Stochastic Control: Theories and Applications in Finance

Fengyi Yuan

University of Michigan

fengyiy@umich.edu

BIMSA
May 22nd, 2025

Table of Contents

Discrete time version: Markov decision process

- State transitions controlled by action a: $X_{t+1} \sim \mathbb{P}(\cdot | t, X_t = x, a)$.
- Collect rewards based on states and actions: $R_t = r(t, X_t, a)$.
- Aim to find a good **policy**: $\phi : \mathcal{X} \to \mathcal{A}$, i.e., at time t, observe state x, then take action $a = \phi(x)$.
- (Discounted) total reward (value function with policy ϕ):

$$J^{\phi}(x) = \sum_{t=0}^{\infty} \delta^t \mathbb{E}^{\phi,x} r(t, X_t, \phi(X_t)).$$

- Problem:

$$\max_{\phi} J^{\phi}(x)$$

Discrete time version: Markov decision process

The tool to find the optimal ϕ:

Dynamic programming principle

With $Q^\pi(x,a) := J^{a\oplus_1\pi}(x)$, we have

$$Q^\phi(x,a) = r(x,a) + \delta\mathbb{E}^{x,a}Q^\phi(X_1,\phi(X_1)).$$

With $V(x) = \sup_\phi Q^\phi(x,\phi(x)) = \sup_\phi J^\phi(x)$, we have

$$V(x) = \sup_a\{r(x,a) + \delta\mathbb{E}^{x,a}V(X_1)\}.$$

Solve V, then determine ϕ^*:

$$\phi^*(x) = \underset{a}{\operatorname{argmax}}\{r(x,a) + \delta\mathbb{E}^{x,a}V(X_1)\}$$

Continuous time version: Stochastic control theory

We also have similar approaches (DPP) for continuous-time problems... But with some specifications and more advanced tools

- State dynamics:

$$dX_t = b(t, X_t, \alpha_t)dt + \sigma(t, X_t, \alpha_t)dW_t.$$

- Objectives (costs or rewards):

$$J(t, x; \alpha) = \mathbb{E}^{t,x,\alpha}\left[\int_0^T e^{-r(s-t)} f(s, X_s, \alpha_s)ds + e^{-r(T-t)}g(X_T)\right].$$

- Problems:

$$\alpha_*^{(t,x)} = \underset{\alpha}{\mathrm{argmax}}\{J(t, x; \alpha)\}$$

Idea: As $\mathbb{P}^{t,x}$ is described by SDE, use tools from stochastic analysis to establish characterizations of critical points!

Stochastic control: Analytical approach

Consider the following **value function:**

$$V(t,x) := \sup_{\alpha \in \mathcal{U}_{t,x}} J(t,x;\alpha) = J(t,x;\alpha_*^{(t,x)})$$

Dynamic Programming principle (continuous time)

For any $t' \geq t$:

$$V(t,x) = \sup_{\alpha \in \mathcal{U}_{t,x}} \mathbb{E}^{t,x,\alpha}\left[\int_t^{t'} e^{-r(s-t)} f(s, X_s, \alpha_s) ds + V(t', X_{t'})\right]$$

Interpretation:
The optimal control $\hat{\alpha}$ for the subproblem on $[t, T]$ remains to be optimal for the subproblem on $[t', T]$.

Stochastic control: Analytical approach

Hamilton-Jacobi-Bellman (HJB) equation

The value function V satisfies

$$\partial_t V + rV + \sup_{\alpha} \left\{ b(\cdot,\alpha)\partial_x V + \frac{1}{2}\sigma(\cdot,\alpha)^2 \partial_{xx} V + f(\cdot,\alpha) \right\} = 0.$$

- Assuming sufficient regularity, the proof is straightforward (classical solution).
- Under mild conditions, value function is always a viscosity solution $\longrightarrow$ existence.
- Uniqueness is challenging and usually relies on comparison principle.
- Extensions: path-dependency (non-Markovian), mean-field equation $\partial_\mu V(t,x,\mu)$...

