Nonconvex Medium-Term Hydropower Scheduling by Stochastic Dual Dynamic Integer Programming

Martin N. Hjelmeland, *Student Member, IEEE,* Jikai Zou, Arild Helseth, *Member, IEEE* and Shabbir Ahmed, *Senior Member, IEEE*

***Abstract*—Hydropower producers rely on stochastic optimization when scheduling their resources over long periods of time. Due to its computational complexity, the optimization problem is normally cast as a stochastic linear program. In a future power market with more volatile power prices, it becomes increasingly important to capture parts of the hydropower operational characteristics that are not easily linearized, e.g. unit commitment and nonconvex generation curves.**

Stochastic dual dynamic programming (SDDP) is a state-of-the-art algorithm for long- and medium-term hydropower scheduling with a linear problem formulation. A recently proposed extension of the SDDP method known as stochastic dual dynamic integer programming (SDDiP) has proven convergence also in the nonconvex case. We apply the SDDiP algorithm to the medium-term hydropower scheduling (MTHS) problem and elaborate on how to incorporate stagewise dependent stochastic variables on the right-hand sides and the objective of the optimization problem. Finally, we demonstrate the capability of the SDDiP algorithm on a case study for a Norwegian hydropower producer.

The case study demonstrates that it is possible but time-consuming to solve the MTHS problem to optimality. However, the case study shows that a new type of cut, known as strengthened Benders cut, significantly contributes to closing the optimality gap compared to classical Benders cuts.

***Index Terms*—Stochastic processes, Dynamic programming, Hydroelectric power generation, Integer programming.**

I. Introduction

The hydropower scheduling problem is difficult given its stochastic and multistage nature, and a variety of different solution techniques have been applied to it, see e.g. [1], [2]. In a liberalized power market, one typically decomposes the overall problem into three hierarchies (long-, medium-, and short-term) according to the system boundary, the level of required details in the hydro system and the representation of different stochastic variables [3], [4], [5]. In this work we consider the medium-term hydropower scheduling (MTHS) problem, aiming at maximizing a single producer's profit. The MTHS problem links the long-term fundamental market model problem with the short-term (ST) operational scheduling problem, as discussed in [3]. The MTHS problem can be formulated as a multistage stochastic programming (MSSP) problem, with many decision stages and uncertainty of inflow and market prices.

Due to the complexity of the MTHS problem, it is typically approximated as a multistage stochastic linear programming (MSSLP) problem. To further exploit the flexibility of hydropower by contributing in ancillary service markets, it becomes increasingly important to capture parts of the operational characteristics that are inherently nonconvex. Nonconvexities arise from binary variables used to model minimum generation limits and unit commitments, and are vital for capturing the generation units' capability to provide reserve capacity. There are also other types of nonconvexities that occur in the MTHS problem, such as the generation-discharge function that is dependent on the water head, discharge from multiple reservoirs to one power station and other topological constraints in the hydropower system. These nonconvexities should not only be represented in the ST operative models but also in the MTHS models that provide the *expected* opportunity cost of water to them.

A. MSSP Problems

The MTHS problem can be formulated as a multistage stochastic program, of the following general extensive form

$$\max_{(x_n,y_n)} \left\{ \sum_{n\in\mathcal{T}} p_n f_n(x_n,y_n) : (x_{a(n)},x_n,y_n)\in X_n, \forall n\in\mathcal{T} \right\}. \tag{1}$$

In the above formulation, $\mathcal{T}$ is a scenario tree given by a set of nodes, n, that describes the underlying stochastic process $\{\tilde{\xi}_t : \forall t = 1,\dots,T\}$. Each node is assigned a probability p_n, and has a unique ancestor node $a(n)$ and set of children nodes $\mathcal{C}(n)$. For each node we define x_n and y_n as vectors comprising the *state* and *stage* variables, respectively. The state variables are those used to carry information from one period to the next. The initial state of the system is x_0, f_n is the objective function and X_n is the set of constraints. Note that the constraint set comprises time-linking constraints, connecting the state variables in $x_{a(n)}$ and x_n.

The size of the MSSP in (1) grows dramatically with the number of decision stages and number of children nodes considered in the scenario tree. Thus, MSSP problems are often solved by decomposition [6]. Stagewise decomposition [7] has become a popular technique for efficiently solving

M. N. Hjelmeland is with the Department of Electric Power Engineering, NTNU Norwegian University of Science and Technology, Trondheim, Norway, e-mail: martin.hjelmeland@ntnu.no.

A. Helseth is with SINTEF Energy Research, Trondheim, Norway.

J. Zou was with the H. Milton Stewart School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA. He is now a Research Scientist at Amazon Web Services.

S. Ahmed is with the H. Milton Stewart School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

Manuscript received March 13, 2017; revised August 12, November 28, 2017 and January 17, 2018; accepted February 1, 2018. This work was supported by the Research Council of Norway under Grant 228731/E20.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2018.2805164, IEEE Transactions on Sustainable Energy

the hydropower scheduling problems [8], [9]. Such a decomposition scheme involves dividing (1) into subproblems for each decision stage. Each subproblem comprises a part of the objective function corresponding to that stage and an expected future profit (EFP)[1] function.

B. Stochastic Dual Dynamic Programming

The stochastic dual dynamic programming algorithm (SDDP) presented in [10] is a stagewise decomposition method for solving the long- and medium-term hydropower scheduling problem. For a scenario tree corresponding to a stagewise independent stochastic process, the SDDP method proceeds by sampling in the scenario tree and sharing Benders cuts between nodes belonging to the same stage. Proof of almost sure convergence for the SDDP algorithm has been given by [11], and for risk averse problems in [12], [13]. The SDDP algorithm has been frequently addressed in the recent literature. Some of the work involves time-inconsistent policies [14], co-optimization of hydro- and wind power [15], joint treatment of energy and ancillary services [16], [17], uncertainty in price [18] and uncertainty in inflow [19].

In spite of considerable research effort, see e.g. [20], [21], [22], [23], [24], the SDDP method does not easily facilitate nonconvexities. As mentioned previously, nonconvexities arise, for example, when representing the detailed relationship between power output and water discharge [25], and the exact unit commitment of generators [26]. The core issue is how the nonconvex EFP function can be modeled. In [21] McCormick envelopes were applied to regions of the nonconvex function between power, discharge and water head. In [22] Lagrangian relaxation was used to convexify the EFP function, whereas [24] applied locally valid cuts to represent the nonconvex EFP function. All the above methods produce solutions to different forms of relaxations rather than the original problem.

Recently [27] proposed a method referred to as stochastic dual dynamic integer programming (SDDiP) allowing integer variables within the SDDP method, i.e. a solution to the multistage stochastic integer programming (MSSIP) problem. By approximating all state variables with binary variables, the authors were able to prove finite convergence as long as the cuts satisfy some sufficient conditions.

Another new and promising method that can be used to solve the nonconvex MTHS problem is reported in [28]. As SDDiP it is an extension of SDDP for handling nonconvexities. The significant difference is that instead of using cutting planes to describe the EFP function, step functions are used and that a monotonic increasing EFP function is required.

