Stochastic Optimal Control and its Applications in Finance

Stochastic Optimal Control and Dynamic Programming

Fenghui Yu

TU Delft

First Meeting of the Dutch Sequential Decision-Making Community

Eindhoven, August 28th, 2025

Contents

A financial example

We consider a market with n assets:

$S_t^i =$ price of asset i, $h_t^i =$ units of asset i in portfolio, $w_t^i =$ portfolio weight on asset i.

Portfolio value and consumption:

$$X_t = \sum_{i=1}^{n} h_t^i S_t^i, \qquad c_t = \text{consumption rate}, \qquad \sum_{i=1}^{n} w_t^i = 1, \quad w_t^i = \frac{h_t^i S_t^i}{X_t}.$$

Self-financing dynamics (in relative weights):

$$\boxed{dX_t = X_t \sum_{i=1}^{n} w_t^i \frac{dS_t^i}{S_t^i} - c_t \, dt}$$

Simplest model

One risky asset and a money market account:

$$dS_t = \alpha S_t\,dt + \sigma S_t\,dW_t, \qquad dB_t = rB_t\,dt.$$

We maximize discounted utility of consumption:

$$\max_{\{w_t^0\},\{w_t^1\},\{c_t\}} \mathbb{E}\left[\int_0^T F(t, X_t, c_t)\,dt + \Phi(X_T)\right].$$

Wealth dynamics with portfolio weights w_t^0, w_t^1 $(w_t^0 + w_t^1 = 1)$:

$$dX_t = X_t\big(w_t^0\,r + w_t^1\,\alpha\big)dt - c_t\,dt + w_t^1\,\sigma X_t\,dW_t.$$

Problem formulation

We consider the stochastic control problem

$$\max_{\{u_t\}_{0\le t\le T}} \mathbb{E}\left[\underbrace{\int_0^T F\big(t, X_t, u_t\big)\,\mathrm{d}t}_{\text{running reward/penalty}} + \underbrace{\Phi\big(X_T\big)}_{\text{terminal reward}} \right]$$

subject to the dynamics (continuous-time controlled SDE)

$$\mathrm{d}X_t = \mu\big(t, X_t, u_t\big)\,\mathrm{d}t \,+\, \sigma\big(t, X_t, u_t\big)\,\mathrm{d}W_t, \qquad X_0 = x_0,$$

with admissible controls $u_t \in U(t, X_t)$ for all $t \in [0, T]$. We restrict attention to feedback control laws of the form

$$u_t = u(t, X_t).$$

Terminology: X = state variable, u = control variable, U = control constraint.

Note: No state space constraints.

How do we solve this optimization problem?

Main idea

- Embed the original problem in a family of problems indexed by (t, x) (start time and state).
- Tie the family together via a PDE: the Hamilton–Jacobi–Bellman (HJB) equation.
- Reduce the stochastic control problem to solving this deterministic PDE.

For notational simplicity in the next slides we first assume X, W and u are scalar.

Some notation

- For any (feedback) control law $u(\cdot,\cdot)$, write

$$\mu^u(t,x) := \mu\big(t,x,u(t,x)\big), \quad \sigma^u(t,x) := \sigma\big(t,x,u(t,x)\big), \quad F^u(t,x) := F\big(t,x,u(t,x)\big).$$

- For a control law $u(\cdot,\cdot)$ the second-order operator $\mathcal{L}^u$ acting on a smooth f is

$$(\mathcal{L}^u f)(t,x) = \mu^u(t,x)\,\partial_x f(t,x) + \tfrac{1}{2}\,\big(\sigma^u(t,x)\big)^2\,\partial_{xx} f(t,x).$$

- Under a control law $u(\cdot,\cdot)$, the controlled state X^u solves

$$\mathrm{d}X_t^u = \mu\big(t,X_t^u,u_t\big)\,\mathrm{d}t + \sigma\big(t,X_t^u,u_t\big)\,\mathrm{d}W_t, \qquad u_t = u(t,X_t^u).$$

Embedding the problem

For each (t, x), define problem $\mathbf{P}(t, x)$: maximize

$$\mathbb{E}_{t,x}\left[\int_t^T F\big(s, X_s^u, u_s\big)\,\mathrm{d}s + \Phi\big(X_T^u\big)\right],$$

subject to

$$\mathrm{d}X_s^u = \mu\big(s, X_s^u, u_s\big)\,\mathrm{d}s + \sigma\big(s, X_s^u, u_s\big)\,\mathrm{d}W_s, \qquad X_t = x,$$

with $u(s, y) \in U$ for all $(s, y) \in [t, T] \times \mathbb{R}^n$.