Stochastic control: Probabilistic approach

Stochastic maximum principle

Suppose f and g are **concave** in x and α. Suppose we have optimal control and trajectory $(\hat{\alpha}, \hat{x})$. Consider the Hamiltonian $H(t,x,\alpha,p,q) = pb(t,x,\alpha) + q\sigma(t,x,\alpha) + f(t,x,\alpha)$. Consider the following **adjoint equation**:

$$\begin{cases} dp(t) = & -\Big(b_x(t,\hat{x}(t),\hat{\alpha}(t))p(t) + \sigma_x(t,\hat{x}(t),\hat{\alpha}(t))q(t) \\ & +f_x(t,\hat{x}(t),\hat{\alpha}(t))\Big)dt + q(t)dW(t), \\ p(T) = & g_x(\hat{x}(T)). \end{cases}$$

Then,

$$\hat{\alpha}(t) = \operatorname*{argmax}_{\alpha} H(t,\hat{x}(t),\alpha,p(t),q(t)).$$

Stochastic control: Probabilistic approach

- The equation of p (the **first-order** adjoint equation) is a Backward Stochastic Differential Equation (BSDE). The solution is a **pair** (p, q)!
- Fixing $\hat{x}$ and $\hat{\alpha}$, p satisfies a linear BSDE. However, $\hat{\alpha}$ depends on p, q, hence we essentially have a nonlinear BSDE (even quadratic growth in portfolio selection case).
- In some cases: $\hat{x}$ and $\hat{\alpha}$ explicitly expressed by p and q. Then, showing the well-posedness of BSDE (sometimes **F**-BSDE) gives the characterization of optimal control.
- Techniques: proper choices of solution space + fixed point theorems.

Portfolio Selection

Returns of assets: R a random variable valued in $\mathbb{R}^d$; a **model** $\mathbb{P}^*$.
Choose $\pi \in \mathbb{R}^d$ to get the best outcome $\pi \cdot R$.

Extensions:

- Discrete time dynamic problems: $R_1, R_2, \cdots, R_T$ take values in $\mathbb{R}^d$, and $\mathbb{P}^*$ is a probability measure on $(\mathbb{R}^d)^T$.
- Continuous time dynamic problems: The asset prices $\{S_t\}_{0 \le t \le T}$ follow different types of SDE under $\mathbb{P}^*$.

The first question to ask: the criterion of optimization with uncertainty $\longrightarrow$ preferences on spaces of r.v. (or distributions)!

Preferences

Suppose X is the outcome r.v..

Key idea: map X to some **numbers**.

- Expectations:
  - $\mathbb{E}[X]$: risk neutral.
  - $\mathbb{E}[U(X)]$: risk averse.
- + Expectation - risk:
  - $\mathbb{E}[X] - \frac{\gamma}{2}\mathrm{Var}[X]$: mean-variance (related to LQ control).

Axiomatization of EU

Subject to some technical conditions, a preference over **distributions** satisfies the following independence axiom iff it is expected utility:

$$\mu_1 \succ \mu_2 \Longrightarrow p\mu_1 + (1-p)\mu_3 \succ p\mu_2 + (1-p)\mu_3$$

Do people make decisions by EU?

An old but famous experiment in behavioral sciences:

Experiment 1				Experiment 2			
Gamble 1A		Gamble 1B		Gamble 2A		Gamble 2B	
Winnings	**Chance**	**Winnings**	**Chance**	**Winnings**	**Chance**	**Winnings**	**Chance**
$1 million	100%	$1 million	89%	Nothing	89%	Nothing	90%
		Nothing	1%	$1 million	11%		
		$5 million	10%			$5 million	10%

Figure: Allais paradox (taken from Wikipedia).

