ENTROPY-REGULARIZED MEAN-VARIANCE PORTFOLIO OPTIMIZATION WITH JUMPS

CHRISTIAN BENDER[1] AND NGUYEN TRAN THUAN[1,2]

Abstract. Motivated by the trade-off between exploitation and exploration in reinforcement learning, we study a continuous-time entropy-regularized mean-variance portfolio selection problem in the presence of jumps. We propose an exploratory SDE for the wealth process associated with multiple risky assets which exhibit Lévy jumps. In contrast to the existing literature, we study the limiting behavior of the natural discrete-time formulation of the wealth process associated with a randomized control in order to derive the continuous-time dynamics. We then show that an optimal distributional control of the continuous-time entropy-regularized exploratory mean-variance problem is still Gaussian despite being in jump models. Moreover, the respective optimal wealth process solves a linear SDE whose representation is explicitly obtained.

2020 Mathematics Subject Classification. Primary: 93E20, 60H30; Secondary: 60F05, 60G51.

Keywords. Entropy regularization, Lévy process, mean-variance portfolio optimization, reinforcement learning, weak convergence.

1. Introduction

1.1. **The problem.** The mean-variance (MV) portfolio optimization problem pioneered by Markowitz [27] is one of the most popular criteria in the portfolio selection theory due to its simple and natural formulation in dealing with the two important aspects of investment, namely, risk and return. In the MV model, investors aim to minimize the variance, which quantifies the risk, of the terminal wealth of their portfolios while targeting a prespecified expected value of the terminal wealth. This criterion therefore effectively reflects a trade-off between the risk and expected return in an intuitive way. After Markowitz's foundational works, the MV approach has attracted considerable attention with numerous extensions and applications. For example, among other works in the continuous-time setting when the financial market is driven by a multidimensional Brownian motion, Zhou and Li [38] investigate the MV problem in terms of stochastic linear-quadratic (LQ) optimization using an embedding method. After that, Li et al. [24] introduce the Lagrange multiplier method to transform the MV problem to an unconstrained stochastic LQ control problem so that standard techniques are applicable. As the literature on the MV criterion is vast, we refer the reader to [37] for a review on this topic.

[1]Department of Mathematics, Saarland University, Germany

[2]Department of Mathematics, Vinh University, 182 Le Duan, Vinh, Nghe An, Viet Nam

E-mail addresses: bender@math.uni-saarland.de, nguyen@math.uni-saarland.de, thuan.tr.nguyen@gmail.com.

Date: February 21, 2025.

The second author dedicates this article to Professor Trần Lộc Hùng for his inspiring life story.

The first preprint version of this manuscript has been made available in December 2023. To the best of our knowledge, it is the first paper discussing the exploratory control problem for reinforcement learning in continuous time with controlled jumps. In the meantime different aspects of reinforcement learning for jump diffusions have been addressed in [12] and in our recent work [4].

The classical model-based MV problem, where model assumptions are predescribed, has been fairly well investigated and quite completely solved in various settings with analytical solutions. To apply these results in practice, one usually needs to estimate model parameters based on historical data of the underlying asset prices accumulated during trading. Nevertheless, it is widely acknowledged that it is difficult to estimate those parameters with an applicable accuracy, and furthermore, classical optimal MV strategies frequently exhibit high sensitivity to those parameters which then might become practically sub-optimal due to estimation error.

In recent years, reinforcement learning (RL) methods, which have increasingly attracted more attention in quantitative finance, become a promising approach to overcome those practical difficulties. By and large, RL algorithms iteratively execute randomized controls for some period (or, episode) and apply the data which has been collected over the previous periods to update the unknown model parameters and the randomized control, see, e.g., [20, 21, 33] for RL algorithms in a continuous-time stochastic control setting. The randomization of the controls reflects the trade-off between exploration (learning the unknown investment environment) and exploitation (optimizing adaptively to the updated model parameters). Thus, RL algorithms can produce (nearly) optimal solutions without the need of statistically estimating the model parameters beforehand. The reader is referred to [16] for an overview to recent developments and applications of RL in finance.

The iterative construction of the randomized controls in the algorithms mentioned above relies on an entropy-regularized formulation of the stochastic control problem. Here, the entropy regularization rewards exploration and leads to the optimality of distribution-valued (or, relaxed) controls. Recently, Wang and Zhou [34] introduced such an entropy-regularized exploratory SDE framework for the MV problem in a Black–Scholes environment. To be more precise and for easier explanation, let us introduce some notations. Let $T > 0$ be a fixed finite time horizon and $W = (W_t)_{t\in[0,T]}$ a standard 1-dimensional Brownian motion. The exploratory SDE for the wealth process $X^\pi = (X^\pi_t)_{t\in[0,T]}$ under an admissible control $\pi = (\pi_t)_{t\in[0,T]}$, which is a *distribution-valued* stochastic process and where π_t is the probability density function of the exploration law at time t, is heuristically derived and has the following form

$$\mathrm{d}X^\pi_t = \mu_t b\mathrm{d}t + \sqrt{\mu_t^2 + \sigma_t^2}\, a\mathrm{d}W_t. \tag{1.1}$$

Here, the drift $b \in \mathbb{R}$ and volatility $a > 0$ are unknown constants, $\mu_t := \int_{\mathbb{R}} u\pi_t(u)\mathrm{d}u$ represents the mean and $\sigma_t^2 := \int_{\mathbb{R}} u^2\pi_t(u)\mathrm{d}u - \mu_t^2$ the variance of the distribution of exploration at time t. We refer to [34, 35] for the motivation and derivation of (1.1). To encourage and quantify the exploration process, Wang and Zhou [34] incorporate a *differential entropy* term to the objective function and the classical MV problem then becomes an *entropy-regularized exploratory MV problem*. The authors then prove that the optimal feedback distributional control is Gaussian with time-decaying variance. Moreover, via a simulation study it is also illustrated in [34] that the RL approach for solving the MV problem significantly improves some other methods such as the traditional maximum likelihood estimate (MLE) and the deep deterministic policy gradient (DDPG). The approach as in [34, 35] has been extended in various contexts, see, e.g., [10, 14, 36].

It is, however, widely acknowledged that models with jumps are more appropriate to describe the fluctuation of asset prices, see, e.g., [1, 8]. Following this direction, many researchers have extensively studied the classical MV problem and its variants in several jump models, see, e.g., [19, 25, 28] and the references therein. Then a question naturally arises: *How would the continuous-time entropy-regularized exploratory MV problem and its solutions be like if the asset*

prices exhibit jumps? To address this question, one first needs to describe the exploratory SDE with jumps for the respective wealth process.

In contrast to the models built upon the Brownian framework by Wang and Zhou [34] and by Wang et al. [35], where the exploratory SDE for the wealth/controlled process can be heuristically inferred from knowing its first two conditional moments only, models with jumps are quite involved as, in general, one has to test against various other functions rather than the linear and quadratic functions to detect the distributional behavior of jumps. In fact, these test functions essentially depend on the jump activities of the underlying asset price process. Hence, the derivation for the exploratory SDE based on first two moments in [34, 35] is seemingly not applicable for jump models, at least in a straightforward way. To deal with this problem, we exploit the linear dependence on controls of the wealth process and propose a different argument to derive the exploratory SDE.

1.2. **Our contributions and discussions.** Let $D \in \mathbb{N}$ and assume that the log-price process of D risky assets is a weak solution of an SDE driven by a D-dimensional Lévy process L. Here L includes, but not necessarily simultaneously, a Brownian motion W and an independent pure-jump Lévy process J, both are D-dimensional. Except the square integrability, there are no additional assumptions imposed on the Lévy measure.

1.2.1. *Continuous-time exploratory SDE with Lévy jumps.* To derive an exploratory SDE for the wealth process, we begin with a discrete-time dynamic of the wealth under an exploration procedure, see Section 3.2.1. In [34, 35], the authors first average out realizations of distributional controls on each discrete-time sub-interval using a law of large numbers, and then combine them all together to infer the dynamic on entire $[0, T]$. Here, unlike the argument in [34, 35], we first explicitly model randomized controls on discrete-time partitions of $[0, T]$ and identify a family of discrete-time integrators which incorporate the additional "exploration noise". To do that, we need to handle the additional randomness caused by exploration differently for the Brownian and for the jump component which can be roughly described as follows:

- For the Brownian part, thanks to the linear structure with respect to the control, one can (partially) separate the original randomness caused by the asset prices and the randomness caused by exploration in an appropriate way, see Section 3.2.3.
- For the jump component, we employ a suitable D^2-dimensional random measure to simultaneously capture both sources of randomness, see Section 3.2.4.

Then, by refining the discrete time points, we show in Theorem 3.5 below that the stochastic *integrators* of our discrete-time scheme converge in distribution to a multidimensional Lévy process. This limit theorem gives rise to a natural continuous-time formulation of the exploratory control problem with entropy regularization. Note that randomized controls on discrete-time grids have recently been considered in [11, 33] for diffusion models. However, [33, Theorem 2.2] investigates the limiting behavior of the cost of such controls and [11, Lemma 4] describes the convergence of the optimal control density, while we apply this discretization to infer the structure of the continuous-time "exploration noise".

We also remark that the heuristic passage to the limit in the existing literature [34, 35] only yields information about the conditional mean and covariance of the continuous-time controlled system. It, thus, allows for many different SDE representations, even in the case of no jumps, as discussed below. In contrast, our derivation identifies a specific SDE formulation, which we consider a natural choice for modeling exploration in the continuous-time framework. Indeed,

as discussed in Section 4.5 below, our formulation of the exploratory SDE, which is derived from discrete-time randomized controls, is closely related to the *sample state process*, which is a key object for the design of learning algorithms in the recent continuous-time RL literature [20, 21]. It can even be interpreted as a mathematically rigorous reformulation of this sample state process which avoids the use of an independent identically distributed sampling mechanism indexed by continuous time.

1.2.2. *Problem formulation in multidimensional setting.* We consider D risky assets with jumps and derive the continuous-time dynamics of the wealth process with exploration, see SDE (3.7) and Remark 3.6 for further discussion.

Let us compare our exploratory SDE with other works in the case of no jumps. Since we use a different argument, our exploratory SDE unsurprisingly takes a different form from (1.1) in [34]. If $D = 1$, then the dynamic of wealth under a distributional control π in our setting particularly becomes

$$\mathrm{d}X_t^\pi = \mu_t b \mathrm{d}t + \mu_t a \mathrm{d}W_t + \sigma_t a \mathrm{d}\mathcal{W}_t, \tag{1.2}$$

where $\mathcal{W}$ is a 1-dimensional Brownian motion independent of W. We notice that X^π in (1.1) and in (1.2) have the same distribution. However, differently from (1.1), in our SDE (1.2) the exploration randomness represented by $\mathcal{W}$ is separated from the noise W caused by asset prices. One also remarks that the SDE in form of (1.2) has been recently considered in [10, 36]. Nevertheless, when $D > 1$, the authors in [10, 36] use an additional 1-dimensional Brownian motion to model the exploration (i.e. $\mathcal{W}$ is 1-dimensional), while, according to our analysis, it suggests to use a D^2-dimensional Brownian motion (i.e. $\mathcal{W}$ is D^2-dimensional).

1.2.3. *Optimal distributional control and wealth process.* Following [34], we first use the Lagrange multiplier method to transform the exploratory MV problem to an entropy-regularized quadratic-loss control problem and then apply the dynamic programing principle to find its solutions.

We show in Theorem 4.10 that, despite the presence of jumps, among admissible distributional controls which are not necessarily in the feedback form, an optimal Gaussian control in feedback form can be obtained. As a feature of our approach, the respective optimal wealth process satisfies a *linear* SDE (see (4.23)) which allows us to find its expression in a closed-form (see (4.27) and [5]). As a consequence, the Lagrange multiplier is also explicitly obtained (see (4.28) and [5]). Moreover, the value function has a quadratic form with respect to the wealth variable whose coefficients are solutions to a system of partial integro-differential equations (PIDEs). In the special case of no jumps and $D = 1$ and with constant coefficients, our value function coincides, of course, with that in [34], see Example 4.16.

1.3. **Structure of the article.** In Section 2, we introduce the notation and recall the classical MV problem. The derivation of the continuous-time exploratory SDE with Lévy jumps is presented in Section 3. In Section 4, we study the entropy-regularized exploratory MV problem, investigate its closed-form solutions, and discuss the Lagrange multipliers. Section 5 is devoted to present the proof of weak convergence of the discrete-time integrators (Theorem 3.5).

2. Preliminaries

2.1. **Notations.** Let $D \in \mathbb{N} := \{1, 2, \ldots\}$. For $a, b \in \mathbb{R}$, we use the usual notations $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$. For $a < b$, let $\int_a^b := \int_{(a,b]}$. Notation log indicates the natural

logarithm. Sub-indexing a symbol by a label means the place where that symbol appears. We also use the conventions $\inf\emptyset := \infty$ and $\sum_{i\in\emptyset} = \int_\emptyset := 0$.

2.1.1. *Vector spaces and matrices.* Let $\|\cdot\|$ be the usual Euclidean norm and $(\mathbf{e}_d)_{d=1}^D$ the natural basis in $\mathbb{R}^D$. For $r > 0$, we set $B_D(r) := \{x \in \mathbb{R}^D : \|x\| < r\}$ and $B_D^c(r) := \mathbb{R}^D \backslash B_D(r)$.

All vectors are written in the column form. For a vector x we use the notation $x^{(i)}$ or $[x]^{(i)}$ to denote its i-th component. For a matrix A,

- $A^{(i,j)}$ or $[A]^{(i,j)}$ is the element in the i-th row and j-th column of A;
- if A is a $D \times D$ matrix, then $\mathbf{tr}[A]$, $\det(A)$, A^{-1} denote the trace, determinant and inverse of A respectively. Let $\mathrm{diag}(A) := \mathrm{diag}(A^{(1,1)}, \dots, A^{(D,D)})$ denote the diagonal matrix with diagonal entries $A^{(1,1)}, \dots, A^{(D,D)}$;
- the usual Euclidean/Frobenius norm of A is also denoted by $\|A\|$, i.e. $\|A\| := \sqrt{\mathbf{tr}[A^\mathsf{T} A]}$.

Notation I_D means the $D \times D$ identity matrix. We also use the following classes of matrices:

- $\mathbb{R}^{D\times D'}$ denotes the family of all real matrices with size $D \times D'$;
- $\mathbb{S}^D$ (resp. $\mathbb{S}_+^D$, $\mathbb{S}_{++}^D$) is the family of all symmetric (resp. positive semidefinite, positive definite) $A \in \mathbb{R}^{D\times D}$. For $A \in \mathbb{S}_+^D$, denote by $A^{\frac{1}{2}} \in \mathbb{S}_+^D$ the unique square root of A, i.e. $A^{\frac{1}{2}} A^{\frac{1}{2}} = A$. If $A \in \mathbb{S}_{++}^D$, then we let $A^{-\frac{1}{2}} := (A^{\frac{1}{2}})^{-1}$.
- $\mathcal{O}_D$ consists of all orthonormal $O \in \mathbb{R}^{D\times D}$, i.e. $O^\mathsf{T} O = I_D$.

For $A \in \mathbb{R}^{D\times D'}$, denote by $\mathrm{vec}(A)$ the vectorization of A defined as an element of $\mathbb{R}^{DD'}$ by stacking the columns of A on top of one another, i.e.

$$\mathrm{vec}(A) := (A^{(1,1)}, \dots, A^{(D,1)}, A^{(1,2)}, \dots, A^{(D,2)}, \dots, A^{(1,D')}, \dots, A^{(D,D')})^\mathsf{T}.$$

For (column) vectors $x_1, \dots, x_n$ with possibly different sizes, $\mathrm{vec}(x_1, \dots, x_n)$ means the vector obtained by stacking x_i on top of x_{i+1}, $1 \le i \le n-1$. To shorten notation at some places we also use the Kronecker product $\otimes\colon \mathbb{R}^D \times \mathbb{R}^{D'} \to \mathbb{R}^{DD'}$ defined by

$$x \otimes y := \mathrm{vec}(x^{(1)} y, \dots, x^{(D)} y).$$

One notices that the operator $\otimes$ is bilinear and $\|x \otimes y\| = \|x\|\|y\|$.

2.1.2. *Function spaces.* For a function $f\colon \mathbb{R}^D \to \mathbb{R}$, we use the following notations:

- $\|f\|_\infty := \sup_{x\in\mathbb{R}^D} |f(x)|$;
- ∂f and $\partial^2 f$ denote usual partial derivatives of f with respect to scalar components;
- ∇f and $\nabla^2 f$ denote the gradient and the Hessian of f respectively, and $\|\nabla f\|_\infty^2 := \sum_{d=1}^D \|\partial_d f\|_\infty^2$, $\|\nabla^2 f\|_\infty^2 := \sum_{d,d'=1}^D \|\partial_{d,d'}^2 f\|_\infty^2$, where partial derivatives $\partial_d f := \partial_{x^{(d)}} f$ and $\partial_{d,d'}^2 f := \partial_{x^{(d)}x^{(d')}}^2 f$;
- When f has several (multivariate) components, we use $\nabla_y f$ and $\nabla_{yy}^2 f$ to indicate the gradient and Hessian of f with respect to component y. If x is a scalar component and y is a multivariate component, then we write $\nabla_{xy}^2 := (\partial_{xy^{(1)}}^2, \dots, \partial_{xy^{(D)}}^2)^\mathsf{T}$.
- $\mathrm{supp}(f)$ stands for the support of f, i.e. the closure of $\{x \in \mathbb{R}^D : f(x) \neq 0\}$.

For $k = 1, 2, \dots$, denote by $C^k(\mathbb{R}^D)$ the family of all k times continuously differentiable functions on $\mathbb{R}^D$. $C_b^k(\mathbb{R}^D)$ consists of all bounded $f \in C^k(\mathbb{R}^D)$ with bounded derivatives (up to the k-th order) and $C_b^\infty(\mathbb{R}^D) := \cap_{k\ge1} C_b^k(\mathbb{R}^D)$. $C_c^k(\mathbb{R}^D)$ denotes the family of all $f \in C^k(\mathbb{R}^D)$ with compact support. We let $f \in C^{1,2}([0,T] \times \mathbb{R}^D)$ if f is (resp. twice) continuously differentiable with respect to $t \in [0,T]$ (resp. to $y \in \mathbb{R}^D$) and its partial derivatives are jointly continuous.

2.2. **Stochastic basis.** Let us fix a time horizon $T \in (0,\infty)$. Assume that $(\Omega, \mathcal{F}, \mathbb{P}; (\mathcal{F}_t)_{t\in[0,T]})$ satisfies the usual conditions, which means that $(\Omega, \mathcal{F}, \mathbb{P})$ is a compete probability space, the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ is right-continuous and $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets. This allows us to assume that every $(\mathcal{F}_t)_{t\in[0,T]}$-adapted local martingale has *càdlàg* (right-continuous with finite left limits) paths. For a random variable ξ, the expectation, variance, and conditional expectation given a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, if it exists under $\mathbb{P}$, is respectively denoted by $\mathbb{E}[\xi]$, $\mathbb{V}[\xi]$, and $\mathbb{E}[\xi|\mathcal{G}]$. We also use $\mathbf{L}_p(\mathbb{P}) := \mathbf{L}_p(\Omega, \mathcal{F}, \mathbb{P})$.

For a càdlàg process $X = (X_t)_{t\in[0,T]}$, we denote $\Delta X_t := X_t - X_{t-}$ for $t \in [0,T]$, where $X_{0-} := X_0$ and $X_{t-} := \lim_{t>s\uparrow t} X_s$ for $t \in (0,T]$. For a time index set $\mathbb{I} \subseteq [0,\infty)$ and for processes $X = (X_t)_{t\in\mathbb{I}}$, $Y = (Y_t)_{t\in\mathbb{I}}$, we write $X = Y$ to indicate that $X_t = Y_t$ for all $t \in \mathbb{I}$ a.s., and the same meaning applied when the relation "=" is replaced by some other standard relations such as "$\leq$", "$>$", etc.

We refer to [30] for unexplained notions such as semimartingales, (optional) quadratic covariation $[X, Y]$ and conditional quadratic covariation $\langle X, Y\rangle$ of semimartingales X, Y.

2.3. **Multidimensional Lévy process.** An $\mathbb{R}^D$-valued process $L = (L_t)_{t\in[0,T]}$ is called a Lévy process if it has independent and stationary increments, has càglàg paths with $L_0 = 0$ a.s. The distributional property of L is characterized by the Lévy–Khintchine formula (see, e.g., [2, Theorem 1.2.14]), for $t \in [0,T]$ and $u \in \mathbb{R}^D$,

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}u^\mathsf{T} L_t}] = \mathrm{e}^{-t\kappa(u)}$$

where the *characteristic exponent* κ is given, for $u \in \mathbb{R}^D$, by

$$\kappa(u) = -\mathrm{i}u^\mathsf{T} b + \frac{u^\mathsf{T} A u}{2} - \int_{z\neq 0} (\mathrm{e}^{\mathrm{i}u^\mathsf{T} z} - 1 - \mathrm{i}u^\mathsf{T} z \mathbb{1}_{\{\|z\|\leq 1\}})\nu(\mathrm{d}z).$$

The characteristic triplet (b, A, ν) associated with the canonical truncation function $h(z) := z\mathbb{1}_{\{\|z\|\leq 1\}}$ is deterministic and consists of the *drift coefficient* $b \in \mathbb{R}^D$, the *Gaussian covariance matrix* $A \in \mathbb{S}_+^D$, and the *Lévy measure* ν, i.e. a measure on $\mathcal{B}(\mathbb{R}^D\backslash\{0\})$ with $\int_{z\neq 0}(\|z\|^2 \wedge 1)\nu(\mathrm{d}z) < \infty$. We call L a *Gaussian Lévy process* if $\nu \equiv 0$, and call L a *purely non-Gaussian Lévy process* if $A = 0$.

2.4. **Classical continuous-time MV portfolio selection.** Assume that the price process of a risk-less asset $S^{(0)} = (S_t^{(0)})_{t\in[0,T]}$ and D risky assets $S = (S_t)_{t\in[0,T]}$ are governed by the following SDEs

$$\begin{aligned}
\mathrm{d}S_t^{(0)} &= rS_t^{(0)}\mathrm{d}t, \quad S_0^{(0)} = 1,\\
\mathrm{d}S_t^{(d)} &= S_{t-}^{(d)}\mathrm{d}R_t^{(d)}, \quad S_0^{(d)} := s_0^{(d)} > 0, \quad d = 1,\dots,D.
\end{aligned}$$

where the interest rate $r \geq 0$ is given and the (stochastic) log-price process R is described by (2.1) and (2.2) below. In the context of stock price modeling, it is natural to assume the condition $\Delta R^{(d)} > -1$ on the jump sizes, which ensures that the stock prices $S^{(d)}$ stay strictly positive. This condition is, however, not required to obtain the main results, and so we do not impose it.

An investment strategy in D risky assets is expressed by a predictable $\mathbb{R}^D$-valued process H where $H_t^{(d)}$ is the *discounted dollar amount* invested in the d-th risky asset at time $t-$, i.e instantly before time t. The resulting discounted wealth process $X^H = (X_t^H)_{t\in[0,T]}$ associated

with H can be written as

$$\mathrm{d}X_t^H = \sum_{d=1}^{D} H_t^{(d)}(\mathrm{d}R_t^{(d)} - r\mathrm{d}t) =: H_t^\intercal \mathrm{d}Y_t, \tag{2.1}$$

where $X_0^H = x_0 \in \mathbb{R}$ is the given initial wealth. From now we will work with the driving process Y and the discounted wealth X^H as in (2.1).

Assume that the log-price of D underlying (discounted) risky assets is represented by a càdlàg and adapted process $Y = (Y_t)_{t\in[0,T]}$ which is Markovian whose infinitesimal generator is given, for sufficiently smooth f, by

$$(\mathcal{L}_Y f)(y) = b(y)^\intercal \nabla f(y) + \frac{1}{2}\mathbf{tr}[A(y)\nabla^2 f(y)] + \int_{z\neq 0} \Big(f(y+\gamma(y)z) - f(y) - \nabla f(y)^\intercal \gamma(y) z\Big)\nu(\mathrm{d}z). \tag{2.2}$$

Here ν is a square integrable Lévy measure and the coefficients $b\colon \mathbb{R}^D \to \mathbb{R}^D$, $A \in \mathbb{S}_+^D$, and $\gamma\colon \mathbb{R}^D \to \mathbb{R}^{D\times D}$ satisfy standard assumptions which will be specified later in Section 3.1.

The classical Markowitz MV portfolio selection problem, parameterized by $\hat{z} \in \mathbb{R}$, is then formulated as

$$\begin{cases} \min_H \mathbb{V}[X_T^H] \\ \text{subject to } X^H \text{ given in (2.1) and } \mathbb{E}[X_T^H] = \hat{z}, \end{cases} \tag{2.3}$$

where the minimum is taken over admissible H which will be specified in our setting later. To deal with the constraint $\mathbb{E}[X_T^H] = \hat{z}$ in (2.3), we follow [38, 34] to consider the objective function parameterized by $w \in \mathbb{R}$,

$$\mathbb{V}[X_T^H] - 2w(\mathbb{E}[X_T^H] - \hat{z}),$$

which is equal to

$$\mathbb{E}[(X_T^H - w)^2] - (\hat{z} - w)^2.$$

Then, to solve (2.3), we consider the following *unconstrained* quadratic-loss minimization problem parameterized by w,

$$\begin{cases} \min_H \mathbb{E}[(X_T^H - w)^2] \\ \text{subject to } X^H \text{ given in (2.1)}. \end{cases} \tag{2.4}$$

Once (2.4) is solved with a minimizer $H^*(w)$, which depends on w, we let $\hat{w}$ be the value such that the constraint $\mathbb{E}\big[X_T^{H^*(\hat{w})}\big] = \hat{z}$ is satisfied. Then such an $H^*(\hat{w})$ solves the original problem (2.3), and $\hat{w}$ is called the *Lagrange multiplier*[1].

