ESAIM: COCV 26 (2020) 87
https://doi.org/10.1051/cocv/2020008

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

A GLOBAL MAXIMUM PRINCIPLE FOR OPTIMAL CONTROL OF GENERAL MEAN-FIELD FORWARD-BACKWARD STOCHASTIC SYSTEMS WITH JUMPS*

Tao Hao[1] and Qingxin Meng[2,**]

Abstract. In this paper we prove a maximum principle of optimal control problem for a class of general mean-field forward-backward stochastic systems with jumps in the case where the diffusion coefficients depend on control, the control set does not need to be convex, the coefficients of jump terms are independent of control as well as the coefficients of mean-field backward stochastic differential equations depend on the joint law of $(X(t), Y(t))$. Since the coefficients depend on measure, higher mean-field terms could be involved. In order to analyse them, two new adjoint equations are brought in and several new generic estimates of their solutions are investigated. Utilizing these subtle estimates, the second-order expansion of the cost functional, which is the key point to analyse the necessary condition, is obtained, and where after the stochastic maximum principle. An illustrative application to mean-field game is considered.

Mathematics Subject Classification. 93E20, 60H10.

Received February 20, 2019. Accepted February 10, 2020.

1. Introduction

For some given measurable mappings $b, \sigma, \beta, f, \phi$, we consider the general mean-field forward-backward stochastic differential equation (FBSDE):

$$\begin{cases} \mathrm{d}X^v(t) = b(t, X^v(t), P_{X^v(t)}, v(t))\mathrm{d}t + \sigma(t, X^v(t), P_{X^v(t)}, v(t))\mathrm{d}W(t) \\ \qquad\qquad + \displaystyle\int_G \beta(t, X^v(t-), P_{X^v(t)}, e)N_\lambda(\mathrm{d}e, \mathrm{d}t), \ t \in [0,T], \\ -\mathrm{d}Y^v(t) = f(t, X^v(t), Y^v(t), Z^v(t), \displaystyle\int_G K^v(t,e)l(e)\lambda(\mathrm{d}e), P_{(X^v(t),Y^v(t))}, v(t))\mathrm{d}t \\ \qquad\qquad - Z^v(t)\mathrm{d}W(t) - \displaystyle\int_G K^v(t,e)N_\lambda(\mathrm{d}e, \mathrm{d}t), \ t \in [0,T], \\ X^v(0) = x_0, \ Y^v(T) = \phi(X^v(T), P_{X^v(T)}), \end{cases} \tag{1.1}$$

*T. Hao work has been supported by National Natural Science Foundation of China (Grant Nos. 71671104, 11871309, 11801315, 11801317, 71803097), the Ministry of Education of Humanities and Social Science Project (Grant No. 16YJA910003), Natural Science Foundation of Shandong Province (No. ZR2018QA001, ZR2019MA013), A Project of Shandong Province Higher Educational Science and Technology Program (Grant Nos. J17KA162, J17KA163), and Incubation Group Project of Financial Statistics and Risk Management of SDUFE. Q.X. Meng work has been supported by the National Natural Science Foundation of China (Grant No. 11871121) and the Natural Science Foundation of Zhejiang Province for Distinguished Young Scholar (Grant No. LR15A010001).

Keywords and phrases: Stochastic control, global maximum principle, general mean-field forward-backward stochastic differential equation with jumps.

[1] School of Statistics, Shandong University of Finance and Economics, Jinan 250014, P.R. China.
[2] Department of Mathematics, Huzhou University, Zhejiang 313000, P.R. China.

[**] Corresponding author: mqx@zjhu.edu.cn

Article published by EDP Sciences

where $P_\eta := P \circ \eta^{-1}$ denotes the probability measure induced by the random variable η; $l : G \to \mathbb{R}^+$ is a Borel function with growth condition $0 < l(e) \leq C(1 \wedge |e|),\ e \in G$. Our control problem is to minimize a cost functional of the form $J(v(\cdot)) = Y^v(0)$. The purpose of this paper is to investigate the second-order necessary condition of the above control problem in the case where σ depends on control and, moreover, the action space is a general space, which means it is needlessly convex.

The motivation comes on the one hand from the rapid development of the theory of mean-field FBSDEs in recent years, in particular, after the appearance of the notion of the derivative of a function with respect to a measure, refer to Lions [24] or Cardaliaguet [8], on the other hand from the work of Hu [17], which solved the Peng's open problem [27] completely.

Stochastic maximum principle (SMP) is an important tool to study stochastic control problem. A lot of papers on this subject have been published. The earliest works can be retrospected to Kushner [19] and Bismut [2]. We refer the subsequent works for first-order SMP to Bensoussan [1], Haussmann [16], Peng [26], Yong and Zhou [32], Framstad, Oksendal and Sulem [14]. In particular, by regarding $Z(\cdot)$ as a control process and taking the terminal condition $Y(T) = \Phi(X(T))$ as a constraint, Yong [31] in 2010 applied the Ekeland variational principle to obtain an optimality variational principle. Using the similar approach, Wu [30] in 2013 investigated the recursive stochastic optimal control problem. However, their maximum principles contain unknown parameters. Later, Hu [17] considered a special case of the model given by Yong [31], and got rid of the unknown parameters in the optimality variational principle [31], [30]. Recently, the second-order SMP also got great attention, see Zhang and Zhang [33, 34, 35], Lu *et al.* [25] and so on. All above works were investigated in classical setting, not in a mean-field framework. As everyone knows, mean-field stochastic differential equations (SDEs) (also named McKean-Vlasov equations) have been considered by Kac [18] as long ago as 1956. However, due to the special structure, mean-field backward stochastic differential equations (BSDEs) were not obtained by Buckdahn, Djehiche, Li and Peng [4] until 2009 with a purely probabilistic approach. From then on, further progresses on mean-field BSDEs were provided by, for example, Buckdahn *et al.* [6], Buckdahn *et al.* BDL. Especially, recently with the pioneer work of Lions [24] to introduce the derivative of a function defined on $\mathcal{P}_2(\mathbb{R}^d)$ with respect to a measure, the theory of general mean-field FBSDEs and related optimal control problems or potential games stirred greatly the zeal of a large number of scholars. For instance, we refer Lasry and Lions [20] for the theory of mean-field game, refer Buckdahn *et al.* [7], Hao and Li [15], Li [21], Chassagneux *et al.* [12] for the investigation of the relationship of the solutions of mean-field FBSDEs and corresponding PDEs, and refer Carmona and Delarue [9, 10], Carmona et al. [11] for the description of the approximate Nash equilibriums of symmetric games, i.e., the probability interpretation of mean-field game. Note that in [9], the authors proved the existence of the approximate Nash equilibriums by making use of the tailor-made form of SMP. However, the assumptions on their SMP are heavy, such as the Hamiltonian being strictly convex in control. A natural question is whether the necessary condition of the optimal control problem for general mean-field forward-backward stochastic systems (1.1) holds still true under some slightly loose assumptions. In this paper, we confirm this declare.

Let us look at the structure of (1.1) and show three main obstacles encountered in investigating the above mean-field optimal control problem systemly:

a) The equation (1.1) is a general mean-field FBSDE. In fact, most of the existing works in mean-field framework can be summarized as two cases:

$$\text{i) } \mathbb{E}[\psi(t,x,X(t),v)]|_{x=X(t),v=v_t}; \qquad \text{ii) } \psi(t,\omega,X(t),\mathbb{E}[X(t)],v_t).$$

However, both of the above cases can be put into the general type (1.1) by the definition of expectation and some simple transformations, for example, for i)

$$\overline{\psi}(t,\omega,X(t),P_{X(t)},v_t) := \mathbb{E}[\psi(t,x,X(t),v)]|_{x=X(t),v=v_t} = \int_{\mathbb{R}^n} \psi(t,x,y,v)P_{X(t)}(\mathrm{d}y)\Big|_{x=X(t),v=v_t}.$$

As we know, it is very difficult to analyse the second-order derivative of a function with respect to a measure. Because a function, which is infinitely differentiable in usual sense, maybe not be twice Fréchet differentiable. Buckdahn *et al.* [5] encountered the same difficulty when studying the SMP for general mean-field systems without jump. By adopting reasonable second-order derivatives of a function with respect to a measure (see [7] or [9]), and by proving two sharpen estimates (see Prop. 4.3 of [5]), the authors solved this difficulty successfully. In this paper, we use the same definition of second-order derivative as [5]. However, the important estimates proved by Buckdahn *et al.* [5] are not suitable for jump and recursive case. Hence, the first obstacle is how to generalize their estimates to jump and recursive case (see Lem. 3.9).

b) A closely related work is Buckdahn, Li and Ma [5], where the cost functional is of the form

$$J(v)=\mathbb{E}\bigg[\int_0^T f(t,X^v(t),P_{X^v(t)},v_t)\mathrm{d}t+\phi(X^v(T),P_{X^v(T)})\bigg].$$

Compared with their work, in this paper we consider the recursive case, *i.e.* f depending on (y,z). Following the scheme of [17], we need to treat the term $Y^1\delta\sigma(t)\mathbb{1}_{E_\varepsilon}(t)$ because its order is $O(\varepsilon)$. By constructing an auxiliary BSDE (4.2) we realize the second-order expansion of Y^ε. Different from [17], the equation (4.2) is a mean-field BSDE with jumps. It is difficult to obtain its explicit expression. Hence, the second obstacle is how to prove the SMP via the equation (4.2).

c) It should be pointed out that although the dynamics involves jump term, the coefficient β does not depend on control. Indeed, Li and Wei [23] proved an useful estimate for the solution of fully coupled FBSDEs as follows: For $p\geq 2, 0\leq t\leq T$, $\xi\in L^p(\Omega,\mathcal{F}_t,P;\mathbb{R}^n)$ there exists a $\delta_0>0$ depending on p and the Lipschitz constant as well as the linear growth constant, such that

$$\mathbb{E}\bigg[\sup_{s\in[t,t+\delta]}|X^{t,\zeta}(s)-\zeta|^p\bigg]\leq C_p\delta(1+|\zeta|^p),\, P\text{-a.s.},\ 0\leq\delta\leq\delta_0 \tag{1.2}$$

(see Prop. 3.2 of [23], p. 1591). According to Remark 3.4 [23], the order of δ at the right hand can not be $\frac{p}{2}$, otherwise the Kolmogorov's Continuity Criterion would imply the continuity of the jump process $X^{t,\zeta}$. But this is impossible. The reader can also refer Remark 357 [28] for similar argument.

Similar to the proof of Proposition 3.2 [23], if β depends on control, for the solution of the first-order variational equation we only have the estimate: for $\ell\geq 1$,

$$\mathbb{E}\bigg[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^{2\ell}\bigg]\leq L_\ell\varepsilon, \tag{1.3}$$

but not

$$\mathbb{E}\bigg[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^{2\ell}\bigg]\leq L_\ell\varepsilon^\ell, \tag{1.4}$$

which is not sufficient to prove the SMP. This is the last obstacle we met.

There are two points, which should be lighten. Firstly, different from the classical case, not only the Taylor expansion of X^ε, $X^\varepsilon=X^*+X^{1,\varepsilon}+X^{2,\varepsilon}+o(\varepsilon)$, but also the second-order expansion of cost functional Y^ε, $Y^\varepsilon=Y^*+\breve{P}+Y^1+Y^2+o(\varepsilon)$ are needed, where the convergence of both of them are in $L^2(\Omega,C[0,T])$ sense. To the best of our knowledge, such second-order expansion has not been seen in the existing literature. Secondly, following the scheme of Proposition 4.3 of [5], we establish two new estimates (see (3.6)). But it should be pointed out that the presence of the jump term makes our proof more technical.

The main result of this paper can be stated roughly as follows: Consider Hamiltonian

$$
\begin{aligned}
&H(t,x,y,z,k,\nu,\mu,v;p,q,P)\\
&\quad= pb(t,x,\nu,v)+q\sigma(t,x,\nu,v)+\frac{1}{2}P\Big(\sigma(t,x,\nu,v)-\sigma(t,X^*(t),P_{X^*(t)},u^*(t))\Big)^2\\
&\qquad+f\left(t,x,y,z+p\Big(\sigma(t,x,\nu,v)-\sigma(t,X^*(t),P_{X^*(t)},u^*(t))\Big),\int_G k(e)l(e)\lambda(\mathrm{d}e),\mu,v\right),
\end{aligned}
\tag{1.5}
$$

$(t,x,y,z,k,\nu,\mu,v,p,q,P)\in[0,T]\times\mathbb{R}^3\times L^2(G,\mathscr{B}(G),\lambda)\times\mathcal{P}_2(\mathbb{R})\times\mathcal{P}_2(\mathbb{R}^2)\times U\times\mathbb{R}^3$.

Let u^* be the optimal control and (X^*,Y^*,Z^*,K^*) be the optimal trajectory. By $(Y^i,Z^i,K^i),\ i=1,2$ we denote the solutions of the first- and second-order adjoint equations, respectively. Under some usual assumptions and the additional assumptions

$$
\widetilde{f}_{\mu_2}(t)=\left(\frac{\partial f}{\partial\mu}\right)_2(t,\widetilde{X}^*(t),\widetilde{Y}^*(t),\widetilde{Z}^*(t),\int_G\widetilde{K}^*(t,e)l(e)\lambda(\mathrm{d}e),P_{(X^*(t),Y^*(t))},\widetilde{u}^*(t);X^*(t),Y^*(t))\geq 0,
$$

$t\in[0,T],\ \widetilde{P}\otimes P$-a.s., and

$$
f_k(t)=\frac{\partial f}{\partial k}(t,X^*(t),Y^*(t),Z^*(t),\int_G K^*(t,e)l(e)\lambda(\mathrm{d}e),P_{(X^*(t),Y^*(t))},u^*(t))\geq 0,
$$

$t\in[0,T], P$-a.s., we have

$$
\begin{aligned}
&H(t,X^*(t),Y^*(t),Z^*(t),K^*(t,\cdot),P_{X^*(t)},P_{(X^*(t),Y^*(t))},v;Y^1(t),Z^1(t),Y^2(t))\\
&\geq H(t,X^*(t),Y^*(t),Z^*(t),K^*(t,\cdot),P_{X^*(t)},P_{(X^*(t),Y^*(t))},u^*(t);Y^1(t),Z^1(t),Y^2(t)),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\forall v\in U,\ \text{a.e.},\ \text{a.s.},
\end{aligned}
\tag{1.6}
$$

where f_{μ_2} and f_k denote the partial derivatives of f with respect to the marginal law P_Y and the variable k along the optimal sextuplet, respectively.

This paper is arranged as follows. Section 2 recalls the notion of the derivative of a function with respect to a measure and some notations. The formulation of the optimal control problem is introduced in Section 3. The variational equations, the adjoint equations and the estimates of their solutions are also given in this section. Section 4 is devoted to the introduction of the first important conclusion of this paper – the second-order expansion of cost functional. The second important conclusion – SMP is proved in Section 5. An illustrative application to mean-field game is considered in Section 6. In the last section some necessary notations and the proof of the auxiliary lemma are shown for closing our paper.

2. Preliminaries

2.1. Derivative of function $h:\mathcal{P}(\mathbb{R}^2)\to\mathbb{R}$

Let $\mathcal{P}(\mathbb{R}^d)$ be the set of all Borel probability measures on $\mathbb{R}^d$. For $1\leq p<+\infty$, let $\mathcal{P}_p(\mathbb{R}^d)$ be the subspace of $\mathcal{P}(\mathbb{R}^d)$ of probability measures having a finite moment of order p over $(\mathbb{R}^d,\mathscr{B}(\mathbb{R}^d))$, and moreover, we endow the space $\mathcal{P}_p(\mathbb{R}^d)$ with the p-Warsserstein metric: for $\nu_1,\nu_2\in\mathcal{P}_p(\mathbb{R}^d)$,

$$
W_p(\nu_1,\nu_2)=\inf\left\{\left[\int_{\mathbb{R}^{2d}}|x-y|^p\varrho(\mathrm{d}x,\mathrm{d}y)\right]^{\frac{1}{p}}:\varrho\in\mathcal{P}_p(\mathbb{R}^{2d}),\ \varrho(\cdot,\mathbb{R}^d)=\nu_1,\ \varrho(\mathbb{R}^d,\cdot)=\nu_2\right\}.
$$

We now recall the derivative of a function h defined on $\mathcal{P}_2(\mathbb{R}^d)$ with respect to a measure, see Cardaliaguet [8], or Buckdahn *et al.* [7] for more details. We call the function $h:\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$ is differentiable in $\nu_0\in\mathcal{P}_2(\mathbb{R}^d)$, if there exists a $\eta_0\in L^2(\mathcal{F};\mathbb{R}^d)$ with $\nu_0=P_{\eta_0}$, such that the lifted function $\bar{h}:L^2(\mathcal{F};\mathbb{R}^d)\to\mathbb{R}$ defined by $\bar{h}(\eta):=h(P_\eta)$ is differentiable at η_0 in Fréchet sense. In other words, there exists a continuous linear functional $D\bar{h}(\eta_0):L^2(\mathcal{F};\mathbb{R}^d)\to\mathbb{R}$, such that for $\eta\in L^2(\mathcal{F};\mathbb{R}^d)$,

$$\bar{h}(\eta_0+\eta)-\bar{h}(\eta_0)=D\bar{h}(\eta_0)(\eta)+o(||\eta||_{L^2}), \tag{2.1}$$

with $||\eta||_{L^2}\to 0$. From Riesz representation theorem and the argument of Cardaliaguet [8], it follows that there exists a Borel measurable function $g:\mathbb{R}^d\to\mathbb{R}^d$ depending only the law of η_0, but not the random variable η_0 itself, such that (2.1) can read as

$$h(P_{\eta_0+\eta})-h(P_{\eta_0})=<g(\eta_0),\eta>+o(||\eta||_{L^2}), \tag{2.2}$$

where $<\cdot,\cdot>$ denotes the "dual product" on $L^2(\mathcal{F};\mathbb{R}^d)$. From (2.2) we can define $\partial_\nu h(P_{\eta_0};a):=g(a), a\in\mathbb{R}^d$, which is called the derivative of h at P_{η_0}. It should be pointed out that the function $\partial_\nu h(P_{\eta_0};a)$ is only $P_{\eta_0}(da)$-a.e. uniquely determined. In our case, for simplicity we just consider those functions $h:\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$ being differentiable in all elements of $\mathcal{P}_2(\mathbb{R}^d)$.

The following spaces have been introduced in [7], [15], [21], [5]. Here we borrow them. We denote

- $C_b^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$ to be all continuously differentiable function h over $\mathcal{P}_2(\mathbb{R}^d)$ with Lipschitz-continuous bounded derivative, *i.e.*, there exists a positive constant C such that,
  (i) $|\partial_\nu h(\nu;a)|\le C,\quad \forall a\in\mathbb{R}^d,\ \nu\in\mathcal{P}_2(\mathbb{R}^d)$;

  (ii) $|\partial_\nu h(\nu_1;a_1)-\partial_\nu h(\nu_2;a_2)|\le C\Big(W_2(\nu_1,\nu_2)+|a_1-a_2|\Big),\ \nu_1,\nu_2\in\mathcal{P}_2(\mathbb{R}^d),\ a_1,a_2\in\mathbb{R}^d.$
- $C_b^{2,1}(\mathcal{P}_2(\mathbb{R}^d))$ to be all measurable function $h\in C_b^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$ satisfying:
  (i) for all $a\in\mathbb{R}^d$, $(\partial_\nu h)_\ell(\cdot;a)\in C_b^{1,1}(\mathcal{P}_2(\mathbb{R}^d)),\ \ell=1,2,\cdot,\cdot,\cdot,d$;

  (ii) for each $\nu\in\mathcal{P}_2(\mathbb{R}^d)$, $\partial_\nu h(\nu;\cdot)$ is differentiable;

  (iii) the second-order derivatives $\partial_a\partial_\nu h:\mathcal{P}_2(\mathbb{R}^d)\times\mathbb{R}^d\to\mathbb{R}^d\otimes\mathbb{R}^d$ and $\partial_\nu^2 h(P_{\nu_0};a,b):=\partial_\nu(\partial_\nu h(\cdot;a))(P_{\nu_0};b):\mathcal{P}_2(\mathbb{R}^d)\times\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^d\otimes\mathbb{R}^d$ are bounded and Lipschitz continuous.

Finally, we use an example to give a more intuitive feeling of Lions' derivative.

Example 2.1. Suppose $\varphi:\mathbb{R}\to\mathbb{R}$ and $\psi:\mathbb{R}\to\mathbb{R}$ are two twice continuously differentiable functions with bounded derivatives, and let $\bar{h}(\xi):=h(P_\xi):=\psi(E[\varphi(\xi)]),\ \xi\in L^2(\mathcal{F};\mathbb{R})$. For given $\xi_0\in L^2(\mathcal{F};\mathbb{R})$, $\bar{h}$ is Fréchet differentiable in ξ_0 and

$$\begin{aligned}\bar{h}(\eta+\xi_0)-\bar{h}(\xi_0)=\psi(E[\varphi(\eta+\xi_0)])-\psi(E[\varphi(\xi_0)])&=\int_0^1\frac{\mathrm{d}}{\mathrm{d}\lambda}\psi(E[\varphi(\lambda\eta+\xi_0)])\mathrm{d}\lambda\\&=\int_0^1\psi'(E[\varphi(\lambda\eta+\xi_0)])E[\varphi'(\lambda\eta+\xi_0)\eta]\mathrm{d}\lambda\\&=\psi'(E[\varphi(\xi_0)])E[\varphi'(\xi_0)\eta]+o(||\eta||_{L^2})\\&=E[\psi'(E[\varphi(\xi_0)])\varphi'(\xi_0)\eta]+o(||\eta||_{L^2}).\end{aligned}$$

According to Lions' work, $h_\mu(P_{\xi_0};a)=\psi'(E[\varphi(\xi_0)])\varphi'(a),\quad a\in\mathbb{R}$, and moreover, $h_{\mu a}(P_{\xi_0};a)=\psi'(E[\varphi(\xi_0)])\varphi''(a),\ a\in\mathbb{R}$. Hence,

i) if $\varphi(x)=x$, then $\bar{h}(\xi_0)=h(P_{\xi_0})=\psi(E[\xi_0])$ and $h_\mu(P_{\xi_0};a)=\psi'(E[\xi_0]),\ \xi_0\in L^2(\mathcal{F};P)$;

ii) if $\psi(x)=x$, then $\bar{h}(\xi_0)=h(P_{\xi_0})=E[\varphi(\xi_0)]$ and $h_\mu(P_{\xi_0};a)=\varphi'(a),\ a\in\mathbb{R}$;
iii) if $\psi(x)=\varphi(x)=x$, then $\bar{h}(\xi_0)=h(P_{\xi_0})=E[\xi_0]$ and $h_\mu(P_{\xi_0};a)=1$.

