Dynamic Programming Principle

Supervisor : Alekos Cecchin

Paul-Emile Galine
21/07/2023

Università
degli Studi
di Padova

Introduction

The Dynamic Programming Principle (DPP) was introduced by Richard E. Bellman in 1952, it is a mathematical optimization technique.

- A powerful method based on a simple concept
- Applications in numerous fields : engineering ; economics ; finance ; computer science
- We focus today on control theory and precisely stochastic control

Outline

During the whole presentation we will deal with finite horizon problems : $T < \infty$

Let $\alpha : [0, T] \to A$ the control and $f : \mathbb{R}^n \times A \to \mathbb{R}$ with $A \subset \mathbb{R}^n$

Definition

We call a control system (S) a system such that :

$$\begin{cases} \dot{x}(s) = f(x(s), \alpha(s)) & s > t \\ x(t) = x \end{cases} \tag{S}$$

where $x(\cdot) \in \mathbb{R}^n$

We define the set of admissible control :
$O := \{\alpha : [0, T] \to A$ a measurable control such that (S) admits a unique solution x(·)$\}$
And, we define set of data :

- Lagrangian : $L : \mathbb{R}^n \times A \to \mathbb{R}$
- Terminal cost function : $g : \mathbb{R}^n \to \mathbb{R}$

Our goal is to find the minimum of the cost functional

Definition (Cost functional and value function)

- Cost functional : $J(\alpha, x, t) = \int_t^T L(x(s), \alpha(s))ds + g(x(T))$
- Value function : V(x,t)= $\inf_{\alpha \in O} J(\alpha, x, t)$

To give the theorem of the Dynamic Programming Principle (DPP) we need to introduce the following hypothesis (H) :

1. g and L are $C^0(\mathbb{R}^n)$ and bounded
2. A is compact
3. $\exists C \in \mathbb{R}\ \forall x, z \quad | g(x) - g(z) | \leq C | x - z |$ (g lipschitz)
4. $\exists C \in \mathbb{R}\ \forall x, z, a \quad | L(x, a) - L(z, a) | \leq C | x - z |$ (L lipschitz uniformly in a)

Under the hypothesis (H) introduced previously we may now give the Dynamic Programming Principle for deterministic control

Theorem - DPP

$$\forall x \in \mathbb{R}^n, t \geq 0, h > 0, t + h < T$$

$$V(x,t) = \inf_{\alpha \in O} \int_t^{t+h} L(x(s), \alpha(s)) ds + v(x(t+h), t+h)$$

We can make a little drawing to be more explicit...

Proof : Omitted

Now, we have seen Dynamic Programming in Deterministic Control, we're interested in the principle in Stochastic Control.

Let :

- $(\Omega,\mathcal{F},\mathbb{F} = (F_t)_{t\geq 0},\mathbb{P})$ a filtered probability space , and W a d-dimensional Brownian motion.
- The control $\alpha = (\alpha_s)$ a progressively measurable (with respect to $\mathbb{F}$) process, valued in A, subset of Rm.

Definition (Controlled Diffusion Process)

We call control diffusion process a process X satifsying the following (SDE) in R^n

$$dX_s = f(X_s, \alpha_s)ds + \sigma(X_s, \alpha_s)dW_s$$

Stochastic Control - Framework

- The coefficients $f : \mathbb{R}^n \times A \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times A \to \mathbb{R}^{nd}$are measurable functions
- Let $L : \mathbb{R}^n \times A \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ measurable functions as seen previously for the Determinstic control
- For $(t, x) \in [0, T] \times \mathbb{R}^n$ we may define $\mathcal{A}(t, x)$ the set of admissible control α such that the controlled diffusion (SDE) starting from x at s=t , has a unique solution denoted by $X^{t,x}$ and such that $\mathbb{E}\left[\int_t^T L(X_s^{t,x}, \alpha_s)ds\right] < \infty$
- Under suitable conditions on b,σ,f and g we may now define the gain function :
  $J(t, x, \alpha) = \mathbb{E}\left[\int_t^T L(X_s^{t,x}, \alpha_s)\, ds + g(X_T^{t,x})\right]$ where $(t, x) \in [0, T] \times \mathbb{R}^n$ and $\alpha \in \mathcal{A}(t, x)$.

Stochastic Control - Framework

The presence of the expectation is crucial in the Stochastic case because on an optimal trajectory, the path taken will depend on $\omega \in \Omega$.
We can make a drawing to illustrate the problem.

In the stochastic case, the objective is to maximize (instead of minimizing) the gain function J over all control processes α, and we introduce the associated value function $V(t, x)$ as the supremum over α:

$$V(t,x) = \sup_{\alpha \in \mathcal{A}(t,x)} J(t,x,\alpha)$$

We may now give the DPP theorem in the Stochastic Case :

Theorem - DPP

$\forall (t,x) \in [0,T] \times \mathbb{R}^n, \forall h > 0, t+h < T,$

$$V(t,x) = \sup_{\alpha \in \mathcal{A}(t,x)} \mathbb{E}\left[\int_t^{t+h} L(X_s^{t,x}, \alpha_s)\, ds + v(t+h, X_{t+h}^{t,x})\right]$$

Proof : Omitted

Università degli Studi di Padova

The main purpose of the Dynamic Programming Principle is that we can now derive a Partial Differential Equation (PDE) over the value function. This equation is called the Hamilton Jacobi Bellman Equation (HJB) .

Idea of the derivation of HJB :

- Let $\alpha^* \in \mathcal{A}(t,x)$ an optimal control (ie a control maximizing the gain function), associated to a control diffusion process X^* starting at x, when s=t.(we can also call it an optimal trajectory).
  So
  $$V(t,x) = J(t,x,\alpha^*) = \sup_{\alpha \in \mathcal{A}(t,x)} J(t,x,\alpha)$$

Idea of the derivation of HJB:

- From DPP we also have,

$$V(t,x) = \mathbb{E}\left[\int_t^{t+h} L(X_s^*, \alpha_s^*)\, ds + V(t+h, X_{t+h}^*)\right]$$

- We assume V to be of class C^1 in t and C^2 in x, we may now apply Ito's lemma :
  $dV(t,X_t^*) = \partial_t V dt + \partial_x V \cdot dX_t^* + \frac{1}{2} Tr(\partial_{xx}^2 V \cdot \sigma\sigma^T(X^*, \alpha^*))dt$

  In the end, we get :

$$dV(t,X_t^*) = (\partial_t V + \partial_x V \cdot f + \frac{1}{2} Tr(\partial_{xx}^2 V \cdot \sigma\sigma^T))dt + (\partial_x V \cdot \sigma) dWt \ (1)$$

Idea of the derivation of HJB:

$$dV(t, X_t^*) = (\partial_t V + \partial_x V \cdot f + \tfrac{1}{2} Tr(\partial_{xx}^2 V \cdot \sigma\sigma^T))dt + (\partial_x V \cdot \sigma) dWt \ (1)$$

We now integrate (1) between t and t+h and then take the expectation and we get :

$$\mathbb{E}\left[V(t+h, X_{t+h}^*)\right] - V(t,x) = \mathbb{E}\left[\int_t^{t+h} (\partial_t V + \partial_x V \cdot f + \frac{1}{2} Tr(\partial_{xx}^2 V \sigma\sigma^T)) ds\right] \ (2)$$

The (LHS) *of* (2) $= -\mathbb{E}\left[\int_t^{t+h} L(X_s^*, \alpha_s^*)\, ds\right]$ from DPP

Idea of the derivation of HJB:

We divde (2) by $\frac{1}{h}$ and let we let it go to 0, by the mean value theorem we finally get :

$$-L(X_t^*, \alpha_t^*) = \partial_t V + [\partial_x V \cdot f + \frac{1}{2} Tr(\partial_{xx}^2 V \cdot \sigma\sigma^T)](X_t^*, \alpha_t^*)$$

Adding the intial condition, we have now that V satisfy the following PDE called the Hamilton Jacobi Bellman Equation for all (t,x) in $[0, T) \times \mathbb{R}^n$:

$$\begin{cases} -\partial_t V + \sup_{\alpha \in \mathcal{A}(t,x)} \{-\partial_x V \cdot f(x,\alpha) - \frac{1}{2} Tr(\partial_{xx}^2 V \cdot \sigma\sigma^T(x,\alpha)) - L(x,\alpha)\} = 0 \\ \text{V(T,x)=g(x)} \end{cases}$$

We often rewrite this HJB Equation in the form :

$$-\partial_t V(t,x) - H(t,x,D_x V(t,x), D_{xx} V(t,x)) = 0, \quad \forall (t,x) \in [0,T) \times \mathbb{R}^n$$

where for $(t,x,p,M) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{S}_n$,

$$H(t,x,p,M) = \sup_{\alpha \in \mathcal{A}} \left[f(x,\alpha) \cdot p + \frac{1}{2} \operatorname{tr}(\sigma \sigma^T (x,\alpha) M) + L(x,\alpha) \right].$$

This function H is called the Hamiltonian of the associated control problem.

Verification Theorem - 1st result - Comparaison Result

Let w be a function in $C^{1,2}([0,T) \times \mathbb{R}^n) \cap C^0([0,T] \times \mathbb{R}^n)$, and satisfying a quadratic growth condition, i.e., there exists a constant C such that

$$|w(t,x)| \leq C(1+|x|^2), \quad \forall (t,x) \in [0,T] \times \mathbb{R}^n.$$

(i) Suppose that

$$-\partial_t w(t,x) - H(t,x,D_x w(t,x), D_{xx} w(t,x)) \geq 0, \quad (t,x) \in [0,T) \times \mathbb{R}^n$$

$$w(T,x) \geq g(x), \quad x \in \mathbb{R}^n.$$

Then $w \geq V$ on $[0,T] \times \mathbb{R}^n$.

Verification Theorem - 2nd result

Given the same hypothesis on w as in the beginning of the 1st result

(ii) Suppose further that :

- $w(T, \cdot) = g$, and there exists a measurable function $\alpha^*(t,x)$, $(t,x) \in [0,T) \times \mathbb{R}^n$, valued in A such that :
  $-\partial_t w(t,x) - H(t,x,D_x w(t,x), D_{xx} w(t,x)) =$
  $-\partial_t w(t,x) - (f(x,\alpha^*) \cdot D_x w(t,x) +$
  $\frac{1}{2} \operatorname{tr}(\sigma\sigma^T(x,\alpha^*) D_{xx} w(t,x)) + L(x,\alpha^*)) = 0,$
- the SDE $dX_s = f(X_s, \alpha^*(s,X_s))ds + \sigma(X_s, \alpha^*(s,X_s))dW_s$ admits a unique solution, denoted by X_t^*, given an initial condition $X_t = x$, and the process $\alpha^*(s, X_s^*) \mid t \le s \le T\}$ lies in $\mathcal{A}(t,x)$

Then $w = V$ on $[0,T] \times \mathbb{R}^n$, and α^* is an optimal control

Proof : The proofs of those two results are omitted, they rely essentially on Ito's formula

We now have a strategy for solving stochastic control problems :

1. Try to solve the HJB Equation : $(t, x) \in [0, T) \times \mathbb{R}^n$:
   $-\partial_t w(t, x) - H(t, x, D_x w(t, x), D_{xx} w(t, x)) = 0$ (HJB)
   with terminal condition w(T,x)=g(x)
2. Fix $(t, x) \in [0, T) \times \mathbb{R}^n$ and find a control $\alpha^*(t, x)$ such that :
   $\alpha^*(t, x) \in \underset{\alpha \in A(t,x)}{\operatorname{argmax}} H(t, x, D_x w(t, x), D_{xx} w(t, x))$
3. Check if the approach is valid under the hypothesis of the Verification theorem, ie the solution w of HJB is smooth $(C^{1,2})$ and satisfy a quadratic growth
4. If the approach is valid under the hypothesis of the verification theorem, the solution of HJB w is the the value function V and α^* is an optimal control

Application in Finance

We are now interested in an application of those results in a concrete case in Finance namely : Merton Portofolio Allocation in Finite Horizon

- An agent invests at any time t a proportion α_t of his wealth in a stock of price S (governed by geometric Brownian motion) and $1-\alpha_t$ in a bond of price S^0 with interest rate r.
- Portfolio constraint : α_t valued in A closed convex subset of R
- The weath process X follows this SDE

$$dX_t = X_t \alpha_t \frac{dS_t}{S_t} + X_t(1-\alpha_t)\frac{dS_t^0}{S_t^0}$$
$$dX_t = X_t(\alpha_t \mu + (1-\alpha_t)r)dt + X_t \alpha_t \sigma dW_t$$

- So if we recall what we have seen in the beginning the wealth process X is a controlled diffusion process :

$$dX_s = f(X_s, \alpha_s)ds + \sigma(X_s, \alpha_s)dW_s$$

where the coefficients are
$f(x, a) := x(a\mu + (1-a)r)$ and $\sigma(x, a) := xa\sigma$

- Given a portfolio strategy $\alpha(t, x) \in \mathcal{A}(t, x)$ associated to $X^{t,x}$ the corresponding wealth process starting from an initial capital $X_t = x > 0$ at time t.
The agent wants to maximize the expected utility from terminal wealth at horizon T.
The value function of the utility maximization problem is then defined by

$$v(t, x) = \sup_{\alpha \in A} \mathbb{E}[U(X_T^{t,x})], \quad (t, x) \in [0, T] \times \mathbb{R}^+$$

- Note that here L=0 in the gain function, we only have terminal cost $g(x) = U(x)$
- The HJB equation for the stochastic control problem is

$$\begin{cases} -\frac{\partial w}{\partial t} - \sup\limits_{a \in A} \left[x(a\mu + (1-a)r)\frac{\partial w}{\partial x} + \frac{1}{2}x^2 a^2 \sigma^2 \frac{\partial^2 w}{\partial x^2} \right] = 0, \\ w(T, x) = U(x), \quad x \in \mathbb{R}^+. \end{cases}$$

- One can consider some power utility function of CRRA ie : $U(x) = \frac{x^p}{p}, \quad x \geq 0, \quad p < 1, \quad p \neq 0.$ which is concave and strictly increasing

- Now, we're looking for a solution of our HJB problem : a candidate solution would be in the form :

$$w(t,x) = \phi(t)U(x),$$

for some positive function ϕ.
We subsitute in the hJB equation, we derive that ϕ and we get the following ODE

$$\phi'(t) + \rho\phi(t) = 0, \quad \phi(T) = 1,$$

where

$$\rho = p \sup_{a \in A} \left[a(\mu - r) + r - \frac{1}{2}a^2(1-p)\sigma^2 \right].$$

Application in Finance

- Solving the ODE we obtain : $\phi(t) = \exp(\rho(T-t))$.
- Hence, our solution of the HJB equation is the function given by :

  $$w(t,x) = \exp(\rho(T-t))U(x), \quad (t,x) \in [0,T] \times \mathbb{R}^+,$$

  w is smooth (in t and x) and satisfy the quadratic growth condition of the verifcation theorem
- Furthermore, the function $a \in A \mapsto a(\mu - r) + r - \frac{1}{2}a^2(1-p)\sigma^2$ is strictly concave on the closed convex set A, and thus attains its maximum at some constant α^*. And by construction $\alpha^*(t,x) \in \underset{\alpha \in A(t,x)}{\operatorname{argmax}} H(t,x,D_x w(t,x), D_{xx} w(t,x))$

- In the end, from the *Verification Theorem*, we have obtained the value function to the utility maximisation problem :

$$v(t,x) = w(t,x) = \exp(\rho(T-t))U(x), \quad (t,x) \in [0,T] \times \mathbb{R}^+$$

  and we have that the optimal proportion of the wealth to invest is a constant α^*

- In the specific case of $A = \mathbb{R}$ (not realistic) :
  The values of α^* and ρ are explicitly given by

$$\alpha^* = \frac{\mu - r}{\sigma^2(1-p)}$$

$$\rho = \frac{(\mu - r)^2}{2\sigma^2}\frac{p}{1-p} + rp$$

- To summarize...
- Some remarks and extension

THANK YOU FOR YOUR ATTENTION