Memo.No. 783

An Introduction to Dynamic Programming

by

Dr. Nadia Makary

June 1967

CAIRO
3. MOHAMED MAZHAR . ZAMALEK

"Opinions Expressed and Positions Taken by Authors are Entirely their Own and do not Necess rily Reflect the Views of the Institute of Nati Planning."

...Allocation Problem ".

Example 2. " Gold-mining : finite-stage case ",

Example 3. " Gold-mining : infinite stage case "

Example 4: " The Toymaker Example "

Example 5: " Seasonal Employment ",

Example 6: " Inventory Problem : finite-stage ca

Example 7: " Inventory Problem : infinite-stage

Preface:

Dynamic programming is an approach for a
multistage decision processes and finding out t
of the optimal policy. This note is a simple in
this approach. It first gives a general descrip
situations where dynamic programming may be app
a number of examples is given to illustrate the
and to classsify the dynamic programming techni

"policy" that maximizes (or minimizes) a predefined
cost) function.

Historically, it was developed mainly throu
papers (1950's) as a result of his trails to solve c
of "programming" problems involving time as a signif
However, the dynamic programming approach is used fo
many static processes that can be formulated as dyna
processes.

In contrast to linear programming, there is
set up for dynamic programming problems. Yet, there
features common to all problems that can be solved b
programming approach.

A general description of the situation where
programming approach can be applied may be presented

A system may be found in one of a possible
At each state there is a number of possible actions.
any of these actions, i.e., by making a decision, th
from one state to another in either a deterministic
way. Consequently, a certain income (or cost) is ea
The process continues for either finite or infinite

future value.

<u>The elements of the dynamic programming problem:</u>

S : set of states.

A : set of prossible actions,

Noticethat the action may depend on the state.

q : "the law of motion" of the system. It associa

each pair (s, a) a probability distribution o

In the deterministic case q(s'/(s, a)) equals

specified state $s' = s_o$ and equals zero for any

$s' \neq s_o$.

i(s,a): the immediate return function. It determines

(or cost) if the system is in state s and act

β : $0 \leq \beta \leq 1$; the discount factor.

The following definitions will be used:

To make a decision: is to choose one of the possible

A policy : is a sequence of decisions.

An optimal policy: is a policy that maximizes (or mi

total expected discounted income (or cost).

Thus the dynamic programming problem as defi

solved if the structure of the optimal policy is kno

an optimal policy for the remaining stages is indepe
policy adopted in previous stages. The direct resul
this principle is the development of the functional
nque which gives the recurrence relation between the
of the return function in successive stages, and thu
the optimal policy for each state with n stages rema
the optimal policy for each state with (n-1) stages

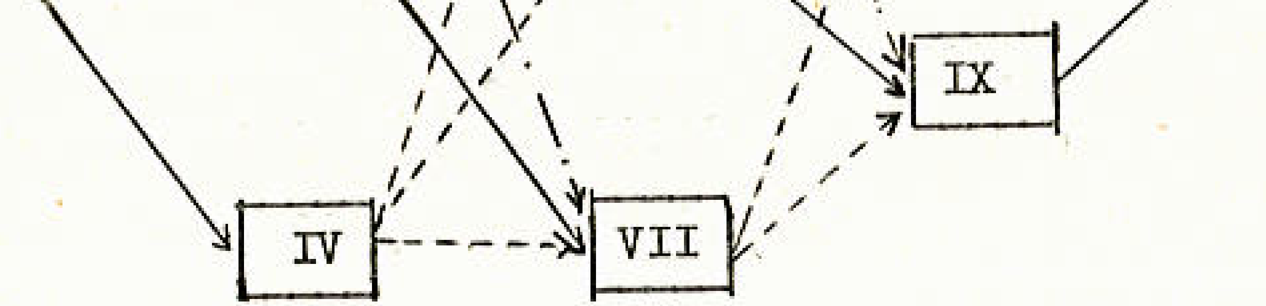
The dynamic programming approach has been su
plied to a wide variety of problems in different fie
follows, a number of examples will be discussed. So
given mainly to clarify the dynamic programming form
technique, others are presented to give an idea of s
applications.

<u>Example 1 :</u> (*)

(finite-stage deterministic process.)

A person wants to travel from city I to city
possible. Owing to the long distance between the two
has to make several stops before reaching his destina
each stop he can choose the route to the next stop as
diagram :

(*) This example is known in the literature by: "Sta



The number of hours necessary to go from one
next depends on the route he chooses as given in the

to / from	II	III	IV
I	2	4	3

to / from	V	VI	VII
II	7	4	6
III	3	2	4
IV	4	1	5

to / from	VI
V	[illegible]
VI	[illegible]
VII	[illegible]

Which routes should he choose in order to g
in the minimum number of hours?

Trail and error may be used for solving this
the dynamic programming approach provides an easier an
solution.

For example: a (II) = a (III) = a(IV) = V or VI

i(s,a) = the immediate "cost" of being in state s and

action a(s) (i.e., the member of hours needed

s to a(s)) .

For example : i(V,IV) = 4.

Take $\beta = 1$.

Notice that the traveller has to make four decisions. Each time he is confronted by a decision, called a "stage". So this problem is a 4-stage decisi

The stages will be numbered in a backward or at the first stage the person is either in state VIII IX and he should make the decision X, while at the four he is in state I and has to choose one of the actions IV.

Now, what is the optimal policy? In other is the sequence of routes that minimizes the total "co

<u>The solution</u> :

Let $f_n(s, a(s))$, n = 1, 2, 3, 4 denote the of being in state s (at the n^{th} stage), taking action following the optimal policy in the remaining (n-1) sta

f_1 (VIII) = 3,

f_1 (IX) = 4,

with a^*(VIII) = a^*(IX) = X

Suppose now that the traveller is in state VI
decides to go to VIII his total "cost" will be :

$$f_2 \text{ (VI, VIII) } = i \text{ (VI, VIII) } + f_1 \text{ (VIII)}$$
$$= 6 + 3 = 9.$$

But if be chooses to go to IX, his total cost

$$f_2 \text{ (VI, IX) } = i\text{(VI, IX) } + f_1 \text{ (IX)}$$
$$= 3 + 4 = 7.$$

Thus, f_2 (VI) = 7 and a^*(VI) = IX.

Similarly, f_2 (s) and a^*(s) can be calculated
state in the second stage. The consequent results are:

s \ a(s)	f_2(s, a(s)) VIII	IX	f_2(s)	a^*(s)
V	4	8	4	VIII
VI	9	7	7	IX
VII	6	7	6	VIII

With this information about the optimal polic
state in the second stage, the optimal decision in ea
the third stage can be found. For example, in state

a(s) \ S	f_3 (s, a(s)) V	VI	VII	f_3(s)	a^*(s)
II	11	11	12	11	V or V
III	7	9	10	7	V
IV	8	8	11	8	V or V

In exactly the same way, the optimal decisio
state in the fourth stage is found; from the followin
be either III or IV:

a(s) \ S	f_4 (s,(s)) II	III	IV	f_4(s)	a^*(s)
I	13	11	11	11	III or IV

These results show that at the initial state
should go to either III or IV. If he chooses III then
from there to V. From V he
should go to VIII and from there to X. His total "co
total number of hours he spends to get from I to X) i
this policy is 11 and it is the minimum possible cost
two more optimal routes that be can follow and still
hours. These alternative routes are:

I ⟶ IV ⟶ V ⟶ VIII ⟶

and I ⟶ IV ⟶ VI ⟶ IX ⟶

that it will mine no gold and be damaged beyond repair. ponding probabilities if it is used to mine gold in G a $(1-P_2)$ with a fraction r_2 $(0 < P_2, r_2 < 1)$.

The process begins by using the machine in eit If the machine is undamaged, another choice for using t in either F or G is made. The process continues in this times if the machine is undamaged, otherwise the proces when the machine is damaged.

What sequence of choices maximizes the amount before the end of the process?

<u>The dynamic programming formulation:</u>

$$S = \{ s=(\alpha,\gamma) : \alpha = (1-r_1)^k x, \ \gamma = (1-r_2)^{\ell} y \ ; \ k,\ell = 0, \dots, n \ ; \ k+\ell = n \ ; \ n = 0, \dots N\text{-}$$

A $= \{F, G\}$, where : a = F means that mine F is to be m
a=G means that mine G should be mined.

$$q : \quad q(s'/(\alpha,\gamma), F) = \begin{cases} P_1 & \text{if } s' = ((1-r_1)\alpha, \gamma) \\ 1-P_1 & \text{if } s' = (\alpha,\gamma) \end{cases}$$

$$q(s'/(\alpha,\gamma), G) = \begin{cases} P_2 & \text{if } s' = (\alpha, (1-r_2)\gamma) \\ 1-P_2 & \text{if } s' = (\alpha,\gamma) \end{cases}$$