- Experiments show 1A$\succ$1B and 2B$\succ$2A exist simultaneously.
- 1B = 0.89*1A + 0.11*C, C = nothing w.p. 1/11 and 5m w.p. 10/11;1A$\succ$1B $\Longrightarrow$ 1A$\succ$C;
- 2A = 0.89*(winning nothing) + 0.11*1A, 2B = 0.89*(winning nothing) + 0.11*C; 1A$\succ$C $\Longrightarrow$ 2A$\succ$2B.

Nonlinear preferences

Whatever the utility function is, the independence axiom issues remain because $\mathbb{E}[U(X)] = \langle U, \mathcal{L}(X) \rangle$ is a linear function.

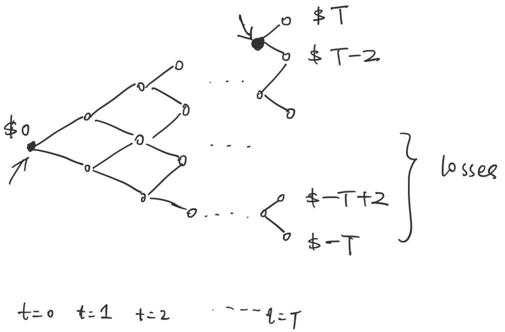
In [Liang, Xia, and Yuan, 2023] we study the stochastic maximum principle for continuous time dynamic portfolio selection problems, with nonlinear preferences.

Examples include:

- Betweenness preference, defined implicitly: $\mathbb{E}[F(X - g(\mathcal{L}(X))] = 0$, $\mathbb{E}[F(X/g(\mathcal{L}(X))] = 0$.
- Weighted utility: $\mathbb{E}[U(X)\lambda(X)]/\mathbb{E}[\lambda(X)]$.
- Mean-variance, or any nonlinear functions outside of moments, $F(\mathbb{E}[X], \mathbb{E}[X^2], \cdots)$.

Optimization?

- Always evaluate the preference value at the **conditional distributions**.
- DPP $\longleftrightarrow$ iterated law of conditional expectations $\longleftrightarrow$ linearity.



- Without DPP, what solutions shall we expect?

Strotz-Pollack equilibirum

General idea: keep doing backward induction anyway; a choice that one has no reason to abandon.

Definition

Suppose under a given strategy or policy π, the pay-off at time t is given by $J(t;\pi)$ (need to condition on $\mathcal{F}_t$). Then π^* is said to an equilibrium if:

- (Discrete time version) $J(t;\pi^* \oplus_1 a) \leq J(t;\pi^*)$ for any t and action a.
- (Continuous time version) $\limsup_{\varepsilon \to 0} \frac{J(t;\pi^* \oplus_\varepsilon a) - J(t;\pi^*)}{\varepsilon} \leq 0$.

See the paper [Liang, Xia, and Yuan, 2023] for detailed definitions of different type of equilibria.

Problem formulation

- We have a model for stock price:

$$\begin{cases} \mathrm{d}S_t^i = S_t^i[\theta^i(t)\mathrm{d}t + \sigma^i(t) \cdot \mathrm{d}W_t^{\mathcal{S}}], \\ S_0^i = s_0^i > 0, \end{cases}$$

  Here θ and σ can be general **stochastic processes** subject to technical conditions. We do need they are adapted to Brownian filtrations, but with possibly more Brownian motions than $W^{\mathcal{S}}$ (incomplete market).

- Wealth dynamic under a (proportional) portfolio π:

$$\begin{cases} \mathrm{d}X_t^{\pi} = X_t^{\pi}\pi_t\theta(t)\mathrm{d}t + X_t^{\pi}\pi_t\sigma(t) \cdot \mathrm{d}W_t^{\mathcal{S}}, \\ X_0^{\pi} = x_0. \end{cases}$$

