# A discretized version of Krylov's estimate and its applications[*]

## Xicheng Zhang[†]

### Abstract

In this paper we prove a discretized version of Krylov's estimate for discretized Itô processes. As applications, we study the weak and strong convergences for Euler's approximation of mean-field SDEs with measurable discontinuous and linear growth coefficients. Moreover, we also show the propagation of chaos for Euler's approximation of mean-field SDEs.

## 1 Introduction

### 1.1 Discretized Krylov's estimate

Let $(\Omega, \mathscr{F}, \mathbb{P}; (\mathscr{F}_t)_{t \geqslant 0})$ be a complete filtration probability space and $(W_t)_{t \geqslant 0}$ a $d$-dimensional standard $\mathscr{F}_t$-Brownian motion. Let $\xi_t$ be a $d$-dimensional Itô process with the following form

$$\xi_t = \xi_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s, \tag{1.1}$$

where $\xi_0 \in \mathscr{F}_0$, $b_s(\omega) : \mathbb{R}_+ \times \Omega \to \mathbb{R}^d$ and $\sigma_s(\omega) : \mathbb{R}_+ \times \Omega \to \mathbb{R}^d \otimes \mathbb{R}^d$ are bounded measurable $\mathscr{F}_t$-adapted processes with bound $\kappa_0$. Suppose that for some $\kappa_1 > 0$,

$$\det(\sigma_s(\omega)\sigma_s^*(\omega)) \geqslant \kappa_1, \ \forall (s, \omega) \in \mathbb{R}_+ \times \Omega,$$

where the asterisk stands for the transpose of a matrix. It is well known that for any $T > 0$ and $p \geqslant d+1$, there exists a constant $C > 0$ depending only on $\kappa_0, \kappa_1, p$ and $d$ such

---

[†]School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, P.R. China.
E-mail: XichengZhang@gmail.com

that for all $f \in L^p([0,T] \times \mathbb{R}^d)$,

$$\mathbb{E}\left(\int_0^T f(s,\xi_s)\mathrm{d}s\right) \leqslant C\|f\|_{L^p([0,T]\times\mathbb{R}^d)}, \tag{1.2}$$

and for time-independent $f \in L^p(\mathbb{R}^d)$ with $p \geqslant d$,

$$\mathbb{E}\left(\int_0^T f(\xi_s)\mathrm{d}s\right) \leqslant C\|f\|_{L^p(\mathbb{R}^d)}. \tag{1.3}$$

Such estimates were proven by Krylov in [8], which plays a basic role in the study of SDEs with measurable coefficients (see also [19] for some extensions).

In this paper we are interesting in showing a discretized version of (1.2). More precisely, for fixed $N \in \mathbb{N}$, we introduce the following discretized Itô process: for $k \in \mathbb{N}$,

$$\xi_k^N := \xi_0^N + \sum_{j=0}^{k-1} b_j/N + \sum_{j=0}^{k-1} \sigma_j \cdot \left(W_{(j+1)/N} - W_{j/N}\right) \tag{1.4}$$
$$= \xi_{k-1}^N + b_{k-1}/N + \sigma_{k-1} \cdot \left(W_{k/N} - W_{(k-1)/N}\right),$$

where $\xi_0^N \in \mathscr{F}_0$, and for each $j \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $b_j \in \mathbb{R}^d$ and $\sigma_j \in \mathbb{R}^d \otimes \mathbb{R}^d$ are $\mathscr{F}_{j/N}$-measurable random variables. We aim to establish a discretized version of Krylov's estimate for $\xi_k^N$ in the following theorem.

**Theorem 1.1.** *Suppose that for some $\kappa_0, \kappa_1 > 0$ and any $j \in \mathbb{N}_0$,*

$$|b_j|, \|\sigma_j\| \leqslant \kappa_0, \quad \det(\sigma_j \sigma_j^*) \geqslant \kappa_1, \quad a.s.$$

*Then for any $p > d+1$, there is a constant $C = C(p,d,\kappa_0,\kappa_1) > 0$ such that for any $N \in \mathbb{N}$ and $f_k \in L^p(\mathbb{R}^d)$, $k = 1, \cdots, N$,*

$$\frac{1}{N}\sum_{k=1}^N \mathbb{E}f_k(\xi_k^N) \leqslant C\left(\frac{1}{N}\sum_{k=1}^N \|f_k\|_{L^p(\mathbb{R}^d)}^p\right)^{1/p}. \tag{1.5}$$

*Moreover, for any $p > d$, we have*

$$\frac{1}{N}\sum_{k=1}^N \mathbb{E}f(\xi_k^N) \leqslant C\|f\|_{L^p(\mathbb{R}^d)}. \tag{1.6}$$

The motivation of studying the above discretized version of Krylov's estimate comes from the study of Euler's scheme for SDEs with measurable discontinuous coefficients. Let us consider the following general SDE in $\mathbb{R}^d$:

$$\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}W_t, \quad X_0 = x, \tag{1.7}$$

where $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ is a Borel measurable function and $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ is a nondegenerate matrix-valued Borel measurable function and continuous in $x$. If $b$ and $\sigma$ are of linear growth in $x$ uniformly in $t$, it is well known that SDE (1.7) admits a unique weak solution $X_t$ (cf. [16]). Moreover, if in addition $\sigma$ is Lipschitz continuous in $x$ uniformly in $t$, then SDE (1.7) admits a unique strong solution (cf. [18]). For $h \in (0,1)$, consider the following Euler approximation of SDE (1.7):

$$\mathrm{d}X_t^h = b_{t_h}(X_{t_h}^h)\mathrm{d}t + \sigma_{t_h}(X_{t_h}^h)\mathrm{d}W_t, \quad X_0^h = x, \tag{1.8}$$

where $t_h := [t/h]h$, which can be solved recursively as follows:

$$X_t^h = X_{t_h}^h + b_{t_h}(X_{t_h}^h)(t - t_h) + \sigma_{t_h}(X_{t_h}^h)(W_t - W_{t_h}), \tag{1.9}$$

or equivalently, for $k = 0, 1, 2, \cdots$ and $t \in [kh, (k+1)h)$,

$$X_t^h = X_{kh}^h + b_{kh}(X_{kh}^h)(t - kh) + \sigma_{kh}(X_{kh}^h)(W_t - W_{kh}). \tag{1.10}$$

One would ask whether it holds

$$\lim_{h \to 0} \mathbb{E} \left( \sup_{t \in [0,T]} |X_t^h - X_t|^2 \right) = 0, \tag{1.11}$$

where the key point of proving the above limit is to show an estimate like (1.5). Notice that if we take $h = 1/N$, then $\xi_k^N := X_{k/N}^{1/N}$ just takes the same form as in (1.4). In fact, when $\sigma$ is Hölder continuous, that is, for some $\alpha \in (0, 1)$ and $c > 0$,

$$\|\sigma_t(x) - \sigma_t(y)\| \leqslant c|x - y|^\alpha,$$

Gyöngy and Krylov [4, Theorem 4.2] proved that $X_t^h$ allows a density $\rho_t^h(y)$ with

$$\left( \int_{\mathbb{R}^d} |\rho_t^h(y)|^q \mathrm{d}y \right)^{1/q} \leqslant C(t^{-d/(2p)} + 1), \quad p = \frac{q}{q-1} > \frac{d}{\alpha},$$

where $C = C(d, p, \kappa_0, \kappa_1) > 0$, see also [9] for two-sides estimates of $\rho_t^h(x)$. From this, it is easy to derive that for any $p > d/\alpha$,

$$\mathbb{E} \int_0^T f(X_{t_h}^h) \mathrm{d}t \leqslant C\|f\|_{L^p(\mathbb{R}^d)}.$$

The above discretized Krylov estimate plays a key role in [4] for showing (1.11) when $b$ is only bounded measurable. However, by (1.6), the above estimate holds for any $p > d$ without any continuity assumption on $\sigma$. In other words, using (1.6) we can drop the *continuity* assumption on $\sigma$ in Theorem 2.8 of [4]. It should be noticed that in the remarkable paper [4], under very broad assumptions, Gyöngy and Krylov used Euler's polygonal approximation to construct the strong solution for SDE (1.7). We mention that if $b$ satisfies some monotonicity condition and $\sigma$ is Lipschitz continuous, Gyöngy [3] showed the rate of almost surely convergence for Euler's scheme. Up to now, there are many works devoted to the study of Euler's approximation for SDEs with irregular coefficients under various assumptions, for examples, see [5, 10, 13, 1] and references therein.

