UGP: Essays on Dynamic Programming

Abhishek Shree

Last updated November 23, 2025

Contents

§1. Introductory Dynamic Programming (Understanding Bellman)

1.1. Introduction

1.1.1. Notation

The standard Dynamic Programming problem desribed throughout will be in metric space $\mathbb{X}$.
Given, $\Phi : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ and $\Gamma : \mathbb{X} \rightrightarrows \mathbb{R}$
For a fixed x^0, find a sequence $x_m = x^1, ..., x^m, ...$ such that

$$\max \Phi(x^0, x^1) + \sum_{i=1}^{\infty} \delta^i \Phi(x^i, x^{i+1}), \text{ where } \delta \in (0,1) \text{ and } x^1 \in \Gamma(x^0), x^{i+1} \in \Gamma(x^i), m = 1, 2, 3, ...$$

Here δ is called the discount factor.

1.1.2. Functional Equations

Considering a space of Real numbers $\mathbb{R}$, we can define a functional equation as follows:

$$f(x) = \max_{0 \leq y \leq x} [g(y) + h(x-y) + f(ay + b(x-y))]$$

where $a, b \in \mathbb{R}$ and g, h are given functions.
This can be seen as an **insurance problem** where functions g and h represent the different places where and initial premium x is invested. The function f represents the value of the policy at the end of the period.

We can define, $R_1(x, y) = g(y) + h(x-y)$, initially the aim is to determine $\max_{0 \leq y \leq x} R_1(x, y)$. The value of x and y obtained from this maximization can be used to generate the next element of the sequence x_m.
i.e. $x_1 = ay + b(x-y) = y_1 + (x_1 - y_1)$
Further we define $R_2(x, y, y_1) = g(y) + h(x-y) + g(y_1) + h(x_1 - y_1)$ where $(y, y_1) \in [0, x] \times [0, x_1]$

This can be generalized to $R_m(x, y, y_1, ..., y_{m-1}) = g(y) + h(x-y) + g(y_1) + h(x_1 - y_1) + g(y_2) + h(x_2 - y_2) + ... + g(y_{m-1}) + h(x_{m-1} - y_{m-1})$
and the maximization problem becomes

$$\max R_m(x, y, y_1, ..., y_{m-1}) \text{ such that}$$

$$\begin{array}{ll} ay + b(x-y) = x_1 & 0 \leq y \leq x \\ ay_1 + b(x_1 - y_1) = x_2 & 0 \leq y_1 \leq x_1 \\ \vdots & \vdots \\ ay_{m-1} + b(x_{m-1} - y_{m-1}) = x_m & 0 \leq y_{m-1} \leq x_{m-1} \end{array}$$

Here, $f_N(x) = \max_{t, y_i} R_N(x, y_1, ..., y_N)$ and $f_N(x)$ can be interpreted as the maximum return obtained from an N-stage decision with initial premium x and $x \geq 0$.

$$\therefore f_N(x) = \max_{0 \leq y \leq x} [g(y) + h(x-y) + f_{N-1}(ay + b(x-y))]$$

or

$$f(x) = \max_{0 \leq y \leq x} [g(y) + h(x-y) + f(ay + b(x-y))] \text{ for } N \to \infty$$

1.2. Existence and Uniqueness Theorems

Under the following assumptions, the functional equation has a unique solution, which is continuous at $x = 0$ and has a value 0 at $x = 0$. Moreover the solution is continuous at $x \geq 0$.

The assumptions are:

1. g and h are continuous on $[0, \infty)$ and g(0) = h(0) = 0
2. If $m(x) = \max_{0 \le y \le x}[\max\{|g(y)|, |h(y)|\}]$ and $c = \max\{a, b\}$, then $\sum_{n=0}^{\infty} m(c^n x) < \infty \forall x \geq 0$

Proof. We define a mapping $T(f, y) = g(y)+h(x-y)+f(ay+b(x-y))$ and $f_{N+1}(x) = \max_{0\le y\le x} T(f_N, y)$, assuming g and h to be non-negative.
The sequence $\{f_N(x)\}$ is an increasing sequence of functions such that $f_N \to f$
Prooving that T is bounded above would suffice.
Method 1: We know that $x_1 \le cx$

$$\begin{aligned} \therefore\ ax_1 + b(x_1 - y) \le cx_1 \le c^2 x \\ \implies T(f_N, y) \le 2(m(x) + m(cx) + ... + m(c^N x)) \end{aligned}$$

Which proves that T is bounded above by a constant.
Method 2:

$$\begin{aligned} &f_{N+1}(x) = \sup_{0\le y\le x} T(f_N, y) \\ &f(x) \ge \sup_{0\le y\le x} T(f_N, y) \\ &f(x) \ge T(f_N, y) \forall y \in [0, x] \\ &f(x) \ge T(f, y) \text{ as } N \to \infty \\ &\therefore \boxed{f(x) = \sup_{0\le y\le x} T(f, y)} \end{aligned}$$

Also,

$$\begin{aligned} T(f_N, y) &= g(y) + h(x-y) + f_N(ay + b(x-y)) \\ &\le g(y) + h(x-y) + f(ay + b(x-y)) \\ &= T(f, y) \end{aligned}$$

$$\begin{aligned} f_{N+1}(x) &= \sup_{0\le y\le x} T(f_N, y) \\ &\le \sup_{0\le y\le x} T(f, y) \\ f_{N+1}(x) &\le \sup_{0\le y\le x} T(f, y) \end{aligned}$$

As $N \to \infty$,

$$\boxed{f(x) \le \sup_{0\le y\le x} T(f, y)}$$

$$\therefore f(x) = \sup_{0\le y\le x} T(f, y) \text{ using the above two inequalities}$$

■

1.2.1. Existence and Uniqueness with g and h not necessarily non-negative

We can write,

$$\begin{aligned} f_1(x) &= \max_{0\le y\le x}[g(y) + h(x-y)] \\ &= \max_{y\in[0,x]}[g(y) + h(x-y) + 0] \\ &= \max_{y\in\Gamma(x)}[g(y) + h(x-y)] \qquad\qquad \Gamma(x) = [0, x] \text{ a set valued map} \end{aligned}$$

Digression

We will use a generalized form of Berge Maximum Theorem in the following proof, which is described here.
Berge Maximum Theorem (Generalized Form)
Let Θ and $\mathbb{X}$ be two metric spaces and $\Gamma : \Theta \rightrightarrows \mathbb{X}$ is a compact valued correspondence. Let $\phi : \mathbb{X} \times \Theta \to \mathbb{R}$ be a continuous function.
Let $\sigma(\theta) = argmax\{\phi(x, \theta) : x \in \Gamma(\theta)\}$ and let $\phi^*(\theta) = \max_x\{\phi(x, \theta) : x \in \Gamma(\theta)\}$

Let $\Gamma : \Theta \rightrightarrows \mathbb{X}$ is continuous at some $\theta \in \Theta$. Then,

1. $\sigma : \Theta \rightrightarrows \mathbb{X}$ is compact-valued, upper semi-continuous and closed at θ.
2. $\phi^* : \Theta \to \mathbb{R}$ is continuous at θ.

