A decomposition method by component for the optimization of maintenance scheduling

Thomas Bittar, Jean-Philippe Chancelier, Pierre Carpentier, Jérôme Lonchampt

A decomposition method by component for the optimization of maintenance scheduling

Une méthode de décomposition par composant pour l'optimisation de la maintenance

Thomas Bittar
Département PRISME, EDF R&D
Chatou, France
CERMICS, École des Ponts ParisTech
Champs sur Marne, France
thomas.bittar@edf.fr

Jean-Philippe Chancelier
CERMICS, École des Ponts ParisTech
Champs sur Marne, France
jean-philippe.chancelier@enpc.fr

Pierre Carpentier
UMA, ENSTA Paris
Palaiseau, France
pierre.carpentier@ensta-paris.fr

Jérôme Lonchampt
Département PRISME, EDF R&D
Chatou, France
jerome.lonchampt@edf.fr

***Abstract*— We present a decomposition method based on the Auxiliary Problem Principle to design optimal maintenance scheduling policies for systems of physical components (turbines, generators, transformers) sharing a common stock of spare parts. The method outperforms a reference blackbox method on a system with 80 components.**

***Résumé*— Nous présentons une méthode de décomposition basée sur le Principe du Problème Auxiliaire afin de déterminer des politiques de maintenance optimales pour des systèmes de composants physiques (turbines, alternateurs, transformateurs) partageant un stock commun de pièces de rechange. Cette méthode obtient de meilleures performances que la méthode boîte noire de référence sur un système de 80 composants.**

***Keywords*— maintenance scheduling, stochastic optimization, decomposition-coordination**

I. INTRODUCTION

In industry, maintenance aims at improving the availability of physical assets and therefore impacts the overall performance of a system. There exists two main kind of maintenance: corrective and preventive. Corrective maintenance (CM) is performed in reaction to a breakdown. Preventive maintenance (PM) consists in repairing or replacing a component before a failure. Maintenance policies have an important economic impact and are therefore studied in various areas such as the electricity sector [1], the manufacturing industry [2] or civil engineering [3]. In the electricity sector, maintenance optimization plays a major role in ensuring a reliable and competitive electricity production.

In this work, we consider components of hydroelectric power plants such as turbines, transformers or generators. We study a system of a given type of components that share a common stock of spare parts. The time horizon is 40 years. Over time, components experience random failures that occur according to known failure distributions. Thus, the dynamics of the system is stochastic. A preventive strategy consists in choosing the dates of replacement for each component of the system. The goal is to find a preventive strategy that minimizes an expected cost depending on maintenance and on the occurrences of forced outages of the system. Operational constraints impose to only look for deterministic maintenance strategies. This means that the dates of PM are chosen at the beginning of the time horizon with only a statistical knowledge of the future dates of failure. This differs from condition-based maintenance [4] where maintenance decisions are taken given the online observation of the degradation state of the components, making the strategy stochastic. The numerical experiments should involve systems constituted of up to 80 components in order to model the most demanding industrial case at EDF. This leads to optimization problems in high dimension that are numerically challenging.

Many studies consider time-based [5], [6] or age-based [7], [8] maintenance policies. Such strategies are only defined with one decision variable per component: either the periodicity of maintenance (time-based) or the age at which a component is replaced (age-based). In this paper, more general strategies are considered as we can decide whether or not to perform a PM at each time step for each component. Suppose that there are T time steps and n components, then our maintenance strategy is defined by nT decision variables instead of n variables for time-based or age-based strategies. The effort is justified as we consider a system on a long-term horizon where the cost incurred by forced outages are of the order of millions of euros. Then, even a minor improvement in the maintenance strategy generates important savings.

As highlighted in [1], a frontal resolution is impracticable for high-dimensional problems and decomposition methods are relevant. Previous works use a linear relaxation to apply decomposition techniques such as Benders [9] or Dantzig-Wolfe decomposition [10]. Our problem is a non-linear mixed integer program. The originality of our work lies in the use of a continuous relaxation of the system on which we apply a decomposition-coordination method based on variational techniques [11]. To our knowledge such a decomposition scheme has not been applied for optimal maintenance scheduling.

Originated from the work of [12]–[15], decomposition-coordination methods consist in splitting the original large-scale optimization problem into several independent subproblems of smaller size that can be solved efficiently. The subproblems are coordinated to ensure that the concatenation of solutions leads to an optimal solution of the original problem. Different types of decomposition-coordination schemes have been designed, by prices, by quantities or by prediction. They have been unified within the Auxiliary Problem Principle [16].

In our setting, each subproblem consists in optimizing the maintenance on a single component. The decomposition algorithm iteratively solves the subproblems with the blackbox algorithm MADS [17] and coordinates the solutions of these subproblems in order to reach a global optimum. We apply the decomposition method on relaxed systems with up to 80 components. The most demanding case takes around 20 hours of computation time. We show that in high dimension the decomposition method outperforms the blackbox algorithm applied directly on the original problem.

The paper is organized as follows: in Section II, we describe the industrial system and formulate the maintenance optimization problem. A decomposition method based on the Auxiliary Problem Principle is applied in Section III. Section IV contains numerical results showing the efficiency of the method in high dimension. Finally, in Section V, we conclude and give directions for future research.

II. System modeling and maintenance optimization problem

We start by describing the model of the studied industrial system and formulate the maintenance optimization problem. In the sequel, the notation $\langle \cdot , \cdot \rangle$ represents the inner product in a Hilbert space and $\|\cdot\|$ is the induced norm. For any vector $v = (v_1, \dots, v_n)$, we denote the first k components of v by:

$$v_{1:k} = (v_1, \dots, v_k)\,.$$

Random variables are always denoted by capital bold letters.

A. Description of the system

We consider a system of $n \in \mathbb{N}^\star$ physical components of a given type (generators, turbines or transformers) from a hydropower plant sharing a common stock of spare parts. A sketch of the system with $n = 2$ components is represented in Fig. 1. A corrective maintenance (CM) consists in the replacement of a component after a failure. A preventive maintenance (PM) is a planned replacement of a component before a failure.

We study this system on a horizon $T \in \mathbb{N}^\star$. In the sequel, $i \in \mathbb{I} = \{1, \dots, n\}$ denotes a component index, $t \in \mathbb{T} = \{0, \dots, T\}$ denotes a time step.

