Stochastic Optimal Control

Bangyan Liao
liaobangyan@westlake.edu.cn

Nov 2024

WESTLAKE UNIVERSITY

Outline

Outline

Stochastic Optimal Control

Definition (Stochastic Optimal Control)

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathcal{P})$ be a fixed filtered probability space on which is defined a Brownian motion $W = (W_t)_{t\geq 0}$. We consider the control-affine problem

$$\min_{u\in\mathcal{U}} \mathbb{E}\Big[\int_0^T f(X_t^u, u_t, t)\,\mathrm{d}t + g(X_T^u)\Big],$$
$$\text{where } \mathrm{d}X_t^u = \big(b(X_t^u, t) + \sigma(t)u(X_t^u, t)\big)\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}W_t, \qquad X_0^u \sim p_0. \tag{1}$$

and where $X_t^u \in \mathbb{R}^d$ is the state, $u : \mathbb{R}^d \times [0, T]$ is the feedback control and belongs to the set of admissible controls $\mathcal{U}$, f is the state cost, $g : \mathbb{R}^d \to \mathbb{R}$ is the terminal cost, $b : \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$ is the base drift, and $\sigma : [0, T] \to \mathbb{R}^{d\times d}$ is the invertible diffusion coefficient and $\lambda \in (0, +\infty)$ is the noise level.

Definition (Cost Functional and Value Function)

The *cost functional* for the control u, point x and time t is defined as $J(u; x, t) := \mathbb{E}\big[\int_t^T f(X_t^u, u_t, t)\,\mathrm{d}t + g(X_T^u)\big|X_t^u = x\big]$. That is, the cost functional is the expected value of the control objective restricted to the times $[t, T]$ with the initial value x at time t. The *value function* or *optimal cost-to-go* at a point x and time t is defined as the minimum value of the cost functional across all possible controls:

$$V(x, t) := \inf_{u \in \mathcal{U}} J(u; x, t) = J(u^*; x, t). \tag{2}$$

Definition (HJB Optimality Condition for SOC)

If we define the infinitesimal generator $\mathcal{L} := \frac{\lambda}{2}\sum_{i,j=1}^{d}(\sigma\sigma^{\top})_{ij}(t)\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{d} b_i(x,t)\partial_{x_i} + \sum_{i=1}^{d}\sigma_i(t)u_i(x,t)\partial_{x_i}$, the value function solves the following Hamilton-Jacobi-Bellman (HJB) partial differential equation:

$$\frac{\partial V(x,t)}{\partial t} + \min_{u\in\mathcal{U}}\{\mathcal{L}V(x,t) + f(x,u,t)\} = 0, V(x,T) = g(x). \tag{3}$$

Definition (Stochastic Maximum Principle)

$$\mathcal{H}(t,x,a,y,z) = b(t,x,a)y + \sigma(t,x,a)z + f(t,x,a).$$

Assume that $(\alpha_t^*) \in \mathcal{A}$ and the pair $((Y_t^*),(Z_t^*))$ is a solution to the BSDE

$$-dY_t = \mathcal{H}_x(t, X_t^*, \alpha_t^*, Y_t, Z_t)dt - Z_t dW_t, \tag{7}$$
$$Y_T = g_x(X_T^*), \tag{8}$$

such that

$$\mathcal{H}(t, X_t^*, \alpha_t^*, Y_t^*, Z_t^*) = \max_{a\in\mathcal{A}} \mathcal{H}(t, X_t, \alpha_t, Y_t^*, Z_t^*) \tag{9}$$

for $0 \le t \le T$ almost surely, where X_t^* is the solution of (5) under the control (α_t^*). If the function

$$(x,a) \mapsto \mathcal{H}(t,x,a,Y_t^*,Z_t^*) \tag{10}$$

is concave for all $t \in [0,T]$ a.s., then (α_t^*) is the solution of the stochastic optimal

Proof.

