A Berry-Esseen Bound for Vector-valued Martingales

Denis Kojevnikov∗ and Kyungchul Song§

Abstract. This note provides a conditional Berry-Esseen bound for the sum of a martingale difference sequence \( \{X_i\}_{i=1}^n \) in \( \mathbb{R}^d \), \( d \geq 1 \), adapted to a filtration \( \{\mathcal{F}_i\}_{i=1}^n \). We approximate the conditional distribution of \( S = \sum_{i=1}^n X_i \) given a sub-\( \sigma \)-field \( \mathcal{F}_0 \subset \mathcal{F}_1 \) by that of a mean zero normal random vector having the same conditional variance given \( \mathcal{F}_0 \) as the vector \( S \). Assuming that the conditional variances \( \mathbb{E}[X_i X_i^\top | \mathcal{F}_{i-1}], i \geq 1 \), are \( \mathcal{F}_0 \)-measurable and non-singular, and the third conditional moments of \( \|X_i\|, i \geq 1 \), given \( \mathcal{F}_0 \) are uniformly bounded, we present a simple bound on the conditional Kolmogorov distance between \( S \) and its approximation given \( \mathcal{F}_0 \) which is of order \( O_{a.s.}(\ln(ed)^{5/4}n^{-1/4}) \).

Keywords. Berry-Esseen bound; Gaussian approximation; Martingale-difference sequence; Vector-valued martingale

1. Introduction

Let \( (\Omega, \mathcal{F}, \mathbb{P}) \) be a probability space and let \( \{X_i\}_{i=1}^n \) be an \( \mathbb{R}^d \)-valued martingale difference sequence with \( d \geq 1 \) adapted to a filtration \( \{\mathcal{F}_i\}_{i=1}^n \), i.e., each \( X_i \) is \( \mathcal{F}_i \)-measurable and \( \mathbb{E}[X_{i+1} | \mathcal{F}_i] = 0 \) a.s. In addition, suppose that we are given a sub-\( \sigma \)-field \( \mathcal{F}_0 \subset \mathcal{F}_1 \), not necessarily trivial, such that \( \mathbb{E}[X_1 | \mathcal{F}_0] = 0 \) a.s. Throughout the paper we assume that each \( X_i \) has finite conditional third moment given \( \mathcal{F}_0 \), i.e., \( \mathbb{E}[\|X_i\|_\infty^3 | \mathcal{F}_0] < \infty \) a.s., where \( \|\cdot\|_\infty \) denotes the maximum norm on \( \mathbb{R}^d \).

The goal of this paper is to establish a uniform distributional approximation of the random vector \( S := \sum_{i=1}^n X_i \) conditionally on \( \mathcal{F}_0 \) by a suitably chosen Gaussian analog. Specifically, we consider a random vector \( T \) whose conditional distribution given \( \mathcal{F}_0 \) is \( \mathcal{N}(0, V) \), where the covariance matrix \( V \) is a version of \( \mathbb{E}[SS^\top | \mathcal{F}_0] \). Namely, the conditional characteristic function of \( T \) is given by

\[
\mathbb{E}[e^{it^\top T} | \mathcal{F}_0] = \exp\left(-\frac{1}{2}t^\top V t\right) \quad \text{a.s.}
\]
for all \( t \in \mathbb{R}^d \). Then we establish a bound on the conditional Kolmogorov distance between \( S \) and \( T \) given \( \mathcal{F}_0 \).

Let \( \mathcal{A} \) denote the collection of sets of the form \( \prod_{j=1}^d (-\infty, r_j] \) with \( r \equiv [r_1, \ldots, r_d]^{\top} \in \mathbb{R}^d \). Also, let \( \mu^G_X \) denote the regular conditional distribution of a vector \( X \) given a sub-\( \sigma \)-field \( \mathcal{G} \subset \mathcal{F} \).\(^1\) The conditional Kolmogorov distance between random vectors \( X \) and \( Y \) in \( \mathbb{R}^d \) given a sub-\( \sigma \)-field \( \mathcal{G} \subset \mathcal{F} \) is defined by

\[
(1.1) \quad d_K(X, Y \mid \mathcal{G})(\omega) := \sup_{A \in \mathcal{A}} |\mu^G_X(\omega, A) - \mu^G_Y(\omega, A)|.
\]