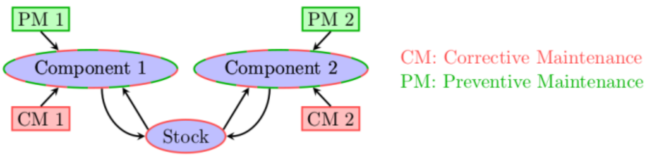

Fig. 1: System of two components sharing the same stock of spare parts

1) Characterization of the stock and the components

The safety stock over time is characterized by the sequence $S = (S_0, \dots, S_T) \in \mathcal{S} \coloneqq [\![0, s]\!]^{T+1}$, where S_t is the number of available spare parts at time t and $s \in \mathbb{N}$ is the maximum number of spare parts. The initial stock is set to $S_0 = s$. The replenishment delay for the parts, that is, the time from order to delivery of a part, is known and denoted by $D \in \mathbb{N}$.

At time t, component i is described by the vector $X_{i,t} \in \mathbb{R}^p$ which contains the following information:

- The regime of the component. A component has only two regimes: in the healthy regime, it runs in its nominal operating point. In the broken regime, it stops working completely. Initially all components are healthy.
- The age of the component (if healthy) or the time for which it has failed (if broken). Initially the components are new.
- The time elapsed since the last D failures of the component, where we recall that D is the number of time steps for the supply of spare parts. This information is useful to compute the dates of replenishment of the stock.

The state of the system is then described at t by the vector $(X_{1,t}, \dots, X_{n,t}, S_t)$. Finally, to describe the components over the whole study period, we introduce

$$X = (X_1, \dots, X_n) = \Big(\big(X_{1,0}, \dots, X_{1,T}\big), \dots, \big(X_{n,0}, \dots, X_{n,T}\big)\Big)\,.$$

For $i \in \mathbb{I}$, we introduce the space $\mathcal{X}_i = (\mathbb{R}^p)^{T+1}$ so that $X_i \in \mathcal{X}_i$ and $X \in \mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$. In order to emphasize that X depends on all the components of the system, we sometimes use the notation $X_{1:n}$ instead of X.

2) Preventive maintenance (PM) strategy

A PM consists in repairing a component although it is in the healthy regime. The dates of PM can be different for each component. They define the preventive maintenance strategy of the system. Operational constraints impose to look for deterministic strategies. This means that the dates of PM are chosen without any knowledge on the state of the system after the beginning of the time horizon and cannot be changed during the study. The maintenance strategy is defined by a vector

$$u = (u_1, \dots, u_n) = \Big(\big(u_{1,0}, \dots, u_{1,T}\big), \dots, \big(u_{n,0}, \dots, u_{n,T}\big)\Big)\,,$$

where for all $(i,t) \in \mathbb{I} \times \mathbb{T}$, $u_{i,t} \in [0,1]$ characterizes the PM for component i at time t. We introduce $\mathbb{U}_i = [0,1]^{T+1}$ so that $u_i \in \mathbb{U}_i$ and $u \in \mathbb{U} = \prod_{i=1}^n \mathbb{U}_i$. A value $u_{i,t} = 1$ means that a PM of component i is performed at t whereas $u_{i,t} = 0$ means that no PM is performed. As we aim at applying a decomposition method which is based on variational techniques, we consider a continuous relaxation of the system. This is why $u_{i,t}$ is allowed to take values in the whole interval

[0,1] and not just in the set {0,1}. Let us give a meaning for the values $u_{i,t} \in [0,1]$. We set a threshold $0 < \nu < 1$: a control $u_{i,t} \geq \nu$ corresponds to a rejuvenation of the component proportional to $u_{i,t}$ and a value $u_{i,t} < \nu$ corresponds to not performing a PM. The threshold ν is introduced for technical reasons linked to the behavior of the optimization. We consider that the maintenance operation is instantaneous and that it does not use parts from the stock. The reason is that PM are planned in advance, hence it is possible to order the parts so that they arrive just on time for the maintenance operation. The modeled stock corresponds only to the safety stock. As PM do not interact with the stock, there is no green border on the stock on Fig. 1.

3) Failures of the components

In our study, the failure distribution of component i is a known Weibull distribution of parameters (β_i, λ_i) denoted by $\mathrm{Weib}(\beta_i, \lambda_i)$, for which the cumulative distribution function F_i is given by:

$$F_i(x) = 1 - e^{-\left(\frac{x}{\lambda_i}\right)^{\beta_i}}$$

The probability of failure of a component at a given time step only depends on its age and its failure distribution. More precisely, assume that component i has age $a \geq 0$ at time t. Then, the probability of failure of component i at $t + \Delta t$ conditionally to the component being healthy at t is given by

$$p_i(a) = \frac{F_i(a + \Delta t) - F_i(a)}{1 - F_i(a)} .$$

In order to model the random failures of the components, we introduce the random process

$$\boldsymbol{W} = \{\boldsymbol{W}_{i,t}\}_{(i,t) \in \mathbb{I} \times \mathbb{T}}$$

defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that all $\boldsymbol{W}_{i,t}$ are independent random variables and follow a uniform distribution on $[0,1]$. Suppose that component i has age a at time t. If $\boldsymbol{W}_{i,t+1} < p_i(a)$, then component i fails at $t + 1$, otherwise no failure occurs.

B. Dynamics of the system

The dynamics of the system is stochastic because it depends on random failures of the components. Hence, the variables that describe the components and the stock are random variables and are represented using the bold characters $\boldsymbol{X}$ and $\boldsymbol{S}$ in the sequel.

1) Dynamics of a component

The dynamics of a component between t and $t + 1$ is described by Fig. 2.

- Suppose that component i is healthy with age a. If $u_{i,t} \geq \nu$, then a PM is performed and the component is rejuvenated proportionally to $u_{i,t}$, its age becomes $(1 - u_{i,t})(a + 1)$. Note that in the case where $u_{i,t} = 1$, a PM makes the component as good as new. If $u_{i,t} < \nu$, then no PM is performed and the component fails with probability $p_i(a)$.

- If component i is broken, it is replaced as soon as there is an available spare part in the stock.

We write the *dynamics of component* i over the whole time horizon as:

$$\Theta_i(\boldsymbol{X}_i, \boldsymbol{S}, u_i, \boldsymbol{W}_i) = 0 \in \mathcal{L}_i,$$

where $\mathcal{L}_i = (\mathbb{R}^p)^{T+1}$ and $\Theta_i = \{\Theta_{i,t}\}_{t \in \mathbb{T}}$ such that:

$$\begin{cases} \Theta_{i,0}(\boldsymbol{X}_i, \boldsymbol{S}, u_i, \boldsymbol{W}_i) = \boldsymbol{X}_{i,0} - x_i \\ \Theta_{i,t+1}(\boldsymbol{X}_i, \boldsymbol{S}, u_i, \boldsymbol{W}_i) = \boldsymbol{X}_{i,t+1} - f(\boldsymbol{X}_{i,t}, \boldsymbol{S}_t, u_{i,t}, \boldsymbol{W}_{i,t+1}), \end{cases}$$

where $x_i \in \mathbb{R}^p$ is the initial state of component i, assumed to be new, and f represents the dynamics described by Fig. 2.

