Multivariate approximation in total variation, I: equilibrium distributions of Markov jump processes

A. D. Barbour[1], M. J. Luczak[2] & A. Xia[3]
Universität Zürich, Queen Mary University of London & University of Melbourne

December 22, 2016

Abstract

For integer valued random variables, the translated Poisson distributions form a flexible family for approximation in total variation, in much the same way that the normal family is used for approximation in Kolmogorov distance. Using the Stein–Chen method, approximation can often be achieved with error bounds of the same order as those for the CLT. In this paper, an analogous theory, again based on Stein's method, is developed in the multivariate context. The approximating family consists of the equilibrium distributions of a collection of Markov jump processes, whose analogues in one dimension are the immigration–death processes with Poisson distributions as equilibria. The method is illustrated by providing total variation error bounds for the approximation of the equilibrium distribution of one Markov

[1]Institut für Mathematik, Universität Zürich, Winterthurertrasse 190, CH-8057 Zürich; e-mail a.d.barbour@math.uzh.ch. 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]School of Mathematical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK; e-mail m.luczak@qmul.ac.uk. 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]School of Mathematics and Statistics, University of Melbourne, Parkville, VIC 3010, Australia; e-mail a.xia@ms.unimelb.edu.au. Work supported in part by Australian Research Council Grants Nos DP120102398 and DP150101459.

jump process by that of another. In a companion paper, it is shown how to use the method for discrete normal approximation in $\mathbb{Z}^d$.

Keywords: Markov jump process; multivariate approximation; total variation distance; infinitesimal generator; Stein's method
AMS subject classification: Primary 62E17; Secondary 62E20, 60J27, 60C05
Running head: Multivariate approximation

1 Introduction

The Stein–Chen method (Chen, 1975) enables the distribution of a sum W of indicator random variables to be approximated by a Poisson distribution in a wide variety of circumstances. In addition, it provides an estimate of the accuracy of the approximation, expressed in terms of the *total variation distance*. Such an approximation is very valuable, since it allows the approximation of the probability $\mathbb{P}[W \in A]$ of an arbitrary subset A of $\mathbb{Z}_+$ by a Poisson probability, and not just of sets A with 'nice' properties. By contrast, the distance classically used for quantifying normal approximation is the Kolmogorov distance, as in the Berry–Esseen theorem, and this measures the largest difference between the probabilities of half lines. Of course, this can easily be extended to (the unions of small numbers of) intervals, but gives no information at all, for instance, about the probability that W is even.

The Poisson family of distributions is, however, too restrictive to be used as widely as the normal distribution for approximation, because mean and variance have to be equal. Starting from the seminal paper of Presman (1983), more general approximations in total variation have been derived, using more flexible families. In particular, for the translated Poisson family, the Stein–Chen method can be adapted in a natural way (Röllin 2005, 2007), allowing for the possibility of treating sums of dependent indicator random variables. What is more, the order of the error in total variation approximation obtained in this way, using the translated Poisson family (Barbour & Xia, 1999) or the discretized normal family (Fang, 2014), need be no worse than that of the error in the normal approximation, measured using Kolmogorov distance. This represents a substantial gain in the scope of the approximation, at relatively small cost.

In this paper, we aim for analogous results in higher dimensions, an undertaking of considerably greater difficulty. The first step is to choose a suitable family of reference distributions. For the Poisson distribution $\mathrm{Po}\,(\lambda)$, there is a Markov jump process, the immigration–death process with constant immigration rate λ and unit *per capita* death rate, whose equilibrium distribution is exactly $\mathrm{Po}\,(\lambda)$, and whose generator can be used as the corresponding Stein operator (Barbour, 1988). Proceeding by analogy, we consider the equilibrium distributions of more general Markov jump processes as possible reference distributions. As in the Poisson case, their generators automatically yield corresponding Stein equations (Barbour, Holst & Janson, Section 10.1). In addition, they come with a probabilistic representation of the solutions to the Stein equation that makes it possible to estimate the quantities needed in exploiting the method. Although there is often no readily available exact representation of the equilibrium distributions of Markov jump processes, they are shown in Theorem 2.3 of Part II, under a weak irreducibility condition, to be close in total variation to discrete multivariate normal distributions, provided that their spread is large. In practice, this allows the discrete normal family to be used instead for approximation, without any material loss of accuracy.