Assuming that the conditional variances \( \mathbb{E}[X_iX_i^{\top} \mid \mathcal{F}_{i-1}] \), \( i \geq 1 \), are \( \mathcal{F}_0 \)-measurable, and the third conditional moments of \( \|X_1\|_{\infty}, \|X_2\|_{\infty}, \ldots \) given \( \mathcal{F}_0 \) are uniformly bounded, we present a simple bound on \( d_K(S, T \mid \mathcal{F}_0) \) of order \( O_{a.s.}((\ln ed)^{5/4}n^{-1/4}) \). In addition, we require that the minimum eigenvalues of \( \mathbb{E}[X_iX_i^{\top} \mid \mathcal{F}_0], i \geq 1 \), are bounded away from zero, that is, the random vectors \( X_1, X_2, \ldots \) are assumed to have non-degenerate conditional distributions given \( \mathcal{F}_0 \).

For scalar-valued martingale difference sequences with constant conditional variances and finite third moments, Grams (1972) showed that \( d_K(S, T) = O(n^{-1/4}) \). If, in addition, \( X_i, i \geq 1 \), are uniformly bounded, Bolthausen (1982) established a bound of order \( O((\ln n) n^{-1/2}) \). Furthermore, he provided examples of martingale difference sequences for which both estimates are sharp. The classical rate of \( O(n^{-1/2}) \) is nevertheless possible under stronger conditions on the conditional moments of \( X_i \)'s. See, for example, Kir'yanova and Rotar' (1991), Renz (1996), and Wu et al. (2020) for recent developments.

In multidimensional settings, extensive research has been focused on sequences of independent random vectors. Chernozhukov et al. (2013) established a Berry-Esseen bound of order \( O((\ln (dn))^{7/8}n^{-1/8}) \) for maxima of sums of such vectors. This result was subsequently improved in Chernozhukov et al. (2017) and Chernozhukov et al. (2019). Recently, Lopes (2020) provided a nearly \( 1/\sqrt{n} \) bound on \( d_K(S, T) \) for i.i.d. sub-Gaussian random vectors, and Kuchibhotla and Rinaldo (2020) improved that result by showing an \( O((\ln en)^{1/2}n^{-1/2}) \) rate of convergence under the weakest possible conditions. This paper relies on the smoothing inequality presented in the latter work.

2. Preliminary Results

Let \( Z_1, \ldots, Z_n \) be i.i.d. standard normal random vectors in \( \mathbb{R}^d \) independent of \( \mathcal{F}_n \). For \( 1 \leq i \leq n \), let \( Y_i = \Sigma_i^{1/2}Z_i \), where \( \Sigma_i \) is a version of \( \mathbb{E}[X_iX_i^{\top} \mid \mathcal{F}_0] \). It is clear that the

\( ^1\)The regular conditional distribution \( \mu^G_Z \) of a random vector \( Z \in \mathbb{R}^d \) given \( \mathcal{G} \subset \mathcal{F} \) satisfies: (i) \( \forall B \in \mathcal{B}(\mathbb{R}^d), \mu^G_Z(\cdot, B) \) is a version of \( \mathcal{P}(Z \in B \mid \mathcal{G})(\cdot) \), and (ii) \( \forall \omega \in \Omega, \mu^G_Z(\omega, \cdot) \) is a distribution on \( \mathbb{R}^d \). In particular, condition (ii) implies that \( d_K(X, Y \mid \mathcal{G}) \) defined in (1.1) is \( \mathcal{G} \)-measurable.
conditional distribution of $T$ given $\mathcal{F}_0$ is the same as that of $\sum_{i=1}^n Y_i$, and so we associate $T$ with the latter sum. In addition, by the properties of conditional distributions,

$$d_K(S, T \mid \mathcal{F}_0) = \sup_{r \in \mathbb{Q}^d} |\mathbb{P}(S \in A_r \mid \mathcal{F}_0) - \mathbb{P}(T \in A_r \mid \mathcal{F}_0)| \quad \text{a.s.,}$$

where $A_r := \prod_{j=1}^d (-\infty, r_j]$ with $r \in \mathbb{R}^d$ is a generic set in $\mathcal{A}$, and $\mathbb{Q}$ is the set of rational numbers.

