Stochastic Optimal Control

Lukas Wessels

www.lukaswessels.org

Research Horizon Seminar, Georgia Tech, November 15, 2023

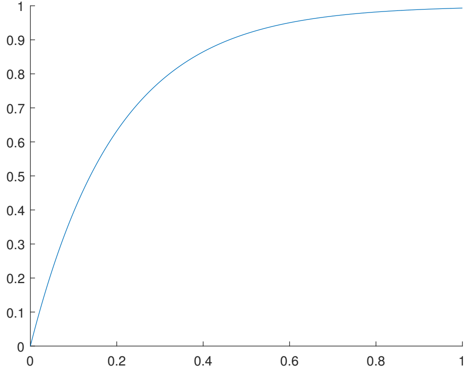
Georgia Institute of Technology

Introduction

Consider the ODE

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}s} X_s = \theta(\mu - X_s), \quad s \in [0,1] \\ X_0 = x \in \mathbb{R}. \end{cases}$$

For the parameters $\theta = 5$, $\mu = 1$, $x = 0$:

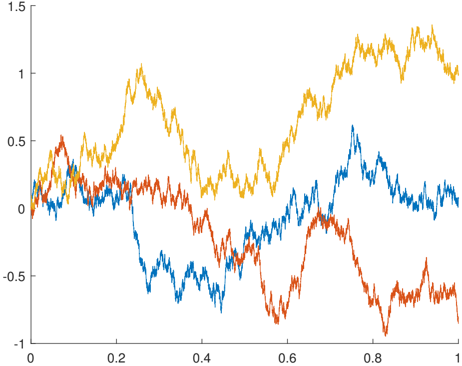


Many phenomena in real world can be modeled with randomness, e.g., stock prices, molecular systems (Langevin dynamics), random growth models, etc.

Brownian Motion

Stochastic process $B : [0, T] \times \Omega \to \mathbb{R}$ such that

- $B_0 = 0$.
- $t \mapsto B_t(\omega)$ almost surely continuous.
- $B_t - B_s$ is independent of B_s for all $0 \leq s \leq t \leq T$.
- $B_t - B_s \sim \mathcal{N}(0, t - s)$ for $0 \leq s \leq t \leq T$.



Remark: Sample paths are nowhere differentiable.

Stochastic Differential Equation

Consider the stochastic differential equation

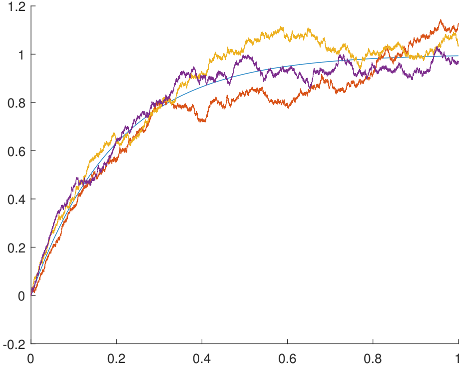
$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}s} X_s = \theta(\mu - X_s) + \sigma \text{“} \frac{\mathrm{d}}{\mathrm{d}s} W_s \text{”}, \quad s \in [0,1] \\ X_0 = x \in \mathbb{R}. \end{cases}$$

Stochastic Differential Equation

Consider stochastic integral equation

$$X_s = x + \int_0^s \theta(\mu - X_r)\mathrm{d}r + \sigma W_s, \quad s \in [0,1]$$

For the parameters $\theta = 5$, $\mu = 1$, $x = 0$, $\sigma = 0.2$:



Control of SDEs

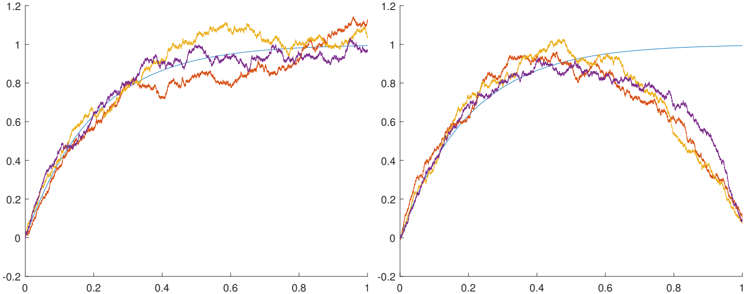
Minimize the cost functional

$$J(a) := \frac{1}{2}\mathbb{E}\left[\nu \int_0^1 a_s^2 \mathrm{d}s + X_1^2\right]$$

over $a : [0,1] \times \Omega \to \mathbb{R}$ subject to the state equation

$$\begin{cases} \mathrm{d}X_s = [\theta(\mu - X_s) + a_s]\,\mathrm{d}s + \sigma \mathrm{d}W_s, & s \in [0,1] \\ X_0 = x \in \mathbb{R}. \end{cases}$$

Parameters: $\nu = 1/100$, $\theta = 5$, $\mu = 1$, $\sigma = 0.2$, $x = 0$.



General Setting

Minimize

$$J(t,x;a) := \mathbb{E}\left[\int_t^T l(X_s,a_s)\mathrm{d}s + g(X_T)\right]$$

over admissible controls $a:(t,T)\times\Omega\to\mathbb{R}$ subject to

$$\begin{cases} \mathrm{d}X_s = b(X_s,a_s)\mathrm{d}s + \sigma(X_s,a_s)\mathrm{d}W_s, & s\in[t,T] \\ X_t = x\in\mathbb{R}, \end{cases}$$

where

- $l:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ are running costs,
- $g:\mathbb{R}\to\mathbb{R}$ are terminal costs,
- b and σ are drift and diffusion coefficient, respectively,
- $(W_s)_{s\in[t,T]}$ is a Brownian motion on $(\Omega,\mathcal{F},(\mathcal{F}_s)_{s\in[t,T]},\mathbb{P})$.

Two approaches:

- Dynamic programming (Richard Bellman 1950s)
- Pontryagin maximum principle (Lev Pontryagin et al. 1956)

Comparison with Reinforcement Learning – I

Minimize

$$J(t,x;a) := \mathbb{E}\left[\int_t^T l(X_s,a_s)\mathrm{d}s + g(X_T)\right]$$

over admissible controls $a : (t,T) \times \Omega \to \mathbb{R}$ subject to

$$\begin{cases} \mathrm{d}X_s = b(X_s,a_s)\mathrm{d}s + \sigma(X_s,a_s)\mathrm{d}W_s, & s \in [t,T] \\ X_t = x \in \mathbb{R}. \end{cases}$$

Markov decision process:

- S – state space
- A – action space
- $P_a(s,s') = \Pr(s_{t+1} = s' | s_t = s, a_t = a)$ – transition probabilities
- $R_a(s,s')$ – reward

Optimization objective:

- $\pi(s)$ – policy
- $\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_{a_t}(s_t, s_{t+1})\right]$ – reward to be maximized

Dynamic Programming

Minimize

$$J(t, x; a) := \mathbb{E}\left[\int_t^T l(X_s, a_s)\mathrm{d}s + g(X_T)\right]$$

over admissible controls $a : (t, T) \times \Omega \to \mathbb{R}$ subject to

$$\begin{cases} \mathrm{d}X_s = b(X_s, a_s)\mathrm{d}s + \sigma(X_s, a_s)\mathrm{d}W_s, & s \in [t, T] \\ X_t = x \in \mathbb{R}. \end{cases}$$

Define value function

$$V(t, x) := \inf_a J(t, x; a).$$

Satisfies dynamic programming principle

$$V(t, x) = \inf_a \mathbb{E}\left[\int_t^\tau l(X_s, a_s)\mathrm{d}s + V(\tau, X_\tau)\right], \qquad \forall \tau \in [t, T].$$

Next: Derive Hamilton–Jacobi–Bellman equation.

