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/TAC.2015.2504867, IEEE Transactions on Automatic Control

Approximation Algorithms for Optimization of Combinatorial Dynamical Systems

Insoon Yang, Samuel A. Burden, Ram Rajagopal, S. Shankar Sastry, and Claire J. Tomlin

***Abstract*—This paper considers an optimization problem for a dynamical system whose evolution depends on a collection of binary decision variables. We develop scalable approximation algorithms with provable suboptimality bounds to provide computationally tractable solution methods even when the dimension of the system and the number of the binary variables are large. The proposed method employs a linear approximation of the objective function such that the approximate problem is defined over the feasible space of the binary decision variables, which is a discrete set. To define such a linear approximation, we propose two different variation methods: one uses continuous relaxation of the discrete space and the other uses convex combinations of the vector field and running payoff. The approximate problem is a 0–1 linear program, which can be exactly or approximately solved by existing polynomial-time algorithms with suboptimality bounds, and does not require the solution of the dynamical system. Furthermore, we characterize a sufficient condition ensuring the approximate solution has a provable suboptimality bound. We show that this condition can be interpreted as the concavity of the objective function or that of a reformulated objective function.**

I. Introduction

The dynamics of critical infrastructures and their system elements—for instance, electric grid infrastructure and their electric load elements—are *interdependent*, meaning that the state of each infrastructure or its system elements influences and is influenced by the state of the others [1]. For example, consider the placement of power electronic actuators, such as high-voltage direct current links, on transmission networks. Such placement requires consideration of the interconnected swing dynamics of transmission grid infrastructures. Furthermore, the ON/OFF control of a large population of electric loads whose system dynamics are coupled with each other, e.g., supermarket refrigeration systems, must take into account their system-system interdependency. These decision-making problems under dynamic interdependencies combine the combinatorial optimization problems of actuator placement and ON/OFF control with the time evolution of continuous system states. Therefore, we seek decision-making techniques that unify combinatorial optimization and dynamical systems theory.

This paper examines a fundamental problem that supports such combinatorial decision-making involving dynamical systems. Specifically, we consider an optimization problem associated with a dynamical system whose state evolution depends on binary decision variables, which we call the *combinatorial dynamical system*. In our problem formulation, the binary decision variables do not change over time, unlike in the optimal control or predictive control of switched systems [2], [3], [4], [5]. Our focus is to develop scalable methods for optimizing the binary variables associated with a dynamical system when the number of the variables is too large to enumerate all possible system 'modes' and when the dimension of the system state is large. However, the optimization problem for combinatorial dynamical system presents a computational challenge because: (i) it is a 0–1 nonlinear program, which is generally NP-hard [6]; and (ii) it requires the solution of a system of ordinary differential equations (ODEs). To provide a computationally tractable solution method that can address large-scale problems, we propose scalable approximation algorithms with provable suboptimality bounds.

The key idea of the proposed methods is to linearize the objective function in the feasible space of binary decision variables. Our first contribution is to propose a linear approximation method for nonlinear optimization of combinatorial dynamical systems. The approximate 0–1 optimization can be efficiently solved because it is a linear 0–1 program and it does not require the solution of the dynamical system. The proposed approximation method allows us to employ polynomial-time exact or approximation algorithms including those for problems with l_0-norm constraints or linear inequality constraints. In particular, the proposed algorithms for an l_0-norm constrained problem are *single-pass*, i.e. they do not require multiple iterations, and are consequently more efficient than applicable greedy algorithms.

The proposed linear approximation approach requires the *derivative* of the objective function, but this is nontrivial to construct because the function's domain is a discrete space, in general. The second contribution of this work is to propose two different derivative concepts. The first concept uses a natural relaxation of the discrete space, whereas for the second concept a novel relaxation method in a function space using convex combinations of the vector fields and running payoffs is developed. We refer to the former construction as the *standard derivative* because it is the same as the derivative concept in continuous space, and the latter as the *nonstandard derivative*. We show the existence and the uniqueness of the nonstandard derivative, and provide an adjoint-based formula for it. The nonstandard derivative is well-defined even when the vector field and the payoff function are undefined on interpolated values of the binary decision variables. Because the two derivatives are different in general, we can solve two instances of the approximate problem (if the problem is well-defined on intermediate values in addition to a 0–1 lattice), one with the standard derivative and another with the nonstandard derivative and then choose the better solution *a posteriori*. If the problem is defined only on a 0–1 lattice, we can utilize the nonstandard derivative.

The third contribution of this paper is to characterize conditions under which the proposed algorithms have provable suboptimality bounds. We show that the concavity of the original problem gives a sufficient condition for the suboptimality bound to hold if the approximation is performed using the standard derivative. On the other hand, the same concavity condition does not suffice when the nonstandard derivative is employed in the approximation. To resolve this difficulty, we propose a reformulated problem and show that its concavity guarantees a suboptimality bound to hold.

In operations research, 0–1 nonlinear optimization problems have been extensively studied over the past five decades, although the problems are not generally associated with dynamical systems. In particular, 0–1 polynomial programming, in which the objective function and the constraints are polynomials in the decision variables, has attracted great attention. Several exact methods that can transform a 0–1 polynomial program into a 0–1 linear program have been developed by introducing new variables that represent the cross terms in the polynomials (e.g., [7]). Roof duality suggests approximation methods for 0–1 polynomial programs [8]. It constructs the best linear function that upperbounds the objective function (in the case of maximization) by solving a dual problem. Its size can be significantly

This work was supported by the NSF CPS project ActionWebs under grant number 0931843, NSF CPS project FORCES under grant number 1239166.

I. Yang is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, insoon@mit.edu

S. A. Burden is with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195, USA, sburden@uw.edu

R. Rajagopal is with the Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94035, USA, ramr@stanford.edu

S. S. Sastry and C. J. Tomlin are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA, {sastry, tomlin}@eecs.berkeley.edu

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/TAC.2015.2504867, IEEE Transactions on Automatic Control

bigger than that of the primal problem because it introduces $O(m^k)$ additional variables, where m and k denote the number of binary variables and the degree of polynomial, respectively. This approach is relevant to our proposed method in the sense that both methods seek a linear function that bounds the objective function. However, the proposed method explicitly constructs such a linear function without solving any dual problems. Furthermore, whereas all the aforementioned methods assume that the objective function is a polynomial in the decision variables, our method does not require a polynomial representation of the objective function. This is a considerable advantage because constructing a polynomial representation of a given function, $J : \{0,1\}^m \to \mathbb{R}$, generally requires 2^m function evaluations (e.g., via multi-linear extension [9]). Even when the polynomial representations of the vector field and the objective function in the decision variables, $\alpha \in \{0,1\}^m$, are given, a polynomial representation of the objective function in α is not readily available because the state of a dynamical system is not, in general, a polynomial in α with a finite degree. For more general 0–1 nonlinear programs, branch-and-bound methods (e.g., [10]) and penalty/smoothing methods (e.g., [11]) have been suggested. However, the branch-and-bound methods cannot find a solution in polynomial time in general. The penalty and smoothing methods do not generally provide any performance guarantee although they perform well in many data sets.