C. Contributions

In this work we apply the SDDiP method presented in [27] to the MTHS problem with a nonconvex function of power and discharge, incorporate stagewise dependencies in stochastic variables and correlations between them (inflow and energy price). We particularly extend the modeling from [27] with stagewise dependency in energy price, since this dependency enters the objective function and thus introduces a nonconvex term that is challenging to handle.

[1] or expected future cost when considering cost minimization problems.

The SDDiP method is tested and verified on a hydropower system in Norway. We evaluate the performance of the SDDiP algorithm on the MTHS problem using different types of cuts and provide recommendations for which types of cuts are most efficient in solving the MTHS problem.

The paper is organized as follows. In the next section, we will describe the basic modeling of the MTHS problem followed by the fundamentals of the SDDiP method. The case study is presented in Section VI and computational results in Section VII followed by concluding remarks in Section VIII.

II. THE MTHS PROBLEM

In a liberalized electricity market the hydropower producer's primary objective is to find operational strategies for each decision stage and maximize the total profit. For the Nordic electricity market, it is assumed that there is a sufficient number of players such that a price-taker assumption is reasonable.

The state variables, x_t, in the MTHS problem considered here are reservoir levels, generator status, inflow and energy price. The stage variables are represented by y_t and include the operational decisions at that stage. First, let us assume that the set of state variables consists of both continuous and binary variables. For the MTHS problem, given by (2), we want to find an operating strategy that maximizes profit in (2a), comprising the revenues from selling energy and capacity reserves, start-up cost and penalty functions ensuring relatively complete recourse. The expectation in (2a) is taken over the stochastic parameter $\tilde{\xi}_t$, representing energy price and inflow.

$$\max_{(x_1,y_1),\ldots,(x_T,y_T)} \mathbb{E}\left\{\sum_{t=1}^{T} f_t(x_t, y_t)\right\} \tag{2a}$$

$$\text{s.t.} \quad Wx_t + Hx_{t-1} + Gy_t = h(\tilde{\xi}_t) \tag{2b}$$

$$By_t = 0 \tag{2c}$$

$$Cy_t - Dx_t \geq 0 \tag{2d}$$

$$Cy_t + Dx_t \leq Cy^{\max} \tag{2e}$$

$$x_t, y_t \in Y_t \tag{2f}$$

$$x_t \in \mathbb{R}^{k_1} \cdot \mathbb{Z}^{k_2}, y_t \in \mathbb{R}^{l_1} \cdot \mathbb{Z}^{l_2} \tag{2g}$$

$$\forall t \in \{1, 2, \ldots, T\}, \tag{2h}$$

where the initial state vector x_0 is given and W, H, G, B, C and D are matrices of suitable dimensions. The right-hand-side parameter vector, $h(\tilde{\xi}_t)$, is dependent on the random data vector $\tilde{\xi}_t$ whose distribution is known, and where ξ_t are the realizations. Due to strong autocorrelation, both inflow and the energy price should be considered as an affine function of state variables. Thus the profit function for stage t is

$$f_t(x_t, y_t) = c_t(x_t)y_t^G + g_t(y_t), \tag{3}$$

comprising the energy price, $c_t(x_t)$, multiplied with generation, y_t^G, and the remaining linear relationships in $g_t(y_t)$. Treatment of the bilinear expression in (3) will be addressed in Section III-B.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2018.2805164, IEEE Transactions on Sustainable Energy

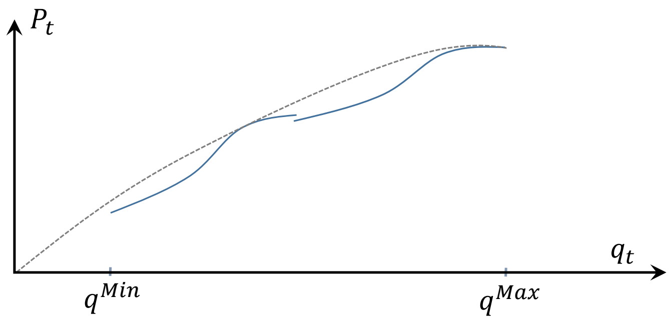

Fig. 1. Illustration of the nonconvex function of power, P_t, and discharge, q_t in a power station with two units in blue. It is conditioned that the first unit is started before the second. The dotted line illustrates the convex relaxation of the function.

The time-linking constraint, (2b), contains all reservoir balances, generator state time-couplings and the Vector Autoregressive model with lag-1 (VAR-1) constraints, outlined in Section III-A. The energy balance is given by (2c), where we assume that all power is sold to the market. Generation and reserve capacity limits are included in (2d) and (2e). We assume that the hydropower producer can offer spinning reserve capacity, meaning that the generator has to be operating to provide this service. The reserve capacity is also assumed symmetric, i.e. an equal amount of both upwards and downwards reserve capacity is sold. For conciseness, we also include constraints that limit spillage, bypass and generation from different reservoirs in (2f). In practice, spillage may only occur when the water head is greater than the bottom of the spillway. If, however, the optimization model finds that the water has better utilization in a lower reservoir, it will spill or bypass water to that reservoir, even-though this might not be feasible in practice. By including binary variables and additional constraints in (2f), we can prohibit this behavior of the model. Similar to when a power station can operate from two different reservoirs with uneven head, we can restrict the model to only operate from one reservoir at the time.

As an extension to the MTHS problem we include a nonconvex relationship between power and discharge for each hydropower station, referred to as the generation function. Much work has been done recently, see [21], [22], to include a nonconvex generation function, also with water head dependencies. Considering the modest water head variations of most Norwegian hydropower stations and to limit the complexity of the modelling, with added computational time, we omit this in our approach. Our experience has shown that it is important to include the nonconvex generation function when selling both energy and reserve capacity in order not to overestimate the sales of capacity reserves [15]. This is especially imperative for periods with low energy price and high capacity price. Then the model wants to operate the units at low power outputs in order to participate in the capacity reserve market. An illustration of a generation function is given in Figure 1.

For presentation simplicity, the reserve capacity price is assumed to be deterministic, but as the work done in [16] shows, it could be extended to a stochastic variable.

III. The SDDiP Formulation

In this section, we reformulate the problem (2) to a form suitable for SDDiP. First, we describe how the stochastic processes are handled.

A. Stochastic Processes

We consider two types of uncertainties in the problem; inflow to the hydro reservoirs and energy price. For the sake of simplicity both processes are assumed to have a normal distribution. Since both stochastic processes exhibit a seasonal pattern and are non-stationary processes, these are normalized by subtracting the seasonal mean and dividing by the seasonal standard deviation. Moreover, we shift the mean of the normalized value to ensure non-negative values. The normalized and shifted series are then assumed to be stationary and fitted to a Vector Autoregressive model with lag-1 (VAR-1), which enables us to account for seasonal effects and capture correlations between the time series within the SDDiP framework. In order to get a manageable number of stochastic outcomes of $\tilde{\xi}_t$, we applied the "Fast forward selection" scenario reduction algorithm in [29], resulting in a confined amount of scenarios with a corresponding probability. Another known method to handle the uncertainty of energy price in SDDP is by an outer Markov chain as proposed in [30]. The drawback is then that there are two independent stochastic processes in the model.

