Available online at www.sciencedirect.com
ScienceDirect
Operations Research Letters 35 (2007) 267–272
Operations Research Letters
www.elsevier.com/locate/orl

# Monotonicity in multidimensional Markov decision processes for the batch dispatch problem

Katerina Papadaki[a,*], Warren B. Powell[b]

[a]*Department of Operational Research, London School of Economics, Houghton Street, London WC2A 2AE, UK*
[b]*Department of Operations Research and Financial Engineering, Princeton University, USA*

Received 10 October 2005; accepted 22 March 2006
Available online 26 May 2006

**Abstract**

Structural properties of stochastic dynamic programs are essential to understanding the nature of the solutions and in deriving appropriate approximation techniques. We concentrate on a class of multidimensional Markov decision processes and derive sufficient conditions for the monotonicity of the value functions. We illustrate our result in the case of the multiproduct batch dispatch (MBD) problem.


*Keywords:* Markov decision processes; Monotone value functions

## 1. Introduction

In stochastic dynamic programming, properties of the one period cost function and value functions (such as monotonicity, convexity, submodularity) are often derived to give insights into the structure of the optimal decision policies. Further, knowledge of the structure of the value functions can aid the design of approximation algorithms that estimate these value functions. This is especially important in the case of multidimensional Markov decision processes where the state space, action space or outcome space are large and optimal solutions become computationally intractable. In this paper we concentrate on monotonicity properties of the value functions. There are various examples in the literature where the monotonicity of the value function was used to prove that there exist threshold-type decision policies. An example where the monotonic structure of the value function was used to aid the estimation of the value functions using an approximate dynamic programming algorithm can be found in [2].

Monotonicity properties have been studied in the literature for various problems formulated as stochastic dynamic programs. Reports on conditions for monotonicity of scalar value functions can be found in [5]. The author in [4, Section 4.7, pp. 102–112] also gives sufficient conditions for monotone value functions for a scalar value function in a finite horizon setting based

* Corresponding author.
*E-mail address:* k.p.papadaki@lse.ac.uk (K. Papadaki).

0167-6377/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.orl.2006.03.013

on the work by [6]. Discussions of monotonicity appear also in [9,1,8]. The paper [7] provides results on general properties of value functions. Most results on the monotonicity of value functions are conditioned on monotonicity properties of the cumulative transition probabilities. Our result for a specific class of Markov decision processes requires a monotonicity property of the state transition function, with no restriction on the distribution of the exogenous stochastic process.

In their paper [3], the authors study the structural properties of the multiproduct batch dispatch (MBD) problem and provide an approximate solution using approximate dynamic programming algorithms. The MBD problem consists of products from different classes arriving at a dispatch station incurring class-dependent holding costs while waiting to be dispatched by a single finite capacity vehicle. There is a fixed cost associated with each dispatch of a vehicle. The problem involves finding the optimal policy of dispatching the vehicle over a discrete finite time horizon. The authors in [3] formulate the problem as a stochastic dynamic program and investigate properties of the value function and the optimal decision policies.

The author in [4] reports structural results on value function properties and optimal decision policies for general stochastic dynamic programs that have a scalar state space. He also provides conditions for monotonicity of the value function for scalar problems. The authors in [3] inappropriately use these scalar results in their paper to report monotonicity of the multidimensional value function of the MBD problem. The value function of the MBD problem is defined on a multidimensional state space $\mathscr{S}$, which does not have a total ordering, but rather a partial ordering. Thus, the property of the value function is only one of *partial monotonicity* based on a partial ordering of the multidimensional state space.

In this paper we define a partial ordering on $\mathscr{S}$ and prove partial monotonicity of the value function for a general Markov decision process. We extend the scalar result for monotonicity of the value function stated in [4], to one of partial monotonicity for the multidimensional case. Then we proceed to apply this result to the MBD problem, thus correcting the structural results of [3].

The paper begins by formulating a general Markov decision model in Section 2. Then, in Section 3 we derive sufficient conditions for the monotonicity of the value functions based on the model formulated in Section 2. In Section 4, we briefly describe the MBD problem and we use the main result to show monotonicity of the value functions.