We begin with a sequence $(X_n,\, n \ge 1)$ of density dependent Markov jump processes on $\mathbb{Z}^d$, where X_n has transition rates

$$X \to X + J \quad \text{at rate} \quad ng^J(n^{-1}X), \qquad X \in \mathbb{Z}^d,\ J \in \mathcal{J}, \tag{1.1}$$

$\mathcal{J}$ is a finite subset of $\mathbb{Z}^d$, and the functions g^J are twice continuously differentiable on $\mathbb{R}^d$. For Poisson approximation in one dimension, we take $\mathcal{J} := \{-1, 1\}$ with $g^{-1}(x) = x$ and $g^1(x) = \mu$ for $x \in \mathbb{R}$, giving a family of immigration–death processes X_n with equilibrium distributions $\mathrm{Po}\,(n\mu)$; n plays the part of the number of summands in the CLT. In higher dimensions, the family is chosen to allow greater flexibility. We initially suppose only that the equations

$$\frac{d\xi}{dt} \;=\; F(\xi) \;:=\; \sum_{J \in \mathcal{J}} J g^J(\xi) \tag{1.2}$$

have an equilibrium point c, so that $F(c) = 0$; that the matrix

$$A \;:=\; DF(c) \tag{1.3}$$

has eigenvalues whose real parts are all negative, making c a strongly stable equilibrium of (1.2); and that the symmetric matrix

$$\sigma^2 \;:=\; \sigma^2(c), \quad \text{where} \quad \sigma^2(x) \;:=\; \sum_{J \in \mathcal{J}} JJ^T g^J(x), \tag{1.4}$$

is positive definite.

The process X_n has generator given by

$$(\mathcal{A}_n h)(X) \;:=\; \sum_{J \in \mathcal{J}} ng^J(n^{-1}X)(h(X+J) - h(X)) \tag{1.5}$$

for bounded $h\colon \mathbb{Z}^d \to \mathbb{R}$. To approximate the distribution of a random vector $W \in \mathbb{Z}^d$ in total variation by the equilibrium distribution Π_n of X_n, should it exist, a key step in using Stein's method is to show that the expectation $\mathbb{E}\{\mathcal{A}_n h(W)\}$ is small for a large class of bounded functions h. In our theorems, we use the functions $h = h_f$ that are determined by solving the Stein equation

$$(\mathcal{A}_n h)(X) \;=\; f(X) \tag{1.6}$$

for h, given any bounded $f\colon \mathbb{Z}^d \to \mathbb{R}$. However, for ease of use, we replace the operator $\mathcal{A}_n$ as Stein operator by the simpler operator

$$\widetilde{\mathcal{A}}_n h(w) \;:=\; \frac{n}{2}\mathrm{Tr}\,(\sigma^2\Delta^2 h(w)) + \Delta h^T(w) A(w - nc), \quad w \in \mathbb{Z}^d, \tag{1.7}$$

where $c \in \mathbb{R}^d$, A and σ^2 are as in (1.3) and (1.4), respectively; here,

$$\Delta_j h(w) \;:=\; h(w + e^{(j)}) - h(w); \quad \Delta^2_{jk} h(w) \;:=\; \Delta_j(\Delta_k h)(w), \tag{1.8}$$

for $1 \le j, k \le d$, where $e^{(j)}$ denotes the j-th coordinate vector. It is shown in Theorem 4.6 that $\widetilde{\mathcal{A}}_n$ is close enough to the original operator $\mathcal{A}_n$ for our purposes.

We also define Σ to be the positive definite symmetric solution of the continuous Lyapounov equation

$$A\Sigma + \Sigma A^T + \sigma^2 \;=\; 0; \tag{1.9}$$

see, for example, Khalil (2002, Theorem 4.6, p.136). Now $n\Sigma$ turns out to be asymptotically equivalent to the covariance matrix of our approximating distribution. For a given random vector W whose distribution we wish to approximate, it is thus clearly a good idea to choose n, A and σ^2 in such a way that, solving (1.9), $n\Sigma \approx \mathrm{Cov}\, W$. There are typically many choices of A and σ^2 that yield the same Σ as solution of (1.9), and which one is best to use in (1.7) is usually dictated by the specific context. Having chosen A and σ^2, it is shown in Theorem 3.2 that there indeed exists a Markov jump process X_n as in (1.1) that yields the corresponding matrices in (1.3) and (1.4).