3. Exploratory SDE with Lévy jumps

3.1. **Setting.** Let us fix $D \in \mathbb{N}$ and set $E := \mathbb{R}^D\backslash\{0\}$. Let φ_D be a probability density of $\xi \sim \mathcal{N}(0, I_D)$ where $\mathcal{N}(0, I_D)$ is the D-dimensional Gaussian distribution with zero mean and covariance I_D.

For b, A, γ and ν appearing in (2.2) we assume throughout this article the following:

Assumption 3.1. The Lévy measure ν and coefficients $b\colon \mathbb{R}^D \to \mathbb{R}^D$, $a, \gamma\colon \mathbb{R}^D \to \mathbb{R}^{D\times D}$, $A := aa^\intercal \in \mathbb{S}_+^D$ satisfy:

(a) (Square integrability) ν is square integrable on E, i.e. $\int_E \|e\|^2 \nu(\mathrm{d}e) < \infty$;

[1]The Lagrange multiplier actually is $2\hat{w}$, but we use $\hat{w}$ to slightly simplify the presentation.

(b) (Growth condition) $\|b(x)\| + \|a(x)\| + \|\gamma(x)\| \leq C_1(1 + \|x\|)$ for all $x \in \mathbb{R}^D$;

(c) (Lipschitz condition) $\|b(x) - b(y)\| + \|a(x) - a(y)\| + \|\gamma(x) - \gamma(y)\| \leq C_2\|x - y\|$ for all $x, y \in \mathbb{R}^D$;

(d) (Non-degeneration) $\Sigma(y) := A(y) + \gamma(y) \int_E ee^\mathsf{T} \nu(\mathrm{d}e)\gamma(y)^\mathsf{T} \in \mathbb{S}^D_{++}$ for all $y \in \mathbb{R}^D$.

3.2. **Continuous-time dynamic of the wealth process with exploration: A heuristic approach.** Let $W = (W_t)_{t\in[0,T]}$ be a D-dimensional standard Brownian motion, and $J = (J_t)_{t\in[0,T]} \subseteq \mathbf{L}_2(\mathbb{P})$ a purely non-Gaussian Lévy process which is independent of W and has the following Lévy–Itô decomposition (see, e.g., [2, Theorem 2.4.26])

$$J_t := \int_0^t \int_E e\widetilde{N}(\mathrm{d}s, \mathrm{d}e).$$

Here $\widetilde{N}$ is the compensated Poisson random measure of J associated with Lévy measure ν. Since b, a, γ in Assumption 3.1 are sufficiently regular, it is known that the following SDE has a unique (strong) solution in $\mathbf{L}_2(\mathbb{P})$ (see, e.g., [22, Theorem 3.1]),

$$\mathrm{d}Y_t = b(Y_{t-})\mathrm{d}t + a(Y_{t-})\mathrm{d}W_t + \gamma(Y_{t-})\mathrm{d}J_t, \quad Y_0 = y_0 \in \mathbb{R}^D,$$

which admits $\mathcal{L}_Y$ provided in (2.2) as the Markov generator.

Let $\{\Pi_n\}_{n\geq 1}$ be a sequence of partitions of $[0, T]$, where $\Pi_n = \{0 =: t^n_0 < t^n_1 < \ldots < t^n_n := T\}$. Denote $\Delta t^n_i := t^n_i - t^n_{i-1}$ and assume that $|\Pi_n| := \max_{1\leq i\leq n} \Delta t^n_i \to 0$ as $n \to \infty$. To shorten the presentation at some places, for a process $(P_t)_{t\in[0,T]}$, we also use the notations

$$P_{n,i} := P_{t^n_i} \quad \text{and} \quad \Delta_{n,i}P := P_{n,i} - P_{n,i-1}.$$

With the convention $\sup \emptyset := 0$, we define

$$\sigma^n_t := \sup\{i \geq 1 : t^n_i \leq t\}, \quad t \in [0, T].$$

For each n, we obtain a process Y^n, which approximates Y along the partition Π_n, given by

$$Y^n_t := Y_0 + \sum_{i=1}^{\sigma^n_t} \Big(b(Y_{n,i-1})\Delta t^n_i + a(Y_{n,i-1})\Delta_{n,i}W + \gamma(Y_{n,i-1})\Delta_{n,i}J\Big), \quad t \in [0, T].$$

3.2.1. *Exploration procedure.* Our main idea is as follows: For $i = 1, \ldots, n$, we draw the control at time t^n_{i-1} from some distribution, which is chosen with the accumulative information available at time t^n_{i-1}. Once the distribution is fixed, the realization is independent of the rest. In addition, since any distribution on $\mathbb{R}^D$ can be represented as $F(\xi)$ for some measurable $F\colon \mathbb{R}^D \to \mathbb{R}^D$ and $\xi \sim \mathcal{N}(0, I_D)$, determining a distribution boils down to find such an F.

Let us specify this idea.

(i) Let $\Xi := \{\xi_{n,i}\}_{n\geq 1, 1\leq i\leq n}$ be a collection of i.i.d. random vectors in $\mathbb{R}^D$ with common distribution $\mathcal{N}(0, I_D)$ and probability density φ_D. Assume that Ξ is independent of (W, J). Family Ξ represents a new source of randomness caused from the exploration along with the randomness generated by (W, J). To capture the information flow, we define the filtration $\mathbb{F}^{\Pi_n} = (\mathcal{F}_{n,i})^n_{i=0}$ as follows

$$\mathcal{F}_{n,i} := \sigma\{(W_s, J_s) : 0 \leq s \leq t^n_i\} \vee \mathcal{G}_{n,i}, \quad \text{where} \quad \mathcal{G}_{n,i} := \sigma\big\{\xi_{n,j} : j \leq i\big\}, \mathcal{G}_{n,0} := \{\emptyset, \Omega\}.$$

(ii) $H\colon \Pi_n \times \Omega \times \mathbb{R}^D \to \mathbb{R}^D$ is admissible in the following sense (here $H_{n,i-1}$ stands for $H_{t^n_{i-1}}$):

(a) For each $i = 1, \ldots, n$, the map $(\omega, u) \mapsto H_{n,i-1}(\omega; u)$ is $\mathcal{F}_{n,i-1} \otimes \mathcal{B}(\mathbb{R}^D)$-measurable;

(b) One has $\mathbb{E}\big[\int_{\mathbb{R}^D} \|H_{n,i-1}(u)\|^2 \varphi_D(u)\mathrm{d}u\big] < \infty$;

(c) As proposed in [34], the exploration cost can be represented in terms of differential entropy which is assumed to be finite to encourage the exploration. Following this idea, we in addition assume that for each $i = 1, \dots, n$ and $\omega \in \Omega$, $H_{n,i-1}(\omega; \zeta)$ has a probability density $p^H_{n,i-1}(\omega; \cdot)$, where $\zeta \sim \mathcal{N}(0, I_D)$ is independent of $\mathcal{F}_{n,i-1}$, such that $\int_{\mathbb{R}^D} p^H_{n,i-1}(u) \log p^H_{n,i-1}(u) \mathrm{d}u$ is an integrable random variable. Then the expected accumulative differential entropy

$$\mathbb{E}\bigg[- \sum_{i=1}^n (t^n_i - t^n_{i-1}) \int_{\mathbb{R}^D} p^H_{n,i-1}(u) \log p^H_{n,i-1}(u) \mathrm{d}u \bigg]$$

is finite.

(iii) The controlled wealth $X^H = (X^H_t)_{t\in[0,T]}$ associated with H along time points of Π_n is

$$X^H_{n,i} = X^H_{n,i-1} + H_{n,i-1}(\xi_{n,i})^\mathsf{T} \Delta_{n,i} Y^n, \quad i = 1, \dots, n, \quad X^H_0 = x_0 \in \mathbb{R}.$$

Proposition 3.2. *For $n \geq 1$, $1 \leq i \leq n$, there exist (uniquely up to a $\mathbb{P}$-null set) a random vector $\mu^H_{n,i-1}$ and a random matrix $\vartheta^H_{n,i-1} \in \mathbb{S}^D_{++}$, both are $\mathcal{F}_{n,i-1}$-measurable and square integrable, and a square integrable random vector $\eta^H_{n,i}$ with*

$$\mathbb{E}[\eta^H_{n,i} | \mathcal{F}_{n,i-1}] = 0 \quad \textit{and} \quad \mathbb{E}[\eta^H_{n,i} (\eta^H_{n,i})^\mathsf{T} | \mathcal{F}_{n,i-1}] = I_D \quad \textit{a.s.} \tag{3.1}$$

such that

$$H_{n,i-1}(\xi_{n,i}) = \mu^H_{n,i-1} + \vartheta^H_{n,i-1} \eta^H_{n,i} \quad \textit{a.s.} \tag{3.2}$$

Proof. See Appendix A. □

We decompose the process X^H with the initial wealth $X^H_0 = x_0 \in \mathbb{R}$ as

$$\begin{aligned} X^H_t &= x_0 + \sum_{i=1}^{\sigma^n_t} H_{n,i-1}(\xi_{n,i})^\mathsf{T} b(Y_{n,i-1}) \Delta t^n_i \\ &\quad + \sum_{i=1}^{\sigma^n_t} H_{n,i-1}(\xi_{n,i})^\mathsf{T} a(Y_{n,i-1}) \Delta_{n,i} W + \sum_{i=1}^{\sigma^n_t} H_{n,i-1}(\xi_{n,i})^\mathsf{T} \gamma(Y_{n,i-1}) \Delta_{n,i} J \\ &=: x_0 + I_{(3.3)} + II_{(3.3)} + III_{(3.3)}. \end{aligned} \tag{3.3}$$

3.2.2. *The drift part $I_{(3.3)}$.* According to the decomposition (3.2), we express, a.s.,

$$\begin{aligned} I_{(3.3)} &= \sum_{i=1}^{\sigma^n_t} (\mu^H_{n,i-1})^\mathsf{T} b(Y_{n,i-1}) \Delta t^n_i + \sum_{i=1}^{\sigma^n_t} (\vartheta^H_{n,i-1} \eta^H_{n,i})^\mathsf{T} b(Y_{n,i-1}) \Delta t^n_i \\ &= \sum_{i=1}^{\sigma^n_t} (\mu^H_{n,i-1})^\mathsf{T} b(Y_{n,i-1}) \Delta t^n_i + \sum_{d=1}^D \sum_{i=1}^{\sigma^n_t} \bigg[\sum_{k=1}^D \vartheta^{H,(k,d)}_{n,i-1} b^{(k)}(Y_{n,i-1}) \bigg] \Big[\eta^{H,(d)}_{n,i} \Delta t^n_i \Big]. \end{aligned}$$

For the discrete-time integrator in the second term, we have the following law of large numbers

$$\sum_{i=1}^{\sigma^n_t} \eta^{H,(d)}_{n,i} \Delta t^n_i \xrightarrow{\mathbf{L}_2(\mathbb{P})} 0 \quad \text{as } n \to \infty$$

for all $d = 1, \dots, D$. Indeed, due to the orthogonality and $\mathbb{E}[|\eta^{H,(d)}_{n,i}|^2] = 1$ it holds that

$$\mathbb{E}\bigg[\Big| \sum_{i=1}^{\sigma^n_t} \eta^{H,(d)}_{n,i} \Delta t^n_i \Big|^2 \bigg] = \sum_{i=1}^{\sigma^n_t} |\Delta t^n_i|^2 \leq t \max_{1\leq i\leq n} \Delta t^n_i \to 0.$$

3.2.3. *The Brownian part $II_{(3.3)}$.* By the same arguments as for the drift part, we decompose $II_{(3.3)}$ as

$$II_{(3.3)} = \sum_{i=1}^{\sigma_t^n} (\mu_{n,i-1}^H)^\mathsf{T} a(Y_{n,i-1}) \Delta_{n,i} W + \sum_{d,d'=1,\dots,D} \sum_{i=1}^{\sigma_t^n} \left[\sum_{k=1}^{D} \vartheta_{n,i-1}^{H,(k,d)} a^{(k,d')}(Y_{n,i-1}) \right] \left[\eta_{n,i}^{H,(d)} \Delta_{n,i} W^{(d')} \right].$$

Define the interpolated process $W^n = (W_t^n)_{t\in[0,T]}$ and the $\mathbb{R}^{D^2}$-valued process $M^n = (M_t^n)_{t\in[0,T]}$ by

$$\begin{aligned} W_t^n &:= \sum_{i=1}^{\sigma_t^n} \Delta_{n,i} W, \\ M_t^{n,(d,d')} &:= \sum_{i=1}^{\sigma_t^n} \eta_{n,i}^{H,(d)} \Delta_{n,i} W^{(d')}, \quad d,d' = 1,\dots,D, \\ M_t^n &= (M_t^{n,(1,1)},\dots,M_t^{n,(1,D)}, M_t^{n,(2,1)},\dots,M_t^{n,(2,D)},\dots,M_t^{n,(D,1)},\dots,M_t^{n,(D,D)})^\mathsf{T}, \end{aligned}$$

so that $\Delta_{n,i} W^n = \Delta_{n,i} W$ and $M_t^n = \sum_{i=1}^{\sigma_t^n} \eta_{n,i}^H \otimes \Delta_{n,i} W$. Then we get

$$II_{(3.3)} = \sum_{i=1}^{\sigma_t^n} (\mu_{n,i-1}^H)^\mathsf{T} a(Y_{n,i-1}) \Delta_{n,i} W^n + \sum_{d,d'=1,\dots,D} \sum_{i=1}^{\sigma_t^n} [\vartheta_{n,i-1}^H a(Y_{n,i-1})]^{(d,d')} \Delta_{n,i} M^{n,(d,d')}.$$

Here W^n, M^n can be respectively regarded as a discrete-time integrator of the first and the second term in the decomposition of $II_{(3.3)}$.

3.2.4. *The jump part $III_{(3.3)}$.* For the jump part in (3.3), a first try is to rewrite

$$III_{(3.3)} = \int_{(0,t]\times E\times\mathbb{R}^D} \sum_{i=1}^{n} \left[H_{n,i-1}(u)^\mathsf{T} \gamma(Y_{n,i-1}) e \right] \mathbb{1}_{(t_{i-1}^n, t_i^n]}(s) \mathbb{m}_n(\mathrm{d}s, \mathrm{d}e, \mathrm{d}u),$$

for the random measure

$$\mathbb{m}_n(\mathrm{d}t, \mathrm{d}e, \mathrm{d}u) := \sum_{i=1}^{n} \delta_{(t_i^n, \Delta_{n,i} J, \xi_{n,i})}(\mathrm{d}t, \mathrm{d}e, \mathrm{d}u), \tag{3.4}$$

on $\mathcal{B}([0,T]\times E\times\mathbb{R}^D)$. Here, δ denotes the Dirac measure. So, we move the Gaussian random variables $\xi_{n,i}$ for the control randomization from the integrand to the random measure $\mathbb{m}_n$, that acts as a new integrator. It is, however, intuitively clear that the limit random measure (in a weak sense) should be

$$\mathbb{m}(\mathrm{d}t, \mathrm{d}e, \mathrm{d}u) = \sum_{j} \delta_{(\tau_j, \Delta J_{\tau_j}, \xi_j)}(\mathrm{d}t, \mathrm{d}e, \mathrm{d}u), \tag{3.5}$$

where $(\tau_j)_{j\in\mathbb{N}}$ are the jump times of J and $(\xi_j)_{j\in\mathbb{N}}$ is a sequence of independent standard Gaussians (independent of J), i.e., in the limit we would like to create independent Gaussian jumps at each jump time of L as additional source of noise. If the original Lévy process has infinite activity, this random measure does not induce a Lévy process, because the squared Gaussian jumps $\sum_{j;\tau_j\le t} \xi_j^2$ do not converge. As a way out, we re-scale the additional Gaussian jumps depending on the jump sizes of the original Lévy process.

To this end, let us fix a $\psi \in C^2(\mathbb{R}^D)$ which satisfies

$$\|\nabla\psi\|_\infty + \|\nabla^2\psi\|_\infty < \infty, \quad \psi \ge 0 \quad \text{and} \quad \psi(x) = 0 \Leftrightarrow x = 0.$$

A prototype example in our context is that, for a given constant $c > 0$,

$$\psi(x) = \sqrt{\|x\|^2 + c^2} - c.$$

Define the random measure $\mathbb{m}_n^\psi$ on $\mathcal{B}([0,T]\times E\times\mathbb{R}^D)$ by setting

$$\mathbb{m}_n^\psi(\mathrm{d}t,\mathrm{d}e,\mathrm{d}u) := \sum_{i=1}^n \delta_{(t_i^n,\,\Delta_{n,i}J,\,\psi(\Delta_{n,i}J)\xi_{n,i})}(\mathrm{d}t,\mathrm{d}e,\mathrm{d}u).$$

Then the third term $III_{(3.3)}$ is finally expressed as

$$III_{(3.3)} = \int_{(0,t]\times E\times\mathbb{R}^D} \sum_{i=1}^n \left[H_{n,i-1}\left(\frac{u}{\psi(e)}\right)^\mathsf{T}\gamma(Y_{n,i-1})e\right]\mathbb{1}_{(t_{i-1}^n,t_i^n]}(s)\mathbb{m}_n^\psi(\mathrm{d}s,\mathrm{d}e,\mathrm{d}u).$$

Note that the random measure $\mathbb{m}_n^\psi$ is characterized by the induced martingale $L^{n,\psi}=(L_t^{n,\psi})_{t\in[0,T]}$ with $L_0^{n,\psi}=0$ and

$$L_t^{n,\psi} := \int_{(0,t]\times E\times\mathbb{R}^D}(e,u)^\mathsf{T}\mathbb{m}_n^\psi(\mathrm{d}s,\mathrm{d}e,\mathrm{d}u) = \sum_{i=1}^{\sigma_t^n}(\Delta_{n,i}J,\psi(\Delta_{n,i}J)\xi_{n,i})^\mathsf{T} = \sum_{i=1}^{\sigma_t^n}\Delta_{n,i}L^{n,\psi}.$$

As explained above, the "damping factor" $\psi(\Delta_{n,i}J)$ in front of $\xi_{n,i}$ in the third coordinate of $\mathbb{m}_n^\psi$ is introduced to ensure that $L^{n,\psi}$ can converge to a Lévy process in the infinite activity case. The smoothness condition for ψ is merely convenient for applying Itô's formula later on.

3.2.5. *Distributional limit of discrete-time integrators.* Set $\mathbf{D} := D^2+3D$. We collect all discrete-time integrators of the Brownian and the jump parts to obtain the triangular array of $\mathbf{D}$-dimensional random vectors $\mathcal{Z}^n=(\mathcal{Z}_t^n)_{t\in[0,T]}$ with

$$\mathcal{Z}^n := \mathrm{vec}(W^n,M^n,L^{n,\psi}).$$

Our purpose is to investigate the distributional limit of $(\mathcal{Z}^n)_{n\geq1}$. To this end, we introduce the Borel measure ν_L^ψ defined on $\mathbb{R}^{2D}$ by setting

$$\nu_L^\psi(\mathrm{d}e,\mathrm{d}u) := \mathbb{1}_{\{\|e\|>0\}}\varphi_D\left(\frac{u}{\psi(e)}\right)\frac{\mathrm{d}u}{\psi(e)^D}\nu(\mathrm{d}e),\quad e,u\in\mathbb{R}^D.$$

Then, by a change of variables, one has

$$\int_{\mathbb{R}^{2D}} f(e,u)\nu_L^\psi(\mathrm{d}e,\mathrm{d}u) = \int_{E\times\mathbb{R}^D} f(e,\psi(e)u)\nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u$$

provided that $f\geq0$ or $\int_{\mathbb{R}^{2D}}|f(e,u)|\nu_L^\psi(\mathrm{d}e,\mathrm{d}u)<\infty$. In particular, choosing $f(e,u)=\|e\|^2+\|u\|^2$ we find that ν_L^ψ is a square integrable Lévy measure on $\mathbb{R}^{2D}\backslash\{0\}$ with $\nu_L^\psi(\{0\}\times\mathbb{R}^D)=0$ as

$$\begin{aligned}&\int_{\mathbb{R}^{2D}\backslash\{0\}}(\|e\|^2+\|u\|^2)\nu_L^\psi(\mathrm{d}e,\mathrm{d}u) = \int_{E\times\mathbb{R}^D}(\|e\|^2+\psi(e)^2\|u\|^2)\nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u\\ &= \int_E\|e\|^2\nu(\mathrm{d}e)+\int_E\psi(e)^2\nu(\mathrm{d}e)\int_{\mathbb{R}^D}\|u\|^2\varphi_D(u)\mathrm{d}u \leq (1+D\|\nabla\psi\|_\infty^2)\int_E\|e\|^2\nu(\mathrm{d}e)<\infty.\end{aligned}$$

We need the following condition to obtain the desired weak convergence.

Assumption 3.3. $\{\widetilde{H}_{n,i-1}(\xi_{n,i})^\mathsf{T}(\Theta_{n,i-1}^H)^{-1}\widetilde{H}_{n,i-1}(\xi_{n,i})\}_{1\leq i\leq n,n\geq1}$ is uniformly integrable.

Remark 3.4. Let us briefly comment on Assumption 3.3.

(1) By the construction of $\eta_{n,i}^H$ in the proof of Proposition 3.2, one has, a.s.,

$$\begin{aligned}&\widetilde{H}_{n,i-1}(\xi_{n,i})^\mathsf{T}(\Theta_{n,i-1}^H)^{-1}\widetilde{H}_{n,i-1}(\xi_{n,i})\\ &= \mathbf{tr}[(\Theta_{n,i-1}^H)^{-\frac12}\widetilde{H}_{n,i-1}(\xi_{n,i})((\Theta_{n,i-1}^H)^{-\frac12}\widetilde{H}_{n,i-1}(\xi_{n,i}))^\mathsf{T}]\\ &= \|\eta_{n,i}^H\|^2,\end{aligned}$$

i.e., Assumption 3.3 is equivalent to the uniform integrability of $\{\|\eta_{n,i}^H\|^2\}_{1\leq i\leq n,n\geq1}$.

(2) Assume, for all n, that $H\colon \Pi_n \times \Omega \times \mathbb{R}^D \to \mathbb{R}^D$ has the form

$$H_{n,i-1}(\omega; u) = \mathsf{m}_{n,i-1}(\omega) + \mathsf{v}_{n,i-1}(\omega)u, \quad i = 1, \dots, n, \tag{3.6}$$

where $\mathsf{m}_{n,i-1}$ and $\mathsf{v}_{n,i-1}$ are respectively $\mathbb{R}^D$-valued and $\mathbb{S}^D_{++}$-valued random variables, both are $\mathcal{F}_{n,i-1}$-measurable and square integrable with $\log(\det(\mathsf{v}_{n,i-1})) \in \mathbf{L}_1(\mathbb{P})$. Then H is linear with respect to the exploration variable and is admissible in the sense given in Section 3.2.1. Moreover, in the notation of Proposition 3.2, one has $\mu^H_{n,i-1} = \mathsf{m}_{n,i-1}$, $\vartheta^H_{n,i-1} = \mathsf{v}_{n,i-1}$, and $\eta^H_{n,i} = \xi_{n,i}$, which obviously implies that Assumption 3.3 holds.

(3) We will see in Theorem 4.10 below that the time discretization of the optimal control process for the associated continuous-time control problem has the form (3.6).

Under the setting of Section 3.2.1, we have the following result whose proof is postponed to Section 5.

Theorem 3.5. *Assume that $\mathcal{W}$ is a D^2-dimensional standard Brownian motion independent of W, and that L^ψ is a square integrable martingale null at 0 which is a $2D$-dimensional purely non-Gaussian Lévy process with Lévy measure ν^ψ_L. Assume that processes $W, \mathcal{W}, L^\psi$ are defined on the same probability space. Then, L^ψ is independent of $(W, \mathcal{W})$, and under Assumption 3.3, the sequence $(\mathcal{Z}^n)_{n\geq 1}$ converges weakly to $\mathrm{vec}(W, \mathcal{W}, L^\psi)$ as $n \to \infty$ in the Skorokhod topology on the space of càdlàg functions $f\colon [0,T] \to \mathbb{R}^{D^2+3D}$.*

By rearranging components of $\mathcal{W}$, we may consider $\mathcal{W}$ as an $\mathbb{R}^{D\times D}$-valued process. Then Theorem 3.5 suggests that the exploratory SDE in the continuous-time setting for the controlled wealth process X^H with an admissible H is as follows

$$\begin{aligned} \mathrm{d}X^H_t = (\mu^H_t)^\mathsf{T} b(Y_{t-})\mathrm{d}t + (\mu^H_t)^\mathsf{T} a(Y_{t-})\mathrm{d}W_t + \mathbf{tr}[(\Theta^H_t)^{\frac{1}{2}} a(Y_{t-})\mathrm{d}\mathcal{W}^\mathsf{T}_t] \\ + \int_{E\times\mathbb{R}^D} H_t\left(\frac{u}{\psi(e)}\right)^\mathsf{T} \gamma(Y_{t-}) e\, \widetilde{N}^\psi_L(\mathrm{d}t, \mathrm{d}e, \mathrm{d}u), \quad X^H_0 = x_0 \in \mathbb{R}, \end{aligned} \tag{3.7}$$

where $\widetilde{N}^\psi_L$ is the compensated Poisson random measure of L^ψ and the underlying process Y is given by

$$\mathrm{d}Y_t = b(Y_{t-})\mathrm{d}t + a(Y_{t-})\mathrm{d}W_t + \gamma(Y_{t-}) \int_{E\times\mathbb{R}^D} e\widetilde{N}^\psi_L(\mathrm{d}t, \mathrm{d}e, \mathrm{d}u), \quad Y_0 = y_0 \in \mathbb{R}^D.$$

One notices that such a Y also admits $\mathcal{L}_Y$ in (2.2) as the generator.

