A Duality Framework for Stochastic Optimal Control of Complex Systems

Andreas A. Malikopoulos, *Member, IEEE*

***Abstract*—We address the problem of minimizing the long-run expected average cost of a complex system consisting of interactive subsystems. We formulate a multiobjective optimization problem of the one-stage expected costs of the subsystems and provide a duality framework to prove that the control policy yielding the Pareto optimal solution minimizes the average cost criterion of the system. We provide the conditions of existence and a geometric interpretation of the solution. For practical situations with constraints consistent to those studied here, our results imply that the Pareto control policy may be of value when we seek to derive online the optimal control policy in complex systems.**

***Index Terms*—Complex systems, multiobjective optimization, Pareto control policy, stochastic optimal control.**

I. Introduction

A. Motivation

COMPLEX systems consist of diverse entities that interact both in space and time. Referring to something as complex implies that it consists of interdependent entities that are connected with each other and can adapt, i.e., they can respond to their local and global environment [1]. Complex systems are encountered in many applications including sustainable transportation, fusion and other alternative energy strategies, and biological systems. For example, the U.S. electricity grid is one of the world's largest complex systems [2] consisting of a dynamic collection of diverse, interacting components that can adapt. These components are also interdependent and operate under an enormous range of physical, reliability, economic, social, and political constraints that need to be satisfied over time scales ranging from seconds, for closed-loop control, to decades, for transmission siting and construction. Hybrid electric vehicles (HEVs) and plug-in HEVs is another complex system [3] consisting of various interdependent subsystems, e.g., the internal combustion engine, the electric machines (motor and generator), and the energy storage system (battery), that are connected and adapt appropriately to provide the power demanded by the driver. Another example of complex system is the hybrid distributed power generation system [4] consisting of wind turbines, photovoltaic generation, energy storage, and the relevant energy conversion control.

Stochastic optimal control of complex systems is a ubiquitous task in engineering. The problem is formulated as sequential decision-making under uncertainty where a controller is faced with the task to select control actions in several time steps to achieve long-term goals efficiently. While the nature of these problems may vary widely, their underlying structure is similar and has two principal features: a discrete-time dynamic system whose state evolves according to given transition probabilities that depend on a decision at each time and a cost function that is additive over time. The objective is to derive an optimal policy that minimizes the long-run expected average cost criterion.

Mathematically, the average cost criterion is prominent as being complex to analyze compared to others; while other classical criteria lead to rational complete solutions, the long-run cost may not [5]. The average cost criterion in Markov chains with finite state and action spaces is well understood [6]–[12]. Dynamic programming (DP) [13] has been widely employed as the principal method for analysis of these problems[14]–[20]. A significant amount of work has focused on inventory problems using linear programming [21], [22], which has been widely used as an alternative to DP method [23]–[30]. Policy iteration [10] has been another method to address problems considering the average cost criterion [31], [32] by adjusting the policy of the system directly rather than using value iteration to derive it. Various other methods proposed in the literature have used matrix decomposition [33], quadratic programming for multiple costs [34], learning algorithms [35], [36], decentralized methods [37], and the risk-sensitive criterion [38].

Despite the significant progress in optimization and control methods within the last decades current techniques, in some instances, may be computationally impractical for online optimal control of large-scale complex systems [39]. One possible approach for ameliorating this difficulty is to develop the framework that exploits the structure of the system interconnections and narrow the range of acceptable solutions.

In this paper, we seek to establish a rigorous framework for the analysis and stochastic optimization of complex systems that will permit online implementation of the optimal control policy with respect to the long-run expected average cost criterion. The contributions of this paper are: 1) the development of a duality framework for the analysis and stochastic optimization of complex systems that can be used to derive the optimal control policy; 2) the formulation and solution of a multiobjective

Manuscript received April 14, 2014; revised August 28, 2014 and October 4, 2015; accepted October 26, 2015. Date of publication December 3, 2015; date of current version September 23, 2016. This work was supported by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory and by the U.S. Department of Energy under Contract DE-AC05-00OR22725. Recommended by Associate Editor A. Nedich.

The author is with the Energy and Transportation Science Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA (e-mail: andreas@ornl.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2015.2504518

optimization problem of the one-stage expected costs of all interactive subsystems yielding an equilibrium operating point among the subsystems that minimizes the long-run expected average cost of the system; and 3) the geometric interpretation of the solution and the formation of the conditions under which the optimal control policy exists.

The remainder of the paper proceeds as follows. In Section II, we introduce our notation and formulate the problem. In Section III, we develop a multiobjective optimization framework to address the problem and introduce the Pareto control policy. In Section IV, we show that the Pareto control policy minimizes the long-run expected average cost criterion. Finally, we present illustrative examples in Section V and concluding remarks in Section VI.

II. System Model and Problem Formulation

A. Notation

We denote random variables with upper case letters, and their realization with lower case letters, e.g., for a random variable X, x denotes its realization. Subscripts denote time, and subscripts in parentheses denote subsystems; for example, $X_{t(i)}$ denotes the random variable of the subsystem i at time t, and $x_{(i)}$ its realization. The shorthand notation $X_{t(1:N)}$ denotes the vector of random variables $(X_{t(1)}, X_{t(2)}, \ldots, X_{t(N)})$ and $x_{(1:N)}$ denotes the vector of their realization $(x_{(1)}, x_{(2)}, \ldots, x_{(N)})$. $\mathbb{P}(\cdot)$ is the transition probability matrix, and $\mathbb{E}[\cdot]$ is the corresponding expectation of a random variable. For a control policy $\boldsymbol{\pi}$, we use $\mathbb{P}^{\pi}(\cdot)$, $\mathbb{E}^{\pi}[\cdot]$ and β^{π} to denote that the transition probability matrix, expectation and stationary distribution depend on the choice of the control policy π.

B. The System Model

We consider a system consisting of N subsystems. The subsystems interact with each other and their environment. At time $t, t = 1, 2, \ldots, T$, the state of each subsystem $i, X_{t(i)}$, takes values in a finite state space $\mathcal{S}_{(i)}$, which is a metric space. For each subsystem i, we also consider a finite control space $\mathcal{U}_{(i)}$, which is also a metric space, from which control actions, $U_{t(i)}$, are chosen.

The initial state of the system $X_{0(1:N)}$ is a random variable taking values in the system's state space, $\mathcal{S} = \prod_{i=1}^{N} \mathcal{S}_{(i)}$. The evolution of the state is imposed by the discrete-time equation

$$X_{t+1(1:N)} = f(X_{t(1:N)}, U_{t(1:N)}, W_{t(1:N)}) \qquad (1)$$

where $W_{t(1:N)}$ is the input from the environment. The system state can be completely observed.

In our formulation, a state-dependent constraint is incorporated; that is, for each realization of the state of the subsystem i, $X_{t(i)} = x_{(i)}$, there is a nonempty and closed set $\mathcal{C}(x_{(i)}) := \{u_{(i)} | X_{t(i)} = x_{(i)}\} \subset \mathcal{U}_{(i)}$ of feasible control actions when the system is in state $x_{(i)}$. For each subsystem i, we denote the set of admissible state/action pairs

$$\Gamma_{(i)} := \left\{(x_{(i)}, u_{(i)}) | x_{(i)} \in \mathcal{S}_{(i)} \text{ and } u_{(i)} \in \mathcal{C}(x_{(i)})\right\}. \qquad (2)$$

The set of admissible state/action pairs for the system is

$$\Gamma := \prod_{i=1}^{N} \Gamma_{(i)} = \left\{(x_{(1:N)}, u_{(1:N)})) | x_{(1:N)} \in \mathcal{S} \right.$$
$$\left. \text{and } u_{(1:N)} \in \mathcal{C}(x_{(1:N)})\right\} \qquad (3)$$

where $\mathcal{C}(x_{(1:N)}) = \prod_{i=1}^{N} \mathcal{C}_{(i)}(x_{(i)})$.

For each state of the system $X_{t(1:N)} = x_{(1:N)}$, we define the functions $\mu : \mathcal{S} \rightarrow \mathcal{U}$, where $\mathcal{U} = \prod_{i=1}^{N} \mathcal{U}_{(i)}$, that map the state space to the control action space defined as the control law. When the system is at state $X_{t(1:N)} = x_{(1:N)}$, the controller chooses action according to the control law $u_{(1:N)} = \mu(x_{(1:N)})$.