Consider a random vector $\eta \sim \mathcal{N}(0, I_d)$, independent of $Z_1, \ldots, Z_n$ and $\mathcal{F}_n$. We approximate the probabilities on the right-hand side of (2.1) with conditional expectations of the following smooth function:

$$\varphi_r(x, \varepsilon) := \mathbb{P}(x + \varepsilon \eta \in A_r),$$

evaluated at $(S, \varepsilon)$ and $(T, \varepsilon)$, respectively, where $\varepsilon$ is a positive, $\mathcal{F}_0$-measurable random variable which will be determined later. Note that for a fixed $\varepsilon > 0$, the function $x \mapsto \varphi_r(x, \varepsilon)$ is infinitely differentiable, and by Lemma 2.3 in Fang and Koike (2021) for each $x, r \in \mathbb{R}^d$ and $s \geq 1$ we have

$$\sum_{j_1, \ldots, j_s=1}^d \left| \frac{\partial}{\partial x_{j_1}} \cdots \frac{\partial}{\partial x_{j_s}} \varphi_r(x, \varepsilon) \right| \leq C_s \varepsilon^{-s} \left[ \ln + d \right]^{s/2},$$

where $C_s > 0$ is a constant depending only on $s$ and $\ln + x \equiv 1 \lor \ln x$. In addition, for an $\mathcal{F}_0$-measurable random variable $\varepsilon$,

$$\mathbb{E}[\varphi_r(S, \varepsilon) - \varphi_r(T, \varepsilon) \mid \mathcal{F}_0] = \mathbb{P}(S + \varepsilon \eta \in A_r \mid \mathcal{F}_0) - \mathbb{P}(T + \varepsilon \eta \in A_r \mid \mathcal{F}_0) \quad \text{a.s.}$$

The following lemma establishes an upper bound on the approximation error due to the use of $\varphi_r$. We define

$$\sigma^2 := \min_{1 \leq j \leq d} |V|_{jj}.$$

**Lemma 2.1.** Suppose that $\sigma > 0$ a.s. There exists a universal constant $C > 0$ such that for any $\varepsilon > 0$,

$$d_K(S, T \mid \mathcal{F}_0) \leq \sup_{r \in \mathbb{Q}^d} |\mathbb{E}[\varphi_r(S, \varepsilon) - \varphi_r(T, \varepsilon) \mid \mathcal{F}_0]| + \frac{C \varepsilon \ln + d}{\sigma} \quad \text{a.s.}$$

**Proof.** Let $\gamma_\varepsilon$ denote a mean zero Gaussian measure on $\mathbb{R}^d$ with covariance matrix $\varepsilon^2 I_d$. By Lemma 1 in Kuchibhotla and Rinaldo (2020), for any $r \in \mathbb{R}^d$ and $\varepsilon > 0$,

$$\left| (\mu^\mathcal{F}_0 - \mu^\mathcal{F}_0) (\omega, A_r) \right| \leq \sup_{r \in \mathbb{Q}^d} \left| (\mu^\mathcal{F}_0 \ast \gamma_\varepsilon - \mu^\mathcal{F}_0 \ast \gamma_\varepsilon) (\omega, A_r) \right| + \frac{C \varepsilon \ln + d}{\sigma(\omega)}$$
for some universal constant $C > 0$. On the other hand, for almost all $\omega \in \Omega$,
\[
P(S + \epsilon \eta \in A_r | F_0)(\omega) - P(T + \epsilon \eta \in A_r | F_0)(\omega) \\
= \int 1_{A_r}(x + \epsilon z)(\mu_{S}^{F_0} \otimes \mu_{\eta} - \mu_{T}^{F_0} \otimes \mu_{\eta})(\omega, dx(z)) \\
= (\mu_{S}^{F_0} * \gamma_{\epsilon} - \mu_{T}^{F_0} * \gamma_{\epsilon})(\omega, A_r).
\]

The next result implies the regularity of the conditional Kolmogorov distance in the sense that for suitable random vectors $X, Y,$ and $Z$, $d_K(X + Z, Y + Z | F_0) \leq d_K(X, Y | F_0)$ a.s. when $Z$ is conditionally independent of $X$ and $Y$ given $F_0$.

