The Annals of Probability
2018, Vol. 46, No. 3, 1405–1440
https://doi.org/10.1214/17-AOP1205

MULTIVARIATE APPROXIMATION IN TOTAL VARIATION, II: DISCRETE NORMAL APPROXIMATION

BY A. D. BARBOUR[1], M. J. LUCZAK[2] AND A. XIA[3]

Universität Zürich, Queen Mary University of London and University of Melbourne

The paper applies the theory developed in Part I to the discrete normal approximation in total variation of random vectors in $\mathbb{Z}^d$. We illustrate the use of the method for sums of independent integer valued random vectors, and for random vectors exhibiting an exchangeable pair. We conclude with an application to random colourings of regular graphs.

1. Introduction. In Theorem 4.8 of Barbour, Luczak and Xia (2018) (Part I), we establish bounds on the total variation distance between the distribution of a random element $W \in \mathbb{Z}^d$ and the equilibrium distribution of a suitably chosen Markov population process X_n. In this paper, we show that the bounds are of order $O(n^{-1/2}\log n)$ as $n \to \infty$ if $W \sim \mathcal{DN}_d(nc, n\Sigma)$, for any $c \in \mathbb{R}^d$ and positive definite symmetric $d \times d$ matrix Σ, where the discrete normal distribution $\mathcal{DN}_d(nc, n\Sigma)$ is obtained from $\mathcal{N}_d(nc, n\Sigma)$ by assigning the probability of the d-box

$$[i_1 - 1/2, i_1 + 1/2) \times \cdots \times [i_d - 1/2, i_d + 1/2)$$

to the integer vector $(i_1, \dots, i_d)$, for each $(i_1, \dots, i_d) \in \mathbb{Z}^d$. From this, we deduce bounds for the discrete normal approximation of any random d-vector W.

To state Theorem 4.8 of Part I in the form that we shall need, we let $c \in \mathbb{R}^d$ be arbitrary, and $A, \sigma^2 \in \mathbb{R}^{d\times d}$ be such that A is a Hurwitz matrix (all its eigenvalues have negative real parts), and that σ^2 is positive definite and symmetric. We let Σ denote the positive definite solution of the continuous Lyapunov equation

$$A\Sigma + \Sigma A^T + \sigma^2 = 0; \tag{1.1}$$

Received December 2016.

[1]Work begun while ADB was Saw Swee Hock Professor of Statistics at the National University of Singapore, carried out in part at the University of Melbourne and at Monash University, and supported in part by Australian Research Council Grants Nos. DP120102728, DP120102398, DP150101459 and DP150103588.

[2]Work carried out in part at the University of Melbourne, and supported by an EPSRC Leadership Fellowship, grant reference EP/J004022/2, and in part by Australian Research Council Grants Nos. DP120102398 and DP150101459.

[3]Work supported in part by Australian Research Council Grants Nos. DP120102398 and DP150101459.

MSC2010 subject classifications. Primary 62E17; secondary 62E20, 60J27, 60C05.

Key words and phrases. Markov population process, multivariate approximation, total variation distance, infinitesimal generator, Stein's method.

for example, if $A = -I$, then $\Sigma = \frac{1}{2}\sigma^2$. We define an associated norm $|x|_\Sigma := \sqrt{x^T\Sigma^{-1}x}$. We then define an operator $\widetilde{\mathcal{A}}$ acting on functions $h\colon \mathbb{Z}^d \to \mathbb{R}$ by

$$\widetilde{\mathcal{A}}_n h(w) := \frac{n}{2}\operatorname{Tr}\bigl(\sigma^2\Delta^2 h(w)\bigr) + \Delta h^T(w)A(w-nc), \qquad w \in \mathbb{Z}^d, \tag{1.2}$$

where

$$\Delta_j h(w) := h\bigl(w+e^{(j)}\bigr) - h(w); \qquad \Delta_{jk}h(w) := \Delta_j(\Delta_k h)(w), \qquad 1 \le j,k \le d.$$

For $f : \mathbb{Z}^d \to \mathbb{R}$, we also define

$$\|f\|^\Sigma_{n\eta,\infty} := \max_{|X-nc|_\Sigma \le n\eta} \bigl|f(X)\bigr|, \tag{1.3}$$

with nc implicit. We then write $\|\Delta h\|^\Sigma_{n\eta,\infty}$ and $\|\Delta^2 h\|^\Sigma_{n\eta,\infty}$ for $\|f\|^\Sigma_{n\eta,\infty}$, when $f(X) = \max_{1\le j\le d}|\Delta_j h(X)|$ and $f(X) = \max_{1\le j,k\le d}|\Delta_{jk}h(X)|$, respectively. For a matrix M, we let $\|M\|$ denote its spectral norm; if it is positive definite and symmetric, we let $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$ denote its largest and smallest eigenvalues, $\rho(M)$ their ratio and $\mathrm{Sp}'(M) := \{\lambda_{\min}(M), \lambda_{\max}(M), d^{-1}\operatorname{Tr}(M)\}$.

THEOREM 1.1 (Theorem 4.8 of Part 1). *Given any c, A and σ^2 as above, there exists an associated sequence of Markov population processes $(X_n, n \ge 1)$, whose restriction X_n^δ to the $n\delta$-ball in $|\cdot|_\Sigma$ with centre nc, for $\delta \le \lambda_{\min}(\sigma^2)/(8\|A\|)$, has equilibrium distribution Π_n^δ concentrated near nc, which is almost the same for all δ. The closeness of $\mathcal{L}(W)$ in total variation to Π_n^δ, for any random vector W in $\mathbb{Z}^d$, can be checked as follows. For $\tilde\delta_0 = \min\{3, \lambda_{\min}(\sigma^2)/(8\|A\|\sqrt{\lambda_{\max}(\Sigma)})$, and for any $v > 0$ and $0 < \delta' < \frac{1}{2}\tilde\delta_0$, there exist constants $C_{1.1}(v,\delta')$ and $n_{1.1}(v,\delta')$, which are continuous functions of v, δ', $\|A\|/\overline{\Lambda}$, $\mathrm{Sp}'(\sigma^2/\overline{\Lambda})$ and $\mathrm{Sp}'(\Sigma)$, where $\overline{\Lambda} := d^{-1}\operatorname{Tr}(\sigma^2)$, but not of n, with the following property: if, for some $v > 0$, $0 < \delta' < \frac{1}{2}\tilde\delta_0$: $n \ge n_{1.1}(v,\delta')$ and $\varepsilon_1, \varepsilon_{20}, \varepsilon_{21}, \varepsilon_{22} > 0$,*

(i) $\mathbb{E}|W-nc|^2_\Sigma \le dvn$;

(ii) $d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathcal{L}(W+e^{(j)})\bigr) \le \varepsilon_1$, *for each* $1 \le j \le d$;

(iii)
$$\bigl|\mathbb{E}\{\widetilde{\mathcal{A}}_n h(W) I[|W-nc|_\Sigma \le n\delta'/3]\}\bigr| \le \overline{\Lambda}\bigl(\varepsilon_{20}\|h\|^\Sigma_{n\tilde\delta_0/4,\infty} + \varepsilon_{21}n^{1/2}\|\Delta h\|^\Sigma_{n\tilde\delta_0/4,\infty} + \varepsilon_{22}n\|\Delta^2 h\|^\Sigma_{n\tilde\delta_0/4,\infty}\bigr),$$

for all $h\colon \mathbb{Z}^d \to \mathbb{R}$, where $\widetilde{\mathcal{A}}_n$ is as defined in (1.2), *then, for any δ such that $2\delta' \le \delta \le \tilde\delta_0$,*

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \Pi_n^\delta\bigr) \le C_{1.1}(v,\delta')\bigl(d^3n^{-1/2} + d^4\varepsilon_1 + \varepsilon_{20} + d^{1/4}\varepsilon_{21} + d^{1/2}\varepsilon_{22}\bigr)\log n.$$

The accuracy of the approximation, for fixed c, A and σ^2, is thus of order $O(\log n\{n^{-1/2} + \varepsilon_1 + \varepsilon_{20} + \varepsilon_{21} + \varepsilon_{22}\})$, and is determined by how small the ε-quantities are. In Section 4, we give examples to show that they can all be of order

$O(n^{-1/2})$, giving an overall bound of order $O(n^{-1/2}\log n)$. The constant $C_{1.1}$ and the quantities $1/\tilde{\delta}_0$ and $d^{-1}\operatorname{Tr}(\sigma^2)$ depend on A and σ^2 in such a way that they do not grow with increasing dimension d, provided that the spectral norm of A and the eigenvalues of σ^2 and Σ remain bounded away from zero and infinity; more detail is given in Part I. Note, however, that n appears in the definition of $\widetilde{\mathcal{A}}_n$ only as a product with σ^2, and so can be chosen to prevent $\operatorname{Tr}(\sigma^2)$ and $\operatorname{Tr}(\Sigma)$ becoming large. Note also that the equilibrium distribution Π_n^δ remains the same if both A and σ^2 are multiplied by a common factor $a > 0$—this merely reflects a new choice of time scale—but the operator $\widetilde{\mathcal{A}}_n$ is multiplied by a. The factor $d^{-1}\operatorname{Tr}(\sigma^2))$ on the right-hand side of the inequality in Condition (iii) of Theorem 1.1 ensures that the constant $C_{1.1}(v, \delta')$ is the same for all choices of a.

The remainder of this paper completes two tasks. The first is to show that, if $W \sim \mathcal{DN}_d(nc, n\Sigma)$, then Conditions (i)–(iii) of Theorem 1.1 are satisfied with all the ε-quantities of order $O(n^{-1/2})$. As a result, Π_n^δ in the above theorem can be replaced by $\mathcal{DN}_d(nc, n\Sigma)$, giving the desired method of proving discrete normal approximation. The second is to show that the theorem can be applied in reasonable generality, yielding good rates of approximation. Note that there are many pairs (A, σ^2) that correspond to the same Σ, and the flexibility of having many pairs (A, σ^2) to use when approximating a single discrete normal distribution $\mathcal{DN}_d(nc, n\Sigma)$ represents a real advantage.

The structure of the paper is as follows. A brief taste of the results to be obtained is given in Section 1.1. In Section 2, the main discrete normal approximation, Theorem 2.4, is established, giving two conditions to be checked in order to conclude discrete normal approximation in total variation. If a "linear regression pair" can be found, these conditions can be substantially simplified; we give a corresponding result in Theorem 3.4 of Section 3. This theorem is applied, in Section 4, to sums of independent random vectors, and then in the more general context of exchangeable pairs, as developed in Stein (1986). We conclude with an application to the joint distribution of the numbers of monochrome edges in a graph colouring problem. A number of proofs that involve lengthy calculations are deferred to Section 5. The form of Theorem 2.4 also lends itself to use under assumptions of local dependence.

1.1. *Illustration.* Theorem 2.4 is somewhat forbidding. Before going into detail, we give a simple corollary of the theorem in the context of exchangeable pairs having the approximate linear regression property, and sketch an example.

Suppose that (W, W') is a pair of random integer valued d-vectors, defined on the same probability space, such that the pairs (W, W') and (W', W) have the same distribution. Assume that $\mathbb{E}\{|W|^3\} < \infty$, and write $\mu := \mathbb{E}W$. Let ξ denote the difference $W' - W$, so that $\mathbb{E}\xi = 0$, and set $\sigma^2 := \mathbb{E}\{\xi\xi^T\}$, assumed positive definite, and $\chi := \mathbb{E}\{|\xi|^3\}$. Assume that, for some $n > 0$ and for some Hurwitz matrix $A \in \mathbb{R}^{d\times d}$ with spectral norm $\|A\|$, we have

$$\begin{aligned}\mathbb{E}\{\xi \mid W\} &= n^{-1}A(W-\mu) + \{\|A\|/n\}^{1/2}R_1(W);\\ \sigma^2(W) &:= \mathbb{E}\{\xi\xi^T \mid W\}.\end{aligned} \tag{1.4}$$

Clearly, $\mathbb{E}\{R_1(W)\} = 0$. Write $L := (\|A\|/n)^{1/2}\chi\{\mathrm{Tr}(\sigma^2)\}^{-3/2}$, let Σ be the solution to (1.1), and assume that

$$\{\mathbb{E}|\Sigma^{-1/2}R_1(W)|^3\}^{1/3} \le \frac{\lambda_{\min}(\sigma^2)}{8\lambda_{\max}(\Sigma)}\sqrt{\frac{d}{2\|A\|}}.$$

Let $\mathcal{J}$ be the set of d-vectors such that $q^J := \mathbb{P}[\xi = J] > 0$. Suppose that $\mathcal{J}$ is finite, and that each of the coordinate vectors $e^{(j)}$, $1 \le j \le d$, can be obtained as a (finite) sum of elements of $\mathcal{J}$. For $Q^J(W) := \mathbb{P}[\xi = J \mid W]$, we set

$$u^J := (q^J)^{-1}\mathbb{E}|Q^J(W) - q^J|,$$

and $u^* := \max_{J\in\mathcal{J}} u^J$.

THEOREM 1.2. *Under the above circumstances, there exist constants n_0 and C, depending on d, σ^2, $\mathcal{J}$ and A, such that, if $n \ge n_0$, we have*

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{DN}_d(\mu, n\Sigma)) \le C\log n\{L(1 + n^{1/2}u^*) + \mathbb{E}|R_1(W)|\}.$$

The key elements in the bound are L, which is the analogue of the Lyapunov ratio appearing in the Berry–Esseen error bound, u^*, which can often be shown to be small by a variance calculation, and the inaccuracy of the linear regression (1.4), expressed by $\mathbb{E}|R_1(W)|$. In examples such as the one that follows, the resulting bound is of order $O(n^{-1/2}\log n)$. The theorem can be deduced from Theorem 3.4, Lemma 4.3 and Corollary 4.4.

As an example, suppose that G_n is an r-regular graph on n vertices. Let the vertices be coloured independently, each with one of m colours, the probability of choosing colour i being $p_i > 0$, $1 \le i \le m$. Let N_i denote the number of vertices having colour i, and let M_i denote the number of edges joining pairs of vertices that both have colour i. We are interested in approximating the joint distribution of

$$W := (M_1, \dots, M_m, N_1, \dots, N_{m-1}) =: (W_1, \dots, W_m, W_{m+1}, \dots, W_{2m-1}),$$

when n becomes large, while r, m and $p_1, \dots, p_m$ remain fixed; the detailed structure of G_n does not appear in the approximation. Multivariate normal approximation in a smooth metric was proved by Rinott and Rotar (1996), and in the convex sets metric by Chen, Goldstein and Shao [(2011), pages 333–334], both with error of order $O(n^{-1/2}\log n)$. Theorem 1.2 shows that the same order of error actually holds in total variation, provided that $m \ge 3$; the details are given in Section 4.2.1. For $m = 2$, the distribution of W is concentrated on a sub-lattice of $\mathbb{Z}^3$, so that discrete normal approximation is not good [but it can be deduced for the pair (M_1, N_1)]. The exchangeable pair is constructed by realizing W from a random colouring of the vertices, and then randomly re-colouring one of the vertices to give W'. The resulting regression is exact, implying that $R_1(w) = 0$ for all w. The set $\mathcal{J}$ is fixed and finite, so that $L = O(n^{-1/2})$, and, for each J, $\mathbb{E}(Q^J(W) - q^J)^2$ can simply be shown to be of order $O(n^{-1})$—the calculation is as for the variance of a sum of n very weakly dependent indicators. If $m \geq 3$, each coordinate vector $e^{(j)}$, $1 \leq j \leq 2m-1$, can be obtained as a sum of elements of $\mathcal{J}$, but this cannot be done if $m = 2$. The analogous problem, in which the proportions of vertices of each colour are held (almost) fixed, but randomly assigned to the vertices, can be treated in much the same way. The exchangeable pair is obtained by swapping the colours of two vertices, and the treatment of $\mathbb{E}(Q^J(W) - q^J)^2$ becomes a little messier.

2. Discrete normal approximation. In this section, we show that Theorem 1.1 can be used to establish approximation by distributions from the discrete normal family. To do so, we need first to establish properties of distributions in the family that are related to the conditions of Theorem 1.1. We always assume that $n \geq d^4$.

We first note the following simple lemma, proved in Section 5.1, in which moments of the discrete normal random variable $W \sim \mathcal{DN}_d(nc, n\Sigma)$ are bounded by expressions similar to those of $\mathcal{N}_d(nc, n\Sigma)$.