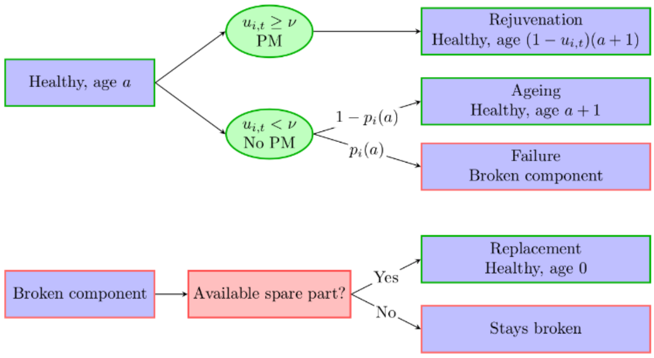

Fig. 2: Dynamics of a component

2) Dynamics of the stock

For the safety stock, the initial number of spare parts is $S_0 = s$. As a PM can be anticipated, we consider that the needed spares are ordered so that they arrive just on time for the scheduled maintenance. A part is used for each CM and a new part is ordered only after the failure of a component. This is a $(s-1; s)$ stock point inventory policy, used for long lead time, expensive and rarely used spare parts. The number of time steps for the supply of a part is D. If there are more broken components than parts in the stock, we replace as many components as possible. The other components stay in a broken state until new parts arrive in the stock. We write the *dynamics of the stock* as:

$$\Theta_S(\boldsymbol{X}_{1:n}, \boldsymbol{S}) = 0 \in \mathcal{L}_S,$$

where $\mathcal{L}_S = \mathbb{R}^{T+1}$ and $\Theta_S = \{\Theta_{S,t}\}_{t \in \mathbb{T}}$ such that:

$$\begin{cases} \Theta_{S,0}(\boldsymbol{X}_{1:n}, \boldsymbol{S}) = \boldsymbol{S}_0 - s \\ \Theta_{S,t+1}(\boldsymbol{X}_{1:n}, \boldsymbol{S}) = \boldsymbol{S}_{t+1} - f_S(\boldsymbol{X}_{1:n,t}, \boldsymbol{S}_t), \end{cases}$$

where f_S represents the dynamics of the stock that has just been described in this part. Note that $\boldsymbol{S}_{t+1}$ depends on the current level of stock $\boldsymbol{S}_t$ but also on $\boldsymbol{X}_{i,t}$ for all $i \in \mathbb{I}$ as the future stock depends on the state of all the components of the system. We say that the stock is coupling all the components.

Finally, the *dynamics of the whole system* is summarized by the equality constraint

$$\Theta(\boldsymbol{X}, \boldsymbol{S}, u, \boldsymbol{W}) = 0 \in \left(\prod_{i=1}^{n} \mathcal{L}_i\right) \times \mathcal{L}_S ,$$

with $\Theta = \{\{\Theta_i\}_{i \in \mathbb{I}}, \Theta_S\}$. We have now completely described the dynamics of the system. In the next section, we specify the costs associated to the system.

C. Costs generated by the system

The costs generated by the system are due to PM, CM and forced outages of the unit. In practice as PM are scheduled in advance, they are cheaper than unpredictable CM. A forced outage of the unit induces a loss of production. It is characterized by a daily cost which is higher than that of a PM or a CM. We consider a discount rate τ meaning that a cost c occurring at time t will be valued $\eta_t c$ with the discount factor $\eta_t = \frac{1}{(1+\tau)^t}$. We introduce the following notations:

- $j^P_{i,t}(u_{i,t})$ is the PM cost incurred at time t for component i. Let C^P_i be the cost of a PM operation on component i. The cost C^P_i covers the cost of the parts for the maintenance and the cost of engineering and maintenance teams labor. As the controls $u_{i,t}$ take values in the interval $[0,1]$, we must define a cost for these "partial" maintenance. We use a quadratic cost as it is strongly convex and should favor numerical convergence of the optimization algorithm:

$$j^P_{i,t}(u_{i,t}) = \eta_t C^P_i u^2_{i,t} \,.$$

In the case where $u_{i,t} = 0$ and $u_{i,t} = 1$, we get back to the real incurred cost. Note that in the case where $0 < u_{i,t} < \nu$, which models a situation where no PM is performed, we have $j^P_{i,t}(u_{i,t}) > 0$.

- $j^C_{i,t}(\boldsymbol{X}_{i,t})$ is the CM cost. It is due at the time of the failure of a component, even if there is no spare part to perform the operation immediately. Let C^C_i be the cost of a CM operation on component i. Similarly as for C^P_i, the cost C^C_i covers the cost of the parts and the cost of engineering and maintenance teams labor. We have

$$j^C_{i,t}(\boldsymbol{X}_{i,t}) = \begin{cases} \eta_t C^C_i, & \text{if component } i \text{ fails at } t \\ 0, & \text{otherwise} \,. \end{cases}$$

- $j^F_t(\boldsymbol{X}_{1:n,t})$ is the forced outage cost. The unit is in forced outage when at least one component is in a failed state and the CM has not occurred immediately because of a lack of spare part. Let C^F be the forced outage cost per time unit. We have

$$j^F_t(\boldsymbol{X}_{1:n,t}) = \begin{cases} \eta_t C^F, & \text{if forced outage} \\ 0, & \text{otherwise} \,. \end{cases}$$

Remark that this cost depends on the coupled state of all components.

In order to consider the costs over the whole study period we introduce:

- The *total maintenance cost* (preventive and corrective) generated by component i on the studied period:

$$j_i(\boldsymbol{X}_i, u_i) = \sum_{t=0}^{T} j^P_{i,t}(u_{i,t}) + j^C_{i,t}(\boldsymbol{X}_{i,t}) \,,$$

- The *total forced outage cost* generated by the system during the studied period:

$$j^F(\boldsymbol{X}_{1:n}) = \sum_{t=0}^{T} j^F_t(\boldsymbol{X}_{1:n,t})$$

No cost on the stock is considered in our model. Note also that the forced outage cost per time unit C^F is fixed but it is possible to change its value across time steps. If the time step used for the simulation is small enough, this feature enables us to take into account seasonality effects or a variation of the electricity price.