## 2. The Markov decision model

In this section we define a fairly general Markov decision model. Based on this model we derive conditions for the monotonicity of the value function. In Section 4, we extend this model to the case of the MBD problem.

We define a discrete finite horizon Markov decision process. Time is divided into discrete intervals called decision epochs that are indexed by $t$, $t \in \{0, 1, \ldots, T\}$. The state, decision, and stochastic processes are $N$-dimensional. We define the state process $\{s_t\}_{t=0}^{T}$, where $s_t = (s_t(1), \ldots, s_t(N))$ and $s_t \in \mathscr{S}_1 \times \cdots \times \mathscr{S}_N = \mathscr{S}$. We assume an exogenous stochastic process whose realization at decision epoch $t$ is denoted by the vector $a_t = (a_t(1), \ldots, a_t(N)) \in \mathscr{A}$, where $\mathscr{A} = \mathscr{A}_1 \times \cdots \times \mathscr{A}_N$. The sets $\mathscr{S}_i$ and $\mathscr{A}_i$, for all $i$, are countable ordered subsets of $\mathbb{R}^+$. Further, we define the decision process denoted at time $t$ by the vector $x_t = (x_t(1), \ldots, x_t(N))$, where $x_t \in \mathscr{X}_1 \times \cdots \times \mathscr{X}_N = \mathscr{X}$. The sets $\mathscr{X}_i$ are finite countable subsets of $\mathbb{R}^+$ and $\mathscr{X}$ is the $N$-dimensional action space, which we assume is independent of $s_t$.

At the beginning of decision epoch $t$, events occur in the following time order: the realization $a_t$ of the stochastic process occurs, the state $s_t$ is measured, the decision $x_t$ is taken.

Let $f : \mathscr{S} \times \mathscr{X} \to \mathscr{S}$ be such that $f(s_t, x_t)$ is the vector that gives the state at the end of decision epoch $t$, just before the realization $a_{t+1}$ of the stochastic process occurs at the beginning of decision epoch $t+1$. We assume that the state at time $t+1$ is as follows:

$$s_{t+1} = f(s_t, x_t) + a_{t+1}. \tag{1}$$

We assume a particular structure of the transition function where the state measured at time $t+1$ is the sum of the state at the end of decision epoch $t$ and the realization of the stochastic process at time $t+1$. This structure is very common to many problems that are formulated as stochastic dynamic problems.

We are now ready to introduce the transition probabilities. First, we define the probabilities for the stochastic process. Let $p_t^a(a_t)$ be the probability that the realization of the stochastic process at the beginning of decision epoch $t$ is given by $a_t$. Further, let $p_{t+1}^s(s_{t+1}|s_t, x_t)$ be the probability that the state at decision epoch $t+1$ is $s_{t+1}$, given that the state and decision variables at epoch $t$ where $s_t$ and $x_t$, respectively. Then using (1) we have

$$p_{t+1}^s(s_{t+1}|s_t, x_t) = p_{t+1}^a(s_{t+1} - f(s_t, x_t)).$$

We let $g_t(s_t, x_t)$ be the one-period cost function at decision epoch $t$ when the state is $s_t$ and decision $x_t$ is taken. We let $g_T(s_T)$ be the terminal cost function. The objective is to minimize over all feasible policies $\pi = (x_0^\pi, \ldots, x_{T-1}^\pi)$ the expected total discounted cost, $F(s_0)$, over the entire time horizon:

$$F(S_0) = \min_{\pi \in \Pi} \mathbb{E}\left\{ \sum_{t=0}^{T-1} \alpha^t g_t(s_t, x_t^\pi) + \alpha^T g_T(S_T) \right\}, \quad (2)$$

where $\Pi$ is the set of all feasible policies and $0 < \alpha < 1$ is a discount factor. The problem described in (2) is equivalent to solving the optimality equations:

$$V_t(s_t) = \min_{x_t \in \mathcal{X}} \left\{ g_t(s_t, x_t) + \alpha \sum_{s' \in \mathcal{S}} p_{t+1}^s(s'|s_t, x_t) V_{t+1}(s') \right\},$$

$$V_T(s_T) = g_T(s_T), \quad (3)$$

for $t = 0, 1, \ldots, T-1$, where $V_t(s_t)$ is the total optimal cost (cost to go) from time period $t$ until the end of the horizon.