Recall the Itô Lemma for SDE $\mathrm{d}X_t^u = \left(b(X_t^u, t) + \sigma(t)u(X_t^u, t)\right)\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}W_t$:

$$\begin{aligned} \mathrm{d}V(X_t^u, t) &= \frac{\partial V(X_t^u, t)}{\partial t}\mathrm{d}t + \frac{\partial V(X_t^u, t)}{\partial x}\mathrm{d}X_t^u + \frac{1}{2}\frac{\partial^2 V(X_t^u, t)}{\partial x^2}(\mathrm{d}X_t^u)^2 \\ &= \frac{\partial V_t}{\partial t}\mathrm{d}t + \mathcal{L}V(x, t)\mathrm{d}t + \nabla V_t(x) \cdot \sqrt{\lambda}\sigma(t)\mathrm{d}W_t \end{aligned} \tag{4}$$

where the $\mathcal{L}V(x, t)$ is the generator which defines as

$$\mathcal{L}V(x, t) = \nabla V_t(x) \cdot \left(b(X_t^u, t) + \sigma(t)u(X_t^u, t)\right) + \frac{\lambda}{2} \mathit{Trace}\left[\sigma(t)\sigma(t)^\top \nabla^2 V_t(x)\right] \tag{5}$$

Proof.

We can derive the HJB equation for SOC through dynamic programming as:

$$
\begin{aligned}
V(X_s^u, s) &= \inf_u \mathbb{E}\left\{\int_s^{s+\Delta s} f(X_t^u, u, t)\mathrm{d}t + V(X_{s+\Delta s}^u, s+\Delta s)\right\} \\
&\approx \inf_u \mathbb{E}\left\{f(X_s^u, u, s)\Delta s + V(X_{s+\Delta s}^u, s+\Delta s)\right\} \\
&\approx \inf_u \mathbb{E}\{f(X_t^u, u, s)\Delta s + V(X_s^u, s) \\
&\quad + \partial_s V(\mathbf{z}, s)\Delta s + \mathcal{L}V(\mathbf{z}, s)\Delta s + \nabla V_s(\mathbf{z}) \cdot \sqrt{\lambda}\sigma(s)\Delta W_s\}
\end{aligned}
\tag{4}
$$

$$
\mathrm{d}X_t^u = \left(b(X_t^u, t) + \sigma(t)u(X_t^u, t)\right)\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}W_t, \quad t \in [s, \tau], \quad X_s^u = \mathbf{z}
$$

□

Outline

Definition (Standard Linear Quadratic-Regularized SOC)

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathcal{P})$ be a fixed filtered probability space on which is defined a Brownian motion $W = (W_t)_{t\geq 0}$. We consider the control-affine problem

$$\begin{aligned} &\min_{u\in\mathcal{U}} \mathbb{E}\Big[\int_0^T \big(\frac{1}{2}\|u(X_t^u, t)\|^2 + f(X_t^u, t)\big)\,\mathrm{d}t + g(X_T^u)\Big], \\ &\text{where } \mathrm{d}X_t^u = \big(b(X_t^u, t) + \sigma(t)u(X_t^u, t)\big)\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}W_t, \qquad X_0^u \sim p_0. \end{aligned} \tag{5}$$

and where $X_t^u \in \mathbb{R}^d$ is the state, $u : \mathbb{R}^d \times [0, T]$ is the feedback control and belongs to the set of admissible controls $\mathcal{U}$, $f : \mathbb{R}^d \times [0, T] \to \mathbb{R}$ is the state cost, $g : \mathbb{R}^d \to \mathbb{R}$ is the terminal cost, $b : \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$ is the base drift, and $\sigma : [0, T] \to \mathbb{R}^{d\times d}$ is the invertible diffusion coefficient and $\lambda \in (0, +\infty)$ is the noise level.

Definition (Linear KL-Regularized SOC)

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathcal{P})$ be a fixed filtered probability space on which is defined a Brownian motion $W = (W_t)_{t\geq 0}$. We consider the control-affine problem

$$\min_{u\in\mathcal{U}} \mathbb{E}\left[\int_0^T (f(X_t^u, t))\,\mathrm{d}t + g(X_T^u)\right] + \lambda \mathbb{E}_{X_0\sim p_0^u}\left[D_{\mathrm{KL}}(p^u(\boldsymbol{X}|X_0)\|p^{base}(\boldsymbol{X}|X_0))\right], \quad (6)$$
$$\text{s.t. } \mathrm{d}X_t^u = (b(X_t^u, t) + \sigma(t)u(X_t^u, t))\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}B_t, \quad X_0^u \sim p_0$$

and where $X_t^u \in \mathbb{R}^d$ is the state, $u : \mathbb{R}^d \times [0, T]$ is the feedback control and belongs to the set of admissible controls $\mathcal{U}$, $f : \mathbb{R}^d \times [0, T] \to \mathbb{R}$ is the state cost, $g : \mathbb{R}^d \to \mathbb{R}$ is the terminal cost, $b : \mathbb{R}^d \times [0, T] \to \mathbb{R}^d$ is the base drift, and $\sigma : [0, T] \to \mathbb{R}^{d\times d}$ is the invertible diffusion coefficient and $\lambda \in (0, +\infty)$ is the noise level.