D. *Formulation of the maintenance optimization problem*

Recall that the dynamics of the system is stochastic as it depends on the failures of the components, modeled by the random vector $\boldsymbol{W}$. The cost function is then stochastic as well. The objective is to find the deterministic maintenance strategy $u \in \mathbb{U}$ that minimizes an expected cost over all failure scenarios:

$$\begin{aligned} \min_{(\boldsymbol{X},\boldsymbol{S},u)\in\mathcal{X}\times\mathcal{S}\times\mathbb{U}} \ & \mathbb{E}\left(\sum_{i=1}^{n} j_i(\boldsymbol{X}_i, u_i) + j^F(\boldsymbol{X}_{1:n})\right) \\ \text{s.t.} \quad & \Theta(\boldsymbol{X}, \boldsymbol{S}, u, \boldsymbol{W}) = 0 \end{aligned} \tag{1}$$

The total maintenance cost is additive with respect to the components meaning that:

$$\mathbb{E}\left(\sum_{i=1}^{n} j_i(\boldsymbol{X}_i, u_i)\right) = \sum_{i=1}^{n} \mathbb{E}\big(j_i(\boldsymbol{X}_i, u_i)\big) \,,$$

whereas the forced outage cost induces a non-additive coupling between the components. We will see that these two terms are treated in a different way for the design of a decomposition-coordination algorithm.

E. *Optimization with Mesh Adaptive Direct Search (MADS) and its limits*

This work is motivated by a real industrial case of maintenance of components of hydroelectric power plants. Problem (1) models a simplified version of the real problem. The number of components n in the system can be up to 80 and T is 40 years so $\mathbb{U}$ has dimension up to 3280 as we have one maintenance decision each year for each component (starting at year 0).

When the number of components is not too large, say $n < 10$, the maintenance problem (1) is solved efficiently by MADS [17]. MADS is an iterative blackbox optimization algorithm that evaluates the objective function at some points lying on a spatial discretization of the admissible space called the mesh. The mesh and the evaluation points are updated at each iteration so as to guarantee the convergence of the algorithm to a critical point. MADS has been designed for continuous optimization and uses the modeling of Section II. In particular, the PM strategies are modeled with a continuous decision variable. The real PM strategy is then obtained by projecting the output of the algorithm on $\{0,1\}$.

MADS is a blackbox algorithm. This means that evaluation points are chosen iteratively without the need for the gradients of the objective function. This feature is particularly appealing as the cost function is not differentiable.

In practice, the objective function is costly to evaluate as the expectation is estimated using Monte-Carlo simulations. When the number of components is large ($n \geq 10$), MADS needs more iterations to explore the high-dimensional space of solutions and the objective function takes more time to evaluate. The algorithm may not be able to find a very effective maintenance strategy. To overcome the difficulty of MADS when dealing with large systems, we use a decomposition of the original optimization problem component by component.

III. A DECOMPOSITION BY COMPONENT BASED ON THE AUXILIARY PROBLEM PRINCIPLE

In this section, we present a decomposition-coordination scheme for the optimal maintenance scheduling problem (1).

This scheme is based on the Auxiliary Problem Principle (APP). The APP has first been introduced in [16] as a unified framework for decomposition methods but also for other classical iterative algorithms. This principle casts the resolution of an optimization problem into the resolution of a sequence of auxiliary problems whose solutions converge to the solution of the original problem. The auxiliary problems are constructed to be decomposable. Thus, the resolution of each auxiliary problem of large dimension boils down to the resolution of independent subproblems of smaller size. Theoretical guarantees on the convergence of the method can be found in [18].

The main advantage of decomposition methods is that the resolution of the small subproblems is much faster than the resolution of the original problem. Decomposition methods are also naturally adapted to parallelization as each subproblem is independent.

The construction of the auxiliary problems and the APP fixed point algorithm is presented in the general case in [11]. In the subsequent sections, we apply this framework for the optimal maintenance scheduling problem (1).

A. *Construction of an auxiliary problem*

The construction of a decomposable auxiliary problem relies on a decomposition of the admissible space $\mathcal{X} \times \mathcal{S} \times \mathbb{U}$ and on a decomposition of the constraint space $\mathcal{L}$. Considering the physical nature of the industrial system composed of n components and a stock, we aim at decomposing the problem in $n+1$ subproblems, one for each component and one for the stock. Hence, we decompose the admissible space and the constraint space as follows:

$$\begin{cases} \mathcal{X} \times \mathcal{S} \times \mathbb{U} = (\mathcal{X}_1 \times \mathbb{U}_1) \times \ldots \times (\mathcal{X}_n \times \mathbb{U}_n) \times \mathcal{S} \\ \mathcal{L} = \mathcal{L}_1 \times \ldots \times \mathcal{L}_n \times \mathcal{L}_S \end{cases}$$

This decomposition is called *decomposition by component*. The subproblem on $\mathcal{X}_i \times \mathbb{U}_i$ is called *subproblem on component* i and the subproblem on $\mathcal{S}$ is called *subproblem on the stock.*

However, problem (1) is not directly decomposable by component because of couplings that we recall now:

- The maintenance cost $j^M = \sum_{i=1}^n j_i$ is additive with respect to the decomposition by component.
- The forced outage cost j^F induces a non-additive coupling between the components.
- The dynamics Θ_i of component i induces a coupling with the stock. The dynamics Θ_S of the stock is coupling the stock with all components.

To construct a decomposable auxiliary problem, we need to introduce an additive auxiliary function K and an block diagonal auxiliary dynamics Φ, see [11] for more details. Let $\overline{\boldsymbol{X}} \in \mathcal{X}$ and $\overline{\boldsymbol{S}} \in \mathcal{S}$, we introduce the auxiliary function $K(\boldsymbol{X}) = \sum_{i=1}^n K_i(\boldsymbol{X}_i)$ with:

$$K_i(\boldsymbol{X}_i) = j^F(\overline{\boldsymbol{X}}_{1:i-1}, \boldsymbol{X}_i, \overline{\boldsymbol{X}}_{i+1:n}) \,.$$

The auxiliary dynamics is defined as $\Phi(\boldsymbol{X}, \boldsymbol{S}, u, \boldsymbol{W}) = \big((\Phi_i(\boldsymbol{X}_i, u_i, \boldsymbol{W}_i))_{i \in \mathbb{I}}, \Phi_S(\boldsymbol{S})\big)$ with:

$$\begin{cases} \Phi_i(\boldsymbol{X}_i, u_i, \boldsymbol{W}_i) = \Theta_i(\boldsymbol{X}_i, \overline{\boldsymbol{S}}, u_i, \boldsymbol{W}_i), \qquad i \in \mathbb{I} \\ \Phi_S(\boldsymbol{S}) = \Theta_S(\overline{\boldsymbol{X}}_{1:n}, \boldsymbol{S}) \,, \end{cases}$$