Even under the condition that all the eigenvalues of A in (1.3) have negative real parts, the process X_n may not have an equilibrium. However, it is shown in Barbour & Pollett (2012, Section 4) that it has a quasi-equilibrium close to nc, and that this is asymptotically extremely close to the equilibrium distribution Π_n^δ of its restriction to a $n\delta$-ball around nc, whatever the value of $\delta > 0$. For technical reasons, we use balls in $\mathbb{R}^d$ derived from the norm $|\cdot|_\Sigma$ defined by

$$|Y|^2_\Sigma \;:=\; Y^T \Sigma^{-1} Y, \tag{1.10}$$

where Σ is as defined above; we let $B_{\delta,\Sigma}(c) := \{\xi \in \mathbb{R}^d \colon |\xi - c|_\Sigma \le \delta\}$. Defining

$$\mathcal{X}_n^\delta(J) := \bigl\{X \in \mathbb{Z}^d \colon \{X, X+J\} \subset B_{n\delta,\Sigma}(nc)\bigr\}, \tag{1.11}$$

we replace X_n with the process X_n^δ having transition rates

$$X \to X + J \;\text{ at rate }\; ng_\delta^J(n^{-1}X) := \begin{cases} ng^J(n^{-1}X), & \text{if } X \in \mathcal{X}_n^\delta(J); \\ 0, & \text{otherwise,} \end{cases} \tag{1.12}$$

for $X \in \mathbb{Z}^d$ and $J \in \mathcal{J}$, with δ to be chosen suitably small and positive; broadly speaking, we choose δ so that c is a strongly attractive equilibrium of the equations (1.2) throughout $B_{\delta,\Sigma}(c)$. Then, if

$$X_n^\delta(0) \in \widetilde{B}_{n,\delta}(c) := \mathbb{Z}^d \cap B_{n\delta,\Sigma}(nc), \tag{1.13}$$

it follows that X_n^δ is a Markov process on the finite state space $\widetilde{B}_{n,\delta}(c)$, and so has an equilibrium distribution; furthermore, if all states in $\widetilde{B}_{n,\delta}(c)$ communicate, this equilibrium distribution Π_n^δ is unique. Assumptions G3 and G4 below guarantee that this is the case: see Lemma 2.1.

Now, if $X_n^\delta \sim \Pi_n^\delta$, it follows by Dynkin's formula and because each set $\mathcal{X}_n^\delta(J)$ is bounded that $\mathbb{E}\{\mathcal{A}_n^\delta h(X_n^\delta)\} = 0$ for all functions $h \colon \mathbb{Z}^d \to \mathbb{R}$, where

$$\mathcal{A}_n^\delta h(X) := n \sum_{J \in \mathcal{J}} g_\delta^J(n^{-1}X)\{h(X+J) - h(X)\}, \qquad X \in \mathbb{Z}^d. \tag{1.14}$$

The essence of Stein's method for total variation approximation is to find a function $h_B = h_{B,n}^\delta$ that solves the equation

$$\mathcal{A}_n^\delta h_B(X) = \mathbb{1}_B(X) - \Pi_n^\delta\{B\}, \quad X \in \widetilde{B}_{n,\delta}(c), \tag{1.15}$$

for each $B \subset \widetilde{B}_{n,\delta}(c)$. Then, if W is any random element of $\mathbb{Z}^d$ and $B \subset \widetilde{B}_{n,\delta}(c)$, it follows that

$$\begin{aligned} \mathbb{P}[W \in B] - \Pi_n^\delta\{B\} &= \mathbb{E}\{(\mathbb{1}_B(W) - \Pi_n^\delta\{B\}) I[W \in \widetilde{B}_{n,\delta'}(c)]\} \\ &\qquad - \Pi_n^\delta\{B\}\mathbb{P}[W \notin \widetilde{B}_{n,\delta'}(c)], \end{aligned}$$

for any $\delta' \le \delta$, so that

$$d_{\mathrm{TV}}(\mathcal{L}(W), \Pi_n^\delta) \le \sup_{B \subset \widetilde{B}_{n,\delta}(c)} |\mathbb{E}\{\mathcal{A}_n^\delta h_B(W) I[W \in \widetilde{B}_{n,\delta'}(c)]\}| + \mathbb{P}[W \notin \widetilde{B}_{n,\delta'}(c)]. \tag{1.16}$$