2.2. Function spaces

Let T be a fixed strictly positive real number and $(\Omega,\mathscr{F},\{\mathscr{F}_t\}_{0\le t\le T},P)$ be a complete filtrated probability space on which a one-dimensional standard Brownian motion $\{W(t),0\le t\le T\}$ is defined. Denote by $\mathscr{P}$ the $\mathscr{F}_t$-predictable σ-field on $[0,T]\times\Omega$ and by $\mathscr{B}(\Lambda)$ the Borel σ-algebra of any topological space Λ. Let $(G,\mathscr{B}(G),\lambda)$ be a measurable space with $\lambda(G)<\infty$ and $\int_G(1\wedge|e|^2)\lambda(\mathrm{d}e)<+\infty$. Let $q:\Omega\times D_q\longrightarrow Z$ be an $\mathscr{F}_t$-adapted stationary Poisson point process with characteristic measure λ, where D_q is a countable subset of $(0,\infty)$. Then the counting measure induced by q is

$$N_q((0,t]\times A):=\#\{s\in D_q;s\le t,q(s)\in A\},\quad for\quad t>0,A\in\mathscr{B}(G).$$

Let

$$N_{\lambda,q}(\mathrm{d}e,\mathrm{d}t):=N_q(\mathrm{d}e,\mathrm{d}t)-\lambda(\mathrm{d}e)\mathrm{d}t \tag{2.3}$$

be a compensated Poisson random martingale measure which is assumed to be independent of Brownian motion $\{W(t),0\le t\le T\}$. In what follows, when no confusion, we always omit the subscript q, and write (2.3) as

$$N_\lambda(\mathrm{d}e,\mathrm{d}t)=N(\mathrm{d}e,\mathrm{d}t)-\lambda(\mathrm{d}e)\mathrm{d}t. \tag{2.4}$$

Assume $\mathbb{F}=\{\mathscr{F}_t\}_{0\le t\le T}$ is P-completed filtration generated by $\{W(t),0\le t\le T\}$ and $\{\iint_{(0,t]\times A}N_\lambda(\mathrm{d}e,\mathrm{d}t),0\le t\le T,A\in\mathscr{B}(G)\}$, and moreover, augmented by a σ-field $\mathscr{F}^o$ with the following property:

(i) the Brownian motion W and the Poisson random measure N_λ are independent of $\mathscr{F}^o$;
(ii) $\mathscr{F}^o$ is "rich enough", *i.e.*, $\mathcal{P}_2(\mathbb{R}^d)=\{P_\eta,\ \eta\in L^2(\mathscr{F}^o;\mathbb{R}^d)\}$;
(iii) $\mathscr{F}^o$ contains the family of all the P-null subsets $\mathcal{N}_P$.

The following several spaces are used frequently.

- By $L^p(\mathscr{F};\mathbb{R}^d)$ we denote the collection of $\mathbb{R}^d$-valued, $\mathscr{F}$-measurable random variables η with $||\eta||_p:=\mathbb{E}[|\eta|^p]^{\frac{1}{p}}<+\infty$.
- By $\mathcal{S}^2_{\mathbb{F}}(0,T;\mathbb{R}^d)$ we denote the space of $\mathbb{R}^d$-valued, $\mathbb{F}$-predictable process φ on $[0,T]$ with $\mathbb{E}[\sup_{0\le s\le T}|\varphi(t)|^2]<+\infty$.
- By $\mathcal{H}^2_{\mathbb{F}}(0,T;\mathbb{R}^d)$ we denote the space of all $\mathbb{R}^d$-valued, $\mathbb{F}$-adapted càdlàg process φ on $[0,T]$, such that $\mathbb{E}[\int_0^T|\varphi(t)|^2\mathrm{d}t]<+\infty$.
- By $\mathcal{K}^2_\lambda(0,T;\mathbb{R}^d)$ we denote the space of all $\mathbb{R}^d$-valued, $\mathscr{P}\times\mathscr{B}(G)$-measure process r on $[0,T]\times G$ satisfying $E[\int_0^T\int_G|r(t,e)|^2\lambda(\mathrm{d}e)\mathrm{d}t]<+\infty$.

Note that we denote $L^2(G,\mathscr{B}(G),\lambda;\mathbb{R}),\mathcal{S}^2_{\mathbb{F}}(0,T;\mathbb{R}),\mathcal{H}^2_{\mathbb{F}}(0,T;\mathbb{R}),\mathcal{K}^2_\lambda(0,T;\mathbb{R})$ by $L^2(G,\mathscr{B}(G),\lambda)$, $\mathcal{S}^2_{\mathbb{F}}(0,T)$, $\mathcal{H}^2_{\mathbb{F}}(0,T)$, $\mathcal{K}^2_\lambda(0,T)$, respectively, for short.

3. Problem formulation, variational equations and adjoint equations

3.1. Problem formulation

Let us first formulate the optimal control problem. Let U be a subset of $\mathbb{R}$. $v(\cdot):[0,T]\times\Omega\to U$ is called an admissible control if $v(\cdot)$ is $\mathcal{F}_t$-progressive measurable process with $\sup\limits_{0\le t\le T}\mathbb{E}[|v(t)|^8]<\infty$. By $\mathcal{U}_{ad}$ we denote the set of all admissible controls. Let the mappings

$$
\begin{aligned}
&b:[0,T]\times\mathbb{R}\times\mathcal{P}_2(\mathbb{R})\times U\to\mathbb{R}, \quad \sigma:[0,T]\times\mathbb{R}\times\mathcal{P}_2(\mathbb{R})\times U\to\mathbb{R},\\
&\beta:[0,T]\times\mathbb{R}\times\mathcal{P}_2(\mathbb{R})\times G\to\mathbb{R}, \quad \phi:\mathbb{R}\times\mathcal{P}_2(\mathbb{R})\to\mathbb{R},\\
&f:[0,T]\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}\times\mathcal{P}_2(\mathbb{R}^2)\times U\to\mathbb{R},
\end{aligned}
$$

satisfy:

Assumption 3.1. There exists a constant $L>0$ such that, for $x,y,z\in\mathbb{R}$, $\nu\in\mathcal{P}_2(\mathbb{R})$, $\mu\in\mathcal{P}_2(\mathbb{R}^2)$, $u\in U$, $e\in G$, and for $h=b,\sigma$,

$$
\begin{aligned}
|h(t,x,\nu,u)|&\le L\left(1+|x|+\left\{\int_{\mathbb{R}}|a|^2\nu(da)\right\}^{\frac12}+|u|\right),\\
|f(t,x,y,z,\mu,u)|&\le L\left(1+|x|+|y|+|z|+\left\{\int_{\mathbb{R}^2}|a|^2\mu(da)\right\}^{\frac12}+|u|\right),\\
|\phi(x,\nu)|&\le L\left(1+|x|+\left\{\int_{\mathbb{R}}|a|^2\nu(da)\right\}^{\frac12}\right),\\
|\beta(t,x,\nu,u,e)|&\le L(1\wedge|e|)\left(1+|x|+\left\{\int_{\mathbb{R}}|a|^2\nu(da)\right\}^{\frac12}+|u|\right),
\end{aligned}
$$

and b,σ,β,f,ϕ are continuous in (t,u);

Assumption 3.2. For $t\in[0,T],u\in U,e\in G$, $(b,\sigma)(t,\cdot,\cdot,u)\in C_b^{1,1}(\mathbb{R}\times\mathcal{P}_2(\mathbb{R}))$, $\beta(t,\cdot,\cdot,e)\in C_b^{1,1}(\mathbb{R}\times\mathcal{P}_2(\mathbb{R}))$, $f(t,\cdot,\cdot,\cdot,\cdot,\cdot,u)\in C_b^{1,1}(\mathbb{R}^4\times\mathcal{P}_2(\mathbb{R}^2))$, $\phi(\cdot,\cdot)\in C_b^{1,1}(\mathbb{R}\times\mathcal{P}(\mathbb{R}))$, *i.e.*,

- (i) For $(t,x,y,z,k,u,e)\in[0,T]\times\mathbb{R}^4\times U\times G$, $(b,\sigma)(t,x,\cdot,u)\in C_b^{1,1}(\mathcal{P}_2(\mathbb{R}))$, $\beta(t,x,\cdot,e)\in C_b^{1,1}(\mathcal{P}_2(\mathbb{R}))$, $f(t,x,y,z,k,\cdot,u)\in C_b^{1,1}(\mathcal{P}_2(\mathbb{R}^2))$, $\phi(x,\cdot)\in C_b^{1,1}(\mathcal{P}_2(\mathbb{R}))$;
- (ii) For $(t,\nu,\mu,u,e)\in[0,T]\times\mathcal{P}_2(\mathbb{R})\times\mathcal{P}_2(\mathbb{R}^2)\times U\times G$, $(b,\sigma)(t,\cdot,\nu,u)\in C_b^1(\mathbb{R})$, $\beta(t,\cdot,\nu,e)\in C_b^1(\mathbb{R})$, $f(t,\cdot,\cdot,\cdot,\cdot,\mu,u)\in C_b^1(\mathbb{R}^4)$, $\phi(\cdot,\nu)\in C_b^1(\mathbb{R})$;
- (iii) All the first-order derivatives $\partial_\ell\psi$, $\psi=b,\sigma,f,\phi$, $\ell=x,y,z,\nu,\mu$ are bounded and Lipschitz continuous in (x,y,z,ν,μ) with the constant independent of $u\in U$, and continuous in u; for each $e\in G$, $\partial_x\beta$, $\partial_\nu\beta$ are bounded by $C(1\wedge|e|)$ and Lipschitz continuous with the constant C independent of $u\in U$ and $e\in G$;

  and, furthermore,

Assumption 3.3. Let b,σ,β,f,ϕ satisfy Assumptions (3.1)–(3.2), and, meanwhile, for $t\in[0,T],u\in U,e\in G$, $(b,\sigma)(t,\cdot,\cdot,u)\in C_b^{2,1}(\mathbb{R}\times\mathcal{P}_2(\mathbb{R}))$, $\beta(t,\cdot,\cdot,e)\in C_b^{2,1}(\mathbb{R}\times\mathcal{P}_2(\mathbb{R}))$, $f(t,\cdot,\cdot,\cdot,\cdot,\cdot,u)\in C_b^{2,1}(\mathbb{R}^4\times\mathcal{P}_2(\mathbb{R}^2))$, $\phi(\cdot,\cdot)\in C_b^{2,1}(\mathbb{R}\times\mathcal{P}(\mathbb{R}))$, *i.e.*, the derivatives of b,σ,β,f,ϕ enjoy the following properties:

- (i) For $(t,u,e)\in[0,T]\times U\times G$, $(\partial_x b,\partial_x\sigma)(t,\cdot,\cdot,u)\in C_b^{1,1}(\mathbb{R}\times\mathcal{P}_2(\mathbb{R}))$, $\partial_x\beta(t,\cdot,\cdot,e)\in C_b^{1,1}(\mathbb{R}\times\mathcal{P}_2(\mathbb{R}))$, $\partial_\ell f(t,\cdot,\cdot,\cdot,\cdot,\cdot,u)\in C_b^{1,1}(\mathbb{R}^4\times\mathcal{P}_2(\mathbb{R}^2))$, $\ell=x,y,z,k$, $\partial_x\phi(\cdot,\cdot)\in C_b^{1,1}(\mathbb{R}\times\mathcal{P}(\mathbb{R}))$;
- (ii) For $(t,u,e)\in[0,T]\times U\times G$, $(\partial_\nu b,\partial_\nu\sigma)(t,\cdot,\cdot,u;\cdot)\in C_b^{1,1}(\mathbb{R}\times\mathcal{P}_2(\mathbb{R})\times\mathbb{R})$, $\partial_\nu\beta(t,\cdot,\cdot,e;\cdot)\in C_b^{1,1}(\mathbb{R}\times\mathcal{P}_2(\mathbb{R})\times\mathbb{R})$, $(\partial_\mu f)_j(t,\cdot,\cdot,\cdot,\cdot,\cdot,u;\cdot,\cdot)\in C_b^{1,1}(\mathbb{R}^4\times\mathcal{P}_2(\mathbb{R}^2)\times\mathbb{R}^2)$, $j=1,2$, $\partial_\nu\phi(\cdot,\cdot;\cdot)\in C_b^{1,1}(\mathbb{R}\times\mathcal{P}(\mathbb{R})\times\mathbb{R})$;
- (iii) All the second-order derivatives of b,σ,f,ϕ are bounded and Lipschitz continuous with the Lipschitz constants independent of $u\in U$ and continuous in u; for each $e\in G$, all the second-order derivatives of β are bounded by $C(1\wedge|e|)$, and Lipschitz continuous with the constant C independent of $(e,u)\in G\times U$.

Remark 3.4. Let us explain Assumption (3.2) and Assumption (3.3) by a simple example. Consider a function $h(x,\nu)=\phi(x)+\psi\Big(\int_{\mathbb{R}}\varphi(x)\nu(\mathrm{d}x)\Big)$, $x\in\mathbb{R}$, $\nu\in\mathcal{P}_2(\mathbb{R})$, where $\phi,\varphi,\psi:\mathbb{R}\to\mathbb{R}$ are continuously differentiable functions. As we know, in the classical case (without mean-field) (see [17], or [26]), the assumption that ϕ is a twice continuously differentiable function with bounded derivative is required in order to study the SMP. Similarly, in the mean-field case the second-order differentiability of h in (x,ν) are also needed.

Since in our setting $\mathscr{F}_0$ is rich enough, *i.e.*, $\mathcal{P}_2(\mathbb{R})=\{P_\eta,\ \eta\in L^2(\mathscr{F}^o;\mathbb{R})\}$, for any $\nu\in\mathcal{P}_2(\mathbb{R})$, we can find a random variable $\xi\in L^2(\Omega,\mathcal{F}_0,P)$ such that $\nu=P_\xi$. Consequently, $h(x,\nu)$ can be written as $h(x,P_\xi)=\phi(x)+\psi E[\varphi(\xi)]$. Assumption (3.2) means that h is once continuously differentiable in (x,P_ξ) with bounded derivatives. Recall Example 2.1, it is not difficult to see Assumption (3.2) can be satisfied if we assume φ,ϕ,ψ are once continuously differentiable functions with bounded derivative. Assumption (3.3) implies that h is a twice continuously differentiable funtion in (x,P_ξ) with bounded derivatives. If the functions φ,ϕ,ψ are twice continuously differentiable with bounded derivative, one can check that Assumption (3.3) holds true.

For $v(\cdot)\in\mathcal{U}_{ad}$, under the Assumptions (3.1)–(3.2), the equation (1.1) possesses a unique solution (X^v,Y^v,Z^v,K^v).

The target of the optimal control problem consists in minimizing $J(v(\cdot))=Y^v(0)$ over $\mathcal{U}_{ad}$, *i.e.*, find an admissible control $u^*(\cdot)\in\mathcal{U}_{ad}$ such that

$$J(u^*(\cdot))=\inf{}_{v\in\mathcal{U}_{ad}}J(v(\cdot)). \tag{3.1}$$

The main purpose of this paper is to study the SMP of optimal control problem (1.1) and (3.1).

Remark 3.5. Throughout this paper, we set $\rho:(0,+\infty)\to(0,+\infty)$ is a function with the property $\rho(\varepsilon)\to0$, as $\varepsilon\to0$, and C is a positive constant, both of which maybe change from one appearance to another.

3.2. Variational equations

This subsection is devoted to the introduction of the first- and second-order variational equations, as well as some estimates of their solutions.

Now let $u^*(\cdot)\in\mathcal{U}_{ad}$ be an optimal control, and by $(X^*(\cdot),Y^*(\cdot),Z^*(\cdot),K^*(\cdot))=(X^{u^*}(\cdot),Y^{u^*}(\cdot),Z^{u^*}(\cdot),K^{u^*}(\cdot))$, the solution of (1.1) with $u^*(\cdot)$ instead of $v(\cdot)$, we denote the optimal state process. It is clear from the definition of a function with respect to a measure, that when studying the first- and second-order derivatives of coefficients with respect to a measure, some auxiliary probability spaces are needed. Hence we would like to introduce first an intermediate probability space and the stochastic processes defined on it as a representative. The other probability spaces and corresponding stochastic processes can be understood in the same sense. For this end, let $(\overline{\Omega},\overline{\mathscr{F}},\overline{P})$ be an intermediate complete probability space, which is independent of $(\Omega,\mathscr{F},P)$. The pair $(\overline{W},\overline{N_\lambda})$ defined on space $(\overline{\Omega},\overline{\mathscr{F}},\overline{P})$ is an independent copy of (W,N_λ), *i.e.*, $(\overline{W},\overline{N_\lambda})$ under $\overline{P}$ has the same law as (W,N_λ) under P. By $\overline{X}^v(\cdot)$ we denote the corresponding state trajectory but driven by $(\overline{W},\overline{N_\lambda})$ instead of (W,N_λ) in (1.1). $\overline{\mathbb{E}}[\cdot]$ only acts on the random variables or/and the stochastic processes with "bar". $(\widetilde{\Omega},\widetilde{\mathscr{F}},\widetilde{P},\widetilde{X}(t),\widetilde{\mathbb{E}}[\cdot])$ and $(\widehat{\Omega},\widehat{\mathscr{F}},\widehat{P},\widehat{X}(t),\widehat{\mathbb{E}}[\cdot])$ can be understood in the same meaning. Note that $(\Omega,\mathscr{F},P),(\widehat{\Omega},\widehat{\mathscr{F}},\widehat{P})$, $(\widetilde{\Omega},\widetilde{\mathscr{F}},\widetilde{P})$ and $(\overline{\Omega},\overline{\mathscr{F}},\overline{P})$ are also independent.

Let $v(\cdot)$ be any given admissible control. For $\phi=b,\sigma,b_x,\sigma_x$ and $\psi=b,\sigma$, define

$$\begin{aligned}
&\delta\phi(t):=\phi(t,X^*(t),P_{X^*(t)},v(t))-\phi(t,X^*(t),P_{X^*(t)},u^*(t)),\\
&(\psi_x,\psi_{xx})(t):=(\frac{\partial\psi}{\partial x},\frac{\partial^2\psi}{\partial x^2})(t,X^*(t),P_{X^*(t)},u^*(t)),\\
&(\psi_\nu,\psi_{\nu a})(t;\widetilde{X}^*(t)):=(\frac{\partial\psi}{\partial\nu},\frac{\partial^2\psi}{\partial\nu\partial a})(t,X^*(t),P_{X^*(t)},u^*(t);\widetilde{X}^*(t)),
\end{aligned}$$

$$
\begin{aligned}
&(\widetilde{\psi}_\nu,\widetilde{\psi}_{\nu a})(t)=(\frac{\partial\psi}{\partial\nu},\frac{\partial^2\psi}{\partial\nu\partial a})(t,\widetilde{X}^*(t),P_{X^*(t)},\widetilde{u}^*(t);X^*(t)),\\
&\psi_{\nu\nu}(t;\widehat{\widetilde{X}}^*(t)):=\frac{\partial^2\psi}{\partial\nu^2}(t,X^*(t),P_{X^*(t)},u^*(t);\widetilde{X}^*(t),\widehat{X}^*(t)).
\end{aligned}
$$

Let $\varepsilon>0$, and $E_\varepsilon\subset[0,T]$ be a Borel set with Borel measure $|E_\varepsilon|=\varepsilon$. For any $v(\cdot)\in\mathcal{U}_{ad}$, we consider the "spike variation" of the optimal control $u^*(\cdot)$: $u^\varepsilon(t):=u^*(t)\mathbb{1}_{(E_\varepsilon)^c}+v(t)\mathbb{1}_{E_\varepsilon}$, and let $(X^\varepsilon,Y^\varepsilon,Z^\varepsilon,K^\varepsilon):=(X^{u^\varepsilon},Y^{u^\varepsilon},Z^{u^\varepsilon},K^{u^\varepsilon})$ be the solution of (1.1) under the control $u^\varepsilon(\cdot)$. Inspired by Peng [26], when the control is involved in the diffusion term and the control domain is not convex, for each $\varepsilon>0$, one can find two processes $X^{1,\varepsilon}$ and $X^{2,\varepsilon}$, such that $X^\varepsilon-X^*-X^{1,\varepsilon}=O(\varepsilon)$, and $X^\varepsilon-X^*-X^{1,\varepsilon}-X^{2,\varepsilon}=o(\varepsilon)$, where the convergence are both in $L^2(\Omega,C[0,T])$ sense. In our case it is easy to check that the first- and second-order variational equations $X^{1,\varepsilon}$ and $X^{2,\varepsilon}$ satisfy

$$
\begin{cases}
\mathrm{d}X^{1,\varepsilon}(t)=\Big\{b_x(t)X^{1,\varepsilon}(t)+\widetilde{\mathbb{E}}[b_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]+\delta b(t)\mathbb{1}_{E_\varepsilon}(t)\Big\}\mathrm{d}t\\
\qquad\qquad+\Big\{\sigma_x(t)X^{1,\varepsilon}(t)+\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]+\delta\sigma(t)\mathbb{1}_{E_\varepsilon}(t)\Big\}\mathrm{d}W(t)\\
\qquad\qquad+\int_G\Big\{\beta_x^-(t,e)X^{1,\varepsilon}(t-)+\widetilde{\mathbb{E}}[\beta_\nu^-(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t-)]\Big\}N_\lambda(\mathrm{d}e,\mathrm{d}t),\ t\in[0,T],\\
X^{1,\varepsilon}(0)=0,
\end{cases}
\tag{3.2}
$$

and

$$
\begin{cases}
\mathrm{d}X^{2,\varepsilon}(t)=\Big\{b_x(t)X^{2,\varepsilon}(t)+\widetilde{\mathbb{E}}[b_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{2,\varepsilon}(t)]+\frac{1}{2}\Big(b_{xx}(t)(X^{1,\varepsilon}(t))^2+\widetilde{\mathbb{E}}[b_{\nu a}(t;\widetilde{X}^*(t))\\
\qquad(\widetilde{X}^{1,\varepsilon}(t))^2]\Big)+\Big(\delta b_x(t)X^{1,\varepsilon}(t)+\widetilde{\mathbb{E}}[\delta b_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]\Big)\mathbb{1}_{E_\varepsilon}(t)\Big\}\mathrm{d}t\\
\quad+\Big\{\sigma_x(t)X^{2,\varepsilon}(t)+\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{2,\varepsilon}(t)]+\frac{1}{2}\Big(\sigma_{xx}(t)(X^{1,\varepsilon}(t))^2+\widetilde{\mathbb{E}}[\sigma_{\nu a}(t;\widetilde{X}^*(t))\\
\qquad(\widetilde{X}^{1,\varepsilon}(t))^2]\Big)+\Big(\delta\sigma_x(t)X^{1,\varepsilon}(t)+\widetilde{\mathbb{E}}[\delta\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]\Big)\mathbb{1}_{E_\varepsilon}(t)\Big\}\mathrm{d}W(t)\\
\quad+\int_G\Big\{\beta_x^-(t,e)X^{2,\varepsilon}(t-)+\widetilde{\mathbb{E}}[\beta_\nu^-(t,e;\widetilde{X}^*(t))\widetilde{X}^{2,\varepsilon}(t-)]\\
\quad+\frac{1}{2}\Big(\beta_{xx}^-(t,e)(X^{1,\varepsilon}(t-))^2+\widetilde{\mathbb{E}}[\beta_{\nu a}^-(t,e;\widetilde{X}^*(t))(\widetilde{X}^{1,\varepsilon}(t-))^2]\Big)\Big\}N_\lambda(\mathrm{d}e,\mathrm{d}t),\quad t\in[0,T],\\
X^{2,\varepsilon}(0)=0.
\end{cases}
\tag{3.3}
$$

Here $(\beta_x^-,\beta_{xx}^-)(t,e)=(\frac{\partial\beta}{\partial x},\frac{\partial^2\beta}{\partial x^2})(t,X^*(t-),P_{X^*(t)},e)$, $(\beta_\nu^-,\beta_{\nu a}^-)(t,e;\widetilde{X}^*(t))=(\frac{\partial\beta}{\partial\nu},\frac{\partial^2\beta}{\partial\nu\partial a})(t,X^*(t-),P_{X^*(t)},e;\widetilde{X}^*(t-))$.

Obviously, under the Assumptions (3.1)–(3.3), the equation (3.2) and the equation (3.3) have unique solutions $\{X^{1,\varepsilon}(t)\}_{t\in[0,T]}$ and $\{X^{2,\varepsilon}(t)\}_{t\in[0,T]}$. Moreover, their solutions satisfy the following estimates:

Proposition 3.6. *Let the Assumptions (3.1)–(3.3) hold true. For $\ell \geq 1$, there exists a constant $L_\ell > 0$ depending only on ℓ such that*

$$\begin{aligned}
&\text{i)}\quad \mathbb{E}\Big[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^{2\ell}\Big] \leq L_\ell\varepsilon^{\ell}, \quad \mathbb{E}\Big[\sup_{t\in[0,T]}|X^{2,\varepsilon}(t)|^{2\ell}\Big] \leq L_\ell\varepsilon^{2\ell};\\
&\text{ii)}\quad \mathbb{E}\Big[\sup_{t\in[0,T]}|X^{\varepsilon}(t)-X^*(t)|^{2\ell}\Big] \leq L_\ell\varepsilon^{\ell};\\
&\text{iii)}\quad \mathbb{E}\Big[\sup_{t\in[0,T]}|X^{\varepsilon}(t)-X^*(t)-X^{1,\varepsilon}(t)|^{2\ell}\Big] \leq L_\ell\varepsilon^{2\ell}.
\end{aligned} \tag{3.4}$$

The proof is similar to Proposition 4.2 in [5]. Hence, we omit it.