The VAR-1 model includes a state variable, δ_t. Since it is a state variable, binary expansion has to be applied to it. After the scenario reduction and sampling of the stochastic parameters are completed, a bound can be computed so that a finite support is guaranteed. Subsequently, we need to shift the mean of the process by U such that δ_t is non-negative. The bound is then given by (4b) and δ_t is the normalized and shifted values. We can derive the following VAR-1 model, which is included in (2b)

$$\delta_t = \Phi\delta_{t-1} + \tilde{\xi}_t \tag{4a}$$

$$\delta_t \in [0, 2U]^2, \tag{4b}$$

where Φ is the correlation matrix. Note that $\mathbb{E}[\delta_t] = U$ and $\mathbb{E}[\tilde{\xi}_t] = U$. The stochastic parameters for inflow and energy price are respectively given by

$$\begin{bmatrix} I_t(\delta_t) \\ c_t(\delta_t) \end{bmatrix} = \sigma_t\big(\delta_t - U\big) + \mu_t. \tag{5}$$

where the inflow, $I_t(\delta_t)$, is the right-hand side of the reservoir constraints and the energy price, $c_t(\delta_t)$, is expressed in the objective function (3). Note that δ_t in the stochastic model is identical with x_t in (3) as the stochastic model was not yet introduced. In order to solve the bilinear term in (3) an approximation is applied.

B. Approximation of the Energy Price

To circumvent the computational complexity introduced by the bilinear term in (3), we apply an approximation by linear relaxation of the objective function term, $c_t(\delta_t)y_t^G$. The relaxation is applied by using the binary expansion method to the energy price state variable, δ_t. The method is based on the fact that if λ_t is an integer variable, $\lambda_t \in \{0,\ldots,L\}$, it can be represented by κ binary variables, $\kappa = \lfloor \log_2(\mathrm{L}) \rfloor + 1$, such that $\lambda_t = \Sigma_{j=1}^{\kappa} 2^{j-1}\lambda_{tj}$. Similarly, for the continuous case, $\lambda_t \in [0, L]$, where λ_t is given with ε accuracy; $\kappa = \lfloor \log_2(\mathrm{L}/\varepsilon) \rfloor + 1$, hence $\lambda_t = \Sigma_{j=1}^{\kappa} 2^{j-1}\varepsilon\lambda_{tj}$. See e.g. [31] for further details about the binary expansion. In the following we use the notation $\lambda_t = \sum_{j=1}^{\kappa} A_{tj}\lambda_{tj}$. Thus, we redefine $\delta_t = \sum_{j=1}^{r} A_{tj}x_{tj}$, with $r = \lfloor \log_2(2\mathrm{U}/\varepsilon) \rfloor + 1$ and substitue δ_t in (5). Moreover, we then replace the bilinear terms $x_{tj}y_t^G$ with auxiliary variables w_{tj}, and constrain these by (6b)-(6e), for all $j \in \{1,\ldots,r\}$. The objective function becomes,

$$
\begin{aligned}
f_t(x_t,y_t) =& \sigma_t \sum_{j=1}^{r} A_{tj}w_{tj} + (\mu_t - \sigma_t U)y_t^G + g(y_t) && \text{(6a)}\\
\text{s.t.}\quad & w_{tj} - y_t^G \le 0 && \text{(6b)}\\
& w_{tj} - M_{tj}x_{tj} \le 0 && \text{(6c)}\\
& w_{tj} - y_t^G + M_{tj}(1 - x_{tj}) \ge 0 && \text{(6d)}\\
& w_{tj} \in \mathbb{R}^r,\ x_{tj} \in \{0,1\}^r && \text{(6e)}\\
& \forall j \in \{1,\ldots,r\}. && \text{(6f)}
\end{aligned}
$$

We fix M_{tj} to $y^{G,\max}$ for all j, in order to get an accurate relaxation.

Note that after the binary expansion is applied to the state variable, δ_t, (4a) becomes a constraint with only binary variables and a stochastic parameter. Thus, in order to ensure feasibility of the model with all possible outcomes of $\tilde{\xi}_t$, the constraint is altered to a less than or equal constraint. Implications of this are discussed in Section VII.

C. Dynamic Programming (DP) Formulation

From the MTHS problem given by (2) and the energy price approximation in (6), we have the following DP equation

$$
\begin{aligned}
\mathcal{Q}_t(x_{t-1},\xi_{t-1}) :=& \max_{x_t,y_t} f_t(x_t,y_t) + \phi_t(x_t) && \text{(7a)}\\
\text{s.t.}\quad & (x_t,y_t) \in X_t(x_{t-1},\tilde{\xi}_t) && \text{(7b)}\\
& x_t \in \mathbb{R}^{k_1}\cdot\mathbb{Z}^{k_2}, y_t \in \mathbb{R}^{l_1}\cdot\mathbb{Z}^{l_2}. && \text{(7c)}
\end{aligned}
$$

Problem (7) consists of the present objective function $f(x_t,y_t)$, the *true* EFP function $\phi_t(x_t)$ and constraint set X_t, with parameters x_{t-1} and $\tilde{\xi}_t$, described by (2b)-(2f) and (6b)-(6f).

As the MTHS problem contains integer stage variables the EFP function is nonconvex with respect to the state variables. Existing nested decomposition methods rely on convex relaxation to approximate the EFP function, and convergence can therefore not be guaranteed.

The traditional approach to solving the nonconvex problem in (7) has been to relax the problem formulation to an LP. We are then able to define a convex relaxation of the EFP function given by,

$$
\begin{aligned}
Q_t^i(x_{t-1},\xi_{t-1}) :=& \max_{x_t,y_t} f_t(x_t,y_t) + \varphi_t^i(x_t) && \text{(8a)}\\
\text{s.t.}\quad & (x_t,y_t) \in X_t^{'}(x_{t-1},\tilde{\xi}_t) && \text{(8b)}\\
& x_t \in \mathbb{R}^{k_3}, y_t \in \mathbb{R}^{l_3}. && \text{(8c)}
\end{aligned}
$$

The EFP function is then $\varphi_t^i(x_t)$, for a given iteration i of the SDDP method, and constraint set $X_t^{'}$ that describes a LP relaxation of X_t. We see that this DP formulation is similar to the SDDP method, where the EFP function is described by hyperplanes, defined as

$$
\begin{aligned}
\varphi_t^i(x_t) := \max\{&\theta_t \le V_t && \text{(9a)}\\
&\theta_t \le v_{t+1}^l + \pi_{t+1}^l x_t,\ \forall l \in \mathcal{H}(i), && \text{(9b)}\\
&x_t \in \mathbb{R}^{k_3}\}. && \text{(9c)}
\end{aligned}
$$

Where the set $\mathcal{H}(i)$ contains the cuts that are used to approximate the EFP function up to iteration i. The cuts are represented with a right hand side parameter v_{t+1}^l, the coefficient π_{t+1}^l and θ_t is a scalar variable representing the value of the EFP function. The different type of cuts that we have used in our implementation will be outlined in Section IV-A.

To solve the nonconvex MTHS problem we apply the SDDiP method proposed in [27]. The key concept of the method is that any function of binary variables can be represented as a convex polyhedral function. This can, therefore, be achieved by applying the above-mentioned binary expansion to all integer and continuous state variables of the original problem.