Definition (HJB equation fot Linear Quadratic-Regularized SOC)

Since the unique optimal control is given in terms of the value function as $u^*(x,t) = -\sigma(t)^\top \nabla V(x,t)$. If we define the infinitesimal generator $L := \frac{\lambda}{2}\sum_{i,j=1}^{d}(\sigma\sigma^\top)_{ij}(t)\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{d} b_i(x,t)\partial_{x_i}$, the value function solves the following Hamilton-Jacobi-Bellman (HJB) partial differential equation:

$$
\begin{aligned}
&(\partial_t + L)V(x,t) - \frac{1}{2}\|(\sigma^\top \nabla V)(x,t)\|^2 + f(x,t) = 0, \\
&V(x,T) = g(x).
\end{aligned}
\tag{7}
$$

Proof of HJB equation.

Recall the HJB equations of general SOC

$$\frac{\partial V(x,t)}{\partial t} + \min_{u \in \mathcal{U}} \{\mathcal{L}V(x,t) + f(x,u,t)\} = 0, V(x,T) = g(x)$$
$$\mathcal{L} := \frac{\lambda}{2} \sum_{i,j=1}^{d} (\sigma\sigma^\top)_{ij}(t)\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{d} b_i(x,t)\partial_{x_i} + \sum_{i=1}^{d} \sigma_i(t)u_i(x,t)\partial_{x_i} \tag{8}$$

with $f(x,u,t) = f(x,t) + \frac{1}{2}\left\|u(x,t)\right\|^2$ Take the gradient and set to zero, we can derive the optimal control:

$$u^*(x,t) = -\sigma(t)\nabla_x V(x,t) \tag{9}$$

Substitute the optimal control into the general SOC problem, we complete the proof.

Theorem (Optimal Conditional Distribution)

$$p^*(\boldsymbol{X}|X_0) = p^{base}(\boldsymbol{X}|X_0)\exp\left(-\lambda^{-1}\int_0^1 f(X_t,t)\mathrm{d}t - \lambda^{-1}g(X_1)\right)/C^1_{tar}$$
$$C^1_{tar} = \mathbb{E}_{\boldsymbol{X}\sim p_{base}(\boldsymbol{X}|X_0)}\left[\exp\left(-\lambda^{-1}\int_0^1 f(X_t,t)\mathrm{d}t - \lambda^{-1}g(X_1)\right)\right] = \exp\left(-\frac{V(X_0,0)}{\lambda}\right) \tag{10}$$

Remark.

We can view the optimal control as a weighted base control.

Optimal Initial Distribution

Linear Quadratic-Regularized SOC

Theorem (Optimal Initial Distribution)

$$
\begin{aligned}
p^*(X_0) &= p^{base}(X_0) \exp\left(-\frac{V(X_0, 0)}{\lambda}\right) / C_{tar}^2, \\
C_{tar}^2 &= \int p^{base}(X_0) \exp\left(-\frac{V(X_0, 0)}{\lambda}\right) = \mathbb{E}_{\boldsymbol{X} \sim p_{base}(X)} \left[\exp\left(-\lambda^{-1} \int_0^1 f(X_t, t)\mathrm{d}t - \lambda^{-1} g(X_1)\right)\right]
\end{aligned}
\tag{11}
$$

Theorem (Optimal Joint Distribution (optimal p_0))

$$p^*(\boldsymbol{X}) = \frac{p^{\mathrm{base}}(\boldsymbol{X}) \exp\left(-\lambda^{-1} \int_0^1 f(X_t, t)\mathrm{d}t - \lambda^{-1} g(X_1)\right)}{\mathbb{E}_{\boldsymbol{X} \sim p_{base}(X)} \left[\exp\left(-\lambda^{-1} \int_0^1 f(X_t, t)\mathrm{d}t - \lambda^{-1} g(X_1)\right)\right]} \tag{12}$$

Theorem (Optimal Joint Distribution (fixed p_0))

$$p^*(\boldsymbol{X}) = p^{\mathrm{base}}(\boldsymbol{X}) \exp\left(-\int_0^1 f(X_t, t)\mathrm{d}t - g(X_1) + V(X_0, 0)\right) \tag{13}$$

Outline

Theorem (Path-integral representation of the optimal control (Feynman-Kac))

$$
\begin{aligned}
u^*(x,t) &= \lambda\sigma(t)^\top \nabla_x \log \mathbb{E}\Big[\exp\Big(-\lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1} g(X_T)\Big)\Big| X_t = x\Big] \\
V(x,t) &= -\lambda \log \mathbb{E}\Big[\exp\Big(-\lambda^{-1}\int_t^T f(X_s,s)\,\mathrm{d}s - \lambda^{-1} g(X_T)\Big)\Big| X_t = x\Big],
\end{aligned}
\tag{14}
$$

where X_t is generated by the uncontrolled process. The optimal control and the value function are related to each other by $u^*(x,t) = -\sigma(t)^\top \nabla V(x,t)$.