### 1.2 Euler's scheme for DDSDEs

Another goal of this paper is to use Theorem 1.1 to derive the same results as in [4] for mean-field (also called McKean-Vlasov or distribution-dependent in literature) SDEs with measurable discontinuous coefficients $b$ and $\sigma$. For $\beta \geqslant 0$, let $\mathcal{P}_\beta(\mathbb{R}^d)$ be the space of all probability measures on $\mathbb{R}^d$ with finite $\beta$-order moment, which is endowed with the weak convergence topology. Let $\beta \geqslant 1$. Consider the following distribution-dependent SDE (abbreviated as DDSDE):

$$\mathrm{d}X_t = b_t(X_t, \mu_{X_t})\mathrm{d}t + \sigma_t(X_t, \mu_{X_t})\mathrm{d}W_t, \quad \text{Law of } X_0 = \nu \in \mathcal{P}_\beta(\mathbb{R}^d), \tag{1.12}$$

where $\mu_{X_t}$ stands for the law of random variable $X_t$ and

$$b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}_\beta(\mathbb{R}^d) \to \mathbb{R}^d, \ \sigma : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}_\beta(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^d$$

are Borel measurable functions. Below we make the following assumptions:

**(H$_\beta$)** For each $x, \mu$, $t \mapsto b_t(x, \mu)$ and $\sigma_t(x, \mu)$ are continuous, and for each $t, x$, $\mu \mapsto b_t(x, \mu)$ and $\sigma_t(x, \mu)$ are weakly continuous. Moreover, for some $\beta \geqslant 1$, there is a constant $c_0 > 0$ such that for all $t \geqslant 0$, $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_\beta(\mathbb{R}^d)$,

$$|b_t(x, \mu)| + \|\sigma_t(x, \mu)\| \leqslant c_0(1 + |x| + \mu(|\cdot|^\beta)^{1/\beta}),$$

and the following nondegenerate condition holds: there is a constant $c_1 > 0$ such that for all $t \geqslant 0$, $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_\beta(\mathbb{R}^d)$,

$$\det(\sigma\sigma^*)(t, x, \mu) \geqslant c_1. \tag{1.13}$$

**(H$'_\beta$)** Let $\bar{b}$ and $\bar{\sigma}$ be two Borel measurable functions on $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d$ with values in $\mathbb{R}^d$ and $\mathbb{R}^d \otimes \mathbb{R}^d$, respectively. Assume that for each $x, y \in \mathbb{R}^d$, $t \mapsto \bar{b}_t(x, y), \bar{\sigma}_t(x, y)$ are continuous, and for some $c_0 > 0$ and all $t \geqslant 0$, $x, y \in \mathbb{R}^d$,

$$|\bar{b}_t(x, y)| + \|\bar{\sigma}_t(x, y)\| \leqslant c_0(1 + |x| + |y|).$$

Moreover, for any $\mu \in \mathcal{P}_\beta(\mathbb{R}^d)$, define

$$b_t(x, \mu) := \int_{\mathbb{R}^d} \bar{b}_t(x, y)\mu(\mathrm{d}y), \quad \sigma_t(x, \mu) := \int_{\mathbb{R}^d} \bar{\sigma}_t(x, y)\mu(\mathrm{d}y),$$

and we also assume the nondegenerate condition (1.13) holds.

The difference between **(H$_\beta$)** and **(H$'_\beta$)** lies in that in the later case,

$$\mu \mapsto b_t(x, \mu), \sigma_t(x, \mu)$$

may be not continuous with respect to the weak convergence. Notice that we do not make any continuity assumptions on $\bar{b}, \bar{\sigma}$ in $x, y$. We now consider the following Euler approximation of DDSDE (1.12):

$$\mathrm{d}X_t^h = b_{t_h}(X_{t_h}^h, \mu_{X_{t_h}^h})\mathrm{d}t + \sigma_{t_h}(X_{t_h}^h, \mu_{X_{t_h}^h})\mathrm{d}W_t, \quad \text{Law of } X_0^h = \nu. \tag{1.14}$$

The following theorem extends [4, Theorem 2.8] to DDSDEs.

**Theorem 1.2.** *Let $\beta > 2$, $\nu \in \mathcal{P}_\beta(\mathbb{R}^d)$ and one of **(H$_\beta$)** and **(H$'_\beta$)** holds.*

(i) *Suppose that weak uniqueness holds for DDSDE* (1.12). *Then there is a unique weak solution $X$ to DDSDE* (1.12) *with initial law $\mathbb{P} \circ X_0^{-1} = \nu$ so that $X^h$ converges to $X$ in distribution. Moreover, for any bounded measurable $f$,*

$$\lim_{h \to 0} \mathbb{E}\left(\int_0^T f(X_{t_h}^h)\mathrm{d}t\right) = \mathbb{E}\left(\int_0^T f(X_t)\mathrm{d}t\right). \tag{1.15}$$

(ii) *Suppose that pathwise uniqueness holds for DDSDE* (1.12). *Then there is a unique strong solution $X$ to DDSDE* (1.12) *with initial law $\mathbb{P} \circ X_0^{-1} = \nu$ so that*

$$\lim_{h \to 0} \mathbb{E}\left(\sup_{t \in [0,T]} |X_t^h - X_t|^2\right) = 0. \tag{1.16}$$

About the weak and strong uniqueness of DDSDE (1.12), by Girsanov's theorem, Li and Min [11] obtained the existence and uniqueness of weak solutions when $b$ is bounded measurable and $\sigma$ is nondegenerate and Lipschitz continuous. While under **(H$_\beta$)** or **(H$'_\beta$)**, when $\sigma$ does not depend on $\mu$ and is Lipschitz continuous in $x$ and $b$ is Lipschitz continuous with respect to $\mu$ in case **(H$_\beta$)**, Mishura and Veretennikov [12] showed the strong uniqueness. In a recent work of the present author with Röckner [14], we established the well-posedness of DDSDEs (1.12) with singular drifts (see also [6]).

## 1.3 Propagation of chaos for Euler's scheme

Below we fix $h \in (0,1)$ and let $\{\xi_j, j \in \mathbb{N}\}$ be a sequence of i.i.d. random variables in $\mathbb{R}^d$ with common distribution $\nu$, and $\{W^j, j \in \mathbb{N}\}$ a sequence of independent $d$-dimensional standard Brownian motions. For numerical reason, we also consider the following interacting particle approximation for Euler's scheme: for fixed $N \in \mathbb{N}$, we define for $j = 1, \cdots, N$,

$$\mathrm{d}X_t^{N,j} = b_{t_h}\big(X_{t_h}^{N,j}, \mu_{t_h}^N\big)\mathrm{d}t + \sigma_{t_h}\big(X_{t_h}^{N,j}, \mu_{t_h}^N\big)\mathrm{d}W_t^j, \ X_0^{N,j} = \xi_j, \tag{1.17}$$

where $\mu_t^N$ is the empirical measure of $\{X_t^{N,j}, j = 1, \cdots, N\}$ defined by

$$\mu_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{N,i}},$$

where $\delta_x$ stands for the Dirac measure concentrated at point $x$. In the following, for simplicity we only consider the case $(\mathbf{H}'_\beta)$, and in this case we have

$$b_{t_h}\big(X_{t_h}^{N,j}, \mu_{t_h}^N\big) = \frac{1}{N}\sum_{i=1}^N \bar{b}_{t_h}\big(X_{t_h}^{N,j}, X_{t_h}^{N,i}\big)$$

and

$$\sigma_{t_h}\big(X_{t_h}^{N,j}, \mu_{t_h}^N\big) = \frac{1}{N}\sum_{i=1}^N \bar{\sigma}_{t_h}\big(X_{t_h}^{N,j}, X_{t_h}^{N,i}\big).$$

For $j \in \mathbb{N}$, let $\bar{X}_t^j$ be the unique solution of the following Euler scheme:

$$\mathrm{d}\bar{X}_t^j = b_{t_h}\left(\bar{X}_{t_h}^j, \mu_{\bar{X}_{t_h}^j}\right)\mathrm{d}t + \sigma_{t_h}\left(\bar{X}_{t_h}^j, \mu_{\bar{X}_{t_h}^j}\right)\mathrm{d}W_t^j, \ \ \bar{X}_0^j = \xi_j. \tag{1.18}$$

Clearly, $\{\bar{X}_\cdot^j, j \in \mathbb{N}\}$ is a family of i.i.d. stochastic processes with common distribution as $X_\cdot^h$.

**Theorem 1.3.** *Let $\beta > 2$ amd $\nu \in \mathcal{P}_\beta(\mathbb{R}^d)$. Suppose that $(\mathbf{H}'_\beta)$ holds and the initial law $\nu$ has a density $\phi \in L_{loc}^q(\mathbb{R}^d)$ for some $q > 1$. Then it holds that for any $T > 0$,*

$$\lim_{N \to \infty} \sup_{j=1,\cdots,N} \mathbb{E}\left(\sup_{t \in [0,T]} |X_t^{N,j} - \bar{X}_t^j|^2\right) = 0. \tag{1.19}$$

For fixed $h \in (0,1)$ and $N \in \mathbb{N}$, we use $\mathscr{E}_h$ and $\mathscr{P}_N$ to denote the operators of Euler's scheme and the interacting particle approximation to DDSDE (1.12), respectively:

$$\mathscr{E}_h : X \to X^h, \ \ \mathscr{P}_N : X \to (X^{N,j})_{j=1,\cdots,N}.$$

From the construction, it is easy to see that

$$\mathscr{E}_h\mathscr{P}_N = \mathscr{P}_N\mathscr{E}_h.$$

Under $(\mathbf{H}'_\beta)$, suppose that the pathwise uniqueness holds for DDSDE (1.12). Then by Theorems 1.2 and 1.3, we have

$$\lim_{h \to 0} \lim_{N \to \infty} \mathbb{E}\|\mathscr{P}_N\mathscr{E}_h X - X\|_{C([0,T])}^2 = 0.$$

Here an open question is to show that

$$\lim_{N \to \infty} \lim_{h \to 0} \mathbb{E}\|\mathscr{P}_N\mathscr{E}_h X - X\|_{C([0,T])}^2 = 0 \tag{1.20}$$

and

$$\lim_{N\to\infty} \mathbb{E}\|\mathscr{P}_N \mathscr{E}_{1/N} X - X\|^2_{C([0,T])} = 0.$$

Obviously, the obstacle is to show the following propagation of chaos under $(\mathbf{H}'_\beta)$:

$$\lim_{N\to\infty} \mathbb{E}\|\mathscr{P}_N X - X\|^2_{C([0,T])} = 0. \tag{1.21}$$

When $b$ and $\sigma$ are Lipschitz continuous in $x$ and $\mu$, the above propagation of chaos (1.21) was proven by Sznitman [17]. Recently, Bao and Huang [2] proved (1.20) by Zvonkin's transformation when $b$ and $\sigma$ are Hölder continuous in $x$ and Lipschitz continuous in $\mu$ with respect to the Wasserstein distance. However, under $(\mathbf{H}'_\beta)$, proving (1.21) seems to be a challenge problem.