Showing that $\mathcal{L}(W)$ is close to Π_n^δ in total variation thus reduces to showing that the right hand side of (1.16) is small. Bounding the probability $\mathbb{P}[W \notin$

$\widetilde{B}_{n,\delta'}(c)]$ typically involves direct estimates, such as Chebyshev's inequality. Thus the main effort goes into bounding $|\mathbb{E}\{\mathcal{A}_n^\delta h_B(W)\}|$.

In order to extract the essential parts of $\mathbb{E}\{\mathcal{A}_n^\delta h_B(W)\}$, we expand the expression for $\mathcal{A}_n^\delta h_B(X)$, using Newton's expansion. To control the remainders in the expansion, we need to be able to control the magnitudes of the first and second differences $\Delta_j h_B(X)$ and $\Delta_{jk}^2 h_B(X)$ for $1 \le j,k \le d$. We obtain bounds for these, given in Theorem 4.1, within a ball $|X - nc|_\Sigma \le n\delta/4$, for δ small enough. They are derived using the explicit representation

$$h_B(X) \;:=\; h_{B,n}^\delta(X) \;=\; -\int_0^\infty (\mathbb{P}[X_n^\delta(t) \in B \,|\, X_n^\delta(0) = X] - \Pi_n^\delta\{B\})\,dt, \tag{1.17}$$

(see Kemeny & Snell (1960, Theorem 5.13(d); 1961, Equation (9))), and depend on careful analysis of the Markov process X_n^δ. This is carried out in Sections 2 and 3. For the remainders in the expansion of $\mathbb{E}\{\mathcal{A}_n^\delta h_B(W)\}$ to be small, we also need to know that $d_{\rm TV}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)}))$ is small for each $1 \le j \le d$, and that $\mathbb{E}|W - nc|_\Sigma^2 \le vn$ for some constant v. This is true if $W \sim \Pi_n^\delta$, as is shown in Proposition 5.2, but needs to be proved separately for any W that is to be approximated by Π_n^δ.

As a result of these considerations, provided that $d_{\rm TV}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)}))$ is small for each $1 \le j \le d$ and that $\mathbb{E}|W - nc|_\Sigma^2 \le vn$, we shall have shown, for suitable $\delta > 0$, that $\mathbb{E}\{\mathcal{A}_n^\delta h_B(W) I_n^\delta(W)\}$ is close to $\mathbb{E}\{\widetilde{\mathcal{A}}_n h_B(W) I_n^\delta(W)\}$, where $I_n^\delta(X) := I[|X - nc|_\Sigma \le n\delta/3]$ and $\widetilde{\mathcal{A}}_n$ is as in (1.7). Hence, for any integer valued random vector W such that $\mathbb{E}\{\widetilde{\mathcal{A}}_n h_B(W) I_n^\delta(W)\}$ is uniformly small for all $B \subset \widetilde{B}_{n,\delta}(c)$, $d_{\rm TV}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)}))$ is small for each $1 \le j \le d$, $\mathbb{P}[W \notin \widetilde{B}_{n,\delta'}(c)]$ is small, and $\mathbb{E}|W - nc|_\Sigma^2 \le vn$, it follows from (1.16) that $d_{\rm TV}(\mathcal{L}(W), \Pi_n^\delta)$ is small. The precise statement of this conclusion, giving a set of quantities that bound $d_{\rm TV}(\mathcal{L}(W), \Pi_n^\delta)$ for an arbitrary integer valued random d-vector W, is presented in Theorem 4.8. An application is given in Section 5.

2 The analysis of X_n^δ: general processes

2.1 Main assumptions

The main arguments of the paper are based on the analysis of a sequence of Markov jump processes X_n, whose transition rates are given in (1.1). For some $\delta_0 > 0$, we make the following assumptions.

Assumption G0. The equations (1.2) have an equilibrium c; thus $F(c) = 0$.

Assumption G1. All eigenvalues of the matrix $A := DF(c)$ have negative real parts.