## 3. Monotonicity of the value function

In this section we state and prove sufficient conditions for the monotonicity of the value function of the MDP defined in Section 2. The main result is summarized in Theorem 3.2. We begin with a few definitions.

**Definition.** We define the partial ordering operator $\preccurlyeq$ or $\succcurlyeq$ on the $N$-dimensional set $\mathcal{S}$. We denote $s \preccurlyeq s'$ ($s \succcurlyeq s'$) for $s, s' \in \mathcal{S}$, if for all $i \in \{1, 2, \ldots, N\}$ we have $s(i) \leqslant s'(i)$ ($s(i) \geqslant s'(i)$).

**Definition.** A real-valued function $F$ defined on an $N$-dimensional set $\mathcal{S}$ is partially nondecreasing (nonincreasing) if for all $s^+, s^- \in \mathcal{S}$ such that $s^+ \succcurlyeq s^-$, we have $F(s^+) \geqslant F(s^-)$ ($F(s^+) \leqslant F(s^-)$).

**Remark.** We would like to note that the expectation in the optimality equations (3) can be rewritten as follows using the partial ordering operators:

$$\sum_{s' \in \mathcal{S}, s' \succcurlyeq f(s_t, x_t)} p_{t+1}^s(s'|s_t, x_t) V_{t+1}(s').$$

This is due to the fact that if for some $i \in \{1, 2, \ldots, N\}$ we have $s'(i) < f(s_t, x_t)(i)$, then $p_{t+1}^s(s'|s_t, x_t) = 0$. Thus, we only need to take the expectation over states $s'$ that satisfy $s' \succcurlyeq f(s_t, x_t)$.

To prove our main result we need the following lemma:

**Lemma 3.1.** *Suppose that $V_{t+1}$ is partially nondecreasing (nonincreasing) in $\mathcal{S}$, and for $e \succcurlyeq 0$ we have $f(s+e, x) \succcurlyeq f(s, x)$ for all $x \in \mathcal{X}$. Then we have,*

$$\sum_{j \in \mathcal{S}, j \succcurlyeq f(s+e, x)} p^s(j|s+e, x) V_{t+1}(j) \geqslant \sum_{j \in \mathcal{S}, j \succcurlyeq f(s,x)} p^s(j|s, x) V_{t+1}(j) \quad (4)$$

(*where the inequality in* (4) *is reversed in the nonincreasing case*).

**Proof.** We can rewrite Eq. (4) using the arrival probabilities instead of the transition probabilities. For $j \succcurlyeq f(s, x)$ we have

$$p^s(j|s, x) = p^a(j - f(s, x)). \quad (5)$$

Using (5), (4) becomes

$$\sum_{j \in \mathcal{S}, j \succcurlyeq f(s+e, x)} p^a(j - f(s+e, x)) V_{t+1}(j) \geqslant \sum_{j \in \mathcal{S}, j \succcurlyeq f(s,x)} p^a(j - f(s, x)) V_{t+1}(j). \quad (6)$$

We substitute $i = j - f(s+e, x)$ in the first sum of (6), and $k = j - f(s, x)$ in the second sum of (6).

Then (6) becomes

$$\sum_{i\in\mathscr{S},i\succcurlyeq 0} p^a(i)V_{t+1}(i+f(s+e,x)) \geqslant \sum_{k\in\mathscr{S},k\succcurlyeq 0} p^a(k)V_{t+1}(k+f(s,x)).$$

The above holds from the assumption that $V_{t+1}$ is partially nondecreasing. □

The following theorem states sufficient conditions for partial monotonicity of the value function for the $N$-dimensional MDP:

**Theorem 3.2.** *Suppose the following conditions hold*:

(i) *For* $e\succcurlyeq 0$ *we have* $f(s+e,x)\succcurlyeq f(s,x)$ *for all* $x\in\mathscr{X}$.
(ii) *The one period cost function* $g_t(s,x)$ *is partially nondecreasing* (*nonincreasing*) *in* $s\in\mathscr{S}$ *for all* $x\in\mathscr{X}$, $t=0,\ldots,T-1$.
(iii) *The terminal cost function* $g_T(s)$ *is partially nondecreasing* (*nonincreasing*) *in* $s\in\mathscr{S}$.

*Then the value function* $V_t(s_t)$ *is partially nondecreasing* (*nonincreasing*) *in* $s_t$ *for all* $t=0,\ldots,T$.

**Proof.** We prove this by induction. From condition (iii) the result holds for $V_T(s_T)=g_T(s_T)$.

Assume now that $V_n$ is partially nondecreasing for $n=t+1,\ldots,T$. We want to prove that $V_t$ is partially nondecreasing. $V_t$ is as follows:

$$V_t(s)=\min_{x_t\in\mathscr{X}}\left\{g_t(s,x_t)+\sum_{j\in\mathscr{S},j\succcurlyeq f(s,x_t)} p^s(j|s,x_t)V_{t+1}(j)\right\}.$$

Given that the action space is finite, $\exists x_t^+\in\mathscr{X}$ which attains the above minimum for state $s=s^+$. Thus, the value function can be written as

$$V_t(s^+)=g_t(s^+,x_t^+)+\sum_{\substack{j\in\mathscr{S}\\ j\succcurlyeq f(s^+,x_t^+)}} p^s(j|s^+,x_t^+)V_{t+1}(j).$$

For $s^+\succcurlyeq s^-$, and due to conditions (i) and (ii) and Lemma 3.1, we have,

$$\begin{aligned}V_t(s^+)&\geqslant g_t(s^-,x_t^+)+\sum_{\substack{j\in\mathscr{S}\\ j\succcurlyeq f(s^-,x_t^+)}} p^s(j|s^-,x_t^+)V_{t+1}(j)\\ &\geqslant \min_{x_t\in\mathscr{X}}\left\{g_t(s^-,x_t)+\sum_{\substack{j\in\mathscr{S}\\ j\succcurlyeq f(s^-,x_t)}} p^s(j|s^-,x_t)V_{t+1}(j)\right\}\\ &=V_t(s^-).\end{aligned}$$

Therefore, $V_t(s)$ is partially nondecreasing for all $t$. □

## 4. Multiproduct batch dispatch problem

In this section we describe the MBD problem and use the result of Theorem 3.2 to show that the value function for this problem is partially nondecreasing.

We consider the problem of multiple types of products manufactured at the supplier's side and waiting to be dispatched in batches to the retailer by a vehicle with finite capacity $K$. We group the products in product classes according to their type and we assume that there is a finite number of $N$ classes. The state variable $s_t(i)$ depicts the number of products of type $i$ that are waiting in the queue.

The products across classes are homogeneous in volume and thus indistinguishable when filling up the vehicle. The differences between product types arise from special storage requirements of the products or from priorities of the product types according to demand. In both of the above cases the holding cost of products differs across product classes either because the cost of inventory is different or because the opportunity cost of shipping different types of products is different. We order the classes according to their holding cost starting from the most expensive type to the least expensive type. In this manner we construct a monotone holding cost structure. If $h=(h_1,\ldots,h_N)$ is the holding cost for each class type, then we have $h_1>h_2>\cdots>h_N$. Further, we assume that there is a fixed cost $c$ of dispatching the vehicle.

At each time epoch arrivals from all product types occur and are given by the vector $a_t$. We assume that

the arrival process is a stochastic process with a general distribution. Given the queue lengths $s_t$, a decision is taken at the beginning of decision epoch $t$ of whether to dispatch the vehicle. Further, if the vehicle is dispatched, a decision is taken on the distribution of product types to be dispatched. We let $x_t(i)$ denote the number of products of type $i$ that are dispatched at time $t$. When the vehicle is not dispatched the decision vector $x_t = 0$. We let $\mathscr{X}(s)$ be the feasible set of decision variables given that we are in state $s$:

$$\mathscr{X}(s) = \left\{ x \in \mathscr{S} : x \preccurlyeq s, \sum_{i=1}^{N} x_i \leqslant K \right\}. \tag{7}$$