Remark 3.6. Let us briefly comment on SDE (3.7). For the Brownian component, the noise caused by exploration, i.e. $\mathcal{W}$, is completely separated from the original noise, i.e. W. While for the jump part, both noises are simultaneously captured by the Poisson random measure generated by a D^2-dimensional Lévy process. Interestingly, for the optimal control H obtained in (4.22), it turns out that one can completely separate these two sources of randomness due to the linearity with respect to the exploration variable.

Remark 3.7. If the control enters in the drift part only, then the reasoning in Section 3.2.2 shows that there is no extra exploration noise in the continuous-time formulation. This is the case, e.g., in [13] where the authors add jumps as uncontrolled Lévy noise.

Remark 3.8. We briefly explain the relation of the jump part in (3.7) to the notion of a *relaxed Poisson measure*, which has been introduced in the context of relaxed controls by [23]. We will

restrict our discussion to the finite activity case (as assumed in [23]), in which we do not need to re-scale the Gaussian jumps and we can formally set $\psi \equiv 1$. Then,

$$\widetilde{N}_L^{\psi}(\mathrm{d}t, \mathrm{d}e, \mathrm{d}u) = \mathbb{m}(\mathrm{d}t, \mathrm{d}e, \mathrm{d}u) - \nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u\mathrm{d}t$$

for the random measure $\mathbb{m}$ defined in (3.5). Hence, $\mathbb{m}$ is a relaxed Poisson measure for the relaxed control $\varphi_D(u)\mathrm{d}u\mathrm{d}t$ in the sense of [23, p. 190]. The approximation $\mathbb{m}_n$, which we consider in (3.4), arises from the modeling of control randomization in RL. Numerically it can be interpreted as a Monte Carlo approximation of the relaxed control $\varphi_D(u)\mathrm{d}u\mathrm{d}t$ on the time grid $t_0^n, \dots, t_n^n$. This Monte Carlo approximation is conceptually different to the numerical approximation considered by Kushner in [23]. Roughly speaking, Kushner's approach would approximate the relaxed control by replacing the multivariate Gaussian distribution by a discrete distribution supported on N points $\alpha_1, \dots, \alpha_N$ and successively calling these N points over a refined time grid.

4. Entropy-regularized exploratory MV problem with Lévy jumps

We work on a fixed complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ carrying the triplet $(W, \mathcal{W}, L^{\psi})$ aforementioned in Theorem 3.5. Let N_L^{ψ} denote the associated Poisson random measure of L^{ψ} with the compensation $\widetilde{N}_L^{\psi} := N_L^{\psi} - \boldsymbol{\lambda}_1 \otimes \nu_L^{\psi}$, where $\boldsymbol{\lambda}_1$ is the 1-dimensional Lebesgue measure. For $0 \le t \le s \le T$, we denote $\mathcal{F}_s^t = \sigma\{W_r - W_t, \mathcal{W}_r - \mathcal{W}_t, L_r^{\psi} - L_t^{\psi} : t \le r \le s\}$ augmented by all $\mathbb{P}$-null sets. Set $\mathcal{F}_s := \mathcal{F}_s^0$.

For $\xi \sim \mathcal{N}(0, I_D)$ we define the family of *deterministic* admissible functions as

$$\mathcal{A} := \left\{ F \;\middle|\; F\colon \mathbb{R}^D \to \mathbb{R}^D \text{ Borel}, \int_{\mathbb{R}^D} \|F(u)\|^2 \varphi_D(u)\mathrm{d}u < \infty, \, F(\xi) \text{ has a probability density } p^F \right\}.$$

Admissible controls in the discrete-time setting are adapted to the continuous-time setting as follows.

Definition 4.1 (Admissible control)**.** For $(t, y) \in [0, T) \times \mathbb{R}^D$, denote by $\mathcal{A}(t, y)$ the family of all admissible controls H for which the following conditions hold:

(H1) (Admissibility) $H\colon [t, T] \times \Omega \times \mathbb{R}^D \to \mathbb{R}^D$ satisfies that
  (a) H is $\mathcal{P}([t, T]) \otimes \mathcal{B}(\mathbb{R}^D)$-measurable, where $\mathcal{P}([t, T])$ is the predictable σ-algebra on $[t, T] \times \Omega$;
  (b) $H_s(\cdot) := H_s(\omega; \cdot) \in \mathcal{A}$ for all $(s, \omega) \in [t, T] \times \Omega$.

(H2) (Integrability) It holds that

$$\mathbb{P}\left(\int_t^T \int_{\mathbb{R}^D} \|H_s(u)\|^2 \varphi_D(u)\mathrm{d}u < \infty \right) = 1, \tag{4.1}$$

and that processes $\mu^H = (\mu_s^H)_{s\in[t,T]}$, $\Theta^H = (\Theta_s^H)_{s\in[t,T]}$ defined on $[t, T] \times \Omega$ by

$$\mu_s^H := \int_{\mathbb{R}^D} H_s(u)\varphi_D(u)\mathrm{d}u, \quad \widetilde{H}_s(u) := H_s(u) - \mu_s^H, \quad \Theta_s^H := \int_{\mathbb{R}^D} \widetilde{H}_s(u)\widetilde{H}_s(u)^{\mathsf{T}}\varphi_D(u)\mathrm{d}u,$$

satisfy that

$$\mathbb{E}\bigg[\int_t^T \bigg((\mu_s^H)^{\mathsf{T}} A(Y_{s-}^{t,y})\mu_s^H + \mathbf{tr}[A(Y_{s-}^{t,y})\Theta_s^H] + \int_{E\times\mathbb{R}^D} |H_s(u)^{\mathsf{T}}\gamma(Y_{s-}^{t,y})e|^2 \nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u \bigg)\mathrm{d}s \bigg] + \mathbb{E}\bigg[\bigg| \int_t^T |(\mu_s^H)^{\mathsf{T}} b(Y_{s-}^{t,y})|\mathrm{d}s \bigg|^2 \bigg] < \infty, \tag{4.2}$$

where $Y^{t,y} = (Y_s^{t,y})_{s\in[t,T]}$ is a (unique) strong solution to the following SDE on $[t,T]$

$$\mathrm{d}Y_s^{t,y} = b(Y_{s-}^{t,y})\mathrm{d}s + a(Y_{s-}^{t,y})\mathrm{d}W_s + \gamma(Y_{s-}^{t,y})\int_{E\times\mathbb{R}^D} e\widetilde{N}_L^{\psi}(\mathrm{d}s,\mathrm{d}e,\mathrm{d}u), \quad Y_t^{t,y} = y. \tag{4.3}$$

(H3) (Finite accumulative differential entropy) There is a kernel $p^H\colon [t,T]\times\Omega\times\mathbb{R}^D \to \mathbb{R}$ such that $p_s^H(\omega;\cdot)$ is a probability density function of $H_s(\omega;\zeta)$ for any $(s,\omega)\in[t,T]\times\Omega$, where $\zeta\sim\mathcal{N}(0,I_D)$ is independent of $\mathcal{F}_s^t$, and that $(s,\omega)\mapsto\int_{\mathbb{R}^D} p_s^H(u)\log p_s^H(u)\mathrm{d}u$ is $(\mathcal{F}_s^t)_{s\in[t,T]}$-predictable with

$$\mathbb{E}\bigg[\int_t^T\bigg|\int_{\mathbb{R}^D} p_s^H(u)\log p_s^H(u)\mathrm{d}u\bigg|\mathrm{d}s\bigg] < \infty. \tag{4.4}$$

For a given control $H\in\mathcal{A}(t,y)$ and $x\in\mathbb{R}$, the dynamic of the controlled wealth process $X^{t,x,y;H} = (X_s^{t,x,y;H})_{s\in[t,T]}$, which is assumed to has càdlàg paths, is described by the exploratory SDE on $[t,T]$ as

$$\begin{aligned}\mathrm{d}X_s^{t,x,y;H} &= (\mu_s^H)^{\mathsf{T}} b(Y_{s-}^{t,y})\mathrm{d}s + (\mu_s^H)^{\mathsf{T}} a(Y_{s-}^{t,y})\mathrm{d}W_s + \mathbf{tr}[(\Theta_s^H)^{\frac12} a(Y_{s-}^{t,y})\mathrm{d}\mathcal{W}_s^{\mathsf{T}}]\\ &\quad + \int_{E\times\mathbb{R}^D} H_s\bigg(\frac{u}{\psi(e)}\bigg)^{\mathsf{T}}\gamma(Y_{s-}^{t,y})e\,\widetilde{N}_L^{\psi}(\mathrm{d}s,\mathrm{d}e,\mathrm{d}u), \quad X_t^{t,x,y;H} = x,\end{aligned} \tag{4.5}$$

where $Y^{t,y}$ solves the SDE (4.3).

Remark 4.2. (1) Processes μ^H and Θ^H are predictable by (H1) and Fubini's theorem.

(2) As a consequence of [6, Theorem X.1.1], there exists a $c_D>0$ such that $\|A^{\frac12}-B^{\frac12}\|\le c_D\|A-B\|^{\frac12}$ for any $A,B\in\mathbb{S}_+^D$. Hence $\mathbb{S}_+^D\ni A\mapsto A^{\frac12}$ is (Hölder) continuous which then ensures that $(\Theta^H)^{\frac12}$ is also a predictable $\mathbb{S}_+^D$-valued process.

(3) Due to (4.2), $X^{t,x,y;H}$ in (4.5) is a square integrable process satisfying

$$\mathbb{E}\Big[\sup_{t\le s\le T}|X_s^{t,x,y;H}|^2\Big] < \infty. \tag{4.6}$$

4.1. **Problem formulation.** We are now in a position to formulate the entropy-regularized exploratory MV problem. Remark that, due to the time inconsistency of the MV problem, we just examine solutions among *precommitted* strategies which are optimal at $t=0$ only.

Let us fix a $\hat{z}\in\mathbb{R}$ which represents the targeted expected terminal wealth. For an initial wealth $x_0\in\mathbb{R}$ and $y_0\in\mathbb{R}^D$, we consider the problem

$$\begin{cases}\min\limits_{H\in\mathcal{A}(0,y_0)} \mathbb{E}\bigg[\Big(X_T^{0,x_0,y_0;H} - \mathbb{E}\Big[X_T^{0,x_0,y_0;H}\Big]\Big)^2 + \lambda\displaystyle\int_0^T\int_{\mathbb{R}^D} p_s^H(u)\log p_s^H(u)\mathrm{d}u\mathrm{d}s\bigg]\\ \text{subject to } X^{0,x_0,y_0;H} \text{ given in (4.5) and } \mathbb{E}\Big[X_T^{0,x_0,y_0;H}\Big] = \hat{z}.\end{cases} \tag{4.7}$$

Here the *exploration weight* $\lambda\ge0$, which is fixed from now on, describes the trade-off between exploitation and exploration and it is also known as the *temperature parameter* in the RL literature.

We follow [34] to apply the Lagrange multiplier method to solve (4.7) (see Section 2.4 for a similar argument in the setting without exploration). In the first step, we examine the following entropy-regularized quadratic-loss minimization problem, parameterized by $\hat{w}\in\mathbb{R}$,

$$\begin{cases}\min\limits_{H\in\mathcal{A}(0,y_0)} \mathbb{E}\bigg[\Big(X_T^{0,x_0,y_0;H} - \hat{w}\Big)^2 + \lambda\displaystyle\int_0^T\int_{\mathbb{R}^D} p_s^H(u)\log p_s^H(u)\mathrm{d}u\mathrm{d}s\bigg]\\ \text{subject to } X^{0,x_0,y_0;H} \text{ given in (4.5)}.\end{cases} \tag{4.8}$$

We solve (4.8) to obtain a solution $H^* := H^*(\hat{w})$ depending on $\hat{w}$. This task is presented in Section 4.2. In the next step, we find the Lagrange multiplier $\hat{w}$ by using the constraint $\mathbb{E}[X_T^{H^*}] = \hat{z}$. Then $H^*(\hat{w})$ is a solution to problem (4.7) where $\hat{w}$ is the obtained Lagrange multiplier. The latter task is done in Section 4.3.

4.2. **The entropy-regularized quadratic-loss optimization problem.** Let us fix $\hat{w} \in \mathbb{R}$. Problem (4.8) is an unconstrained control problem and we will find its solutions via the dynamic programing approach. Define the function $V^H(\cdot|\hat{w})$ associated with a control $H \in \mathcal{A}(t,y)$ and $x \in \mathbb{R}$ by setting

$$V^H(t,x,y|\hat{w}) := \mathbb{E}\bigg[\Big(X_T^{t,x,y;H} - \hat{w}\Big)^2 + \lambda \int_t^T \int_{\mathbb{R}^D} p_s^H(u) \log p_s^H(u) \mathrm{d}u\mathrm{d}s\bigg].$$

We consider the following system of problems which particularly yields to (4.8) when $(t,x,y) = (0,x_0,y_0)$.

Problem 4.3. *For given $(t,x,y) \in [0,T) \times \mathbb{R} \times \mathbb{R}^D$, find an $H^* \in \mathcal{A}(t,y)$ such that*

$$V^*(t,x,y|\hat{w}) := V^{H^*}(t,x,y|\hat{w}) = \min_{H \in \mathcal{A}(t,y)} V^H(t,x,y|\hat{w}) \tag{4.9}$$

subject to the state equation (4.5).

Definition 4.4. For a given initial triple (t,x,y), any $H^* \in \mathcal{A}(t,y)$ satisfying (4.9) is called an *optimal control*, the corresponding controlled state process $X^{t,x,y;*} := X^{t,x,y;H^*}$ is called an *optimal state/wealth process*, and $V^*(\cdot|\hat{w})$ satisfying the terminal condition $V^*(T,x,y|\hat{w}) = (x-\hat{w})^2$ is called the *value function*.

4.2.1. *Entropy-regularized Hamilton–Jacobi–Bellman (HJB) equation.* As we use the dynamic programming approach to solve Problem 4.3, it is useful to consider the associated HJB equation. Let us first introduce some notations. For $F \in \mathcal{A}$, we define $m^F \in \mathbb{R}^D$ and $\theta^F \in \mathbb{S}_+^D$ by

$$\begin{aligned} m^F &:= \int_{\mathbb{R}^D} F(u)\varphi_D(u)\mathrm{d}u, \\ \theta^F &:= \int_{\mathbb{R}^D} (F(u)-m^F)(F(u)-m^F)^\mathsf{T}\varphi_D(u)\mathrm{d}u = \int_{\mathbb{R}^D} F(u)F(u)^\mathsf{T}\varphi_D(u)\mathrm{d}u - m^F(m^F)^\mathsf{T}, \end{aligned}$$

and the *differential entropy* of F is denoted by

$$\mathsf{Ent}(F) := -\int_{\mathbb{R}^D} p^F(u) \log p^F(u)\mathrm{d}u.$$

Using the classical Bellman's principle of optimality and a standard verification argument (see the proof of Theorem 4.10 below) we find that the HJB type formula in our setting is stated in form of a (possibly degenerate) second-order PIDE as follows:

$$\begin{aligned} 0 = {} & \partial_t v(t,x,y) + b(y)^\mathsf{T}\nabla_y v(t,x,y) + \frac{1}{2}\mathbf{tr}[A(y)\nabla_{yy}^2 v(t,x,y)] \\ & + \min_{F\in\mathcal{A}} \bigg\{ \frac{1}{2}\partial_{xx}^2 v(t,x,y)\Big((m^F)^\mathsf{T}A(y)m^F + \mathbf{tr}[A(y)\theta^F]\Big) \\ & \qquad + (m^F)^\mathsf{T}\Big(A(y)\nabla_{xy}^2 v(t,x,y) + \partial_x v(t,x,y)b(y)\Big) \\ & \qquad + \int_{E\times\mathbb{R}^D} \Big(v(t,x+F(u)^\mathsf{T}\gamma(y)e, y+\gamma(y)e) - v(t,x,y) \\ & \qquad\qquad - \partial_x v(t,x,y)F(u)^\mathsf{T}\gamma(y)e - \nabla_y v(t,x,y)^\mathsf{T}\gamma(y)e\Big)\nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u \end{aligned}$$

$$- \lambda \mathsf{Ent}(F)\Big\}, \quad (t,x,y) \in [0,T) \times \mathbb{R} \times \mathbb{R}^D, \tag{4.10}$$

with the terminal condition $v(T,x,y) = (x-\hat{w})^2$ for $(x,y) \in \mathbb{R} \times \mathbb{R}^D$.

Remark 4.5. By [9, Theorem 8.6.5], one has $\mathsf{Ent}(F) = -\infty$ if $\det(\theta^F) = 0$. Hence, it suffices to consider the above minimization over $F \in \mathcal{A}$ with $\det(\theta^F) > 0$, i.e. over $F \in \mathcal{A}$ with $\theta^F \in \mathbb{S}^D_{++}$.

Remark 4.6. In the case of no jumps, i.e. $\nu = 0$, (4.10) simplifies to

$$\begin{aligned} 0 = \partial_t v(t,x,y) + b(y)^\mathsf{T} \nabla_y v(t,x,y) + \frac{1}{2}\mathbf{tr}[A(y)\nabla^2_{yy} v(t,x,y)] \\ + \min_{m \in \mathbb{R}^D, \theta \in \mathbb{S}^D_{++}} \Big\{ \frac{1}{2}\partial^2_{xx} v(t,x,y)\Big(m^\mathsf{T} A(y) m + \mathbf{tr}[A(y)\theta]\Big) \\ + m^\mathsf{T}\Big(A(y)\nabla^2_{xy} v(t,x,y) + \partial_x v(t,x,y) b(y)\Big) \\ - \lambda \max_{F \in \mathcal{A};\, m^F = m \text{ and } \theta^F = \theta} \mathsf{Ent}(F)\Big\}, \quad (t,x,y) \in [0,T) \times \mathbb{R} \times \mathbb{R}^D. \end{aligned}$$

By the entropy-maximizing property of the Gaussian distribution, the maximum over F is achieved at the linear function $F(u) = m + \theta^{\frac{1}{2}} u$ and the HJB-equation becomes a second-order PDE for the unknown value function v. This PDE can the be solved by a quadratic ansatz as done e.g. in [34] for the case $D = 1$ and constant coefficients. In the presence of jumps, this "separation argument" (solving first for F given its first two moments) a priori does not work anymore, because F explicitly enters the integral term of the HJB-PIDE (4.10). As we will show in the next section, (4.10) can nonetheless be solved by a quadratic ansatz, and, then, the separation step can be performed a posteriori, leading to the Gaussianity of the optimal control law.

4.2.2. *Quadratic ansatz.* We first introduce the following function classes in relation to the coefficient γ and Lévy measure ν.

Definition 4.7. For a Borel function $g\colon [0,T] \times \mathbb{R}^D \to \mathbb{R}$ we let $g \in \Upsilon(0)$ (resp. $g \in \Upsilon(1)$, $g \in \Upsilon(2)$) if there exists a (jointly) continuous function $\Upsilon^{(0)}_g$(resp. $\Upsilon^{(1)}_g, \Upsilon^{(2)}_g$)$\colon [0,T] \times \mathbb{R}^D \to [0,\infty)$ such that

$$\begin{aligned} &\int_E |g(t, y+\gamma(y)e)|\, \|e\|^2 \nu(\mathrm{d}e) \le \Upsilon^{(0)}_g(t,y), \\ \text{resp.} \quad &\int_E |g(t, y+\gamma(y)e) - g(t,y)|\, \|e\| \nu(\mathrm{d}e) \le \Upsilon^{(1)}_g(t,y), \\ \text{resp.} \quad &\int_E \Big|g(t, y+\gamma(y)e) - g(t,y) - \nabla_y g(t,y)^\mathsf{T} \gamma(y) e\Big| \nu(\mathrm{d}e) \le \Upsilon^{(2)}_g(t,y), \end{aligned}$$

for all $(t,y) \in [0,T] \times \mathbb{R}^D$, where we additionally assume that $\nabla_y g$ exists and measurable for $g \in \Upsilon(2)$. Then $\Upsilon^{(k)}_g$ is called an Υ-*dominating function* of $g \in \Upsilon(k)$.

Remark 4.8. A standard calculation shows that $g \in \Upsilon(0) \cap \Upsilon(1) \cap \Upsilon(2)$ if $\int_E \|e\|^2 \nu(\mathrm{d}e) < \infty$ and one of the following holds:

(a) g is twice continuously differentiable with respect to y with

$$\sup_{(t,y) \in [0,T] \times \mathbb{R}^D} (|g(t,y)| + \|\nabla_y g(t,y)\| + \|\nabla^2_{yy} g(t,y)\|) < \infty.$$

(b) $\sup_{(t,y) \in [0,T] \times \mathbb{R}^D} |g(t,y)| < \infty$, $\nabla_y g$ is jointly continuous on $[0,T] \times \mathbb{R}^D$, and $\nu(E) < \infty$.

For $\alpha \in \Upsilon(0)\cap\Upsilon(1)$, $\alpha > 0$ on $[0,T]\times\mathbb{R}^D$, and $\nabla_y\alpha$ exists, we define the functions $\mathcal{M}_\alpha\colon [0,T]\times\mathbb{R}^D \to \mathbb{R}^D$ and $\mathcal{S}_\alpha\colon [0,T]\times\mathbb{R}^D \to \mathbb{S}^D_{++}$ as

$$\mathcal{M}_\alpha(t,y) := \alpha(t,y)b(y) + A(y)\nabla_y\alpha(t,y) + \gamma(y)\int_E (\alpha(t,y+\gamma(y)e) - \alpha(t,y))e\,\nu(\mathrm{d}e), \tag{4.11}$$

$$\mathcal{S}_\alpha(t,y) := \alpha(t,y)A(y) + \gamma(y)\bigg(\int_E \alpha(t,y+\gamma(y)e)ee^\mathsf{T}\nu(\mathrm{d}e)\bigg)\gamma(y)^\mathsf{T}. \tag{4.12}$$

In particular, if $\alpha \equiv 1$ on $[0,T]\times\mathbb{R}^D$ then $\mathcal{M}_\alpha = b$ and $\mathcal{S}_\alpha = \Sigma$. One also remarks that the mapping $\mathcal{S}_\alpha$ is well-defined. Indeed, for any $(t,y) \in [0,T]\times\mathbb{R}^D$ and $u \in \mathbb{R}^D\backslash\{0\}$, one has $u^\mathsf{T}\mathcal{S}_\alpha(t,y)u > 0$ because of $\alpha > 0$ and the non-degenerate condition (see Section 3.1). As a consequence, the inverse $\mathcal{S}_\alpha^{-1}(t,y)$ exists and also belongs to $\mathbb{S}^D_{++}$ which can be easily derived from the spectral decomposition of $\mathcal{S}_\alpha(t,y)$.

