Sufficient Conditions for Monotone Value Functions in Multidimensional Markov Decision Processes: The Multiproduct Batch Dispatch Problem

Katerina Papadaki * Warren B. Powell †

October 7, 2005

Abstract

Structural properties of stochastic dynamic programs are essential to understanding the nature of the solutions and in deriving appropriate approximation techniques. 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. 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.

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. 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 Papadaki and Powell [1].

Monotonicity properties have been studied in the literature for various problems formulated as stochastic dynamic programs. Ross [4] reports on conditions for monotonicity

*K. Papadaki is with the Department of Operational Research in London School of Economics. E-mail: k.p.papadaki@lse.ac.uk

†W. B. Powell is with the Department of Operations Research and Financial Engineering at Princeton University. E-mail: powell@princeton.edu

of scalar value functions. Puterman [3] also gives sufficient conditions for monotone value functions for a scalar value function in a finite horizon setting based on the work by Serfozo [6]. Discussions of monotonicity appear also in Topkis [8] and Heyman and Sobel [7]. Smith and McCardle [5] provide 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 [2], Papadaki and Powell 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. Papadaki and Powell [2] formulate the problem as a stochastic dynamic program and investigate properties of the value function and the optimal decision policies.

Puterman [3] 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. Papadaki and Powell 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 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 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 [3], 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 Papadaki and Powell [2].

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 multiproduct batch dispatch problem and we use the main result to show monotonicity of the value functions. Finally, in section 5 we conclude.

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 multiproduct batch dispatch 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 $s_t = (s_t(1), ..., s_t(N))$ and $s_t \in \mathcal{S}_1 \times \ldots \times \mathcal{S}_N = \mathcal{S}$. We assume an exogenous stochastic process whose realization at decision epoch t is denoted by the vector $a_t = (a_t(1), ..., a_t(N)) \in \mathcal{A}$, where $\mathcal{A} = \mathcal{A}_1 \times \ldots \times \mathcal{A}_N$. The sets $\mathcal{S}_i$ and $\mathcal{A}_i$, for all i, are discrete ordered sets. Further, we define the decision process denoted at time t by the vector $x_t = (x_t(1), ..., x_t(N)) \in \mathcal{X}$, where $\mathcal{X}$ is the N-dimensional action space.

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 $f(s_t, x_t)$ be 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:

$$p_{t+1}^s(s_{t+1}|s_t, x_t) = p_{t+1}^a\left(s_{t+1} - f(s_t, x_t)\right) \tag{2}$$

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:

$$F(S_0) = \min_{\pi \in \Pi} E\left\{\sum_{t=0}^{T-1} \alpha^t g_t(s_t, x_t^\pi) + \alpha^T g_T(S_T)\right\} \tag{3}$$

where Π is the set of all feasible policies and $0 < \alpha < 1$ is a discount factor. The problem described in (3) is equivalent to solving the optimality equations:

$$\begin{aligned} V_t(s_t) &= \min_{x_t \in \mathcal{X}}\left\{g_t(s_t, x_t) + \alpha \sum_{s' \in \mathcal{S}} p_{t+1}^s(s'|s_t, x_t) V_{t+1}(s')\right\} \\ V_T(s_T) &= g_T(s_T) \end{aligned} \tag{4}$$

for $t = 0, 1, ..., T-1$, 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 proposition 3.2. We begin with a few definitions.

Definition We define the partial ordering operator $\preceq$ or $\succeq$ on the N-dimensional set $\mathcal{S}$: We denote $x \preceq y$ ($x \succeq y$) for $x, y \in \mathcal{S}$, if for all $i \in \{1, 2, \ldots, N\}$ we have $x(i) \leq y(i)$ ($x(i) \geq y(i)$).

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

Remark We would like to note that the expectation in the optimality equations (4) can be rewritten as follows using the partial ordering operators:

$$\sum_{s' \in \mathcal{S}, s' \succeq f(s_t, x_t)} p^s_{t+1}(s'|s_t, x_t) V_{t+1}(s') \tag{5}$$

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 $p^s_{t+1}(s'|s_t, x_t) = 0$. Thus we only need to take the expectation over states s' that satisfy $s' \succeq f(s_t, x_t)$.

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 $e \succeq 0$ we have $f(s + e, x) \succeq f(s, x)$ for all $x \in \mathcal{X}$. Then we have,*

$$\sum_{j \in \mathcal{S}, j \succeq f(s+e, x)} p^s\left(j|s + e, x\right) V_{t+1}(j) \geq \sum_{j \in \mathcal{S}, j \succeq f(s,x)} p^s\left(j|s, x\right) V_{t+1}(j) \tag{6}$$

(where the inequality in (6) is reversed in the nonincreasing case).