Note: The original problem is $\mathbf{P}(0, x_0)$.

The optimal value function

Define the (controlled) performance for a law u by

$$J(t,x;u) := \mathbb{E}_{t,x}\left[\int_t^T F\big(s, X_s^u, u_s\big)\,\mathrm{d}s + \Phi\big(X_T^u\big)\right].$$

The optimal value function is

$$V(t,x) := \sup_{u\in\mathcal{U}} J(t,x;u), \qquad (t,x)\in[0,T]\times\mathbb{R}^n.$$

We seek a PDE for V.

Assumptions

We assume (for the derivation):

- There exists an optimal feedback control $\hat{u}$.
- The optimal value V is sufficiently regular: $V \in C^{1,2}$.
- Interchange/limit steps used below are justified.

The Bellman optimality principle

Dynamic programming relies heavily on the following basic result.

Proposition

If $\hat{u}$ is optimal on $[t, T]$, then it is optimal on every subinterval $[s, T]$ with $t \leq s \leq T$.

Proof idea: Law of iterated expectations.

Basic strategy to derive the PDE

For simplicity of notations, we demonstrate with $x \in \mathbb{R}$.

- Fix (t, x) and a small $h > 0$.
- Pick an arbitrary control law u.
- Define a new control u^* by

$$u^*(s, y) = \begin{cases} u(s, y), & (s, y) \in [t, t+h] \times \mathbb{R}, \\ \hat{u}(s, y), & (s, y) \in (t+h, T] \times \mathbb{R}. \end{cases}$$

That is, use u on $[t, t+h]$ and then switch to the (unknown) optimal law $\hat{u}$ for the remainder.

Basic idea

Consider two strategies on $[t, T]$ starting from (t, x):

I: Use the optimal law $\hat{u}$ throughout. Then $J(t, x; \hat{u}) = V(t, x)$.

II: Use u^* defined above. The total value is

$$J(t, x; u^*) = \mathbb{E}_{t,x}\left[\int_t^{t+h} F\big(s, X_s^u, u_s\big)\,\mathrm{d}s \;+\; V\big(t+h, X_{t+h}^u\big)\right].$$

By optimality, Strategy I is at least as good as Strategy II.

Dynamic programming principle

Optimality gives

$$V(t,x) \;\geq\; \mathbb{E}_{t,x}\left[\int_t^{t+h} F\big(s, X_s^u, u_s\big)\,\mathrm{d}s + V\big(t+h, X_{t+h}^u\big)\right],$$

for all u with equality if and only if $u = \hat{u}(t,x)$.
We also get the reverse inequality since

$$J(t,x;u^*) \leq \sup_{u\in\mathcal{U}} \mathbb{E}_{t,x}\left[\int_t^{t+h} F\big(s, X_s^u, u_s\big)\,\mathrm{d}s \;+\; V\big(t+h, X_{t+h}^u\big)\right].$$

and hence the Dynamic Programming Principle (DPP):

$$\boxed{V(t,x) \;=\; \sup_{u\in\mathcal{U}} \mathbb{E}_{t,x}\left[\int_t^{t+h} F\big(s, X_s^u, u_s\big)\,\mathrm{d}s + V\big(t+h, X_{t+h}^u\big)\right]}$$

Comparing strategies

By Itô's formula applied to $V(s, X_s^u)$ on $[t, t+h]$,

$$V\big(t+h, X_{t+h}^u\big) = V(t,x) + \int_t^{t+h} \Big(\partial_t V + \mathcal{L}^u V\Big)(s, X_s^u)\,\mathrm{d}s \\ + \int_t^{t+h} \partial_x V(s, X_s^u)\,\sigma^u(s, X_s^u)\,\mathrm{d}W_s.$$

Taking expectations and rearranging yields

$$\mathbb{E}_{t,x}\left[\int_t^{t+h} \Big(F^u + \partial_t V + \mathcal{L}^u V\Big)(s, X_s^u)\,\mathrm{d}s\right] \;\leq\; 0.$$

Remark: We have equality above if and only if $u = \hat{u}$.