Note that K and Φ are designed so that K_i and Φ_i (resp. Φ_S) only depend on variables of the space $\mathcal{X}_i \times \mathbb{U}_i$ (resp. $\mathcal{S}$) while the other variables are fixed to a value defined by $\overline{\boldsymbol{X}}$ and $\overline{\boldsymbol{S}}$. Let $\bar{u} \in \mathbb{U}, \overline{\boldsymbol{\Lambda}} \in \mathcal{L}$,[1] the auxiliary problem that results from this choice of K and Φ is:

$$\begin{aligned} \min_{(\boldsymbol{X},\boldsymbol{S},u) \in \mathcal{X}\times\mathcal{S}\times\mathbb{U}} \mathbb{E}\Bigg(& \sum_{i=1}^n \big(j_i(\boldsymbol{X}_i, u_i) + K_i(\boldsymbol{X}_i)\big) \\ & + \langle \overline{\boldsymbol{\Lambda}}, \big(\Theta'(\overline{\boldsymbol{X}}, \overline{\boldsymbol{S}}, \bar{u}, \boldsymbol{W}) \\ & - \Phi'(\overline{\boldsymbol{X}}, \overline{\boldsymbol{S}}, \bar{u}, \boldsymbol{W})\big) \cdot (\boldsymbol{X}, \boldsymbol{S}, u) \rangle \Bigg) \\ \text{s.t.} \quad & \Phi(\boldsymbol{X}, \boldsymbol{S}, u, \boldsymbol{W}) = 0 \,. \end{aligned} \tag{2}$$

The inner product in the auxiliary problem is a coordination term. It can be interpreted as a penalty, weighted by $\overline{\boldsymbol{\Lambda}}$, that quantifies the non-respect of the original dynamics when using the auxiliary dynamics. The idea of the APP is to iteratively solve the auxiliary problem (2), with an appropriate update of the values of $\overline{\boldsymbol{X}}, \overline{\boldsymbol{S}}, \bar{u}, \overline{\boldsymbol{\Lambda}}$ at each iteration so that the sequence of solutions of the auxiliary problems converges to the solution of the original problem (1).

By construction the auxiliary problem is decomposable by component. Solving (2) amounts to solve $n+1$ independent subproblems. The subproblem on component $i \in \mathbb{I}$ is:

$$\begin{aligned} \min_{(\boldsymbol{X}_i,u_i) \in \mathcal{X}_i \times \mathbb{U}_i} \mathbb{E}\big(& j_i(\boldsymbol{X}_i, u_i) + K_i(\boldsymbol{X}_i) \\ & + \langle \overline{\boldsymbol{\Lambda}}_S, \partial_{\boldsymbol{X}_i}\Theta_S(\overline{\boldsymbol{X}}_{1:n}, \overline{\boldsymbol{S}}) \cdot \boldsymbol{X}_i \rangle \big) \\ \text{s.t.} \quad & \Phi_i(\boldsymbol{X}_i, u_i, \boldsymbol{W}_i) = 0 \end{aligned}$$

The subproblem on the stock is:

$$\begin{aligned} \min_{\boldsymbol{S} \in \mathcal{S}} \mathbb{E} & \left(\sum_{i=1}^n \langle \overline{\boldsymbol{\Lambda}}_i, \partial_{\boldsymbol{S}}\Theta_i(\overline{\boldsymbol{X}}_i, \overline{\boldsymbol{S}}, \bar{u}_i, \boldsymbol{W}_i) \cdot \boldsymbol{S} \rangle \right) \\ \text{s.t.} \quad & \Phi_S(\boldsymbol{S}) = 0 \end{aligned}$$

The theory of the APP is based on the following fundamental theorem.

Theorem 1. *Suppose that the maintenance cost j^M and the forced outage cost j^F are convex and lower semi-continuous. Assume moreover that $j^M + j^F$ is coercive and that j^F is differentiable. Suppose also that Θ is linear and differentiable. Let $(\boldsymbol{X}^\#, \boldsymbol{S}^\#, u^\#)$ be a solution of the auxiliary problem* (2) *and $\boldsymbol{\Lambda}^\#$ be an optimal multiplier for its constraint. If $(\boldsymbol{X}^\#, \boldsymbol{S}^\#, u^\#, \boldsymbol{\Lambda}^\#) = (\overline{\boldsymbol{X}}, \overline{\boldsymbol{S}}, \bar{u}, \overline{\boldsymbol{\Lambda}})$, then $(\boldsymbol{X}^\#, \boldsymbol{S}^\#, u^\#)$ is a solution of the original problem* (1) *and $\boldsymbol{\Lambda}^\#$ is an optimal multiplier for its constraint.*

The proof consists in checking that if $(\boldsymbol{X}^\#, \boldsymbol{S}^\#, u^\#, \boldsymbol{\Lambda}^\#) = (\overline{\boldsymbol{X}}, \overline{\boldsymbol{S}}, \bar{u}, \overline{\boldsymbol{\Lambda}})$ then it solves the variational inequalities that are satisfied by an optimal solution and an optimal multiplier of the master problem (1). More details can be found in [11].

B. *The APP fixed-point algorithm*

Theorem 1 suggests to use a fixed-point algorithm to solve the original problem, leading to the APP fixed-point Algorithm 1. The maximum number of iterations is $M \in \mathbb{N}$.

[1] Formally, $\overline{\boldsymbol{\Lambda}}$ is an element of the dual cone of the cone of constraints $\{0\}_{\mathcal{L}} \subset \mathcal{L}$ where $\{0\}_{\mathcal{L}}$ denotes the cone which only contains the element $0 \in \mathcal{L}$. The dual cone of $\{0\}_{\mathcal{L}}$ is the cone $\mathcal{L}^\star$ where $\mathcal{L}^\star$ is the dual space of $\mathcal{L}$. As $\mathcal{L}$ is reflexive, we can identify $\mathcal{L}$ and $\mathcal{L}^\star$ so that $\overline{\boldsymbol{\Lambda}} \in \mathcal{L}$.

The subproblems on the components are solved with the blackbox algorithm MADS [17]. At iteration k, MADS only outputs a primal solution $(\boldsymbol{X}_i^{k+1}, u_i^{k+1})$ of subproblem $i \in \mathbb{I}$. The optimal multiplier $\boldsymbol{\Lambda}_i^{k+1}$ is computed afterwards using the adjoint state.