An important class of 0–1 nonlinear programs is the minimization or the maximization of a submodular set-function, which has the property of *diminishing returns*. Unconstrained submodular function minimization can be solved in polynomial time using a convex extension (e.g., [12]) or a combinatorial algorithm (e.g., [13], [14]). However, constrained submodular function minimization is NP-hard in general, and approximation algorithms with performance guarantees are available only in special cases (e.g., [15]). On the other hand, our proposed method can handle a large class of linear constraints with a guaranteed suboptimality bound. In the case of submodular function maximization, a standard greedy algorithm can obtain a provably near-optimal solution [16]. Our algorithm for l_0-norm constrained problems has, in general, lower computational complexity than the greedy algorithm.

The rest of this paper is organized as follows. The problem setting for the optimization of combinatorial dynamical systems is specified in Section II. In Section III, a linear approximation approach for this problem is proposed. Furthermore, we provide a condition under which the proposed approximate problem gives a solution with a guaranteed suboptimality bound and show that the condition can be interpreted as the concavity of the objective function or that of a reformulated objective function. In Section IV, algorithms to solve the approximate problems with several types of linear inequality constraints are suggested.

II. Problem Setting

Consider the following dynamical system in the continuous state space $\mathcal{X} \subseteq \mathbb{R}^n$:

$$\dot{x}(t) = f(x(t), \mathbf{a}), \quad x(0) = \mathbf{x} \in \mathcal{X}, \tag{1}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$. Let $x^{\mathbf{a}} := (x_1^{\mathbf{a}}, \cdots, x_n^{\mathbf{a}})$ denote the solution of the ordinary differential equation (ODE) (1) given $\mathbf{a} \in \mathbb{R}^m$. We call (1) a *combinatorial dynamical system* when $\mathbf{a}$ is chosen as an m-dimensional binary vector variable $\alpha := \{\alpha_1, \cdots, \alpha_m\} \in \{0,1\}^m$. We later view α as a decision variable that does not change over time in a given time interval $[0,T]$. We consider the following assumptions on the vector field.

Assumption 1. *For each $\alpha \in \{0,1\}^m$, $f(\cdot, \alpha) : \mathbb{R}^n \to \mathbb{R}^n$ is twice differentiable, has a continuous second derivative and is globally Lipschitz continuous in $\mathcal{X}$.*

Assumption 2. *For any $\boldsymbol{x} \in \mathcal{X}$, $f(\boldsymbol{x}, \cdot) : \mathbb{R}^m \to \mathbb{R}^n$ is continuously differentiable in $[0,1]^m$.*

Under Assumption 1, the solution of (1) satisfies the following property (Proposition 5.6.5 in [17]): for any $\alpha \in \{0,1\}^m$, $\|x^\alpha\|_2 := \left(\int_0^T \|x^\alpha(t)\|^2 dt\right)^{\frac{1}{2}} < \infty$. In other words, $x^\alpha : [0,T] \to \mathbb{R}^n$ is such that $x^\alpha \in L^2([0,T]; \mathbb{R}^n)$. Furthermore, Assumption 1 guarantees that the system admits a unique solution, which is continuous in time, for each $\alpha \in \{0,1\}^m$.

Our aim is to determine the binary vector $\alpha \in \{0,1\}^m$ that maximizes a payoff (or utility) function, $J : \mathbb{R}^m \to \mathbb{R}$, associated with the dynamical system (1). More specifically, we want to solve the following combinatorial optimization problem:

$$\max_{\alpha \in \{0,1\}^m} \quad J(\alpha) := \int_0^T r(x^\alpha(t), \alpha) dt + q(x^\alpha(T)) \tag{2a}$$

$$\text{subject to} \quad \mathbf{A}\alpha \leq \mathbf{b}, \tag{2b}$$

where x^α is the solution of (1) and $r : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $q : \mathbb{R}^n \to \mathbb{R}$ are running and terminal payoff functions, respectively. Here, $\mathbf{A}$ is an $l \times m$ matrix, $\mathbf{b}$ is an l-dimensional vector and the inequality constraint (2b) holds entry-wise. Note that the objective function J and the solution α^{OPT} of the optimization problem depend on the initial value $\mathbf{x}$ of the dynamical system. For notational simplicity, we suppress the dependency, i.e., $J(\alpha) = J(\alpha, \mathbf{x})$ and $\alpha^{\text{OPT}} = \alpha^{\text{OPT}}(\mathbf{x})$.

This optimization problem, in general, presents a computational challenge because (i) it is NP-hard; and (ii) it requires the solution to the system of ODEs (1). Therefore, we seek a scalable approximation method that gives a suboptimal solution with a provable suboptimality bound. The key idea of our proposed method is to take a first-order linear approximation of the objective function (2a) with respect to the binary decision variable α. This linear approximation should also take into account the dependency of the state on the binary decision variable. If the payoff function in (2a) is replaced with its linear approximation, which is linear in the decision variable, the approximate problem is a 0–1 linear program. Therefore, existing polynomial-time exact or approximation algorithms for 0–1 linear programs can be employed, as shown in Section IV. To obtain the linear approximations of the payoff function J, in the following section we formulate two different derivatives of J with respect to the discrete decision variable. Furthermore, we suggest a sufficient condition under which the approximate solution has a guaranteed suboptimality bound in Section III-C.

III. Linear Approximation for Optimization of Combinatorial Dynamical Systems

Suppose for a moment that the derivative of the objective function with respect to the binary decision variable is given, and that the derivative is well-defined in $\{0,1\}^m$, which is the feasible space of the decision variable. The derivative can be used to obtain the first-order linear approximation of the objective function, i.e., for $\alpha \in \{0,1\}^m$, $J(\alpha) \approx J(\bar{\alpha}) + DJ(\bar{\alpha})^\top(\alpha - \bar{\alpha})$. If the objective function (2a) is substituted with the right-hand side of the Taylor expansion, then we obtain the following approximate problem:

$$\max_{\alpha \in \{0,1\}^m} \quad DJ(\bar{\alpha})^\top \alpha \tag{3a}$$

$$\text{subject to} \quad \mathbf{A}\alpha \leq \mathbf{b}. \tag{3b}$$

This approximate problem is a 0–1 linear program, which can be solved by several polynomial-time exact or approximation algorithms

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/TAC.2015.2504867, IEEE Transactions on Automatic Control

(see Section IV). We characterize a bound on the suboptimality of the approximate solution in Section III-C.

We propose two different variation approaches for defining the derivatives in the discrete space $\{0,1\}^m$. The first uses the variation of the binary decision variable in a relaxed continuous space; the second uses the variation of the vector field of dynamical systems. The first and second concepts of the derivatives are called the *standard* and *nonstandard* derivatives, respectively. It is advantageous to have two different derivative concepts: we solve the approximate problem (3) twice, one with the standard derivative $D^{\text{S}}J$ and another with the nonstandard derivative $D^{\text{NS}}J$ and then choose the better solution. The one of two approximate solutions that outperforms another is problem-dependent, in general.

Remark 1. *As we will see in the following subsection, the nonstandard derivative requires less restrictive assumptions than the standard derivative. One important distinction is that the nonstandard derivative can be well-defined even when the problem is only defined on the* 0–1 *lattice* $\{0,1\}^m$*, i.e.,* $f : \mathbb{R}^n \times \{0,1\}^m \to \mathbb{R}^n$ *and* $J : \{0,1\}^m \to \mathbb{R}$ *(i.e.,* $r : \mathbb{R}^n \times \{0,1\}^m \to \mathbb{R}$*). However, we can use the standard derivative only if the problem is well-defined on the relaxed space* $[0,1]^m$.

A. Standard and Nonstandard Derivatives

We first define the derivative of the payoff function, J, with respect to discrete variation of the decision variable by relaxing the discrete space $\{0,1\}^m$ into the continuous space $\mathbb{R}^m$. This definition of derivatives in discrete space is exactly the same as the standard definition of derivatives in continuous space. Therefore, it requires the differentiability of the vector field and the running payoff with respect to α.

Assumption 3. *The functions* $r(\,\cdot\,,\alpha) : \mathbb{R}^n \to \mathbb{R}$ *and* $q : \mathbb{R}^n \to \mathbb{R}$ *are continuously differentiable for any* $\alpha \in \{0,1\}^m$.

Assumption 4. *For each* $\boldsymbol{x} \in \mathcal{X}$, $r(\boldsymbol{x},\cdot) : \mathbb{R}^m \to \mathbb{R}^n$ *is continuously differentiable in* $[0,1]^m$.

More precisely, Assumptions 3 and 4 are needed for the standard derivative while the nonstandard derivative does not require Assumption 4. Throughout this paper, we let $\mathbf{1}_i$ denote the m-dimensional vector whose ith entry is one and all other entries are zero. For notational convenience, we introduce a functional, $\mathcal{J} : L^2([0,T];\mathbb{R}^n) \times \mathbb{R}^m \to \mathbb{R}$, defined as

$$\mathcal{J}(z,\beta) := \int_0^T r(z(t),\beta)dt + q(z(T)). \tag{4}$$

Note that $J(\alpha) = \mathcal{J}(x^\alpha,\alpha)$, where x^α is defined as the solution to the ODE (1) with α.

Definition 1. *Suppose that Assumptions 1, 2, 3 and 4 hold. The standard derivative* $D^{S}J : \{0,1\}^m \to \mathbb{R}^m$ *of the payoff function* J *in* (2a) *at* $\bar\alpha \in \{0,1\}^m$ *is defined as*

$$[D^{S}J(\bar\alpha)]_i := \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\mathcal{J}(x^{\bar\alpha+\epsilon\mathbf{1}_i},\bar\alpha+\epsilon\mathbf{1}_i) - \mathcal{J}(x^{\bar\alpha},\bar\alpha)\right]$$

for $i = 1,\cdots,m$*, where the functional* $\mathcal{J} : L^2([0,T];\mathbb{R}^n)\times\mathbb{R}^m \to \mathbb{R}$ *is defined in* (4) *and* $x^{\bar\alpha}$ *is the solution of* (1) *with* $\bar\alpha$.

The standard derivative can be computed by direct and adjoint-based methods [18], [17]. We summarize the adjoint-based method in the following proposition.

Proposition 1. *Suppose that Assumptions 1, 2, 3 and 4 hold. The derivative in Definition 1 can be obtained as*

$$D^{S}J(\bar\alpha) = \int_0^T \left(\frac{\partial f(x^{\bar\alpha}(t),\bar\alpha)}{\partial\boldsymbol{\alpha}}^\top \lambda^{\bar\alpha}(t) + \frac{\partial r(x^{\bar\alpha}(t),\bar\alpha)}{\partial\boldsymbol{\alpha}}^\top\right)dt,$$

where $x^{\bar\alpha}$ *is the solution of* (1) *with* $\bar\alpha$ *and* $\lambda^{\bar\alpha}$ *solves the following adjoint system:*

$$-\dot\lambda^{\bar\alpha}(t) = \frac{\partial H(x^{\bar\alpha}(t),\lambda^{\bar\alpha}(t),\bar\alpha)}{\partial\boldsymbol{x}}^\top, \quad \lambda^{\bar\alpha}(T) = \frac{\partial q(x^{\bar\alpha}(T))}{\partial\boldsymbol{x}}^\top \tag{5}$$

with the Hamiltonian $H : \mathbb{R}^n\times\mathbb{R}^n\times\{0,1\}^m \to \mathbb{R}$,

$$H(\boldsymbol{x},\boldsymbol{\lambda},\alpha) := \boldsymbol{\lambda}^\top f(\boldsymbol{x},\alpha) + r(\boldsymbol{x},\alpha).$$

We now define the derivative of the payoff function using variations in vector fields and running payoffs. This nonstandard definition of derivatives does not require Assumptions 2 and 4, i.e., the differentiability of the vector field and the running payoff with respect to α. Furthermore, the nonstandard derivative is well-defined even when the vector field and the payoff function are not defined on the interpolated values of the binary decision variable, i.e., $f(\,\cdot\,,\alpha)$ and $r(\,\cdot\,,\alpha)$ are defined only at $\alpha \in \{0,1\}^m$. This is a practical advantage of the nonstandard derivative over the standard derivative. The proposed variation procedure for the nonstandard derivative is as follows.