Assumption G2. For each $J \in \mathcal{J}$, the function g^J is of class C^2 in the Euclidean ball $B_{\delta_0}(c) := \{x\colon |x-c| \le \delta_0\}$.

Assumption G3. There exists $\varepsilon_0 > 0$ such that

$$\inf_{x\in B_{\delta_0}(c)} g^J(x) \;\ge\; \varepsilon_0 g^J(c) \;=:\; \mu_0^J \;>\; 0, \quad J \in \mathcal{J}.$$

Assumption G4. For each unit vector $e^{(j)} \in \mathbb{R}^d$, $1 \le j \le d$, there exists a finite sequence of elements $J_1^{(j)}, \ldots, J_{r(j)}^{(j)}$ of $\mathcal{J}$ such that

$$e^{(j)} \;=\; \sum_{l=1}^{r(j)} J_l^{(j)}.$$

For d-vectors, we use $|\cdot|$ to denote the Euclidean norm, $|\cdot|_1$ to denote the ℓ_1-norm, and $|X|_\Sigma$ to denote $|\Sigma^{-1/2}X|$. For a $d \times d$ matrix B, we define the spectral norm

$$\|B\| \;:=\; \sup_{y\in\mathbb{R}^d\colon |y|=1} |By|,$$

and use $\|B\|_1$ to denote $\sum_{i=1}^d \sum_{j=1}^d |B_{ij}|$. Note that, for any d-vector b and $d \times d$ matrix B, the inequalities

$$d^{-1}|b|_1 \;\le\; \sqrt{d^{-1}b^Tb} \quad\text{and}\quad d^{-2}\|B\|_1 \;\le\; \sqrt{d^{-2}\mathrm{Tr}\,(B^TB)} \;\le\; \sqrt{d^{-1}\|B\|^2}$$

yield

$$|b|_1 \;\le\; d^{1/2}|b| \quad\text{and}\quad \|B\|_1 \;\le\; d^{3/2}\|B\|. \tag{2.1}$$

For a $d\times d$ positive definite symmetric matrix M, we write $\bar\lambda(M)$ for $d^{-1}\mathrm{Tr}\,(M)$, $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ for its smallest and largest eigenvalues, respectively, and $\rho(M) := \lambda_{\max}(M)/\lambda_{\min}(M)$ for its condition number; we use $\mathrm{Sp}'(M)$ to denote the triple $(\bar\lambda(M), \lambda_{\min}(M), \lambda_{\max}(M))$.

For a real function $h\colon \mathbb{Z}^d \to \mathbb{R}$, we define

$$\|\Delta h(X)\|_\infty \;:=\; \max_{1\le i\le d} |\Delta_i h(X)|; \quad \|\Delta^2 h(X)\|_\infty \;:=\; \max_{1\le i,j\le d} |\Delta_{ij} h(X)|.$$

For any $a > 0$, we then set

$$\begin{aligned} \|h\|_{a,\infty} \;&:=\; \max\{|h(X)|\colon X \in \mathbb{Z}^d, |X-nc| \le a\}; \\ \|\Delta h\|_{a,\infty} \;&:=\; \max\{\|\Delta h(X)\|_\infty\colon X \in \mathbb{Z}^d, |X-nc| \le a\}; \\ \|\Delta^2 h\|_{a,\infty} \;&:=\; \max\{\|\Delta^2 h(X)\|_\infty\colon X \in \mathbb{Z}^d, |X-nc| \le a\}, \end{aligned} \tag{2.2}$$

for c as above. For $g\colon \mathbb{R}^d \to \mathbb{R}$ twice differentiable, we set

$$\|D^2 g(x)\| \;:=\; \limsup_{t\to 0} \sup_{y\colon |y|=1} t^{-1}|Dg(x+ty) - Dg(x)|,$$

where D denotes the differential operator. We then define the quantities

$$L_0 \;:=\; \max_{J\in\mathcal{J}} \frac{|g^J|_{\delta_0}}{g^J(c)}; \quad L_1 \;:=\; \max_{J\in\mathcal{J}} \frac{|Dg^J|_{\delta_0}}{g^J(c)}; \quad L_2 \;:=\; \max_{J\in\mathcal{J}} \frac{\|D^2 g^J\|_{\delta_0}}{g^J(c)}, \quad (2.3)$$

finite in view of Assumptions G2 and G3, where $\|H\|_\delta := \sup_{x\in B_\delta(c)} \|H(x)\|$, for any vector- or matrix-valued function H and for any choice of norm $\|\cdot\|$.