Itô's Formula

Theorem

Let $f \in C^{1,2}((0,T) \times \mathbb{R})$ *and let*

$$X_s = X_t + \int_t^s b(r, X_r)\mathrm{d}r + \int_t^s \sigma(r, X_r)\mathrm{d}W_r.$$

Then

$$\begin{aligned} f(s, X_s) = f(t, X_t) &+ \int_t^s f_t(r, X_r)\mathrm{d}r + \int_t^s f_x(r, X_r) b(r, X_r)\mathrm{d}r \\ &+ \int_t^s f_x(r, X_r)\sigma(r, X_r)\mathrm{d}W_r + \frac{1}{2}\int_t^s f_{xx}(r, X_r)\sigma^2(r, X_r)\mathrm{d}r. \end{aligned}$$

Derivation of the HJB Equation

Assume $V \in C^{1,2}([t, T] \times \mathbb{R})$. Then

$$\begin{aligned} V(\tau, X_\tau) &= V(t,x) + \int_t^\tau V_t(s, X_s)\mathrm{d}s + \int_t^\tau V_x(s, X_s)\mathrm{d}X_s + \frac{1}{2}\int_t^\tau V_{xx}(s, X_s)\mathrm{d}s \\ &= V(t,x) + \int_t^\tau V_t(s, X_s) + V_x(s, X_s)b(X_s, a_s) + \frac{1}{2}V_{xx}(s, X_s)\mathrm{d}s \\ &\quad + \int_t^\tau V_x(s, X_s)\sigma(X_s, a_s)\mathrm{d}W_s. \end{aligned}$$

Plugging this into the dynamic programming principle, we obtain

$$\begin{aligned} 0 &= \inf_a \mathbb{E}\left[\int_t^\tau l(X_s, a_s)\mathrm{d}s + V(\tau, X_\tau) - V(t,x)\right] \\ &= \inf_a \mathbb{E}\left[\int_t^\tau l(X_s, a_s) + V_t(s, X_s) + V_x(s, X_s)b(X_s, a_s) + \frac{1}{2}V_{xx}(s, X_s)\sigma^2(X_s, a_s)\mathrm{d}s\right]. \end{aligned}$$

Dividing by $\tau - t$ and sending $\tau \downarrow t$ yields ...

Hamilton–Jacobi–Bellman Equation

Dividing by $\tau - t$ and sending $\tau \downarrow t$ yields

$$\begin{cases} V_t(t,x) + \inf_a \left\{ l(x,a) + V_x(t,x) b(x,a) + \frac{1}{2} V_{xx}(t,x) \sigma^2(x,a) \right\} = 0 \\ V(T,x) = g(x). \end{cases}$$

Remarks:

- Fully nonlinear PDE for value function V.
- Does not posses a classical solution (in general).
- Crandall, Evans, Lions and others developed solution theory in 1980s.

Verification Theorem

Theorem

Assume $V \in C^{1,2}((t,T)\times\mathbb{R})$. *If*

$$a_s^* \in \arg\min \left\{ l(X_s^*, a) + V_x(s, X_s^*) b(X_s^*, a) + \frac{1}{2} V_{xx}(s, X_s^*) \sigma^2(X_s^*, a) \right\},$$

then a^* *is optimal.*

Proof: By Itô's formula, we have

$$\begin{aligned} &\mathbb{E}\left[V(T, X_T^*)\right] \\ &= V(t,x) + \mathbb{E}\left[\int_t^T V_t(s, X_s^*) + V_x(s, X_s^*) b(X_s^*, a_s^*) + \frac{1}{2} V_{xx}(s, X_s^*)\sigma^2(X_s^*, a_s^*) \mathrm{d}s\right]. \end{aligned}$$

Verification Theorem

Theorem

Assume $V \in C^{1,2}((t,T) \times \mathbb{R})$. If

$$a_s^* \in \arg\min \left\{ l(X_s^*, a) + V_x(s, X_s^*) b(X_s^*, a) + \frac{1}{2} V_{xx}(s, X_s^*) \sigma^2(X_s^*, a) \right\},$$

then a^* is optimal.

Proof: Hence,

$$V(t,x) = \mathbb{E}\left[g(X_T^*) + \int_t^T l(X_s^*, a_s^*) \mathrm{d}s \right]$$

$$-\mathbb{E}\left[\int_t^T V_t(s, X_s^*) + l(X_s^*, a_s^*) + V_x(s, X_s^*) b(X_s^*, a_s^*) + \frac{1}{2} V_{xx}(s, X_s^*) \sigma^2(X_s^*, a_s^*) \mathrm{d}s \right].$$

Verification Theorem

Theorem

Assume $V \in C^{1,2}((t,T) \times \mathbb{R})$. If

$$a_s^* \in \arg\min \left\{ l(X_s^*, a) + V_x(s, X_s^*) b(X_s^*, a) + \frac{1}{2} V_{xx}(s, X_s^*) \sigma^2(X_s^*, a) \right\},$$

then a^* is optimal.

Proof: Hence,

$$V(t,x) = \mathbb{E}\left[g(X_T^*) + \int_t^T l(X_s^*, a_s^*) \mathrm{d}s \right]$$

$$\cancel{- \mathbb{E}\left[\int_t^T \underbrace{V_t(s, X_s^*) + l(X_s^*, a_s^*) + V_x(s, X_s^*) b(X_s^*, a_s^*) + \frac{1}{2} V_{xx}(s, X_s^*) \sigma^2(X_s^*, a_s^*)}_{=0} \mathrm{d}s \right]}.$$

$\square$

Optimal Synthesis

Assume

- $l(x,a) = l_1(x) + l_2(a)$
- $b(x,a) = b_1(x) - a$
- $\sigma(x,a) = \sigma_1(x)$

In this case,

$$a_s^* \in \arg\min \left\{ l_1(X_s^*) + l_2(a) + V_x(s,X_s^*)b_1(X_s^*) - V_x(s,X_s^*)a \right\},$$

and the minimum is attained at

$$a_s^* = Dl_2^{-1}(V_x(s,X_s^*)).$$

$\implies$ optimal control is of feedback form.

Comparison with Reinforcement Learning – II

Markov decision process:

- S – state space
- A – action space
- $P_a(s, s') = \Pr(s_{t+1} = s' | s_t = s, a_t = a)$ – transition probabilities
- $R_a(s, s')$ – reward

Optimization objective:

- $\pi(s)$ – policy
- $\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_{a_t}(s_t, s_{t+1})\right]$ – reward to be maximized

Dynamic programming:

- $V(s) = \sum_{s'} P_{\pi(s)}(s, s')(R_{\pi(s)}(s, s') + \gamma V(s'))$ – value
- $\pi(s) = \operatorname{argmax}_a \left\{\sum_{s'} P_a(s, s')(R_a(s, s') + \gamma V(s'))\right\}$

Viscosity Solutions – Motivation

Let $\varphi \in C^\infty((0,T)\times\mathbb{R})$ be such that $V-\varphi$ has global maximum at (t,x), i.e.,

$$V(s,y)-V(t,x)\le\varphi(s,y)-\varphi(t,x)$$

for all $(s,y)\in[0,T]\times\mathbb{R}$. Then,

$$\begin{aligned}0&=\inf_a\mathbb{E}\left[\int_t^\tau l(X_s,a_s)\mathrm{d}s+V(\tau,X_\tau)-V(t,x)\right]\\&\le\inf_a\mathbb{E}\left[\int_t^\tau l(X_s,a_s)\mathrm{d}s+\varphi(\tau,X_\tau)-\varphi(t,x)\right].\end{aligned}$$

Thus,

$$\varphi_t(t,x)+\inf_a\left\{l(x,a)+\varphi_x(t,x)b(x,a)+\frac{1}{2}\varphi_{xx}(t,x)\sigma^2(x,a)\right\}\ge 0.$$

Viscosity Solutions – Definition

Consider the HJB equation

$$\begin{cases} V_t(t,x) + \inf_a \left\{ l(x,a) + V_x(t,x)b(x,a) + \frac{1}{2} V_{xx}(t,x)\sigma^2(x,a) \right\} = 0 \\ V(T,x) = g(x). \end{cases} \tag{1}$$

Definition (Viscosity Solution)

$V \in C([0,T] \times \mathbb{R})$ is a viscosity subsolution of (1), if $V(T,y) \leq g(y)$ and $\forall \varphi \in C^\infty((0,T) \times \mathbb{R})$ such that $V - \varphi$ has a global maximum at (t,x), it holds

$$\varphi_t(t,x) + \inf_a \left\{ l(x,a) + \varphi_x(t,x)b(x,a) + \frac{1}{2}\varphi_{xx}(t,x)\sigma^2(x,a) \right\} \geq 0$$

Viscosity Solutions – Consistency

Let $V \in C^{1,2}((0,T)\times\mathbb{R})$ be a classical solution of

$$\begin{cases} V_t(t,x) + \inf_a \left\{ l(x,a) + V_x(t,x)b(x,a) + \frac{1}{2}V_{xx}(t,x)\sigma^2(x,a) \right\} = 0 \\ V(T,x) = g(x) \end{cases}$$

and let $\varphi \in C^\infty((0,T)\times\mathbb{R})$ be such that $V-\varphi$ has a global maximum at (t,x). Then

$$V_t(t,x) = \varphi_t(t,x), \quad V_x(t,x) = \varphi_x(t,x), \quad V_{xx}(t,x) \le \varphi_{xx}(t,x).$$

Hence,

$$\begin{aligned} &\varphi_t(t,x) + \inf_a \left\{ l(x,a) + \varphi_x(t,x)b(x,a) + \frac{1}{2}\varphi_{xx}(t,x)\sigma^2(x,a) \right\} \\ &\ge V_t(t,x) + \inf_a \left\{ l(x,a) + V_x(t,x)b(x,a) + \frac{1}{2}V_{xx}(t,x)\sigma^2(x,a) \right\} = 0 \end{aligned}$$

$\implies$ Every classical solution is a viscosity solution.

Viscosity Solutions and Optimal Control

We have shown:

- Existence: The value function is a viscosity solution.
- Consistency: Every classical solution is a viscosity solution.

What about uniqueness?

- Uniqueness: Proof is more involved (uses analytical methods).

Open problems:

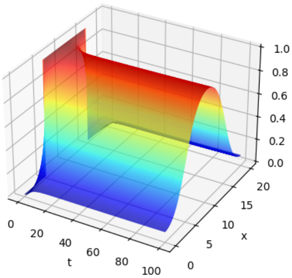
- Prove higher regularity of the viscosity solution (not possible in general).
- How to construct optimal feedback controls without differentiability of the value function?

Stochastic PDEs

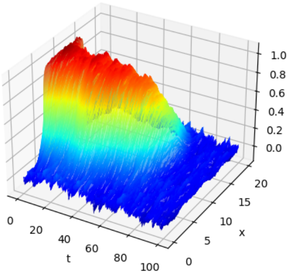
Consider the Nagumo equation

$$\begin{cases} \mathrm{d}X_s = \left[\Delta X_s + X_s(X_s - 1)\left(\frac{1}{2} - X_s\right)\right]\mathrm{d}s + \sigma \mathrm{d}W_s, \quad s \in [0, 100] \\ X_0 = x \in L^2(0, 20) \end{cases}$$

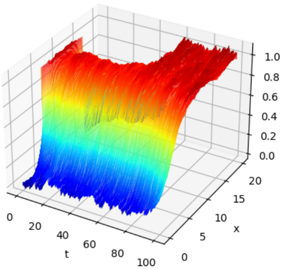
with Neumann boundary conditions and $u = 1_{[5,15]}$.