After the binary expansion is applied to the state variables in (2g), $x_t \in \mathbb{R}^{k_1} \times \mathbb{Z}^{k_2}$, it is reformulated to $x_t \in \{0,1\}^k$, such that all state variables in the MTHS problem are binary and representing the original continuous state variables to ε-accuracy. Note that $k \neq k_1 + k_2$, as it depends on the binary expansion.

Since the possible state variables solutions are given by a finite set of binary variables, we can generate a finite number of cuts to approximate it, as the convex hull of the binary state space ensures that we can compute tight bounds. See Figure 2 and its explanation below.

Another reformulation used in the SDDiP method is to generate local copies of the state variables. This reformulation is crucial to ensure that one is able to generate cuts that accurately approximate the EFP function. The DP formulation of the problem with binary state variables becomes

$$
\begin{aligned}
\underline{Q}_t^i(x_{t-1},\xi_{t-1}) :=& \max_{x_t,y_t,z_t} f_t(x_t,y_t) + \psi_t^i(x_t) && \text{(10a)}\\
\text{s.t.}\quad & (x_t,z_t,y_t) \in X_t^{''}(\tilde{\xi}_t) && \text{(10b)}\\
& z_t = x_{t-1} && \text{(10c)}\\
& z_t \in [0,1]^k && \text{(10d)}\\
& x_t \in \{0,1\}^k, y_t \in \mathbb{R}^{l_1}\cdot\mathbb{Z}^{l_2}. && \text{(10e)}
\end{aligned}
$$

Where (10c) is the essential copy constraint, connecting the previous state solution, x_{t-1}, and the copy variable z_t. The set $X_t^{''}$ is obtained by transforming all state variables in X_t into binaries. The EFP function, $\psi_t^i(x_t)$, is defined as (9), with the differentiation that all the state variables has to be binary.

IV. REVIEW OF THE SDDIP APPROACH

The SDDiP approach builds on the same principles as SDDP, where each main iteration comprises a forward and a backward iteration, as shown in Algorithm 1. We can summarize the main differences between SDDiP and SDDP into

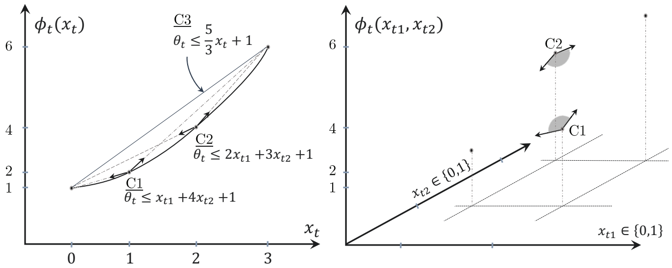

Fig. 2. Illustration of the nonconvex EFP function (left) and the transformation into the binary state (right). Note that we are solving a maximization problem and thus the function is nonconcave, but in line with standard terminology, we refer to it as a nonconvex function. Assume that the continuous variable x_t is transformed by two binary variables x_{t1} and x_{t2}, such that $x_t = x_{t1} + 2x_{t2}$. By constructing two hyperplanes (C1 and C2) in the binary state space, we can illustrate how they are projected to the continuous space. This illustrates how by using binary state variables we are able to construct cuts that approximate a nonconvex function. The hyperplane C3 is added to show the tightest cut we are able to generate from the convex relaxation of the nonconvex function.

three pillars; the requirement of discretization of state variables into binary variables, introduction of the copy variable and the different cut families it facilitates.

Scenarios are sampled from the underlying stochastic process in line 4 in Algorithm 1. The forward iteration solves the problem given in (10) for each stage and scenario, providing candidate solutions to the problem in line 9 and an α confidence lower bound on the optimal value.

The backward iteration operates along the N state trajectories obtained in the previous forward iteration to compute cuts that are passed backward in time. The different types of cuts considered are described in the next section. The value of the objective function in the first stage will then represent an upper bound of the true optimal solution in line 23. Convergence is achieved if the lower bound is within a statistical confidence interval of the upper bound.

Algorithm 1 The SDDiP Method

1: Set x_0^i, ξ_0^i, $i \leftarrow 1$, UB $= +\infty$ and LB $= -\infty$,
2: Apply binary expansion on continuous and integer state variables
3: **while** $i < i^{\max}$ **or** some other stopping criteria **do**
4: Sample N scenarios $\Omega^i = \xi_1^k, \ldots, \xi_T^k{}_{k=1,\ldots,N}$
5: /* Forward Iteration */
6: **for** k=1,. . . ,N **do**
7: **for** t=1,. . . ,T **do**
8: Solve $\underline{Q}_t^{ik}(x_{t-1}^{ik}, \xi_{t-1}^{ik})$
9: Collect solution $f_t(x_t^{ik}, y_t^{ik}, \xi_t^{ik}), x_t^{ik}, y_t^{ik}, z_t^{ik}$
10: $\text{lb}^k \leftarrow \Sigma_{t=1,\ldots,T} f_t(x_t^{ik}, y_t^{ik}, \xi_t^{ik})$
11: /* Compute lower bound */
12: $\mu \leftarrow \frac{1}{N}\Sigma_{k=1}^N \text{lb}^\text{k}$ and $\sigma^2 \leftarrow \frac{1}{N-1}\Sigma_{k=1}^N(\text{lb}^\text{k} - \mu)^2$
13: $\text{LB} \leftarrow \mu + \text{z}_\alpha \frac{\sigma}{\sqrt{\text{N}}}$
14: /* Backward Iteration */
15: **for** t=T,. . . , 2 **do**
16: **for** k=1,. . . , N **do**
17: **for** m$\in \mathcal{C}(t)$ **do**
18: Solve a suitable relaxation of
19: $\underline{Q}_m^{ik}(x_t^{ik}, z_t^{ik}, \xi_t^{ik})$
20: Collect cut coefficients and parameters
21: Add desired cut(s) as described in Sec. IV-A
22: /* Compute upper bound */
23: $\text{UB} \leftarrow \underline{\text{Q}}_1^\text{i}(\text{x}_0^\text{i}, \xi_0^\text{i})$
24: $i \leftarrow i + 1$

A. Cut Families

SDDiP is defined on a notion of *valid* and *tight* cuts. A cut is *valid* and *tight* if it supports the EFP function. For maximization problems, it is only *valid* if it provides an upper approximation of the EFP function. An illustration is given in Figure 2, where the cuts C1, C2 and C3 are all valid, but only C1 is tight when the state variable has the value 1.

The different cut families that could be used within SDDiP are briefly described below.

1) Benders Cuts (B): Benders cuts (see [32]) are constructed by solving the LP relaxation of problem (10). We obtain the following Benders cut for stage t in iteration i

$$\theta_t \leq \sum_{m \in \mathcal{C}(t)} q_{tm} \left[\underline{Q}_m^{\text{LP}} + (\pi_m^i)^\top (x_t - x_t^i) \right]. \tag{11}$$

Where π_m^i is a vector of dual values corresponding to (10c), q_{tm} is the probability of going to a child node m, in decision stage t and $\underline{Q}_m^{\text{LP}}$ is the optimal LP relaxation value. Convergence is not guaranteed with only Benders cuts for SDDiP. In some practical scheduling cases, the Benders cuts may be far from tight, as we will see in the Section VI-B.

