2015 American Control Conference
Palmer House Hilton
July 1-3, 2015. Chicago, IL, USA

A Multiobjective Optimization Framework for Stochastic Control of Complex Systems

Andreas A. Malikopoulos, *Member, IEEE*, Vasileios Maroulas and Jie Xiong

Abstract— **This paper addresses the problem of minimizing the long-run expected average cost of a complex system consisting of subsystems that interact with each other and the environment. We treat the stochastic control problem as a multiobjective optimization problem of the one-stage expected costs of the subsystems, and we show that the control policy yielding the Pareto optimal solution is an optimal control policy that minimizes the average cost criterion for the entire system. For practical situations with constraints consistent to those we study here, our results imply that the Pareto control policy may be of value in deriving online an optimal control policy in complex systems.**

I. Introduction

Complex systems encountered in virtually many engineering applications involve many distinct physical processes. For example, a complete computational model of a large-scale fusion device is a complex system involving issues of fluid dynamics, deformation of solid materials, thermal effects, ablation, fracture, corrosion and aging of materials, radiation, and many other phenomena. The predictability of large-scale models of complex systems is dependent upon correctly taking into account the uncertainty that results from experimental errors and the random nature of the data defining the systems. Uncertainties must be quantified and propagated throughout the models to produce accurate results. A complex network such as the electrical grid is a challenging system incurring uncertainties. Modeling the behavior of such a system, addressing issues of reliability and security, can be onerous, e.g., a system-level model of this complex network at scales ranging from individual power plants to international energy distribution systems that can be used in policy making and risk analysis.

Stochastic optimization of complex dynamic systems is a ubiquitous task in engineering. The problem is formulated as sequential decision-making under uncertainty where an intelligent system, e.g., power grid, or wind plant, is faced with the task to select those actions in several time steps to achieve long-term goals efficiently using the long-run expected average cost per unit time. The average cost criterion for Markov chains with finite state and arbitrary action spaces has been extensively reported in the literature (see, e.g., [1] and references therein). The problem of minimizing the average cost in a controlled Markov chain with a finite state and control space by solving a dual constrained optimization problem was addressed in [2]. It was shown that the control policy that yields higher probability distribution to the states with low cost and lower probability distribution to the states with the high cost is an optimal solution and it is defined as an equilibrium control policy. The average cost criterion in Markov chains with finite state and action spaces is well understood and has been extensively reported in the literature [3], [4], [5], [6], [7].

In this paper, we address the problem of controlling a system of interacting subsystems to minimize the long-run expected average cost per unit time. We propose an optimization framework that can be used online to derive the optimal control policy while the subsystems interacting with their environment. For example, such a system could be a hybrid electric vehicle (HEV) consisting of four main subsystems: 1) an internal combustion engine, 2) a motor, 3) a generator, and 4) a battery. These subsystems interact with each other to deliver the power demanded by the driver (environment) under the imposed physical constraints. Deriving online the control policy that distributes the power demanded by the driver optimally to the engine, motor/generator and battery to maximize the efficiency of the HEV has been the object of intense study since 1998, and it still remains a challenging control problem [8]. In the proposed approach, we treat the stochastic control problem as a multiobjective optimization problem of the one-stage expected costs of the subsystems, and we show that the control policy yielding the Pareto optimal solution for the one-stage costs is an optimal control policy that minimizes the long-run expected average cost criterion.

The remainder of the paper proceeds as follows. In Section II, we introduce our notation and formulate the problem. In Section III, we correlate the stationary probability distribution of each subsystem with the entire system. In Section IV, we provide the solution of the stochastic control problem with the Pareto control policy. Finally, we present an

This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the US Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

This research was supported by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for DOE. Research of VM is supported by the Simons Foundation Grant # 279870. Research of JX is supported by Macao Science and Technology Fund FDCT076/2012/A3 and Multi-Year Research Grants of the University of Macau No.MYRG2014-00015-FST and MYRG2014-00034-FST. This support is gratefully acknowledged.

A.A. Malikopoulos is with the Energy & Transportation Science Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA (phone: 865-946-1529; fax: 865-946-1354; e-mail: andreas@ornl.gov).

V. Maroulas is with the Faculty of the Department of Mathematics, University of Tennessee, Knoxville, TN.

J. Xiong is with the Faculty of the Department of Mathematics, University of Macau, Macau, China.

978-1-4799-8686-6/$31.00 ©2015 AACC

illustrative example in Section V, and we draw concluding remarks in Section VI.

II. Problem Formulation

A. Notation

We denote random variables with upper case letters, their realization with lower case letters, and their space of realizations by script letters, e.g., for a random variable X, x denotes its realization. Subscripts denote time, and subscripts in parenthesis denote a subsystem; for example, $X_{t(i)}$ denote the random variable of the subsystem i at time t, and $x_{(i)}$ its realization. The short hand notation $X_{t(1:N)}$ denotes the vector $\{X_{t(1)}, X_{t(2)}, ..., X_{t(N)}\}$. Superscripts in parenthesis denote the interactions among the subsystems; for example, $Z_t^{(ij)}$ denotes the input to subsystem j from the subsystem i. $\mathbb{P}(\cdot)$ is the transition probability matrix and $\mathbb{E}[\cdot]$ is the corresponding expectation of a random variable. For a control policy π, we use $\mathbb{P}^\pi(\cdot)$ and $\mathbb{E}^\pi[\cdot]$ to denote that the transition probability matrix and expectation depend on the choice of the control policy π. For different control policies of a subsystem, we use $\pi_{(i)}^j$ to denote the j^{th} control policy of the subsystem i.

B. The Model

We consider a system consisting of N subsystems. The subsystems interact with each other and the environment. At time $t, t = 1, 2, ..., T$, the state of each subsystem $i, X_{t(i)}$, takes values in a finite state space $\mathcal{S}_{(i)}$, which is a metric space and $(\mathcal{S}_{(i)}, \mathcal{B}(S_{(i)}))$ denotes the corresponding measurable space, where $\mathcal{B}(S_{(i)})$ is the smallest σ-algebra generated by open sets.

For each subsystem i, we also consider a finite control space $\mathcal{U}_{(i)}$ from which control actions, $U_{t(i)}$, are chosen. We assume $\mathcal{U}_{(i)}$ is a metric space and we denote $(\mathcal{U}_{(i)}, \mathcal{B}(\mathcal{U}_{(i)}))$ the corresponding measurable space. Furthermore, we assume that both $\mathcal{S}_{(i)}$ and $\mathcal{U}_{(i)}$ are compact spaces for each subsystem i.

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

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

where the input from the environment, $W_{t(1:N)}$, is a sequence of independent random variables, independent of the initial state $X_{0(1:N)}$ and takes values in a set $\mathcal{W}$. Furthermore, the system has N observations, each for each subsystem, which are generated according to

$$Y_{t(1:N)} = h(X_{t(1:N)}, V_{t(1:N)}) \tag{2}$$

where the error from the sensors, $V_{t(1:N)}$, is a sequence of independent random variables, $\{V_{t(1:N)}, t = 1, 2, ..., T\}$, independent of the initial state $X_{0(1:N)}$ and $\{W_{t(1:N)}, t = 1, 2, ..., T\}$, and takes values in a set $\mathcal{V}$. The state of the system can be observed.

Each subsystem interacts with each other. The input of the subsystem j from subsystem i, $Z_t^{(ij)}$, is a function of the subsystem i output $Y_{t(i)}$

$$Z_t^{(ij)} = g(Y_{t(i)}). \tag{3}$$

For example, in a HEV, $Z_t^{(ij)}$ corresponds to the amount of power transmitted from the subsystem i to subsystem j, which is a portion of the subsystem's i output.

In our formulation a state-depedent constraint is incorporated; that is, for each realization of the state of the subsystem i, $X_{t(i)} = x_{(i)}$, there is a nonempty set $\mathcal{C}(x_{(i)}) := \{u_{(i)} | X_{t(i)} = x_{(i)}\} \subset \mathcal{U}_{(i)}$ of admissible control actions.

For each subsystem i, we denote the set of admissible state/action pairs

$$\Gamma_{(i)} : = \{(X_{t(i)}, U_{t(i)}) | X_{t(i)} \in \mathcal{S}_{(i)} \text{ and } U_{t(i)} \in \mathcal{C}(x_{(i)})\} \tag{4}$$

such that it is a measurable subset of $(\mathcal{S}_{(i)} \times \mathcal{U}_{(i)})$. For each subsystem i, we define the Borel measurable functions $\mu_{(i)} : \mathcal{S}_{(i)} \to \mathcal{U}_{(i)}$ that map the state space to the control action space defined as the control law. When the subsystem i is at state $X_{t(i)} = x_{(i)}$, the centralized controller chooses action $U_{t(i)}$ according to the control law $U_{t(i)} = \mu_{(i)}(x_{(i)})$. Each sequence of the measurable functions $\mu_{(i)}$ is called a stationary control policy or stationary control strategy

$$\pi_{(i)} : = (\mu_{(i)}(1), \mu_{(i)}(2), ..., \mu_{(i)}(|\mathcal{S}_{(i)}|)) \tag{5}$$

where $|\mathcal{S}_{(i)}|$ is the cardinality of the subsystem's i state space $\mathcal{S}_{(i)}$. Let Π_i denote the set of the collection of the stationary control policies for each subsystem i

$$\Pi_i : = \{\pi_{(i)} | \pi_{(i)} = \{\mu_{(i)}(1), \mu_{(i)}(2), ..., \mu_{(i)}(|\mathcal{S}_{(i)}|)\}. \tag{6}$$

To ensure that the set Π_i is nonempty we assume that the set $\Gamma_{(i)}$ for each subsystem i contains the graph of all Borel measurable functions $(\mu_{(i)}(1), \mu_{(i)}(2), ..., \mu_{(i)}(|\mathcal{S}_{(i)}|))$.

C. The Average Cost Criterion

At time t, an one-stage expected cost $k_{t(i)}^\pi(X_{t(1:N)}, U_{t(1:N)})$ is incurred for each subsystem i that depends on the state, $X_{t(1:N)}$, and control action, $U_{t(1:N)}$, of the entire system. Similarly, an one-stage expected cost $k_t^\pi(X_{t(1:N)}, U_{t(1:N)})$ is incurred for the entire system.

Assumption 2.1: The one-stage expected costs for each subsystem i, $k_{t(i)}^\pi(X_{t(1:N)}, U_{t(1:N)})$ and for the system $k_t^\pi(X_{t(1:N)}, U_{t(1:N)})$ are nonzero, positive and bounded real numbers.

Assumption 2.2: At each state, $X_{t(1:N)}$, and control action, $U_{t(1:N)}$, of the entire system the relationship between the one-stage expected cost of each subsystem i, $k_{t(i)}^\pi(X_{t(1:N)}, U_{t(1:N)})$, and the cost of the system, $k_t^\pi(X_{t(1:N)}, U_{t(1:N)})$ is given by

$$k_{t(i)}^\pi(X_{t(1:N)}, U_{t(1:N)}) = \frac{\lambda_{(i)} \cdot \nu_{(i)}^{\lambda_{(i)}}}{k_t^\pi(X_{t(1:N)}, U_{t(1:N)})^{\lambda_{(i)}}} \tag{7}$$

where $\lambda_{(i)}, \nu_{(i)}$ are two positive real numbers.

Similarly at each state, $X_{t(1:N)}$, and control action, $U_{t(1:N)}$, of the entire system, the relationship between the one-stage expected costs, $k^{\pi}_{t(i)}\big(X_{t(1:N)}, U_{t(1:N)}\big)$ and $k^{\pi}_{t(j)}\big(X_{t(1:N)}, U_{t(1:N)}\big)$, corresponding to any two subsystems i and j, $i \neq j$, is given by

$$k^{\pi}_{t(i)}\big(X_{t(1:N)}, U_{t(1:N)}\big) = \frac{\phi_{(ij)} \cdot \theta_{(ij)}^{\phi_{(ij)}}}{k^{\pi}_{t(j)}\big(X_{t(1:N)}, U_{t(1:N)}\big)^{\phi_{(ij)}}} \quad (8)$$

where $\phi_{(ij)}, \theta_{(ij)}$ are two positive real numbers.

Assumption (2.2) essentially imposes a tradeoff between the one-stage expected costs of the subsystems as well as a tradeoff between the expected costs of each subsystem and the system. These tradeoffs are deemed characteristic in many engineering applications. For example, in HEVs there is a tradeoff between the efficiencies of the subsystems (e.g., engine, motor, generator, and battery) and also a tradeoff between the efficiency of each subsystem and the efficiency of the HEV.

We are concerned with deriving a stationary optimal control policy $\pi = \big(\pi_{(1)}, \pi_{(2)}, ..., \pi_{(N)}\big)$, where $\pi \in \prod_{i=1}^{N} \Pi_i$, to minimize the long-run expected average cost of the system

$$J(\pi) = \lim_{T\to\infty} \frac{1}{T+1} \mathbb{E}^{\pi} \left[\sum_{0}^{T} k^{\pi}_{t}\big(X_{t(1:N)}, U_{t(1:N)}\big) \right]. \quad (9)$$

To guarantee that the limit in (9) exists, we impose the following assumption.

Assumption 2.3: For each stationary control policy π, the Markov chain $\big\{X_{t(1:N)} | t = 1, 2, ...\big\}$ has a single ergodic class.

Namely, for each stationary policy $\pi \in \Pi$, there is a unique probability distribution (row vector) $\beta^{\pi} = \big(\beta(1)^{\pi}, \beta(2)^{\pi}, ..., \beta(|\mathcal{S}|)^{\pi}\big)$, with $\sum_{l=1}^{|\mathcal{S}|} \beta(l)^{\pi} = 1$ [9, p. 227] where $\beta(l)^{\pi}$ denotes to the stationary probability of the state $X_{t(1:N)} = l, l \in \mathbb{N}$, such that $\beta^{\pi} = \beta^{\pi} \cdot \mathbb{P}^{\pi}$, and $|\mathcal{S}|$ denotes the cardinality of the system's state space $\mathcal{S}$. Under Assumption (2.3), it is known [10, p. 175] that

$$\lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} [\mathbb{P}^{\pi}]^{t} = \mathbf{1} \cdot \beta^{\pi}, \quad (10)$$

where $\mathbf{1} = [1, 1, ..., 1]^{T}$. Substituting (10) into (9) 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

$$J(\pi) = \beta^{\pi} \cdot \mathbf{k}^{\pi}, \quad (11)$$

where

$$\mathbf{k}^{\pi} = \bigg(k^{\pi}_{t}\big(1, U_{t(1:N)}\big), \cdots, k^{\pi}_{t}\big(|\mathcal{S}|, U_{t(1:N)}\big) \bigg)^{T}, \quad (12)$$

is the column vector of the entire system's one-stage expected cost and $U_{t(1:N)}$ is the control action as specified by the control law $\mu_{(1:N)}$ of the control policy π.

Consequently, a stationary control policy is optimal if

$$J^{*} = J(\pi) = \inf \{J(\pi) | \pi \in \Pi\}. \quad (13)$$

We are concerned with deriving a stationary optimal control policy that minimizes the long-run expected average cost of the entire system.

III. Preliminary results

In this section, we provide some preliminary results which will be useful for our analysis later on. We begin by recalling the Kronecker product and its properties (see [11], [12]).

Definition 3.1: 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 $A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$.

Assumption 3.2: The stationary probability distribution of each subsystem i depends only on its control policy $\pi_{(i)}$.

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 3.3: Consider N evolving subsystems with corresponding transition probability matrix $\mathbb{P}_{(i)}$, $i = 1, \cdots, N$ defined by $\mathbb{P}_{(i)}(X_{t+1(i)} = x'_{(i)} | X_{t(i)} = x_{(i)}, U_{t(i)} = u_{(i)})$. Furthermore, let consider that each subsystem i operates under the control policy $\pi_{(i)}$. Then the transition probability matrix of the entire system satisfies

$$\mathbb{P}^{\pi} = \mathbb{P}^{\pi(1)}_{(1)} \otimes \mathbb{P}^{\pi(2)}_{(2)} \otimes \cdots \otimes \mathbb{P}^{\pi(N)}_{(N)} \quad (14)$$

where $\pi = (\pi_{(1)}, \cdots, \pi_{(N)})$.

Proof: The transition probability matrix of the entire system is defined as the following conditional probability:

$$\begin{aligned} &\mathbb{P}(X_{t+1(1:N)} = (x'_{(1)}, x'_{(2)}, \cdots, x'_{(N)}) | \\ &\qquad X_{t(1:N)} = (x'_{(1)}, x'_{(2)}, \cdots, x_{N})) \\ &= \mathbb{P}(X_{t+1(1)} = x'_{(1)}, \cdots, X_{t+1(N)} = x'_{(N)} | \\ &\qquad X_{t(1)} = x'_{(1)}, \cdots, X_{t(N)} = x'_{(N)}) \end{aligned}$$

However, the subsystems evolve independently, hence the transition probability matrix equals to $\prod_{i=1}^{N} \mathbb{P}(X_{t+1(i)} = x'_{(i)} | X_{t(i)} = x_{(i)})$ for any arbitrary N-tuples $(x'_{(1)}, x'_{(2)}, \cdots, x'_{(N)})$ and $(x_{(1)}, x_{(2)}, \cdots, x_{(N)})$. Now, let consider a control policy $\pi = (\pi_{(1)}, \cdots, \pi_{(N)})$ for the entire system and the corresponding probability matrices $\mathbb{P}^{\pi(1)}_{(1)}, \cdots, \mathbb{P}^{\pi(N)}_{(N)}$ with respect to corresponding policies. Following the definition of Kronecker product (Definition 3.1) equation (14) is true for $N = 2$. Next, consider $\mathbb{P}^{\pi(N)}_{(N)} = \mathbb{P}^{\pi(N)}_{(N)}(X_{t+1(N)} = x'_{(N)} | X_{t(N)} = x_{(N)})$ for any possible combination of $(x_{(N)}, x'_{(N)})$. Assuming that (14) holds for $N - 1$ such that $\mathbb{P}^{\pi(1:N-1)} \doteq \mathbb{P}^{\pi(1)}_{(1)} \otimes \mathbb{P}^{\pi(2)}_{(2)} \otimes \cdots \otimes \mathbb{P}^{\pi(N-1)}_{(N-1)}$ and by Definition 3.1, we observe that $\mathbb{P}^{\pi} = \mathbb{P}^{\pi(1:N-1)}_{(N-1)} \otimes \mathbb{P}^{\pi(N)}_{(N)}$ (as in the case for $N = 2$). Therefore by induction (14) holds. ■

Proposition 3.4: Consider a controlled Markov chain with a single ergodic class [Assumption (2.3)] for the entire system and a single ergodic class 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, $\beta^{\pi}_{(i)}$, $i = 1, \cdots, N$, i.e.,

$$\beta^{\pi} = \beta^{\pi}_{(1)} \otimes \beta^{\pi}_{(2)} \otimes \cdots \otimes \beta^{\pi}_{(N)}, \quad (15)$$

where $\pi = (\pi_{(1)}, \cdots, \pi_{(N)})$ is the control policy of the entire system and $\pi_{(i)}$ is the corresponding one of subsystem i.

Proof: From Assumption (2.3) for each subsystem and the entire system, we have that

$$\beta^{\pi} = \beta^{\pi}\mathbb{P}^{\pi}, \text{ and} \tag{16}$$

$$\beta_{(i)}^{\pi(i)} = \beta_{(i)}^{\pi(i)}\mathbb{P}_{(i)}^{\pi(i)} \tag{17}$$

where $\mathbb{P}^{\pi}$ and $\mathbb{P}_{(i)}^{\pi(i)}$ are the transition probability matrices for the entire system and for each subsystem i, respectively. Let consider the Kronecker product of subsystems, $\beta_{(1)}^{\pi} \otimes \beta_{(2)}^{\pi} \otimes \cdots \otimes \beta_{(N)}^{\pi}$. Then according to (17) and the properties of Kronecker product [11], it can be written as

$$\begin{aligned}
&\Big[\beta_{(1)}^{\pi(1)}\mathbb{P}_{(1)}^{\pi(1)}\Big] \otimes \Big[\beta_{(2)}^{\pi(2)}\mathbb{P}_{(2)}^{\pi(2)}\Big] \otimes \cdots \otimes \Big[\beta_{(N)}^{\pi(N)}\mathbb{P}_{(N)}^{\pi(N)}\Big] \\
&= \Big[\beta_{(1)}^{\pi(1)} \otimes \beta_{(2)}^{\pi(2)} \otimes \cdots \otimes \beta_{(N)}^{\pi(N)}\Big] \cdot \\
&\quad \Big[\mathbb{P}_{(1)}^{\pi(1)} \otimes \mathbb{P}_{(2)}^{\pi(2)} \otimes \cdots \otimes \mathbb{P}_{(N)}^{\pi(N)}\Big] \\
&= \Big[\beta_{(1)}^{\pi(1)} \otimes \beta_{(2)}^{\pi(2)} \otimes \cdots \otimes \beta_{(N)}^{\pi(N)}\Big] \cdot \mathbb{P}^{\pi}.
\end{aligned}$$

However, the stationary probability is unique and thus the result follows. ■

IV. Pareto Optimal Control Policy

Various methods can be used to solve (9) offline and derive the optimal control policy that minimizes the long-run expected average cost J. In this paper, we seek the theoretical framework that will yield the optimal control policy online while the subsystems interact with each other. In our proposed approach, a centralized controller attempts to establish an equilibrium among the subsystems, which is the Pareto optimal solution of the one-stage expected costs of the subsystems, that minimizes the long-run average cost of the entire system.

Let consider the function $f\colon \mathcal{X} \to \mathbb{R}^n$, $f = \big(f_1(x), f_2(x)), ..., f_N(x)\big)$ and the following multiobjective optimization problem

$$\min_{x} \big(f_1(x), f_2(x), ..., f_N(x)\big) \tag{18}$$
$$\text{s.t. } x \in \mathcal{X}.$$

The result of the optimization problem (18) 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.

Definition 4.1: [13] A solution $x^o \in \mathcal{X}$ is called Pareto optimal if there is no $x \in \mathcal{X}$ such that $f(x) \leq f(x^o)$. If x^o is Pareto optimal $f(x^o)$ is called efficient. If $x^1, x^2 \in \mathcal{X}$ and $f(x^1) < f(x^2)$, we say x^1 dominates x^2 and $f(x^1)$ dominates $f(x^2)$. The set of all Pareto optimal solutions $x^o \in \mathcal{X}$ is the Pareto set, $\mathcal{X}_P$. The set of all efficient points $y = f(x^*) \in \mathcal{Y}$ where $x^* \in \mathcal{X}_P$, is $\mathcal{Y}_e$ the efficient set.

Definition 4.2: Let $\rho : \mathcal{X} \mapsto \mathbb{R}^+$ be the map defined as

$$\rho(x) = \|f(x) - f^I(x)\| \tag{19}$$

where $f^I(x) = (\min_{x\in\mathcal{X}} f_1(x), \cdots, \min_{x\in\mathcal{X}} f_p(x))$. If there exists $x^* \in \mathcal{X}$ such that $\rho(x^*) = \min_{x\in\mathcal{X}} \rho(x) \doteq \rho^*$, then x^* is said to be the strong Pareto optimal solution.

Theorem 4.3: The control policy π^o that at each time,t, yields the dominant Pareto optimal solution of the one-stage expected cost, $k_{t(i)}^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)$, of the subsystems yields also the minimum one-stage expected cost of the entire system, $k_t^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)$.

Proof: Let $\mathbf{k}_s^{\pi}$ be the column vector of the one-stage expected costs of the subsystems,

$$\mathbf{k}_s^{\pi} = \Bigg(k_{t(1)}^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big), k_{t(2)}^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big) \\ ..., k_{t(N)}^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big)\Bigg)^T.$$

Let π^o be the Pareto control policy that yields the dominant Pareto optimal solution among the subsystems. Namely, for each time t the Pareto control policy is the result of the following multiobjective optimization problem

$$\max_{\pi} \mathbf{k}_s^{\pi} \tag{20}$$

Then by Definition (4.1) there is no other control policy $\pi' \in \Pi$ such that $\mathbf{k}_s^{\pi'} > \mathbf{k}_s^{\pi^o}$, and thus from (7), there is no other control policy $\pi' \in \Pi$ such that $k_t^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big) < k_t^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)$.

Suppose that there is a control policy π' such that for the one-stage expected cost of the system we have $k_t^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big) < k_t^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)$. Then from (7) there is a subsystem i that the control policy π' yields

$$k_{t(i)}^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big) = \frac{\lambda_{(i)} \cdot \nu_{(i)}^{\lambda_{(i)}}}{k_t^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big)^{\lambda_{(i)}}} >$$
$$k_{t(i)}^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big) = \frac{\lambda_{(i)} \cdot \nu_{(i)}^{\lambda_{(i)}}}{k_t^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)^{\lambda_{(i)}}},$$

and a subsystem j such that from (8) we have

$$k_{t(i)}^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big) = \frac{\phi_{(ij)} \cdot \theta_{(ij)}^{\phi_{(ij)}}}{k_{t(j)}^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big)^{\phi_{(ij)}}} >$$
$$\frac{\phi_{(ij)} \cdot \theta_{(ij)}^{\phi_{(ij)}}}{k_{t(j)}^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)^{\phi_{(ij)}}} = k_{t(i)}^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big), \tag{21}$$

where $\lambda_{(i)}, \nu_{(i)}, \phi_{(ij)}$, and $\theta_{(ij)}$ are all positive real numbers. Hence $k_{t(j)}^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big) < k_{t(j)}^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)$, and from (7) we have

$$k_{t(j)}^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big) = \frac{\lambda_{(j)} \cdot \nu_{(j)}^{\lambda_{(j)}}}{k_t^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big)^{\lambda_{(j)}}} <$$
$$\frac{\lambda_{(j)} \cdot \nu_{(j)}^{\lambda_{(j)}}}{k_t^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)^{\lambda_{(j)}}} = k_{t(j)}^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big). \tag{22}$$

However, the last equation implies $k_t^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big) > k_t^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)$, which contradicts the hypothesis.

■

Theorem 4.4: The control policy π^o that yields the Pareto optimal solution of the one-stage expected cost between the subsystems is the optimal control policy π^* that minimizes the long-run expected average cost criterion.

Proof: Let π^o be the Pareto control policy that yields the Pareto optimal solution among the subsystems. From Theorem (4.3) we have that for each realization of the state $X_{t(1:N)} = x_{(1:N)}$

$$k_t^{\pi^o}\big(x_{(1:N)}, U_{t(1:N)}\big) < k_t^{\pi'}\big(x_{(1:N)}, U_{t(1:N)}\big), \forall t$$

for any other control policy $\pi' \in \Pi$. Since the system's one-stage cost is bounded by Assumption (2.1), taking the expected average sum from $t = 0$ up to a finite time $T \in \mathbb{N}$ is well-defined. Thus

$$\frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=0}^{T} k_t^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right] < \frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=1}^{T} k_t^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right]. \tag{23}$$

Taking the liminf as T goes to infinity

$$\liminf_{T\to\infty} \frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=0}^{T} k_t^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right] < \liminf_{T\to\infty} \frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=1}^{T} k_t^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right]. \tag{24}$$

From Assumption (2.3) the limit in (24) is well defined, hence for all control policies $\pi' \in \Pi$

$$J(\pi^o) = \lim_{T\to\infty} \frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=0}^{T} k_t^{\pi^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right] < J(\pi') = \lim_{T\to\infty} \frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=1}^{T} k_t^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right]. \tag{25}$$

■

V. EXAMPLE

A. Case with Two Subsystems

We consider a system with two subsystems. Each subsystem has two states, i.e., $\mathcal{S}_i = \{1, 2\}$, and two control actions $\mathcal{U}_i = \{a, b\}$. Thus there are four control policies for each subsystem. For example, for the subsystem 1, we have $\pi^1_{(1)} = \{a, a\}, \pi^2_{(1)} = \{a, b\}, \pi^3_{(1)} = \{b, a\}$, and $\pi^4_{(1)} = \{b, b\}$. The transition probability matrices associated with each control policy $\pi^j_{(1)}$ for the first subsystem are : $\mathbb{P}^{\pi^1_{(1)}}_{(1)} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$, $\mathbb{P}^{\pi^2_{(1)}}_{(1)} = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$, $\mathbb{P}^{\pi^3_{(1)}}_{(1)} = \begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{bmatrix}$, and $\mathbb{P}^{\pi^4_{(1)}}_{(1)} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$. Similarly, the transition probability matrices associated for the second subsystem are: $\mathbb{P}^{\pi^1_{(2)}}_{(2)} = \begin{bmatrix} 0.5 & 0.5 \\ 0.45 & 0.55 \end{bmatrix}$, $\mathbb{P}^{\pi^2_{(2)}}_{(2)} = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{bmatrix}$, $\mathbb{P}^{\pi^3_{(2)}}_{(2)} = \begin{bmatrix} 0.6 & 0.4 \\ 0.45 & 0.55 \end{bmatrix}$, and $\mathbb{P}^{\pi^4_{(2)}}_{(2)} = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}$.

The output for each subsystem is given by four 2×2 matrices as we have two states and two actions for each subsystem. For the fist subsystem corresponding to each control policy the output is given (we use the superscript $\pi^j_{(i)}$ in the output to emphasize the dependency on the control policy: $Y^{\pi^1_{(1)}}_{t(1)} = \begin{bmatrix} 4.8 & 4.0 \\ 5.6 & 9.6 \end{bmatrix}$, $Y^{\pi^2_{(1)}}_{t(1)} = \begin{bmatrix} 4.8 & 4.0 \\ 11.2 & 10.4 \end{bmatrix}$, $Y^{\pi^3_{(1)}}_{t(1)} = \begin{bmatrix} 8.0 & 6.4 \\ 5.6 & 9.6 \end{bmatrix}$, and $Y^{\pi^4_{(1)}}_{t(1)} = \begin{bmatrix} 8.0 & 6.4 \\ 11.2 & 10.4 \end{bmatrix}$. The output of the second subsystem is: $Y^{\pi^1_{(2)}}_{t(2)} = \begin{bmatrix} 4.9 & 4.2 \\ 6.3 & 7.0 \end{bmatrix}$, $Y^{\pi^2_{(2)}}_{t(2)} = \begin{bmatrix} 4.9 & 4.2 \\ 7.7 & 9.8 \end{bmatrix}$, $Y^{\pi^3_{(2)}}_{t(2)} = \begin{bmatrix} 6.3 & 8.4 \\ 6.3 & 7.0 \end{bmatrix}$, and $Y^{\pi^4_{(2)}}_{t(2)} = \begin{bmatrix} 6.3 & 8.4 \\ 7.7 & 9.8 \end{bmatrix}$. We assume that 25% of the subsystem's output goes to subsystem 2, i.e., $Z^{(12)}_t = 0.25 \cdot Y_{t(1)}$; and also 43% percent of the subsystem's output goes to subsystem 1, i.e., $Z^{(21)}_t = 0.43 \cdot Y_{t(2)}$. 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

$$c_{t(1)}\big(X_{t(1)} | X_{t-1(1)}, U_{t-1(1)}\big) = \frac{W_{t-1(1)} + Z^{(21)}_{t-1}}{Y_{t-1(1)} + Z^{(12)}_{t-1}}, \tag{26}$$

and

$$c_{t(2)}\big(X_{t(2)} | X_{t-1(2)}, U_{t-1(2)}\big) = \frac{W_{t-1(2)} + Z^{(12)}_{t-1}}{Y_{t-1(2)} + Z^{(12)}_{t-1}} \tag{27}$$

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

$$c_t\big(X_{t(1:2)} | X_{t-1(1:2)}, U_{t-1(1:2)}\big) = \frac{W_{t-1(1)} + W_{t-1(2)}}{Y_{t-1(1)} + Y_{t-1(2)}}. \tag{28}$$

The transition cost matrix for each subsystem and for entire system is a 4×4 since we have 4 states in total (two for each subsystem) and the cost depends on each other state and control action. Similar to the cost matrix, the transition probability matrix is also a 4×4 for the four states. When the subsystem 1 follows the control policy $\pi^1_{(1)}$ and the subsystem 2 follows the control policy $\pi^1_{(2)}$ the transition probability matrix is given from Proposition 3.3, i.e., $\mathbb{P}^{(\pi^1_{(1)}, \pi^1_{(2)})} = \mathbb{P}^{\pi^1_{(1)}} \otimes \mathbb{P}^{\pi^1_{(2)}}$.

Therefore,

$$\mathbb{P}^{(\pi^1_{(1)}, \pi^1_{(2)})} = \begin{bmatrix} 0.35 & 0.35 & 0.15 & 0.15 \\ 0.315 & 0.385 & 0.135 & 0.165 \\ 0.2 & 0.2 & 0.3 & 0.3 \\ 0.18 & 0.22 & 0.27 & 0.33 \end{bmatrix}.$$

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

$$\mathbf{k}^{(\pi^j_{(1)}, \pi^l_{(2)})}_{t(i)}\big(X_{t(1:2)}, U_{t(1:2)}\big)$$

$$= \sum_{k=1}^{4} \left[[\mathbb{P}^{(\pi^j_{(1)}, \pi^l_{(2)})}]_{1k} \cdot [\mathbb{C}_{(i)}^{(\pi^j_{(1)}, \pi^l_{(2)})}]_{1k} \right]. \qquad (29)$$

For example, to compute the one-stage expected cost for subsystem 1 following the control policy $\pi^1_{(1)}$ when the subsystem 2 follows the control policy $\pi^1_{(2)}$ we have

$$\mathbf{k}_{(1)}^{(\pi^1_{(1)}, \pi^2_{(2)})} \big(X_{t(1:2)}, U_{t(1:2)}\big)$$

$$= \begin{bmatrix} \sum_{k=1}^4 [\mathbb{P}^{(\pi^1_{(1)}, \pi^1_{(2)})}]_{1k} [\mathbb{C}_{(1)}^{(\pi^1_{(1)}, \pi^1_{(2)})}]_{1k} \\ \sum_{k=1}^4 [\mathbb{P}^{(\pi^1_{(1)}, \pi^1_{(2)})}]_{2k} [\mathbb{C}_{(1)}^{(\pi^1_{(1)}, \pi^1_{(2)})}]_{2k} \\ \sum_{k=1}^4 [\mathbb{P}^{(\pi^1_{(1)}, \pi^1_{(2)})}]_{3k} [\mathbb{C}_{(1)}^{(\pi^1_{(1)}, \pi^1_{(2)})}]_{3k} \\ \sum_{k=1}^4 [\mathbb{P}^{(\pi^1_{(1)}, \pi^1_{(2)})}]_{4k} [\mathbb{C}_{(1)}^{(\pi^1_{(1)}, \pi^1_{(2)})}]_{4k} \end{bmatrix} = \begin{bmatrix} 2.9945 \\ 3.1562 \\ 1.8170 \\ 1.9154 \end{bmatrix}. \qquad (30)$$

Thus the correlations between the one-stage costs of the subsystems, $\mathbf{k}_{(1)}^{(\pi^j_{(1)}, \pi^l_{(2)})} \big(X_{t(1:2)}, U_{t(1:2)}\big)$ and $\mathbf{k}_{(2)}^{(\pi^j_{(1)}, \pi^l_{(2)})} \big(X_{t(1:2)}, U_{t(1:2)}\big)$, and between these costs and the cost of the entire system, $\mathbf{k}^{(\pi^j_{(1)}, \pi^l_{(2)})} \big(X_{t(1:2)}, U_{t(1:2)}\big)$, are given by functions of the form of (7) and (8) (Assumption 2.2).

The stationary probability distribution is given by (15). For example, the stationary distribution imposed by the control policy $\pi = \big(\pi^1_{(1)}, \pi^1_{(2)}\big)$, is $\beta^\pi = \beta_{(1)}^{\pi(1)} \otimes \beta_{(2)}^{\pi(2)} = [\ 0.2707 \quad 0.3008 \quad 0.2030 \quad 0.2256\ ]$. Hence the average cost of subsystem 1 with respect to policy $(\pi^1_{(1)}, \pi^1_{(2)})$ is given by (11), $J(\pi) = \beta(\pi) \cdot \mathbf{k}_{(1)}^{(\pi^1_{(1)}, \pi^1_{(2)})} = 2.5602$. 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.

We note that the subsystem 1 reaches its minimum average cost J_1 when the policy $(\pi^4_{(1)}, \pi^1_{(2)})$ is used. For the subsystem 2, the optimal cost is attained with the policy $(\pi^1_{(1)}, \pi^4_{(2)})$. Finally, for the entire system, optimality occurs when policy $(\pi^4_{(1)}, \pi^4_{(2)})$ is imposed. Therefore, both subsystems and the entire system attain optimal average costs at different combinations of policies. Thus to determine the trade-offs of average costs $J_{(1)}, J_{(2)}$ of two subsystems we incorporate the Pareto optimal theory [14]. We note that $J^I_{(1)} = J_{(1)}(\pi^4_{(1)}, \pi^1_{(2)}) = 1.6317$ and $J^I_{(2)}(\pi^1_{(1)}, \pi^4_{(2)}) = 1.5235$. The strong Pareto optimal control policy is $\pi^o = (\pi^4_{(1)}, \pi^4_{(2)})$ and minimizes the long-run expected average cost of the system system.

B. Automotive Control Application

HEVs have attracted considerable attention due to their potential to reduce petroleum consumption and greenhouse gas emissions [15]. The theoretical results presented here has been used in the problem of optimizing online the power management control in HEVs. The effectiveness of the efficiency of the Pareto control policy was validated through simulation and it was compared with the control policy derived offline using dynamic programming for the long-run expected average cost criterion. Both control policies achieved the same cumulative fuel consumption [16], demonstrating that the Pareto control policy minimizes the average cost criterion. This framework has been extended and considered the battery in the problem formulation [17] aiming at enhancing our understanding of the associated tradeoffs among the HEV subsystems, e.g., the engine, the motor, and the battery.

VI. CONCLUDING REMARKS

The results presented here addressed the problem of controlling a system of interacting subsystems to minimize the long-run expected average cost per unit time. We showed that the control policy yielding the Pareto optimal solution for the one-stage expected costs of the subsystems is an optimal control policy that minimizes the average cost criterion of the entire system. For practical situations with constraints consistent to those we studied here, our results imply that the Pareto control policy may be of value in deriving online an optimal control policy, e.g., online optimization of the power management control of HEVs.

REFERENCES

[1] 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 Journal on Control and Optimization*, vol. 31, no. 2, pp. 282–344, 1993.

[2] A. A. Malikopoulos, "Equilibrium Control Policies for Markov Chains," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, Orlando, Florida, December 12-14 2011.

[3] R. A. Howard, *Dynamic Programming and Markov Processes*. The MIT Press, June 1960.

[4] H. J. Kushner, *Introduction to Stochastic Control*. Holt, Rinehart and Winston, 1971.

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

[6] P. R. Kumar and P. Varaiya, *Stochastic systems*. Prentice Hall, June 1986.

[7] J. L. Doob, *Stochastic Processes*. Wiley-Interscience, January 1990.

[8] A. A. Malikopoulos, "Supervisory Power Management Control Algorithms for Hybrid Electric Vehicles: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 1869–1885, 2014.

[9] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford University Press, August 2001.

[10] S. M. Ross, *Stochastic Processes*, 2nd ed. Wiley, January 1995.

[11] S. Searle, G. Casella, and C. McCulloch, *Variance Components*, ser. Wiley Series in Probability And Statistics. Wiley, 2006.

[12] R. Horn and C. Johnson, *Topics in Matrix Analysis*, ser. Topics in Matrix Analysis. Cambridge University Press, 1994.

[13] M. Ehrgott, *Multicriteria Optimization*. Springer, 2nd edition, 2005.

[14] M. K. Ghosh, "Markov decision processes with multiple costs," *Operations Research Letters*, vol. 9, no. 4, pp. 257–260, 1990.

[15] A. A. Malikopoulos, "Stochastic optimal control for series hybrid electric vehicles," in *Proceedings of the 2013 American Control Conference*, 2013, pp. 1189 – 1194.

[16] ——, "A multiobjective optimization framework for online stochastic optimal control in hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, 2015.

[17] M. Shaltout, A. A. Malikopoulos, S. Pannala, and D. Chen, "A consumer-oriented control framework for performance analysis in hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, 2014.