Proposition 4.9 (Quadratic value function)**.** *Let $\alpha,\beta \in C^{1,2}([0,T]\times\mathbb{R}^D)\cap\Upsilon(2)$. Assume that $\alpha \in \Upsilon(0)\cap\Upsilon(1)$ and $\alpha > 0$, and that α,β solve the following system of PIDEs pointwise on $[0,T)\times\mathbb{R}^D$,*

$$\begin{cases} \partial_t\alpha(t,y) + \mathcal{L}_Y\alpha(t,y) - (\mathcal{M}_\alpha^\mathsf{T}\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha)(t,y) = 0, \\ \partial_t\beta(t,y) + \mathcal{L}_Y\beta(t,y) - \dfrac{\lambda}{2}\log\left(\dfrac{(\lambda\pi)^D}{\det(\mathcal{S}_\alpha(t,y))}\right) = 0, \\ \alpha(T,\cdot)\equiv 1 \quad \textit{and} \quad \beta(T,\cdot)\equiv 0, \end{cases} \tag{4.13}$$

where $\mathcal{L}_Y\phi(t,y) := (\mathcal{L}_Y\phi(t,\cdot))(y)$ for $\phi \in \{\alpha,\beta\}$. Then, for $(t,x,y)\in[0,T]\times\mathbb{R}\times\mathbb{R}^D$,

$$v^{\mathrm{opt}}(t,x,y) := \alpha(t,y)(x-\hat{w})^2 + \beta(t,y) \tag{4.14}$$

solves the HJB equation (4.10). Moreover, a minimizer $F^{\mathrm{opt}} = F^{\mathrm{opt}}_\alpha(t,x,y;\lambda;\cdot) \in \mathcal{A}$ is

$$F^{\mathrm{opt}}_\alpha(t,x,y;\lambda;u) = m^{\mathrm{opt}}_\alpha(t,x,y) + \theta^{\mathrm{opt}}_\alpha(t,y;\lambda)^{\frac{1}{2}}u, \tag{4.15}$$

where

$$m^{\mathrm{opt}}_\alpha(t,x,y) := -(x-\hat{w})(\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha)(t,y) \quad \textit{and} \quad \theta^{\mathrm{opt}}_\alpha(t,y;\lambda) := \frac{\lambda}{2}\mathcal{S}_\alpha^{-1}(t,y). \tag{4.16}$$

Proof. One first notices that $\mathcal{L}_Y\alpha(t,\cdot)$ and $\mathcal{L}_Y\beta(t,\cdot)$ are well-defined functions for $t\in[0,T]$. To simplify the presentation, we omit the argument y of coefficient functions b, A, γ, and for fixed (t,y), we formally use the following notations for α (and analogously for β),

$$\alpha := \alpha(t,y), \quad \tilde{\alpha}(e) := \alpha(t,y+\gamma(y)e), \quad \mathcal{L}_Y\alpha := \mathcal{L}_Y\alpha(t,y).$$

Plugging the ansatz (4.14) into the HJB equation (4.10) and rearranging terms we get the following which holds pointwise on $[0,T)\times\mathbb{R}\times\mathbb{R}^D$,

$$\begin{aligned} 0 = {} & (x-\hat{w})^2(\partial_t\alpha + \mathcal{L}_Y\alpha) + (\partial_t\beta + \mathcal{L}_Y\beta) \\ & + \min_{F\in\mathcal{A},\,\theta^F\in\mathbb{S}^D_{++}} \bigg\{\alpha\Big((m^F)^\mathsf{T}Am^F + \mathbf{tr}[A\theta^F]\Big) + 2(x-w)(m^F)^\mathsf{T}(A\nabla_y\alpha + \alpha b) \\ & + \int_E \Big(\tilde{\alpha}(e)e^\mathsf{T}\gamma^\mathsf{T}(\theta^F + m^F(m^F)^\mathsf{T})\gamma e + 2(x-w)(\tilde{\alpha}(e)-\alpha)(m^F)^\mathsf{T}\gamma e\Big)\nu(\mathrm{d}e) - \lambda\mathsf{Ent}(F)\bigg\}, \end{aligned} \tag{4.17}$$

where the minimization is taken over $F\in\mathcal{A}$ with $\theta^F \in \mathbb{S}^D_{++}$ due to Remark 4.5. Remark that given any $m\in\mathbb{R}^D$, $\theta\in\mathbb{S}^D_{++}$, there always exists an $F\in\mathcal{A}$ such that $m^F = m$ and $\theta^F = \theta$, for example, one might take $F(u) = m + \theta^{\frac{1}{2}}u$. Then the minimum over $F\in\mathcal{A}$ with $\theta^F\in\mathbb{S}^D_{++}$ in (4.17) can be separated into two individual minimization problems, one is over $m^F\in\mathbb{R}^D$ and

the other is over $\theta^F \in \mathbb{S}^D_{++}$. Specifically, let $\Psi^F_{(4.17)}$ denote the expression inside the minimum in (4.17), then one has

$$\begin{aligned}\min_{F\in\mathcal{A},\,\theta^F\in\mathbb{S}^D_{++}} \Psi^F_{(4.17)} &= \min_{m\in\mathbb{R}^D}\Big\{\alpha m^\mathsf{T} A m + 2(x-w)m^\mathsf{T}(A\nabla_y\alpha + \alpha b)\\ &\qquad + \int_E \Big(\tilde\alpha(e)(m^\mathsf{T}\gamma e)^2 + 2(x-w)(\tilde\alpha(e)-\alpha)m^\mathsf{T}\gamma e\Big)\nu(\mathrm{d}e)\Big\}\\ &\quad + \min_{\theta\in\mathbb{S}^D_{++}}\Big\{\alpha\mathbf{tr}[A\theta] + \int_E \tilde\alpha(e)e^\mathsf{T}\gamma^\mathsf{T}\theta\gamma e\,\nu(\mathrm{d}e) - \lambda\max_{F\in\mathcal{A},\,\theta^F=\theta}\mathsf{Ent}(F)\Big\}\\ &=: \min_{m\in\mathbb{R}^D} f_{(4.18)}(m) + \min_{\theta\in\mathbb{S}^D_{++}} g_{(4.18)}(\theta). \end{aligned} \tag{4.18}$$

It is known that the differential entropy is translation invariant and it is maximized over all distributions with a given covariance matrix by Gaussian distribution, see, e.g., [9, Theorem 8.6.5]. Hence, $g_{(4.18)}$ can be expressed as

$$g_{(4.18)}(\theta) = \alpha\mathbf{tr}[A\theta] + \int_E \tilde\alpha(e)e^\mathsf{T}\gamma^\mathsf{T}\theta\gamma e\,\nu(\mathrm{d}e) - \frac{\lambda}{2}\log(\det(\theta)) - \frac{\lambda D}{2}\log(2\pi\mathrm{e}).$$

Combining (4.17) with (4.18) yields the equation

$$(x-\hat w)^2(\partial_t\alpha + \mathcal{L}_Y\alpha) + (\partial_t\beta + \mathcal{L}_Y\beta) + \min_{m\in\mathbb{R}^D} f_{(4.18)}(m) + \min_{\theta\in\mathbb{S}^D_{++}} g_{(4.18)}(\theta) = 0. \tag{4.19}$$

We first consider the minimization problem

$$\min_{\theta\in\mathbb{S}^D_{++}} g_{(4.18)}(\theta).$$

By vectorization, $\mathbb{S}^D_{++}$ can be regarded as an open subset of $\mathbb{R}^{D(D+1)/2}$, where the openness (under the Euclidean norm) can be inferred from Sylvester's criterion, so that $g_{(4.18)}$ becomes a function defined on $\mathbb{S}^D_{++} \subset \mathbb{R}^{D(D+1)/2}$. Since $\theta \mapsto -\log(\det(\theta))$ is a convex and differentiable function on $\mathbb{S}^D_{++}$, it implies that $g_{(4.18)}$ is also convex and differentiable. Hence, solutions of $\nabla g_{(4.18)}$ globally minimize $g_{(4.18)}$ on $\mathbb{S}^D_{++}$. To find its solutions, we represent $\theta = (\theta^{(1,1)},\dots,\theta^{(D,1)},\theta^{(2,2)},\dots\theta^{(D,2)},\dots,\theta^{(D,D)})^\mathsf{T} \in \mathbb{R}^{D(D+1)/2}$. Then, for $1 \le j \le i \le D$, according to [17, p.311, Eq. (8.12)] one has

$$\frac{\partial\log\det(\theta)}{\partial\theta^{(i,j)}} = [2\theta^{-1} - \mathrm{diag}(\theta^{-1})]^{(i,j)}$$

so that the partial derivatives of $g_{(4.18)}$ are computed by

$$\begin{aligned}&\frac{\partial g_{(4.18)}}{\partial\theta^{(i,j)}}(\theta)\\ &= \alpha[2A - \mathrm{diag}(A)]^{(i,j)} + \int_E \tilde\alpha(e)[2\gamma ee^\mathsf{T}\gamma^\mathsf{T} - \mathrm{diag}(\gamma ee^\mathsf{T}\gamma^\mathsf{T})]^{(i,j)}\nu(\mathrm{d}e) - \frac{\lambda}{2}[2\theta^{-1} - \mathrm{diag}(\theta^{-1})]^{(i,j)}.\end{aligned}$$

Solving $\nabla g_{(4.18)}(\theta) = 0$ we get the solution $\theta = \theta^{\mathrm{opt}}_\alpha(t,y;\lambda)$ as provided in (4.16). Hence, $\theta^{\mathrm{opt}}_\alpha(t,y;\lambda)$ is a global minimizer of $g_{(4.18)}$ on $\mathbb{S}^D_{++}$. We next investigation the problem

$$\min_{m\in\mathbb{R}^D} f_{(4.18)}(m).$$

Solving $\nabla f_{(4.18)}(m) = 0$ yields the solution $m = m^{\mathrm{opt}}_\alpha(t,x,y)$ which is provided in (4.16). Moreover, since

$$\nabla^2 f_{(4.18)} = 2\alpha A + 2\gamma\Big(\int_E \tilde\alpha(e)ee^\mathsf{T}\nu(\mathrm{d}e)\Big)\gamma^\mathsf{T} = 2\mathcal{S}_\alpha$$

and $\mathcal{S}_\alpha \in \mathbb{S}^D_{++}$ as claimed above, we infer that $m^{\mathrm{opt}}_\alpha(t,x,y)$ is a global minimizer of $f_{(4.18)}$ on $\mathbb{R}^D$. Plugging these minimizers back into (4.19) and noticing that

$$\mathbf{tr}[\alpha A \mathcal{S}_\alpha^{-1}] = \mathbf{tr}\left[I_D - \int_E \tilde{\alpha}(e)\gamma e e^\mathsf{T}\gamma^\mathsf{T}\mathcal{S}_\alpha^{-1}\nu(\mathrm{d}e)\right] = D - \int_E \tilde{\alpha}(e) e^\mathsf{T}\gamma^\mathsf{T}\mathcal{S}_\alpha^{-1}\gamma e\,\nu(\mathrm{d}e)$$

we eventually arrive at the equation

$$0 = (x-\hat{w})^2\Big(\partial_t\alpha + \mathcal{L}_Y\alpha - \mathcal{M}_\alpha^\mathsf{T}\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha\Big) + \left(\partial_t\beta + \mathcal{L}_Y\beta - \frac{\lambda}{2}\log\left(\frac{(\lambda\pi)^D}{\det(\mathcal{S}_\alpha)}\right)\right)$$

which holds true according to assumption (4.13). As a consequence, the function provided in (4.15) is an optimal solution of (4.17). □

4.2.3. *Verification argument.* In the following result, the coefficients $K \in \{b, a, \gamma, A, \Sigma\}$ are conveniently extended to be defined on $[0,T]\times\mathbb{R}^D$ by setting $K(t,y) := K(y)$. We recall $\mathcal{M}_\alpha$ and $\mathcal{S}_\alpha$ from (4.11) and (4.12) respectively.

Theorem 4.10. *Let α, β satisfy the assumptions of Proposition 4.9. Let $(t,x,y) \in [0,T)\times\mathbb{R}\times\mathbb{R}^D$ and recall $Y^{t,y}$ in (4.3). Assume furthermore that $\{\beta(\tau, Y^{t,y}_\tau)\,|\,\tau\colon\Omega\to[t,T]$ is a stopping time$\}$ is uniformly integrable and that α is bounded on $[0,T]\times\mathbb{R}^D$ and satisfies*

$$\int_t^T |(\mathcal{M}_\alpha^\mathsf{T}\mathcal{S}_\alpha^{-1}b)(s,Y^{t,y}_{s-})|^2\mathrm{d}s + \sup_{s\in(t,T)}|(\mathcal{M}_\alpha^\mathsf{T}\mathcal{S}_\alpha^{-1}\Sigma\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha)(s,Y^{t,y}_{s-})| \le c_{(4.20)} \quad a.s., \tag{4.20}$$

$$\mathbb{E}\left[\int_t^T \mathbf{tr}[(\Sigma\mathcal{S}_\alpha^{-1})(s,Y^{t,y}_{s-})]\mathrm{d}s\right] + \mathbb{E}\left[\int_t^T \left|\log\Big(\det(\mathcal{S}_\alpha(s,Y^{t,y}_{s-}))\Big)\right|\mathrm{d}s\right] < \infty, \tag{4.21}$$

for some non-random constant $c_{(4.20)} > 0$. Then a solution for Problem 4.3 is

$$H^{t,x,y;*}_s(u) = -(X^{t,x,y;*}_{s-} - \hat{w})(\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha)(s,Y^{t,y}_{s-}) + \sqrt{\frac{\lambda}{2}}\mathcal{S}_\alpha^{-\frac{1}{2}}(s,Y^{t,y}_{s-})u, \quad s\in(t,T], u\in\mathbb{R}^D, \tag{4.22}$$

with $H^{t,x,y;*}_t(u) := -(x-\hat{w})(\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha)(t,y) + \sqrt{\frac{\lambda}{2}}\mathcal{S}_\alpha^{-\frac{1}{2}}(t,y)u$, and the corresponding optimal wealth process $X^{t,x,y;*} = (X^{t,x,y;*}_s)_{s\in[t,T]}$ is a unique càdlàg (strong) solution to the SDE on $[t,T]$,

$$\mathrm{d}X^{t,x,y;*}_s = -(X^{t,x,y;*}_{s-} - \hat{w})\mathrm{d}Z^{t,y}_s + \sqrt{\frac{\lambda}{2}}\mathrm{d}M^{t,y}_s, \quad X^{t,x,y;*}_t = x. \tag{4.23}$$

Here $Z^{t,y} = (Z^{t,y}_s)_{s\in[t,T]}$, $M^{t,y} = (M^{t,y}_s)_{s\in[t,T]}$ are càdlàg with $Z^{t,y}_t = 0, M^{t,y}_t = 0$ given by

$$\begin{cases} \mathrm{d}Z^{t,y}_s = (\mathcal{M}_\alpha^\mathsf{T}\mathcal{S}_\alpha^{-1})(s,Y^{t,y}_{s-})\mathrm{d}Y^{t,y}_s, \\ \mathrm{d}M^{t,y}_s = \mathbf{tr}[(\mathcal{S}_\alpha^{-\frac{1}{2}}a)(s,Y^{t,y}_{s-})\mathrm{d}\mathcal{W}_s^\mathsf{T}] + \displaystyle\int_{E\times\mathbb{R}^D}\left(\mathcal{S}_\alpha^{-\frac{1}{2}}(s,Y^{t,y}_{s-})\frac{u}{\psi(e)}\right)^\mathsf{T}\gamma(Y^{t,y}_{s-})e\,\widetilde{N}^\psi_L(\mathrm{d}s,\mathrm{d}e,\mathrm{d}u). \end{cases}$$

The value function is $V^*(\cdot|\hat{w}) = v^{\mathrm{opt}}$, where v^{opt} is provided in (4.14).

Remark 4.11. The RL algorithms developed in [20, 21] learn the value function and the optimal measure-valued control in parametric classes of functions and probability measures (which have to be chosen beforehand). The structural results on the optimal value and the optimal control obtained in Proposition 4.9 and Theorem 4.10 facilitate such a parametrization. Indeed, Proposition 4.9 shows that the optimal value function is quadratic in the portfolio wealth with coefficients which can be computed in terms of the functions α and β. Moreover, the optimal control law is Gaussian with mean $-(X^{t,x,y;*}_- - \hat{w})(\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha)(\cdot,Y^{t,y}_-)$ and covariance matrix $\frac{\lambda}{2}\mathcal{S}_\alpha^{-1}(\cdot,Y^{t,y}_-)$ by Theorem 4.10 (and, hence, mean and covariance matrix do not depend on β). Under additional structural assumptions on the stock price model, see Example 4.14 below, the

PIDE for α can be solved in closed form, leading to an explicit parametrization for the optimal control law.

We also remark that, in general, the mean of the optimal control linearly depends on the associated portfolio wealth, while its covariance matrix is independent of the portfolio wealth.

Proof of Theorem 4.10. Let us fix $(t,x,y) \in [0,T)\times\mathbb{R}\times\mathbb{R}^D$. For the sake of notational simplicity, in the presentation below we omit the super-scripts (t,y) and (t,x,y) in relevant processes such as $Y^{t,y}$ in (4.3), $X^{t,x,y;H}$ in (4.5), and $M^{t,y}$, $Z^{t,y}$. Since $\mathbb{E}\big[\int_t^T \mathbf{tr}[(\Sigma\mathcal{S}_\alpha^{-1})(s,Y_{s-})]\mathrm{d}s\big] < \infty$ by (4.21), it implies that M is a uniformly square integrable martingale with $\mathbb{E}[\max_{t\le s\le T}|M_s|^2] < \infty$ due to Doob's maximal inequality. By assumption (4.20), we apply Lemma B.2 to infer that the SDE (4.23) has a unique càdlàg solution X^* with

$$\mathbb{E}\bigg[\sup_{t\le s\le T}|X_s^*|^2\bigg] < \infty. \tag{4.24}$$

Step 1. Take $H \in \mathcal{A}(t,y)$ arbitrarily. For v^{opt} given in (4.14), one has

$$V^H(t,x,y|\hat{w}) = \mathbb{E}\bigg[v^{\mathrm{opt}}(T,X_T^H,Y_T) + \lambda\int_t^T\int_{\mathbb{R}} p_s^H(u)\log p_s^H(u)\mathrm{d}u\mathrm{d}s\bigg].$$

Applying Itô's formula (see, e.g., [22, Theorem 2.5]) for $v^{\mathrm{opt}} \in C^{1,2}([0,T]\times\mathbb{R}^{1+D})$ and X^H, Y we obtain, a.s., for $t < r \le T$,

$$\begin{aligned}
v^{\mathrm{opt}}(r,X_r^H,Y_r) - v^{\mathrm{opt}}(t,x,y) &= \int_t^r \partial_t v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})\mathrm{d}s\\
&+ \int_t^r \partial_x v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})(\mu_s^H)^\mathsf{T} b(Y_{s-})\mathrm{d}s + \int_t^r \nabla_y v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})^\mathsf{T} b(Y_{s-})\mathrm{d}s\\
&+ \int_t^r \partial_x v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})\Big((\mu_s^H)^\mathsf{T} a(Y_{s-})\mathrm{d}W_s + \mathbf{tr}[(\Theta_s^H)^{\frac12}a(Y_{s-})\mathrm{d}\mathcal{W}_s^\mathsf{T}]\Big)\\
&+ \int_t^r \nabla_y v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})^\mathsf{T} a(Y_{s-})\mathrm{d}W_s\\
&+ \frac12\int_t^r \partial_{xx}^2 v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})\Big((\mu_s^H)^\mathsf{T} A(Y_{s-})\mu_s^H + \mathbf{tr}[A(Y_{s-})\Theta_s^H]\Big)\mathrm{d}s\\
&+ \int_t^r \Big((\mu_s^H)^\mathsf{T} A(Y_{s-})\nabla_{xy}^2 v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-}) + \frac12\mathbf{tr}[\nabla_{yy}^2 v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})A(Y_{s-})]\Big)\mathrm{d}s\\
&+ \int_{(t,r]\times E\times\mathbb{R}^D}\bigg[v^{\mathrm{opt}}\bigg(s,X_{s-}^H + H_s\Big(\frac{u}{\psi(e)}\Big)^\mathsf{T}\gamma(Y_{s-})e, Y_{s-}+\gamma(Y_{s-})e\bigg)\\
&\qquad\qquad - v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})\bigg]\widetilde{N}_L^\psi(\mathrm{d}s,\mathrm{d}e,\mathrm{d}u)\\
&+ \int_{(t,r]\times E\times\mathbb{R}^D}\bigg[v^{\mathrm{opt}}\bigg(s,X_{s-}^H + H_s\Big(\frac{u}{\psi(e)}\Big)^\mathsf{T}\gamma(Y_{s-})e, Y_{s-}+\gamma(Y_{s-})e\bigg) - v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})\\
&\quad - \partial_x v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})H_s\Big(\frac{u}{\psi(e)}\Big)^\mathsf{T}\gamma(Y_{s-})e - \nabla_y v^{\mathrm{opt}}(s,X_{s-}^H,Y_{s-})^\mathsf{T}\gamma(Y_{s-})e\bigg]\nu_L^\psi(\mathrm{d}e,\mathrm{d}u)\mathrm{d}s.
\end{aligned} \tag{4.25}$$

We let $z := \frac{u}{\psi(e)}$ and denote by $P(s,e,z)$ the integrand against $\nu_L^\psi(\mathrm{d}e,\mathrm{d}u)\mathrm{d}s$ in (4.25). It follows from the explicit form of v^{opt} that

$$\begin{aligned}
P(s,e,z) &= \alpha(s,Y_{s-}+\gamma(Y_{s-})e)(X_{s-}^H + H_s(z)^\mathsf{T}\gamma(Y_{s-})e - \hat{w})^2 - \alpha(s,Y_{s-})(X_{s-}^H-\hat{w})^2\\
&\quad + \beta(s,Y_{s-}+\gamma(Y_{s-})e) - \beta(s,Y_{s-}) - 2\alpha(s,Y_{s-})(X_{s-}^H-\hat{w})H_s(z)^\mathsf{T}\gamma(Y_{s-})e
\end{aligned}$$

$$
\begin{aligned}
&- \nabla_y \alpha(s, Y_{s-})^\mathsf{T}(X^H_{s-} - \hat{w})^2 \gamma(Y_{s-})e - \nabla_y \beta(s, Y_{s-})^\mathsf{T}\gamma(Y_{s-})e \\
&= (X^H_{s-} - \hat{w})^2 \Big[\alpha(s, Y_{s-} + \gamma(Y_{s-})e) - \alpha(s, Y_{s-}) - \nabla_y \alpha(s, Y_{s-})^\mathsf{T}\gamma(Y_{s-})e\Big] \\
&\quad + 2(X^H_{s-} - \hat{w})\Big[\alpha(s, Y_{s-} + \gamma(Y_{s-})e) - \alpha(s, Y_{s-})\Big] H_s(z)^\mathsf{T}\gamma(Y_{s-})e \\
&\quad + \alpha(s, Y_{s-} + \gamma(Y_{s-})e)(H_s(z)^\mathsf{T}\gamma(Y_{s-})e)^2 \\
&\quad + \beta(s, Y_{s-} + \gamma(Y_{s-})e) - \beta(s, Y_{s-}) - \nabla_y \beta(s, Y_{s-})^\mathsf{T}\gamma(Y_{s-})e.
\end{aligned}
$$

Let $\Upsilon^{(0)}_\alpha, \Upsilon^{(1)}_\alpha, \Upsilon^{(2)}_\alpha$ and $\Upsilon^{(2)}_\beta$ respectively be (continuous) Υ-dominating functions of α and β in the sense of Definition 4.7. Then, for some constant $c_D > 0$ depending only on D, we get, a.s,

$$
\begin{aligned}
&\int_t^T \int_{E\times\mathbb{R}^D} |P(s,e,z)|\nu^\psi_L(\mathrm{d}e,\mathrm{d}u)\mathrm{d}s = \int_t^T \int_{E\times\mathbb{R}^D} |P(s,e,u)|\nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u\mathrm{d}s \\
&\le \int_t^T (X^H_{s-} - \hat{w})^2 \Upsilon^{(2)}_\alpha(s, Y_{s-})\mathrm{d}s \\
&\quad + 2c_D \int_t^T |X^H_{s-} - \hat{w}| \bigg(\int_{\mathbb{R}^D} \|H_s(u)\|\varphi_D(u)\mathrm{d}u \bigg) \|\gamma(Y_{s-})\| \Upsilon^{(1)}_\alpha(s, Y_{s-})\mathrm{d}s \\
&\quad + c_D \int_t^T \|\gamma(Y_{s-})\|^2 \Upsilon^{(0)}_\alpha(s, Y_{s-}) \bigg(\int_{\mathbb{R}^D} \|H_s(u)\|^2 \varphi_D(u)\mathrm{d}u \bigg) \mathrm{d}s + \int_t^T \Upsilon^{(2)}_\beta(s, Y_{s-})\mathrm{d}s \\
&< \infty,
\end{aligned}
$$

where we use the càdlàg property of X^H, Y and assumption (4.1) to deduce the finiteness. Let $Q(s,e,z)$ denote the integrand against $\widetilde{N}^\psi_L$ in (4.25) and define

$$
\begin{aligned}
R(s,e,z) &:= \Big[2\alpha(s, Y_{s-})(X^H_{s-} - \hat{w})H_s(z) + \nabla_y \alpha(s, Y_{s-})(X^H_{s-} - \hat{w})^2 + \nabla_y \beta(s, Y_{s-})\Big]^\mathsf{T} \gamma(Y_{s-})e \\
&=: \widetilde{R}(s,z)^\mathsf{T}\gamma(Y_{s-})e
\end{aligned}
$$

so that

$$
Q(s,e,z) = P(s,e,z) + R(s,e,z).
$$

Then, there is a constant $c'_D > 0$ such that, a.s.,

$$
\begin{aligned}
&\int_t^T \int_{E\times\mathbb{R}^D} |R(s,e,z)|^2 \nu^\psi_L(\mathrm{d}e,\mathrm{d}u)\mathrm{d}s = \int_t^T \int_{E\times\mathbb{R}^D} |R(s,e,u)|^2 \nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u\mathrm{d}s \\
&\le c'_D \bigg(\int_E \|e\|^2 \nu(\mathrm{d}e) \bigg) \int_t^T \int_{\mathbb{R}^D} \|\widetilde{R}(s,u)\|^2 \|\gamma(Y_{s-})\|^2 \varphi_D(u)\mathrm{d}u\mathrm{d}s \\
&< \infty.
\end{aligned}
$$

On the other hand, by rearranging terms we get a predictable process ϕ^H and a local martingale U^H null at t such that

$$
v^{\mathrm{opt}}(r, X^H_r, Y_r) - v^{\mathrm{opt}}(t,x,y) = \int_t^r \phi^H_s \mathrm{d}s + U^H_r.
$$

Since v^{opt} solve the HJB equation (4.17) and any $H \in \mathcal{A}(t,y)$ is sub-optimal in general, we arrive at, a.s,

$$
v^{\mathrm{opt}}(r, X^H_r, Y_r) - v^{\mathrm{opt}}(t,x,y) \ge -\lambda \int_t^r \int_{\mathbb{R}^D} p^H_s(u) \log p^H_s(u)\mathrm{d}u\mathrm{d}s + U^H_r. \tag{4.26}
$$

To deal with U^H, we define the localizing sequence $(\tau_n)_{n\ge1}$ as follows

$$
\tau_n := T \wedge \inf \bigg\{ r \in (t,T] : \int_t^r \bigg(\int_{E\times\mathbb{R}^D} (|P(s,e,u)| + |R(s,e,u)|^2)\nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u
$$

$$+\nabla_y v^{\rm opt}(s,X^H_{s-},Y_{s-})^{\mathsf T}A(Y_{s-})\nabla_y v^{\rm opt}(s,X^H_{s-},Y_{s-})$$
$$+|\partial_x v^{\rm opt}(s,X^H_{s-},Y_{s-})|^2\Big((\mu^H_s)^{\mathsf T}A(Y_{s-})\mu^H_s+\mathbf{tr}[A(Y_{s-})\Theta^H_s]\Big)\Big)\mathrm{d}s\geq n\Big\}.$$

Since the integrand against $\mathrm{d}s$ in the definition of τ_n is integrable on $[t,T]$ a.s., the integral $\int_t^r(\cdots)\mathrm{d}s$ is finite and non-decreasing in r a.s., and hence $(\tau_n)_{n\geq 1}$ is a non-decreasing sequence of stopping times converging a.s. to T as $n\to\infty$. We note that the local martingale U^H on the right-hand side of (4.26) is an integrable martingale null at t when stopping at τ_n, and hence, vanishes when taking the expectation. Therefore,

$$\begin{aligned}v^{\rm opt}(t,x,y)&\leq\mathbb{E}\bigg[v^{\rm opt}(\tau_n,X^H_{\tau_n},Y_{\tau_n})+\lambda\int_t^{\tau_n}\int_{\mathbb{R}^D}p^H_s(u)\log p^H_s(u)\mathrm{d}u\bigg]\\&=\mathbb{E}\bigg[\alpha(\tau_n,Y_{\tau_n})(X^H_{\tau_n}-\hat w)^2+\beta(\tau_n,Y_{\tau_n})+\lambda\int_t^{\tau_n}\int_{\mathbb{R}^D}p^H_s(u)\log p^H_s(u)\mathrm{d}u\bigg].\end{aligned}$$

By assumption, α is continuous and bounded, β is continuous and $\{\beta(\tau_n,Y_{\tau_n})\}_{n\geq1}$ is uniformly integrable, and the entropy term is also uniform integrable for $H\in\mathcal{A}(t,y)$, we exploit (4.6) and use the dominated convergence theorem with keeping in mind that $(\tau_n)_{n\geq1}$ is a.s. eventually constant T to get

$$v^{\rm opt}(t,x,y)\leq\mathbb{E}\bigg[v^{\rm opt}(T,X^H_T,Y_T)+\lambda\int_t^T\int_{\mathbb{R}^D}p^H_s(u)\log p^H_s(u)\mathrm{d}u\bigg]=V^H(t,x,y|\hat w).$$

Since $H\in\mathcal{A}(t,y)$ is arbitrary, it implies that $v^{\rm opt}(t,x,y)\leq V^*(t,x,y|\hat w)$.