Remark 3.7. If β depends on control, we only have for $\ell \geq 1$,

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^{2\ell}\Big] \leq L_\ell\varepsilon. \tag{3.5}$$

In fact, in this case we have to estimate, for $\ell \geq 1$,

$$\mathbb{E}\left[\left(\int_0^T\int_G (1\wedge|e|^2)|X^{1,\varepsilon}(t)|^2\mathbb{1}_{E_\varepsilon}(t)N(\mathrm{d}t,\mathrm{d}e)\right)^{\ell}\right].$$

Similar to the proof of Proposition 3.2 [22], we just have (3.5).
Besides, as we state in introduction, in our framework if the estimate

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^{2\ell}\Big] \leq L_\ell\varepsilon^{\ell},$$

holds true when β depends on control. Then the Kolmogorovs Continuity Criterion would imply the continuity of the jump process $X^{1,\varepsilon}$. This is impossible!

An extra assumption is the need to prove the following lemma.

Assumption 3.8. Let $1+\beta_x(t,e) \geq \delta$, $(t,e)\in[0,T]\times G$, where δ is some given positive constant.

Lemma 3.9. *Let the Assumptions (3.1), (3.2) and (3.8) hold true and let $(\overline{\Omega},\overline{\mathcal{F}},\overline{P})$ be an intermediate probability space and independent of space $(\Omega,\mathcal{F},P)$, and let $(\overline{\psi}_3(t,e))_{(t,e)\in[0,T]\times G}$, $(\overline{\psi}_1(t))_{t\in[0,T]}$ be two progressively measurable stochastic processes defined on the product space $(\Omega\times\overline{\Omega},\mathcal{F}\times\overline{\mathcal{F}},P\otimes\overline{P})$ and $(\overline{\psi}_2(t))_{t\in[0,T]}$ be a progressively measurable stochastic process defined on the space $(\overline{\Omega},\overline{\mathcal{F}},\overline{P})$. Moreover, assume $(\overline{\psi}_i(t))_{t\in[0,T]}$, $i=1,2,3$ satisfies the following properties:*

$$\begin{aligned}
&\text{a) for } e\in G,\ t\in[0,T],\quad |\overline{\psi}_1(t)|\leq C,\quad |\overline{\psi}_3(t,e)|\leq C(1\wedge|e|),\ P\otimes\overline{P}\text{-a.s.},\\
&\text{b) for } \ell\geq 1,\quad \overline{\mathbb{E}}\Big[\sup_{t\in[0,T]}|\overline{\psi}_2(t)|^{2\ell}\Big]\leq C_\ell.
\end{aligned}$$

Then

$$
\begin{aligned}
&\text{i) } \mathbb{E}\Big[\int_0^T \Big|\overline{\mathbb{E}}[\overline{\psi}_1(t)\overline{\psi}_2(t)\overline{X}^{1,\varepsilon}(t)]\Big|^4 \mathrm{d}t\Big] \le \varepsilon^2\rho(\varepsilon),\\
&\text{ii) } \mathbb{E}\Big[\int_0^T \int_G \Big|\overline{\mathbb{E}}[\overline{\psi}_3(t,e)\overline{\psi}_2(t)\overline{X}^{1,\varepsilon}(t)]\Big|^4 \lambda(\mathrm{d}e)\mathrm{d}t\Big] \le \varepsilon^2\rho(\varepsilon).
\end{aligned}
\tag{3.6}
$$

Proof. Under the Assumptions (3.1), (3.2) and (3.8), the proof of i) follows that of Proposition 4.3 [5]. Hence, we mainly estimate ii). The proof of ii) is an adaptation of that for Proposition 4.3 [5]. Let us state it in detail. Denote

$$
\begin{aligned}
S(t) = &\int_0^t \Big(-b_x(s) + \frac{1}{2}|\sigma_x(s)|^2 + \int_G (\beta_x(s,e) - \ln(1+\beta_x(s,e)))\lambda(\mathrm{d}e)\Big)\mathrm{d}s\\
&- \int_0^t \sigma_x(s)\mathrm{d}W(s) - \int_0^t\int_G \ln(1+\beta_x^-(s,e))N_\lambda(\mathrm{d}e,\mathrm{d}s),
\end{aligned}
\tag{3.7}
$$

and consider $m(t) = e^{S(t)}$. Obviously, $(m(s))_{s\in[0,T]}$ satisfies

$$
\begin{cases}
dm(t) = m(t)\Big(-b_x(t) + |\sigma_x(t)|^2 + \displaystyle\int_G \frac{|\beta_x(t,e)|^2}{1+\beta_x(t,e)}\lambda(\mathrm{d}e)\Big)\mathrm{d}t - m(t)\sigma_x(t)\mathrm{d}W(t)\\
\qquad\qquad - \displaystyle\int_G m(t-)\frac{\beta_x^-(t,e)}{1+\beta_x^-(t,e)}N_\lambda(\mathrm{d}e,\mathrm{d}t),\ t\in[0,T],\\
m(0) = 1.
\end{cases}
\tag{3.8}
$$

Let $n(t) = m(t)^{-1} = e^{-S(t)}$. Due to $\beta_x(t,e) \ge \delta - 1 > -1$, $(t,e)\in[0,T]\times G$, the boundness of b_x, σ_x, β_x implies that, for $\ell \ge 1$,

$$
\mathbb{E}\Big[\sup_{t\in[0,T]} \big(|n(t)|^\ell + |m(t)|^\ell\big)\Big] \le C_\ell,
\tag{3.9}
$$

where C_ℓ is a positive constant only depending on ℓ.

On the other hand, it follows from Itô's formula for semi-martingale with jumps (see [28], Thm. 93) that

$$
\begin{aligned}
\mathrm{d}X^{1,\varepsilon}(t)m(t) = &\, m(t)\Big(\widetilde{\mathbb{E}}[\sigma_\nu(t,\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)] + \delta\sigma(t)\mathbb{1}_{E_\varepsilon}(t)\Big)\mathrm{d}W(t)\\
&+ \int_G m(t-)\frac{1}{1+\beta_x^-(t,e)}\widetilde{\mathbb{E}}[\beta_\nu^-(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t-)]N_\lambda(\mathrm{d}t,\mathrm{d}e)\\
&+ \Big\{m(t)\big(\widetilde{\mathbb{E}}[b_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)] + \delta b(t)\mathbb{1}_{E_\varepsilon}(t)\big)\\
&- m(t)\sigma_x(t)\big(\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)] + \delta\sigma(t)\mathbb{1}_{E_\varepsilon}(t)\big)\\
&- \int_G m(t)\frac{\beta_x(t,e)}{1+\beta_x(t,e)}\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]\lambda(\mathrm{d}e)\Big\}\mathrm{d}t.
\end{aligned}
\tag{3.10}
$$

Hence,

$$
X^{1,\varepsilon}(t) = n(t)\Theta_1^\varepsilon(t) + n(t)\Theta_2^\varepsilon(t) + \Theta_3^\varepsilon(t),
\tag{3.11}
$$

where

$$
\begin{aligned}
\Theta_1^\varepsilon(t) &:= \int_0^t m(s)\Big(\widetilde{\mathbb{E}}[\sigma_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)] + \delta\sigma(s)\mathbb{1}_{E_\varepsilon}(s)\Big)\mathrm{d}W(s),\\
\Theta_2^\varepsilon(t) &:= \int_0^t\int_G m(s-)\frac{1}{1+\beta_x^-(s,e)}\widetilde{\mathbb{E}}[\beta_\nu^-(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s-)]N_\lambda(\mathrm{d}s,\mathrm{d}e),\\
\Theta_3^\varepsilon(t) &:= n(t)\int_0^t \Big\{m(s)\big(\widetilde{\mathbb{E}}[b_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)] + \delta b(s)\mathbb{1}_{E_\varepsilon}(s)\big)\\
&\quad - m(s)\sigma_x(s)\big(\widetilde{\mathbb{E}}[\sigma_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)] + \delta\sigma(s)\mathbb{1}_{E_\varepsilon}(s)\big)\\
&\quad - \int_G m(s)\frac{\beta_x(s,e)}{1+\beta_x(s,e)}\widetilde{\mathbb{E}}[\beta_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]\lambda(\mathrm{d}e)\Big\}\mathrm{d}s.
\end{aligned}
\tag{3.12}
$$

We are now ready to calculate $\overline{\mathbb{E}}[\overline{\psi}_3(t,e')\overline{\psi}_2(t)\overline{X}^{1,\varepsilon}(t)]$ with the help of (3.11) and (3.12). For each given $e'\in G$,

$$
\begin{aligned}
\overline{\mathbb{E}}[\overline{\psi}_3(t,e')\overline{\psi}_2(t)\overline{X}^{1,\varepsilon}(t)] &= \overline{\mathbb{E}}[\overline{\psi}_3(t,e')\overline{\psi}_2(t)\overline{n}(t)\overline{\Theta}_1^\varepsilon(t)] + \overline{\mathbb{E}}[\overline{\psi}_3(t,e')\overline{\psi}_2(t)\overline{n}(t)\overline{\Theta}_2^\varepsilon(t)]\\
&\quad + \overline{\mathbb{E}}[\overline{\psi}_3(t,e')\overline{\psi}_2(t)\overline{\Theta}_3^\varepsilon(t)] := \Xi_1^\varepsilon(t,e') + \Xi_2^\varepsilon(t,e') + \Xi_3^\varepsilon(t,e').
\end{aligned}
\tag{3.13}
$$

Since, for $\ell\geq 1$,

$$
\begin{aligned}
&\overline{\mathbb{E}}\Big[\sup_{t\in[0,T]}|\overline{\psi}_3(t,e')\overline{\psi}_2(t)\overline{n}(t)|^\ell\Big] \leq CE\Big[\sup_{t\in[0,T]}|\overline{\psi}_2(t)\overline{n}(t)|^\ell\Big]\\
&\leq C\Big\{\mathbb{E}[\sup_{t\in[0,T]}|\overline{\psi}_2(t)|^{2\ell}]\Big\}^{\frac12}\cdot\Big\{\mathbb{E}[\sup_{t\in[0,T]}|\overline{n}(t)|^{2\ell}]\Big\}^{\frac12} \leq C_\ell,
\end{aligned}
\tag{3.14}
$$

where C_ℓ does not depends on e' because of $|\psi_3(t,e')|\leq C(1\wedge|e'|)\leq C$, and observe that $\mathbb{F}=\mathbb{F}^W\bigvee\mathbb{F}^N$, according to the martingale representation theorem for jump process (see [29]), we have, for each $t\in[0,T],\ e'\in G$, there exists a unique pair $(\overline{\theta}_{\cdot,t,e'},\overline{\gamma}_{\cdot,t,e'})\in\mathcal{H}^2_{\overline{\mathbb{F}}}(0,t)\times\overline{K}^2_\lambda(0,t)$, such that $\overline{P}$-a.s.,

$$
\overline{\psi}_3(t,e')\overline{\psi}_2(t)\overline{n}(t) = \mathbb{E}[\overline{\psi}_3(t,e')\overline{\psi}_2(t)\overline{n}(t)] + \int_0^t\overline{\theta}_{s,t,e'}d\overline{W}(s) + \int_0^t\overline{\gamma}_{s,t,e'}(e)\overline{N}_\lambda(\mathrm{d}e,\mathrm{d}s).
\tag{3.15}
$$

We argue that, for $\ell\geq 1$ and for each $e'\in G$, there exists a constant $C_\ell>0$ depending on ℓ, but independent of e', such that

$$
\overline{\mathbb{E}}\Bigg[\Big(\int_0^t|\overline{\theta}_{s,t,e'}|^2\mathrm{d}s\Big)^{\frac{\ell}{2}} + \Big(\int_0^t\int_G|\overline{\gamma}_{s,t,e'}(e)|^2\lambda(\mathrm{d}e)\mathrm{d}s\Big)^{\frac{\ell}{2}}\Bigg] \leq C_\ell.
\tag{3.16}
$$

Indeed, for $\ell \geq 2$, from Burkholder-Davis-Gundy, Doob's maximal inequality and Hölder inequality, we have, for $t \in [0,T]$,

$$\begin{aligned}
&\overline{\mathbb{E}}\Big[\Big(\int_0^t |\overline{\theta}_{s,t,e'}|^2 \mathrm{d}s + \int_0^t \int_G |\overline{\gamma}_{s,t,e'}(e)|^2 \overline{N}(\mathrm{d}e,\mathrm{d}s)\Big)^{\frac{\ell}{2}}\Big] \\
&\leq C_\ell \overline{\mathbb{E}}\Big[\sup_{s\in[0,t]}\Big|\int_0^s \overline{\theta}_{\tau,s,e'} d\overline{W}(\tau) + \int_0^s \int_G \overline{\gamma}_{\tau,s,e'}(e)\overline{N}_\lambda(\mathrm{d}e,d\tau)\Big|^\ell\Big] \\
&\leq C_\ell(\frac{\ell}{\ell-1})^\ell \overline{\mathbb{E}}\Big[|\int_0^t \overline{\theta}_{s,t,e'} d\overline{W}(s) + \int_0^t\int_G \overline{\gamma}_{s,t,e'}(e)\overline{N}_\lambda(\mathrm{d}e,\mathrm{d}s)|^\ell\Big] \\
&\leq C_\ell\Big\{\overline{\mathbb{E}}\Big[|\overline{\psi}_3(t,e')\overline{\psi}_2(t)\overline{n}(t)|^\ell\Big] + |\mathbb{E}[\overline{\psi}_3(t,e')\overline{\psi}_2(t)\overline{n}(t)]|^\ell\Big\} \leq C_\ell,
\end{aligned} \tag{3.17}$$

where C_ℓ is independent of e' because of (3.14).
Clearly, (3.17) implies, for each $e' \in G$,

$$\overline{\mathbb{E}}\Big[\Big(\int_0^t \int_G |\overline{\gamma}_{s,t,e'}(e)|^2 \overline{N}(\mathrm{d}e,\mathrm{d}s)\Big)^{\frac{\ell}{2}}\Big] \leq C_\ell.$$

Recall Lemma 3.1 [23], it follows

$$\overline{\mathbb{E}}\left[\left(\int_0^t \int_G |\overline{\gamma}_{s,t,e'}(e)|^2 \lambda(\mathrm{d}e)\mathrm{d}s\right)^{\frac{\ell}{2}}\right] \leq (\frac{\ell}{2})^{\frac{\ell}{2}} \overline{\mathbb{E}}\left[\left(\int_0^t \int_G |\overline{\gamma}_{s,t,e'}|^2 \overline{N}(\mathrm{d}e,\mathrm{d}s)\right)^{\frac{\ell}{2}}\right] \leq C_\ell. \tag{3.18}$$

Hence, for $\ell \geq 2$, (3.16) holds true.

If $1 \leq \ell < 2$, for each $e' \in G$, the fact $(\overline{\theta}_{\cdot,t,e'}, \overline{\gamma}_{\cdot,t,e'}) \in \mathcal{H}^2_{\overline{\mathbb{F}}}(0,t) \times \overline{K}^2_\lambda(0,t)$ and Hölder inequality allow to show (3.16).

We now estimate $\Xi_1^\varepsilon(t,e'), \Xi_2^\varepsilon(t,e'), \Xi_3^\varepsilon(t,e')$ one after another.

First, as for $\Xi_1^\varepsilon(t,e')$, following (3.15) we have

$$\Xi_1^\varepsilon(t,e') = \overline{\mathbb{E}}\Big[\int_0^t \overline{\theta}_{s,t,e'}\Big(\overline{m}(s)\widetilde{\mathbb{E}}[\overline{\sigma}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)] + \overline{m}(s)\delta\overline{\sigma}(s)\mathbb{1}_{E_\varepsilon}(s)\Big)\mathrm{d}s\Big].$$

From this, the boundness of $\delta\sigma$ and Hölder inequality, it yields

$$\begin{aligned}
&|\Xi_1^\varepsilon(t,e')|^2 \\
&\leq 2\overline{\mathbb{E}}\left[\left(\int_0^t |\overline{\theta}_{s,t,e'}\overline{m}(s)\widetilde{\mathbb{E}}[\overline{\sigma}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|\mathrm{d}s\right)^2\right] + 2\overline{\mathbb{E}}\left[\left(\int_0^t |\overline{\theta}_{s,t,e'}\overline{m}(s)\delta\overline{\sigma}(s)\mathbb{1}_{E_\varepsilon}(s)|\mathrm{d}s\right)^2\right] \\
&\leq 2\left\{\overline{\mathbb{E}}[(\int_0^t |\overline{\theta}_{s,t,e'}|^2\mathrm{d}s)^3]\right\}^{\frac{1}{3}} \cdot \left\{\overline{\mathbb{E}}\left[\sup_{s\in[0,T]}|\overline{m}(s)|^{12}\right]\right\}^{\frac{1}{6}} \cdot \left\{\overline{\mathbb{E}}[(\int_0^t |\widetilde{\mathbb{E}}[\overline{\sigma}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^3 ds)^{\frac{4}{3}}]\right\}^{\frac{1}{2}} \\
&\quad + 2\varepsilon\left\{\overline{\mathbb{E}}\left[\sup_{s\in[0,T]}|\overline{m}(s)|^4\right]\right\}^{\frac{1}{2}} \cdot \left\{\overline{\mathbb{E}}[(\int_0^t |\mathbb{1}_{E_\varepsilon}(s)\overline{\theta}_{s,t,e'}|^2\mathrm{d}s)^2]\right\}^{\frac{1}{2}}.
\end{aligned}$$

Hence, thanks to (3.9), (3.16), there exists a constant $C>0$ independent of e' such that

$$|\Xi_1^\varepsilon(t,e')|^4 \le C\overline{\mathbb{E}}[\int_0^t |\widetilde{\mathbb{E}}[\overline{\sigma}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\mathrm{d}s] + C\varepsilon^2\rho^*(\varepsilon),$$

where $\rho^*(\varepsilon) := \overline{\mathbb{E}}[(\int_0^t |\mathbb{1}_{E_\varepsilon}(s)\overline{\theta}_{s,t,e'}|^2\mathrm{d}s)^2]$. Obviously, the Dominated Convergence Theorem implies $\rho^*(\varepsilon)\to 0$ as $\varepsilon\to 0$. Then it follows $\lambda(G)<\infty$ and (3.6)-i) that

$$\int_0^r\int_G \mathbb{E}|\Xi_1^\varepsilon(t,e')|^4\lambda(\mathrm{d}e')\mathrm{d}t \le C\int_0^r\Big(\int_0^t \overline{\mathbb{E}}[|\widetilde{\mathbb{E}}[\overline{\sigma}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4]\mathrm{d}s\Big)\mathrm{d}t + C\varepsilon^2\rho^*(\varepsilon) \le \varepsilon^2\rho(\varepsilon). \tag{3.19}$$

Second, we now pay attention to $\Xi_2^\varepsilon(t,e')$. Due to

$$\Xi_2^\varepsilon(t,e') = \overline{\mathbb{E}}\Big[\int_0^t\int_G \overline{\gamma}_{s,t,e'}(e)\overline{m}(s)\widetilde{\mathbb{E}}[\overline{\beta}_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]\frac{1}{1+\overline{\beta}_x(s,e)}\lambda(\mathrm{d}e)\mathrm{d}s\Big],$$

and from (3.9), (3.16) we get

$$\begin{aligned}|\Xi_2^\varepsilon(t,e')|^2 \le C\Big\{\overline{\mathbb{E}}[\int_0^t\int_G |\widetilde{\mathbb{E}}[\overline{\beta}_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e)\mathrm{d}s]\Big\}^{\frac12}\cdot\\ \Big\{\overline{\mathbb{E}}[(\int_0^t\int_G |\overline{\gamma}_{s,t,e'}(e)|^2\lambda(\mathrm{d}e)\mathrm{d}s)^4]\Big\}^{\frac14}\cdot\Big\{\overline{\mathbb{E}}\Big[\sup_{s\in[0,T]}|\overline{m}(s)|^8\Big]\Big\}^{\frac14}.\end{aligned}$$

Hence,

$$\int_0^r\int_G \mathbb{E}[|\Xi_2^\varepsilon(t,e')|^4]\lambda(\mathrm{d}e')\mathrm{d}t \le C\int_0^r \overline{\mathbb{E}}[\int_0^t\int_G |\widetilde{\mathbb{E}}[\overline{\beta}_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e)\mathrm{d}s]\mathrm{d}t. \tag{3.20}$$

Third, as for $\Xi_3^\varepsilon(t,e')$, since $|\psi_3(t,e')|\le C(1\wedge|e'|)\le C$, we have

$$\begin{aligned}|\Xi_3^\varepsilon(t,e')| &\le C\overline{\mathbb{E}}\Big[|\overline{\psi}_2(t)\overline{\Theta}_3^\varepsilon(t)|\Big]\\ &\le C\Big\{\overline{\mathbb{E}}[|\overline{\psi}_2(t)\overline{n}(t)\int_0^t(\overline{m}(s)\widetilde{\mathbb{E}}[\overline{b}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)] + \overline{m}(s)\delta\overline{b}(s)\mathbb{1}_{E_\varepsilon}(s))\mathrm{d}s|]\\ &\quad+\overline{\mathbb{E}}[|\overline{\psi}_2(t)\overline{n}(t)\int_0^t(\overline{m}(s)\overline{\sigma}_x(s)\widetilde{\mathbb{E}}[\overline{\sigma}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)] + \overline{m}(s)\overline{\sigma}_x(s)\delta\overline{\sigma}(s)\mathbb{1}_{E_\varepsilon}(s))\mathrm{d}s|]\\ &\quad+\overline{\mathbb{E}}[|\overline{\psi}_2(t)\overline{n}(t)\int_0^t\int_G \overline{m}(s)\frac{\overline{\beta}_x(s,e)}{1+\overline{\beta}_x(s,e)}\widetilde{\mathbb{E}}[\overline{\beta}_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]\lambda(\mathrm{d}e)\mathrm{d}s|]\Big\}.\end{aligned} \tag{3.21}$$

The boundness of b,σ,σ_x, (3.9) and the assumption $\overline{\mathbb{E}}[\sup_{t\in[0,T]}|\overline{\psi}_2(t)|^{2\ell}]\le C_\ell,\ \ell\ge 1$ allow to show

$$\overline{\mathbb{E}}\Big[|\overline{\psi}_2(t)\overline{n}(t)\int_0^t \overline{m}(s)(\delta\overline{b}(s)+\overline{\sigma}_x(s)\delta\overline{\sigma}(s))\mathbb{1}_{E_\varepsilon}(s)\mathrm{d}s|\Big] \le C\varepsilon. \tag{3.22}$$

On the other hand, notice that

$$\begin{aligned}
&\text{i) } \overline{\mathbb{E}}\Big[|\overline{\psi}_2(t)\overline{n}(t)\int_0^t \overline{m}(s)\widetilde{\mathbb{E}}[\overline{b}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]\mathrm{d}s|\Big]\\
&\quad\le \Big\{\overline{\mathbb{E}}\big[\int_0^t |\widetilde{\mathbb{E}}[\overline{b}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^2\mathrm{d}s\big]\Big\}^{\frac{1}{2}}\cdot\Big\{\overline{\mathbb{E}}\big[\sup_{t\in[0,T]}|\overline{\psi}_2(t)\overline{n}(t)|^4\big]\Big\}^{\frac{1}{4}}\cdot\Big\{\overline{\mathbb{E}}\big[\sup_{t\in[0,T]}|\overline{m}(t)|^4\big]\Big\}^{\frac{1}{4}}\\
&\quad\le C\Big\{\overline{\mathbb{E}}\big[\int_0^t |\widetilde{\mathbb{E}}[\overline{b}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^2\mathrm{d}s\big]\Big\}^{\frac{1}{2}};\\
&\text{ii) } \overline{\mathbb{E}}\Big[|\overline{\psi}_2(t)\overline{n}(t)\int_0^t \overline{m}(s)\overline{\sigma}_x(s)\widetilde{\mathbb{E}}[\overline{\sigma}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]\mathrm{d}s|\Big]\le C\Big\{\overline{\mathbb{E}}\big[\int_0^t |\widetilde{\mathbb{E}}[\overline{\sigma}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^2\mathrm{d}s\big]\Big\}^{\frac{1}{2}};\\
&\text{iii) } \overline{\mathbb{E}}\Big[|\overline{\psi}_2(t)\overline{n}(t)\int_0^t\int_G \overline{m}(s)\frac{\overline{\beta}_x(s,e)}{1+\overline{\beta}_x(s,e)}\widetilde{\mathbb{E}}[\overline{\beta}_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]\lambda(\mathrm{d}e)\mathrm{d}s|\Big]\\
&\quad\le C\Big\{\overline{\mathbb{E}}\big[\int_0^t\int_G |\widetilde{\mathbb{E}}[\overline{\beta}_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^2\lambda(\mathrm{d}e)\mathrm{d}s\big]\Big\}^{\frac{1}{2}}.
\end{aligned}$$

Combining (3.21), (3.22) and the above i), ii), iii), we know that there exists a constant $C>0$ independent of e', such that

$$\begin{aligned}
|\Xi_3^\varepsilon(t,e')|^4 \le{}& C\varepsilon^4 + C\overline{\mathbb{E}}\big[\int_0^t |\widetilde{\mathbb{E}}[\overline{b}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\mathrm{d}s\big] + C\overline{\mathbb{E}}\big[\int_0^t |\widetilde{\mathbb{E}}[\overline{\sigma}_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\mathrm{d}s\big]\\
&+ C\overline{\mathbb{E}}\big[\int_0^t\int_G |\widetilde{\mathbb{E}}[\overline{\beta}_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e)\mathrm{d}s\big].
\end{aligned} \tag{3.23}$$

Hence, from (3.6)-i)

$$\begin{aligned}
&\int_0^r\int_G \mathbb{E}|\Xi_3^\varepsilon(t,e')|^4\lambda(\mathrm{d}e')\mathrm{d}t\\
&\quad\le \varepsilon^2\rho(\varepsilon) + C\int_0^r\int_G \overline{\mathbb{E}}\Big[\int_0^t\int_G |\widetilde{\mathbb{E}}[\overline{\beta}_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e)\mathrm{d}s\Big]\lambda(\mathrm{d}e')\mathrm{d}t\\
&\quad\le \varepsilon^2\rho(\varepsilon) + C\lambda(G)\int_0^r \overline{\mathbb{E}}\Big[\int_0^t\int_G |\widetilde{\mathbb{E}}[\overline{\beta}_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e)\mathrm{d}s\Big]\mathrm{d}t.
\end{aligned} \tag{3.24}$$

Thanks to (3.13), (3.19), (3.20), (3.24), we have

$$\begin{aligned}
&\mathbb{E}\Big[\int_0^r\int_G |\overline{\mathbb{E}}[\overline{\psi}_3(s,e')\overline{\psi}_2(s)\overline{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e')\mathrm{d}s\Big]\\
&\le \int_0^r\int_G \mathbb{E}|\Xi_1^\varepsilon(s,e')+\Xi_2^\varepsilon(s,e')+\Xi_3^\varepsilon(s,e')|^4\lambda(\mathrm{d}e')\mathrm{d}s\\
&\le \varepsilon^2\rho(\varepsilon) + C\int_0^r \mathbb{E}\Big[\int_0^t\int_G |\widetilde{\mathbb{E}}[\beta_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e)\mathrm{d}s\Big]\mathrm{d}t.
\end{aligned} \tag{3.25}$$

Notice that $(\overline{\Omega},\overline{\mathcal{F}},\overline{P})$ is an intermediate probability space. So if we take $\widetilde{\psi}_3(s,e)=\beta_\nu(s,e;\widetilde{X}^*(s))$, $\widetilde{\psi}_2(s)=1$, the Gronwall inequality can show, for $t\in[0,T]$,

$$\mathbb{E}\Big[\int_0^t\int_G |\widetilde{\mathbb{E}}[\beta_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e)\mathrm{d}s\Big]\le\varepsilon^2\rho(\varepsilon),$$

which and (3.25) imply the desired result, *i.e.*, ii) of (3.6). □

Corollary 3.10. *In (3.6), if* $\widetilde{\psi}_2(t)=1$, $\widetilde{\psi}_1(t)=b_\nu(t;\widetilde{X}^*(t)),\sigma_\nu(t;\widetilde{X}^*(t))$ *and* $\widetilde{\psi}_3(t,e)=\beta_\nu(t,e;\widetilde{X}^*(t))$, *separately, one has*

$$\begin{aligned}&\text{i)}\quad \mathbb{E}\bigg[\int_0^T|\widetilde{\mathbb{E}}[b_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\mathrm{d}s\bigg]+\mathbb{E}\bigg[\int_0^T|\widetilde{\mathbb{E}}[\sigma_\nu(s;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\mathrm{d}s\bigg]\le\varepsilon^2\rho(\varepsilon);\\&\text{ii)}\quad \mathbb{E}\bigg[\int_0^T\int_G|\widetilde{\mathbb{E}}[\beta_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e)\mathrm{d}s\bigg]\le\varepsilon^2\rho(\varepsilon).\end{aligned}\tag{3.26}$$

Proposition 3.11. *Let the Assumptions (3.1–(3.3)), 3.8 hold true, then*

$$\mathbb{E}[\sup_{t\in[0,T]}|X^\varepsilon(t)-X^*(t)-X^{1,\varepsilon}(t)-X^{2,\varepsilon}(t)|^2]\le\varepsilon^2\rho(\varepsilon).\tag{3.27}$$

Proposition 3.12. *Let us define*

$$(\phi_x,\phi_{xx})(T):=\left(\frac{\partial\phi}{\partial x},\frac{\partial^2\phi}{\partial x^2}\right)(X^*(T),P_{X^*(T)}),(\phi_\nu,\phi_{\nu a})(T;\widetilde{X}^*(T)):=\left(\frac{\partial\phi}{\partial\nu},\frac{\partial^2\phi}{\partial\nu\partial a}\right)(X^*(T),P_{X^*(T)};\widetilde{X}^*(T)),$$

and

$$\begin{aligned}&\Lambda_T^\varepsilon(\phi):=\phi(X^\varepsilon(T),P_{X^\varepsilon(T)})-\phi(X^*(T),P_{X^*(T)})-\Big\{\phi_x(T)(X^{1,\varepsilon}(T)+X^{2,\varepsilon}(T))\\&+\widetilde{\mathbb{E}}[\phi_\nu(T;\widetilde{X}^*(T))(\widetilde{X}^{1,\varepsilon}(T)+\widetilde{X}^{2,\varepsilon}(T))]+\frac{1}{2}\phi_{xx}(T)(X^{1,\varepsilon}(T))^2+\frac{1}{2}\widetilde{\mathbb{E}}[\phi_{\nu a}(T;\widetilde{X}^*(T))(\widetilde{X}^{1,\varepsilon}(T))^2]\Big\},\end{aligned}\tag{3.28}$$

then

$$\mathbb{E}[|\Lambda_T^\varepsilon(\phi)|^2]\le\varepsilon^2\rho(\varepsilon).\tag{3.29}$$

The similar proofs of the above two propositions can be found in Buckdahn, Li and Ma [5].

3.3. Adjoint equations

In order to apply duality method to investigate our stochastic maximum principle, two adjoint equations are brought in. Compared with the classical case, see Hu [17], a remarkable difference is that the first-order adjoint equation is a mean-field BSDE with jumps. But the second-order adjoint equation is a classical linear BSDE with jumps, but not mean-field type.

Let us first introduce some notations, which are used time to time, for $\ell = x, y, z, k$, $\theta = x, y, z, k, a_i$, $i, j = 1, 2$,

$$\begin{aligned} &\Pi^*(s) = (X^*(s), Y^*(s), Z^*(s), \int_G K^*(s,e)l(e)\lambda(\mathrm{d}e)), \quad \Lambda^*(s) = (X^*(s), Y^*(s)),\\ &f_\ell(s) = \frac{\partial f}{\partial \ell}(s, \Pi^*(s), P_{\Lambda^*(s)}, u^*(s)),\\ &(f_{\mu_i}, f_{\mu_i\theta})(s; \widetilde{\Lambda}^*(s)) = \left(\left(\frac{\partial f}{\partial \mu}\right)_i, \frac{\partial}{\partial \theta}\left(\left(\frac{\partial f}{\partial \mu}\right)_i\right)\right)(s, \Pi^*(s), P_{\Lambda^*(s)}, u^*(s); \widetilde{\Lambda}^*(s)),\\ &(\widetilde{f}_{\mu_i}, \widetilde{f}_{\mu_i\theta})(s) = \left(\left(\frac{\partial f}{\partial \mu}\right)_i, \frac{\partial}{\partial \theta}\left(\left(\frac{\partial f}{\partial \mu}\right)_i\right)\right)(s, \widetilde{\Pi}^*(s), P_{\Lambda^*(s)}, \widetilde{u}^*(s); \Lambda^*(s)),\\ &f_{\mu_i\mu_j}(s; \widehat{\widetilde{\Lambda}}^*(s)) = \left(\frac{\partial}{\partial \mu}\left(\left(\frac{\partial f}{\partial \mu}\right)_i\right)\right)_j(s, \Pi^*(s), P_{\Lambda^*(s)}, u^*(s); \widetilde{\Lambda}^*(s), \widehat{\Lambda}^*(s)). \end{aligned} \tag{3.30}$$

With these concise notations in hand, the first-order adjoint equation can be read as

$$\begin{cases} -\mathrm{d}Y^1(s) = F(s)\mathrm{d}s - Z^1(s)\mathrm{d}W(s) - \int_G R^1(s,e)N_\lambda(\mathrm{d}e, \mathrm{d}s), \ s \in [0,T],\\ \quad Y^1(T) = \phi_x(T) + \widetilde{\mathbb{E}}[\widetilde{\phi}_\nu(T)], \end{cases} \tag{3.31}$$

where

$$\begin{aligned} F(s) = &\, Y^1(s)\Big(f_y(s) + \widetilde{\mathbb{E}}[\widetilde{f}_{\mu_2}(s)] + f_z(s)\sigma_x(s) + \int_G f_k(s)l(e)\beta_x(s,e)\lambda(\mathrm{d}e) + b_x(s)\Big)\\ &+ \widetilde{\mathbb{E}}\Big[\widetilde{Y}^1(s)\Big(\widetilde{f}_z(s)\widetilde{\sigma}_\nu(s) + \int_G \widetilde{f}_k(s)l(e)\widetilde{\beta}_\nu(s,e)\lambda(\mathrm{d}e) + \widetilde{b}_\nu(s)\Big)\Big]\\ &+ Z^1(s)\Big(f_z(s) + \sigma_x(s)\Big) + \widetilde{\mathbb{E}}[\widetilde{Z}^1(s)\widetilde{\sigma}_\nu(s)] + f_x(s) + \widetilde{\mathbb{E}}[\widetilde{f}_{\mu_1}(s)]\\ &+ \int_G \Big(R^1(s,e)\big(f_k(s)l(e) + \beta_x(s,e)\big) + \widetilde{\mathbb{E}}[\widetilde{R}^1(s,e)\widetilde{\beta}_\nu(s,e)]\Big)\lambda(\mathrm{d}e). \end{aligned}$$

Under the Assumptions (3.1)–(3.2) the unique solution of equation (3.31), (Y^1, Z^1, R^1), satisfies, for $\ell \geq 2$,

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |Y^1(t)|^\ell + \Big(\int_0^T |Z^1(t)|^2\mathrm{d}t\Big)^{\frac{\ell}{2}} + \Big(\int_0^T \int_G |R^1(t,e)|^2\lambda(\mathrm{d}e)\mathrm{d}t\Big)^{\frac{\ell}{2}}\Big] \leq C_\ell \tag{3.32}$$

(see Li [21], Prop. 4.1).

Once obtaining the solution (Y^1, Z^1, R^1) of the equation (3.31), we can consider the following second-order adjoint equation

$$\begin{cases} -\mathrm{d}Y^2(s) = G(s)\mathrm{d}s - Z^2(s)\mathrm{d}W(s) - \int_G R^2(s,e)N_\lambda(\mathrm{d}e, \mathrm{d}s), \ s \in [0,T],\\ \quad Y^2(T) = \phi_{xx}(T) + \widetilde{\mathbb{E}}[\widetilde{\phi}_{\nu y}(T)], \end{cases} \tag{3.33}$$

where

$$
\begin{aligned}
G(s) = {} & Y^2(s)\Big(f_y(s) + \widetilde{\mathbb{E}}[\widetilde{f}_{\mu_2}(s)] + 2f_z(s)\sigma_x(s) + \int_G f_k(s)l(e)\big(2\beta_x(s,e) + (\beta_x(s,e))^2\big)\lambda(\mathrm{d}e) \\
& + 2b_x(s) + (\sigma_x(s))^2 + \int_G (\beta_x(s,e))^2\lambda(\mathrm{d}e)\Big) \\
& + Z^2(s)\Big(f_z(s) + 2\sigma_x(s)\Big) + \int_G R^2(s,e)\Big(f_k(s)l(e) + 2\beta_x(s,e) + (\beta_x(s,e))^2\Big)\lambda(\mathrm{d}e) \\
& + Y^1(s)\Big(f_z(s)\sigma_{xx}(s) + \int_G f_k(s)l(e)\beta_{xx}(s,e)\lambda(\mathrm{d}e) + b_{xx}(s)\Big) + \widetilde{\mathbb{E}}\Big[\widetilde{Y}^1(s)\Big(\widetilde{f}_z(s)\widetilde{\sigma}_{\nu a}(s) \\
& + \int_G \widetilde{f}_k(s)l(e)\widetilde{\beta}_{\nu a}(s,e)\lambda(\mathrm{d}e) + \widetilde{b}_{\nu a}(s)\Big)\Big] + Z^1(s)\sigma_{xx}(s) + \widetilde{\mathbb{E}}[\widetilde{Z}^1(s)\widetilde{\sigma}_{\nu a}(s)] \\
& + \int_G \Big(R^1(s,e)\beta_{xx}(s,e) + \widetilde{\mathbb{E}}[\widetilde{R}^1(s,e)\widetilde{\beta}_{\nu a}(s,e)]\Big)\lambda(\mathrm{d}e) + O(s)D^2f(s)O^\intercal(s) \\
& + \widetilde{\mathbb{E}}[\widetilde{f}_{\mu_1 a_1}(s)] + (Y^1(s))^2\widetilde{\mathbb{E}}[\widetilde{f}_{\mu_2 a_2}(s)],
\end{aligned}
$$

and $O(s) = (1, Y^1(s), Y^1(s)\sigma_x(s) + Z^1(s), \int_G l(e)(Y^1(s)\beta_x(s,e) + R^1(s,e))\lambda(\mathrm{d}e))$, $D^2f(s)$ denotes the Hessian matrix of f with respect to (x,y,z,k), *i.e.*, $D^2 f = \begin{pmatrix} f_{xx}(t) & f_{xy}(t) & f_{xz}(t) & f_{xk}(t) \\ f_{yx}(t) & f_{yy}(t) & f_{yz}(t) & f_{yk}(t) \\ f_{zx}(t) & f_{zy}(t) & f_{zz}(t) & f_{zk}(t) \\ f_{kx}(t) & f_{ky}(t) & f_{kz}(t) & f_{kk}(t) \end{pmatrix}$. Since we have known (Y^1, Z^1, R^1), the equation (3.33) is a classical linear BSDE with jumps. From the well-known existence and uniqueness theorem of BSDEs with jumps, under the Assumptions (3.1)–(3.3) the equation (3.33) admits a unique solution (Y^2, Z^2, R^2) and, moreover, for $\ell \geq 2$,

$$
\mathbb{E}\Big[\sup_{t\in[0,T]}|Y^2(t)|^\ell + \Big(\int_0^T |Z^2(t)|^2\mathrm{d}t\Big)^{\frac{\ell}{2}} + \Big(\int_0^T\int_G |R^2(t,e)|^2\lambda(\mathrm{d}e)\mathrm{d}t\Big)^{\frac{\ell}{2}}\Big] \leq C_\ell. \tag{3.34}
$$

From Lemma 3.9, the following estimates hold true.

Corollary 3.13. *Let the Assumptions (3.1)–(3.8) hold true, and set for $\ell = x, y, z$,*

$$
\begin{aligned}
&\widetilde{M_1}(s) = (f_{\mu_1}, f_{\mu_2}, f_{\mu_1\ell}, f_{\mu_2\ell})(s;\widetilde{\Lambda}^*(s)), \quad \widetilde{M_2}(s) = (f_{k\mu_1}, f_{k\mu_2})(s;\widetilde{\Lambda}^*(s)), \\
&\widehat{\widetilde{M}}_3(s) = (f_{\mu_1\mu_1}, f_{\mu_1\mu_2}, f_{\mu_2\mu_2})(s;\widehat{\widetilde{\Lambda}}^*(s)).
\end{aligned}
$$

Moreover, let $X^{1,\varepsilon}$ and Y^1 be the solutions of (3.2) and (3.31), respectively, then

$$
\begin{aligned}
&\text{i) } \mathbb{E}\Big[\int_0^T |\widetilde{\mathbb{E}}[\widetilde{M_1}(s)\widetilde{Y}^1(s)\widetilde{X}^{1,\varepsilon}(s)]|^4\mathrm{d}s + \int_0^T\int_G |\widetilde{\mathbb{E}}[\widetilde{M_2}(s)\widetilde{Y}^1(s)\widetilde{X}^{1,\varepsilon}(s)|^4\lambda(\mathrm{d}e)\mathrm{d}s\Big] \leq \varepsilon^2\rho(\varepsilon); \\
&\text{ii) } \mathbb{E}\widehat{\mathbb{E}}\Big[\int_0^T |\widetilde{\mathbb{E}}[\widehat{\widetilde{M}}_3(s)\widetilde{Y}^1(s)\widetilde{X}^{1,\varepsilon}(s)]|^4\mathrm{d}s\Big] \leq \varepsilon^2\rho(\varepsilon).
\end{aligned}
\tag{3.35}
$$

4. The second-order expansion of cost functional Y^ε

The second-order expansion of cost functional Y^ε is stated in this section, which plays an important role in proving our stochastic maximum principle. More precisely, we prove that there exists a stochastic process

$\breve{P} = (\breve{P}(t))_{t\in[0,T]}$ with $\breve{P}(T) = 0$, such that, for all $t \in [0, T]$,

$$Y^\varepsilon(t) = Y^*(t) + Y^1(t)(X^{1,\varepsilon}(t) + X^{2,\varepsilon}(t)) + \frac{1}{2}Y^2(t)(X^{1,\varepsilon}(t))^2 + \breve{P}(t) + o(\varepsilon), \tag{4.1}$$

where the convergence is in $L^2(\Omega, C[0,T])$ sense.

For this purpose, let us first introduce the following linear mean-field BSDE with jumps:

$$\begin{cases} -\mathrm{d}\breve{P}(t) = \Big(f_y(t)\breve{P}(t) + f_z(t)\breve{Q}(t) + \int_G f_k(t)l(e)\breve{K}(t,e)\lambda(\mathrm{d}e) + \widetilde{\mathbb{E}}[f_{\mu_2}(t;\widetilde{\Lambda}^*(t))\widetilde{\breve{P}}(t)] \\ \qquad + \big(A_1(t) + \Delta f(t)\big)\mathbb{1}_{E_\varepsilon}(t)\Big)\mathrm{d}t - \breve{Q}(t)\mathrm{d}W(t) - \int_G \breve{K}(t,e)N_\lambda(\mathrm{d}e, \mathrm{d}t),\ t \in [0,T], \\ \breve{P}(T) = 0, \end{cases} \tag{4.2}$$

where

$$\begin{aligned} A_1(t) &= Y^1(t)\delta b(t) + Z^1(t)\delta\sigma(t) + \frac{1}{2}Y^2(t)(\delta\sigma(t))^2, \\ \Delta f(t) &= f(t, X^*(t), Y^*(t), Z^*(t) + Y^1(t)\delta\sigma(t), \int_G K^*(t,e)l(e)\lambda(\mathrm{d}e), P_{(X^*(t),Y^*(t))}, v(t)) \\ &\quad - f(t, X^*(t), Y^*(t), Z^*(t), \int_G K^*(t,e)l(e)\lambda(\mathrm{d}e), P_{(X^*(t),Y^*(t))}, u^*(t)). \end{aligned}$$

Obviously, under the Assumptions (3.1)–(3.2) the equation (4.2) possesses a unique solution $(\breve{P}, \breve{Q}, \breve{K}) \in \mathcal{S}^2_{\mathbb{F}}(0,T) \times \mathcal{H}^2_{\mathbb{F}}(0,T) \times K^2_\lambda(0,T)$ (see [21]). Moreover,

Proposition 4.1. *Let the Assumptions (3.1)–(3.2) hold true, then for $\ell \geq 2$,*

$$E\Big[\sup_{t\in[0,T]} |\breve{P}(t)|^\ell + \Big(\int_0^T |\breve{Q}(t)|^2\mathrm{d}t\Big)^{\frac{\ell}{2}} + \Big(\int_0^T \int_G |\breve{K}(t,e)|^2\lambda(\mathrm{d}e)\mathrm{d}t\Big)^{\frac{\ell}{2}}\Big] \leq \varepsilon^{\frac{\ell}{2}}\rho_\ell(\varepsilon), \tag{4.3}$$

where $\rho_\ell : (0, +\infty) \to (0, +\infty)$ depending only on ℓ with $\rho_\ell(\varepsilon) \to$ as $\varepsilon \to 0$.

Proof. From the standard argument for the solutions of classical BSDEs with jumps, we have, for $\ell \geq 2$,

$$\begin{aligned} &\mathbb{E}\Big[\sup_{t\in[0,T]} |\breve{P}(t)|^\ell + \Big(\int_0^T |\breve{Q}(t)|^2\mathrm{d}t\Big)^{\frac{\ell}{2}} + \Big(\int_0^T \int_G |\breve{K}(t,e)|^2\lambda(\mathrm{d}e)\mathrm{d}t\Big)^{\frac{\ell}{2}}\Big] \\ &\leq C_\ell\mathbb{E}\Big[(\int_0^T |A_1(t) + \Delta f(t)|\mathbb{1}_{E_\varepsilon}(t)\mathrm{d}t)^\ell\Big]. \end{aligned} \tag{4.4}$$

The reader can refer to [22], [23] for more details.

On the other hand, thanks to the boundness of b, σ and the Lipschitz property of f, Hölder inequality implies that

$$\begin{aligned} &\mathbb{E}\Big[\Big(\int_0^T |A(t) + \Delta f(t)|\mathbb{1}_{E_\varepsilon}(t)\mathrm{d}t\Big)^\ell\Big] \\ &\leq C\mathbb{E}\Big[\Big(\int_0^T (|Y^1(s)\delta b(s) + Z^1(s)\delta\sigma(s) + \frac{1}{2}Y^2(s)(\delta\sigma(s))^2 + Y^1(s) + 1|\mathbb{1}_{E_\varepsilon}(s))\mathrm{d}s\Big)^\ell\Big] \\ &\leq \varepsilon^{\frac{\ell}{2}}\rho_\ell(\varepsilon), \end{aligned}$$

where $\rho_\ell(\varepsilon) := \varepsilon^{\frac{\ell}{2}}(E[\sup_{s\in[0,T]}|Y^1(s)|^\ell + \sup_{s\in[0,T]}|Y^2(s)|^\ell] + 1) + \mathbb{E}[(\int_0^T |Z^1(s)|^2 \mathbb{1}_{\mathbb{E}_\varepsilon}(s)\mathrm{d}s)^{\frac{\ell}{2}}]$. Clearly, $\rho_\ell(\varepsilon) \to 0$ as $\varepsilon \to 0$. The proof is completed. □

The following theorem shows the second-order expansion of cost functional Y^ε.

Theorem 4.2. *Let the Assumptions (3.1)–(3.8) hold true, then there exists an adapted stochastic process over $[0,T]$, $\breve{P} = (\breve{P}(t))_{t\in[0,T]}$ with $\breve{P}(T) = 0$, such that*

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |Y^\varepsilon(t) - Y^*(t) - \breve{P}(t) - Y^1(t)(X^{1,\varepsilon}(t) + X^{2,\varepsilon}(t)) - \frac{1}{2}Y^2(t)(X^{1,\varepsilon}(t))^2|^2\Big] \le \varepsilon^2\rho(\varepsilon). \tag{4.5}$$

Proof. Like investigating classical Pontryagin Maximum Principle, an important element of proving Theorem 4.2 is to apply Itô's formula to

$$M(t) := Y^1(t)(X^{1,\varepsilon}(t) + X^{2,\varepsilon}(t)) + \frac{1}{2}Y^2(t)(X^{1,\varepsilon}(t))^2. \tag{4.6}$$

For this, we have

$$\begin{aligned} Y^1(t)(X^{1,\varepsilon}(t) + X^{2,\varepsilon}(t)) + \frac{1}{2}Y^2(t)(X^{1,\varepsilon}(t))^2 &= Y^1(T)(X^{1,\varepsilon}(T) + X^{2,\varepsilon}(T)) + \frac{1}{2}Y^2(T)(X^{1,\varepsilon}(T))^2 \\ - \int_t^T \Big(A(s) + A_4(s)\mathbb{1}_{E_\varepsilon}(s) + A_5(s)\Big)\mathrm{d}s &- \int_t^T \Big(B(s) + B_4(s)\mathbb{1}_{E_\varepsilon}(s) + B_5(s)\Big)\mathrm{d}W(s) \\ - \int_t^T \int_G \Big(C^-(s,e) + C_4^-(s,e)\Big)N_\lambda(\mathrm{d}e,\mathrm{d}s), \end{aligned} \tag{4.7}$$

where $A, B, C, A_4, \cdots, C_4$ are given in Appendix.

Let us first admit the following lemma for a moment. Lemma 4.3 argues the powers of $\int_0^T |A_4(t)\mathbb{1}_{E_\varepsilon}(t)|\mathrm{d}t$, $\int_0^T |A_5(t)|\mathrm{d}t$, $\int_0^T |B_4(t)\mathbb{1}_{E_\varepsilon}(t)|\mathrm{d}t$, $\int_0^T |B_5(t)|\mathrm{d}t$, $\int_0^T\int_G |C_4(t,e)|\lambda(\mathrm{d}e)\mathrm{d}t$, as the elements of $L^2(\Omega)$, are all $o(\varepsilon)$. Note that due to the structures of A_4, A_5, B_4, B_5, C_4 involving the first- and second-order derivatives of the coefficients with respect to a measure, hence, the proof is not trivial and far from the classical case. In the proof of Lemma 4.3 we mainly borrow the new estimates given in Corollary 3.10 and Corollary 3.13. We place the proof of Lemma 4.3 in Appendix.

Lemma 4.3. *We make the same Assumptions as in Theorem 4.2, then the following estimates hold true:*

$$\begin{aligned} &\text{i)}\quad \mathbb{E}\Big[\Big(\int_0^T |A_4(t)\mathbb{1}_{E_\varepsilon}(t)|\mathrm{d}t\Big)^2\Big] + \mathbb{E}\Big[\Big(\int_0^T |A_5(t)|\mathrm{d}t\Big)^2\Big] \le \varepsilon^2\rho(\varepsilon); \\ &\text{ii)}\quad \mathbb{E}\Big[\Big(\int_0^T |B_4(t)\mathbb{1}_{E_\varepsilon}(t)|\mathrm{d}t\Big)^2\Big] + \mathbb{E}\Big[\Big(\int_0^T |B_5(t)|\mathrm{d}t\Big)^2\Big] \le \varepsilon^2\rho(\varepsilon); \\ &\text{iii)}\quad \mathbb{E}\Big[\Big(\int_0^T\int_G |C_4(t,e)|\lambda(\mathrm{d}e)\mathrm{d}t\Big)^2\Big] \le \varepsilon^2\rho(\varepsilon); \\ &\text{iv)}\quad E\Big[\int_0^T \Big(|M(t)|^2 + |B(t)|^2 + \int_G |C(t,e)|^2\lambda(\mathrm{d}e)\Big)\mathbb{1}_{E_\varepsilon}(t)\mathrm{d}t\Big] \le \varepsilon\rho(\varepsilon). \end{aligned} \tag{4.8}$$

With the help of Lemma 4.3, (4.7) can be written as

$$\begin{aligned}&Y^1(t)(X^{1,\varepsilon}(t)+X^{2,\varepsilon}(t))+\frac{1}{2}Y^2(t)(X^{1,\varepsilon}(t))^2=Y^1(T)(X^{1,\varepsilon}(T)+X^{2,\varepsilon}(T))+\frac{1}{2}Y^2(T)(X^{1,\varepsilon}(T))^2\\&-\int_t^T A(s)\mathrm{d}s-\int_t^T B(s)\mathrm{d}W(s)-\int_t^T\int_G C^-(s,e)N_\lambda(\mathrm{d}e,\mathrm{d}s)+o(\varepsilon).\end{aligned}\tag{4.9}$$

For convenience, let us set

$$\begin{aligned}&\Delta X(t)=X^\varepsilon(t)-X^*(t)-X^{1,\varepsilon}(t)-X^{2,\varepsilon}(t),\quad \Delta Y(t)=Y^\varepsilon(t)-Y^*(t)-\breve{P}(t)-M(t),\\&\Delta Z(t)=Z^\varepsilon(t)-Z^*(t)-\breve{Q}(t)-B(t),\quad \Delta K(t,e)=K^\varepsilon(t,e)-K^*(t,e)-\breve{K}(t,e)-C(t,e).\end{aligned}\tag{4.10}$$

By (4.2), (4.9), (4.10) and the definition of $A(s)$, see Appendix, we have

$$\begin{aligned}\Delta Y(t)=&\int_t^T\Big\{f(s,X^\varepsilon(s),Y^\varepsilon(s),Z^\varepsilon(s),\int_G K^\varepsilon(s,e)l(e)\lambda(\mathrm{d}e),P_{(X^\varepsilon(s),Y^\varepsilon(s))},u^\varepsilon(s))\\&-f(s,X^*(s),Y^*(s),Z^*(s),\int_G K^*(s,e)l(e)\lambda(\mathrm{d}e),P_{(X^*(s),Y^*(s))},u^*(s))+A_2(s)+\frac{1}{2}A_3(s)\\&-\Big(f_y(s)\breve{P}(s)+f_z(s)\breve{Q}(s)+\int_G f_k(s)l(e)\breve{K}(s,e)\lambda(\mathrm{d}e)+\widetilde{\mathbb{E}}[f_{\mu_2}(s;\widetilde{\Lambda}^*(s))\widetilde{\breve{P}}(s)]+\Delta f(s)\mathbb{1}_{E_\varepsilon}(s)\Big)\Big\}\mathrm{d}s\\&-\int_t^T\Delta Z(s)\mathrm{d}W(s)-\int_t^T\int_G\Delta K(s,e)N_\lambda(\mathrm{d}s,\mathrm{d}e)+o(\varepsilon),\quad t\in[0,T].\end{aligned}\tag{4.11}$$

We now analyse $f(s,X^\varepsilon(s),Y^\varepsilon(s),Z^\varepsilon(s),\int_G K^\varepsilon(s,e)l(e)\lambda(\mathrm{d}e),P_{(X^\varepsilon(s),Y^\varepsilon(s))},u^\varepsilon(s))-f(s,X^*(s),\ Y^*(s),$ $Z^*(s),\int_G K^*(s,e)l(e)\lambda(\mathrm{d}e),P_{(X^*(s),Y^*(s))},u^*(s))$. To facilitate the presentation, let $\Lambda^\varepsilon(s)=(X^\varepsilon(s),Y^\varepsilon(s))$. First, inspired by (3.27) and the definitions of $\Delta Y,\Delta Z,\Delta K$, we write

$$\begin{aligned}&f(s,X^\varepsilon(s),Y^\varepsilon(s),Z^\varepsilon(s),\int_G K^\varepsilon(s,e)l(e)\lambda(\mathrm{d}e),P_{\Lambda^\varepsilon(s)},u^\varepsilon(s))\\&-f(s,X^*(s),Y^*(s),Z^*(s),\int_G K^*(s,e)l(e)\lambda(\mathrm{d}e),P_{\Lambda^*(s)},u^*(s))\\&=\Delta f(s)\mathbb{1}_{E_\varepsilon}(s)+I_1(s)+I_2(s),\end{aligned}\tag{4.12}$$

where

$$\begin{aligned}I_1(s)=&f(s,X^\varepsilon(s),Y^\varepsilon(s),Z^\varepsilon(s),\int_G K^\varepsilon(s,e)l(e)\lambda(\mathrm{d}e),P_{\Lambda^\varepsilon(s)},u^\varepsilon(s))-f(s,X^*(s)+X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s),\\&Y^*(s)+\breve{P}(s)+M(s),Z^*(s)+\breve{Q}(s)+B(s),\int_G(K^*(s,e)+\breve{K}(s,e)+C(s,e))l(e)\lambda(\mathrm{d}e),\\&P_{(X^*(s)+X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s),Y^*(s)+\breve{P}(s)+M(s))},u^\varepsilon(s)),\\I_2(s)=&f(s,X^*(s)+X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s),Y^*(s)+\breve{P}(s)+M(s),Z^*(s)+\breve{Q}(s)+B(s),\\&\int_G(K^*(s,e)+\breve{K}(s,e)+C(s,e))l(e)\lambda(\mathrm{d}e),P_{(X^*(s)+X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s),Y^*(s)+\breve{P}(s)+M(s))},u^\varepsilon(s))\\&-f(s,X^*(s),Y^*(s),Z^*(s)+Y^1(s)\delta\sigma(s)\mathbb{1}_{E_\varepsilon}(s),\int_G K^*(s,e)l(e)\lambda(\mathrm{d}e),P_{\Lambda^*(s)},u^\varepsilon(s)).\end{aligned}$$

Thanks to the Lipschitz assumption on f and the definitions of $\Delta Y, \Delta Z, \Delta K$ one can obtain

$$
\begin{aligned}
|I_1(s)| \leq C\Big(&|\Delta X(s)| + |\Delta Y(s)| + |\Delta Z(s)| + \int_G |\Delta K(s,e)| l(e)\lambda(\mathrm{d}e) \\
&+ \big(\mathbb{E}[|\Delta X(s)|^2]\big)^{\frac{1}{2}} + \big(\mathbb{E}[|\Delta Y(s)|^2]\big)^{\frac{1}{2}}\Big).
\end{aligned} \tag{4.13}
$$

Now we focus on $I_2(s)$. Obviously, from the definition of $B(s)$, see Appendix, $I_2(s)$ can be written as

$$
I_2(s) = I_3(s) + (I_4(s) - I_3(s))\mathbb{1}_{E_\varepsilon}(s),
$$

where

$$
\begin{aligned}
I_3(s) = f(&s, X^*(s) + X^{1,\varepsilon}(s) + X^{2,\varepsilon}(s), Y^*(s) + \check{P}(s) + M(s), Z^*(s) + \check{Q}(s) + B_2(s) + \frac{1}{2}B_3(s), \\
&\int_G (K^*(s,e) + \check{K}(s,e) + C(s,e)) l(e)\lambda(\mathrm{d}e), P_{(X^*(s)+X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s), Y^*(s)+\check{P}(s)+M(s))}, u^*(s)) \\
&- f(s, X^*(s), Y^*(s), Z^*(s), \int_G K^*(s,e) l(e)\lambda(\mathrm{d}e), P_{\Lambda^*(s)}, u^*(s)), \\
I_4(s) = f(&s, X^*(s) + X^{1,\varepsilon}(s) + X^{2,\varepsilon}(s), Y^*(s) + \check{P}(s) + M(s), Z^*(s) + \check{Q}(s) + B(s), \\
&\int_G (K^*(s,e) + \check{K}(s,e) + C(s,e)) l(e)\lambda(\mathrm{d}e), P_{(X^*(s)+X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s), Y^*(s)+\check{P}(s)+M(s))}, v(s)) \\
&- f(s, X^*(s), Y^*(s), Z^*(s) + Y^1(s)\delta\sigma(s), \int_G K^*(s,e) l(e)\lambda(\mathrm{d}e), P_{(X^*(s), Y^*(s))}, v(s)).
\end{aligned}
$$

According to Proposition 3.6, Proposition 4.1, Lemma 4.3-iv), the definition of $M(s), B_2(s), B_3(s), C(s,e)$, see Appendix, as well as the fact $W_2(P_\xi, P_\eta) \leq (E|\xi - \eta|^2)^{\frac{1}{2}},\ \xi, \eta \in L^2(\Omega, \mathcal{F}_T, P)$, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\int_0^T |I_4(s) - I_3(s)|\mathbb{1}_{E_\varepsilon}(s)\mathrm{d}s\Big)^2\Big] \\
&\quad \leq C\varepsilon\mathbb{E}\Big[\int_0^T \Big(|\check{P}(s) + M(s)|^2 + |\check{Q}(s) + B_2(s) + \frac{1}{2}B_3(s)|^2 + \int_G |\check{K}(s,e) + C(s,e)|^2\lambda(\mathrm{d}e) \\
&\qquad + |X^{1,\varepsilon}(s) + X^{2,\varepsilon}(s)|^2 + \mathbb{E}[|X^{1,\varepsilon}(s) + X^{2,\varepsilon}(s)|^2] + \mathbb{E}[|\check{P}(s) + M(s)|^2]\Big)\mathbb{1}_{\mathbb{E}_\varepsilon}(s)\mathrm{d}s\Big] \\
&\quad \leq C\varepsilon^2\rho(\varepsilon) + C\varepsilon\mathbb{E}\Big[\int_0^T (|M(s)|^2 + |B_2(s) + \frac{1}{2}B_3(s)|^2 + \int_G |C(s,e)|^2\lambda(\mathrm{d}e))\mathbb{1}_{E_\varepsilon}(s)\mathrm{d}s\Big] \\
&\quad \leq C\varepsilon^2\rho(\varepsilon).
\end{aligned} \tag{4.14}
$$

So we now analyse $I_3(s)$. Applying the second-order expansion to $I_3(s)$, see Appendix for further details, we get

$$\begin{aligned}
I_3(s) &= f_y(s)\check{P}(s) + f_z(s)\check{Q}(s) + \int_G f_k(s)\check{K}(s,e)l(e)\lambda(\mathrm{d}e) + \widetilde{\mathbb{E}}[f_{\mu_2}(s;\widetilde{\Lambda}^*(s))\widetilde{\check{P}}(s)] \\
&\quad + (X^{1,\varepsilon}(s) + X^{2,\varepsilon}(s))\Big(f_x(s) + f_y(s)Y^1(s) + f_z(s)(Y^1(s)\sigma_x(s) + Z^1(s)) \\
&\quad + \int_G f_k(s)l(e)(Y^1(s)\beta_x(s,e) + R^1(s,e))\lambda(\mathrm{d}e)\Big) + f_z(s)Y^1(s)\widetilde{\mathbb{E}}[\widetilde{\sigma}_\nu(s;\widetilde{X}^*(s))(\widetilde{X}^{1,\varepsilon}(s) + \widetilde{X}^{2,\varepsilon}(s))] \\
&\quad + \int_G f_k(s)l(e)Y^1(s)\widetilde{\mathbb{E}}[\widetilde{\beta}_\nu(s,e;\widetilde{X}^*(s))(\widetilde{X}^{1,\varepsilon}(s) + \widetilde{X}^{2,\varepsilon}(s))]\lambda(\mathrm{d}e) \\
&\quad + \widetilde{\mathbb{E}}[f_{\mu_1}(s;\widetilde{\Lambda}^*(s))(\widetilde{X}^{1,\varepsilon}(s) + \widetilde{X}^{2,\varepsilon}(s))] + \widetilde{\mathbb{E}}[f_{\mu_2}(s;\widetilde{\Lambda}^*(s))\widetilde{Y^1}(s)(\widetilde{X}^{1,\varepsilon}(s) + \widetilde{X}^{2,\varepsilon}(s))] \\
&\quad + \frac{1}{2}(X^{1,\varepsilon}(s))^2\Big(f_y(s)Y^2(s) + f_z(s)\Big(Y^1(s)\sigma_{xx}(s) + 2Y^2(s)\sigma_x(s) + Z^2(s)\Big) \\
&\quad + \int_G f_k(s)l(e)\big(Y^1(s)\beta_{xx}(s,e) + Y^2(s)(2\beta_x(s,e) + (\beta_x(s,e))^2) + R^2(s,e)\big)\lambda(\mathrm{d}e)\Big) \\
&\quad + \frac{1}{2}\Big(f_z(s)Y^1(s)\widetilde{\mathbb{E}}[\sigma_{\nu a}(s;\widetilde{X}^*(s))(\widetilde{X}^{1,\varepsilon}(s))^2] + \int_G f_k(s)l(e)Y^1(s)\widetilde{\mathbb{E}}[\beta_{\nu a}(s,e;\widetilde{X}^*(s))(\widetilde{X}^{1,\varepsilon}(s))^2]\lambda(\mathrm{d}e) \\
&\quad + \widetilde{\mathbb{E}}[f_{\mu_2}(s;\widetilde{\Lambda}^*(s))\widetilde{Y}^1(s)(\widetilde{X}^{1,\varepsilon}(s))^2]\Big) \\
&\quad + O(s)D^2f(s)O^\intercal(s)(X^{1,\varepsilon}(s))^2 + \frac{1}{2}\widetilde{\mathbb{E}}\Big[\Big(f_{\mu_2 a_2}(s;\widetilde{\Lambda}^*(s))(\widetilde{Y}^1(s))^2 + f_{\mu_1 a_1}(s;\widetilde{\Lambda}^*(s))\Big)(\widetilde{X}^{1,\varepsilon}(s))^2\Big] \\
&\quad + I_5(s),
\end{aligned} \tag{4.15}$$

where $\mathbb{E}[(\int_0^T I_5(s)\mathrm{d}s)^2] \le \varepsilon^2\rho(\varepsilon)$.

Consequently, combing all the above analyses and recall the definitions of $A_2(s), A_3(s)$, see Appendix, the equation (4.11) can read as

$$\begin{aligned}
\Delta Y(t) = &\int_t^T \Big(I_1(s) + (I_4(s) - I_3(s))\mathbb{1}_{\mathbb{E}_\varepsilon}(s) + I_5(s)\Big)\mathrm{d}s - \int_t^T \Delta Z(s)\mathrm{d}W(s) \\
&- \int_t^T\int_G \Delta K(s,e)N_\lambda(\mathrm{d}e,\mathrm{d}s) + o(\varepsilon),\ t\in[0,T].
\end{aligned} \tag{4.16}$$

It follows from (4.13), (4.14), (4.15) and Gronwall inequality that

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{s\in[0,T]}|\Delta Y(s)|^2 + \int_0^T |\Delta Z(s)|^2\mathrm{d}s + \int_0^T\int_G |\Delta K(s,e)|^2\lambda(\mathrm{d}e)\mathrm{d}s\Big] \\
&\le C\mathbb{E}\Big[(\int_0^T |(I_4(s) - I_3(s))\mathbb{1}_{\mathbb{E}_\varepsilon}(s) + I_5(s)|\mathrm{d}s)^2\Big] + o(\varepsilon^2) \le \varepsilon^2\rho(\varepsilon).
\end{aligned}$$

The proof is completed. □

Remark 4.4. If f does not depend on (y,z,k) and just depends on the law of $X^*(t)$, not on that of $Y^*(t)$, as well as $\beta \equiv 0$, then (4.2) is of the form

$$\begin{cases} -\mathrm{d}\breve{P}(t) = \Big(Y^1(t)\delta b(t) + Z^1(t)\delta\sigma(t) + \dfrac{1}{2}Y^2(t)(\delta\sigma(t))^2 + f(t, X^*(t), P_{X^*(t)}, v(t)) \\ \qquad\qquad - f(t, X^*(t), P_{X^*(t)}, u^*(t))\Big)\mathrm{d}t - \breve{Q}(t)\mathrm{d}W(t),\ t \in [0,T], \\ \breve{Y}(T) = 0, \end{cases} \tag{4.17}$$

which is just right the case investigated by Buckdahn *et al.* [5], and, accordingly, our stochastic maximum principle is consistent with theirs.

5. Stochastic maximum principle

In this section, the main result of this paper – SMP is proved.

5.1. Hamiltonian function

We define

$$\begin{aligned} &H(t,x,y,z,k,\nu,\mu,v;p,q,P) \\ &\quad = pb(t,x,\nu,v) + q\sigma(t,x,\nu,v) + \frac{1}{2}P\Big(\sigma(t,x,\nu,v) - \sigma(t,X^*(t),P_{X^*(t)},u^*(t))\Big)^2 \\ &\quad\ \ + f\Big(t,x,y,z+p\Big(\sigma(t,x,\nu,v) - \sigma(t,X^*(t),P_{X^*(t)},u^*(t))\Big), \int_G k(e)l(e)\lambda(\mathrm{d}e), \mu, v\Big), \end{aligned} \tag{5.1}$$

$(t,x,y,z,k,\nu,\mu,v,p,q,P) \in [0,T] \times \mathbb{R}^3 \times L^2(G,\mathscr{B}(G),\lambda) \times \mathcal{P}_2(\mathbb{R}) \times \mathcal{P}_2(\mathbb{R}^2) \times U \times \mathbb{R}^3$.

We now state the SMP.

Theorem 5.1. *Let the Assumptions (3.1)–(3.8) hold true, and, furthermore, let*

$$\widetilde{f}_{\mu_2}(t) = \left(\frac{\partial f}{\partial \mu}\right)_2 (t, \widetilde{X}^*(t), \widetilde{Y}^*(t), \widetilde{Z}^*(t), \int_G \widetilde{K}^*(t,e)l(e)\lambda(\mathrm{d}e), P_{(X^*(t),Y^*(t))}, \widetilde{u}^*(t); X^*(t), Y^*(t)) \geq 0,$$

$t \in [0,T]$, $\widetilde{P} \otimes P$*-a.s., and*

$$f_k(t) = \frac{\partial f}{\partial k}(t, X^*(t), Y^*(t), Z^*(t), \int_G K^*(t,e)l(e)\lambda(\mathrm{d}e), P_{(X^*(t),Y^*(t))}, u^*(t)) \geq 0,$$

$t \in [0,T], P$*-a.s. Suppose that* $u^*(\cdot)$ *is the optimal control and* (X^*, Y^*, Z^*, K^*) *is the corresponding solution of (1.1). Then there exist two pairs of stochastic processes* (Y^1, Z^1, R^1) *and* (Y^2, Z^2, R^2) *satisfying (3.31) and (3.33), respectively, such that*

$$\begin{aligned} &H(t, X^*(t), Y^*(t), Z^*(t), K^*(t,\cdot), P_{X^*(t)}, P_{(X^*(t),Y^*(t))}, v; Y^1(t), Z^1(t), Y^2(t)) \\ &\quad \geq H(t, X^*(t), Y^*(t), Z^*(t), K^*(t,\cdot), P_{X^*(t)}, P_{(X^*(t),Y^*(t))}, u^*(t); Y^1(t), Z^1(t), Y^2(t)), \\ &\qquad\qquad \forall v \in U,\ \text{a.e.},\ \text{a.s.} \end{aligned} \tag{5.2}$$

Proof. According to $J(v(\cdot)) = Y^v(0)$, (4.5) and $X^{1,\varepsilon}(0) = X^{2,\varepsilon}(0) = 0$ we have

$$J(u^\varepsilon(\cdot)) - J(u^*(\cdot)) = Y^\varepsilon(0) - Y^*(0) = \breve{P}(0) + o(\varepsilon) \geq 0. \tag{5.3}$$

Recall that

$$\begin{cases} -\mathrm{d}\breve{P}(t) = \Big\{ f_y(t)\breve{P}(t) + f_z(t)\breve{Q}(t) + \displaystyle\int_G f_k(t)l(e)\breve{K}(t,e)\lambda(\mathrm{d}e) + \widetilde{\mathbb{E}}[f_{\mu_2}(t;\widetilde{\Lambda}^*(t))\widetilde{\breve{P}}(t)] \\ \qquad + (A_1(t) + \Delta f(t))\mathbb{1}_{E_\varepsilon}(t)\Big\}\mathrm{d}t - \breve{Q}(t)\mathrm{d}W(t) - \displaystyle\int_G \breve{K}(t,e)N_\lambda(\mathrm{d}e,\mathrm{d}t),\ t \in [0,T], \\ \breve{P}(T) = 0, \end{cases} \tag{5.4}$$

which, however, inspires us to consider the dual mean-field SDE with jumps:

$$\begin{cases} d\Upsilon(t) = \Big(f_y(t)\Upsilon(t) + \widetilde{\mathbb{E}}[\widetilde{f}_{\mu_2}(t)\widetilde{\Upsilon}(t)]\Big)\mathrm{d}t + f_z(t)\Upsilon(t)\mathrm{d}W(t) + \displaystyle\int_G f_k(t)l(e)\Upsilon(t)N_\lambda(\mathrm{d}e,\mathrm{d}t),\ t \in [0,T], \\ \Upsilon(0) = 1. \end{cases} \tag{5.5}$$

Applying Itô's formula to $\Upsilon(t)\breve{P}(t),\ t \in [0,T]$, it follows

$$\breve{P}(0) = \mathbb{E}\Big[\int_0^T \Upsilon(s)(A_1(s) + \Delta f(s))\mathbb{1}_{E_\varepsilon}(s)\mathrm{d}s\Big]. \tag{5.6}$$

But, P-a.s.

$$\Upsilon(s) > 0,\ s \in [0,T]. \tag{5.7}$$

In fact, consider the auxiliary SDE with jumps

$$\begin{cases} d\Upsilon^1(t) = f_y(t)\Upsilon^1(t)\mathrm{d}t + f_z(t)\Upsilon^1(t)\mathrm{d}W(t) + \displaystyle\int_G f_k(t)l(e)\Upsilon^1(t)N_\lambda(\mathrm{d}e,\mathrm{d}t),\ t \in [0,T], \\ \Upsilon^1(0) = 1. \end{cases} \tag{5.8}$$

Denote $\Delta\Upsilon(t) = \Upsilon^1(t) - \Upsilon(t)$, then from Itô's formula we have

$$\begin{aligned} d((\Delta\Upsilon(t))^+)^2 = {}& 2\mathbb{1}_{\{\Delta\Upsilon(t)>0\}}\Delta\Upsilon(t)\Big\{ f_y(t)\Delta\Upsilon(t) - \widetilde{\mathbb{E}}[\widetilde{f}_{\mu_2}(t)\widetilde{\Upsilon}(t)] + f_z(t)\Delta\Upsilon(t)\mathrm{d}W(t) \\ & + \int_G f_k(t)l(e)\Delta\Upsilon(t)N_\lambda(\mathrm{d}e,\mathrm{d}t)\Big\} + \mathbb{1}_{\{\Delta\Upsilon(t)>0\}}(f_z(t))^2(\Delta\Upsilon(t))^2\mathrm{d}t \\ & + \int_G \mathbb{1}_{\{\Delta\Upsilon(t)>0\}}(f_k^-(t))^2(\Delta\Upsilon(t-))^2|l(e)|^2N(\mathrm{d}e,\mathrm{d}t), \end{aligned} \tag{5.9}$$

where $f_k^-(t) = \frac{\partial f}{\partial k}(s,\Pi^*(s-),P_{\Lambda^*(s)},u^*(s))$.

Thanks to [28] (see p. 87), the assumption $f_k(t) \geq 0,\ t \in [0,T]$ and $l(e) > 0,\ e \in G$ allows to show $\Upsilon^1(s) > 0,\ s \in [0,T],\ P$-a.s. Moreover, notice $\widetilde{f}_{\mu_2}(s) \geq 0,\ s \in [0,T],\ \widetilde{P} \otimes P$-a.s., hence, for $s \in [0,T]$, $\widetilde{\mathbb{E}}[\widetilde{f}_{\mu_2}(s)\widetilde{\Upsilon}^1(s)] \geq 0$, and then

$$\begin{aligned}\mathbb{E}[((\Delta\Upsilon(t))^+)^2] \leq & \mathbb{E}\Big[\int_0^t 2\mathbb{1}_{\{\Delta\Upsilon(s)>0\}}\Delta\Upsilon(s)\{f_y(s)\Delta\Upsilon(s)+\widetilde{\mathbb{E}}[\widetilde{f}_{\mu_2}(s)\Delta\widetilde{\Upsilon}(s)]\}\mathrm{d}s\Big]\\ &+\mathbb{E}\Big[\int_0^t \mathbb{1}_{\{\Delta\Upsilon(s)>0\}}(f_z(s))^2(\Delta\Upsilon(s))^2\mathrm{d}s\Big]\\ &+\mathbb{E}\Big[\int_0^t\int_G \mathbb{1}_{\{\Delta\Upsilon(s)>0\}}(f_k(s))^2(\Delta\Upsilon(s))^2|l(e)|^2\lambda(\mathrm{d}e)\mathrm{d}s\Big]\\ \leq & C\mathbb{E}\Big[\int_0^t((\Delta\Upsilon(s))^+)^2\mathrm{d}s\Big].\end{aligned} \tag{5.10}$$

Then the Gronwall lemma implies $\Delta\Upsilon(s)=\Upsilon^1(s)-\Upsilon(s)\leq 0,\ s\in[0,T]$. On the other hand, the assumption $f_k(t)\geq 0, t\in[0,T]$ can show $\Upsilon^1(s)>0, s\in[0,T],\ P$-a.s. Hence, $\Upsilon(s)\geq\Upsilon^1(s)>0,\ s\in[0,T],\ P$-a.s. Combining (5.3), (5.6) and (5.7) we have the desired result. □

Remark 5.2. If our model does not contain jump term, and the coefficient f just depends on the law $P_{X^*(t)}$, but not the joint law $P_{(X^*(t),Y^*(t))}$, then $\widetilde{f}_{\mu_2}(t)=f_k(t)=0,\ t\in[0,T]$. In this case our SMP is consistent with that of Buckdahn *et al.* [5].

6. Application to mean-field game

In this section we investigate an illustrative mean-field game without jumps to state our theoretical results. Suppose there are two types of investment possibilities in a market:

i) a risk-free security (*e.g.* a bond): $\mathrm{d}s_0(t)=r(t)S_0(t)\mathrm{d}t,\ S_0(0)>0$, where $r(t)$ is a bounded deterministic function;
ii) a risky security (*e.g.* a stock):

$$\begin{cases}\mathrm{d}s_1(t)=S_1(t)\mu(t)\mathrm{d}t+S_1(t)\sigma(t)\mathrm{d}W(t),\\ S_1(0)=s_1>0,\end{cases}\tag{6.1}$$

where $\mu(t)\neq 0,\ \sigma(t)\neq 0,\ \beta(t,e)$ are bounded deterministic function. We assume $\mu(t)>r(t),\ t\in[0,T]$.

Let $\theta(t)=(\theta_0(t),\theta_1(t))\in\mathbb{R}^2$ be the number of units possessed at time t the risk-free and the risky security. If the portfolio is self-finance, the wealth dynamics can be described as

$$\begin{cases}\mathrm{d}X(t)=\Big\{r(t)X(t)+(\mu(t)-r(t))v(t)\Big\}\mathrm{d}t+\sigma(t)v(t)\mathrm{d}W(t),\\ X(0)=x_0>0.\end{cases}\tag{6.2}$$

where $v(t):=\theta_1(t)S_1(t)$, see [30] for more details.

Next we introduce the corresponding mean-field game. Assume there exist N-individual agents in the market. The own control of i-th agent v_i has instantaneous and immediate effects on its own state. The controls of all other agents $v_j,\ i\neq j$ don't yield an immediate effect on the i-th state, but in some averaged manners. Inspired by Lions' works, we can assume that the wealth dynamics of the i-th agent is

$$\begin{cases}\mathrm{d}X^{i,v_i}(t)=\Big\{r^i(t)X^{i,v_i}(t)+(\mu^i(t)-r^i(t))v^i(t)+g^i(\bar{\nu}^N(t))\Big\}\mathrm{d}t+\sigma^i(t)v^i(t)\mathrm{d}W^i(t),\\ X^{i,v_i}(0)=x_0^i>0,\end{cases}\tag{6.3}$$

where $\bar{\nu}^N(t)=\frac{1}{N}\sum_{i=1}^{N}\delta_{X^{i,v_i}(s)}$, δ_y denotes the Dirac measure in y; $g^i:\mathbb{R}\to\mathbb{R}$ is a twice continuously differentiable function.

The i-th agent wants to minimize the utility $Y^{i,v_i}(0)$ resulting from v^i:

$$Y^{i,v_i}(t)=E^{\mathcal{F}_t}\Big[X^{i,v^i}(T)+\int_t^T f^i(s,X^{i,v_i}(s),\bar{\bar{\nu}}^N(s),v^i(s))\mathrm{d}s\Big], \tag{6.4}$$

where $\bar{\bar{\nu}}^N(s)=\frac{1}{N}\sum_{i=1}^{N}\delta_{(X^{i,v_i}(s),Y^{i,v_i}(s))}$. We suppose that the running cost function f^i depends not only on the immediate wealth X^{i,v_i}, but also on the average value of all the joint pair (X^{i,v_i},Y^{i,v_i}).
Thanks to [13], the recursive utility $Y^{i,v_i}(t)$ solves the following BSDE:

$$\begin{cases}\mathrm{d}Y^{i,v_i}(t)=-f^i(t,X^{i,v_i}(t),\bar{\bar{\nu}}^N(t),v^i(t))\mathrm{d}t+Z^{i,v_i}(t)\mathrm{d}W^i(t),\\ \ Y^{i,v_i}(t)=X^{i,v_i}(T).\end{cases} \tag{6.5}$$

Consequently, our system can be described as

$$\begin{cases}\mathrm{d}X^{i,v_i}(t)=\Big\{r^i(t)X^{i,v_i}(t)+(\mu^i(t)-r^i(t))v^i(t)+g^i(\bar{\nu}^N(t))\Big\}\mathrm{d}t+\sigma^i(t)v^i(t)\mathrm{d}W^i(t),\ t\in[0,T],\\ \mathrm{d}Y^{i,v_i}(t)=-f^i(t,X^{i,v_i}(t),\bar{\bar{\nu}}^N(t),v^i(t))\mathrm{d}t+Z^{i,v_i}(t)\mathrm{d}W^i(t),\ t\in[0,T],\\ \ X^{i,v_i}(0)=x^i,\ Y^{i,v_i}(T)=X^{i,v_i}(T).\end{cases} \tag{6.6}$$

Now we suppose the game is symmetric, *i.e.*, $r^i(\cdot)=r(\cdot),\ \mu^i(\cdot)=\mu(\cdot),v^i(\cdot)=v(\cdot),g^i(\cdot)=g(\cdot),W^i(\cdot)=W(\cdot),f^i(\cdot)=f(\cdot),x^i=x$. As $N\to\infty$, the system (6.6) can be characterized as a mean-field system

$$\begin{cases}\mathrm{d}X^v(t)=\Big\{r(t)X^v(t)+(\mu(t)-r(t))v(t)+g(P_{X^v(t)})\Big\}\mathrm{d}t+\sigma(t)v(t)\mathrm{d}W(t),\ t\in[0,T],\\ \mathrm{d}Y^v(t)=-f(t,X^v(t),P_{(X^v(t),Y^v(t))},v(t))\mathrm{d}t+Z^v(t)\mathrm{d}W(t),\ t\in[0,T],\\ \ X^v(0)=x,\ Y^v(T)=X^v(T).\end{cases} \tag{6.7}$$

In order to give a more concise form of our SMP, we suppose $g(P_{X^v(t)})=\psi(E[\varphi(X^v(t))])$ and $f(t,X^v(t),P_{(X^v(t),Y^v(t))},v(t))=\alpha(t)X^v(t)+\beta(t)v(t)+\psi\Big(E[\phi(X^v(t),Y^v(t))]\Big)$, where $\psi,\varphi:\mathbb{R}\to\mathbb{R},\ \phi:\mathbb{R}^2\to\mathbb{R}$ are twice continuously differentiable functions with bounded derivatives; $\alpha(\cdot),\beta(\cdot)$ are deterministic functions and bounded. Note that our method is suitable for the general case of f and g.

From Example 2.1, we know

$$\begin{aligned}g_\nu(P_{X^*(t)};a)&=\psi'(E[\varphi(X^*(t))])\varphi'(a),\ a\in\mathbb{R},\\ g_{\nu a}(P_{X^*(t)};a)&=\psi'(E[\varphi(X^*(t))])\varphi''(a),\ a\in\mathbb{R}.\end{aligned} \tag{6.8}$$

Define $h(P_{X^*(t),Y^*(t)})=\psi\Big(E[\phi(X^*(t),Y^*(t))]\Big)$. Similar to Example 2.1 we also have

$$\begin{aligned}h_{\mu_1}(P_{(X^*(t),Y^*(t))};a_1,a_2)&=\psi'(E[\phi(X^*(t),Y^*(t))])\phi_{a_1}(a_1,a_2),\ a_1,a_2\in\mathbb{R},\\ h_{\mu_1a_1}(P_{(X^*(t),Y^*(t))};a_1,a_2)&=\psi'(E[\phi(X^*(t),Y^*(t))])\phi_{a_1a_1}(a_1,a_2),\ a_1,a_2\in\mathbb{R},\\ h_{\mu_2}(P_{(X^*(t),Y^*(t))};a_1,a_2)&=\psi'(E[\phi(X^*(t),Y^*(t))])\phi_{a_2}(a_1,a_2),\ a_1,a_2\in\mathbb{R},\\ h_{\mu_2a_2}(P_{(X^*(t),Y^*(t))};a_1,a_2)&=\psi'(E[\phi(X^*(t),Y^*(t))])\phi_{a_2a_2}(a_1,a_2),\ a_1,a_2\in\mathbb{R}.\end{aligned} \tag{6.9}$$

Corollary 6.1. *Let Assumption (3.1), Assumption (3.2) and Assumption (3.3) hold true, and let* $\psi'(a)\phi_{a_2}(a_1,a_2) \geq 0\ \ a, a_1, a_2 \in \mathbb{R}$. *By* $u^*(\cdot)$ *and* (X^*, Y^*, Z^*) *we denote the optimal control and optimal trajectory. Then there exist two functions* Y^1 *and* Y^2 *such that, for* $t \in [0,T]$,

$$Y^1(t)(\mu(t) - r(t))(v - u^*(t)) + \frac{1}{2}Y^2(t)\sigma^2(t)(v - u^*(t))^2 \geq 0, \tag{6.10}$$

where Y^1 *and* Y^2 *satisfy the following backward ordinary differential equations, separately,*

$$\begin{cases} \mathrm{d}Y^1(t) = -\Big\{Y^1(s)\Big[\psi'(E[\phi(X^*(t),Y^*(t))])E[\phi_{a_2}(X^*(t),Y^*(t))] + r(t) \\ \qquad\qquad + \psi'(E[\varphi(X^*(t))])E[\varphi'(X^*(t))]\Big] + \alpha(t) \\ \qquad\qquad + \psi'(E[\phi(X^*(t),Y^*(t))])E[\phi_{a_1}(X^*(t),Y^*(t))]\Big\}\mathrm{d}t, \quad t \in [0,T], \\ Y^1(T) = 1. \end{cases} \tag{6.11}$$

$$\begin{cases} \mathrm{d}Y^2(t) = -\Big\{Y^2(t)\Big[\psi'(E[\phi(X^*(t),Y^*(t))])E[\phi_{a_2}(X^*(t),Y^*(t))] + 2r(t)\Big] \\ \qquad\qquad + Y^1(t)\Big[\psi'(E[\varphi(X^*(t))])E[\varphi''(X^*(t))]\Big] + \psi'(E[\phi(X^*(t),Y^*(t))])\cdot \\ \qquad\qquad \Big[E[\phi_{a_1a_1}(X^*(t),Y^*(t))] + (Y^1(t))^2E[\phi_{a_2a_2}(X^*(t),Y^*(t))]\Big], t \in [0,T], \\ Y^2(T) = 0. \end{cases} \tag{6.12}$$

Appendix A.

A.1 Some notations

The aim of this subsection is to collect some notations used in this paper, in particular, in Section 4:

$$\begin{aligned} M(t) &= Y^1(t)(X^{1,\varepsilon}(t) + X^{2,\varepsilon}(t)) + \frac{1}{2}Y^2(t)(X^{1,\varepsilon}(t))^2, \\ A_1(t) &= Y^1(t)\delta b(t) + Z^1(t)\delta\sigma(t) + \frac{1}{2}Y^2(t)(\delta\sigma(t))^2, \\ A_2(t) &= \Big(Y^1(t)b_x(t) + Z^1(t)\sigma_x(t) + \int_G R^1(t,e)\beta_x(t,e)\lambda(\mathrm{d}e) - F(t)\Big)(X^{1,\varepsilon}(t) + X^{2,\varepsilon}(t)) \\ &\quad + \widetilde{\mathbb{E}}\Big[\Big(Y^1(t)b_\nu(t;\widetilde{X}^*(t)) + Z^1(t)\sigma_\nu(t;\widetilde{X}^*(t)) + \int_G R^1(t,e)\beta_\nu(t,e;\widetilde{X}^*(t))\lambda(\mathrm{d}e)\Big) \\ &\quad (\widetilde{X}^{1,\varepsilon}(t) + \widetilde{X}^{2,\varepsilon}(t))\Big], \end{aligned}$$

$$
\begin{aligned}
A_3(t) &= (X^{1,\varepsilon}(t))^2\Big(b_{xx}(t)Y^1(t)+\sigma_{xx}(t)Z^1(t)+\int_G\beta_{xx}(t,e)R^1(t,e)\lambda(\mathrm{de})+2Y^2(t)b_x(t)\\
&\quad+Y^2(t)(\sigma_x(t))^2+\int_G Y^2(t)(\beta_x(t,e))^2\lambda(\mathrm{de})+2Z^2(t)\sigma_x(t)+\int_G R^2(t,e)\big(2\beta_x(s,e)\\
&\quad+(\beta(t,e))^2\big)\lambda(\mathrm{de})-G(t)\Big)+\widetilde{\mathbb{E}}\Big[\Big(Y^1(t)b_{\nu a}(t;\widetilde{X}^*(t))+Z^1(t)\sigma_{\nu a}(t;\widetilde{X}^*(t))\\
&\quad+\int_G R^1(t,e)\beta_{\nu a}(t,e;\widetilde{X}^*(t))\lambda(\mathrm{de})\Big)(\widetilde{X}^{1,\varepsilon}(t))^2\Big],\\
A_4(t) &= X^{1,\varepsilon}(t)\Big\{Y^1(t)\delta b_x(t)+Z^1(t)\delta\sigma_x(t)+Y^2(t)\Big(\delta b(t)+\delta\sigma(t)\sigma_x(t)\Big)+Z^2(t)\delta\sigma(t)\Big\}\\
&\quad+\widetilde{\mathbb{E}}\Big[\widetilde{X}^{1,\varepsilon}(t)\Big\{Y^1(t)\delta b_\nu(t;\widetilde{X}^*(t))+Z^1(t)\delta\sigma_\nu(t;\widetilde{X}^*(t))+Y^2(t)\delta\sigma(t)\sigma_\nu(t;\widetilde{X}^*(t))\Big\}\Big],\\
A_5(t) &= Y^2(t)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[b_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]+\frac{1}{2}Y^2(t)\Big(\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]\Big)^2\\
&\quad+Y^2(t)\sigma_x(t)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]+Z^2(t)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]\\
&\quad+\frac{1}{2}\int_G\Big(Y^2(t)\Big(\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]\Big)^2+Y^2(t)\beta_x(t,e)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\\
&\qquad\widetilde{X}^{1,\varepsilon}(t)]\Big)\lambda(\mathrm{de})+\int_G\Big(R^2(t,e)(\beta_x(t,e)+1)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]\Big)\lambda(\mathrm{de}),\\
B_1(t) &= Y^1(t)\delta\sigma(t),\\
B_2(t) &= (Y^1(t)\sigma_x(t)+Z^1(t))(X^{1,\varepsilon}(t)+X^{2,\varepsilon}(t))+Y^1(t)\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))(\widetilde{X}^{1,\varepsilon}(t)+\widetilde{X}^{2,\varepsilon}(t))],
\end{aligned}
$$

$$
\begin{aligned}
B_3(t) &= (Y^1(t)\sigma_{xx}(t)+2Y^2(t)\sigma_x(t)+Z^2(t))(X^{1,\varepsilon}(t))^2+Y^1(t)\widetilde{\mathbb{E}}[\sigma_{\nu a}(t,\widetilde{X}^*(t))(\widetilde{X}^{1,\varepsilon}(t))^2],\\
B_4(t) &= (Y^1(t)\delta\sigma_x(t)+Y^2(t)\delta\sigma(t))X^{1,\varepsilon}(t)+\widetilde{\mathbb{E}}[Y^1(t)\delta\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)],\\
B_5(t) &= Y^2(t)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\sigma_\nu(t,\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)],\\
C_2(t,e) &= \Big(Y^1(t)\beta_x(t,e)+R^1(t,e)\Big)(X^{1,\varepsilon}(t)+X^{2,\varepsilon}(t))+Y^1(t)\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))(\widetilde{X}^{1,\varepsilon}(t)+\widetilde{X}^{2,\varepsilon}(t))],\\
C_3(t,e) &= \Big(Y^1(t)\beta_{xx}(t,e)+Y^2(t)\big(2\beta_x(t,e)+(\beta_x(t,e))^2\big)+R^2(t,e)\Big)(X^{1,\varepsilon}(t))^2\\
&\quad+Y^1(t)\widetilde{\mathbb{E}}[\beta_{\nu a}(t,e;\widetilde{X}^*(t))(\widetilde{X}^{1,\varepsilon}(t))^2],\\
C_4(t,e) &= Y^2(t)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]+\frac{1}{2}Y^2(t)(\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)])^2\\
&\quad+Y^2(t)\beta_x(t,e)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)],
\end{aligned}
$$

and we denote

$$
\begin{aligned}
&A(s)=A_1(s)\mathbb{1}_{E_\varepsilon}(s)+A_2(s)+\frac{1}{2}A_3(s),\quad C(s,e)=C_2(s,e)+\frac{1}{2}C_3(s,e),\\
&B(s)=B_1(s)\mathbb{1}_{E_\varepsilon}(s)+B_2(s)+\frac{1}{2}B_3(s).
\end{aligned}
$$

Here $F(t)$ and $G(t)$ are given in (3.31) and (3.33), respectively, and $C^-(s,e)$ denotes the time s for the stochastic processes in $C(s,e)$ instead by $s-$.

A.2 Proof of Lemma 4.3

As for i) of (4.8), by observing the structures of $A_4(s)$ and $A_5(s)$, we mainly work out the central ingredients, *i.e.*, those terms involving the derivatives of the coefficients with respect to the measure.

a$_1$) From the boundness of σ_ν, (3.4) and Dominated Convergence Theorem, it follows

$$\begin{aligned}&\mathbb{E}\bigg[\bigg(\int_0^T \mathbb{1}_{E_\varepsilon}(t)\widetilde{\mathbb{E}}[\widetilde{X}^{1,\varepsilon}(t)Z^1(t)\delta\sigma_\nu(t;\widetilde{X}^*(t))]\mathrm{d}t\bigg)^2\bigg] \leq \varepsilon\mathbb{E}\bigg[\int_0^T \mathbb{1}_{E_\varepsilon}(t)|Z^1(t)|^2\mathbb{E}[|X^{1,\varepsilon}(t)|^2]\mathrm{d}t\bigg]\\&\leq \varepsilon\mathbb{E}\bigg[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^2\bigg]\mathbb{E}\bigg[\int_0^T \mathbb{1}_{E_\varepsilon}(t)|Z^1(t)|^2\mathrm{d}t\bigg] \leq \varepsilon^2\rho_1(\varepsilon),\end{aligned}\tag{A.1}$$

where $\rho_1(\varepsilon) := \mathbb{E}\Big[\int_0^T \mathbb{1}_{E_\varepsilon}(t)(Z^1(t))^2\mathrm{d}t\Big] \to 0$, as $\varepsilon \to 0$.

a$_2$) According to Assumption (3.1), Hölder inequality and the estimates (3.26), (3.34), we obtain

$$\begin{aligned}&\mathbb{E}\bigg[\bigg(\int_0^T \mathbb{1}_{E_\varepsilon}(t)\widetilde{\mathbb{E}}[\widetilde{X}^{1,\varepsilon}(t)Y^2(t)\delta\sigma(t)\sigma_\nu(t;\widetilde{X}^*(t))]\mathrm{d}t\bigg)^2\bigg]\\&\leq C\varepsilon\mathbb{E}\bigg[\sup_{t\in[0,T]}|Y^2(t)|^2\sup_{t\in[0,T]}\Big(|1+X^*(t)+\{\mathbb{E}[|X^*(t)|^2]\}^{\frac12}+|u(t)|+|u^*(t)|\Big)^2\cdot\\&\qquad\qquad\int_0^T \mathbb{1}_{E_\varepsilon}(t)|\widetilde{\mathbb{E}}[\widetilde{X}^{1,\varepsilon}(t)\sigma_\nu(t;\widetilde{X}^*(t))]|^2\mathrm{d}t\bigg]\end{aligned}\tag{A.2}$$

$$\begin{aligned}&\leq C\varepsilon^{\frac32}\Big\{\mathbb{E}[\sup_{t\in[0,T]}|Y^2(t)|^8]\Big\}^{\frac14}\Big\{\mathbb{E}\Big[\sup_{t\in[0,T]}|1+|X^*(t)|^8+\mathbb{E}[|X^*(t)|^8]+|u(t)|^8+|u^*(t)|^8\Big]\Big\}^{\frac14}\cdot\\&\qquad\qquad\Big\{\mathbb{E}\int_0^T|\widetilde{\mathbb{E}}[\widetilde{X}^{1,\varepsilon}(t)\sigma_\nu(t;\widetilde{X}^*(t))]|^4\mathrm{d}t\Big\}^{\frac12}\\&\leq C\varepsilon^{\frac52}\rho(\varepsilon).\end{aligned}$$

a$_3$) Thanks to the boundness of σ_ν, (3.4) and (3.26), one can check

$$\begin{aligned}&\mathbb{E}\bigg[\Big(\int_0^T Y^2(t)(\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)])^2\mathrm{d}t\Big)^2\bigg]\\&\leq CE\Big[\Big(\int_0^T|Y^2(t)|\mathbb{E}[|X^{1,\varepsilon}(t)|]|\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|\mathrm{d}t)^2\Big]\\&\leq C\varepsilon\mathbb{E}\Big[\sup_{t\in[0,T]}|Y^2(t)|^2\cdot\int_0^T|\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|^2\mathrm{d}t\Big]\\&\leq C\varepsilon\Big\{\mathbb{E}\Big[\sup_{t\in[0,T]}|Y^2(t)|^4\Big]\Big\}^{\frac12}\Big\{\mathbb{E}[\int_0^T|\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|^4\mathrm{d}t]\Big\}^{\frac12}\\&\leq C\varepsilon^2\rho(\varepsilon).\end{aligned}\tag{A.3}$$

a_4) The Assumptions (3.1)–(3.2), (3.4) and (3.26)-ii) allow to show

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\int_0^T\int_G Y^2(t)\beta_x(t,e)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t)\widetilde{X}^{1,\varepsilon}(t))]\lambda(\mathrm{d}e)\mathrm{d}s\Big)^2\Big]\\
&\quad\le \mathbb{E}\Big[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^2\sup_{t\in[0,T]}|Y^2(t)|^2\Big(\int_0^T\int_G(1\wedge|e|)|\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|\lambda(\mathrm{d}e)\mathrm{d}t\Big)^2\Big]\\
&\quad\le C\mathbb{E}\Big[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^2\sup_{t\in[0,T]}|Y^2(t)|^2\Big(\int_0^T\int_G|\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t)\widetilde{X}^{1,\varepsilon}(t))]|^4\lambda(\mathrm{d}e)\mathrm{d}t\Big)^{\frac12}\Big]\\
&\quad\le C\Big\{\mathbb{E}[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^8]\Big\}^{\frac14}\Big\{\mathbb{E}[\sup_{t\in[0,T]}|Y^2(t)|^8]\Big\}^{\frac14}\Big\{\mathbb{E}\int_0^T\int_G|\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|^4\lambda(\mathrm{d}e)\mathrm{d}s\Big\}^{\frac12}\\
&\quad\le\varepsilon^2\rho(\varepsilon).
\end{aligned}\tag{A.4}$$

a_5) Notice that for each $e\in G$, $|\beta_x(t,e)|\le C(1\wedge|e|)\le C$ and $\lambda(G)<+\infty$, (3.4) and (3.26) imply

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\int_0^T\int_G R^2(t,e)(\beta_x(t,e)+1)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]\lambda(\mathrm{d}e)\mathrm{d}t\Big)^2\Big]\\
&\quad\le C\mathbb{E}\Big[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^2\Big(\int_0^T\int_G|R^2(t,e)\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|\lambda(\mathrm{d}e)\mathrm{d}t\Big)^2\Big]\\
&\quad\le C\Big\{\mathbb{E}\Big[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^8\Big]\Big\}^{\frac14}\Big\{\mathbb{E}\Big[\int_0^T\int_G|\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|^4\lambda(\mathrm{d}e)\mathrm{d}t\Big]\Big\}^{\frac12}\\
&\quad\le C\varepsilon^2\rho(\varepsilon).
\end{aligned}\tag{A.5}$$

Combining the above estimates a_1)–a_5), we have

$$\mathbb{E}\Big[\Big(\int_0^T|A_4(t)\mathbb{1}_{E_\varepsilon}(t)|\mathrm{d}t\Big)^2\Big]+\mathbb{E}\Big[\Big(\int_0^T|A_5(t)|\mathrm{d}t\Big)^2\Big]\le C\varepsilon^2\rho(\varepsilon).\tag{A.6}$$

ii) of (4.8) can be calculated with the similar argument.

Let us now turn to $C_4(s,e)$. Through analysing the definition of $C_4(s,e)$, in order to prove iii) in (4.8) we just need to estimate the following terms.

b_1) By (3.4), (3.26) and $\lambda(G)<+\infty$, one knows

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\int_0^T\int_G Y^2(t)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]\lambda(\mathrm{d}e)\mathrm{d}t\Big)^2\Big]\\
&\quad\le C\Big\{\mathbb{E}\Big[\sup_{t\in[0,T]}|Y^2(t)|^8\Big]\Big\}^{\frac14}\Big\{\mathbb{E}\Big[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^8\Big]\Big\}^{\frac14}\cdot\\
&\qquad\qquad\Big\{\mathbb{E}\Big[\int_0^T\int_G|\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t))]|^4\lambda(\mathrm{d}e)\mathrm{d}t\Big]\Big\}^{\frac12}\\
&\quad\le C\varepsilon^2\rho(\varepsilon).
\end{aligned}\tag{A.7}$$

b$_2$) On the other hand, from (3.4), (3.26) again, it also yields

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\int_0^T\int_G Y^2(t)\beta_x(t,e)X^{1,\varepsilon}(t)\widetilde{\mathbb{E}}[\beta_\nu(s,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]\lambda(\mathrm{d}e)\mathrm{d}t\Big)^2\Big]\\
&\leq CE\Big[\sup_{t\in[0,T]}|Y^2(t)|^2\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^2\Big(\int_0^T\int_G(1\wedge|e|)|\widetilde{\mathbb{E}}[\beta_\nu(s,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|\lambda(\mathrm{d}e)\mathrm{d}t\Big)^2\Big]\\
&\leq C\varepsilon^2\rho(\varepsilon).
\end{aligned}
\tag{A.8}
$$

According to the above estimates, iii) of (4.8) can be obtained.

We are now ready to investigate iv) of (4.8), *i.e.*,

$$
\mathbb{E}\Big[\int_0^T(|M(s)|^2+|B(s)|^2+\int_G|C(s,e)|^2\lambda(\mathrm{d}e))\mathbb{1}_{E_\varepsilon}(s)\mathrm{d}s\Big]\leq C\varepsilon\rho(\varepsilon).
$$

Recall the definitions of $M(s),B(s),C(s,e)$, it is feasible to consider some central estimates. Let us now show them one by one.

To begin with, the following two estimates are the need for proving $\mathbb{E}\Big[\int_0^T|B(s)|^2\mathbb{1}_{E_\varepsilon}(s)\mathrm{d}s\Big]\leq C\varepsilon\rho(\varepsilon)$. From (3.26) and the boundness of $\sigma_{\nu a}$, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\int_0^T\mathbb{1}_{E_\varepsilon}(t)|Y^1(t)\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|^2\mathrm{d}t\Big]\\
&\leq\mathbb{E}\Big[\sup_{t\in[0,T]}|Y^1(t)|^2\int_0^T\mathbb{1}_{E_\varepsilon}(t)|\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|^2\mathrm{d}t\Big]\\
&\leq\Big\{\mathbb{E}\big[\sup_{t\in[0,T]}|Y^1(t)|^4\big]\Big\}^{\frac12}\Big\{\mathbb{E}\Big[\big(\int_0^T\mathbb{1}_{E_\varepsilon}(t)|\widetilde{\mathbb{E}}\big[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)\big]|^2\mathrm{d}t)^2\Big]\Big\}^{\frac12}\\
&\leq C\varepsilon^{\frac12}\Big\{\mathbb{E}\Big[\int_0^T|\widetilde{\mathbb{E}}[\sigma_\nu(t;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}(t)]|^4dt\Big]\Big\}^{\frac12}\leq C\varepsilon^{\frac32}\rho(\varepsilon),
\end{aligned}
\tag{A.9}
$$

and

$$
\begin{aligned}
&\mathbb{E}\Big[\int_0^T\mathbb{1}_{E_\varepsilon}(t)|Y^1(t)|^2|\widetilde{\mathbb{E}}[\sigma_{\nu a}(t;\widetilde{X}^*(t))(\widetilde{X}^{1,\varepsilon}(t))^2]|^2\mathrm{d}t\Big]\\
&\leq\varepsilon\mathbb{E}\Big[\sup_{t\in[0,T]}|Y^1(t)|^2\Big]\mathbb{E}\Big[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^4\Big]\leq C\varepsilon^3.
\end{aligned}
\tag{A.10}
$$

What's more, let us show $\mathbb{E}\Big[\int_0^T\int_G|C(t,e)|^2\mathbb{1}_{E_\varepsilon}(t)\lambda(\mathrm{d}e)\mathrm{d}t\Big]\leq C\varepsilon\rho(\varepsilon)$. As in the preceding proof we are primarily concerned with expectation terms. As for the expectation term in $C_2(t,e)$, due to $\lambda(G)<+\infty$, (3.26) allows to show

$$\begin{aligned}
&\mathbb{E}\Big[\int_0^T\int_G \mathbb{1}_{E_\varepsilon}(t)|Y^1(t)\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}]|^2\lambda(\mathrm{de})\mathrm{d}t\Big]\\
&\le \mathbb{E}\Big[\sup_{t\in[0,T]}|Y^1(t)|^2\int_0^T\int_G \mathbb{1}_{E_\varepsilon}(t)|\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}]|^2\lambda(\mathrm{de})\mathrm{d}t\Big]\\
&\le \Big\{\mathbb{E}[\sup_{t\in[0,T]}|Y^1(t)|^4]\Big\}^{\frac12}\Big\{\mathbb{E}[(\int_0^T\int_G \mathbb{1}_{E_\varepsilon}(t)|\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}]|^2\lambda(\mathrm{de})\mathrm{d}t)^2]\Big\}^{\frac12}\\
&\le C\varepsilon^{\frac12}\Big\{\mathbb{E}[\int_0^T\int_G |\widetilde{\mathbb{E}}[\beta_\nu(t,e;\widetilde{X}^*(t))\widetilde{X}^{1,\varepsilon}]|^4\lambda(\mathrm{de})\mathrm{d}t)]\Big\}^{\frac12}\\
&\le C\varepsilon^{\frac32}\rho(\varepsilon).
\end{aligned}\tag{A.11}$$

As regards the expectation term in $C_3(t,e)$, the boundness of $\beta_{\nu a}$ can imply

$$\begin{aligned}
&\mathbb{E}\Big[\int_0^T\int_G \mathbb{1}_{E_\varepsilon}(t)|Y^1(t)|^2|\widetilde{\mathbb{E}}[\beta_{\nu a}(t,e;\widetilde{X}^*(t))(\widetilde{X}^{1,\varepsilon}(t))^2]|^2\lambda(\mathrm{de})\mathrm{d}t\Big]\\
&\quad\le \varepsilon\Big\{\mathbb{E}[\sup_{t\in[0,T]}|Y^1(t)|^4]\Big\}^{\frac12}\Big\{\mathbb{E}[\sup_{t\in[0,T]}|X^{1,\varepsilon}(t)|^8]\Big\}^{\frac12}\le C\varepsilon^3.
\end{aligned}\tag{A.12}$$

Finally, the proof of $\mathbb{E}\Big[\int_0^T|M(s)|^2\mathbb{1}_{E_\varepsilon}(s)\mathrm{d}s\Big]\le\varepsilon\rho(\varepsilon)$ is analogous to that of $\mathbb{E}\Big[\int_0^T|B(s)|^2\mathbb{1}_{E_\varepsilon}(s)\mathrm{d}s\Big]\le\varepsilon\rho(\varepsilon)$. So here we omit it. Combining all the above estimates, we can get iv) of (4.8). The proof is completed.

A.3 The second order expansion of $I_3(s)$

Making the first-order expansion of f and according to the definitions of $M(s), B_2(s), B_3(s), C(s,e)$, we obtain

$$\begin{aligned}
I_3(s) &= f_y(s)\breve{P}(s)+f_z(s)\breve{Q}(s)+\int_G f_k(s)\breve{K}(s,e)l(e)\lambda(\mathrm{de})+\widetilde{\mathbb{E}}[f_{\mu_2}(s;\widetilde{\Lambda}^*(s))\widetilde{\breve{P}}(s)]\\
&\quad+(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))\Big(f_x(s)+f_y(s)Y^1(s)+f_z(s)(Y^1(s)\sigma_x(s)+Z^1(s))\\
&\quad+\int_G f_k(s)l(e)(Y^1(s)\beta_x(s,e)+R^1(s,e))\lambda(\mathrm{de})\Big)+f_z(s)Y^1(s)\widetilde{\mathbb{E}}[\sigma_\nu(s;\widetilde{X}^*(s))(\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))]\\
&\quad+\widetilde{\mathbb{E}}[f_{\mu_1}(s;\widetilde{\Lambda}^*(s))(\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))]+\widetilde{\mathbb{E}}[f_{\mu_2}(s;\widetilde{\Lambda}^*(s))\widetilde{Y}^1(s)(\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))]\\
&\quad+\frac12(X^{1,\varepsilon}(s))^2\Big(f_y(s)Y^2(s)+f_z(s)(Y^1(s)\sigma_{xx}(s)+2Y^2(s)\sigma_x(s)+Z^2(s))\\
&\quad+\int_G f_k(s)l(e)(Y^1(s)\beta_{xx}(s,e)+Y^2(s)(2\beta_x(s,e)+(\beta_x(s,e))^2+R^2(s,e)))\lambda(\mathrm{de})\Big)\\
&\quad+\frac12 f_z(s)Y^1(s)\widetilde{\mathbb{E}}[\sigma_{\nu a}(s;\widetilde{X}^*(s))(\widetilde{X}^{1,\varepsilon}(s))^2]+\frac12\int_G f_k(s)l(e)Y^1(s)\widetilde{\mathbb{E}}[\beta_{\nu a}(s,e;\widetilde{X}^*(s))(\widetilde{X}^{1,\varepsilon}(s))^2]\lambda(\mathrm{de})\\
&\quad+\frac12\widetilde{\mathbb{E}}[f_{\mu_2}(s;\widetilde{\Lambda}^*(s))\widetilde{Y}^1(s)(\widetilde{X}^{1,\varepsilon}(s))^2]+\Delta(s),
\end{aligned}\tag{A.13}$$

where

$$\begin{aligned}\Delta(s)=&\int_0^1(f_x^\theta(s)-f_x(s))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))\mathrm{d}\theta+\int_0^1(f_y^\theta(s)-f_y(s))(\breve{P}(s)+M(s))\mathrm{d}\theta\\&+\int_0^1(f_z^\theta(s)-f_z(s))(\breve{Q}(s)+B_2(s)+\frac{1}{2}B_3(s))\mathrm{d}\theta+\int_0^1\int_G(f_k^\theta(s)-f_k(s))l(e)(\breve{K}(s,e)\\&+C(s,e))\lambda(\mathrm{d}e)\mathrm{d}\theta+\int_0^1\widetilde{\mathbb{E}}[(f_{\mu_1}^\theta(s;\widetilde{\Lambda}^*(s))-f_{\mu_1}(s;\widetilde{\Lambda}^*(s)))(\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))]\mathrm{d}\theta\\&+\int_0^1\widetilde{\mathbb{E}}[(f_{\mu_2}^\theta(s;\widetilde{\Lambda}^*(s))-f_{\mu_2}(s;\widetilde{\Lambda}^*(s)))(\widetilde{\breve{P}}(s)+\widetilde{M}(s))]\mathrm{d}\theta,\end{aligned}\tag{A.14}$$

and we define, for $0<\varrho<1$, and $l=x,y,z,k$,

$$\begin{aligned}f_l^\varrho(s):=&\frac{\partial f}{\partial l}(s,X^*(s)+\varrho(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s)),Y^*(s)+\varrho(\breve{P}(s)+M(s)),\\&Z^*(s)+\varrho(\breve{Q}(s)+B_2(s)+\frac{1}{2}B_3(s)),\int_G(K^*(s,e)+\varrho(\breve{K}(s,e)+C(s,e)))l(e)\lambda(\mathrm{d}e),\\&P_{(X^*(s)+\varrho(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s)),Y^*(s)+\varrho(\breve{P}(s)+M(s)))},u^*(s)).\end{aligned}$$

$f_{\mu_1}^\theta(s;\widetilde{\Lambda}^*(s))$ and $f_{\mu_2}^\theta(s;\widetilde{\Lambda}^*(s))$ can be understood in the same sense.

Let us now focus on $\Delta(s)$. To start with, we argue that the first term on the right hand side can be written as

$$\begin{aligned}\bullet\quad&\int_0^1(f_x^\theta(s)-f_x(s))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))\mathrm{d}\theta=\frac{1}{2}(X^{1,\varepsilon}(s))^2\Big(f_{xx}(s)+f_{xy}(s)Y^1(s)\\&+f_{xz}(s)(Y^1(s)\sigma_x(s)+Z^1(s))+\int_G f_{xk}(s)l(e)(Y^1(s)\beta_x(s,e)+R^1(s,e))\lambda(\mathrm{d}e)\Big)+I_1(s),\end{aligned}\tag{A.15}$$

where $\mathbb{E}[\int_0^T|I_1(s)|^2\mathrm{d}s]\le\varepsilon^2\rho(\varepsilon)$.

In fact, according to Taylor expansion one has

$$\begin{aligned}&\int_0^1(f_x^\theta(s)-f_x(s))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))\mathrm{d}\theta=\frac{1}{2}\Big\{f_{xx}(s)(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))^2+f_{xy}(s)(X^{1,\varepsilon}(s)\\&+X^{2,\varepsilon}(s))(\breve{P}(s)+M(s))+f_{xz}(s)(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))(\breve{Q}(s)+B_2(s)+\frac{1}{2}B_3(s))\\&+\int_G f_{xk}(s)l(e)(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))(\breve{K}(s,e)+C(s,e))\lambda(\mathrm{d}e)\Big\}+\Theta_1(s)+\Theta_2(s),\end{aligned}\tag{A.16}$$

where

$$\begin{aligned}\Theta_1(s)=&\widetilde{\mathbb{E}}[f_{x\mu_1}(s;\widetilde{\Lambda}^*(s))(\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))](X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))\\&+\widetilde{\mathbb{E}}[f_{x\mu_2}(s;\widetilde{\Lambda}^*(s))(\widetilde{\breve{P}}(s)+\widetilde{M}(s))](X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s));\end{aligned}$$

$$
\begin{aligned}
\Theta_2(s)=\int_0^1\theta\mathrm{d}\theta\int_0^1 d\rho\Big\{&(f_{xx}^{\rho\theta}(s)-f_{xx}(s))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))^2\\
&+(f_{xy}^{\rho\theta}(s)-f_{xy}(s))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))(\breve{P}(s)+M(s))\\
&+(f_{xz}^{\rho\theta}(s)-f_{xz}(s))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))(\breve{Q}(s)+B_2(s)+\frac{1}{2}B_3(s))\\
&+\int_G(f_{xk}^{\rho\theta}(s)-f_{xk}(s))l(e)(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))(\breve{K}(s,e)+C(s,e))\lambda(\mathrm{d}e)\Big\}.
\end{aligned}
$$

For proving (A.15), we will show six auxiliary estimates:

$$
\begin{aligned}
&\text{i) }\mathbb{E}[\int_0^T|f_{xx}(s)(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))^2-f_{xx}(s)(X^{1,\varepsilon}(s))^2|^2\mathrm{d}s]\le\varepsilon^2\rho(\varepsilon);\\
&\text{ii) }\mathbb{E}[\int_0^T|f_{xy}(s)(\breve{P}(s)+M(s))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))-f_{xy}(s)Y^1(s)(X^{1,\varepsilon}(s))^2|^2\mathrm{d}s]\le\varepsilon^2\rho(\varepsilon);\\
&\text{iii) }\mathbb{E}[\int_0^T|f_{xz}(s)(\breve{Q}(s)+B_2(s)+\frac{1}{2}B_3(s))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))\\
&\qquad\qquad-f_{xz}(s)(Y^1(x)\sigma_x(s)+Z^1(s))(X^{1,\varepsilon}(s))^2|^2\mathrm{d}s]\le\varepsilon^2\rho(\varepsilon);\\
&\text{iv) }\mathbb{E}[\int_0^T\int_G|f_{xk}(s)l(e)(\breve{K}(s,e)+C(s,e))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))\\
&\qquad\qquad-f_{xk}(s)l(e)(Y^1(x)\beta_x(s,e)+R^1(s,e))(X^{1,\varepsilon}(s))^2|^2\lambda(\mathrm{d}e)\mathrm{d}s]\le\varepsilon^2\rho(\varepsilon);\\
&\text{v) }\mathbb{E}[\int_0^T|\Theta_1(s)|^2\mathrm{d}s]\le\varepsilon^2\rho(\varepsilon);\quad\text{vi) }\mathbb{E}[\int_0^T|\Theta_2(s)|^2\mathrm{d}s]\le\varepsilon^2\rho(\varepsilon).
\end{aligned}
\tag{A.17}
$$

The proofs of i)–iii) are very analogous to that of iv). So it is now the main work to estimate iv)–vi) in (A.17). As for iv), we are just concerned with the following expectation term because the other terms can be dealt with similarly. Notice the boundness of f_{xk} and (3.26), we get

$$
\begin{aligned}
&\mathbb{E}\Big[\int_0^T\int_G|f_{xk}(s)l(e)X^{1,\varepsilon}(s)Y^1(s)\widetilde{\mathbb{E}}[\beta_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^2\lambda(\mathrm{d}e)\mathrm{d}s\Big]\\
&\le C\mathbb{E}\Big[\sup_{s\in[0,T]}|X^{1,\varepsilon}(s)|^2\cdot\sup_{s\in[0,T]}|Y^1(s)|^2\cdot\int_0^T\int_G(1\wedge|e|)|\widetilde{\mathbb{E}}[\beta_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^2\lambda(\mathrm{d}e)\mathrm{d}s\Big]\\
&\le C\Big\{\mathbb{E}\Big[\sup_{s\in[0,T]}|X^{1,\varepsilon}(s)|^8\Big]\Big\}^{\frac14}\Big\{\mathbb{E}\Big[\sup_{s\in[0,T]}|Y^1(s)|^8\Big]\Big\}^{\frac14}\\
&\times\Big\{\mathbb{E}\Big[\int_0^T\int_G|\widetilde{\mathbb{E}}[\beta_\nu(s,e;\widetilde{X}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e)\mathrm{d}s\Big]\Big\}^{\frac12}\\
&\le\varepsilon^2\rho(\varepsilon).
\end{aligned}
\tag{A.18}
$$

Now observe Θ_1, the central work of proving v) is to estimate the following mean-field term,

$$\begin{aligned}
&\mathbb{E}\Big[\int_0^T |\widetilde{\mathbb{E}}[f_{x\mu_2}(s;\widetilde{Y}^*(s))\widetilde{Y}^1(s)\widetilde{X}^{1,\varepsilon}(s)]X^{1,\varepsilon}(s)|^2\mathrm{d}s\Big]\\
&\le C\mathbb{E}\Big[\sup_{s\in[0,T]}|X^{1,\varepsilon}(s)|^2\cdot\int_0^T|\widetilde{\mathbb{E}}[f_{x\mu_2}(s;\widetilde{\Lambda}^*(s))\widetilde{Y}^1(s)\widetilde{X}^{1,\varepsilon}(s)]|^2\mathrm{d}s\Big]\\
&\le C\Big\{\mathbb{E}\Big[\sup_{s\in[0,T]}|X^{1,\varepsilon}(s)|^4\Big]\Big\}^{\frac{1}{2}}\Big\{\mathbb{E}\Big[\int_0^T|\widetilde{\mathbb{E}}[f_{x\mu_2}(s;\widetilde{\Lambda}^*(s))\widetilde{Y}^1(s)\widetilde{X}^{1,\varepsilon}(s)]|^4\mathrm{d}s\Big]\Big\}^{\frac{1}{2}}\\
&\le \varepsilon^2\rho(\varepsilon).
\end{aligned}\tag{A.19}$$

The last step comes from (3.4) and (3.35).