We also define

$$\begin{aligned}
&\Lambda \;:=\; \textstyle\sum_{J\in\mathcal{J}} g^J(c)|J|^2 \;=\; \mathrm{Tr}\,(\sigma^2); && \gamma \;:=\; \textstyle\sum_{J\in\mathcal{J}} g^J(c)|J|^3;\\
&J_{\max} \;:=\; \max_{J\in\mathcal{J}} |J|; && J^{\Sigma}_{\max} \;:=\; \max_{J\in\mathcal{J}} |\Sigma^{-1/2}J|;\\
&\sigma^2_\Sigma \;:=\; \Sigma^{-1/2}\sigma^2\Sigma^{-1/2}; && \alpha_1 \;:=\; \tfrac{1}{2}\lambda_{\min}(\sigma^2_\Sigma);\\
&\overline{\Lambda} \;:=\; \bar{\lambda}(\sigma^2) \;=\; d^{-1}\Lambda; \qquad \bar{\gamma} \;:=\; d^{-3/2}\gamma; && \mu_* \;:=\; \min_{J\in\mathcal{J}} \mu^J_0,
\end{aligned} \quad (2.4)$$

where σ^2 is defined in (1.4), and Σ in (1.9). In the sections that follow, we establish many bounds that depend on these basic parameters. They are mainly expressed as continuous functions of the elements of the set

$$\mathcal{K} \;:=\; \{L_0,\, L_1,\, L_2,\, \varepsilon_0,\, \mathrm{Sp}'(\sigma^2/\overline{\Lambda}),\, \mathrm{Sp}'(\Sigma),\, d^{-1}J_{\max},\, \|A\|/\overline{\Lambda},\, \overline{\Lambda}/\mu_*,\, \delta_0\}, \quad (2.5)$$

and, with slight abuse of notation, are said to belong to the set $\mathcal{K}$. If they are also continuous functions of another parameter, such as δ, they are said to belong to $\mathcal{K}(\delta)$. The $\overline{\Lambda}$-factors ensure that the quantities remain invariant if all the transition rates g^J are multiplied by the same constant. In particular, constants of the form κ_i and K_i belong to $\mathcal{K}$, and the implied constants in any order expressions also belong to $\mathcal{K}$.

The d-dependence in $\bar{\lambda}(\sigma^2)$ and $d^{-1}J_{\max}$ is put in to ensure that the quantities do not automatically have to grow with the dimension d. It is chosen in this way for the latter in view of Lemma 3.1, and for the former by comparison with $\sigma^2 = I$. In order to avoid many provisos in the bounds, we shall assume throughout that $d \le n^{1/4}$, which is ultimately no restriction, since our bounds are typically of no use unless d is rather smaller than $n^{1/7}$.

We note two immediate consequences of Assumptions G3 and G4.

Lemma 2.1. *Assumptions G3 and G4 imply that σ^2 is positive definite, and that, for any $\delta > 0$, there exists $n_{2.1}(\delta) < \infty$ such that the process X^δ_n is irreducible on $\widetilde{B}_{n,\delta}(c)$, defined in (1.13), as long as $n \ge n_{2.1}(\delta)$.*

Proof. For the first statement, if $x^T\sigma^2 x = 0$, then $x^T J = 0$ for all $J \in \mathcal{J}$, because of Assumption G3. This, from Assumption G4, implies that $x^T e^{(j)} = 0$ for all $1 \le j \le d$, so that $x = 0$.

For the second statement, setting $r_{\max} := \max_{1\le j\le d} r(j)$, it is immediate that, under the transitions for the Markov process X_n^δ, the states X and $X \pm e^{(j)}$ communicate, for all $1 \le j \le d$, as long as $|X - nc|_\Sigma < n\delta - r_{\max} J^\Sigma_{\max}$. Hence, starting from an X with $|X - nc|_\Sigma \le \max_{1\le j\le d} |e^{(j)}|_\Sigma$, it follows that all states X with $|X - nc|_\Sigma < n\delta - r_{\max} J^\Sigma_{\max}$ intercommunicate.