Problem formulation

- The preference is characterized by a function $g : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$.
- For each t the agent “optimizes” the value of **conditional distribution**: $J(t; \pi) = g\big(\mathcal{L}(X_T^\pi | \mathcal{F}_t)\big)$.
- The first order condition relies on the notion of derivatives with respect to **probability measures**: $\nabla g : \mathcal{P}(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ is such that

$$\frac{d}{ds} g(s\nu + (1-s)\mu) = \langle \nabla g(s\nu + (1-s)\mu, \cdot), \nu - \mu \rangle, \forall \nu, \mu.$$

- In terms of random variables, and use the fundamental theorem of calculus,

$$\begin{aligned} g(\mathcal{L}(Y)) - g(\mathcal{L}(X)) = \int_0^1 & \mathbb{E}[\nabla g(s\mathcal{L}(Y) + (1-s)\mathcal{L}(X), Y) \\ & - \nabla g(s\mathcal{L}(Y) + (1-s)\mathcal{L}(X), X)] ds. \end{aligned}$$

Main result

Equilibrium condition

Suppose $\xi^t = \partial_x \nabla g(\mathbb{P}^t_{\bar{X}_T}, \bar{X}_T)$. Under certain technical assumptions on θ, σ, $\bar{\pi}$, ξ^t, $\mathbb{E}_s[\bar{X}_T \xi^t]$, $Z^{\bar{X}_T \xi^t, \mathcal{S}}$, M_0 and M_1, if we have

$$\boxed{(\kappa(t) - \sigma^\dagger(t)\bar{\pi}_t)\mathbb{E}_t[\bar{X}_T \xi^t] + Z^{\bar{X}_T \xi^t, \mathcal{S}}(t) = 0}, \quad t \in [0, T],$$

then $\bar{\pi}$ is a Type-I or Type-II equilibrium, depending on the technical assumptions we impose.

Remark: This is the first-order condition from the stochastic maximum principle, with $p^t(s) = \bar{X}_s^{-1}\mathbb{E}_s[\bar{X}_T \xi^t]$, $q^{t,\mathcal{S}}(s) = Z^{\bar{X}_T \xi^t, \mathcal{S}}(s)$, $q^{t,\mathcal{O}}(s) = Z^{\bar{X}_T \xi^t, \mathcal{O}}(s) - \sigma(s)\bar{\pi}(s)_s p^t(s)$.

Proof idea

Under some convexity conditions:

$$g(\mathbb{P}^t_{\bar{X}_T^{t,\varepsilon,\varphi}}) - g(\mathbb{P}^t_{\bar{X}_T}) \lesssim \mathbb{E}_t\left[\xi^t(\bar{X}_T^{t,\varepsilon,\varphi} - \bar{X}_T)\right] = \mathbb{E}_t[p^t(T)(\bar{X}_T^{t,\varepsilon,\varphi} - \bar{X}_T)]$$

Using Itô and the BSDE of p and q (adjoint equations!):

$$\mathbb{E}_t[p^t(T)(\bar{X}_T^{t,\varepsilon,\varphi} - \bar{X}_T)] \lesssim -\int_t^{t+\varepsilon} (\bar{X}_s^{t,\varepsilon,\varphi} - \bar{X}_s)(p^t(s)\theta(s) + \sigma(s)q^{t,\mathcal{S}}(s))\mathrm{d}s + o(\varepsilon)$$

Specific examples

- EU example: Suppose we are in the dollar amount case, the market is complete and U is smooth (say, exponential utility). Then $\xi^t = U'(\bar{X}_T)$, and FOC$\Leftrightarrow Z_t^{U'(\bar{X}_T)} = -\kappa(t)\mathbb{E}_t[U'(\bar{X}_T)] \Leftrightarrow \mathbb{E}_s[U'(\bar{X}_T)] = \lambda Y_s$, with Y the (unique) pricing kernel: $\mathrm{d}Y_s = -\kappa(s)Y_s\mathrm{d}W_s$. Thus, $\bar{X}_T = U'^{(-1)}(\lambda Y_T)$.
- MV example: Because $g(\mathbb{P}_X) = \mathbb{E}[X] - \frac{\gamma}{2}\mathbb{E}[X^2] + \frac{\gamma}{2}(\mathbb{E}[X])^2$, we have $\partial_x \nabla g(\mathbb{P}_X, x) = 1 - \gamma x + \gamma\mathbb{E}[X]$, and

$$\xi^t = 1 - \gamma\bar{X}_T + \gamma\mathbb{E}_t[\bar{X}_T], \quad Z^{\xi^t,\mathcal{S}}(s) = -\gamma Z^{\bar{X}_T,\mathcal{S}}(s).$$