2) Lagrangian Cuts (L): This cut family is based on Lagrangian relaxation, where we relax the copy constraint (10c). The Lagrangian multiplier is obtained by

$$\bar{\pi}_t^i := \underset{\bar{\pi}_t}{\operatorname{argmin}} \left\{ \mathcal{L}_t^i(\bar{\pi}_t) + \bar{\pi}_t^\top x_{t-1} \right\}, \tag{12}$$

and $\mathcal{L}_t^i$ is defined as

$$\mathcal{L}_t^i(\bar{\pi}_t) := \max_{x_t, y_t, z_t} f_t(x_t, y_t, \xi_t) + \psi_t^i(x_t) - \bar{\pi}_t^\top z_t \tag{13a}$$

$$\text{s.t.} \quad (z_t, x_t, y_t) \in X_t^{''}(\tilde{\xi}_t) \tag{13b}$$

$$z_t \in [0,1]^k \tag{13c}$$

$$x_t \in \{0,1\}^k, y_t \in \mathbb{R}^{l_1} \cdot \mathbb{Z}^{l_2}. \tag{13d}$$

Lagrangian relaxation often aims to relax complicating constraints and divide the problem into smaller subproblems that can more easily be computed. Our aim, however, is to find good multipliers that make the Lagrangian cuts as tight as possible. It is essential for the convergence of SDDiP that the Lagrangian cuts can be generated in a sufficient manner, in regard to both computation time and tightness. By repeatedly solving (12) and updating the Lagrange multipliers $\bar{\pi}_t$, e.g. using the subgradient [33] or level [34] methods, one can construct tight and valid Lagrangian cuts, as described in [27]. Nevertheless whichever method we use to get the Lagrangian multiplier, $\bar{\pi}_t^i$, the cut we construct is of the following form,

$$\theta_t \leq \sum_{m \in \mathcal{C}(t)} q_{tm} [\mathcal{L}_m^i(\bar{\pi}_m^i) + (\bar{\pi}_m^i)^\top x_t]. \tag{14}$$

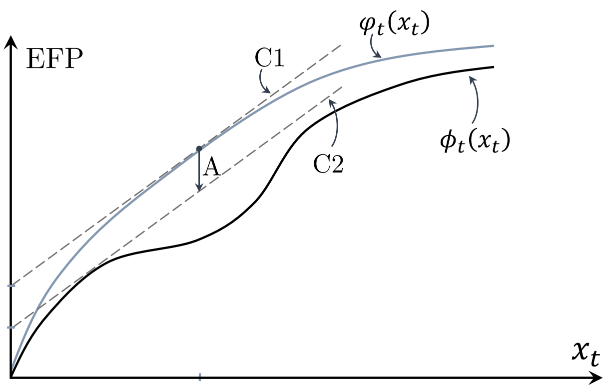

Fig. 3. Illustration of improvement for the strengthened Benders cut (C2) compared to Benders cut (C1) by generating a tighter right-hand-side parameter, equaling the distance of A. $\varphi_t(x_t)$ is the LP relaxation of the original nonconvex function $\phi_t(x_t)$.

3) Strengthened Benders Cuts (SB): The *strengthened Benders cuts* presented in [27] are parallel to the regular Benders cut and are at least as tight. They are not guaranteed to be tight, but they may improve the solution quality significantly compared to the Benders cuts, as illustrated in Figure 3. The increased precision comes at a modest increase in computation time.

Strengthened Benders cuts are constructed as follows. First, we solve the LP relaxation of problem (10). Then problem (13) is solved with $\bar{\pi}_t$ equals to the optimal LP dual solution with respect to constraint (10c), π_t^i. From the solution of the latter problem we obtain the right-hand side $\mathcal{L}_t^i(\pi_t^i)$ and then construct the following strengthened Benders cut

$$\theta_t \leq \sum_{m \in \mathcal{C}(t)} q_{tm}[\mathcal{L}_m^i(\pi_m^i) + (\pi_m^i)^\top x_t]. \quad (15)$$

The strengthened Benders cuts do not require the state variables to be binary and can such be used to achieve good solutions to problems that do not possess highly nonconvex properties.

4) Integer Optimality Cuts (I): The last family of cuts we have used in our method is the *integer optimality cuts* [35], [36]. The integer optimality cuts are valid, tight and finite and will, therefore, guarantee convergence. They are also very fast to generate, as they only rely on the solution obtained in the forward iteration.

B. Cut Discussion

As mentioned in Section I-B, the Lagrangian cuts proposed by [27], and used in this paper, differentiate from the Lagrangian cuts proposed by [22], [21]. Instead of introducing copy constraints they dualize the time-linking constraints and are therefore not able to guarantee tight cuts, even-though they in many cases are stronger than the Benders cuts. This can be shown by a small example. Consider,

$$Q(x) = \min_{y_1, y_2}\{y_1 + y_2 : 2y_1 + y_2 \geq 3x, \quad (16)$$
$$y_1 \in \{0, 1, 2\}, y_2 \in [0, 3]\}$$

Where $Q(x)$ is a value function dependent on the binary state x, and y_1 and y_2 are local variables. First, by dualizing the time-linking constraint and providing $x = 1$ as a candidate solution one gets the following Lagrangian cut $\theta \geq 1.5x$, this is essentially the method proposed by [22], [21] and as the small example shows it does not necessarily provide tight cuts. Furthermore, let's add a copy constraint $z = x, z \in [0, 1]$ and swap x with z in the time-linking constraint. By dualizing the copy constraint we can compute the Lagrangian cut $\theta \geq -1 + 3x$, with $x = 1$ as a candidate solution. For $x = 0$ we get that $\theta \geq 0$, such that we are able to provide a tight approximation of $Q(x)$.

As proven in [27] and partially illustrated above, the SDDiP algorithm will converge if Lagrangian or integer optimality cuts are added. Benders and strengthened Benders cuts will not guarantee convergence, but will often serve to improve convergence rate in concert with Lagrangian or integer optimality cuts. Benders cuts are by far the least computational demanding type of cuts since their construction only involves solving LP problems. For this reason, we only add Benders cuts until a stable upper bound is reached. Subsequently, we add the other cut families to further tighten the EFP functions.

C. The Approximate SDDiP Problem

The SDDiP method depends on the aforementioned three pillars to ensure finite convergence. However, we are still able to use some of its features if we do not restrict the state variables to be binary. Finite convergence can then no longer be guaranteed, with the trade-off of reduced computational complexity. We reformulate (7) by applying the same concepts of copy variables and copy constraints. For this problem we can only compute Benders and Strengthened Benders cuts. This is done in the same manner as for the (10) in Section IV-A1 and IV-A3. Results from this approach is also reported in Section VI-B

V. Bounding the EFP Function

Due to the bilinear objective term introduced when considering uncertain energy price, we relaxed the problem by artificial variables and constraints containing the big-M notation in (6). As we will elaborate, it is difficult to obtain tight bounds for that problem formulation. Consider the case where a given energy price state variable is $x_{tj} = 1$, we then have

$$\pi_{t,j}^E = \frac{\partial Q_t^i}{\partial x_{tj}} = \frac{\partial Q_t^i}{\partial w_{tj}} \cdot \frac{\partial w_{tj}}{\partial x_{tj}} = \frac{\partial Q_t^i}{\partial w_{tj}} \cdot M_{tj}, \quad (17)$$

where $\pi_{t,j}^E$ is the dual value for the copy constraints for that energy price state variable in (10c).