Letting $h \to 0$

Divide by h, move h inside the expectation, and let $h \downarrow 0$ to obtain the pointwise inequality

$$F(t,x,u) + \partial_t V(t,x) + (\mathcal{L}^u V)(t,x) \ \leq \ 0, \qquad \text{for all } u,$$

with equality if and only if $u = \hat{u}(t,x)$. Thus,

$$\partial_t V(t,x) + \sup_{u \in U} \left\{ F(t,x,u) + (\mathcal{L}^u V)(t,x) \right\} = 0.$$

The HJB equation

Thoerem

Under suitable regularity assumptions:

- V solves the Hamilton–Jacobi–Bellman PDE

$$\partial_t V(t,x) + \sup_{u \in U} \left\{ F(t,x,u) + (\mathcal{L}^u V)(t,x) \right\} = 0, \qquad V(T,x) = \Phi(x).$$

- For each (t,x), the supremum is attained at $u = \hat{u}(t,x)$.

Multi-dimensional generator and dynamics

For $u \in \mathbb{R}^k$ define

$$\boldsymbol{\mu}_u(t,\boldsymbol{x}) := \boldsymbol{\mu}(t,\boldsymbol{x},u), \quad \boldsymbol{\sigma}_u(t,\boldsymbol{x}) := \boldsymbol{\sigma}(t,\boldsymbol{x},u), \quad C_u(t,\boldsymbol{x}) := \boldsymbol{\sigma}_u(t,\boldsymbol{x})\,\boldsymbol{\sigma}_u(t,\boldsymbol{x})^\top.$$

For smooth f and fixed u, the generator is

$$(\mathcal{L}^u f)(t,\boldsymbol{x}) = \sum_{i=1}^n \mu_u^i(t,\boldsymbol{x})\,\partial_{x_i} f + \tfrac{1}{2}\sum_{i,j=1}^n C_u^{ij}(t,\boldsymbol{x})\,\partial_{x_i x_j} f.$$

Under a control law u the state satisfies

$$\mathrm{d}\boldsymbol{X}_t^u = \boldsymbol{\mu}\big(t,\boldsymbol{X}_t^u,u_t\big)\,\mathrm{d}t + \boldsymbol{\sigma}\big(t,\boldsymbol{X}_t^u,u_t\big)\,\mathrm{d}\boldsymbol{W}_t, \qquad u_t = u(t,\boldsymbol{X}_t^u).$$

Logic and problem

We derived HJB as a *necessary* condition assuming V is the optimal value and sufficiently smooth.

Question: If we solve the HJB PDE, have we found the optimal value and an optimal control?
Answer: Yes — this is guaranteed by the Verification Theorem.

The verification theorem

Suppose $H(t,x)$ and $g(t,x)$ satisfy

- H is sufficiently integrable and solves

$$\partial_t H + \sup_{u \in U}\{F(t,x,u) + (\mathcal{L}^u H)(t,x)\} = 0, \qquad H(T,x) = \Phi(x).$$

- For each (t,x) the supremum is attained at $u = g(t,x)$.

Then

1. $V(t,x) = H(t,x)$ is the optimal value function, and
2. there exists an optimal control $\hat{u}$ given by $\hat{u}(t,x) = g(t,x)$.