Algorithm 1 APP fixed-point algorithm

> Start with $(\bar{\mathbf{X}}, \bar{\mathbf{S}}, \bar{u}) = (\mathbf{X}^0, \mathbf{S}^0, u^0)$, $\bar{\mathbf{\Lambda}} = \mathbf{\Lambda}^0$
> **for** $k = 0, \dots, M-1$ **do**
>   **for** $i = 1, \dots, n$ **do**
>   Solve the subproblem on component i.
>   Let $(\mathbf{X}_i^{k+1}, u_i^{k+1})$ be a solution and $\mathbf{\Lambda}_i^{k+1}$ be an optimal multiplier.
>   **end for**
>   Solve the subproblem on the stock.
>   Let $\mathbf{S}^{k+1}$ be a solution and $\mathbf{\Lambda}_S^{k+1}$ be an optimal multiplier.
>   Set $(\bar{\mathbf{X}}, \bar{\mathbf{S}}, \bar{u}) = ((\mathbf{X}_1^{k+1}, \dots, \mathbf{X}_n^{k+1}), \mathbf{S}^{k+1}, (u_1^{k+1}, \dots, u_n^{k+1}))$
>   Set $(\bar{\mathbf{\Lambda}}_1, \dots, \bar{\mathbf{\Lambda}}_n, \bar{\mathbf{\Lambda}}_S) = (\mathbf{\Lambda}_1^{k+1}, \dots, \mathbf{\Lambda}_n^{k+1}, \mathbf{\Lambda}_S^{k+1})$
> **end for**
> **return** the maintenance strategy u^M

The subproblem on the stock is very easy to solve numerically. The constraint $\Phi_S(\boldsymbol{S}) = 0$ represents the dynamics of the stock with $\bar{\boldsymbol{X}} = (\bar{\boldsymbol{X}}_1, \dots, \bar{\boldsymbol{X}}_n)$ being fixed. The value of $\bar{\boldsymbol{X}}$ completely determines the dynamics of the stock. Hence, solving the subproblem on the stock just boils down to simulate its dynamics. The optimal multiplier is also computed using the adjoint state.

In practice, the assumptions of Theorem 1 are not satisfied as j^F is not continuous and therefore not differentiable. The dynamics Θ is neither linear nor differentiable. However, the decomposition method may still give good results in practice. In order to be able to compute the gradients that appear in the auxiliary problem, we use a continuous relaxation of the dynamics. A continuous relaxation of the cost function is also used as the gradients of the cost appear in the multiplier update step.

IV. Numerical results

In this part, we present the results of the decomposition methodology applied to the optimal maintenance scheduling problem. The expectation in (1) cannot be evaluated exactly, so we solve a Monte-Carlo approximation of the problem with $Q = 100$ fixed failure scenarios $\omega_1, \dots, \omega_Q \in \Omega$:

$$\min_{(\boldsymbol{X},\boldsymbol{S},u) \in \mathcal{X} \times \mathcal{S} \times \mathbb{U}} \frac{1}{Q} \sum_{q=1}^{Q} \sum_{i=1}^{n} j_i\big(\boldsymbol{X}_i(\omega_q), u_i\big) + j^F\left(\boldsymbol{X}_{1:n}(\omega_q)\right)$$

$$\text{s.t. } \Theta\big(\boldsymbol{X}(\omega_q), \boldsymbol{S}(\omega_q), u, \boldsymbol{W}(\omega_q)\big) = 0, \quad q \in \{1, \dots Q\}$$

The reference algorithm is the blackbox algorithm MADS applied directly on the original optimization problem. The maintenance strategies given by the two algorithms are then evaluated on a set of 10^5 failure scenarios, distinct from those used for the optimization. For the numerical experiments, we consider a system with the characteristics given in Table I. Parameters of the computation are given in Table II.

Remark 1. The APP fixed-point algorithm solves a decomposable auxiliary problem at each iteration, this algorithm is designed to be parallelized. It runs on 80 processors so that the subproblems on the components are solved in parallel. The reference algorithm MADS runs only on one processor. Note that it is also possible to parallelize MADS [17], although the implementation is not as immediate as for the decomposition method. The parallel version of MADS has not been tested.

TABLE I. Characteristics of the industrial system

Parameter	**Value**		
Number of components n	80		
Initial number of spare parts $\boldsymbol{S}_0$	16		
Horizon T	40 years		
Time step Δt	1 year		
Number of time steps for supply D	2		
Discount rate τ	0.08		
Maintenance threshold ν	0.9		
Yearly forced outage cost C^F	10000 k€/ year		
	Comp. 1	**Comp.** 2	**Comp.** $i \geq 3$
PM cost C^P	50 k€	50 k€	50 k€
CM cost C^C	100 k€	250 k€	200 k€
Failure distribution	Weib(2.3, 10)	Weib(4, 20)	Weib(3, 10)
Mean time to failure	8.85 years	18.13 years	8.93 years

TABLE II. Parameters of the computation

	Decomposition	MADS
Fixed-point iterations	50	/
Cost function calls	10^3/subproblem/iteration	8×10^5
Processor model	Intel® Xeon® Processor E5-2680 v4, 2.4 GHz	
Computation time	18h24min	22h30min

The output of the two algorithms is a maintenance strategy with $u_{i,t} \in [0,1]$ for $(i,t) \in \mathbb{I} \times \mathbb{T}$. From the operational perspective, a PM makes a component as good as new. Hence, for the evaluation of the strategy, the controls are projected on $\{0,1\}$: we consider that if $u_{i,t} \geq \nu$, then the PM makes the component as good as new, otherwise no PM is performed. The comparison between the two maintenance strategies is fair as we use the same procedure for their evaluation.

The mean cost of the best solution is 12902 k€ with MADS and 11483 k€ with the decomposition which represents a gain of 11%. The values of some quantiles are gathered in Table III. Fig. 3 represents the distribution of the cost. Fig. 4 outlines that the average CM cost is higher with the decomposition strategy. However, a much lower PM cost makes the decomposition more efficient than MADS. This is due to the fact that fewer PM are performed with the decomposition strategy than with MADS strategy (Table IV). The counterpart is that failures and forced outages occur more often with the decomposition strategy (Table IV). The forced outage cost is not visible on Fig. 4 as it represents only 0.05 k€ for MADS strategy and 4.09 k€ for the decomposition strategy, showing that both methods give efficient PM strategies. There are more forced outages with the decomposition strategy (63 occurrences in 10^5 failure scenarios versus 1 for MADS) but they almost all occur in the last two time steps of the study horizon. Therefore the cost of forced outages is low because of the discount factor.