### 1.4 Plan and notations

This paper is organized as follows: In Section 2, we prove Theorem 1.1. In Section 3, we prove Theorem 1.2. In Section 4, we prove Theorem 1.3. Throughout this paper we use the following conventions:

- For a matrix $\sigma$, we use $\|\sigma\|$ to denote the Hilbert-Schmidt norm of $\sigma$.
- For $R > 0$, we use $B_R$ to denote the ball in $\mathbb{R}^d$ with radius $R$ and center $0$.
- We use $A \lesssim B$ (resp. $\asymp$) to denote $A \leqslant CB$ (resp. $C^{-1}B \leqslant A \leqslant CB$) for some unimportant constant $C \geqslant 1$, whose dependence on the parameters can be traced from the context.

## 2 Proof of Theorem 1.1

To prove (1.5), we shall use the classical Krylov estimate (1.2). For this we need to embed $\xi_k^N$ into a continuous Itô process. For $k \in \mathbb{N}_0$ and $t \in [k/N, (k+1)/N)$, we define

$$\tilde{b}_t^N := b_k, \quad \tilde{\sigma}_t^N := \sigma_k$$

and

$$X_t^N := \xi_k^N + \tilde{b}_t^N \cdot (t - k/N) + \tilde{\sigma}_t^N \cdot (W_t - W_{k/N}).$$

In this way, it is easy to see that $X_{k/N}^N = \xi_k^N$ and

$$X_t^N = \xi_0^N + \int_0^t \tilde{b}_s^N \mathrm{d}s + \int_0^t \tilde{\sigma}_s^N \mathrm{d}W_s, \quad t \geqslant 0.$$

Similarly, let $(f_k)_{k\in\mathbb{N}}$ be a family of nonnegative measurable functions in $\mathbb{R}^d$. If we define

$$\tilde{f}_N(t,x) := \sum_{k=0}^{\infty} f_{k+1}(x)\mathbf{1}_{t\in[k/N,(k+1)/N)}, \quad t \geqslant 0, x \in \mathbb{R}^d,$$

then we can write

$$\frac{1}{N}\sum_{k=1}^{N} \mathbb{E}f_k(\xi_k^N) = \int_0^1 \mathbb{E}\tilde{f}_N\big(t, X_{t_+^N}^N\big)\mathrm{d}t,$$

where $t_+^N := ([tN]+1)/N$. Moreover, by (1.2) we have for any $p \geqslant d+1$ and $y \in \mathbb{R}^d$,

$$\int_0^1 \mathbb{E}\left(\tilde{f}_N(t, X_t^N + y)\right)\mathrm{d}t \leqslant C\|\tilde{f}_N\|_{L^p([0,1]\times\mathbb{R}^d)} = C\left(\frac{1}{N}\sum_{k=1}^{N}\|f_k\|_p^p\right)^{1/p}. \tag{2.1}$$

At this moment, we can not immediately conclude (1.5) because we need to treat $\mathbb{E}\tilde{f}_N\big(t, X^N_{t^k_+}\big)$ rather than $\mathbb{E}\tilde{f}_N(t, X^N_t)$. Notice that for $t \in [(k-1)/N, k/N)$,

$$\xi^N_k = X^N_t + b_{k-1} \cdot \ell^k_t + \sigma_{k-1} \cdot \big(W_{k/N} - W_t\big),$$

where

$$\ell^k_t := k/N - t.$$

Since $W_{k/N} - W_t$ is independent with $X^N_t$ and $b_{k-1}, \sigma_{k-1}$, by the change of variable we have

$$\frac{1}{N}\mathbb{E}f_k(\xi^N_k) = \mathbb{E}\int^{k/N}_{(k-1)/N} f_k\big(X^N_t + b_{k-1}\ell^k_t + \sigma_{k-1}\cdot(W_{k/N} - W_t)\big)\mathrm{d}t$$

$$= \int^{k/N}_{(k-1)/N}\int_{\mathbb{R}^d} \mathbb{E}f_k\big(X^N_t + b_{k-1}\ell^k_t + \sigma_{k-1}y\big)\varphi_{\ell^k_t}(y)\mathrm{d}y\mathrm{d}t$$

$$= \int^{k/N}_{(k-1)/N} \mathbb{E}\int_{\mathbb{R}^d} f_k\big(X^N_t + y\big)\varphi_{\ell^k_t}\big(\sigma^{-1}_{k-1}(y - b_{k-1}\ell^k_t)\big)|\det\sigma^{-1}_{k-1}|\mathrm{d}y\mathrm{d}t,$$

where $\varphi_t(y) := (2\pi t)^{-d/2}\mathrm{e}^{-\frac{|y|^2}{2t}}$ is the distributional density of Brownian motion $W_t$. Noting that

$$|\det\sigma^{-1}_{k-1}| = 1/|\det\sigma_{k-1}| \leqslant 1/\sqrt{\kappa_1},$$

and

$$|\sigma^{-1}_{k-1}(y - b_{k-1}\ell^k_t)|^2 \geqslant c_0|y - b_{k-1}\ell^k_t|^2 \geqslant \frac{c_0}{2}\big(|y|^2 - 2|b_{k-1}|^2|\ell^k_t|^2\big),$$

we have for $\lambda = \sqrt{c_0/2}$ and $t \in [(k-1)/N, k/N)$,

$$\varphi_{\ell^k_t}(\sigma^{-1}_{k-1}(y - b_{k-1}\ell^k_s)) \leqslant \mathrm{e}^{c_0\kappa^2_0|\ell^k_t|/2}\varphi_{\ell^k_t}(\lambda y) \leqslant \mathrm{e}^{c_0\kappa^2_0/(2N)}\varphi_{\ell^k_t}(\lambda y).$$

Hence, for $\gamma = \frac{p}{d+1}$ and $q = \frac{\gamma}{\gamma-1}$, by Hölder's inequality we further have

$$\frac{1}{N}\sum^N_{k=1}\mathbb{E}f_k(\xi^N_k) \lesssim \int_{\mathbb{R}^d}\left(\sum^N_{k=1}\int^{k/N}_{(k-1)/N}\mathbb{E}f_k\big(X^N_t + y\big)\varphi_{\ell^k_t}(\lambda y)\mathrm{d}t\right)\mathrm{d}y$$

$$\leqslant \int_{\mathbb{R}^d}\left(\sum^N_{k=1}\int^{k/N}_{(k-1)/N}\mathbb{E}f_k\big(X^N_t + y\big)^\gamma\mathrm{d}t\right)^{\frac{1}{\gamma}}\left(\sum^N_{k=1}\int^{k/N}_{(k-1)/N}|\varphi_{\ell^k_t}(\lambda y)|^q\mathrm{d}t\right)^{\frac{1}{q}}\mathrm{d}y$$

$$= \int_{\mathbb{R}^d}\left(\mathbb{E}\int^1_0\tilde{f}_N(t, X^N_t + y)^\gamma\mathrm{d}t\right)^{\frac{1}{\gamma}}\left(N\int^{1/N}_0|\varphi_t(\lambda y)|^q\mathrm{d}t\right)^{\frac{1}{q}}\mathrm{d}y.$$

By Krylov's estimate (2.1), we obtain

$$\frac{1}{N}\sum^N_{k=1}\mathbb{E}f_k(\xi^N_k) \lesssim \|\tilde{f}_N\|_{L^{\gamma(d+1)}([0,1]\times\mathbb{R}^d)}\int_{\mathbb{R}^d}\left(N\int^{1/N}_0|\varphi_t(\lambda y)|^q\mathrm{d}t\right)^{\frac{1}{q}}\mathrm{d}y. \qquad (2.2)$$