Step 2. As suggested by (4.15), H^* provided in (4.22) is a candidate for optimal controls. If H^* is admissible, then we can apply the arguments in ***Step 1*** for H^*, where inequality (4.26) becomes an equality, to obtain

$$v^{\rm opt}(t,x,y)=V^{H^*}(t,x,y|\hat w).$$

Hence $v^{\rm opt}(t,x,y)=V^*(t,x,y|\hat w)$. It remains to show that H^* is admissible by verifying the requirements in Definition 4.1. Condition (H1) is obvious from the definition of H^*. For (H2), one has

$$\mu^{H^*}_s=-(X^*_{s-}-\hat w)(\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha)(s,Y_{s-})\quad\text{and}\quad\Theta^{H^*}_s=\frac{\lambda}{2}\mathcal{S}_\alpha^{-1}(s,Y_{s-}).$$

Condition (4.1) is straightforward due to the càdlàg property of X^*,Y and the continuity of $\mathcal{M}_\alpha$, $\mathcal{S}_\alpha^{-1}$ on $[0,T]\times\mathbb{R}^D$. For (4.2), expanding the square and using $\int_{\mathbb{R}^D}u\varphi_D(u)\mathrm{d}u=0$ and $\int_{\mathbb{R}^D}uu^{\mathsf T}\varphi_D(u)\mathrm{d}u=I_D$ in the jump part we get

$$\begin{aligned}&(\mu^{H^*}_s)^{\mathsf T}A(Y_{s-})\mu^{H^*}_s+\mathbf{tr}[A(Y_{s-})\Theta^{H^*}_s]+\int_{E\times\mathbb{R}^D}|H^*_s(u)^{\mathsf T}\gamma(Y_{s-})e|^2\nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u\\&=(X^*_{s-}-\hat w)^2(\mathcal{M}_\alpha^{\mathsf T}\mathcal{S}_\alpha^{-1}A\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha)(s,Y_{s-})+\frac{\lambda}{2}\mathbf{tr}[(A\mathcal{S}_\alpha^{-1})(s,Y_{s-})]\\&+(X^*_{s-}-\hat w)^2\bigg(\mathcal{M}_\alpha^{\mathsf T}\mathcal{S}_\alpha^{-1}\gamma\int_E ee^{\mathsf T}\nu(\mathrm{d}e)\gamma^{\mathsf T}\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha\bigg)(s,Y_{s-})+\frac{\lambda}{2}\mathbf{tr}\bigg[\bigg(\gamma\int_E ee^{\mathsf T}\nu(\mathrm{d}e)\gamma^{\mathsf T}\mathcal{S}_\alpha^{-1}\bigg)(s,Y_{s-})\bigg]\\&=(X^*_{s-}-\hat w)^2(\mathcal{M}_\alpha^{\mathsf T}\mathcal{S}_\alpha^{-1}\Sigma\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha)(s,Y_{s-})+\frac{\lambda}{2}\mathbf{tr}[(\Sigma\mathcal{S}_\alpha^{-1})(s,Y_{s-})].\end{aligned}$$

In addition, using Hölder's inequality yields

$$\bigg|\int_t^T|(\mu^{H^*}_s)^{\mathsf T}b(s,Y_{s-})|\mathrm{d}s\bigg|^2\leq\bigg(\int_t^T(X^*_{s-}-\hat w)^2\mathrm{d}s\bigg)\bigg(\int_t^T|(\mathcal{M}_\alpha^{\mathsf T}\mathcal{S}_\alpha^{-1}b)(s,Y_{s-})|^2\mathrm{d}s\bigg).$$

Hence (4.2) is satisfied by using (4.20), (4.21) and (4.24). To verify (H3), we might take $p_s^{H^*}(\cdot)$ to be the continuous density function of the Gaussian distribution $\mathcal{N}(\mu_s^{H^*}, \Theta_s^{H^*})$ with mean $\mu_s^{H^*}$ and covariance matrix $\Theta_s^{H^*}$, and then (4.4) follows from (4.21). □

4.3. **Explicit solutions of optimal exploratory SDEs and Lagrange multipliers.** As an advantage of our approach, the optimal exploratory dynamic (4.23) is a linear SDE with jumps which enables us to find its solutions in a closed-form. As a consequence, we can also explicitly determine the Lagrange multiplier $\hat{w}$ using the constraint $\mathbb{E}[X_T^{H^*}] = \hat{z}$, where H^* is given in (4.22).

We consider Problem 4.3 and assume the assumptions of Theorem 4.10 for $(t, x, y) = (0, x_0, y_0)$, and omit super-scripts $(0, y_0)$ and $(0, x_0, y_0)$ in relevant processes.

Proposition 4.12. *Under the assumptions of Theorem 4.10, if $\Delta Z \neq 1$ on $[0,T]$ then the optimal wealth process $X^* = (X_r^*)_{r\in[0,T]}$ in (4.23) is given by*

$$X_r^* = \hat{w} + \left[x_0 - \hat{w} + \sqrt{\frac{\lambda}{2}}\left(\int_0^r \frac{\mathrm{d}M_s}{\mathcal{E}(-Z)_{s-}} + \int_0^r \frac{\mathrm{d}[M,Z]_s}{\mathcal{E}(-Z)_{s-}}\right)\right]\mathcal{E}(-Z)_r, \quad r \in [0,T], \tag{4.27}$$

where $\mathcal{E}(-Z) = (\mathcal{E}(-Z)_r)_{r\in[0,T]}$ denotes the Doléans–Dade exponential[2] of $-Z$, i.e.

$$\mathcal{E}(-Z)_0 = 1, \quad \mathcal{E}(-Z)_r = \exp\left(-Z_r - \frac{1}{2}\int_0^r \mathrm{d}[Z,Z]_s^c\right)\prod_{0<s\le r}(1-\Delta Z_s)\mathrm{e}^{\Delta Z_s}, \quad r \in (0,T].$$

Here the quadratic covariation terms are explicitly expressed as follows

$$\begin{aligned}\mathrm{d}[M,Z]_s &= \int_{E\times\mathbb{R}^D} \frac{1}{\psi(e)} e^{\intercal}\left[\gamma^{\intercal}\mathfrak{S}_\alpha^{-1}\mathcal{M}_\alpha(\mathfrak{S}_\alpha^{-\frac{1}{2}}u)^{\intercal}\gamma\right](s,Y_{s-})e\, N_L^{\psi}(\mathrm{d}s,\mathrm{d}e,\mathrm{d}u),\\ \mathrm{d}[Z,Z]_s^c &= (\mathcal{M}_\alpha^{\intercal}\mathfrak{S}_\alpha^{-1}A\mathfrak{S}_\alpha^{-1}\mathcal{M}_\alpha)(s,Y_{s-})\mathrm{d}s.\end{aligned}$$

Moreover, if $\mathbb{E}[\mathcal{E}(-Z)_T] \neq 1$ then the Lagrange multiplier $\hat{w}$ (such that $\mathbb{E}[X_T^*] = \hat{z}$) is given by

$$\hat{w} = \frac{1}{1-\mathbb{E}[\mathcal{E}(-Z)_T]}\left(\hat{z} - \sqrt{\frac{\lambda}{2}}\mathbb{E}\left[\left(\int_0^T \frac{\mathrm{d}M_s}{\mathcal{E}(-Z)_{s-}} + \int_0^T \frac{\mathrm{d}[M,Z]_s}{\mathcal{E}(-Z)_{s-}}\right)\mathcal{E}(-Z)_T\right] - x_0\mathbb{E}[\mathcal{E}(-Z)_T]\right). \tag{4.28}$$

Proof. For Z given in Theorem 4.10, we write

$$-(X_{s-}^* - \hat{w})\mathrm{d}Z_s = (X_{s-}^* - \hat{w})\mathrm{d}(-Z_s).$$

Since $\Delta Z \neq 1$ by assumption, it implies that $\inf\{s \in (0,T] : 1 - \Delta Z_s = 0\} = \infty$ a.s. We then apply [30, Exercise V.27] to obtain the explicit representation for X^* as in (4.27).

For the Lagrange multiplier $\hat{w}$, we first notice that $\mathcal{E}(-Z)$ satisfies the following SDE on $[0,T]$

$$\mathcal{E}(-Z)_r = 1 + \int_0^r \mathcal{E}(-Z)_{s-}\mathrm{d}(-Z_s).$$

Since the *conditional* quadratic variation[3] of the integrator $-Z$ is

$$\langle -Z, -Z\rangle_T = \int_0^T (\mathcal{M}_\alpha^{\intercal}\mathfrak{S}_\alpha^{-1}\Sigma\mathfrak{S}_\alpha^{-1}\mathcal{M}_\alpha)(s,Y_{s-})\mathrm{d}s \le c_{(4.20)}T,$$

it follows from condition (4.20) and Lemma B.2 that $\mathcal{E}(-Z)$ is square integrable. Since X^* is also square integrable by (4.24), letting $r = T$ and taking the expectation both sides of (4.27) we rearrange terms and use the constraint $\mathbb{E}[X_T^*] = \hat{z}$ to obtain (4.28). □

[2] See, e.g., [30, Section II.8].

[3] See, e.g., [30, Chapter III, p.124].

Remark 4.13. (1) If ν is absolutely continuous with respect to the D-dimensional Lebesgue measure $\boldsymbol{\lambda}_D$, then $\Delta Z \neq 1$ on $[0,T]$. Indeed, by letting $J_t := \int_0^t \int_{E\times\mathbb{R}^D} e\widetilde{N}_L^{\psi}(\mathrm{d}s,\mathrm{d}e,\mathrm{d}u)$ so that J admits ν as its Lévy measure we get

$$\begin{aligned}\mathbb{E}[\#\{s\in(0,T]:\Delta Z_s=1\}] &= \mathbb{E}[\#\{s\in(0,T]:(\mathcal{M}_\alpha^{\mathsf{T}}\mathcal{S}_\alpha^{-1}\gamma)(s,Y_{s-})\Delta J_s=1\}]\\ &= \mathbb{E}\bigg[\int_0^T\int_E \mathbb{1}_{\{(\mathcal{M}_\alpha^{\mathsf{T}}\mathcal{S}_\alpha^{-1}\gamma)(s,Y_{s-})e=1\}}\nu(\mathrm{d}e)\mathrm{d}s\bigg]\\ &= 0,\end{aligned}$$

where we combine Fubini's theorem with the fact that hyperplanes have Lebesgue measure zero to obtain the last equality. Hence, $\#\{s\in(0,T]:\Delta Z_s=1\}=0$ a.s.

(2) If the condition "$\Delta Z\neq 1$ on $[0,T]$" in Proposition 4.12 is not satisfied, then we can still obtain explicit representations of X^* and $\hat{w}$. However, as these expressions are rather technical, we refer the interested readers to [5] for more details.

4.4. **Illustrative examples.** Let us consider some situations in which assumptions of Proposition 4.9 and Theorem 4.10 are validated. For matrices $P,Q\in\mathbb{S}^D$ we write $P\preceq Q$ or $Q\succeq P$ if $Q-P\in\mathbb{S}_+^D$.

Example 4.14 (Proportional coefficients)**.** Let b,a,γ in Section 3.1 satisfy $b^{(i)},a^{(i,j)},\gamma^{(i,j)}\in C_b^\infty(\mathbb{R}^D)$ for all $i,j=1,\dots,D$. Assume that there are constants $K>0$ and $\varepsilon>0$ such that, for all $y\in\mathbb{R}^D$,

$$\begin{cases} A(y)\succeq \varepsilon I_D,\\ (b^{\mathsf{T}}\Sigma^{-1}b)(y)=K.\end{cases}\tag{4.29}$$

For example, if there exist $U\colon\mathbb{R}^D\to\mathbb{S}_{++}^D$ with $U^{(i,j)}\in C_b^\infty(\mathbb{R}^D)$, $i,j=1,\dots,D$, and a constant $\delta>0$ such that $U(y)\succeq\delta I_D$ for all $y\in\mathbb{R}^D$ and that, for some constant $\tilde{b}\in\mathbb{R}^D$, $\tilde{a},\tilde{\gamma}\in\mathbb{R}^{D\times D}$ with $\tilde{b}\neq 0$ and $\det(\tilde{a})\neq 0$,

$$b(y)=U(y)\tilde{b},\quad a(y)=U(y)\tilde{a},\quad \gamma(y)=U(y)\tilde{\gamma},$$

then condition (4.29) holds true with $K=\tilde{b}^{\mathsf{T}}(\tilde{a}\tilde{a}^{\mathsf{T}}+\tilde{\gamma}\int_E ee^{\mathsf{T}}\nu(\mathrm{d}e)\tilde{\gamma}^{\mathsf{T}})^{-1}\tilde{b}$ and $\varepsilon=\delta^2\tilde{\varepsilon}$, where $\tilde{\varepsilon}>0$ is sufficiently small such that $\tilde{a}\tilde{a}^{\mathsf{T}}\succeq\tilde{\varepsilon}I_D$.

Now, under (4.29), Assumption 3.1 is obviously satisfied. Moreover, since $\Sigma(y)\succeq A(y)$, it follows from the ellipticity condition $A(y)\succeq\varepsilon I_D$ that $\Sigma(y)\succeq\varepsilon I_D$. Hence, Lemma B.1 gives

$$\Sigma^{-1}(y)\preceq\frac{1}{\varepsilon}I_D,\quad\forall y\in\mathbb{R}^D.$$

Consequently, one has $\sup_{y\in\mathbb{R}^D}\|\Sigma^{-1}(y)\|<\infty$.

We first find solution α of the PIDEs (4.13) which does not depend on y. For $\alpha(t,\cdot)=\alpha(t)$, we get

$$\mathcal{S}_\alpha(t,y)=\alpha(t)\Sigma(y),\quad \mathcal{M}_\alpha(t,y)=\alpha(t)b(y)$$

so that $(\mathcal{M}_\alpha^{\mathsf{T}}\mathcal{S}_\alpha^{-1}\mathcal{M}_\alpha)(t,y)=\alpha(t)K$. Then the PIDE for α in (4.13) boils down to the following ordinary differential equation (ODE)

$$\begin{cases}\alpha'(t)-\alpha(t)K=0,\quad t\in[0,T),\\ \alpha(T)=1,\end{cases}$$

whose solution is given by

$$\alpha(t)=\mathrm{e}^{-(T-t)K},\quad t\in[0,T].$$

It is easy to check that the assumptions of Proposition 4.9 and Theorem 4.10 are satisfied for α. Next, plugging this α into the PIDE for β in (4.13) we obtain

$$\begin{cases} \partial_t \beta(t,y) + \mathcal{L}_Y \beta(t,y) - \dfrac{\lambda}{2} \log\left(\dfrac{(\lambda\pi)^D}{\det(\alpha(t)\Sigma(y))} \right) = 0, \quad t \in [0,T), \\ \beta(T,\cdot) = 0. \end{cases} \tag{4.30}$$

We apply [26, Theorem 1] to conclude that the PIDE (4.30) has a unique classical solution $\beta \in C^{1,2}([0,T]\times\mathbb{R}^D)$. Moreover, β and its partial derivatives $\partial_t\beta, \nabla_y\beta, \nabla^2_{yy}\beta$ are uniformly bounded on $[0,T]\times\mathbb{R}^D$. Then β also satisfies the assumptions of Proposition 4.9 and Theorem 4.10. The Feynman–Kac representation for β (see, e.g., [3, Theorems 3.4 and 3.5]) is

$$\beta(t,y) = \mathbb{E}\bigg[\int_t^T f(s, Y_s^{t,y}) \mathrm{d}s \bigg], \quad (t,y) \in [0,T]\times\mathbb{R}^d,$$

where $f(t,y) := -\frac{\lambda}{2}\log(\frac{(\lambda\pi)^D}{\det(\alpha(t)\Sigma(y))})$ and $Y^{t,y}$ is given in (4.3). The value function is

$$V^*(t,x,y|\hat{w}) = \mathrm{e}^{-(T-t)K}(x-\hat{w})^2 + \beta(t,y).$$

The associated exploratory SDE for $X^* = (X^*_r)_{r\in[0,T]}$ is given by $X_0 = x_0$ and

$$\begin{aligned} \mathrm{d}X^*_r = &-(X^*_{r-} - \hat{w})\bigg(K\mathrm{d}r + (b^\mathsf{T}\Sigma^{-1}a)(Y_{r-})\mathrm{d}W_r + (b^\mathsf{T}\Sigma^{-1}\gamma)(Y_{r-}) \int_{E\times\mathbb{R}^D} e\widetilde{N}^\psi_L(\mathrm{d}r,\mathrm{d}e,\mathrm{d}u) \bigg) \\ &+ \sqrt{\frac{\lambda}{2}}\mathrm{e}^{\frac{1}{2}(T-r)K}\bigg(\mathbf{tr}[(\Sigma^{-\frac{1}{2}}a)(Y_{r-})\mathrm{d}\mathcal{W}_r^\mathsf{T}] + \int_{E\times\mathbb{R}^D} \frac{u^\mathsf{T}}{\psi(e)}(\Sigma^{-\frac{1}{2}}\gamma)(Y_{r-}) e\, \widetilde{N}^\psi_L(\mathrm{d}r,\mathrm{d}e,\mathrm{d}u) \bigg), \end{aligned} \tag{4.31}$$

whose explicit expression can be derived either from [5] or from (4.27) provided that $\Delta Z \neq 1$ on $[0,T]$.

Regarding the Lagrange multiplier $\hat{w}$, due to the condition $(b^\mathsf{T}\Sigma^{-1}b)(y) = K$ in (4.29), we can simply calculate its value by taking the expectation of X^*_r with noting that the martingale terms in the expression (4.31) of X^*_r are square integrable null at 0, and then using Fubini's theorem to get

$$\mathbb{E}[X^*_r] = x_0 - K\int_0^r (\mathbb{E}[X^*_s] - \hat{w})\mathrm{d}s,$$

which then gives

$$\mathbb{E}[X^*_r] = \hat{w} + (x_0 - \hat{w})\mathrm{e}^{-Kr}, \quad r \in [0,T].$$

By the constraint $\mathbb{E}[X^*_T] = \hat{z}$, we arrive at

$$\hat{w} = \frac{\hat{z}\mathrm{e}^{KT} - x_0}{\mathrm{e}^{KT} - 1}.$$

Example 4.15 (Constant coefficients)**.** Let b, a, γ be constants on $\mathbb{R}^D$ with $b \neq 0$ and $\Sigma \in \mathbb{S}^D_{++}$, where a might be degenerate. In this situation we can find solutions α, β of the PIDEs (4.13) which do not depend on y. Namely, by letting $\alpha(t,\cdot) = \alpha(t)$, $\beta(t,\cdot) = \beta(t)$ and plugging them into (4.13) we obtain a system of ODEs for α, β which possesses the following solutions on $[0,T]$,

$$\begin{cases} \alpha(t) = \mathrm{e}^{-(T-t)K}, \\ \beta(t) = -(T-t)^2\dfrac{\lambda D}{4}K - (T-t)\dfrac{\lambda}{2}\log\left(\dfrac{(\lambda\pi)^D}{\det(\Sigma)}\right), \end{cases}$$

where $K := b^\intercal \Sigma^{-1} b > 0$. It is also easy to check that the assumptions of Proposition 4.9 and Theorem 4.10 are fulfilled for α, β. Then the value function is explicitly given by

$$V^*(t,x,y|\hat{w}) = V^*(t,x|\hat{w}) := \mathrm{e}^{-(T-t)K}(x-\hat{w})^2 - (T-t)^2\frac{\lambda D}{4}K - (T-t)\frac{\lambda}{2}\log\left(\frac{(\lambda\pi)^D}{\det(\Sigma)}\right).$$

The SDE for the optimal wealth X^* and the Lagrange multiplier $\hat{w}$ are respectively the same as those in Example 4.14 where one notices here that coefficients b, a, γ, Σ are constant[4].

In the following we continue to specialize Example 4.15 to the case of no jumps.

Example 4.16 (Constant coefficients, $\nu \equiv 0$ and $D = 1$)**.** This is the setting considered by Wang and Zhou [34]. For $a > 0$, $b \neq 0$, letting $\sigma := a$ and $\rho := \frac{b}{a}$ we get the value function

$$V^*(t,x|\hat{w}) = \mathrm{e}^{-\rho^2(T-t)}(x-\hat{w})^2 - \frac{\lambda}{2}\left(\frac{\rho^2}{2}(T-t)^2 + (T-t)\log\left(\frac{\lambda\pi}{\sigma^2}\right)\right)$$

which coincides with that in [34, Theorem 3.1]. The associated SDE for the optimal wealth X^* in our setting is

$$\mathrm{d}X_s^* = -\rho^2(X_s^* - \hat{w})\mathrm{d}s - \rho(X_s^* - \hat{w})\mathrm{d}W_s + \sqrt{\frac{\lambda}{2}}\mathrm{e}^{\frac{\rho^2}{2}(T-t)}\mathrm{d}\mathcal{W}_s, \quad X_0^* = x_0, \tag{4.32}$$

whose explicit representation is given, according to Proposition 4.12, by

$$X_r^* = \hat{w} + \left[x_0 - \hat{w} + \sqrt{\frac{\lambda}{2}}\int_0^r \mathrm{e}^{\rho W_s + \frac{3}{2}\rho^2 s}\mathrm{e}^{\frac{1}{2}\rho^2(T-s)}\mathrm{d}\mathcal{W}_s\right]\mathrm{e}^{-\rho W_r - \frac{3}{2}\rho^2 r}, \quad r \in [0,T].$$

We emphasize that the optimal exploratory SDE (4.32) is different from that in [34, Eq. (27)] which is formulated in our notation as

$$\mathrm{d}\widetilde{X}_s^* = -\rho^2(\widetilde{X}_s^* - \hat{w})\mathrm{d}s + \sqrt{\rho^2(\widetilde{X}_s^* - \hat{w})^2 + \frac{\lambda}{2}\mathrm{e}^{\rho^2(T-s)}}\mathrm{d}W_s, \quad \widetilde{X}_0^* = x_0. \tag{4.33}$$

However, solutions X^* of (4.32) and $\widetilde{X}^*$ of (4.33) have the same (finite-dimensional) distribution because of the uniqueness in law of (4.33).

4.5. **Relation to the sample state process.** The continuous-time RL algorithms designed in [20, 21] rely on the sample state process for the actual learning task, which is the solution of an SDE that models the state dynamics evaluated along a randomized control. In this subsection, we explain the relation between our exploratory dynamics and the sample state process. To avoid technicalities and to highlight the key ideas, we focus on the situation in Example 4.16.

The construction of the sample state process in [20, 21] starts with a feedback control $\pi(\cdot|t,x)$ with values in the space of probability density functions. If the portfolio value is in state $X_t = x$ at time t, the portfolio position H_t is randomly drawn from the probability density $\pi(\cdot|t,x)$. The actual drawing of the portfolio position is performed based on a family $(Z_t)_{t\in[0,T]}$ of independent random variables in [21], which are uniformly distributed on $[0,1]$. This family is supposed to be independent of the stochastic processes driving the stock price, hence, of the Brownian motion W in the context of Example 4.16. Denoting by $h_\pi(t,x;\cdot)$ the quantile function of the distribution with density $\pi(\cdot|t,x)$, the random drawing of the portfolio position can be made explicit by letting $H_t = h_\pi(t,X_t,Z_t)$, which formally leads to the SDE

$$\mathrm{d}X_t^{h_\pi} = h_\pi(t,X_t^{h_\pi},Z_t)(b\mathrm{d}t + \sigma\mathrm{d}W_t), \quad X_0^{h_\pi} = x. \tag{4.34}$$

[4]If $\nu(\{e \in E : b^\intercal\Sigma^{-1}\gamma e = 1\}) = 0$, then applying the same argument as in Remark 4.13(1) yields $\Delta Z \neq 1$ on $[0,T]$, and hence, (4.27) is usable.

In the terminology of [21], the solution of this SDE is the *sample state process* corresponding to the *action process* $h_\pi(t, X_t, Z_t)$, which is sampled from the given density π.[5]

Given the optimality of Gaussian randomization, which in the context of Example 4.16 has first been derived in [34], we now specialize to the case that $\pi(\cdot|t,x)$ is a Gaussian density with mean $\mu(t,x)$ and standard deviation $\vartheta(t,x) := \theta(t,x)^{\frac{1}{2}}$. Then, $h_\pi(t, X_t, Z_t) = \mu(t, X_t) + \vartheta(t, X_t)\xi_t$, where $(\xi_t)_{t\in[0,T]}$ is an independent family of standard Gaussians constructed from $(Z_t)_{t\in[0,T]}$ via $\xi_t = \Phi^{-1}(Z_t)$ (Φ denoting the cumulative distribution function of the standard normal distribution). Thus, (4.34) becomes (suppressing the superscript h_π)

$$\mathrm{d}X_t = b\mu(t, X_t)\mathrm{d}t + b\vartheta(t, X_t)\xi_t\mathrm{d}t + \sigma\mu(t, X_t)\mathrm{d}W_t + \sigma\vartheta(t, X_t)\xi_t\mathrm{d}W_t, \quad X_0 = x. \tag{4.35}$$

If we write $\mathbb{G} = (\mathcal{G}_t)_{t\in[0,T]}$ for the filtration generated by W and ξ, then ξ becomes an adapted process, but it is well-known that non-constant families of independent, identically distributed random variables indexed by continuous time cannot be measurable with respect to the standard product σ-field, see, e.g., [31, Proposition 2.1]. Hence, ξ fails to be progressively measurable in the usual sense of stochastic calculus, and so the SDE (4.35) cannot be studied in the classical SDE framework. To deal with this problem, the authors in [21] refer to the framework of rich Fubini extensions developed in [31, 32]. Roughly speaking, one can construct an extension $\bar{\boldsymbol{\lambda}}$ of the Lebesgue measure on $[0,T]$ beyond the σ-field of Lebesgue-measurable sets and a suitable probability space such that the process ξ becomes measurable with respect to some appropriate Fubini extension of the classical product measure space, see [29, Theorem 2] for a precise statement. Here the notion of a *Fubini extension* refers to the property that a suitable reformulation of Fubini's theorem on iterated integration is still valid, see [31]. Then, the Lebesgue integrals in (4.35) can be replaced by integrals with respect to the extension $\bar{\boldsymbol{\lambda}}$ of the Lebesgue measure (but we will write $\mathrm{d}t$ in place of $\bar{\boldsymbol{\lambda}}(\mathrm{d}t)$ below to simplify the notation). However, with this construction, it is still not clear to us, how to extend the Itô integral to integrands which only satisfy this weaker measurability property ensured by the rich Fubini construction. In the following informal discussion, we make the conjecture that Itô's integral can be properly extended such that the standard results of stochastic calculus are still in force.

Under this conjecture, we may consider

$$A^\xi_t = \int_0^t \xi_s\mathrm{d}s, \quad W^\xi_t = \int_0^t \xi_s\mathrm{d}W_s.$$

Then, (4.35) can be rewritten as

$$\mathrm{d}X_t = b\mu(t, X_t)\mathrm{d}t + b\vartheta(t, X_t)\mathrm{d}A^\xi_t + \sigma\mu(t, X_t)\mathrm{d}W_t + \sigma\vartheta(t, X_t)\mathrm{d}W^\xi_t, \quad X_0 = x.$$

By the above conjecture, W^ξ is a continuous martingale with

$$\langle W^\xi\rangle_t = \int_0^t \xi_s^2\mathrm{d}s, \quad \langle W^\xi, W\rangle_t = \int_0^t \xi_s\mathrm{d}s.$$

Sun's exact law of large numbers [31, Theorem 2.6] developed in the framework of rich Fubini extensions now implies

$$\int_0^t \xi_s^q\mathrm{d}s = \int_0^t \mathbb{E}[\xi_s^q]\mathrm{d}s = \begin{cases} 0, & q = 1, \\ t, & q = 2. \end{cases}$$

[5]The authors in [21] do not give an explicit construction of the action process, whereas we use the construction based on the quantile function in this subsection. It is, however, clear from the presentation in [21] that iid uniform random variables $(Z_t)_{t\in[0,T]}$ independent of W are applied for the control randomization mechanism in [21].

Hence $A^\xi \equiv 0$ and, by Lévy's characterization, W^ξ is a Brownian motion independent of W. Thus, the SDE (4.35) for the sample state process with Gaussian randomization becomes

$$\mathrm{d}X_t = b\mu(t, X_t)\mathrm{d}t + \sigma\mu(t, X_t)\mathrm{d}W_t + \sigma\vartheta(t, X_t)\mathrm{d}W_t^\xi, \quad X_0 = x. \tag{4.36}$$

This is exactly our form of the exploratory SDE (3.7) (with a different notation for the additional independent Brownian motion), when specialized to the setting of Example 4.16 and applied to feedback controls with Gaussian randomization.

Let us now look at the special case, when the control randomization is performed according to the standard Gaussian distribution independent of time and state, i.e., $\mu(t, x) = 0$ and $\vartheta(t, x) = 1$. Then, the corresponding portfolio wealth process $X_t = x + \sigma W_t^\xi$ in (4.36) is independent of W, and, hence, independent of the stock price dynamics. This appears to be counter-intuitive and illustrates that the SDE for the sample state process may not be able properly describe the portfolio wealth along a randomized portfolio (with continuous re-sampling in the randomization mechanism). In order justify the use of this SDE, we discretize the portfolio process $H_t = \xi_t$ on a time grid $0 = t_0 < t_1 < \cdots < t_n = T$ via $H_t^n = \xi_{t_j}$, for $t \in (t_j, t_{j+1}]$. Then, H^n is $\mathbb{G}$-predictable (since it is adapted and left-continuous), and, hence, its wealth process with initial endowment x

$$X_t^n = x + \int_0^t H_s^n(b\mathrm{d}s + \sigma\mathrm{d}W_s) = x + b\sum_{j=0}^{n-1} \xi_{t_j}(t_{j+1} \wedge t - t_j \wedge t) + \sigma\sum_{j=0}^{n-1} \xi_{t_j}(W_{t_{j+1}\wedge t} - W_{t_j\wedge t})$$

is well-defined in the framework of the classical stochastic integration theory. It is not difficult to check that the first sum converges to zero by the law of large numbers (cp. Section 3.2.2) and the second sum weakly converges to a Brownian motion independent of W by Donsker's invariance principle. Hence, the "wealth process" $X_t = x + \sigma W_t^\xi$ for the non-predictable portfolio position process $H_t = \xi_t$ suggested by the sample state process in the continuous-time RL literature can be properly interpreted as the weak limit of the wealth processes of the approximating sequence of predictable randomized portfolio positions H^n. The limit result in this illustrative example is, of course, the simplest special case of our general result, Theorem 3.5, which motivates our formulation of the exploratory SDE.

Summarizing, the above discussion suggests that our formulation (3.7) of the exploratory SDE is one way to give a mathematically rigorous meaning to the SDE which models the sample state process in the recent RL literature. Moreover, Theorem 3.5 provides a justification for the use of this SDE formulation as a limit of a natural control randomization mechanism in discrete time. While the results in this paper are presented for the mean-variance portfolio selection problem, it is obvious, how to transfer the derivation of our exploratory SDE based on Theorem 3.5 to more general problems with controlled diffusion and jumps, provided the control enters the diffusion part linearly. The case of general dependence of the diffusion coefficient on the control requires more advanced tools from the theory of random measures and is discussed in our follow-up work [4].

Remark 4.17. In our general setting with D stocks, the Gaussian randomization leads to an action process of the form

$$\mu(t, X_t, Y_t) + \vartheta(t, X_t, Y_t)\xi_t$$

where the mean $\mu(t, x, y)$ takes values in $\mathbb{R}^D$, $\vartheta(t, x, y)$ is the positive definite root of the positive definite $D \times D$ covariance matrix $\theta(t, x, y)$ and each ξ_t is a vector of D independent standard

Gaussians. Following the same argument as above, we will consider the "processes"

$$W_t^{\xi,(d,d')} = \int_0^t \xi_s^{(d)} \mathrm{d}W_s^{(d')},$$

which additionally drive the SDE for the sample state process. Then, by the Lévy characterization as above, $(W^{d'}, W^{\xi,(d,d')}; d = 1, \dots D,\ d' = 1, \dots D)$ is a $(D^2 + D)$-dimensional Brownian motion. Thus, making the sample state process rigorous by the same reasoning as above, the diffusion part is driven by a $(D^2 + D)$-dimensional Brownian motion as in our formulation (3.7) of the exploratory SDE.

5. Weak convergence of discrete-time integrators

This section provides the proof of Theorem 3.5. Throughout this part, let c_D denote a positive constant depending only on D, and its value might vary in each appearance. The time-change σ^n is extended constantly over $t \in [T, \infty)$. To cover necessary test functions for the proof of Theorem 3.5, we use the following function space.

Definition 5.1. For $\mathbf{D} = D^2 + 3D$, we let $g \in C_*^2(\mathbb{R}^{\mathbf{D}})$ if the following conditions hold:

(G1): $g \in C^2(\mathbb{R}^{\mathbf{D}})$ with $g(0) = 0$ and $\|\nabla^2 g\|_\infty < \infty$;

(G2): for $1 \le d \vee d' \le D^2 + D$, the function $\partial^2_{d,d'} g$ takes value 0 in a neighborhood of 0;

(G3): $c_{(\mathrm{G3})} := \max_{1 \le d \le D^2+D} \|\partial_d g(0_{D^2+D}, \cdot)\|_\infty < \infty$, where 0_{D^2+D} is the vector 0 in $\mathbb{R}^{D^2+D}$;

(G4): $c_{(\mathrm{G4})} := \max_{D^2+D+1 \le d \le \mathbf{D}} \|\partial_d g\|_\infty < \infty$ and $\partial_d g(0) = 0$ for any $D^2 + D + 1 \le d \le \mathbf{D}$.

Proposition 5.2. *For any $g \in C_*^2(\mathbb{R}^{\mathbf{D}})$, one has when $n \to \infty$ that*

$$\sum_{i=1}^n \left| \mathbb{E}[g(\Delta_{n,i}\mathcal{Z}^n) | \mathcal{F}_{n,i-1}] - (t_i^n - t_{i-1}^n) \int_{\mathbb{R}^{2D}} g(0, e, u) \nu_L^\psi(\mathrm{d}e, \mathrm{d}u) \right| \xrightarrow{\mathbf{L}_1(\mathbb{P})} 0, \tag{5.1}$$

where $\mathcal{Z}^n$ is given in Section 3.2.5.

Proof. With a slight abuse of notation, in the sequel we use symbols η, ξ without any sub-indices to denote *deterministic* vectors in $\mathbb{R}^D$, whereas $\eta_{n,i}^H$ and $\xi_{n,i}$ are random vectors introduced in Section 3.2.1. Recall that

$$\Delta_{n,i}\mathcal{Z}^n = \mathrm{vec}(\Delta_{n,i}W^n, \Delta_{n,i}M^n, \Delta_{n,i}L^{n,\psi}) = \mathrm{vec}(\Delta_{n,i}W,\ \eta_{n,i}^H \otimes \Delta_{n,i}W,\ \Delta_{n,i}J,\ \psi(\Delta_{n,i}J)\xi_{n,i}).$$

Step 1. Since $g(0) = 0$ by (G1) and $\partial_d g(0) = 0$ for $D^2 + D + 1 \le d \le \mathbf{D}$ by (G4), an argument using Taylor expansion shows

$$|g(0, e, u)| \le c_D \|\nabla^2 g\|_\infty (\|e\|^2 + \|u\|^2), \quad e, u \in \mathbb{R}^D. \tag{5.2}$$

Since ν_L^ψ is a square integrable Lévy measure, it ensures that $\int_{\mathbb{R}^{2D}} |g(0, e, u)| \nu_L^\psi(\mathrm{d}e, \mathrm{d}u) < \infty$. Moreover, for any n, i, since

$$\begin{aligned}\mathbb{E}[\|\Delta_{n,i}\mathcal{Z}^n\|^2] &= \mathbb{E}[\|\Delta_{n,i}W\|^2 + \|\eta_{n,i}^H \otimes \Delta_{n,i}W\|^2 + \|\Delta_{n,i}J\|^2 + \psi(\Delta_{n,i}J)^2\|\xi_{n,i}\|^2] \\ &\le (t_i^n - t_{i-1}^n)\left(D + D^2 + \int_E \|e\|^2 \nu(\mathrm{d}e) + D\|\nabla\psi\|_\infty^2 \int_E \|e\|^2 \nu(\mathrm{d}e)\right) < \infty,\end{aligned}$$

together with the fact that g has at most quadratic growth at infinity as $\|\nabla^2 g\|_\infty < \infty$ by (G1), it implies that $\mathbb{E}[|g(\Delta_{n,i}\mathcal{Z}^n)|] < \infty$.

Step 2. To shorten the notation, for each $\eta, \xi \in \mathbb{R}^D$, we define $g_{\eta,\xi} \colon \mathbb{R}^{2D} \to \mathbb{R}$ by

$$g_{\eta,\xi}(w, j) := g(w, \eta \otimes w, j, \psi(j)\xi), \quad w, j \in \mathbb{R}^D.$$

Then, $g_{\eta,\xi} \in C^2(\mathbb{R}^{2D})$. Furthermore, for any $d, d' = 1, \dots, D$, the partial derivatives of $g_{\eta,\xi}$ are given, with the convention $\eta^{(0)} := 1$ and $z := (w, \eta \otimes w, j, \psi(j)\xi) \in \mathbb{R}^{\mathbf{D}}$, by

$$\partial_d g_{\eta,\xi}(w,j) = \sum_{k=0}^{D} \eta^{(k)} \partial_{d+kD} g(z), \quad \partial^2_{d,d'} g_{\eta,\xi}(w,j) = \sum_{k,l=0}^{D} \eta^{(k)}\eta^{(l)} \partial^2_{d+kD,d'+lD} g(z), \tag{5.3}$$

$$\partial_{D+d} g_{\eta,\xi}(w,j) = \partial_{D^2+D+d} g(z) + \partial_d \psi(j) \sum_{k=1}^{D} \xi^{(k)} \partial_{D^2+2D+k} g(z), \tag{5.4}$$

$$\begin{aligned}\partial^2_{D+d',D+d} g_{\eta,\xi}(w,j) &= \partial^2_{D^2+D+d',D^2+D+d} g(z) + \partial^2_{d',d}\psi(j) \sum_{k=1}^{D} \xi^{(k)} \partial_{D^2+2D+k} g(z) \\ &\quad + \partial_d \psi(j) \sum_{k=1}^{D} \xi^{(k)} \Big[\partial^2_{D^2+D+d',D^2+2D+k} g(z) + \partial_{d'}\psi(j) \sum_{l=1}^{D} \xi^{(l)} \partial^2_{D^2+2D+l,D^2+2D+k} g(z) \Big].\end{aligned}$$

Hence, there exists a constant $c_{(5.5)} := c(D, \|\nabla\psi\|_\infty, \|\nabla^2\psi\|_\infty, \|\nabla^2 g\|_\infty, c_{(\mathrm{G4})}) > 0$ such that

$$\max_{1 \le d,d' \le D} \|\partial^2_{D+d',D+d} g_{\eta,\xi}\|_\infty \le c_{(5.5)}(1 + \|\xi\|^2). \tag{5.5}$$

We also define the function R_1^g, which represents the remainder term in a Taylor expansion of $g_{\eta,\xi}$, by setting for $w, j, \eta, \xi, e \in \mathbb{R}^D$ that

$$R_1^g(w,j;\eta,\xi;e) := g_{\eta,\xi}(w, j+e) - g_{\eta,\xi}(w,j) - \sum_{d=1}^{D} e^{(d)} \partial_{D+d} g_{\eta,\xi}(w,j).$$

Due to condition (G3), Taylor expansion implies for any $a \in \mathbb{R}^{D^2+D}$, $a' \in \mathbb{R}^{2D}$ that

$$|g(a,a') - g(0,a')| \le c_D(c_{(\mathrm{G3})}\|a\| + \|\nabla^2 g\|_\infty \|a\|^2) \le c_{(5.6)}(\|a\| + \|a\|^2) \tag{5.6}$$

for some constant $c_{(5.6)} := c_{(5.6)}(D, \|\nabla^2 g\|_\infty, c_{(\mathrm{G3})}) > 0$. Hence,

$$\begin{aligned}&|R_1^g(w,j;\eta,\xi;e) - R_1^g(w,j;0,\xi;e)| \\ &\le |g(w,\eta\otimes w, j+e, \psi(j+e)\xi) - g(0, j+e, \psi(j+e)\xi)| \\ &\quad + |g(w,0,j+e,\psi(j+e)\xi) - g(0,j+e,\psi(j+e)\xi)| \\ &\quad + |g(w,\eta\otimes w, j, \psi(j)\xi) - g(0,j,\psi(j)\xi)| + |g(w,0,j,\psi(j)\xi) - g(0,j,\psi(j)\xi)| \\ &\quad + \sum_{d=1}^{D} |e^{(d)}| \Big[\Big| \partial_{D^2+D+d} g(w,\eta\otimes w, j, \psi(j)\xi) - \partial_{D^2+D+d} g(w,0,j,\psi(j)\xi) \Big| \\ &\qquad + |\partial_d \psi(j)| \sum_{k=1}^{D} |\xi^{(k)}| \Big| \partial_{D^2+2D+k} g(w,\eta\otimes w, j, \psi(j)\xi) - \partial_{D^2+2D+k} g(w,0,j,\psi(j)\xi) \Big| \Big] \\ &\le 2c_{(5.6)}(\|(w,\eta\otimes w)\| + \|(w,\eta\otimes w)\|^2 + \|w\| + \|w\|^2) + c_{(5.7)}\|e\|(1+\|\xi\|)\|\eta\otimes w\| \qquad (5.7) \\ &\le 4c_{(5.6)}(\|w\| + \|w\|^2 + \|\eta\otimes w\| + \|\eta\otimes w\|^2) + c_{(5.7)}\|e\|(1+\|\xi\|)\|\eta\otimes w\|,\end{aligned}$$

where $c_{(5.7)} := c(D, \|\nabla\psi\|_\infty, \|\nabla^2 g\|_\infty) > 0$. Moreover, the Taylor remainder R_1^g is estimated by

$$\sup_{(w,j)\in\mathbb{R}^{2D}} |R_1^g(w,j;\eta,\xi;e)| \le c_D \max_{1\le d,d'\le D} \|\partial^2_{D+d,D+d'} g_{\eta,\xi}\|_\infty \|e\|^2 \le c_{(5.8)}(1+\|\xi\|^2)\|e\|^2, \tag{5.8}$$

where $c_{(5.8)} := c_D c_{(5.5)}$.

Step 3. For $n \ge 1$ and $1 \le i \le n$, since $\eta^H_{n,i}$ is $\mathcal{F}_{n,i-1} \vee \sigma\{\xi_{n,i}\}$-measurable and $(\Delta_{n,i}W, \Delta_{n,i}J)$ is independent of $\mathcal{F}_{n,i-1} \vee \sigma\{\xi_{n,i}\}$, we get, a.s.,

$$\mathbb{E}[g(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] = \mathbb{E}\Big[\mathbb{E}\Big[g(\Delta_{n,i}W, \eta^H_{n,i}\otimes\Delta_{n,i}W, \Delta_{n,i}J, \psi(\Delta_{n,i}J)\xi_{n,i}) \Big| \mathcal{F}_{n,i-1} \vee \sigma\{\xi_{n,i}\} \Big] \Big| \mathcal{F}_{n,i-1} \Big]$$

$$= \mathbb{E}[G_{n,i}(\eta^H_{n,i}, \xi_{n,i})\,|\mathcal{F}_{n,i-1}],$$

where $G_{n,i}$ is a non-random and measurable function defined as

$$G_{n,i}(\eta,\xi) := \mathbb{E}[g(\Delta_{n,i}W, \eta\otimes\Delta_{n,i}W, \Delta_{n,i}J, \psi(\Delta_{n,i}J)\xi)], \quad \eta,\xi\in\mathbb{R}^D.$$

Given $\eta,\xi\in\mathbb{R}^D$, applying Itô's formula for $g_{\eta,\xi}\in C^2(\mathbb{R}^{2D})$ yields, a.s.,

$$\begin{aligned}
&g(\Delta_{n,i}W, \eta\otimes\Delta_{n,i}W, \Delta_{n,i}J, \psi(\Delta_{n,i}J)\xi) = g_{\eta,\xi}(W_{t_i^n}-W_{t_{i-1}^n}, J_{t_i^n}-J_{t_{i-1}^n})\\
&= \sum_{d=1}^D \int_{t_{i-1}^n}^{t_i^n} \partial_d g_{\eta,\xi}(W_s - W_{t_{i-1}^n}, J_{s-}-J_{t_{i-1}^n})\mathrm{d}W_s^{(d)}\\
&\quad + \frac{1}{2}\sum_{d,d'=1}^D \int_{t_{i-1}^n}^{t_i^n} \partial^2_{d,d'} g_{\eta,\xi}(W_s - W_{t_{i-1}^n}, J_{s-}-J_{t_{i-1}^n})\mathrm{d}s\\
&\quad + \int_{t_{i-1}^n}^{t_i^n}\int_E \Big(g_{\eta,\xi}(W_s - W_{t_{i-1}^n}, J_{s-}-J_{t_{i-1}^n}+e) - g_{\eta,\xi}(W_s - W_{t_{i-1}^n}, J_{s-}-J_{t_{i-1}^n})\Big)\widetilde{N}(\mathrm{d}e,\mathrm{d}s)\\
&\quad + \int_{t_{i-1}^n}^{t_i^n}\int_E R_1^g(W_s - W_{t_{i-1}^n}, J_{s-}-J_{t_{i-1}^n};\eta,\xi;e)\nu(\mathrm{d}e)\mathrm{d}s. \qquad (5.9)
\end{aligned}$$

For $d=1,\dots,D$, we derive from (5.3) that $(w,j)\mapsto\partial_d g_{\eta,\xi}(w,j)$ has at most linear growth at infinity which hence implies that the stochastic integrals with respect to the Brownian motions are square integrable martingales. Moreover, for $w,j,j'\in\mathbb{R}^D$, due to (5.4) and (G4) one has

$$|g_{\eta,\xi}(w,j)-g_{\eta,\xi}(w,j')| \le c_D \max_{1\le d\le D}\|\partial_d g_{\eta,\xi}(w,\cdot)\|_\infty \|j-j'\| \le c_D c_{(\mathrm{G4})}(1+\|\nabla\psi\|_\infty\|\xi\|)\|j-j'\|.$$

Then, due to the assumption $\int_E \|e\|^2\nu(\mathrm{d}e)<\infty$, the stochastic integral with respect to the compensated Poisson random measure $\widetilde{N}$ in (5.9) is also a square integrable martingale which then vanishes after taking the expectation. Hence,

$$G_{n,i}(\eta,\xi) = G^W_{n,i}(\eta,\xi) + G^J_{n,i}(\eta,\xi),$$

where the integrability condition is satisfied so that Fubini's theorem enables us to define

$$\begin{aligned}
G^W_{n,i}(\eta,\xi) &:= \frac{1}{2}\sum_{d,d'=1}^D \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}\Big[\partial^2_{d,d'} g_{\eta,\xi}(W_s - W_{t_{i-1}^n}, J_{s-}-J_{t_{i-1}^n})\Big]\mathrm{d}s,\\
G^J_{n,i}(\eta,\xi) &:= \int_{t_{i-1}^n}^{t_i^n}\int_E \mathbb{E}\Big[R_1^g(W_s - W_{t_{i-1}^n}, J_{s-}-J_{t_{i-1}^n};\eta,\xi;e)\Big]\nu(\mathrm{d}e)\mathrm{d}s.
\end{aligned}$$