For Θ_2, since the second-order derivatives of f is Lipschitz continuous, hence, it is easy to check that the power of ε for each term of Θ_2 is not less than $\frac{3}{2}$. So, vi) in (A.17) holds true.

Next, with the preceding argument, we also have

$$\begin{aligned}
\bullet\quad &\int_0^1(f_y^\theta(s)-f_y(s))(\breve{P}(s)+M(s))\mathrm{d}\theta=\frac{1}{2}(X^{1,\varepsilon}(s))^2Y^1(s)\Big(f_{yx}(s)+f_{yy}(s)Y^1(s)\\
&\quad+f_{yz}(s)(Y^1(s)\sigma_x(s)+Z^1(s))+\int_G f_{yk}(s)l(e)(Y^1(s)\beta_x(s,e)+R^1(s,e))\lambda(\mathrm{d}e)\Big)+I_2(s),\\
\bullet\quad &\int_0^1(f_z^\theta(s)-f_z(s))(\breve{Q}(s)+B_2(s)+\frac{1}{2}B_3(s))\mathrm{d}\theta\\
&\quad=\frac{1}{2}(X^{1,\varepsilon}(s))^2(Y^1(s)\sigma_x(s)+Q^1(s))\Big(f_{zx}(s)+f_{zy}(s)Y^1(s)+f_{zz}(s)(Y^1(s)\sigma_x(s)+Z^1(s))\\
&\quad+\int_G f_{zk}(s)l(e)(Y^1(s)\beta_x(s,e)+R^1(s,e))\lambda(\mathrm{d}e)\Big)+I_3(s),\\
\bullet\quad &\int_0^1\int_G(f_k^\theta(s)-f_k(s))l(e)(\breve{K}(s,e)+C(s,e))\lambda(\mathrm{d}e)\mathrm{d}\theta\\
&\quad=\frac{1}{2}(X^{1,\varepsilon}(s))^2\int_G l(e)(Y^1\beta_x(s,e)+R^1(s,e))\lambda(\mathrm{d}e)\Big(f_{kx}(s)+f_{ky}(s)Y^1(s)+f_{kz}(s)(Y^1(s)\sigma_x(s)\\
&\quad+Z^1(s))+\int_G f_{kk}(s)l(e)(Y^1(s)\beta_x(s,e)+R^1(s,e))\lambda(\mathrm{d}e)\Big)+I_4(s),
\end{aligned}\tag{A.20}$$

where $\mathbb{E}[\int_0^T|I_2(s)|^2+|I_3(s)|^2+|I_4(s)|^2\mathrm{d}s]\le\varepsilon^2\rho(\varepsilon)$.

In addition, we now switch to analysing the mean-field term $\int_0^1\widetilde{\mathbb{E}}[(f_{\mu_1}^\theta(s;\widetilde{\Lambda}^*(s))-f_{\mu_1}(s;\widetilde{\Lambda}^*(s)))$ $(\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))]\mathrm{d}\theta$. It follows Taylor expansion that

$$\bullet\quad\int_0^1\widetilde{\mathbb{E}}[(f_{\mu_1}^\theta(s;\widetilde{\Lambda}^*(s))-f_{\mu_1}(s;\widetilde{\Lambda}^*(s)))(\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))]\mathrm{d}\theta=\frac{1}{2}\widetilde{\mathbb{E}}[f_{\mu_1a_1}(s;\widetilde{\Lambda}^*(s))(\widetilde{X}^{1,\varepsilon}(s))^2]+I_5(s),\tag{A.21}$$

where

$$
\begin{aligned}
I_5(s) &= I_{5,1}(s) + I_{5,2}(s) + I_{5,3}(s),\\
I_{5,1}(s) &= \frac{1}{2}\widetilde{\mathbb{E}}\Big[f_{\mu_1 a_1}(s;\widetilde{\Lambda}^*(s))\Big((\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))^2-(\widetilde{X}^{1,\varepsilon}(s))^2\Big)\Big],\\
I_{5,2}(s) &= \frac{1}{2}\Big\{\widetilde{\mathbb{E}}\Big[(\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))\Big(f_{\mu_1 x}(s;\widetilde{\Lambda}^*(s))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))+f_{\mu_1 y}(s;\widetilde{\Lambda}^*(s))(\breve{P}(s)+M(s))\\
&\quad + f_{\mu_1 z}(s;\widetilde{\Lambda}^*(s))(\breve{Q}(s)+B_2(s)+\frac{1}{2}B_3(s))+\int_G f_{\mu_1 k}(s;\widetilde{\Lambda}^*(s))l(e)(\breve{K}(s,e)+C(s,e))\lambda(\mathrm{d}e)\\
&\quad + \widehat{\mathbb{E}}[f_{\mu_1\mu_1}(s;\widehat{\widetilde{\Lambda}}^*(s))(\widehat{X}^{1,\varepsilon}(s)+\widehat{X}^{2,\varepsilon}(s))]+\widehat{\mathbb{E}}[f_{\mu_1\mu_2}(s;\widehat{\widetilde{\Lambda}}^*(s))(\widehat{\breve{P}}(s)+\widehat{M}(s))]\Big)\Big]\Big\},\\
I_{5,3}(s) &= \int_0^1\theta\mathrm{d}\theta\int_0^1 d\rho\Big\{\widetilde{\mathbb{E}}\Big[(\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))\Big((f^{\rho\theta}_{\mu_1 x}(s;\widetilde{\Lambda}(s))-f_{\mu_1 x}(s;\widetilde{\Lambda}(s)))(X^{1,\varepsilon}(s)+X^{2,\varepsilon}(s))\\
&\quad + (f^{\rho\theta}_{\mu_1 y}(s;\widetilde{\Lambda}(s))-f_{\mu_1 y}(s;\widetilde{\Lambda}(s)))(\breve{P}(s)+M(s))\\
&\quad + (f^{\rho\theta}_{\mu_1 z}(s;\widetilde{\Lambda}(s))-f_{\mu_1 z}(s;\widetilde{\Lambda}(s)))(\breve{Q}(s)+B_2(s)+\frac{1}{2}B_3(s))\\
&\quad + \int_G(f^{\rho\theta}_{\mu_1 k}(s;\widetilde{\Lambda}(s))-f_{\mu_1 k}(s;\widetilde{\Lambda}(s)))l(e)(\breve{K}(s,e)+C(s,e))\lambda(\mathrm{d}e)\\
&\quad + \widehat{\mathbb{E}}[(f^{\rho\theta}_{\mu_1\mu_1}(s;\widehat{\widetilde{\Lambda}}^*(s))-f_{\mu_1\mu_1}(s;\widehat{\widetilde{\Lambda}}^*(s))(\widehat{X}^{1,\varepsilon}(s)+\widehat{X}^{2,\varepsilon}(s))]\\
&\quad + \widehat{\mathbb{E}}[(f^{\rho\theta}_{\mu_1\mu_2}(s;\widehat{\widetilde{\Lambda}}^*(s))-f_{\mu_1\mu_2}(s;\widehat{\widetilde{\Lambda}}^*(s))(\widehat{\breve{P}}(s)+\widehat{M}(s))]\\
&\quad + (f^{\rho\theta}_{\mu_1 a_1}(s;\widetilde{\Lambda}(s))-f_{\mu_1 a_1}(s;\widetilde{\Lambda}(s)))(\widetilde{X}^{1,\varepsilon}(s)+\widetilde{X}^{2,\varepsilon}(s))\Big)\Big]\Big\}.
\end{aligned}
$$

We now want to show $\mathbb{E}[(\int_0^T I_5(s)\mathrm{d}s)^2]\le\varepsilon^2\rho(\varepsilon)$. Indeed, From the boundness of $f_{\mu_1 a_1}$ and (3.4), it is easy to get $\mathbb{E}[(\int_0^T I_{5,1}(s)\mathrm{d}s)^2]\le C\varepsilon^3$. Next let us prove $\mathbb{E}[(\int_0^T I_{5,2}(s)\mathrm{d}s)^2]\le\varepsilon^2\rho(\varepsilon)$. According to the structure of $I_{5,2}(s)$, we know that it is enough to only handle the following jump term and expectation term:

$$
\begin{aligned}
&\text{i) } \mathbb{E}\Big[\Big(\int_0^T\int_G\widetilde{\mathbb{E}}[f_{\mu_1 k}(s;\widetilde{\Lambda}^*(s))\widetilde{X}^{1,\varepsilon}(s)]l(e)(\breve{K}(s,e)+C(s,e))\lambda(\mathrm{d}e)\mathrm{d}s\Big)^2\Big]\le\varepsilon^2\rho(\varepsilon),\\
&\text{ii) } \mathbb{E}\Big[\Big(\int_0^T\widetilde{\mathbb{E}}\Big[\widetilde{X}^{1,\varepsilon}(s)\widehat{\mathbb{E}}[f_{\mu_1\mu_2}(s;\widetilde{X}^*(s),\widehat{Y}^*(s))(\widehat{\breve{P}}(s)+\widehat{M}(s))\Big]\mathrm{d}s\Big)^2\Big]\le\varepsilon^2\rho(\varepsilon).
\end{aligned}
$$

For i), from Hölder inequality, (4.3), (3.4), (3.6)-ii) with $\widetilde{\psi}_3(t)=f_{\mu_1 k}(s;\widetilde{\Lambda}^*(s))$, $\widetilde{\psi}_2(t)\equiv 1$ and the definition of $C(s,e)$, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\int_0^T\int_G\widetilde{\mathbb{E}}[f_{\mu_1 k}(s;\widetilde{\Lambda}^*(s))\widetilde{X}^{1,\varepsilon}(s)]l(e)(\breve{K}(s,e)+C(s,e))\lambda(\mathrm{d}e)\mathrm{d}s\Big)^2\Big]\\
&\le\mathbb{E}\Big[\int_0^T\int_G|l(e)|^2|\widetilde{\mathbb{E}}[f_{\mu_1 k}(s;\widetilde{\Lambda}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^2\lambda(\mathrm{d}e)\mathrm{d}s\cdot\int_0^T\int_G|\breve{K}(s,e)+C(s,e)|^2\lambda(\mathrm{d}e)\mathrm{d}s\Big]\\
&\le\Big\{\mathbb{E}\Big[(\int_0^T\int_G(1\wedge|e|)|\widetilde{\mathbb{E}}[f_{\mu_1 k}(s;\widetilde{\Lambda}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^2\lambda(\mathrm{d}e)\mathrm{d}s)^2\Big]\Big\}^{\frac{1}{2}}\Big\{\mathbb{E}\Big[(\int_0^T\int_G|\breve{K}(s,e)+C(s,e)|^2\lambda(\mathrm{d}e)\mathrm{d}s)^2\Big]\Big\}^{\frac{1}{2}}
\end{aligned}
$$

$$
\begin{aligned}
&\leq C\Big\{\mathbb{E}\Big[\int_0^T\int_G|\widetilde{\mathbb{E}}[f_{\mu_1 k}(s;\widetilde{\Lambda}^*(s))\widetilde{X}^{1,\varepsilon}(s)]|^4\lambda(\mathrm{d}e)\mathrm{d}s\Big]\Big\}^{\frac{1}{2}}\cdot\Big\{\mathbb{E}\Big[\Big(\int_0^T\int_G|\breve{K}(s,e)+C(s,e)|^2\lambda(\mathrm{d}e)\mathrm{d}s\Big)^2\Big]\Big\}^{\frac{1}{2}}\\
&\leq \varepsilon\rho(\varepsilon)\Big(\varepsilon\rho(\varepsilon)+\Big\{\mathbb{E}\Big[\Big(\int_0^T\int_G|C(s,e)|^2\lambda(\mathrm{d}e)\mathrm{d}s\Big)^2\Big]\Big\}^{\frac{1}{2}}\Big)\\
&\leq \varepsilon\rho(\varepsilon)(\varepsilon\rho(\varepsilon)+\varepsilon)=\varepsilon^2\rho(\varepsilon).
\end{aligned}
$$

Let us calculate ii). From (3.35)-ii), (4.3), (3.4) and the boundness of $f_{\mu_1\mu_2}$, it is easy to check

$$
\begin{aligned}
&\mathbb{E}\Big[\Big(\int_0^T\widetilde{\mathbb{E}}\Big[\widetilde{X}^{1,\varepsilon}(s)\widehat{\mathbb{E}}[f_{\mu_1\mu_2}(s;\widehat{\widetilde{\Lambda}}^*(s))(\widehat{\breve{P}}(s)+\widehat{M}(s))]\Big]\mathrm{d}s\Big)^2\Big]\\
&\quad\leq\Big\{\mathbb{E}\Big[\sup_{t\in[0,T]}|X^{1,\varepsilon}(s)|^4\Big]\Big\}^{\frac{1}{2}}\Big\{\mathbb{E}\widetilde{\mathbb{E}}\Big[\int_0^T|\widehat{\mathbb{E}}[f_{\mu_1\mu_2}(s;\widehat{\widetilde{\Lambda}}^*(s))(\widehat{\breve{P}}(s)+\widehat{M}(s))]|^4\Big]\Big\}^{\frac{1}{2}}\\
&\quad\leq\varepsilon^2\rho(\varepsilon).
\end{aligned}
$$

According to the Assumption (3.3), (3.4), (3.26), (3.35) and (4.3) and the above two estimates, applying Hölder inequality again, it yields $\mathbb{E}[(\int_0^T I_{5,2}(s)\mathrm{d}s)^2]\leq\varepsilon^2\rho(\varepsilon)$. Besides, thanks to the continuous property of the second-order derivatives of f, one can check the validity of $\mathbb{E}[(\int_0^T I_{5,3}(s)\mathrm{d}s)^2]\leq\varepsilon^2\rho(\varepsilon)$ easily. Hence, we prove $\mathbb{E}[(\int_0^T I_5(s)\mathrm{d}s)^2]\leq\varepsilon^2\rho(\varepsilon)$.

Finally, following the above argument, it also yields that

$$
\begin{aligned}
\bullet\ &\int_0^1\widetilde{\mathbb{E}}[(f^\theta_{\mu_2}(s;\widetilde{\Lambda}^*(s))-f_{\mu_2}(s;\widetilde{\Lambda}^*(s)))(\widetilde{\breve{P}}(s)+\widetilde{M}(s))]\mathrm{d}\theta\\
&\qquad=\frac{1}{2}\widetilde{\mathbb{E}}[f_{\mu_2 a_2}(s;\widetilde{\Lambda}^*(s))(\widetilde{Y}^1(s))^2(\widetilde{X}^{1,\varepsilon}(s))^2]+I_6(s),
\end{aligned}
\tag{A.22}
$$

here $I_6(s)$ satisfying $\mathbb{E}[(\int_0^T I_6(s)\mathrm{d}s)^2]\leq\varepsilon^2\rho^*(\varepsilon)$.
Combining (A.13), (A.15), (A.20), (A.21), (A.22), we have (4.15).

Acknowledgements. The authors thank the associate editor and the referees for their helpful comments.

References

[1] A. Bensoussan, Maximum principle and dynamic programming approaches to the optimal control of partially observed diffusions. *Stochastics* **9** (1983) 169–222.
[2] J.M. Bismut, Conjugate convex functions in optimal stochastic control. *J. Math. Analysis Appl.* **44** (1973) 384–404.
[3] R. Buckdahn, B. Djehiche and J. Li, A general stochastic maximum principle for SDEs of mean-field type. *Appl. Math. Optim.* **64** (2011) 197–216.
[4] R. Buckdahn, B. Djehiche, J. Li and S. Peng, Mean-field backward stochastic differential equations: a limit approach. *Ann. Probab.* **37** (2009) 1524–1565.
[5] R. Buckdahn, J. Li and J. Ma, A stochastic maximum principle for general mean-field systems. *Appl. Math. Optim.* **74** (2016) 507–534.
[6] R. Buckdahn, J. Li and S. Peng, Mean-field backward stochastic differential equations and related partial differential equations. *Stoch. Proc. Appl.* **119** (2009) 3133–3154.
[7] R. Buckdahn, J. Li, S. Peng and C. Rainer, Mean-field stochastic differential equations and associated PDEs. *Ann. Probab.* **45** (2014) 824–874.
[8] P. Cardaliaguet, Notes on Mean Field Games (from P.-L. Lions' lectures at Collège de France) (2013). Available at https://www.ceremade.dauphine.fr/~cardalia.
[9] R. Carnoma and F. Delarue, Probabilistic analysis of mean-field games. *SIAM J. Control Optim.* **51** (2013) 2705–2734.
[10] R. Carnoma and F. Delarue, Forward-backward stochastic differential equations and controlled McKean-Vlasov dynamics. *Ann. Probab.* **43** (2015) 2647–2700.
[11] R. Carnoma, F. Delarue and A. Lachapelle, Control of McKean-Vlasov dynamic versus mean field games. *Math. Financ. Econ.* **7** (2013) 131–166.

[12] J.F. Chassagneux, D. Crisan and F. Delarue, A probabilistic approach to classical solutions of the master equation for large population equilibria (2015). Available at http://arxiv.org/abs/1411.3009v2.

[13] N. El Karoui, S. Peng and M. Quenez, Backward Stochastic differential equation in finance. *Math. Finance* **7** (1997) 1–71.

[14] N.C. Framstad, B. Oksendal and A. Sulem, Sufficient stochastic maximum principle for the optimal control of jump diffusions and applications to finance. *J. Optim. Theory Appl.* **121** (2004) 77–98.

[15] T. Hao and J. Li, Mean-field SDEs with jumps and nonlocal integral-PDEs. *Nonlinear Differ. Equ. Appl.* **23** (2017) 1–51.

[16] U.G. Haussmann, A Stochastic Maximum Principle for Optimal Control of Diffusions. Longman Scientific and Technical, Harlow, England (1986).

[17] M. Hu, Stochastic global maximum principle for optimization with recursive utilities. *Proba. Unce. Quanti. Risk* **2** (2017) 1–20.

[18] M. Kac, Foundations of kinetic theory. In *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability* **3** (1956) 171–197.

[19] H.J. Kushner, Necessary conditions for continuous parameter stochastic optimization problems. *SIAM J. Control Optim.* **10** (1972) 550–565.

[20] J.M. Lasry and P.L. Lions, Mean field games. *Jpn. J. Math.* **2** (2007) 229–260.

[21] J. Li, Mean-field forward and backward SDEs with jumps. Associated nonlocal quasi-linear integral-PDEs. *Stoch. Proc. Appl.* **128** (2018) 3118–3180.

[22] J. Li and Q. Wei, Stochastic differential games for fully coupled FBSDEs with jumps. *Appl. Math. Optim.* **71** (2013) 411–448.

[23] J. Li and Q. Wei, L^p estimates for fully coupled FBSDEs with jumps. *Stoch. Proc. Appl.* **124** (2014) 1582–1611.

[24] P.L. Lions, Cours au Collège de France : Théorie des jeu à champs moyens (2013). Available at http://www.college-de-france.fr/default/EN/all/equ[1]der/audiovideo.jsp.

[25] Q. Lu, H. Zhang and X. Zhang, Second order optimality conditions for optimal control problems of stochastic evolution equations (2018). Available at https://arxiv.org/abs/1811.07337.

[26] S. Peng, A general stochastic maximum principle for optimal control problems. *SIAM J. Control Optim.* **28** (1990) 966–979.

[27] S. Peng, Open problems on backward stochastic differential equations, In S. Chen, X. Li, J. Yong, X.Y. Zhou (Eds.), Contorl of distributed parameter and stochstic systems. Kluwer Acad Pub., Boston (1998) 265–273.

[28] R. Situ, Theory of stochastic differential equations with jumps and applications. Springer (2012).

[29] S. Tang and X. Li, Necessary conditions for optimal control of stochastic systems with random jumps. *SIAM J. Control Optim.* **32** (1994) 1447–1475.

[30] Z. Wu, A general maximum principle for optimal control problems of forward-backward stochastic control systems. *Automatica* **49** (2013) 1473–1480.

[31] J. Yong, Optimality variational principle for controlled forward-backward stochastic differential equations with mixed initial-terminal conditions. *SIAM J. Control Optim.* **48** (2010) 4119–4156.

[32] J. Yong, X. Zhou, Stochastic controls. Springer Verlag, New York, NY (1999).

[33] H. Zhang and X. Zhang, Pointwise second-order necessary conditions for stochastic optimal controls, Part I: The case of convex control constraint. *SIAM J. Control Optim,* **53** (2015) 2267–2296.

[34] H. Zhang and X. Zhang, Pointwise second-order necessary conditions for stochastic optimal controls, Part II: The general case. *SIAM J. Control Optim.* **53** (2015) 2841–2875.

[35] H. Zhang and X. Zhang, Some results on pointwise second-order necessary conditions for stochastic optimal controls. *Sci. China Math.* **59** (2016) 227–238.