Some definitions that follow from above are:

1. **Upper semi-continuous Map:** Let $\Gamma : \mathbb{X} \rightrightarrows \mathbb{Y}$, then Γ is upper semi-continuous at $x \in \mathbb{X}$ if for any $\{x^m\} \in \mathbb{X}$ and $\{y^m\} \in \mathbb{Y}$ with $x_0^m \to x$ and $y^m \in \Gamma(x^m)$, the sequence $\{y^m\}$ has a convergent subsequence such that its limit is in $\Gamma(x)$.
2. **Compact valued Map:** $\Gamma : \mathbb{X} \rightrightarrows \mathbb{Y}$ is a compact valued correspondence which is upper semi-continuous if for a compact subset $S \subset \mathbb{X}$, the set $\Gamma(S)$ is compact.
3. **Closed Map:** $\Gamma : \mathbb{X} \rightrightarrows \mathbb{Y}$ is closed at $x \in \mathbb{X}$ if for any $\{x^m\} \in \mathbb{X}$ with $x^m \to x$ and $\{y^m\} \in \mathbb{Y}$ with $y^m \to y$ such that $y^m \in \Gamma(x^m)$ for each m, we have $y \in \Gamma(x)$.

By applying the generalized form of Berge Maximum Theorem, we can show that $f_1(x)$ is continuous.

$T(f_N, y) = g(y) + h(x - y) + f_N(ay + b(x - y))$ and $f_{N+1}(x) = \max_{0 \le y \le x} T(f_N, y)$.

$$\begin{aligned} f_{N+1}(x) = T(f_N, y_N) &\geq T(f_N, y_N + 1) \\ f_{N+2}(x) = T(f_N + 1, y_N + 1) &\geq T(f_N + 1, y_N) \end{aligned}$$

Define,

$$u_N(x) = \sup_{0 \le z \le x} |f_N(z) - f_N + 1(z)|$$

and

$$\begin{aligned} f_{N+1}(x) - f_{N+2}(x) &\geq T(f_N, y_{N+1}) - T(f_N + 1, y_{N+1}) \\ f_{N+1}(x) - f_{N+2}(x) &\leq T(f_N, y_N) - T(f_N + 1, y_N) \end{aligned}$$

From these two inequalities, we get

$$|f_{N+1}(x) - f_{N+2}(x)| \leq \max\{|T(f_N, y_{N+1}) - T(f_N + 1, y_{N+1})|, |T(f_N, y_N) - T(f_N + 1, y_N)|\}$$

$$|T(f_N, y_{N+1}) - T(f_N + 1, y_{N+1})| = |f_N(ay_N + b(x - y_N)) - f_{N+1}(ay_{N+1} + b(x - y_{N+1})) \leq u_N(cx)$$

$$|T(f_N, y_N) - T(f_N + 1, y_N)| = |f_N(ay_N + b(x - y_N)) - f_{N+1}(ay_N + b(x - y_N)) \leq u_N(cx)$$

$\because ay + b(x - y) \leq cx$ where $c = \max\{a, b\}$

$$\therefore \boxed{u_{N+1}(x) \leq u_N(cx)} \text{ for } N \in \mathbb{N}$$

We now need to estimate $u_1(x)$

$$\begin{aligned} f_2(x) &= T(f_1, y_1) \\ f_1(x) - f_2(x) &\geq g(y_1) + h(x - y_1) - T(f_1, y_1) \\ f_2(x) - f_1(x) &\leq f_1(ay_1 + b(x - y_1)) \end{aligned}$$

Let y_0 be the maximizing point to obtain $f_1(x)$ i.e. $f_1(x) = g(y_0) + h(x - y_0)$ and $f_2(x) \geq T(f_1, y_0)$

$$\begin{aligned} f_1(x) - f_2(x) &\geq g(y_0) + h(x - y_0) - T(f_1, y_0) \\ &= -f_1(ay_0 + b(x - y_0)) \\ f_2(x) - f_1(x) &\geq f_1(ay_0 + b(x - y_0)) \end{aligned}$$

$$|f_2(x) - f_1(x)| \leq \max\{f_1(ay_1 + b(x - y_1)), |f_1(ay_0 + b(x - y_0))|\} \implies |f_2(x) - f_1(x)| \leq 2m(cx) \forall z \in [0, x]$$

$$\therefore \boxed{u_1(x) \leq 2m(cx)}$$

$$\begin{gathered} u_{N+1}(x) \leq u_N(cx) \leq u_{N-1}(c^2x) \leq \ldots \leq u_1(c^N x) \leq 2m(c^{N+1}x) \\ \therefore u_{N+1}(x) \leq 2m(c^{N+1}x) \text{ and } u_1(x) \leq 2m(cx) \text{ for } N \in \mathbb{N} \\ \implies \boxed{u_N(x) \leq 2m(c^N x) \text{ for } N \in \mathbb{N}} \end{gathered}$$

From the above result we can say that,

$$\sum_{N=1}^{\infty} u_N(x) \leq \sum_{N=1}^{\infty} 2m(c^N x)$$

that is

$$\begin{gathered} \sum_{N=1}^{\infty} u_N(x) \leq \infty (\text{convergent}) \\ \implies u_N(x) \to 0 \text{ as } N \to \infty \end{gathered}$$

$\implies f_N(x)$ is a cauchy sequence (which then implies that $f_N - f_M$ can be expressed as a sum of differences).

$$\therefore f(x) = \lim_{N \to \infty} f_N(x) \text{ exists and is continuous } \forall x \geq 0$$

Now we know that,

$$\begin{aligned} f_{N+1}(x) &\geq T(f_N, y) \\ &\geq g(y) + h(x - y) + f_N(ay + b(x - y)) \text{ for } y \in [0, x] \end{aligned}$$