For the remainder, we note that, because the set $\mathcal{J}$ is finite, the infimum $\inf_{u\in\mathbb{R}^d:\, |u|_\Sigma=1} \min_{J\in\mathcal{J}} u^T\Sigma^{-1}J$ is attained at some u_*. Then $\min_{J\in\mathcal{J}} u_*^T\Sigma^{-1}J \ge 0$ together with $F(c) = 0$ would imply that $u_*^T\Sigma^{-1}J = 0$ for all $J \in \mathcal{J}$; and this is impossible, as argued above. Hence there exists $k_* > 0$ such that, for all u with $|u|_\Sigma = 1$, $\min_{J\in\mathcal{J}} u^T\Sigma^{-1}J < -k_*$; without loss of generality, we can also take $k_* \le 1$.

Taking any X with $|X - nc|_\Sigma \le n\delta$, write $X - nc = xu$, for $u \in \mathbb{R}^d$ with $|u|_\Sigma = 1$ and $x \ge 0$. Then, noting that $\sqrt{1-y} \le 1 - y/2$ in $0 \le y \le 1$, we have

$$\begin{aligned}\min_{J\in\mathcal{J}} |X + J - nc|_\Sigma &= \min_{J\in\mathcal{J}} \Big\{ |X - nc|^2_\Sigma + 2(X - nc)^T\Sigma^{-1}J + |J|^2_\Sigma \Big\}^{1/2} \\ &\le x\Big\{1 - 2x^{-1}k_* + x^{-2}\{J^\Sigma_{\max}\}^2\Big\}^{1/2} \\ &\le x - k_*/2,\end{aligned}$$

provided that $x \ge \max\{k_*, \{J^\Sigma_{\max}\}^2/k_*\}$. Thus each state with $|X - nc|_\Sigma \le n\delta$ communicates with some state X' for which $|X' - nc|_\Sigma \le |X - nc|_\Sigma - k_*/2$, and hence, repeating this step, with one such that $|X - nc|_\Sigma < n\delta - r_{\max}J^\Sigma_{\max}$. Combining these results, we see that X_n^δ is irreducible, provided that

$$n \;\ge\; n_{2.1}(\delta) \;:=\; \delta^{-1}\big\{(r_{\max}+1)J^\Sigma_{\max} + \max\{k_*, \{J^\Sigma_{\max}\}^2/k_*\}\big\}. \qquad \square$$

If Assumption G4 is *not* satisfied, then the lattice generated by the jumps in $\mathcal{J}$ is a proper sub-lattice of $\mathbb{Z}^d$.

2.2 X_n^δ stays close to nc

In this section, we show that, whatever its initial value $X_n^\delta(0)$, the process X_n^δ rapidly gets close to nc. Thereafter, it remains close to nc with high probability for a very long time. To formulate our results, we define the hitting times

$$\begin{aligned}\tau_n^\delta(\eta) &:= \inf\{u \ge 0\colon\ |X_n^\delta(u) - nc|_\Sigma \ge n\eta\};\\ \tilde\tau_n^\delta(\eta) &:= \inf\{u \ge 0\colon\ |X_n^\delta(u) - nc|_\Sigma \le n\eta\},\end{aligned} \tag{2.6}$$

for any $0 < \eta \leq \delta \leq \delta_0$.

We begin by establishing some Lyapunov–Foster–Tweedie drift conditions, showing that X_n^δ has a strong tendency to drift towards nc in the $|\cdot|_\Sigma$ norm.

Lemma 2.2. *Let X_n be a sequence of Markov jump processes, whose transition rates are given in (1.1), and such that Assumptions G0–G4 are satisfied. Define*

$$\begin{aligned} h_0(X) &:= (X-nc)^T\Sigma^{-1}(X-nc) \;=\; |X-nc|_\Sigma^2; \\ h_\theta(X) &:= \exp\{n^{-1}\theta h_0(X)\},\;\; \theta > 0. \end{aligned}$$