LEMMA 2.1. *For $l \in \mathbb{Z}_+$, we have*

$$\text{(a)} \quad \mathbb{E}|W - nc|_\Sigma^l \leq C(l)(nd)^{l/2},$$

whenever $n \geq 1/\lambda_{\min}(\Sigma)$, for universal constants $C(l)$ given in Section 5.1. In addition, for each $1 \leq j \leq d$ and $n \geq 1$,

$$\text{(b)} \quad \mathbb{E}(W_j - nc_j)^2 \leq \frac{1}{2} + 2n\Sigma_{jj},$$

and, for $l \in \mathbb{Z}_+$ and for universal constants $C'(l)$ given in Section 5.1,

$$\text{(c)} \quad \mathbb{E}\{[\Sigma^{-1}(W - nc)]_j^{2l}\} \leq n^l C'(l)(1 + (\Sigma^{-1})_{jj}^l),$$

whenever $n \geq d/\{4(\lambda_{\min}(\Sigma))^2\}$.

The next lemma, proved in Section 5.2, establishes an approximate integration by parts formula for multivariate discrete normal distributions. We write $I_n^\eta(X) := I[|X - nc|_\Sigma \leq n\eta/3]$ for any $\eta > 0$, and we say that $C \in \mathcal{K}_\Sigma$ if C is an increasing function of $\lambda_{\max}(\Sigma)$, $1/\lambda_{\min}(\Sigma)$, and $C(\delta) \in \mathcal{K}_\Sigma(\delta)$ if $C(\delta) \in \mathcal{K}_\Sigma$ for each fixed δ. We also define

$$\psi_\Sigma(n) := \frac{6}{n\sqrt{\lambda_{\min}(\Sigma)}}, \tag{2.1}$$

noting that its inverse is $\psi_\Sigma^{-1}(\delta) = \psi_\Sigma(\delta)$.

LEMMA 2.2. *Suppose that $W \sim \mathcal{DN}_d(nc, n\Sigma)$. Then there exist constants $n_{2.2} \in \mathcal{K}_\Sigma$ and $C^{(1)}_{2.2}(\delta), C^{(2)}_{2.2}(\delta), C^{(3)}_{2.2}(\delta) \in \mathcal{K}_\Sigma(\delta)$, such that, for any $n \ge \max\{n_{2.2}, \psi_\Sigma(\delta)\}$ and for any function $f : \mathbb{Z}^d \to \mathbb{R}$, we have*:

(a) $$\bigl|\mathbb{E}\{\Delta f(W)^T b I_n^\delta(W)\} - n^{-1}\mathbb{E}\{(f(W)(W-nc)^T \Sigma^{-1} b I_n^\delta(W)\}\bigr| \le d^{1/2} C^{(1)}_{2.2}(\delta) n^{-1} |b|_1 \|f\|^\Sigma_{n\delta/2,\infty};$$

(b) $$\begin{aligned}&\bigl|\mathbb{E}\{\Delta f(W)^T B(W-nc) I_n^\delta(W)\}\\ &\quad - \mathbb{E}\{f(W)[n^{-1}(W-nc)^T \Sigma^{-1} B(W-nc) - \operatorname{Tr} B] I_n^\delta(W)\}\bigr|\\ &\quad\le d^{1/2} C^{(2)}_{2.2}(\delta) n^{-1/2} \|B\|_1 \|f\|^\Sigma_{n\delta/2,\infty} + \sum_{j=1}^d |B_{jj}| \|\Delta f\|^\Sigma_{n\delta/2,\infty};\end{aligned}$$

(c) $$\begin{aligned}&Z\bigl|\mathbb{E}\{\Delta f(W)^T B(W-nc) I_n^\delta(W)\}\\ &\quad - \mathbb{E}\{f(W)[n^{-1}(W-nc)^T \Sigma^{-1} B(W-nc) - \operatorname{Tr} B] I_n^\delta(W)\}\bigr|\\ &\quad\le d C^{(3)}_{2.2}(\delta) n^{-1/2} \sum_{j=1}^d \bigl|(e^{(j)})^T B\bigr| \|f\|^\Sigma_{n\delta/2,\infty} + \sum_{j=1}^d |B_{jj}| \|\Delta f\|^\Sigma_{n\delta/2,\infty},\end{aligned}$$

for any d-vector b and any $d \times d$ matrix B. The constants $n_{2.2}$, $C^{(1)}_{2.2}(\delta)$, $C^{(2)}_{2.2}(\delta)$ and $C^{(3)}_{2.2}(\delta)$ are defined in (5.8), (5.9), (5.15) *and following* (5.16), *respectively.*

With the help of the lemmas above, we can now show that, if W has the discrete normal distribution $\mathcal{DN}_d(nc, n\Sigma)$, then it satisfies the conditions of Theorem 1.1, with $\varepsilon_1 \le c_1 n^{-1/2}$, $\max\{\varepsilon_{20}, \varepsilon_{21}\} \le c_2 d^{5/2} n^{-1/2}$ and $\varepsilon_{22} = 0$, and hence that the conditions of Theorem 1.1 imply a bound on the error of approximating the distribution of a random d-vector by $\mathcal{DN}_d(nc, n\Sigma)$.

THEOREM 2.3. *For Σ positive definite, suppose that σ^2, positive definite, and A are such that $A\Sigma + \Sigma A^T + \sigma^2 = 0$; write $\overline{\Lambda} := d^{-1}\operatorname{Tr}(\sigma^2)$. Then, if $W \sim \mathcal{DN}_d(nc, n\Sigma)$, for any $n \ge \max\{n_{2.2}, \psi_\Sigma(\delta)\}$, we have*:

(i) $\mathbb{E}|W - nc|^2_\Sigma \le dC(2)n$;

(ii) $d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \le C^{(1)}_{2.3} n^{-1/2}$ *for each $1 \le j \le d$*;

(iii) $$\bigl|\mathbb{E}\{\widetilde{\mathcal{A}}_n h(W) I[|W - nc|_\Sigma \le n\delta/3]\}\bigr| \le d^{5/2} n^{-1/2} \overline{\Lambda} C^{(2)}_{2.3}(\delta)\bigl(\|h\|^\Sigma_{n\delta/2,\infty} + n^{1/2}\|\Delta h\|^\Sigma_{n\delta/2,\infty}\bigr),$$

where $\widetilde{\mathcal{A}}_n$ is as defined in (1.2), *$C(2)$ is as in Lemma* 2.1, *and $C^{(1)}_{2.3}$ and $C^{(2)}_{2.3}(\delta)$ are continuous functions of $\|A\|/\overline{\Lambda}$, $\operatorname{Sp}'(\sigma^2/\overline{\Lambda})$ and $\operatorname{Sp}'(\Sigma)$; $C^{(1)}_{2.3}$ is given in* (2.2),

$C^{(2)}_{2.3}(\delta)$ *implicitly in* (2.8). *Hence a random d-vector satisfying the conditions of Theorem* 1.1 *with $n \ge n_{2.3}$ and $0 < \delta < \frac{1}{2}\tilde{\delta}_0(A, \sigma^2)$ has*

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathcal{DN}_d(nc, n\Sigma)\big) \\ &\qquad \le C_{2.3}(v, \delta)\big(d^4\big(n^{-1/2} + \varepsilon_1\big) + \varepsilon_{20} + d^{1/4}\varepsilon_{21} + d^{1/2}\varepsilon_{22}\big)\log n, \end{aligned}$$

with

$$\begin{aligned} C_{2.3}(v, \delta) &:= C_{1.1}(v, \delta) + C_{1.1}\big(C(2), \delta\big)\big(1 + C^{(1)}_{2.3} + C^{(2)}_{2.3}(\delta)\big); \\ n_{2.3} &:= \max\big\{n_{1.1}(v, \delta), n_{2.2}, \psi_\Sigma(\delta)\big\}. \end{aligned}$$

PROOF. Part (i) is immediate from (5.1), with $v = C(2)$. For Part (ii), we pick $\delta = 1$, and then take $b = e^{(j)}$ and any function f with $\|f\|_\infty \le 1$ in Lemma 2.2(a). This gives

$$\mathbb{E}\big|\Delta_j f(W) I^1_n(W)\big| \le d^{1/2} C^{(1)}_{2.2}(1) n^{-1} + n^{-1/2}\sqrt{C'(1)\big(1 + \big(\Sigma^{-1}\big)_{jj}\big)},$$

in view of Lemma 2.1(c). For the remaining part of $|\mathbb{E}\{\Delta_j f(W)\}|$, using $\|\Delta_j f\|_\infty \le 2$, we have

$$\mathbb{E}\big|\Delta_j f(W) I\big[|W - nc|_\Sigma > n/3\big]\big| \le 18 d C(2)/n,$$

by Chebyshev's inequality and from Part (i), and the estimate follows because $n \ge d^2$, with

$$C^{(1)}_{2.3} := C^{(1)}_{2.2}(1) + \sqrt{C'(1)\big(1 + \big(\Sigma^{-1}\big)_{jj}\big)} + 18 C(2). \tag{2.2}$$

For Part (iii), we use Lemma 2.2(b). This gives

$$\begin{aligned} &\big|\mathbb{E}\big\{\Delta h(W)^T A(W - nc) I^\delta_n(W)\big\} \\ &\qquad - \mathbb{E}\big\{h(W)\big[n^{-1}(W - nc)^T \Sigma^{-1} A(W - nc) - \operatorname{Tr} A\big] I^\delta_n(W)\big\}\big| \\ &\qquad \le d^{1/2} C^{(2)}_{2.2}(\delta) n^{-1/2} \|A\|_1 \|h\|^\Sigma_{n\delta/2,\infty} + \sum_{j=1}^d |A_{jj}| \|\Delta h\|^\Sigma_{n\delta/2,\infty}. \end{aligned} \tag{2.3}$$

Then, since

$$\operatorname{Tr}\big(\sigma^2 \Delta^2 h(W)\big) = \sum_{i=1}^d \sum_{j=1}^d \sigma^2_{ij} \Delta_j f_i(W),$$

where $f_i(W) := \Delta_i h(W)$, it follows from Lemma 2.2(a), with $f = f_i$ and with b the ith column of σ^2, that

$$\begin{aligned}
&\Bigg| n\mathbb{E}\{\mathrm{Tr}(\sigma^2\Delta^2 h(W)) I_n^\delta(W)\} \\
&\qquad - \mathbb{E}\Bigg\{\sum_{i=1}^d\sum_{j=1}^d \sigma_{ij}^2 \Delta_i h(W)\{\Sigma^{-1}(W-nc)\}_j I_n^\delta(W)\Bigg\}\Bigg| \\
&\qquad \le d^{1/2} C_{2.2}^{(1)}(\delta)\|\sigma^2\|_1 \|\Delta h\|_{n\delta/2,\infty}^{\Sigma};
\end{aligned} \tag{2.4}$$

note also that

$$\begin{aligned}
&\mathbb{E}\Bigg\{\sum_{i=1}^d\sum_{j=1}^d \sigma_{ij}^2 \Delta_i h(W)\{\Sigma^{-1}(W-nc)\}_j I_n^\delta(W)\Bigg\} \\
&\qquad = \mathbb{E}\{\Delta h(W)^T\sigma^2\Sigma^{-1}(W-nc) I_n^\delta(W)\}.
\end{aligned} \tag{2.5}$$

But now, from Lemma 2.2(c),

$$\begin{aligned}
&\big|\mathbb{E}\{\Delta h(W)^T\sigma^2\Sigma^{-1}(W-nc) I_n^\delta(W)\} \\
&\quad - \mathbb{E}\{h(W)[n^{-1}(W-nc)^T\Sigma^{-1}\sigma^2\Sigma^{-1}(W-nc) - \mathrm{Tr}(\sigma^2\Sigma^{-1})] I_n^\delta(W)\}\big| \\
&\quad \le d C_{2.2}^{(3)}(\delta) n^{-1/2} \sum_{j=1}^d |(e^{(j)})^T\Sigma^{-1}\sigma^2| \|h\|_{n\delta/2,\infty}^{\Sigma} \\
&\qquad + \sum_{j=1}^d |[\sigma^2\Sigma^{-1}]_{jj}| \|\Delta h\|_{n\delta/2,\infty}^{\Sigma} \\
&\quad \le d C_{2.2}^{(3)}(\delta) n^{-1/2}\{\lambda_{\min}(\Sigma)\}^{-1}\|\sigma^2\|_1 \|h\|_{n\delta/2,\infty}^{\Sigma} \\
&\qquad + \|\sigma^2\Sigma^{-1}\|_1 \|\Delta h\|_{n\delta/2,\infty}^{\Sigma}.
\end{aligned} \tag{2.6}$$

Hence, and since

$$\|A\|_1 \le d^{3/2}\|A\|; \qquad \|\sigma^2\|_1 \le d^{3/2}\lambda_{\max}(\sigma^2)$$

and

$$\|\sigma^2\Sigma^{-1}\|_1 \le d^{3/2}\|\sigma^2\Sigma^{-1}\| \le d^{3/2}\lambda_{\max}(\sigma^2)/\lambda_{\min}(\Sigma),$$

it follows from (2.3), (2.4) and (2.6) that

$$
\begin{aligned}
&\mathbb{E}\{\widetilde{\mathcal{A}}_n h(W) I_n^\delta(W)\} \\
&\quad = \mathbb{E}\biggl\{\Bigl(\mathrm{Tr}\bigl\{A(W-nc)\Delta h(W)^T\bigr\} + \frac{1}{2} n\,\mathrm{Tr}\bigl\{\sigma^2 \Delta^2 h(W)\bigr\}\Bigr) I_n^\delta(W)\biggr\} \\
&\quad = \mathbb{E}\biggl\{h(W)\biggl[\frac{1}{2} n^{-1}(W-nc)^T\bigl(2\Sigma^{-1}A + \Sigma^{-1}\sigma^2\Sigma^{-1}\bigr)(W-nc) \\
&\qquad - \mathrm{Tr}\,A - \frac{1}{2}\,\mathrm{Tr}\bigl(\sigma_\Sigma^2\bigr)\biggr] I_n^\delta(W)\biggr\} + \theta,
\end{aligned}
\tag{2.7}
$$

where

$$
\begin{aligned}
|\theta| &\le d^{1/2} C_{2.2}^{(2)}(\delta) n^{-1/2} \|A\|_1 \|h\|_{n\delta/2,\infty}^\Sigma \\
&\quad + \|A\|_1 \|\Delta h\|_{n\delta/2,\infty}^\Sigma + \frac{1}{2} d^{1/2} C_{2.2}^{(1)}(\delta) \|\sigma^2\|_1 \|\Delta h\|_{n\delta/2,\infty}^\Sigma \\
&\quad + \frac{1}{2} d C_{2.2}^{(3)}(\delta) n^{-1/2} \bigl\{\lambda_{\min}(\Sigma)\bigr\}^{-1} \|\sigma^2\|_1 \|h\|_{n\delta/2,\infty}^\Sigma \\
&\quad + \|\sigma^2\Sigma^{-1}\|_1 \|\Delta h\|_{n\delta/2,\infty}^\Sigma \\
&\le d^{5/2} n^{-1/2} \overline{\Lambda} C_{2.3}^{(2)}(\delta)\bigl(\|h\|_{n\delta/2,\infty}^\Sigma + n^{1/2}\|\Delta h\|_{n\delta/2,\infty}^\Sigma\bigr),
\end{aligned}
\tag{2.8}
$$

and $C_{2.3}^{(2)}(\delta)$ is a function of $\|A\|/\overline{\Lambda}$ and the elements of $\mathrm{Sp}'(\Sigma)$, $\mathrm{Sp}'(\sigma^2/\overline{\Lambda})$.