Handling the HJB equation

1. Start from the HJB for V.
2. For fixed (t, x) solve the static maximization

$$\max_{u \in U} \left\{ F(t, x, u) + (\mathcal{L}^u V)(t, x) \right\},$$

   treating t, x and the (unknown) V and its derivatives as parameters.
3. Denote the maximizer $\hat{u} = \hat{u}(t, x; V)$. This is the *candidate* optimal law.
4. Substitute $\hat{u}(t, x; V)$ back into HJB to obtain a PDE for V only:

$$\partial_t V + F^{\hat{u}}(t, x) + (\mathcal{L}^{\hat{u}} V)(t, x) = 0, \quad V(T, x) = \Phi(x).$$

5. Solve this PDE. Then set the feedback law to $\hat{u}(t, x; V)$.

Making an Ansatz

- The HJB is generally nonlinear and hard; closed forms are rare.
- In applications one often *guesses* a parametric form (Ansatz) for V and identifies the parameters from the PDE.
- Heuristic: V often inherits structure from Φ and the running criterion F.
- Many classical solved problems are crafted to be analytically tractable.

Recall the simplest model

One risky asset and a money market account:

$$dS_t = \alpha S_t \, dt + \sigma S_t \, dW_t, \qquad dB_t = r B_t \, dt.$$

We maximize discounted utility of consumption:

$$\max_{\{w_t^0\},\{w_t^1\},\{c_t\}} \mathbb{E}\left[\int_0^T F(t, X_t, c_t)\, dt + \Phi(X_T)\right].$$

Wealth dynamics with portfolio weights w_t^0, w_t^1 $(w_t^0 + w_t^1 = 1)$:

$$dX_t = X_t \big(w_t^0 \, r + w_t^1 \, \alpha\big) dt - c_t \, dt + w_t^1 \, \sigma X_t \, dW_t.$$

Issue: with no constraint on X_t one can push wealth negative and obtain unbounded utility by consuming arbitrarily large amounts.

What are the problems?

- Unbounded objective: consume "arbitrarily large" amounts.
- Wealth X_t can become negative; no prohibition in the naïve setup.
- Natural constraint $X_t \geq 0$ is a *state constraint* and classical dynamic programming does not allow it directly.

Good news: Dynamic Programming can be generalized to handle such problems.

Generalized problem (with exit at the boundary)

Let D be a nice open subset of $[0,T]\times\mathbb{R}^n$ and consider

$$\max_{u\in\mathcal{U}}\ \mathbb{E}\left[\int_0^{\tau} F(s,X_s^u,u_s)\,ds\ +\ \Phi(\tau,X_\tau^u)\right],$$

with controlled dynamics

$$dX_t=\mu(t,X_t,u_t)\,dt+\sigma(t,X_t,u_t)\,dW_t,\qquad X_0=x_0,$$

and stopping time (exit or terminal time)

$$\tau=\inf\{\,t\ge 0:(t,X_t)\in\partial D\,\}\wedge T.$$

Generalized HJB

Under suitable regularity, the value function V solves

$$\partial_t V(t,x) + \sup_{u\in U}\Big\{F(t,x,u) + \mathcal{L}^u V(t,x)\Big\} = 0, \quad (t,x)\in D,$$

with boundary condition $V(t,x) = \phi(t,x)$ for $(t,x)\in\partial D$, where

$$\mathcal{L}^u V := \mu(t,x,u)\,\partial_x V + \frac{1}{2}\sigma^2(t,x,u)\,\partial_{xx} V.$$

A standard verification theorem applies.

Applications in trading problems

Reformulated consumption–investment problem

Exit when wealth hits zero:

$$\max_{c_t \geq 0,\ w_t \in \mathbb{R}} \mathbb{E}\left[\int_0^\tau F(t, c_t)\, dt + \Phi(X_\tau)\right], \qquad \tau = \inf\{t \geq 0 : X_t = 0\} \wedge T,$$

with notation $w_t^1 = w_t$, $w_t^0 = 1 - w_t$ and dynamics

$$\boxed{dX_t = w_t(\alpha - r)X_t\, dt + (rX_t - c_t)\, dt + w_t\, \sigma X_t\, dW_t.}$$

HJB equation

Take $F(t,c) = e^{-\beta t}\,\dfrac{c^{1-\gamma}}{1-\gamma}$ (CRRA utility, $\gamma \neq 1$). The HJB reads

$$\partial_t V + \sup_{c\geq 0,\, w\in\mathbb{R}} \left\{ e^{-\beta t}\frac{c^{1-\gamma}}{1-\gamma} + wx(\alpha - r)V_x + (rx - c)V_x + \frac{1}{2}x^2 w^2 \sigma^2 V_{xx} \right\} = 0,$$

with $V(T,x) = 0$ and $V(t,0) = 0$.

