Stochastic Lagrangian Relaxation applied to Power Scheduling in a Hydro-Thermal System under Uncertainty

M.P. Nowak and W. Römisch

Institut für Mathematik
Humboldt-Universität Berlin
10099 Berlin, Germany

Abstract

A dynamic (multi-stage) stochastic programming model for the weekly cost-optimal generation of electric power in a hydro-thermal generation system under uncertain load is developed. The model involves a large number of mixed-integer (stochastic) decision variables and constraints linking time periods and operating power units. A stochastic Lagrangian relaxation scheme is designed by assigning (stochastic) multipliers to all constraints coupling power units. It is assumed that the stochastic load process is given (or approximated) by a finite number of realizations (scenarios) in scenario tree form. Solving the dual by a bundle subgradient method leads to a successive decomposition into stochastic single (thermal or hydro) unit subproblems. The stochastic thermal and hydro subproblems are solved by a stochastic dynamic programming technique and by a specific descent algorithm, respectively. A Lagrangian heuristics that provides approximate solutions for the first stage (primal) decisions starting from the optimal (stochastic) multipliers is developed. Numerical results are presented for realistic data from a German power utility and for numbers of scenarios ranging from 5 to 100 and a time horizon from 7 to 9 days. The sizes of the corresponding optimization problems go up to 200.000 binary and 350.000 continuous variables, and more than 500.000 constraints.

1 Introduction

Mathematical models for the efficient operation of electric power generation systems often lead to rather complex optimization problems. In particular, they are characterized by combinations of challenges like mixed-integer decisions, nonlinear costs, large dimensions and data uncertainty. The latter aspect mostly concerns uncertainties of electrical load forecasts, of generator failures, of flows to hydro reservoirs or plants, and of fuel or electricity prices (cf. [12, 13, 15, 18, 29] for earlier relevant work). The present paper aims at treating power optimization in a hydro-thermal system under uncertain electrical load. More precisely, a generation system comprising thermal units and pumped hydro storage plants as encountered at the German utility VEAG Vereinigte Energiewerke AG Berlin is considered. The relevant mathematical optimization model contains a large number of binary and continuous variables, constraints and uncertainty appearing in the load constraints. The time horizon is 7 to 9 days as it is needed for the efficient weekly operation of hydro-thermal systems involving weekly load and pumping cycles.

The machinery of stochastic programming offers modelling and solution techniques for such optimization problems under uncertainty. In the present paper, a multi-stage stochastic programming model in which the expected production costs are minimized and stages refer to the availability of further observations of the load is developed. In particular, the first stage refers to the time period for which a reliable load forecast is available. The attention is focused on the (deterministic) first-stage scheduling decisions (on/off and outputs), which are obtained by minimizing the total expected generation costs and, hence, hedge against uncertainty. Since the stochastic programming model contains mixed-integer decisions in all stages and is large-scale, new questions on the design of solution algorithms are raised.

Nowadays, solution methods are well developed for linear multi-stage stochastic programs without integrality constraints (cf. the monographs [3, 17, 16, 38] and the state-of-the-art surveys [2, 34]). Recently, progress has been made for mixed-integer stochastic programming models and applications to power optimization. The following algorithmic approaches for mixed-integer multi-stage models seem to be known in the literature: (a) Stochastic branch and bound methods ([26]), (b) scenario decomposition by splitting methods combined with suitable heuristics ([32], [25], [36, 37]), (c) scenario decomposition combined with branch and bound ([6, 7]), (d) stochastic (augmented) Lagrangian relaxation of coupling constraints ([8, 30], [10, 33]). The approaches in (b) and (c) are based on a successive decomposition of the stochastic program into finitely many deterministic (or scenario) programs, which may be solved by available conventional techniques. The idea of (d) is a successive decomposition into finitely many smaller stochastic subproblems for which (efficient) solution techniques have to be developed eventually. Due to the non-convexity of the underlying stochastic program, the successive decompositions in (b)-(d) have to be combined with certain global optimization techniques (branch-and-bound, heuristics etc.).

The approach followed in the present paper consists in a *stochastic version* of the classical Lagrange relaxation idea ([23]), which is very popular in power optimization ([1, 11, 14, 24, 35, 39, 40]). Since the corresponding coupling constraints contain

random variables, stochastic multipliers are needed for the dualization, and the dual problem represents a nondifferentiable stochastic program. Subsequently, the approach is based on the same, but *stochastic*, ingredients as in the classical case: a solver for the nondifferentiable dual, subproblem solvers, and a Lagrange heuristics. It turns out that, with a state-of-the-art bundle method for solving the dual, efficient stochastic subproblem solvers based on a specific descent algorithm and stochastic dynamic programming, respectively, and a specific Lagrange heuristics for determining a nearly optimal first-stage solution, this *stochastic Lagrangian relaxation* algorithm becomes efficient.
The paper is organized as follows. In Section 2 a detailed description of the hydro-thermal generation system is given and the stochastic programming model is developed. Section 3 describes the stochastic Lagrangian relaxation approach together with its components: algorithms for solving the stochastic dual, single-unit subproblems and economic dispatch problems, and the Lagrange heuristics. For all (sub)algorithms, numerical experience is provided. Finally, numerical results for the stochastic Lagrangian relaxation based algorithm are reported in Section 4 for realistic data and the VEAG-owned hydro-thermal generation system.

2 Model

We consider a power generation system comprising (coal-fired and gas-burning) thermal units, pumped hydro storage plants and delivery contracts, and describe a model for its weekly cost-optimal generation under uncertainty on the electrical load (cf. [10], [28]). Let T denote the number of time intervals obtained by discretizing the operation horizon. This discretization may be chosen uniformly (e. g. hourly) or non-uniformly. Let I and J denote the number of thermal and pumped hydro storage units in the system. Delivery contracts are regarded as particular thermal units. The decision variables in the model correspond to the outputs of units, i. e., the electric power generated or consumed by each unit of the system. These decision variables are denoted by

$$\mathbf{u}_i^t,\ \mathbf{p}_i^t,\ i=1,\ldots,I,\ \text{ and } \mathbf{s}_j^t, \mathbf{w}_j^t,\ j=1,\ldots,J,\ t=1,\ldots,T,$$

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

$$\begin{array}{l} p_i^{min}\mathbf{u}_i^t \leq p_i^t \leq p_i^{max}\mathbf{u}_i^t,\ \mathbf{u}_i^t \in \{0,1\},\ i=1,\ldots,I,\ t=1,\ldots,T, \\ 0 \leq \mathbf{s}_j^t \leq s_j^{max},\ 0 \leq \mathbf{w}_j^t \leq w_j^{max}, \\ 0 \leq \mathbf{l}_j^t \leq l_j^{max},\ j=1,\ldots,J,\ t=1,\ldots,T. \end{array} \tag{1}$$

The constants $p_i^{min}, p_i^{max}, s_j^{max}, w_j^{max}$, and l_j^{max} denote the minimal/maximal outputs of the units and the maximal storage volumes in the upper reservoirs, respectively. The dynamics of the storage volume, which is measured in electrical energy, is modelled by the equations:

$$\begin{array}{rcll} \mathbf{l}_j^t & = & \mathbf{l}_j^{t-1} - \mathbf{s}_j^t + \eta_j \mathbf{w}_j^t & ,\quad t=1,\ldots,T, \\ \mathbf{l}_j^0 & = & l_j^{in},\ \ \mathbf{l}_j^T = l_j^{end} & ,\quad j=1,\ldots,J. \end{array} \tag{2}$$

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 (or pumping) 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) show, in particular, that there occur no in- or outflows in the upper reservoirs and, hence, that the storage plants of the system operate with a constant amount of water. Together with the upper and lower bounds for $\mathbf{l}_j^t$ the equations (2) mean that certain reservoir constraints have to be maintained for all 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 thermal stress and 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:

$$\mathbf{u}_i^{t-1} - \mathbf{u}_i^t \leq 1 - \mathbf{u}_i^\tau, \quad \tau = t+1,\ldots,\min\{t+\tau_i-1,\ T\},\ t=1,\ldots,T. \tag{3}$$

Analogous constraints can be formulated describing minimum up-times. The next constraints are coupling across power units: the load and reserve constraints. The first constraints are essential for the operation of the power system and express that the sum of the output powers is greater than or equal to the load demand in each time period. Denoting by $\mathbf{d}^t$ the load demand during period t, the load constraints are described by the inequalities:

$$\sum_{i=1}^{I} \mathbf{p}_i^t + \sum_{j=1}^{J} (\mathbf{s}_j^t - \mathbf{w}_j^t) \geq \mathbf{d}^t,\ t=1,\ldots,T. \tag{4}$$

In order to compensate unexpected events (e.g. sudden load increases or decreases, outages of units) 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. The corresponding constraints are given by the following inequalities:

$$\sum_{i=1}^{I}(p_i^{max}\mathbf{u}_i^t - \mathbf{p}_i^t) \geq \mathbf{r}^t,\ t = 1, \ldots, T, \tag{5}$$

where $\mathbf{r}^t > 0$ is the spinning reserve in period t, which is assumed to be proportional to $\mathbf{d}^t$. 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.,

$$\sum_{i=1}^{I}\sum_{t=1}^{T}[FC_i(\mathbf{p}_i^t,\ \mathbf{u}_i^t) + SC_i^t(\mathbf{u}_i)], \tag{6}$$

where FC_i are the fuel costs for the operation of the thermal unit i during period t and SC_i^t are the start-up costs for getting the unit on-line in this period. We assume that each FC_i is piecewise linear convex, strictly monotonically increasing and of the form

$$FC_i(\mathbf{p},\ \mathbf{u}) = \max_{l=1,\ldots,L}\{a_{il}\,\mathbf{p} + b_{il}\,\mathbf{u}\}\ , \tag{7}$$

where a_{il} and b_{il} are fixed cost coefficients. The start-up costs $SC_i^t(\mathbf{u}_i)$ may vary from a maximum cold-start value to a much smaller value when the unit i is still relatively close to its operation temperature. The following description of start-up costs reflects this dependence on the down-time:

$$SC_i^t(\mathbf{u}_i) = \max_{\tau=0,\ldots,\tau_i^c} c_i^\tau(\mathbf{u}_i^t - \sum_{\kappa=1}^{\tau}\mathbf{u}_i^{t-\kappa}),$$

where $c_i^0 = 0$ and c_i^τ, $\tau = 0, \ldots, \tau_i^c$ are fixed increasing cost coefficients, τ_i^c is the time the unit i needs to cool down, and $c_i^{\tau_i^c}$ its maximum cold-start costs. Altogether, minimizing the objective function (6) subject to the constraints (1)-(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 $(\mathbf{u},\ \mathbf{p},\ \mathbf{s},\ \mathbf{w})$ is optimal, then the load constraints (4) are typically satisfied with equality and we have $\mathbf{s}_j^t\mathbf{w}_j^t = 0$ for all $j = 1, \ldots, J$, $t = 1, \ldots, T$, i. e., generation and pumping do not occur simultaneously (cf. [15]).
The minimization problem (1)-(6) represents a mixed-integer program with linear constraints, and IT binary and $(I+2J)T$ continuous decision variables, respectively. For a typical configuration of the VEAG-owned generation system with $I = 25$ (thermal), $J = 7$ (hydro) and $T = 168$ (i. e., 7 days with hourly discretization), the dimension of the model is shown in the first row of Figure 3. Figure 1 shows a typical load curve of a peak load week and a corresponding cost-optimal hydro-thermal

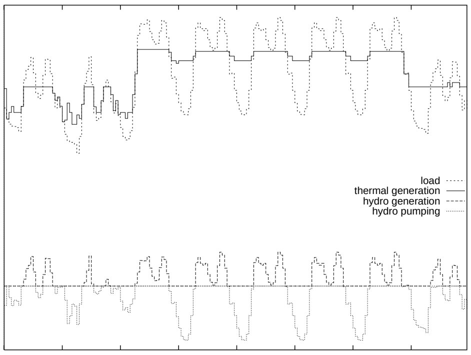

Figure 1: Load curve and hydro-thermal schedule

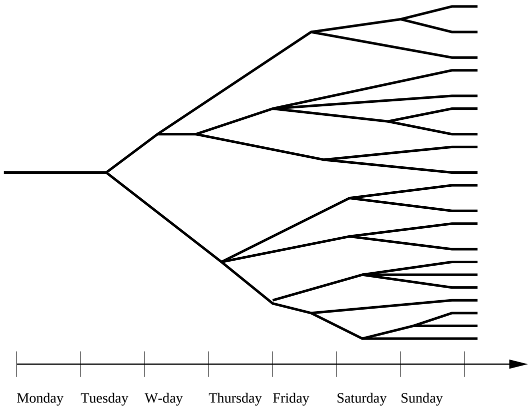

Figure 2: Example for a Scenario Tree

schedule. The load curve in Figure 1 exhibits two overlapping cycles: a daily and weekly cycle. Pumped hydro storage plants are designed to exploit these two cycles by saving fuel costs when serving the peak load with hydro-energy and 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 hydro storage plants. The remaining load, i. e., the difference between the original system load and the hydro schedule, shows a more uniform structure than the original load. This portion of the load is covered by the total output of thermal units. So far we have tacitly assumed that the electrical load is given and 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. But, clearly, the actual electrical load may deviate from the predicted load at any time period due to various unforeseeable (random) influences (temperature, daylight, switch off of local consumers etc.). This gives rise to a stochastic model of the electrical load $\{\mathbf{d}^t : t = 1, \ldots, T\}$ as a (discrete-time) stochastic process on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ reflecting 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 σ-field defined by the random vector $(\mathbf{d}^1, \ldots, \mathbf{d}^t)$. Hence, we have $\{\emptyset, \Omega\} = \mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \ldots \subseteq \mathcal{A}_t \subseteq \ldots \subseteq \mathcal{A}_T \subseteq \mathcal{A}$. The sequence of scheduling decisions $\{(\mathbf{u}^t, \mathbf{p}^t, \mathbf{s}^t, \mathbf{w}^t) : t = 1, \ldots, T\}$ also forms a stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$, which is assumed to be adapted to the filtration of σ-fields, i.e., *non-anticipative*. The latter condition means that the decision $(\mathbf{u}^t, \mathbf{p}^t, \mathbf{s}^t, \mathbf{w}^t)$ depends only on the data history $(\mathbf{d}^1, \ldots, \mathbf{d}^t)$ or, equivalently, that $(\mathbf{u}^t, \mathbf{p}^t, \mathbf{s}^t, \mathbf{w}^t)$ is $\mathcal{A}_t$-measurable. Since all decision variables are uniformly bounded, we may restrict our attention to decisions $(\mathbf{u}, \mathbf{p}, \mathbf{s}, \mathbf{w})$ belonging to $L^\infty(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^m)$, where $m := 2(I + J)T$. Then the non-anticipativity condition can be formulated equivalently as

$$(\mathbf{u}, \mathbf{p}, \mathbf{s}, \mathbf{w}) \in \mathop{\times}_{t=1}^{T} L^\infty \left(\Omega, \mathcal{A}_t, \mathbb{P}; \mathbb{R}^{2(I+J)}\right), \tag{8}$$

and the (stochastic) optimization problem consists in minimizing the expected cost (cf. (6))

$$\mathbb{E}\{\sum_{i=1}^{I}\sum_{t=1}^{T}[FC_i(\mathbf{p}_i^t, \mathbf{u}_i^t) + SC_i^t(\mathbf{u}_i)]\} \tag{9}$$

over all decisions ($\mathbf{u}$, $\mathbf{p}$, $\mathbf{s}$, $\mathbf{w}$) satisfying the non-anticipativity constraint (8), and $\mathbb{P}$-almost surely, the constraints (1)-(5). Among the constraints (1)-(5), (2) and (3) reflect the dynamics of the model and (4), (5) couple power units. Altogether, the stochastic program involves $2(I + J)T$ stochastic decision variables. It is a discrete-time dynamic or multi-stage stochastic recourse problem, where the *stages* correspond to steps in the decision process at which new observations of the stochastic load are taken into account. For the numerical solution of the dynamic recourse model we now assume that an (approximate) *discrete* multivariate probability distribution of the stochastic load vector $\mathbf{d} = (\mathbf{d}^1, \ldots, \mathbf{d}^T)$ is given, such that its support consists of finitely many atoms or *scenarios* and that the non-anticipativity constraint (8) is satisfied. This approximation of the load can be represented in the form of a *scenario tree*. Each path of the tree from the root to a leaf corresponds to one scenario; each node of the tree corresponds to a component of the decision ($\mathbf{u}$, $\mathbf{p}$, $\mathbf{s}$, $\mathbf{w}$). Figure 2 shows an example of a load scenario tree over a weekly time horizon, where new observations of the electrical load lead to a number of additional daily scenarios. Since the decision variable ($\mathbf{u}$, $\mathbf{p}$, $\mathbf{s}$, $\mathbf{w}$) exhibits the same tree structure as the load, the model may easily become extremely large if the number of nodes in the scenario tree increases. Figure 3 shows how the dimension of the model (1)-(5), (8), (9) increases with the number of scenarios for a scenario tree with equidistant binary branches (without taking into account the constraints of type (3) and the objective function).

Scenarios	Nodes	Variables		Constraints	Nonzeros
		binary	continuous		
1	168	4200	6652	13441	19657
5	462	11550	18018	36965	54059
10	756	18900	29484	60490	88462
20	1176	29400	45864	94100	137612
30	1663	41575	64857	133070	194601
50	2478	61950	96642	198290	289976
80	3696	92400	144144	295760	432512
100	4200	105000	163800	336100	491500

Figure 3: Dimension of the mixed-integer LP depending on the number of scenarios with T=168, I=25 and J=7

3 Stochastic Lagrangian Relaxation

The huge size of the model, described in the previous section, prevents the application of state-of-the-art mixed-integer LP solvers. However, decomposition techniques may provide a practicable alternative. Here, we make use of the fact that the model is loosely coupled with respect to the operation of different units. Associating stochastic Lagrange multipliers with the coupling constraints (4) and (5) leads to the Lagrangian L and the dual function D:

$$L(\mathbf{u}, \mathbf{p}, \mathbf{s}, \mathbf{w}; \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbb{E} \sum_{t=1}^{T} \{ \sum_{i=1}^{I} [FC_i(\mathbf{p}_i^t, \mathbf{u}_i^t) + SC_i^t(\mathbf{u}_i)] \tag{10}$$

$$+ \boldsymbol{\lambda}^t (\mathbf{d}^t - \sum_{i=1}^{I} \mathbf{p}_i^t - \sum_{j=1}^{J} (\mathbf{s}_j^t - \mathbf{w}_j^t)) + \boldsymbol{\mu}^t (\mathbf{r}^t - \sum_{i=1}^{I} (\mathbf{u}_i^t p_i^{max} - \mathbf{p}_i^t) \} \tag{11}$$

$$D(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{(\mathbf{u}, \mathbf{p}, \mathbf{s}, \mathbf{w})} L(\mathbf{u}, \mathbf{p}, \mathbf{s}, \mathbf{w}; \boldsymbol{\lambda}, \boldsymbol{\mu}) \ , \tag{12}$$

where the minimization in (12) is subject to the remaining single unit constraints (1), (2), (3) and (8). Justified by general duality results for convex multi-stage stochastic programs (see [31] and the review in Section 4 of [10]) we consider the dual problem

$$\max \{ D(\boldsymbol{\lambda}, \boldsymbol{\mu}) : (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathop{\times}_{t=1}^{T} L^1(\Omega, \mathcal{A}_t, \mathbb{P}; \mathbb{R}_+^2) \} . \tag{13}$$

In particular, this means that the stochastic multiplier processes $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are non-negative $\mathbb{P}$-almost surely and adapted to the filtration $\{\mathcal{A}_t : t = 1, \ldots, T\}$ generated by the load process. Hence, they exhibit the same tree structure as $\mathbf{d}$. Furthermore, the dimension of the dual problem (13) is twice the number N of nodes in the scenario tree. The optimal value of the dual problem (13) provides a lower bound for the optimal costs of the nonconvex (primal) model. For a discussion of the (relative) duality gap in our context of power optimization, the reader is referred to [1], [24] and Section 4 in [10]. Due to the relaxation of the coupling constraints (4) and (5), the minimization in (12) decomposes into stochastic single unit subproblems and the dual function takes the form

$$D(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \sum_{i=1}^{I} D_i(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \sum_{j=1}^{J} \hat{D}_j(\boldsymbol{\lambda}) + \mathbb{E} \sum_{t=1}^{T} [\boldsymbol{\lambda}^t \mathbf{d}^t + \boldsymbol{\mu}^t \mathbf{r}^t] \ , \tag{14}$$

where $D_i(\boldsymbol{\lambda}, \boldsymbol{\mu})$ and $\hat{D}_j(\boldsymbol{\lambda})$ refer to the optimal values of the thermal subproblem (21) and the hydro storage subproblem (15), respectively. The dual function D is concave and nondifferentiable on $\mathbb{R}^{2N}$, and, in fact, polyhedral due to (7).
Similar to the deterministic case, the stochastic Lagrangian relaxation algorithm for solving the model in Section 2 consists of the following ingredients:

(a) Maximization of the dual function D by a proximal bundle method using function and subgradient information (Sect. 3.3);
(b) Efficient solvers for the stochastic single unit subproblems: stochastic dynamic programming (Sect. 3.2) and a specific descent algorithm (Sect. 3.1);
(c) Lagrange heuristics for determining a feasible first-stage decision (Sect. 3.4);
(d) Economic dispatch for determining an approximate solution for the optimal first-stage decision (Sect. 3.5).

In the remaining part of this section we provide a detailed description of all these ingredients.

3.1 Descent Algorithm for Stochastic Storage Problems

The subproblem (15) - (17), which corresponds to the hydro storage plant j, is a linear multi-stage stochastic program:

$$\hat{D}_j(\boldsymbol{\lambda}) = \min_{(\mathbf{s}_j,\mathbf{w}_j)} \left\{ \mathbb{E} \sum_{t=1}^{T} [\boldsymbol{\lambda}^t(\mathbf{w}_j^t - \mathbf{s}_j^t)] : \ (\mathbf{s}_j, \mathbf{w}_j) \in \mathop{\times}_{t=1}^{T} L^\infty(\Omega, \mathcal{A}_t, \mathbb{P}; \mathbb{R}^2)\,, \right. \tag{15}$$

$$0 \le \mathbf{s}_j^t \le s_j^{max}, \ 0 \le \mathbf{w}_j^t \le w_j^{max}, \ t = 1, \ldots, T, \tag{16}$$

$$\left. 0 \le \mathbf{l}_j^t \le l_j^{max}, \ \mathbf{l}_j^t = \mathbf{l}_j^{t-1} - \mathbf{s}_j^t + \eta_j \mathbf{w}_j^t, \ t = 1, \ldots, T, \ \mathbf{l}_j^0 = l_j^{in}, \ \mathbf{l}_j^T = l_j^{end} \right\}. \tag{17}$$

In this problem the storage volume variable $\mathbf{l}_j^t$ plays the role of a resource state variable, which means that the variables for $t > t_0$ and $t < t_0$ do not influence each other when $\mathbf{l}_j^t$ for $t = t_0$ is fixed. The equation (17) describing the dynamics of the system is one-dimensional. Hence, the storage volume can be increased and decreased using $\mathbf{w}$ and $\mathbf{s}$, respectively. The costs of changing the storage volume, i.e., $\boldsymbol{\lambda}^{t_0}(\Delta\mathbf{w}_j^{t_0} - \Delta\mathbf{s}_j^{t_0})$, have to be compared with the changes of costs in all subsequent time periods, i.e.,

$$\mathbb{E} \sum_{t=t_0+1}^{T} [\boldsymbol{\lambda}^t(\Delta\mathbf{w}_j^t - \Delta\mathbf{s}_j^t)], \tag{18}$$

to find out whether an alteration of the storage volume leads to a decrease of the objective function or not. If such a change of the storage volume $\mathbf{l}_j^t$ in any node does not lead to a decrease of the objective function, then the current point is optimal. The subsequent costs are caused by changes of variables in the subtree in order to satisfy the balance (17). Since the problem has a special structure, elements $(\Delta\mathbf{s}_j^t, \Delta\mathbf{w}_j^t)$ yielding a minimal value of (18) have many zero components. In [27] it is shown that the search for descent directions may be restricted to such elements. Moreover, the non-zero components describe a subtree of the scenario tree. Then, the conditions on step lengths and on steps to be descent directions take simpler forms. The construction of these subtrees is done in a systematic way starting at

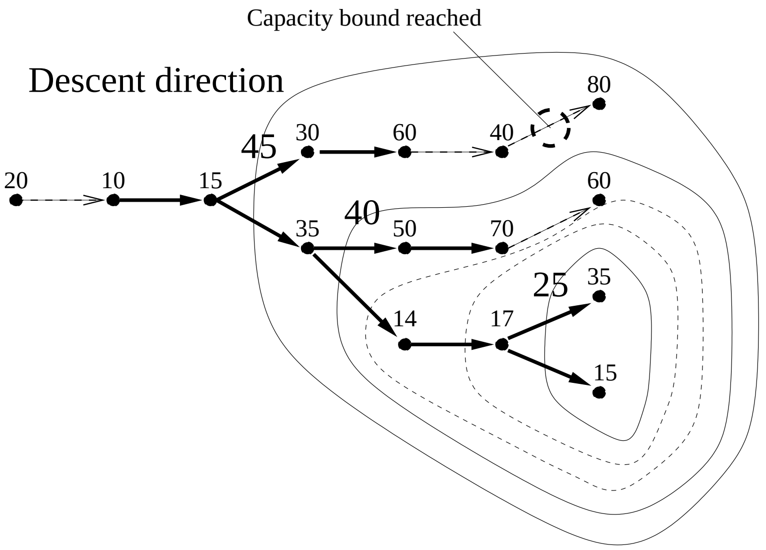

Figure 4: Example of a simple descent direction

the leaves and determining which nodes should be leaves in such subtrees. This is explained next. Figure 4 shows an example with 4 scenarios having identical probabilities (i.e. $\frac{1}{4}$) and 7 stages. The small numbers at the nodes represent the values $\boldsymbol{\lambda}^t$. The subtree mentioned above is marked with thick arrows (starting at the node with value 10, ending at nodes with 60, 70, 35, and 15). For the moment and without loss of generality we assume that $\eta_j = 1$, thus there is just one variable x for changing the storage volume. To simplify the presentation, we take up the position of the storage operator. Then, the $\boldsymbol{\lambda}$-values represent prices for buying and selling a certain commodity, and the aim is to maximize the profit. In addition, increasing and decreasing the storage volume by a certain amount could be understood as buying and selling a certain amount, respectively. Assume that a certain amount is bought at the second stage to a price of 10. The price paid at this node has to be compared with the gain from selling the amount at some nodes in the subtree in order to keep the balance (17). Each node in the subtree is examined in order to determine whether the amount should be sold at this node or should be kept for the subsequent nodes. If the amount is kept up to the last stage, it has to be sold in any case. In our example, the gain is 15 in the lowest scenario. If this happens, the amount is also sold at the last stage of the second lowest scenario due to the stochastic nature of the problem. Hence, the gain is 35. The average gain of these two scenarios is 25, shown at the lowest inner solid surrounding. The comparison of this average gain with the price at the node before, i.e. 17, leads to the decision to keep the amount up to the last stage. Hence, it follows that the gain of selling the amount at this node or later is 25. This is denoted by a surrounding with a dotted line, indicating that it has the same value as the inner one. The decision at the node before, i.e., the result of the comparing the value of the surrounded subtree (25) with the value at the node (14), leads to the same decision (to keep the amount) except for the fact that, at this point, it is out of interest at which node the amount is actually sold. In the last but one stage of the second scenario the comparison of the value for the last stage (60) with the one of the stage before (70) yields the decision to sell the amount at that node, i.e. at the last but one stage. The uppermost scenario indicates the case where keeping up to the last stage is not feasible due to capacity bounds. Hence, the comparison leads to selling at the node with value 60. Applying the same analysis to all nodes yields where the amount should be kept and where it should be sold in order to get maximal gain from buying at the second stage. A flow from the second stage to subsequent stages is associated with this maximal gain, which corresponds to a subtree denoted by thick arrows in Figure 4. Note that the leaves of this subtree correspond to nodes where the decision is selling. Further, these decisions are independent of the node at which this subtree starts. In case the storage is not empty at the first stage, it is also feasible to sell first and to buy back the amount later. However, this can be treated in a similar way and leads to a second set of binary decisions. After this analysis has been applied to all nodes of the scenario tree, a descent direction examining all nodes just once can be found. For technical details of this method and for the case of $\eta_j \neq 1$ the reader is referred to [27]. Here, we only sketch the conditions on the existence of a descent direction. The variables and decisions for the case of an increased storage are denoted by the superscript *up*, while *down* refers to the case of a decreased storage. The decision to reduce the storage is denoted by $b_k^{up} = 1$, whereas $b_k^{up} = 0$ refers to the decision to keep it. Similarly, the notations $b_k^{down} = 1$ and $b_k^{down} = 0$ are used. Let π_k be the probability and Succ(k) the set of all successors of the node k, and introduce the following auxiliary variables:

- d_k^{up} and d_k^{down} denote upper bounds for the step length:

$$d_k^{up} = \begin{cases} x_k - x^{min}, & \text{if } b_k^{up} = 1 \\ \min\{l^{max} - l_k, \min_{\kappa \in Succ(k)} d_\kappa^{up}\}, & \text{if } b_k^{up} = 0 \end{cases}$$

$$d_k^{down} = \begin{cases} x^{max} - x_k, & \text{if } b_k^{down} = 1 \\ \min\{l_k, \min_{\kappa \in Succ(k)} d_\kappa^{down}\}, & \text{if } b_k^{down} = 0 \end{cases} .$$

- r_k^{up} and r_k^{down} denote the best average values for the subtrees:

$$r_k^{up} = \begin{cases} \lambda_k \pi_k, & \text{if } b_k^{up} = 1 \\ \sum_{\kappa \in Succ(k)} r_\kappa^{up}, & \text{if } b_k^{up} = 0 \end{cases}$$

$$r_k^{down} = \begin{cases} \lambda_k \pi_k, & \text{if } b_k^{down} = 1 \\ \sum_{\kappa \in Succ(k)} r_\kappa^{down}, & \text{if } b_k^{down} = 0 \end{cases} .$$

Now, the conditions on the existence of a descent direction, to which a subtree starting at node k is associated, read:

$$\text{Case of increasing the level:} \qquad \min\{x^{max} - x_k, d_k^{up}\}\{\lambda_k \pi_k + r_k^{up}\} \le 0\,, \qquad (19)$$

$$\text{Case of decreasing the level:} \quad \min\{x_k - x^{min}, d_k^{down}\}\{\lambda_k \pi_k + r_k^{down}\} \le 0\,. \qquad (20)$$

Having these conditions in mind, the algorithm can be outlined as follows:

Step 1: Input and initialization;
Step 2: Determine a feasible point;
Step 3: Compute d_k^{up}, d_k^{down}, r_k^{up}, and r_k^{down} at all nodes;
Step 4: Search for the node (root of the subtree) with the steepest descent; unless it can be found, then the current iterate is optimal $\rightarrow$ **STOP**;
Step 5: Update x_k and l_k at all nodes;
Step 6: Goto Step 3.

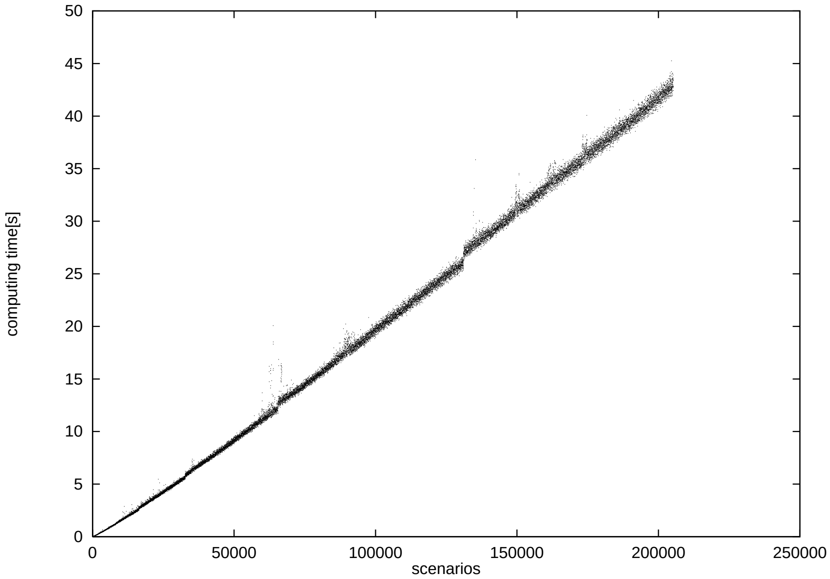

Figure 5: Computation times [s] of *EXCHA*

This descent algorithm *EXCHA* was implemented and tested for the case of $\eta_j < 1$. Each step of the algorithm requires only a few elementary computations, and in each step some variable attains an upper or lower bound. Hence, the algorithm is very efficient, as can also be seen in Figure 5, where the computing times in seconds of *EXCHA* on an HP-workstation are shown for a stochastic hydro storage problem with $T \le 18$ and binary trees branching at all time periods with numbers of scenarios ranging up to 200.000. Notice that, in case a sequence of such problems with slightly different dual variables has to be solved, the last iterate of the previous problem can be used as the next initial point.

3.2 Stochastic Dynamic Programming

The subproblem (21) - (22), that corresponds to the thermal unit i, is a mixed-integer multi-stage stochastic program. But, since the inner minimization with respect to the one-dimensional continuous variable $\mathbf{p}_i^t$ can be carried out explicitly by examining the kinks of the fuel costs FC_i, it reduces to a combinatorial multi-stage stochastic program:

$$D_i(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{u}_i} \{ \sum_{t=1}^{T} (\min_{\mathbf{p}_i^t} [FC_i(\mathbf{p}_i^t, \mathbf{u}_i^t) - (\boldsymbol{\lambda}^t - \boldsymbol{\mu}^t)\, \mathbf{p}_i^t] - \boldsymbol{\mu}^t \mathbf{u}_i^t p_i^{max} + SC_i^t(\mathbf{u}_i)) : \tag{21}$$

$$\mathbf{u}_i^t p_i^{min} \leq \mathbf{p}_i^t \leq \mathbf{u}_i^t p_i^{max},\ t = 1, \ldots, T,\ \ (\mathbf{u}_i, \mathbf{p}_i) \in \mathop{\times}_{t=1}^{T} L^\infty(\Omega,\, \mathcal{A}_t,\, I\!P; I\!R^2) \} \,. \tag{22}$$

The start-up costs $SC_i^t(\mathbf{u}_i)$ depend on the components $\mathbf{u}_i^\tau$ of $\mathbf{u}_i$, $\tau = t, t-1, \ldots, t - \tau_i^c$, where τ_i^c is the time the unit i needs to cool down. In order to apply the dynamic programming algorithm to stochastic programs, the state space is extended by including the recent history such that minimum up/down-times and start-up costs depend just on the current and the previous state. Figure 6 shows a part of the state transition graph of a thermal unit having a minimum up-time of 6 hours, a minimum down-time of 5 hours, and a cooling down time of 8 hours. It shows possible and feasible transitions on some fixed arc of the scenario tree, where the arrows refer to feasible transitions. From now on, the index i of the (fixed) thermal unit is omitted.

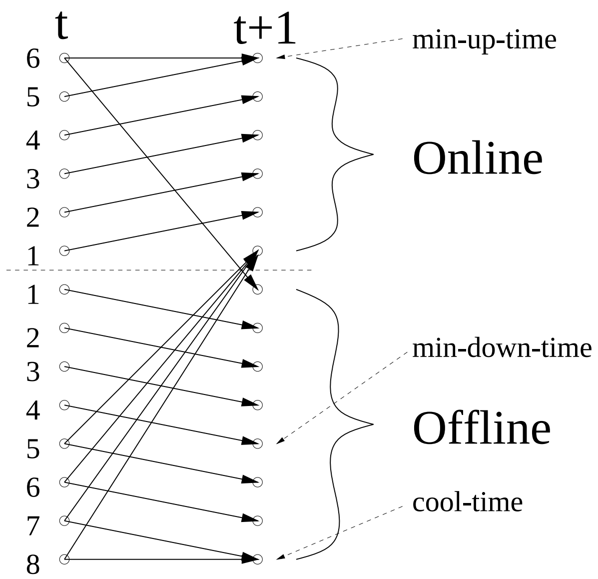

Figure 6: Transition graph for 2 time periods

Let $\alpha_t(s)$ denote the node weight at time t and state s and $\widehat{SC}(s, \tilde{s})$ the arc weight for the arc from state s to state $\tilde{s}$ in the state transition graph. The node weights

$\alpha_t(s)$ are equal to 0 for off-line states s and correspond to the linearly perturbed fuel costs for on-line states s, i.e.,

$$\alpha_t(s) = \min_p \{FC(p,1) - (\boldsymbol{\lambda}_t - \boldsymbol{\mu}_t)p \ : p^{min} \le p \le p^{max}\} \ . \tag{23}$$

The arc weights $\widehat{SC}(s,\tilde{s})$ describe start-up costs for the thermal unit. They are independent of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, and are non-zero only for arcs leading from off-line states to on-line states. The cost-to-go functions are given as

$$\gamma_t(s) = \alpha_t(s) + \mathbb{E}\left(\min_{\tilde{s}}\{\widehat{SC}(s,\tilde{s}) + \gamma_{t+1}(\tilde{s})\}|\mathcal{A}_t\right). \tag{24}$$

Having the formula for the costs-to-go, the dynamic programming algorithm can be applied. As each node of the scenario tree is considered only twice, the algorithm is reasonably fast. For one thermal unit, one load scenario, and one week with an hourly discretization the algorithm needs just 40 milliseconds on an HP-workstation.

3.3 Proximal Bundle Method

We consider the minimization of a convex function f on a nonempty closed convex set X in $\mathbb{R}^n$ and assume that the optimal set X_* is nonempty and we can compute $f(x)$ and a subgradient $g(x) \in \partial f(x)$ for each $x \in X$. The proximal bundle method [19, 21] generates a sequence (x^k) in X converging to some element of X_*, and trial points $y^k \in X$ starting with $y^1 = x^1$ for evaluating subgradients $g(y^k)$ of f and its polyhedral lower approximation

$$\tilde{f}_k(x) = \max_{j \in J^k}\left\{f(y^j) + < g(y^j), x - y^j >\right\} , \tag{25}$$

where J^k is a subset of $\{1,...,k\}$. In the iteration k the next trial point y^{k+1} is selected by

$$y^{k+1} \in \arg\min\{\tilde{f}_k(x) + \frac{1}{2}u_k\|x - x^k\|^2 : x \in X\} \tag{26}$$

where u_k is a proximity weight. A descent step to $x^{k+1} = y^{k+1}$ occurs if $f(y^{k+1}) \le f(x^k) + \alpha v_k$, where $\alpha \in (0,1)$ is fixed and $v_k = \tilde{f}_k(y^{k+1}) - f(x^k) \le 0$. If $v_k = 0$, then x^k is optimal. Otherwise, a null step $x^{k+1} = x^k$ improves the next polyhedral function $\tilde{f}_k$. Strategies for updating u_k and choosing J^{k+1} are discussed in [19, 21]. The method is implemented such that the cardinality of J^k is bounded (by some natural number NGRAD) and that it terminates if $-v_k$ is less than a given (relative) optimality tolerance.

This technique is applied to solve the dual stochastic problem (13) by putting $f = -D$ and $X = \mathbb{R}_+^{2N}$, where N is the number of nodes of the load scenario tree. Our computational experience with the proximal bundle code NOA 3.0 ([20]) for solving (13) is very encouraging (cf. Section 4). In our test runs, for instance, NOA 3.0 applied to solving (13) performed in 300 iterations as good as a standard subgradient method (with step lengths $\frac{1}{k}$) in 10.000 iterations.

3.4 Lagrange Heuristics

After having solved the dual stochastic program (13) we obtain a lower bound to the optimal costs of the power scheduling model in Section 2, and together with the optimal $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ we have scheduling decisions $(\mathbf{u}, \mathbf{p}, \mathbf{s}, \mathbf{w})$. In general, however, the scheduling decisions violate the load and reserve constraints (4) and (5).
In the following, we describe a technique for determining a feasible approximate solution for the optimal first-stage decision of the multi-stage stochastic power scheduling problem. Since this technique starts from the available dual information $(\boldsymbol{\lambda}, \boldsymbol{\mu})$, it is called *Lagrange heuristics* as in the deterministic setting.
In a first step, the mean value functions of the (discrete-time) stochastic processes $\mathbf{d}$ (load), $\mathbf{r}$ (reserve), $\mathbf{l}$ (storage volumes), $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are computed. Clearly, they coincide with their realizations (scenarios) during all time periods belonging to the first stage. Next, generation and pumping decisions, s_j and w_j, are determined from the constraints (2), where $\mathbf{l}_j$ is replaced by its expectation $\mathbb{E}[\mathbf{l}_j]$. Furthermore, on/off decisions u_i are computed by dynamic programming as solutions of the thermal subproblems (21), where the stochastic multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are replaced by their expectations $\lambda = \mathbb{E}[\boldsymbol{\lambda}]$ and $\mu = \mathbb{E}[\boldsymbol{\mu}]$, respectively. For one of the test runs explained in Section 4, Figure 7 shows the results after the first step of the heuristics: the mean *load* and *reserve* curves $\mathbb{E}[\mathbf{d}^t]$ and $\mathbb{E}[\mathbf{r}^t]$, the hydro genera-

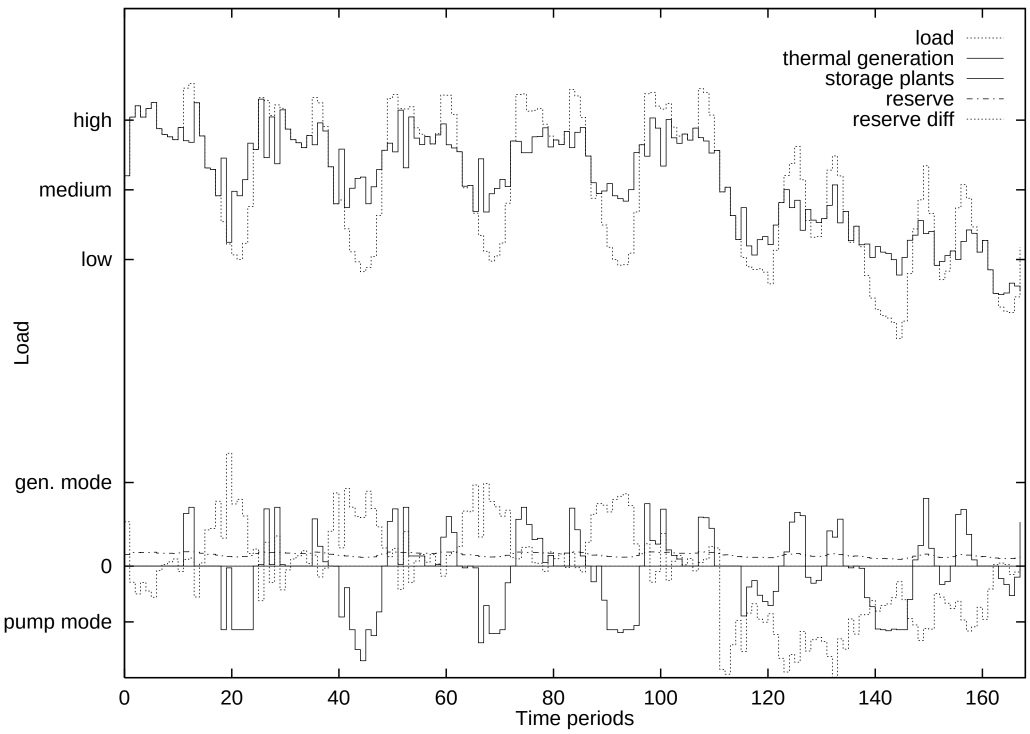

Figure 7: Schedules after averaging

tion and pumping curves $\sum_{j=1}^{J} s_j^t$ and $\sum_{j=1}^{J} w_j^t$, and the reduced mean load curve $\mathbb{E}[\mathbf{d}_t] - \sum_{j=1}^{J}(s_j^t - w_j^t)$ for $t = 1, ..., T$. Furthermore, it shows that the reserve constraint (28) is violated e.g. during $1 \leq t \leq 12$ and $110 \leq t \leq 168$.
In order to find scheduling decisions $(u,\, p,\, s,\, w)$ that are feasible for the reserve constraint (28), the schedules of the pumped hydro storage and the thermal plants are modified during the next two steps. The second step consists in applying a *water rescheduling* procedure, which is taken from [9]. Its idea is to reduce the value

$$\mathbb{E}[\mathbf{d}^t] + \mathbb{E}[\mathbf{r}^t] + \sum_{j=1}^{J}(w_j^t - s_j^t) \tag{27}$$

by modifying the schedule of the hydro plants if the (modified) reserve constraint

$$\sum_{i=1}^{I} u_i^t p_i^{max} \geq \mathbb{E}[\mathbf{d}^t] + \mathbb{E}[\mathbf{r}^t] + \sum_{j=1}^{J}(w_j^t - s_j^t) \tag{28}$$

is violated at time t and the value (27) is the largest in a certain set of neighbouring time intervals.
In the third step, the hydro schedules are kept fixed and binary variables u_i^t satisfying the reserve constraint (28) are determined by the algorithm described in [40]. Its main idea consists in determining the time t, where the constraint (28) is violated the most, and in computing the necessary increase of μ^t to switch on (by dynamic programming) just as many thermal units as needed to satisfy (28). This procedure is repeated until the reserve constraint (28) is satisfied in all time intervals. Since this technique does not distinguish between identical units which appear in real-life power generation systems quite often, the start-up costs of such units are slightly modified. In our computational experiments this modification led to improved computational results (cf. Section 4).
Altogether, the *Lagrange heuristics* takes the following form:

Step 1 Determine the mean values of $\mathbf{d}$, $\mathbf{l}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$,
Step 2 Use the hydro plants as in [9] to reduce
- the violation of the reserve constraints,
- the difference between maximal and minimal thermal load,

Step 3 Search for a reserve feasible schedule by the procedure in [40] after having modified start-up costs for identical units.

3.5 Economic Dispatch Algorithm

The Lagrange heuristics ends with a binary schedule u_i^t for the thermal units such that a feasible schedule $(u,\, p,\, s,\, w)$ exists for the original power optimization problem in Section 2 when replacing the stochastic load $\mathbf{d}$ and reserve $\mathbf{r}$ by their expected values. In a final step, a cost-optimal schedule $(p,\, s,\, w)$ is determined for fixed u by solving the corresponding primal problem, in which the start-up costs are fixed and, hence, negligible. The aim of this section is to develop an algorithmic approach for solving this economic dispatch problem. The approach also applies to multi-stage stochastic power scheduling models with fixed stochastic binary decisions $\mathbf{u}$. Since this may be of independent interest, we consider the model:

$$\min_{(\mathbf{p},\mathbf{s},\mathbf{w})} \left\{ \mathbb{E} \sum_{i=1}^{I} \sum_{t=1}^{T} FC_i(\mathbf{p}_i^t, \mathbf{u}_i^t) \; : \; (\mathbf{p},\, \mathbf{s},\, \mathbf{w}) \in \mathop{\times}_{t=1}^{T} L^\infty(\Omega,\, \mathcal{A}_t,\, I\!P; I\!R^{I+2J}) \right., \tag{29}$$

$$p_i^{min} \le \mathbf{p}_i^t \le p_i^{max},\; t = 1, \ldots, T,\; i = 1, \ldots, I, \;\; \text{if } \mathbf{u}_i^t = 1 \tag{30}$$

$$0 \le \mathbf{s}_j^t \le s_j^{max},\; 0 \le \mathbf{w}_j^t \le w_j^{max},\; 0 \le \mathbf{l}_j^t \le l_j^{max}, \tag{31}$$

$$\mathbf{l}_j^t = \mathbf{l}_j^{t-1} - \mathbf{s}_j^t + \eta_j \mathbf{w}_j^t,\; t = 1, \ldots, T,\; \mathbf{l}_j^0 = l_j^{in},\; \mathbf{l}_j^T = l_j^{end},\; j = 1, \ldots, J, \tag{32}$$

$$\sum_{i=1}^{I} \mathbf{p}_i^t + \sum_{j=1}^{J} (\mathbf{s}_j^t - \mathbf{w}_j^t) \ge \mathbf{d}^t \;,\; t = 1, \ldots, T, \tag{33}$$

$$\left. \sum_{i=1}^{I} (\mathbf{u}_i^t p_i^{max} - \mathbf{p}_i^t) \ge \mathbf{r}^t \;,\; t = 1, \ldots, T \; \right\}. \tag{34}$$

The stochastic program (29)-(34) has the same structure as (15)-(17) except for the appearance of thermal units. This motivates the idea to apply the same technique as in Section 3.1. Thermal units and storage plants are coupled by the constraints (33). Moving the sum $\sum_{j=1}^{J} (\mathbf{s}_j^t - \mathbf{w}_j^t)$ to the right-hand side in (33) and taking the right-hand side as a parameter, the optimization problem (29), (30), (33) decomposes into parametric programs for each time period t and scenario ω. Denoting the parameter by v the parametric programs are of the form:

$$P_{t,\omega}(v) : \; \min_{\mathbf{p^t}} \{ \sum_{i=1}^{I} FC_i(p_i^t, \mathbf{u}_i^t(\omega)) \; : \sum_{i=1}^{I} p_i^t \ge v \,,\; \mathbf{u}_i^t(\omega) p_i^{min} \le p_i^t \le \mathbf{u}_i^t(\omega) p_i^{max} \}. \tag{35}$$

Such programs were studied in [4, 5] for piecewise linear and quadratic fuel costs and, viewed as parametric programs, in [22] for the case of (piecewise) quadratic costs. The solution method for $P_{t,\omega}(v)$ starts with all units at their minimum output level. Then, a priority list is used to determine in which order the units have to increase their output level until the parameter $v \in [\sum_{i=1}^{I} \mathbf{u}_i^t(\omega) p_i^{min} \,,\, \sum_{i=1}^{I} \mathbf{u}_i^t(\omega) p_i^{max} - \mathbf{r}^t(\omega)]$ is met. The priority list can be built beforehand, and then used for all time periods. Denoting the optimal value function of $P_{t,\omega}(v)$ by $\phi_{t,\omega}(v)$, the objective function (29) reads

$$\min_{(\mathbf{s},\mathbf{w})} \mathbb{E} \sum_{t=1}^{T} \phi_{t,\cdot} (\mathbf{d}^t - \sum_{j=1}^{J} (\mathbf{s}_j^t - \mathbf{w}_j^t)). \tag{36}$$

The objective function (36) now has to be minimized with respect to the constraints (31) - (32). This reformulation of the model allows to study how the objective function varies when altering the operation of the hydro plants. Since the linearization of the model (36), (31)-(32) coincides with the problem (15)-(17), the search for descent directions of the objective function can be done in the same way as in Section 3.1. The place of the dual parameter $\boldsymbol{\lambda}$ is now taken by the derivative of $\phi_{t,\omega}(\,.\,)$ if it exists. This leads to

$$\sum_{j=1}^{J} \min_{(\mathbf{s}_j,\mathbf{w}_j)} \mathbb{E} \sum_{t=1}^{T} \lambda^t(\mathbf{w}_j^t - \mathbf{s}_j^t) \;, \tag{37}$$

where λ^t is defined by

$$\lambda^t = \frac{\mathrm{d}\phi_{t,\cdot}}{\mathrm{d}v}(\mathbf{d}^t - \sum_{j=1}^{J}(\mathbf{s}_j^t - \mathbf{w}_j^t)) \;. \tag{38}$$

The calculation of step-lengths has to take care of the structure of $\phi_{t,\omega}(\,.\,)$, i.e., steps should not go further than the next kink of $\phi_{t,\omega}(\,.\,)$. In non-differentiability points of $\phi_{t,\omega}(\,.\,)$ directional derivatives have to be used instead. Then, the determination of descent steps has to deal with several cases. Another possibility consists in smoothing the function $\phi_{t,\omega}(v)$ and in reducing the smoothing intervals as the number of iterations increases. This descent method was coded, tested and compared with

CPLEX-function	Pricing strategy primal/dual					
	-1	0	1	2	3	4
Simplex/primal	1232.4	1188.4	1918.1	2664.1	2440.7	1696.9
Simplex/dual		1086.1	946.2	1103.4	1466.5	1083.8
baropt	94.78					
hybbaropt/primal	114.7	114.3	114.3	486.5	114.4	114.3
hybbaropt/dual		115.0	114.6	693.0	1424.8	114.8
hybnetopt/primal	957.6	910.3	1298.0	2252.8	1960.9	1162.6
hybnetopt/dual		1393.8	1253.7	1412.0	1833.9	1392.3

Table 1: Computing times [s] for different CPLEX-functions and options

CPLEX 4.0. Test runs of our code *ECDISP* were performed for the VEAG system with 25 thermal units and 7 pumped hydro storage plants. Table 1 contains results for a test example with one load scenario and 192 time periods, which is equivalent to an LP with 14200 columns, 17856 rows, and 46256 non-zeros. The table shows computing times of *CPLEX 4.0* on a SPARCstation IPX (4/50) with 64 MB main memory and 40 MHz, which have to be compared with the *ECDISP* computing time of 50.95 seconds. Since the barrier method performs significantly better than the simplex method, and even better than the network simplex method, further comparisons were made with the barrier method only. Table 2 contains results for test problems with $T = 192$ and up to 22 scenarios. *CPLEX 4.0* ran out of memory for problems with a higher number of scenarios. The advantage of using *ECDISP* ranges from 1.8 up to 18.7, and in average *ECDISP* is 5-6 times faster. The Figures

scen's	nodes	columns	rows	non-zeros	ECDISP[s]	CPLEX[s]	adv'
3	336	24840	31248	80944	18.69	97.61	5.22
5	462	34148	42966	111294	29.48	162.47	5.51
7	588	43456	54684	141644	47.93	206.00	4.30
9	687	50766	63891	165487	43.09	305.43	7.09
11	792	58520	73656	190776	67.17	500.30	7.45
13	930	68716	86490	224018	86.73	461.54	5.32
15	1035	76470	96255	249307	98.04	569.18	5.81
17	1036	76528	96348	249532	117.42	620.65	5.29
19	1120	82728	104160	269760	91.63	1720.33	18.7
21	1232	91000	114576	296736	131.94	243.27	1.84
22	1260	93064	117180	303476	128.18	794.93	6.20

Table 2: Comparison of *ECDISP* with *CPLEX*

8 and 9 show that the number of steps and the computing times of *ECDISP* grow almost linearly with respect to the number of scenarios.

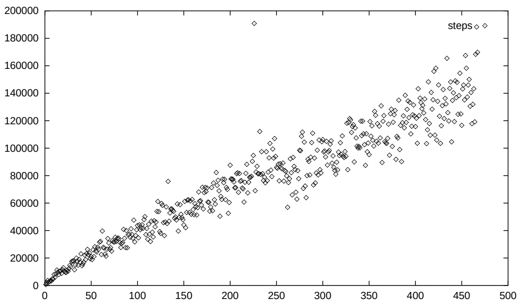

Figure: 8: Number of steps versus number of scenarios

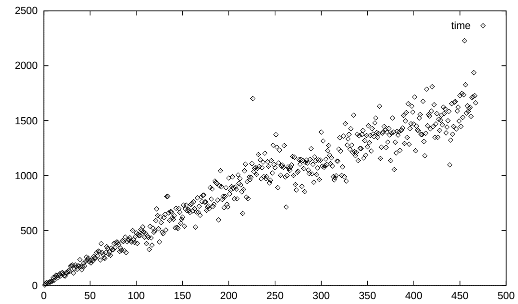

Figure: 9: Computing times [s] versus number of scenarios

4 Numerical Results

The stochastic Lagrangian relaxation algorithm described in the previous section has the following structure:

Step 1: Input and initialization
Step 2: Solve the dual problem by the proximal bundle method (Sect. 3.3)
  Step 2.1: Solve the thermal unit subproblems by stochastic dynamic programming (Sect. 3.2)
  Step 2.2: Solve the storage subproblems by a descent algorithm (Sect. 3.1)
Step 3: Apply the Lagrange heuristics (Sect. 3.4)

Step 4: Solve a final economic dispatch problem (Sect. 3.5)

The algorithm was implemented and coded in *C++* except for the proximal bundle method where the FORTRAN-package *NOA 3.0* ([20]) was used as a callable library. For testing the implementation, a test bunch of load scenario trees was generated as follows. The tree structure was built by generating scenarios successively from randomly chosen base scenarios and branching points. Each load scenario was generated by adding a (discretized) Brownian motion to a reference load process obtained from real-life data of the utility VEAG. In a final step, randomly chosen probabilities were assigned to each scenario. Test runs were performed for the hydro-thermal power generation system of VEAG comprising 25 thermal units and 7 pumped storage plants on an HP 9000 (780/J280) Compute-Server with 180 MHz frequency and 768 MByte main memory under HP-UX 10.20.

Nr.	modified cost functions				original cost functions			
	opt.tol: 10^{-3}		opt.tol: 10^{-4}		opt.tol: 10^{-3}		opt.tol: 10^{-4}	
	gap/%	time/s	gap/%	time/s	gap/%	time/s	gap/%	time/s
1	0.18	34.08	0.10	89.84	0.67	31.32	0.56	86.24
2	0.25	47.82	0.12	109.92	0.60	42.67	0.61	100.07
3	0.43	44.81	0.26	111.75	0.44	35.10	0.25	102.61
4	0.34	53.86	0.14	119.84	0.94	47.40	0.96	115.55
5	0.20	78.42	0.11	157.31	0.98	73.76	0.93	151.64
6	0.27	41.24	0.19	85.12	1.23	34.09	1.13	80.42
7	0.39	39.52	0.11	88.35	0.66	37.42	0.54	79.88

Table 3: Influence of modified costs and of bundle method optimality tolerances on the gap and computing times.

For test runs with 10 scenarios Table 3 shows the influence of modifying the start-up costs and of changing the optimality tolerance of the bundle method on the gap and computing times. It shows that a (slight) modification of start-up costs of former identical units leads to smaller gaps. Here, the gap refers to the (relative) difference of the costs for the primal feasible (approximate) solution and the optimal value of the dual stochastic problem. Moreover, improving optimality tolerances leads to smaller gaps paid by increased computing times.
Figure 10 provides the final output of the algorithm and contains, in particular, the approximately optimal first-stage solution for the total thermal and hydro generation (for the time periods $t = 1, \ldots, 24$). Table 4 shows how the computing time grows with increasing numbers of scenarios. Since the complexity of the model is higher compared to the stochastic programs in Sections 3.1 and 3.4, the variance of the computing time is larger than the variances expressed in the Figures 5 and 9. The reason is that the iteration numbers in the bundle method, in the method for

Scenarios	Nodes	time[s]/gap[%]	Nodes	time[s]/gap[%]
10	781	31.2 / 0.274	1043	52.93 / 0.138
10	1232	50.36 / 0.201	975	54.21 / 0.723
20	1982	89.13 / 0.149	1627	93.62 / 0.101
20	1651	67.94 / 0.367	1805	84.73 / 0.066
30	2643	139.71 / 0.528	2643	138.61 / 0.528
30	2548	147.51 / 0.849	2515	162.14 / 0.175
50	4530	475.29 / 0.175	4060	274.43 / 0.096
50	4041	312.86 / 0.099	4457	288.03 / 0.430
80	7190	804.04 / -0.018	7120	698.19 / 0.379
80	6548	537.28 / 0.137	6501	597.04 / 0.114
100	9230	1183.25 / 0.108	9224	1072.18 / 0.131
100	7727	929.68 / 0.087	8867	1234.12 / 0.304

Table 4: Computing times and gaps (NOA 3.0: $Opt.tol = 10^{-3},\ NGRAD = 50$)

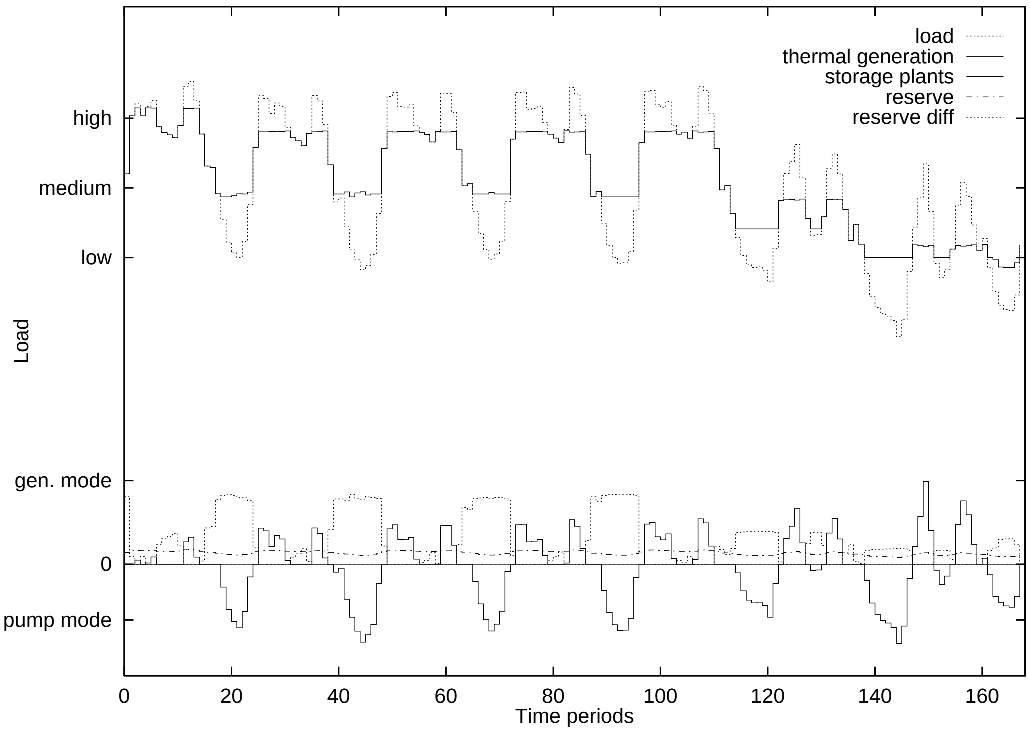

Figure 10: Approximate solution

searching a reserve feasible solution and in the economic dispatch solver depend on the input data in a very involved way. Another observation is that the gap seems to be independent of the number of scenarios.

5 Conclusions

We have elaborated a mixed-integer multi-stage stochastic programming model for power scheduling in a hydro-thermal generation system under uncertainty on the electrical load. Due to the huge size of the model, an application of state-of-the-art mixed-integer LP solvers is prevented. Therefore, we have developed a novel approach based on stochastic Lagrangian relaxation of coupling constraints. It consists of proximal bundle iterations for solving a stochastic dual followed by a Lagrange heuristics to determine a nearly optimal primal first-stage solution. The stochastic dual decomposes into stochastic thermal and hydro subproblems, which are solved by specific fast algorithms. Our computational experience indicates that the stochastic Lagrangian relaxation algorithm is able to produce good approximate first-stage solutions for medium-size realistic power systems and 20 (100) load scenarios within less than 2 (20) minutes on a modern HP-workstation. It also indicates that the algorithm bears potential for solving more complex real-life power scheduling models under uncertainty in reasonable time.

Acknowledgement

The authors are grateful to K.C. Kiwiel (Polish Academy of Sciences, Warsaw) for the permission to use the NOA 3.0 package, to Darinka Dentcheva (Humboldt-University Berlin) for useful discussions on the Lagrange heuristics described in Section 3.4 and to G. Schwarzbach and J. Thomas (VEAG Vereinigte Energiewerke AG, Berlin) for providing us with the data used in our test runs. This research was supported by the Schwerpunktprogramm *Echtzeit-Optimierung großer Systeme* of the Deutsche Forschungsgemeinschaft and the Graduiertenkolleg *Geometrie und Nichtlineare Analysis* at the Humboldt-University Berlin.

References

[1] 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.

[2] J.R. Birge: Stochastic programming computations and applications, *INFORMS Journal on Computing* 9 (1997), 111-133.

[3] J.R. Birge and F. Louveaux: *Introduction to Stochastic Programming, Springer*, New York 1997.

[4] P.P.J. van den Bosch: Optimal static dispatch with linear, quadratic and nonlinear functions of the fuel costs, *IEEE Transactions on Power Apparatus and Systems* PAS-104 (1985), 3402-3408.

[5] P.P.J. van den Bosch and F.A. Lootsma: Scheduling of power generation via large-scale nonlinear optimization, *Journal of Optimization and Applications* 55(1987), 1684-1690.

[6] C.C. Carøe and R. Schultz: Dual decomposition in stochastic integer programming, Preprint SC 96-46, Konrad-Zuse-Zentrum für Informationstechnik Berlin, 1996 and to appear in *Operations Research Letters.*

[7] C.C. Carøe and R. Schultz: A two-stage stochastic program for unit commitment under uncertainty in a hydro-thermal power system, DFG-Schwerpunktprogramm *Echtzeit-Optimierung großer Systeme*, Preprint 98-13, 1998.

[8] P. Carpentier, G. Cohen, J.-C. Culioli and A. Renaud: Stochastic optimization of unit commitment: a new decomposition framework, *IEEE Transactions on Power Systems* 11(1996), 1067-1073.

[9] D. Dentcheva, R. Gollmer, A. Möller, W. Römisch and R. Schultz: Solving the unit commitment problem in power generation by primal and dual methods, in: *Progress in Industrial Mathematics at ECMI 96* (M. Brøns, M.P. Bendsøe and M.P. Sørensen Eds.), Teubner, Stuttgart 1997, 332-339.

[10] D. Dentcheva and W. Römisch: Optimal power generation under uncertainty via stochastic programming, in: *Stochastic Programming Methods and Technical Applications* (K. Marti and P. Kall Eds.), Lecture Notes in Economics and Mathematical Systems Vol. 458, Springer-Verlag, Berlin 1998, 22-56.

[11] S. Feltenmark and K.C. Kiwiel: Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems, Report TRITA/MAT-98-OS1, Department of Mathematics, Royal Institute of Technology, Stockholm, January 1998, revised November 1998 and submitted to *SIAM Journal on Optimization.*

[12] S.-E. Fleten, S.W. Wallace and W.T. Ziemba: Portfolio management in a deregulated hydropower based electricity market, in: *Hydropower' 97* (E. Broch, D.K. Lysne, N. Flatabø and E. Helland-Hansen Eds.), Balkema, Rotterdam 1997.

[13] A. Gjelsvik, T.A. Røtting and J. Røynstrand: Long-term scheduling of hydro-thermal power systems, in: *Hydropower '92* (E. Broch and D.K. Lysne Eds.), A.A. Balkema, Rotterdam 1992, 539-546.

[14] R. Gollmer, A. Möller, W. Römisch, R. Schultz, G. Schwarzbach and J. Thomas: Optimale Blockauswahl bei der Kraftwerkseinsatzplanung der VEAG, in: *Optimierung in der Energieversorgung II*, VDI-Berichte 1352, Düsseldorf 1997, 71-85.

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

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

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

[18] 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.

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

[20] 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.

[21] K.C. Kiwiel: Approximations in proximal bundle methods and decomposition of convex programs, *Journal of Optimization Theory and Applications* 84(1995), 529-548.

[22] P. Kleinmann and R. Schultz: A simple procedure for optimal load dispatch using parametric programming, *ZOR - Methods and Models of Operations Research* 34(1990), 219-229.

[23] C. Lemarechal: Lagrangian decomposition and nonsmooth optimization, in: *Nonsmooth Optimization Methods and Applications* (F. Gianessi Ed.), Gordon and Breach 1992, 201-216.

[24] C. Lemarechal and A. Renaud: Dual equivalent convex and nonconvex problems, Research Report, INRIA, Rocquencourt 1996 and submitted to *Mathematical Programming.*

[25] A. Løkketangen and D.L. Woodruff: Progressive hedging and tabu search applied to mixed integer (0,1) multi-stage stochastic programming, *Journal of Heuristics* 2(1996), 111-128.

[26] V.I. Norkin, G. Pflug and A. Ruszczyński: A branch and bound method for stochastic global optimization, *Mathematical Programming* 83(1998), 425-450.

[27] 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.

[28] M.P. Nowak and W. Römisch: Optimal power dispatch via multistage stochastic programming, in: *Progress in Industrial Mathematics at ECMI 96* (M. Brøns, M.P. Bendsøe and M.P. Sørensen Eds.), Teubner, Stuttgart 1997, 324-331.

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

[30] A. Renaud: Optimizing short-term operation of power generation: a new stochastic model, Lecture presented at the VIII International Conference on Stochastic Programming, Vancouver (Canada), August 1998.

[31] 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.

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

[33] 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.

[34] A. Ruszczyński: Decomposition methods in stochastic programming, *Mathematical Programming, Series B*, 79(1997), 333-353.

[35] G.B. Sheble and G.N. Fahd: Unit commitment literature synopsis, *IEEE Transactions on Power Systems* 9(1994), 128-135.

[36] 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.

[37] S. Takriti, B. Krasenbrink and L.S.-Y. Wu: Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem, IBM Research Report RC 21066, Yorktown Heights, New York 1997.

[38] R. J.-B. Wets: Stochastic Programming, in: it 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.

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

[40] 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.