TABLE III. Quantiles of the cost of the two maintenance strategies (k€)

	1%	5%	25%	50%	75%	95%	99%
Decomposition	10385	10683	11136	11472	11809	12326	12713
MADS	11858	12151	12588	12894	13211	13674	14010

The cumulative number of PM is shown on Fig. 5. As already noticed there are fewer PM with the decomposition strategy. A striking feature with the decomposition strategy is that there are almost no PM in the first three years. This exploits the fact that the components are new. The reference algorithm MADS applied directly on the original problem does not detect this feature. In fact, the region of the space corresponding to not doing any PM in the first three years jointly for all components is a very small subset of the admissible space of the original problem and is not explored by MADS. On the other hand, the subproblems in the APP fixed-point algorithm act on an individual component, it is then easier to figure out that doing no PM in the first three years is profitable.

TABLE IV. OVERVIEW OF THE NUMBER OF PM, FAILURES AND FORCED OUTAGES FOR EACH STRATEGY

	Decomposition	MADS
Total nb. of PM	447	558
Mean nb. of PM/component	5.6	7.0
Mean time between PM	6.1 years	5.0 years
Mean nb. of failures/component	1.40	1.18
Nb. of forced outages/Nb. of scenarios	63/10000	1/10000

There is also a significant reduction of the number of PM in the last five years of the study horizon. It is indeed useless to invest money to repair a component for the last few years. Moreover, the discount factor at the end of the horizon greatly reduces the incurred cost so that a forced outage is not too penalizing. This is why some forced outages occur with the decomposition strategy at the end of the study period.

Another indicator that is monitored by decision makers is the level of stock. A necessary condition for the occurrence of a forced outage is that the stock is empty. Hence, we look at the probability of having an empty stock. The higher this

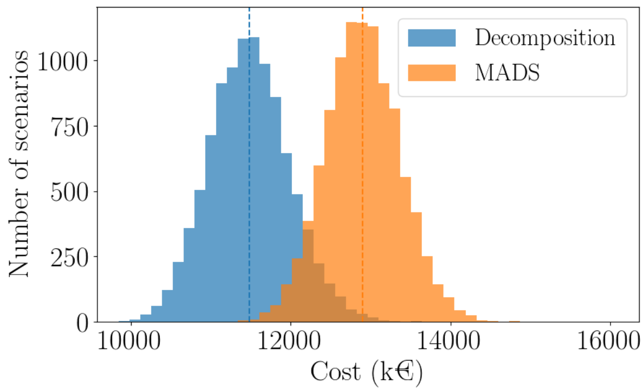

Fig. 3: Distribution of the cost for the two maintenance strategies

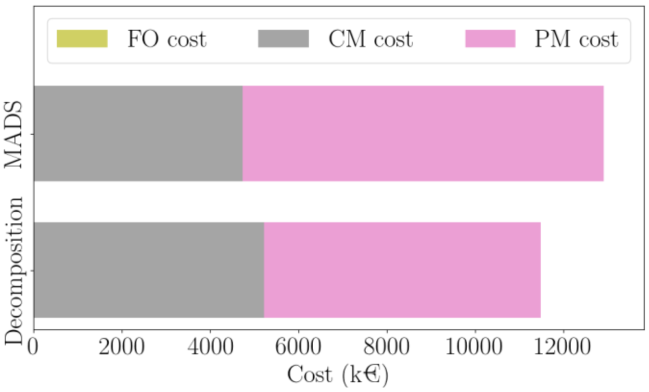

Fig. 4: Part of the PM, CM and forced outage cost in the total expected cost

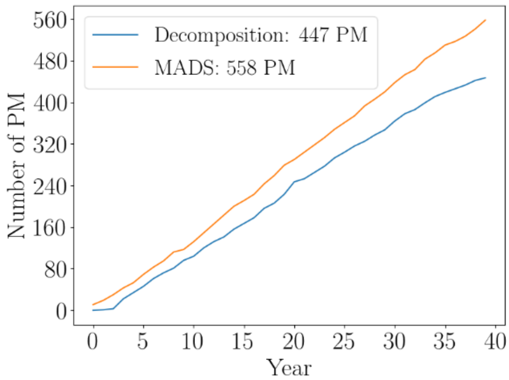

Fig. 5: Cumulative number of PM

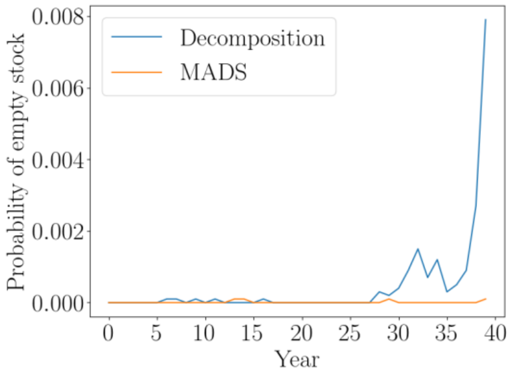

Fig. 6: Evolution of the probability of having an empty stock

probability, the higher the probability of forced outage. The probability of having an empty stock is very low for both strategies in the first 30 years and then increases for the decomposition strategy (Fig. 6). Again, because of the discount factor, forced outages in the last few years do not have important financial consequences. It is then more profitable to do fewer PM and allow for a higher risk of failure. This is what the decomposition strategy does.

Overall the strategy obtained by decomposition is more cost effective than MADS strategy. For a decision maker the decomposition strategy requires less investment as we do fewer PM. It also has the best expected cost. Even in the case of extreme events, it is more robust than MADS strategy, as shown by the 99% quantile in Table III. Indeed the forced outages may occur only at the end of the horizon.

V. CONCLUDING REMARKS AND OUTLOOK

In this work we study a maintenance scheduling optimization problem for hydropower plants management but the methodology that has been presented can be applied to any kind of physical assets as long as the failure distribution and the dynamics of the system are known. We set up a decomposition method to find a deterministic preventive maintenance strategy for a system of physical components sharing a common stock of spare parts. The decomposition relies on the Auxiliary Problem Principle. We construct a sequence of auxiliary problems that are solved iteratively. The auxiliary problems are decomposable into independent subproblems of smaller dimension that are solved in parallel. Each subproblem involves only one component of the system or the stock.

The main advantage of the decomposition method against classical optimization methods is that it is scalable to problems of arbitrary size as it is designed to be parallelized. It is then theoretically possible to jointly optimize the maintenance scheduling of a whole fleet of dependent components at the price of some modeling effort. From the implementation perspective, the most challenging part is the tuning of the relaxation of the system in order to compute the gradients of the dynamics that appear in the auxiliary problem. The performance of the method is sensitive to this tuning that must be done carefully.

On the industrial system, the decomposition method outperforms the blackbox algorithm MADS applied directly on the full problem. The mean cost of the best solution is 12902 k€ with MADS and 11483 k€ with the decomposition meaning that our new approach generates a gain of 11% in the total Life Cycle Cost when compared with off the shelf methods. Moreover, if the gain in Life Cycle Cost is important, the decrease by 20% of the number of preventive actions, directly linked to scheduled investments budget, will be highly appealing to decision makers seeking to optimize and prioritize Capex.

The strategy given by the decomposition involves fewer PM especially at the beginning and the end of the time horizon. More forced outages occur but only at the end of the time horizon so without heavy financial consequences. It is also robust to extreme events as the 99% quantile is better for the decomposition strategy than for MADS. This work proves the interest of the modeling effort needed to apply the decomposition method.

However, these results raise the question about the relevance of the discount factor in the modeling when studying systems on a long term horizon. In the model, failures at the end of the horizon only incur very low cost and it is more profitable to let the system fail. But is it really profitable in practice? Some interesting thoughts on this point can be found in [19].

Moreover, some challenges still remain for an application of this decomposition method in an operational context. Here, the dynamics is simulated with a time step of one year. This means that the state of a component cannot change within a year and we have to estimate a yearly cost of forced outage, resulting in an inaccurate evaluation of the real costs. A smaller time step must be used for the simulation of the dynamics in order to get an accurate evaluation of the costs. Moreover, smaller time steps enable us to take into account seasonality effects on the components by varying the forced outage cost for example. It should be noted that smaller time steps for the simulation of the dynamics will increase the computation time for the evaluation of the cost function in the maintenance optimization problem. However, the complexity of the problem stays the same as maintenance decisions are always made on a yearly basis, so the space of admissible maintenance strategies does not change. In other words, the number of iterations to solve the problem should stay the same but each one of them is longer.

On the other hand, to speed up the resolution, we can consider doing PM only on a two-year basis or even focusing only on periodic strategies. This reduces the complexity of the problem. Preliminary results show that the performance of periodic strategies is close to those presented in this paper, showing that if computation time is limited, looking for periodic strategies is a good alternative.

Finally, it is also possible to model more complex systems, by adding a control on the time of the order of spare parts or dependence between the failures of the components for instance. We could also consider imperfect preventive maintenance. However, a balance must be found between the simplicity of the model and its adequation to reality given the industrial application in mind.

References

[1] A. Froger, M. Gendreau, J. E. Mendoza, E. Pinson, and L.-M. Rousseau, “Maintenance Scheduling in the Electricity Industry: A Literature Review,” *Eur. J. Oper. Res.*, vol. 251, no. 3, pp. 695–706, 2016.

[2] S.-H. Ding and S. Kamaruddin, “Maintenance policy optimization—literature review and directions,” *Int. J. Adv. Manuf. Technol.*, vol. 76, no. 5–8, pp. 1263–1283, 2015.

[3] M. Sánchez-Silva, D. M. Frangopol, J. Padgett, and M. Soliman, “Maintenance and Operation of Infrastructure Systems: Review,” *J. Struct. Eng.*, vol. 142, no. 9, p. F4016004, 2016.

[4] M. C. A. Olde Keizer, S. D. P. Flapper, and R. H. Teunter, “Condition-based maintenance policies for systems with multiple dependent components: A review,” *Eur. J. Oper. Res.*, vol. 261, no. 2, pp. 405–420, 2017.

[5] O. Roux, M. A. Jamali, D. A. Kadi, and E. Châtelet, “Development of simulation and optimization platform to analyse maintenance policies performances for manufacturing systems,” *Int. J. Comput. Integr. Manuf.*, vol. 21, no. 4, pp. 407–414, 2008.

[6] R. Sarker and A. Haque, “Optimization of maintenance and spare provisioning policy using simulation,” *Appl. Math. Model.*, vol. 24, no. 10, pp. 751–760, 2000.

[7] M.-C. Chen, C.-M. Hsu, and S.-W. Chen, “Optimizing joint maintenance and stock provisioning policy for a multi-echelon spare parts logistics network,” *J. Chin. Inst. Ind. Eng.*, vol. 23, no. 4, pp. 289–302, 2006.

[8] N. Rezg, X. Xie, and Y. Mati, “Joint optimization of preventive maintenance and inventory control in a production line using simulation,” *Int. J. Prod. Res.*, vol. 42, no. 10, pp. 2029–2046, 2004.

[9] R. Lusby, L. F. Muller, and B. Petersen, “A solution approach based on Benders decomposition for the preventive maintenance scheduling problem of a stochastic large-scale energy system,” *J. Sched.*, vol. 16, no. 6, pp. 605–628, 2013.

[10] A. Grigoriev, J. van de Klundert, and F. C. R. Spieksma, “Modeling and solving the periodic maintenance problem,” *Eur. J. Oper. Res.*, vol. 172, no. 3, pp. 783–797, 2006.

[11] P. Carpentier and G. Cohen, *Décomposition-coordination en optimisation déterministe et stochastique*, vol. 81. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017.

[12] K. J. Arrow and L. Hurwicz, *Decentralization and Computation in Resource Allocation*. Stanford University, Department of Economics, 1960.

[13] L. S. Lasdon and J. D. Schoeffler, “A multi-level technique for optimization,” *Jt. Autom. Control Conf.*, vol. 3, pp. 85--92, 1965.

[14] M. D. Mesarović, D. Macko, and Y. Takahara, *Theory of Hierarchical, Multilevel Systems*, vol. 68. Academic Press, 1970.

[15] Y. Takahara, “Multilevel Approach to Dynamic Optimization,” Case Western Reserve University, Systems Research Center, SRC-50-C-64-18, 1964.

[16] G. Cohen, “Optimization by decomposition and coordination: A unified approach,” *IEEE Trans. Autom. Control*, vol. 23, no. 2, pp. 222–232, 1978.

[17] C. Audet and J. E. Dennis Jr, “Mesh Adaptive Direct Search Algorithms for Constrained Optimization,” *SIAM J. Optim.*, vol. 17, no. 1, pp. 188–217, 2006.

[18] G. Cohen, “Auxiliary problem principle and decomposition of optimization problems,” *J. Optim. Theory Appl.*, vol. 32, no. 3, pp. 277–305, 1980.

[19] I. Ekeland, “Discounting the Future: the Case of Climate Change,” *Cah. Chaire Finance Dév. Durable Univ. Paris-Dauphine*, vol. 27, p. 31, 2009.