From a practical standpoint, it is obvious that there are time periods and scenarios where y_t^G is far from $y^{G,\max}$. In this sense the proposed relaxation method has the same drawback as the integer optimality cut, but now the big-M notation cascades into all types of cuts, making it hard to compute cuts that are tight for not only the given candidate solution but nearby solutions as well. A modest upside is that the big-Ms depend on a physical value and not the expected future

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2018.2805164, IEEE Transactions on Sustainable Energy

profit, contributing to a tighter formulation than for the integer optimality cuts.

Given the nature of the problem at hand, different techniques could be applied in order to make the big-Ms tighter by adding some heuristics on the bounds of y_t. Nevertheless, for an MTHS problem with a price taker assumption and no demand, this has proven difficult, as the only limiting constraints are generation limits and available water in the reservoirs.

Observe that the big-M formulation is only required for our implementation of the uncertainty of the energy price. The following theorem establishes a method for constructing an upper bound to this problem so as to be able to evaluate the policy quality.

Theorem 1. *Let v^* be the optimal value of the MTHS problem in* (2) *with price uncertainty, and v^{LP} be the optimal value of its LP relaxation. Let $\bar{v}^{\mathrm{LP}}$ be the optimal value of the LP relaxation where the uncertain energy price is replaced by its expected value. Then*

$$v^* \leq v^{\mathrm{LP}} \leq \bar{v}^{\mathrm{LP}}.$$

For brevity, we omit a detailed proof of the above result and provide a sketch. Note that the first inequality is simply owing to the LP relaxation. Since the price uncertainties are only on the right-hand side (recall the modeling approach discussed in Section III-B), we know that the optimal value of LP relaxation is a concave function of the uncertain parameters [6]. It follows by Jensen's inequality that $v^{\mathrm{LP}} \leq \bar{v}^{\mathrm{LP}}$, hence an upper bound to the original problem can be constructed by using the upper bound computed with only Benders cuts applied to the problem where energy price is modeled by its expected value.

VI. Case Study

For the case study, we considered a Norwegian hydropower reservoir system that consists of reservoirs with both short- and long-term storage capacity. The case study is documented in previous publications [15], [26]. The system contains three hydropower reservoirs, with two power stations and a total installed capacity of 414 MW. One of the modeling issues with this system is that the two lower reservoirs, with uneven water head, are connected to the same power station, such that the station can only generate from one of them at a time. Considering that one of the reservoir's size is 2% of the other with an equal amount of inflow, linear SDDP tended to overestimate its ability to avoid spillage. This is handled more precisely in the SDDiP framework.

The mean and standard deviation parameters for the stochastic processes were extracted from price forecasts obtained from a fundamental market model using inflow from 70 years of historical data as input. For simplicity, we only use a single inflow and energy price series in this work, as extensions to multiple are straightforward.

The stagewise decision problem of the SDDiP formulation consists of 137 constraints, 182 variables (120 continuous and 62 binary), 52 decision stages and 9 branches for each decision stage. In each forward iteration we sampled 2 scenarios and restricted the number of forward and backward iterations to 50. Lastly, a final simulation of 300 forward scenarios was performed to evaluate the policy by calculating a 95% confidence interval of the lower bound. For an even comparison between the different methods we use the same scenarios in the forward iteration. Computations were performed on a Dell Latitude E7240 with an Intel Core i7-4600U processor with 2.7 GHz clock rate and 16 GB RAM. The problem was formulated in Python 3.5 with Gurobi 7.0 as the mathematical solver. We did not implement any parallel processing except from default settings in the MIP solver.

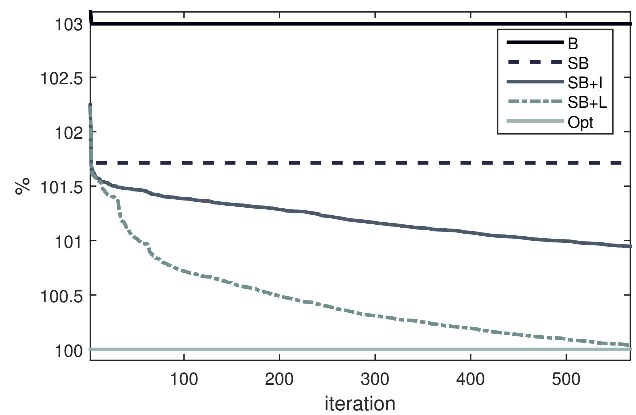

Fig. 4. Illustration of the upper bound for SDDiP of the MTHS problem with 6 stages, convex generation function and expected values for the stochastic parameters. Values are given with the optimal value as base.

A. Validation

In order to validate the convergence properties of our implementation of the SDDiP method for the MTHS problem, a case was constructed with only 6 decision stages, convex generation function and the stochastic parameters modeled by their expected value. The upper bounds for different combinations of cut families are illustrated in Figure 4. Due to the reduced scale of the problem it is possible to compute the true optimal value, illustrated by the lower line in the figure. The figure shows the results from iteration 3 to 567, where convergence is obtained using strengthened Benders and Lagrangian cuts (SB+L). The reason to use Lagrangian and integer optimality cuts combined with strengthened Benders is that strengthened Benders cuts rapidly find a good stable solution, that the other cuts can start from. From the figure, it is clear that Benders (B) and strengthened Benders (SB) does improve only up to a certain point, from there either Lagrangian or integer optimality cuts are required to obtain convergence. This does, however, come at a cost of significant increase of computational time. The benefit of using strengthened Benders cuts compared to Benders cuts is also evident, as we see the gap is almost reduced by half.

B. Results

Now we turn to the 52-stage MTHS problem defined in Section II where two different case studies were performed. First, in Case I we modeled only uncertainty of inflow, then in Case II we also included uncertainty of the energy price.

TABLE I
CASE I, INFLOW UNCERTAINTY. SDDIP (TOP) AND APPROXIMATE PROBLEM (BOTTOM).

	UB [kEUR]	stat. LB [kEUR]	gap [%]	time [s]
B	47 353.72	40 776.87	13.89	16 792
B-SB	42 907.58	41 135.43	4.13	17 716
B-SB-I	42 854.80	41 197.25	3.87	30 277
B	47 030.08	40 973.28	12.88	742
B-SB	42 906.90	41 350.85	3.63	1 176

TABLE II
CASE II, INFLOW AND ENERGY PRICE UNCERTAINTY. SDDIP (TOP) AND APPROXIMATE PROBLEM (BOTTOM).

	UB [kEUR]	stat. LB [kEUR]	gap [%]	$\underline{\text{gap}}$ [%]	time [s]
B	107 469.94	42 276.74	60.66	11.24	11 788
B-SB	84 380.16	43 802.85	48.09	7.37	13 896
B-SB-I	84 291.78	43 748.69	48.10	7.50	18 966
B	107 449.36	42 444.20	60.50	10.80	2 279
B-SB	84 367.26	43 891.43	47.98	7.15	2 869

Subsequently, we tested the SDDiP method using Benders, strengthened Benders and integer optimality cuts. Lagrangian cuts were omitted due to computational requirement. We also computed the approximate problem, as defined in Section IV-C, with the Benders and strengthened Benders cut families. Recall that for the approximate problem the state variables are not required to be binary. The solution of the approximate problem, when only using Benders cuts, will also give an indication on how well the generic SDDP method performs for the MTHS problem.