(i) The 0–1 vector variable $\bar\alpha$ in the discrete space $\{0,1\}^m$ is mapped to $x^{\bar\alpha}$ in the continuous metric space $L^2([0,T];\mathbb{R}^n)$ via the original dynamical system (1);

(ii) In $L^2([0,T];\mathbb{R}^n)$, we construct a new state $x^{\epsilon(\bar\alpha,\alpha)}$ as the solution to the ϵ-*variational system* associated with $(\bar\alpha,\alpha)$ for $\epsilon\in[0,1]$,

$$\dot x(t) = f^{\epsilon(\bar\alpha,\alpha)}(x(t)), \quad x(0) = \mathbf{x}\in\mathcal{X}, \tag{6}$$

where the new vector field is obtained as the convex combination of the two vector fields with $\bar\alpha$ and α, i.e.,

$$f^{\epsilon(\bar\alpha,\alpha)}(\,\cdot\,) := (1-\epsilon)f(\,\cdot\,,\bar\alpha) + \epsilon f(\,\cdot\,,\alpha).$$

Set the distance between α and its ϵ-variation $\epsilon(\bar\alpha,\alpha)$ as ϵ; and

(iii) The nonstandard derivative of J is defined in the following:

Definition 2. *Suppose that Assumptions 1 and 3 hold. We define the nonstandard derivative* $D^{NS}J : \{0,1\}^m \to \mathbb{R}^m$ *of the payoff function* J *at* $\bar\alpha\in\{0,1\}^m$ *as*

$$[D^{NS}J(\bar\alpha)]_i := \begin{cases} \lim_{\epsilon\to0^+}\frac{1}{\epsilon}\left[\mathcal{J}^{\epsilon(\bar\alpha,\bar\alpha+\mathbf{1}_i)}(x^{\epsilon(\bar\alpha,\bar\alpha+\mathbf{1}_i)}) - \mathcal{J}(x^{\bar\alpha},\bar\alpha)\right] & \text{if } \bar\alpha_i = 0\\ \lim_{\epsilon\to0^+}\frac{1}{\epsilon}\left[\mathcal{J}(x^{\bar\alpha},\bar\alpha) - \mathcal{J}^{\epsilon(\bar\alpha,\bar\alpha-\mathbf{1}_i)}(x^{\epsilon(\bar\alpha,\bar\alpha-\mathbf{1}_i)})\right] & \text{if } \bar\alpha_i = 1,\end{cases}$$

where $\mathcal{J} : L^2([0,T];\mathbb{R}^n)\times\mathbb{R}^m\to\mathbb{R}$ *is given by* (4) *and* $\mathcal{J}^{\epsilon(\bar\alpha,\alpha)} : L^2([0,T];\mathbb{R}^n)\to\mathbb{R}$ *is defined as*

$$\mathcal{J}^{\epsilon(\bar\alpha,\alpha)}(\,\cdot\,) := (1-\epsilon)\mathcal{J}(\,\cdot\,,\bar\alpha) + \epsilon\mathcal{J}(\,\cdot\,,\alpha).$$

Here, $x^{\bar\alpha}$ *is the solution of* (1) *with* $\bar\alpha$ *and* $x^{\epsilon(\bar\alpha,\alpha)}$ *is the solution of* (6).

Note that we separately consider the cases with $\bar\alpha_i = 0$ and $\bar\alpha_i = 1$. This is because $\bar\alpha + \mathbf{1}_i$ is out of the feasible space of the binary decision variable when $\bar\alpha_i = 1$ and similarly for $\bar\alpha - \mathbf{1}_i$ when $\bar\alpha_i = 0$. Unlike a classical derivative with respect to continuous variable, the allowed directions for discrete variation depend on the base point $\bar\alpha$. Here, the new payoff functional uses the convex combination of the running payoff because $\mathcal{J}^{\epsilon(\bar\alpha,\alpha)}(z) = \int_0^T (1-\epsilon)r(z,\bar\alpha) + \epsilon r(z,\alpha)dt + q(z(T))$. The ϵ-variational system is used as a continuation tool of the discrete variation from one decision variable to another. The analytical properties of its solution are discussed in our previous work [19] and also summarized in [20].

This nonstandard definition of derivatives raises the two following questions: (i) *is the nonstandard derivative well-defined?*; and (ii) *is there a method to compute the nonstandard derivative?* We answer these two questions using the adjoint system (5) associated with the combinatorial optimization problem (2).

Theorem 1. *Suppose that Assumptions 1 and 3 hold. The nonstandard derivative $D^{NS}J : \{0,1\}^m \to \mathbb{R}^m$ satisfies*

$$[D^{NS}J(\bar{\alpha})]_i := \int_0^T \left(f(x^{\bar{\alpha}}(t), \bar{\alpha} + \mathbf{1}_i) - f(x^{\bar{\alpha}}(t), \bar{\alpha})\right)^\top \lambda^{\bar{\alpha}}(t) + r(x^{\bar{\alpha}}(t), \bar{\alpha} + \mathbf{1}_i) - r(x^{\bar{\alpha}}(t), \bar{\alpha})dt$$

when $\bar{\alpha}_i = 0$, and

$$[D^{NS}J(\bar{\alpha})]_i := \int_0^T \left(f(x^{\bar{\alpha}}(t), \bar{\alpha}) - f(x^{\bar{\alpha}}(t), \bar{\alpha} - \mathbf{1}_i)\right)^\top \lambda^{\bar{\alpha}}(t) + r(x^{\bar{\alpha}}(t), \bar{\alpha}) - r(x^{\bar{\alpha}}(t), \bar{\alpha} - \mathbf{1}_i)dt$$

when $\bar{\alpha}_i = 1$. Here, $x^{\bar{\alpha}}$ and $\lambda^{\bar{\alpha}}$ are the solutions of (1) *and* (5) *with $\bar{\alpha}$, respectively. The derivative uniquely exists and is bounded.*

The proof of Theorem 1 is contained in [20]. We also provide detailed comparisons between the standard and nonstandard derivative concepts in [20].

B. Complexity of Computing Derivatives

To solve the 0–1 linear program (3), we first need to compute the standard derivative $D^S J(\bar{\alpha})$ or the nonstandard derivative $D^{NS}J(\bar{\alpha})$. Recall that the dimensions of the system state and the binary decision variable are n and m, respectively. Let N_T be the number of time points in the time interval $[0, T]$ used to integrate the system (1) and the adjoint system (5). Then the complexity of computing the trajectories of $x^{\bar{\alpha}}$ and $\lambda^{\bar{\alpha}}$ is $O(nN_T)$ if the first-order forward Euler scheme is employed. Note that the computation of the adjoint state trajectory $\lambda^{\bar{\alpha}}$ requires the state trajectory $x^{\bar{\alpha}}$ in $[0, T]$. Given $x^{\bar{\alpha}}$ and $\lambda^{\bar{\alpha}}$, calculating all the entries of either the standard derivative or the nonstandard derivative requires $O(mnN_T)$ if a first-order approximation scheme for the integral over time is used. Therefore, the total complexity of computing either the standard derivative or the nonstandard derivative is $O(mnN_T)$.

C. Suboptimality Bounds

We now characterize the condition in which the solution to the approximate problem (3) has a guaranteed suboptimality bound. The suboptimality bound is obtained by showing that the optimal value of the payoff function is bounded by an affine function of the solution to the approximate problem (3). This motivates the following concavity-like assumption:

Assumption 5. *Let $\bar{\alpha} \in \{0,1\}^m$ be the point at which the original problem* (2) *is linearized. The following equality holds*

$$DJ(\bar{\alpha})^\top(\alpha - \bar{\alpha}) \geq J(\alpha) - J(\bar{\alpha}) \quad \forall \alpha \in \{0,1\}^m. \tag{7}$$

Here, DJ represents $D^S J$ if the standard derivative used in the approximate problem (3), and it represents $D^{NS}J$ if the nonstandard derivative is adopted in the approximate problem.

For notational convenience, we let $\mathcal{A}$ denote the feasible set of the optimization problem (2), i.e., $\mathcal{A} := \{\alpha \in \{0,1\}^m \mid \mathbf{A}\alpha \leq \mathbf{b}\}$. By subtracting $J(\bar{\alpha})$ from the payoff function, we normalize the payoff function such that, given $\bar{\alpha} \in \{0,1\}^m$ at which the original problem (2) is linearized, $J(\bar{\alpha}) = 0$. Note that $J(\alpha^{\text{OPT}}) \geq 0$, where α^{OPT} is a solution of the original optimization problem (2), if $\bar{\alpha} \in \mathcal{A}$.

Theorem 2 (Performance Guarantee)**.** *Suppose that Assumption 5 holds. Let*

$$\begin{aligned} &\alpha^{OPT} \in \arg\max_{\alpha \in \mathcal{A}} \ J(\alpha), \\ &\alpha^* \in \arg\max_{\alpha \in \mathcal{A}} \ D^S J(\bar{\alpha})^\top \alpha, \quad \hat{\alpha}^* \in \arg\max_{\alpha \in \mathcal{A}} \ D^{NS} J(\bar{\alpha})^\top \alpha. \end{aligned} \tag{8}$$

If $D^S J(\bar{\alpha})^\top(\alpha^* - \bar{\alpha}) \neq 0$ and $D^{NS}J(\bar{\alpha})^\top(\hat{\alpha}^* - \bar{\alpha}) \neq 0$, set

$$\rho := \frac{J(\alpha^*)}{D^S J(\bar{\alpha})^\top(\alpha^* - \bar{\alpha})}, \quad \hat{\rho} := \frac{J(\hat{\alpha}^*)}{D^{NS}J(\bar{\alpha})^\top(\hat{\alpha}^* - \bar{\alpha})}. \tag{9}$$

and we have the following suboptimality bounds for the solutions of the approximate problems, i.e., α^* and $\hat{\alpha}^*$:

$$\rho J(\alpha^{OPT}) \leq J(\alpha^*), \quad \hat{\rho} J(\alpha^{OPT}) \leq J(\hat{\alpha}^*). \tag{10}$$

Otherwise, $J(\alpha^{OPT}) = J(\bar{\alpha}) = 0$ when $\bar{\alpha} \in \mathcal{A}$.

Proof: Due to Assumption 5, we have

$$J(\alpha^{\text{OPT}}) = J(\alpha^{\text{OPT}}) - J(\bar{\alpha}) \leq D^S J(\bar{\alpha})^\top(\alpha^{\text{OPT}} - \bar{\alpha}). \tag{11}$$

On the other hand, because $\alpha^* \in \arg\max_{\alpha \in \mathcal{A}} \ D^S J(\bar{\alpha})^\top \alpha$ and $\alpha^{\text{OPT}} \in \mathcal{A}$,

$$D^S J(\bar{\alpha})^\top \alpha^{\text{OPT}} \leq D^S J(\bar{\alpha})^\top \alpha^*. \tag{12}$$

Suppose that $D^S J(\bar{\alpha})^\top(\alpha^* - \bar{\alpha}) \neq 0$. Then, $D^S J(\bar{\alpha})^\top(\alpha^* - \bar{\alpha}) > 0$ due to (8). Combining (11) and (12), we obtain the first inequality in (10); the second inequality can be derived using a similar argument. If $D^S J(\bar{\alpha})^\top(\alpha^* - \bar{\alpha}) = 0$ or $D^{NS}J(\bar{\alpha})^\top(\hat{\alpha}^* - \bar{\alpha}) = 0$, we have $J(\alpha^{\text{OPT}}) \leq 0 = J(\bar{\alpha})$. Due to the optimality of α^{OPT}, the inequality must be binding. ■

The coefficients ρ and $\hat{\rho}$ must be computed *a posteriori* because they require the solutions, α^* and $\hat{\alpha}^*$, respectively, of the approximate problems. They do not require the solution, α^{OPT}, of the original optimization problem. Note that ρ is, in general, different from $\hat{\rho}$. If $\bar{\alpha}$ is feasible, i.e., $\bar{\alpha} \in \mathcal{A}$, then we can improve the approximate solution by a simple post-processing that replaces it with $\bar{\alpha}$ if it is worse than $\bar{\alpha}$. The payoff functions evaluated at the post-processed approximate solutions are guaranteed to be greater than or equal to zero because $J(\bar{\alpha}) = 0$.

Corollary 1 (Post-Processing)**.** *Suppose that Assumption 5 holds and $\bar{\alpha} \in \mathcal{A}$. Let α^{OPT}, α^* and $\hat{\alpha}^*$ be given by* (8)*. Assume that $D^S J(\bar{\alpha})^\top(\alpha^* - \bar{\alpha}) \neq 0$ and $D^{NS}J(\bar{\alpha})^\top(\hat{\alpha}^* - \bar{\alpha}) \neq 0$. Define $\alpha_* \in \arg\max\{J(\alpha^*), J(\bar{\alpha})\}$, $\hat{\alpha}_* \in \arg\max\{J(\hat{\alpha}^*), J(\bar{\alpha})\}$, $\rho_* := \max\{\rho, 0\}$ and $\hat{\rho}_* := \max\{\hat{\rho}, 0\}$, where ρ and $\hat{\rho}$ are given by* (10)*. Then, the following suboptimality bounds for α_* and $\hat{\alpha}_*$ hold:*

$$\rho_* J(\alpha^{OPT}) \leq J(\alpha_*), \quad \hat{\rho}_* J(\alpha^{OPT}) \leq J(\hat{\alpha}_*).$$

The complexity of checking (7) in Assumption 5 for all $\alpha \in \{0,1\}^m$ increases exponentially as the dimension of the decision variable α increases. Therefore, we provide sufficient conditions, which are straightforward to check in some applications of interest, for Assumption 5. Note that the inequality condition (7) with $DJ = D^S J$ is equivalent to the concavity of the payoff function at $\bar{\alpha}$ if the space in which α lies is $[0,1]^m$ instead of $\{0,1\}^m$. This observation is summarized in the following proposition.

Proposition 2. *Suppose that Assumptions 1, 2, 3 and 4 hold. We also assume that the payoff function $J : \mathbb{R}^m \to \mathbb{R}$ in* (2a) *with x^α defined by* (1) *is concave in $[0,1]^m$, i.e.,*

$$J(\alpha) := \int_0^T r(x^\alpha(t), \alpha)dt + q(x^\alpha(T)),$$

with x^α satisfying $\dot{x}^\alpha(t) = f(x^\alpha(t), \alpha)$ and $x^\alpha(0) = \mathbf{x}$ is concave for all $\alpha \in [0,1]^m$. Then, the inequality condition (7) *with $DJ = D^S J$ holds for any $\bar{\alpha} \in \{0,1\}^m$.*

Recall that we view x^α as a function of α. Therefore, the concavity of J is affected by how the system state depends on α.

The inequality condition (7) with $DJ = D^{NS}J$ is difficult to interpret due to the nonstandard derivative. We reformulate the dynamical system and the payoff function such that (i) the standard derivative of the reformulated payoff function corresponds to the nonstandard derivative of the original payoff function and (ii) the reformulated and original payoff functions have the same values at any $\alpha \in \{0,1\}^m$. Then, the concavity of the reformulated payoff function guarantees the inequality (7). To be more precise, we consider the following *reformulated vector field and running payoff*:

$$\begin{aligned} \hat{f}(\,\cdot\,,\alpha) &:= f(\,\cdot\,,0) + \sum_{i=1}^m \alpha_i (f(\,\cdot\,,\mathbf{1}_i) - f(\,\cdot\,,0)), \\ \hat{r}(\,\cdot\,,\alpha) &:= r(\,\cdot\,,0) + \sum_{i=1}^m \alpha_i (r(\,\cdot\,,\mathbf{1}_i) - r(\,\cdot\,,0)). \end{aligned} \tag{13}$$

In general, $\hat{f}(\,\cdot\,,\alpha)$ (resp. $\hat{r}(\,\cdot\,,\alpha)$) and $f(\,\cdot\,,\alpha)$ (resp. $r(\,\cdot\,,\alpha)$) are different even when α is in the discrete space $\{0,1\}^m$. One can show that they are the same when $\alpha \in \{0,1\}^m$ if the following additivity assumption holds.

Assumption 6. *The functions $f(\boldsymbol{x},\,\cdot\,)$ and $r(\boldsymbol{x},\,\cdot\,)$ are additive in the entries of $\alpha \in \{0,1\}^m$ for all $\boldsymbol{x} \in \mathcal{X}$, i.e.,*

$$\begin{aligned} f(\,\cdot\,,\alpha) &= f(\,\cdot\,,0) + \sum_{i=1}^m (f(\,\cdot\,,\alpha_i\mathbf{1}_i) - f(\,\cdot\,,0)), \\ r(\,\cdot\,,\alpha) &= r(\,\cdot\,,0) + \sum_{i=1}^m (r(\,\cdot\,,\alpha_i\mathbf{1}_i) - r(\,\cdot\,,0)). \end{aligned}$$

Note that these additivity conditions are less restrictive than the conditions that both of the functions are affine in α as shown in [20].

This reformulation and Assumption 6 play an essential role in interpreting the nontrivial inequality condition (7) (with $DJ = D^{\text{NS}}J$) as the concavity of a reformulated payoff function, $\hat{J}$, defined in the next theorem. The standard derivative of the reformulated payoff function is equivalent to the nonstandard derivative of the original payoff function under Assumption 6, i.e., $D^{\text{S}}\hat{J} \equiv D^{\text{NS}}J$. Furthermore, the two payoff functions have the same values when α is in the discrete space $\{0,1\}^m$, i.e., $J|_{\{0,1\}^m} \equiv \hat{J}|_{\{0,1\}^m}$. Therefore, the inequality condition (7) with nonstandard derivative can be interpreted as the concavity of the reformulated payoff function.

Theorem 3. *Suppose that Assumptions 1, 3 and 6 hold. Define the reformulated payoff function $\hat{J} : \mathbb{R}^m \to \mathbb{R}$ as*

$$\hat{J}(\alpha) := \int_0^T \hat{r}(y^\alpha(t),\alpha)dt + q(y^\alpha(T)), \tag{14}$$

with y^α satisfying

$$\dot{y}^\alpha(t) = \hat{f}(y^\alpha(t),\alpha), \quad y^\alpha(0) = \mathbf{x} \in \mathcal{X},$$

where $\hat{f}$ and $\hat{r}$ are the reformulated vector field and running payoff, respectively, given in (13). *If the reformulated payoff function $\hat{J}$ is concave in $[0,1]^m$, then the inequality condition* (7) *with $DJ = D^{\text{NS}}J$ holds for any $\bar{\alpha} \in \{0,1\}^m$.*

Proof: Fix $\boldsymbol{x} \in \mathbb{R}^n$ and $i \in \{1,\cdots,m\}$. If $\alpha_i = 0$, then $\alpha_i(f(\boldsymbol{x},\mathbf{1}_i) - f(\boldsymbol{x},0)) = 0 = f(\boldsymbol{x},\alpha_i\mathbf{1}_i) - f(\boldsymbol{x},0)$. If $\alpha_i = 1$, then $\alpha_i(f(\boldsymbol{x},\mathbf{1}_i) - f(\boldsymbol{x},0)) = f(\boldsymbol{x},\alpha_i\mathbf{1}_i) - f(\boldsymbol{x},0)$. On the other hand, due to Assumption 6, we have $f(\boldsymbol{x},\alpha) = f(\boldsymbol{x},0) + \sum_{i=1}^m (f(\boldsymbol{x},\alpha_i\mathbf{1}_i) - f(\boldsymbol{x},0))$. Therefore, $\hat{f}(\boldsymbol{x},\alpha) = f(\boldsymbol{x},\alpha)$ for any $\alpha \in \{0,1\}^m$. Using a similar argument, we can show that $\hat{r}(\boldsymbol{x},\alpha) = r(\boldsymbol{x},\alpha)$ for any $\alpha \in \{0,1\}^m$. These imply that

$$\hat{J}(\alpha) = J(\alpha) \quad \forall \alpha \in \{0,1\}^m. \tag{15}$$

Furthermore, using the adjoint-based formula in Proposition 1 for the standard derivative of the reformulated payoff function $\hat{J}$, we obtain

$$\begin{aligned} [D^{\text{S}}\hat{J}(\alpha)]_i := \int_0^T & (f(x^\alpha(t),\mathbf{1}_i) - f(x^\alpha(t),0))^\top \lambda^\alpha(t) \\ & + r(x^\alpha(t),\mathbf{1}_i) - r(x^\alpha(t),0)dt. \end{aligned} \tag{16}$$