Then there exist positive constants $K_{2.2}, \delta_{2.2}$ and θ_1 in $\mathcal{K}$ and $\delta'_{2.2}(d) \in \mathcal{K}(d)$ such that, for any $\delta \leq \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$ and any $X \in \widetilde{B}_{n,\delta}(c)$ with $|X-nc|_\Sigma \geq K_{2.2}\sqrt{nd}$, we have

$$\mathcal{A}_n^\delta h_0(X) \;\leq\; -\alpha_1 h_0(X); \quad \mathcal{A}_n^\delta h_\theta(X) \;\leq\; -\tfrac{1}{2}n^{-1}\alpha_1\theta h_0(X)h_\theta(X),\; 0 < \theta \leq \theta_1;$$

for the latter inequality, we also require that $n \geq n_{2.2} \in \mathcal{K}$. The quantities $K_{2.2}, \delta_{2.2}$, $\delta'_{2.2}(d)$ and θ_1 are given in (2.12), (2.14) and (2.19).

Proof. It is immediate that, for the above choice of h_0,

$$h_0(X+J) - h_0(X) \;=\; J^T\Sigma^{-1}(X-nc) + (X-nc)^T\Sigma^{-1}J + J^T\Sigma^{-1}J.$$

Multiplying by $ng_\delta^J(x)$, where $x := n^{-1}X$, and adding over J, we have

$$\mathcal{A}_n^\delta h_0(X) \;=\; n\{F(x)^T\Sigma^{-1}(X-nc) + (X-nc)^T\Sigma^{-1}F(x) + \mathrm{Tr}\,(\Sigma^{-1}\sigma^2(x))\}, \tag{2.7}$$

as long as $|X-nc|_\Sigma < n\delta - J^\Sigma_{\max}$, where F is as defined in (1.2), and σ^2 as in (1.4). For $|X-nc|_\Sigma \geq n\delta - J^\Sigma_{\max}$, the truncation (1.12) may change this expression: see below. Now, using (2.3), for $x, y \in B_{\delta_0}(c)$, we have

$$\begin{aligned} |F(x) - F(y) - A(x-y)| &\leq \frac{1}{2}\sum_{J\in\mathcal{J}} |J|g^J(c)L_2|x-y|\{|x-y| + 2|y-c|\} \\ &\leq \Lambda L_2|x-y|\{|x-y| + |y-c|\}. \end{aligned} \tag{2.8}$$

Substituting (2.8), with $y = c$, into (2.7), and using (1.9), we have

$$\begin{aligned} \mathcal{A}_n^\delta h_0(X) &\leq -(X-nc)^T\Sigma^{-1}\sigma^2\Sigma^{-1}(X-nc) + n\mathrm{Tr}\,(\Sigma^{-1}\sigma^2(x)) \\ &\quad + 2\Lambda L_2 n^{-1}\|\Sigma^{-1/2}\|\,|X-nc|^2\,|X-nc|_\Sigma. \end{aligned} \tag{2.9}$$

Using the inequalities

$$
\begin{aligned}
(X-nc)^T\Sigma^{-1}\sigma^2\Sigma^{-1}(X-nc) &\ \geq\ \lambda_{\min}(\sigma^2_\Sigma)\,|X-nc|_\Sigma; \\
\lambda_{\min}(\Sigma)\,|X-nc|^2_\Sigma \ \leq\ |X-nc|^2 &\ \leq\ \lambda_{\max}(\Sigma)\,|X-nc|^2_\Sigma, \qquad (2.10)
\end{aligned}
$$

it first follows that $(X-nc)^T\Sigma^{-1}\sigma^2\Sigma^{-1}(X-nc) \geq 2\alpha_1|X-nc|^2_\Sigma$. Then

$$
n\mathrm{Tr}\,(\Sigma^{-1}\sigma^2(x)) \ \leq\ nL_0\mathrm{Tr}\,(\sigma^2_\Sigma) \ \leq\ \tfrac{1}{2}\alpha_1|X-nc|^2_\Sigma \qquad (2.11)
$$

if $|X-nc|_\Sigma \geq K_{2.2}\sqrt{nd}$, where

$$
K^2_{2.2} \ :=\ \frac{2L_0}{d\alpha_1}\,\mathrm{Tr}\,(\sigma^2_\Sigma) \ \leq\ 4L_0\rho(\sigma^2)\rho(\Sigma), \qquad (2.12)
$$

since $(1/2d\alpha_1)\mathrm{Tr}\,(\sigma^2_\Sigma) \leq \rho(\sigma^2_\Sigma) \leq \rho(\sigma^2)\rho(\Sigma)$. