for the underlying probability distribution, a compromise between the quality of approximation and the size of the approximate problem has to be taken into consideration, too. The size of the scenario based multi-stage model easily grows out of hand with increasing number of scenarios and stages. In order to illustrate this fact, let $u_{i,n}$, $p_{i,n}$, $s_{j,n}$, $w_{j,n}$, and $\ell_{j,n}$, denote the n-th scenario of the variables u_i, p_i, s_j, w_j, and ℓ_j. Then the scenario based model consists in minimizing the objective function

$$\sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{t=1}^{T} \pi_n \left[FC_{it}\left(p_{i,n}^t,\, u_{i,n}^t\right) + SC_{it}\left(u_{i,n}(t)\right)\right] \tag{2.11 }$$

over all decisions $\{(u_n,\, p_n\, s_n\, w_n) : n = 1, \ldots, N\}$ satisfying the bound and integrality constraints (2.1), the system dynamics

$$\begin{aligned} &\ell_{j,n}^t = \ell_{j,n}^{t-1} - s_{j,n}^t + \eta_j w_{j,n}^t,\ \ \ell_{j,n}^0 = \ell_j^{in},\ \ \ell_{j,n}^T = \ell_j^{end}, j = 1, \ldots, J, \\ &u_{i,n}^{t-1} - u_{i,n}^t \leq 1 - u_{i,n}^{\tau},\ \ \tau = t+1, \ldots, \min\{t + \tau_i - 1, T\}, \\ &t = 1, \ldots, T,\ \ n = 1, \ldots, N, \end{aligned} \tag{2.12 }$$

the loading and reserve constraints

$$\begin{aligned} &\sum_{i=1}^{I} p_{i,n}^t + \sum_{j=1}^{J} (s_{i,n}^t - w_{i,n}^t) \geq d_n^t,\ \ \sum_{i=1}^{I} \left(p_{it}^{\max} u_{i,n}^t - p_i^t\right) \geq r^t, \\ &\qquad t = 1, \ldots, T,\ \ n = 1, \ldots, N, \end{aligned} \tag{2.13 }$$

and the scenario nonanticipativity constraints, which can be formulated analogously to (2.10).
When regarding the nonanticipativity constraints and introducing decision variables at each node of the scenario tree, the number of decisions in the (deterministic) optimization model (2.11)-(2.13) amounts to $2(I+J)\sum_{k=1}^{K} n_k\,(t_{k+1}-t_k)$. Hence, the model may easily become extremely large if the scenario tree contains too many paths. Even for the (very) small scenario tree in Figure 2 (i. e., with $K=7$, $n_K=K$ and $t_{k+1}-t_k=24$) the model involves $672\cdot I$ binary and $672\cdot(I+2J)$ continuous variables and standard methods including those reviewed in Section 1, may not be able to solve the problem in reasonable time. This requires other techniques that exploit the underlying structure of the original stochastic model.

2.3 Two-Stage Stochastic Model

Again we assume the load $\{d^t : t=1,\dots,T\}$ to be given as a (discrete-time) stochastic process on some probability space $(\Omega,\,\mathcal{A},\,\mu)$. However, this time the load process does not involve an information structure and the decision process consists of two stages where the first-stage decisions correspond to the here-and-now schedules for all power generation units over the whole time horizon. The second-stage decisions correspond to future compensation or recourse actions of each unit in each time period in response to the environment created by the chosen first-stage decision and the load realization in that specific time period. Hence, the aim of such a two-stage dynamic model can be formulated as follows: Find an optimal schedule for the whole power system and planning horizon such that the uncertain demand can be compensated by the system, all system constraints are satisfied and the sum of the total generation costs and the expected compensation costs is minimal.
In order to give a mathematical formulation of the model, let $(u,\,p,\,s,\,w)$ denote the first-stage scheduling decisions as in Section 2.1 and $(\hat{u},\,\hat{p},\,\hat{s},\,\hat{w})$ denote the stochastic compensation decisions having the components $\hat{u}_i^t,\,\hat{p}_i^t,\,\hat{s}_j^t,\,\hat{w}_j^t,\ i=1,\dots,I,\ j=1,\dots,J,\ t=1,\dots,T$, which correspond to the compensation actions of each unit at time period t.
In addition to the (non-stochastic) constraints for $(u,\,p,\,s,\,w)$, (2.1) (capacity limits), (2.2) (storage dynamics), (2.3) (minimum down-time constraints) and (2.5) (reserve constraints), we have to require that the compensation actions also satisfy certain system constraints. These are the unit capacity limits, minimum-down time constraints and reservoir capacity bounds :

$$p_{it}^{\min}\hat{u}_i^t \le p_i^t\hat{u}_i^t + \hat{p}_i^t \le p_{it}^{\max}\hat{u}_i^t,\ \ \hat{u}_i^t \in \{0,\,1\},\ \ i=1,\dots,I, \tag{2.14 }$$

$$\hat{u}_i^{t-1} - \hat{u}_i^t \le 1-\hat{u}_i^\tau,\ \ \tau=t+1,\dots,\min\{t+\tau_i-1,\,T\},\ i=1,\dots,I, \tag{2.15 }$$

$$0 \le s_j^t + \hat{s}_j^t \le s_{jt}^{\max}, \qquad 0 \le w_j^t + \hat{w}_j^t \le w_{jt}^{\max},$$

$$0 \le \ell_j^t + \hat{\ell}_j^t \le \ell_j^{\max}, \qquad \hat{\ell}_j^0 = \hat{\ell}_j^T = 0 \tag{2.16 }$$

$$\hat{\ell}_j^t = \hat{\ell}_j^{t-1} - \hat{s}_j^t + \eta_j \hat{w}_j^t, \qquad j = 1, \ldots, J, \ t = 1, \ldots, T, \ \mu - \text{a. s.}$$

In other words, the constraints (2.16) for the hydro scheduling decisions mean that the sum of first-stage decisions and recourse actions is feasible, too. The formulation (2.14) of the thermal unit capacity limits for the compensation stage becomes more involved because the term $p_i^t \hat{u}_i^t$ introduces a nonlinear constraint connecting first- and second-stage variables. The nonlinearity in (2.14) is avoided when requiring that a thermal unit, which is scheduled to be on-line in the first-stage, must not be off-line in the compensation action. In this case, (2.14) can be replaced by the (linear) constraints:

$$p_{it}^{\min} \hat{u}_i^t \le p_i^t + \hat{p}_i^t \le p_{it}^{\max} \hat{u}_i^t, \ \ u_i^t \le \hat{u}_i^t, \ \ \hat{u}_i^t \in \{0, 1\}, i = 1, \ldots, I. \tag{2.17 }$$

This formulation of the thermal unit capacity limits seems to be quite natural and realistic because generation systems often possess sufficient flexibility to compensate load decreases by lowering output levels of thermal units. However, there might be a need for new on-line units in order to compensate unpredictable load increases. Another possible compensation strategy could be based on a subdivision of the set of available thermal units into two sets $\mathcal{I}_1$ and $\mathcal{I}_2$ such that $\mathcal{I}_1 \cup \mathcal{I}_2 = \{1, \ldots, I\}$ and the conditions

$$u_i^t = \hat{u}_i^t, \ i \in \mathcal{I}_1, \ \text{ and } u_i^t \le \hat{u}_i^t, \ \ i \in \mathcal{I}_2, t = 1, \ldots, T, \ \mu - \ \text{a. s. },$$

are satisfied. This means that only some of the available thermal units may change their on/off state when compensating uncertain load. From a modelling point of view this strategy would lead to a reduction of the number of binary variables.

In the following, we always assume that (2.17) instead of (2.14) is satisfied. Observe that the conditions (2.15) and (2.17) imply (2.3).

The loading constraints (2.4) are modified by requiring that the sum of the first-stage power outputs of all generation units satisfies the load with some probability $\pi_t \in (0, 1)$ in period t, $t = 1, \ldots, T$, and that the sum of the total power outputs satisfies the load with probability one. Denoting by F_{d^t} the distribution function of d^t, the (modified) loading constraints are given by the following inequalities:

$$\sum_{i=1}^{I} p_i^t + \sum_{j=1}^{J} (s_j^t - w_j^t) \ge F_{d^t}^{-1}(\pi_t), \ \ t = 1, \ldots, T, \tag{2.18 }$$

$$\sum_{i=1}^{I} (p_i^t + \hat{p}_i^t) + \sum_{j=1}^{J} (s_j^t + \hat{s}_j^t - (w_j^t + \hat{w}_j^t)) \ge d^t, \ \ t = 1, \ldots, T, \mu - \text{a.s.} \tag{2.19 }$$

A variant of (2.18), which will be considered in Section 5, is that the term $F_{d^t}^{-1}(\pi_t)$ is replaced by the expected load $\mathbb{E}(d_t)$, $t = 1, \ldots, T$. In both cases, the constraint

(2.18) means that the sum of the first-stage output power satisfies a certain predicted or approximated load and the second-stage decisions take care of satisfying the stochastic load with probability one.
Since the real operation of the system takes place during the compensation action, the objective function corresponds to the total average costs for operating the termal units, i. e.,

$$I\!E\left\{\sum_{i=1}^{I}\sum_{t=1}^{T}\left[FC_{it}\left(p_i^t+\hat{p}_i^t,\,\hat{u}_i^t\right)+SC_{it}\left(\hat{u}_i(t)\right)\right]\right\} \tag{2.20 }$$

where FC_{it} and SC_{it} denote the fuel cost and start-up cost functions, respectively, for the operation of unit i during period t, and $\hat{u}_i(t) := (\hat{u}_i^1, \ldots, \hat{u}_i^t)$.
The stochastic power production planning model consists then in minimizing the objective function (2.20) over all deterministic decisions $(u,\, p,\, s,\, w)$ and all stochastic decisions $(\hat{u},\, \hat{p},\, \hat{s},\, \hat{w}) \in L^\infty(\Omega,\, \mathcal{A},\, \mu;\, I\!R^m)$ satisfying the constraints (2.1), (2.2), (2.5), (2.15)-(2.19). The model represents a two-stage stochastic mixed-integer program involving $2(I+J)T$ deterministic and $2(I+J)T$ stochastic decision variables. Similar to the dynamic model in the previous section, only the loading constraints (2.18), (2.19) and the reserve constraints (2.5) are coupling across units.

3 Lagrangian relaxation approach

Lagrangian relaxation is a solution technique primarily for minimizing a nonsmooth function. We would like to recall the basic ideas and some facts in order to clarify the reasons that make this approach appropriate for solving the problems introduced in the previous section. Our presentation is inspired by [40]. Let us consider an optimization problem

$$\min f(x) \text{ subject to } x \in C,\ \ g(x) \le 0, \tag{3.1 }$$

where $f : I\!R^n \to I\!R$, $C \subseteq I\!R^n$, $g : I\!R^n \to I\!R^m$.
We suppose that the functions f and g and the set C have some special structure, which makes the Lagrangian problem

$$\min\left[L(x,\lambda) = f(x) + \lambda g(x)\right] \text{ subject to } x \in C \tag{3.2 }$$

much easier to solve than the problem (3.1), where $\lambda \in I\!R^m_+$. Let us assume the following:

(A) $\qquad \Theta(\lambda) = \min_{x\in C} L(x,\,\lambda) = L(x_\lambda,\,\lambda)$ is finite for all $\lambda \in I\!R^m_+$.

Furthermore, we set $\tilde{g}(\lambda) = -g(x_\lambda)$.
Be aware that $L(\cdot,\lambda)$ may have several minima for some λ, but $\Theta(\lambda)$ is well-defined, since the minimal value is non-ambiguous. By the weak duality theorem, we have

$$\Theta(\lambda) \le \min_{x\in C} L(x,\,\lambda) \le f(x)$$

for all feasible points x in (3.1). The following statement is straight-forward but important.

Proposition 3.1 *([18]) Any solution $\bar{x}$ of the Lagrangian problem (3.2) solves the perturbed problem (3.3):*

$$\min f(x) \text{ subject to } x \in C,\ -\tilde{g}(\lambda) \geq g(x). \tag{3.3 }$$

Proof: For any feasible x in (3.3) and $\lambda \in I\!R^m_+$ we have

$$\begin{aligned} f(x) &\geq f(x) + \lambda[g(x) + \tilde{g}(\lambda)] \\ &= L(x,\,\lambda) + \lambda\tilde{g}(\lambda) \geq L(\bar{x},\,\lambda) - \lambda \cdot g(\bar{x}) \\ &= f(\bar{x}) \end{aligned}$$

□

We conclude that if $\bar{x}$ is "almost feasible", it is "almost a solution" of (3.1). If we succeed in finding a feasible point to (3.1), then we have a solution to (3.1), because the inequality of (3.3) is satisfied. Having in mind the weak duality theorem, it is clear that any feasible point $\bar{x}$ of (3.1) produces an upper bound $f(\bar{x})$ for $\Theta(\lambda)$. Hence, to solve (3.1) via (3.2) it is necessary to maximize Θ on $I\!R^m_+$.
We call $\Theta(\cdot)$ the *dual function*, λ the *dual variable*, and the problem

$$\max \Theta(\lambda) \text{ subject to } \lambda \in I\!R^m_+ \tag{3.4 }$$

the *dual problem* to (3.1). We show that Θ is a concave function having subgradients at all λ by virtue of the assumption (A). Let us denote a solution of (3.2) for $\bar{\lambda}$ by $\bar{x}$.

$$\begin{aligned} \Theta(\lambda) &= \min_{x\in C} L(x,\lambda) \leq L(\bar{x},\lambda) \\ &= L(\bar{x},\bar{\lambda}) + (\lambda - \bar{\lambda})g(\bar{x}) \\ &= \Theta(\bar{\lambda}) - (\lambda - \bar{\lambda})\tilde{g}(\bar{\lambda}) \end{aligned}$$

The latter inequality characterizes concavity and implies

$$\tilde{g}(\bar{\lambda}) \in \partial[-\Theta(\bar{\lambda})]$$

where $\partial[-\Theta(\cdot)]$ stands for the subdifferential of $-\Theta$ with respect to λ calculated at the point $\bar{\lambda}$.
Let us suppose that the problem under consideration has a separable structure, i. e., the problem is of the following form:

the variables	$x = (x_1, \ldots, x_{\bar{n}})$ and $x_i \in I\!R^{n_i}\quad i = 1,\ldots,\bar{n}$,
the objective function	$f(x) = \sum_{i=1}^{\bar{n}} f_i(x_i) + f_0$,
the related constraints	$g_j(x) = \sum_{i=1}^{\bar{n}} g_j^i(x_i) + g_j^0,\quad j = 1,\ldots,m$,

where f_0 and g_j^0 $(j = 1,\ldots,m)$ are constants.
Let us further suppose some special structure of the set C. We assume the set C to be the following product

$$C = \left(\mathop{\times}_{i=1}^{i_0} \{0,\,1\}^{n_i}\right) \times \left(\mathop{\times}_{i=i_0+1}^{\bar{n}} B_i\right),$$

where $B_i \subseteq I\!R^{n_i}$ are compact convex sets. This means that $x_1, \ldots, x_{i_0}$ are binary variables and we consider a mixed-integer problem.
Furthermore, let us assume the functions f_i and g_j^i to be convex piecewise linear or (piecewise) quadratic functions. Then $L(\cdot, \lambda)$ is a convex function, too.
The strong duality theorem does not apply due to the presence of integrality, i. e. the structure of the set C. However, we are in a favourable situation to have

- the assumption (A) is satisfied,
- decomposable structure of the relaxed problem,
- description of the subgradients of $\Theta(\lambda)$.

We call the following optimization problem a continuous relaxation of the problem (3.1)

$$\min f(x) \text{ subject to } x \in \tilde{C},\ g(x) \leq 0,$$
$$\text{where } \tilde{C} = \left(\times_{i=1}^{i_0} [0,1]^{n_i}\right) \times \left(\times_{i=i_0+1}^{\bar{n}} B_i\right).$$

Proposition 3.2 *The Lagrangian relaxation provides a better lower bound of the optimal value of (3.1) than the continuous relaxation of the problem.*

Proof: The following sequence of inequalities holds true for each $\lambda \in I\!R^m_+$:

$$\min_{\substack{x \in C \\ g(x) \leq 0}} f(x) \geq \min_{x \in C} L(x,\, \lambda) \geq \min_{x \in \tilde{C}} L(x,\, \lambda) = \min_{\substack{x \in \tilde{C} \\ g(x) \leq 0}} f(x)$$

The last equality results from the strong duality theorem. This proves the assertion. □
Observe that $L(x, \lambda)$ has a separable structure with respect to the components x_i, which together with the special structure of C leads to a decomposition of the problem 3.2 into $\bar{n}$ subproblems of dimension n_i each. The subproblems read

$$P_i(\lambda) : \min f_i(x_i) + \sum_{j=1}^{m} \lambda_j g_j^i(x_i) \text{ subject to } x_i \in C_i,$$

$$\text{where:} \qquad C_i = \begin{cases} \{0,\, 1\}^{n_i} & \text{if} \quad 1 \leq i \leq i_0 \\ B_i & \text{if} \quad i_0 \leq i \leq n_i \end{cases}$$

Denoting the marginal functions of the problems above by $\Theta_i(\lambda)$ $(i = 1, \ldots, \bar{n})$ we obtain for the dual function

$$\Theta_i(\lambda) = \sum_{i=1}^{\bar{n}} \Theta_i(\lambda) + f_0 + \sum_{j=1}^{m} \lambda_j g_j^0$$

Consequently, the dual problem has a separable structure, too. The latter observations make an approach to problems with decomposable structure via Lagrangian relaxation attractive. A solution procedure should include:

- a method for solving the non-smooth concave optimization problem (3.4).
- fast algorithms for minimizing the Lagrange-function $L(x, \lambda)$ at a given point λ, i. e., for solving the subproblems $P_i(\lambda), i = 1, \dots, \bar{n}$. The solution provides then the value of Θ and its subgradients.
- a technique to obtain a primal feasible solution.

The latter point needs separate investigations. As already mentioned, a dual method does not provide a primal feasible solution due to the integrality conditions. Thus, we have to use the information on the dual solution to calculate a primal feasible point close to the dual solution efficiently. Due to the first proposition, such a procedure will obtain a fairly good point. In [2], it is shown that the relative duality gap for mixed integer problems with special structure becomes small under certain assumptions. We will see later how the estimate given there is modified for the dynamic recourse problem.

Methods for nonsmooth optimization have been subject of intensive development during the last 15 years. An algorithm for minimizing a convex function known for a long time is the cutting-plane method given in [10], [33]. It develops the natural idea to use subgradient-information and to generate a linear approximation of the function associated with it. Let us suppose that, at a certain moment, values $f(x_1), \dots, f(x_k)$ and subgradients $y_1 \in \partial f(x_1), \dots, y_k \in \partial f(x_k)$ are available. We define

$$\tilde{f}_k(x) = \max\{f(x_i) + < y_i, x - x_i >, \quad i = 1, \dots, k\}$$

and, minimizing $\tilde{f}_k$, obtain a further point x_{k+1}. It is assumed that $\tilde{f}_k$ is bounded from below on C and we are able to compute values and subgradients of f.

However, this algorithm has some well-known drawbacks. The initial iterations are inefficient. The number of cuts increases after each iteration and there is no reliable rule for deleting them. The minimization of the approximate function is sensitive when approaching a point of nondifferentiability. Further developments have led to the so-called bundle methods which offer a stabilizing device based on the following ingredients:

- a sequence $\{x_n\}$ of stabilized iterates;
- a criterion (test) deciding whether a new iterate has been found and (or) whether the bundle of information, i.e., the approximation $\tilde{f}_k$, should be enriched;
- a sequence $\{M_n\}$ of positive definite matrices used for a stabilizing term.

Bundle methods are pioneered by Wolfe and Lemarechal. A detailed study on the subject can be found in [35] and [28]. A comprehensive review is given in [42]. One description of the main idea of (first-order) bundle methods is the following:

Suppose iterate x_n and a bundle of subgradients y_k have been computed. As above, we use the bundle of information to formulate a lower approximation of the function f, i. e., $\tilde{f}_n(x) = \max\{f(x_i) + < y_i, x - x_i >, \quad i = 1, \dots, k\}$, and

1. minimize $\tilde{f}_n(x) + \frac{1}{2} < M_n(x - x_n), x - x_n >$
   and let the point $\bar{x}$ be its minimal point.

2. compute a nominal decrease
   $\alpha_n = f(x_n) - \tilde{f}_k(\bar{x}) - \frac{1}{2} < M_n(\bar{x} - x_n), \bar{x} - x_n >$.
   A constant $c \in (0, 1)$ being chosen, we perform the descent test:
   $f(\bar{x}) \leq f(x_n) - c\alpha_n$
   If the inequality is satisfied we set $x_{n+1} = \bar{x};\ y_{k+1} = \bar{x}$
   and increase n and k by 1.
   Otherwise, n is kept fixed, we set $y_{k+1} = \bar{x}$ and increase k by 1. In some versions (cf. [41]) an additional test is made before increasing k.

3. The choice of $\{M_n\}$ given in the literature is:
   - an abstract sequence, as in [39],
   - $M_n \equiv I$, as in [34],
   - $M_n = \mu_n I$ with heuristic rules for computing μ_n, in [36], [63],
   - solving a quasi-Newton equation in [41].

This description of the bundle methods corresponds to the proximal point concept (i. e., the Moreau-Yosida regularization). Recall that, given a positive semi-definite matrix M,

$$F(x) = \inf \left\{ f(y) + \frac{1}{2} < M(y - x),\ y - x > \right\} \tag{3.5 }$$

is the Moreau-Yosida regularization of the function f. In the classical framework M should be positive definite. In [41], it is suggested to allow a degenerate proximal term and it is shown there that the essential proprties can be reproduced also in this case. A relationship between these concepts and certain first order bundle methods was observed by several authors, e.g. [28]. Methods of order higher than one are studied in [36] and [63] where a single stabilizing parameter is varied.
In [36] the choice of weights μ for updating the matrix in the proximal term is considered. The matrix M is intended to accumulate information about the curvature of f around the point $\bar{x}$. Safeguarded quadratic interpolation is proposed for choosing the weights μ_{n+1} so that the curvature of f between x_n and $\bar{x}$ is estimated. The algorithm computes a direction for the next iterate x_{n+1} by solving a quadratic program, then the descent test and the update of the bundle of subgradients are modified accordingly. The reported computational experiments indicate that this technique can decrease the number of objective evaluations necessary for reaching a desired accuracy in the optimal value significantly.
The algorithms presented in [7], [41], [43], referred to as variable metric bundle methods, make use of the Moreau-Yosida regularization of the objective function and develop some quasi-Newton formulas. Two strategies for updating the matrix M in the minimization procedure are suggested in [41]. In the first version, called diagonal quasi-Newton method, M is proportional to the identity matrix, while the second version uses a full quasi-Newton matrix. The matrix is updated at the end

of a descent-step, when a new stabilizing iterate point is computed. The updating procedure corresponds to a regularizing scheme for the gradient of F.
In [43] M is a positive definite matrix and, thus, there is a unique solution of (3.5), which is denoted by $y(x)$. The main idea is to approximate $y(x)$ and to vary the matrix M in order to use the information gathered in finding one approximation to help in finding the next one. Let J be some approximation of the Jacobian $J(x)$ of $y(x)$. A Newton step $-\left[\nabla^2 F_\mu(x)\right]^{-1} \nabla F(x)$ is approximated there by

$$[M(I-J)]^{-1} M(y(x)-x) = [I-J]^{-1}(y(x)-y),$$

where I is the identity matrix. M could be fixed or updated by

$$M_n = \mu_n G_n,$$

where μ_n is some constant and G_n is an estimate of ∇F computed by information from previous iterations . How to compute the necessary estimate J of the Jacobian matrix of $y(x)$ is discussed in detail in [43]. The method developed there is called approximate Newton-method.
A precise study of the second-order properties of the Moreau-Yosida regularization is presented in [44] for the problem of minimizing a closed proper convex function, which is a selection of a finite number of twice continuously differentiable functions. It is proved that under certain constraint qualification the gradient ∇F_M is piecewise smooth. Further conditions are formulated that guarantee a superlinear (quadratic) convergence of an approximate Newton method for minimizing F.
Generally, one can consider any Newton-type method for nonsmooth equations in order to solve optimization problems. Newton-type methods in such a generality are considered in e. g. [53], [54], [50], [38]. The methods presented there are applied to solving optimization problems via augmented Lagrangians [54], via the Karush-Kuhn-Tucker equations [53] or via the Moreau-Yosida regularization [5].
Our review is not an attempt to comment all recent developments of solution techniques for nonsmooth optimization problems. We only wish to present the main ideas of the well-established methods in order to clarify which of them are appropriate for solving the nonsmooth problems studied in the next two sections.

4 Lagrangian Relaxation for the Dynamic Recourse Problem

In this section, we consider the Lagrangian relaxation approach for the dynamic recourse model (2.1)-(2.9) in detail and sketch a conceptual algorithm for solving the problem. The decision variables are uniformly bounded functions $(u, p, s, w) \in \times_{t=1}^{T} L^\infty(\Omega, \mathcal{A}_t, \mu; I\!R^{m_t})$, $m_t = 2(I+J)$. The variables (u_i, p_i), $i = 1, \dots, I$, and (s_j, w_j), $j = 1, \dots, J$, are associated with one single operation unit i, and j, respectively. All constraints except for (2.4) and (2.5) are associated with a single operation unit. Thus, natural candidates for the relaxation are the coupling constraints (2.4) and (2.5). We associate Lagrange multipliers λ_1 and λ_2 with the load- and reserve-constraints, respectively. Setting $x = (u, p, s, w)$ and

$$
\begin{aligned}
L(x, \lambda) \quad = \quad & I\!E \Bigg\{ \sum_{i=1}^{I} \sum_{t=1}^{T} \left[FC_{it} \left(p_i^t, u_i^t\right) + SC_{it} \left(u_i(t)\right) \right] \\
& + \lambda_1^t \left(d^t - \sum_{i=1}^{I} p_i^t - \sum_{j=1}^{J} \left(s_j^t - w_j^t \right) \right) \\
& + \lambda_2^t \left(r^t - \sum_{i=1}^{I} \left(u_i^t p_{it}^{\max} - p_i^t \right) \right) \Bigg\},
\end{aligned}
\tag{4.1 }
$$

we have to clarify what kind of objects λ_1 and λ_2 are. Duality theorems for dynamic models that are relevant for our setting are considered in [56], [58]. We utilize the results of [56]. For stating a duality result we neglect integrality and substitute $u_i^t \in \{0, 1\}$ by $u_i^t \in [0, 1]$ in (2.1) for a moment. We denote the modified constraint by $(2.1\ )^*$.

First, let us recall that the dynamic recourse problem has *relatively complete recourse* if the following procedure leads to a choice of decisions x_k, $k = 1, \ldots, K$, almost surely for all stages k: Let x_1 be a feasible solution of the first stage. In the second stage (having a new observation of the load), we can choose x_2 satisfying the restrictions and the dynamics of the system, i. e., in particular, (2.2) and (2.3) hold true with the corresponding components of x_1 and x_2. And so forth: In the k-th stage, we are able to choose a feasible decision x_k.

Nonanticipativity and relatively complete recourse provide sufficient conditions for considering L^1 to be the space of Lagrange multipliers λ, instead of working with esoteric objects from $(L^\infty)^*$ (cf. [56]).

Suppose, additionally, that *strict feasibility* holds true. It means, that the feasible set determined by $(2.1\ )^*$-(2.5) has a non-empty interior in $\times_{i=1}^{T} L^\infty(\Omega, \mathcal{A}_t, \mu; I\!R^{m_t})$, i. .e., there exists a positive real number ε, a point $\bar{x} \in \times_{i=1}^{T} L^\infty(\Omega, \mathcal{A}_t, \mu; I\!R^{m_t})$ and a neighbourhood U of $\bar{x}$ such that any point $x = (u, p, s, w) \in U$ satisfies $(2.1\ )^*$-(2.3) and the inequalities:

$$
\sum_{i=1}^{I} p_i^t + \sum_{j=1}^{J} \left(s_j^t - w_j^t \right) \geq d^t + \varepsilon, \quad t = 1, \ldots, T,
$$

$$
\sum_{i=1}^{I} \left(p_i^t - u_i^t p_{it}^{\max} \right) \geq r^t + \varepsilon, \quad t = 1, \ldots, T.
$$

In terms of a power generation system, strict feasibility means that the generation system should have the capacity to produce power that satisfies every slightly changed demand and reserve-condition regarding the other constraints. This is a reasonable and acceptable restriction, which can be assumed to be satisfied.

We denote

$$
X = \left\{ x \in \mathop{\times}_{i=1}^{T} L^\infty(\Omega, \mathcal{A}_t, \mu; I\!R^{m_t}) : \ (2.1\ )^* - \ (2.5\ ) \text{ are fulfilled} \right\};
$$

$$\Lambda = \left\{ \lambda \in \mathop{\times}_{i=1}^{T} L^1 \left(\Omega,\, \mathcal{A}_t,\, \mu; I\!R^2 \right) : \lambda_1^t,\, \lambda_2^t \geq 0 \; \mu - \text{ a.s. for } t = 1, \ldots, T \right\}.$$

The following duality statement holds true.

Proposition 4.1 *The Langrange function (4.1) has at least one saddle point* $(\bar{x}, \bar{\lambda}) \in X \times \Lambda$ *assuming (2.6), (2.7), relatively complete recourse and strict feasibility. In order that the function* $\bar{x} \in X$ *be an optimal solution of the problem (2.1) - (2.9) it is necessary and sufficient that the following conditions be satisfied a.s. for some* $\bar{\lambda} \in \Lambda$ *:*

$$\begin{array}{ll} \bar{\lambda}_1^t & \left[d^t - \sum\limits_{i=1}^{I} \bar{p}_i^t - \sum\limits_{j=1}^{J} \left(\bar{s}_j^t - \bar{w}_j^t \right) \right] = 0 \\ \bar{\lambda}_2^t & \left[r^t + \sum\limits_{i=1}^{I} \left(\bar{p}_i^t - p_{it}^{\max} \bar{u}_i^t \right) \right] = 0 \\ 0 & \in \partial_x L \left(\bar{x}^t,\, \bar{\lambda}^t \right), \qquad t = 1, \ldots, T. \end{array} \tag{4.2 }$$

Proof: The assertion follows by Theorem 1 and the arguments of Theorem 7 from [56]. □

Now we consider the relaxed problem:

$$\min_{(u,\, p,\, s,\, w)} L(u,\, p,\, s,\, w) \;\text{ subject to } (2.1) \text{ - } (2.3). \tag{4.3 }$$

Denoting the marginal function of the latter problem by $\Theta(\lambda)$, the dual problem reads

$$\max \Theta(\lambda) \;\text{ subject to } \lambda \in \Lambda. \tag{4.4 }$$

Let us check the properties of Θ discussed in the previous section. The concavity of Θ and Proposition 3.1 follow trivially. Proposition 3.3. follows due to the duality statement discussed above.
Observe that the assumption (A) of Section 3 is satisfied, i. e., the feasible set with respect to the continuous variables is a compact (box-constrained) set because of (2.1).
Now, we show that the dual problem is decomposable with respect to the single units. Using the notations of the previous section, we define

$$x_i = (u_i,\, p_i),\; i = 1, \ldots, I, \quad x_{I+j} = (s_j,\, w_j), \quad j = 1, \ldots, J, \quad \bar{n} = I + J,$$

and observe that all functions are separable with respect to $x_i,\; i = 1, \ldots, \bar{n}$.
We define functions $\Theta_i(\cdot)$ and $\tilde{\Theta}_j(\cdot)$:

$$\begin{aligned} \Theta_i(\lambda) &= \min_{(u_i, p_i)} I\!E \sum_{t=1}^{T} \left[FC_{it}\left(p_i^t,\, u_i^t\right) + SC_{it}(u_i(t)) - \lambda_1^t p_i^t - \lambda_2^t \left(u_i^t\, p_{it}^{\max} - p_i^t\right) \right] \\ &= \min_{u_i} I\!E \sum_{t=1}^{T} \left[\min_{p_i^t} \left\{ FC_{it}\left(p_I^t,\, u_i^t\right) - \left(\lambda_1^t - \lambda_2^t\right) p_i^t \right\} + SC_{it}(u_i(t)) - \lambda_2^t\, u_i^t\, p_{it}^{\max} \right] \end{aligned}$$

The latter equality holds by the separable structure of the functions FC_{it} with respect to p_i^t and u_i^t (cf.(2.7)) and the possibility to exchange min and $\mathbb{E}$ in the above expression.

$$\tilde{\Theta}_j(\lambda) = \min_{(s_j, w_j)} \mathbb{E} \sum_{t=1}^{T} \left[-\lambda_1^t (s_j^t - w_j^t) \right]$$

Consequently, the function $\Theta(\lambda)$ can be expressed as:

$$\Theta(\lambda) = \sum_{i=1}^{I} \Theta_i(\lambda) + \sum_{j=1}^{J} \tilde{\Theta}_j(\lambda) + \mathbb{E} \sum_{t=1}^{T} \left[\lambda_1^t \, d^t + \lambda_2^t \, r^t \right]$$

It has a separable structure with respect to the single units as well as the constraints (2.1) - (2.3), (2.6) - (2.8) do. Thus, the value and subgradients of $\Theta(\lambda)$ can be computed for a given argument λ by solving the subproblems $P_i(\lambda), \ i = 1, \ldots, I$ and $\tilde{P}_j(\lambda) \ j = 1, \ldots, J$:

$$P_i(\lambda): \quad \min_{u_i} \mathbb{E} \sum_{t=1}^{T} \left[\min_{p_i^t} \{ FC_{it} \, (p_i^t, \, u_i^t) - (\lambda_1^t - \lambda_2^t) \, p_i^t \} + SC_{it}(u_i(t)) - \lambda_2^t \, u_i^t \, p_{it}^{\max} \right]$$

subject to (2.1), (2.3)

$$\tilde{P}_j(\lambda): \quad \min_{(s_j, w_j)} \mathbb{E} \sum_{t=1}^{T} \left[-\lambda_1^t \left(s_j^t - w_j^t \right) \right]$$

subject to (2.1), (2.2), (2.6)

Note that these are dynamic recourse problems themselves associated with the single generation units. As noticed in the previous section, the subgradients of $\Theta(\lambda)$ with respect to λ_1 and λ_2 are given by
$d^t - \sum_{i=1}^{I} p_i^t - \sum_{j=1}^{J} \left(s_j^t - w_j^t \right)$ and
$r^t - \sum_{i=1}^{I} (u_i^t \, p_{it}^{\max} - p_i^t)$,
where (u_i^t, p_i^t) and (s_j^t, w_j^t) are solutions of $P_i(\lambda)$, $i = 1 \ldots, I$, and $\tilde{P}_j(\lambda), \ j = 1, \ldots, J$, respectively.
Therefore, the necessary information for a nonsmooth optimization method of the kind discussed in Section 3 is at hand providing efficient algorithms for solving the subproblems are available. Consequently, we shall have established an algorithm for solving the problem (2.1) - (2.9) if the following points are clarified:

- approximation of the stochastic process d^t by a scenario tree;
- choice of an appropriate method for solving the dual problem (4.4);
- efficient algorithms for solving the subproblems $P_i(\cdot)$, $\tilde{P}_j(\cdot)$,

- gaining information from the solution of the dual problem (4.4) for computing a primal feasible solution and providing an estimation of the occurring relative duality gap.

Let us comment all these points. The stochastic process d^t can be approximated by means of an analysis of statistical data using also expert knowledge. The first thing to clarify is the nature of the demand randomness. In order to estimate the load of the system one usually uses the data of the same week from previous years, data of days with similar weather conditions, and the experience of experts. The strategy of creating scenarios has to reflect truly all possible future demands. The number of scenarios that approximate the demand has to be chosen in such a way that a fairly good approximation is obtained but the speed of the optimization procedure is not affected critically since the execution time of the algorithm grows rapidly as the number of scenarios included increases. The probability assigned to each scenario can be calculated according to the likelihood of their occurrence.
The functions FC_{it} and SC_{it} $i = 1, \dots, I,\ t = 1, \dots, T$ are assumed to be piecewise linear or quadratic. Consequently, the function $\Theta(\lambda)$ is piecewise twice continuously differentiable. Therefore, any method of non-smooth optimization of those discussed in the previous section could be applied. The methods developed as bundle methods of order higher than one could be applied successfully, e. g. [36], [41], [43]. Unfortunately, for those guaranteeing superlinear convergence ([44]), no computational code is available up to now. The variable metric bundle methods [36], [41], [43] provide convergence but no estimate of the rate is given. We would like to emphasize that those methods are finite for piecewise linear convex functions. The published experience with NOA Version 3.0 ([37]) reports fast convergence in practice (cf. [36]).
The efficiency of the optimization algorithm depends to great extent on the fast computation of the values and subgradients of the objective function $\Theta(\lambda)$. Therefore, the development of fast algorithms for solving the problems $P_i(\lambda)$ and $\tilde{P}_j(\lambda),\ i = 1, \dots, I,\ \ j = 1, \dots, J$, is important. An algorithm for solving the problems $\tilde{P}_j(\lambda),\ j = 1, \dots, J$, has been developed by Nowak [48]. It regards $\tilde{P}_j(\lambda)$ as a network-flow problem and suggests a procedure adapted to the structure called EXCHA. The crucial point in this procedure is the selection of a proper direction from a prescribed subset of descent directions for minimizing the objective of $\tilde{P}_j(\lambda)$. Let us consider the problems $P_i(\lambda),\ i = 1, \dots, I$. The inner minimization (with respect to p_i) can be done explicitly or by one-dimensional optimization. Further, a dynamic programming procedure can be used to minimize the expected costs with respect to the integer variables u_i. A state transition graph of the unit to each scenario regarding the nonanticipativity constraint can be considered. Then the solution corresponds to a tree in this graph that has minimal weighted length. In order to reduce the number of nodes, we can include the constraints (2.3) into the process of generating the state transition graph by setting nodes "off" for at least τ_i periods.
Another substantial part of the solution procedure for the dynamic recourse problem consists in developing an algorithm for the determination of a primal feasible solution after one has found a solution of the dual problem. As already established, if we find

an "almost" feasible point, it is "almost" a solution (Proposition 3.1). In addition, the optimal value $\Theta(\lambda)$ of the dual problem is a better lower bound of the objective function of the primal problem than the value of its continuous relaxation. It is possible to use some modification of the heuristic procedure presented for this purpose in [69] and further modified as in [14]. More precisely, the algorithm should work as follows:

- try to satisfy the reserve-constraints by using pumped-storage hydro plants in those time intervals, where the largest values of $d^t + r^t$ occur. If the reserve-constraints are still violated, use the procedure of [69].
- improve the feasible solution found at the end of the procedure above by solving the problem keeping the integer variables fixed. An algorithm for the latter problem is suggested in [49] that is a modification of the network-flow algorithm in [48]. The problem is considered as a network-flow problem again and the algorithm makes use of its special structure.

Summarizing, the presented solution technique includes the following basic steps:

- generation of a scenario tree (discrete approximation of d)
- solving the problem (4.4) e.g. by NOA Version 3.0,
  solving the problems $P_i(\lambda), i = 1, \dots, I$, by dynamic programming and
  $\tilde{P}_j(\lambda),\ j = 1, \dots, J$, by EXCHA.
- determination of a primal feasible solution by the procedure described above.

An illustrative example for an approximation of the load is given in Figure 3 and Figure 4 expresses the corresponding stochastic schedule for fixed binary variables. The values of the approximative load are generated by using the value of a given load, and a standard normal random variable (see [49] for details).
A final remark is due. There is an estimate for the occurring duality gap. We use the description of the problem (2.11)-(2.13) based on scenarios. At this place, we incorporate the nonanticipativity condition into the representation of the model. More precisely, we consider decisions $x^t_{in} = (u^t_{in}, p^t_{in}, s^t_{in}, w^t_{in})$ and $x^t_{i\hat{n}}$ that correspond to scenarios n and $\hat{n}$ fulfilling $d^t_n = d^t_{\hat{n}}$ for all $t = 1, \dots, t_k$ as indistinguishable up to the stage k. We use only one notation for the decisions at stage k for all scenarios that are indistinguishable up to that stage. Recall that the number of scenarios at the stage k is denoted by n_k and the number of load and reserve-constraints amounts to $2 \sum\limits_{k=1}^{K} n_k(t_{k+1} - t_k)$.