Finally, for any y and B, we have $y^T B y = y^T B^T y = \frac{1}{2} y^T (B + B^T) y$, so that

$$
\begin{aligned}
y^T\bigl(2\Sigma^{-1}A + \Sigma^{-1}\sigma^2\Sigma^{-1}\bigr)y &= y^T\bigl(\Sigma^{-1}A + A^T\Sigma^{-1} + \Sigma^{-1}\sigma^2\Sigma^{-1}\bigr)y \\
&= y^T\Sigma^{-1}\bigl(A\Sigma + \Sigma A^T + \sigma^2\bigr)\Sigma^{-1}y = 0,
\end{aligned}
$$

from (1.1), and

$$
\mathrm{Tr}\bigl(\sigma_\Sigma^2\bigr) = -\,\mathrm{Tr}\bigl(\Sigma^{-1/2}A\Sigma^{1/2} + \Sigma^{1/2}A^T\Sigma^{-1/2}\bigr) = -2\,\mathrm{Tr}\,A.
$$

This, with (2.7), establishes that

$$
\begin{aligned}
&\bigl|\mathbb{E}\{\widetilde{\mathcal{A}}_n h(W) I_n^\delta(W)\}\bigr| \\
&\qquad \le d^{5/2} n^{-1/2} C_{2.3}^{(2)}(\delta)\bigl\{\|h\|_{n\delta/2,\infty}^\Sigma + n^{1/2}\|\Delta h\|_{n\delta/2,\infty}^\Sigma\bigr\},
\end{aligned}
\tag{2.9}
$$

as required. The final conclusion follows from the triangle inequality. □

Discrete normal approximation using Theorem 2.3 involves checking the conditions of Theorem 1.1. These can be replaced with analogous conditions in which the norm $|\cdot|_\Sigma$ is replaced by the Euclidean norm. Here, the parameter n is also chosen to standardize $d^{-1}\,\mathrm{Tr}(\Sigma)$; we omit the routine proof.

THEOREM 2.4. *Let W be a random vector in $\mathbb{Z}^d$ with mean $\mu := \mathbb{E}W$ and positive definite covariance matrix $V := \mathbb{E}\{(W - \mu)(W - \mu)^T\}$; define $n := \lceil d^{-1}\operatorname{Tr} V\rceil$, $c := n^{-1}\mu$ and $\Sigma := n^{-1}V$. Let A be a $d \times d$ Hurwitz matrix such that $\sigma^2 := -(A\Sigma + \Sigma A^T)$ is positive definite, and write $\overline{\Lambda} := d^{-1}\operatorname{Tr}\sigma^2$. Set*

$$\eta_0 := \frac{1}{6}\min\left\{\sqrt{\lambda_{\min}(\Sigma)}, \frac{\lambda_{\min}(\sigma^2)}{24\|A\|\sqrt{\rho(\Sigma)}}\right\} = \frac{1}{6}\tilde{\delta}_0\sqrt{\lambda_{\min}(\Sigma)}.$$

Then, for any $0 < \eta \le \eta_0$, there exist continuous functions $C_{2.4}(\eta), n_{2.4}(\eta)$ of $\|A\|/\overline{\Lambda}$, $\mathrm{Sp}'(\sigma^2/\overline{\Lambda})$, $\mathrm{Sp}'(\Sigma)$ and η, not depending on d or n, with the following property: if, for some ε_1, ε_{20}, ε_{21} and ε_{22}, and for some $n \ge n_{2.4}(\eta)$,

(a) $d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \le \varepsilon_1$ *for each* $1 \le j \le d$;

(b) $$\begin{aligned}&|\mathbb{E}\{\widetilde{\mathcal{A}}_n h(W)\}I[|W - nc| \le n\eta]| \\ &\quad\le \overline{\Lambda}(\varepsilon_{20}\|h\|_{3n\eta_0/2,\infty} + \varepsilon_{21}n^{1/2}\|\Delta h\|_{3n\eta_0/2,\infty} + \varepsilon_{22}n\|\Delta^2 h\|_{3n\eta_0/2,\infty}),\end{aligned}$$

for all $h\colon \mathbb{Z}^d \to \mathbb{R}$, then it follows that

$$\begin{aligned}&d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{DN}_d(nc, n\Sigma)) \\ &\quad\le C_{2.4}(\eta)(d^3 n^{-1/2} + d^4\varepsilon_1 + \varepsilon_{20} + d^{1/4}\varepsilon_{21} + d^{1/2}\varepsilon_{22})\log n.\end{aligned}$$

The estimate required in Condition (b), apart from the truncation to $|W - nc| \le n\eta/6$, is typical of those that are needed for multivariate normal approximation using Stein's method. The extra work needed, to translate multivariate normal approximation into discrete normal approximation in total variation, lies in establishing Condition (a) with a suitably small ε_1. Since, from Theorem 2.3, Condition (a) is satisfied with $\varepsilon_1 = O(n^{-1/2})$ if $W \sim \mathcal{DN}_d(nc, n\Sigma)$ and Σ is nonsingular, the triangle inequality for a general W yields

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \le 2d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{DN}_d(nc, n\Sigma)) + O(n^{-1/2}),$$

so that $d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)}))$ has to be small if total variation approximation of $\mathcal{L}(W)$ by the discrete normal is to be accurate.

We make some effort to make explicit the typical dependence of the error bounds on the dimension d. This is largely for comparison with the error bounds derived by Bentkus (2003) and Fang (2014) for approximation, with respect to the convex sets metric, of standardized sums of independent random vectors by the standard d-dimensional normal distribution. Here, since multiplicative standardization makes no sense in the domain of random vectors with integer coordinates, there are more quantities than just dimension that may affect the sizes of the approximation errors. Nonetheless, we attempt some comparison with the above approximations. To do so, we think of many quantities, such as the eigenvalues of σ^2,

A and Σ, as being bounded away from zero and infinity as d varies, and the traces of these matrices thus being thought of as having order d. This is because, in the standardized setting, using the Stein approach as in Götze (1991) or Fang (2014), one has $\sigma^2 = 2I$, $A = -I$ and $\Sigma = I$. Our bounds then also involve the values of other parameters, in particular $\|A\|$ and the elements of $\mathrm{Sp}'(\sigma^2)$ and $\mathrm{Sp}'(\Sigma)$, in a way that can be deduced from our arguments, but that we do not attempt to make explicit, other than that their dependence on these parameters is continuous. However, we always work in terms of approximations for fixed values of n and the parameters of a problem, so that implicit orders of magnitude play no direct part in the results that we obtain.

3. Linear regression pairs. In this section, we establish a discrete normal approximation theorem for the distribution of a random vector W, when a copy W' can be defined on the same probability space, in such a way that $\mathbb{E}\{W' \mid W\}$ is approximately a linear function of W. There are many examples where this is the case, including those given in Rinott and Rotar (1996) and Reinert and Röllin (2009).

Suppose then that (W, W') is a pair of random integer valued d-vectors, defined on the same probability space and having the same distribution. Assume that $\mathbb{E}\{|W|^3\} < \infty$, and write $\mu := \mathbb{E}W$. Let ξ denote the difference $W' - W$, so that $\mathbb{E}\xi = 0$, and set $\sigma^2 := \mathbb{E}\{\xi\xi^T\}$, assumed positive definite. Suppose that ξ exhibits an almost linear regression on W, and that the conditional variance $\sigma^2(W) := \mathbb{E}\{\xi\xi^T \mid W\}$ is more or less constant as a function of W. Specifically, assume that, for some $n > 0$ and for some $d \times d$ Hurwitz matrix A with spectral norm $\|A\|$, we have

$$\begin{aligned} \mathbb{E}\{\xi \mid W\} &= n^{-1}A(W-\mu) + n^{-1/2}\|A\|^{1/2}R_1(W); \\ \sigma^2(W) &:= \mathbb{E}\{\xi\xi^T \mid W\} = \sigma^2 + R_2(W), \end{aligned} \tag{3.1}$$

where $\mathbb{E}|R_1(W)|$ and $\mathbb{E}\|R_2(W)\|_1$ are to be thought of as small. These two quantities appear explicitly in the bound on the error in our discrete normal approximation, and clearly, $\mathbb{E}\{R_1(W)\} = 0$ and $\mathbb{E}\{R_2(W)\} = 0$. Let Σ be the positive definite solution to $A\Sigma + \Sigma A^T + \sigma^2 = 0$.

REMARK 3.1. Note that, in (3.1), multiplying n and A by the same positive constant c does not change the regression, but Σ is divided by c. This leaves both $n\Sigma$, the asymptotic approximation to $\mathrm{Var}\, W$, and $\|A\|/n$ unchanged, the latter implying that $R_1(W)$ remains the same also. The effective data for the problem are the distributions of ξ and W, and in particular σ^2 and $\mathrm{Var}\, W$, and also $\widehat{A} := A/n$, which is typically "small." In order to circumvent the indeterminacy, one can compute $\widehat{\Sigma} := n\Sigma$, typically "large," by solving $\widehat{A}\widehat{\Sigma} + \widehat{\Sigma}\widehat{A}^T + \sigma^2 = 0$. Then $\tilde{n} := n/\|A\|$, $\widetilde{A} := \tilde{n}\widehat{A} = A/\|A\|$ and $\widetilde{\Sigma} := \widehat{\Sigma}/\tilde{n}$ are the same for all c, yield the same regression matrix $\widetilde{A}/\tilde{n} = \widehat{A}$, and can be used as a standard version, if required.

We now define further parameters:

$$
\begin{aligned}
&\alpha_1 := \frac{1}{2}\lambda_{\min}(\Sigma); \qquad \nu := \mathrm{Tr}(\sigma_\Sigma^2)/(d\alpha_1);\\
&\chi := \mathbb{E}\{|\xi|^3\}; \qquad L := (\|A\|/n)^{1/2}\chi\{\mathrm{Tr}(\sigma^2)\}^{-3/2};\\
&\chi_\Sigma := \mathbb{E}|\Sigma^{-1/2}\xi|^3; \qquad L_\Sigma := (\|A\|/n)^{1/2}\chi_\Sigma\{\mathrm{Tr}(\sigma_\Sigma^2)\}^{-3/2} \le L\rho(\Sigma)^{3/2},
\end{aligned} \tag{3.2}
$$

and set $Z := z(W)$, where $z(w) := (nd\nu)^{-1/2}\Sigma^{-1/2}(w-\mu)$. L, L_Σ and Z all involve A, n and Σ only through the standardized quantities $n/\|A\|$ and $n\Sigma$. We then assume that the following inequalities hold:

$$
\{\|A\|/\alpha_1\}^{1/2}\mathbb{E}\{(1+|Z|)|\Sigma^{-1/2}R_1(W)|\} \le \frac{1}{2}(\mathrm{Tr}(\sigma_\Sigma^2))^{1/2}(1+\mathbb{E}|Z|^2); \tag{3.3}
$$

$$
\{\|A\|/\alpha_1\}^{1/2}\mathbb{E}\{|Z|(1+|Z|)|\Sigma^{-1/2}R_1(W)|\} \le \frac{1}{4}(\mathrm{Tr}(\sigma_\Sigma^2))^{1/2}(1+\mathbb{E}|Z|^3). \tag{3.4}
$$

They can reasonably be expected to be satisfied if $|R_1(W)|$ is indeed small. In particular, (3.3)–(3.4) are satisfied if

$$
\{\mathbb{E}|\Sigma^{-1/2}R_1(W)|^3\}^{1/3} \le \frac{1}{8}(\alpha_1\,\mathrm{Tr}(\sigma_\Sigma^2)/\|A\|)^{1/2}. \tag{3.5}
$$

Under the above conditions, the second and third moments of $|Z|$ can be suitably bounded; the proof is given in Section 5.3.

LEMMA 3.2. *If* (3.3) *and* (3.4) *hold, and if* $n/\alpha_1 \ge 1$, *then*

$$
\mathbb{E}|Z|^2 \le 2; \qquad \mathbb{E}|Z|^3 \le m_3 := 2\left(1 + \frac{10\chi_\Sigma}{(\mathrm{Tr}(\sigma_\Sigma^2))^{3/2}}\right),
$$

where $Z = z(W)$, *with* $z(w)$ *as defined above. In particular, for any* $\delta > 0$,

$$
\begin{aligned}
&\frac{n}{\|A\|}\mathbb{P}[|W-\mu|_\Sigma > n\delta\|A\|^{-1/2}]\\
&\qquad \le 2d^{3/2}\delta^{-3}((\|A\|/n)^{1/2} + 10L_\Sigma)\left\{\frac{2\bar\lambda(\sigma_\Sigma^2)}{\lambda_{\min}(\sigma_\Sigma^2)}\right\}^{3/2}.
\end{aligned}
$$

REMARK 3.3. Note that

$$
\left\{|W-\mu|_\Sigma > \frac{n\delta}{\sqrt{\|A\|}}\right\} = \left\{|Z| > \delta\sqrt{\frac{n}{\|A\|}}\sqrt{\frac{\lambda_{\min}(\sigma_\Sigma^2)}{2\bar\lambda(\sigma_\Sigma^2)}}\right\} \tag{3.6}
$$

involves only standardized quantities.

We are now in a position to prove a discrete normal approximation theorem. To state it, we introduce some further notation:

$$\begin{aligned}
\varepsilon_1 &:= \max_{1\le j\le d} d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathcal{L}\bigl(W + e^{(j)}\bigr)\bigr);\\
\varepsilon_1(\xi) &:= \max_{1\le j\le d} d_{\mathrm{TV}}\bigl(\mathcal{L}(W \mid \xi), \mathcal{L}\bigl(W + e^{(j)} \mid \xi\bigr)\bigr).
\end{aligned}\tag{3.7}$$

THEOREM 3.4. *Assume that (W, W') is a pair of random integer valued d-vectors, such that $\mathcal{L}(W) = \mathcal{L}(W')$ and that $\mathbb{E}|W|^3 < \infty$; write $\mu := \mathbb{E}W$. Suppose that $\xi := W' - W$ satisfies the regression condition (3.1), for matrices A and σ^2 such that A is Hurwitz and σ^2 is positive definite; let Σ be the positive definite solution of $A\Sigma + \Sigma A^T + \sigma^2 = 0$. Define $\mathbb{E}|\xi|^3 := \chi$, $\overline{\Lambda} := d^{-1}\operatorname{Tr}(\sigma^2)$ and $L := (\|A\|/n)^{1/2}\chi\{\operatorname{Tr}(\sigma^2)\}^{-3/2}$, and assume that (3.3) and (3.4) hold. Let $\widetilde{A}$ and $\widetilde{\Sigma}$ be as in Remark 3.1. Then there exist constants n_0 and C, depending on $\|\widetilde{A}\|$ and σ^2, such that, if $n/\|A\| \ge n_0$, we have*

$$\begin{aligned}
&d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathcal{DN}_d(\mu, n\Sigma)\bigr)\\
&\quad\le C\log n\bigl\{d^3\bigl(\|A\|/n\bigr)^{1/2} + d^4\varepsilon_1 + d^{1/4}\mathbb{E}\bigl|R_1(W)\bigr|\\
&\qquad + d^{1/2}\mathbb{E}\bigl\|R_2(W)\bigr\|_1 + d^3L + d^2\mathbb{E}\bigl\{|\xi|^3\varepsilon_1(\xi)\bigr\}\bigr\}.
\end{aligned}$$

PROOF. Because $\mathcal{L}(W) = \mathcal{L}(W')$, we have

$$\begin{aligned}
0 &= \bigl(n/\|A\|\bigr)\mathbb{E}\bigl\{h\bigl(W'\bigr)I\bigl[\bigl|W' - \mu\bigr|_\Sigma \le M\bigr] - h(W)I\bigl[|W - \mu|_\Sigma \le M\bigr]\bigr\}\\
&= \bigl(n/\|A\|\bigr)\mathbb{E}\bigl\{\bigl(h\bigl(W'\bigr) - h(W)\bigr)I\bigl[|W - \mu|_\Sigma \le M\bigr]\bigr\}\\
&\quad + \bigl(n/\|A\|\bigr)\mathbb{E}\bigl\{h\bigl(W'\bigr)\bigl(I\bigl[\bigl|W' - \mu\bigr|_\Sigma \le M\bigr] - I\bigl[|W - \mu|_\Sigma \le M\bigr]\bigr)\bigr\},
\end{aligned}\tag{3.8}$$

for any function $h\colon \mathbb{Z}^d \to \mathbb{R}$ and $M > 0$. We shall take $M = n\eta/6\sqrt{\|A\|}$, for η to be prescribed later, in view of (3.6). For bounded functions h, the second term can be simply estimated, using Lemma 3.2, by

$$\begin{aligned}
\theta_0 &:= 2\bigl(n/\|A\|\bigr)\|h\|_\infty\mathbb{P}\bigl[|W - \mu|_\Sigma > M\bigr]\\
&\le 864d^{3/2}\eta^{-3}\bigl(\bigl(\|A\|/n\bigr)^{1/2} + 10L_\Sigma\bigr)\biggl\{\frac{2\bar{\lambda}(\sigma_\Sigma^2)}{\lambda_{\min}(\sigma_\Sigma^2)}\biggr\}^{3/2}\|h\|_\infty.
\end{aligned}\tag{3.9}$$