Note that by the change of variable and the scaling property of $\varphi_t(y)$,

$$\int_{\mathbb{R}^d}\left(N\int^{1/N}_0\varphi_t(\lambda y)^q\mathrm{d}t\right)^{\frac{1}{q}}\mathrm{d}y = \int_{\mathbb{R}^d}\left(\int^1_0\varphi_{t/N}(\lambda y)^q\mathrm{d}t\right)^{\frac{1}{q}}\mathrm{d}y$$

$$= N^{d/2}\int_{\mathbb{R}^d}\left(\int^1_0\varphi_t(\lambda y\sqrt{N})^q\mathrm{d}t\right)^{\frac{1}{q}}\mathrm{d}y = \lambda^{-d}\int_{\mathbb{R}^d}\left(\int^1_0\varphi_t(y)^q\mathrm{d}t\right)^{\frac{1}{q}}\mathrm{d}y.$$

The desired estimate (1.5) now follows by (2.2) and showing that the last integral is finite. In fact, by the change of variable, for some $c = c(d, q) > 0$,

$$\int_{\mathbb{R}^d} \left( \int_0^1 \varphi_t(y)^q \mathrm{d}t \right)^{\frac{1}{q}} \mathrm{d}y = c \int_{\mathbb{R}^d} |y|^{\frac{2}{q}-d} \left( \int_{|y|^2}^\infty t^{\frac{dq}{2}-2} \mathrm{e}^{-t} \mathrm{d}t \right)^{\frac{1}{q}} \mathrm{d}y$$

$$\lesssim \int_{B_1} |y|^{\frac{2}{q}-d} \left( \int_0^\infty t^{\frac{dq}{2}-2} \mathrm{e}^{-t} \mathrm{d}t \right)^{\frac{1}{q}} \mathrm{d}y + \int_{B_1^c} |y|^{\frac{2}{q}-d} \left( \int_{|y|^2}^\infty t^{\frac{dq}{2}-2} \mathrm{e}^{-t} \mathrm{d}t \right)^{\frac{1}{q}} \mathrm{d}y < \infty.$$

As for (1.6) it follows by using (1.3) in the above proof.

**Remark 2.1.** When $b_j$ and $\sigma_j$ are nonrandom, the estimate (1.5) is trivial because $\xi_k^N$, $k = 1, \cdots, N$ are Gaussian random variables. However, in the general case, we only know that $\xi_k^N$ is a nondegenerate semimartingale with respect to $\mathscr{F}_{k/N}$. Here it is quite interesting to give a purely probabilistic proof for Theorem 1.1. It should be noticed that (1.2) can be derived from (1.5) by discretized approximation.

## 3 Proof of Theorem 1.2

The following lemma is standard by Burkholder and Gronwall's inequalities.

**Lemma 3.1.** Let $\beta > 2$. Under **($H_\beta$)** or **($H_\beta'$)**, for any $T > 0$, there is a constant $C > 0$ such that for all $h \in (0,1)$,

$$\mathbb{E}\left( \sup_{t \in [0,T]} |X_t^h|^\beta \right) \leqslant C(1 + \mathbb{E}|X_0|^\beta), \tag{3.1}$$

and for any $s, t \in [0, T]$,

$$\mathbb{E}|X_s^h - X_t^h|^\beta \leqslant C|s - t|^{\beta/2}. \tag{3.2}$$

*Proof.* Note that

$$X_t^h = X_0 + \int_0^t b_{s_h}\left(X_{s_h}^h, \mu_{X_{s_h}^h}\right)\mathrm{d}s + \int_0^t \sigma_{s_h}\left(X_{s_h}^h, \mu_{X_{s_h}^h}\right)\mathrm{d}W_s. \tag{3.3}$$

For simplicity, we let $|X_t^h|_* := \sup_{s \in [0,t]} |X_s^h|$. By Burkholder's inequality and the linear growth of $b$ and $\sigma$, we have

$$\mathbb{E}|X_t^h|_*^\beta \lesssim \mathbb{E}|X_0|^\beta + \int_0^t \left( 1 + \mathbb{E}|X_{s_h}^h|^\beta + \mu_{X_{s_h}^h}(|\cdot|^\beta) \right)\mathrm{d}s$$

$$\lesssim \mathbb{E}|X_0|^\beta + \int_0^t \left( 1 + \mathbb{E}|X_s^h|_*^\beta \right)\mathrm{d}s,$$

which implies (3.1) by Gronwall's inequality. As for (3.2), it follows by (3.3) and (3.1). □

Let $\mathbb{Q}_h$ be the law of $(X_\cdot^h, W_\cdot)$ in product space $\mathbb{C} \times \mathbb{C}$, where $\mathbb{C}$ is the continuous functions space. By (3.2), since $\beta > 2$, $(\mathbb{Q}_h)_{h \in (0,1)}$ is tight. Therefore, by Prokhorov's theorem, there are a subsequence $h_n \to 0$ as $n \to \infty$ and $\mathbb{Q} \in \mathcal{P}(\mathbb{C} \times \mathbb{C})$ so that

$$\mathbb{Q}_n := \mathbb{Q}_{h_n} \to \mathbb{Q} \text{ weakly.}$$

Now, by Skorokhod's representation theorem, there are a probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$ and random variables $(\tilde{X}^n, \tilde{W}^n)$ and $(\tilde{X}, \tilde{W})$ defined on it such that

$$(\tilde{X}^n, \tilde{W}^n) \to (\tilde{X}, \tilde{W}), \quad \tilde{\mathbb{P}} - a.s. \tag{3.4}$$

and

$$\tilde{\mathbb{P}} \circ (\tilde{X}^n, \tilde{W}^n)^{-1} = \mathbb{Q}_n = \mathbb{P} \circ (X^{h_n}, W)^{-1}, \quad \tilde{\mathbb{P}} \circ (\tilde{X}, \tilde{W})^{-1} = \mathbb{Q}. \tag{3.5}$$

Define $\tilde{\mathscr{F}}_t^n := \sigma(\tilde{W}_s^n, \tilde{X}_s^n; s \leqslant t)$. Notice that

$$\mathbb{P}(W_t - W_s \in \cdot|\mathscr{F}_s) = \mathbb{P}(W_t - W_s \in \cdot) \Rightarrow \tilde{\mathbb{P}}(\tilde{W}_t^n - \tilde{W}_s^n \in \cdot|\tilde{\mathscr{F}}_s^n) = \tilde{\mathbb{P}}(\tilde{W}_t^n - \tilde{W}_s^n \in \cdot).$$

In other words, $\tilde{W}_t^n$ is an $\tilde{\mathscr{F}}_t^n$-Brownian motion. Thus, by (3.3) and (3.5) we have

$$\tilde{X}_t^n = \tilde{X}_0^n + \int_0^t b_{s_n}\big(\tilde{X}_{s_n}^n, \mu_{\tilde{X}_{s_n}^n}\big)\mathrm{d}s + \int_0^t \sigma_{s_n}\big(\tilde{X}_{s_n}^n, \mu_{\tilde{X}_{s_n}^n}\big)\mathrm{d}\tilde{W}_s^n, \tag{3.6}$$

where $s_n := s_{h_n} = [s/h_n]h_n$.

To take the limits, we recall a result of Skorokhod [15, p.32].

**Lemma 3.2.** *Let $\{f_n(t), t \geqslant 0, n \in \mathbb{N}\}$ be a sequence of measurable $\tilde{\mathscr{F}}_t^n$-adapted processes. Suppose that*

*(i) For every $T, \varepsilon > 0$, there is an $M_\varepsilon > 0$ such that for all $n$,*

$$\tilde{\mathbb{P}}\left\{\sup_{t \in [0,T]} |f_n(t)| > M_\varepsilon\right\} \leqslant \varepsilon.$$

*(ii) For each $t$, $f_n(t) \to f(t)$ in probability as $n \to \infty$, and for every $T, \varepsilon > 0$,*

$$\lim_{\delta \to 0} \lim_{n \to \infty} \sup_{|t-s| \leqslant \delta, s, t \in [0,T]} \tilde{\mathbb{P}}(|f_n(t) - f_n(s)| > \varepsilon) = 0,$$

*or for every $T, \varepsilon > 0$,*

$$\lim_{n \to \infty} \tilde{\mathbb{P}}\left\{\sup_{t \in [0,T]} |f_n(t) - f(t)| > \varepsilon\right\} = 0.$$

*Then it holds that for every $T > 0$,*

$$\int_0^T f_n(t)\mathrm{d}\tilde{W}_t^n \overset{n \to \infty}{\Rightarrow} \int_0^T f(t)\mathrm{d}\tilde{W}_t, \text{ in probability.}$$

Using the above lemma we can show the following limits by the discretized Krylov estimate.

**Lemma 3.3.** *Under ($\mathbf{H}_\beta$) or ($\mathbf{H}_\beta'$), for each $t > 0$, the following limits hold*

$$\int_0^t b_{s_n}\left(\tilde{X}_{s_n}^n, \mu_{\tilde{X}_{s_n}^n}\right)\mathrm{d}s \to \int_0^t b\left(s, \tilde{X}_s, \mu_{\tilde{X}_s}\right)\mathrm{d}s, \tag{3.7}$$

$$\int_0^t \sigma_{s_n}\left(\tilde{X}_{s_n}^n, \mu_{\tilde{X}_{s_n}^n}\right)\mathrm{d}\tilde{W}_s^n \to \int_0^t \sigma\left(s, \tilde{X}_s, \mu_{\tilde{X}_s}\right)\mathrm{d}\tilde{W}_s \tag{3.8}$$

*in probability as $n \to 0$,*

*Proof.* We only prove (3.8) in case ($\mathbf{H}_\beta'$). The others are similar and easier. Below for simplicity we shall drop the tilde. In case ($\mathbf{H}_\beta'$), define

$$\bar{\sigma}_t^\varepsilon(x, y) := \bar{\sigma}_t(\cdot, \cdot) * \varrho_\varepsilon(x, y), \quad \sigma_t^\varepsilon(x, \mu) := \int_{\mathbb{R}^d} \bar{\sigma}_t^\varepsilon(x, y)\mu(\mathrm{d}y),$$

where $(\varrho_\varepsilon)_{\varepsilon\in(0,1)}$ is a family of mollifiers in $\mathbb{R}^d\times\mathbb{R}^d$ with support in $B_\varepsilon\times B_\varepsilon$. For fixed $\varepsilon\in(0,1)$, since $\bar\sigma_\varepsilon$ is continuous and linear growth in $x,y$, by (3.4) and Lemma 3.2 (see also [4, Lemma 3.1]), it is easy to see that for fixed $\varepsilon\in(0,1)$,

$$\int_0^t \sigma_{s_n}^\varepsilon\left(X_{s_n}^n,\mu_{X_{s_n}^n}\right)\mathrm{d}W_s^n \to \int_0^t \sigma_s^\varepsilon\left(X_s,\mu_{X_s}\right)\mathrm{d}W_s$$

in probability as $n\to\infty$. Indeed, it suffices to prove the following two limits:

$$\int_0^t \sigma_{s_n}^\varepsilon\left(X_{s_n}^n,\mu_{X_{s_n}^n}\right)\mathrm{d}W_s^n \to \int_0^t \sigma_s^\varepsilon\left(X_s^n,\mu_{X_s^n}\right)\mathrm{d}W_s^n, \tag{3.9}$$

$$\int_0^t \sigma_s^\varepsilon\left(X_s^n,\mu_{X_s^n}\right)\mathrm{d}W_s^n \to \int_0^t \sigma_s^\varepsilon\left(X_s,\mu_{X_s}\right)\mathrm{d}W_s \tag{3.10}$$

in probability as $n\to\infty$. Limit (3.9) follows by (3.2) and the continuity of $t\mapsto\sigma_t^\varepsilon(x,y)$, and limit (3.10) follows by (3.4) and Lemma 3.2. Therefore, it remains to prove that

$$\int_0^t \sigma_{s_n}^\varepsilon\left(X_{s_n}^n,\mu_{X_{s_n}^n}\right)\mathrm{d}W_s^n \to \int_0^t \sigma_{s_n}\left(X_{s_n}^n,\mu_{X_{s_n}^n}\right)\mathrm{d}W_s^n \tag{3.11}$$

in probability uniformly in $n$ as $\varepsilon\to0$, and

$$\int_0^t \sigma_s^\varepsilon\left(X_s,\mu_{X_s}\right)\mathrm{d}W_s \to \int_0^t \sigma_s\left(X_s,\mu_{X_s}\right)\mathrm{d}W_s \text{ in probability as } \varepsilon\to0. \tag{3.12}$$

We only show (3.11). By Itô's isometric formula, we have

$$\mathbb{E}\left|\int_0^t\left[\sigma_{s_n}^\varepsilon\left(X_{s_n}^n,\mu_{X_{s_n}^n}\right)-\sigma_{s_n}\left(X_{s_n}^n,\mu_{X_{s_n}^n}\right)\right]\mathrm{d}W_s^n\right|^2$$

$$=\int_0^t\mathbb{E}\left\|\sigma_{s_n}^\varepsilon\left(X_{s_n}^n,\mu_{X_{s_n}^n}\right)-\sigma_{s_n}\left(X_{s_n}^n,\mu_{X_{s_n}^n}\right)\right\|^2\mathrm{d}s$$

$$\leqslant\int_0^t\mathbb{E}\left\|\bar\sigma_{s_n}^\varepsilon\left(X_{s_n}^n,\bar X_{s_n}^n\right)-\bar\sigma_{s_n}\left(X_{s_n}^n,\bar X_{s_n}^n\right)\right\|^2\mathrm{d}s=:J_\varepsilon^n(t),$$

where $\bar X^n$ is an independent copy of $X^n$. More precisely, $(X^n,\bar X^n)$ solves the following equation (the Euler scheme):

$$\begin{cases}\mathrm{d}X_t^n = b_{t_n}\left(X_{t_n}^n,\mu_{X_{t_n}^n}\right)\mathrm{d}t + \sigma_{t_n}\left(X_{t_n}^n,\mu_{X_{t_n}^n}\right)\mathrm{d}W_t,\\ \mathrm{d}\bar X_t^n = b_{t_n}\left(\bar X_{t_n}^n,\mu_{\bar X_{t_n}^n}\right)\mathrm{d}t + \sigma_{t_n}\left(\bar X_{t_n}^n,\mu_{\bar X_{t_n}^n}\right)\mathrm{d}\bar W_t,\end{cases} \tag{3.13}$$

where $(W,X_0^n)$ and $(\bar W,\bar X_0^n)$ are independent and have the same distributions. In order to use the discretized Krylov estimate to show

$$\lim_{\varepsilon\to0}\sup_n J_\varepsilon^n(t)=0, \tag{3.14}$$

we use a standard stopping time technique. For $R>0$, we define a stopping time

$$\tau_R^n := \inf\{t>0: |X_t^n|\vee|\bar X_t^n|>R\},$$

and make the following decomposition:

$$J_\varepsilon^n(t)=\int_0^t\mathbb{E}\left(\mathbf{1}_{t\geqslant\tau_R^n}\|\bar\sigma_{s_n}(X_{s_n}^n,\bar X_{s_n}^n)-\bar\sigma_{s_n}^\varepsilon(X_{s_n}^n,\bar X_{s_n}^n)\|^2\right)\mathrm{d}s$$

$$+\int_0^t\mathbb{E}\left(\mathbf{1}_{t<\tau_R^n}\|\bar\sigma_{s_n}(X_{s_n}^n,\bar X_{s_n}^n)-\bar\sigma_{s_n}^\varepsilon(X_{s_n}^n,\bar X_{s_n}^n)\|^2\right)\mathrm{d}s=:J_{R,\varepsilon}^{n,1}(t)+J_{R,\varepsilon}^{n,2}(t).$$

For $J_{R,\varepsilon}^{n,1}(t)$, by Hölder's inequality, (3.1) and Chebyshev's inequality we have

$$J_{R,\varepsilon}^{n,1}(t) \leqslant \mathbb{P}(t \geqslant \tau_R^n)^{\frac{\beta-2}{\beta}} \left( \int_0^t \mathbb{E}\|\bar{\sigma}_{s_n}(X_{s_n}^n, \bar{X}_{s_n}^n) - \bar{\sigma}_{s_n}^\varepsilon(X_{s_n}^n, \bar{X}_{s_n}^n)\|^\beta \mathrm{d}s \right)^{\frac{2}{\beta}}$$

$$\lesssim \left( \frac{\mathbb{E}\big(\sup_{t\in[0,T]}|X_t^n| \vee |\bar{X}_t^n|^\beta\big)}{R^\beta} \right)^{\frac{\beta-2}{\beta}} \left( \int_0^t \big(1 + \mathbb{E}|X_{s_n}^n|^\beta + \mathbb{E}|\bar{X}_{s_n}^n|^\beta\big) \mathrm{d}s \right)^{\frac{2}{\beta}}$$

$$\leqslant C/R^{\beta-2} \to 0 \text{ uniformly in } n, \varepsilon \text{ as } R \to \infty. \tag{3.15}$$

For $J_{R,\varepsilon}^{n,2}(t)$, we can not directly use the discretized Krylov estimate to conclude

$$\lim_{\varepsilon\to 0} \sup_n J_{R,\varepsilon}^{n,2}(t) = 0, \quad \forall t, R > 0, \tag{3.16}$$

because the Euler scheme (3.13) has unbounded coefficients. We need to cutoff the coefficients. Let $\chi_R(x)$ be a nonnegative smooth cutoff function with $\chi_R(x) = 1$ for $|x| < R$ and $\chi_R(x) = 0$ for $|x| > R + 1$. Define

$$b_t^{n,R}(x) := b_t(x, \mu_{X_{t_n}^n})\chi_R(x), \ \sigma_t^{n,R}(x) := \sigma_t(x\chi_R(x), \mu_{X_{t_n}^n}).$$

Let $(X^{n,R}, \bar{X}^{n,R})$ solve the following equation in $\mathbb{R}^{2d}$ (no coupling):

$$\begin{cases} \mathrm{d}X_t^{n,R} = b_{t_n}^{n,R}(X_{t_n}^{n,R})\mathrm{d}t + \sigma_{t_n}^{n,R}(X_{t_n}^{n,R})\mathrm{d}W_t, \ X_0^{n,R} = X_0^n, \\ \mathrm{d}\bar{X}_t^{n,R} = b_{t_n}^{n,R}(\bar{X}_{t_n}^{n,R})\mathrm{d}t + \sigma_{t_n}^{n,R}(\bar{X}_{t_n}^{n,R})\mathrm{d}\bar{W}_t, \ \bar{X}_0^{n,R} = \bar{X}_0^n, \end{cases}$$

where $(W, X_0^n)$ and $(\bar{W}, \bar{X}_0^n)$ are the same as in (3.13). From the construction, one sees that

$$(X_t^n, \bar{X}_t^n) = (X_t^{n,R}, \bar{X}_t^{n,R}), \ \ t < \tau_R^n. \tag{3.17}$$

Moreover, it is easy to see that $(X^{n,R}, \bar{X}^{n,R})$ is a discretized $\mathbb{R}^{2d}$-valued Itô process with coefficients satisfying the assumptions in Theorem 1.1 uniformly in $n$. Thus, for fixed $R > 0$ and any $p > 2d + 1$, by (3.17) and (1.5) we have

$$J_{R,\varepsilon}^{n,2}(t) = \int_0^t \mathbb{E}\Big(\mathbf{1}_{t<\tau_R^n}\|\bar{\sigma}_{s_n}(X_{s_n}^{n,R}, \bar{X}_{s_n}^{n,R}) - \bar{\sigma}_{s_n}^\varepsilon(X_{s_n}^{n,R}, \bar{X}_{s_n}^{n,R})\|^2\Big)\mathrm{d}s$$

$$\leqslant \int_0^t \mathbb{E}\Big(\mathbf{1}_{|X_{s_n}^{n,R}|\vee|\bar{X}_{s_n}^{n,R}|<R}\|\bar{\sigma}_{s_n}(X_{s_n}^{n,R}, \bar{X}_{s_n}^{n,R}) - \bar{\sigma}_{s_n}^\varepsilon(X_{s_n}^{n,R}, \bar{X}_{s_n}^{n,R})\|^2\Big)\mathrm{d}s$$

$$= \int_0^{h_n} \mathbb{E}\Big(\mathbf{1}_{|X_0^n|\vee|\bar{X}_0^n|<R}\|\bar{\sigma}_0(X_0^n, \bar{X}_0^n) - \bar{\sigma}_0^\varepsilon(X_0^n, \bar{X}_0^n)\|^2\Big)\mathrm{d}s$$

$$+ \int_{h_n}^t \mathbb{E}\Big(\mathbf{1}_{|X_{s_n}^{n,R}|\vee|\bar{X}_{s_n}^{n,R}|<R}\|\bar{\sigma}_{s_n}(X_{s_n}^{n,R}, \bar{X}_{s_n}^{n,R}) - \bar{\sigma}_{s_n}^\varepsilon(X_{s_n}^{n,R}, \bar{X}_{s_n}^{n,R})\|^2\Big)\mathrm{d}s$$

$$\lesssim h_n + \left( \int_0^t \|\bar{\sigma}_{s_n}(\cdot, \cdot) - \bar{\sigma}_{s_n}^\varepsilon(\cdot, \cdot)\|_{L^{2p}(B_R \times B_R)}^{2p} \mathrm{d}s \right)^{1/p}$$

$$\lesssim h_n + \left( \int_0^t \|\bar{\sigma}_s(\cdot, \cdot) - \bar{\sigma}_s^\varepsilon(\cdot, \cdot)\|_{L^{2p}(B_R \times B_R)}^{2p} \mathrm{d}s \right)^{1/p}$$

$$+ \left( \int_0^t \|\bar{\sigma}_s(\cdot, \cdot) - \bar{\sigma}_{s_n}(\cdot, \cdot)\|_{L^{2p}(B_{R+1} \times B_{R+1})}^{2p} \mathrm{d}s \right)^{1/p}$$

$$=: h_n + I_\varepsilon^R(t) + K_n^R(t),$$

A discretized version of Krylov's estimate and its applications

where $h_n \downarrow 0$ as $n \to \infty$, and the constants contained in the above $\lesssim$ may depend on $R$. By the dominated convergence theorem and the continuity of $t \mapsto \sigma_t(x,y)$, we have

$$\lim_{\varepsilon \to \infty} I_\varepsilon^R(t) = 0, \quad \lim_{n \to \infty} K_n^R(t) = 0,$$

which in turn implies the limit (3.16), and so (3.14). Thus we complete the proof. □

*Proof of (i) of Theorem 1.2.* Using the above lemma and taking limits for both sides of (3.6), one finds that $(\tilde{X}, \tilde{W})$ solves the following SDE:

$$\tilde{X}_t = \tilde{X}_0 + \int_0^t b_s\left(\tilde{X}_s, \mu_{\tilde{X}_s}\right) \mathrm{d}s + \int_0^t \sigma_s\left(\tilde{X}_s, \mu_{\tilde{X}_s}\right) \mathrm{d}\tilde{W}_s. \tag{3.18}$$

Since the weak uniqueness holds for DDSDE (1.12), any weak solutions have the same distribution. Hence, the whole Euler approximation $X^h$ weakly converges to the unique weak solution $X$ in distribution. As for (1.15), it follows by Krylov's estimate (1.5). □

In order to show (ii) of Theorem 1.2, we need the following important observation due to [4, Lemma 1.1], which has the root of Yamada-Watanabe's theorem.

**Lemma 3.4.** *Let $(Z_h)_{h \in (0,1)}$ be a family of random elements in a Polish space $(E, \rho)$. Then $Z_h$ converges in probability to an $E$-valued random element as $h \to 0$ if and only if for every pair of subsequences $(Z_{h_n}, Z_{\ell_n})_{n \in \mathbb{N}}$, there exists a subsubsequence $(Z_{h_{n(k)}}, Z_{\ell_{n(k)}})_{k \in \mathbb{N}}$ that converges in distribution to a random element in $E \times E$, which supports on the diagonal $\{(x,y) \in E \times E : x = y\}$.*

*Proof.* We use a contradiction method. Suppose that $Z_h$ does not converge in probability. Then there is an $\varepsilon > 0$ such that for any $\delta > 0$, there are $h_\delta$ and $\ell_\delta$ less than $\delta$ such that

$$\mathbb{P}\{\rho(Z_{h_\delta}, Z_{\ell_\delta}) > \varepsilon\} \geqslant \varepsilon.$$

Thus we can choose two subsequences $Z_{h_n}$ and $Z_{\ell_n}$ such that

$$\inf_{n \in \mathbb{N}} \mathbb{P}\{\rho(Z_{h_n}, Z_{\ell_n}) > \varepsilon\} \geqslant \varepsilon. \tag{3.19}$$

By the assumption, there is a subsubsequence $(Z_{h_{n(k)}}, Z_{\ell_{n(k)}})_{k \in \mathbb{N}}$ such that

$$\lim_{k \to \infty} \mathbb{E}\left(\rho(Z_{h_{n(k)}}, Z_{\ell_{n(k)}}) \wedge 1\right) = 0.$$

Clearly, this is contradict with (3.19). By the completeness of $(E, \rho)$, we complete the proof. □

Now we are in a position to give

*Proof of (ii) of Theorem 1.2.* Let $X^{h_n}$ and $X^{\ell_n}$ be two subsequences of $X^h$. Clearly, by Lemma 3.1, the law of $(X^{h_n}, X^{\ell_n}, W)_{n \in \mathbb{N}}$ in $\mathbb{C} \times \mathbb{C} \times \mathbb{C}$ is tight. As above, by Skorokhod's embedding theorem, there exist subsequences $n(k)$, a probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$, carrying stochastic processes $(\tilde{X}^{h_{n(k)}}, \hat{X}^{\ell_{n(k)}}, \tilde{W}^k)$ and $(\tilde{X}, \hat{X}, \tilde{W})$ such that

$$\left(\tilde{X}^{h_{n(k)}}, \hat{X}^{\ell_{n(k)}}, \tilde{W}^k\right) \overset{k \to \infty}{\Rightarrow} \left(\tilde{X}, \hat{X}, \tilde{W}\right) \quad \tilde{\mathbb{P}} - a.s.$$

and for each $k \in \mathbb{N}$,

$$\tilde{\mathbb{P}} \circ \left(\tilde{X}^{h_{n(k)}}, \hat{X}^{\ell_{n(k)}}, \tilde{W}^k\right)^{-1} = \mathbb{P} \circ \left(X^{h_{n(k)}}, X^{\ell_{n(k)}}, W\right)^{-1}.$$

As in showing (3.18), one sees that $(\tilde{X}, \tilde{W})$ and $(\hat{X}, \tilde{W})$ are two solutions of DDSDE (1.12) defined on the same probability space with the same initial values $\tilde{X}_0 = \hat{X}_0$. The latter point is due to

$$\tilde{\mathbb{P}}(\tilde{X}_0 = \hat{X}_0) \geqslant \lim_{k \to \infty} \tilde{\mathbb{P}}\left(\tilde{X}_0^{h_{n(k)}} = \hat{X}_0^{\ell_{n(k)}}\right) = \lim_{k \to \infty} \mathbb{P}\left(X_0^{h_{n(k)}} = X_0^{\ell_{n(k)}}\right) = 1.$$

By the pathwise uniqueness, we obtain $\tilde{X} = \hat{X}$. Thus by Lemma 3.4, we conclude that $X^h$ converges in probability to a random elelment $X$ in $\mathbb{C}$ as $h \downarrow 0$. Using Lemma 3.3, one sees that $X$ is a solution of DDSDE (1.12). Moreover, the convergence (1.16) follows by (3.1) and the dominated convergence theorem. $\qquad\square$

## 4 Propagation of chaos: Proof of Theorem 1.3

In this section we use induction to prove Theorem 1.3. First of all, we prepare several lemmas. The following lemma is the same as in Lemma 3.1. We omit the details.