For large N,

$$\begin{aligned} f(x) &\geq g(y) + h(x - y) + f(ay + b(x - y)) \\ &\geq \max_{0 \leq y \leq x} T(f, y) \end{aligned} \qquad (A)$$

For a bounded sequence $\{y_N\}$ with $y_N \to y^*$ for some $y^* \in [0, x]$ and $N \to \infty$ we have,

$$\begin{aligned} f(x) &= T(f, y^*) \\ &\leq \max_{0 \leq y \leq x} T(f, y) \qquad (B) \end{aligned}$$

using (A) and (B),

$$\boxed{f(x) = \max_{0 \leq y \leq x} T(f, y)}$$

1.2.2. Uniqueness

By proving the existence, we know that,

$$f(x) = \max_{0 \leq y \leq x} T(f, y)$$

Assume on the contrary that there are two such solutions $f \& F$

$$\begin{aligned} f(x) = T(f, y) \geq T(f, w) \\ F(x) = T(F, w) \geq T(F, y) \end{aligned}$$

Now,

$$\begin{aligned} |f(x) - F(x)| &\leq \max\{|T(f, y) - T(F, y)|, |T(f, w) - T(F, w)|\} \\ &\leq \max\{|f(ay + b(x - y) - F(ay + b(x - y))|, |f(aw + b(x - w) - F(aw + b(x - w))|\} \end{aligned}$$

From previous definition, $u(x) = \sup_{0 \leq z \leq x} f(z) - F(z)$ We know that $u(x) \geq 0$ and $u(x) \to 0$ as $x \to \infty$.

$$\begin{aligned} \therefore 0 \geq u(x) \geq |f(x) - F(x)| \\ \implies |f(x) - F(x)| \to 0 \text{ as } x \to \infty \\ \implies f(x) = F(x) \text{ for all } x \geq 0 \end{aligned}$$

$\therefore$ The solution obtained is unique.

1.3. Approximations

Theorem 1.1. Let $f_0(x)$ satisfy the following conditions:

- $f_0(x)$ is continuous for $x \geq 0$
- $f_0(0) = 0$

Then if the conditions of existence are fulfilled, the sequence defined by

$$f_{n+1}(x) = \max_{0 \leq y \leq x} T(f_n, y) \text{ for } n \in \mathbb{W}$$

converges to the solution $f(x)$ of the problem, uniformly in any finite interval.

1.3.1. Approximation in the policy space

We call a sequence of allocations, i.e. a sequence of admissible choices of y, a policy and a policy which yields $f(x)$ an optimal policy.
The duality that exists in the theory of dynamic programmingarises from the interconnection between the functions $f(x)$ which measure the maximum return and the policies which yield these maximum returns. Actually a policy is a function, since a policy is a determination of y as a function of x. If the policy is not unique, y will not be a single-valued function of x.
Just as we can approximate in the space of the functions $f(x)$, so we can approximate in the space of policies, $y(x)$.

Theorem 1.2. Let $f_0(x)$ be the result of an initial approximation in the policy space, that is, $f_0(x) = T(f_0, y_0(x))$ where $y_0(x)$ is any continuous function of x satisfying $0 \leq y_0(x) \leq x$, the sequence defined by

$$f_{n+1}(x) = \max_{0 \leq y \leq x} T(f_n, y) \text{ for } n \in \mathbb{N}$$

converges to the solution $f(x)$, uniformly and in a finite interval.

Proof. By the definition of $f_n(x)$, we have

$$f_{n+1} \geq f_n \geq f_0 \ \forall \ \mathbb{N}$$

Proving that $f_0(x)$ is continuous, will imply continuity of $f_n(x)$.

$$f_0(x) = g(y_0) + h(x - y_0) + \cdots$$

where g and h are continuous functions.
Hence, $f_0(x)$ which is iteratively obtained from $g(x)$ and $h(x)$ is also continuous in a finite interval. ■

We can construct examples for such a solution, say

Example. $f(x) = T(f, y(x))$ be an iterative form of the solution.
If the optimal policy consisted of the choice $y = 0$ continually, the solution can be represented by the equation $f(x) = h(x) + f(bx)$, with the value of b such that $h(b^n x) < \infty$
i.e.,

$$\begin{aligned} f(x) &= h(x) + f(bx) \\ &= h(x) + h(bx) + f(b^2 x) \\ &= h(x) + h(bx) + h(b^2 x) + f(b^3 x) \\ &= \cdots \\ &= \sum_{n=0}^{\infty} h(b^n x) \end{aligned}$$

Here, $h(b^n x) < \infty$ for all $n \in \mathbb{N}$, so the series converges and $f(x)$ is a continuous function of x.

1.4. Properties of the solution

1.4.1. Convexity

Theorem 1.3. If, along with existence assumptions, we impose the conditions that g and h be convex functions of x, then $f(x)$ will be a convex function, and for each value of x, y will equal 0 or x.

Proof. g and h are convex functions and,

$$f_1(x) = \max_{0 \leq y \leq x} (g(y) + h(x - y))$$

In the interval $[0, x]$, g and h are convex, hence $g(y) + h(x - y)$ is also convex.
$\because$ Maximum of a convex function must occur at the endpoints of the interval, i.e. $y = 0$ or $y = x$. We can say that,

$$f_1(x) = \max(g(y), h(x))$$

Then,

$$f_2(x) = \max_{0 \le y \le x} (g(y) + h(x - y) + f_1(ay + b(x - y)))$$

similarly follows,

$$f_2(x) = \max(g(x) + f_1(ax), h(x) + f_1(bx))$$

Therefore we can inductively conclude that $f_n(x)$ is convex thus, the limit of the sequence $f(x)$ is also convex. ∎

1.4.2. Concavity

Theorem 1.4. [Weak Version] If, along with existence assumptions, we impose the conditions that g and h be strictly concave functions of x, then $f(x)$ will be a strictly concave function, with an optimal policy which is unique.

Lemma 1.5. If $G(x, y)$ is a concave function of x and y for $x, y \ge 0$, then $f(x)$ as defined by $f(x) = \max\limits_{0 \le y \le x} G(x, y)$ is a concave function of x for $x > 0$.

Theorem 1.6. [Strong Version] Assuming that,

1. $g(x)$ and $h(x)$ are both strictly concave for $x > 0$, monotone increasing with continuous derivatives and that $g(0) = h(0) = 0$.
2. $\dfrac{g'(0)}{(1-a)} > \dfrac{h'(0)}{(1-b)}, h'(0) > g'(\infty), b > a$.

Then the optimal policy has the following form:

1. $y = x$ for $0 < x < \bar{x}$, where $\bar{x}$ is the root of $h'(0) = g'(x) + (b - a)g'(ax) + (b - a)ag'(a^2x) + \cdots$
2. $y = y(x)$ for $x > \bar{x}$ where $y(x)$ is a function satisfying the inequalities $0 < y(x) < x$, and $y(x)$ is the solution of $\boxed{g'(y) - h'(x - y) + (a - b)f'(ay + b(x - y)) = 0}$

A Result on Concavity

Let $g(x) = \sup\limits_{0 \le y \le x} \phi(x, y)$ and $\epsilon > 0$ $\exists y_1 \in [0, x]$ such that,

$$\phi(x_1, y_1) > g(x_1) - \epsilon$$
$$\phi(x_2, y_2) > g(x_2) - \epsilon$$

$$\begin{aligned}
\therefore \lambda(\phi(x_1, y_1)) + (1 - \lambda)(\phi(x_2, y_2)) &> \lambda g(x_1) + (1 - \lambda)g(x_2) - \epsilon \\
\implies \phi(\lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2)) &> \lambda g(x_1) + (1 - \lambda)g(x_2) - \epsilon \\
\implies \phi(\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2) &> \lambda g(x_1) + (1 - \lambda)g(x_2) - \epsilon
\end{aligned}$$

We can also say that,

$$\begin{aligned}
\max_{0 \le y \le \lambda x_1 + (1-\lambda)x_2} \phi(\lambda x_1 + (1 - \lambda)x_2, y) &\ge \lambda g(x_1) + (1 - \lambda)g(x_2) - \epsilon \\
\implies g(\lambda x_1 + (1 - \lambda)x_2) &\ge \lambda g(x_1) + (1 - \lambda)g(x_2) - \epsilon \\
\implies g \text{ is concave.}
\end{aligned}$$

Proof. It is known that,

$$f_1(x) = \max_{0 \le y \le x} (g(y) + h(x - y))$$

$\because g'(0) > h'(0) \implies g'(y) - h'(x-y) > 0$ for $y \in [0, x]$ with the roots at $y = x$ and in the interval $[0, x'], \forall x' > y$
At $y = x$, the value x^* for which $g'(x^*) - h'(0) = 0$ is the root of the equation $g'(x) - h'(0) = 0$.
This equation has precisely one solution due to the fact that g and h are strictly concave.
For $x \ge x^*$, let $y_1(x)$ be the unique solution of $g'(y) = h'(x - y)$. Then, $f_1(x) = g(y_1) + h(x - y_1)$ and $f'(x) = [g'(y_1) - h'(x - y_1)]\frac{dy_1}{dx} + h'(x - y_1) = h'(x - y_1)$ for $x > x_1$.
Further,

$$f_2(x) = \begin{cases} g'(x), & \text{if } 0 \le x \le x_2 \\ h'(x - y_2) + bf_1(ay_2 + b(x - y_2)), & \text{if } x > x_2 \end{cases}$$

where $y_2(x)$ is the unique solution of $g'(y) = h'(x - y) + (b - a)f_1(ay + b(x - y))$ Inductively,

$$f_{n+1}(x) = \max_{0 \le y \le x} \{g(y) + h(x - y) + f_n(ay + b(x - y))\}$$

The solution space hence can be divided into three parts from, $[0, x_2]$, $[x_2, x_1]$ and $[x_1, \infty)$, wherein the inequality $f'_{n+1}(x) \ge f'_n(x)$ holds for all x.
For $n = 1$,

$$f'_2(x) = \frac{bg'(y_2) - ah'(x - y_2)}{b - a} \text{ for } x \ge x_2$$

and

$$f'_1(x) = \frac{bg'(y_1) - ah'(x - y_1)}{b - a} \text{ for } x \ge x_1$$

For $[x_1, \infty)$, and f' is monotonically decreasing and $y_2 < y_1$, $f'_2(x) > f'_1(x)$. In, $[0, x_2]$, $f'_2(x) = f'_1(x)$.

In $[x_2, x_1]$,

$$f'_1(x) = g'(x)$$

$$f'_2(x) = \frac{bg'(y_2) - ah'(x - y_2)}{b - a} \text{ for } x \ge x_2$$

Hence, in the interval, $0 \le y_2 \le x$ for $x \in [x_2, x_1]$,

$$f'_2(x) \ge \frac{bg'(x) - ah'(0)}{b - a} > g'(x)$$

And we assume that, $g'(x) \ge h'(0) \therefore f'_2(x) > f'_1(x)$
Or, for $x_1 > x_2 > x_3 > \cdots > x_n > 0$, $f'_1(x) \le f'_2(x) \le f'_3(x) \le \cdots \le f'_n(x) \le \cdots$

Since $f_n(x)$ converges to $f(x)$, $f'_n(x)$ to $f'(x)$, $y_n(x)$ to $y(x)$ and x_n to $\bar{x}$, the solution has the form,

$$f(x) = \max_{0 \le y \le x} \{g(y) + h(x - y) + f(ay + b(x - y))\}$$

$$\frac{df}{dy} = 0 \implies g'(y) - h'(x - y) + (a - b)f'(ay + b(x - y)) = 0$$

■

1.5. Examples

Example 1. Show that, The continuous solution of $f(x) = \max[cx^d + f(ax), ex^g + f(bx)], f(0) = 0$, subject to

1. $a, b \in (0,1)$, $c, d, e, g > 0$,
2. $0 < d < g$

is given by,

$$f(x) = \begin{cases} \frac{cx^d}{1-a^d}, & \text{if } 0 \le x \le \bar{x} \\ ex^g + f(bx), & \text{if } x \ge \bar{x} \end{cases}$$

where, $\bar{x} = \left(\dfrac{c/(1-a^d)}{e/(1-b^d)}\right)^{\frac{1}{g-d}}$.