For the first term, we write

$$h\bigl(W'\bigr) - h(W) = \xi^T\Delta h(W) + \frac{1}{2}\xi^T\Delta^2 h(W)\xi + e_2(W, \xi, h),\tag{3.10}$$

thus defining $e_2(X, J, h)$. From (3.1), its first element yields

$$
\begin{aligned}
&\frac{n}{\|A\|}\big|\mathbb{E}\{\xi^T \Delta h(W) I[|W-\mu|_\Sigma \le M]\} \\
&\quad - \mathbb{E}\{n^{-1}(W-\mu)^T A^T \Delta h(W) I[|W-\mu|_\Sigma \le M]\}\big| \\
&\qquad \le (n/\|A\|)^{1/2}\mathbb{E}\{|R_1(W)^T \Delta h(W)| I[|W-\mu|_\Sigma \le M]\} \\
&\qquad \le (n/\|A\|)^{1/2}\mathbb{E}|R_1(W)| \|\Delta h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty} =: \theta_1.
\end{aligned}
\tag{3.11}
$$

Then

$$
\begin{aligned}
&\frac{n}{2\|A\|}\big|\mathbb{E}\{\xi^T \Delta^2 h(W) \xi I[|W-\mu|_\Sigma \le M]\} \\
&\quad - \mathbb{E}\{\mathrm{Tr}(\sigma^2 \Delta^2 h(W)) I[|W-\mu|_\Sigma \le M]\}\big| \\
&\qquad \le \frac{1}{2}\mathbb{E}\{\|R_2(W)\|_1\}(n/\|A\|)\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty} =: \theta_2.
\end{aligned}
\tag{3.12}
$$

It remains to bound $(n/\|A\|)\mathbb{E}\{e_2(W, \xi, h) I[|W-\mu|_\Sigma \le M]\}$. We first consider $|\xi| > \sqrt{n/\|A\|}$, and use the bound

$$
\begin{aligned}
\mathbb{E}\{|\xi|_1^r I[|\xi| > \sqrt{n/\|A\|}]\} &\le d^{r/2}\mathbb{E}\{|\xi|^r I[|\xi| > \sqrt{n/\|A\|}]\} \\
&\le d^{r/2}\chi(n/\|A\|)^{-(3-r)/2}
\end{aligned}
\tag{3.13}
$$

for $r = 0, 1, 2$. Since

$$
\begin{aligned}
&|e_2(W, \xi, h)| I[|W-\mu|_\Sigma \le M] \\
&\qquad \le 2\|h\|_\infty + |\xi|_1 \|\Delta h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty} + \frac{1}{2}|\xi|_1^2 \|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty},
\end{aligned}
$$

it follows, using (3.13), that

$$
\begin{aligned}
\theta_3 &:= \frac{n}{\|A\|}\mathbb{E}\{|e_2(W, \xi, h)| I[|W-\mu|_\Sigma \le M] I[|\xi| > \sqrt{n/\|A\|}]\} \\
&\le \chi\sqrt{\frac{\|A\|}{n}}\left\{2\|h\|_\infty + \left(\frac{dn}{\|A\|}\right)^{1/2}\|\Delta h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty} + \frac{dn}{2\|A\|}\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty}\right\} \\
&\le 2L\{\mathrm{Tr}(\sigma^2)\}^{3/2}\left\{\|h\|_\infty + \left(\frac{dn}{\|A\|}\right)^{1/2}\|\Delta h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty}\right. \\
&\qquad \left. + \frac{dn}{\|A\|}\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty}\right\}.
\end{aligned}
\tag{3.14}
$$

For $|\xi| \le \sqrt{n/\|A\|}$, we split $e_2(W, \xi, h)$ into a sum of third differences and a remainder:

$$
e_2(W, \xi, h) = E_2(W, \xi, h) - \frac{1}{2}\sum_{j=1}^{d} \xi_j \Delta_{jj} h(W).
\tag{3.15}
$$

For the contribution from the second term in (3.15), we have at most

$$
\begin{aligned}
\theta_4 &:= \frac{n}{2\|A\|}\left|\mathbb{E}\left\{\sum_{j=1}^{d} \xi_j \Delta_{jj} h(W) I\big[|W-\mu|_\Sigma \le M\big] I\big[|\xi| \le \sqrt{n/\|A\|}\big]\right\}\right| \\
&\le \frac{n}{2\|A\|}\left|\mathbb{E}\left\{\sum_{j=1}^{d} \xi_j \Delta_{jj} h(W) I\big[|W-\mu|_\Sigma \le M\big]\right\}\right| \\
&\quad + \frac{1}{2}\mathbb{E}\big\{|\xi|_1 I\big[|\xi| > \sqrt{n/\|A\|}\big]\big\} \frac{n}{\|A\|}\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty} \\
&=: \theta_4' + \theta_4'',
\end{aligned}
\tag{3.16}
$$

say. Now, recalling $\nu := \mathrm{Tr}(\sigma_\Sigma^2)/(d\alpha_1)$ and $Z := (nd\nu)^{-1/2}\Sigma^{-1/2}(W-\mu)$, (3.1) and (3.3) give

$$
\begin{aligned}
\theta_4' &= \frac{1}{2\|A\|}\sum_{j=1}^{d}\big|\mathbb{E}\big\{\big(\big[A(W-\mu)\big]_j + n^{1/2}\|A\|^{1/2}\big[R_1(W)\big]_j\big) \\
&\qquad \times \Delta_{jj} h(W) I\big[|W-\mu|_\Sigma \le M\big]\big\}\big| \\
&\le \frac{1}{2} n^{-1/2}\big\{(d\nu)^{1/2}\mathbb{E}\big|A\Sigma^{1/2}Z\big|_1 + \|A\|^{1/2}\mathbb{E}\big|R_1(W)\big|_1\big\} \\
&\qquad \times \frac{n}{\|A\|}\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty} \\
&\le \frac{1}{2}\big(\|A\|/n\big)^{1/2} d\sqrt{\nu}\big(\|A\|^{-1/2}\|\Sigma^{1/2}\|\big) \\
&\qquad \times \left\{\sqrt{2}\|A\| + \frac{3}{2}\alpha_1\right\}\big(n/\|A\|\big)\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty}.
\end{aligned}
\tag{3.17}
$$

Then, from (3.13),

$$
\theta_4'' \le \frac{1}{2}\big\{d^{1/2}\big(\|A\|/n\big)\chi\big\}\big(n/\|A\|\big)\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{6\sqrt{\|A\|}},\infty}.
\tag{3.18}
$$

For the first term in (3.15), we use Lemma 4.4(i) and Remark 4.5 of Part I to conclude that, if $|\xi| \le \sqrt{n/\|A\|}$ and $n\eta/24\sqrt{\|A\|} \ge \sqrt{n/\|A\|\lambda_{\min}(\Sigma)}$, then

$$
\begin{aligned}
\theta_5(\xi) &:= \big(n/\|A\|\big)\big|\mathbb{E}\big\{E_2(W,\xi,h) I\big[|W-\mu|_\Sigma \le M\big] \mid \xi\big\}\big| \\
&\le \big\{d^{3/2}|\xi|^3\varepsilon_1(\xi) + 2d|\xi|^2\mathbb{P}\big[|W-\mu|_\Sigma \ge M/4 \mid \xi\big]\big\} \\
&\qquad \times \big(n/\|A\|\big)\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{4\sqrt{\|A\|}},\infty}.
\end{aligned}
$$

Taking expectations, and then using Lemma 3.2, this gives

$$
\begin{aligned}
\theta_5 &:= \mathbb{E}\{|\theta_5(\xi)| I[|\xi| \le \sqrt{n/\|A\|}]\} \\
&\le \biggl\{d^{3/2}\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\} + \frac{2dn}{\|A\|}\mathbb{P}[|W-\mu|_\Sigma \ge M/4]\biggr\}\frac{n}{\|A\|}\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{4\sqrt{\|A\|}},\infty} \\
&\le \biggl\{d^{3/2}\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\} + \eta^{-3}d^{5/2}C\{\rho(\sigma^2_\Sigma)\}^{3/2}\biggl(\sqrt{\frac{\|A\|}{n}} + L_\Sigma\biggr)\biggr\} \\
&\quad \times \frac{n}{\|A\|}\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{4\sqrt{\|A\|}},\infty},
\end{aligned}
\tag{3.19}
$$

for C a universal constant.

Let

$$
\widetilde{\mathcal{A}}_{\tilde n} h(w) := \frac{1}{2}\tilde n \operatorname{Tr}(\sigma^2\Delta^2 h(w)) + (w-\mu)^T\widetilde{A}^T\Delta h(w). \tag{3.20}
$$

Then, combining the estimates (3.9) and (3.11)–(3.19) with (3.8) and (3.10), we have shown that

$$
\begin{aligned}
&|\mathbb{E}\{\widetilde{\mathcal{A}}_{\tilde n} h(W) I[|W-\mu|_{\widetilde\Sigma} \le \tilde n\eta/6]\}| \\
&\quad = \biggl|\frac{1}{2}\tilde n\mathbb{E}\{\operatorname{Tr}(\sigma^2\Delta^2 h(W)) I[|W-\mu|_\Sigma \le n\eta/6\sqrt{\|A\|}]\} \\
&\qquad + \mathbb{E}\{(W-\mu)^T\widetilde{A}^T\Delta h(W) I[|W-\mu|_\Sigma \le n\eta/6\sqrt{\|A\|}]\}\biggr|
\end{aligned}
\tag{3.21}
$$

$$
\begin{aligned}
&\le \sum_{l=0}^{3}\theta_l + \theta_4' + \theta_4'' + \theta_5 \\
&\le \varepsilon_{20}\|h\|_\infty + \varepsilon_{21}\tilde n^{1/2}\|\Delta h\|^{\Sigma}_{\frac{n\eta}{4\sqrt{\|A\|}},\infty} + \varepsilon_{22}\tilde n\|\Delta^2 h\|^{\Sigma}_{\frac{n\eta}{4\sqrt{\|A\|}},\infty} \\
&\le \overline{\Lambda}\{\varepsilon_{20}'\|h\|_\infty + \varepsilon_{21}'\tilde n^{1/2}\|\Delta h\|^{\widetilde\Sigma}_{n\eta/4,\infty} + \varepsilon_{22}'\tilde n\|\Delta^2 h\|^{\widetilde\Sigma}_{n\eta/4,\infty}\},
\end{aligned}
\tag{3.22}
$$

with

$$
\begin{aligned}
&\varepsilon_{20}' = C_0(\eta)d^{3/2}(\tilde n^{-1/2} + L); \qquad \varepsilon_{21}' = \overline{\Lambda}^{-1}(\mathbb{E}|R_1(W)| + 2Ld^2\overline{\Lambda}^{3/2}); \\
&\varepsilon_{22}' = C_2(\eta)(\mathbb{E}\|R_2(W)\|_1 + Ld^{5/2} + d^{5/2}\tilde n^{-1/2} + d^{3/2}\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\}),
\end{aligned}
$$

where the constants $C_l(\eta)$ depend on η, $\|\widetilde{A}\|$ and the elements of $\mathrm{Sp}'(\sigma^2)$ and $\mathrm{Sp}'(\widetilde\Sigma)$. Since, if $\tilde n\eta/12 > 2\{\lambda_{\min}(\widetilde\Sigma)\}^{-1/2}$, the quantity in (3.21) does not change if $h(X)$ is replaced by zero for $|X-\mu|_{\widetilde\Sigma} > \tilde n\eta/4$, the norm $\|h\|_\infty$ can be replaced by $\|h\|^{\widetilde\Sigma}_{\tilde n\eta/4,\infty}$ for such $\tilde n$ and η. Thus Condition (iii) of Theorem 1.1 is satisfied, for $\widetilde{\mathcal{A}}_n$ as defined in (3.20), if we take $\eta = \tilde\delta_0$, for $\tilde\delta_0$ as defined in Theorem 1.1, and for $\tilde n$ such that $\tilde n \ge \max\{n_{1.1}, 24/(\tilde\delta_0\{\lambda_{\min}(\widetilde\Sigma)\}^{1/2})\}$. The remaining conditions of

Theorem 1.1, with $\widetilde{\Sigma}$ for Σ and with $\tilde{n}$ for n, are easily checked: Condition (i) is implied by Lemma 3.2, with $v = 2\nu$, and Condition (ii) is just (3.7). This proves the theorem. □

REMARK 3.5. Direct computation of the quantities $\mathbb{E}|R_1(W)|$ and $\mathbb{E}\|R_2(W)\|_1$ can be awkward. It may be easier to find bounds on

$$\widetilde{R}_1 := n^{1/2}\{\mathbb{E}(\xi \mid \mathcal{F}) - n^{-1}A(W-\mu)\} \quad \text{and} \quad \widetilde{R}_2 := \mathbb{E}(\xi\xi^T \mid \mathcal{F}) - \sigma^2,$$

for a σ-field $\mathcal{F}$ such that W is $\mathcal{F}$-measurable. From the properties of conditional expectation and Jensen's inequality, it follows that, for any nonnegative random variable $Y(W)$, we have

$$\mathbb{E}\{Y(W)|R_1(W)|\} \le \mathbb{E}\{Y(W)|\widetilde{R}_1|\};$$
$$\mathbb{E}\{Y(W)\|R_2(W)\|_1\} \le \mathbb{E}\{Y(W)\|\widetilde{R}_2\|_1\}.$$

Hence we can use $\widetilde{R}_1$ and $\widetilde{R}_2$ in place of $R_1(W)$ and $R_2(W)$ when computing the bounds in the theorem and in verifying conditions (3.3)–(3.4).

4. Examples. In Part I, following the proof of Theorem 5.3, it was remarked that, using Theorem 2.3, error bounds of order $O(n^{-1/2}\log n)$ for the (quasi-)equilibrium distributions of rather general Markov jump processes can be proved. Here, we concentrate on examples exhibiting the linear regression structure of the previous section.

4.1. *Sums of independent integer valued random vectors.* Let Y_i, $1 \le i \le m$, be independent $\mathbb{Z}^d$-valued random vectors, with means μ_i and covariance matrices S_i, and let $\gamma_i := \mathbb{E}|Y_i - \mu_i|^3$. Write $\mathbb{P}[Y_i = X] =: p_{i,X}$, $X \in \mathbb{Z}^d$, and define $u_i := \min_{1\le j\le d}\{1 - d_{\mathrm{TV}}(\mathcal{L}(Y_i), \mathcal{L}(Y_i + e^{(j)}))\}$. Let

$$W := \sum_{i=1}^m Y_i; \qquad \mu := \mathbb{E}W = \sum_{i=1}^m \mu_i; \qquad s_m := \sum_{i=1}^m u_i;$$

$$S := \mathbb{E}\{(W-\mu)(W-\mu)^T\} = \sum_{i=1}^m S_i; \qquad \Gamma := \sum_{i=1}^m \gamma_i.$$

We apply Theorem 3.4 to approximate the distribution of W.

To start with, we need to define a W' on the same probability space, in such a way that $\mathcal{L}(W') = \mathcal{L}(W)$, and such that $\xi = W' - W$ is not too large. The canonical way to do this [Stein (1986), page 16] is to let $(Y_1', \dots, Y_m')$ be an independent copy of $(Y_1, \dots, Y_m)$, and to let K be uniformly distributed on $\{1, 2, \dots, m\}$, independently of the Y_i and the Y_i'; then W' is taken to be $W - Y_K + Y_K'$. It is clear that $\mathcal{L}(W') = \mathcal{L}(W)$, and also, writing $\xi := W' - W = Y_K' - Y_K$, that

$$\begin{aligned}\mathbb{E}(\xi \mid W) &= \mathbb{E}\{\mathbb{E}(\xi \mid Y_1, \dots, Y_m) \mid W\}\\ &= \mathbb{E}\left\{m^{-1}\sum_{i=1}^m(\mu_i - Y_i) \,\Big|\, W\right\} = -m^{-1}(W-\mu),\end{aligned}$$

so that the regression condition in (3.1) is satisfied with $A/n = -I/m$, and with $R_1(W) = 0$. Then $\sigma^2 = \mathbb{E}\{\xi\xi^T\} = 2S/m$, giving, for the standardized quantities of Remark 3.1, $\Sigma_0 = S$ and $\alpha_0 = 1/m$, and hence $\tilde{n} = m$, $\widetilde{A} = -I$ and $\widetilde{\Sigma} = S/m$. Note also that

$$\chi = \mathbb{E}|\xi|^3 = m^{-1}\sum_{i=1}^{m}\mathbb{E}|Y_i - Y_i'|^3 \le 4m^{-1}\Gamma.$$

As a next step in applying Theorem 3.4, we show that the quantity ε_1 of (3.7) can be suitably bounded.

LEMMA 4.1. *For W as defined above,*

$$\varepsilon_1 := \max_{1\le j\le d} d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})\big) = O\big(s_m^{-1/2}\big).$$

PROOF. Fix any $1 \le j \le d$, and, for $X \in \mathbb{Z}^d$, define

$$p_{i,X}^- := \frac{1}{2}(p_{i,X} \wedge p_{i,X-e^{(j)}}); \qquad p_{i,X}^+ := \frac{1}{2}(p_{i,X} \wedge p_{i,X+e^{(j)}}).$$

Then define the pair $(Y_i, \widetilde{Y}_i)$ jointly, for $1 \le i \le m$, by

$$(Y_i, \widetilde{Y}_i) = \begin{cases} (X, X - e^{(j)}) & \text{with probability } p_{i,X}^-; \\ (X, X + e^{(j)}) & \text{with probability } p_{i,X}^+; \\ (X, X) & \text{with probability } p_{i,X} - p_{i,X}^- - p_{i,X}^+, \end{cases} \qquad X \in \mathbb{Z}^d.$$

Set $Z_i := Y_i - \widetilde{Y}_i$. Then Z_i takes the values $e^{(j)}$ and $-e^{(j)}$ each with probability $\sum_{X\in\mathbb{Z}^d} p_{i,X}^+$, and takes the value 0 with probability $1 - \sum_{X\in\mathbb{Z}^d} p_{i,X} \wedge p_{i,X+e^{(j)}}$. Hence, for $T_0 := 0$ and $T_k := \sum_{i=1}^k Z_i$, the process $\{T_k, 0 \le k \le m\}$ is a lazy symmetric random walk. Define

$$Y_i' := \begin{cases} \widetilde{Y}_i, & i \le \tau; \\ Y_i, & i > \tau, \end{cases}$$

where $\tau := \min\{k\colon 1 \le k \le m, T_k = e^{(j)}\}$ if this is defined, and with $\tau = m$ otherwise. Set $W' = \sum_{i=1}^m Y_i'$. Then, by the Mineka coupling argument [Lindvall (2002), Section II.14], it follows that

$$d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})\big) \le \mathbb{P}\big[W \ne W' + e^{(j)}\big] \le \mathbb{P}[\tau > m] = O\big(s_m^{-1/2}\big). \qquad \square$$

As a result of this lemma, it is clear that the quantity ε_1 of (3.7) is of order $O(s_m^{-1/2})$. Defining $W^{(i)} := W - Y_i$ and $\tilde{s}_m := s_m - \max_{1\le i\le m} u_i$, we now observe that, for any $X \in \mathbb{Z}^d$, the conditional quantity $\varepsilon_1(X)$ is bounded by

$$\tilde{\varepsilon}_1 := \max_{1\le i\le m}\max_{1\le j\le d} d_{\mathrm{TV}}\big(\mathcal{L}(W^{(i)}), \mathcal{L}(W^{(i)} + e^{(j)})\big) = O\big((\tilde{s}_m)^{-1/2}\big), \tag{4.1}$$

with the final order statement following directly from Lemma 4.1. This is because, for any $X \in \mathbb{Z}^d$,

$$\begin{aligned}
&d_{\mathrm{TV}}\big(\mathcal{L}(W + e^{(j)} \mid \xi = X), \mathcal{L}(W \mid \xi = X)\big) \\
&\qquad \le m^{-1} \sum_{i=1}^{m} d_{\mathrm{TV}}\big(\mathcal{L}\big(W + e^{(j)} \mid \xi = X, K = i\big), \mathcal{L}(W \mid \xi = X, K = i)\big),
\end{aligned}$$

and because, by independence,

$$\begin{aligned}
&d_{\mathrm{TV}}\big(\mathcal{L}(W + e^{(j)} \mid \xi_i = X), \mathcal{L}(W \mid \xi_i = X)\big) \\
&\quad \le \mathbb{E}\big\{d_{\mathrm{TV}}\big(\mathcal{L}\big(W^{(i)} + \xi_i + e^{(j)} \mid \xi_i, \xi_i' = \xi_i + X\big), \\
&\qquad\qquad \mathcal{L}\big(W^{(i)} + \xi_i \mid \xi_i, \xi_i' = \xi_i + X\big)\big)\big\} \\
&\quad = \mathbb{E}\big\{d_{\mathrm{TV}}\big(\mathcal{L}\big(W^{(i)} + \xi_i + e^{(j)} \mid \xi_i\big), \mathcal{L}\big(W^{(i)} + \xi_i \mid \xi_i\big)\big)\big\} \\
&\quad = \mathbb{E}\big\{d_{\mathrm{TV}}\big(\mathcal{L}\big(W^{(i)} + e^{(j)} \mid \xi_i\big), \mathcal{L}\big(W^{(i)} \mid \xi_i\big)\big)\big\} \\
&\quad = \mathbb{E}\big\{d_{\mathrm{TV}}\big(\mathcal{L}\big(W^{(i)} + e^{(j)}\big), \mathcal{L}\big(W^{(i)}\big)\big)\big\} \le \tilde{\varepsilon}_1.
\end{aligned}$$

Thus a number of the elements appearing in the bound given in Theorem 3.4 can be successfully handled. We now show that a multivariate discrete normal approximation can indeed be established. We write

$$\overline{\Lambda} := d^{-1}\operatorname{Tr}(\sigma^2) = 2\operatorname{Tr}(S/m) \quad \text{and} \quad L := m^{-1/2}\frac{\chi}{\{\operatorname{Tr}(\sigma^2)\}^{3/2}} \ge m^{-1/2};$$

the latter quantity, introduced in (3.2), is of order $O(m^{-1/2})$ if the ratio $\mathbb{E}|\xi|^3/\{\mathbb{E}|\xi|^2\}^{3/2}$ remains bounded.

THEOREM 4.2. *Under the above circumstances*,

$$d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathcal{DN}_d(\mu, S)\big) \le C d^{7/2} \log m\big(L + (d/m)^{1/2}\big)\sqrt{\frac{m}{\tilde{s}_m}},$$

for a suitable constant C, depending only on $\mathrm{Sp}'(S/m)$.

PROOF. With the definitions of W' and W given above, the regression condition in (3.1) is satisfied with $R_1(w) = 0$ for all $w \in \mathbb{Z}^d$, so that Conditions (3.3) and (3.4) are trivially satisfied. Then $\varepsilon_1 = O(s_m^{-1/2})$, by Lemma 4.1, and

$$\mathbb{E}\big\{|\xi|^3\varepsilon_1(\xi)\big\} = O\big((\tilde{s}_m)^{-1/2}\chi\big), \tag{4.2}$$

from the observations above. Note that

$$\chi = L\sqrt{m}d^{3/2}\overline{\Lambda}^{3/2}. \tag{4.3}$$

For $\mathbb{E}\|R_2(W)\|_1$, for any $X, w \in \mathbb{Z}^d$, we write $p(X) := \mathbb{P}[\xi = X]$, obtaining

$$\sigma_{il}^2(w) = \sum_{X \in \mathbb{Z}^d} X_i X_l \mathbb{P}[\xi = X \mid W = w] = \sum_{X \in \mathbb{Z}^d} p(X) X_i X_l \frac{\mathbb{P}[W = w \mid \xi = X]}{\mathbb{P}[W = w]}.$$

Hence

$$\begin{aligned}
&\mathbb{E}|\sigma_{il}^2(W) - \sigma_{il}^2| \\
&\quad = \sum_{w \in \mathbb{Z}^d} \left| \sum_{X \in \mathbb{Z}^d} p(X) X_i X_l \bigl(\mathbb{P}[W = w \mid \xi = X] - \mathbb{P}[W = w]\bigr) \right| \\
&\quad = \sum_{w \in \mathbb{Z}^d} \Biggl| \sum_{X \in \mathbb{Z}^d} p(X) X_i X_l \\
&\qquad \times \sum_{y \in \mathbb{Z}^d} p(y) \bigl(\mathbb{P}[W = w \mid \xi = X] - \mathbb{P}[W = w \mid \xi = y]\bigr) \Biggr| \\
&\quad \le \sum_{X \in \mathbb{Z}^d} p(X) |X_i||X_l| \sum_{y \in \mathbb{Z}^d} p(y) 2 d_{\mathrm{TV}}\bigl(\mathcal{L}(W \mid \xi = X), \mathcal{L}(W \mid \xi = y)\bigr).
\end{aligned} \tag{4.4}$$

Now, by independence,

$$\begin{aligned}
&d_{\mathrm{TV}}\bigl(\mathcal{L}(W \mid \xi = X), \mathcal{L}(W \mid \xi = y)\bigr) \\
&\quad \le \frac{1}{m} \sum_{i=1}^m d_{\mathrm{TV}}\bigl(\mathcal{L}\bigl(W^{(i)} + Y_i \mid Y_i' - Y_i = x\bigr), \mathcal{L}\bigl(W^{(i)} + Y_i \mid Y_i' - Y_i = y\bigr)\bigr) \\
&\quad \le \frac{1}{m} \sum_{i=1}^m \mathbb{E}\bigl\{d_{\mathrm{TV}}\bigl(\mathcal{L}\bigl(W^{(i)} + Y_i \mid Y_i', Y_i = Y_i' - x\bigr), \\
&\qquad\qquad \mathcal{L}\bigl(W^{(i)} + Y_i \mid Y_i', Y_i = Y_i' - y\bigr)\bigr)\bigr\} \\
&\quad = \frac{1}{m} \sum_{i=1}^m \mathbb{E}\bigl\{d_{\mathrm{TV}}\bigl(\mathcal{L}\bigl(W^{(i)}\bigr), \mathcal{L}\bigl(W^{(i)} - y + x\bigr)\bigr)\bigr\} \\
&\quad \le \tilde{\varepsilon}_1 |y - x|_1.
\end{aligned}$$

Substituting this bound into (4.4) and adding over $1 \le i, l \le d$ thus gives

$$\begin{aligned}
\mathbb{E}\bigl\|R_2(W)\bigr\|_1 &\le 2 \sum_{i=1}^d \sum_{l=1}^d \sum_{X \in \mathbb{Z}^d} p(X) |X_i||X_l| \sum_{y \in \mathbb{Z}^d} p(y) \tilde{\varepsilon}_1 |x - y|_1 \\
&\le 2\tilde{\varepsilon}_1 \sum_{X \in \mathbb{Z}^d} p(X) |X|_1^2 \bigl\{|X|_1 + \mathbb{E}|\xi|_1\bigr\} \\
&\le 4\tilde{\varepsilon}_1 \mathbb{E}|\xi|_1^3 \le 4\tilde{\varepsilon}_1 d^{3/2} \chi.
\end{aligned} \tag{4.5}$$

It only remains to collect the elements needed for Theorem 3.4. From (4.2) and (4.5), and from the definition of L, we have

$$d^{1/2}\mathbb{E}\|R_2(W)\|_1 + d^2\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\} = O\bigl(d^2\chi\tilde{s}_m^{-1/2}\bigr) = O\bigl(Ld^{7/2}(m/\tilde{s}_m)^{1/2}\overline{\Lambda}^{3/2}\bigr).$$

Combining this with the remaining elements of the bound given in Theorem 3.4, and noting that $\tilde{s}_m \le m$, the theorem follows. □

Except for the logarithmic factors, the bound obtained in the theorem is of the same order in m as would be expected for weaker metrics, such as the convex sets metric [Bentkus (2003), Fang and Röllin (2015)], if $\tilde{s}_m \asymp m$. The latter asymptotic equivalence holds, for example, for identically distributed summands whose common distribution has nontrivial overlap with its unit translates in each direction. It is possible, however, for $\tilde{s}_m$ to be significantly smaller than m. For instance, if all the summands making up W are on $2\mathbb{Z} \times \mathbb{Z}^{d-1}$, then $s_m = 0$, and the discrete normal is not a good approximation to W in total variation, since it puts about half its probability mass on points whose first coordinate is an odd integer, whereas $\mathcal{L}(W)$ puts zero mass on this set.

The best approximation order with respect to the convex sets metric, for sums of independent and identically distributed random variables with finite third moment, is $O(d^{7/4}L)$. Thus our rate is weaker in m by a factor of $\log m$, and in dimension by a factor of $d^{9/4}$. If the distributions are not identical, the best known d-dependence for approximation in the convex sets metric is rather worse, unless the random variables are also assumed to be bounded. Since the total variation metric is substantially stronger than the convex sets metric, our bounds are of encouragingly small order in d, too.

4.2. *Exchangeable pairs*. If the pair (W, W') is also *exchangeable*, so that $\mathcal{L}((W, W')) = \mathcal{L}((W', W))$, a neat argument of Röllin and Ross (2015) delivers bounds on the quantities ε_1 and $\varepsilon_1(\xi)$ of (3.7), which appear in the bound given in Theorem 3.4. These can be of considerable practical use in deriving explicit bounds from the general expressions given in Theorem 3.4.

For $\xi := W' - W$, let $\mathcal{J}$ be the set of d-vectors such that $q^J := \mathbb{P}[\xi = J] > 0$, and suppose that each of the coordinate vectors $e^{(j)} \in \mathbb{R}^d$ can be obtained as a (finite) sum of elements of $\mathcal{J}$. For $Q^J(W) := \mathbb{P}[\xi = J \mid W]$, set

$$u^J := (q^J)^{-1}\mathbb{E}|Q^J(W) - q^J|, \tag{4.6}$$

to be thought of as small. Note that, by exchangeability,

$$q^J = \mathbb{E}\{I[W' - W = J]\} = \mathbb{E}\{I[W - W' = J]\} = q^{-J}. \tag{4.7}$$

We then write

$$\tilde{u}_j := \sum_{l=1}^{r(j)}\bigl(u^{J_l^{(j)}} + u^{-J_l^{(j)}}\bigr) \qquad \text{where } \sum_{l=1}^{r(j)} J_l^{(j)} = e^{(j)},$$

and then set $\tilde{u}^* := \max_{1\le j\le d} \tilde{u}_j$ and $u^* := \sup_{J\in\mathcal{J}} u^J$. With the help of these quantities, we can bound the differences $d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)}))$ between the distribution of W and its translates.

LEMMA 4.3. *For each* $1 \le j \le d$, *we have*

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W + e^{(j)}), \mathcal{L}(W)\bigr) \le \tilde{u}_j,$$

and

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W + e^{(j)} \mid \xi = J), \mathcal{L}(W \mid \xi = J)\bigr) \le \tilde{u}_j + 2u^J.$$

Hence, *in particular*, *for each* $J \in \mathcal{J}$,

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W + e^{(j)} \mid \xi = J), \mathcal{L}(W \mid \xi = J)\bigr) \le \tilde{u}^* + 2u^*,$$

and $d_{\mathrm{TV}}(\mathcal{L}(W + e^{(j)}), \mathcal{L}(W)) \le \tilde{u}^*$. *Furthermore*, *for* $R_2(W)$ *as defined in* (3.1), *we have*

$$\mathbb{E}\|R_2(W)\|_1 \le d\,\mathrm{Tr}(\sigma^2)u^*.$$

PROOF. For any $J \in \mathcal{J}$ and any f with $\|f\|_\infty = 1$, we use exchangeability to give

$$\mathbb{E}\{f(W')I[W' - W = J] - f(W)I[W - W' = J]\} = 0.$$

As in the proof of Theorem 3.6 of Röllin and Ross (2015), we divide by q^J, using (4.7), and evaluate the expectation by conditioning on W, giving

$$\begin{aligned}
0 &= (q^J)^{-1}\mathbb{E}\{f(W + J)Q^J(W) - f(W)Q^{-J}(W)\} \\
&= \mathbb{E}\{f(W + J) - f(W)\} + (q^J)^{-1}\mathbb{E}\{f(W + J)(Q^J(W) - q^J)\} \\
&\quad - (q^{-J})^{-1}\mathbb{E}\{f(W)(Q^{-J}(W) - q^{-J})\},
\end{aligned}$$

from which it follows that

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W + J), \mathcal{L}(W)\bigr) \le u^J + u^{-J}.$$

The first statement now follows by the triangle inequality.

