A NEW APPROACH TO STRONG EMBEDDINGS

SOURAV CHATTERJEE

ABSTRACT. We revisit strong approximation theory from a new perspective, culminating in a proof of the Komlós-Major-Tusnády embedding theorem for the simple random walk. The proof is almost entirely based on a series of soft arguments and easy inequalities. The new technique, inspired by Stein's method of normal approximation, is applicable to any setting where Stein's method works. In particular, one can hope to take it beyond sums of independent random variables.

1. INTRODUCTION

Let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. random variables with $\mathbb{E}(\varepsilon_1) = 0$ and $\mathbb{E}(\varepsilon_1^2) = 1$. For each k, let

$$S_k = \sum_{i=1}^{k} \varepsilon_i.$$

Suppose we want to construct a standard Brownian motion $(B_t)_{t \geq 0}$ on the same probability space so as to minimize the growth rate of

$$\max_{1 \leq k \leq n} |S_k - B_k|. \tag{1}$$

Since S_n and B_n both grow like $\sqrt{n}$, one would typically like to have the above quantity growing like $o(\sqrt{n})$, and preferably, as slowly as possible. This is the classical problem of coupling a random walk with a Brownian motion, usually called an 'embedding problem' because the most common approach is to start with a Brownian motion and somehow extract the random walk as a process embedded in the Brownian motion.

The study of such embeddings began with the works of Skorohod [19, 20] and Strassen [22], who showed that under the condition $\mathbb{E}(\varepsilon_1^4) < \infty$, it is possible to make (1) grow like $n^{1/4}(\log n)^{1/2}(\log \log n)^{1/4}$. In fact, this was shown to be the best possible rate under the finite fourth moment assumption by Kiefer [12].

For a long time, this remained the best available result in spite of numerous efforts by a formidable list of authors to improve on Skorohod's idea. For a detailed account of these activities, see the comprehensive recent survey of Obłój [16] and the bibliography of the monograph by Csörgő and

Révész [6]. Therefore it came as a great surprise when Komlós, Major, and Tusnády [13], almost fifteen years after Skorohod's original work, proved by a completely different argument that one can actually have

$$\max_{k \leq n} |S_k - B_k| = O(\log n)$$

when ε_1 has a finite moment generating function in a neighborhood of zero. Moreover, they showed that this is the best possible result that one can hope for in this situation.

Theorem 1.1 (Komlós-Major-Tusnády [13]). *Let* $\varepsilon_1, \varepsilon_2, \ldots$ *be i.i.d. random variables with* $\mathbb{E}(\varepsilon_1) = 0$, $\mathbb{E}(\varepsilon_1^2) = 1$, *and* $\mathbb{E} \exp \theta |\varepsilon_1| < \infty$ *for some* $\theta > 0$. *For each* k, *let* $S_k := \sum_{i=1}^{k} \varepsilon_i$. *Then for any* n, *it is possible to construct a version of* $(S_k)_{0 \leq k \leq n}$ *and a standard Brownian motion* $(B_t)_{0 \leq t \leq n}$ *on the same probability space such that for all* $x \geq 0$,

$$\mathbb{P}\big(\max_{k \leq n} |S_k - B_k| \geq C \log n + x\big) \leq K e^{-\lambda x},$$

where C, K, *and* λ *do not depend on* n.

The paper [13] also contains another very important result, a similar embedding theorem for uniform empirical processes. However, this will not be discussed in this article. See the recent articles by Mason [15] and Csörgő [4] as well as the book [5] for more on the KMT embedding theorem for empirical procceses.

One problem with the proof of Theorem 1.1, besides being technically difficult, is that it is very hard to generalize. Indeed, even the most basic extension to the case of non-identically distributed summands by Sakhanenko [17] is so complex that some researchers are hesitant to use it (see also Shao [18]). A nearly optimal multivariate version of the KMT theorem was proved by Einmahl [10]; the optimal result was obtained by Zaitsev [23] at the end of an extraordinary amount of hard work. More recently, Zaitsev has established multivariate versions of Sakhanenko's theorem [24, 25, 26]. For further details and references, let us refer to the survey article by Zaitsev [27] in the Proceedings of the ICM 2002.

The investigation in this paper is targeted towards a more conceptual understanding of the problem that may allow one to go beyond sums of independent random variables. It begins with the following abstract method of coupling an arbitrary random variable W with a Gaussian random variable Z so that $W - Z$ has exponentially decaying tails at the appropriate scale. (Such a coupling will henceforth be called a *strong coupling*, to distinguish it from the 'weak' couplings given by bounds on total variation or Wasserstein metrics.)

Theorem 1.2. *Suppose* W *is a random variable with* $\mathbb{E}(W) = 0$ *and finite second moment. Let* T *be another random variable, defined on the same probability space as* W, *such that whenever* φ *is a Lipschitz function and* φ'

is a derivative of φ a.e., we have

(2) $$\mathbb{E}(W\varphi(W)) = \mathbb{E}(\varphi'(W)T).$$

Suppose $|T|$ is almost surely bounded by a constant. Then, given any $\sigma^2 > 0$, we can construct $Z \sim N(0, \sigma^2)$ on the same probability space such that for any $\theta \in \mathbb{R}$,

$$\mathbb{E}\exp(\theta|W - Z|) \leq 2\,\mathbb{E}\exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2}\right).$$

Let us make a definition here, for the sake of convenience. Whenever (W, T) is a pair of random variables satisfying (2), we will say that T is a *Stein coefficient* for W.

The key idea, inspired by Stein's method of normal approximation [21], is that if $T \simeq \sigma^2$ with high probability, then one can expect that W is approximately Gaussian with mean zero and variance σ^2. This conclusion is heuristically justified because a random variable Z follows the $N(0, \sigma^2)$ distribution if and only if $\mathbb{E}(Z\varphi(Z)) = \sigma^2\mathbb{E}(\varphi'(Z))$ for all continuously differentiable φ such that $\mathbb{E}|\varphi'(Z)| < \infty$. Stein's method is a process of getting rigorous bounds out of this heuristic.

However, classical Stein's method can only give bounds on quantities like

$$\sup_{f \in \mathcal{F}} |\mathbb{E}f(W) - \mathbb{E}f(Z)|,$$

for various classes of functions $\mathcal{F}$. This includes, for example, bounds on the total variation distance and the Wasserstein distance, and the Berry-Esséen bounds. Theorem 1.2 seems to be of a fundamentally different nature.

To see how Stein coefficients can be constructed in a large array of situations, let us consider a few examples.

Example 1. Suppose X is a random variable with $\mathbb{E}(X) = 0$, $\mathbb{E}(X^2) < \infty$, and following a density ρ that is positive on an interval (bounded or unbounded) and zero outside. Let

(3) $$h(x) := \frac{\int_x^\infty y\rho(y)dy}{\rho(x)}$$

on the support of ρ. Then, assuming ideal conditions and applying integration by parts, we have $\mathbb{E}(X\varphi(X)) = \mathbb{E}(\varphi'(X)h(X))$ for all Lipschitz φ. Thus, $h(X)$ is a Stein coefficient for X. The above computation is carried out more precisely in Lemma 2.3 in Section 2.

Example 2. Suppose $X_1, \ldots, X_n$ are i.i.d. copies of the random variable X from the above example, and let $W = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$. Then by Example 1,

$$\mathbb{E}(W\varphi(W)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}(X_i\varphi(W))$$

$$= \frac{1}{n}\sum_{i=1}^n \mathbb{E}(h(X_i)\varphi'(W)) = \mathbb{E}\left(\varphi'(W)\frac{1}{n}\sum_{i=1}^n h(X_i)\right).$$

Thus, $\frac{1}{n}\sum_i h(X_i)$ is a Stein coefficient for W. Note that this becomes more and more like a constant as n increases, and so we can use Theorem 1.2 to get more and more accurate couplings.

Example 3. Suppose $\varepsilon_1,\ldots,\varepsilon_n$ are i.i.d. symmetric ± 1-valued r.v. Let $S_n = \sum_{i=1}^n \varepsilon_i$. Let $Y \sim \text{Uniform}[-1,1]$. Let $W_n = S_n + Y$. Let

$$T_n = n - S_n Y + \frac{1 - Y^2}{2}.$$

It will be shown in the proof of Theorem 3.1 in Section 3 that T_n is a Stein coefficient for W_n. (The construction of this T_n is somewhat ad hoc. The author has not yet found a general technique for smoothening of discrete random variables in a way that can automatically generate a Stein coefficient.) Letting $\sigma^2 = n$, Lemma 1.2 tells us that it is possible to construct $Z_n \sim N(0,n)$ such that

$$\mathbb{E}\exp(\theta|W_n - Z_n|) \le 2\,\mathbb{E}\exp\left(\frac{2\theta^2(T_n - n)^2}{n}\right).$$

Since $T_n = n + O(\sqrt{n})$ and $|W_n - S_n| \le 1$, it is now clear how to use Theorem 1.2 to construct S_n and Z_n on the same probability space such that irrespective of n,

$$\mathbb{E}\exp(\theta|S_n - Z_n|) \le C$$

for some fixed constants θ and C. By Markov's inequality, for all $x \ge 0$,

$$\mathbb{P}(|S_n - Z_n| \ge x) \le Ce^{-\theta x}.$$

This is the first step in our proof of the KMT embedding theorem for the simple random walk.

Example 4. Suppose $\mathbf{X} = (X_1,\ldots,X_n)$ is a vector of i.i.d. standard Gaussian random variables. Let $W = f(\mathbf{X})$, where f is absolutely continuous. Suppose $\mathbb{E}(W) = 0$. Let $\mathbf{X}' = (X_1',\ldots,X_n')$ be an independent copy of $\mathbf{X}$. Let

$$T = \int_0^1 \frac{1}{2\sqrt{t}}\sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{X})\frac{\partial f}{\partial x_i}(\sqrt{1-t}\mathbf{X} + \sqrt{t}\mathbf{X}')dt.$$

Then one can show that T is a Stein coefficient for W (see [3], Lemma 5.3). This has been used to prove CLTs for linear statistics of eigenvalues of random matrices [3].

Example 5. Theorem 1.2 can be used to construct strong couplings for sums of dependent random variables. An example of such a result is the following.

Theorem 1.3. *Suppose $X_1,\ldots,X_n,X_{n+1}$ are i.i.d. random variables with mean zero, variance 1, and probability density ρ. Suppose ρ is bounded above and below by positive constants on a compact interval, and zero outside. Let $S_n := \sum_{i=1}^n X_i X_{i+1}$. Then it is possible to construct S_n and a Gaussian*

random variable $Z_n \sim N(0,n)$ on the same probability space such that for all $x \geq 0$,

$$\mathbb{P}(|S_n - Z_n| \geq x) \leq e^{-C(\rho)x},$$

where $C(\rho)$ is a positive constant depending only on the density ρ (and not on n).

The process $\{S_n\}$, upon proper scaling, is sometimes called the 'autocorrelation process' for the sequence $\{X_n\}$. It may be possible to use the above result to prove a KMT type coupling for autocorrelation processes. The proof of Theorem 1.3 is short enough to be presented right here.

Proof of Theorem 1.3. Let $X_0 \equiv 0$. Let h be defined as in (3). Then note that for any φ, the definition of h and Example 1 show that

$$\mathbb{E}(S_n\varphi(S_n)) = \sum_{i=1}^{n} \mathbb{E}(X_i X_{i+1}\varphi(S_n))$$

$$= \sum_{i=1}^{n} \mathbb{E}(X_{i+1}(X_{i-1} + X_{i+1})h(X_i)\varphi'(S_n)).$$

This shows that if

$$D_i := h(X_i)X_{i+1}(X_{i-1} + X_{i+1}),$$

then $T_n := \sum_{i=1}^{n} D_i$ is a Stein coefficient for S_n. Now, for any $1 \leq i \leq n$,

$$\mathbb{E}(D_i - 1 \mid X_1, \ldots, X_{i-1}) = \mathbb{E}(h(X_i))\mathbb{E}(X_{i+1}^2) - 1 = 0,$$

since $\mathbb{E}(h(X_i)) = \mathbb{E}(X_i^2) = 1$. Moreover it is easy to show that by the assumed conditions on ρ that $|D_i|$ is almost surely bounded by a constant depending on ρ. Therefore by the Azuma-Hoeffding inequality [11, 1] for sums of bounded martingale differences, we get that for each $\alpha \in \mathbb{R}$,

$$\mathbb{E}(e^{\alpha(T_n-n)}) \leq e^{C_1(\rho)\alpha^2 n}$$

where $C_1(\rho)$ is some constant depending only on ρ. Thus if Z is a standard Gaussian random variable, independent of all else, then for any $\alpha \in \mathbb{R}$

$$\mathbb{E}(e^{\alpha Z(T_n-n)/\sqrt{n}}) \leq \mathbb{E}(e^{C_1(\rho)Z^2\alpha^2}).$$

Therefore choosing $\alpha = C_2(\rho)$ small enough, one gets

$$\mathbb{E}(e^{C_2(\rho)Z(T_n-n)/\sqrt{n}}) \leq 2.$$

On the other hand, first conditioning on T_n we get

$$\mathbb{E}(e^{C_2(\rho)Z(T_n-n)/\sqrt{n}}) = \mathbb{E}(e^{C_2(\rho)^2(T_n-n)^2/2n}).$$

By Theorem 1.2, this completes the proof. $\square$

Sketch of the proof of Theorem 1.2. (Full details are given in Section 2.) First, let $h(W) := \mathbb{E}(T|W)$. Then $h(W)$ is again a Stein coefficient for W. Moreover, one can show that the function h is non-negative a.e. on the

support of W. It is not difficult to verify that to prove Theorem 1.2 it suffices to construct a coupling such that for all θ,

$$\mathbb{E}\exp(\theta|W - Z|) \leq 2\,\mathbb{E}\exp\bigl(2\theta^2\bigl(\sqrt{h(W)} - \sigma\bigr)^2\bigr).$$

Fix a function $r : \mathbb{R}^2 \to \mathbb{R}$. For $f \in C^2(\mathbb{R}^2)$, let

$$\mathcal{L}f(x, y) := h(x)\frac{\partial^2 f}{\partial x^2} + 2r(x, y)\frac{\partial^2 f}{\partial x \partial y} + \sigma^2\frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$

Suppose there exists a probability measure μ on $\mathbb{R}^2$ such that for all f,

$$(4) \qquad\qquad \int_{\mathbb{R}^2} \mathcal{L}f\, d\mu = 0.$$

The main idea is as follows: *every* choice of r that admits a μ satisfying (4) gives a coupling of W and Z. Indeed, suppose μ is as above and (X, Y) is a random vector with law μ. Take any $\Phi \in C^2(\mathbb{R})$, and let $\varphi = \Phi'$. Putting $f(x, y) = \Phi(x)$ in (4) gives

$$\mathbb{E}(h(X)\varphi'(X)) = \mathbb{E}(X\varphi(X)).$$

Since this holds for all φ (which is a property that characterizes W) it is possible to argue that X must have the same law as W. Similarly, putting $f(x, y) = \Phi(y)$, we get $\mathbb{E}(Y\varphi(Y)) = \sigma^2\mathbb{E}(\varphi'(Y))$, and thus, $Y \sim N(0, \sigma^2)$. Note that this argument did not depend on the choice of r at all, except through the assumption that there exists a μ satisfying (4).

Now the question is, for what choices of r does there exist a μ satisfying (4)? In Lemma 2.1 it is proved that this is possible whenever the matrix

$$\begin{pmatrix} h(x) & r(x, y) \\ r(x, y) & \sigma^2 \end{pmatrix}$$

is positive semidefinite for all (x, y), plus some extra conditions. Note that this is the same as saying that the operator $\mathcal{L}$ is elliptic.

Intuitively, the 'best' coupling of W and Z is obtained when the choice of $r(x, y)$ is such that the matrix displayed above is the 'most singular'. This choice is given by the geometric mean

$$r(x, y) = \sigma\sqrt{h(x)}.$$

With this choice of r and $f(x, y) = \frac{1}{2k}(x - y)^{2k}$ (where k is an arbitrary positive integer), a small computation gives

$$\mathcal{L}f(x, y) = (2k - 1)(x - y)^{2k-2}\bigl(\sqrt{h(x)} - \sigma\bigr)^2 - (x - y)^{2k}.$$

Since (4) holds for this f, we get

$$\mathbb{E}(X - Y)^{2k} = (2k - 1)\mathbb{E}((X - Y)^{2k-2}(\sqrt{h(X)} - \sigma)^2)$$
$$\leq (2k - 1)(\mathbb{E}(X - Y)^{2k})^{(k-1)/k}(\mathbb{E}(\sqrt{h(X)} - \sigma)^{2k})^{1/k}.$$

This gives

$$\mathbb{E}(X - Y)^{2k} \leq (2k - 1)^k\mathbb{E}(\sqrt{h(X)} - \sigma)^{2k}.$$

It is now easy to complete the proof by combining over $k \geq 1$.

The KMT theorem for the SRW. As an application of Theorem 1.2, we give a new proof of Theorem 1.1 for the simple random walk. Although this is just a special case of the full theorem, it is important in its own right due to the importance of the SRW in various areas of science and mathematics. For instance, within the last ten years, the KMT embedding for the SRW played a pivotal role in the solution of a series of long-standing open questions about the simple random walk by the quartet of authors Dembo, Peres, Rosen, and Zeitouni [7, 8].

The proof of the KMT theorem for the SRW is obtained using a combination of Theorem 1.2, Example 3, and an induction argument. The induction step involves proving the following theorem about sums of exchangeable binary variables. This seems to be a new result.

Theorem 1.4. *There exist positive universal constants C, K and λ_0 such that the following is true. Take any integer $n \geq 2$. Suppose $\varepsilon_1, \ldots, \varepsilon_n$ are exchangeable ± 1 random variables. For $k = 0, 1, \ldots, n$, let $S_k = \sum_{i=1}^{k} \varepsilon_i$ and let $W_k = S_k - \frac{k}{n} S_n$. It is possible to construct a version of $W_0, \ldots, W_n$ and a standard Brownian bridge $(\widetilde{B}_t)_{0 \leq t \leq 1}$ on the same probability space such that for any $0 < \lambda < \lambda_0$,*

$$\mathbb{E} \exp(\lambda \max_{k \leq n} |W_k - \sqrt{n} \widetilde{B}_{k/n}|) \leq \exp(C \log n) \mathbb{E} \exp\left(\frac{K \lambda^2 S_n^2}{n}\right).$$

Note that by Example 2, it is possible to use Theorem 1.2 and induction whenever the summands have a density with respect to Lebesgue measure and the function h is reasonably well-behaved. This holds, for instance, for log-concave densities, or densities of the type considered in Theorem 1.3. In such cases it is not very difficult (although technically messier than the binary case) to prove a version of Theorem 1.4 using the method of this paper. However, we do not know yet how to use Theorem 1.2 to prove the KMT theorem in its full generality, because we do not know how to generalize the smoothing technique of Example 3.

The theorem that we prove about the KMT coupling for the SRW, stated below, is somewhat stronger than existing results.

Theorem 1.5. *Let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. symmetric ± 1-valued random variables. For each k, let $S_k := \sum_{i=1}^{k} \varepsilon_i$. It is possible to construct a version of the sequence $(S_k)_{k \geq 0}$ and a standard Brownian motion $(B_t)_{t \geq 0}$ on the same probability space such that for all n and all $x \geq 0$,*

$$\mathbb{P}\left(\max_{k \leq n} |S_k - B_k| \geq C \log n + x\right) \leq K e^{-\lambda x},$$

where C, K, and λ do not depend on n.

The above result is stronger than the corresponding statement about the SRW implied by Theorem 1.1 because it gives a single coupling for the whole process, instead of giving different couplings for different n. Such

results have been recently established in the KMT theorem for summands with finite pth moment [14, 28].

The paper is organized as follows. In Section 2, we prove Theorem 1.2. Two versions of Example 3 are worked out in Section 3. The main induction step, which proves Theorem 1.4, is carried out in Section 4. Finally, the proof of Theorem 1.5 is completed in Section 5.

2. Proof of Theorem 1.2

The proof will proceed as a sequence of lemmas. The lemmas will not be used in the subsequent sections, and only Theorem 1.2 is relevant for the future steps.

Lemma 2.1. *Let n be a positive integer, and suppose A is a continuous map from $\mathbb{R}^n$ into the set of $n \times n$ positive semidefinite matrices. Suppose there exists a constant $b \geq 0$ such that for all $x \in \mathbb{R}^n$,*

$$\|A(x)\| \leq b.$$

Then there exists a probability measure μ on $\mathbb{R}^n$ such that if X is a random vector following the law μ, then

$$(5) \qquad \qquad \mathbb{E}\exp\langle\theta, X\rangle \leq \exp(b\|\theta\|^2)$$

for all $\theta \in \mathbb{R}^n$, and

$$(6) \qquad \qquad \mathbb{E}\langle X, \nabla f(X)\rangle = \mathbb{E}\operatorname{Tr}(A(X)\operatorname{Hess} f(X))$$

for all $f \in C^2(\mathbb{R}^n)$ such that the expectations $\mathbb{E}|f(X)|^2$, $\mathbb{E}\|\nabla f(X)\|^2$, and $\mathbb{E}|\operatorname{Tr}(A(X)\operatorname{Hess} f(X))|$ are finite. Here ∇f and $\operatorname{Hess} f$ denote the gradient and Hessian of f, and Tr stands for the trace of a matrix.

Proof. Let K denote the set of all probability measures μ on $\mathbb{R}^n$ satisfying

$$\int x\mu(dx) = 0 \quad \text{and} \quad \int \exp\langle\theta, x\rangle\mu(dx) \leq \exp(b\|\theta\|^2) \text{ for all } \theta \in \mathbb{R}^n.$$

It is easy to see by the Skorokhod representation theorem and Fatou's lemma that K is a (nonempty) compact subset of the space V of all finite signed measures on $\mathbb{R}^n$ equipped with the topology of weak-* convergence (that is, the locally convex Hausdorff topology generated by the separating family of seminorms $|\mu|_f := |\int f d\mu|$, where f ranges over all continuous functions with compact support). Also, obviously, K is convex.

Now fix $\varepsilon \in (0, 1)$. Define a map $T_\varepsilon : K \to V$ as follows. Given $\mu \in K$, let X and Z be two independent random vectors, defined on some probability space, with $X \sim \mu$ and Z following the standard gaussian law on $\mathbb{R}^n$. Let $T_\varepsilon\mu$ be the law of the random vector

$$(1 - \varepsilon)X + \sqrt{2\varepsilon A(X)}Z,$$

where $\sqrt{A(X)}$ denotes the positive semidefinite square root of the matrix $A(X)$. Then for any $\theta \in \mathbb{R}^n$,

$$\int \exp\langle\theta, x\rangle T_\varepsilon\mu(dx) = \mathbb{E}\exp\big\langle\theta, (1-\varepsilon)X + \sqrt{2\varepsilon A(X)}Z\big\rangle$$

$$= \mathbb{E}\exp\big(\langle\theta, (1-\varepsilon)X\rangle + \varepsilon\langle\theta, A(X)\theta\rangle\big)$$

$$\le \exp(b\varepsilon\|\theta\|^2)\mathbb{E}\exp\langle\theta, (1-\varepsilon)X\rangle$$

$$\le \exp(b\varepsilon\|\theta\|^2 + b(1-\varepsilon)^2\|\theta\|^2).$$

For $\varepsilon \in (0,1)$, $1-\varepsilon+\varepsilon^2 \le 1$. Hence, $b\varepsilon+b(1-\varepsilon)^2 \le b$, and therefore T_ε maps K into K. Since A is a continuous map, and the transformation $A \mapsto \sqrt{A}$ is continuous (see e.g. [2], page 290, equation (X.2)), it is easy to see that T_ε is continuous under the weak-* topology. Hence, by the Schauder-Tychonoff fixed point theorem for locally convex topological vector spaces (see e.g. [9], Chapter V, 10.5), we see that T_ε must have a fixed point in K. For each $\varepsilon \in (0,1)$, let μ_ε be a fixed point of T_ε, and let X_ε denote a random vector following the law μ_ε.