The expected profit for the given case study with a final simulation of 300 forward scenarios is shown in Table I and Table II for Case I and II, respectively. The reported computation time includes final simulation and the gap is given as $\frac{UB-LB}{UB}$.

We obtained good solutions after 50 iterations in both the SDDiP and for the approximate problem. The approximate problem is solved faster as there are less binary variables, and the convergence gap is better for the given amount of iterations. When comparing the solutions from the Benders to the strengthened Benders cut families, for both SDDiP and the approximate problem, a significant improvement of the convergence gap is observed. This can also be seen in the convergence plot in Figure 5. We see that after iteration 20 when the strengthened Benders cuts were added, both the upper and lower bound for both cases improves.

The case with uncertain energy price results in a large gap. When using strengthened Benders cuts a significant improvement is observed compared to only Benders cuts. By using the upper bound from Theorem 1 we can use the approximate problem's value with Benders cuts and compute a tighter gap, referred to as $\underline{\text{gap}}$ in Table II. It shows that eventhough the upper bound initially was very high, the policy that it provides is quite good. Notably is the improvement with strengthened Benders cuts for the approximate problem with superior policy and computation time compared to the other results.

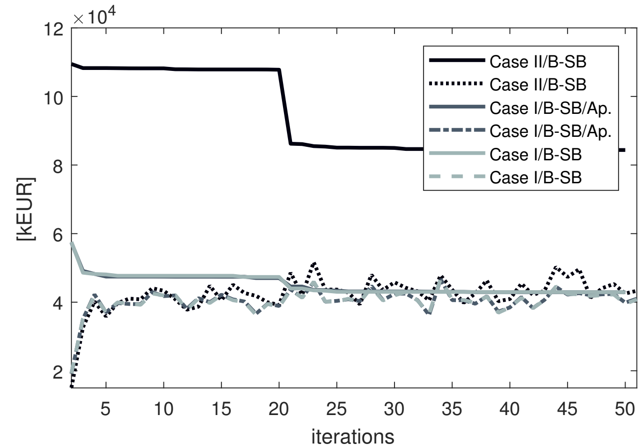

Fig. 5. Convergence plot for the different case studies with Benders and strengthened Benders cuts. The solid line indicates the upper bound and the other the expected lower bound. One can clearly observe how large the gap difference is between Case I and II. The approximate method also follows SDDiP close as seen for the dashed and dash-dot lines.

VII. DISCUSSION

We found that the integer cuts are not very computationally efficient compared to solution improvement when applied for the given case study. Lagrangian cuts were omitted from the large case due for the same reason, and are therefore not represented. On the other hand, the strengthened Benders cut significantly improves solution quality compared to ordinary Benders cuts with only a modest increase in computational effort. We believe this is due to the characteristics of the presented MTHS case; it seems more important to adjust the right-hand side of the EFP function than to fine-tune its shape by adjusting the cut coefficients. Many Norwegian hydropower stations have high water head and their generation functions are not significantly affected by head variations. It may be that a nonconvex generation function with water head dependencies found in other systems, e.g. as reported in [22], will lead to a more pronounced shape of the EFP function, further justifying the need for Lagrangian and integer cuts.

Due to the introduction of very large numbers in the constraint matrix when adding integer optimality cuts, we have observed increased computation time for the subproblems. When comparing the computation time between the first stage subproblem for Case II after 50 iterations with SB and SB+I we observe an increase from 0.02 to 0.08 seconds. It is clear that the added constraints with large coefficients make it difficult for the solver to find a solution quickly. Subsequently, the integer optimality cuts do not provide good enough solution improvements to justify the added computational burden we observe, as seen in Table I and II.

Lastly, we observed that the lower bounds are higher for the approximate problem, when comparing computations with the same cut families, e.g. for Case I with Benders cuts the approximate problem has a lower bound of 40 973.28 and the SDDiP problem has 40 776.87. This follows from how the VAR-1 model is altered to a less than or equal constraint to ensure feasibility of the model, as discussed in Section III-B. In the final simulation, where we compute the lower bound, there are some extreme scenarios where the model are required to use an artificial penalty variable that ensures relative complete recourse. Since the inflow state in the SDDiP formulation is somewhat less than that of the approximate, due to the relaxation of the equality constraint, a subsequently higher penalty is observed in the order of the difference of the lower bounds that explains this observation.

VIII. Conclusion

With the earlier mentioned challenges flexible power producers encounter, we believe that binary state expansion and the SDDiP algorithm gives the producers a useful tool to address new challenges that involve nonconvexities in the modeling.

We observed that our method performs very well when omitting the uncertainty modeling of energy price, due to the drawback of the big-M formulation. Nonetheless, it allowed us to model the stochastic processes within a unified framework and we showed that the resulting policy from SDDiP is very good. Methods to improve the implementation of the uncertainty of the energy price will be investigated in future research, including methods on how to avoid the relaxation of the VAR-1 constraint.

In the case study, we found that the strengthened Benders cuts are superior to Benders cuts in finding good policies and bounds. We also found that the strengthened Benders cuts can be used in an approximate SDDiP method to provide satisfying results for nonconvex MTHS problems, that also includes a nonconvex generation function. To eventually close the optimality gap, however, one would have to use SDDiP and add either Lagrangian or Integer optimality cuts. Further, we observed that adding Integer optimality cuts did not improve the results enough to justify the increased computational time. We emphasize that these findings are case-specific, and expect that the efficiency of Lagrangian and Integer optimality cuts will be higher in scheduling problems with more pronounced nonconvexities.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

References

[1] J. W. Labadie, "Optimal operation of multireservoir systems: State-of-the-art review," *Journal of Water Resources Planning and Management*, vol. 130, no. 2, pp. 93–111, 2004.

[2] G. Steeger, L. A. Barroso, and S. Rebennack, "Optimal bidding strategies for hydro-electric producers: A literature survey," *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1758–1766, 2014.

[3] O. B. Fosso, A. Gjelsvik, A. Haugstad, B. Mo, and I. Wangensteen, "Generation scheduling in a deregulated system. the norwegian case," *IEEE Transactions on Power Systems*, vol. 14, no. 1, pp. 75–81, 1999.

[4] R. Scharff, J. Egerer, and L. Söder, "A description of the operative decision-making process of a power generating company on the Nordic electricity market," *Energy Systems*, vol. 5, pp. 349–369, 2014.

[5] S.-E. Fleten and T. K. Kristoffersen, "Short-term hydropower production planning by stochastic programming," *Computers & Operations Research*, vol. 35, no. 8, pp. 2656–2671, 2008.

[6] J. R. Birge and F. Loveaux, *Introduction to Stochastic Programming*. Springer, 2 ed., 2011.

[7] J. R. Birge, "Decomposition and partitioning methods for multi-stage stochastic linear programs," *Operations Research*, vol. 33, pp. 989–1007, 1985.

[8] M. V. F. Pereira and L. M. V. G. Pinto, "Stochastic optimization of a multireservoir hydroelectric system: A decomposition approach," *Water Resources Research*, vol. 21, no. 6, pp. 779–792, 1985.