For the second, we have

$$\begin{aligned}
&\mathbb{E}\{f(W + e^{(j)}) - f(W) \mid \xi = J\} \\
&\quad = (q^J)^{-1}\mathbb{E}\{(f(W + e^{(j)}) - f(W))I[\xi = J]\} \\
&\quad = (q^J)^{-1}\mathbb{E}\{(f(W + e^{(j)}) - f(W))Q^J(W)\} \\
&\quad = \mathbb{E}\{f(W + e^{(j)}) - f(W)\} \\
&\qquad + (q^J)^{-1}\mathbb{E}\{(f(W + e^{(j)}) - f(W))(Q^J(W) - q^J)\}.
\end{aligned}$$

Hence we have

$$d_{\mathrm{TV}}\big(\mathcal{L}(W \mid \xi = J), \mathcal{L}(W + e^{(j)} \mid \xi = J)\big) \leq d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})\big) + 2u^J, \tag{4.8}$$

and the second part follows; note that exchangeability was not used in proving (4.8).

Finally, from the definition of $R_2(W)$ in (3.1), we have

$$\{R_2(w)\}_{il} = \sigma_{il}^2(w) - \sigma_{il}^2 = \sum_{J,J' \in \mathcal{J}} J_i J_l' \big(Q^J(w) - q^J\big),$$

for any $1 \leq i, l \leq d$, so that

$$\mathbb{E}\big|\sigma_{il}^2(W) - \sigma_{il}^2\big| \leq \sum_{J,J' \in \mathcal{J}} q^J |J_i| |J_l'| u^J \leq \mathbb{E}\{|\xi_i||\xi_l|\} u^*.$$

This in turn implies that

$$\mathbb{E}\|R_2(W)\|_1 = \sum_{i=1}^d \sum_{l=1}^d \mathbb{E}\big|\sigma_{il}^2(W) - \sigma_{il}^2\big| \leq \mathbb{E}\{|\xi|_1^2\} u^* \leq d \operatorname{Tr}(\sigma^2) u^*,$$

as claimed. $\square$

The following corollary is immediate.

COROLLARY 4.4. *Under the above assumptions*,

$$d^{1/2}\mathbb{E}\|R_2(W)\|_1 \leq C\overline{\Lambda}\tilde{n}^{-1/2} d^{5/2}\{\tilde{n}^{1/2}u^*\}$$

and

$$d^2\mathbb{E}\{|\xi|^3 \varepsilon_1(\xi)\} \leq C'\overline{\Lambda}^{3/2} d^{7/2} L\{\tilde{n}^{1/2}(\tilde{u}^* + 2u^*)\},$$

for constants C and C' that depend only on $\mathrm{Sp}'(\sigma^2/\overline{\Lambda})$.

REMARK 4.5. Note that, by the argument in Remark 3.5, we can bound the quantities u^J above by $(q^J)^{-1}\mathbb{E}|\mathbb{P}[\xi = J \mid \mathcal{F}] - q^J|$, for any σ-field $\mathcal{F}$ such that W is $\mathcal{F}$-measurable. Such quantities may be easier to bound in practice.

REMARK 4.6. For an exchangeable pair (W, W'), we see that

$$\begin{aligned}
\mathbb{E}\{\xi\xi^T\} &= \mathbb{E}\{(W' - \mu)(W' - W)^T - (W - \mu)(W' - W)^T\} \\
&= -\mathbb{E}\{-(W - \mu)(W - W')^T + (W - \mu)(W' - W)^T\} \\
&= -2\mathbb{E}\{(W - \mu)(W' - W)^T\} = -2\mathbb{E}\{(W - \mu)\mathbb{E}(\xi^T \mid W)\} \\
&= -2\mathbb{E}\{\mathbb{E}(\xi \mid W)(W - \mu)^T\},
\end{aligned} \tag{4.9}$$

the last equality following because $\mathbb{E}\{\xi\xi^T\}$ is symmetric. If the remainders $R_1(W)$ and $R_2(W)$ in (3.1) were exactly zero, this would give

$$\frac{1}{2}\sigma^2 = -n^{-1}A\,\mathrm{Cov}(W) = -n^{-1}\,\mathrm{Cov}(W)A^T,$$

and hence also $A^{-1}\sigma^2 = \sigma^2(A^T)^{-1}$. If this is the case, we can easily solve for Σ, since then $\Sigma := -\frac{1}{2}A^{-1}\sigma^2 = -\frac{1}{2}\sigma^2(A^T)^{-1}$ satisfies $A\Sigma + \Sigma A^T + \sigma^2 = 0$ and is symmetric.

4.2.1. *Monochrome edges in regular graphs.* As an example of the application of Theorem 3.4 in the exchangeable setting, suppose that G_n is an r-regular graph on n vertices (so that one of n and r is even); thus there are $nr/2$ edges in the graph. Let the vertices be coloured independently, each with one of m colours, the probability of choosing colour i being $p_i > 0$, $1 \le i \le m$. Let N_i denote the number of vertices having colour i, and let M_i denote the number of edges joining pairs of vertices that both have colour i. We approximate the joint distribution of

$$W := (M_1, \dots, M_m, N_1, \dots, N_{m-1}) =: (W_1, \dots, W_m, W_{m+1}, \dots, W_{2m-1}),$$

when n becomes large, while r, m and $p_1, \dots, p_m$ remain fixed; the detailed structure of G_n does not appear in the approximation. Of course, the value of $N_m = n - \sum_{i=1}^{m-1} N_i$ is implied by knowledge of W. This problem, in the context of multivariate normal approximation, was considered by Rinott and Rotar (1996) and in Chen, Goldstein and Shao (2011), pages 333–334.

THEOREM 4.7. *For $m \ge 3$, r and $p_1, \dots, p_m$ fixed, we can find $\nu \in \mathbb{R}^{2m-1}$ and a $(2m-1)\times(2m-1)$ covariance matrix Σ such that, as $n \to \infty$,*

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathcal{DN}_{2m-1}(n\nu, n\Sigma)\bigr) = O\bigl(n^{-1/2}\log n\bigr).$$

PROOF. We use the notation of Theorem 3.4 throughout. We begin by observing that

$$\mathbb{E}M_i = nrp_i^2/2; \qquad \mathbb{E}N_i = np_i,$$

determining $\nu := n^{-1}\mathbb{E}W$. After rather more calculation, the covariances are given, for $1 \le i \ne l \le m$, by

$$\begin{aligned}
&\mathrm{Var}(M_i) = \frac{1}{2}nrp_i^2(1-p_i)\{1 + (2r-1)p_i\};\\
&\mathrm{Cov}(N_i, N_l) = -np_ip_l;\\
&\mathrm{Cov}(M_i, M_l) = -\frac{1}{2}nr(2r-1)p_i^2p_l^2; \qquad \mathrm{Cov}(M_i, N_l) = -nrp_i^2p_l;\\
&\mathrm{Cov}(M_i, N_i) = nrp_i^2(1-p_i); \qquad \mathrm{Var}(N_i) = np_i(1-p_i),
\end{aligned} \tag{4.10}$$

in turn determining Σ.

We now construct an exchangeable pair (W, W') by first realizing a colouring $(C(j), 1 \le j \le n)$, and using it to define

$$M_i := \sum_{\{j,j'\} \in G} I[C(j) = C(j') = i] \quad \text{and} \quad N_i := \sum_{j=1}^{n} I[C(j) = i], \tag{4.11}$$

for each $1 \le i \le m$, thus defining W. We then choose a vertex K uniformly at random, independently of $(C(j), 1 \le j \le m)$, and then replace $C(K)$ by C', where C' is independently sampled from $1, 2, \dots, m$ with $\mathbb{P}[C' = i] = p_i$, $1 \le i \le m$. If this new colouring is denoted by $(C'(j), 1 \le j \le m)$, then we define M_i' and N_i' as in (4.11), but with the $C'(j)$ in place of $C(j)$, and hence deduce W'. Of course, $\mathcal{L}(W, W') = \mathcal{L}(W', W)$, and W' differs from W only through the (possibly) new colour at the vertex K, and through its impact in changing which edges incident to K are monochrome:

$$M_i' - M_i = \sum_{j\,:\,\{j,K\} \in G} \bigl(I[C(j) = C'(K) = i] - I[C(j) = C(K) = i]\bigr)$$

$$N_i' - N_i = \{I[C'(K) = i] - I[C(K) = i]\}.$$

Hence, for $1 \le l \le m$, we have

$$\begin{aligned} &\mathbb{E}\{\xi_l \mid C(1), \dots, C(n)\} \\ &\qquad = n^{-1} \sum_{k=1}^{n} \sum_{j\,:\,\{j,k\} \in G} \{p_l I[C(k) = l] - I[C(j) = C(k) = l]\} \\ &\qquad = n^{-1}\{p_l r N_l - 2M_l\} = \mathbb{E}\{\xi_l \mid W\}, \end{aligned}$$

and, for $m + 1 \le l \le 2m - 1$,

$$\mathbb{E}\{\xi_l \mid C(1), \dots, C(n)\} = n^{-1}\{np_{l-m} - N_{l-m}\} = \mathbb{E}\{\xi_l \mid W\}.$$

This gives an exact linear regression as in (3.1), with $R_1(w) = 0$ for all w, and with A having nonzero elements given by

$$\begin{aligned} A_{ll} &:= -2, \qquad A_{l,l+m} := rp_l, \qquad 1 \le l \le m - 1; \\ A_{mm} &:= -2, \qquad A_{m,m+t} := -rp_m, \qquad 1 \le t \le m - 1; \\ A_{ll} &:= -1, \qquad m + 1 \le l \le 2m - 1. \end{aligned}$$

Since A is upper triangular, its eigenvalues are -2, with multiplicity m, and -1, with multiplicity $m - 1$, so that it is indeed spectrally negative.

The set $\mathcal{J}$, consisting of the possible values that can be taken by ξ, is finite, and does not depend on n. If $C(K) = i \ne l = C'(K)$, then the $m + i$ and $m + l$ components of ξ each have modulus one (though, if i or l are equal to m, one of these components is not present in W), and the i and l components are in modulus at most r; all other components of ξ are zero. Hence $|\xi|^2 \le 2(r^2 + 1)$ a.s., and $\mathbb{E}|\xi|^3$

remains bounded as n increases; L is thus of strict order $n^{-1/2}$. The components of $\sigma^2 := \mathbb{E}\{\xi\xi^T\}$ can be explicitly calculated: for $1 \le l \ne l' \le m$, they are given by

$$\mathbb{E}\xi_l^2 = 2p_l^2(1-p_l)\{r(r-1)p_l + r\}; \qquad \mathbb{E}\{\xi_l\xi_{l'}\} = -2r(r-1)p_l^2 p_{l'}^2;$$
$$\mathbb{E}\{\xi_l\xi_{m+l}\} = 2rp_l^2(1-p_l); \qquad \mathbb{E}\{\xi_l\xi_{m+l'}\} = -2rp_l^2 p_{l'};$$
$$\mathbb{E}\{\xi_{m+l}^2\} = 2p_l; \qquad \mathbb{E}\{\xi_{m+l}\xi_{m+l'}\} = -2p_l p_{l'},$$

where terms with subscript $2m$ are to be ignored.

In order to apply Theorem 3.4, we now just need to find bounds for ε_1, $\mathbb{E}\{|\xi|^3\varepsilon_1(\xi)\}$ and $\mathbb{E}\|R_2(W)\|_1$. From Lemma 4.3 and Corollary 4.4, these are all bounded by fixed multiples of u^* and $\tilde{u}^*$. For each J in the fixed finite set $\mathcal{J}$, the probability q^J in the denominator of u^J is fixed and positive, and hence bounded away from zero. To bound the numerator, we condition on a larger σ-field $\mathcal{F}$, with respect to which W is measurable, as in Remark 4.5. Let $T_{m,r}$ denote the set of all m-tuples of nonnegative integers $t_1, \dots, t_m$ such that $\sum_{i=1}^m t_i = r$, and, for $t := (t_1, \dots, t_m) \in T_{m,r}$, let $E_j(i_0; t)$ denote the event that $C(j) = i_0$, and that t_i of the r neighbours of j have colour i, $1 \le i \le m$. For each fixed j, these are disjoint events whose union over $1 \le i_0 \le m$ and $t \in T_{m,r}$ is the sure event. We let $\mathcal{F}$ be the σ-field generated by the events

$$\{E_j(i_0; t); 1 \le j \le n, 1 \le i_0 \le m, t \in T_{m,r}\}.$$

Then, if $K = j$, the value $J \in \mathcal{J}$ taken by ξ is determined by which of the events $(E_j(i_0; t); 1 \le i_0 \le m, t \in T_{m,r})$ occurs. For each J, there is a collection $S(J)$ of possible choices, consisting of just one possible $i_0 = i_0(J)$, the index for which $J_{m+i_0} = -1$ (if there is none, then $i_0 = m$), but of all t that satisfy $t_{i_0} = -J_{i_0}$ and $t_{i_1} = J_{i_1}$, where i_1 is the index for which $J_{m+i_1} = 1$ (or m, if there is none such). Thus

$$\mathbb{P}[\xi = J \mid \mathcal{F}] = n^{-1} \sum_{j=1}^{n} \sum_{t \in T_{m,r}:\ (i_0(J),t) \in S(J)} I[E_j(i_0(J); t)].$$

Now, if $j' \ne j$ is such that the set of neighbours $\mathcal{N}(j)$ (including j) in G is disjoint from the set $\mathcal{N}(j')$, the events $I[E_j(i_0(J); t)]$ and $I[E_{j'}(i_0(J); t')]$ are independent. Since, for each j, there are no more than $r + r^2$ choices of $j' \ne j$ for which this is not the case, it follows that

$$\mathrm{Var}\{\mathbb{P}[\xi = J \mid \mathcal{F}]\} = O(n^{-1}).$$

Hence $\mathrm{Var}\{Q^J(W)\} = O(n^{-1})$ also, and so $\mathbb{E}|Q^J(W) - q^J| = O(n^{-1/2})$ for all $J \in \mathcal{J}$, implying that $u^* = O(n^{-1/2})$.

The argument for $\tilde{u}^*$ is not yet finished, since, for each $1 \le l \le 2m-1$, it is necessary to find a chain $J^{(1)}, J^{(2)}, \dots, J^{(R)}$ such that each $J^{(i)} \in \mathcal{J}$ and $\sum_{i=1}^R J^{(i)} = e^{(l)}$. For $m+1 \le l \le 2m-1$, this is easy: $\xi = e^{(l)}$ if, when W is constructed, a vertex has colour m and no neighbours of colours m or l, and its colour is replaced by l when resampling to obtain W'. Note that, to do this, we need at least three colours: $m \geq 3$. To get $e^{(l)}$ for $1 \leq l \leq m-1$, a chain of length 2 is needed: a vertex of colour m with no neighbours of colour m and with exactly one of colour l is recoloured with colour l, giving $J = e^{(l)} + e^{(l+m)}$. Then $J = -e^{(l+m)}$ can be attained by reversing the order of the choices in the example for $m+1 \leq l \leq 2m-1$. To get $e^{(m)}$, a vertex of colour $l \neq m$ with no neighbours of colour l and exactly one of colour m is recoloured m, yielding $e^{(m)} - e^{(m+l)}$, and then adding $e^{(m+l)}$ as before completes the chain. Thus, for $m \geq 3$, we have $\tilde{u}^* = O(n^{-1/2})$ also, and applying Theorem 3.4, the result follows. $\square$

There remains the case of $m = 2$. Here, discrete normal approximation in total variation is not good, since it can be seen that $M_1 - M_2 = r(N_1 - n/2)$, so that W is degenerate; what is more, reducing to (W_1, W_2) gives an integer vector living on a proper sub-lattice of $\mathbb{Z}^2$. However, the pair (M_1, N_1) can be approximated using the method above, and the remaining components of M and N follow from $N_2 = n - N_1$ and $M_2 = M_1 - r(N_1 - n/2)$.

5. Technicalities.

5.1. *Proof of Lemma* 2.1. Let φ_n denote the density of the multivariate normal distribution $\mathcal{N}_d(nc, n\Sigma)$, and, for $X \in \mathbb{Z}^d$, let $[X]$ denote the box

$$[X] := \left\{x \in \mathbb{R}^d : X_i - \frac{1}{2} < x_i \leq X_i + \frac{1}{2}, 1 \leq i \leq d\right\}.$$

Let N_d, $d \geq 1$, denote a standard d-dimensional normal random vector. For (a), the bound on $\mathbb{E}|W - nc|_\Sigma^l$ is obtained by first writing

$$|X - nc|_\Sigma^l \leq \big(|X - t|_\Sigma + |t - nc|_\Sigma\big)^l \leq 2^{l-1}\big(|X - t|_\Sigma^l + |t - nc|_\Sigma^l\big).$$

Taking this inside the integral, we have

$$\begin{aligned}
\mathbb{E}|W - nc|_\Sigma^l &= \sum_{X \in \mathbb{Z}^d} |X - nc|_\Sigma^l \int_{[X]} \varphi_n(t)\,dt \\
&\leq \sum_{X \in \mathbb{Z}^d} \int_{[X]} \varphi_n(t) 2^{l-1}\left(\left(\frac{1}{2}\sqrt{d/\lambda_{\min}(\Sigma)}\right)^l + |t - nc|_\Sigma^l\right) dt \\
&\leq \mathbb{E}\left\{2^{l-1}\left(\left(\frac{1}{2}\sqrt{d/\lambda_{\min}(\Sigma)}\right)^l + n^{l/2}|N_d|^l\right)\right\} \\
&\leq 2^l \mathbb{E}|N_d|^l n^{l/2},
\end{aligned}$$

for

$$n \geq \frac{d}{4(\mathbb{E}|N_d|)^2 \lambda_{\min}(\Sigma)} = \frac{d}{8\lambda_{\min}(\Sigma)}\big\{\Gamma(d/2)/\Gamma\big((d+1)/2\big)\big\}^2.$$

Part (a) follows, taking $C(l) := 2^l\sqrt{k(l)}$, where

$$k(l) := \mathbb{E}N_1^{2l} = \frac{(2l)!}{2^l l!},$$

since $2^l\mathbb{E}|N_d|^l \le 2^l\sqrt{\mathbb{E}N_d^{2l}} \le 2^l d^{l/2}\sqrt{\mathbb{E}N_1^{2l}}$, and by noting that, in $d \ge 1$, $\frac{d}{8}\{\Gamma(d/2)/\Gamma((d+1)/2)\}^2 \le 1$.

For (c), the bound on $\mathbb{E}\{[\Sigma^{-1}(W-nc)]_j^{2l}\}$, we first note that

$$\mathbb{E}\{(a^T N_d)^{2l}\} = (a^T a)^l \mathbb{E}\{N_1^{2l}\} = k(l)(a^T a)^l,$$

for any $a \in \mathbb{R}^d$. So, since

$$[\Sigma^{-1}(X-nc)]_j^{2l} \le 2^{2l-1}\left\{\left(\frac{d}{4\{\lambda_{\min}(\Sigma)\}^2}\right)^l + [\Sigma^{-1}(t-nc)]_j^{2l}\right\}$$

for $t \in [X]$, it follows that

$$\begin{aligned}
\mathbb{E}\{[\Sigma^{-1}(W-nc)]_j^{2l}\} &= \sum_{X\in\mathbb{Z}^d}[\Sigma^{-1}(X-nc)]_j^{2l}\int_{[X]}\varphi_n(t)\,dt \\
&\le 2^{2l-1}\left\{\left(\frac{d}{4\{\lambda_{\min}(\Sigma)\}^2}\right)^l + n^l\mathbb{E}\{\{(e^{(j)})^T\Sigma^{-1/2}N_d\}^{2l}\}\right\} \\
&\le 2^{2l-1}\left\{\left(\frac{d}{4\{\lambda_{\min}(\Sigma)\}^2}\right)^l + n^l k(l)(\Sigma^{-1})_{jj}^l\right\}
\end{aligned}$$

and the stated bound follows, with $C'(l) = 2^{2l-1}k(l)$. Part (b) is similar, but simpler.

5.2. *Proof of Lemma* 2.2. We note first that, from Lemma 2.1(a),

$$\mathbb{E}|W-nc|_\Sigma^i \le C(i)(nd)^{i/2}, \tag{5.1}$$

if $n \ge 1/\lambda_{\min}(\Sigma)$. For (a), bounding the difference between $\mathbb{E}\{\Delta f(W)^T bI_n^\delta(W)\}$ and $n^{-1}\mathbb{E}\{(f(W)(W-nc)^T\Sigma^{-1}bI_n^\delta(W)\}$, we begin by observing that

$$\begin{aligned}
&\mathbb{E}\{\Delta_j f(W)I_n^\delta(W)\} \\
&\quad= \sum_{X\in\mathbb{Z}^d} f(X)\{\mathbb{P}[W = X-e^{(j)}] - \mathbb{P}[W=X]\}I_n^\delta(X) \\
&\qquad+ \sum_{X\in\mathbb{Z}^d} f(X)\mathbb{P}[W=X-e^{(j)}]\{I_n^\delta(X-e^{(j)}) - I_n^\delta(X)\}.
\end{aligned} \tag{5.2}$$

Because, from the definition of $I_n^\eta(X)$, $|I_n^\delta(X-e^{(j)}) - I_n^\delta(X)| = 1$ requires $|X-e^{(j)}-nc|_\Sigma > n\delta/3$ and $|X-nc|_\Sigma \le n\delta/3$, or vice versa, the last term in (5.2) is in modulus at most

$$\mathbb{P}[|W-nc|_\Sigma > n\delta/3 - 1/\sqrt{\lambda_{\min}(\Sigma)}]\max_{|X-nc|_\Sigma\le n\delta/3+1/\sqrt{\lambda_{\min}(\Sigma)}}|f(X)|.$$

Thus it follows from (5.1) and a fourth moment Markov inequality that, if $n \geq \max\{1/\lambda_{\min}(\Sigma), 6/(\delta\sqrt{\lambda_{\min}(\Sigma)})\} = \max\{1/\lambda_{\min}(\Sigma), \psi_\Sigma(\delta)\}$, then

$$\begin{aligned}
&\left|\sum_{X\in\mathbb{Z}^d} f(X)\mathbb{P}[W = X - e^{(j)}]\{I_n^\delta(X - e^{(j)}) - I_n^\delta(X)\}\right| \\
&\qquad \leq \|f\|_{n\delta/2,\infty}^{\Sigma}\mathbb{P}[|W - nc|_\Sigma > n\delta/6] \\
&\qquad \leq (6/\delta)^4 d^2 C(4) n^{-2}\|f\|_{n\delta/2,\infty}^{\Sigma} \leq d^2 C_1(\delta) n^{-2}\|f\|_{n\delta/2,\infty}^{\Sigma},
\end{aligned} \tag{5.3}$$

where $C_1(\delta) = (6/\delta)^4 C(4) \in \mathcal{K}_\Sigma(\delta)$.

For the remainder of (5.2), we write

$$\mathbb{P}[W = X - e^{(j)}] - \mathbb{P}[W = X] = \int_{[X]} \varphi_n(t) D_j(t)\,dt,$$

where

$$D_j(t) := \exp\left\{-\frac{1}{2n}\{-2[\Sigma^{-1}(t - nc)]_j + (\Sigma^{-1})_{jj}\}\right\} - 1.$$

Since $|e^x - 1 - x| \leq \frac{1}{2}x^2 e^{|x|}$, it follows that, for $|X - nc|_\Sigma \leq n\delta/3$,

$$\begin{aligned}
&\left|D_j(t) - \frac{1}{n}[\Sigma^{-1}(t - nc)]_j\right| \\
&\qquad \leq \frac{1}{2n}|(\Sigma^{-1})_{jj}| + \frac{1}{n^2}\left\{([\Sigma^{-1}(t - nc)]_j)^2 + \frac{1}{4}(\Sigma^{-1})_{jj}^2\right\} e^{\xi_j(\delta)},
\end{aligned}$$

where

$$\begin{aligned}
\xi_j(\delta) &:= \frac{1}{n}\max_{|X-nc|_\Sigma \leq n\delta/3}\left\{|[\Sigma^{-1}(X - nc)]_j| + |(\Sigma^{-1})_{jj}| + \frac{1}{2}d^{1/2}\|\Sigma^{-1}\|\right\} \\
&\leq \frac{1}{3}\|\Sigma^{-1/2}\|\delta + \frac{3}{2\lambda_{\min}(\Sigma)} =: \xi^*(\delta),
\end{aligned} \tag{5.4}$$

if $n \geq d^{1/2}/\lambda_{\min}(\Sigma)$, true in turn if $n \geq n_1 := (\lambda_{\min}(\Sigma))^{-8/7}$, because $n \geq d^4$. Note also that $n_1 \geq 1/\lambda_{\min}(\Sigma)$. Hence, fixing δ, for such X and for $t \in [X]$,

$$\left|D_j(t) - \frac{1}{n}[\Sigma^{-1}(X - nc)]_j\right| \leq C_2(\delta) n^{-1}(d^{1/2} + n^{-1}[\Sigma^{-1}(X - nc)]_j^2), \tag{5.5}$$

for $C_2(\delta) := 2e^{\xi^*(\delta)}/\lambda_{\min}(\Sigma) \in \mathcal{K}_\Sigma(\delta)$, again if $n \geq n_1$. This in turn implies that, for $|X - nc|_\Sigma \leq n\delta/3$,

$$\begin{aligned}
&|\{\mathbb{P}[W = X - e^{(j)}] - \mathbb{P}[W = X]\} - n^{-1}\mathbb{P}[W = X][\Sigma^{-1}(X - nc)]_j| \\
&\qquad \leq C_2(\delta) n^{-1}(d^{1/2} + n^{-1}[\Sigma^{-1}(X - nc)]_j^2)\mathbb{P}[W = X],
\end{aligned} \tag{5.6}$$

and hence that

$$\begin{aligned}&\Big|\sum_{X\in\mathbb{Z}^d} f(X)\{\mathbb{P}[W=X-e^{(j)}]-\mathbb{P}[W=X]\}I_n^\delta(X)\\ &\qquad -n^{-1}\mathbb{E}\{f(W)[\Sigma^{-1}(W-nc)]_j I_n^\delta(W)\}\Big|\\ &\qquad \le C_2(\delta)n^{-1}\mathbb{E}\{d^{1/2}+n^{-1}[\Sigma^{-1}(W-nc)]_j^2\}\|f\|^{\Sigma}_{n\delta/2,\infty}.\end{aligned}\tag{5.7}$$

Now, writing $b=\sum_{j=1}^d b_j e^{(j)}$ and using linearity and Lemma 2.1(c), requiring $n\ge d/\{4(\lambda_{\min}(\Sigma))^2\}$, the inequality (a) follows, if $n\ge\max\{n_{2.2},\psi_\Sigma(\delta)\}$, where

$$n_{2.2}:=\max\{d^4,n_1,\{4(\lambda_{\min}(\Sigma))^2\}^{-4/3}\},\tag{5.8}$$

with

$$C^{(1)}_{2.2}(\delta):=C_1(\delta)+C_2(\delta)\{1+C'(1)(1+1/\lambda_{\min}(\Sigma))\}.\tag{5.9}$$

For (b), bounding the difference between $\mathbb{E}\{\Delta f(W)^T B(W-nc)I_n^\delta(W)\}$ and $\mathbb{E}\{f(W)[n^{-1}(W-nc)^T\Sigma^{-1}B(W-nc)-\operatorname{Tr}B]I_n^\delta(W)\}$, we argue in similar style. For $i\ne j$, writing $E^{(ji)}:=e^{(j)}(e^{(i)})^T$, we have

$$\begin{aligned}&\mathbb{E}\{\Delta f(W)^T E^{(ji)}(W-nc)I_n^\delta(W)\}\\ &\quad=\mathbb{E}\{\Delta_j f(W)(W_i-nc_i)I_n^\delta(W)\}\\ &\quad=\sum_{X\in\mathbb{Z}^d} f(X)(X_i-nc_i)\{\mathbb{P}[W=X-e^{(j)}]-\mathbb{P}[W=X]\}I_n^\delta(X)\\ &\qquad+\sum_{X\in\mathbb{Z}^d} f(X)(X_i-nc_i)\mathbb{P}[W=X-e^{(j)}]\{I_n^\delta(X-e^{(j)})-I_n^\delta(X)\}.\end{aligned}\tag{5.10}$$

For $n\ge\max\{n_{2.2},\psi_\Sigma(\delta)\}$, we bound the second element in (5.10) much as for (5.3), using a Markov inequality, Cauchy–Schwarz and Lemma 2.1(a,b), giving

$$\begin{aligned}&\Big|\sum_{X\in\mathbb{Z}^d} f(X)(X_i-nc_i)\mathbb{P}[W=X-e^{(j)}]\{I_n^\delta(X-e^{(j)})-I_n^\delta(X)\}\Big|\\ &\quad\le\mathbb{E}\{|W_i-nc_i|I[|W-nc|_\Sigma>n\delta/6]\}\|f\|^{\Sigma}_{n\delta/2,\infty}\\ &\quad\le(6/n\delta)^3\mathbb{E}\{|W_i-nc_i||W-nc|^3_\Sigma\}\|f\|^{\Sigma}_{n\delta/2,\infty}\\ &\quad\le(6/n\delta)^3\sqrt{\mathbb{E}|W_i-nc_i|^2}\sqrt{\mathbb{E}|W-nc|^6_\Sigma}\|f\|^{\Sigma}_{n\delta/2,\infty}\\ &\quad\le(6/\delta)^3 n^{-1}d^{3/2}\sqrt{2(1+\Sigma_{ii})C(6)}\|f\|^{\Sigma}_{n\delta/2,\infty}\\ &\quad\le d^{3/2}C_3(\delta)n^{-1}\|f\|^{\Sigma}_{n\delta/2,\infty},\end{aligned}\tag{5.11}$$

where $C_3(\delta) = (6/\delta)^3\sqrt{2(1+\lambda_{\max}(\Sigma))C(6)} \in \mathcal{K}_\Sigma(\delta)$. The first element in (5.10) is treated using (5.6), Cauchy–Schwarz and Lemma 2.1(b,c), giving

$$\begin{aligned}
&\Bigg|\sum_{X\in\mathbb{Z}^d} f(X)(X_i - nc_i)\{\mathbb{P}[W = X - e^{(j)}] - \mathbb{P}[W = X]\}I_n^\delta(X) \\
&\qquad - n^{-1}\mathbb{E}\{f(W)(W_i - nc_i)[\Sigma^{-1}(W - nc)]_j I_n^\delta(W)\}\Bigg| \\
&\qquad \le C_2(\delta)n^{-1}\mathbb{E}\{|W_i - nc_i|(d^{1/2} + n^{-1}[\Sigma^{-1}(W - nc)]_j^2)\}\|f\|_{n\delta/2,\infty}^{\Sigma} \\
&\qquad \le C_2(\delta)n^{-1/2}\sqrt{2(1+\Sigma_{ii})}\Big(d^{1/2} + \sqrt{C'(2)(1+(\Sigma^{-1})_{ii}^2)}\Big)\|f\|_{n\delta/2,\infty}^{\Sigma} \\
&\qquad \le d^{1/2}C_4(\delta)n^{-1/2}\|f\|_{n\delta/2,\infty}^{\Sigma},
\end{aligned} \tag{5.12}$$

with

$$C_4(\delta) := C_2(\delta)\sqrt{2(1+\lambda_{\max}(\Sigma)}\big(1 + \sqrt{C'(2)(1+\lambda_{\min}(\Sigma)^{-2})}\big) \in \mathcal{K}_\Sigma(\delta).$$

Note that

$$(W_i - nc_i)[\Sigma^{-1}(W - nc)]_j = (W - nc)^T\Sigma^{-1}E^{(ji)}(W - nc).$$

For $i = j$, there is an extra term:

$$\begin{aligned}
&\mathbb{E}\{\Delta f(W)^T E^{(ii)}(W - nc)I_n^\delta(W)\} \\
&\qquad = \sum_{X\in\mathbb{Z}^d} f(X)(X_i - nc_i)\{\mathbb{P}[W = X - e^{(i)}] - \mathbb{P}[W = X]\}I_n^\delta(X) \\
&\qquad\quad + \sum_{X\in\mathbb{Z}^d} f(X)(X_i - nc_i)\mathbb{P}[W = X - e^{(i)}]\{I_n^\delta(X - e^{(i)}) - I_n^\delta(X)\} \\
&\qquad\quad - \sum_{X\in\mathbb{Z}^d} f(X)\mathbb{P}[W = X - e^{(i)}]I_n^\delta(X - e^{(i)}).
\end{aligned} \tag{5.13}$$

Now

$$\begin{aligned}
&\sum_{X\in\mathbb{Z}^d} f(X)\mathbb{P}[W = X - e^{(i)}]I_n^\delta(X - e^{(i)}) \\
&\qquad = \mathbb{E}\{\Delta_i f(W)I_n^\delta(W)\} + \mathbb{E}\{f(W)I_n^\delta(W)\},
\end{aligned}$$

and $|\mathbb{E}\{\Delta_i f(W)I_n^\delta(W)\}| \le \|\Delta f\|_{n\delta/2,\infty}^\Sigma$, giving

$$\begin{aligned}\tag{5.14} &\big|\mathbb{E}\{\Delta f(W)^T E^{(ii)}(W-nc)I_n^\delta(W)\} \\ &\quad - n^{-1}\mathbb{E}\{f(W)(W-nc)^T\Sigma^{-1}E^{(ii)}(W-nc)I_n^\delta(W)\} \\ &\quad - \mathbb{E}\{f(W)I_n^\delta(W)\}\big| \\ &\quad\le d^{3/2}C_3(\delta)n^{-1}\|f\|_{n\delta/2,\infty}^\Sigma + d^{1/2}C_4(\delta)n^{-1/2}\|f\|_{n\delta/2,\infty}^\Sigma \\ &\qquad + \|\Delta f\|_{n\delta/2,\infty}^\Sigma.\end{aligned}$$

The second estimate now follows for general $B = \sum_{i=1}^d \sum_{j=1}^d B_{ij}E^{(ij)}$, by linearity, with

$$C_{2.2}^{(2)}(\delta) := C_4(\delta) + C_3(\delta), \tag{5.15}$$

provided that $n \ge d^2$.

The proof of the final part of Lemma 2.2, bounding the difference between $\mathbb{E}\{\Delta f(W)^T B(W-nc)I_n^\delta(W)\}$ and

$$\mathbb{E}\{f(W)[n^{-1}(W-nc)^T\Sigma^{-1}B(W-nc) - \operatorname{Tr} B]I_n^\delta(W)\},$$

proceeds in very much the same way, but starting with $e^{(j)}b^T$ in place of $E^{(ji)}$ in (5.10) and (5.13), for any $b \in \mathbb{R}^d$, and then writing $B = \sum_{j=1}^d e^{(j)}b(j)^T$ with $b(j) := B^T e^{(j)}$. The quantities $(X_i - nc_i)$ and $(W_i - nc_i)$ are replaced in the computations by $b^T(X-nc)$ and $b^T(W-nc) = b^T\Sigma^{1/2}\Sigma^{-1/2}(W-nc)$, respectively. The error terms corresponding to (5.11) and (5.12) then yield the bounds $d^2C_3'(\delta)|b|n^{-1}\|f\|_{n\delta/2,\infty}^\Sigma$ and $dC_4'(\delta)|b|n^{-1/2}\|f\|_{n\delta/2,\infty}^\Sigma$, with

$$C_3'(\delta) := (6/\delta)^3 C(4)\sqrt{\lambda_{\max}(\Sigma)}; \tag{5.16}$$

$$C_4'(\delta) := C_2(\delta)\sqrt{C(2)\lambda_{\max}(\Sigma)}\{1 + \sqrt{C'(2)(1+\lambda_{\min}(\Sigma)^{-2})}\}, \tag{5.17}$$

giving $C_{2.2}^{(3)}(d) = C_4'(\delta) + C_3'(\delta)$. The analogue of (5.13) yields an error bounded by $|b_j|\|\Delta f\|_{n\delta/2,\infty}^\Sigma$, and Part (c) now follows.

5.3. *Proof of Lemma* 3.2. To bound the moments of $Z := (nd\nu)^{-1/2} \times \Sigma^{-1/2}(W-\mu)$, we use the equation $\mathbb{E}h(W') - \mathbb{E}h(W) = 0$ for suitably chosen real functions h. First, we take $h(w) = (w-\mu)^T\Sigma^{-1}(w-\mu)$, giving

$$\mathbb{E}\{2\xi^T\Sigma^{-1}(W-\mu) + \xi^T\Sigma^{-1}\xi\} = 0.$$

Noting that $\xi^T\Sigma^{-1}\xi = \operatorname{Tr}(\Sigma^{-1/2}\xi\xi^T\Sigma^{-1/2})$, and using (3.1), we have

$$\begin{aligned}&-\mathbb{E}\{2n^{-1}(W-\mu)^T\Sigma^{-1}A(W-\mu) + 2n^{-1/2}\|A\|^{1/2}R_1(W)^T\Sigma^{-1}(W-\mu)\} \\ &\quad = \mathbb{E}\operatorname{Tr}(\sigma_\Sigma^2(W)) = \operatorname{Tr}(\sigma_\Sigma^2),\end{aligned}$$

where $\sigma_\Sigma^2(W) := \Sigma^{-1/2}\sigma^2(W)\Sigma^{-1/2}$. Writing $s_n^2 := \mathbb{E}|Z|^2$, it follows from (3.3) and because $A\Sigma + \Sigma A^T + \sigma^2 = 0$ that

$$2\alpha_1 d\nu s_n^2 \le (d\nu\alpha_1)^{1/2}\bigl(\mathrm{Tr}(\sigma_\Sigma^2)\bigr)^{1/2}\bigl(1+s_n^2\bigr) + \mathrm{Tr}\bigl(\sigma_\Sigma^2\bigr).$$

From the definition of ν in (3.2), it thus follows directly that $s_n^2 \le 2$, establishing the first part.

For the third moment, we start with $h(z) = (1+z^Tz)^{3/2}$. The function h has derivatives

$$Dh(z) = 3\bigl(1+z^Tz\bigr)^{1/2}z$$

and

$$D^2h(z) = \frac{3zz^T}{(1+z^Tz)^{1/2}} + 3\bigl(1+z^Tz\bigr)^{1/2}I.$$

Furthermore,

$$\begin{aligned}&\biggl|\bigl\{h(z+\zeta)-h(z)\bigr\} - 3\bigl(1+z^Tz\bigr)^{1/2}\zeta^Tz\\ &\quad - \frac{3(\zeta^Tz)^2}{2(1+z^Tz)^{1/2}} - \frac{3}{2}\bigl(1+z^Tz\bigr)^{1/2}|\zeta|^2\biggr|\\ &\quad =: d_3(h,z,\zeta) \le k_{3,h}|\zeta|^3,\end{aligned} \tag{5.18}$$

for a constant $k_{3,h} \le 22$ that does not depend on d. This can be seen by considering separately the cases where $|\zeta| \ge (|z| \vee 1)$, $|\zeta| \le |z|$ and $1 \ge |\zeta| \ge |z|$.

For $|\zeta| \ge (|z| \vee 1)$, simply take the terms one by one, giving

$$d_3(h,z,\zeta) \le |\zeta|^3\biggl(\bigl\{5^{3/2}+2^{3/2}\bigr\} + 3\cdot 2^{1/2} + \frac{3}{2} + \frac{3}{2}\cdot 2^{1/2}\biggr) \le 22|\zeta|^3.$$

For $1 \ge |\zeta| \ge |z|$, use the bounds

$$\bigl|(1+x)^{1/2}-1\bigr| \le \frac{1}{2}x^{1/2}; \qquad \biggl|(1+x)^{3/2} - 1 - \frac{3}{2}x\biggr| \le \frac{3}{8}x^{3/2}$$

in $0 \le x \le 1$ to give

$$\bigl|\bigl(1+z^Tz\bigr)^{1/2}-1\bigr| \le \frac{1}{2}|\zeta|; \qquad \biggl|h(z+\zeta)-h(z) - \frac{3}{2}\bigl(2\zeta^Tz+\zeta^T\zeta\bigr)\biggr| \le 27|\zeta|^3/8.$$

Then the first, second and fourth terms in $d_3(h,z,\zeta)$ together give at most

$$|\zeta|^3\biggl(\frac{27}{8}+\frac{3}{2}+\frac{3}{4}\biggr) \le \frac{45}{8}|\zeta|^3,$$

and the third adds at most $\frac{3}{2}|\zeta|^3$ to this. For $|\zeta| \le |z|$, Taylor's expansion gives

$$\biggl|(1+x+y)^{3/2} - (1+x)^{3/2} - \frac{3}{2}y(1+x)^{1/2} - \frac{3y^2}{8\sqrt{1+x}}\biggr| \le \frac{|y|^3}{16(1+x)^{3/2}}.$$

We take $x = z^T z$ and $y = 2\zeta^T z + \zeta^T \zeta$, for which $|y| \le 3|\zeta||z|$. The first, second and fourth terms in $d_3(h, z, \zeta)$ together thus give

$$\frac{3(2\zeta^T z + \zeta^T \zeta)^2}{8(1 + z^T z)^{1/2}},$$

up to an error of at most

$$\frac{|2\zeta^T z + \zeta^T \zeta|^3}{16(1 + z^T z)^{3/2}} \le \frac{27|\zeta|^3|z|^3}{16|z|^3} \le \frac{27}{16}|\zeta|^3.$$

Then

$$\left|\frac{3(2\zeta^T z + \zeta^T \zeta)^2}{8(1 + z^T z)^{1/2}} - \frac{3(\zeta^T z)^2}{2(1 + z^T z)^{1/2}}\right| \le \frac{12|\zeta|^3|z| + 3|\zeta|^4}{8|z|} \le \frac{15}{8}|\zeta|^3,$$

giving an overall bound of $\frac{57}{16}|\zeta|^3$.

We now substitute $z = Z = z(W)$ and $\zeta = (nd\nu)^{-1/2}\Sigma^{-1/2}\xi$ into (5.18), and take expectations. Since

$$\mathbb{E}h\bigl(Z(W + \xi)\bigr) = \mathbb{E}h\bigl(Z(W)\bigr),$$

this immediately gives

$$\begin{aligned}
&\mathbb{E}\bigl\{-3\bigl(1 + Z^T Z\bigr)^{1/2}(nd\nu)^{-1/2}\xi^T \Sigma^{-1/2} Z\bigr\} \\
&\qquad \le n^{-1}\mathbb{E}\left\{\frac{3Z^T \Sigma^{-1/2}\xi\xi^T \Sigma^{-1/2} Z}{2d\nu(1 + Z^T Z)^{1/2}} + \frac{3}{2}\bigl(1 + Z^T Z\bigr)^{1/2}\frac{|\Sigma^{-1/2}\xi|^2}{d\nu}\right\} \\
&\qquad\quad + \frac{k_{3,h}\chi_\Sigma}{(nd\nu)^{3/2}} \\
&\qquad \le n^{-1}\frac{3}{2d\nu}\mathbb{E}\bigl\{\bigl(2|Z| + 1\bigr)\bigl|\Sigma^{-1/2}\xi\bigr|^2\bigr\} + \frac{k_{3,h}\chi_\Sigma}{(nd\nu)^{3/2}}.
\end{aligned} \tag{5.19}$$

Now

$$\begin{aligned}
&\mathbb{E}\bigl\{-3\bigl(1 + Z^T Z\bigr)^{1/2}(nd\nu)^{-1/2}\xi^T \Sigma^{-1/2} Z\bigr\} \\
&\qquad = \mathbb{E}\bigl\{-3\bigl(1 + Z^T Z\bigr)^{1/2}(nd\nu)^{-1/2} \\
&\qquad\qquad \times \bigl(n^{-1}(W - \mu)^T A^T + n^{-1/2}\|A\|^{1/2} R_1(W)^T\bigr)\Sigma^{-1/2} Z\bigr\} \\
&\qquad = n^{-1}\mathbb{E}\biggl\{3\bigl(1 + Z^T Z\bigr)^{1/2} \\
&\qquad\qquad \times \left(\frac{1}{2} Z^T \sigma_\Sigma^2 Z - (d\nu)^{-1/2}\|A\|^{1/2} R_1(W)^T \Sigma^{-1/2} Z\right)\biggr\},
\end{aligned} \tag{5.20}$$

and, using (3.4),

$$(5.21)\quad (d\nu)^{-1/2}\|A\|^{1/2}\mathbb{E}\{(1+Z^TZ)^{1/2}|R_1(W)^T\Sigma^{-1/2}Z)|\} \le \frac{1}{4}\alpha_1(1+\mathbb{E}|Z|^3).$$

Then, by the arithmetic and geometric means inequality, for any $a > 0$,

$$|Z||\Sigma^{-1/2}\xi|^2 \le \frac{1}{3}\{(a|Z|)^3 + 2(a^{-1/2}|\Sigma^{-1/2}\xi|)^3\},$$

so that, taking $a = (d\nu\alpha_1)^{1/3}$,

$$(5.22)\qquad (d\nu)^{-1}\mathbb{E}\{|Z||\Sigma^{-1/2}\xi|^2\} \le \frac{\alpha_1}{3}\{\mathbb{E}|Z|^3 + 2(n/\|A\|)^{1/2}L_\Sigma\}.$$

Combining (5.19)–(5.22), recalling that $\mathrm{Tr}(\sigma_\Sigma^2) = d\nu\alpha_1$, and multiplying by n, it follows that

$$3\alpha_1\mathbb{E}|Z|^3 \le \frac{3}{4}\alpha_1(1+\mathbb{E}|Z|^3) + \alpha_1\{\mathbb{E}|Z|^3 + 2(n/\|A\|)^{1/2}L_\Sigma\} + \frac{3}{2}\alpha_1 + k_{3,h}\alpha_1 L_\Sigma\sqrt{\frac{\alpha_1}{\|A\|}},$$

giving $\mathbb{E}|Z^3| \le 2(1+10(n/\|A\|)^{1/2}L_\Sigma)$ if $n/\alpha_1 \ge 1$, because $k_{3,h} \le 22$. The final inequality is then immediate.

REFERENCES

BARBOUR, A. D., LUCZAK, M. J. and XIA, A. (2018). Multivariate approximation in total variation, I: Equilibrium distributions of Markov jump processes. *Ann. Probab.* **46** 1351–1404.

BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117

CHEN, L. H. Y., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal Approximation by Stein's Method*. Springer, Heidelberg. MR2732624

FANG, X. (2014). Discretized normal approximation by Stein's method. *Bernoulli* **20** 1404–1431. MR3217448

FANG, X. and RÖLLIN, A. (2015). Rates of convergence for multivariate normal approximation with applications to dense graphs and doubly indexed permutation statistics. *Bernoulli* **21** 2157–2189. MR3378463

GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

LINDVALL, T. (2002). *Lectures on the Coupling Method*. Dover, Mineola, NY. MR1924231

REINERT, G. and RÖLLIN, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** 2150–2173. MR2573554

RINOTT, Y. and ROTAR, V. (1996). A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Anal.* **56** 333–350. MR1379533

RÖLLIN, A. and ROSS, N. (2015). Local limit theorems via Landau–Kolmogorov inequalities. *Bernoulli* **21** 851–880. MR3338649

STEIN, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. IMS, Hayward, CA. MR0882007

A. D. BARBOUR
INSTITUT FÜR MATHEMATIK
UNIVERSITÄT ZÜRICH
WINTERTHURERTRASSE 190
CH-8057 ZÜRICH
SWITZERLAND
E-MAIL: a.d.barbour@math.uzh.ch

M. J. LUCZAK
SCHOOL OF MATHEMATICAL SCIENCES
QUEEN MARY UNIVERSITY OF LONDON
MILE END ROAD
LONDON E1 4NS
UNITED KINGDOM
E-MAIL: malwinajluczak@gmail.com

A. XIA
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MELBOURNE
PARKVILLE, VICTORIA 3010
AUSTRALIA
E-MAIL: a.xia@ms.unimelb.edu.au