$\sigma = 0$



$\sigma = 0.05$



$\sigma = 0.05$

Control of Stochastic PDEs

Introduce control $a : [0, 100] \times [0, 20] \times \Omega \to \mathbb{R}$

$$\begin{cases} \mathrm{d}X_s = \left[\Delta X_s + X_s(X_s - 1)\left(\frac{1}{2} - X_s\right) + a_s\right]\mathrm{d}s + \sigma \mathrm{d}W_s, \quad s \in [0, 100] \\ X_0 = x \in L^2(0, 20) \end{cases} \tag{$\star$}$$

and cost functional

$$J(a) = \mathbb{E}\left[\int_0^{100}\int_\Lambda \left(X_s(\xi) - X^{\mathrm{ref}}(s,\xi)\right)^2 + a^2(s,\xi)\mathrm{d}\xi\mathrm{d}s + \int_\Lambda \left(X_{100}(\xi) - X^T(\xi)\right)^2 \mathrm{d}\xi\right]$$

where X^{ref}, X^T are desired reference profiles.

Goal:

$$\boxed{\text{Minimize} \quad J \quad \text{subject to } (\star)}$$

General Setting

Minimize

$$J(t,x;a) := \mathbb{E}\left[\int_t^T l(X_s,a_s)\mathrm{d}s + g(X_T)\right]$$

over admissible controls $a : [t,T] \times \Omega \to H$, subject to the abstract SDE

$$\begin{cases} \mathrm{d}X_s = [AX_s + b(X_s,a_s)]\mathrm{d}s + \sigma(X_s,a_s)\mathrm{d}W_s, & s \in [t,T] \\ X_t = x \in H, \end{cases}$$

where

- $l : H \times H \to \mathbb{R}$ are running costs,
- $g : H \to \mathbb{R}$ are terminal costs,
- $A : \mathcal{D}(A) \subset H \to H$ is a "nice" linear unbounded operator.
- b and σ are drift and diffusion coefficient, respectively.
- $(W_s)_{s\in[0,T]}$ is a cylindrical Wiener process on $(\Omega, \mathcal{F}, (\mathcal{F}_s)_{s\in[0,T]}, \mathbb{P})$.

Dynamic Programming

Value function

$$V(t,x) := \inf_a J(t,x;a)$$

satisfies dynamic programming principle

$$V(t,x) = \inf_a \mathbb{E}\left[\int_t^\tau l(X_s, a_s)\mathrm{d}s + V(\tau, X_\tau)\right], \qquad \forall \tau \in [t, T].$$

In this case, the HJB equation reads

$$\begin{cases} V_t(t,x) + \langle Ax, DV(t,x)\rangle_H \\ \quad + \inf_a \left\{ l(x,a) + \langle DV(t,x), b(x,a)\rangle_H + \frac{1}{2}\mathrm{Tr}(\sigma(x,a)\sigma^*(x,a)D^2V(t,x) \right\} = 0 \\ V(T,\cdot) = g, \end{cases}$$

where DV and D^2V are first and second order Fréchet differential of V with respect to x.

Remarks on the Infinite Dimensional Case

Difficulties:

- The term $\langle Ax, DV(t,x)\rangle$ is only defined for $x \in \mathcal{D}(A) \subset H$.
- Bounded, closed sets in H are not compact.
- ...

$\rightsquigarrow$ resolved by Crandall, Ishii, Lions, Święch, and others.

Open problems:

- Higher regularity of viscosity solution (not possible in general).
- Efficient numerical approximation of viscosity solution ("curse of dimensionality").