

# Stein’s method and approximating the quantum harmonic oscillator

IAN W. MCKEAGUE<sup>1</sup>, EROL A. PEKÖZ<sup>2</sup> and YVIK SWAN<sup>3</sup>

<sup>1</sup>Department of Biostatistics, Columbia University, Room R639, 722 West 168th Street, New York, NY 10032, USA. E-mail: im2131@columbia.edu

<sup>2</sup>Questrom School of Business, Boston University, 595 Commonwealth Avenue, Room 607, Boston, MA 02215, USA. E-mail: pekoz@bu.edu

<sup>3</sup>Département de Mathématique, Université de Liège, B37 12 allée de la découverte, B-4000 Liège, Belgium. E-mail: yswan@uliege.be

Hall *et al.* [*Phys. Rev. X* **4** (2014) 041013] recently proposed that quantum theory can be understood as the continuum limit of a deterministic theory in which there is a large, but finite, number of classical “worlds.” A resulting Gaussian limit theorem for particle positions in the ground state, agreeing with quantum theory, was conjectured in Hall *et al.* [*Phys. Rev. X* **4** (2014) 041013] and proven by McKeague and Levin [*Ann. Appl. Probab.* **26** (2016) 2540–2555] using Stein’s method. In this article we show how quantum position probability densities for higher energy levels beyond the ground state may arise as distributional fixed points in a new generalization of Stein’s method. These are then used to obtain a rate of distributional convergence for conjectured particle positions in the first energy level above the ground state to the (two-sided) Maxwell distribution; new techniques must be developed for this setting where the usual “density approach” Stein solution (see Chatterjee and Shao [*Ann. Appl. Probab.* **21** (2011) 464–483]) has a singularity.

*Keywords:* higher energy levels; interacting particle system; Maxwell distribution; Stein’s method

## 1. Introduction

Hall *et al.* [8] proposed a many interacting worlds (MIW) theory for interpreting quantum mechanics in terms of a large but finite number of classical “worlds.” In the case of the MIW harmonic oscillator, an energy minimization argument was used to derive a recursion giving the location of the oscillating particle as viewed in each of the worlds. Hall *et al.* conjectured that the empirical distribution of these locations converges to Gaussian as the total number of worlds  $N$  increases. McKeague and Levin [10] recently proved such a result and provided a rate of convergence. More specifically, McKeague and Levin showed that if  $x_1, \dots, x_N$  is a decreasing, zero-mean sequence of real numbers satisfying the recursion relation

$$x_{n+1} = x_n - \frac{1}{x_1 + \dots + x_n}, \tag{1}$$

then the empirical distribution of the  $x_n$  tends to standard Gaussian when  $N \rightarrow \infty$ . Here  $x_n$  represents the location of the oscillating particle in the  $n$ th world, and the Gaussian limit distribution agrees with quantum theory for a particle in the lowest energy (ground) state.

The hypothesized correspondence with quantum theory suggests that stable configurations should also exist at higher energies in the MIW theory. Moreover, the empirical distributions of these configurations should converge to distributions with densities of the form

$$p_k(x) = \frac{(\text{He}_k(x))^2}{k!} \varphi(x), \quad x \in \mathbb{R}, \quad (2)$$

where  $\varphi(x)$  is the standard normal density,

$$\text{He}_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}$$

is the (probabilist's)  $k$ th Hermite polynomial, and  $k$  is a non-negative integer. The ground state discussed above corresponds to  $k = 0$  and has the standard Gaussian limit. However, the question of how to characterize higher energy MIW states corresponding to  $k \geq 1$  is still unresolved as far as we know.

The energy minimization approach of Hall *et al.* [8] starts with an analysis of the Hamiltonian for the MIW harmonic oscillator:

$$H_0(\mathbf{x}, \mathbf{p}) = E(\mathbf{p}) + V(\mathbf{x}) + U_0(\mathbf{x}),$$

where the locations of particles (having unit mass) in the  $N$  worlds are specified by  $\mathbf{x} = (x_1, \dots, x_N)$  with  $x_1 > x_2 > \dots > x_N$ , and their momenta by  $\mathbf{p} = (p_1, \dots, p_N)$ . Here  $E(\mathbf{p}) = \sum_{n=1}^N p_n^2/2$  is the kinetic energy,  $V(\mathbf{x}) = \sum_{n=1}^N x_n^2$  is the potential energy (for the parabolic trap), and

$$U_0(\mathbf{x}) = \sum_{n=1}^N \left( \frac{1}{x_{n+1} - x_n} - \frac{1}{x_n - x_{n-1}} \right)^2$$

is called the ‘‘interworld’’ potential, where  $x_0 = \infty$  and  $x_{N+1} = -\infty$ . In the ground state, there is no movement because all the momenta  $p_n$  have to vanish for the total energy to be minimized. In this case, as mentioned above, Hall *et al.* [8] showed that the particle locations  $x_n$  satisfy (1) and McKeague and Levin [10] showed that the empirical distribution tends to a standard Gaussian distribution.

Our contribution in the present article is to derive an interworld potential for the second energy state ( $k = 1$ ) and show that the empirical distribution of the configuration that minimizes the corresponding Hamiltonian has a limit distribution that again agrees with quantum theory. The interworld potential in this case is shown to be

$$U_1(\mathbf{x}) = 9 \sum_{n=1}^N \left( \frac{1}{x_{n+1}^3 - x_n^3} - \frac{1}{x_n^3 - x_{n-1}^3} \right)^2 x_n^4 \quad (3)$$

and the minimizer of the corresponding Hamiltonian  $H_1(\mathbf{x}, \mathbf{p}) = E(\mathbf{p}) + V(\mathbf{x}) + U_1(\mathbf{x})$  is shown to satisfy the recursion

$$x_{n+1}^3 = x_n^3 - 3 \left( \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}. \quad (4)$$

Further, we show that if  $x_1, \dots, x_N$  is a decreasing, zero-mean solution, then the empirical distribution of the  $x_n$  converges to the (two-sided) Maxwell distribution having density  $p_1(x) = x^2 e^{-x^2/2} / \sqrt{2\pi}$ . The entire sequence  $x_1, \dots, x_N$  should be viewed as indexed by  $N$ , though we suppress notation for this dependence and write  $x_1, \dots, x_N$  instead of  $x_{1,N}, \dots, x_{N,N}$ . We also give a rate of convergence using a new extension of Stein's method. Our approach is generalizable to recursions that converge to the distributions of other higher energy states of the quantum harmonic oscillator, although we do not pursue such extensions here.

We initially thought that the MIW interpretation could be based on a "universal" interworld potential function  $U_0$  that applies to all energy levels, with the densities  $p_k(x)$  then arising as limits of *local* minima of  $H_0$ . However, this idea turned out to be analytically unworkable. Here we propose an alternative approach in terms of adapting the interworld potential to each higher energy level. Minimizing the resulting Hamiltonian is then tractable and the solution can be shown to converge to  $p_k(x)$ , at least in the case  $k = 1$ . Hall *et al.* [8] derived their interworld potential  $U_0$  as a discretization of Bohm's quantum potential summed over the particle ensemble, see Bohm [1]. The challenge in general is to extend this derivation to higher-energy wave functions in a way that leads to an explicit recursion minimizing the resulting Hamiltonian, and to show that it agrees with  $p_k(x)$  in the limit. A major contribution here, in addition to providing a rate of convergence, is a general method for finding such interworld potential functions and their associated particle recursions.

Stein's method (see Stein [17], Chen *et al.* [5] and Ross [16]) is a well-established technique for obtaining explicit error bounds for distributional limit theorems. However, the usual "density approach" (see Chatterjee and Shao [4]) for applying Stein's method does not seem to work in cases where the density function vanishes at a point in the interior of the support of the target distribution (here we have  $p_1(0) = 0$  and the support is the whole real line). As we elaborate later, in this case the solution to the Stein equation will have a singularity and also unbounded derivatives. This motivates the new technique we will develop to handle such distributions. While there are plenty of examples of Stein's method applied to distributions with a density having a zero on the boundary of the support (the gamma and beta distributions, for example), there have been no examples (that we know of) with a zero in the interior of the support; the higher energy distributions  $p_k(x)$ , for  $k \geq 1$ , appear to be the first such distributions considered. The price one has to pay with our approach for handling these zeros is more complicated estimates involving couplings. In our case, however, analytical properties of the recursion (4) can fortunately be exploited to establish such estimates.

In Section 2, we generalize the argument of Hall *et al.* [8] to derive the interworld potential, and show how it leads to the solution (4). In Section 3, we introduce the notion of a generalized zero-bias transformation, and show that the distributional properties of eigenstates of the quantum harmonic oscillator can be characterized in terms of fixed points of this transformation. Also, we derive the generalized zero-bias distribution for the empirical distribution of general configurations. Section 4 develops our results based on the new extension of Stein's method to show convergence of the configuration that minimizes the Hamiltonian of the second energy state.

## 2. Interworld potentials for higher energy states

Hall *et al.* [8] introduced their MIW theory from the perspective of the de Broglie–Bohm interpretation of quantum mechanics, which is mathematically equivalent to standard quantum theory. They used this approach to construct an ansatz for the conjectured interworld potential  $U_0$  governing the ground state wave function of the quantum harmonic oscillator. In this section, we introduce an extended version of this ansatz aimed at providing a MIW characterization of the higher energy eigenstates.

Our argument follows along the lines of Section IIIA of Hall *et al.* [8] with the major difference being that we now need to introduce a more general way of approximating the density of particle location for a stationary wave function  $\psi(x)$ , namely for a density of the form  $p(x) = |\psi(x)|^2 = b(x)\varphi(x)$ , where  $b(x)$  is a non-negative, even, smooth function having finitely many zeros. Here  $b$  represents a “baseline” that varies more rapidly than  $\varphi(x)$ . Let  $x_1 > x_2 > \dots > x_N$ . Bohm’s quantum potential summed over the ensemble  $\{x_n\}$  is defined by

$$U_\psi(\mathbf{x}) = \sum_{n=1}^N [p'(x_n)/p(x_n)]^2, \quad (5)$$

where we are using dimensionless units. An approximation to  $p(x_n)$  based on ignoring  $\varphi(x)$  is given (up to a normalizing constant) by

$$\tilde{p}(x_n) = \frac{b(x_n)}{B(x_n) - B(x_{n+1})},$$

where  $B(x) = \int_0^x b(t) dt$  is the cumulative baseline function. This suggests

$$\frac{p'(x_n)}{p(x_n)} \approx \frac{\tilde{p}(x_n) - \tilde{p}(x_{n-1})}{(x_n - x_{n-1})\tilde{p}(x_n)} \approx \left[ \frac{1}{B(x_n) - B(x_{n+1})} - \frac{1}{B(x_{n-1}) - B(x_n)} \right] b(x_n),$$

where we set  $B(x_0) = \infty$  and  $B(x_{N+1}) = -\infty$ . Our proposed ansatz for the interworld potential is then based on inserting the above expression into (5) to obtain

$$U_b(\mathbf{x}) = \sum_{n=1}^N \left[ \frac{1}{B(x_{n+1}) - B(x_n)} - \frac{1}{B(x_n) - B(x_{n-1})} \right]^2 b(x_n)^2. \quad (6)$$

Note that our earlier assumptions about  $b$  imply that  $B$  is strictly increasing, so  $U_b$  is well-defined. In the simplest cases  $b(x) = 1$  and  $b(x) = x^2$  the above expression for  $U_b$  agrees with the interworld potentials  $U_0$  and  $U_1$  defined in the [Introduction](#). We conjecture that the interworld potential  $U_b$  is suitable for obtaining MIW approximations to the class of target distributions of the form  $p_k(x)$ . Indeed, there may be a natural affinity between our new version of Stein’s method and the densities  $p_k(x)$  for all the energy levels of the quantum harmonic oscillator.

Specializing to the case  $b(x) = x^2$ , the following argument characterizes the minimizer of the Hamiltonian  $H_1$  (i.e., the ground state when the interworld potential is  $U_1$ ) in terms of a

solution to the recursion (4). In any ground state the particles do not move, so the kinetic energy  $E$  vanishes. Then, adapting the argument of Hall *et al.* [8] to apply to  $H_1$ , we have

$$\begin{aligned}
 9(N-1)^2 &= 9 \left[ \sum_{n=1}^{N-1} \frac{x_{n+1}^3 - x_n^3}{x_{n+1}^3 - x_n^3} \right]^2 \\
 &= 9 \left[ \sum_{n=1}^N \left( \frac{1}{x_{n+1}^3 - x_n^3} - \frac{1}{x_n^3 - x_{n-1}^3} \right) x_n^2 (x_n - \overline{x_N^3/x_n^2}) \right]^2 \\
 &\leq 9 \left[ \sum_{n=1}^N \left( \frac{1}{x_{n+1}^3 - x_n^3} - \frac{1}{x_n^3 - x_{n-1}^3} \right)^2 x_n^4 \right] \left[ \sum_{n=1}^N (x_n - \overline{x_N^3/x_n^2})^2 \right] \\
 &\leq U_1(\mathbf{x})V(\mathbf{x}),
 \end{aligned}$$

where the first inequality is Cauchy–Schwarz. So  $U_1 \geq 9(N-1)^2/V$ , leading to

$$H_1 = U_1 + V \geq 9(N-1)^2/V + V \geq 6(N-1)$$

with the last inequality being equality for  $V = 3(N-1)$ . It follows that  $H_1$  is minimized when  $V = 3(N-1)$ , the mean  $\overline{x_N^3}$  of  $\{x_n^3, n = 1, \dots, N\}$  vanishes, and

$$\frac{1}{x_n} = \alpha \left[ \frac{1}{x_{n+1}^3 - x_n^3} - \frac{1}{x_n^3 - x_{n-1}^3} \right]$$

for some constant  $\alpha$ . The sum of the right of the above display telescopes, leading to the recursion (4) by rearranging and noting that  $\alpha = -V/(N-1) = -3$ .

The following lemma provides the basic properties we need to ensure the existence of a solution of the Maxwell recursion (4) that minimizes the Hamiltonian  $H_1$ , as well as ensuring that the solution is unique. This result is analogous to Lemma 1 of McKeague and Levin [10] concerning solutions of (1), but the difference here is that the variance is 3, agreeing with the Maxwell distribution (rather than close to standard normal in the case of (1)).

**Lemma 2.1.** *Suppose  $N$  is even. Every zero-median solution  $x_1, \dots, x_N$  of (4) satisfies:*

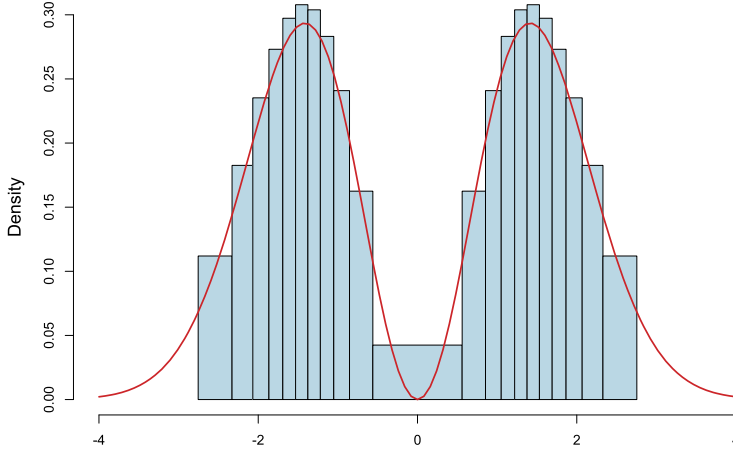
- (P1) *Zero-mean:*  $x_1 + \dots + x_N = 0$ .
- (P2) *Maxwell variance:*  $x_1^2 + \dots + x_N^2 = 3(N-1)$ .
- (P3) *Symmetry:*  $x_n = -x_{N+1-n}$  for  $n = 1, \dots, N$ .

Further, there exists a unique solution  $x_1, \dots, x_N$  such that (P1) and

- (P4) *Strictly decreasing:*  $x_1 > \dots > x_N$

hold. This solution has the zero-median property, and thus also satisfies (P2) and (P3).

**Proof.** The proof follows identical steps to the proof of Lemma 1 of McKeague and Levin [10], apart from the variance property (P2), which is proved using (P1) and (P3) as follows. Denote



**Figure 1.** Example with  $b(x) = x^2$ ,  $N = 22$ , showing the piecewise constant density having mass  $1/(N - 1)$  uniformly distributed over the intervals between successive  $x_n$  compared with the Maxwell density, where the breaks in the histogram are the successive  $x_n$  satisfying the recursion (4).