We also define the dispatch variables $z_t$ which are functions of the decision variables indicating whether the vehicle is dispatched or not:

$$z_t(x_t) = \begin{cases} 1 & \text{if } \sum_{i=1}^{N} x_t(i) > 0, \\ 0 & \text{if } \sum_{i=1}^{N} x_t(i) = 0. \end{cases}$$

We assume that at the beginning of decision epoch $t$ the arrivals occur immediately before the state variable $s_t$ is measured and the decisions $x_t$ are taken immediately after. Thus, the system dynamics for the MBD problem are as follows:

$$s_{t+1} = s_t - x_t + a_{t+1}.$$

The one period cost function $g_t(s_t, x_t)$ consists of the dispatch cost and the holding cost:

$$g_t(s_t, x_t) = cz_t(x_t) + h^T(s_t - x_t).$$

The objective is to determine optimal dispatch policies over the finite time horizon to minimize expected total costs. The objective function to be minimized and the optimality equations are given by (2) and (3) described in Section 2.

We note that the action space defined in (7) depends on the state variable. The results of Section 3 apply to MDPs where the action space at each time epoch is the same. However, [3] proved the following result on the structure of the optimal decision policies: the optimal way to fill up the vehicle is to sort the product types according to their holding cost and iteratively fill the vehicle starting from the class with the highest holding cost. More formally, when the state variable is $s$ and the decision is to dispatch the vehicle then the dispatch vector should be $\chi(s)$, where its $i$th component is as follows:

$$\chi_i(s) = \begin{cases} s(i) & \text{if } \sum_{k=1}^{i} s(k) < K, \\ K - \sum_{k=1}^{i-1} s(k) & \text{if } \sum_{k=1}^{i-1} s(k) < K \\ & \quad \leqslant \sum_{k=1}^{i} s(k), \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

The authors in [3] proved the following result for the MBD problem:

**Proposition 4.1.** *Given that the state at time $t$ is $s_t$ then the optimal decision $x_t$ is either 0 or $\chi(s_t)$, for all $t = 0, \ldots, T-1$.*

Using the above result we redefine the action space to be $\{0, 1\}$. We let our decision variables to be $z_t \in \{0, 1\}$. The transition function becomes: $s_{t+1} = s_t - z_t\chi(s_t) + a_{t+1}$. The state at the end of decision epoch $t$ just before the arrivals of decision epoch $t+1$ occur was denoted in Section 2 by $f(s_t, x_t)$. Here we define the function $f$ as follows:

$$f(s_t, z_t) \equiv s_t - z_t\chi(s_t).$$

Thus the cost function, transition probabilities and optimality become

$$g_t s_t z_t = cz_t + h^T f(s_t, z_t),$$

$$p^s_{t+1}(s'|s_t, z_t) = p^a_{t+1}(s' - f(s_t, z_t)),$$

$$V_t(s_t) = \min_{z_t \in \{0,1\}} \left\{ cz_t + h^T(s_t - z_t\chi(s_t)) + \alpha \sum_{s' \in \mathscr{S}, s' \succcurlyeq f(s_t, z_t)} p^s_{t+1}(s'|s_t, z_t) V_{t+1}(s') \right\}. \tag{9}$$

In the next two lemmas we show that the MBD problem satisfies the necessary conditions of Theorem 3.2.

**Lemma 4.2.** *For all $z \in \{0, 1\}$ and for all $s^+, s^- \in \mathscr{S}$ such that $s^+ \succcurlyeq s^-$ we have:*

$$f(s^+, z) \succcurlyeq f(s^-, z). \tag{10}$$

**Proof.** We consider two cases. For $z = 0$, (10) follows from $s^+ \succcurlyeq s^-$.

Now consider the case $z=1$: We compare the vectors $s^- - \chi(s^-)$ and $s^+ - \chi(s^+)$. In the case that they are both zero then (10) is trivial. In the case that one of them is zero then it has to be $s^- - \chi(s^-)$ since $s^-$ is a smaller state and it would empty out faster after a dispatch than $s^+$ would. In this case (10) is satisfied since $s^+ - \chi(s^+) \geqslant 0$.

Now, consider the case that both vectors $s^- - \chi(s^-)$ and $s^+ - \chi(s^+)$ are nonzero. From the definition of $\chi$, (8), there exists an $i$ such that

$$\begin{cases} s^+(k) - \chi_k(s^+) = 0 & \text{for } k < i, \\ s^+(k) - \chi_k(s^+) > 0 & \text{for } k = i, \\ s^+(k) - \chi_k(s^+) = s^+(k) & \text{for } k > i \end{cases} \tag{11}$$

and there exists a $j$ such that

$$\begin{cases} s^-(k) - \chi_k(s^-) = 0 & \text{for } k < j, \\ s^-(k) - \chi_k(s^-) > 0 & \text{for } k = j, \\ s^-(k) - \chi_k(s^-) = s^-(k) & \text{for } k > j. \end{cases} \tag{12}$$

Since only $K$ units are dispatched and the entries of $s^-$ are not greater than $s^+$, the dispatch vector $\chi(s^-)$ will have the capacity to dispatch from lower holding cost classes than the dispatch vector $\chi(s^+)$. Thus, $j$ must be greater than $i$. Using this and (11) and (12) we get the desired result. □

**Lemma 4.3.** *The one period cost function $g_t(s, z)$ is partially nondecreasing in $s \in \mathscr{S}$ for all $z \in \{0, 1\}$ and for all $t = 0, \ldots, T-1$.*

**Proof.** Let $s^+, s^- \in \mathscr{S}$ such that $s^+ \geqslant s^-$. From Lemma 4.2 we have that:

$$h^T f(s^+, z) \geqslant h^T f(s^-, z). \tag{13}$$

Thus, from (9) and (13) we get that $g_t(s^+, z) \geqslant g_t(s^-, z)$ for all $t = 0, \ldots, T-1$. □

If we assume that the terminal cost function $g_T(s_T)$ is partially nondecreasing, and since the sets $\mathscr{S}_i$ and $\mathscr{A}_i$ are countable ordered subsets of $\mathbb{R}^+$, and the action space $\{0, 1\}$ is finite, then we can apply Theorem 3.2 and we have the following result:

**Theorem 4.4.** *The value function $V_t$ for the MBD problem is partially nondecreasing for all $t=0, \ldots, T$.*

## Acknowledgements

We would like to thank Diego Klabjan for highlighting the inconsistency of using the scalar results from [4] in a multidimensional setting. The second author was supported in part by Grant AFOSR-FA9550-05-1-0121 from the Air Force Office of Scientific Research and NSF Grant CMS-0324380.

## References

[1] D.P. Heyman, M.J. Sobel, Stochastic Models in Operations Research, vol. II, McGraw-Hill, New York, 1984.
[2] K. Papadaki, W.B. Powell, A monotone adaptive dynamic programming algorithm for a stochastic batch service problem, European J. Oper. Res. 142 (1) (2002) 108–127.
[3] K. Papadaki, W.B. Powell, An adaptive dynamic programming algorithm for a stochastic multiproduct batch dispatch problem, Naval Res. Logist. 50 (7) (2003) 742–769.
[4] M.L. Puterman, Markov Decision Processes, Wiley, New York, 1994, pp. 102–112 (Section 4.7).
[5] S.M. Ross, Introduction to Stochastic Dynamic Programming, Academic Press, New York, 1983.
[6] R.F. Serfozo, Monotone optimal policies for Markov decision processes, Math. Programming Study 6 (1976) 202–215.
[7] J.E. Smith, K.F. McCardle, Structural properties of stochastic dynamic programs, Oper. Res. 50 (5) (2002) 796–809.
[8] N.L. Stokey, R.E. Lucas, Recursive Methods in Economic Dynamics, Harvard University Press, Cambridge, MA, London, England, 1989.
[9] D.M. Topkis, Supermodularity and Complementarity, Princeton University Press, Princeton, NJ, 1998.