Definition 1: Each sequence of the functions μ is defined as a stationary control policy of the system

$$\pi := (\mu(1), \mu(2), \ldots, \mu(|\mathcal{S}|)) \qquad (4)$$

where $|\mathcal{S}|$ is the cardinality of the system's state space $\mathcal{S}$.

Let Π denote the set of the collection of the stationary control policies

$$\Pi := \{\pi | \pi = (\mu(1), \mu(2), \ldots, \mu(|\mathcal{S}|)\}. \qquad (5)$$

The stationary control policy π operates as follows. Associated with each state $X_{t(1:N)} = x_{(1:N)}$ is the function $\mu(x_{(1:N)}) \in \mathcal{C}(x_{(1:N)})$. If at any time the controller finds the system in state $x_{(1:N)}$, then the controller always chooses the action based on the function $\mu(x_{(1:N)})$. A stationary policy depends on the history of the process only through the current state, and thus to implement it, the controller only needs to know the current state of the system. The advantages for implementation of a stationary policy are apparent as it requires the storage of less information than required to implement a general policy.

At each stage t, the controller observes the state of the system, $X_{t(1:N)} = x_{(1:N)} \in \mathcal{S}$, and an action, $u_{t(1:N)} = \mu(X_{t(1:N)})$, is realized from the feasible set of actions at that state. At the same stage t, an uncertainty, $W_{t(1:N)}$, is incorporated in the system. At the next stage, $t+1$, the system transits to the state $X_{t+1(1:N)} = x'_{(1:N)} \in \mathcal{S}$ and a transition cost for each subsystem i, $c_{t(i)}(X_{t+1(i)} | X_{t(i)}, U_{t(i)})$, where $c_{t(i)} : \mathcal{S}_{(i)} \times \mathcal{C}(x_{(i)}) \times \mathcal{S}_{(i)} \rightarrow \mathbb{R}$, and for the system, $c_t(X_{t+1(1:N)} | X_{t(1:N)}, U_{t(1:N)})$, where $c_t : \mathcal{S} \times \mathcal{C}(x_{(1:N)}) \times \mathcal{S} \rightarrow \mathbb{R}$, are incurred.

C. Assumptions

In the model described above, we consider the following assumptions:

(A1) There exists μ such that the graph of μ is included in Γ.

(A2) The input from the uncertainty $W_{t(1:N)}$ is a sequence of independent random variables, independent of the initial state $X_{0(1:N)}$, and takes values in the finite set $\mathcal{W}$.

(A3) For each stationary control policy π, the Markov chain $\{X_{t(1:N)} | t = 1, 2, \ldots\}$ has a unique probability distribution (row vector).

(A4) The one-stage expected cost of the system, $k_t^\pi : \Gamma \to \mathbb{R}$

$$k_t^\pi\left(X_{t(1:N)}, U_{t(1:N)}\right) = \sum_{x'_{(1:N)} \in \mathcal{S}} P\Big(X_{t+1(1:N)} = x'_{(1:N)} | X_{t(1:N)} = x_{(1:N)}, U_{t(1:N)}\Big) \cdot c_t\left(X_{t+1(1:N)} = x'_{(1:N)} | X_{t(1:N)} = x_{(1:N)}, U_{t(1:N)}\right)$$

is a continuous function of the one-stage costs of the subsystems and it is uniformly bounded.

(A5) The control action realized at each subsystem doesn't affect the transition probability matrix of the other subsystems.

We briefly comment on the above assumptions. A1 ensures that the set of the collection of the stationary control policies, Π, is nonempty. A2 imposes a condition yielding that the state $X_{t+1(1:N)}$ depends only on $X_{t(1:N)}$ and $U_{t(1:N)}$. Namely, the evolution of the state is a Markov chain [9]. A3 implies that for each stationary policy $\pi \in \Pi$, there is a unique probability distribution (row vector) $\beta^\pi = (\beta(1)^\pi, \beta(2)^\pi, \ldots, \beta(k)^\pi, \ldots, \beta(|\mathcal{S}|)^\pi)$, with $\sum_{k=1}^{|\mathcal{S}|} \beta(k)^\pi = 1$ [40, p. 227] such that $\beta^\pi = \beta^\pi \cdot \mathbb{P}^\pi$. Under this assumption, it is known [41, p. 175] that

$$\lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} [\mathbb{P}^\pi]^t = \mathbf{1} \cdot \beta^\pi \tag{6}$$

where $\mathbb{P}^\pi$ is the transition probability matrix and $\mathbf{1} = [1, 1, \ldots, 1]^T$. A4 imposes that the interaction of the subsystems has an impact on one-stage expected cost of the system. Finally, A5 implies that the subsystems evolve independently.

D. Problem Formulation

We are concerned with deriving a stationary optimal control policy π to minimize the long-run expected average cost of the system

$$J(\pi) = \lim_{T\to\infty} \frac{1}{T+1} \mathbb{E}^\pi \left[\sum_{0}^{T} k_t^\pi\left(X_{t(1:N)}, U_{t(1:N)}\right) \right]. \tag{7}$$

Since for each control policy the Markov chain has a unique probability distribution (A3), it follows that the limit in (7) exists. Substituting (6) into (7) shows that the long-run average cost, $J(\pi)$, does not depend on the initial state $X_{0(1:N)}$ and is given simply as

$$J(\pi) = \beta^\pi \cdot k^\pi \tag{8}$$

where β^π the stationary probability distribution of the entire system and

$$k^\pi = \left(k_t^\pi\left(1, U_{t(1:N)}\right), k_t^\pi\left(2, U_{t(1:N)}\right), \ldots, k_t^\pi\left(|\mathcal{S}|, U_{t(1:N)}\right)\right)^T \tag{9}$$

is the column vector of the system's one-stage expected cost.

Various methods that discussed in the Introduction can be used to solve (7) or (8) offline and derive the optimal control policy that minimizes the long-run expected average cost J of the system. In this paper, we seek the theoretical framework that will implement the optimal control policy online while the subsystems interact with each other. The intention here is to identify an equilibrium operating point among the subsystems; if the systems operate at this equilibrium, then the average cost of the system will be minimized.

III. Multiobjective Optimization Analysis

A. Pareto Control Policy

To identify an equilibrium operating point among the subsystems, we formulate a multiobjective optimization problem for the one-stage cost of the subsystems. Let us consider the function $f : \Gamma \to \mathbb{R}^N$

$$f = \Big(k_{t(1)}^\pi\left(X_{t(1:N)}, U_{t(1:N)}\right), k_{t(2)}^\pi\left(X_{t(1:N)}, U_{t(1:N)}\right), \ldots, k_{t(N)}^\pi\left(X_{t(1:N)}, U_{t(1:N)}\right)\Big) \tag{10}$$

where $k_{t(i)}^\pi(X_{t(1:N)}, U_{t(1:N)})$ is the one-stage expected cost for each subsystem i and the following multiobjective optimization problem:

$$\min_{U_{t(1:N)} \in \mathcal{C}(x_{(1:N)})} \Big(k_{t(1)}^\pi\left(X_{t(1:N)}, U_{t(1:N)}\right), k_{t(2)}^\pi\big(X_{t(1:N)}, U_{t(1:N)}\big), \ldots, k_{t(N)}^\pi\big(X_{t(1:N)}, U_{t(1:N)}\big)\Big). \tag{11}$$

The result of the problem (11) is called Pareto efficiency. In a Pareto efficiency allocation among agents, no one can be made better without making at least one other agent worse. The following result provides the conditions that the Pareto efficiency exists.

Proposition 1 [42]: Let Γ be a nonempty and compact set, and the one-stage expected cost for each subsystem i, $k_{t(i)}^\pi(X_{t(i)}, U_{t(i)}) : \Gamma \to \mathbb{R}$, be lower semicontinuous for all $i = 1, \ldots, N$. Then the Pareto efficiency is not empty.

In our problem, the set of admissible state/action pairs, Γ, is a nonempty compact set (A1). Furthermore, the one-stage expected cost for each subsystem i, $k_{t(i)}^\pi(X_{t(i)}, U_{t(i)})$, is a continuous function (A4). Consequently, the Pareto efficiency exists.

Definition 2: The Pareto control policy π^o is defined as the policy that yields the minimum one-stage expected cost of the system, $k_t^{\pi^o}(X_{t(1:N)}, U_{t(1:N)}^o)$, at each realization of the system state $X_{t(1:N)} = x_{(1:N)}$.

B. Impact of the Pareto Control Policy on the System's Expected Cost

To simplify notation, in the rest of the paper the one-stage expected cost of each subsystem i, $k_{t(i)}^\pi(X_{t(1:N)}, U_{t(1:N)})$, and the one-stage expected cost of the system, $k_t^\pi(X_{t(1:N)}, U_{t(1:N)})$, incurred when the system operates under the control policy π, will be denoted by $k_{t(i)}^\pi$ and k_t^π, respectively.

Definition 3: In a system consisting of N interactive subsystems, the group of subsystems whose expected costs are a decreasing function with respect to the cost of the system is defined as the *minor* group.

Definition 4: In a system consisting of N interactive subsystems, the group of subsystems whose expected costs are an increasing function with respect to the cost of the system is defined as the *principal* group.

Without loss of generality, we assume that the minor group consists of the subsystems $1, 2, \ldots, m, m \in \mathbb{N}$, and the principal group consists of the subsystems $m+1, \ldots, N$. Thus, since the one-stage expected cost of the system is a function δ of the one-stage cost of the subsystems (A4)

$$k_t^\pi = \delta\left(k_{t(1)}^\pi, k_{t(2)}^\pi, \ldots, k_{t(N)}^\pi\right) \tag{12}$$

from Definition 3, for each subsystem i in the minor group and for any two control policies $\pi, \pi' \in \Pi$ such that $k_{t(i)}^\pi \le k_{t(i)}^{\pi'}$, if we fix the one-stage cost of the other subsystems in both minor and principal groups we have

$$k_t^\pi = \delta\left(\ldots, k_{t(i)}^\pi, \ldots\right) \ge k_t^{\pi'} = \delta\left(\ldots, k_{t(i)}^{\pi'}, \ldots\right). \tag{13}$$

Similarly, from Definition 4, for each subsystem j in the principal group and for any two control policies $\pi, \pi' \in \Pi$ such that $k_{t(j)}^\pi \le k_{t(j)}^{\pi'}$, if we fix the one-stage cost of the other subsystems in both minor and principal groups, we have

$$k_t^\pi = \delta\left(\ldots, k_{t(j)}^\pi, \ldots\right) \le k_t^{\pi'} = \delta\left(\ldots, k_{t(j)}^{\pi'}, \ldots\right). \tag{14}$$

Problem 1: We consider the special case where the system consists of N subsystems of a minor group only.

Proposition 2: The solution of the following multiobjective optimization problem at each realization of the state $X_{t(1:N)} = x_{(1:N)}$ yields the Pareto control policy of the system

$$\max_{U_{t(1:N)} \in \mathcal{C}(x_{(1:N)})} \left(k_{t(1)}^\pi, \ldots, k_{t(N)}^\pi\right)$$
$$\text{subject to} \quad X_{t(1:N)} \in \mathcal{S}. \tag{15}$$

Proof: Let $u_{(1:N)}^* \in \mathcal{C}(x_{(1:N)})$ be the solution of (15) at each realization of the state $X_{t(1:N)} = x_{(1:N)}$ under the control policy π. Thus, if we operate the system under π, then $k_{t(i)}^\pi \ge k_{t(i)}^{\pi'}, i = 1, \ldots, N$, for all $\pi' \in \Pi$. Therefore from Definition 3 we have $k_t^\pi \le k_t^{\pi'}$, for all $\pi' \in \Pi$, and hence from Definition 2 π is the Pareto control policy. ■

Problem 2: We consider the special case where the system consists of N subsystems of a principal group only.

Proposition 3: The solution of the following multiobjective optimization problem at each realization of the state $X_{t(1:N)} = x_{(1:N)}$ yields the Pareto control policy of the system.

$$\min_{U_{t(1:N)} \in \mathcal{C}(x_{(1:N)})} \left(k_{t(1)}^\pi, \ldots, k_{t(N)}^\pi\right)$$
$$\text{subject to} \quad X_{t(1:N)} \in \mathcal{S}. \tag{16}$$

Proof: Let $u_{(1:N)}^* \in \mathcal{C}(x_{(1:N)})$ be the solution of (16) at each realization of the state $X_{t(1:N)} = x_{(1:N)}$ under the control policy π. Thus, if we operate the system under π, then $k_{t(i)}^\pi \le k_{t(i)}^{\pi'}, i = 1, \ldots, N$, for all $\pi' \in \Pi$. Therefore from Definition 4 we have $k_t^\pi \le k_t^{\pi'}$, for all $\pi' \in \Pi$, and hence from Definition 2 π is the Pareto control policy. ■

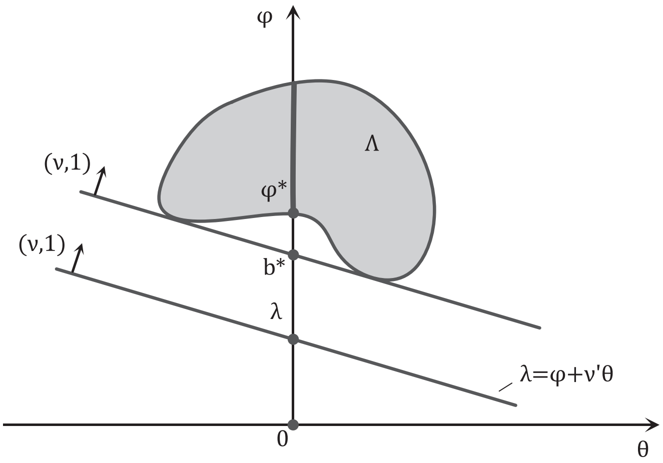

Fig. 1. Geometric framework for duality analysis.

Problem 3: We consider the general case where the system consists of N subsystems of both a minor and principal group.

In this case, to derive the Pareto control policy, we formulate the following optimization problem for the one-stage cost of the system:

$$\min_{U_{t(1:N)} \in \mathcal{C}(x_{(1:N)})} k_t^\pi =$$
$$\min_{U_{t(1:N)} \in \mathcal{C}(x_{(1:N)})} \delta\left(k_{t(1)}^\pi, k_{t(2)}^\pi, \ldots, k_{t(N)}^\pi\right)$$
$$\text{subject to } X_{t(1:N)} \in \mathcal{S}. \tag{17}$$

The Pareto control policy is derived by computing at each realization of the system state $X_{t(1:N)} = x_{(1:N)} \in \mathcal{S}$, the control action $u_{(1:N)}^o$ that yields the minimum one-stage expected cost of the system in (17).

IV. Duality Framework

A. Geometric Framework for Duality Analysis

We use a geometric framework from duality analysis, referred to as *min common/max crossing point* problems (see [43, p. 120]), to show that the Pareto control policy is an optimal control policy that minimizes the long-run expected average cost of the system, and provide a geometric interpretation of the solution.

The *min common/max crossing point* framework captures the most essential elements of duality by considering two geometric problems. Let's consider a nonempty subset Λ of $\mathbb{R}^{n+1}$ as shown in Fig. 1. The axis θ corresponds to $\mathbb{R}^n$ and the axis φ corresponds to $\mathbb{R}$.

The first geometric problem, the *min common point*, seeks to find the minimum value φ^* of the subset Λ in φ axis. The second geometric problem, the *max crossing point*, seeks to find the nonvertical hyperplane that contains Λ in its corresponding upper closed half space and crosses φ axis at a maximum point b^*.

Mathematically, the *min common point* problem can be written as

$$\begin{aligned}&\min \varphi\\ &\text{subject to} : (0,\varphi) \in \Lambda.\end{aligned} \tag{18}$$

A nonvertical hyperplane in $\mathbb{R}^{n+1}$ is specified by its normal $(\nu, 1) \in \mathbb{R}^{n+1}$, where $\nu \in \mathbb{R}^n$, and a scalar $\lambda \in \mathbb{R}$ as

$$\varphi + \nu'\theta = \lambda. \tag{19}$$

Such a hyperplane crosses the $(n+1)$st axis, φ, at $(0,\lambda)$. The hyperplane contains Λ in its upper closed half plane if and only if for all $(\theta,\varphi) \in \Lambda$

$$\varphi + \nu'\theta \geq \lambda. \tag{20}$$

Similarly

$$\inf_{(\theta,\varphi)\in\Lambda}\{\varphi + \nu'\theta\} \geq \lambda. \tag{21}$$

Thus, the *max crossing point* problem can be written

$$\begin{aligned}&\max \inf_{(\theta,\varphi)\in\Lambda}\{\varphi + \nu'\theta\}\\ &\text{subject to} : \nu \in \mathbb{R}^n.\end{aligned} \tag{22}$$

The function $b(\nu) = \inf\{\varphi + \nu'\theta\}$ is the dual function.

Definition 5: If $(\bar{\theta}, \bar{\varphi})$ belongs to the closure of Λ and for all $(\theta,\varphi) \in \Lambda$, $\bar{\theta} + \nu' \cdot \bar{\varphi} \leq \theta + \nu' \cdot \varphi$, we say that the hyperplane supports Λ at $(\bar{\theta}, \bar{\varphi})$.

Proposition 4—(See [43, p. 123]): The *max crossing point* of the dual function is less than or equal to the *min common point*, namely $b^* \leq \varphi^*$.

Proof: For all $(\theta,\varphi) \in \Lambda$ and $\nu \in \mathbb{R}^n$, we have

$$b(\nu) = \inf_{(\theta,\varphi)\in\Lambda}\{\varphi + \nu'\theta\} \leq \inf_{(0,\varphi)\in\Lambda}\{\varphi\}. \tag{23}$$

Taking the supremum over $\nu \in \mathbb{R}^n$, we have

$$b^* = \sup_{\nu\in\mathbb{R}^n}\inf_{(\theta,\varphi)\in\Lambda}\{\varphi + \nu'\theta\} \leq \varphi^* = \sup_{\nu\in\mathbb{R}^n}\inf_{(0,\varphi)\in\Lambda}\{\varphi\}. \tag{24}$$

■

B. Strong Duality of the Pareto Control Policy

We want to investigate the impact of the Pareto control policy on the long-run expected average cost of the system. This will involve characterizing the solution of the Pareto control policy within a duality framework. We recall that k^π is the column vector of the system's one-stage expected cost for each state, $1, 2, \ldots, |\mathcal{S}|$, under the control policy $\pi = (\mu(1), \mu(2), \ldots, \mu(|\mathcal{S}|))$, namely

$$k^\pi = \left(k^\pi\left(1,\mu(1)\right), k^\pi\left(2,\mu(2)\right), \ldots, k^\pi\left(|\mathcal{S}|,\mu(|\mathcal{S}|)\right)\right)^T.$$

We formulate the following problem:

$$\begin{aligned}&\min_{\pi\in\Pi} \|k^\pi + \mathbb{M}^\pi \cdot q\|\\ &\text{subject to} : \beta^\pi \cdot \mathbb{M}^\pi = 0\end{aligned} \tag{25}$$

where $\mathbb{M}^\pi = \mathbb{P}^\pi - I$, $q \in \mathbb{R}^{|\mathcal{S}|}$ such that $\mathbb{M}^\pi \cdot q > 0$, and $\beta^\pi = (\beta(1)^\pi, \beta(2)^\pi, \ldots, \beta(|\mathcal{S}|)^\pi)$ is the probability distribution corresponding to the control policy π.

We refer to this problem as the primal problem, and we denote by $\|k^\pi + \mathbb{M}^\pi \cdot q\|^*$ its optimal value. The Lagrangian function of the above minimization problem is

$$L(\pi,\nu) = \|k^\pi + \mathbb{M}^\pi \cdot q\| + (\beta^\pi \cdot \mathbb{M}^\pi) \cdot \nu \tag{26}$$

where $\nu \in \mathbb{R}^{|\mathcal{S}|}$ is the vector of the Lagrange multipliers.

We use the *min common/max crossing point* framework described above to visualize the duality in (26). We consider the following set:

$$\Lambda := \left\{(\beta^\pi \cdot \mathbb{M}^\pi, \|k^\pi + \mathbb{M}^\pi \cdot q\| | \pi \in \Pi\right\}. \tag{27}$$

Lemma 1: The hyperplane with norm $(\nu, 1)$ that passes through the vector $(\beta^\pi \cdot \mathbb{M}^\pi, \|k^\pi + \mathbb{M}^\pi \cdot q\|)$ intercepts the vertical axis φ at the value of $L(\pi,\nu)$.

Proof: The hyperplane with norm $(\nu, 1)$ that passes through $(\beta^\pi \cdot \mathbb{M}^\pi, \|k^\pi + \mathbb{M}^\pi \cdot q\|)$ satisfies

$$\varphi + \theta' \cdot \nu = \|k^\pi + \mathbb{M}^\pi \cdot q\| + (\beta^\pi \cdot \mathbb{M}^\pi) \cdot \nu = L(\pi,\nu). \tag{28}$$

■

Lemma 2: The hyperplane that passes through $\|k^\pi + \mathbb{M}^\pi \cdot q\|^*$ supports Λ.

Proof: From Lemma 1, we have

$$\varphi + \theta' \cdot \nu = \|k^\pi + \mathbb{M}^\pi \cdot q\| + (\beta^\pi \cdot \mathbb{M}^\pi) \cdot \nu = \lambda. \tag{29}$$

Since for each stationary control policy we have a unique probability distribution (A3)

$$\beta^\pi = \beta^\pi \cdot \mathbb{P}^\pi \Rightarrow \beta^\pi \cdot (\mathbb{P}^\pi - I) = 0 \Rightarrow (\beta^\pi \cdot \mathbb{M}^\pi) = 0. \tag{30}$$

Thus, for each control policy $\pi \in \Pi$, the elements of the set Λ are located only on the axis φ, and

$$\|k^\pi + \mathbb{M}^\pi \cdot q\| = \lambda. \tag{31}$$

Thus

$$\|k^{\pi^*} + \mathbb{M}^{\pi^*} \cdot q\|^* = \varphi^* \leq \|k^\pi + \mathbb{M}^\pi \cdot q\| = \lambda, \quad \forall\, \pi \in \Pi. \tag{32}$$

■

Theorem 1: The Pareto control policy π^o is the optimal control policy that minimizes the long-run expected average cost criterion of the system, under the assumption (A3) and (A4).

Proof: Let

$$\mathbf{1} \cdot \psi = k^\pi + \mathbb{M}^\pi \cdot q, \quad \forall\, \pi \in \Pi \tag{33}$$

where $\mathbf{1} = (1, 1, \ldots, 1)^T$, and $\psi^\pi \in \mathbb{R}$. Recall that $q \in \mathbb{R}^{|\mathcal{S}|}$ such that $\mathbb{M}^\pi \cdot q > 0$.

Multiplying the above equation by $\beta^\pi = (\beta(1)^\pi, \beta(2)^\pi, \ldots, \beta(k)^\pi, \ldots, \beta(|\mathcal{S}|)^\pi)$ from the left we have

$$\begin{aligned}\psi^\pi &= \beta^\pi \cdot k^\pi + \beta^\pi \cdot \mathbb{M}^\pi \cdot q &(34)\\ &= \beta^\pi \cdot k^\pi + \beta^\pi \cdot (\mathbb{P}^\pi - I) \cdot q &(35)\\ &= \beta^\pi \cdot k^\pi + \beta^\pi \cdot \mathbb{P}^\pi \cdot q - \beta^{\pi^o} \cdot q &(36)\\ &= \beta^\pi \cdot k^\pi + \beta^\pi \cdot q - \beta^\pi \cdot q = \beta^\pi \cdot k^\pi &(37)\end{aligned}$$

since $\mathbb{M}^\pi = \mathbb{P}^\pi - \mathbb{I}$ and $\beta^\pi = \beta^\pi \cdot \mathbb{P}^\pi$. So from (8), ψ^π is the long-run expected average cost corresponding to the control policy π.

From the Definition 2 of the Pareto control policy

$$k^{\pi^o} \leq k^{\pi}, \quad \forall\, \pi \in \Pi \tag{38}$$

and since $\mathbb{M}^{\pi} \cdot q > 0$, (38) through (33) can be written

$$k^{\pi^o} \leq k^{\pi} + \mathbb{M}^{\pi} \cdot q = \mathbf{1} \cdot \psi^{\pi} \tag{39}$$

where ψ^{π} is the long-run expected average cost corresponding to any control policy $\pi \in \Pi$. Multiplying (39) by β^{π^o} from the left, we have

$$\psi^{\pi^o} = \beta^{\pi^o} \cdot k^{\pi^o} \leq \psi^{\pi}, \quad \forall\, \pi \in \Pi. \tag{40}$$

Thus, the Pareto control policy is the optimal control policy that minimizes the long-run expected average cost. ■

Theorem 2: The Pareto control policy π^o supports Λ.