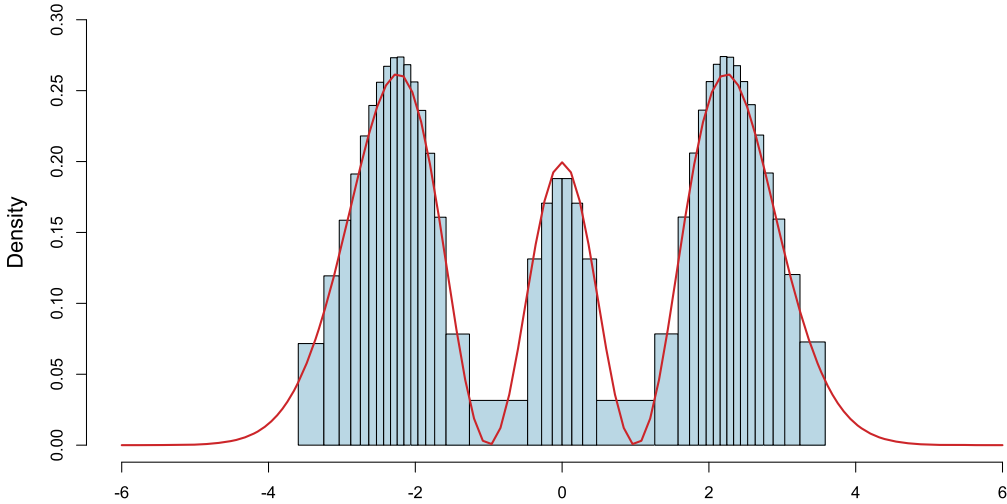
$S_n = \sum_{i=1}^n x_i^{-1}$  for  $n = 1, \dots, N$ , and set  $S_0 = 0$ . Using (4), we can write

$$\begin{aligned} 3(N - 1) &= 3 \sum_{n=1}^{N-1} S_n S_n^{-1} = \sum_{n=1}^{N-1} S_n (x_n^3 - x_{n+1}^3) = \sum_{n=1}^{N-1} [(S_{n-1} + x_n^{-1})x_n^3 - S_n x_{n+1}^3] \\ &= \sum_{n=1}^{N-1} [S_{n-1} x_n^3 - S_n x_{n+1}^3 + x_n^2] \\ &= x_1^2 + \dots + x_{N-1}^2 - S_{N-1} x_N^3, \end{aligned}$$

where we used the recursion in the second equality, and the last equality is from a telescoping sum. (P3) implies  $S_N = 0$ , so  $-S_{N-1} = 1/x_N$ , and (P2) follows.  $\square$

Although in the sequel we concentrate on the case  $k = 1$  (see Figure 1), to conclude this section we briefly discuss general densities of the form  $p_k$  given in (2). The above argument for  $b(x) = x^2$  can be extended to general  $U_b$  under the condition that  $B(x)$  is proportional to  $x b(x)$ , which is the case when  $b(x)$  is proportional to  $x^r$  for some even non-negative integer  $r$  (but not for the square of the  $k$ th Hermite polynomial unless  $k = 0$  or 1). Under this condition, it can be shown that the minimizer of the Hamiltonian based on  $U_b$  is a symmetric solution of the recursion

$$B(x_{n+1}) = B(x_n) - \left( \sum_{i=1}^n \frac{x_i}{b(x_i)} \right)^{-1}. \quad (7)$$



**Figure 2.** Example with  $b(x) = \text{He}_k(x)^2/k!$  for  $k = 2$ ,  $N = 41$ , where the breaks in the histogram are the successive  $x_n$  satisfying the recursion (7) and the red curve is  $p_k(x)$ .

We have not been able to show that this recursion minimizes the Hamiltonian for general  $b$ , but our numerical results suggest that it is very close if not identical to a minimizer. With  $k = 2$  we have  $b(x) = (x^2 - 1)^2/2$ ,  $B(x) = x^5/10 - x^3/3 + x/2$ , and the symmetric solution of the resulting recursion produces a remarkably good agreement with  $p_k$ , see Figure 2.

### 3. Generalized zero-bias transformations

Let  $W$  be a symmetric random variable and  $b: \mathbb{R} \rightarrow \mathbb{R}$  a non-negative function such that  $\sigma^2 = \mathbb{E}[W^2/b(W)] < \infty$ . Goldstein and Reinert [7] gives a distributional fixed point characterization of the Gaussian distribution, which we generalize in the definition below.

**Definition 3.1.** If there is a random variable  $W^*$  such that

$$\sigma^2 \mathbb{E} \left[ \frac{f'(W^*)}{b(W^*)} \right] = \mathbb{E} \left[ \frac{Wf(W)}{b(W)} \right]$$

for all absolutely continuous functions  $f: \mathbb{R} \rightarrow \mathbb{R}$  such that  $\mathbb{E}|Wf(W)/b(W)| < \infty$ , we say that  $W^*$  has the *b-generalized-zero-bias distribution* of  $W$ .

**Remark 3.2.** Goldstein and Reinert [7] study the case  $b(x) = 1$  and show that  $W^*$  has the same distribution as  $W$  if and only if  $W$  has a Gaussian distribution. Distributional fixed point characterizations for exponential, gamma and other nonnegative distributions and the connection with Stein's method have been studied in Peköz and Röllin [13], Peköz *et al.* [14,15].

**Remark 3.3.** By a routine extension of the proof of Proposition 2.1 of Chen *et al.* [5], it can be shown that there exists a unique distribution for  $W^*$ , and it is absolutely continuous with density

$$p^*(x) \propto b(x) \mathbb{E} \left[ \frac{W}{b(W)} 1_{W \geq x} \right].$$

We note in passing that the  $\sigma^2$  should be on the other side of the equality in the first display of Chen *et al.*'s proposition, which corresponds to  $b(x) = 1$ , the usual zero-bias distribution of  $W$ . The composition of the  $b$ -generalized-zero-bias transformation with the  $(1/b)$ -generalized-zero-bias transformation is the usual zero-bias transformation.

**Remark 3.4.** With  $\varphi$  the standard normal density and  $b$  a  $\varphi$ -integrable function, if  $W$  has density

$$p(x) = b(x)\varphi(x), \tag{8}$$

then its distribution is a fixed point for the  $b$ -generalized-zero-bias transformation since

$$p^*(x) = b(x) \int_x^\infty \frac{t}{b(t)} p(t) dt = b(x) \int_x^\infty t \varphi(t) dt = p(x).$$

The following result gives the  $b$ -generalized-zero-bias distribution of the uniform distribution on  $N$  points.

**Proposition 3.5.** *Given an integer  $N > 1$ , let  $x_1 > x_2 > \dots > x_N$  be such that  $b(x_n) > 0$  for all  $n$ . Let  $\mathbb{P}_N$  be the empirical distribution of the  $x_n$ :*

$$\mathbb{P}_N(A) = \frac{\#\{n: x_n \in A\}}{N}$$

*for any Borel set  $A \subset \mathbb{R}$ . Under the symmetry condition  $x_n = x_{N-n+1}$  for  $n = 1, \dots, N$ , the  $b$ -generalized-zero-bias distribution  $\mathbb{P}_N^*$  of  $\mathbb{P}_N$  is defined, and has density*

$$p^*(x) \propto b(x) \left[ \sum_{i=1}^n \frac{x_i}{b(x_i)} \right]$$

*for  $x_{n+1} < x \leq x_n$  ( $n = 1, \dots, N - 1$ ), and  $p^*(x) = 0$  if  $x > x_1$  or  $x \leq x_N$ .*

**Proof.** Immediate from Remark 3.3. □

Recall the following distances between distribution functions  $F$  and  $G$ . The Kolmogorov distance is

$$d_K(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|,$$



and the Wasserstein distance is

$$d_W(F, G) = \sup_{h \in \mathcal{H}} \left| \int_{\mathbb{R}} h dF - \int_{\mathbb{R}} h dG \right|,$$

where

$$\mathcal{H} = \{h: \mathbb{R} \rightarrow \mathbb{R} \text{ Lipschitz with } \|h'\| \leq 1\}$$

and  $\|\cdot\|$  is the supremum norm. Using Proposition 1.2 in Ross [16], these two metrics are seen to be related by

$$d_K(F, G) \leq \sqrt{2C d_W(F, G)}$$

if  $G$  has density bounded by  $C$ .

Restricting attention to the special case  $b(x) = x^2$ , we can now state our main result, along with an important corollary.

**Theorem 3.6.** *Suppose  $W^*$  is constructed on the same probability space as the zero-mean random variable  $W$  and is distributed according to the  $x^2$ -generalized-zero-bias distribution of  $W$ . Let  $M$  have the two-sided Maxwell density  $x^2 e^{-x^2/2} / \sqrt{2\pi}$ . Then there exist positive finite constants  $\lambda_1, \lambda_2, \lambda_3$  and  $\lambda_4$  such that*

$$\begin{aligned} d_W(\mathcal{L}(W), \mathcal{L}(M)) &\leq \lambda_1 \mathbb{E}|W - W^*| + \lambda_2 \mathbb{E}[|W| |W - W^*|] \\ &\quad + \lambda_3 \mathbb{E} \left| \frac{1}{W} - \frac{1}{W^*} \right| + \lambda_4 \mathbb{E} \left| 1 - \frac{W^*}{W} \right|. \end{aligned} \tag{9}$$

**Proof.** The inequality follows immediately from Theorem 4.4. Finiteness of the constants (along with explicit upper bounds) is detailed in Proposition 4.5.  $\square$

The following corollary gives a rate of convergence of the solution to (4) to the two-sided Maxwell distribution in terms of the Wasserstein distance; we postpone the proof until Section 4.3.

**Corollary 3.7.** *Suppose  $x_1, \dots, x_N$  is a monotonic, zero-mean, finite sequence of real numbers satisfying (4), let  $\mathbb{P}_N$  be the empirical distribution of these values, and let  $M$  be as in Theorem 3.6. Then there is a constant  $C > 0$  such that*

$$d_W(\mathbb{P}_N, \mathcal{L}(M)) \leq C \sqrt{\frac{\log N}{N}}.$$

## 4. The Stein equation and its solutions

### 4.1. General considerations on Stein's method and the problem with Stein's density approach

Let  $F$  and  $G$  be two cumulative distribution functions which one wishes to compare. Denote  $L^1(F)$  (resp.,  $L^1(G)$ ) the class of Borel measurable functions  $h: \mathbb{R} \rightarrow \mathbb{R}$  such that  $\int |h| dF < \infty$  (resp.,  $\int |h| dG < \infty$ ). A discrepancy measure between  $F$  and  $G$  is an *integral probability metric* if it can be written in the form

$$d_{\mathcal{H}}(F, G) := \sup_{h \in \mathcal{H}} \left| \int |h| dF - \int |h| dG \right|$$

for some class of test functions  $\mathcal{H} \subset L^1(F) \cap L^1(G)$ . The aforementioned Kolmogorov and Wasserstein distances are two important examples of integral probability metrics.

Suppose that  $F$  is absolutely continuous with density  $p$  on the real line, and introduce the operator  $h \rightarrow f_h$  which, to each  $h \in L^1(F)$ , assigns the function

$$f_h(x) = \frac{1}{p(x)} \int_x^\infty (h(u) - F(h)) p(u) du, \quad (10)$$

where  $F(h) = \int h dF$ . The integrability condition  $h \in L^1(F)$  guarantees that  $f_h$  is the unique absolutely continuous solution to the differential equation

$$f'_h(x) + \frac{p'(x)}{p(x)} f_h(x) = h(x) - F(h)$$

to also satisfy the boundary conditions  $\lim_{x \rightarrow \pm\infty} p(x) f_h(x) = 0$ . Under the assumption that  $\mathcal{H} \subset L^1(G)$ , we can integrate with respect to  $G$  on both sides of the differential equation (known, in the Stein community argot, as a ‘‘Stein equation’’) to get

$$d_{\mathcal{H}}(F, G) = \sup_{h \in \mathcal{H}} \left| \mathbb{E} \left[ f'_h(W) + \frac{p'(W)}{p(W)} f_h(W) \right] \right|, \quad (11)$$

with  $W$  a random variable distributed according to  $G$ . This last expression provides a means of bounding integral probability metrics (and thus in particular the Kolmogorov and Wasserstein distances) in terms of the action of a differential operator over a class of functions.

The steps outlined above form the basis of what is known as the ‘‘density approach’’ to Stein's method (see, e.g., Chatterjee and Shao [4]), which is the most intuitive extension of Stein's method of normal approximation (as described in Chen *et al.* [5]) to arbitrary continuous target distributions. In order for (11) to be of practical use, however, it is crucial that the functions  $p'/p$ ,  $f_h$  and  $f'_h$  be amenable to computations; it is particularly important that  $f_h$  and its first derivatives be bounded. Such conditions are not met in the case of the two-sided Maxwell distribution  $p(x) = x^2 \varphi(x)$  with which we are concerned in this paper. Indeed, for such a density, we have on the one hand  $p'(x)/p(x) = 2/x - x$  and, on the other hand,

$f_h(x) = x^{-2}e^{x^2/2} \int_x^\infty (h(u) - F(h))u^2e^{-u^2/2} du$ , both of which have a singularity at  $x = 0$ . Because of this, applying the classical Stein's method toolkit to the right-hand side of (11) will ultimately lead to trivial upper bounds and more elaborate methods need to be devised. This will be performed in the coming sections.

Before proceeding to the description of our proposal, we stress that the classical version of the ‘‘density approach’’ to Stein's method that we have just described actually breaks down for any target density  $p$  such that  $p(x_0) = 0$  at some  $x_0$  not on the edges of the support. Indeed, in general, the Stein solution (10) is the product of two terms: one term with a singularity wherever the density has a zero, and a second term that vanishes at the endpoints of the range of the target random variable. This results in the peculiar behavior of singularities inside the range of the random variable when the density has a zero there. Note that for the one-sided Maxwell distribution the solution  $f_h(x)$  has the same form as for its two-sided counterpart, though it is now only defined for  $x \geq 0$ ; since in this one-sided case when  $x = 0$  the second term  $\int_0^\infty (h(u) - F(h))u^2e^{-u^2/2} du$  vanishes, we would have  $f_h(0) = 0$ . This term doesn't vanish (unless  $h$  is an even function) for the two-sided Maxwell case, thus giving rise to the singularity at  $x = 0$ .

## 4.2. Coupling based Stein's method for densities of the form (8)

Let  $X$  be a random variable with probability density function  $p$  which we assume to be of the form (8). Let  $W$  be a symmetric random variable whose distribution we want to compare with that of  $X$ . First, we introduce the random variable  $W^*$  proposed in Definition 3.1 and write

$$\begin{aligned} \mathbb{E}[h(W)] - \mathbb{E}[h(X)] &= \mathbb{E}\left[\frac{f'(W)}{b(W)} - W\frac{f(W)}{b(W)}\right] \\ &= \mathbb{E}\left[\frac{f'(W)}{b(W)} - \frac{f'(W^*)}{b(W^*)}\right] \end{aligned} \quad (12)$$

for  $f = f_h$  solutions to the differential equation

$$\frac{f'(w)}{b(w)} - w\frac{f(w)}{b(w)} = h(w) - \mathbb{E}[h(X)]. \quad (13)$$

Taking suprema over all  $h \in \mathcal{H}$ , we deduce

$$d_{\mathcal{H}}(\mathcal{L}(W), \mathcal{L}(X)) = \sup_{h \in \mathcal{H}} \left| \mathbb{E}\left[\frac{f'(W)}{b(W)} - \frac{f'(W^*)}{b(W^*)}\right] \right| \quad (14)$$

for all  $\mathcal{H}$  such that the solutions to (13) are well-defined. Expression (14) provides an alternative to (11) which we will now prove to be useful to our purpose.

At this stage, the next typical ‘‘Stein-method’’ step is to write  $W^* = W + (W^* - W)$  and Taylor expand the integrand in (14) around  $W$  to deduce a bound on  $d_{\mathcal{H}}(\mathcal{L}(W), \mathcal{L}(X))$  expressed in terms of the difference between  $W^*$  and  $W$ . Unfortunately, for similar reasons as those described in Section 4.1, the solutions to (13) also have singularities which make this intuition unexploitable directly. We propose to bypass this difficulty by introducing intermediate functions

$\tau_X$  and  $g$  – to be defined later on in the text – for which

$$\begin{aligned} \mathbb{E}[h(W)] - \mathbb{E}[h(X)] &= \mathbb{E}\left[\frac{f'(W)}{b(W)} - \frac{f'(W^*)}{b(W^*)}\right] \\ &= \mathbb{E}\left[W(\tau_X(W) - 1)g(W) - W^*(\tau_X(W^*) - 1)g(W^*)\right] \\ &\quad + \mathbb{E}\left[\tau_X(W)g'(W) - \tau_X(W^*)g'(W^*)\right]. \end{aligned} \quad (15)$$

Bounding integral probability metrics  $d_{\mathcal{H}}(\cdot, \cdot)$  between  $\mathcal{L}(W)$  and  $\mathcal{L}(X)$  then boils down to finding bounds on the four terms provided in (15). Obviously this will only lead to reasonable results if the intermediate functions  $\tau_X$  and  $g$  are chosen wisely.

### 4.3. The Stein kernel equation for densities of the form (8)

We start by introducing the integral operator

$$h - \Phi(h) \mapsto \mathcal{T}_\varphi^{-1}(h - \Phi(h))(w) := \frac{1}{\varphi(w)} \int_w^\infty (h(u) - \Phi(h))\varphi(u) du \quad (16)$$

with  $\Phi(h) = \int h d\Phi$  and  $\Phi$  the standard Gaussian cumulative distribution function. (The notation  $\mathcal{T}_\varphi^{-1}$  is taken from Ley, Reinert and Swan [9].) We also introduce the function

$$\tau_X(x) = \frac{1}{p(x)} \int_x^\infty up(u) du,$$

which is called the ‘‘Stein kernel’’ of  $X$  (or, equivalently, of  $p$ ) – again we refer to Ley, Reinert and Swan [9] for intuition and first properties.

**Remark 4.1.** Stein kernels were introduced in Stein [17], Cacoullos and Papathanasiou [2], and have proven to be of great use in Gaussian analysis, see, for example, Nourdin and Peccati [11] and Chatterjee [3]. Their importance in the abstract approach to Stein’s method has been investigated in Döbler [6], where it is shown that they have a regularizing effect on the solutions to general Stein equations.

**Lemma 4.2.** *Let  $x \mapsto b(x)$  be a nonnegative even function with support a subset of  $(-\infty, \infty)$  and such that  $\lim_{x \rightarrow \pm\infty} b(x)\varphi(x) = 0$ . Suppose furthermore that  $b$  is absolutely continuous and integrable w.r.t.  $\varphi$  with integral  $\int_{-\infty}^\infty b(x)\varphi(x) dx = 1$ . Let  $X$  be a random variable with density  $x \mapsto b(x)\varphi(x)$ . Then*

$$\tau_X(x) = 1 + \frac{\mathcal{T}_\varphi^{-1}b'(x)}{b(x)} \quad (17)$$

*under the convention that the ratio is set to zero at all points  $x$  such that  $b(x) = 0$  and  $\mathcal{T}_\varphi^{-1}b'(x) \neq 0$ . Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be a Borel function such that  $E|h(X)| < \infty$ , and set  $\tilde{h} =$*

$h - \mathbb{E}[h(X)]$ . Then

$$g_h(x) = \frac{\int_x^\infty b(u)\tilde{h}(u)\varphi(u) du}{b(x)\varphi(x) + \int_x^\infty b'(u)\varphi(u) du} \quad (18)$$

is the unique solution  $g$  of the ODE

$$\tau_X(x)g'(x) - xg(x) = \tilde{h}(x), \quad (19)$$

which satisfies the asymptotic property  $\lim_{x \rightarrow \pm\infty} \tau(x)\varphi(x)b(x)g(x) = 0$ .

**Proof.** Integrating by parts in the definition of the Stein kernel for  $p = b\varphi$  we get (assuming that  $\lim_{x \rightarrow \pm\infty} b(x)\varphi(x) = 0$ )

$$\begin{aligned} \int_x^{+\infty} yp(y) dy &= \int_x^\infty b(y)(-\varphi'(y)) dy \\ &= b(x)\varphi(x) + \int_x^\infty b'(y)\varphi(y) dy \end{aligned}$$

so that (17) follows by definition (16) of the inverse Stein operator. For the second claim, we follow Nourdin and Peccati [12], Proposition 3.2.2, and note how

$$\tau_X(x)g'(x) - xg(x) = \frac{(\tau_X(x)g(x)p(x))'}{p(x)}$$

so that any solution to (19) has the form

$$g(x) = \frac{1}{\tau(x)p(x)} \int_{-\infty}^x \tilde{h}(u)p(u) du + \frac{d}{\tau(x)p(x)}, \quad (20)$$

where  $d \in \mathbb{R}$ . By dominated convergence, one infers that

$$\lim_{x \rightarrow \pm\infty} \int_{-\infty}^x \tilde{h}(y)b(y)\varphi(y) dy = 0,$$

so that the first summand in (20) has the announced form (18) and the asymptotic property is satisfied if and only if  $d = 0$ .  $\square$

Our next result provides the connection between the Stein equations (13) and (19).

**Lemma 4.3.** *Suppose that  $b$  only has isolated zeros. Let all notations be as above and introduce the function  $g = g_f$  defined at all  $x$  such that  $b(x) > 0$  through*

$$\frac{f'(x) - xf(x)}{b(x)} = \tau_X(x)g'(x) - xg(x).$$

Then

$$\begin{aligned} \mathbb{E}\left[\frac{f'(W)}{b(W)} - \frac{f'(W^*)}{b(W^*)}\right] &= \mathbb{E}\left[W(\tau_X(W) - 1)g(W) - W^*(\tau_X(W^*) - 1)g(W^*)\right] \\ &\quad + \mathbb{E}\left[\tau_X(W)g'(W) - \tau_X(W^*)g'(W^*)\right]. \end{aligned} \quad (21)$$

**Proof.** Since

$$\frac{f'(x) - xf(x)}{b(x)} = \frac{(f(x)\varphi(x))'}{b(x)\varphi(x)}$$

and

$$\tau_X(x)g'(x) - xg(x) = \frac{(b(x)\tau_X(x)g(x)\varphi(x))'}{b(x)\varphi(x)}$$

at all  $x$  for which  $b(x) \neq 0$ , we deduce that  $f$  and  $g$  are mutually defined by  $f = (b\tau_X)g$ . This in turn gives

$$\frac{f'(x)}{b(x)} = \left(\frac{b'(x)}{b(x)}\tau_X(x) + \tau_X'(x)\right)g(x) + \tau_X(x)g'(x) =: \psi(x)g(x) + \tau_X(x)g'(x),$$

which, combined with  $\psi(x) = x(\tau_X(x) - 1)$  (that is easily derived using the various definitions involved), leads to the useful identity

$$\frac{f'(x)}{b(x)} = x(\tau_X(x) - 1)g(x) + \tau_X(x)g'(x) \quad (22)$$

from which (21) is directly derived.  $\square$

Combining identities (12) and (21) we get (15), as promised. As already mentioned in the [Introduction](#), the price to pay for circumventing the singularities is the necessity to bound several additional quantities concerning the couplings we obtain. The explicit nature of the recursion described in Section 2 nevertheless allows us to compute the resulting quantities satisfactorily, leading to the bounds claimed in Theorem 3.6 and Corollary 3.7. This we perform in the Maxwell case in the next sections.

#### 4.4. Approximating the two-sided Maxwell distribution

**Theorem 4.4.** *Let  $p(x) = x^2\varphi(x)$ , and take  $f$  a solution to the Stein equation*

$$f'(w)/w^2 - wf(w)/w^2 = \tilde{h}(w), \quad (23)$$

where  $\tilde{h}$  is a function having bounded first derivative and zero-mean under  $p$ . Set  $c = \|\tilde{h}'\|$ . Then for any coupling of  $W$  and  $W^*$  on a joint probability space such that  $W^*$  has the  $x^2$ -generalized

zero biased distribution for  $W$ ,

$$\begin{aligned} \left| \mathbb{E} \left[ \frac{f'(W)}{(W)^2} - \frac{f'(W^*)}{(W^*)^2} \right] \right| &\leq \lambda_1 \mathbb{E} |W - W^*| + \lambda_2 \mathbb{E} [|W| |W - W^*|] \\ &\quad + \lambda_3 \mathbb{E} \left| \frac{1}{W} - \frac{1}{W^*} \right| + \lambda_4 \mathbb{E} \left| 1 - \frac{W^*}{W} \right| \end{aligned} \quad (24)$$

with

$$\lambda_1 \leq 6c, \quad \lambda_2 \leq 7c, \quad \lambda_3 \leq 18c \quad \text{and} \quad \lambda_4 \leq 22c. \quad (25)$$

**Proof.** With  $b(x) = x^2$  we have  $\tau_X(x) = 1 + 2/x^2$  and  $\psi(x) = 2/x$ , so that (21) becomes

$$\begin{aligned} &= \mathbb{E} \left[ \frac{2}{W^*} g(W^*) - \frac{2}{W} g(W) \right] + \mathbb{E} \left[ \left( 1 + \frac{2}{(W^*)^2} \right) g'(W^*) - \left( 1 + \frac{2}{(W)^2} \right) g'(W) \right] \\ &= 2 \mathbb{E} \left[ \left( \frac{1}{W^*} - \frac{1}{W} \right) g(W^*) \right] + 2 \mathbb{E} \left[ \frac{1}{W} (g(W^*) - g(W)) \right] \\ &\quad + \mathbb{E} [g'(W^*) - g'(W)] + 2 \mathbb{E} \left[ \frac{1}{(W^*)^2} g'(W^*) - \frac{1}{(W)^2} g'(W) \right]. \end{aligned}$$

The first two terms are dealt with easily to get

$$\begin{aligned} &2 \left| \mathbb{E} \left[ \left( \frac{1}{W^*} - \frac{1}{W} \right) g(W^*) \right] + 2 \mathbb{E} \left[ \frac{1}{W} (g(W^*) - g(W)) \right] \right| \\ &\leq 2 \|g\| \mathbb{E} \left[ \left| \frac{1}{W^*} - \frac{1}{W} \right| \right] + 2 \|g'\| \mathbb{E} \left[ \frac{1}{|W|} |W^* - W| \right]. \end{aligned}$$

For the last two terms, we introduce the function

$$\chi(x) = g'(x)/x$$

to get on the one hand

$$\begin{aligned} &\mathbb{E} [g'(W^*) - g'(W)] \\ &= \mathbb{E} \left[ W^* \frac{g'(W^*)}{W^*} - W \frac{g'(W)}{W} \right] \\ &= \mathbb{E} [(W^* - W) \chi(W^*)] + \mathbb{E} [W(\chi(W^*) - \chi(W))] \end{aligned}$$

so that

$$|\mathbb{E} [g'(W^*) - g'(W)]| \leq \|\chi\| \mathbb{E} [|W^* - W|] + \|\chi'\| \mathbb{E} [|W(W^* - W)|]$$

and, on the other hand

$$\begin{aligned} & \mathbb{E}\left[\frac{1}{(W^*)^2}g'(W^*) - \frac{1}{(W)^2}g'(W)\right] \\ &= \mathbb{E}\left[\frac{1}{W^*}\chi(W^*) - \frac{1}{W}\chi(W)\right] \\ &= \mathbb{E}\left[\left(\frac{1}{W^*} - \frac{1}{W}\right)\chi(W^*)\right] + \mathbb{E}\left[\frac{1}{W}(\chi(W^*) - \chi(W))\right] \end{aligned}$$

so that

$$2\left|\mathbb{E}\left[\frac{1}{(W^*)^2}g'(W^*) - \frac{1}{(W)^2}g'(W)\right]\right| \leq 2\|\chi\|\mathbb{E}\left[\left|\frac{1}{W^*} - \frac{1}{W}\right|\right] + 2\|\chi'\|\mathbb{E}\left[\frac{1}{|W|}|W^* - W|\right].$$

Combining these different estimates we obtain (24), with  $\lambda_1, \lambda_2, \lambda_3$  and  $\lambda_4$  expressed in terms of  $\|\chi\|, \|\chi'\|, \|g\|$  and  $\|g'\|$  as follows:

$$\lambda_1 = \|\chi\|, \quad \lambda_2 = \|\chi'\|, \quad \lambda_3 = 2(\|g\| + \|\chi\|) \quad \text{and} \quad \lambda_4 = 2(\|g'\| + \|\chi'\|).$$

The inequalities in (25) are proved in the Proposition 4.5 below.  $\square$

The next step is to bound  $\|\chi\|, \|\chi'\|, \|g\|$  and  $\|g'\|$  in a non trivial way; this we achieve in the next proposition.

**Proposition 4.5.** *Let  $h : \mathbb{R} \rightarrow \mathbb{R}$  be absolutely continuous and integrable with respect to  $p(x) = x^2\varphi(x)$ . Set  $c = \|h'\|$  which we suppose to be finite. Let  $X \sim p$ , define*

$$g_0(x) = \begin{cases} e^{x^2/2} \int_x^\infty y^2(h(y) - \mathbb{E}[h(X)])e^{-y^2/2} dy & \text{if } x > 0, \\ e^{x^2/2} \int_{-\infty}^x y^2(h(y) - \mathbb{E}[h(X)])e^{-y^2/2} dy & \text{if } x \leq 0, \end{cases} \quad (26)$$

and set

$$g(x) = \frac{g_0(x)}{x^2 + 2} \quad \text{and} \quad \chi(x) = \frac{g'(x)}{x}. \quad (27)$$

Then

$$\|g\| \leq 3c, \quad \|g'\| \leq 4c, \quad \|\chi\| \leq 6c \quad \text{and} \quad \|\chi'\| \leq 7c.$$

**Remark 4.6.** The function  $g_0$  defined in (26) satisfies

$$g_0'(x) - xg_0(x) = x^2(h(x) - \mathbb{E}[Z^2h(Z)]) \quad (28)$$

with  $Z \sim \varphi$  a standard Gaussian random variable.



**Remark 4.7.** The function  $g$  defined in (27) satisfies

$$\frac{(g(x)\tau(x)p(x))'}{p(x)} = h(x) - \mathbb{E}[h(X)]$$

with  $X \sim p$  and  $\tau(x) = 1 + 2/x^2$ .

**Proof.** In order to simplify future notations, we introduce  $\Phi(x) = \int_{-\infty}^x \varphi(t) dt$ ,  $\bar{\Phi}(x) = \int_x^{\infty} \varphi(t) dt$ ,  $\Upsilon(x) = e^{x^2/2} \int_x^{\infty} t^2 e^{-t^2/2} dt$  and  $\bar{\Upsilon}(x) = e^{x^2/2} \int_{-\infty}^x t^2 e^{-t^2/2} dt$ . Using the identity

$$\int_a^b t^2 e^{-t^2/2} dt = ae^{-a^2/2} - be^{-b^2/2} + \int_a^b e^{-t^2/2} dt, \quad -\infty \leq a < b \leq \infty, \quad (29)$$

we deduce that  $\Upsilon(x) = x + e^{x^2/2} \int_x^{\infty} e^{-t^2/2} dt$  and  $\bar{\Upsilon}(x) = -x + e^{x^2/2} \int_{-\infty}^x e^{-t^2/2} dt$  and thus

$$\Upsilon(x), \bar{\Upsilon}(x) \leq |x| + \sqrt{\frac{\pi}{2}} \quad \text{at all } x \in \mathbb{R} \quad \text{and} \quad \lim_{x \rightarrow \infty} \frac{\Upsilon(x)}{x} = \lim_{x \rightarrow -\infty} \frac{\bar{\Upsilon}(x)}{x} = 1. \quad (30)$$

The proof is now broken down into several steps.

*Step 1:* rewrite the solutions. Following Chen *et al.* [5], page 39, we rewrite the test functions in term of their derivatives (still with  $Z$  a standard normal random variable)

$$\begin{aligned} h(y) - E[h(X)] &= h(y) - E[Z^2 h(Z)] \\ &= \int_{-\infty}^{\infty} z^2 (h(y) - h(z)) \varphi(z) dz \\ &= \int_{-\infty}^y z^2 \left( \int_z^y h'(t) dt \right) \varphi(z) dz - \int_y^{\infty} z^2 \left( \int_y^z h'(t) dt \right) \varphi(z) dz. \end{aligned}$$

Changing the order of integration then using (29) leads to the rhs becoming

$$\begin{aligned} &\int_{-\infty}^y h'(t) \left[ \int_{-\infty}^t z^2 \varphi(z) dz \right] dt - \int_y^{\infty} h'(t) \left[ \int_t^{\infty} z^2 \varphi(z) dz \right] dt \\ &= \int_{-\infty}^y h'(t) \left[ -t\varphi(t) + \int_{-\infty}^t \varphi(z) dz \right] dt - \int_y^{\infty} h'(t) \left[ t\varphi(t) + \int_t^{\infty} \varphi(z) dz \right] dt \\ &= - \int_{-\infty}^{\infty} h'(t)t\varphi(t) dt + \int_{-\infty}^y h'(t)\Phi(t) dt - \int_y^{\infty} h'(t)\bar{\Phi}(t) dt, \end{aligned}$$

and thus

$$h(y) - E[h(X)] = \int_{-\infty}^y h'(t)\Phi(t) dt - \int_y^{\infty} h'(t)\bar{\Phi}(t) dt - E[Zh'(Z)]. \quad (31)$$

We deduce the following useful bound

$$\frac{h(x) - E[h(X)]}{x^2 + 2} \leq c \frac{2(x + 1/\sqrt{2\pi}) + \sqrt{2/\pi}}{x^2 + 2} \leq 2c. \quad (32)$$

Plugging (31) in (26) leads to (we restrict the discussion to  $x > 0$ , the other case following by symmetry)

$$\begin{aligned} g_0(x) &= -\mathbb{E}[Zh'(Z)]\Upsilon(x) && =: I(x) \\ &+ e^{x^2/2} \int_x^\infty \int_{-\infty}^y y^2 e^{-y^2/2} h'(t) \Phi(t) dt dy && =: II(x) \\ &- e^{x^2/2} \int_x^\infty \int_y^\infty y^2 e^{-y^2/2} h'(t) \bar{\Phi}(t) dt dy && =: III(x). \end{aligned}$$

To deal with the quantities  $II(x)$  and  $III(x)$ , we again interchange integrations to get

$$\begin{aligned} II(x) &= e^{x^2/2} \int_{-\infty}^x \left( \int_x^\infty y^2 e^{-y^2/2} dy \right) h'(t) \Phi(t) dt \\ &+ e^{x^2/2} \int_x^\infty \left( \int_t^\infty y^2 e^{-y^2/2} dy \right) h'(t) \Phi(t) dt \\ &= \Upsilon(x) \int_{-\infty}^x h'(t) \Phi(t) dt + e^{x^2/2} \int_x^\infty e^{-t^2/2} \Upsilon(t) h'(t) \Phi(t) dt \end{aligned}$$

and

$$\begin{aligned} III(x) &= e^{x^2/2} \int_x^\infty \left( \int_x^t y^2 e^{-y^2/2} dy \right) h'(t) \bar{\Phi}(t) dt \\ &= e^{x^2/2} \int_x^\infty (e^{-x^2/2} \Upsilon(x) - e^{-t^2/2} \Upsilon(t)) h'(t) \bar{\Phi}(t) dt \\ &= \Upsilon(x) \int_x^\infty h'(t) \bar{\Phi}(t) dt - e^{x^2/2} \int_x^\infty e^{-t^2/2} \Upsilon(t) h'(t) \bar{\Phi}(t) dt \end{aligned}$$

and thus if  $x \geq 0$  we have

$$\begin{aligned} g_0(x) &= -\mathbb{E}[Zh'(Z)]\Upsilon(x) + \Upsilon(x) \int_{-\infty}^x h'(t) \Phi(t) dt \\ &- \Upsilon(x) \int_x^\infty h'(t) \bar{\Phi}(t) dt + e^{x^2/2} \int_x^\infty e^{-t^2/2} \Upsilon(t) h'(t) dt. \end{aligned} \quad (33)$$

By a similar argument we deduce that if  $x < 0$  then

$$\begin{aligned} g_0(x) &= -\mathbb{E}[Zh'(Z)]\Upsilon(x) + \Upsilon(x) \int_{-\infty}^x h'(t)\bar{\Phi}(t) dt \\ &\quad - \Upsilon(x) \int_x^{\infty} h'(t)\Phi(t) dt + e^{x^2/2} \int_{-\infty}^x e^{-t^2/2}\Upsilon(t)h'(t) dt. \end{aligned} \quad (34)$$

*Step 2:* a bound on  $\|g\|$ . Supposing  $\|h'\| \leq c$  we can use (33) and the first claim in (30) to deduce that for  $x \geq 0$ :

$$\begin{aligned} |g_0(x)| &\leq cE|Z|(x + \sqrt{\pi/2}) + c(x + \sqrt{\pi/2}) \int_{-\infty}^x \Phi(t) dt \\ &\quad + c(x + \sqrt{\pi/2}) \int_x^{\infty} \bar{\Phi}(t) dt + ce^{x^2/2} \int_x^{\infty} e^{-t^2/2}(t + \sqrt{\pi/2}) dt. \end{aligned}$$

The last two terms decrease strictly to 0 as  $x \rightarrow \infty$ , with maximum value  $c/2$  and  $c(1 + \pi/2)$ , respectively. The first term is equal to  $c(\sqrt{2/\pi}x + 1)$  and the second one is equal to

$$\begin{aligned} c(x + \sqrt{\pi/2}) \int_{-\infty}^x \Phi(t) dt &= c(x + \sqrt{\pi/2})(x\Phi(x) + \varphi(x)) \\ &\leq c(x^2 + (\sqrt{\pi/2} + 1/\sqrt{2\pi})x + 1/2). \end{aligned}$$

Similar (symmetric) bounds hold for  $x \leq 0$  and thus, collecting all these estimates, we may conclude:

$$|g(x)| = \frac{|g_0(x)|}{x^2 + 2} \leq 3c. \quad (35)$$

*Step 3:* a bound on  $\|g'\|$ . Here we start by rewriting the derivative as

$$g'(x) = \frac{g'_0(x)}{x^2 + 2} - \frac{2x}{(x^2 + 2)^2} g_0(x). \quad (36)$$

Using (35), the second summand is easily seen to be uniformly bounded (by  $3c$ ). We are left with the first summand for which we start by rewriting the numerator, for  $x \geq 0$ , using (33):

$$\begin{aligned} g'_0(x) &= -\Upsilon'(x)\mathbb{E}[Zh'(Z)] + \Upsilon'(x) \int_{-\infty}^x h'(t)\Phi(t) dt \\ &\quad - \Upsilon'(x) \int_x^{\infty} h'(t)\bar{\Phi}(t) dt \\ &\quad + \Upsilon(x)(h'(x)\Phi(x) + h'(x)\bar{\Phi}(x)) \\ &\quad + xe^{x^2/2} \int_x^{\infty} \Upsilon(t)h'(t)e^{-t^2/2} dt - e^{x^2/2}\Upsilon(x)h'(x)e^{-x^2/2}, \end{aligned}$$

which leads to

$$\begin{aligned} g'_0(x) &= -\Upsilon'(x)\mathbb{E}[Zh'(Z)] \\ &\quad + \Upsilon'(x) \int_{-\infty}^x h'(t)\Phi(t) dt - \Upsilon'(x) \int_x^{\infty} h'(t)\bar{\Phi}(t) dt \\ &\quad + xe^{x^2/2} \int_x^{\infty} \Upsilon(t)h'(t)e^{-t^2/2} dt. \end{aligned} \quad (37)$$

Now we can use the fact that  $\Upsilon'(x) = xe^{x^2/2} \int_x^{\infty} e^{-t^2/2} dt \leq 1$  for all  $x \geq 0$  as well as all the arguments outlined at the previous step to deduce the bound:  $|g'_0(x)| \leq c(\sqrt{\frac{2}{\pi}} + 2(x+1/\sqrt{2\pi}) + \frac{1}{\sqrt{2\pi}}) \leq 2c(x+1)$  whence

$$\frac{|g'_0(x)|}{x^2+2} \leq c \frac{2x+2}{x^2+2} \leq c. \quad (38)$$

Similar (symmetric) arguments hold also for negative  $x$  and thus  $|g'(x)| \leq 4c$ .

*Step 4:* a bound on  $\chi(x) = g'(x)/x$ . Using (36), we know that

$$\chi(x) = \frac{g'_0(x)}{x(x^2+2)} - \frac{1}{(x^2+2)^2}g_0(x). \quad (39)$$

The second summand in (39) is bounded using (34) to get

$$\frac{1}{(x^2+2)^2}|g_0(x)| \leq 3c. \quad (40)$$

For the first summand, we use (37) to deduce

$$\begin{aligned} \frac{g'_0(x)}{x} &= -\frac{\Upsilon'(x)}{x}\mathbb{E}[Zh'(Z)] \\ &\quad + \frac{\Upsilon'(x)}{x} \int_{-\infty}^x h'(t)\Phi(t) dt - \frac{\Upsilon'(x)}{x} \int_x^{\infty} h'(t)\bar{\Phi}(t) dt \\ &\quad + e^{x^2/2} \int_x^{\infty} \Upsilon(t)h'(t)e^{-t^2/2} dt. \end{aligned}$$

At this stage it is useful to remark that, for  $x \geq 0$ , the function  $\Upsilon'(x)/x$  is strictly decreasing with maximal value  $\sqrt{\pi}/2$  and hence  $|\frac{g'_0(x)}{x}| \leq c(1 + 2(x + \frac{1}{\sqrt{2\pi}}) + \frac{1}{\sqrt{2\pi}}) \leq c(2x+3)$  and thus  $|\frac{g'_0(x)}{x(x^2+2)}| \leq 3c$  which, combined with (40), leads (after applying the symmetric arguments for  $x \leq 0$ ) to  $|\chi(x)| \leq 6c$ .

*Step 5:* a bound on  $\|\chi'\|$ . Direct computations using (28)

$$\chi(x) = \frac{1}{x^2+2} \left( 1 - \frac{1}{x^2+2} \right) g_0(x) - \frac{x}{x^2+2} (h(x) - \mathbb{E}[Z^2h(Z)])$$

and thus

$$\begin{aligned} \chi'(x) = & -\frac{2x^3}{(x^2+2)^2} \frac{g_0(x)}{x^2+2} + \left(1 - \frac{2}{x^2+2}\right) \frac{g_0'(x)}{x^2+2} \\ & - \frac{2-x^2}{x^2+2} \frac{h(x) - \mathbb{E}[Z^2 h(Z)]}{x^2+2} - \frac{x}{x^2+2} h'(x). \end{aligned}$$

Using the bounds  $|2x^3/(x^2+2)^2| \leq 1$ ,  $|1 - 2/(x^2+2)| \leq 1$ ,  $|(2-x^2)/(x^2+2)| \leq 1$  and  $|x/(x^2+2)| \leq 1$  as well as (35), (38) and (32) we conclude (after applying the symmetric arguments for  $x \leq 0$ )  $|\chi'(x)| \leq 7c$ .  $\square$

#### 4.5. Verifying bounds on expectations

In this section, we find bounds on the expectations in Theorem 3.6 in order to prove Corollary 3.7. We will make use of the following lemma.

**Lemma 4.8.** *If  $x_1, \dots, x_N$  is the unique strictly decreasing zero-mean solution of (4), then  $x_1 = O(\sqrt{\log N})$ .*

**Proof.** To simplify the notation, note that it suffices to consider the rescaled recursion  $x_{n+1}^3 = x_n^3 - S_n^{-1}$ , where  $S_n$  is defined in the proof of Lemma 2.1. By expressing  $x_1^3$  as a telescoping sum,

$$x_1^3 = \sum_{n=1}^{m-1} (x_n^3 - x_{n+1}^3) + x_m^3 = \sum_{n=1}^{m-1} S_n^{-1} + x_m^3 \leq \sum_{n=1}^{m-1} (n/x_1)^{-1} + x_m^3 \leq x_1(1 + \log m) + x_m^3,$$

where we have used Euler's approximation to the harmonic sum for the last inequality. By the variance property (P2) (in this rescaled case  $x_1^2 + \dots + x_N^2 = N - 1$ ), we have that  $x_1$  is bounded away from zero (as a sequence indexed by  $N$ ) and  $x_m$  is bounded, so  $x_m^2/x_1$  is bounded. Dividing the above display by  $x_1$ , we then obtain  $x_1 = O(\sqrt{\log N})$ .  $\square$

**Proof of Corollary 3.7.** From Proposition 3.5 and the recursion (4), note that  $p^*(x)$  puts mass  $1/(N-1)$  on each interval between successive  $x_n$ , so it is easy to create a coupling of  $W \sim \mathbb{P}_N$  with  $W^* \sim p^*(x)$  such that

$$|W - W^*| \leq |x_n - x_{n+1}|$$

when  $W \in [x_{n+1}, x_n]$ . For a detailed proof of such a coupling, see the construction given in McKeague and Levin [10]. From (P3) (see Lemma 2.1) and Lemma 4.8 we then have

$$\mathbb{E}|W - W^*| \leq \frac{1}{N-1} \sum_{n=1}^{N-1} (x_n - x_{n+1}) = \frac{2x_1}{N-1} = O\left(\frac{\sqrt{\log N}}{N}\right). \quad (41)$$

Second, using  $|W| \leq x_1 = O(\sqrt{\log N})$  it follows immediately that

$$\mathbb{E}[|W||W - W^*|] = O\left(\frac{\log N}{N}\right).$$

Third, the zero-median property gives

$$2x_m^3 = x_m^3 - x_{m+1}^3 = S_m^{-1} \geq (m/x_m)^{-1} = x_m/m,$$

where  $m = N/2 + 1$ , so  $x_m \geq 1/\sqrt{N}$ . By symmetry

$$\mathbb{E}\left|\frac{1}{W} - \frac{1}{W^*}\right| = \mathbb{E}\left|\frac{1}{W} - \frac{1}{W^*}\right| 1_{W^* \in (x_{m+1}, x_m]} + 2 \sum_{n=1}^{m-1} \mathbb{E}\left|\frac{1}{W} - \frac{1}{W^*}\right| 1_{W^* \in (x_{n+1}, x_n]}.$$

From Proposition 3.5, note that  $p^*(x) \propto x^2$  for  $x \in (x_{m+1}, x_m]$ . Also using the fact that  $p^*(x)$  puts mass  $1/(N-1)$  on this interval, the first term above can be written

$$\frac{6}{x_m^3(N-1)} \int_0^{x_m} \left(\frac{1}{x} - \frac{1}{x_m}\right) x^2 dx \leq \frac{3}{x_m(N-1)} = O\left(\frac{1}{\sqrt{N}}\right).$$

The second term is bounded above by the telescoping sum

$$\frac{2}{N-1} \sum_{n=1}^{m-1} \left(\frac{1}{x_{n+1}} - \frac{1}{x_n}\right) = \frac{2}{N-1} \left(\frac{1}{x_m} - \frac{1}{x_1}\right) = O\left(\frac{1}{\sqrt{N}}\right),$$

so we have

$$\mathbb{E}\left|\frac{1}{W} - \frac{1}{W^*}\right| = O\left(\frac{1}{\sqrt{N}}\right).$$

Fourth,

$$\mathbb{E}\left|1 - \frac{W^*}{W}\right| \leq \sqrt{N} \mathbb{E}|W - W^*| = O\left(\sqrt{\frac{\log N}{N}}\right)$$

using  $|W| \geq x_m \geq 1/\sqrt{N}$  and (41). The corollary now follows from Theorem 3.6.  $\square$

## Acknowledgements

The research of Ian McKeague was partially supported by NSF Grant DMS-1307838 and NIH Grant 2R01GM095722-05. The research of Yvik Swan was partially supported by the Fonds de la Recherche Scientifique – FRS under Grant no F.4539.16. We also thank the Institute for Mathematical Sciences at National University of Singapore for support during the *Workshop on New Directions in Stein's Method* (May 18–29, 2015) where work on the paper was initiated.

## References

- [1] Bohm, D. (1952). A suggested interpretation of the quantum theory in terms of “hidden” variables. I. *Phys. Rev.* (2) **85** 166–179. [MR0046287](#)
- [2] Cacoullos, T. and Papathanasiou, V. (1989). Characterizations of distributions by variance bounds. *Statist. Probab. Lett.* **7** 351–356. [MR1001133](#)
- [3] Chatterjee, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. [MR2449121](#)
- [4] Chatterjee, S. and Shao, Q.-M. (2011). Nonnormal approximation by Stein’s method of exchangeable pairs with application to the Curie–Weiss model. *Ann. Appl. Probab.* **21** 464–483. [MR2807964](#)
- [5] Chen, L.H.Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein’s Method. Probability and Its Applications (New York)*. Heidelberg: Springer. [MR2732624](#)
- [6] Döbler, C. (2015). Stein’s method of exchangeable pairs for the beta distribution and generalizations. *Electron. J. Probab.* **20** no. 109, 34. [MR3418541](#)
- [7] Goldstein, L. and Reinert, G. (1997). Stein’s method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7** 935–952. [MR1484792](#)
- [8] Hall, M.J.W., Deckert, D.A. and Wiseman, H.M. (2014). Quantum phenomena modeled by interactions between many classical worlds. *Phys. Rev. X* **4** 041013.
- [9] Ley, C., Reinert, G. and Swan, Y. (2017). Stein’s method for comparison of univariate distributions. *Probab. Surv.* **14** 1–52. [MR3595350](#)
- [10] McKeague, I.W. and Levin, B. (2016). Convergence of empirical distributions in an interpretation of quantum mechanics. *Ann. Appl. Probab.* **26** 2540–2555. [MR3543904](#)
- [11] Nourdin, I. and Peccati, G. (2009). Stein’s method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. [MR2520122](#)
- [12] Nourdin, I. and Peccati, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein’s Method to Universality. Cambridge Tracts in Mathematics* **192**. Cambridge: Cambridge Univ. Press. [MR2962301](#)
- [13] Peköz, E.A. and Röllin, A. (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.* **39** 587–608. [MR2789507](#)
- [14] Peköz, E.A., Röllin, A. and Ross, N. (2013). Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Probab.* **23** 1188–1218. [MR3076682](#)
- [15] Peköz, E.A., Röllin, A. and Ross, N. (2016). Generalized gamma approximation with rates for urns, walks and trees. *Ann. Probab.* **44** 1776–1816. [MR3502594](#)
- [16] Ross, N. (2011). Fundamentals of Stein’s method. *Probab. Surv.* **8** 210–293. [MR2861132](#)
- [17] Stein, C. (1986). *Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes – Monograph Series* **7**. Hayward, CA: IMS. [MR0882007](#)

Received November 2016 and revised March 2017