Proof. (Path-integral Control).

Let us recall the HJB optimality condition

$$\begin{aligned} &(\partial_t + L)V(x,t) - \frac{\lambda}{2}\|(\sigma^\top \nabla V)(x,t)\|^2 + f(x,t) = 0, \\ &L = \frac{\lambda}{2}\sum_{i,j=1}^{d}(\sigma\sigma^\top)_{ij}(t)\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{d} b_i(x,t)\partial_{x_i} \\ &V(x,T) = g(x). \end{aligned} \tag{15}$$

and perform the Cole-Hopf transform $V(x,t) = -\lambda \ln \Psi(x,t)$.

Proof. (Path-integral Control).

and perform the Cole-Hopf transform $V(x,t) = -\lambda \ln \Psi(x,t)$.

$$
\begin{aligned}
& -\lambda \frac{\partial_t \Psi + L\Psi}{\Psi}(x,t) + \frac{\lambda^2}{2} \| \frac{\sigma^\top \nabla \Psi}{\Psi}(x,t)\|^2 - \frac{\lambda^2}{2} \| \frac{\sigma^\top \nabla \Psi}{\Psi}(x,t)\|^2 + f(x,t) = 0 \\
& L = \frac{\lambda}{2} \sum_{i,j=1}^{d} (\sigma\sigma^\top)_{ij}(t) \partial_{x_i}\partial_{x_j} + \sum_{i=1}^{d} b_i(x,t)\partial_{x_i} \\
& \Psi(x,T) = \exp(-\lambda^{-1} g(x)).
\end{aligned} \tag{15}
$$

Proof. (Path-integral Control).

After some canceling processes, we have

$$
\begin{aligned}
&\partial_t \Psi(x,t) + L\Psi(x,t) - \lambda^{-1}\Psi(x,t) f(x,t) = 0 \\
&L = \frac{\lambda}{2}\sum_{i,j=1}^{d} (\sigma\sigma^\top)_{ij}(t)\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{d} b_i(x,t)\partial_{x_i} \\
&\Psi(x,T) = \exp(-\lambda^{-1} g(x)).
\end{aligned} \tag{15}
$$

Then, let us recall the Feynman-Kac formulation:

$$
\begin{cases}
\frac{\partial u(x,t)}{\partial t} + \mu(x,t)\frac{\partial u(x,t)}{\partial x} + \frac{1}{2}\sigma^2(x,t)\frac{\partial^2 u(x,t)}{\partial x^2} - q(x,t)u(x,t) = -g(x,t) \\
u(x,T) = f(x)
\end{cases} \tag{16}
$$

with its conclusion

Proof. (Path-integral Control).

$$u(x,t) = \mathbb{E}\left[f(\xi_T)e^{-\int_t^T q(\theta,\xi_\theta)d\theta} + \int_t^T g(s,\xi_s)e^{-\int_t^s q(\theta,\xi_\theta)d\theta} ds | \xi_t = x\right] \tag{15}$$

Then, substitute it into the original formula,

$$\Psi(x,t) = \mathbb{E}\left[\exp(-\lambda^{-1}g(x))\exp(-\lambda^{-1}\int_t^T f(s,X_s)ds)|X_t = x\right] \tag{16}$$

□

Theorem (Path-integral representation of the optimal control (Kappen))

$$\begin{aligned} u^*(x,t) &= \lambda\sigma(t)^\top \nabla_x \log \mathbb{E}_{\boldsymbol{X}\sim p^{WFR}} \left[\exp\left(-\lambda^{-1} g(X_T)\right) \middle| X_t = x\right] \\ V(x,t) &= -\lambda \log \mathbb{E}_{\boldsymbol{X}\sim p^{WFR}} \left[\exp\left(-\lambda^{-1} g(X_T)\right) \middle| X_t = x\right], \end{aligned} \tag{17}$$

where X_t is generated by the uncontrolled fisher-rao process.

$$\begin{cases} X_t = X_t + b(X_t,t)\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}B_t, & 1 - f(x,t)\,dt/\lambda \\ X_t = \dagger, & f(x,t)\,dt/\lambda \end{cases}, \quad X_t = x \tag{18}$$

The optimal control and the value function are related to each other by $u^*(x,t) = -\sigma(t)^\top \nabla V(x,t)$.