On the other hand, under Assumption 6, the adjoint-based formula for the nonstandard derivative of the original payoff function J can be rewritten as

$$\begin{aligned} [D^{\text{NS}}J(\alpha)]_i := \int_0^T & (f(x^\alpha(t),(\alpha_i+1)\mathbf{1}_i) - f(x^\alpha(t),\alpha_i\mathbf{1}_i))^\top \lambda^\alpha(t) \\ & + r(x^\alpha(t),(\alpha_i+1)\mathbf{1}_i) - r(x^\alpha(t),\alpha_i\mathbf{1}_i)dt \end{aligned}$$

when $\alpha_i = 0$, and

$$\begin{aligned} [D^{\text{NS}}J(\alpha)]_i := \int_0^T & (f(x^\alpha(t),\alpha_i\mathbf{1}_i) - f(x^\alpha(t),(\alpha_i-1)\mathbf{1}_i))^\top \lambda^\alpha(t) \\ & + r(x^\alpha(t),\alpha_i\mathbf{1}_i) - r(x^\alpha(t),(\alpha-1)\mathbf{1}_i)dt \end{aligned}$$

when $\alpha_i = 1$. Plugging $\alpha_i = 0$ and $\alpha_i = 1$ into the two formulae, respectively, and comparing them with (16), we conclude that

$$D^{\text{S}}\hat{J}(\alpha) = D^{\text{NS}}J(\alpha) \quad \forall \alpha \in \{0,1\}^m. \tag{17}$$

Suppose now that $\hat{J}$ is concave in $[0,1]^m$. Then, we have

$$D^{\text{S}}\hat{J}(\bar{\alpha})^\top(\alpha - \bar{\alpha}) \geq \hat{J}(\alpha) - \hat{J}(\bar{\alpha}) \quad \forall \bar{\alpha},\alpha \in \{0,1\}^m.$$

Combining this inequality with (15) and (17), we confirm that the inequality condition (7) with $DJ = D^{\text{NS}}J$ holds for any $\bar{\alpha} \in \{0,1\}^m$. ■

When the system is linear, the concavity of r (resp. $\hat{r}$) guarantees that the optimization problem (2) (resp. (14)) is concave. When the system is nonlinear, the results on convex control systems [21], [22] can be used to provide a sufficient condition for the concavity. In more general cases, we admit that it is nontrivial to check the concavity. Further studies on characterizing the conditions for the concavity will be performed in the future.

IV. Algorithms

We now propose approximation algorithms for the optimization (2) of combinatorial dynamical systems using the linear approximation proposed in the previous section. Formulating the approximate problem (3) only requires the computation of the standard or nonstandard derivative with computational complexity $O(mnN_T)$ as suggested in Section III-B, i.e., it is linear in the dimension of the decision variable. Because the approximate problem (3) is a 0–1 linear program, several polynomial time exact or approximation algorithms can be employed.

A. l_0-Norm Constraints

An important class of combinatorial optimization problems relevant to (2) is to maximize the payoff function, given that the l_0-norm of the decision variable is bounded. More specifically, instead of the original linear constraint (2b), we consider the constraint, $\underline{K} \leq \|\alpha\|_0 \leq \overline{K}$, where $\underline{K}$ and $\overline{K}$ are given constants. We consider the following first-order approximation of the combinatorial optimization problem:

$$\max_{\alpha \in \{0,1\}^m} \{DJ(\bar{\alpha})^\top \alpha \mid \underline{K} \leq \|\alpha\|_0 \leq \overline{K}\}. \tag{18}$$

A simple algorithm to solve (18) can be designed based on the ordering of the entries of $DJ(\bar{\alpha})$, where DJ is equal to either $D^{\text{S}}J$ or $D^{\text{NS}}J$. Let $\mathbf{d}(\cdot)$ denote the map from $\{1,\cdots,m\}$ to $\{1,\cdots,m\}$ such that $[DJ(\bar{\alpha})]_{\mathbf{d}(i)} \geq [DJ(\bar{\alpha})]_{\mathbf{d}(j)}$ for any $i,j \in \{1,\cdots,m\}$ such that $i \leq j$. Such a map can be constructed using a sorting algorithm with $O(m\log m)$ complexity. Note that such a map may not be unique. We let $\alpha_{\mathbf{d}(i)} = 1$ for $i = 1,\cdots,\underline{K}$. We then assign 1 on $\alpha_{\mathbf{d}(i)}$ if $[DJ(\bar{\alpha})]_{\mathbf{d}(i)} > 0$ and $\underline{K}+1 \leq i \leq \overline{K}$. Therefore, the total computational complexity to solve the approximate problem (18) is $O(mnN_T) + O(m\log m)$.

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/TAC.2015.2504867, IEEE Transactions on Automatic Control

B. General Linear Constraints

A totally unimodular (TU) matrix is defined as an integer matrix for which the determinant of every square non-singular sub-matrix is either $+1$ or -1. TU matrices play an important role in integer programs because they are invertible over the integers (e.g., Chapter III.1. of [23]). Suppose that $\mathbf{A}$ is TU and $\mathbf{b}$ is integral. Let $\bar{\mathbf{A}} := \begin{bmatrix} \mathbf{A} \\ I_{m\times m} \end{bmatrix}$ and $\bar{\mathbf{b}} := \begin{bmatrix} \mathbf{b} \\ \mathbf{1} \end{bmatrix}$, where $\mathbf{1}$ is the m-dimensional vector whose entries are all 1's. The new matrix $\bar{\mathbf{A}}$ is also TU. Therefore, the solution of the approximate problem can be obtained as the solution of the linear program, whose feasible region is relaxed to $\mathbb{R}^m$, of the form $\max_{\alpha\in\mathbb{R}^m}\{DJ(\bar{\alpha})^\top\alpha \mid \bar{\mathbf{A}}\alpha \leq \bar{\mathbf{b}}\}$. The proof of the exactness of this continuous relaxation can be found in [23].

Suppose that $l = 1$, i.e., $\mathbf{A} \in \mathbb{R}^{1\times l}$ is a vector and $\mathbf{b} \in \mathbb{R}$ is a scalar and that all the entries of $\mathbf{A}$ and $[DJ(\bar{\alpha})]_i$ are non-negative.[1] In this case, the approximate problem (3) is a 0–1 *knapsack problem*. Several algorithms and computational experiments for 0–1 knapsack problems can be found in the monograph [24] and the references therein. If no assumptions are imposed on the linear inequality constraints, then successive linear or semidefinite relaxation methods for a 0–1 polytope can provide approximation algorithms with suboptimality bounds [25], [26], [27].

Remark 2. *Our proposed* 0–1 *linear program approximation does not have any dynamical system constraints, while the original problem* (2) *does. This is advantageous because the approximate problem does not require any computational effort to solve the dynamical system once the standard or nonstandard derivative is calculated. In other words, the complexity of any algorithm applied to the approximate problem is independent of the time horizon* $[0, T]$ *of the dynamical system or the number,* N_T*, of discretization points in* $[0, T]$ *used to approximate* $DJ(\bar{\alpha})$.

We validated the performance of the proposed approximation algorithms by solving ON/OFF control problems of commercial interconnected refrigeration systems. In our numerical experiments, the suboptimality bound is greater than 74% though in practice the performance of the proposed approximation algorithm is greater than 90% of the oracle's performance. See [20] for more details.

V. Conclusion

We have proposed approximation algorithms for optimization of combinatorial dynamical systems, in which the decision variable is a binary vector and a payoff is evaluated along the solution of the systems. The key idea of the approximation is to replace the payoff function with a first-order approximation obtained from a derivative that is well-defined in the feasible space of the binary decision variable. We proposed two different variation methods to define such derivatives. The approximate problem has three major advantages: (i) the approximate problem is a 0–1 linear program and, therefore, can be exactly or approximately solved by polynomial time algorithms with suboptimality bounds; (ii) it does not require us to repeatedly simulate the dynamical system; and (iii) its solution has a provable suboptimality bound under certain concavity conditions.

References

[1] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdepencies," *IEEE Control Systems*, vol. 21, no. 6, pp. 11–25, 2001.

[2] M. S. Branicky, V. S. Borkar, and S. K. Mitter, "A unified framework for hybrid control: model and optimal control theory," *IEEE Transactions on Automatic Control*, vol. 43, no. 1, pp. 31–45, 1998.

[3] X. Xu and P. J. Antsaklis, "Optimal control of switched systems based on parametrization of the switching instants," *IEEE Transactions on Automatic Control*, vol. 49, no. 1, pp. 2–16, 2004.

[4] R. Vasudevan, H. Gonzalez, R. Bajcsy, and S. S. Sastry, "Consistent approximations for the optimal control of constrained switched systems–Part 1: A conceptual algorithm," *SIAM Journal on Control and Optimization*, vol. 51, no. 6, pp. 4463–4483, 2013.

[5] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints," *Automatica*, vol. 35, pp. 407–427, 1999.

[6] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Dover, 1998.

[7] L. J. Watters, "Reduction of integer polynomial programming problems to zero-one linear programming problems," *Operations Research*, vol. 15, no. 6, pp. 1171–1174, 1967.

[8] P. L. Hammer, P. Hansen, and B. Simeone, "Roof duality, complementation and persistency in quadratic 0–1 optimization," *Mathematical Programming*, vol. 28, pp. 121–155, 1984.

[9] P. L. Hammer and S. Rudeanu, *Boolean Methods in Operations Research*. New York: Springer-Verlag, 1968.

[10] A. H. Land and A. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, no. 3, pp. 497–520, 1960.

[11] W. Murray and K.-M. Ng, "An algorithm for nonlinear optimization problems with binary variables," *Computational Optimization and Applications*, vol. 47, no. 2, pp. 257–288, 2010.

[12] M. Grötschel, L. Lovász, and A. Schrijver, "The ellipsoid method and its consequences in combinatorial optimization," *Combinatorica*, vol. 1, no. 2, pp. 169–197, 1981.

[13] A. Schrijver, "A combinatorial algorithm minimizing submodular functions in strongly polynomial time," *Journal of Combinatorial Theory, Series B*, vol. 80, pp. 346–355, 2000.

[14] S. Iwata, L. Fleischer, and S. Fujishige, "A combinatorial strongly polynomial algorighm for minimizing submodular functions," *Journal of the Association for Computing Machinery*, vol. 48, no. 4, pp. 761–777, 2001.

[15] S. Jegelka and J. Bilmes, "Approximation bounds for inference using cooperative cuts," in *Proceedings of the 28th International Conference on Machine Learning*, 2011.

[16] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions–I," *Mathematical Programming*, vol. 14, pp. 265–294, 1978.

[17] E. Polak, *Optimization: algorithms and consistent approximations*. New York: Springer-Verlag, 1997.

[18] P. Kokotović and J. Heller, "Direct and adjoint sensitivity equations for parameter optimization," *IEEE Transactions on Automatic Control*, vol. 12, no. 5, pp. 609–610, 1967.

[19] I. Yang, S. A. Burden, S. S. Sastry, and C. J. Tomlin, "Infinitesimal interconnection variations in nonlinear networked systems," in *Proceedings of the 52nd IEEE Conference on Decision and Control*, 2013, pp. 1417–1422.

[20] I. Yang, "Risk management and combinatorial optimization for large-scale demand response and renewable energy integration," Ph.D. dissertation, University of California, Berkeley, 2015.

[21] V. Azhmyakov and J. Raisch, "Convex control systems and convex optimal control problems with constraints," *IEEE Transactions on Automatic Control*, vol. 53, no. 4, pp. 993–998, 2008.

[22] P. Colaneri, R. H. Middleton, Z. Chen, D. Caporale, and F. Blanchini, "Convexity of the cost functional in an optimal control problem for a class of positive switched systems," *Automatica*, vol. 50, pp. 1227–1234, 2014.

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

[24] S. Martello and P. Toth, *Knapsack Problems*. New York: John Wiley & Sons, 1990.

[25] L. Lovász and A. Schrijver, "Cones of matrices and set-fuctions and 0–1 optimization," *SIAM Journal on Optimization*, vol. 1, no. 2, pp. 166–190, 1991.

[26] H. D. Sherali and W. P. Adams, "A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems," *SIAM Journal on Discrete Mathematics*, vol. 3, no. 3, pp. 411–430, 1990.

[27] J. B. Lasserre, "An explicit equivalent positive semidefinite program for nonlinear 0–1 programs," *SIAM Journal on Optimization*, vol. 12, no. 3, pp. 756–769, 2002.

[1] The latter non-negativity assumption can easily be relaxed by fixing $\alpha_j = 0$ for j such that $[DJ(\bar{\alpha})]_j < 0$.