Proof: From Theorem 1, we have

$$k^{\pi^o} + \mathbb{M}^{\pi^o} \cdot q \leq k^{\pi} + \mathbb{M}^{\pi} \cdot q, \quad \forall\, \pi \in \Pi. \tag{41}$$

Hence

$$\|k^{\pi^o} + \mathbb{M}^{\pi^o} \cdot q\| \leq \|k^{\pi} + \mathbb{M}^{\pi} \cdot q\| \tag{42}$$

and from Lemma 2

$$\|k^{\pi^*} + \mathbb{M}^{\pi^*} \cdot q\|^* = \|k^{\pi^o} + \mathbb{M}^{\pi^o} \cdot q\|. \tag{43}$$

■

Corollary 1: There is no duality gap in (26), and thus the Pareto control policy π^o yields the global optimal solution.

V. Illustrative Examples

A. Preliminary Results

In this section, we provide some results that we need to use for the illustrative examples in the next subsection. We begin by recalling the Kronecker product and its properties (see [44], [45]).

Definition 6: If A is an m-by-n matrix and B is a p-by-q matrix, then the Kronecker product $A \otimes B$ is the mp-by-np block matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}.$$

The next proposition provides an expression of the transition probability of the entire system as a Kronecker product of the transition probabilities of each subsystem.

Proposition 5 [46]: Consider N evolving subsystems with corresponding transition probability matrices $\mathbb{P}_{(i)}$, $i = 1, \ldots, N$ defined by $\mathbb{P}_{(i)}(X_{t+1(i)} = x'_{(i)} | X_{t(i)} = x_{(i)}, U_{t(i)} = u_{(i)})$. Now consider that the system operates under the control policy π. Then the transition probability matrix of the entire system satisfies

$$\mathbb{P}^{\pi} = \mathbb{P}^{\pi}_{(1)} \otimes \mathbb{P}^{\pi}_{(2)} \otimes \cdots \otimes \mathbb{P}^{\pi}_{(N)}. \tag{44}$$

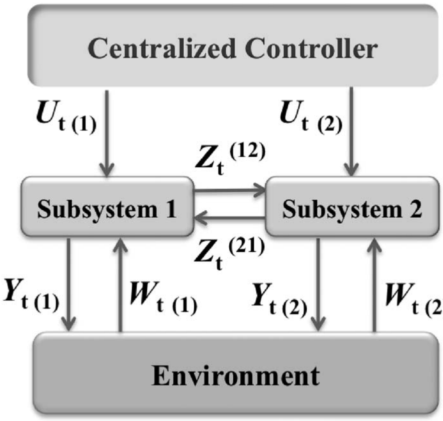

Fig. 2. System of two subsystems.

Proposition 6 [46]: Consider a controlled Markov chain with a unique probability distribution for each control policy π (A3) for the entire system and another one for each subsystem. Then the stationary probability of the entire system, β^{π}, can be expressed as the Kronecker product of each stationary probability of each corresponding subsystem i, $\beta^{\pi}_{(i)}$, $i = 1, \ldots, N$, i.e.,

$$\beta^{\pi} = \beta^{\pi}_{(1)} \otimes \beta^{\pi}_{(2)} \otimes \cdots \otimes \beta^{\pi}_{(N)}. \tag{45}$$

B. A System With Subsystems of a Minor Group

We consider a system of two interactive subsystems of a minor group [46], illustrated in Fig. 2.

Each subsystem has two states, i.e., $\mathcal{S}_{(i)} = \{1, 2\}$, and two control actions $\mathcal{U}_{(i)} = \{a, b\}$. Thus, the system has four states $\mathcal{S} = \{1, 2, 3, 4\} = \left\{ \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix} \right\}$, and there are sixteen control policies. The transition probability matrices associated with the control policies for the first subsystem are $\mathbb{P}^{\pi^1}_{(1)} = \mathbb{P}^{\pi^2}_{(1)} = \mathbb{P}^{\pi^3}_{(1)} = \mathbb{P}^{\pi^4}_{(1)} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$, $\mathbb{P}^{\pi^5}_{(1)} = \mathbb{P}^{\pi^6}_{(1)} = \mathbb{P}^{\pi^7}_{(1)} = \mathbb{P}^{\pi^8}_{(1)} = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$, $\mathbb{P}^{\pi^9}_{(1)} = \mathbb{P}^{\pi^{10}}_{(1)} = \mathbb{P}^{\pi^{11}}_{(1)} = \mathbb{P}^{\pi^{12}}_{(1)} = \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}$, and $\mathbb{P}^{\pi^{13}}_{(1)} = \mathbb{P}^{\pi^{14}}_{(1)} = \mathbb{P}^{\pi^{15}}_{(1)} = \mathbb{P}^{\pi^{16}}_{(1)} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$. Similarly, the transition probability matrices for the second subsystem are $\mathbb{P}^{\pi^1}_{(2)} = \mathbb{P}^{\pi^5}_{(2)} = \mathbb{P}^{\pi^9}_{(2)} = \mathbb{P}^{\pi^{13}}_{(2)} = \begin{bmatrix} 0.5 & 0.5 \\ 0.45 & 0.55 \end{bmatrix}$, $\mathbb{P}^{\pi^2}_{(2)} = \mathbb{P}^{\pi^6}_{(2)} = \mathbb{P}^{\pi^{10}}_{(2)} = \mathbb{P}^{\pi^{14}}_{(2)} = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix}$, $\mathbb{P}^{\pi^3}_{(2)} = \mathbb{P}^{\pi^7}_{(2)} = \mathbb{P}^{\pi^{11}}_{(2)} = \mathbb{P}^{\pi^{15}}_{(2)} = \begin{bmatrix} 0.6 & 0.4 \\ 0.45 & 0.55 \end{bmatrix}$, and $\mathbb{P}^{\pi^4}_{(2)} = \mathbb{P}^{\pi^8}_{(2)} = \mathbb{P}^{\pi^{12}}_{(2)} = \mathbb{P}^{\pi^{16}}_{(2)} = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}$.

The output for each subsystem with respect to each control policy is given by four 2×2 matrices as we have two states and two actions for each subsystem. For the first subsystem corresponding to each control policy the output is given: $Y^{\pi^1}_{t(1)} = Y^{\pi^2}_{t(1)} = Y^{\pi^3}_{t(1)} = Y^{\pi^4}_{t(1)} = \begin{bmatrix} 4.8 & 4.0 \\ 5.6 & 9.6 \end{bmatrix}$, $Y^{\pi^5}_{t(1)} = Y^{\pi^6}_{t(1)} = Y^{\pi^7}_{t(1)} = Y^{\pi^8}_{t(1)} = \begin{bmatrix} 4.8 & 4.0 \\ 11.2 & 10.4 \end{bmatrix}$, $Y^{\pi^9}_{t(1)} = Y^{\pi^{10}}_{t(1)} = Y^{\pi^{11}}_{t(1)} = Y^{\pi^{12}}_{t(1)} = \begin{bmatrix} 8.0 & 6.4 \\ 5.6 & 9.6 \end{bmatrix}$, and $Y_{t(1)}^{\pi^{13}} = Y_{t(1)}^{\pi^{14}} = Y_{t(1)}^{\pi^{15}} = Y_{t(1)}^{\pi^{16}} = \begin{bmatrix} 8.0 & 6.4 \\ 11.2 & 10.4 \end{bmatrix}$. The output of the second subsystem with respect to each control policy is $Y_{t(2)}^{\pi^1} = Y_{t(2)}^{\pi^5} = Y_{t(2)}^{\pi^9} = Y_{t(2)}^{\pi^{13}} = \begin{bmatrix} 4.9 & 4.2 \\ 6.3 & 7.0 \end{bmatrix}$, $Y_{t(2)}^{\pi^2} = Y_{t(2)}^{\pi^6} = Y_{t(2)}^{\pi^{10}} = Y_{t(2)}^{\pi^{14}} = \begin{bmatrix} 4.9 & 4.2 \\ 7.7 & 9.8 \end{bmatrix}$, $Y_{t(2)}^{\pi^3} = Y_{t(2)}^{\pi^7} = Y_{t(2)}^{\pi^{11}} = Y_{t(2)}^{\pi^{15}} = \begin{bmatrix} 6.3 & 8.4 \\ 6.3 & 7.0 \end{bmatrix}$, and $Y_{t(2)}^{\pi^4} = Y_{t(2)}^{\pi^8} = Y_{t(2)}^{\pi^{12}} = Y_{t(2)}^{\pi^{16}} = \begin{bmatrix} 6.3 & 8.4 \\ 7.7 & 9.8 \end{bmatrix}$.

We assume that 25% of the subsystem's output goes to subsystem 2, i.e., $Z_t^{(12)} = 0.25 \cdot Y_{t(1)}$ and also 43% percent of the subsystem's output goes to subsystem 1, i.e., $Z_t^{(21)} = 0.43 \cdot Y_{t(2)}$. The input for each subsystem is $W_{t(1)} = 15$ and $W_{t(2)} = 16$ respectively. Furthermore, we assume that the transition cost for each subsystem is given by

$$c_{t(1)}(X_{t(1)}, U_{t(1)}) = \frac{W_{t(1)} + Z_t^{(21)}}{Y_{t(1)} + Z_t^{(12)}} \tag{46}$$

$$c_{t(2)}(X_{t(2)}, U_{t(2)}) = \frac{W_{t(2)} + Z_t^{(12)}}{Y_{t(2)} + Z_t^{(12)}} \tag{47}$$

respectively. The transition cost for the entire system is given by

$$c_t(X_{t(1:2)}, U_{t(1:2)}) = \frac{W_{t(1)} + W_{t(2)}}{Y_{t(1)} + Y_{t(2)}}. \tag{48}$$

The transition cost matrix for each subsystem and for the entire system is a 4 × 4 matrix since we have four states in total (two for each subsystem), and the cost depends on each state and control action. For example, if we want to compute the transition cost matrices for each subsystem, $\mathbb{C}_{(1)}^{\pi^1}, \mathbb{C}_{(2)}^{\pi^1}$, and for the system, $\mathbb{C}^{\pi^1}$, when the system operates under the control policy π^1, substituting $W_{t(1)}, W_{t(2)}, Y_{t(1)}^{\pi^1}, Z_t^{(12)}, Y_{t(2)}^{\pi^1}, Z_t^{(21)}$ in (46), (47), and (48) yields

$$\mathbb{C}_{(1)}^{\pi^1} = \begin{bmatrix} 2.85 & 2.80 & 3.42 & 3.36 \\ 2.95 & 3.00 & 3.54 & 3.60 \\ 2.44 & 2.40 & 1.43 & 1.40 \\ 2.53 & 2.57 & 1.48 & 1.50 \end{bmatrix}$$

$$\mathbb{C}_{(2)}^{\pi^1} = \begin{bmatrix} 2.45 & 2.87 & 2.43 & 2.83 \\ 1.56 & 1.23 & 1.55 & 1.21 \\ 2.69 & 3.13 & 2.66 & 3.10 \\ 1.71 & 1.34 & 1.69 & 1.33 \end{bmatrix}, \text{ and}$$

$$\mathbb{C}^{\pi^1} = \begin{bmatrix} 3.19 & 3.44 & 3.48 & 3.78 \\ 2.48 & 2.12 & 2.64 & 2.24 \\ 1.92 & 2.01 & 2.02 & 2.12 \\ 1.64 & 1.47 & 1.71 & 1.53 \end{bmatrix}.$$

The entry (3, 2) in the transition cost matrices $\mathbb{C}_{(1)}^{\pi^1}, \mathbb{C}_{(2)}^{\pi^1}$, and $\mathbb{C}^{\pi^1}$ corresponds to the costs incurred when the subsystem 1 resides at state 2 and transits to state 1 while the subsystem 2 resides at state 1 and transits to state 2 following the control policy π^1.

TABLE I
LONG-RUN AVERAGE COSTS FOR SUBSYSTEM 1

2.5602	2.6712	2.6390	2.7255
2.0249	2.1127	2.0872	2.1556
1.8029	1.8811	1.8584	1.9193
1.6317	1.7025	1.6820	1.7371

TABLE II
LONG-RUN AVERAGE COSTS FOR SUBSYSTEM 2

2.2511	1.8617	1.7921	1.5235
2.3194	1.9182	1.8464	1.5697
2.3102	1.9106	1.8391	1.5634
2.3383	1.9338	1.8615	1.5825

Similar to the cost matrix, the transition probability matrix is also a 4 × 4 for the four states. When the system operates under the control policy π^1, the transition probability matrix is given from Proposition 5, i.e., $\mathbb{P}^{\pi^1} = \mathbb{P}_{(1)}^{\pi^1} \otimes \mathbb{P}_{(2)}^{\pi^1}$. Therefore

$$\mathbb{P}^{\pi^1} = \begin{bmatrix} 0.35 & 0.35 & 0.15 & 0.15 \\ 0.315 & 0.385 & 0.135 & 0.165 \\ 0.2 & 0.2 & 0.3 & 0.3 \\ 0.18 & 0.22 & 0.27 & 0.33 \end{bmatrix}.$$

The one-stage expected cost, $k^\pi(X_{t(1:2)}, U_{t(1:2)})$, of each subsystem i is a 4 × 1 vector, and the value of the element m is computed as follows:

$$k_{t(i)}^\pi(X_{t(1:2)}, U_{t(1:2)}) = \sum_{k=1}^{4} [\mathbb{P}^\pi]_{\mathrm{mk}} \cdot [\mathbb{C}^\pi]_{\mathrm{mk}}. \tag{49}$$

For example, to compute the one-stage expected cost for subsystem 1 following the control policy π^1, we have:

$$k_{(1)}^{\pi^1}(X_{t(1:2)}, U_{t(1:2)}) = \begin{bmatrix} \sum_{k=1}^4 [\mathbb{P}^{\pi^1}]_{1k} [\mathbb{C}_{(1)}^{\pi^1}]_{1k} \\ \sum_{k=1}^4 [\mathbb{P}^{\pi^1}]_{2k} [\mathbb{C}_{(1)}^{\pi^1}]_{2k} \\ \sum_{k=1}^4 [\mathbb{P}^{\pi}]_{3k} [\mathbb{C}_{(1)}^{\pi^1}]_{3k} \\ \sum_{k=1}^4 [\mathbb{P}^{\pi^1}]_{4k} [\mathbb{C}_{(1)}^{\pi^1}]_{4k} \end{bmatrix} = \begin{bmatrix} 2.9945 \\ 3.1562 \\ 1.8170 \\ 1.9154 \end{bmatrix}. \tag{50}$$

The stationary probability distribution is given by (45). For example, the stationary distribution imposed by the control policy π^1, is $\beta^{\pi^1} = \beta_{(1)}^{\pi^1} \otimes \beta_{(2)}^{\pi^1} = [0.2707 \;\; 0.3008 \;\; 0.2030 \;\; 0.2256]$. Hence the average cost of subsystem 1 with respect to policy π^1 is given by (8), $J(\pi) = \beta^\pi \cdot k_{(1)}^{\pi^1} = 2.5602$. In a similar way we can compute the corresponding one-stage cost vectors and probability distributions for the subsystems 1, 2, and the entire system for all 16 control policies. The average costs for the subsystems and the system corresponding to each control policy are summarized in Tables I, II, and III. Each value in the table (reading the table row by row) corresponds to the long-run expected average cost for the control policies from π^1 to π^{16}. We note that subsystem 1 reaches its minimum average cost J_1 when the policy π^{13} is used. For subsystem 2, the optimal cost is attained with the policy π^4. Finally, for the entire system optimality occurs under the control policy π^{16} which is the Pareto control policy as it corresponds to the Pareto efficiency one-stage expected cost for each subsystem.

TABLE III
LONG-RUN AVERAGE COSTS FOR ENTIRE SYSTEM

2.7557	2.4427	2.4607	2.2307
2.3801	2.1328	2.1522	1.9695
2.3178	2.0876	2.1108	1.9398
2.1821	1.9746	1.9977	1.8431

C. A System With Subsystems of a Minor Group With Varying Transition Probability and Cost Matrices

In this example we use synthetic data to examine the Pareto control policy of the systems. First, we use DP to compute the optimal control policy, denoted by π^*, that minimizes the average cost of the entire system. We anticipate that the Pareto control policy π^0 will yield the same result.

Let the subsystems' inputs be $W_{t(1)} = 15$, $W_{t(2)} = 16$ as in the previous example. Next, a random output of each subsystem is considered. The total output of the first subsystem 1 associated with action, a, is a matrix with random entries distributed according to a uniform distribution, $Y(1,3)$. Similarly, the total output of the same subsystem with respect to action b, is a matrix with entries distributed according to $Y(8,10)$. For the second subsystem, the entries of the matrix associated with action a are independent and identically distributed (i.i.d.) $Y(2,4)$, and the ones associated with action b i.i.d. $Y(9,12)$.

Next, let $\rho^* \doteq \rho(\pi^*)$ be the map defined as

$$\rho^\pi = \|f - f^s\| \tag{51}$$

where

$$f = \left(k^\pi_{t(1)}(X_{t(1:2)}, U_{t(1:2)}), k^\pi_{t(2)}(X_{t(1:2)}, U_{t(1:2)})\right) \tag{52}$$

$$f^s = \left(\min_{U_{t(1:2)} \in \mathcal{C}(x_{(1:2)})} k^\pi_{t(1)}(X_{t(1:2)}, U_{t(1:2)}),\right.$$
$$\left.\min_{U_{t(1:2)} \in \mathcal{C}(x_{(1:2)})} k^\pi_{t(2)}(X_{t(1:2)}, U_{t(1:2)})\right). \tag{53}$$

We perform 1000 replications and we observe in Fig. 3 that the absolute difference between $\rho(\pi^0)$ and $\rho(\pi^*)$ is zero. This indicates that π^0 yields in fact the strong Pareto solution for the one-stage expected costs. Furthermore, Fig. 3 shows that $\min_{\pi \in \Pi} \rho(\pi) = \rho(\pi^0)$, where π^0 is such that $J^* = J(\pi^0)$. Hence, based on these synthetic data, one can conclude that the optimal control policy of the entire system is the Pareto control policy.

D. Power Management Control of a Hybrid Electric Vehicle: A System With Subsystems of a Principal Group

The results presented here have been used in the problem of optimizing online the power management control in a HEV [47] consisting of subsystems of a principal group. The Pareto control policy was validated through simulation and it was compared with the control policy derived offline by DP using the long-run expected average cost. Both control policies achieved the same cumulative fuel consumption as illustrated in Fig. 4, demonstrating that the Pareto control policy is the optimal control policy with respect to the average cost criterion and can be implemented online. This work has been extended [48] by considering the battery in the problem formulation in addition to the engine's and motor's efficiency that can provide insights on how to prioritize these objectives based on consumers' needs and preferences.

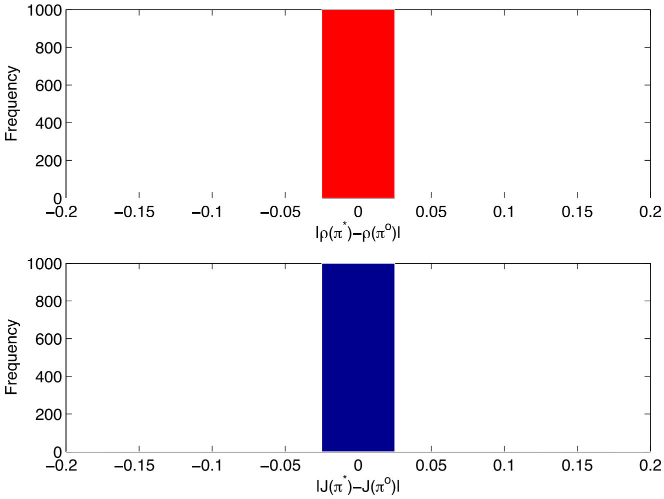

Fig. 3. Histograms of the difference between the average costs corresponding to the optimal and Pareto control policies π^* and π^o, respectively.

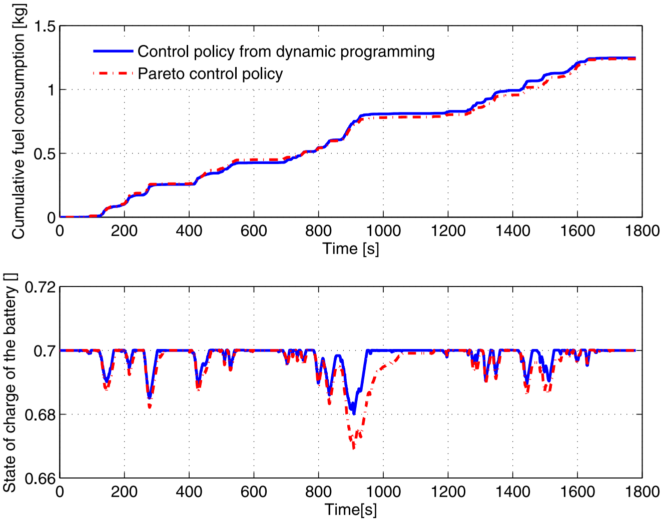

Fig. 4. Cumulative fuel consumption and state of charge of the battery for a parallel hybrid electric vehicle using the control policy derived from dynamic programming and the Pareto control policy over the city-suburban heavy duty vehicle route driving cycle [47].

VI. CONCLUDING REMARKS

In this paper, we established a framework for the analysis and stochastic optimization of complex systems consisting of interactive subsystems. We formulated the stochastic control problem as a multiobjective optimization problem of the one-stage expected costs of the subsystems and developed a duality framework to prove that the Pareto control policy minimizes the long-run expected average cost criterion of the system. We provided a geometric interpretation of the solution and conditions for its existence. The Pareto control policy identifies an equilibrium operating point among the subsystem. If the system operates at this equilibrium, then the long-run expected average cost per unit time is minimized. For practical situations with constraints consistent to those studied here, our results imply that the Pareto control policy may be of value when we seek to derive online the optimal control policy in complex systems.

One potential extension of this work could be to investigate whether a similar analysis can yield the desired *emergence* in a complex system from a decentralized perspective. Emergence refers to the spontaneous creation of order and functionality from the bottom up. Wherever we see complex systems in the physical world, we see emergent patterns at every level, both in structure and functionality. Emergence occurs without a central planner, from the bottom up, based on the interaction of the individual entities in a system. As a simple example from the natural world of how emergence arises, we can consider the flying patterns created by a flock of birds following three simple rules: 1) stay close but don't bomb into birds around me; 2) fly as fast as birds near me; and 3) move towards the center of the group. The fact that a rule applied locally leads to a macro-level property is what is meant by the term bottom up. Another example of a bottom-up emergent phenomenon is the traffic jam resulting from a specific sequence of vehicle-to-vehicle and vehicle-to-infrastructure interactions. If we could develop the framework to characterize emergence, then we would be able to designate the rules for the interactions of the individual subsystems so that the desired emergent phenomena would occur.

Acknowledgment

The author would like to thank Y. Shen for her assistance in running the simulation of the first illustrative example (Section V-B) with varying transition probability and cost matrices. The support of the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory is gratefully acknowledged.

References

[1] J. H. Miller and S. E. Page, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, NJ, USA: Princeton Univ. Press, Mar. 2007.

[2] F. Alexander, M. Anitescu, J. Bell, D. Brown, M. Ferris, M. Luskin, S. Mehrotra, B. Moser, A. Pinar, A. Tartakovsky, K. Willcox, S. Wrigh, and V. Zavala, "A multifaceted mathematical approach for complex systems," DOE Workshop on Mathematics for the Analysis, Simulation, Optimization of Complex Systems, Department of Energy (DOE), Office of Science, Tech. Rep., 2011.

[3] A. A. Malikopoulos, "Supervisory power management control algorithms for hybrid electric vehicles: A survey," *IEEE Trans. Intell. Transport. Syst.*, vol. 15, no. 5, pp. 1869–1885, Oct. 2014.

[4] P. R. Prasanna and A. Rathore, "Analysis, design, experimental results of a novel soft-switching snubberless current-fed half-bridge front-end converter-based pv inverter," *IEEE Trans. Power Electron.*, vol. 28, no. 7, pp. 3219–3230, 2013.

[5] A. Arapostathis, V. Borkar, E. Fernandez-Gaucherand, M. K. Ghosh, and S. I. Marcus, "Discrete-time controlled Markov processes with average cost criterion: A survey," *SIAM J. Control Optim.*, vol. 31, no. 2, pp. 282–344, 1993.

[6] P. Varaiya, "Optimal and suboptimal stationary controls for Markov chains," *IEEE Trans. Autom. Control*, vol. AC-23, no. 3, pp. 388–394, 1978.

[7] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete-Time Case*, 1st ed. New York, NY, USA: Athena Scientific, Feb. 2007.

[8] H. J. Kushner, *Introduction to Stochastic Control*. Austin, TX, USA: Holt, Rinehart and Winston, 1971.

[9] P. R. Kumar and P. Varaiya, *Stochastic Systems*. Upper Saddle River, NJ, USA: Prentice–Hall, Jun. 1986.

[10] R. A. Howard, *Dynamic Programming and Markov Processes*. Cambridge, MA, USA: The MIT Press, Jun. 1960.

[11] J. L. Doob, *Stochastic Processes*. New York, NY, USA: Wiley-Interscience, Jan. 1990.

[12] A. A. Malikopoulos, "Equilibrium control policies for Markov chains," in *Proc. IEEE Conf. Decision Control Eur. Control Conf.*, Orlando, FL, USA, Dec. 12–14, 2011, pp. 7093–7098.

[13] R. Bellman, Ed., *Dynamic Programming*. Princeton, NJ, USA: Princeton Univ. Press, 1957.

[14] D. J. White, "Dynamic programming, Markov chains, the method of successive approximations," *J. Mathemat. Anal. Appl.*, vol. 6, no. 3, pp. 373–376, 1963.

[15] J. Bather, "Optimal decision procedures for finite Markov chains. Part I. Examples," *Advan. Appl. Probabil.*, vol. 5, no. 2, pp. 328–339, 1973.

[16] J. Bather, "Optimal decision procedures for finite Markov chains. Part II: Communicating systems," *Advan. Appl. Probabil.*, vol. 5, no. 3, pp. 521–540, 1973.

[17] J. Bather, "Optimal decision procedures for finite Markov chains. Part III. General convex systems," *Advan. Appl. Probabil.*, vol. 5, no. 3, pp. 541–553, 1973.

[18] G. Hübner, "On the fixed points of the optimal reward operator in stochastic dynamic programming with discount factor greater than one," *ZAMM—J. Appl. Mathemat. and Mechan./Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 57, no. 8, pp. 477–480, 1977.

[19] A. Federgruen, P. J. Schweitzer, and H. C. Tijms, "Contraction mappings underlying undiscounted Markov decision problems," *J. Mathemat. Anal. Appl.*, vol. 65, no. 3, pp. 711–730, 1978.

[20] D. P. Bertsekas, "A new value iteration method for the average cost dynamic programming problem," *SIAM J. Control Optim.*, vol. 36, no. 2, pp. 742–759, 1998.

[21] A. S. Manne, "Linear programming and sequential decisions," *Manage Sci.*, vol. 6, no. 3, pp. 259–267, 1960.

[22] H. M. Wagner, "On the optimality of pure strategies," *Manage Sci.*, vol. 6, no. 3, pp. 268–269, 1960.

[23] C. Derman, "On sequential decisions and Markov chains," *Manage Sci.*, vol. 9, no. 1, pp. 16–24, 1962.

[24] C. Derman and M. Klein, "Some remarks on finite horizon Markovian decision models," *Oper. Res.*, vol. 13, no. 2, pp. 272–278, 1965.

[25] A. Hordijk and L. C. M. Kallenberg, "Linear programming and Markov decision chains," *Manage Sci.*, vol. 25, no. 4, pp. 352–362, 1979.

[26] A. Hordijk and L. C. M. Kallenberg, "Constrained undiscounted stochastic dynamic programming," *Mathemat. Oper. Res.*, vol. 9, no. 2, pp. 276–289, 1984.

[27] K. W. Ross and R. Varadarajan, "Markov decision processes with sample path constraints: The communicating case," *Oper. Res.*, vol. 37, no. 5, pp. 780–790, 1989.

[28] K. W. Ross, "Randomized and past-dependent policies for Markov decision processes with multiple constraints," *Oper. Res.*, vol. 37, no. 3, pp. 474–477, 1989.

[29] J. B. Lasserre, "Detecting optimal and non-optimal actions in average-cost Markov decision processes," *J. Appl. Probabil.*, vol. 31, no. 4, pp. 979–990, 1994.

[30] A. Zadorojniy and A. Shwartz, "Robustness of policies in constrained Markov decision processes," *IEEE Trans. Autom. Control*, vol. 51, no. 4, pp. 635–638, Apr. 2006.

[31] B. L. Miller and A. F. Veinott Jr., "Discrete dynamic programming with a small interest rate," *Ann. Mathemat. Statist.*, vol. 40, no. 2, pp. 366–370, 1969.

[32] H. S. Chang, "A policy improvement method for constrained average Markov decision processes," *Oper. Res. Lett.*, vol. 35, no. 4, pp. 434–438, 2007.

[33] B. F. Lamond and M. L. Puterman, "Generalized inverses in discrete time Markov decision processes," *SIAM J. Matrix Anal. and Applic.*, vol. 10, no. 1, pp. 118–134, 1989.

[34] M. K. Ghosh, "Markov decision processes with multiple costs," *Oper. Res. Lett.*, vol. 9, no. 4, pp. 257–260, 1990.

[35] Z. Ren and B. H. Krogh, "Adaptive control of Markov chains with average cost," *IEEE Trans. Autom. Control*, vol. 46, no. 4, pp. 613–617, Apr. 2001.

[36] J. Abounadi, D. Bertsekas, and V. S. Borkar, "Learning algorithms for Markov decision processes with average cost," *SIAM J. Control Optim.*, vol. 40, no. 3, pp. 681–698, 2001.

[37] H. S. Chang, "Decentralized learning in finite Markov chains: Revisited," *IEEE Trans. Autom. Control*, vol. 54, no. 7, pp. 1648–1653, Jul. 2009.

[38] R. Cavazos-Cadena, "Solutions of the average cost optimality equation for finite Markov decision chains: Risk-sensitive and risk-neutral criteria," *Mathemat. Meth. Oper. Res.*, vol. 70, no. 3, pp. 541–566, 2009.

[39] B. Hendrickson and M. Wright, “Mathematical research challenges in optimization of complex systems,” Sandia National Lab. and Courant Inst. of Mathemat. Sci., Department of Energy, Tech. Rep., 2006.

[40] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes* 3rd ed. Oxford, U.K.: Oxford Univ. Press, Aug. 2001.

[41] S. M. Ross, *Stochastic Processes*, 2nd ed. New York, NY, USA: Wiley, Jan. 1995.

[42] M. Ehrgott, *Multicriteria Optimization*, 2nd ed. New York, NY, USA: Springer, 2005.

[43] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. New York, NY, USA: Athena Scientific, Apr. 2003.

[44] S. Searle, G. Casella, and C. McCulloch, *Variance Components*, ser. Probability and Statistics. New York, NY, USA: Wiley, 2006.

[45] R. Horn and C. Johnson, *Topics in Matrix Analysis*, ser. Topics in Matrix Analysis. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[46] A. A. Malikopoulos, V. Maroulas, and J. Xiong, “A multiobjective optimization framework for stochastic control of complex systems,” in *Proc. 2015 Amer. Control Conf.*, 2015, pp. 4263–4268.

[47] A. A. Malikopoulos, “A multiobjective optimization framework for online stochastic optimal control in hybrid electric vehicles,” *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 2, pp. 440–450, Mar. 2016.

[48] M. Shaltout, A. A. Malikopoulos, S. Pannala, and D. Chen, “A consumer-oriented control framework for performance analysis in hybrid electric vehicles,” *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 4, pp. 1451–1464, Apr. 2015.

Andreas A. Malikopoulos (M’06) received the Diploma in mechanical engineering from the National Technical University of Athens, Athens, Greece, in 2000, and the M.S. and Ph.D. degrees from the Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA, in 2004 and 2008, respectively.

He is currently the Deputy Director of the Urban Dynamics Institute and an Alvin M. Weinberg Fellow with the Energy and Transportation Science Division, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN. Before joining ORNL, he was a Senior Researcher with General Motors Global Research and Development, conducting research in the area of stochastic optimization and control of advanced propulsion systems. His research spans several fields, including analysis, optimization, and control of complex systems; decentralized systems; and stochastic scheduling and resource allocation problems. His emphasis is on applications related to energy, transportation, and operations research.