A note on dynamic programming with unbounded rewards

TECHNOLOGICAL UNIVERSITY EINDHOVEN

Department of Mathematics

STATISTICS AND OPERATIONS RESEARCH GROUP

Memorandum COSOR 75-13

A note on dynamic programming with unbounded rewards

by

J.A.E.E. van Nunen and J. Wessels

Eindhoven, September 1975

A note on dynamic programming with unbounded rewards

by

J.A.E.E. van Nunen and J. Wessels

Summary. In a recent paper, Lippman presents sufficient conditions for Denardo's N-stage contraction in discounted semi-Markov decision processes with unbounded rewards. In this note it is demonstrated that Lippman's conditions may be replaced by weaker conditions which even imply 1-stage contraction. The verification of the conditions of this note is somewhat easier.

Lippman [2] considers a discounted semi-Markov decision process with general state space S and action space A. He presents sufficient conditions for the existence of a normed Banach space of realvalued functions on S in which Denardo's N-stage contraction approach [1] may be used.
In Lippman's notation $q(\cdot|x,a)$, $r(x,a)$ denote the transition probability and one period reward respectively for state $x \in S$ and action $a \in A$; $\alpha > 0$ is the discountfactor; $t(\cdot|x,a)$ is the probability distribution function of the time until the next transition (given state $x \in S$, action $a \in A$).

The conditions in [2] are the following:
A function w on S exists with $w(x) \geq 1$, an integer $m \geq 1$ exists, a number β $(0 \leq \beta < 1)$ exists, positive numbers b and M exist, such that for all $x \in S$, $a \in A$:

$$\beta(x,a) := \int_0^\infty e^{-\alpha\tau} t(d\tau|x,a) \leq \beta \ ,$$

$$|r(x,a)|w^{-m}(x) \leq M \ ,$$

$$\int_S w^n(y)q(dy|x,a) \leq [w(x) + b]^n \quad \text{for } n = 1,\ldots,m \ .$$

Lippman's Banach space consists of realvalued functions u on S with the following norm:

$$\|u\| := \sup_x |u(x)|w^{-m}(x) \ .$$

Hence Lippman uses weighted supremum norms as introduced more generally for Markov decision processes in [3].
In [2] it is proved that under these conditions there exists an integer $J \geq 1$, such that for any sequence of policies $f_1, \ldots, f_J$ the operator $T_{f_1}, \ldots, T_{f_J}$ is a contraction. Here a policy f maps S into A, and T_f is defined as an operator in the Banach space with

$$(T_f u)(x) := r(x, f(x)) + \beta(x, f(x)) \int_S u(y) q(dy \mid x, f(x)) .$$

Lemma. Under Lippman's conditions the following holds: For any $\rho > \beta$ there exists a positive function v on S, such that

$$\beta \int_S v(y) q(dy|x,a) \leq \rho v(x) \quad \text{for all } x \in S,\ a \in A .$$

Proof. Choose a real number c with $c \geq b[(\frac{\rho}{\beta})^{1/m} - 1]^{-1}$ or $b + c \leq (\frac{\rho}{\beta})^{1/m} c$. Define $v(x) := [w(x) + c]^m$. Then

$$\int_S v(y) q(dy|x,a) = \int_S [w(y) + c]^m q(dy|x,a) =$$

$$= \sum_{n=0}^{m} \binom{m}{n} c^{m-n} \int_S w^n(y) q(dy|x,a) \leq$$

$$\leq \sum_{n=0}^{m} \binom{m}{n} c^{m-n} [w(x) + b]^n = [w(x) + b + c]^m \leq$$

$$\leq [w(x) + (\frac{\rho}{\beta})^{1/m} c]^m \leq \frac{\rho}{\beta} v(x) .$$

This lemma enables us to introduce a new weighted supremum norm (and hence a new Banach space, which actually contains the old one if $v = (w + c)^m$) in which T_f itself is already a contraction:

$$\| u \|_v := \sup_x |u(x)| v^{-1}(x) \quad \text{if } v(x) > 0 .$$

Consider the Banach space of realvalued functions u on S with $\| u \|_v < \infty$.

<u>Theorem</u>. Under Lippman's conditions the following holds: For any ρ $(\beta < \rho < 1)$ there exists a function v on S with $v(x) > 0$, such that for any policy f

$$\| T_f u_1 - T_f u_2 \|_v \le \rho \| u_1 - u_2 \|_v$$

$$\| r_f \|_v \le M \ ,$$

where $r_f(x) := r(x,f(x))$.

<u>Proof</u>. Choose c and v as in the lemma. Then

$$\begin{aligned} |(T_f u_1 - T_f u_2)(x)| &\le \beta \int_S |u_1(y) - u_2(y)| q(dy|x,f(x)) \\ &\le \beta \| u_1 - u_2 \|_v \int_S v(y) q(dy|x,f(x)) \\ &\le \rho \| u_1 - u_2 \|_v \, v(x) \ . \end{aligned}$$

Furthermore: $|r(x,a)| v^{-1}(x) \le |r(x,a)| w^{-m}(x) \le M$.

Now Lippman's conditions may be replaced by the following weaker and simpler conditions: A function v on S exists with $v(x) > 0$, a number β $(0 \le \beta < 1)$ exists, a number ρ $(\beta < \rho < 1)$ exists, a positive number M exists, such that for all $x \in S$, $a \in A$:

$$\beta(x,a) := \int_{0^-}^{\infty} e^{-\alpha\tau} t(d\tau|x,a) \le \beta \ ,$$

$$|r(x,a)| v^{-1}(x) \le M \ ,$$

$$\beta \int_S v(y) q(dy|x,a) \le \rho v(x) \ .$$

Namely, if our conditions are satisfied T_f is a ρ-contraction with respect to the norm $\| . \|_v$ and $\| r_f \|_v \le M$.

Remarks.

1) In order that T_f is contracting it is not necessary that $v(x) \geq 1$; in [2] the condition $w(x) \geq 1$ is essential. Actually we proved that, if Lippman's conditions are satisfied, with $w(x) > 0$ instead of $w(x) \geq 1$, than still a v-norm may be found satisfying our conditions.
2) As demonstrated in [3], the discounting requirement is not essential in our analysis: if we replace $\beta(x,a)q(\cdot|x,a)$ by $p(\cdot|x,a)$ then our conditions become:

$$|r(x,a)|v^{-1}(x) \leq M < \infty$$

$$\int_S v(y)p(dy|x,a) \leq \rho v(x) \quad \text{with } \rho < 1 \text{ .}$$

These conditions allow the situation $\alpha = 0$ in certain cases and give some weakening for $\alpha > 0$.

References

[1] E.V. Denardo, Contraction mappings in the theory underlying dynamic programming.
SIAM Review 9 (1967), 165-177.

[2] S.A. Lippman, On dynamic programming with unbounded rewards.
Management Science 21 (1975), 1225-1233.

[3] J. Wessels, Markov programming by successive approximations with respect to weighted supremum norms.
J. Math. Anal. Appl. (to appear).