Path-integral Control (Kappen).

We first perform the Cole-Hopf transform $V(x,t) = -\lambda \ln \Psi(x,t)$to HJB equation.

$$
\begin{aligned}
-\lambda \frac{\partial_t \Psi + \mathcal{L}\Psi}{\Psi}(x,t) + \frac{\lambda^2}{2}\left\|\frac{\sigma^\top \nabla \Psi}{\Psi}(x,t)\right\|^2 - \frac{\lambda^2}{2}\left\|\frac{\sigma^\top \nabla \Psi}{\Psi}(x,t)\right\|^2 + f(x,t) = 0 \\
\partial_t \Psi(x,t) + \mathcal{L}\Psi(x,t) - \lambda^{-1}\Psi(x,t) f(x,t) = 0 \\
\mathcal{L} = \frac{\lambda}{2}\sum_{i,j=1}^{d} (\sigma\sigma^\top)_{ij}(t)\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{d} b_i(x,t)\partial_{x_i} \\
\Psi(x,1) = \exp(-\lambda^{-1} g(x))
\end{aligned}
\tag{19}
$$

Path-integral Control (Kappen).

HJB equation is also a Kolmogorov backward equation, which has its adjoint Kolmogorov forward equation which describes the forward distribution evolution.

$$
\begin{aligned}
\partial_t \rho(x,t) &= \mathcal{L}^\dagger \rho(x,t) - \lambda^{-1}\rho(x,t) f(x,t) \\
\mathcal{L}^\dagger \rho(x,t) &= \frac{\lambda}{2}\sum_{i,j=1}^{d} \partial_{x_i}\partial_{x_j}\left((\sigma\sigma^\top)_{ij}(t)\,\rho\right) - \sum_{i=1}^{d} \partial_{x_i}\left(b_i(x,t)\,\rho\right) \\
\rho(y,t|x,t) &= \delta(y-x)
\end{aligned}
\tag{19}
$$

Then, according to the generator definition, we have

$$
\Psi(x,t) = \int \rho(y,T|x,t)\exp(-\lambda^{-1} g(x))\,dy \tag{20}
$$

Outline

Given the Hamilton-Jacobi-Bellman equation and the Fokker-Plank equation:

$$
\begin{aligned}
&(\partial_t + \mathcal{L})V(x,t) - \frac{1}{2}\|(\sigma^T \nabla V)(x,t)\|^2 + f(x,t) = 0, \quad V(x,T) = g(x). \\
&\mathcal{L} := \frac{\lambda}{2}\sum_{i,j=1}^{d}(\sigma\sigma^T)_{ij}(t)\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{d} b_i(x,t)\partial_{x_i}
\end{aligned}
\tag{HJB}
$$

$$
\begin{aligned}
&(\partial_t - \mathcal{L}^*)p(x,t) + \nabla \cdot [(\sigma\sigma^T \nabla V p)(x,t)] = 0, \quad p(x,0) = p_0 \\
&\mathcal{L}^* p(x,t) := \frac{\lambda}{2}\sum_{i,j=1}^{d}\partial_{x_i}\partial_{x_j}[(\sigma\sigma^T p)(x,t)] - \sum_{i=1}^{d}\partial_{x_i}[(bp)(x,t)]
\end{aligned}
\tag{FK}
$$

We can get the forward-backward PDEs system through Cole-Hopf transformation:

$$
\Psi(x,t) = \exp\left(-\frac{V(x,t)}{\lambda}\right), \quad \hat{\Psi}(x,t) = p(x,t)\exp\left(\frac{V(x,t)}{\lambda}\right) \tag{21}
$$

Theorem (Forward and Backward PDEs)

We can get the corresponding forward-backward PDEs system:

$$\begin{cases} \dfrac{\partial \Psi(x,t)}{\partial t} = -\nabla \Psi^\top b - \dfrac{\lambda}{2}\sigma^2 \Delta \Psi + \lambda^{-1} f \Psi \\ \dfrac{\partial \hat{\Psi}(x,t)}{\partial t} = -\nabla \cdot (\hat{\Psi} b) + \dfrac{\lambda}{2}\sigma^2 \Delta \hat{\Psi} - \lambda^{-1} f \hat{\Psi} \end{cases} \quad s.t. \quad \begin{array}{c} \Psi(\cdot,0)\hat{\Psi}(\cdot,0) = p_0 \\ \Psi(\cdot,T)\hat{\Psi}(\cdot,T) = p_T \end{array}. \tag{22}$$

Proof. (Proof of Forward-Backward PDEs system).