To derive (5.1) it suffices to prove that the following three convergences hold:

$$G^n_{(5.10)} := \sum_{i=1}^n \mathbb{E}[|G^W_{n,i}(\eta^H_{n,i},\xi_{n,i})|] \to 0, \qquad (5.10)$$

$$G^n_{(5.11)} := \sum_{i=1}^n \mathbb{E}[|G^J_{n,i}(\eta^H_{n,i},\xi_{n,i}) - G^J_{n,i}(0,\xi_{n,i})|] \to 0, \qquad (5.11)$$

$$G^n_{(5.12)} := \sum_{i=1}^n \mathbb{E}\left[\left|G^J_{n,i}(0,\xi_{n,i}) - (t_i^n - t_{i-1}^n)\int_{E\times\mathbb{R}^D} g(0,e,\psi(e)u)\nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u\right|\right] \to 0. \qquad (5.12)$$

Step 4. We show $G^n_{(5.10)}\to 0$. For $1\le d,d'\le D$, by (5.3) one has

$$G^W_{n,i}(\eta,\xi) = \frac{1}{2}\sum_{d,d'=1}^D\sum_{k,l=0}^D \eta^{(k)}\eta^{(l)}\int_{t_{i-1}^n}^{t_i^n}\mathbb{E}\Big[\partial^2_{d+kD,d'+lD} g(W_s - W_{t_{i-1}^n}, \eta\otimes(W_s - W_{t_{i-1}^n}),$$

$$J_{s-} - J_{t^n_{i-1}}, \psi(J_{s-} - J_{t^n_{i-1}})\xi)\Big]\mathrm{d}s.$$

Let $(\bar{W}, \bar{J})$ be an independent copy of (W, J) with the corresponding expectation $\bar{\mathbb{E}}$. Applying Fubini's theorem we get

$$\begin{aligned} G^n_{(5.10)} &\le \frac{1}{2}\sum_{d,d'=1}^{D}\sum_{k,l=0}^{D}\sum_{i=1}^{n}\mathbb{E}\Big[|\eta^{H,(k)}_{n,i}\eta^{H,(l)}_{n,i}|\int_{t^n_{i-1}}^{t^n_i}\bar{\mathbb{E}}\Big[\Big|\partial^2_{d+kD,d'+lD}g(\bar{W}_s - \bar{W}_{t^n_{i-1}}, \eta^H_{n,i}\otimes(\bar{W}_s - \bar{W}_{t^n_{i-1}}), \\ &\qquad \bar{J}_{s-} - \bar{J}_{t^n_{i-1}}, \psi(\bar{J}_{s-} - \bar{J}_{t^n_{i-1}})\xi_{n,i})\Big|\Big]\mathrm{d}s\Big] \\ &= \frac{1}{2}\sum_{d,d'=1}^{D}\sum_{k,l=0}^{D}\int_0^T\mathbb{E}\Big[\sum_{i=1}^{n}|\eta^{H,(k)}_{n,i}\eta^{H,(l)}_{n,i}|\Big|\partial^2_{d+kD,d'+lD}g(W_s - W_{t^n_{i-1}}, \eta^H_{n,i}\otimes(W_s - W_{t^n_{i-1}}), \\ &\qquad J_{s-} - J_{t^n_{i-1}}, \psi(J_{s-} - J_{t^n_{i-1}})\xi_{n,i})\Big|\mathbb{1}_{(t^n_{i-1},t^n_i]}(s)\Big]\mathrm{d}s \\ &=: \frac{1}{2}\sum_{d,d'=1}^{D}\sum_{k,l=0}^{D}\int_0^T\mathbb{E}\big[G^n_{(5.13)}(s)\big]\mathrm{d}s. \end{aligned} \tag{5.13}$$

In order to derive (5.10), we prove for any $1 \le d, d' \le D$, $0 \le k, l \le D$ that

$$\int_0^T \mathbb{E}\big[G^n_{(5.13)}(s)\big]\mathrm{d}s \to 0 \quad \text{as } n \to \infty.$$

By the dominated convergence theorem, it is sufficient to show that

$$\lim_{n\to\infty}\mathbb{E}[G^n_{(5.13)}(s)] = 0 \quad \text{for all } s \in (0,T), \quad \text{and} \quad \int_0^T \sup_{n\ge1}\mathbb{E}[G^n_{(5.13)}(s)]\mathrm{d}s < \infty. \tag{5.14}$$

Indeed, for each fixed $s \in (0,T)$ one has

$$\eta^H_{n,i}\otimes(W_s - W_{t^n_{i-1}}) \xrightarrow{\mathbf{L}_2(\mathbb{P})} 0 \quad \text{and} \quad \psi(J_{s-} - J_{t^n_{i-1}})\xi_{n,i} \xrightarrow{\mathbf{L}_2(\mathbb{P})} 0$$

when $n \to \infty$ because of the independence, $t^n_{i-1} \to s$, and

$$\begin{aligned} &\mathbb{E}[\|\eta^H_{n,i}\otimes(W_s - W_{t^n_{i-1}})\|^2] = \mathbb{E}[\|\eta^H_{n,i}\|^2]\,\mathbb{E}[\|W_s - W_{t^n_{i-1}}\|^2] = D^2(s - t^n_{i-1}), \\ &\mathbb{E}[\|\psi(J_{s-} - J_{t^n_{i-1}})\xi_{n,i}\|^2] \le D\|\nabla\psi\|^2_\infty\mathbb{E}[\|J_{s-} - J_{t^n_{i-1}}\|^2] = (s - t^n_{i-1})D\|\nabla\psi\|^2_\infty\int_E\|e\|^2\nu(\mathrm{d}e). \end{aligned}$$

Since $\partial^2_{d+kD,d'+lD}g$ is continuous and is equal to 0 in a neighborhood of 0 by (G2), we get

$$G^n_{(5.13)}(s) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty,$$

where the convergence in probability can be asserted by showing that any subsequence has a further subsequence converging a.s. to 0. Moreover, since g has bounded second-order partial derivatives by (G1) and $\{\|\eta^H_{n,i}\|^2\}_{1\le i\le n, n\ge1}$ is uniformly integrable by Assumption 3.3, it implies that $\{G^n_{(5.13)}(s)\}_{n\ge1}$ is also uniformly integrable. Hence, the dominated convergence theorem is applicable to obtain the first assertion in (5.14). The integrability condition in (5.14) is easily verified by noting that

$$\begin{aligned} \sup_{n\ge1}\mathbb{E}[G^n_{(5.13)}(s)] &\le \|\nabla^2 g\|_\infty \sup_{1\le i\le n, n\ge1}\mathbb{E}[|\eta^{H,(k)}_{n,i}\eta^{H,(l)}_{n,i}|] \\ &\le \frac{1}{2}\|\nabla^2 g\|_\infty \sup_{1\le i\le n, n\ge1}\mathbb{E}[|\eta^{H,(k)}_{n,i}|^2 + |\eta^{H,(l)}_{n,i}|^2] = \|\nabla^2 g\|_\infty. \end{aligned}$$

Hence, (5.10) is proved.

Step 5. We prove $G^n_{(5.11)} \to 0$. By the independence and Fubini's theorem we obtain

$$\begin{aligned} G^n_{(5.11)} &\le \int_E \int_0^T \mathbb{E}\Big[\sum_{i=1}^n \Big| R_1^g(W_s - W_{t^n_{i-1}}, J_{s-} - J_{t^n_{i-1}}; \eta^H_{n,i}, \xi_{n,i}; e) \\ &\qquad - R_1^g(W_s - W_{t^n_{i-1}}, J_{s-} - J_{t^n_{i-1}}; 0, \xi_{n,i}; e) \Big| \mathbb{1}_{(t^n_{i-1}, t^n_i]}(s) \Big] \mathrm{d}s\nu(\mathrm{d}e) \\ &=: \int_E \int_0^T \mathbb{E}\big[G^n_{(5.15)}(s;e)\big] \mathrm{d}s\nu(\mathrm{d}e). \end{aligned} \tag{5.15}$$

By dominated convergence, it suffices to show that

$$\forall (s,e) \in (0,T) \times E : \lim_{n\to\infty} \mathbb{E}[G^n_{(5.15)}(s;e)] = 0, \tag{5.16}$$

$$\text{and} \quad \int_E \int_0^T \sup_{n\ge 1} \mathbb{E}[G^n_{(5.15)}(s;e)] \mathrm{d}s\nu(\mathrm{d}e) < \infty. \tag{5.17}$$

Indeed, for each $(s,e) \in (0,T) \times E$, using (5.7) yields

$$\begin{aligned} G^n_{(5.15)}(s;e) \le \sum_{i=1}^n \Big[& 4c_{(5.6)} \Big(\|W_s - W_{t^n_{i-1}}\| + \|W_s - W_{t^n_{i-1}}\|^2 + \|\eta^H_{n,i} \otimes (W_s - W_{t^n_{i-1}})\| \\ & + \|\eta^H_{n,i} \otimes (W_s - W_{t^n_{i-1}})\|^2 \Big) + c_{(5.7)} \|e\| (1 + \|\xi_{n,i}\|) \|\eta^H_{n,i} \otimes (W_s - W_{t^n_{i-1}})\| \Big] \mathbb{1}_{(t^n_{i-1}, t^n_i]}(s). \end{aligned}$$

Then, by Hölder's inequality we get

$$\begin{aligned} \mathbb{E}[G^n_{(5.15)}(s;e)] \le \sum_{i=1}^n \Big[& 4c_{(5.6)} \Big(\sqrt{D}\sqrt{s - t^n_{i-1}} + D(s - t^n_{i-1}) + D\sqrt{s - t^n_{i-1}} + D^2 (s - t^n_{i-1}) \Big) \\ & + c_{(5.7)} \|e\| \sqrt{\mathbb{E}[|1 + \|\xi_{n,i}\||^2]} \sqrt{\mathbb{E}[\|\eta^H_{n,i} \otimes (W_s - W_{t^n_{i-1}})\|^2]} \Big] \mathbb{1}_{(t^n_{i-1}, t^n_i]}(s) \\ \to 0 \quad & \text{as } n \to \infty, \end{aligned}$$

which then verifies (5.16). To show (5.17), we use the estimate (5.8) to get

$$\sup_{n\ge 1} \mathbb{E}[G^n_{(5.15)}(s;e)] \le 2c_{(5.8)} \sup_{n\ge 1, 1\le i\le n} \mathbb{E}[(1 + \|\xi_{n,i}\|^2)\|e\|^2] = 2c_{(5.8)}(D+1)\|e\|^2.$$

Since $\int_E \|e\|^2 \nu(\mathrm{d}e) < \infty$ by assumption, (5.17) follows.

Step 6. We show $G^n_{(5.12)} \to 0$. By the independence and Fubini's theorem one has

$$\begin{aligned} & G^n_{(5.12)} \\ & \le \sum_{i=1}^n \int_E \int_{t^n_{i-1}}^{t^n_i} \mathbb{E}\Big[\Big| R_1^g(W_s - W_{t^n_{i-1}}, J_{s-} - J_{t^n_{i-1}}; 0, \xi_{n,i}; e) - \int_{\mathbb{R}^D} g(0, e, \psi(e)u) \varphi_D(u) \mathrm{d}u \Big| \Big] \mathrm{d}s\nu(\mathrm{d}e) \\ & \le \int_{\mathbb{R}^D} \int_E \int_0^T \mathbb{E}\Big[\sum_{i=1}^n \Big| R_1^g(W_s - W_{t^n_{i-1}}, J_{s-} - J_{t^n_{i-1}}; 0, u; e) - g(0, e, \psi(e)u) \Big| \mathbb{1}_{(t^n_{i-1}, t^n_i]}(s) \Big] \\ & \qquad \times \mathrm{d}s\nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u \\ & =: \int_{\mathbb{R}^D} \int_E \int_0^T \mathbb{E}\big[G^n_{(5.18)}(s;e,u)\big] \mathrm{d}s\nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u. \end{aligned} \tag{5.18}$$

For any $(s,e,u) \in (0,T] \times E \times \mathbb{R}^D$, since the first two arguments in R_1^g converge to 0 a.s. as $n \to \infty$, we obtain that $G^n_{(5.18)}(s;e,u) \to 0$ a.s. Moreover, one has

$$\begin{aligned} \mathbb{E}\Big[\sup_{n\ge 1} |G^n_{(5.18)}(s;e,u)| \Big] &\le \mathbb{E}\Big[\sup_{n\ge 1, 1\le i\le n} |R_1^g(W_s - W_{t^n_{i-1}}, J_{s-} - J_{t^n_{i-1}}; 0, u; e)| \Big] + |g(0, e, \psi(e)u)| \\ &\le c_{(5.8)}(1 + \|u\|^2)\|e\|^2 + |g(0, e, \psi(e)u)|. \end{aligned}$$

Since, by (5.2),

$$\int_{\mathbb{R}^D}\int_E \Big((1+\|u\|^2)\|e\|^2 + |g(0,e,\psi(e)u)|\Big)\nu(\mathrm{d}e)\varphi_D(u)\mathrm{d}u < \infty,$$

the dominated convergence theorem implies that $G^n_{(5.12)} \to 0$ as $n \to \infty$. □

We first deal with the jump part of the limit of $(\mathcal{Z}^n)_{n\geq 1}$. To do this, we recall from [18, p.395] the function space $C_2(\mathbb{R}^{\mathbf{D}})$, which consists of all continuous bounded functions $g\colon \mathbb{R}^{\mathbf{D}} \to \mathbb{R}$ with $0 \notin \mathrm{supp}(g)$.

Lemma 5.3. *The assertion (5.1) holds true for $g \in C_2(\mathbb{R}^{\mathbf{D}})$. Consequently, for any $t \in [0,\infty)$ one has when $n \to \infty$ that*

$$\sum_{i=1}^{\sigma_t^n} \mathbb{E}[g(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] \xrightarrow{\mathbf{L}_1(\mathbb{P})} (t\wedge T)\int_{\mathbb{R}^{2D}} g(0,e,u)\nu_L^{\psi}(\mathrm{d}e,\mathrm{d}u).$$

Proof. It suffices to show the convergence for $t \in [0,T]$. Let $g \in C_2(\mathbb{R}^{\mathbf{D}})$ and assume that $\mathrm{supp}(g) \cap B_{\mathbf{D}}(r_g) = \emptyset$ for some $r_g > 0$. Let $\varepsilon > 0$ be arbitrarily small and $K > r_g$ a sufficiently large constant which is specified later. Since g is continuous and bounded, there is a continuous function g_K with compact support such that $\|g_K\|_\infty \leq \|g\|_\infty$ and $g_K = g$ on $B_{\mathbf{D}}(K)$. Moreover, by convolution approximation, there is a function $\hat{g}_{\varepsilon,K} \in C_2(\mathbb{R}^{\mathbf{D}}) \cap C_c^2(\mathbb{R}^{\mathbf{D}})$ such that $\mathrm{supp}(g_K - \hat{g}_{\varepsilon,K}) \cap B_{\mathbf{D}}(r_g/2) = \emptyset$ and $\|g_K - \hat{g}_{\varepsilon,K}\|_\infty \leq \varepsilon$. For $t \in (0,T]$, we denote

$$I^g_{(5.19)} := \sum_{i=1}^{\sigma_t^n} \left|\mathbb{E}[g(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] - (t_i^n - t_{i-1}^n)\int_{\mathbb{R}^{2D}} g(0,e,u)\nu_L^{\psi}(\mathrm{d}e,\mathrm{d}u)\right| \tag{5.19}$$

and then get by the triangle inequality that

$$I^g_{(5.19)} \leq I^{g-g_K}_{(5.19)} + I^{g_K-\hat{g}_{\varepsilon,K}}_{(5.19)} + I^{\hat{g}_{\varepsilon,K}}_{(5.19)}.$$

Since $\hat{g}_{\varepsilon,K} \in C_2(\mathbb{R}^{\mathbf{D}}) \cap C_c^2(\mathbb{R}^{\mathbf{D}}) \subset C_*^2(\mathbb{R}^{\mathbf{D}})$, according to Proposition 5.2 one has

$$I^{\hat{g}_{\varepsilon,K}}_{(5.19)} \xrightarrow{\mathbf{L}_1(\mathbb{P})} 0.$$

For the stochastic term in $I^{g-g_K}_{(5.19)}$, we have, a.s.,

$$\begin{aligned}
&\sum_{i=1}^{\sigma_t^n} \mathbb{E}[|(g-g_K)(\Delta_{n,i}\mathcal{Z}^n)|\,|\mathcal{F}_{n,i-1}] \leq \|g-g_K\|_\infty \sum_{i=1}^{n} \mathbb{E}[\mathbb{1}_{\{\|\Delta_{n,i}\mathcal{Z}^n\|\geq K\}}|\mathcal{F}_{n,i-1}]\\
&\leq \frac{2\|g\|_\infty}{K^2}\sum_{i=1}^{n}\mathbb{E}[\|\Delta_{n,i}\mathcal{Z}^n\|^2|\mathcal{F}_{n,i-1}]\\
&= \frac{2\|g\|_\infty}{K^2}\sum_{i=1}^{n}\mathbb{E}\Big[\|\Delta_{n,i}W\|^2 + \|\eta^H_{n,i}\otimes\Delta_{n,i}W\|^2 + \|\Delta_{n,i}J\|^2 + \psi(\Delta_{n,i}J)^2\|\xi_{n,i}\|^2\,\Big|\mathcal{F}_{n,i-1}\Big]\\
&\leq \frac{2\|g\|_\infty}{K^2}\sum_{i=1}^{n}\left[(t_i^n - t_{i-1}^n)(D+D^2) + (1+D\|\nabla\psi\|_\infty^2)(t_i^n - t_{i-1}^n)\int_E \|e\|^2\nu(\mathrm{d}e)\right]\\
&= \frac{2T\|g\|_\infty}{K^2}\left[D + D^2 + (1+D\|\nabla\psi\|_\infty^2)\int_E\|e\|^2\nu(\mathrm{d}e)\right].
\end{aligned}$$

For the stochastic term in $I^{g_K-\hat{g}_{\varepsilon,K}}_{(5.19)}$, we use the same arguments as for $I^{g-g_K}_{(5.19)}$ to obtain, a.s.,

$$\sum_{i=1}^{\sigma_t^n} \mathbb{E}[|(g_K-\hat{g}_{\varepsilon,K})(\Delta_{n,i}\mathcal{Z}^n)|\,|\mathcal{F}_{n,i-1}] \leq \|g_K-\hat{g}_{\varepsilon,K}\|_\infty \sum_{i=1}^{n} \mathbb{E}[\mathbb{1}_{\{\|\Delta_{n,i}\mathcal{Z}^n\|\geq r_g/2\}}|\mathcal{F}_{n,i-1}]$$

$$\leq \frac{4T\varepsilon}{r_g^2}\left[D + D^2 + (1 + D\|\nabla\psi\|_\infty^2)\int_E \|e\|^2\nu(\mathrm{d}e)\right].$$

Then, by the triangle inequality,

$$I^{g-g_K}_{(5.19)} \leq \frac{2T\|g\|_\infty}{K^2}\left[D + D^2 + (1 + D\|\nabla\psi\|_\infty^2)\int_E \|e\|^2\nu(\mathrm{d}e)\right] + 2T\|g\|_\infty \int_{B^c_{2D}(K)} \nu^\psi_L(\mathrm{d}e, \mathrm{d}u)$$

which can be made arbitrarily small as long as we choose a sufficiently large $K > 0$. Analogously,

$$I^{g_K-\hat{g}_{\varepsilon,K}}_{(5.19)} \leq \varepsilon\left[\frac{4T}{r_g^2}\left(D + D^2 + (1 + D\|\nabla\psi\|_\infty^2)\int_E \|e\|^2\nu(\mathrm{d}e)\right) + T\int_{B^c_{2D}(r_g/2)} \nu^\psi_L(\mathrm{d}e, \mathrm{d}u)\right].$$

Eventually, since $\varepsilon > 0$ is arbitrarily small, it implies that $I^g_{(5.19)} \xrightarrow{\mathbf{L}_1(\mathbb{P})} 0$. □

We continue to investigate the continuous and the drift components of the limit of $(\mathcal{Z}^n)_{n\geq 1}$. To this end, let us fix a truncation function $h\colon \mathbb{R}^{\mathbf{D}} \to \mathbb{R}^{\mathbf{D}}$ in the sense of [18, Definition II.2.3], i.e. h is bounded and $h(z) = z$ in a neighborhood of 0. As we will see later that the limit of $(\mathcal{Z}^n)_{n\geq 1}$ does not depend on the particular form of truncation function, we assume that $h = (h^{(d)})_{d=1}^{\mathbf{D}}$ with $h^{(d)} \in C_b^2(\mathbb{R}^{\mathbf{D}})$.

Lemma 5.4. *For any $t \in [0, \infty)$, one has when $n \to \infty$ that*

$$\sup_{s\leq t}\left\|\sum_{i=1}^{\sigma^n_s} \mathbb{E}[h(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] - B_s\right\| \xrightarrow{\mathbf{L}_1(\mathbb{P})} 0,$$

where $B := B(h)$ given by

$$B_t := (t \wedge T)\int_{\mathbb{R}^{2D}} (h(0, e, u) - (0, e, u)^{\mathsf{T}})\nu^\psi_L(\mathrm{d}e, \mathrm{d}u).$$

Proof. It is sufficient to consider $t \in [0, T]$ and prove that for any $d = 1, \ldots, \mathbf{D}$ one has

$$I^{(d)}_{(5.20)} := \sup_{s\leq t}\left|\sum_{i=1}^{\sigma^n_s} \mathbb{E}[h^{(d)}(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] - sB^{(d)}_1\right| \xrightarrow{\mathbf{L}_1(\mathbb{P})} 0, \tag{5.20}$$

Let $\tilde{h}^{(d)}(z) := h^{(d)}(z) - z^{(d)}$ for $z = (z^{(1)}, \ldots, z^{(\mathbf{D})}) \in \mathbb{R}^{\mathbf{D}}$. It follows from $\mathbb{E}[\Delta_{n,i}\mathcal{Z}^n|\mathcal{F}_{n,i-1}] = 0$ a.s. that

$$\mathbb{E}[h^{(d)}(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] = \mathbb{E}[\tilde{h}^{(d)}(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] \quad \text{a.s.} \tag{5.21}$$

Hence we now prove (5.20) for $\tilde{h}^{(d)}$ in place of $h^{(d)}$. We remark that there is no problem regarding $\mathbb{P}$-null sets for that replacement as only countably many random variables are considered in (5.20). On the other hand, since $h^{(d)} \in C_b^2(\mathbb{R}^{\mathbf{D}})$ and $h^{(d)}(z) = z^{(d)}$ in a neighborhood of 0, it is straightforward to check that $\tilde{h}^{(d)} \in C_*^2(\mathbb{R}^{\mathbf{D}})$. By the triangle inequality, a.s.,

$$\begin{aligned} I^{(d)}_{(5.20)} &\leq \sup_{s\leq t}\left|\sum_{i=1}^{\sigma^n_s} \mathbb{E}[\tilde{h}^{(d)}(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] - t^n_{\sigma^n_s}B^{(d)}_1\right| + \sup_{s\leq t}\left|t^n_{\sigma^n_s}B^{(d)}_1 - sB^{(d)}_1\right| \\ &\leq \sum_{i=1}^{n}\left|\mathbb{E}[\tilde{h}^{(d)}(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] - (t^n_i - t^n_{i-1})B^{(d)}_1\right| + \max_{1\leq i\leq n}(t^n_i - t^n_{i-1})|B^{(d)}_1|. \end{aligned}$$

According to Proposition 5.2, the first term on the right-hand side converges to 0 in $\mathbf{L}_1(\mathbb{P})$. The second term $\max_{1\leq i\leq n}(t^n_i - t^n_{i-1})|B^{(d)}_1|$ obviously tends to 0 as $n \to \infty$. Hence, (5.20) follows. □

We now investigate the continuous part of the limit of $(\mathcal{Z}^n)_{n\geq 1}$. For $t\in[0,\infty)$, we define the matrices $\boldsymbol{C}_t=(C_t^{(k,l)})\in\mathbb{R}^{\mathbf{D}}\times\mathbb{R}^{\mathbf{D}}$ and its modification $\widetilde{\boldsymbol{C}}_t=(\widetilde{C}_t^{(k,l)})\in\mathbb{R}^{\mathbf{D}}\times\mathbb{R}^{\mathbf{D}}$ by

$$C_t^{(k,l)}:=\begin{cases} t\wedge T & \text{if } 1\leq k=l\leq D^2+D\\ 0 & \text{otherwise,}\end{cases} \tag{5.22}$$

and

$$\widetilde{C}_t^{(k,l)}:=C_t^{(k,l)}+(t\wedge T)\int_{\mathbb{R}^{2D}}(h^{(k)}h^{(l)})(0,e,u)\nu_L^{\psi}(\mathrm{d}e,\mathrm{d}u).$$

Lemma 5.5. *For any $t\in[0,\infty)$ and $1\leq k,l\leq\mathbf{D}$, one has when $n\to\infty$ that*

$$I_{(5.23)}:=\sum_{i=1}^{\sigma_t^n}\mathbb{E}[h^{(k)}(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}]\,\mathbb{E}[h^{(l)}(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}]\xrightarrow{\mathbf{L}_1(\mathbb{P})}0, \tag{5.23}$$

$$I_{(5.24)}:=\sum_{i=1}^{\sigma_t^n}\mathbb{E}[(h^{(k)}h^{(l)})(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}]\xrightarrow{\mathbf{L}_1(\mathbb{P})}\widetilde{C}_t^{(k,l)}. \tag{5.24}$$

Proof. It suffices to prove for $t\in[0,T]$. We first show that $I_{(5.23)}\xrightarrow{\mathbf{L}_1(\mathbb{P})}0$ as $n\to\infty$. In the sequel we employ the notation as in the proof of Lemma 5.4. According to (5.21) one has, a.s.,

$$\begin{aligned}I_{(5.23)}=&\sum_{i=1}^{\sigma_t^n}\mathbb{E}\Big[\tilde{h}^{(k)}(\Delta_{n,i}\mathcal{Z}^n)-(t_i^n-t_{i-1}^n)B_1^{(k)}\Big|\mathcal{F}_{n,i-1}\Big]\mathbb{E}[h^{(l)}(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}]\\&+B_1^{(k)}\sum_{i=1}^{\sigma_t^n}(t_i^n-t_{i-1}^n)\mathbb{E}\Big[\tilde{h}^{(l)}(\Delta_{n,i}\mathcal{Z}^n)-(t_i^n-t_{i-1}^n)B_1^{(l)}\Big|\mathcal{F}_{n,i-1}\Big]+B_1^{(k)}B_1^{(l)}\sum_{i=1}^{\sigma_t^n}(t_i^n-t_{i-1}^n)^2.\end{aligned}$$

Then, a.s.,

$$\begin{aligned}|I_{(5.23)}|\leq&\|h^{(l)}\|_\infty\sum_{i=1}^{n}\Big|\mathbb{E}\Big[\tilde{h}^{(k)}(\Delta_{n,i}\mathcal{Z}^n)-(t_i^n-t_{i-1}^n)B_1^{(k)}\Big|\mathcal{F}_{n,i-1}\Big]\Big|\\&+|B_1^{(k)}|\max_{1\leq i\leq n}(t_i^n-t_{i-1}^n)\sum_{i=1}^{n}\Big|\mathbb{E}\Big[\tilde{h}^{(l)}(\Delta_{n,i}\mathcal{Z}^n)-(t_i^n-t_{i-1}^n)B_1^{(l)}\Big|\mathcal{F}_{n,i-1}\Big]\Big|\\&+t|B_1^{(k)}B_1^{(l)}|\max_{1\leq i\leq n}(t_i^n-t_{i-1}^n).\end{aligned}$$

Since $\max_{1\leq i\leq n}(t_i^n-t_{i-1}^n)\to 0$, applying Proposition 5.2 yields (5.23).