**Lemma 2.2.** Let $X, Y,$ and $Z$ be random vectors in $\mathbb{R}^d$ defined on $(\Omega, \mathcal{F}, P)$ such that $Z$ is conditionally independent of $X$ and $Y$ given $F_0$. Then for any $A \in \mathcal{A}$,
\[
|P(X + Z \in A | F_0) - P(Y + Z \in A | F_0)| \leq d_K(X, Y | F_0) \quad a.s.
\]

**Proof.** Let $\mathcal{G} := F_0 \lor \sigma(Z)$. Then
\[
P(X + Z \in A | F_0) - P(Y + Z \in A | F_0) \\
= E[P(X + Z \in A | \mathcal{G}) - P(Y + Z \in A | \mathcal{G}) | F_0] \quad a.s.,
\]
and for almost all $\omega \in \Omega$,
\[
|P(X + Z \in A | \mathcal{G})(\omega) - P(Y + Z \in A | \mathcal{G})(\omega)| \\
= \left| \int 1_A(x + Z(\omega))(\mu_X^{F_0} - \mu_Y^{F_0})(\omega, dx) \right| \leq d_K(X, Y | F_0)(\omega).
\]

Finally, we give an upper bound on the moments of the maximum norm of a Gaussian random vector.

**Lemma 2.3.** Let $Y \equiv [Y_1, \ldots, Y_d]^T$ be a zero-mean Gaussian vector in $\mathbb{R}^d$, $d \geq 1$, with $\sigma_j^2 := EY_j^2 > 0$ for all $1 \leq j \leq d$, and let $\bar{\sigma} := \max_{1 \leq j \leq d} \sigma_j$. Then for any $s \geq 2$,
\[
E\|Y\|_\infty^s \leq C_s \bar{\sigma}^s (\ln d)^{s/2},
\]
where $C_s > 0$ is a constant depending only on $s$.

**Proof.** Let $f : [a, \infty) \to \mathbb{R}$, $a \geq 0$, be a strictly increasing convex function. Using Jensen’s inequality, we have
\[
E\|Y\|_\infty^s \leq E[a \lor \|Y\|_\infty^s] \leq f^{-1}(E[f(a \lor \|Y\|_\infty^s)]).
\]
First, for $s > 2$ consider $f(x) = \exp (c_s (x/a)^{2/s})$ with $a > 0$ and $c_s := s/2 - 1$, which is convex on $[a, \infty)$. Letting $a = (2\sqrt{c_s\sigma})^s$, we find that

$$
E[f(a \vee \|Y\|_{\infty}^s)] = E \exp \left( c_s \left( 1 \vee \frac{\|Y\|_{\infty}^2}{a^{2/s}} \right) \right) \leq e^{c_s} E \exp \left( \frac{\|Y\|_{\infty}^2}{4\sigma^2} \right)
$$

$$
\leq e^{c_s} \sum_{j=1}^{p} E \exp \left( \frac{|Y_j|^2}{4\sigma^2} \right) = e^{c_s} \sum_{j=1}^{d} \sqrt{\frac{2\sigma^2}{2\sigma^2 - \sigma_j^2}} \leq \sqrt{2} e^{c_s} d,
$$

and, therefore,

$$
E\|Y\|_{\infty}^s \leq \left[ \ln \left( \sqrt{2} e^{c_s} d \right) \right]^{s/2} (2\sigma)^s \leq C_s \sigma^s (\ln d)^{s/2}
$$

for some $C_s > 0$ depending only on $s$. For $s = 2$ we take $f(x) = \exp(x/(2\sigma^2))$ and $a = 0$ which similarly yield (2.5). $\blacksquare$

3. Main Results

In this section we derive a Berry-Esseen bound for the random vector $S$. Let $\| \cdot \|_{e,p}$ denote the element-wise $p$-norm in $\mathbb{R}^{k \times l}$, i.e., for a $k \times l$ matrix $A$, $\|A\|_{e,p} = \|\text{vec}(A)\|_p$, $p \in [1, \infty]$, and let

$$
\lambda^2 := \min_{1 \leq i \leq n} \lambda_{\min}(\Sigma_i),
$$

where $\lambda_{\min}(A)$ is the smallest eigenvalue of $A$.

**Lemma 3.1.** Suppose that $\lambda > 0$ a.s. There exists a universal constant $C > 0$ such that for any $\epsilon > 0$,

$$
\sup_{r \in Q} \left| E[\varphi_r(S, \epsilon) - \varphi_r(T, \epsilon) \mid \mathcal{F}_0] \right| 
\leq C[\ln d]^{3/2} \lambda (\gamma_1 + \gamma_3) \epsilon^{-1} + C[\ln d] \beta \ln \left( 1 + n \lambda^2 \epsilon^{-2} \right) \quad \text{a.s.,}
$$

where

$$
\gamma_s := \max_{1 \leq i \leq n} \left( E[\|X_i\|_{\infty}^s \mid \mathcal{F}_0] + \bar{\sigma}_s^s [\ln d]^{s/2} \right) / \lambda_s, \quad s > 0,
$$

$$
\beta := \max_{1 \leq i \leq n} E[\|E[X_i X_i^\top \mid \mathcal{F}_{i-1}] - \Sigma_i\|_{e,\infty} \mid \mathcal{F}_0] / \lambda^2,
$$

and $\bar{\sigma}_i^2 := \max_{1 \leq j \leq d} [\Sigma_i]_{jj}$.

**Proof.** First, letting

$$
U_i := \sum_{j=1}^{i-1} X_j + \sum_{j=i+1}^{n} Y_j,
$$

...
1 \leq i \leq n, we write

\begin{equation}
\left| E[\varphi_r(S, \epsilon) - \varphi_r(T, \epsilon) \mid F_0] \right| \\
\leq \sum_{i=1}^{n} \left| E[\varphi_r(U_i + X_i, \epsilon) - \varphi_r(U_i + Y_i, \epsilon) \mid F_0] \right| \text{ a.s.}
\end{equation}

(3.1)

Consider the right hand side of the preceding display. For each \(1 \leq i \leq n\), let \(S_i = S_{i-1} + X_i\) and \(T_i = T_{i-1} + Y_i\) with \(S_0 \equiv 0\) and \(T_0 \equiv 0\). We also define

\[ \varepsilon_i := \left( \epsilon^2 + (n-i)^2 \right)^{1/2} \text{ and } V_i := \left( \sum_{k=i+1}^{n} \Sigma_k - (n-i)^2 I_d \right)^{1/2}. \]

Since \(Y_1, \ldots, Y_n\) are conditionally independent of \(F_n\) given \(F_0\), by Lemma 2.2 we have

\[ |E[\varphi_r(U_i + X_i, \epsilon) - \varphi_r(U_i + Y_i, \epsilon) \mid F_0]| \]

\[ = |P(S_{i-1} + X_i + \varepsilon_i \eta \in A_{r-V_i \eta^r} \mid F_0) - P(S_{i-1} + Y_i + \varepsilon_i \eta \in A_{r-V_i \eta^r} \mid F_0)| \]

\[ \leq \sup_{\eta \in Q^d} |E[\varphi_r(S_{i-1} + X_i, \varepsilon_i) - \varphi_r(S_{i-1} + Y_i, \varepsilon_i) \mid F_0]| \text{ a.s.} \]

for each \(1 \leq i < n\), where \(\eta^r\) is an independent copy of \(\eta\).

**Claim 3.1.** There exists a universal constant \(C > 0\) such that for each \(r \in R\),

\begin{equation}
E[\varphi_r(S_{i-1} + X_i, \varepsilon_i) - \varphi_r(S_{i-1} + Y_i, \varepsilon_i) \mid F_0] \\
\leq C \varepsilon_i^{-2} [\ln d] \alpha^2 \beta + C \varepsilon_i^{-3} [\ln d]^{3/2} \alpha^3 \gamma_3,
\end{equation}

(3.2)

if \(1 \leq i < n\), and

\begin{equation}
E[\varphi_r(S_{n-1} + X_n, \epsilon) - \varphi_r(S_{n-1} + Y_n, \epsilon) \mid F_0] \\
\leq C \varepsilon^{-1} [\ln d]^{1/2} \alpha \gamma_1.
\end{equation}

(3.3)

**Proof.** We show (3.2). The inequality (3.3) follows using similar arguments. Let \(h_{1i}(\tau) := \varphi_r(S_{i-1} + \tau X_i, \varepsilon_i)\) and \(h_{2i}(\tau) := \varphi_r(S_{i-1} + \tau Y_i, \varepsilon_i)\). Using Taylor’s expansion up to terms of the third order,

\[ h_{1i}(1) - h_{2i}(1) = \sum_{j=1}^{2} \frac{1}{j!} \left( h_{1i}^{(j)}(0) - h_{2i}^{(j)}(0) \right) + \frac{1}{3!} \left( h_{1i}^{(3)}(\tau_1) - h_{2i}^{(3)}(\tau_2) \right), \]

where \(|\tau_1|, |\tau_2| \leq 1\). First, it is clear that

\[ E[E[h_{1i}^{(j)}(0) - h_{2i}^{(j)}(0) \mid F_{i-1}] \mid F_0] = 0 \text{ a.s.,} \]

and, using (2.2),

\[ |E[h_{1i}^{(j)}(0) - h_{2i}^{(j)}(0) \mid F_0]| \leq E[|E[h_{1i}^{(j)}(0) - h_{2i}^{(j)}(0) \mid F_{i-1}]| \mid F_0] \]

\[ \leq C' \varepsilon_i^{-2} [\ln d] E[|E[X_i X_i^T \mid F_{i-1}] - \Sigma_i| |\epsilon, \infty \mid F_0] \text{ a.s.,} \]
where $C' > 0$ is a universal constant. Finally, using (2.2) and Lemma 2.3,

$$|\mathbb{E}[h_{i_1}^{(3)}(\tau_1) - h_{i_2}^{(3)}(\tau_2) | \mathcal{F}_0] | \leq \mathbb{E}[|h_{i_1}^{(3)}(\tau_1)| | \mathcal{F}_0] + \mathbb{E}[|h_{i_2}^{(3)}(\tau_2)| | \mathcal{F}_0]$$

$$\leq C''\varepsilon^{-3} \left( \mathbb{E}[|\|X_i\|_\infty| | \mathcal{F}_0] + \sigma_i^3[\ln |\ln + d|^{3/2}] \right) \text{ a.s.,}$$

where $C'' > 0$ is a universal constant.

Using Claim 3.1, the result follows from (3.1) by noticing that

$$\sum_{i=1}^{n-1} \varepsilon_i^{-2} \leq \int_0^1 \frac{n - 1}{\epsilon^2 + (n - 1)\lambda^2 x} \, dx \leq \frac{1}{\lambda^2} \ln \left( 1 + \frac{n\lambda^2}{\varepsilon^2} \right)$$

and

$$\sum_{i=1}^{n-1} \varepsilon_i^{-3} \leq \int_0^1 \frac{n - 1}{\epsilon^2 + (n - 1)\lambda^2 x}^{3/2} \, dx \leq \frac{2}{\lambda^2\epsilon}. \tag{3.2}$$

**Theorem 3.1.** Suppose that $\lambda > 0$ a.s. There exists a universal constant $C > 0$ such that

$$d_K(S, T | \mathcal{F}_0) \leq C[\ln + d]^{5/4}(\gamma\lambda/\sigma)^{1/2}$$

$$+ C[\ln + d]^{3/2}\beta\ln \left( 1 + \frac{n\lambda}{[\ln + d]^{1/2}\gamma} \right) \text{ a.s.,} \tag{3.4}$$

where $\gamma \equiv \gamma_1 + \gamma_3$, and $\sigma$ is defined in (2.3).

**Proof.** Using Lemmas 2.1 and 3.1, we find that for any $\epsilon > 0$,

$$d_K(S, T | \mathcal{F}_0) \leq \frac{C[\ln + d]^{3/2}\lambda\gamma}{\epsilon}$$

$$+ C[\ln + d]^{3/2}\beta\ln \left( 1 + \frac{n\lambda^2}{\varepsilon^2} \right) + C\varepsilon[\ln + d] \text{ a.s.,}$$

where $C$ is a universal constant. Since this inequality holds for all $\epsilon > 0$, it also holds for random $\epsilon$ a.s. on the event $\{\epsilon \in (0, \infty)\}$. Consequently, the result follows by choosing $\epsilon = [\ln + d]^{1/4}(\lambda\sigma\gamma)^{1/2}$ and noticing that $\sigma \geq \lambda$. \rule{5mm}{5mm}

**Remark.** (1) If the conditional variances $\mathbb{E}[X_iX_i^\top | \mathcal{F}_{i-1}]$, $1 < i \leq n$, are $\mathcal{F}_0$-measurable, then $\beta = 0$ a.s., and the bound in Theorem 3.1 becomes

$$d_K(S, T | \mathcal{F}_0) \leq C[\ln + d]^{5/4}(\gamma\lambda/\sigma)^{1/2}$$

$$\leq C\gamma^{1/2}[\ln + d]^{5/4}n^{-1/4} \text{ a.s.}$$

because

$$\sigma^2/n \geq \min_{1 \leq i \leq n} \min_{1 \leq j \leq d} |\Sigma_{i,j}| \geq \lambda^2 \text{ a.s.}$$

In this case, when $\sup_{i \geq 1} \mathbb{E}[|X_i|_\infty | \mathcal{F}_0] < \infty$ a.s., and the smallest eigenvalues of $\Sigma_1, \Sigma_2, \ldots$ are uniformly bounded away from zero, the bound is of order $O_{a.s.}( [\ln + d]^{5/4}n^{-1/4} )$. 


(2) Noticing that \( \ln(1 + x) \leq \sqrt{x} \) for \( x \geq 0 \), the second term on the right hand side of (3.4) can be further bounded by

\[
\frac{\ln_4 d^{3/4} \sqrt{n\beta}}{\sqrt{\gamma\bar{\sigma}/\bar{\lambda}}}.
\]

The latter quantity is similar to the corresponding term of the bound given in Theorem 2 in Section 9.3 of Chow and Teicher (1997) for scalar-valued martingales. The corresponding first term is, however, of order \( O(n^{-1/8}) \) under the conditions of part (1).

(3) The bound in (3.4) trivially applies to maxima of vector-valued martingales because for \( r \in \mathbb{R} \) and a random vector \( \xi \in \mathbb{R}^d, \{\xi \in A_{ri}\} = \{\max_{1 \leq j \leq d} \xi_j \leq r\} \), where \( i \) is a vector of ones, and therefore, letting \( M(\xi) := \max_{1 \leq j \leq d} \xi_j \),

\[
d_K(M(S), M(T) \mid \mathcal{F}_0) \leq d_K(S, T \mid \mathcal{F}_0) \quad \text{a.s.}
\]

References

Bolthausen, E., 1982. Exact convergence rates in some martingale central limit theorems. The Annals of Probability 10 (3), 672–688.

Chernozhukov, V., Chetverikov, D., Kato, K., 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high dimensional random vectors. Annals of Statistics 41 (6), 2786–2819.

Chernozhukov, V., Chetverikov, D., Kato, K., 2017. Central limit theorems and bootstrap in high dimensions. The Annals of Probability 45 (4), 2309–2352.

Chernozhukov, V., Chetverikov, D., Kato, K., Koike, Y., 2019. Improved central limit theorem and bootstrap approximations in high dimensions.

Chow, Y. S., Teicher, H., 1997. Probability Theory: Independence, Interchangeability, Martingales, 3rd Edition. Springer Texts in Statistics. Springer New York.

Fang, X., Koike, Y., 2021. High-dimensional central limit theorems by Stein’s method. The Annals of Applied Probability 31 (4), 1660–1686.

Grams, W. F., 1972. Rate of convergence in the central limit theorem for dependent variables. Ph.D. thesis, Florida State University.

Kir’yanova, L. V., Rotar’, V. I., 1991. Estimates for the rate of convergence in the central limit theorem for martingales. Theory of Probability & Its Applications 36 (2), 289–302.

Kuchibhotla, K., Rinaldo, A., 2020. High-dimensional CLT for sums of non-degenerate random vectors: \( n^{1/2} \)-rate.

Lopes, M. E., 2020. Central limit theorem and bootstrap approximation in high dimensions with near \( 1/\sqrt{n} \) rates.
Renz, J., 1996. A note on exact convergence rates in some martingale central limit theorems. The Annals of Probability 24 (3), 1616–1637.

Wu, S., Ma, X., Sang, H., Fan, X., 2020. A Berry–Esseen bound of order $\frac{1}{\sqrt{n}}$ for martingales. Comptes Rendus. Mathématique 358 (6), 701–712.