Solving the embedded static problem

First order conditions give (where $V_x = \partial_x V$, $V_{xx} = \partial_{xx} V$)

$$\boxed{c^*(t,x) = \left(\frac{e^{-\beta t}}{V_x(t,x)}\right)^{1/\gamma} = h(t)^{-1/\gamma}\, x,} \qquad \boxed{w^*(t,x) = -\frac{V_x}{xV_{xx}} \cdot \frac{\alpha - r}{\sigma^2} = \frac{\alpha - r}{\gamma\, \sigma^2}.}$$

Motivated by homotheticity, use the ansatz

$$\boxed{V(t,x) = e^{-\beta t}\, \frac{h(t)\, x^{1-\gamma}}{1-\gamma},} \qquad h(T) = 0.$$

ODE for the scaling function $h(t)$

Plugging the ansatz and c^*, w^* into HJB yields the Bernoulli-type ODE

$$\dot{h}(t) = \left[\beta - (1-\gamma)\left(r + \frac{(\alpha - r)^2}{2\gamma\,\sigma^2}\right)\right] h(t) - (1-\gamma)\,h(t)^{\,1-1/\gamma}, \qquad h(T) = 0$$

Thus

$$c_t^* = h(t)^{-1/\gamma} X_t, \qquad w_t^* = \frac{\alpha - r}{\gamma\,\sigma^2} \quad \text{(Merton proportion)}.$$

The ODE can be solved in closed form (Bernoulli equation).

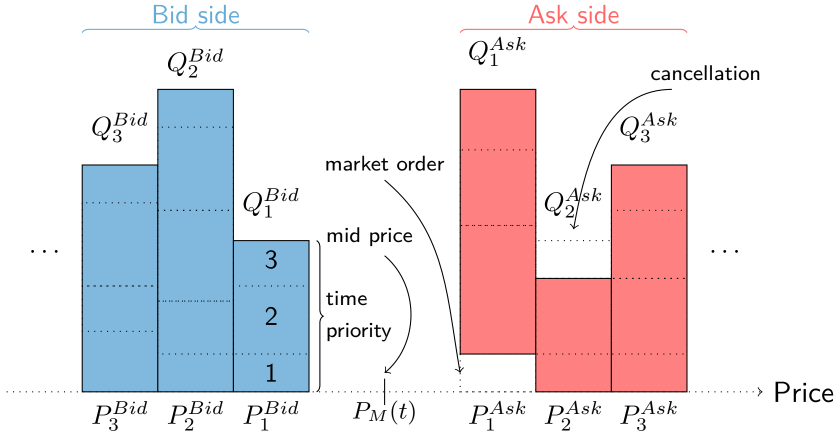
Observations

- State constraints (e.g. $X_t \geq 0$) can be handled via a generalized HJB with exit times.
- With CRRA utility and Black–Scholes returns:

$$w_t^* = \frac{\alpha - r}{\gamma\,\sigma^2} \quad \text{(constant in } t \text{ and } x\text{)}.$$

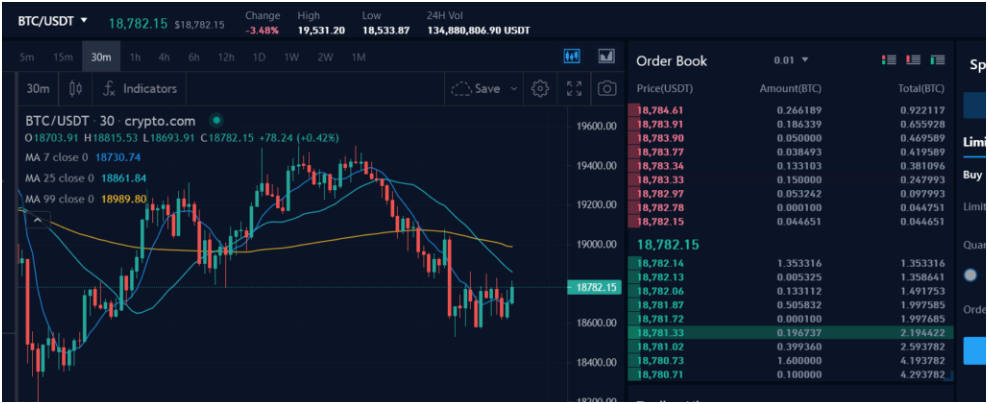
- Optimal consumption is proportional to wealth: $c_t^* = m(t)\,X_t$ with $m(t) = h(t)^{-1/\gamma}$ and h from a Bernoulli ODE.