[9] J. Jacobs, G. Freeman, J. G. Grygier, D. Morton, G. Schultz, K. Staschus, and J. Stedinger, "SOCRATES: A system for scheduling hydroelectric generation under uncertainty," *Annals of Operations Research*, vol. 59, no. 1, pp. 99–133, 1995.

[10] M. V. F. Pereira and L. M. V. G. Pinto, "Multi-stage stochastic optimization applied to energy planning," *Mathematical Programming*, vol. 52, pp. 359–375, 1991.

[11] A. B. Philpott and Z. Guan, "On the convergence of stochastic dual dynamic programming and related methods," *Operations Research Letters*, vol. 36, no. 4, pp. 450–455, 2008.

[12] A. Shapiro, W. Tekaya, J. P. da Costa, and M. P. Soares, "Risk neutral and risk averse stochastic dual dynamic programming method," *European Journal of Operational Research*, vol. 224, no. 2, pp. 375–391, 2013.

[13] A. Philpott and V. de Matos, "Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion," *European Journal of Operational Research*, vol. 218, no. 2, pp. 470 – 483, 2012.

[14] A. Brigatto, A. Street, and D. M. Valladao, "Assessing the cost of time-inconsistent operation policies in hydrothermal power systems," *IEEE Transactions on Power Systems*, 2017. DOI: 10.1109/TPWRS.2017.2672204.

[15] M. N. Hjelmeland, M. Korpås, and A. Helseth, "Combined SDDP and simulator model for hydropower scheduling with sales of capacity," in *2016 13th International Conference on the European Energy Market (EEM)*, pp. 1–5, June 2016.

[16] A. Helseth, M. Fodstad, and B. Mo, "Optimal medium-term hydropower scheduling considering energy and reserve capacity markets," *IEEE Transactions on Sustainable Energy*, vol. PP, no. 99, pp. 1–9, 2016.

[17] A. Street, A. Brigatto, and D. M. Valladao, "Co-optimization of energy and ancillary services for hydrothermal operation planning under a general security criterion," *IEEE Transactions on Power Systems*, 2017. DOI: 10.1109/TPWRS.2017.2672555.

[18] S. Rebennack, "Combining sampling-based and scenario-based nested Benders decomposition methods: application to stochastic dual dynamic programming," *Mathematical Programming*, vol. 156, no. 1, pp. 343–389, 2016.

[19] H. Poorsepahy-Samian, V. Espanmanesh, and B. Zahraie, "Improved inflow modeling in stochastic dual dynamic programming," *Journal of Water Resources Planning and Management*, vol. 142, no. 12, 2016.

[20] B. Flach, L. Barroso, and M. Pereira, "Long-term optimal allocation of hydro generation for a price-maker company in a competitive market: latest developments and a stochastic dual dynamic programming approach," *IET Generation, Transmission & Distribution*, vol. 4, no. 2, pp. 299–314, 2010.

[21] S. Cerisola, J. M. Latorre, and A. Ramos, "Stochastic dual dynamic programming applied to nonconvex hydrothermal models," *European Journal of Operational Research*, vol. 218, pp. 687–897, 2012.

[22] F. Thome, M. V. F. Pereira, and M. Granville, Sergio Fampa, "Nonconvexities representation on hydrothermal operation planning using SDDP." http://www.psr-inc.com/publicacoes/, Working paper, 2013.

[23] G. Steeger and S.Rebennack, "Dynamic convexification within nested benders decomposition using lagrangian relaxation: An application to the strategic bidding problem," *European Journal of Operational Research*, vol. 257, no. 2, pp. 669–686, 2017.

[24] H. Abgottspon, K. Njálsson, M. A. Bucher, and G. Andersson, "Risk-averse medium-term hydro optimization considering provision of spinning reserves," in *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, (Durham, England), 2014.

[25] A. Diniz and M. E. P. Maceira, "A four-dimensional model of hydro generation for the short-term hydrothermal dispatch problem considering head and spillage effects," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1298 – 1308, 2008.

[26] M. N. Hjelmeland, A. Helseth, and M. Korpås, "A case study on medium-term hydropower scheduling with sales of capacity," *Energy Procedia*, vol. 87, pp. 124 – 131, 2016.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2018.2805164, IEEE Transactions on Sustainable Energy

[27] J. Zou, S. Ahmed, and X. A. Sun, "Stochastic dual dynamic integer programming," Submitted for publication, 2016. Available at Optimization Online: http://www.optimization-online.org/DB_HTML/2016/05/5436.html.

[28] A. B. Philpott, F. Wahid, and F. Bonnans, "MIDAS: A mixed integer dynamic approximation scheme," *Optimization-online*, 2016.

[29] H. Heitsch and W. Römisch, "Scenario reduction algorithms in stochastic programming," *Computational Optimization and Applications*, vol. 24, pp. 187–206, 2003.

[30] A. Gjelsvik, B. Mo, and A. Haugstad, *Long- and Medium-term Operations Planning and Stochastic Modelling in Hydro-dominated Power Systems Based on Stochastic Dual Dynamic Programming*, book section 2, pp. 33–55. Energy Systems, Springer Berlin Heidelberg, 2010.

[31] F. Glover, "Improved linear integer programming formulations of nonlinear integer problems," *Management Science*, vol. 22, no. 4, pp. 455–460, 1975.

[32] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numerische mathematik*, vol. 4, no. 1, pp. 238–252, 1962.

[33] M. Held, P. Wolfe, and H. P. Crowder, "Validation of subgradient optimization," *Mathematical programming*, vol. 6, no. 1, pp. 62–88, 1974.

[34] C. Lemaréchal, A. Nemirovskii, and Y. Nesterov, "New variants of bundle methods," *Mathematical programming*, vol. 69, no. 1, pp. 111–147, 1995.

[35] G. Laporte and F. V. Louveaux, "The integer l-shaped method for stochastic integer programs with complete recourse," *Operations research letters*, vol. 13, no. 3, pp. 133–142, 1993.

[36] G. Angulo, S. Ahmed, and S. S. Dey, "Improving the integer l-shaped method," *INFORMS Journal on Computing*, vol. 28, no. 3, pp. 483–499, 2016.

Martin N. Hjelmeland received the M.Sc. degree in electric power engineering from the Norwegian University of Science and Technology (NTNU) in 2015. He is currently pursuing a Ph.D. at the same location where his research-topic is on medium-term hydropower scheduling in multiple markets.

Jikai Zou received the B.S degree in Mathematics from Peking University, Beijing, China, in 2000; the M.S. and PhD degrees in Operations Research, respectively, from Purdue University, West Lafayette, IN, in 2013, and Georgia Institute of Technology, Atlanta, GA, in 2017.

Arild Helseth (M'10) was born on Stord, Norway in 1977. He received the M.Sc. and Ph.D. degrees in electrical power engineering from the Norwegian University of Science and Technology. Currently he works at SINTEF Energy Research with hydro-thermal and hydropower scheduling models and methods.

Shabbir Ahmed is the Anderson-Interface Chair and Professor in the H. Milton Stewart School of Industrial & Systems Engineering at the Georgia Institute of Technology. His research interests are in stochastic and discrete optimization. He is a Fellow of INFORMS and a Senior Member of IEEE.