Solution: Assuming, A as the choice $cx^d + f(ax)$ and B as the choice $ex^g + f(bx)$, we have any solution of the form $A^{k_1^t}B^{k_2^t}$ or the number of times A and B are chosen to reach a solution at the t^{th} step.
At optimal time, represented by say ∞, any choice of A or B would not change the solution. Mathematically,

$$BA^\infty = A^\infty$$

Now,

$$\begin{aligned} A^\infty &= cx^d + f(ax) = cx^d + c(ax)^d + f(a^2x) + f(a^3x) + \cdots \\ &= \frac{cx^d}{(1-a^d)} \end{aligned}$$

Similarly, $BA^\infty \implies f(x) = ex^g + \frac{cb^dx^d}{(1-a^d)}$.
Equating them we get,

$$ex^g + \frac{cb^dx^d}{(1-a^d)} = \frac{cx^d}{(1-a^d)} \implies x = \left(\frac{c/(1-a^d)}{e/(1-b^d)}\right)^{\frac{1}{g-d}}$$

The value of x is the optimal x, $\bar{x}$.
Case I: When $0 \le x \le \bar{x}$ We have to show that $\frac{cx^d}{1-a^d}$ is the solution. At $t = \infty$,

$$\max\left[\frac{cx^d}{(1-a^d)}, ex^g + \frac{cb^dx^d}{(1-a^d)}\right] = \frac{cx^d}{(1-a^d)}$$

This is true for small x when $g > d > 0$ and $b \in (0,1)$. If we proceed for more values of x, the smallest value at which solution B becomes optimal is attained at $BA^\infty = A^\infty$ which occurs at $x = \bar{x}$ as calculated above.

Case II: When $x \ge \bar{x}$, we have to show that $ex^g + f(bx)$ is the solution. At $t = \infty$,
Define, $f_{AB}(x) = cx^d + ea^gx^g + f(abx)$ and $f_{BA}(x) = ex^g + cb^dx^d + f(abx)$. These are compositions of the two choices A and B. It can be interpreted as the dominant choice at each step, as f_{AB} will signify the choice of A at the first step and B at the second step and so on. Similarly, f_{BA} will signify the choice of B at the first step and A at the second step and so on.

Now, their point of intersection occurs at $p = \left(\frac{c/(1-b^d)}{e/(1-a^g)}\right)^{\frac{1}{g-d}}$.
Now, $p < \bar{x}$ as $g > d$. At this point, $f_{AB}(x) < f_{BA}(x)$ for $x > p$ for $x > \bar{x}$. This arises a contradiction.
$\therefore f(x) = ex^g + f(bx)$ is the optimal choice only when $x \ge \bar{x}$.

Example 2. Let us define the function $f_N(a) = \max_R[x_1 x_2 \cdots x_n]$ where R is the region determined by the conditions

1. $x_1 + x_2 + \cdots + x_n = a, a > 0$
2. $x_i \geq 0$

Prove that $f_N(a)$ satisfies the recurrence relation $f_N(a) = \max\limits_{0 \leq x \leq a} x f_{N-1}(a - x), N \geq 2$ with $f_1(a) = a$.

Solution: We can solve this using mathematical induction.
$N = 1$: $f_1(a) = \max[x_1] = a$ as $x_1 = a$. Hence the relation holds.
Let $N = k$ be true. That is, $f_k(a) = \max\limits_{0 \leq x \leq a} x f_{k-1}(a - x)$.
For $N = k + 1$,

$$\begin{aligned} f_{k+1}(x) &= \max[x_1 x_2 \cdots x_{k+1}] \\ &= \max[\max[x_1 x_2 \cdots x_k] x_{k+1}] \\ &= \max[x_{k+1} f_k(a - x_{k+1})] \qquad \text{as } x_1 + x_2 + \cdots + x_k = a - x_{k+1} \end{aligned}$$

$\because$ The relationship hold for all $N \in \mathbb{N}$. The recurrence relation holds.
This completes the proof.

Example 3. For the previous f, show inductively that $f_N(a) = \frac{a^N}{N^N}$, and hence establish the arithmetic-geometric mean inequality, for $x_i > 0$.

Solution: Given, $f_N(a) = \max\limits_{0 \leq x \leq a} x f_{N-1}(a - x)$.
For $N = 1$ $f_1(a) = a = \frac{a^1}{1^1}$ which is true.
Let $f_k(a) = \frac{a^k}{k^k}$ hold as well.
For, $N = k + 1$, $f_{k+1}(a) = \max\limits_{0 \leq x \leq a} x f_k(a - x) = \max\limits_{0 \leq x \leq a} x \frac{(a-x)^k}{k^k}$ which is maximized at $x = \frac{a}{k+1}$ resulting in $f_{k+1}(x) = \frac{a^{k+1}}{(k+1)^{k+1}}$
$\because$ The relationship hold for all $N \in \mathbb{N}$. The recurrence relation holds.
This completes the proof.

Example 4. Define a function,

$$f_N(a) = \min_R \sum_{i=1}^{N} x_i^p, p > 0$$

where R is a region defined by,

1. $\sum\limits_{i=1}^{N} x_i \geq a, a > 0$
2. $x_i \geq 0$

Show that $f_N(a)$ satisfies the recurrence relation

$$f_N(a) = \min_{0 \leq x \leq a} [x^p + f_{N-1}(a - x)], N \geq 2 \text{ with } f_1(a) = a^p$$

Solution: We can solve this inductively as,
For $N = 1$, $f_1(a) = a^p$, which is true.
Let for $N = k$, $f_k(a) = \min\limits_{0 \leq x \leq a} [x^p + f_{k-1}(a - x)]$
For, $N = k + 1$,

$$
\begin{aligned}
f_{k+1}(a) &= \min_R \sum_{i=1}^{k+1} x_i^p \\
&= \min_R[\min_R \sum_{i=1}^{k} x_i^p + x_{k+1}^p] \\
&= \min_{0\le x\le a}[x_{k+1}^p + f_k(a-x)]
\end{aligned}
$$

$\because$ The relationship hold for all $N \in \mathbb{N}$. The recurrence relation holds.
This completes the proof.

Example 5. Let $f(x)$ and $F(x)$ be the continuous solutions of the above equations under the assumptions that $u(x,y)$ and $v(x,y)$ are continuous in x and $y \forall x, y > 0$, with $0 < a, b < 1$, and that $\sum_{n=0}^{\infty} m(c^n z) < \infty$ where $m(z) = \max_{0\le x\le z}[\max_{0\le y\le x} \max |u(x,y)|, |v(x,y)|]$. If $\max_{0\le x\le z} \max_{0\le y\le x} |u(x,y) - v(x,y)| = D(z)$ and $\sum_{n=0}^{\infty} D(c^n z) < \infty$, $c = \max(a,b)$, show that,

$$|f(x) - F(x)| < \sum_{n=0}^{\infty} D(c^n z)$$

Solution: We define,

$$
\begin{aligned}
f_1(x) &= \max_{0\le y\le x} u(x,y) \\
f_{N+1}(x) &= \max_{0\le y\le x} \max\left\{u(x,y), f_N(ay + b(x-y))\right\} \\
F_1(x) &= \max_{0\le y\le x} v(x,y) \\
F_{N+1}(x) &= \max_{0\le y\le x} \max\left\{v(x,y), F_N(ay + b(x-y))\right\}
\end{aligned}
$$

Also, $\lim_{N\to\infty} f_N(x) = f(x)$ and $\lim_{N\to\infty} F_N(x) = F(x)$.
Now,

$$|f_1(x) - F_1(x)| \le \max_{0\le y\le x} |u(x,y) - v(x,y)| \le D(x)$$

Similarly,

$$
\begin{aligned}
|f_{N+1}(x) - F_{N+1}(x)| &\le \max_{0\le y\le x} \max\left\{|u(x,y) - v(x,y)|, |f_N(ay + b(x-y)) - F_N(ay + b(x-y))|\right\} \\
&\le D(x) + \max_{0\le y\le x} |f_N(ay + b(x-y)) - F_N(ay + b(x-y))|
\end{aligned}
$$

$$\therefore |f_{N+1}(x) - F_{N+1}(x)| \le \sum_{n=0}^{\infty} D(c^n x)$$

When, $N \to \infty$

$$|f(x) - F(x)| \le \sum_{n=0}^{\infty} D(c^n x)$$

This completes the proof.

§2. A Modern Outlook to Dynamic Programming

2.1. Introduction

This chapter discusses a more formal approach to dealing with dynamic programming problems and considers the formalism adapted by Dimitri Bertsekas in his book Dynamic Programming and Optimal Control Volume 1. In this chapter, we will be setting up problems as formal dynamic programming problems and discuss approaches to find solutions to those.

2.2. The Basic Model

The basic model assumed by Bertsekas takes two assumptions into account:

1. The underlying dynamic time system is discrete in nature.
2. There is a cost function that is additive over time.

The dynamic system expresses the evolution of some variables referred to as **states**. The system is also under the influence of certain decisions made at discrete time instances.
The system is of the form:

$$x_{k+1} = f_k(x_k, u_k, w_k)$$

where

k	indexes in discrete time,
x_k	state of the system with past information,
u_k	control/decision variable,
w_k	disturbance or noise,
f_k	describes how the system is changing

The cost function helps to formulate that total cost over time t, and hence, it is expected to be additive in nature. The total cost incurred becomes,

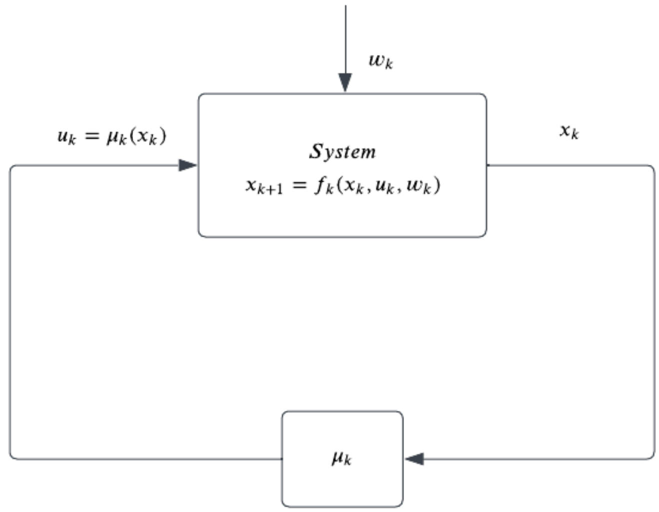
$$g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)$$

Due to w_k the cost is a random variable, hence it cannot be optimized. We, therefore tend to optimize the expected total cost where the expectation is with respect to the joint distribution of the random variables, i.e.,

$$\mathbb{E}\left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)\right]$$

Question 1. How to formulate a problem as a dynamic programming problem?
The basic components towards formulating the problem would be:

1. Deducing the problem as a discrete-time system of the form $x_{k+1} = f_k(x_k, u_k, w_k)$.
2. *Independent random parameters* w_k. This will be generalized by allowing the probability distribution of w_k to be dependent on x_k and u_k.
3. A *control constraint*. The constraint set depends on x_k and $u_k \in U_k(x_k)$.
4. A *additive cost* of the form:
$$\mathbb{E}\left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)\right]$$
5. *Optimization over policies*, generally closed-loop. In layman's terms, it means the rule for choosing u_k for each k and each possible value of x_k.



Information gathering in the basic problem. At each time k the controller observes the current state x_k and applies a control $u_k = \mu_k(x_k)$ that depends on that state.

2.3. Discrete-State and Finite-State Problems

Generally, there are many situations where the state is naturally discrete and there is no continuous counterpart of the problem. Such situations are often conveniently specified in terms of probabilities of transition between the states, i.e.,

$$p_{ij}(u,k) = P\{x_{k+1} = j | x_k = i, u_k = u\}$$

This type of translate transition can alternatively be described in terms of discrete time system equation $\boxed{x_{k+1} = w_k}$ where the probability distribution of random parameter w_k is

$$P\{w_k = j | x_k = i, u_k = u\} = p_{ij}(u,k)$$

Conversely, given a discrete-state system in the form $x_{k+1} = f_k(x_k, u_k, w_k)$ together with the probability distribution $P_k(w_k | x_k, u_k)$ we can provide an equivalent transition probability description. The corresponding transition probabilities are given by

$$p_{ij}(u,k) = P\{W_k(i,u,j) | x_k = i, u_k = u\}$$

where the $W_k(i,u,j)$ is the set,

$$W_k(i,u,j) = \{w | j = f_k(i,u,w)\}$$

Thus, a discrete state system can equivalently be described in terms of a difference equation or in terms of transition probabilities.

2.4. Solution for the Basic Model

For a discrete-time dynamic system given as $x_{k+1} = f_k(x_k, u_k, w_k)$ with state $x_k \in S_k$, control $u_k \in C_k$ and a random disturbance $w_k \in D_k$.
The control constraint u_k takes values from a subset $U_k(x_k) \subset C_k$ which depends on x_k; i.e., $u_k \in U_k(x_k) \forall x_k \in S_k$.
We consider the class of policies that consist of a sequence of functions.

$$\pi = \{\mu_0, \dots, \mu_{N-1}\}$$

where μ_k maps states to control, i.e., $u_k = \mu_k(x_k)$ and such that $\mu_k(x_k) \subset U_k(x_k) \forall x_k \in S_k$ as *admissible policies*. Given an initial state x_0 and an admissible policy π, we can generate the system as $x_{k+1} = f_k(x_k, \mu_k(x_k), w_k)$.