**Lemma 4.1.** *Let $\beta \geqslant 2$. Under ($\mathbf{H}'_\beta$), for any $T > 0$, there is a constant $C > 0$ such that for all $N \in \mathbb{N}$ and $j = 1, \cdots, N$,*

$$\mathbb{E}\left(\sup_{t \in [0,T]} |X_t^{N,j}|^\beta\right) \leqslant C(1 + \mathbb{E}|\xi_j|^\beta). \tag{4.1}$$

**Lemma 4.2.** *Let $\beta \geqslant 2$. Under ($\mathbf{H}'_\beta$), for any $T > 0$, there is a constant $C > 0$ such that for all $N \in \mathbb{N}$ and $j = 1, \cdots, N$,*

$$\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^{N}\left(b_t(\bar{X}_t^j, \mu_{\bar{X}_t^j}) - \bar{b}_t(\bar{X}_t^j, \bar{X}_t^i)\right)\right|^2 \leqslant C/N. \tag{4.2}$$

*Proof.* Notice that

$$\mu_{\bar{X}_t^1} = \cdots = \mu_{\bar{X}_t^N}.$$

For simplicity, if we define

$$\hat{b}_t(x, y) := b_t(x, \mu_{\bar{X}_t^1}) - \bar{b}_t(x, y),$$

then the left hand side of (4.2) denoted by $\mathscr{I}$ can be written as

$$\mathscr{I} = \mathbb{E}\left|\frac{1}{N}\sum_{i=1}^{N}\hat{b}_t(\bar{X}_t^j, \bar{X}_t^i)\right|^2 = \frac{1}{N^2}\sum_{i,k=1}^{N}\mathbb{E}\langle\hat{b}_t(\bar{X}_t^j, \bar{X}_t^i), \hat{b}_t(\bar{X}_t^j, \bar{X}_t^k)\rangle.$$

For $i \neq j \neq k$, since $\bar{X}_t^i, \bar{X}_t^j, \bar{X}_t^k$ are independent, we have

$$\mathbb{E}\langle\hat{b}_t(\bar{X}_t^j, \bar{X}_t^i), \hat{b}_t(\bar{X}_t^j, \bar{X}_t^k)\rangle = 0.$$

Therefore, by Lemma 4.1 and the linear growth of $\hat{b}$, we get

$$\mathscr{I} \leqslant \frac{2}{N^2}\sum_{i=1}^{N}\mathbb{E}|\hat{b}_t(\bar{X}_t^j, \bar{X}_t^i)|^2 \leqslant \frac{C}{N^2}\sum_{i=1}^{N}\left(1 + \mathbb{E}|\bar{X}_t^j|^2 + \mathbb{E}|\bar{X}_t^i|^2\right) \leqslant C/N.$$

The proof is complete. $\qquad\square$

**Lemma 4.3.** *Let $m \in \mathbb{N}$ and $f : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^m$ be a locally bounded measurable function with*

$$|f_t(x, y)| \leqslant c(1 + |x| + |y|).$$

# A discretized version of Krylov's estimate and its applications

Let $f_t^\varepsilon(x,y) := f_t(\cdot) * \varrho_\varepsilon(x,y)$ be the mollifying approximation. Define

$$\mathscr{A}_{N,\varepsilon}^f(t) := \sup_{j=1,\cdots,N} \mathbb{E}\int_0^t |f_{s_h} - f_{s_h}^\varepsilon|^2(X_{s_h}^{N,j}, \mu_{s_h}^N)\mathrm{d}s, \tag{4.3}$$

$$\bar{\mathscr{A}}_{N,\varepsilon}^f(t) := \sup_{j=1,\cdots,N} \mathbb{E}\int_0^t |f_{s_h} - f_{s_h}^\varepsilon|^2(\bar{X}_{s_h}^j, \mu_{\bar{X}_{s_h}^j})\mathrm{d}s. \tag{4.4}$$

Let $\beta > 2$ and $\nu \in \mathcal{P}_\beta(\mathbb{R}^d)$. Suppose that $(\mathbf{H}'_\beta)$ holds and the initial law $\nu$ has a density $\phi \in L^q_{loc}(\mathbb{R}^d)$ for some $q > 1$. Then we have

$$\lim_{\varepsilon\to 0}\overline{\lim_{N\to\infty}} \mathscr{A}_{N,\varepsilon}^f(t) = 0, \quad \lim_{\varepsilon\to 0}\overline{\lim_{N\to\infty}} \bar{\mathscr{A}}_{N,\varepsilon}^f(t) = 0.$$

*Proof.* We only prove the first limit. For simplicity, we write

$$F_t^\varepsilon(x,y) := |f_t(x,y) - f_t^\varepsilon(x,y)|^2.$$

Without loss of generality we assume $t > h$. Notice that

$$\mathbb{E}\int_0^t F_{s_h}^\varepsilon(X_{s_h}^{N,j},\mu_{s_h}^N)\mathrm{d}s = h\mathbb{E}F_0^\varepsilon(X_0^{N,j},\mu_0^N) + \mathbb{E}\int_h^t F_{s_h}^\varepsilon(X_{s_h}^{N,j},\mu_{s_h}^N)\mathrm{d}s =: I_{N,\varepsilon}^{(1)} + I_{N,\varepsilon}^{(2)}.$$

For $I_{N,\varepsilon}^{(1)}$, by the assumption we have

$$I_{N,\varepsilon}^{(1)} = \frac{h}{N}\sum_{i=1}^N \mathbb{E}F_0^\varepsilon(X_0^{N,j},X_0^{N,i}) = \frac{h}{N}\int_{\mathbb{R}^d}F_0^\varepsilon(x,x)\phi(x)\mathrm{d}x$$
$$+ \frac{h(N-1)}{N}\int_{\mathbb{R}^{2d}}F_0^\varepsilon(x,y)\phi(x)\phi(y)\mathrm{d}x\mathrm{d}y =: I_{N,\varepsilon}^{(11)} + I_{N,\varepsilon}^{(12)}.$$

For $I_{N,\varepsilon}^{(11)}$, we clearly have

$$\lim_{N\to\infty}\sup_\varepsilon I_{N,\varepsilon}^{(11)} \leqslant \lim_{N\to\infty}\frac{C}{N}\int_{\mathbb{R}^d}(1+|x|^2)\phi(x)\mathrm{d}x = 0.$$

For $I_{N,\varepsilon}^{(12)}$, if we define $B_R := \{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d : |x| < R, |y| < R\}$ for $R > 0$, then by Hölder's inequality and $\phi \in L^q_{loc}(\mathbb{R}^d)$,

$$I_{N,\varepsilon}^{(12)} \leqslant \int_{B_R}F_0^\varepsilon(x,y)\phi(x)\phi(y)\mathrm{d}x\mathrm{d}y + \int_{B_R^c}F_0^\varepsilon(x,y)\phi(x)\phi(y)\mathrm{d}x\mathrm{d}y$$
$$\lesssim \left(\iint_{B_R}|F_0^\varepsilon(x,y)|^{\frac{q}{q-1}}\mathrm{d}x\mathrm{d}y\right)^{\frac{q-1}{q}}\left(\iint_{B_R}|\phi(x)\phi(y)|^q\mathrm{d}x\mathrm{d}y\right)^{\frac{1}{q}}$$
$$+ \int_{B_R^c}(1+|x|^2+|y|^2)\phi(x)\phi(y)\mathrm{d}x\mathrm{d}y$$
$$\lesssim \left(\iint_{B_R}|F_0^\varepsilon(x,y)|^{\frac{q}{q-1}}\mathrm{d}x\mathrm{d}y\right)^{\frac{q-1}{q}}\left(\int_{|x|<R}|\phi(x)|^q\mathrm{d}x\right)^{\frac{2}{q}}$$
$$+ \frac{1}{R^{\beta-2}}\int_{\mathbb{R}^{2d}}(1+|x|^\beta+|y|^\beta)\phi(x)\phi(y)\mathrm{d}x\mathrm{d}y,$$

where the constant $C$ contained in $\lesssim$ is independent of $N, R, \varepsilon$. By the dominated convergence theorem and first letting $\varepsilon \to 0$ and then $R \to \infty$, we get

$$\lim_{\varepsilon\to 0}\sup_N I_{N,\varepsilon}^{(12)} = 0.$$

Next we treat $I_{N,\varepsilon}^{(2)}$ and write

$$I_{N,\varepsilon}^{(2)} = \frac{1}{N}\mathbb{E}\int_h^t F_{s_h}^\varepsilon(X_{s_h}^{N,1}, X_{s_h}^{N,1})\mathrm{d}s + \frac{1}{N}\sum_{i=2}^N \mathbb{E}\int_h^t F_{s_h}^\varepsilon(X_{s_h}^{N,1}, X_{s_h}^{N,i})\mathrm{d}s =: I_{N,\varepsilon}^{(21)} + I_{N,\varepsilon}^{(22)}.$$