1) Given the HJB equation, we first substitute the Cole-Hopf transform $V(x,t) = -\lambda \ln \Psi(x,t)$ into it and get

$$-\lambda \frac{\partial_t \Psi + \mathcal{L}\Psi}{\Psi}(x,t) + \frac{\lambda^2}{2} \left\| \frac{\sigma^\top \nabla \Psi}{\Psi}(x,t) \right\|^2 - \frac{\lambda^2}{2} \left\| \frac{\sigma^\top \nabla \Psi}{\Psi}(x,t) \right\|^2 + f(x,t) = 0 \tag{23}$$

After some calculation, we get

$$\begin{gathered} \partial_t \Psi(x,t) + \mathcal{L}\Psi(x,t) - \lambda^{-1}\Psi(x,t) f(x,t) = 0 \\ \mathcal{L} = \frac{\lambda}{2} \sum_{i,j=1}^{d} (\sigma\sigma^\top)_{ij}(t) \partial_{x_i} \partial_{x_j} + \sum_{i=1}^{d} b_i(x,t) \partial_{x_i} \\ \Psi(x,1) = \exp(-\lambda^{-1} g(x)) \end{gathered} \tag{24}$$

Proof. (Proof of Forward-Backward PDEs system).

2) Starting directly from $\rho = \Psi\hat{\Psi}$, differentiate ρ with respect to time. Using the product rule and substituting equations:

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= \frac{\partial(\Psi\hat{\Psi})}{\partial t} = \frac{\partial \Psi}{\partial t}\hat{\Psi} + \Psi\frac{\partial \hat{\Psi}}{\partial t} \\ &= \left(-\nabla\Psi^\top f - \frac{1}{2}\sigma^2\Delta\Psi + F\Psi\right)\hat{\Psi} + \Psi\left(-\nabla\cdot(\hat{\Psi}f) + \frac{1}{2}\sigma^2\Delta\hat{\Psi} - F\hat{\Psi}\right)\end{aligned}$$

We regroup the above expression by the terms associated with f, σ^2, and F:

$$\frac{\partial \rho}{\partial t} = \underbrace{-\left[(\nabla\Psi\cdot f)\hat{\Psi} + \Psi\nabla\cdot(\hat{\Psi}f)\right]}_{\text{Term A}} + \underbrace{\frac{1}{2}\sigma^2\left(\Psi\Delta\hat{\Psi} - \hat{\Psi}\Delta\Psi\right)}_{\text{Term B}} + \underbrace{(F\Psi\hat{\Psi} - F\Psi\hat{\Psi})}_{\text{Term C=0}}$$

Proof. (Proof of Forward-Backward PDEs system).

(Simplify Term A)

Using the vector identity $\nabla \cdot (A\mathbf{B}) = (\nabla A) \cdot \mathbf{B} + A(\nabla \cdot \mathbf{B})$, we note that Term A is the negative of the divergence of $\rho f = (\Psi \hat{\Psi}) f$:

$$\text{Term A} = -\left[\nabla \cdot ((\Psi \hat{\Psi}) f)\right] = -\nabla \cdot (\rho f)$$

Proof. (Proof of Forward-Backward PDEs system).

(Simplify Term B)
We use the divergence form of Green's second identity:
$\nabla \cdot (A\nabla B - B\nabla A) = A\Delta B - B\Delta A$.

$$\text{Term B} = \frac{1}{2}\sigma^2 \nabla \cdot (\Psi \nabla \hat{\Psi} - \hat{\Psi} \nabla \Psi)$$

Now, we compute the expression inside the parentheses, based on the definitions $\Psi = e^{-u}$ and $\hat{\Psi} = \rho e^{u}$:

$$\Psi \nabla \hat{\Psi} = e^{-u} \nabla (\rho e^{u}) = \nabla \rho + \rho \nabla u$$
$$\hat{\Psi} \nabla \Psi = \rho e^{u} \nabla (e^{-u}) = -\rho \nabla u$$

Proof. (Proof of Forward-Backward PDEs system).

Thus:

$$\Psi \nabla \hat{\Psi} - \hat{\Psi} \nabla \Psi = (\nabla \rho + \rho \nabla u) - (-\rho \nabla u) = \nabla \rho + 2\rho \nabla u$$

Substituting this back into Term B:

$$\text{Term B} = \frac{1}{2}\sigma^2 \nabla \cdot (\nabla \rho + 2\rho \nabla u) = \frac{1}{2}\sigma^2 \Delta \rho + \sigma^2 \nabla \cdot (\rho \nabla u)$$

Proof. (Proof of Forward-Backward PDEs system).

(Combine the Results)
Substituting the simplified Term A and Term B into the equation:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho f) + \frac{1}{2}\sigma^2 \Delta \rho + \sigma^2 \nabla \cdot (\rho \nabla u)$$

Rearranging the terms on the right-hand side of the above equation:

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \left(\sigma^2 \nabla \cdot (\rho \nabla u) - \nabla \cdot (\rho f)\right) + \frac{1}{2}\sigma^2 \Delta \rho \\ &= \nabla \cdot (\sigma^2 \rho \nabla u - \rho f) + \frac{1}{2}\sigma^2 \Delta \rho \\ \Longrightarrow \frac{\partial \rho}{\partial t} &- \nabla \cdot (\rho(\sigma^2 \nabla u - f)) - \frac{1}{2}\sigma^2 \Delta \rho = 0 \end{aligned}$$

Theorem (Forward and Backward SDEs (PDE-inspired))

$$dX_t = (b + \sigma^2 \lambda \nabla \log \Psi)\, dt + \sqrt{\lambda}\sigma\, dW_t, \quad X_0 \sim p_0^*, \quad X_1 \sim p_1^* \tag{23}$$

$$d\bar{X}_s = (-b + \sigma^2 \lambda \nabla \log \hat{\Psi})\, ds + \sqrt{\lambda}\sigma\, d\bar{W}_s, \quad \bar{X}_0 \sim p_1^*, \quad \bar{X}_1 \sim p_0^* \tag{24}$$

where $\bar{X}$ and $\bar{W}$ represents the reverse process. In fact, the stochastic processes in Eq. 23 and Eq. 24 share the same marginal densities $p_t^{(23)} = p_s^{(24)} = p_t^*$. Besides, its marginal density obeys a factorization principle $p_t^*(X_t) = \Psi(t, X_t)\, \hat{\Psi}(s, X_s)$.

Theorem (Forward and Backward SDEs)

Consider the pair of SDEs

$$\begin{aligned} \mathrm{d}X_t &= b(X_t, t)\,\mathrm{d}t + \sqrt{\lambda}\sigma(t)\mathrm{d}B_t, && X_0 \sim p_0, \\ \mathrm{d}Y_t &= (-f(X_t, t) + \frac{1}{2}\|Z_t\|^2)\,\mathrm{d}t + \sqrt{\lambda}\langle Z_t, \mathrm{d}B_t\rangle, && Y_T = g(X_T). \end{aligned} \tag{25}$$

where $Y : \Omega \times [0, T] \to \mathbb{R}$ *and* $Z : \Omega \times [0, T] \to \mathbb{R}^d$ *are progressively measurable random processes. It turns out that* Y_t *and* Z_t *defined as* $Y_t := V(X_t, t)$ *and* $Z_t := \sigma(t)^\top \nabla V(X_t, t) = -u^*(X_t, t)$ *satisfy the HJB optimality condition.*

Proof of Forward and Backward SDEs.

We apply Itô's lemma to $Y_t = V(X_t, t)$:

$$\begin{aligned} dY_t &= \frac{\partial V}{\partial t} dt + (\nabla V)^\top dX_t + \frac{1}{2} \mathrm{Tr}\left((d_B X_t)(d_B X_t)^\top H(V) \right) \\ &= \frac{\partial V}{\partial t} dt + (\nabla V)^\top (b\, dt + \sqrt{\lambda}\sigma\, dB_t) + \frac{1}{2} \mathrm{Tr}\left((\sqrt{\lambda}\sigma)(\sqrt{\lambda}\sigma)^\top H(V) \right) dt \\ &= \left(\frac{\partial V}{\partial t} + b \cdot \nabla V + \frac{\lambda}{2} \mathrm{Tr}(\sigma\sigma^\top H(V)) \right) dt + \sqrt{\lambda} (\nabla V)^\top \sigma\, dB_t \end{aligned}$$

The term in the parenthesis is the spatial part of the HJB operator. From the HJB equation, we can substitute this term:

$$\frac{\partial V}{\partial t} + b \cdot \nabla V + \frac{\lambda}{2} \mathrm{Tr}(\sigma\sigma^\top H(V)) = \frac{1}{2} \|\sigma^\top \nabla V\|^2 - f$$

Proof of Forward and Backward SDEs.