The expected cost then becomes,

$$J_\pi(x_0) = \mathbb{E}\left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k)\right]$$

Hence, an optimal policy π^* is one that minimizes this cost J over all $\pi \in \Pi$, i.e.,

$$\boxed{J_{\pi^*}(x_0) = \min_{\pi\in\Pi} J_\pi(x_0)}$$

Note: The optimal cost here depends upon x_0 or the choice of initial state. However, it is generally possible to find a policy π^* optimal for all choices of $x_0 \in S_k$.
The optimal cost, hence is,

$$\boxed{J^*(x_0) = \min_{\pi\in\Pi} J_\pi(x_0)}$$

Here, J^* is a function that assigns an optimal cost to each initial state $x_0 \in S_k$ and is called the optimal cost function or the optimal value function.

Note: The min used in the description of the optimal value denotes the greatest lower bound of the set $\{J_\pi(x_0)|\pi \in \Pi\}$ or $J^*(x_0) = \inf_{\pi\in\Pi} J_\pi(x_0)$

2.5. Principle of Optimality

Intuitively speaking, a concept of truncated policy is introduced here. We start with a truncated policy $\mu_i^*, \mu_{i+1}^*, \ldots, \mu_{N-1}^*$. If it is not optimal, we reduce it further by switching to another policy in the space $\subset U_k(x_k)$ until we reach an optimal policy for a subproblem. This means that the optimal policy can be *constructed* in a piece-wise fashion, by constructing optimal policies for tail subproblems first. Formally,

Principle of Optimality

Let $\pi^* = \{\mu_0^*, \mu_1^*, \ldots, \mu_{N-1}^*\}$ be an optimal policy for the basic problem, and assume that when using π^*, a given state x_i occurs at time i with a positive probability. Consider the subproblem whereby we are at x_i at time i and wish to minimize the "cost-to-go" from time i to time N:

$$\mathbb{E}\left[g_N(x_N) + \sum_{k=i}^{N-1} g_k(x_k, \mu_k(x_k), w_k)\right]$$

Then the truncated policy $\mu_i^*, \mu_{i+1}^*, \ldots, \mu_{N-1}^*$ is optimal for this subproblem.

The algorithm for solving the Basic Model revolves around the Principle of Optimality, or breaking the optimization into subproblems. It can be stated as follows,

Proposition 2.1. For every initial state x_0, the optimal cost $J^*(x_0)$ of the basic problem is equal to $J_0(x_0)$, given by the following algorithm, which proceeds backward in time from period $N-1$ to period 0:

$$J_N(x_N) = g_N(x_N) \tag{2.1}$$

$$J_k(x_k) = \min_{u_k\in U_k(x_k)} \mathbb{E}_{w_k}\left[g_k(x_k, \mu_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k))\right], k = 0, 1, \ldots, N-1, \tag{2.2}$$

where the expectation is taken with respect to the probability distribution of w_k, which depends on x_k and u_k. Furthermore, if $u_k^* = \mu_k^*(x_k)$ minimizes the right side of Equation (2.2) for each x_k and k, the policy $\pi^* = \{\mu_0^*, \mu_1^*, \ldots, \mu_{N-1}^*\}$ is optimal.

2.6. Some Problems

Problem 2.6.1. (Exponential Cost Function)
Given a cost function of the form:

$$\mathbb{E}_{w_k}\left[\exp\left(g_N(x_N)+\sum_{k=0}^{N-1} g_k(x_k,u_k,w_k)\right)\right]$$

Show that the optimal cost and optimal policy can be obtained from the DP-like algorithm:

1. For $k=0,1,\ldots,N-1$, define the optimal cost functions as follows:

$$\begin{aligned} J_N(x_N) &= \exp(g_N(x_N)) \\ J_k(x_k) &= \min_{u_k\in U_k(x_k)} \mathbb{E}_{w_k}\left[J_{k+1}\left(f_k(x_k,u_k,w_k)\right)\exp\left(g_k(x_k,u_k,w_k)\right)\right] \end{aligned}$$

2. Define the function $V_k(x_k)$ as the natural logarithm of $(J_k(x_k)$:

$$V_k(x_k)=\ln(J_k(x_k))$$

   Assuming that g_k depends only on x_k and u_k (not on w_k), show that the above algorithm can be rewritten as:

$$\begin{aligned} V_N(x_N) &= g_N(x_N) \\ V_k(x_k) &= \min_{u_k\in U_k(x_k)}\left\{g_k(x_k,u_k)+\ln\mathbb{E}_{w_k}\left[\exp\left(V_{k+1}(f_k(x_k,u_k,w_k))\right)\right]\right\} \end{aligned}$$

Solution.
We know that the exponential function is strictly increasing in nature. We solve this problem by induction.

$$\begin{aligned} \pi^k &= \{\mu_k,\mu_{k+1},\ldots,\mu_{N-1}\}\forall\mu_k(x_k)\subset U_k(x_k) \\ J_k^*(x_k) &= \min_{\pi_k}\mathbb{E}_{w_i}\left(\exp\left(g_N(x_N)+\sum_{i=k}^{N-1} g_i(x_i,\mu_i,w_i)\right)\right) \\ J_N^*(x_N) &= \exp(g_N(x_N)) \end{aligned}$$

At $k=N$,

$$\begin{aligned} J_N^*(x_N) &= \min_{\pi_N}\mathbb{E}_{w_N}\left(\exp\left(g_N(x_N)+0\right)\right) \\ &= \exp(g_N(x_N)) = J_N(x_N) \quad \text{(by definition)} \end{aligned}$$

Assuming the identity to be true for some k and all x_{k+1}, $J_{k+1}^*(x_{k+1})=J_{k+1}(x_{k+1})$
$\therefore$ For x_k,

$$\begin{aligned} J_k^*(x_k) &= \min_{\pi_k}\mathbb{E}_{w_i}\left(\exp\left(g_N(x_N)+\sum_{i=k+1}^{N-1} g_i(x_i,\mu_i,w_i)+g_k(x_k,\mu_k,w_k)\right)\right) \\ &= \min_{\pi_{k+1}}\mathbb{E}_{w_i}\left(\exp\left(g_N(x_N)+\sum_{i=k+1}^{N-1} g_i(x_i,\mu_i,w_i)\right)\right)\cdot\min_{\mu_k}\mathbb{E}_{w_k}\left(\exp g_k(x_k,\mu_k,w_k)\right) \quad \text{(Principle of Optimality)} \\ &= J_{k+1}^*(x_{k+1})\cdot\min_{\mu_k}\mathbb{E}_{w_k}\left(\exp g_k(x_k,\mu_k,w_k)\right) \\ &= \min_{\mu_k}\mathbb{E}_{w_k}\left(J_{k+1}^*(x_{k+1})\cdot\exp g_k(x_k,\mu_k,w_k)\right) \\ &= \min_{\mu_k\in\mathbb{U}_k}\mathbb{E}_{w_k}\left(J_{k+1}(x_{k+1})\cdot\exp g_k(x_k,\mu_k,w_k)\right) = J_k(x_k) \end{aligned}$$

Hence, the proof is complete.

Part (b),
For $V_N(x_N)$:

$$\begin{aligned} V_N(x_N) &= \ln J_N(x_N) \\ &= \ln\Big(\exp\left(g_N(x_N)\right)\Big) \\ &= g_N(x_N) \end{aligned}$$

For $V_k(x_k)$:

$$\begin{aligned} V_k(x_k) &= \ln J_k(x_k) \\ &= \ln\left(\min_{u_k \in U_k} \mathbb{E}_{w_k}\Big(J_{k+1}(x_{k+1}) \cdot \exp\left(g_k(x_k, u_k)\right)\Big)\right) \end{aligned}$$

We can move the $\ln$ term inside the minimum as it is increasing and positive,

$$\begin{aligned} V_k(x_k) &= \min_{u_k \in U_k} \ln\Big(\exp\left(g_k(x_k, u_k)\right) \cdot \mathbb{E}_{w_k}\left(J_{k+1}(x_{k+1})\right)\Big) \\ &= \min_{u_k \in U_k} \Big\{g_k(x_k, u_k) + \ln\Big(\mathbb{E}_{w_k}\left(J_{k+1}(x_{k+1})\right)\Big)\Big\} \\ &= \min_{u_k \in U_k(x_k)} \Big\{g_k(x_k, u_k) + \ln \mathbb{E}_{w_k}\Big[\exp\left(V_{k+1}(f_k(x_k, u_k, w_k))\right)\Big]\Big\} \end{aligned}$$

This completes the proof for part (b).

Problem 2.6.2. (Minimization over a Subset of Policies)
Consider a variation of the basic problem whereby we seek

$$\min_{\pi \in \tilde{\Pi}} J_\pi(x_0)$$

, where $\tilde{\Pi}$ is some given subset of the set of sequences $\{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$ of functions $\mu_k : S_k \to C_k$ and $\mu_k \subset \mathcal{U}_k(k)$ for all $k \in S_k$. Assume that $\pi^* = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$ belongs to $\tilde{\Pi}$ and attains the minimum in the DP algorithm; that is, for all $k = 0, 1, \ldots, N-1$ and $x_k \in S_k$:

$$J_k(x_k) = \min_{\mu_k(x_k) \in \mathcal{U}_k(x_k)} \mathbb{E}\{g_k(x_k, \mu_k(x_k), w_k) + J_{k+1}(f_k(x_k, \mu_k(x_k), w_k))\}$$

with $J_N(x_N) = g_N(x_N)$. Assume further that the functions J_k are real-valued, and that the preceding expectations are well-defined and finite. Show that π^* is optimal within $\tilde{\Pi}$ and that the corresponding optimal cost is $J_0(x_0)$.

Solution.
We look at $\mu_k(x_k) \in \tilde{\Pi} \subset \mathcal{U}_k(x_k)$,

$$J_k(x_k) = \min_{\mu_k(x_k) \in \tilde{\Pi}} \mathbb{E}\{g_k(x_k, \mu_k(x_k), w_k) + J_{k+1}(f_k(x_k, \mu_k(x_k), w_k))\}$$

$\because \pi^*$ is optimal within $\mathcal{U}_k(x_k) \forall k \in \mathbb{N}$

$\implies$ We can always pick a subset with $\mu_k^* \in \pi^*$ such that μ_k^* is also optimal control in $\mathcal{U}_k(x_k)$.

Within $\tilde{\Pi}$, we can say that,

$$J_k(x_0) \leq J_k^*(x_0) + (N-1)\epsilon \quad \text{(for some } \epsilon\text{)}$$

We also have,

$$J_k^*(x_k) \leq J_k^i(x_k) \quad \forall x_k, i \in \mathbb{W}$$

Using $i = 0$ we get,

$$J_0^*(x_k) \leq J_k^0(x_k) = J_0^*(x_k)$$

This situation is similar to optimization in a controlled environment, i.e., $\mu_k \in \Omega_k(x_k) \subset \mathcal{U}_k(k)$
$\therefore \pi' = [\mu_0', \ldots, \mu_{N-1}']$

$$\therefore J' = \min_{\mu_0} \left[g_0 + \min_{\mu_1} \left[g_1 + \ldots \min_{\mu_{N-1}} [g_{N-1}] \right] \right] + g_N(x_N)$$

$$\implies J_{N-1}'(x_{N-1}) = \min_{\mu_{N-1}} [g_{N-1}] + g_N(x_N) \quad \text{and,}$$

$$\implies J_0'(x_0) = \min_{\mu_0} [g_0 + J_1'(x_1)]$$

$$\implies \boxed{J' = J_0'(x_0)} \quad \forall \mu_k \in \Omega_k(x_k)$$

This completes the proof.
Some utility of using a subset like that can be seen as,

- $\to$ Each block of min can be seen as a block in a cascade.
- $\to$ Transforming μ from $\mathcal{U}_k(k)$ to $\Omega_k(x_k)$ reduces the number of blocks within the cascade.
- $\to$ This reduction in dimensionality is better for greedy computational methods to work out.

Note: This can also be seen as a simpler example of the Berge Maximum Theorem but in one dimension space where the set Π is reduced and the optimal is still found.

[Bel57] [Ber95]

References

[Bel57] Richard Bellman. *Dynamic Programming*. Dover Publications, 1957.

[Ber95] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 1st edition, 1995.