We next show that $I_{(5.24)}\to\widetilde{C}_t^{(k,l)}$ in $\mathbf{L}_1(\mathbb{P})$. For $z=(z^{(1)},\ldots,z^{(\mathbf{D})})\in\mathbb{R}^{\mathbf{D}}$, we define

$$q^{(k,l)}(z):=z^{(k)}z^{(l)}\quad\text{and}\quad\hat{h}^{(k,l)}(z):=\begin{cases}(h^{(k)}h^{(l)})(z)-q^{(k,l)}(z) & \text{if } 1\leq k\vee l\leq D^2+D\\ (h^{(k)}h^{(l)})(z) & \text{otherwise.}\end{cases}$$

We now verify that $\hat{h}^{(k,l)}\in C_*^2(\mathbb{R}^{\mathbf{D}})$ for any $k,l=1,\ldots,\mathbf{D}$:

- $\hat{h}^{(k,l)}$ obviously satisfies (G1).
- Let $1\leq d\vee d'\leq D^2+D$. If $k\vee l\leq D^2+D$, then $\hat{h}^{(k,l)}$, and thus $\partial^2_{d,d'}\hat{h}^{(k,l)}$, are 0 in a neighborhood of 0. If $k\vee l\geq D^2+D+1$, then $\partial^2_{d,d'}\hat{h}^{(k,l)}=\partial^2_{d,d'}(h^{(k)}h^{(l)}-q^{(k,l)})$, which also shows that $\partial^2_{d,d'}\hat{h}^{(k,l)}$ is 0 around 0. Hence, (G2) is satisfied.
- For $d=1,\ldots,D^2+D$ and for any $j\in\mathbb{R}^{2D}$, one has

$$\partial_d\hat{h}^{(k,l)}(0,j)=\begin{cases}\partial_d(h^{(k)}h^{(l)})(0,j)-\partial_d q^{(k,l)}(0,j) & \text{if } 1\leq k\vee l\leq D^2+D\\ \partial_d(h^{(k)}h^{(l)})(0,j) & \text{otherwise}\end{cases}=\partial_d(h^{(k)}h^{(l)})(0,j).$$

Hence, $\max_{1\le d\le D^2+D}\|\partial_d\hat{h}^{(k,l)}(0_{D^2+D},\cdot)\|_\infty \le \|\nabla(h^{(k)}h^{(l)})\|_\infty < \infty$, which verifies (G3).

- For $d = D^2+D+1,\dots,\mathbf{D}$, since $\partial_d q^{(k,l)} = 0$ if $k\vee l \le D^2+D$ we infer that $\partial_d\hat{h}^{(k,l)} = \partial_d(h^{(k)}h^{(l)})$ and $\partial_d\hat{h}^{(k,l)}(0) = h^{(l)}(0)\partial_d h^{(k)}(0) + h^{(k)}(0)\partial_d h^{(l)}(0) = 0$. Thus, (G4) is satisfied.

Applying Proposition 5.2 and noticing that, for any $1\le k,l\le \mathbf{D}$,

$$\int_{\mathbb{R}^{2D}} \hat{h}^{(k,l)}(0,e,u)\nu_L^\psi(\mathrm{d}e,\mathrm{d}u) = \int_{\mathbb{R}^{2D}} (h^{(k)}h^{(l)})(0,e,u)\nu_L^\psi(\mathrm{d}e,\mathrm{d}u)$$

we obtain

$$\sum_{i=1}^{\sigma_t^n}\left|\mathbb{E}[\hat{h}^{(k,l)}(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] - (t_i^n - t_{i-1}^n)\int_{\mathbb{R}^{2D}}(h^{(k)}h^{(l)})(0,e,u)\nu_L^\psi(\mathrm{d}e,\mathrm{d}u)\right| \xrightarrow{\mathbf{L}_1(\mathbb{P})} 0. \tag{5.25}$$

On the other hand, for $1\le k\vee l\le D^2+D$, a direct calculation exploiting the independence and (3.1) gives the following convergence as $n\to\infty$, particularly in $\mathbf{L}_1(\mathbb{P})$,

$$\begin{aligned}\sum_{i=1}^{\sigma_t^n}\mathbb{E}[q^{(k,l)}(\Delta_{n,i}\mathcal{Z}^n)|\mathcal{F}_{n,i-1}] &= \sum_{i=1}^{\sigma_t^n}\mathbb{E}[\Delta_{n,i}\mathcal{Z}^{n,(k)}\Delta_{n,i}\mathcal{Z}^{n,(l)}|\mathcal{F}_{n,i-1}]\\ &= \begin{cases} t^n_{\sigma_t^n} & \text{if } 1\le k=l\le D^2+D\\ 0 & \text{otherwise}\end{cases} \to tC_1^{(k,l)}.\end{aligned}$$

Therefore, (5.24) follows from (5.25), and the proof is completed. □

Proof of Theorem 3.5. We combine [18, Theorem VIII.2.29] with Lemmas 5.3 to 5.5 to obtain that $(\mathcal{Z}^n_{t\wedge T})_{t\in[0,\infty)} \to \mathcal{Z}$ weakly in the Skorokhod topology on the space of càdlàg functions $:[0,\infty)\to\mathbb{R}^{\mathbf{D}}$. Here, $\mathcal{Z}$ is a semimartingale with the predictable characteristic[6] $(B,C,m_{\mathcal{Z}})$ associated with the truncation function h, where

- $\mathcal{Z}_0 = 0$ as $\mathcal{Z}_0^n = 0$ for all n;
- h is taken as in the paragraph right before Lemma 5.4;
- B is provided in Lemma 5.4;
- C is defined in (5.22);
- $m_{\mathcal{Z}}(\mathrm{d}t,\mathrm{d}z) = \nu_{\mathcal{Z}}(\mathrm{d}z)\boldsymbol{\lambda}_{[0,T]}(\mathrm{d}t)$, where $\boldsymbol{\lambda}_{[0,T]}$ is the restriction of the Lebesgue measure on $[0,T]$, $\nu_{\mathcal{Z}}$ is a Lévy measure on $\mathbb{R}_0^{\mathbf{D}} := \mathbb{R}^{\mathbf{D}}\backslash\{0\}$ with support on $\{0\}\times\mathbb{R}_0^{2D}$, i.e. $\nu_{\mathcal{Z}}(\mathbb{R}_0^{D^2+D}\times\mathbb{R}_0^{2D}) = 0$, and such that $\nu_{\mathcal{Z}}(\{0\}\times B) = \nu_L^\psi(B)$ for $B\in\mathcal{B}(\mathbb{R}_0^{2D})$.

Note that $((W_{t\wedge T},\mathcal{W}_{t\wedge T}))_{t\in[0,\infty)}$ and $(L^\psi_{t\wedge T})_{t\in[0,\infty)}$ are independent due to Lemma B.3. Then a standard calculation using Lévy–Khintchine formula shows that $(\mathrm{vec}(W_{t\wedge T},\mathcal{W}_{t\wedge T},L^\psi_{t\wedge T}))_{t\in[0,\infty)}$ is a (time-inhomogeneous) Lévy process with characteristic triplet $(B,C,m_{\mathcal{Z}})$ with respect to the truncation function h. Hence, we derive from [18, Theorem VIII.2.29] the weak convergence

$$(\mathcal{Z}^n_{t\wedge T})_{t\in[0,\infty)} \to (\mathrm{vec}(W_{t\wedge T},\mathcal{W}_{t\wedge T},L^\psi_{t\wedge T}))_{t\in[0,\infty)}.$$

Eventually, since the limit process has no fixed time of discontinuity, we apply [7, Theorem 16.7] to obtain the weak convergence on the time interval $[0,T]$ as desired. □

[6] in the sense of [18, Definition II.2.6].

Appendix A. Proof of Proposition 3.2

Condition $\mathbb{E}\big[\int_{\mathbb{R}^D} \|H_{n,i-1}(u)\|^2 \varphi_D(u)\mathrm{d}u\big] < \infty$ allows us to define

$$\mu^H_{n,i-1} := \int_{\mathbb{R}^D} H_{n,i-1}(u)\varphi_D(u)\mathrm{d}u, \quad \widetilde{H}_{n,i-1}(u) := H_{n,i-1}(u) - \mu^H_{n,i-1},$$
$$\Theta^H_{n,i-1} := \int_{\mathbb{R}^D} \widetilde{H}_{n,i-1}(u)\widetilde{H}_{n,i-1}(u)^\mathsf{T}\varphi_D(u)\mathrm{d}u.$$

Obviously $\mu^H_{n,i-1} \in \mathbf{L}_2(\mathbb{P})$. In addition, the finiteness of accumulative entropy implies that $\det(\Theta^H_{n,i-1}) > 0$ a.s. for all n, i. Since $\Theta^H_{n,i-1} \in \mathbb{S}^D_{++}$, we apply the spectral theorem for symmetric matrices to obtain a real diagonal matrix $\Lambda^H_{n,i-1} = \mathrm{diag}(\lambda_1(\Theta^H_{n,i-1}), \ldots, \lambda_D(\Theta^H_{n,i-1}))$ with $\lambda_1(\Theta^H_{n,i-1}) \geq \cdots \geq \lambda_D(\Theta^H_{n,i-1}) > 0$ and a $U^H_{n,i-1} \in \mathcal{O}_D$, such that

$$\Theta^H_{n,i-1} = U^H_{n,i-1}\Lambda^H_{n,i-1}(U^H_{n,i-1})^\mathsf{T}.$$

One remarks that $U^H_{n,i-1}$ and $\Lambda^H_{n,i-1}$ are matrices whose entries are $\mathcal{F}_{n,i-1}$-measurable random variables. Now, by adjusting on a $\mathbb{P}$-null set, we define

$$\vartheta^H_{n,i-1} := (\Theta^H_{n,i-1})^{\frac{1}{2}} = U^H_{n,i-1}(\Lambda^H_{n,i-1})^{\frac{1}{2}}(U^H_{n,i-1})^\mathsf{T},$$
$$\eta^H_{n,i} := U^H_{n,i-1}\hat{\eta}^H_{n,i}, \quad \text{where} \quad \hat{\eta}^{H,(d)}_{n,i} := \frac{1}{\sqrt{\lambda_d(\Theta^H_{n,i-1})}}(U^H_{n,i-1}\mathbf{e}_d)^\mathsf{T}\widetilde{H}_{n,i-1}(\xi_{n,i}), \quad d = 1, \ldots, D.$$

Then it is easy to check that $\vartheta^H_{n,i-1} \in \mathbf{L}_2(\mathbb{P})$. Moreover, for $d = 1, \ldots, D$, one has, a.s.,

$$[\vartheta^H_{n,i-1}\eta^H_{n,i}]^{(d)} = \sum_{k=1}^D U^{H,(d,k)}_{n,i-1}\sqrt{\lambda_k(\Theta^H_{n,i-1})}\hat{\eta}^{H,(k)}_{n,i} = \sum_{k,l=1}^D U^{H,(d,k)}_{n,i-1}U^{H,(l,k)}_{n,i-1}\widetilde{H}^{(l)}_{n,i-1}(\xi_{n,i})$$
$$= \sum_{l=1}^D [U^H_{n,i-1}(U^H_{n,i-1})^\mathsf{T}]^{(d,l)}\widetilde{H}^{(l)}_{n,i-1}(\xi_{n,i}) = \widetilde{H}^{(d)}_{n,i-1}(\xi_{n,i}),$$

which shows $\vartheta^H_{n,i-1}\eta^H_{n,i} = \widetilde{H}_{n,i-1}(\xi_{n,i})$ a.s. For any $d = 1, \ldots, D$, we let $\hat{\eta}^{H,(d)}_{n,i}(\varepsilon)$ be the random variable obtained by adding $\varepsilon > 0$ to $\lambda_d(\Theta^H_{n,i-1})$ in the definition of $\hat{\eta}^{H,(d)}_{n,i}$. Then one has, a.s.,

$$\mathbb{E}\Big[\big|\hat{\eta}^{H,(d)}_{n,i}(\varepsilon)\big|^2\Big|\mathcal{F}_{n,i-1}\Big] = \frac{1}{\lambda_d(\Theta^H_{n,i-1}) + \varepsilon}\mathbf{e}_d^\mathsf{T}(U^H_{n,i-1})^\mathsf{T}\mathbb{E}\Big[\widetilde{H}_{n,i-1}(\xi_{n,i})(\widetilde{H}_{n,i-1}(\xi_{n,i}))^\mathsf{T}\Big|\mathcal{F}_{n,i-1}\Big]U^H_{n,i-1}\mathbf{e}_d$$
$$= \frac{1}{\lambda_d(\Theta^H_{n,i-1}) + \varepsilon}\mathbf{e}_d^\mathsf{T}(U^H_{n,i-1})^\mathsf{T}\Theta^H_{n,i-1}U^H_{n,i-1}\mathbf{e}_d = \frac{\mathbf{e}_d^\mathsf{T}\Lambda^H_{n,i-1}\mathbf{e}_d}{\lambda_d(\Theta^H_{n,i-1}) + \varepsilon}$$
$$= \frac{\lambda_d(\Theta^H_{n,i-1})}{\lambda_d(\Theta^H_{n,i-1}) + \varepsilon}.$$

Letting $\varepsilon \downarrow 0$ yields $\mathbb{E}[|\hat{\eta}^{H,(d)}_{n,i}|^2|\mathcal{F}_{n,i-1}] = 1$ a.s. by the monotone convergence theorem, and thus, $\|\hat{\eta}^H_{n,i}\| \in \mathbf{L}_2(\mathbb{P})$ as a by-product. Analogously, we can show that $\mathbb{E}[\hat{\eta}^{H,(d)}_{n,i}\hat{\eta}^{H,(d')}_{n,i}|\mathcal{F}_{n,i-1}] = \mathbb{1}_{\{d=d'\}}$ a.s., which means that $\mathbb{E}[\hat{\eta}^H_{n,i}(\hat{\eta}^H_{n,i})^\mathsf{T}|\mathcal{F}_{n,i-1}] = I_D$. Then we get $\mathbb{E}[\eta^H_{n,i}(\eta^H_{n,i})^\mathsf{T}|\mathcal{F}_{n,i-1}] = I_D$ a.s., and hence, (3.2) follows. The uniqueness is straightforward.

Appendix B. Some auxiliary results

B.1. **Positive semidefinite matrices.** For matrices $A, B \in \mathbb{S}^D$ we write $A \preceq B$ if $B - A \in \mathbb{S}^D_+$.

Lemma B.1 ([15], Sec.82, Exercises 12 and 13)**.**

(1) *For $A, B \in \mathbb{S}^D_+$ with $A \preceq B$ one has $\det(A) \leq \det(B)$.*

(2) *Let $A, B \in \mathbb{S}^D_{++}$ with $A \preceq B$. Then $B^{-1} \preceq A^{-1}$ and $\mathbf{tr}[AC] \leq \mathbf{tr}[BC]$ for any $C \in \mathbb{S}^D_+$.*

B.2. **Integrability for solutions of SDEs with jumps.** Although the following fact can be easily extended to a multidimensional setting, however, we formulate it in the one-dimensional case for the sake of simplicity. We refer to [5] for its proof.

Lemma B.2. *Let $\xi = (\xi_t)_{t\in[0,T]}$ be càdlàg and adapted with $\|\xi\|^2_{\mathcal{S}_2([0,T])} := \mathbb{E}[\sup_{0\le t\le T}\xi_t^2] < \infty$. Assume that $\mathrm{d}Z_t = \phi_t\mathrm{d}t + \mathrm{d}K_t$, where $K = (K_t)_{t\in[0,T]}$ is a càdlàg $\mathbf{L}_2(\mathbb{P})$-martingale satisfying $\mathrm{d}\langle K,K\rangle_t = \eta_t^2\mathrm{d}t$, where η and ϕ are progressively measurable with $\sup_{0<t<T}\eta_t^2 + \int_0^T\phi_t^2\mathrm{d}t \le C$ a.s. for some (non-random) constant $C>0$. Then, for a Lipschitz function $\sigma\colon\mathbb{R}\to\mathbb{R}$, the SDE*

$$X_t = \xi_t + \int_0^t \sigma(X_{u-})\mathrm{d}Z_u, \quad X_0 = \xi_0 = x_0 \in \mathbb{R},$$

has a unique càdlàg strong solution $X = (X_t)_{t\in[0,T]}$ satisfying $\mathbb{E}[\sup_{0\le t\le T}X_t^2] \le C' < \infty$ for some constant $C' = C'(\|\xi\|_{\mathcal{S}_2([0,T])}, T, \sigma, C) > 0$.

B.3. **Independence of Gaussian and purely non-Gaussian Lévy processes.** Lévy processes in the following assertion are considered with the canonical truncation function $h(x) = x\mathbb{1}_{\{\|x\|\le 1\}}$. We refer to [5] for its proof.

Lemma B.3. *Let $D, D' \in \mathbb{N}$. Assume that W is a D-dimensional Gaussian Lévy process and L is a D'-dimensional purely non-Gaussian Lévy process, both defined on the same probability space. Then W and L are independent.*

References

1. Aït-Sahalia, Y., Jacod, J.: High-frequency financial econometrics. Princeton University Press (2014)
2. Applebaum, D.: Lévy processes and stochastic calculus, 2nd ed. University Press, Cambridge (2009)
3. Barles, G., Buckdahn, R., Pardoux, E.: Backward stochastic differential equations and integral-partial differential equations. Stochastics 60, 57–83 (1997)
4. Bender, C., Thuan, N.T.: Continuous time reinforcement learning: a random measure approach. Preprint (2024)
5. Bender, C., Thuan, N.T.: Supplementary material for "Entropy-regularized mean-variance portfolio optimization with jumps".
6. Bhatia, R.: Matrix analysis. Springer-Verlag New York (1997)
7. Billingsley, P.: Convergence of probability measures, 2nd ed. John Wiley & Sons, Inc. (1999)
8. Cont, R., Tankov, P.: Financial modeling with jump processes. Chapman & Hall/CRC Press (2003)
9. Cover, T., Thomas, J.: Elements of information theory, 2nd ed. John Wiley & Sons (2006)
10. Dai, M., Dong, Y., Jia, Y.: Learning equilibrium mean-variance strategy. Math. Finance 33, 1166–1212 (2023)
11. Donnelly, R., Jaimungal, S.: Exploratory control with Tsallis entropy for latent factor models. SIAM J. Financial Math. 15, 26–53 (2024)
12. Gao, X., Li, L., Zhou, X.Y.: Reinforcement learning for jump-diffusions, with financial applications. Preprint (2024)
13. Guo, X., Hu, A., Zhang, Y.: Reinforcement learning for linear-convex models with jumps via stability analysis of feedback controls. SIAM J. Control Optim. 61, 755–787 (2023)
14. Guo, X., Xu, R., Zariphopoulou, T.: Entropy regularization for mean field games with learning. Math. Oper. Res. 47, 3239–3260 (2022)
15. Halmos, P.R.: Finite-dimensional vector spaces. Springer New York (1974)
16. Hambly, B., Xu, R., Yang, H.: Recent advances in reinforcement learning in finance. Math. Finance 33, 437–503 (2023)
17. Harville, D.A.: Matrix algebra from a statistician's perspective. Springer-Verlag, New York (1997)
18. Jacod, J., Shiryaev, A.: Limit theorems for stochastic processes, 2nd ed. Springer, Berlin Heidelberg (2003)
19. Jeanblanc, M., Mania, M., Santacroce, M., Schweizer, M.: Mean-variance hedging via stochastic control and BSDEs for general semimartingales. Ann. Appl. Probab. 22, 2388–2428 (2012)

20. Jia, Y., Zhou, X.Y.: Policy gradient and actor-critic learning in continuous time and space: Theory and algorithms. J. Mach. Learn. Res. 23, 1–50 (2022)
21. Jia, Y., Zhou, X.Y.: q-learning in continuous time. J. Mach. Learn. Res. 24, 1–61 (2023)
22. Kunita, H.: Stochastic differential equations based on Lévy processes and stochastic flows of diffeomorphisms. In: Rao, M.M. (eds.): Real and stochastic analysis. Trends in Mathematics, pp. 305–373. Birkhäuser Boston (2004)
23. Kushner, H.: Jump-diffusions with controlled jumps: existence and numerical methods. J. Math. Anal. Appl. 249, 179–198 (2000)
24. Li, X., Zhou, X.Y., Lim, A.E.: Dynamic mean-variance portfolio selection with no-shorting constraints. SIAM J. Control Optim. 40, 1540–1555 (2002)
25. Lim, A.E.: Mean-variance hedging when there are jumps. SIAM J. Control Optim. 44, 1893–1922 (2005)
26. Ma, J., Yong, J., Zhao, Y.: Four step scheme for general Markovian forward-backward SDEs. J. Syst. Sci. Complex 23, 546–571 (2010)
27. Markowitz, H.: Portfolio selection. Journal of Finance 7, 77–91 (1952)
28. Øksendal, B., Sulem, A.: Applied stochastic control of jump diffusions, 3rd ed. Springer Cham (2019)
29. Podczeck, K.: On existence of rich Fubini extensions. Econ. Theory 45, 1–22 (2010)
30. Protter, P.: Stochastic integration and differential equations, 2nd ed. Springer Berlin, Heidelberg (2005)
31. Sun, Y.: The exact law of large numbers via Fubini extension and characterization of insurable risks. J. Econ. Theory 126, 31–69 (2006)
32. Sun, Y., Zhang, Y.: Individual risk and Lebesgue extension without aggregate uncertainty. J. Econ. Theory 144, 432–443 (2009)
33. Szpruch, L., Treetanthiploet, T., Zhang, Y.: Optimal scheduling of entropy regularizer for continuous-time linear-quadratic reinforcement learning. SIAM J. Control Optim. 62, 135–166 (2024)
34. Wang, H., Zhou, X.Y.: Continuous-time mean-variance portfolio selection: A reinforcement learning framework. Math. Finance 30, 1–36 (2020)
35. Wang, H., Zariphopoulou, T., Zhou, X.Y.: Reinforcement learning in continuous time and space: A stochastic control approach. J. Mach. Learn. Res. 21, 1–34 (2020)
36. Wu, B., Li, L.: Reinforcement learning for continuous-time mean-variance portfolio selection in a regime-switching market. J. Econ. Dyn. Control. 158, Article 104787 (2024)
37. Zhang, Y., Li, X., Guo, S.: Portfolio selection problems with Markowitz's mean-variance framework: a review of literature. Fuzzy Optim. Decis. Making 17, 125–158 (2018)
38. Zhou, X.Y., Li, D.: Continuous-time mean-variance portfolio selection: A stochastic LQ framework. Appl. Math. Optim. 42, 19–33 (2000)