Substituting this into the expression for dY_t yields:

$$dY_t = \left(\frac{1}{2}\|\sigma^\top \nabla V(X_t, t)\|^2 - f(X_t, t)\right) dt + \sqrt{\lambda}(\sigma(t)^\top \nabla V(X_t, t))^\top dB_t$$

Now, we use the definitions from our ansatz (**??**), $Z_t = \sigma^\top \nabla V$:

$$dY_t = \left(-f(X_t, t) + \frac{1}{2}\|Z_t\|^2\right) dt + \sqrt{\lambda} Z_t \cdot dB_t$$

This is exactly the required BSDE dynamics (**??**). Finally, the terminal condition is met since:

$$Y_T = V(X_T, T) = g(X_T)$$

$\square$

Definition (Verification Theorem for Linear Quadratic SOC)

The *verification theorem* states that if a function V solves the HJB equation above and has certain regularity conditions, then V is the value function (2) of the problem (6). An implication of the verification theorem is that for every $u \in \mathcal{U}$,

$$V(x,t) + \mathbb{E}\Big[\frac{1}{2}\int_t^T \|\sigma^\top \nabla V + u\|^2(X_s^u, s)\,\mathrm{d}s \,\Big|\, X_t^u = x\Big] = J(u,x,t). \tag{26}$$

Equation (26) can be deduced by integrating the HJB equation (7) over $[t, T]$, and taking the conditional expectation with respect to $X_t^u = x$.

Proof. (Verification Theorem).

By Itô Lemma, we have that

$$
\begin{aligned}
V(X_T^u, T) - V(X_t^u, t) = \int_t^T \Big(&\partial_s V(X_s^u, s) + \langle b(X_s^u, s) + \sigma(X_s^u, s) u(X_s^u, s), \nabla V(X_s^u, s)\rangle \\
&+ \frac{\lambda}{2} \sum_{i,j=1}^{d} (\sigma\sigma^\top)_{ij}(X_s^u, s)\partial_{x_i}\partial_{x_j} V(X_s^u, s)\Big)\,\mathrm{d}s + S_t^u,
\end{aligned}
\tag{27}
$$

where $S_t^u = \sqrt{\lambda}\int_t^T \nabla V(X_s^u, s)^\top \sigma(X_s^u, s)\,\mathrm{d}B_s$. Note that by (7),

$$
\begin{aligned}
&\partial_s V(X_s^u, s) + \langle b(X_s^u, s) + \sigma(X_s^u, s) u(X_s^u, s), \nabla V(X_s^u, s)\rangle \\
&+ \frac{\lambda}{2} \sum_{i,j=1}^{d} (\sigma\sigma^\top)_{ij}(X_s^u, s)\partial_{x_i}\partial_{x_j} V(X_s^u, s)
\end{aligned}
\tag{28}
$$

Proof. (Verification Theorem).

$$
\begin{aligned}
&= \frac{1}{2}\|(\sigma^\top \nabla V)(X_s^u, s)\|^2 - f(X_s^u, s) + \langle \sigma(X_s^u, s)u(X_s^u, s), \nabla V(X_s^u, s)\rangle \\
&= \frac{1}{2}\|(\sigma^\top \nabla V)(X_s^u, s) + u(X_s^u, s)\|^2 - \frac{1}{2}\|u(X_s^u, s)\|^2 - f(X_s^u, s),
\end{aligned} \tag{27}
$$

and this implies that

$$
\begin{aligned}
g(X_T^u) - V(X_t^u, t) = \int_t^T \Big(&\frac{1}{2}\|(\sigma^\top \nabla V)(X_s^u, s) + u(X_s^u, s)\|^2 - \\
&\frac{1}{2}\|u(X_s^u, s)\|^2 - f(X_s^u, s)\Big)\, \mathrm{d}s + S_t^u
\end{aligned} \tag{28}
$$

Since $\mathbb{E}[S_t^u \mid X_t^u = x] = 0$, rearranging and taking the conditional expectation with respect to X_t^u yields the final result. $\square$

Reference

- Stochastic Optimal Control Matching
- An optimal control approach to particle filtering
- Stochastic Optimal Control
- Continuous Time Stochastic Optimal Control: Lagrange-Chow Redux