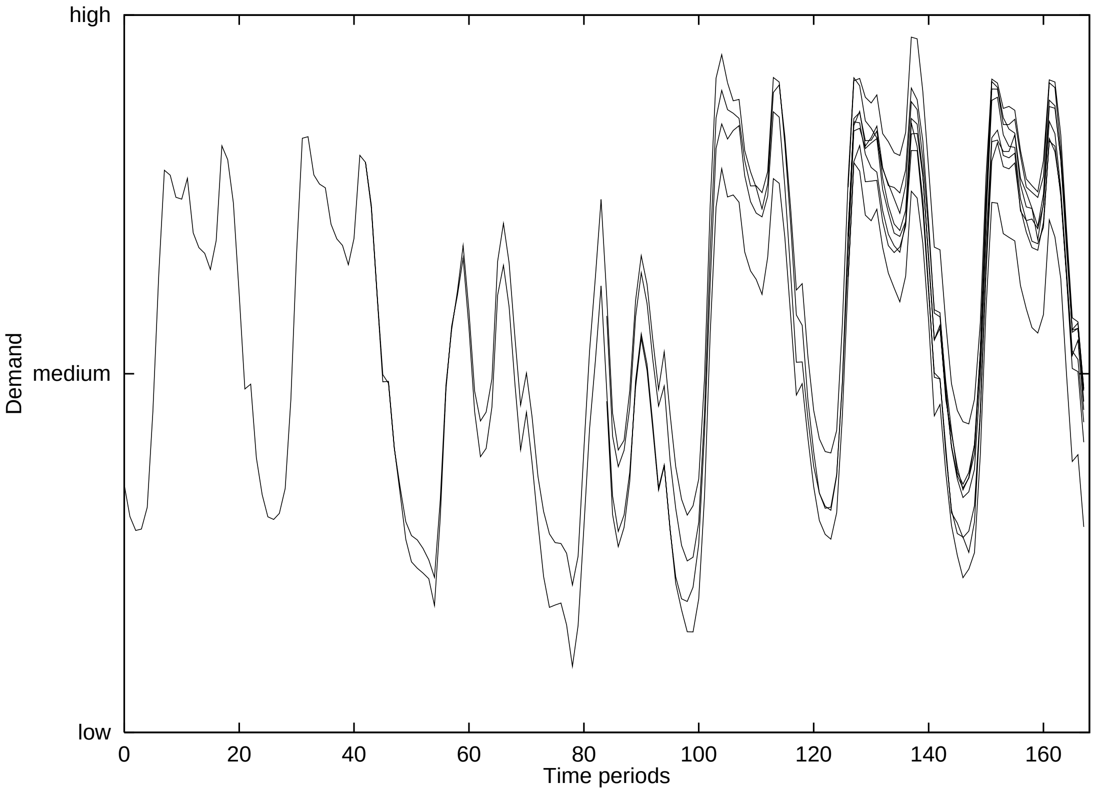

Fig. 3: sampled load scenarios

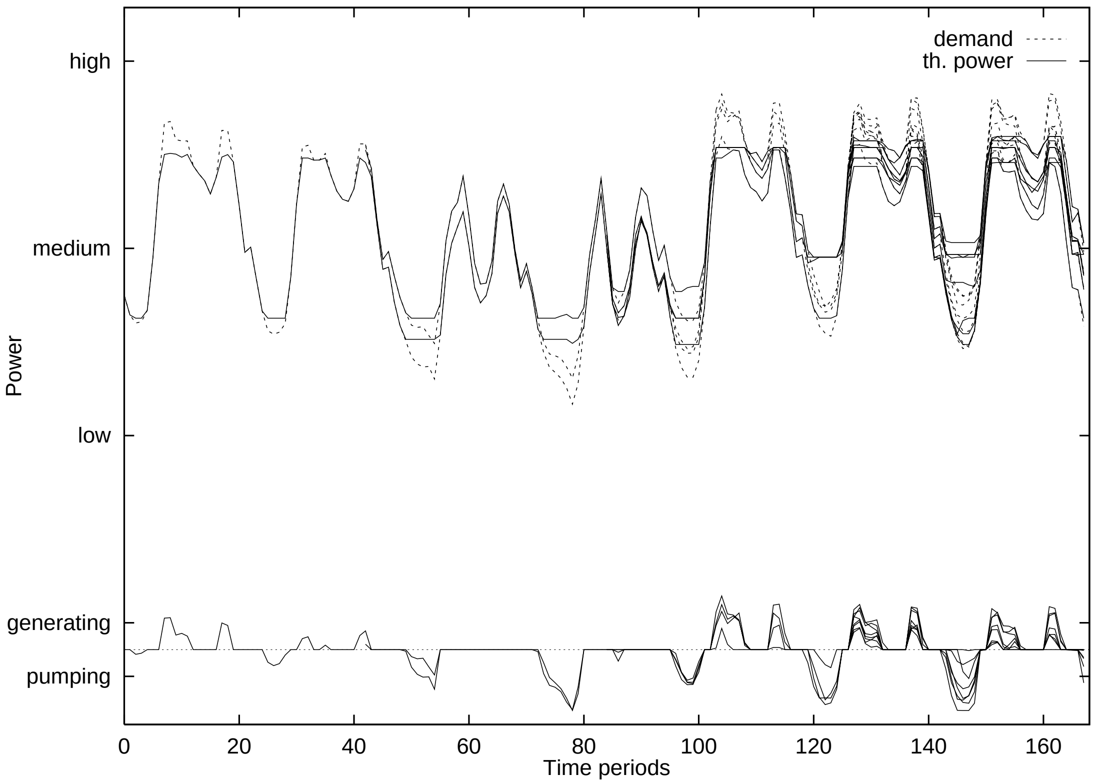

Fig. 4: solution for the load given in Figure 3

Proposition 4.2 *Assume relatively complete recourse for the dynamic recourse problem. Let its optimal value be denoted by F^* and the optimal value of its dual problem by Θ^*. Then there exists a constant ρ such that the following estimate holds true:*

$$F^* - \Theta^* \leq (2\sum_{k=1}^{K} n_k(t_{k+1} - t_k) + 1)\rho$$

Proof: The proof follows from Proposition 5.26 in [2]. We only have to show that the assumptions (A1)-(A3) made there are satisfied in our situation. (A1) is just the feasibility of the problem, which holds due to relatively complete recourse. (A2) and (A3) are easily checked specifying the required conditions. □

We consider the same dynamic recourse problem with a modified objective function:

$$\frac{1}{I\,N} I\!E \sum_{i=1}^{I}\sum_{t=1}^{T}\left[FC_{it}(p_i^t,\, u_i^t) + SC_{it}\,(u_i(t))\right]$$

The objective function in this case represents the *average costs per scenario-term.* We have the same optimal solution for both problems and the duality gap becomes

$$F^* - \Theta^* \leq \frac{2\sum\limits_{k=1}^{K} n_k(t_{k+1} - t_k) + 1}{I\,N}\rho$$

The latter inequality implies that the duality gap goes to zero as $I \to \infty$. Consequently, the duality gap becomes small for large systems independently of making the discrete approximation of the load finer ($N \to \infty$).

5 Lagrangian Relaxation for the Two-Stage Model

We consider the two-stage stochastic power production planning model elaborated in Section 2.3 under the assumption that the fuel cost functions exhibit the form (2.7). Setting $x := (u, p, s, w)$ and $\hat{x} := (\hat{u},\, \hat{p},\, \hat{s},\, \hat{w}) \in L^\infty(\Omega,\, \mathcal{A},\, \mu;\, I\!R^m)$ the optimization problem consists in minimizing the objective function

$$F(x,\, \hat{x}) := I\!E\left\{\sum_{i=1}^{I}\sum_{t=1}^{T}\left[FC_{it}\left(p_i^t + \hat{p}_i^t,\, \hat{u}_i^t\right) + SC_{it}\,(\hat{u}_i(t))\right]\right\} \tag{5.1 }$$

over all decisions $x \in I\!R^m$and $\hat{x} \in L^\infty(\Omega,\, \mathcal{A},\, \mu,\, I\!R^m)$ such that the unit capacity limits (2.1), (2.2), (2.16), (2.17) the minimum down-time constraints (2.15) and the loading and reserve constraints

$$\begin{aligned}
&\sum_{i=1}^{I} p_i^t + \sum_{j=1}^{J}(s_j^t - w_j^t) \geq I\!E(d^t),\\
&\sum_{i=1}^{I}(p_i^t + \hat{p}_i^t) + \sum_{j=1}^{J}\left(s_j^t + \hat{s}_j^t - (w_j^t + \hat{w}_j^t)\right) \geq d^t,\ \ \mu - \text{ a. s.,}\\
&\sum_{i=1}^{I}(u_i^t\, p_{it}^{\max} - p_i^t) \geq r^t,\ \ t = 1, \ldots T,
\end{aligned} \tag{5.2 }$$

respectively, are satisfied. The constraints (5.2) are coupling across units while all remaining constraints are associated with the operation of single (thermal or hydro) units. With a similar argument based on a duality statement as in the previous section, we relax the constraints (5.2) by introducing Lagrange multipliers $\lambda = (\lambda_1, \lambda_2, \lambda_3)$, where $\lambda_1, \lambda_3 \in I\!R^T$ and $\lambda_2 \in L^1(\Omega, \mathcal{A}, \mu; I\!R^T)$. The dual problem is then of the following form:

$$\max \left\{ \Theta(\lambda) : \quad \lambda \in I\!R^T \times L^1(\Omega, \mathcal{A}, \mu; I\!R^T) \times I\!R^T, \ \lambda \geq 0, \ \mu - \text{ a. s.} \right\} \tag{5.3 }$$

where

$$\Theta(\lambda) := \inf \{ L(x, \hat{x}; \lambda) : x \text{ and } \hat{x} \text{ satisfy (2.1), (2.2) and (2.15) - (2.17) } \},$$

$$\begin{aligned}
L(x, \hat{x}; \lambda) \quad := \quad & F(x, \hat{x}) + \sum_{t=1}^{T} \lambda_1^t \left(I\!E(d^t) - \sum_{i=1}^{I} p_i^t - \sum_{j=1}^{J} (s_j^t - w_j^t) \right) \\
& + I\!E \left\{ \sum_{t=1}^{T} \lambda_2^t \left(d^t - \sum_{i=1}^{I} (p_i^t + \hat{p}_i^t) - \sum_{j=1}^{J} \left(s_j^t + \hat{s}_j^t - (w_j^t + \hat{w}_j^t) \right) \right) \right\} \\
& + \sum_{t=1}^{T} \lambda_3^t \left(r^t - \sum_{i=1}^{I} (u_i^t p_{it}^{\max} - p_i^t)) \right) \\
= \quad & \sum_{i=1}^{I} \sum_{t=1}^{T} [I\!E \{ FC_{it} (p_i^t + \hat{p}_i^t, \hat{u}_i^t) + SC_{it} (\hat{u}_i(t)) - \lambda_2^t (p_i^t + \hat{p}_i^t) \} \\
& - (\lambda_1^t - \lambda_3^t) p_i^t - \lambda_3^t u_i^t p_{it}^{\max}] \\
& - \sum_{j=1}^{J} \sum_{t=1}^{T} \left[\lambda_1^t (s_j^t - w_j^t) + I\!E \left\{ \lambda_2^t \left(s_j^t + \hat{s}_j^t - (w_j^t + \hat{w}_j^t) \right) \right\} \right] \\
& + \sum_{t=1}^{T} [\lambda_1^t I\!E(d^t) + I\!E(\lambda_2^t d^t) + \lambda_3^t r^t] .
\end{aligned}$$

Hence, the dual function Θ decomposes into the form

$$\Theta(\lambda) = \sum_{i=1}^{I} \Theta_i(\lambda) + \sum_{j=1}^{J} \tilde{\Theta}_j(\lambda) + \sum_{t=1}^{T} \left[\lambda_1^t I\!E(d^t) + I\!E(\lambda_2^t d^t) + \lambda_3^t r^t \right] . \tag{5.4 }$$

Here $\Theta_i(\lambda)$ is the optimal value of a two-stage stochastic program for the (single) thermal unit i, which has the form:

$$\begin{aligned}
\min \quad & \left\{ I\!E \left\{ \sum_{t=1}^{T} [FC_{it} (p_i^t + \hat{p}_i^t, \hat{u}_i^t) - \lambda_2^t (p_i^t + \hat{p}_i^t) + SC_{it} (\hat{u}_i(t)) \right. \right. \\
& - \ (\lambda_1^t - \lambda_3^t) p_i^t - \lambda_3^t u_i^t p_{it}^{\max}] \} : \ p_{it}^{\min} \hat{u}_i^t \leq p_i^t + \hat{p}_i^t \leq p_{it}^{\max} \hat{u}_i^t, \\
& \left. p_{it}^{\min} u_i^t \leq p_i^t \leq p_{it}^{\max} u_i^t, \ \text{ and minimum down-times (2.15)} \right\}
\end{aligned} \tag{5.5 }$$

Introducing the optimal value function for the second-stage problem and taking into account the special form (2.7) of the fuel costs, the two-stage mixed-integer stochastic program (5.5) may be rewritten as

$$
\begin{aligned}
\min \{ \quad & \sum_{t=1}^{T} [(\lambda_3^t - \lambda_1^t) p_i^t - \lambda_3^t u_i^t p_{it}^{\max}] + I\!\!E \{\Phi_i(u_i, p_i; \lambda_2)\} : \\
& p_{it}^{\min} u_i^t \le p_i^t \le p_{it}^{\max} u_i^t, \ t = 1, \ldots, T, \text{ and } (2.3\ ) \Big\},
\end{aligned}
\tag{5.6 }
$$

where

$$
\begin{aligned}
\Phi_i(u_i, p_i; \lambda_2) := \quad & \inf \left\{ \sum_{t=1}^{T} \left[\max_{\ell=1,\ldots,L} f_{i\ell}(p_i^t + \hat{p}_i^t) - \lambda_2^t (p_i^t + \hat{p}_i^t) + SC_{it}(\hat{u}_i(t)) + c_i \hat{u}_i^t \right] : \right. \\
& \left. p_{it}^{\min} \hat{u}_i^t \le p_i^t + \hat{p}_i^t \le p_{it}^{\max} \hat{u}_i^t, t = 1, \ldots, T, \text{ and } (2.15\ ) \right\}.
\end{aligned}
$$

Since the minimization with respect to p_i and $\hat{p}_i$ (μ- a. s.) in (5.5) or (5.6) can be performed explicitly, the models represent two-stage stochastic combinatorial programs and can be solved by dynamic stochastic programming. Problem (5.6) simplifies essentially for the case of $\mathcal{I}_2 = \emptyset$, i. e., $\hat{u}_i^t = u_i^t$ ($i = 1, \ldots, I$, $t = 1, \ldots, T$), because the compensation program does not contain binary decisions, Φ_i enjoys a separability structure and can be computed explicitly. In the latter case (5.6) takes the form

$$
\begin{aligned}
\min \quad & \left\{ \sum_{t=1}^{T} [SC_{it}(u_i(t)) + (\lambda_3^t - \lambda_1^t) p_i^t - (\lambda_3^t p_{it}^{\max} - c_i) u_i^t] + I\!\!E \left\{ \sum_{t=1}^{T} \hat{\Phi}_{it}(u_i^t, p_i^t; \lambda_2^t) \right\} : \right. \\
& \left. p_{it}^{\min} u_i^t \le p_i^t \le p_{it}^{\max} u_i^t, t = 1, \ldots, T, \text{ and } (2.3\ ) \right\},
\end{aligned}
$$

where

$$
\hat{\Phi}_{it}(u_i^t, p_i^t; \lambda_2^t) := \inf \left\{ \max_{\ell=1,\ldots,L} f_{i\ell}(p_i^t + \hat{p}_i^t) - \lambda_2^t (p_i^t + \hat{p}_i^t) : p_{it}^{\min} u_i^t \le p_i^t + \hat{p}_i^t \le p_{it}^{\max} u_i^t \right\}.
$$

The term $\tilde{\Theta}_j(\lambda)$ in the representation (5.4) of the dual function Θ is the optimal value of the following stochastic pumped-storage subproblem for the plant j:

$$
\begin{aligned}
\min \quad & \left\{ - \sum_{t=1}^{T} (\lambda_1^t + I\!\!E(\lambda_2^t)) (s_j^t - w_j^t) + I\!\!E \left[\sum_{t=1}^{T} \lambda_2^t \left(\hat{s}_j^t - \hat{w}_j^t \right) \right] : \ (s_j, w_j) \text{ and } (\hat{s}_j, \hat{w}_j) \right. \\
& \text{satisfy } \ 0 \le s_j^t \le s_{jt}^{\max}, \ 0 \le w_j^t \le w_{jt}^{\max}, \ 0 \le \ell_j^t \le \ell_{jt}^{\max}, \\
& t = 1, \ldots, T, \text{ and } (2.2\ ), (2.16\ ) \}.
\end{aligned}
\tag{5.7 }
$$

Problem (5.7) represents a linear two-stage stochastic program, which can be solved by standard solution techniques (cf. [17], [19]).
These facts motivate a Lagrangian relaxation-based conceptual solution method for the two-stage stochastic model, which is similar to the algorithm developed in the previous section. Its basic steps are:

- Generation of scenarios $d_n, n = 1, \ldots, N$, for the load process d and replacing d by this discrete approximation;

- solving the concave dual problem (5.3) by applying appropriate nondifferentiable optimization methods (cf. Section 3), where function values and subgradients of Θ are computed by solving the single unit subproblems (5.5) and (5.7). Note that (5.3) has dimension $2TN$ and Θ is piecewise linear or quadratic,

- determining a primal feasible solution for the first-stage variables by a procedure that is similar to the method described in Section 4.

Acknowledgements
The authors wish to thank G. Scheibner, G. Schwarzbach and J. Thomas (VEAG Vereinigte Energiewerke AG) for the fruitful and productive cooperation. Further thanks are due to M.P. Nowak (Humboldt-University Berlin) for valuable comments and to C. Carøe (University of Copenhagen), R. Gollmer (Humboldt-University Berlin) and R. Schultz (Konrad-Zuse-Zentrum für Informationstechnik Berlin) for helpful discussions on the two-stage model in Section 2.3.

References

[1] K. Aoki, M. Itoh, T. Satoh, K. Nara and M. Kanezashi: Optimal long-term unit commitment in large scale systems including fuel constrained thermal and pumped-storage hydro, IEEE Transactions on Power Systems 4(1989),1065-1073.

[2] D.P. Bertsekas: Constrained Optimization and Lagrange Multiplier Methods, Academic Press, New York 1982.

[3] D.P. Bertsekas, G.S. Lauer, N.R. Sandell Jr. and T.A. Posbergh: Optimal short-term scheduling of large-scale power systems, IEEE Transactions on Automatic Control AC-28(1983),1-11.

[4] J.R. Birge: Decomposition and partitioning methods for multi-stage stochastic linear programs, Operations Research 33(1985), 989-1007.

[5] J.R. Birge, L. Qi and Z. Wei: Two methods for nonsmooth optimization, Applied Mathematics Report AMR95/4, School of Mathematics, The University of New South Wales (Sydney, Australia), 1995.

[6] J.R. Birge and C.H. Rosa: Parallel decomposition of large-scale stochastic nonlinear programs, Annals of Operations Research 64(1996), 39-65.

[7] J.F. Bonnans, J.C. Gilbert, C. Lemaréchal and C.A. Sagastizábal: A family of variable metric proximal methods, Mathematical Programming 68(1995), 15-47.

[8] D.W. Bunn and S.N. Paschentis: Development of a stochastic model for the economic dispatch of electric power, European Journal of Operational Research 27(1986),179-191.

[9] C.C. Carøe and R. Schultz: Dual decomposition in stochastic integer programming, Preprint SC 96-46, Konrad-Zuse-Zentrum für Informationstechnik Berlin, 1996.

[10] E.W. Cheney and A.A. Goldstein: Newton's method for convex programming and Tchebycheff approximation, Numerische Mathematik 1(1959), 253-268.

[11] G.B. Dantzig and G. Infanger: Approaches to stochastic programming with application to electric power systems, in [21], 125-138.

[12] M.A.H. Dempster: On stochastic programming II: Dynamic problems under risk, Stochastics 25(1988), 15-42.

[13] M.A.H. Dempster and R.T. Thompson: EVPI-based importance sampling solution procedures for multistage stochastic linear programmes on parallel MIMD architectures, Annals of Operations Research (submitted 1996).

[14] D. Dentcheva, R. Gollmer, A. Möller, W. Römisch and R. Schultz: Solving the unit commitment problem in power generation by primal and dual methods, Humboldt-Universität Berlin, Institut für Mathematik, Preprint Nr. 96-25, 1996, to appear in ECMI 96 Proceedings.

[15] J. Dupačová: Multistage stochastic programs: The state-of-the-art and selected bibliography, Kybernetika 31(1995), 151-174.

[16] J. Dupačová: Stochastic programming: Approximation via scenarios, Working Paper, Charles University Prague, 1996 and Aportaciones Matematicas (Mexican Society of Mathematics)(to appear).

[17] Y. Ermoliev and R.J.-B. Wets (Eds.): Numerical Techniques for Stochastic Optimization, Springer-Verlag, Berlin 1988.

[18] H. Everett: Generalized Lagrange multiplier method for solving problems of optimal allocation of resources, Operations Research 11(1963), 399-417.