Thus, FOC turns into a simple form $Z^{\bar{X}_T,\mathcal{S}}(t) = \frac{\kappa(t)}{\gamma}$. This reproduces results in [Basak and Chabakauri, 2010].

Specific examples

Weighted CRRA utility:

$$g(\mathbb{P}_X) = \frac{\mathbb{E}[X^{1-\rho} \cdot X^{\gamma}]}{(1-\rho)\mathbb{E}[X^{\gamma}]}.$$

- To make g monotone and concave, we require $-1 < \gamma \leq 0$, $\gamma \leq \rho < \gamma + 1$.
- Decreasing weight function: put more weights on bad scenarios.
- Extensions to other types of weight functions are possible.

Weighted utility

Derivatives of weighted utility

Let g be given by weighted utility. Then standing assumptions are satisfied. Moreover,

$$\nabla g(\mu, x) = \frac{1}{1-\rho} \cdot \frac{x^{1-\rho+\gamma} \int_0^\infty x^\gamma \mu(\mathrm{d}x) - x^\gamma \int_0^\infty x^{1-\rho+\gamma} \mu(\mathrm{d}x)}{\left(\int_0^\infty x^\gamma \mu(\mathrm{d}x)\right)^2}.$$

Next question: how to transform FOC to something we can solve? Both powers of terminal endowments should be important:

$$Y_i(s) := \mathbb{E}_s[\bar{X}_T^{r_i}], \quad Z_i(s) := Z^{Y_i(T)}(s),$$

with $r_1 = \gamma$, $r_2 = 1 - \rho + \gamma$.

Weighted utility

Verification theorem

Let g be given by weighted utility. Then a Type-II equilibrium is given by

$$\sigma(t)\bar{\pi}_t = \frac{1}{\rho - 2\gamma}\kappa(t) + \frac{1}{\rho - 2\gamma}[\lambda_1 \bar{Z}_1(t) + \lambda_2 \bar{Z}_2(t)],$$

in which $\bar{Z}_1$ and $\bar{Z}_2$ is a solution of

$$\begin{cases} \mathrm{d}\bar{Y}_i(s) = -\dfrac{1}{2}\Big[\bar{Z}(s)^\dagger \mathbf{C}^i \bar{Z}(s) + \mathbf{c}_{i,i}\bar{Z}_i(s)\kappa(s) \\ \qquad\qquad + \mathbf{c}_{-i,i}\bar{Z}_{-i}\kappa(s) + \mathbf{b}_i|\kappa(s)|^2\Big]\,\mathrm{d}s \\ \qquad\qquad + \bar{Z}_i(s)\mathrm{d}W_s,\ i = 1, 2, \\ \bar{Y}_1(T) = \bar{Y}_2(T) = 0, \end{cases}$$

So far we assume $\mathbb{P}^*$ is known and derive solutions from it.

Fact: most model-based solutions perform very badly in empirical studies...