For $I_{N,\varepsilon}^{(21)}$, by (4.1) we have

$$\lim_{N\to\infty}\sup_\varepsilon I_{N,\varepsilon}^{(21)} \leqslant \lim_{N\to\infty}\frac{C}{N}\mathbb{E}\int_h^t(1+|X_{s_h}^{N,1}|^2)\mathrm{d}s = 0.$$

For $I_{N,\varepsilon}^{(22)}$, using the discretized Krylov estimate and the same argument as in showing (3.7), we also have

$$\lim_{\varepsilon\to 0}\sup_N I_{N,\varepsilon}^{(22)} = 0.$$

Combining the above limits, we complete the proof. $\qquad\qquad\square$

Now we can give

*Proof of Theorem 1.3.* By equations (1.17), (1.18) and Burkholder's inequality we have

$$\mathbb{E}\left(\sup_{s\in[0,t]}|X_s^{N,j} - \bar{X}_s^j|^2\right) \lesssim \mathbb{E}\int_0^t \left|b_{s_h}\left(X_{s_h}^{N,j}, \mu_{s_h}^N\right) - b_{s_h}\left(\bar{X}_{s_h}^j, \mu_{\bar{X}_{s_h}^j}\right)\right|^2 \mathrm{d}s$$

$$+ \mathbb{E}\int_0^t \left|\sigma_{s_h}\left(X_{s_h}^{N,j}, \mu_{s_h}^N\right) - \sigma_{s_h}\left(\bar{X}_{s_h}^j, \mu_{\bar{X}_{s_h}^j}\right)\right|^2 \mathrm{d}s$$

$$\lesssim \mathbb{E}\int_0^t \left|b_{s_h}^\varepsilon\left(X_{s_h}^{N,j}, \mu_{s_h}^N\right) - b_{s_h}^\varepsilon\left(\bar{X}_{s_h}^j, \mu_{\bar{X}_{s_h}^j}\right)\right|^2 \mathrm{d}s$$

$$+ \mathbb{E}\int_0^t \left|\sigma_{s_h}^\varepsilon\left(X_{s_h}^{N,j}, \mu_{s_h}^N\right) - \sigma_{s_h}^\varepsilon\left(\bar{X}_{s_h}^j, \mu_{\bar{X}_{s_h}^j}\right)\right|^2 \mathrm{d}s$$

$$+ \mathscr{A}_{N,\varepsilon}^b(t) + \bar{\mathscr{A}}_{N,\varepsilon}^b(t) + \mathscr{A}_{N,\varepsilon}^\sigma(t) + \bar{\mathscr{A}}_{N,\varepsilon}^\sigma(t),$$

where for $f = b$ and $\sigma$, $f_t^\varepsilon(x,y) := f_t * \varrho_\varepsilon(x,y)$ is the mollifying approximation, and $\mathscr{A}_{N,\varepsilon}^f(t)$ and $\bar{\mathscr{A}}_{N,\varepsilon}^f(t)$ are defined by (4.3) and (4.4), respectively. Now let $\chi_R(x,y)$ be a smooth function with

$$\chi_R(x,y) = 1, \ |x|\vee|y| < R, \ \chi_R(x,y) = 0, \ \ |x|\vee|y| > R+1.$$

For $f = b$ or $\sigma$, by definition, we make the following decomposition:

$$\mathbb{E}\int_0^t \left|f_{s_h}^\varepsilon\left(X_{s_h}^{N,j}, \mu_{s_h}^N\right) - f_{s_h}^\varepsilon\left(\bar{X}_{s_h}^j, \mu_{\bar{X}_{s_h}^j}\right)\right|^2 \mathrm{d}s$$

$$= \mathbb{E}\int_0^t \left|\frac{1}{N}\sum_{i=1}^N \left(f_{s_h}^\varepsilon\left(X_{s_h}^{N,j}, X_{s_h}^{N,i}\right) - f_{s_h}^\varepsilon\left(\bar{X}_{s_h}^j, \bar{X}_{s_h}^i\right)\right)\right|^2 \mathrm{d}t$$

$$\leqslant \frac{1}{N}\sum_{i=1}^N \mathbb{E}\int_0^t \left|f_{s_h}^\varepsilon\left(X_{s_h}^{N,j}, X_{s_h}^{N,i}\right) - f_{s_h}^\varepsilon\left(\bar{X}_{s_h}^j, \bar{X}_{s_h}^i\right)\right|^2 \mathrm{d}t$$

$$\leqslant \frac{2}{N}\sum_{i=1}^N \mathbb{E}\int_0^t \left|f_{s_h}^{\varepsilon,R}\left(X_{s_h}^{N,j}, X_{s_h}^{N,i}\right) - f_{s_h}^{\varepsilon,R}\left(\bar{X}_{s_h}^j, \bar{X}_{s_h}^i\right)\right|^2 \mathrm{d}s$$

$$+ \frac{4}{N}\sum_{i=1}^N \mathbb{E}\int_0^t |f_{s_h}^\varepsilon - f_{s_h}^{\varepsilon,R}|^2\left(X_{s_h}^{N,j}, X_{s_h}^{N,i}\right)\mathrm{d}s$$

$$+ \frac{4}{N}\sum_{i=1}^N \mathbb{E}\int_0^t |f_{s_h}^\varepsilon - f_{s_h}^{\varepsilon,R}|^2\left(\bar{X}_{s_h}^j, \bar{X}_{s_h}^i\right)\mathrm{d}s$$

$$=: \frac{1}{N}\sum_{i=1}^N \left(\mathscr{I}_{R,1}^{i,j} + \mathscr{I}_{R,2}^{i,j} + \mathscr{I}_{R,3}^{i,j}\right),$$

## References

[1] J. Bao, X. Huang, and C. G. Yuan, Convergence rate of Euler-Maruyama scheme for SDEs with Hölder–Dini continuous drifts. *J. Theoret. Probab.*, **32** (2019), 848–871. MR-3959630

[2] J. Bao and X. Huang, Approximations of Mckean-Vlasov SDEs with irregular coefficients. arXiv:1905.08522

[3] I. Gyöngy, A note on Euler's approximation. *Potential Anal.* **8** (1998), 205–216. MR-1625576

[4] I. Gyöngy and N. Krylov, Existence of strong solutions for Ito's stochastic equations via approximations, *Probab. Theory Relat. Fields.*, **105** (1996), 143–158. MR-1392450

[5] I. Gyöngy and M. Rásonyi, A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients. *Stochastic Process. Appl.*, **121** (2011), 2189–2200. MR-2822773

[6] X. Huang and F. Y. Wang, Distribution dependent SDEs with singular coefficients. arXiv:1805.01682v1.

[7] P. E. Kloeden and E. Platen, *Numerical solutions of stochastic differential equations*. Stochastic modelling and applied probability. Springer, Berlin, 1992. MR-1214374

[8] N. V. Krylov, *Controlled diffusion processes*. Translated from the Russian by A. B. Aries. Applications of Mathematics, Vol. 14. Springer-Verlag, New York-Berlin, 1980. MR-0601776

[9] V. Lemaire and S. Menozzi, On some Non Asymptotic Bounds for the Euler Scheme. *Electronic Journal Probability.*, **15** (2010), 1645–1681. MR-2735377

[10] G. Leobacher and M. Szölgyenyi, Convergence of the Euler-Maruyama method for multidimensional SDEs with discontinuous drift and degenerate diffusion coefficient. *Numer. Math.*, **138** (2018), 219–239. MR-3745015

[11] J. Li and H. Min, Weak solutions of mean-field stochastic differential equations and application to zero-sum stochastic differential games. *SIAM Journal on Control and Optimization*, **54**(3) (2016), 1826–1858. MR-3517577

[12] Y. S. Mishura and A. Y. Veretennikov, Existence and uniqueness theorems for solutions of McKean-Vlasov stochastic equations. arXiv:1603.02212v4.

[13] H.-L. Ngo and D. Taguchi, Strong rate of convergence for the Euler-Maruyama approximation of stochastic differential equations with irregular coefficients. *Math. Comp.*, **85** (2016), 1793–1819. MR-3471108

[14] M. Röckner and X. Zhang, Well-posedness of distribution dependent SDEs with singular drifts. arXiv:1803.02212v4.

[15] A. V. Skorokhod, *Studies in the theory of random processes*. Dover, New York, 1982. MR-0185620

[16] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*. Grundlehren der Mathematischen Wissenschaften, Vol. 233. Springer-Verlag, Berlin-New York, 1979. MR-0532498

[17] A. S. Sznitman, *Topics in propagation of chaos*. In École d'Été de Prob. de Saint-Flour XIX-1989, Vol. 1464, *Lect. Notes in Math.*, pages 165–251. Springer-Verlag, 1991. MR-1108185

[18] A. J. Veretennikov, On the strong solutions of stochastic differential equations. *Theory Probab. Appl.*, **24** (1979), 354–366. MR-0532447

[19] L. Xie and X. Zhang, Ergodicity of stochastic differential equations with jumps and singular coefficients. To appear in *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*. arXiv:1705.01402v1

# Electronic Journal of Probability
# Electronic Communications in Probability

## Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

## Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4], ProjectEuclid[5]

- Purely electronic

## Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm