Reprinted from JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS
All Rights Reserved by Academic Press, New York and London

Vol. 112, No. 1, November 15, 1985
Printed in Belgium

Optimal Policies for Controlled Markov Chains with a Constraint

FREDERICK J. BEUTLER AND KEITH W. ROSS*

Computer, Information and Control Engineering Program, The University of Michigan, Ann Arbor, Michigan 48109

Submitted by S. M. Meerkov

The time average reward for a discrete-time controlled Markov process subject to a time-average cost constraint is maximized over the class of al causal policies. Each epoch, a reward depending on the state and action, is earned, and a similarly constituted cost is assessed; the time average of the former is maximized, subject to a hard limit on the time average of the latter. It is assumed that the state space is finite, and the action space compact metric. An accessibility hypothesis makes it possible to utilize a Lagrange multiplier formulation involving the dynamic programming equation, thus reducing the optimization problem to an unconstrained optimization parametrized by the multiplier. The parametrized dynamic programming equation possesses compactness and convergence properties that lead to the following: If the constraint can be satisfied by any causal policy, the supremum over time-average rewards respective to all causal policies is attained by either a simple or a mixed policy; the latter is equivalent to choosing independently at each epoch between two specified simple policies by the throw of a biased coin.

I. INTRODUCTION

Markov systems have frequently served as models for computer-communication networks, production operations, computer operating systems, and macroeconomic system behavior—among many other applications. These models have naturally suggested the potential for optimization engendered by the Markov property. Accordingly, many journal articles, text, and monographs have dealt with stochastic dynamic programming as an optimization technique (see References for a small recent sample) applicable to the Markov setting.

Optimizations over a finite horizon and /or with a discounted cost structure are appropriate to many applications, and are the most easily amenable to analysis; therefore, finite horizon and discounted cost problems appear to predominate in the literature. Nevertheless, systems that perform a large number of operations over a short time span involve

* Current address: University of Pennsylvania, Philadelphia, Pennsylvania 19104.

0022-247X/85 $3.00

negligible discounting over their useful lifespan, while moving into an equilibrium mode relatively quickly. It follows that a long-term average reward criterion is more suitable to the latter type of system.

Communication systems commonly handle a large number of message blocks in a very short period, just as computers typically process a tremendous number of jobs rapidly. At the same time, such systems can often be represented by Markovian queueing networks [11]. These characteristics suggest that a time-average reward be used as being the most relevant for applications to such systems. Indeed, a number of recent works (e.g., [3, 13, 15]) take precisely this approach, applying dynamic stochastic optimization with a time-average reward to a system of queues.

However, there is little literature pertinent to dynamic optimization of systems respective to a time-average reward while subject to global constraints (but see [5, 9, 12, 10]). For instance, there could be absolute limits on the prevalence of system crashes, the average throughput of certain subsystems, the resources suitable for a specialized component, etc.

In this work, we study the dynamic optimization of discrtete-time Markovian systems by Lagrangian multiplier techniques. We assume a finite state space, a compact action space, continuity of probabilities and rewards respective to the actions, plus an accessibility condition. These hypotheses lead to the existence of an optimal policy. The optimal policy is always stationary. It is either non-randomized stationary (i.e., simple) or consist of a mix of two non-randomized policies, equivalent to choosing independently one of two simple policies at each epoch by the toss of a (biased) coin.[1] Moreover, the optimization procedure requires only repeated solutions of the time-average dynamic programming equation (see [1, 14] for solution methodology), so that the optimum policy subject to the global constraint can be found in principle.

The plan of this paper is as follows. In the present section, we shall establish our notation and present the problem statement. The second section discusses the dynamic programming equation (DPE), with special reference to properties required in the remainder of the paper. The Lagrange multiplier and the consequent parametrized version of the DPE are introduced and analyzed in Section III. These results enable us to exhibit the specific form of the optimum policy for the most interesting problem setting; this is the theme of the fourth section. The final section briefly treats cases other than the one analyzed in Section IV.

The basic random process, $\{X_k\}$, is defined on the finite *state space*, $\mathbf{S} = \{0, 1, ..., N\}$. An *action*, designated by a, is a parametrization of the transition law, which is described more fully below. An action belongs to the *action space*, designated by **A**, and assumed throughout to be a fixed compact metric space equipped with the σ-algebra generated by its open sets.

[1] We shall show in a future paper that additional (reasonable) assumptions met in realistic models of queueing networks can assure a non-randomized stationary policy, or even bang-bang control.

To describe the transition law, first let

$$H_n = (X_0, A_0, X_1, A_1, ..., X_n, A_n) \tag{1.1}$$

be a *history* of $\{X_k\}$; such a history consists not only of the variates up to the "present" (i.e., epoch n), but also the corresponding actions A_k taken at the respective epochs. We require that $\{X_k\}$ be a *controlled Markov process* by demanding that (compare [14])

$$P(X_{n+1} = y \mid H_{n-1}, X_n = x; A_n = a) = P(X_{n+1} = y \mid X_n = x; a). \tag{1.2}$$

We call the right side of (1.2) the *law of motion*, which we write as $P_{xy}(a)$. We shall consistently require that the controlled transition probability is continuous in a for all $x, y \in \mathbf{S}$; this hypothesis not only takes care of all measurability requirements, but it is also essential for convergence purposes.

Control of the process appears through the application of a *policy* or *strategy* to the transition function. In general, the policy u in the *policy space* (designated **U**) can be described as $\mathbf{u} = \{u_0, u_1, ...\}$, where u_k is applied at epoch k. Specifically, $u_{k+1}(\cdot \mid H_k, X_k = x)$ is a conditional probability measure in the wide sense (see [6, p. 29*ff.*]) over **A**. It is seen that this notion of policy inherently demands causality.

There are also subspaces of **U** representing less complex policies. A policy is said to be a *stationary policy* **f** in space **F** if it is constituted by a probability measure over **A** that is conditioned only on the preceding state. More precisely, we write $m_{\mathbf{f}}(\cdot, x)$ to represent **f**, and obtain (for instance)

$$P_{xy}(\mathbf{f}) = \int_A P_{xy}(a)\, m_{\mathbf{f}}(da, x). \tag{1.3}$$

It is known (see [14, p. 30]) that application of a stationary policy to a controlled Markov process yields a Markov process with stationary transition probabilities.