Proof We can rewrite equation (6) using the arrival probabilities instead of the transition probabilities. For $j \succeq f(s, x)$ we have:

$$p^s(j|s, x) = p^a(j - f(s, x)) \tag{7}$$

Using (7), (6) becomes:

$$\sum_{j \in \mathcal{S}, j \succeq f(s+e, x)} p^a\left(j - f(s + e, x)\right) V_{t+1}(j) \geq \sum_{j \in \mathcal{S}, j \succeq f(s,x)} p^a\left(j - f(s, x)\right) V_{t+1}(j) \tag{8}$$

We substitute $i = j - f(s + e, x)$ in the first sum of (8), and $k = j - f(s, x)$ in the second sum of (8). Then (8) becomes:

$$\sum_{i \in \mathcal{S}, i \succeq 0} p^a(i) V_{t+1}\left(i + f(s + e, x)\right) \geq \sum_{k \in \mathcal{S}, k \succeq 0} p^a(k) V_{t+1}\left(k + f(s, x)\right) \tag{9}$$

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

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

Proposition 3.2 *Suppose the following conditions hold:*
(i) For $e \succeq 0$ we have $f(s+e,x) \succeq f(s,x)$ for all $x \in \mathcal{X}$.
(ii) The one period cost function $g_t(s,x)$ is partially nondecreasing (nonincreasing) in $s \in \mathcal{S}$ for all $x \in \mathcal{X}$, $t = 0,\ldots,T-1$.
(iii) The terminal cost function $g_T(s)$ is partially nondecreasing (nonincreasing) in $s \in \mathcal{S}$.

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:

$$V_t(s) = \min_{x_t \in \mathcal{X}} \left\{ g_t(s,x_t) + \sum_{j \in \mathcal{S}, j \succeq f(s,x_t)} p^s\left(j|s,x_t\right) V_{t+1}(j) \right\} \tag{10}$$

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

$$V_t(s^+) = g_t\left(s^+, x_t^+\right) + \sum_{\substack{j \in \mathcal{S} \\ j \succeq f(s^+, x_t^+)}} p^s\left(j|s^+, x_t^+\right) V_{t+1}(j) \tag{11}$$

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

$$\begin{aligned} V_t(s^+) &\geq g_t\left(s^-, x_t^*\right) + \sum_{\substack{j \in \mathcal{S} \\ j \succeq f(s^-, x_t^*)}} p^s\left(j|s^-, x_t^*\right) V_{t+1}(j) \\ &\geq \min_{x_t \in \mathcal{X}} \left\{ g_t\left(s^-, x_t\right) + \sum_{\substack{j \in \mathcal{S} \\ j \succeq f(s^-, x_t)}} p^s\left(j|s^-, x_t\right) V_{t+1}(j) \right\} \\ &= V_t(s^-) \end{aligned} \tag{12}$$

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

4 Multiproduct Batch Dispatch Problem

In this section we describe the Multiproduct Batch Dispatch (MBD) problem and use the result of proposition 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 $h_1 > h_2 > \ldots > h_N$. 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 $\mathcal{X}(s)$ to be the feasible set of decision variables given that we are in state s:

$$\mathcal{X}(s) = \left\{ x \in \mathcal{S} : x \leq s, \sum_{i=1}^{m} x_i = 0 \;\; OR \;\; \min(\sum_{i=1}^{m} s_i, K) \right\} \tag{13}$$

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

$$z_t(x_t) = \begin{cases} 1 & \text{if } \sum_{i=1}^{N} x_t(i) > 0 \\ 0 & \text{if } \sum_{i=1}^{N} x_t(i) = 0 \end{cases} \tag{14}$$

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:

$$s_{t+1} = s_t - x_t + a_{t+1} \tag{15}$$

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

$$g_t(s_t, x_t) = cz_t(x_t) + h^T(s_t - x_t) \tag{16}$$

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 (3) and (4) described in section 2.

We note that the action space defined in (13) 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,

Papadaki and Powell [2] 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 ith component is as follows:

$$\chi_i(s) = \begin{cases} s(i) & \text{if } \sum_{k=1}^{i} s(k) < K \\ K - \sum_{k=1}^{i-1} s(k) & \text{if } \sum_{k=1}^{i-1} s(k) < K \leq \sum_{k=1}^{i} s(k) \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Papadaki and Powell [2] proved the following theorem for the MBD problem:

Theorem 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* $t = 0, ..., T-1$.

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:

$$f(s_t, z_t) \equiv s_t - z_t\chi(s_t) \tag{18}$$

Thus the cost function, transition probabilities and optimality equations become:

$$g_t(s_t, z_t) = cz_t + h^T f(s_t, z_t) \tag{19}$$

$$p^s_{t+1}(s'|s_t, z_t) = p^a_{t+1}\left(s' - f(s_t, z_t)\right) \tag{20}$$

$$V_t(s_t) = \min_{z_t \in \{0,1\}} \left\{ cz_t + h^T\left(s_t - z_t\chi(s_t)\right) + \alpha \sum_{s' \in \mathcal{S}, s' \succeq f(s_t, z_t)} p^s_{t+1}(s'|s_t, z_t) V_{t+1}(s') \right\} \tag{21}$$

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

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

$$f(s^+, z) \succeq f(s^-, z) \tag{22}$$

Proof We consider two cases. For $z = 0$, (22) follows from $s^+ \succeq s^-$.

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 (22) 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 (22) is satisfied since $s^+ - \chi(s^+) \geq 0$.

Now, consider the case that both vectors $s^- - \chi(s^-)$ and $s^+ - \chi(s^+)$ are non-zero. From the definition of χ, (17), there exists an i such that:

$$\begin{cases} s^+(k) - \chi_k(s^+) = 0 & \text{for } k < i \\ s^+(k) - \chi_k(s^+) > 0 & \text{for } k = i \\ s^+(k) - \chi_k(s^+) = s^+(k) & \text{for } k > i \end{cases} \tag{23}$$

and there exists a j such that:

$$\begin{cases} s^-(k) - \chi_k(s^-) = 0 & \text{for } k < j \\ s^-(k) - \chi_k(s^-) > 0 & \text{for } k = j \\ s^-(k) - \chi_k(s^-) = s^-(k) & \text{for } k > j \end{cases} \tag{24}$$

Since only K units are dispatched and the entries of s^- are smaller 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 (23) and (24) we get the desired result. □

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

Proof Let $s^+, s^- \in \mathcal{S}$ such that $s^+ \succeq s^-$. From lemma 4.2 we have that:

$$h^T f(s^+, z) \geq h^T f(s^-, z) \tag{25}$$

Thus from (19) and (25) we get that $g_t(s^+, z) \geq g_t(s^-, z)$ for all $t = 0, \ldots, T-1$. □

Thus the sufficient conditions of proposition 3.2 are satisfied if we assume that the terminal cost function $g_T(s_T)$ is partially nondecreasing. Therefore, the value function V_t for the MBD problem is partially nondecreasing for all $t = 0, \ldots, T$.

5 Conclusions

We have provided sufficient conditions for partial monotonicity of value functions for a class of multidimensional Markov decision processes. The result is based on the assumption that the transition function of the Markov decision model has a particular structure. Contrary to previous results, we impose no conditions on the distribution of the exogenous stochastic process. We use the result in the multiproduct batch dispatch problem to show partial monotonicity of the value functions.

Acknowledgements

We would like to thank Diego Klabjan for highlighting the inconsistency of using the scalar results of Puterman [3] in a multidimensional setting.

References

[1] K. Papadaki, W. B. Powell, "A monotone adaptive dynamic programming algorithm for a stochastic batch service problem", *European Journal of Operational Research, Vol. 142, No. 1, pp.108-127, 2002.*

[2] K. Papadaki, W. B. Powell, "An Adaptive Dynamic Programming Algorithm for a Stochastic Multiproduct Batch Dispatch Problem", *Naval Research Logistics, Vol. 50, No. 7, pp.742-769, 2003.*

[3] M. L. Puterman, "Markov Decision Processes", *John Wiley and Sons, Inc., New York, 1994.*

[4] S. M. Ross, "Introduction to Stochastic Dynamic Programming", *Academic Press, New York, 1983.*

[5] J. E. Smith, K. F. McCardle, "Structural Properties of Stochastic Dynamic Programs", *Operations Research, Vol. 50, No. 5, pp.796-809, 2002.*

[6] R. F. Serfozo, "Monotone Optimal Policies for Markov Decision Processes", *Mathematical Programming Study, Vol. 6, pp. 202215, 1976.*

[7] D. P. Heyman, M. J. Sobel, "Stochastic Models in Operations Research: Vol II", *McGraw-Hill, New York, 1984.*

[8] D. M. Topkis, "Supermodularity and Complementarity", *Princeton, NJ: Princeton University Press, 1998.*