[19] K. Frauendorfer: Stochastic Two-Stage Programming, Lecture Notes in Economics and Mathematical Systems Vol. 392, Springer-Verlag, Berlin 1992.

[20] K. Frauendorfer: Barycentric scenario trees in convex multistage stochastic programming, Mathematical Programming 75(1996), 277-294.

[21] K. Frauendorfer, H. Glavitsch and R. Bacher (Eds.): Optimization in Planning and Operation of Electric Power Systems, Physica-Verlag, Heidelberg 1993.

[22] S. Früholz: Stochastische Optimierungsmodelle zu Lastverteilungsproblemen bei unsicherem Bedarf, Diplomarbeit, Humboldt-Universität Berlin, Institut für Mathematik 1995.

[23] H.I. Gassmann: MSLiP: A computer code for the multistage stochastic linear programming problem, Mathematical Programming 47(1990), 407-423.

[24] N. Gröwe and W. Römisch: A stochastic programming model for optimal power dispatch: Stability and numerical treatment; in: Stochastic Optimization (K. Marti, Ed.), Lecture Notes in Economics and Mathematical Systems Vol. 379, Springer-Verlag, Berlin 1992, 111-139.

[25] N. Gröwe, W. Römisch and R. Schultz: A simple recourse model for power dispatch under uncertain demand, Annals of Operations Research 59(1995), 135-164.

[26] J. Guddat, W. Römisch and R. Schultz: Some applications of mathematical programming techniques in optimal power dispatch, Computing 49(1992), 193-200.

[27] J.L. Higle and S. Sen: Stochastic Decomposition - A Statistical Method for Large Scale Stochastic Linear Programming, Kluwer Academic Publishers, Dordrecht 1996.

[28] J.-B. Hiriart-Urruty and C. Lemaréchal: Convex Analysis and Minimization Algorithms I and II, Springer-Verlag, Berlin 1993.

[29] G. Infanger: Planning Under Uncertainty - Solving Large-Scale Stochastic Linear Programs, Boyd and Fraser, 1994.

[30] T. Ishikida and P.P. Varaiya: Pricing of electric power under uncertainty: information and efficiency, IEEE Transactions on Power Systems 10(1995), 884-890.

[31] J. Jacobs, G. Freeman, J. Grygier, D. Morton, G. Schultz, K. Staschus and J. Stedinger: SOCRATES: A system for scheduling hydroelectric generation under uncertainty, Annals of Operations Research 59 (1995), 99-133.

[32] P. Kall and S.W. Wallace: Stochastic Programming, Wiley, Chichester 1994.

[33] J.E. Kelley: The cutting plane method for solving convex programs, Journal of SIAM 8(1960), 703-712.

[34] K.C. Kiwiel: An aggregate subgradient method for nonsmooth convex minimization, Mathematical Programming 27(1983), 320-341.

[35] K.C. Kiwiel: Methods of Descent for Nondifferentiable Optimization, Lecture Notes in Mathematics Vol. 1133, Springer-Verlag, Berlin 1985.

[36] K.C. Kiwiel: Proximity control in bundle methods for convex nondifferentiable minimization, Mathematical Programming 46(1990), 105-122.

[37] K.C. Kiwiel: User's Guide for NOA 2.0/3.0: A Fortran package for convex nondifferentiable optimization, Polish Academy of Sciences, Systems Research Institute, Warsaw (Poland), 1993/94.

[38] B. Kummer: Newton's method based on generalized derivatives for non-smooth functions: convergence analysis, in: Advances in Optimization (W. Oettli and D. Pallaschke Eds.), Lecture Notes in Economics and Mathematical Systems Vol. 382, Springer-Verlag, Berlin 1992, 171-194.

[39] C. Lemaréchal: Bundle methods in nonsmooth optimization, in: C. Lemaréchal and R. Mifflin (Eds.), Nonsmooth Optimization, Pergamon Press, Oxford 1978.

[40] C. Lemaréchal: Lagrangian decomposition and nonsmooth optimization: Bundle algorithm, prox iteration, augmented Lagrangian, in: Nonsmooth Optimization Methods and Applications (F. Gianessi Ed.), Gordon and Breach, Amsterdam 1992, 201-216.

[41] C. Lemaréchal and C.A. Sagastizábal: An approach to variable metric bundle methods, in: System Modelling and Optimization (J. Henry and J.P. Yvor Eds.), Lecture Notes in Control and Information Sciences Vol. 197, Springer-Verlag, New York 1994, 144-162.

[42] C. Lemaréchal and J. Zowe: A condensed introduction to bundle methods in nonsmooth optimization, DFG-Schwerpunktprogramm "Anwendungsbezogene Optimierung und Steuerung", Report No. 495, 1994.

[43] R. Mifflin: A quasi-second-order proximal bundle algorithm, Mathematical Programming 73(1996), 51-72.

[44] R. Mifflin, L. Qi and D. Sun: Properties of the Moreau-Yosida regularization of a piecewise C^2 convex function, manuscript, School of Mathematics, The University of New South Wales (Sydney, Australia), 1995.

[45] J.M. Mulvey and A. Ruszczyński: A new scenario decomposition method for large-scale stochastic optimization, Operations Research 43(1995), 477-490.

[46] G.L. Nemhauser and L.A. Wolsey: Integer and Combinatorial Optimization, Wiley, New York 1988.

[47] V.I. Norkin, Y. Ermoliev and A. Ruszczyński: On optimal allocation of indivisibles under uncertainty, International Institute of Applied Systems Analysis (Laxenburg, Austria), Working Paper WP-94-21, 1994.

[48] M.P. Nowak: A fast descent method for the hydro storage subproblem in power generation, International Institute for Applied Systems Analysis (Laxenburg, Austria), Working Paper WP-96-109, 1996.

[49] M.P. Nowak and W. Römisch: Optimal power dispatch via multistage stochastic programming, Humboldt-Universität Berlin, Institut für Mathematik, Preprint Nr. 96-26, 1996, to appear in ECMI 96 Proceedings.

[50] J.-S. Pang and L. Qi: Nonsmooth equations: Motivation and algorithms, SIAM Journal on Optimization 3(1993), 443-465.

[51] M.V.F. Pereira and L.M.V.G. Pinto: Multi-stage stochastic optimization applied to energy planning, Mathematical Programming 52 (1991), 359-375.

[52] A. Prékopa: Stochastic Programming, Kluwer Academic Publishers, Dordrecht 1995.

[53] L. Qi and H. Jiang: Karush-Kuhn-Tucker equations and convergence analysis of Newton methods and Quasi-Newton methods for solving these equations, Applied Mathematics Report AMR94/5, School of Mathematics, The University of New South Wales (Sydney, Australia), 1994.

[54] L. Qi and J. Sun: A nonsmooth version of Newton's method, Mathematical Programming 58(1993), 353-367.

[55] R.T. Rockafellar and R.J.-B. Wets: Nonanticipativity and L^1-martingales in stochastic optimization problems, Mathematical Programming Study 6(1976), 170-187.

[56] R.T. Rockafellar and R.J.-B. Wets: The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints, SIAM Journal on Control and Optimization 16(1978), 16-36.

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

[58] W. Römisch and R. Schultz: Decomposition of a multi-stage stochastic program for power dispatch, ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik 76(1996) Suppl.3, 29-32.

[59] C.H. Rosa and A. Ruszczyński: On augmented Lagrangian decomposition methods for multistage stochastic programs, Annals of Operations Research 64(1996), 289-309.

[60] A. Ruszczyński: A regularized decomposition method for minimizing a sum of polyhedral functions, Mathematical Programming 35(1986), 309-333.

[61] R. Schultz: Discontinuous optimization problems in stochastic integer programming, ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik 76(1996) Suppl.3, 33-36.

[62] R. Van Slyke and R. J.-B. Wets: L-shaped linear programs with applications to optimal control and stochastic programming, SIAM Journal on Applied Mathematics 17(1969), 638-663.

[63] H. Schramm and J. Zowe: A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results, SIAM Journal on Optimization 2(1992), 121-152.

[64] S. Takriti, J.R. Birge and E. Long: A stochastic model for the unit commitment problem, IEEE Transactions on Power Systems 11(1996), 1497-1508.

[65] S. Takriti, J.R. Birge and E. Long: Intelligent unified control of unit commitment and generation allocation, Technical Report 94-26, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor 1994.

[66] H. Wacker (Ed.): Applied Optimization Techniques in Energy Problems, Teubner, Stuttgart 1985.

[67] R. J.-B. Wets: Stochastic Programming, in: Handbooks in Operations Research and Management Science, Vol. 1, Optimization (G.L. Nemhauser, A.H.G. Rinnooy Kan, M.J. Todd Eds.), North-Holland, Amsterdam 1989, 573-629.

[68] A.J. Wood and B.F. Wollenberg: Power Generation, Operation, and Control (Second Edition), Wiley, New York 1996.

[69] F. Zhuang and F.D. Galiana: Towards a more rigorous and practical unit commitment by Lagrangian relaxation, IEEE Transactions on Power Systems 3(1988), 763-773.