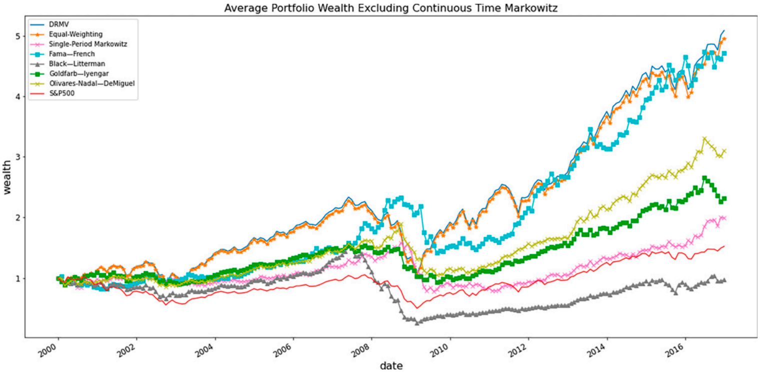


Figure: Taken from [Blanchet, Chen, and Zhou, 2021].

Resolving model uncertainty (misspecification): learning

Q-learning:

- $Q^\pi(x,a) := J(t,x;\pi \oplus_1 a)$;
- For some π_n, try to use some parametrized function F_θ (say, a network) to approximate Q^π: $F_{\theta_n} \approx Q^{\pi_n}$;
- Update π_{n+1} so that $\pi_{n+1}(x) \in \mathrm{argmax}_a F_{\theta_n}(x,a)$.

In continuous time...

- Q^π is not well-defined: changing the strategy at a single time instance does not affect pay-offs!
- The key step to continuous-time (reinforcement) learning is to find a function to learn.
- [Jia and Zhou, 2023]:

$$q^\pi(t,x,a) := \lim_{\triangle t \to 0} \frac{Q^\pi_{\triangle t}(t,x,a) - Q^\pi(t,x,a)}{\triangle t}$$

- Stochastic control perspective:

$$q^\pi(t,x,a) = \partial_t J(t,x;\pi) + H(t,x,a,\partial_x J(t,x;\pi), \partial_{xx} J(t,x;\pi))$$

q-learning of MFG/MFC

Because stochastic control theory of MFG/MFC has been developed, q-learning framework of these problems can also be devised.

- [Wei, Yu, and Yuan, 2024]: Use the representative agent's perspective to construct q-function and update policy $\pi(t, x, \mu)$.
- But still need $\bar{J}(t, \mu)$ (aggregate value function) in iterations.
- Allow μ to be estimated from empirical samplings in algorithms (decentralized learning).
- Several examples with exact parameterizations, including LQ and non-LQ.

Resolving model uncertainty (misspecification): DRO

Model uncertainty: $\mathbb{P}^*$ unknown. $\mathbb{P}^n := \frac{1}{n}\sum \delta_{X_i}$ (empirical measure) known.

DRO: find a " robust " solution for *any* model $\mathbb{P}'$ in a certain candidate set. [Blanchet, Chen, and Zhou, 2021]: $\mathcal{W}(\mathbb{P}', \mathbb{P}^n) < \delta$.

Questions:

- Is $\mathbb{P}^*$ in the candidate set? How many samples (n) do we need to achieve "robustness"?
- Dynamic problems? ($\mathcal{AW}(\cdot, \cdot)$?)
- Figure out why $1/n$ portfolio performs quite well in many empirical tests?

References

Zongxia Liang, Jianming Xia, Fengyi Yuan (2023)
Dynamic portfolio selection for nonlinear law-dependent preferences.
Preprint arXiv: 2311.06745.

Jose Blanchet, Lin Chen, Xun Yu Zhou (2021)
Distributionally robust mean-Variance portfolio selection with Wasserstein distances
Management Science 68(9): 6382-6410.

Yanwei Jia, Xun Yu Zhou (2023)
q-Learning in Continuous Time.
Journal of Machine Learning Research 24:1-61.

Xiaoli Wei, Xiang Yu, Fengyi Yuan (2024)
Unified continuous-time q-learning for mean-field game and mean-field control problems.
Preprint arXiv: 2407.04521.

Suleyman Basak, Georgy Chabakauri (2010)
Dynamic Mean-Variance Asset Allocation.
The Review of Financial Studies 23(8): 2970-3016.

Thank you!

Personal Webpage: https://fy-yuan.github.io