A still more restricted class **G** is that of *simple* or *non-randomized stationary* policies. These are obtained by specializing the measure $m_{\mathbf{f}}$ to consist of a single atom. Consequently, **G** can be characterized by a simple mapping, namely, $\mathbf{g}: S \to \mathbf{A}$; hence, $g(x)$ acquires meaning as an element of **A**, and **g** is viewed as a deterministic vector.

Finally, we shall need a mixed policy, whose space is indicated by ${}_m\mathbf{F}$. A mixed policy is merely a stationary policy with atoms of mass q and $1-q$ for each $x \in S$. Such a policy may be written symbolically for convenience $\mathbf{f}_q = q\mathbf{g}_1 + (1-q)\,\mathbf{g}_2$, with $q \in [0, 1]$.

At each epoch, the system earns a *reward* (or *payoff*) that depends on both state and action taken at that time; thus, the reward at epoch k is $C(X_k, A_k)$, where A_k is the action taken at that epoch. It is assumed that $C(x, \cdot)$ is continuous on $\mathbf{A}$. Since C is then a continuous function on a compact set, we may assume that $C: \mathbf{S} \times \mathbf{A} \to \mathbf{R}^+$ is not only bounded, but also non-negative.

The reward just mentioned gives rise to an *average reward* defined by

$$R_x(\mathbf{u}) \triangleq \liminf_n n^{-1} E_{\mathbf{u}} \left[\sum_{k=0}^{n-1} C(X_k, A_k) \mid X_0 = x \right], \tag{1.4}$$

where $E_{\mathbf{u}}$ is the (conditional) expectation when the policy $\mathbf{u}$ is applied to the system. Ordinarily, it is desired to find the $\mathbf{u} \in \mathbf{U}$ that yields the supremum of the average reward (1.4) over $\mathbf{u} \in \mathbf{U}$ for all $x \in S$, if such exists; however, we have yet to consider the supremum only over those policies that satisfy a specified constraint.

To this end, let the system incur a *cost* called $D(\cdot, \cdot)$ with definition and properties entirely analogous to those of the reward. The *average cost* is given by

$$K_x(\mathbf{u}) \triangleq \limsup_n n^{-1} E_{\mathbf{u}} \left[\sum_{k=0}^{n-1} D(X_k, A_k) \mid X_0 = x \right], \tag{1.5}$$

Our *constraint* is on the average cost, in the sense that we require

$$K_x(\mathbf{u}) \leqslant \alpha \tag{1.6}$$

for all x. Then, if $\mathbf{U}_0$ is the subspace of $\mathbf{U}$ on which the constraint (1.6) is satisfied, we shall discuss the attainment of

$$R_x = \sup_{\mathbf{u} \in \mathbf{U}_0} R_x(\mathbf{u}). \tag{1.7}$$

Any policy $\mathbf{u}$ that attains R_x for each x while simultaneously satisfying the constraint (1.6) is termed a *constrained optimal policy*, or more simply, an *optimal policy*.

Specifically, we shall attempt to answer each of these questions as completely as possible:[2]

Question 1.1. What are necessary and sufficient conditions for $\mathbf{U}_0$ to be non-empty?

Question 1.2. When does there exist an optimal policy?

[2] We consider these questions only in light of the accessibility assumption (Hypothesis 2.2).

Question 1.3. If such a policy exists, are there simpler optimal policies, that is, optimal policies belonging to $\mathbf{G}$ or ${}_{\mathrm{m}}\mathbf{F}$?

Question 1.4. If an optimal policy exists, how can it be determined?

For the most part, we shall be able to resolve each of the above questions completely. In the next section, we will establish certain facts on the dynamic programming equation (DPE) which probably have little independent interest, but are needed to analyze the equation as a function of a running parameter. Then, Section III introduces the Lagrangian, and shows how the DPE behaves in terms of the parameter. This leads into Section IV, where the properties of the preceding sections enable one to show the existence of optimal policies belonging to $\mathbf{G}$ and ${}_{\mathrm{m}}\mathbf{F}$. Finally, Section V discusses cases not amenable to the techniques of Section IV.

II. Preliminaries

In this section, we outline a number of properties of the dynamic programming equation (hereafter called DPE) which do not depend on the Lagrangian formulation of the constraint. As indicated in the preceding section, $P_{xy}(a)$ is the transition probability when action a is apllied, and $C(x, a)$ is the reward corresponding to state $x \in \mathbf{S}$ and action $a \in \mathbf{A}$. In terms of this notation, a central result (see [14, 1, 7]) is

THEOREM 2.1. *Suppose there exists a scalar c and a bounded vector $\mathbf{h}$ such that the DPE*

$$c + h(x) = \sup_{a \in \mathbf{A}} \left[C(x, a) + \sum_{y \in S} P_{xy}(a)\, h(y) \right] \tag{2.1}$$

is satisfied for each $x \in S$. Then any policy $\mathbf{g} \in \mathbf{G}$ specified by

$$g(x) = \arg\sup_{a \in \mathbf{A}} \left[C(x, a) + \sum_{y \in S} P_{xy}(a)\, h(y) \right] \tag{2.2}$$

attains

$$J = \sup_{\mathbf{u} \in \mathbf{U}} R_x(\mathbf{u}) \tag{2.3}$$

for all $x \in S$. Moreover, the constant c in (2.1) *satisfies $c = J$.*

It is clear that no solutions to the DPE can exist unless the suprema on the right of (1.2) can be attained. More difficult, however, is the determination of simple sufficiency conditions under which Theorem 2.1 is

applicable. The most transparent of these necessitates that the process $\{X_n\}$ has only one recurrent class, into which it inevitably moves. In fact, for some state—call it 0—one requires for a sufficiency condition the stronger property that T_x, the hitting time for 0 starting from $x \in S$, satisfy

$$\sup_{x \in \mathbf{S}} \sup_{\mathbf{g} \in \mathbf{G}} E_{\mathbf{g}}(T_x) < \infty, \tag{2.4}$$

where $E_{\mathbf{g}}$ denotes the probability measure corresponding to policy $\mathbf{g}$. This standard condition [14, 1] is usually applied to finite action spaces; however, no change in proof is needed to extend it to compact $\mathbf{A}$. If S and $\mathbf{A}$ are of finite cardinality, the accessibility hypothesis below suffices to satisfy (2.3).

HYPOTHESIS 2.2. *For every* $\mathbf{g} \in \mathbf{G}$. *the state* 0 *is accessible from each* $x \in S$. *In fact, weaker assumptions that divide* $\mathbf{S}$ *into fixed recurrent classes are possible, but these complicate the theory without providing additional insights.*

When $\mathbf{A}$ is not finite, as here, it is not a priori obvious that Hypothesis 2.2 implies the validity of (2.4). Accordingly, we must prove that (2.4) continues to hold for compact $\mathbf{A}$. This is turn requires some subsidiary results on convergence and continuity. Let us first define $\mathbf{P}$ as the space of transition matrices parametrized by the elements of $\mathbf{G}$. Also, we shall take, for non-periodic $\mathbf{P}(\mathbf{g})$,

$$\mathbf{P}^*(\mathbf{g}) \triangleq \lim_n [\mathbf{P}(\mathbf{g})]^n; \tag{2.5}$$

the definition could easily be extended to the periodic case by substituting the Cesaro means on the right of (2.5). In terms of the above notation, we now state and prove certain technical results that will imply (2.4) under Hypothesis 2.2

LEMMA 2.3. *If Hypothesis* 2.2 *holds for* $\mathbf{G}$, *it is also true for* $\mathbf{F}$.

Proof. Suppose Hypothesis is true for $\mathbf{G}$, but fails for some $\mathbf{f} \in \mathbf{F}$. Then there is a closed (see [8, p. 384]) set $\mathbf{C} \subset \mathbf{S}$ such that $0 \notin \mathbf{C}$ and

$$\sum_{y \notin \mathbf{C}} P_{xy}(\mathbf{f}) = 0$$

for each $x \in \mathbf{C}$. Consequently, there exists for each such x a $a_x \in \mathbf{A}$ for which

$$\sum_{y \notin \mathbf{C}} P_{xy}(a_x) = 0.$$

Then for any $\mathbf{g}$ with $g(x) = a_x$, the set $\mathbf{C}$ is closed also with respect to this $\mathbf{g}$, which is inconsistent with Hypothesis 2.2

LEMMA 2.4. $\mathbf{F}$ *and* $\mathbf{G}$ *are sequentially compact. Both* $\mathbf{P}(\cdot)$ *and* $\mathbf{P}^*(\cdot)$ *are continuous functions on* $\mathbf{F}$.

Remark. Since $\mathbf{P}$ and $\mathbf{P}^*$ have only a finite number of components, all metrics are equivalent to componentwise convergence on the reals. There are also equivalent metrics on $\mathbf{F}$ such as, for example, that induced by the metric $\tilde{d}$ of $\mathbf{A}$ by

$$d(\mathbf{f}, \mathbf{f}') = \sum_{x \in \mathbf{S}} \tilde{d}(m_{\mathbf{f}}(\cdot, x), m_{\mathbf{f}'}(\cdot, x)).$$

The metric $\tilde{d}$ is determined by the weak convergence of probability measures; see [2]. For $\mathbf{g} \in \mathbf{G}$, the latter is equivalent to the pointwise convergence of each component.

Proof. The first statement follows because $\mathbf{A}$ is compact and $\mathbf{F}$ therefore tight (see [2, Sect. 6 of Chap. 1]). For the second claim, it need only be observed that $\mathbf{G} \subset \mathbf{F}$, where $\mathbf{G}$ is closed because the same is true of $\mathbf{A}$. Since the respective elements of $\mathbf{P}(\cdot)$ are continuous functions of $\mathbf{f}$ by the definition of weak convergence of probability measures, the third assertion is also valid.

Suppose now that $\mathbf{P}^*(\cdot)$ is not continuous respective to $\mathbf{F}$. Then for some $\mathbf{f}_n \to \mathbf{f}_0$ there is a subsequence—which we shall also denote by $\{\mathbf{f}_n\}$ for convenience—such that $\mathbf{P}^*(\mathbf{f}_n) \to \mathbf{P}^*$, the latter not being equal to $\mathbf{P}^*(\mathbf{f}_0)$. But $\mathbf{P}^*(\mathbf{f}_n)\,\mathbf{P}(\mathbf{f}_n) = \mathbf{P}^*(\mathbf{f}_n)$ implies $\mathbf{P}^*\mathbf{P}(\mathbf{f}_0) = \mathbf{P}^*$. On the other hand, the preceding Lemma assures that there is only one recurrent class, whence the latter equation has the unique transition probability solution $\mathbf{P}^*(\mathbf{f}_0)$; thus, the desired contradiction is reached.

With the aid of the Lemma, we now prove

THEOREM 2.5. *Let* $\mathbf{S}$ *be finite and* $\mathbf{A}$ *compact, as previously assumed, and let Hypothesis* 2.2 *apply. Then* (2.4) *is valid, and in fact*

$$\sup_{x \in \mathbf{S}} \sup_{\mathbf{f} \in \mathbf{S}} E_{\mathbf{f}}(T_x) < \infty.$$

Proof. It is enough to show that for arbitrary $x \in \mathbf{S} - \{0\}$

$$\sup_{\mathbf{f} \in \mathbf{F}} P_{\mathbf{f}}(T_x > n) = \beta(x) < 1 \tag{2.6}$$

since this implies

$$\sup_{\mathbf{f} \in \mathbf{F}} P_{\mathbf{f}}(T_x > kn) \leqslant \beta^k$$

where $\beta \triangleq \max[\beta(x)] < 1$. In turn, the latter inequality implies

$$\sup_{\mathbf{f} \in \mathbf{F}} E_{\mathbf{f}}(T_x) \leqslant \frac{n}{1-\beta} < \infty.$$

For the proof of (2.6), observe that

$$P_{\mathbf{f}}(T_x > n) = \sum_{y=1}^{n} \hat{P}^n_{xy}(\mathbf{f}), \tag{2.7}$$

in which $\hat{\mathbf{P}}$ is the matrix obtained from $\mathbf{P}$ by deleting the row and column corresponding to state 0. Suppose then that (2.6) is false. Then there exists a sequence $\{\mathbf{f}_m\} \in \mathbf{F}$ and hence a convergent subsequence (which we shall also call $\{\mathbf{f}_m\}$) such that $\mathbf{f}_m \to \mathbf{f}_0$ (say), and

$$\sum_{y=1}^{n} \hat{P}^n_{xy}(\mathbf{f}_m) > 1 - m^{-1}. \tag{2.8}$$

The continuity result of Lemma 2.4 therefore implies that

$$P_{\mathbf{f}_0}(T_x > n) = 1.$$

Because the cardinality card $\mathbf{S} = N+1$, we obtain $P_{\mathbf{f}_0}(T_x = \infty) = 1$, which is in clear contradiction to Lemma 2.3.

Theorem 2.5 leads to the eventual conclusion that our hypotheses guarantee at least one solution to the DPE. However, the presence of the Lagrange multiplier complicates the situation sufficiently to create a need for further discussion of the DPE. In particular, we must obtain a better understanding of the (vector) $\mathbf{h}$ appearing in the DPE. To this end, we adopt the notation $\mathbf{C}(\mathbf{g})$ for the vector whose components are $C(x, g(x))$, and $\mathbf{e}$ the $(N+1)$-vector all of whose entries are unity. Also, take $\tilde{\mathbf{G}}$ to be the subspace of $\mathbf{G}$ defined by

$$\tilde{\mathbf{G}} = \{\mathbf{g}: \mathbf{g} \in \mathbf{C},\ g(x) \text{ satisfies (2.2) for all } x \in \mathbf{S}\}.$$

Then for any $\mathbf{g} \in \tilde{\mathbf{G}}$

$$J\mathbf{e} = \mathbf{P}^*(\tilde{\mathbf{g}})\,\mathbf{C}(\tilde{\mathbf{g}}); \tag{2.9}$$

this is easily derived from (2.1) by setting the right side equal to its supremum, writing the result in the vector form

$$J\mathbf{e} + \mathbf{h}(\tilde{\mathbf{g}}) = \mathbf{C}(\tilde{\mathbf{g}}) + \mathbf{P}(\tilde{\mathbf{g}})\,\mathbf{h}(\tilde{\mathbf{g}}), \tag{2.10}$$

and premultiplying (2.10) by $\mathbf{P}^*(\tilde{\mathbf{g}})$.[3]

[3] A $\mathbf{g} \in \mathbf{G}$ may satisfy (2.9) and hence attain the supremum without belonging to $\tilde{\mathbf{G}}$. For instance, if the x column of $\mathbf{P}^*(\mathbf{g})$ consists of zeros, the value of $C(x, g(x))$ is irrelevant to J.

The principal Lemma regarding $\mathbf{h}$ is now the following:

LEMMA 2.6. *Let the DPE be satisfied. Then the* $\mathbf{h}$ *appearing in* (2.1) *is unique up to a constant vector. If we require*

$$h_0(\mathbf{g}) = 0, \tag{2.11}$$

$\mathbf{h}$ *is uniquely specified (independently of* $\tilde{\mathbf{g}} \in \tilde{\mathbf{G}}$*) by*

$$\mathbf{h} = \mathbf{A}(\mathbf{I} - \mathbf{P}(\tilde{\mathbf{g}}) + \mathbf{P}^*(\tilde{\mathbf{g}}))^{-1}(\mathbf{I} - \mathbf{P}^*(\tilde{\mathbf{g}}))\,\mathbf{C}(\tilde{\mathbf{g}}). \tag{2.12}$$

Here $\mathbf{I}$ *is the identity matrix, and* $\mathbf{A}$ *denotes the square matrix all of whose elements are zero, except that* $A_{i0} = -1$ *and* $A_{ii} = 1$ *for* $i = 1, 2, \ldots, n$.

Proof. For fixed $\tilde{\mathbf{g}} \in \tilde{\mathbf{G}}$, it is shown in [1, p. 331] that $\mathbf{h}(\tilde{\mathbf{g}})$ takes a form (2.12), but without the premultiplication by $\mathbf{A}$. Furthermore, $\mathbf{h}(\tilde{\mathbf{g}})$ may be uniquely specified up to an arbitrary constant vector, as indicated on page 340 of the same reference. Now $\mathbf{A}$ has the effect of subtracting $h_0(\tilde{\mathbf{g}})$ from each component, so that the $\mathbf{h}(\tilde{\mathbf{g}})$ of (2.12) is valid in the DPE, and moreover satisfies (2.11) also.

Suppose now that $\tilde{\mathbf{g}}$ and $\hat{\mathbf{g}}$ both belong to $\tilde{\mathbf{G}}$ To complete the proof of the Lemma, we must then verify that $\mathbf{h}(\tilde{\mathbf{g}}) = \mathbf{h}(\hat{\mathbf{g}})$ modulo a constant vector. To this end, compare (2.10) with

$$J\mathbf{e} + \mathbf{h}(\hat{\mathbf{g}}) \geqslant \mathbf{C}(\tilde{\mathbf{g}}) + \mathbf{P}(\tilde{\mathbf{g}})\,\mathbf{h}(\hat{\mathbf{g}}), \tag{2.13}$$

where "$\geqslant$" in a vector relation means "$\geqslant$" for each component. Here (2.13) applies because $\hat{\mathbf{g}}$ satisfies (2.2) and (2.10) with $\mathbf{h}(\hat{\mathbf{g}})$. Next, take $\mathbf{v} \triangleq \mathbf{h}(\tilde{\mathbf{g}}) - \mathbf{h}(\tilde{\mathbf{g}})$, so that subtracting (2.10) from (2.13) produces $\mathbf{v} \geqslant \mathbf{P}(\tilde{\mathbf{g}})\,\mathbf{v}$ and therefore also

$$\mathbf{v} \geqslant \mathbf{P}^*(\tilde{\mathbf{g}})\,\mathbf{v}. \tag{2.14a}$$

Interchanging $\tilde{\mathbf{g}}$ and $\hat{\mathbf{g}}$ in the above argument also yields

$$\mathbf{P}^*(\hat{\mathbf{g}})\,\mathbf{v} \geqslant \mathbf{v}. \tag{2.14b}$$

Letting $\pi(\mathbf{g})$ denote any row of $\mathbf{P}^*$, we obtain from (2.14) that

$$\pi(\tilde{\mathbf{g}})\,\mathbf{v} \leqslant \min_{x \in \mathbf{S}}[v(x)], \qquad \max_{x \in \mathbf{S}}[v(x)] \leqslant \pi(\hat{\mathbf{g}})\,\mathbf{v}. \tag{2.15}$$

However, the accessibility hypothesis requires that $\pi_0(\tilde{\mathbf{g}}) > 0$, so that the first inequality of (2.15) implies $v(0) \leqslant \min[v(x)]$; similarly, $\max[v(x)] \leqslant v(0)$. Thus $v(x) = v(0)$ for all $x \in \mathbf{S}$, i.e., $\mathbf{v}$ is a constant vector.

This completes our discussion of the properties of the DPE; while some of these are not needed for the usual dynamic programming considerations, they will turn out to be crucial in the Lagrangian constraint formulation of the constrained optimization problem.

III. Lagrange Formulation for Optimization

A constrained optimization problem can often be reduced to one without constraints through the introduction of parameters called *Lagrange multipliers* (see [4, Chap. IV]). This technique turns out to be useful in the solution of our problem. The methodology again utilizes the DPE, which is now applied to the reward

$$B^\lambda(x, a) \triangleq C(x, a) - \lambda D(x, a), \tag{3.1}$$

in which λ is the Lagrange multiplier.

According to the preceding section, the DPE is solvable for each $\lambda \geqslant 0$ with the *constrained reward* furnished by (3.1). In fact, for each such λ, the supremum J^λ over $\mathbf{u} \in \mathbf{U}$ is attained by at least one $\mathbf{g}^\lambda \in \tilde{\mathbf{G}}^\lambda$, where $\mathbf{g}^\lambda$ satisfies the DPE for parameter λ, and $\tilde{\mathbf{G}}^\lambda$ is the set of all such elements of $\mathbf{G}$.

It is desirable at this point to introduce further notation to clarify the role of the multiplier. We take

$$J_x^\lambda(\mathbf{u}) \triangleq \liminf_n n^{-1} E_{\mathbf{u}}\left[\sum_{k=0}^{n-1} B^\lambda(X_k, A_k) \mid X_0 = x\right].$$

Because of the accessibility hypothesis, rewards and costs are the same for any initial state for each $\mathbf{f} \in \mathbf{F}$, so that no reference to the state need be reflected in the self-explanatory notation $J^\lambda = J^\lambda(\mathbf{g}^\lambda)$. We also write R^λ for $R(\mathbf{g}^\lambda)$ and K^λ for $K(\mathbf{g}^\lambda)$, where it must be recognized that R^λ and K^λ may be multiple valued functions if $\tilde{\mathbf{G}}^\lambda$ is not a singleton. Finally, it makes sense to use $\mathbf{h}^\lambda$ as the unique $\mathbf{h}$ (see (2.12)) occuring in the constrained DPE for parameter λ.

The first results of this section consist of a series of inequalities, namely,

Lemma 3.1. *J^λ, R^λ, and K^λ are all monotone non-increasing in λ.*

Proof. These assertions are all a consequence of the fundamental inequality that reads

$$\begin{aligned} J^{\lambda+\eta}(\mathbf{g}^\lambda) - J^\lambda(\mathbf{g}^\lambda) &\leqslant J^{\lambda+\eta}(\mathbf{g}^{\lambda+\eta}) - J^\lambda(\mathbf{g}^\lambda) \\ &\leqslant J^{\lambda+\eta}(\mathbf{g}^{\lambda+\eta}) - J^\lambda(\mathbf{g}^{\lambda+\eta}) < 0 \end{aligned} \tag{3.2}$$

for any positive $\lambda \geqslant 0$, $\eta > 0$. Then

$$-\eta K^\lambda \leqslant J^{\lambda+\eta} - J^\lambda \leqslant -\eta K^{\lambda+\eta} \leqslant 0. \tag{3.3}$$

This proves all but the last claim. For the latter, assume R^λ is not monotone non-increasing. Then there exist λ, η such that $R^\lambda < R^{\lambda+\eta}$. But $K^\lambda \geqslant K^{\lambda+\eta}$, whence,

$$R^\lambda - \lambda K^\lambda < R^{\lambda+\eta} - \lambda K^{\lambda+\eta}.$$

Consequently, we have the contradiction $J^\lambda < J^\lambda(\mathbf{g}^{\lambda+\eta})$.

A technical result that will soon prove useful is

Lemma 3.2. *J^λ is uniformly absolutely continuous, with*

$$-K^\lambda \leqslant \left(\frac{dJ^\lambda}{d\lambda}\right)^+ \leqslant -\lim_{\eta \downarrow 0} K^{\lambda+\eta}. \tag{3.4}$$

Also, the derivative

$$\frac{dJ^\lambda}{d\lambda} = -K^\lambda \tag{3.5}$$

exists for almost all $\lambda \geqslant 0$.

Proof. From (3.3) and the monotonicity of K^λ, J^λ satisfies the Lipschitz condition

$$|J^{\lambda+\eta} - J^\lambda| \leqslant \eta K^\lambda \leqslant \eta K^0. \tag{3.6}$$

Note that K^0 is bounded (for any $\mathbf{u} \in \mathbf{U}$) because $D(\cdot, \cdot)$ is a continuous function on a compact set.

As for the right derivative (3.4), one need only divide by η in (3.3), observing that the bounded monotone function K^λ possesses limits from the right. Since, moreover, K^λ is continuous almost everywhere, one obtains an equality in (3.4) for almost all λ. And, by the absolute continuity, the right derivative must coincide with the ordinary derivative.

The monotonicity properties just demonstrated are relevant to the Lagrangian use of the constraint. It will be recalled that in the classical use of the Lagrange multiplier technique, the multiplier λ is chosen so that the constraint is met in a fashion consistent with the desired optimization. A similar phenomenon is seen here, although the situation is naturally more complicated. To use this idea, we introduce

$$\gamma \triangleq \inf\{\lambda : K^\lambda \leqslant \alpha\}, \tag{3.7}$$

where the α in (3.7) is the constraint constant mentioned in (1.6). By the monotonicity of K^λ (according to Lemma 3.1), this γ is well defined. Then we have

LEMMA 3.3. *Let*

$$K(\mathbf{g}) < \alpha \tag{3.8}$$

for some $\mathbf{g} \in \mathbf{G}$. *Then* $\gamma < \infty$.

Proof. Suppose that the assertion of the Lemma is false. Then $K^\lambda > \alpha$ for all λ, and consequently $J^\lambda < R^0 - \lambda\alpha$ for $\lambda > 0$. On the other hand, there exists a $\delta > 0$ and a $\mathbf{g}$ for which $K(\mathbf{g}) = \alpha - \delta$. For this $\mathbf{g}$, $J(\mathbf{g}) = R(\mathbf{g}) - \lambda(\alpha - \delta)$. Hence, $J(\mathbf{g}) > J^\lambda$ for all sufficiently large λ, which is clearly a contradiction.

We shall also need some results on continuity and compactness.

LEMMA 3.4. $R(\mathbf{g})$ *and* $K(\mathbf{g})$ *are continuous on* $\mathbf{G}$; $J^\lambda(\mathbf{g})$ *in continuous on* $(R^+ \times \mathbf{G})$.[4]

Proof. The continuity of $C(x, \cdot)$, together with that of $\mathbf{P}^*(\cdot)$ (see Lemma 2.4), implies $\mathbf{P}^*(\mathbf{g}_n)\,\mathbf{C}(\mathbf{g}_n) \to \mathbf{P}^*(\mathbf{g}_0)\,\mathbf{C}(\mathbf{g}_0)$; a similar argument applies to $K(\cdot)$. As for the last claim, note that $J^\lambda(\mathbf{g}) = R(\mathbf{g}) - \lambda K(\mathbf{g})$.

The compactness result requires a certain amount of new notation. It is natural to use $\tilde{\mathbf{G}}^\lambda$ to denote those $\mathbf{g} \in \mathbf{G}$ satisfying the constrained DPE with parameter λ. Then let

$$\tilde{\mathbf{G}}_\eta \triangleq \bigcup_{\lambda \leq \eta} (\lambda, \tilde{\mathbf{G}}^\lambda), \tag{3.9}$$

with the obvious product space topology on $\tilde{\mathbf{G}}_\eta$.

THEOREM 3.5. *For any* η, $\tilde{\mathbf{G}}_\eta$ *is compact.*

Proof. Since $\tilde{\mathbf{G}}_\eta$ is a subspace of $([0, \eta] \times \mathbf{G})$, it is already totally bounded, and we need only to show $\tilde{\mathbf{G}}_\eta$ to be closed. To this end, let $\lambda_n \to \lambda_0$, and assume $\mathbf{g}^{\lambda_n} \to \mathbf{g}_0$, with $\mathbf{g}^{\lambda_n} \in \tilde{\mathbf{G}}^{\lambda_n}$. We must prove that $\mathbf{g}_0 \in \tilde{\mathbf{G}}^{\lambda_0}$.

In the first place, we have

$$J^{\lambda_n}\mathbf{e} + \mathbf{h}^{\lambda_n} = \mathbf{B}^{\lambda_n}(\mathbf{g}^{\lambda_n}) + \mathbf{P}(\mathbf{g}^{\lambda_n})\,\mathbf{h}^{\lambda_n}, \tag{3.10}$$

in which the right side is maximal for every component. Each term in (3.10) converges to the corresponding limit, that is,

$$J^{\lambda_0}\mathbf{e} + \mathbf{h} = \mathbf{B}^{\lambda_0}(\mathbf{g}^{\lambda_0}) + \mathbf{P}(\mathbf{g}^{\lambda_0})\,\mathbf{h}. \tag{3.11}$$

[4] Lemma 3.4 applies equally well to $\mathbf{F}$, but its use in the subsequent portions of the paper are confined to $\mathbf{G}$.

To move from (3.10) to (3.11), we first observe that J and $\mathbf{P}$ are continuous in their arguments, and $\mathbf{B}$ converges with the joint convergence of $\{\lambda_n\}$ and $\{\mathbf{g}^{\lambda_n}\}$ convergence of both sequences.

Next, let us look at the convergence of $\{\mathbf{h}^{\lambda_n}\}$. We make use of the following fact: if the normed linear operators $\{A_n\}$ converge to A_0, where A_0 has a bounded inverse, we have $A_n^{-1} \to A_0^{-1}$. Indeed, we use (2.12), together with the continuity of $\mathbf{P}$ and $\mathbf{P}^*$, and convergence of $\mathbf{B}$. The result is that $\mathbf{h}^{\lambda_n}$ converges to some $\mathbf{h}$ which, together with the other variates, satisfies (3.11).

It only remains to show that the right side of (3.11) is maximal. Fix (any) x, and define for convenience

$$f(n, a) \triangleq C(x, a) - \lambda_n D(x, a) + \sum_{y \in \mathbf{S}} P_{xy}(a)\, h^{\lambda_n}(y). \tag{3.12}$$

Then the x coordinate on the right side of (3.11) reads $\lim_n \sup_a f(n, a)$. However, $f(n, a)$ converges uniformly in n with respect to $\mathbf{A}$. We also see that $f(n, \cdot)$ is uniformly continuous, and note that $\{\mathbf{g}_n(x)\}$ converges. These facts enable us to conclude that

$$\lim_n \sup_{a \in \mathbf{A}} f(n, a) = \sup_{a \in \mathbf{A}} \lim_n f(n, a), \tag{3.13}$$

which shows that the right side of (3.11) is the supremum over $\mathbf{A}$ for each $x \in \mathbf{S}$.

IV. THE OPTIMAL POLICY

The policy space $\mathbf{U}$ can consist only of policies incapable of meeting the constraint (1.6), or be composed solely of policies that do satisfy the constraint. However, the case characterized by the greatest theoretical as well as applied interest for realistic problems is the one described by

> The simple policy that manifests the largest reward J given by (2.3) fails to satisfy the constraint condition, and
>
> There is at least one simple policy that meets the constraint with something to spare, that is, satisfies $K(\mathbf{g}) < \alpha$.

We defer the analysis of the other possible cases until the next section. At this time, we pursue the situation just described via the following more formal assumption:

HYPOTHESIS 4.1. *Let* $\mathbf{g}^0$ *be an unconstrained supremum, i.e.,*

$$\mathbf{g}^0 \triangleq \arg \sup_{\mathbf{g} \in \mathbf{G}} R(\mathbf{g}), \tag{4.1}$$

and assume that

$$K(\mathbf{g}^0) > \alpha \tag{4.2}$$

for every such $\mathbf{g}^0$. *Suppose further there exists a* $\mathbf{g} \in \mathbf{G}$ *such that*

$$K(\mathbf{g}) < \alpha. \tag{4.3}$$

Remark 4.2. An alternative assumption, which is somewhat weaker but leads to the same result, is

$$0 < \gamma < \infty. \tag{4.4}$$

In the present section, we shall assume that Hypothesis 4.1 holds throughout; Section V is devoted to constrained optimizations when this hypothesis fails to hold. The rationale for Hypothesis 4.1 becomes more evident in light of the key optimization criterion, namely,

THEOREM 4.3. *Suppose that for some* $\lambda \geqslant 0$ *and some* $\mathbf{f} \in \mathbf{F}$ *we have*

$$K(\mathbf{f}) = \alpha \qquad \text{and} \qquad J^\lambda(\mathbf{f}) = J^\lambda \tag{4.5}$$

for all $x \in \mathbf{S}$. *Then* $\mathbf{f}$ *solves the constrained optimization problem* (1.6) *and* (1.7).

Proof. From the applicability of the DPE, it follows that $J^\lambda(\mathbf{f}) \geqslant J_x^\lambda(\mathbf{u})$ for all $\mathbf{u} \in \mathbf{U}$ and all $x \in \mathbf{S}$. Therefore

$$R(\mathbf{f}) \geqslant R_x(\mathbf{u}) + \lambda[\alpha - K_x(\mathbf{u})]. \tag{4.6}$$

But the term in brackets in non-negative, since $\mathbf{u} \in \mathbf{U}_0$ requires that (1.7) is satisfied. Consequently, $R(\mathbf{f}) \geqslant R_x(\mathbf{u})$ for each $\mathbf{u} \in \mathbf{U}_0$, and each $x \in \mathbf{S}$.

We now find an $\mathbf{f} \in \mathbf{F}$ meeting the conditions of the Theorem. It will be seen that this $\mathbf{f}$ is actually of even simpler form. In fact, the optimal constrained policy belongs either to $\mathbf{G}$ or to the mixture policy set ${}_{\mathrm{m}}\mathbf{F}$. It is emphasized again that this policy in ${}_{\mathrm{m}}\mathbf{F}$ is actually optimal over the class of *all* causal policies meeting the constraint, that is, attaining the suppremum in (1.7).

THEOREM 4.4. *Under Hypothesis* 4.1, *there exists a constrained optimal policy in* ${}_{\mathrm{m}}\mathbf{F}$.

Proof. We use Theorem 4.3 to verify the existence of an optimal policy as claimed. In the first place, if $K^\lambda = \alpha$ for some λ, any corresponding $\mathbf{g}^\lambda \in \tilde{\mathbf{G}}$ satisfies the conditions of that Theorem, and is therefore the optimal policy.

Suppose no such λ as the above exists. Then, since K^λ is non-increasing, and $\gamma \in (0, \infty)$, we have

$$\lim_{\lambda \uparrow \gamma} K^\lambda = \alpha^0 \qquad \text{and} \qquad \lim_{\lambda \downarrow \gamma} K^\lambda = \alpha_0 \tag{4.7}$$

in which

$$\alpha_0 < \alpha < \alpha^0. \tag{4.8}$$

Let $\{\lambda_n\}$ be a sequence that increases to γ, along which the corresponding $\mathbf{g}^{\lambda_n} \in \tilde{\mathbf{G}}$ converges; this is possible because of the compactness of $\mathbf{G}$. Another compactness property (see Theorem 3.5) assures that $\bar{\mathbf{g}} \in \tilde{\mathbf{G}}^\gamma$, where $\bar{\mathbf{g}} \triangleq \lim \mathbf{g}^{\lambda_n}$. Also, $K(\bar{\mathbf{g}}) = \alpha^0$, from Lemma 3.4. An entirely similar procedure, but for a decreasing sequence, yields $\underline{\mathbf{g}} \in \tilde{\mathbf{G}}^\gamma$, with $K(\underline{\mathbf{g}}) = \alpha_0$.

Take

$$\mathbf{f}_q \triangleq q\underline{\mathbf{g}} + (1-q)\,\bar{\mathbf{g}}. \tag{4.9}$$

Clearly, $\mathbf{f}_q \in {}_{\mathrm{m}}\mathbf{F}$, so that it remains to demonstrate that there exists a $\gamma \in [0, 1]$ such that

$$J^\gamma = J^\gamma(\mathbf{f}_q) \qquad \text{and} \qquad K(\mathbf{f}_q) = \alpha. \tag{4.10}$$

For the first part of (4.10), we turn to the DPE in the vector form. By Lemma 3.4, $J^\gamma = J^\gamma(\underline{\mathbf{g}}) = J^\gamma(\bar{\mathbf{g}})$. Next, $B^\gamma(\mathbf{f}_q) = qB^\gamma(\underline{\mathbf{g}}) + (1-q)B^\gamma(\bar{\mathbf{g}})(\bar{\mathbf{g}})$, by an argument analogous to (1.3). The same reasoning applies to $\mathbf{P}(\mathbf{f}_q) = q\mathbf{P}(\underline{\mathbf{g}}) + (1-q)\,\mathbf{P}(\bar{\mathbf{g}})$. Finally, Lemma 2.6 states that $\mathbf{h}^\gamma$ is the same for $\underline{\mathbf{g}}$ and $\bar{\mathbf{g}}$. In other words, we have for $\underline{\mathbf{g}}$

$$J^\gamma \mathbf{e} + \mathbf{h}^\gamma = B^\gamma(\underline{\mathbf{g}}) + \mathbf{P}(\underline{\mathbf{g}})\,\mathbf{h}^\gamma \tag{4.11}$$

together with a similar equation for $\bar{\mathbf{g}}$. The algebraic sum of these two vector equalities yields

$$J^\gamma \mathbf{e} + \mathbf{h}^\gamma = B^\gamma(\mathbf{f}_q) + \mathbf{P}(\mathbf{f}_q)\,\mathbf{h}^\gamma. \tag{4.12}$$

If both sides are premultiplied by $\mathbf{P}^*(\mathbf{f}_q)$, we obtain for arbitrary γ the equality $J^\gamma = J^\gamma(\mathbf{f}_q)$ as required by (4.10).

To verify the second part of (4.10), observe the continuity in q of $\mathbf{f}_q$. This in turn leads to the continuity of $\mathbf{D}(\mathbf{f}_q) = q\mathbf{D}(\underline{\mathbf{g}}) + (1-q)\,\mathbf{D}(\bar{\mathbf{g}})$ and (by Lemma 2.4) also that of $\mathbf{P}(\mathbf{f}_q)$ and $\mathbf{P}^*(\mathbf{f}_q)$.[5] But

$$K(\mathbf{f}_q)\,\mathbf{e} = \mathbf{P}^*(\mathbf{f}_q)\,\mathbf{D}(\mathbf{f}_q), \tag{4.13}$$

[5] It is tempting—but erroneous—to conclude that $\mathbf{P}^*(\mathbf{f}_q) = q\mathbf{P}^*(\underline{\mathbf{g}}) + (1-q)\,\mathbf{P}^*(\bar{\mathbf{g}})$. Further, a related example demonstrates that $J(\mathbf{g}_1) = J(\mathbf{g}_2) = J$ is consistent with $J(q\mathbf{g}_1 + (1-q)\,\mathbf{g}_2) < J$ unless $\mathbf{g}_1$ and $\mathbf{g}_2$ both satisfy the DPE.

which is now continuous in q by the continuity of each of its constituents. Further, $K(\mathbf{f}_q) = \alpha_0 < \alpha$ for $q = 0$ and $K(\mathbf{f}_q) = \alpha^0 > \alpha$ for $q = 1$. Therefore, there exists a $q \in (0, 1)$ such that the second half of (4.10) is satisfied.

Thus we have demonstrated that the optimal constrained policy may belong to $\mathbf{G}$, and is in no case more complex than a convex combination of two policies in $\mathbf{G}$, with a selection to be made between the two policies by the throw of a (biased) coin whose faces have probabilities q and $1-q$. However, it may well be that Hypothesis 4.1 is not valid, in which case neither the methodologies nor the conclusions of this section necessarily hold. Section V will contain a systematic exposition of these cases and the resultant policies.

V. Other Constraint Hypotheses

When Hypothesis 4.1 fails to apply, there are a number of other possibilities, most of which are not as interesting. Nevertheless, we shall briefly summarize the various situations that may occur. At the same time, we shall attempt to respond to the Questions posed at the end of Section I. We shall be able to provide answers in all except one case, where the nature of the optimal policy is not completely understood.

Let us assume there is an unconstrained solution of DPE that meets the constraint (1.6), that is, $\mathbf{g} \in \mathbf{G}$ solves the DPE, while at the same time $K(\mathbf{g}) \leqslant \alpha$. This case can be identified directly; it is also evidenced by $\gamma = 0$, or equivalently $\mathring{J}^{\lambda}|_{\lambda=0} \geqslant -\alpha$. It follows from Theorem 2.1 that no $\mathbf{u} \in \mathbf{U}$ exhibits a greater average reward than the $\mathbf{g}$ that solves (2.1).

Another possibility is that no $\mathbf{g} \in \mathbf{G}$ satisfies the constraint $K(\mathbf{g}) \leqslant \alpha$. In that case, $K^{\lambda} > \alpha$, and it is necessary (but not sufficient) that $\gamma = \infty$. We again apply the DPE, this time to find $\inf_{g \in \mathbf{G}} K(\mathbf{g})$. The resulting infimum will, of course, exceed α, as expected. However, it is more important to note in this instance that no policy in $\mathbf{U}$ can operate at a lower average cost than the simple policy in $\mathbf{G}$ that attains the infimum.

The discussion of the last paragraph answers Question 1.1: $\mathbf{U}_0$ is nonempty *iff* one of the conditions on $\mathbf{G}$ mentioned in the preceding paragraph is met. Further, when $\gamma < \infty$, the optimal policy is well defined, and must belong to either $\mathbf{G}$ or ${}_{\mathrm{m}}\mathbf{F}$, depending on the character of J^{λ} in the vicinity of γ. The solutions for $\gamma < \infty$ can always be implemented by solving parametrized versions of the DPE.

Finally, there is the case

$$\inf_{g \in \mathbf{G}} K(\mathbf{g}) = \alpha. \tag{5.1}$$

If $\gamma < \infty$ (equivalently, $J^{\lambda} = \alpha$ from some finite λ on), Theorem 4.3 prescribes the solution. It is possible, however, to exhibit examples where

$\gamma = \infty$. This means that the Lagrangian formulation is incapable of treating the problem. Although the infimum in (5.1) can be attained by some $\mathbf{g}$ (as is seen by using the usual compactness arguments), it seems unclear how one ought to proceed to find such a $\mathbf{g}$ explicitly, or to optimize relative to the average reward. Nor is it clear whether there is some $\mathbf{u} \in \mathbf{U}_0$ that reaps a larger reward than the best $\mathbf{g}$.

Note added in proof. The authors have recently shown that Theorem 4.4 can be strengthened to assert the existence of an optimal constrained policy that is randomized between two actions at no more than one state, and is nonrandom on all the other states of $\mathbf{S}$.

References

1. D. P. Bertsekas, "Dynamic Programming and Stochastic Control," Academic Press, New York, 1976.
2. P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1968.
3. V. S. Borkar, Controlled Markov chains and stochastic networks, *SIAM J. Control Optim.* **21** (1983), 652–666.
4. R. Courant and D. Hilbert, "Methods of Mathematical Physics," Vol. I, Interscience, New York, 1953.
5. C. Derman, "Finite State Markovian Decision Processes," Academic Press, New York, 1970.
6. J. L. Doob, "Stochastic Processes," Wiley, New York, 1953.
7. E. B. Dynkin and A. A. Yushkevich, "Controlled Markov Processes," Springer-Verlag, New York, 1979.
8. W. Feller, "An Introduction to Probability Theory and Its Applications," Vol. 1, 3rd ed., Wiley, New York, 1968.
9. E. B. Frid, On optimal strategies in control problems with constraints, *Theory Probab. Appl.* **18** (1972), 182–192.
10. L. C. Kallenberg, "Linear Programming and Finite Markovian Control Problems," North-Holland, New York, 1983.
11. L. Kleinrock, "Queueing Systems," Vol. 2, "Computer Applications," Wiley, New York, 1976.
12. M. Robin, On optimal stochastic control with constraints, *in* "Game Theory and Related Topics," pp. 187–202, North-Holland, New York, 1982.
13. Z. Rosberg, P. Varaiya, and J. Walrand, Optimal control of service in tandem queues, *IEEE Trans. Automat. Control.* **AC-27** (1982), 600–610.
14. S. M. Ross, "Introduction to Stochastic Dynamic Programming," Academic Press, New York, 1983.
15. A. Segall, Dynamic file assignment in a computer network, *IEEE Trans. Automat. Control* **AC – 21** (1976), 161–173.

Printed by the St. Catherine Press Ltd., Tempelhof 41, Bruges, Belgium