Limit order book



[1]

[1] Picture credit: C. Lehalle, O. Mounjid, and M. Rosenbaum. Optimal liquidity-based trading tactics. *Stochastic Systems*. 11(4), 2018.


BTC/USDT
18,782.15 $18,782.15
Change -3.48%
High 19,531.20
Low 18,533.87
24H Vol 134,880,806.90 USDT
5m 15m 30m 1h 4h 6h 12h 1D 1W 2W 1M
30m
Indicators
Save
BTC/USDT · 30 · crypto.com
O18703.91 H18815.53 L18693.91 C18782.15 +78.24 (+0.42%)
MA 7 close 0 18730.74
MA 25 close 0 18861.84
MA 99 close 0 18989.80
19600.00
19400.00
19200.00
19000.00
18782.15
18600.00
18400.00
Order Book
0.01
Price(USDT)
Amount(BTC)
Total(BTC)
18,784.61 0.266189 0.922117
18,783.91 0.186339 0.655928
18,783.90 0.050000 0.469589
18,783.77 0.038493 0.419589
18,783.34 0.133103 0.381096
18,783.33 0.150000 0.247993
18,782.97 0.053242 0.097993
18,782.78 0.000100 0.044751
18,782.15 0.044651 0.044651
18,782.15
18,782.14 1.353316 1.353316
18,782.13 0.005325 1.358641
18,782.06 0.133112 1.491753
18,781.87 0.505832 1.997585
18,781.72 0.000100 1.997685
18,781.33 0.196737 2.194422
18,781.02 0.399360 2.593782
18,780.73 1.600000 4.193782
18,780.71 0.100000 4.293782
Buy

Market Making

- Provide liquidity by posting bid/ask in the LOB and earn the spread.
- Goal: profit from spread while controlling inventory risk.
- Classical approach: stochastic control $\Rightarrow$ HJB for optimal quotes.

A Canonical MM Model

- Mid-price: $dS_t = \sigma\, dW_t$.
- Quotes: post S_t^b, S_t^a; define spreads $\delta_t^b = S_t - S_t^b$, $\delta_t^a = S_t^a - S_t$.
- Order arrivals (independent of W):

$$\lambda^b(\delta) = \lambda^a(\delta) = Ae^{-k\delta}.$$

- Inventory: $q_t = N_t^b - N_t^a$.
- Cash:

$$dX_t = (S_t - \delta_t^a)\, dN_t^a - (S_t - \delta_t^b)\, dN_t^b.$$

- CARA utility at T:

$$V(s,x,q,t) = \sup_{\{\delta_u^a, \delta_u^b\}} \mathbb{E}\Big[-e^{-\gamma(X_T + q_T S_T)} \mid X_t{=}x, S_t{=}s, q_t{=}q\Big].$$

Analytical limits & RL opportunity

- HJB admits (semi) closed-form solutions only under strong assumptions (e.g. CARA/CRRA/quadratic utility, specific dynamics).
- Real markets $\Rightarrow$ specification risk.
- Reinforcement learning for MM: Q-learning, SARSA, deep policy gradients; states: quotes/LOB features, inventory, volatility, order-flow; actions: spreads/quotes; rewards: P&L with inventory penalties, etc.
- Multi-agent RL to model competition and interaction effects.

T. Björk. (2020) Arbitrage Theory in Continuous Time, 4th edition. *Oxford University Press*.

M. Avellaneda and S. Stoikov (2008). High-frequency trading in a limit order book. *Quantitative Finance*, 8(3),217–224.