Now take any $f \in C^2(\mathbb{R}^n)$ with ∇f and Hess f bounded and uniformly continuous. Fix $\varepsilon \in (0,1)$, and let

$$Y_\varepsilon = -\varepsilon X_\varepsilon + \sqrt{2\varepsilon A(X_\varepsilon)}Z.$$

By the definition of $T_\varepsilon\mu$, note that

$$\text{(7)} \qquad\qquad \mathbb{E}\big(f(X_\varepsilon + Y_\varepsilon) - f(X_\varepsilon)\big) = 0.$$

Now let

$$\mathcal{R}_\varepsilon = f(X_\varepsilon + Y_\varepsilon) - f(X_\varepsilon) - \big\langle Y_\varepsilon, \nabla f(X_\varepsilon)\big\rangle - \frac{1}{2}\big\langle Y_\varepsilon, \text{Hess } f(X_\varepsilon)\,Y_\varepsilon\big\rangle.$$

First, note that

$$\text{(8)} \qquad\qquad \mathbb{E}\big\langle Y_\varepsilon, \nabla f(X_\varepsilon)\big\rangle = -\varepsilon\mathbb{E}\big\langle X_\varepsilon, \nabla f(X_\varepsilon)\big\rangle.$$

By the definition of K, all moments of $\|X_\varepsilon\|$ are bounded by constants that do not depend on ε. Hence, as $\varepsilon \to 0$, we have

(9)

$$\mathbb{E}\big\langle Y_\varepsilon, \text{Hess } f(X_\varepsilon)\,Y_\varepsilon\big\rangle = 2\varepsilon\mathbb{E}\,\text{Tr}\big(\sqrt{A(X_\varepsilon)}\,\text{Hess } f(X_\varepsilon)\sqrt{A(X_\varepsilon)}\big) + O(\varepsilon^{3/2})$$

$$= 2\varepsilon\mathbb{E}\,\text{Tr}\big(A(X_\varepsilon)\,\text{Hess } f(X_\varepsilon)\big) + O(\varepsilon^{3/2}).$$

Now, by the boundedness and uniform continuity of Hess f, one can see that

$$|\mathcal{R}_\varepsilon| \le \|Y_\varepsilon\|^2\delta(\|Y_\varepsilon\|),$$

where $\delta : [0,\infty) \to [0,\infty)$ is a bounded function satisfying $\lim_{t\to 0}\delta(t) = 0$. Now, by the nature of K, it is easy to verify that the moments of $\varepsilon^{-1}\|Y_\varepsilon\|^2$ can be bounded by constants that do not depend on ε. Combining this with the above-mentioned properties of δ and the fact that $\|Y_\varepsilon\| \to 0$ in probability as $\varepsilon \to 0$, we get

$$\text{(10)} \qquad\qquad \lim_{\varepsilon\to 0}\varepsilon^{-1}\mathbb{E}|\mathcal{R}_\varepsilon| = 0.$$

Now let μ be a cluster point of the collection $\{\mu_\varepsilon\}_{0<\varepsilon<1}$ as $\varepsilon \to 0$, and let X denote a random variable following the law μ. Such a cluster point exists because K is a compact set. By uniform integrability, equations (7), (8), (9), (10), and the continuity of A, we get

$$\mathbb{E}\langle X, \nabla f(X) \rangle = \mathbb{E}\operatorname{Tr}(A(X)\operatorname{Hess} f(X)).$$

This completes the proof for $f \in C^2(\mathbb{R}^n)$ with ∇f and $\operatorname{Hess} f$ bounded and uniformly continuous. Next, take any $f \in C^2(\mathbb{R}^n)$. Let $g : \mathbb{R}^n \to [0,1]$ be a C^∞ function such that $g(x) = 1$ if $\|x\| \le 1$ and $g(x) = 0$ if $\|x\| \ge 2$. For each $a > 1$, let $f_a(x) = f(x)g(a^{-1}x)$. Then $f_a \in C^2$ with ∇f_a and $\operatorname{Hess} f_a$ bounded and uniformly continuous. Moreover, f_a and its derivatives converge pointwise to those of f as $a \to \infty$, as is seen from the expressions

$$\frac{\partial f_a}{\partial x_i} = \frac{\partial f}{\partial x_i}(x)g(a^{-1}x) + a^{-1}f(x)\frac{\partial g}{\partial x_i}(a^{-1}x),$$

$$\frac{\partial^2 f_a}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)g(a^{-1}x) + a^{-1}\frac{\partial f}{\partial x_i}(x)\frac{\partial g}{\partial x_j}(a^{-1}x)$$

$$+ a^{-1}\frac{\partial f}{\partial x_j}(x)\frac{\partial g}{\partial x_i}(a^{-1}x) + a^{-2}f(x)\frac{\partial^2 g}{\partial x_i \partial x_j}(a^{-1}x).$$

Since $\mathbb{E}\|X\|^2 < \infty$ and $\|A(x)\| \le b$, the above expressions also show that if the expectations $\mathbb{E}|f(X)|^2$, $\mathbb{E}\|\nabla f(X)\|^2$, and $\mathbb{E}|\operatorname{Tr}(A(X)\operatorname{Hess} f(X))|$ are finite, then we can apply the dominated convergence theorem to conclude that

$$\lim_{a\to\infty} \mathbb{E}\langle X, \nabla f_a(X) \rangle = \mathbb{E}\langle X, \nabla f(X) \rangle \text{ and}$$

$$\lim_{a\to\infty} \mathbb{E}\operatorname{Tr}(A(X)\operatorname{Hess} f_a(X)) = \mathbb{E}\operatorname{Tr}(A(X)\operatorname{Hess} f(X)).$$

This completes the proof.   □

Lemma 2.2. *Let A and X be as in Lemma* 2.1. *Take any $1 \le i < j \le n$. Let*

$$v_{ij}(x) := a_{ii}(x) + a_{jj}(x) - 2a_{ij}(x),$$

where a_{ij} denotes the (i,j)th element of A. Then for all $\theta \in \mathbb{R}$,

$$\mathbb{E}\exp(\theta|X_i - X_j|) \le 2\mathbb{E}\exp(2\theta^2 v_{ij}(X)).$$

Proof. Take any positive integer k. Define $f : \mathbb{R}^n \to \mathbb{R}$ as

$$f(x) := (x_i - x_j)^{2k}.$$

Then a simple calculation shows that

$$\langle x, \nabla f(x) \rangle = 2k(x_i - x_j)^{2k},$$

and

$$\operatorname{Tr}(A(x)\operatorname{Hess} f(x)) = 2k(2k-1)(x_i - x_j)^{2k-2}v_{ij}(x).$$

The positive definiteness of A shows that v_{ij} is everywhere nonnegative. An application of Hölder's inequality now gives

$$\mathbb{E}|\operatorname{Tr}(A(X)\operatorname{Hess} f(X))| \leq 2k(2k-1)\big(\mathbb{E}(X_i-X_j)^{2k}\big)^{\frac{k-1}{k}}\big(\mathbb{E}v_{ij}(X)^k\big)^{\frac{1}{k}}.$$

From the identity (6) we can now conclude that

$$\mathbb{E}(X_i-X_j)^{2k} \leq (2k-1)\big(\mathbb{E}(X_i-X_j)^{2k}\big)^{\frac{k-1}{k}}\big(\mathbb{E}v_{ij}(X)^k\big)^{\frac{1}{k}}.$$

This shows that

$$\mathbb{E}(X_i-X_j)^{2k} \leq (2k-1)^k\mathbb{E}v_{ij}(X)^k.$$

To complete the proof, note that

$$\mathbb{E}\exp(\theta|X_i-X_j|) \leq 2\mathbb{E}\cosh(\theta(X_i-X_j))$$
$$= 2\sum_{k=0}^{\infty}\frac{\theta^{2k}\mathbb{E}(X_i-X_j)^{2k}}{(2k)!}$$
$$\leq 2 + 2\sum_{k=1}^{\infty}\frac{(2k-1)^k\theta^{2k}\mathbb{E}(v_{ij}(X)^k)}{(2k)!}.$$

By the slightly crude but easy inequality

$$\frac{(2k-1)^k}{(2k)!} \leq \frac{2^k}{k!},$$

the proof is done. $\qquad\square$

Lemma 2.3. *Suppose ρ is a probability density function on $\mathbb{R}$ which is positive on an interval (bounded or unbounded) and zero outside. Suppose $\int_{-\infty}^{\infty} x\rho(x)dx = 0$. For each x in the support of ρ, let*

$$h(x) := \frac{\int_x^{\infty} y\rho(y)dy}{\rho(x)}.$$

Outside the support, let $h \equiv 0$. Let X be a random variable with density ρ and finite second moment. Then

$$(11) \qquad\qquad \mathbb{E}(X\varphi(X)) = \mathbb{E}(h(X)\varphi'(X))$$

for each absolutely continuous φ such that both sides are well defined and $\mathbb{E}|h(X)\varphi(X)| < \infty$. Moreover, if h_1 is another function satisfying (11) for all Lipschitz φ, then $h_1 = h$ a.e. on the support of ρ.

Conversely, if Y is a random variable such that (11) holds with Y in place of X, for all φ such that $|\varphi(x)|$, $|x\varphi(x)|$, and $|h(x)\varphi'(x)|$ are uniformly bounded, then Y must have the density ρ.

Proof. Let $u(x) = h(x)\rho(x)$. Note that u is continuous, positive on the support of ρ, and $\lim_{x\to-\infty} u(x) = \lim_{x\to\infty} u(x) = 0$ since

$$u(x) = \int_x^{\infty} y\rho(y)dy = -\int_{-\infty}^x y\rho(y)dy.$$

Note that the above identity holds because $\int_{-\infty}^{\infty} x\rho(x)dx = 0$. Again, by the assumption that $\mathbb{E}(X^2) < \infty$, it is easy to verify that

$$\mathbb{E}(h(X)) = \int_{-\infty}^{\infty} u(x)dx = \mathbb{E}(X^2) < \infty.$$

When φ is a bounded Lipschitz function, then (11) is just the integration by parts identity

$$\int_{-\infty}^{\infty} x\varphi(x)\rho(x)dx = \int_{-\infty}^{\infty} \varphi'(x)u(x)dx.$$

Now take any absolutely continuous φ and a C^∞ map $g : \mathbb{R} \to [0,1]$ such that $g(x) = 1$ on $[-1,1]$ and $g(x) = 0$ outside $[-2,2]$. For each $a > 1$, let

$$\varphi_a(x) := \varphi(x)g(a^{-1}x).$$

Then

$$\varphi_a'(x) = \varphi'(x)g(a^{-1}x) + a^{-1}\varphi(x)g'(a^{-1}x).$$

It is easy to see that φ_a and φ_a' are bounded, and they converge to φ and φ' pointwise as $a \to \infty$. Moreover, $|x\varphi_a(x)| \le |x\varphi(x)|$ and

$$|h(x)\varphi_a'(x)| \le |h(x)\varphi'(x)| + a^{-1}\|g'\|_\infty|h(x)\varphi(x)|.$$

Since we have assumed that $\mathbb{E}|X\varphi(X)|$, $\mathbb{E}|h(X)\varphi'(X)|$, and $\mathbb{E}|h(X)\varphi(X)|$ are finite, we can now apply the dominated convergence theorem to conclude that (11) holds for φ.

Suppose h_1 is another function satisfying (11) for all Lipschitz φ and $\mathbb{E}(X^2) < \infty$. Let $\varphi(x)$ be a Lipschitz function such that $\varphi'(x) = \text{sign}(h_1(x) - h(x))$. Then

$$0 = \mathbb{E}(\varphi'(X)(h_1(X) - h(X))) = \mathbb{E}|h_1(X) - h(X)|.$$

This shows that $h_1 = h$ a.e. on the support of ρ.

For the converse, let X have density ρ and take any bounded continuous function $v : \mathbb{R} \to \mathbb{R}$, let $m = \mathbb{E}v(X)$, and define

$$\varphi(x) := \frac{1}{u(x)}\int_{-\infty}^{x} \rho(y)(v(y) - m)dy = -\frac{1}{u(x)}\int_{x}^{\infty} \rho(y)(v(y) - m)dy$$

on the support of ρ. Since u is nonzero and absolutely continuous everywhere on the support of ρ, therefore φ is well-defined and absolutely continuous. Next, we prove that $|x\varphi(x)|$ is uniformly bounded. If $x \ge 0$, then

$$|x\varphi(x)| = \left| \frac{x}{u(x)} \int_{x}^{\infty} \rho(y)(v(y) - m)dy \right|$$

$$\le \frac{2\|v\|_\infty}{|u(x)|} \int_{x}^{\infty} y\rho(y)dy = 2\|v\|_\infty.$$

Similarly, the same bound holds for $x < 0$. A direct verification shows that

$$h(x)\varphi'(x) - x\varphi(x) = v(x) - m.$$

Thus, $|h(x)\varphi'(x)|$ is uniformly bounded. Finally, by the continuity of φ, $|\varphi(x)| \leq \sup_{|t| \leq 1} |\varphi(t)| + |x\varphi(x)|$ is also uniformly bounded.

So, if Y is a random variable such that (11) holds for Y in place of X and every φ such that $|\varphi(x)|$, $|x\varphi(x)|$, and $|h(x)\varphi'(x)|$ are uniformly bounded, then

$$\mathbb{E}v(Y) - \mathbb{E}v(X) = \mathbb{E}(v(Y) - m) = \mathbb{E}(h(Y)\varphi'(Y) - Y\varphi(Y)) = 0.$$

Thus, Y must have the same distribution as X. $\qquad\square$

Proof of Theorem 1.2. First, assume W has a density ρ with respect to Lebesgue measure which is positive and continuous everywhere. Define h in terms of ρ as in the statement of Lemma 2.3. Then by the second assertion of Lemma 2.3,

$$h(w) = \mathbb{E}(T|W = w) \text{ a.s.}$$

Note that h is nonnegative by definition. So we can define a function A from $\mathbb{R}^2$ into the set of 2×2 positive semidefinite matrices as

$$A(x_1, x_2) := \begin{pmatrix} h(x_1) & \sigma\sqrt{h(x_1)} \\ \sigma\sqrt{h(x_1)} & \sigma^2 \end{pmatrix}.$$

Note that $A(x_1, x_2)$ does not depend on x_2 at all. It is easy to see that A is positive semidefinite. Also, since ρ is assumed to be continuous, therefore so are h and A. Since T is bounded by a constant, so is h. Let $X = (X_1, X_2)$ be a random vector satisfying (5) and (6) of Lemma 2.1 with this A. Take any absolutely continuous $\varphi : \mathbb{R} \to \mathbb{R}$ such that $|\varphi(x)|$, $|x\varphi(x)|$, and $|h(x)\varphi'(x)|$ are uniformly bounded. Let Φ denote an antiderivative of φ, i.e. a function such that $\Phi' = \varphi$. We can assume that $\Phi(0) = 0$. Define $f : \mathbb{R}^2 \to \mathbb{R}$ as $f(x_1, x_2) := \Phi(x_1)$. Then for some constant C, for all x_1, x_2,

$$|f(x_1, x_2)| \leq C|x_1|, \ \|\nabla f(x_1, x_2)\| \leq C,$$
$$\text{and} \ |\operatorname{Tr}(A(x_1, x_2)\operatorname{Hess} f(x_1, x_2))| \leq C.$$

Thus, we can apply Lemma 2.1 to conclude that for this f,

$$\mathbb{E}\langle X, \nabla f(X)\rangle = \mathbb{E}\operatorname{Tr}(A(X)\operatorname{Hess} f(X)),$$

which can be written as

$$\mathbb{E}(X_1\varphi(X_1)) = \mathbb{E}(h(X_1)\varphi'(X_1)).$$

Since this holds for all φ such that $|\varphi(x)|$, $|x\varphi(x)|$, and $|h(x)\varphi'(x)|$ are uniformly bounded, Lemma 2.3 tells us that X_1 must have the same distribution as W.

Similarly, taking any φ such that $|\varphi(x)|$, $|x\varphi(x)|$, and $|\varphi'(x)|$ are uniformly bounded, letting Φ be an antiderivative of φ, and putting $f(x_1, x_2) = \Phi(x_2)$, we see that

$$\mathbb{E}(X_2\varphi(X_2)) = \sigma^2\mathbb{E}(\varphi'(X_2)),$$

which implies that $X_2 \sim N(0, \sigma^2)$. We now wish to apply Lemma 2.2 to the pair (X_1, X_2). Note that

$$v_{12}(x_1, x_2) = h(x_1) + \sigma^2 - 2\sigma\sqrt{h(x_1)} = \left(\sqrt{h(x_1)} - \sigma\right)^2$$

Since $h(x_1) \geq 0$, we have

$$\left(\sqrt{h(x_1)} - \sigma\right)^2 = \frac{\left(h(x_1) - \sigma^2\right)^2}{\left(\sqrt{h(x_1)} + \sigma\right)^2} \leq \frac{\left(h(x_1) - \sigma^2\right)^2}{\sigma^2}.$$

Since $h(X_1)$ has the same distribution as $h(W)$, and $h(W) = \mathbb{E}(T|W)$, the required bound can now be obtained using Lemma 2.2 and Jensen's inequality.

So we have finished the proof when W has a probability density ρ with respect to Lebesgue measure which is positive and continuous everywhere. Let us now drop that assumption, but keep all others. For each $\varepsilon > 0$, let $W_\varepsilon := W + \varepsilon Y$, where Y is an independent standard gaussian random variable. If ν denotes the law of W on the real line, then W_ε has the probability density function

$$\rho_\varepsilon(x) = \int_{-\infty}^\infty \frac{e^{-(x-y)^2/2\varepsilon^2}}{\sqrt{2\pi}\varepsilon} d\nu(y).$$

From the above representation, it is easy to deduce that ρ_ε is positive and continuous everywhere. Again, note that for any Lipschitz φ,

$$\begin{aligned}
\mathbb{E}(W_\varepsilon \varphi(W_\varepsilon)) &= \mathbb{E}(W\varphi(W + \varepsilon Y)) + \varepsilon\mathbb{E}(Y\varphi(W + \varepsilon Y)) \\
&= \mathbb{E}(T\varphi'(W + \varepsilon Y)) + \varepsilon^2\mathbb{E}(\varphi'(W + \varepsilon Y)) \\
&= \mathbb{E}((T + \varepsilon^2)\varphi'(W_\varepsilon)).
\end{aligned}$$

(Note that in the second step, we required that (2) holds for *any* derivative of φ instead of just one.) Thus, by what we have already proved, we can construct a version of W_ε and a $N(0, \sigma^2 + \varepsilon^2)$ r.v. Z_ε on the same probability space such that for all θ,

$$\mathbb{E}\exp(\theta|W_\varepsilon - Z_\varepsilon|) \leq 2\mathbb{E}\exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2 + \varepsilon^2}\right).$$

Let μ_ε be the law of the pair $(W_\varepsilon, Z_\varepsilon)$ on $\mathbb{R}^2$. Clearly, $\{\mu_\varepsilon\}_{\varepsilon > 0}$ is a tight family. Let μ_0 be a cluster point as $\varepsilon \to 0$, and let $(W_0, Z_0) \sim \mu_0$. Then W_0 has the same distribution as W, and $Z_0 \sim N(0, \sigma^2)$. By the Skorokhod representation, Fatou's lemma, and the monotone convergence theorem, it is clear that

$$\mathbb{E}\exp(\theta|W_0 - Z_0|) \leq \liminf_{\varepsilon \to 0} \mathbb{E}\exp(\theta|W_\varepsilon - Z_\varepsilon|) \leq 2\mathbb{E}\exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2}\right).$$

This completes the proof. $\qquad\square$

3. ELABORATIONS ON EXAMPLE 3

The goal of this section is to prove the following two theorems. The first one is simply Example 3 from Section 1. The second one can be called a conditional version of the same thing (which is harder to prove).

Theorem 3.1. *There exist universal constants κ and $\theta_0 > 0$ such that the following is true. Let n be a positive integer and let $\varepsilon_1, \ldots, \varepsilon_n$ be i.i.d. symmetric ± 1 random variables. Let $S_n = \sum_{i=1}^n \varepsilon_i$. It is possible to construct a version of S_n and $Z_n \sim N(0, n)$ on the same probability space such that*

$$\mathbb{E} \exp(\theta_0 |S_n - Z_n|) \leq \kappa.$$

Note that by Markov's inequality, this implies exponentially decaying tails for $|S_n - Z_n|$, with a rate of decay that does not depend on n.

Theorem 3.2. *Let $\varepsilon_1, \ldots, \varepsilon_n$ be n arbitrary elements of $\{-1, 1\}$. Let π be a uniform random permutation of $\{1, \ldots, n\}$. For each $1 \leq k \leq n$, let $S_k = \sum_{\ell=1}^k \varepsilon_{\pi(\ell)}$, and let*

$$W_k = S_k - \frac{kS_n}{n}.$$

There exist universal constants $c > 1$ and $\theta_0 > 0$ satisfying the following. Take any $n \geq 3$, any possible value of S_n, and any $n/3 \leq k \leq 2n/3$. It is possible to construct a version of W_k and a gaussian random variable Z_k with mean 0 and variance $k(n-k)/n$ on the same probability space such that for any $\theta \leq \theta_0$,

$$\mathbb{E} \exp(\theta |W_k - Z_k|) \leq \exp\left(1 + \frac{c\theta^2 S_n^2}{n}\right).$$

Both of the above theorems will be proved using Theorem 1.2. We proceed as before in a sequence of lemmas that are otherwise irrelevant for the rest of the manuscript (except Lemma 3.5, which has an important application later on).

Lemma 3.3. *Suppose X and Y are two independent random variables, with X following the symmetric distribution on $\{-1, 1\}$ and Y following the uniform distribution on $[-1, 1]$. Then for any Lipschitz φ, we have*

$$\mathbb{E}(X\varphi(X+Y)) = \mathbb{E}((1 - XY)\varphi'(X+Y)),$$

and

$$\mathbb{E}(Y\varphi(X+Y)) = \frac{1}{2}\mathbb{E}((1 - Y^2)\varphi'(X+Y)).$$

Proof. We have

$$\mathbb{E}((1 - XY)\varphi'(X+Y)) = \frac{1}{4}\int_{-1}^{1}(1+y)\varphi'(-1+y)dy$$
$$+ \frac{1}{4}\int_{-1}^{1}(1-y)\varphi'(1+y)dy.$$

Integrating by parts, we see that

$$\int_{-1}^{1} (1+y)\varphi'(-1+y)dy = 2\varphi(0) - \int_{-1}^{1} \varphi(-1+y)dy,$$

and

$$\int_{-1}^{1} (1-y)\varphi'(1+y)dy = -2\varphi(0) + \int_{-1}^{1} \varphi(1+y)dy.$$

Adding up, we get

$$\mathbb{E}((1-XY)\varphi'(X+Y)) = \frac{1}{4}\int_{-1}^{1} \varphi(1+y)dy - \frac{1}{4}\int_{-1}^{1} \varphi(-1+y)dy$$
$$= \mathbb{E}(X\varphi(X+Y)).$$

For the second part, just observe that for any x, integration by parts gives

$$\frac{1}{2}\int_{-1}^{1} y\varphi(x+y)dy = \frac{1}{2}\int_{-1}^{1} \frac{1-y^2}{2}\varphi'(x+y)dy.$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Proof of Theorem 3.1. For simplicity, let us write S for S_n. Let Y be a random variable independent of $\varepsilon_1,\dots,\varepsilon_n$ and uniformly distributed on the interval $[-1,1]$. Suppose we are given the values of $\varepsilon_1,\dots,\varepsilon_{n-1}$. Let $\mathbb{E}^-$ denote the conditional expectation given this information. Let

$$S^- = \sum_{i=1}^{n-1} \varepsilon_i, \quad X = \varepsilon_n.$$

Then Lemma 3.3 gives

$$\mathbb{E}^-(X\varphi(S^- + X + Y)) = \mathbb{E}^-((1-XY)\varphi'(S+Y))$$
$$= \mathbb{E}^-((1-\varepsilon_n Y)\varphi'(S+Y)).$$

Taking expectation on both sides we get

$$\mathbb{E}(\varepsilon_n\varphi(S+Y)) = \mathbb{E}((1-\varepsilon_n Y)\varphi'(S+Y)).$$

By symmetry, this gives

$$\mathbb{E}(S\varphi(S+Y)) = \mathbb{E}((n-SY)\varphi'(S+Y)).$$

Again, by Lemma 3.3, we have

$$\mathbb{E}(Y\varphi(S+Y)) = \frac{1}{2}\mathbb{E}((1-Y^2)\varphi'(S+Y)).$$

Thus, putting $\widetilde{S} = S+Y$ and

$$T = n - SY + \frac{1-Y^2}{2},$$

we have

(12) $$\mathbb{E}(\widetilde{S}\varphi(\widetilde{S})) = \mathbb{E}(T\varphi'(\widetilde{S})).$$

Let $\sigma^2 = n$. Then

$$\frac{(T - \sigma^2)^2}{\sigma^2} \leq \frac{2S^2 + \frac{1}{2}}{n}.$$

Now, clearly, $\mathbb{E}(\widetilde{S}) = 0$ and $\mathbb{E}(\widetilde{S}^2) < \infty$. The equation (12) holds and the random variable T is a.s. bounded. Therefore, all conditions for applying Theorem 1.2 to $\widetilde{S}$ are met, and hence we can conclude that it is possible to construct a version of $\widetilde{S}$ and a $N(0, \sigma^2)$ random variable Z on the same space such that for all θ,

$$\mathbb{E}\exp(\theta|\widetilde{S} - Z|) \leq 2\mathbb{E}\exp(2\theta^2\sigma^{-2}(T - \sigma^2)^2).$$

Since the value of S is determined if we know $\widetilde{S}$, we can now construct a version of S on the same probability space satisfying $|S - \widetilde{S}| \leq 1$. It follows that

$$\mathbb{E}\exp(\theta|S - Z|) \leq 2\mathbb{E}\exp(|\theta| + 2\theta^2\sigma^{-2}(T - \sigma^2)^2).$$

Using the bound on $(T - \sigma^2)^2/\sigma^2$ obtained above, we have

$$\mathbb{E}\exp(\theta|S - Z|) \leq 2\exp(|\theta| + \theta^2/n)\mathbb{E}\exp(4\theta^2 S^2/n).$$

To complete the argument, note that if V is a standard gaussian r.v., independent of S, then

$$\begin{aligned}
\mathbb{E}\exp(4\theta^2 S^2/n) &= \mathbb{E}\exp(\sqrt{8}\theta V S/\sqrt{n}) \\
&= \mathbb{E}(\mathbb{E}(\exp(\sqrt{8}\theta V \varepsilon_1/\sqrt{n})|V)^n) \\
&= \mathbb{E}(\cosh^n(\sqrt{8}\theta V/\sqrt{n})).
\end{aligned}$$

Using the simple inequality $\cosh x \leq \exp x^2$, this gives

$$(13) \qquad \mathbb{E}\exp(4\theta^2 S^2/n) \leq \mathbb{E}\exp(8\theta^2 V^2) = \frac{1}{\sqrt{1 - 16\theta^2}} \quad \text{if } 16\theta^2 < 1.$$

The conclusion now follows by choosing θ_0 sufficiently small. $\square$

Lemma 3.4. *Let all notation be as in the statement of Theorem 3.2. Then for any $\theta \in \mathbb{R}$ and any $1 \leq k \leq n$, we have*

$$\mathbb{E}\exp(\theta W_k/\sqrt{k}) \leq \exp\theta^2.$$

Remark. Note that the bound does not depend on the value of S_n. This is crucial for the next lemma and the induction step later on. Heuristically, this phenomenon is not mysterious because the centered process $(W_k)_{k \leq n}$ has maximum freedom to fluctuate when $S_n = 0$.

Proof. Fix k, and let $m(\theta) := \mathbb{E}\exp(\theta W_k/\sqrt{k})$. Since W_k is a bounded random variable, there is no problem in showing that m is differentiable and

$$m'(\theta) = \frac{1}{\sqrt{k}}\mathbb{E}(W_k\exp(\theta W_k/\sqrt{k})).$$

Now note that

$$\frac{1}{n}\sum_{i=1}^{k}\sum_{j=k+1}^{n}\left(\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)}\right) = \frac{(n-k)\sum_{i=1}^{k}\varepsilon_{\pi(i)} - k\sum_{j=k+1}^{n}\varepsilon_{\pi(j)}}{n}$$

$$= \frac{(n-k)\sum_{i=1}^{k}\varepsilon_{\pi(i)} - k(S_n - \sum_{i=1}^{k}\varepsilon_{\pi(i)})}{n}$$

$$= \sum_{i=1}^{k}\varepsilon_{\pi(i)} - \frac{kS_n}{n} = W_k.$$

Thus,

$$(14) \qquad m'(\theta) = \frac{1}{n\sqrt{k}}\sum_{i=1}^{k}\sum_{j=k+1}^{n}\mathbb{E}((\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})\exp(\theta W_k/\sqrt{k})).$$

Now fix $i \leq k < j$. Let $\pi' = \pi \circ (i,j)$, so that $\pi'(i) = \pi(j)$ and $\pi'(j) = \pi(i)$. Then π' is again uniformly distributed on the set of all permutations of $\{1,\ldots,n\}$. Moreover, (π, π') is an *exchangeable pair* of random variables. Let

$$W_k' = \sum_{\ell=1}^{k}\varepsilon_{\pi'(\ell)} - \frac{kS_n}{n}.$$

Then

$$\mathbb{E}((\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})\exp(\theta W_k/\sqrt{k})) = \mathbb{E}((\varepsilon_{\pi'(i)} - \varepsilon_{\pi'(j)})\exp(\theta W_k'/\sqrt{k}))$$

$$= \mathbb{E}((\varepsilon_{\pi(j)} - \varepsilon_{\pi(i)})\exp(\theta W_k'/\sqrt{k})).$$

Averaging the two equal quantities, we get

$$\mathbb{E}((\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})\exp(\theta W_k/\sqrt{k}))$$

$$= \frac{1}{2}\mathbb{E}((\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})(\exp(\theta W_k/\sqrt{k}) - \exp(\theta W_k'/\sqrt{k}))).$$

Thus, from the inequality

$$|e^x - e^y| \leq \frac{1}{2}|x - y|(e^x + e^y)$$

and the fact that $W_k - W_k' = \varepsilon_{\pi(i)} - \varepsilon_{\pi(j)}$, we get

$$\left|\mathbb{E}((\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})\exp(\theta W_k/\sqrt{k}))\right|$$

$$\leq \frac{|\theta|}{4\sqrt{k}}\mathbb{E}((\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})^2(\exp(\theta W_k/\sqrt{k}) + \exp(\theta W_k'/\sqrt{k})))$$

$$\leq \frac{|\theta|}{\sqrt{k}}\mathbb{E}(\exp(\theta W_k/\sqrt{k}) + \exp(\theta W_k'/\sqrt{k}))$$

$$= \frac{2|\theta|}{\sqrt{k}}\mathbb{E}\exp(\theta W_k/\sqrt{k}) = \frac{2|\theta|}{\sqrt{k}}m(\theta).$$

Using this estimate in (14), we get

$$|m'(\theta)| \leq \frac{2|\theta|}{nk} \sum_{i=1}^{k} \sum_{j=k+1}^{n} m(\theta) \leq 2|\theta|m(\theta).$$

Using that $m(0) = 1$, it is now easy to complete the proof. $\qquad \square$

Lemma 3.5. *Let us continue with the notation of Theorem* 3.2. *There exists a universal constant $\alpha_0 > 0$ such that for all n, all possible values of S_n, all k such that $k \leq 2n/3$, and all $\alpha \leq \alpha_0$, we have*

$$\mathbb{E} \exp(\alpha S_k^2/k) \leq \exp\left(1 + \frac{3\alpha S_n^2}{4n}\right).$$

Remark. The exact value of the constant $3/4$ in the above bound is not important; what is important is that the constant is < 1 as long as we take $k \leq 2n/3$. This is why the induction argument can be carried out in Section 4. However, there is no mystery; the fact that one can always get a constant < 1 can be explained via simple heuristic arguments once Lemma 3.4 is known.

Proof. Let Z be an independent standard gaussian random variable. Then

$$\mathbb{E} \exp(\alpha S_k^2/k) = \mathbb{E} \exp\left(\sqrt{\frac{2\alpha}{k}} Z S_k\right)$$
$$= \mathbb{E} \exp\left(\sqrt{\frac{2\alpha}{k}} Z W_k + \sqrt{\frac{2\alpha}{k}} \frac{k S_n}{n} Z\right).$$

Now, by Lemma 3.4 we have

$$\mathbb{E}\left(\exp\left(\sqrt{\frac{2\alpha}{k}} Z W_k\right)\Big| Z\right) \leq \exp(2\alpha Z^2).$$

Thus, we have

$$\mathbb{E} \exp(\alpha S_k^2/k) \leq \mathbb{E} \exp\left(2\alpha Z^2 + \sqrt{\frac{2\alpha}{k}} \frac{k S_n}{n} Z\right).$$

Since S_n is nonrandom, the right hand side is just the expectation of a function of a standard gaussian random variable, which can be easily computed. This gives, for $0 < \alpha < 1/4$,

$$\mathbb{E} \exp(\alpha S_k^2/k) \leq \frac{1}{\sqrt{1 - 4\alpha}} \exp\left(\frac{\alpha k S_n^2}{(1 - 4\alpha)n^2}\right).$$

The lemma is now proved by bounding k by $2n/3$ and choosing α_0 small enough to ensure that $1/(1 - 4\alpha_0)$ is sufficiently close to 1. $\qquad \square$

Proof of Theorem 3.2. For simplicity, we shall write W for W_k and S for S_n, but S_k will be written as usual.

Let Y be a random variable independent of π and uniformly distributed on the interval $[-1, 1]$. Fix $1 \leq i \leq k$ and $k < j \leq n$. Suppose we are given

the values of $\{\pi(\ell), \ell \neq i, j\}$. Let $\mathbb{E}^-$ denote the conditional expectation given this information. Let

$$S^- = \sum_{\ell \neq i,j} \varepsilon_{\pi(\ell)}, \quad W^- = \sum_{\ell \leq k, \ell \neq i} \varepsilon_{\pi(\ell)} - \frac{kS}{n}.$$

If $S \neq S^-$, then we must have $\varepsilon_{\pi(i)} = \varepsilon_{\pi(j)}$, and hence in that case

$$\mathbb{E}^-((\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})\varphi(W + Y)) = 0.$$

Next let us consider the only other possible scenario, $S = S^-$. Then the conditional distribution of $\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)}$ is symmetric over $\{-2, 2\}$. Let

$$X = \frac{\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)}}{2} = \varepsilon_{\pi(i)},$$

and note that

$$W = W^- + X.$$

Thus, under $S = S^-$, Lemma 3.3 shows that for all Lipschitz φ,

$$\begin{aligned}
\mathbb{E}^-((\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})\varphi(W + Y)) &= 2\mathbb{E}^-(X\varphi(W^- + X + Y)) \\
&= 2\mathbb{E}^-((1 - XY)\varphi'(W + Y)) \\
&= \mathbb{E}^-((2 - (\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})Y)\varphi'(W + Y)).
\end{aligned}$$

Next, let

$$a_{ij} := 1 - \varepsilon_{\pi(i)}\varepsilon_{\pi(j)} - (\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})Y.$$

A simple verification shows that

$$a_{ij} = \begin{cases} 2 - (\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})Y & \text{if } \varepsilon_{\pi(i)} \neq \varepsilon_{\pi(j)} \\ 0 & \text{if } \varepsilon_{\pi(i)} = \varepsilon_{\pi(j)}. \end{cases}$$

Thus, irrespective of whether $S = S^-$ or $S \neq S^-$, we have

$$\mathbb{E}^-((\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)})\varphi(W + Y)) = \mathbb{E}^-(a_{ij}\varphi'(W + Y)).$$

Clearly, we can now replace $\mathbb{E}^-$ by $\mathbb{E}$ in the above expression. Now, as in the proof of Lemma 3.4, observe that

$$W = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=k+1}^{n}(\varepsilon_{\pi(i)} - \varepsilon_{\pi(j)}).$$

Combining the last two observations, we have

$$\mathbb{E}(W\varphi(W + Y)) = \mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{k}\sum_{j=k+1}^{n}a_{ij}\right)\varphi'(W + Y)\right).$$

Again, by Lemma 3.3, we have

$$\mathbb{E}(Y\varphi(W + Y)) = \frac{1}{2}\mathbb{E}((1 - Y^2)\varphi'(W + Y)).$$

Thus, putting $\widetilde{W} = W + Y$ and

$$T = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=k+1}^{n} a_{ij} + \frac{1 - Y^2}{2},$$

we have

(15) $$\mathbb{E}(\widetilde{W}\varphi(\widetilde{W})) = \mathbb{E}(T\varphi'(\widetilde{W})).$$

Now

$$\frac{1}{n} \sum_{i=1}^{k} \sum_{j=k+1}^{n} a_{ij} = \frac{k(n-k)}{n} - \frac{(\sum_{i=1}^{k} \varepsilon_{\pi(i)})(\sum_{j=k+1}^{n} \varepsilon_{\pi(j)})}{n} - WY.$$

Let $\sigma^2 = k(n-k)/n$. Since $n/3 \le k \le 2n/3$ and $|W| \le |S_k| + \frac{2}{3}|S|$, a simple computation gives

$$\frac{(T - \sigma^2)^2}{\sigma^2} \le \frac{n}{k(n-k)}(|S_k| + |W| + 1/2)^2$$

$$\le C\left(\frac{S_k^2}{k} + \frac{S^2}{n} + 1\right),$$

where C is a universal constant.

Now, clearly, $\mathbb{E}(\widetilde{W}) = 0$ and $\mathbb{E}(\widetilde{W}^2) < \infty$. The equation (15) holds and the random variable T is a.s. bounded. Therefore, all conditions for applying Theorem 1.2 to $\widetilde{W}$ are met, and hence we can conclude that it is possible to construct a version of $\widetilde{W}$ and a $N(0, \sigma^2)$ random variable Z on the same space such that for all θ,

$$\mathbb{E}\exp(\theta|\widetilde{W} - Z|) \le 2\mathbb{E}\exp(2\theta^2\sigma^{-2}(T - \sigma^2)^2).$$

Since the value of W is determined if we know $\widetilde{W}$, we can now construct a version of W on the same probability space satisfying $|W - \widetilde{W}| \le 1$. It follows that

$$\mathbb{E}\exp(\theta|W - Z|) \le 2\mathbb{E}\exp(|\theta| + 2\theta^2\sigma^{-2}(T - \sigma^2)^2).$$

Using the bound on $(T - \sigma^2)^2/\sigma^2$ obtained above, we have

$$\mathbb{E}\exp(\theta|W - Z|) \le 2\exp(|\theta| + C\theta^2 S^2/n + C\theta^2)\mathbb{E}\exp(C\theta^2 S_k^2/k),$$

where, again, C is a universal constant. The conclusion now follows from Lemma 3.5 by choosing θ sufficiently small. $\qquad\square$

4. The induction step

The goal of this section is to prove the following theorem, which couples a pinned random walk with a Brownian Bridge. The tools used are Theorem 3.2 and induction. The induction hypothesis, properly formulated, allows us to get rid of the dyadic construction of the usual KMT proofs. The following is an alternative statement of Theorem 1.4, given here for the convenience of the reader.

Theorem 4.1. *Let us continue with the notation of Theorem 3.2. There exist positive universal constants C, K and λ_0 such that the following is true. For any $n \geq 2$, and any possible value of S_n, it is possible to construct a version of $W_0, W_1, \ldots, W_n$ and gaussian r.v. $Z_0, Z_1, \ldots, Z_n$ with mean zero and*

$$(16) \qquad \mathrm{Cov}(Z_i, Z_j) = \frac{(i \wedge j)(n - (i \vee j))}{n}$$

on the same probability space such that for any $\lambda \in (0, \lambda_0)$,

$$\mathbb{E} \exp(\lambda \max_{i \leq n} |W_i - Z_i|) \leq \exp\left(C \log n + \frac{K\lambda^2 S_n^2}{n}\right).$$

Proof. Recall the universal constants α_0 from Lemma 3.5 and c and θ_0 from Theorem 3.2. We contend that for carrying out the induction step, it suffices to take

$$(17) \qquad K = 8c, \ \lambda_0 \leq \sqrt{\frac{\alpha_0}{16c}} \wedge \frac{\theta_0}{2}, \ \text{and} \ C \geq \frac{1 + \log 2}{\log(3/2)}.$$

Choosing the constants to satisfy these constraints, we will now prove the claim by induction on n. Now, for each n, and each possible value a of S_n, let $f_a^n(\mathbf{s})$ denote the discrete probability density function of the sequence $(S_0, S_1, \ldots, S_n)$. Note that this is just the uniform distribution over $\mathcal{A}_a^n$, where

$$(18) \qquad \mathcal{A}_a^n := \{\mathbf{s} \in \mathbb{Z}^{n+1} : s_0 = 0, \ s_n = a, \ \text{and} \ |s_i - s_{i-1}| = 1 \ \text{for all} \ i.\}$$

Thus, for any $\mathbf{s} \in \mathcal{A}_a^n$,

$$(19) \qquad f_a^n(\mathbf{s}) = \frac{1}{|\mathcal{A}_a^n|}.$$

Let $\phi^n(\mathbf{z})$ denote the probability density function of a gaussian random vector $(Z_0, \ldots, Z_n)$ with mean zero and covariance (16).

We want to show that for each n, and each possible value a of S_n, we can construct a joint probability density $\rho_a^n(\mathbf{s}, \mathbf{z})$ on $\mathbb{Z}^{n+1} \times \mathbb{R}^{n+1}$ such that

$$(20) \qquad \int \rho_a^n(\mathbf{s}, \mathbf{z}) \, d\mathbf{z} = f_a^n(\mathbf{s}), \quad \int \rho_a^n(\mathbf{s}, \mathbf{z}) \, d\mathbf{s} = \phi^n(\mathbf{z}),$$

and for each $\lambda < \lambda_0$,

$$\int \exp\left(\lambda \max_{i \leq n}\left|s_i - \frac{ia}{n} - z_i\right|\right) \rho_a^n(\mathbf{s}, \mathbf{z}) \, d\mathbf{s} \, d\mathbf{z} \leq \exp\left(C \log n + \frac{K\lambda^2 a^2}{n}\right).$$

Suppose ρ_a^k can be constructed for $k = 1, \ldots, n - 1$, for allowed values of a in each case. We will now demonstrate a construction of ρ_a^n when a is an allowed value for S_n.

First, fix a possible value a of S_n and an index k such that $n/3 \leq k \leq 2n/3$ (for definiteness, take $k = [n/2]$). Given $S_n = a$, let $g_a^{n,k}(s)$ denote the

density function of S_k. Recall the definition (18) of $\mathcal{A}_a^n$ and note that for all allowed values of s of S_k, an elementary counting argument gives

$$(21) \qquad g_a^{n,k}(s) = \frac{|\mathcal{A}_s^k||\mathcal{A}_{a-s}^{n-k}|}{|\mathcal{A}_a^n|}.$$

Let $h^{n,k}(z)$ denote the density function of the gaussian distribution with mean 0 and variance $k(n-k)/n$. By Theorem 3.2 and the inequality $\exp|x| \leq \exp(x) + \exp(-x)$, we see that there exists a joint density function $\psi_a^{n,k}(s,z)$ on $\mathbb{Z} \times \mathbb{R}$ such that

$$(22) \qquad \int \psi_a^{n,k}(s,z)\,dz = g_a^{n,k}(s), \quad \int \psi_a^{n,k}(s,z)\,ds = h^{n,k}(z),$$

and for all $0 < \theta \leq \theta_0$,

$$(23) \qquad \int \exp\left(\theta\left|s - \frac{ka}{n} - z\right|\right)\psi_a^{n,k}(s,z)\,ds\,dz \leq \exp\left(1 + \frac{c\theta^2 a^2}{n}\right).$$

Now define a function $\gamma_a^n : \mathbb{Z} \times \mathbb{R} \times \mathbb{Z}^{k+1} \times \mathbb{R}^{k+1} \times \mathbb{Z}^{n-k+1} \times \mathbb{R}^{n-k+1} \to \mathbb{R}$ as follows:

$$(24) \qquad \gamma_a^n(s,z,\mathbf{s},\mathbf{z},\mathbf{s}',\mathbf{z}') := \psi_a^{n,k}(s,z)\rho_s^k(\mathbf{s},\mathbf{z})\rho_{a-s}^{n-k}(\mathbf{s}',\mathbf{z}').$$

By integrating over $\mathbf{s}',\mathbf{z}'$, then $\mathbf{s},\mathbf{z}$, and finally s,z, it is easy to verify that γ_a^n is a probability density function (if either a or s is not an allowed value, then $\psi_a^{n,k}(s,z) = 0$, so there is no problem).

Let $(S,Z,\mathbf{S},\mathbf{Z},\mathbf{S}',\mathbf{Z}')$ denote a random vector following the density γ_a^n. In words, this means the following: We are first generating (S,Z) from the joint distribution $\psi_a^{n,k}$; given $S = s, Z = z$, we are independently generating the pairs $(\mathbf{S},\mathbf{Z})$ and $(\mathbf{S}',\mathbf{Z}')$ from the joint densities ρ_s^k and ρ_{a-s}^{n-k} respectively.

Now define two random vectors $\mathbf{Y} \in \mathbb{R}^{n+1}$ and $\mathbf{U} \in \mathbb{Z}^{n+1}$ as follows. For $i \leq k$, let

$$Y_i = Z_i + \frac{i}{k}Z,$$

and for $i \geq k$, let

$$Y_i = Z'_{i-k} + \frac{n-i}{n-k}Z.$$

Note that the two definitions match at $i = k$ because $Z_k = Z'_0 = 0$. Next, define $U_i = S_i$ for $i \leq k$ and $U_i = S + S'_{i-k}$ for $i \geq k$. Again, the definitions match at $i = k$ because $S_k = S$ and $S'_0 = 0$. *We claim that the joint density of $(\mathbf{U}, \mathbf{Y})$ is a valid candidate for ρ_a^n.* The claim is proved in several steps.

1. *Marginal distribution of $\mathbf{U}$.* From equations (20) and (22) it is easy to see that

$$\int \gamma_a^n(s,z,\mathbf{s},\mathbf{z},\mathbf{s}',\mathbf{z}')\,d\mathbf{z}\,d\mathbf{z}'\,dz = g_a^{n,k}(s)f_s^k(\mathbf{s})f_{a-s}^{n-k}(\mathbf{s}').$$

In other words, the distribution of the triplet $(S,\mathbf{S},\mathbf{S}')$ can be described as follows: Generate S from the distribution of S_k given $S_n = a$; then independently generate $\mathbf{S}$ and $\mathbf{S}'$ from the conditional distributions f_s^k and

f_{a-s}^{n-k}. It should now be intuitively clear that $\mathbf{U}$ has marginal density f_a^n. Still, to be completely formal, we apply equations (19) and (21) to get

$$g_a^{n,k}(s)f_s^k(\mathbf{s})f_{a-s}^{n-k}(\mathbf{s}') = \frac{|\mathcal{A}_s^k||\mathcal{A}_{a-s}^{n-k}|}{|\mathcal{A}_a^n|}\frac{1}{|\mathcal{A}_s^k|}\frac{1}{|\mathcal{A}_{a-s}^{n-k}|} = \frac{1}{|\mathcal{A}_a^n|},$$

and observe that there is a one-to-one correspondence between $(S, \mathbf{S}, \mathbf{S}')$ and $\mathbf{U}$, and $\mathbf{U}$ can take any value in $\mathcal{A}_a^n$.

2. *Marginal distribution of* $\mathbf{Y}$. First, we claim that Z, $\mathbf{Z}$, and $\mathbf{Z}'$ are independent with densities $h^{n,k}$, ϕ^k, and ϕ^n respectively. Again, using (20) and (22), this is easily seen as follows.

$$\begin{aligned}
\int \gamma_a^n(s, z, \mathbf{s}, \mathbf{z}, \mathbf{s}', \mathbf{z}')\, d\mathbf{s}'\, d\mathbf{s}\, ds &= \int \psi_a^{n,k}(s, z)\rho_s^k(\mathbf{s}, \mathbf{z})\rho_{a-s}^{n-k}(\mathbf{s}', \mathbf{z}')\, d\mathbf{s}'\, d\mathbf{s}\, ds \\
&= \phi^{n-k}(\mathbf{z}')\int \psi_a^n(s, z)\rho_s^k(\mathbf{s}, \mathbf{z})d\mathbf{s}\, ds \\
&= \phi^{n-k}(\mathbf{z}')\phi^k(\mathbf{z})\int \psi_a^n(s, z)ds \\
&= \phi^{n-k}(\mathbf{z}')\phi^k(\mathbf{z})h^{n,k}(z).
\end{aligned}$$

Thus, $\mathbf{Y}$ is a gaussian random vector with mean zero. It only remains to compute $\mathrm{Cov}(Y_i, Y_j)$. Considering separately the cases $i \le j \le k$, $k \le i \le j$, and $i \le k \le j$, it is now straightforward to verify that $\mathrm{Cov}(Y_i, Y_j) = i(n - j)/n$ in each case. Thus, $\mathbf{Y} \sim \phi^n$.

3. *The exponential bound.* For $0 \le i \le n$, let

$$W_i = U_i - \frac{ia}{n}.$$

We have to show that for $0 < \lambda < \lambda_0$,

$$\mathbb{E}\exp(\lambda\max_{i \le n}|W_i - Y_i|) \le \exp\left(C\log n + \frac{K\lambda^2 a}{n}\right),$$

where C, K, and λ_0 are as in (17). Now let

$$T_L := \max_{i \le k}\left|S_i - \frac{iS}{k} - Z_i\right|, \quad T_R := \max_{i \ge k}\left|S'_{i-k} - \frac{i - k}{n - k}(a - S) - Z'_{i-k}\right|,$$

and

$$T := \left|S - \frac{ka}{n} - Z\right|.$$

We claim that

(25) $$\max_{i \le n}|W_i - Y_i| \le \max\{T_L, T_R\} + T.$$

To prove this, first take any $i \leq k$. Then

$$
\begin{aligned}
|W_i - Y_i| &= \left| S_i - \frac{ia}{n} - \left(Z_i + \frac{iZ}{k} \right) \right| \\
&\leq \left| S_i - \frac{iS}{k} - Z_i \right| + \left| \frac{iS}{k} - \frac{ia}{n} - \frac{iZ}{k} \right| \\
&\leq T_L + \frac{i}{k} T \leq T_L + T.
\end{aligned}
$$

Similarly, for $i \geq k$,

$$
\begin{aligned}
|W_i - Y_i| &= \left| S + S'_{i-k} - \frac{ia}{n} - \left(Z'_{i-k} + \frac{n-i}{n-k} Z \right) \right| \\
&\leq \left| S'_{i-k} - \frac{i-k}{n-k}(a - S) - Z'_{i-k} \right| \\
&\quad + \left| S + \frac{i-k}{n-k}(a - S) - \frac{ia}{n} - \frac{n-i}{n-k} Z \right| \\
&= \left| S'_{i-k} - \frac{i-k}{n-k}(a - S) - Z'_{i-k} \right| + \frac{n-i}{n-k} \left| S - \frac{ka}{n} - Z \right| \\
&\leq T_R + T.
\end{aligned}
$$

This proves (25). Now fix $\lambda < \lambda_0$. Using the crude bound $\exp(x \vee y) \leq \exp x + \exp y$, we get

$$
(26) \qquad \exp(\lambda \max_{i \leq n} |W_i - Y_i|) \leq \exp(\lambda T_L + \lambda T) + \exp(\lambda T_R + \lambda T).
$$

Now, by the construction (24), it is easy to check that given $(S, Z) = (s, z)$, the conditional density of $(\mathbf{S}, \mathbf{Z})$ is simply ρ_s^k. By the induction hypothesis, this implies that

$$
\mathbb{E}(\exp(\lambda T_L)|S, Z) \leq \exp\left(C \log k + \frac{K\lambda^2 S^2}{k} \right).
$$

It is easy to see that the moment generating functions of both T_L and T are finite everywhere, and hence there is no problem in applying the Cauchy-Schwarz inequality to get

$$
\begin{aligned}
\mathbb{E}\exp(\lambda T_L + \lambda T) &\leq \left[\mathbb{E}\left(\mathbb{E}(\exp(\lambda T_L)|S, Z)^2 \right) \mathbb{E}(\exp(2\lambda T)) \right]^{1/2} \\
&\leq \exp(C \log k) \left[\mathbb{E}\exp\left(\frac{2K\lambda^2 S^2}{k} \right) \mathbb{E}\exp(2\lambda T) \right]^{1/2}.
\end{aligned}
$$

We wish to apply Lemma 3.5 to bound the first term inside the bracket. Observe that by (17), we have

$$
2K\lambda^2 \leq 16c \cdot \frac{\alpha_0}{16c} = \alpha_0,
$$

and also $n/3 \leq k \leq 2n/3$ by assumption. Hence Lemma 3.5 can indeed be applied to get

$$\mathbb{E}\exp\left(\frac{2K\lambda^2 S^2}{k}\right) \leq \exp\left(1 + \frac{3K\lambda^2 a^2}{2n}\right).$$

Next, note that by (17), $2\lambda \leq \theta_0$. Hence by inequality (23) with $\theta = 2\lambda$, we get the bound

$$\mathbb{E}\exp(2\lambda T) \leq \exp\left(1 + \frac{4c\lambda^2 a^2}{n}\right).$$

Combining the last three steps, we have

$$\mathbb{E}\exp(\lambda T_L + \lambda T) \leq \exp\left(C\log k + 1 + \frac{(3K + 8c)\lambda^2 a^2}{4n}\right).$$

Now, by (17), $3K + 8c = 4K$. Again, since $n/3 \leq k \leq 2n/3$, we have

$$\log k = \log n - \log(n/k) \leq \log n - \log(3/2).$$

Thus,

$$\mathbb{E}\exp(\lambda T_L + \lambda T) \leq 2^{1/2}\exp\left(C\log n - C\log(3/2) + 1 + \frac{K\lambda^2 a^2}{n}\right).$$

By the symmetry of the situation, we can get the exact same bound on $\mathbb{E}\exp(\lambda T_R + \lambda T)$. Combined with (26), this gives

$$\mathbb{E}\exp(\lambda \max_{i \leq n}|W_i - Y_i|) \leq 2\exp\left(C\log n - C\log(3/2) + 1 + \frac{K\lambda^2 a^2}{n}\right).$$

Finally, from the condition on C in (17), we see that

$$-C\log(3/2) + 1 + \log 2 \leq 0.$$

This completes the induction step. To complete the argument, we just choose C so large and λ_0 so small that the result is true for $n = 2$ even if the vectors (W_0, W_1, W_2) and (Z_0, Z_1, Z_2) are chosen to be independent of each other. □

5. COMPLETING THE PROOFS OF THE MAIN THEOREMS

In this final section, we put together the pieces to complete the proofs of Theorem 1.4 and Theorem 1.5. The following lemma combines Theorem 4.1 and Theorem 3.1 to give a 'finite n version' of Theorem 1.5.

Lemma 5.1. *There exist universal constants $B > 1$ and $\lambda > 0$ such that the following is true. Let n be a positive integer and let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ be i.i.d. symmetric ± 1 random variables. Let $S_k = \sum_{i=1}^{k}\varepsilon_i$, $k = 0, 1, \ldots, n$. It is possible to construct a version of the sequence $(S_k)_{k \leq n}$ and gaussian random variables $(Z_k)_{k \leq n}$ with mean 0 and $\mathrm{Cov}(Z_i, Z_j) = i \wedge j$ on the same probability space such that $\mathbb{E}\exp(\lambda|S_n - Z_n|) \leq B$ and*

$$\mathbb{E}\exp(\lambda \max_{k \leq n}|S_k - Z_k|) \leq B\exp(B\log n).$$

Proof. Recall the universal constants θ_0 and κ from Theorem 3.1 and C, K, and λ_0 from Theorem 4.1. Choose λ so small that

$$\lambda < \frac{\theta_0 \wedge \lambda_0}{2} \quad \text{and} \quad 16K\lambda^2 < 1.$$

Let the probability densities f_a^n, ρ_a^n, and ϕ^n be as in the proof of Theorem 4.1. Let g^n and h^n denote the densities of S_n and Z_n respectively. By Theorem 3.1 and the choice of λ, there is a joint density ψ^n on $\mathbb{Z} \times \mathbb{R}$ such that

$$\int \psi^n(s,z)\,dz = g^n(s), \quad \int \psi^n(s,z)\,ds = h^n(z),$$

and

(27) $$\int \exp(2\lambda|s-z|)\psi^n(s,z)\,ds\,dz \le \kappa.$$

Now define a function $\gamma^n : \mathbb{Z} \times \mathbb{R} \times \mathbb{Z}^{n+1} \times \mathbb{R}^{n+1} \to \mathbb{R}$ as

$$\gamma^n(s,z,\mathbf{s},\mathbf{z}) := \psi^n(s,z)\rho_s^n(\mathbf{s},\mathbf{z}).$$

It is easy to check that this is a probability density function. Let $(S,Z,\mathbf{S},\mathbf{Z})$ be a random vector following this density. As in the proof of Theorem 4.1, an easy integration shows that the joint density of $(Z,\mathbf{Z})$ is simply

$$h^n(z)\phi^n(\mathbf{z}).$$

Define a random vector $\mathbf{Y} = (Y_0, \ldots, Y_n)$ as

$$Y_i = Z_i + \frac{i}{n}Z.$$

By the independence of Z and $\mathbf{Z}$ and their distributions, it follows that $\mathbf{Y}$ is a mean zero gaussian random vector with $\mathrm{Cov}(Y_i, Y_j) = i \wedge j$.

Next, integrating out z and $\mathbf{z}$ we see that the joint density of $(S,\mathbf{S})$ is

$$g^n(s)f_s^n(\mathbf{s}).$$

Elementary probabilistic reasoning now shows that the marginal distribution of $\mathbf{S}$ is the same as that of a simple random walk up to time n.

Let us now show that the law of the pair $(\mathbf{S}, \mathbf{Y})$ satisfies the conditions of the theorem. First, let $W_i = S_i - iS/n$. Note that for any $i \le n$,

$$|S_i - Y_i| = \left| S_i - \left(Z_i + \frac{i}{n}Z \right) \right|$$

$$\le |W_i - Z_i| + \frac{i}{n}|S - Z|.$$

Note that the conditional distribution of $(\mathbf{S}, \mathbf{Z})$ given $(S,Z) = (s,z)$ is simply ρ_s^n. Since $\lambda < \lambda_0$, we have by the construction of ρ_s^n that

$$\mathbb{E}\left(\exp(\lambda \max_{i \le n} |W_i - Z_i|) \big| S, Z \right) \le \exp\left(C\log n + \frac{K\lambda^2 S^2}{n} \right).$$

Thus, using the Cauchy-Schwarz inequality and (27), we can now get

$$\mathbb{E}\exp(\lambda \max_{i\leq n}|S_i - Y_i|)$$

$$\leq \left[\mathbb{E}\big(\mathbb{E}\big(\exp(\lambda \max_{i\leq n}|W_i - Z_i|)\big|S,Z\big)^2\big)\mathbb{E}\exp(2\lambda|S-Z|)\right]^{1/2}$$

$$\leq \exp(C\log n)\left[\kappa\mathbb{E}\exp(2K\lambda^2 S^2/n)\right]^{1/2}.$$

By inequality (13) and the choice of λ, the proof of the maximal inequality is done. For the other inequality, note that we have (27) and $Y_n = Z$ since $Z_n = 0$. $\qquad\qquad\square$

Proofs of Theorems 1.4 *and* 1.5. The proof of Theorem 1.4 follows trivially from Theorem 4.1. The proof of Theorem 1.5 also follows quite easily from Lemma 5.1, but some more work is required. We carry out the few remaining steps below.

For $r = 1, 2, \ldots$ let $m_r = 2^{2^r}$, and $n_r = m_r - m_{r-1}$. For each r $(S_k^{(r)}, Z_k^{(r)})_{0\leq k\leq n_r}$ be a random vector satisfying the conclusions of Lemma 5.1, and suppose these random vectors are independent. Inductively define an infinite sequence $(S_k, Z_k)_{k\geq 0}$ as follows. Let $S_k = S_k^{(1)}$ and $Z_k = Z_k^{(1)}$ for $k \leq m_1$. Having defined $(S_k, Z_k)_{k\leq m_{r-1}}$, define $(S_k, Z_k)_{m_{r-1}<k\leq m_r}$ as

$$S_k := S_{k-m_{r-1}}^{(r)} + S_{m_{r-1}}, \quad Z_k := Z_{k-m_{r-1}}^{(r)} + Z_{m_{r-1}}.$$

Clearly, since the increments are independent, S_k and Z_k are indeed random walks with binary and gaussian increments respectively.

Now recall the constants B and λ in Lemma 5.1. First, note that for each r, by Lemma 5.1 and independence we have

(28)
$$\mathbb{E}\exp(\lambda|S_{m_r} - Z_{m_r}|) \leq \mathbb{E}\exp\left(\lambda\sum_{\ell=1}^{r}|S_{n_\ell}^{(\ell)} - Z_{n_\ell}^{(\ell)}|\right)$$

$$= \prod_{\ell=1}^{r}\mathbb{E}\exp\big(\lambda|S_{n_\ell}^{(\ell)} - Z_{n_\ell}^{(\ell)}|\big) \leq B^r.$$

Next, let

$$C = \frac{1}{1 - \frac{\exp(-\frac{1}{2}B\log 4)}{B}}.$$

We will show by induction that for each r,

(29)
$$\mathbb{E}\exp(\lambda \max_{k\leq m_r}|S_k - Z_k|) \leq CB^r\exp(B\log m_r).$$

By Lemma 5.1 and the facts that $B > 1$ and $C > 1$, this holds for $r = 1$. Suppose it holds for $r - 1$. By the inequality $\exp(x \vee y) \leq \exp x + \exp y$, we have

(30)
$$\mathbb{E}\exp(\lambda \max_{k\leq m_r}|S_k - Z_k|) \leq \mathbb{E}\exp(\lambda \max_{m_{r-1}\leq k\leq m_r}|S_k - Z_k|)$$

$$+ \mathbb{E}\exp(\lambda \max_{k\leq m_{r-1}}|S_k - Z_k|).$$

Let us consider the first term. We have

$$\max_{m_{r-1} \leq k \leq m_r} |S_k - Z_k| \leq \max_{1 \leq j \leq n_r} |S_j^{(r)} - Z_j^{(r)}| + |S_{m_{r-1}} - Z_{m_{r-1}}|.$$

Thus, by independence and Lemma 5.1, and the inequality (28), we get

$$\mathbb{E} \exp(\lambda \max_{m_{r-1} \leq k \leq m_r} |S_k - Z_k|) \leq B^r \exp(B \log m_r).$$

By the induction hypothesis and the relation $m_r = m_{r-1}^2$, we see that the second term in (30) has the bound

$$\mathbb{E} \exp(\lambda \max_{k \leq m_{r-1}} |S_k - Z_k|) \leq CB^{r-1} \exp(B \log m_{r-1})$$

$$= CB^{r-1} \exp\left(\frac{B \log m_r}{2}\right).$$

Combining, we get

$$\mathbb{E} \exp(\lambda \max_{k \leq m_r} |S_k - Z_k|) \leq B^r \exp(B \log m_r)\left(1 + \frac{C}{B} \exp\left(-\frac{B \log m_r}{2}\right)\right).$$

From the definition of C, it easy to verify (since $m_r \geq 4$), that the term within the parentheses in the above expression is bounded by C. This completes the induction step.

So we have now shown (29). Since $r \leq const. \log m_r$, this shows that there exists a constant K such that for all r,

$$\mathbb{E} \exp(\lambda \max_{k \leq m_r} |S_k - Z_k|) \leq K \exp(K \log m_r).$$

Now let us prove such an inequality for arbitrary n instead of m_r. Take any $n \geq 2$. Let r be such that $m_{r-1} \leq n \leq m_r$. Then $m_r = m_{r-1}^2 \leq n^2$. Thus,

$$\mathbb{E} \exp(\lambda \max_{k \leq n} |S_k - Z_k|) \leq \mathbb{E} \exp(\lambda \max_{k \leq m_r} |S_k - Z_k|)$$

$$\leq K \exp(K \log m_r) \leq K \exp(2K \log n).$$

It is now easy to complete the argument using Markov's inequality. □

References

[1] Azuma, K. (1967). Weighted sums of certain dependent random variables. *Tôhoku Math. Journal* **19** 357–367

[2] Bhatia, R. (1997). *Matrix analysis.* Springer-Verlag, New York.

[3] CHATTERJEE, S. (2008). Fluctuations of eigenvalues and second order Poincaré inequalities. To appear in *Probab. Theory Related Fields.* Available at http://arxiv.org/abs/0705.1224

[4] CSÖRGŐ, M. (2007). A glimpse of the KMT (1975) approximation of empirical processes by Brownian bridges via quantiles. *Acta Sci. Math. (Szeged)* **73** no. 1-2, 349–366.

[5] CSÖRGŐ, M. and HORVÁTH, L. (1993). *Weighted Approximations in Probability and Statistics.* John Wiley & Sons, Chichester.

[6] CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

[7] DEMBO, A., PERES, Y., ROSEN, J., and ZEITOUNI, O. (2001). Thick points for planar Brownian motion and the Erdős-Taylor conjecture on random walk. *Acta Math.* **186** no. 2, 239–270.

[8] DEMBO, A., PERES, Y., ROSEN, J., and ZEITOUNI, O. (2004). Cover times for Brownian motion and random walks in two dimensions. *Ann. of Math. (2)* **160** no. 2, 433–464.

[9] DUNFORD, N. and SCHWARTZ, J. T. (1958). *Linear Operators. I.* Interscience Publishers, Inc., New York.

[10] EINMAHL, U. (1989). Extensions of results of Komlós, Major and Tusnády to the multivariate case. *J. Multivar. Anal.,* **28** 20–68.

[11] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Stat. Assoc.* **58** 13–30.

[12] KIEFER, J. (1969). On the deviations in the Skorokhod-Strassen approximation scheme. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **13** 321–332.

[13] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 111–131.

[14] LIFSHITS, M. A. (2007). *Lectures on the strong approximation.* St. Petersburg University Press, St. Petersburg, 1–32 (in Russian).

[15] MASON, D. M. (2007). Some observations on the KMT dyadic scheme. *J. Statist. Plann. Inference* **137** no. 3, 895–906.

[16] OBŁÓJ, J. (2004). The Skorokhod embedding problem and its offspring. *Probab. Surv.* **1** 321–390 (electronic).

[17] SAKHANENKO, A. I. (1984). Rate of convergence in the invariance principle for variables with exponential moments that are not identically distributed. (Russian) *Limit theorems for sums of random variables,* 4–49, Trudy Inst. Mat., **3** "Nauka" Sibirsk. Otdel., Novosibirisk.

[18] SHAO, Q.-M. (1995). Strong approximation theorems for independent random variables and their applications. *J. Multivariate Anal.* **52** no. 1, 107–130.

[19] SKOROHOD, A. V. (1961). *Issledovaniya po teorii sluchainykh protsessov (Stokhasticheskie differentsialnye uravneniya i predelnye teoremy d lya protsessov Markova).* Izdat. Kiev. Univ., Kiev.

[20] SKOROKHOD, A. V. (1965). *Studies in the theory of random processes.* Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass.

[21] STEIN, C. (1986). *Approximate computation of expectations.* IMS Lecture Notes—Monograph Series, **7**.

[22] STRASSEN, V. (1967). Almost sure behavior of sums of independent random variables and martingales. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66),* Vol. II, Part 1, 315–343. Univ. California Press, Berkeley, Calif.

[23] ZAITSEV, A. YU. (1998). Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *ESAIM : Probability and Statistics,* **2** 41–108.

[24] ZAITSEV, A. YU. (2000). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. I. *Teor. Veroyatnost. i Primenen.* **45** no. 4, 718–738; *translation in Theory Probab. Appl.* **45** (2002), no. 4, 624–641.

[25] ZAITSEV, A. YU. (2001). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. II. *Teor. Veroyatnost. i Primenen.* **46** no. 3, 535–561; *translation in Theory Probab. Appl.* **46** (2003), no. 3, 490–514.

[26] ZAITSEV, A. YU. (2001). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. III. *Teor. Veroyatnost. i Primenen.* **46** no. 4, 744–769; *translation in Theory Probab. Appl.* **46** (2003), no. 4, 676–698.

[27] ZAITSEV, A. YU. (2002). Estimates for the strong approximation in multidimensional central limit theorem. *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002),* 107–116, *Higher Ed. Press, Beijing.*

[28] ZAITSEV, A. YU. (2008). Personal communication.

COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, NY 10012. *E-mail:* `sourav@cims.nyu.edu`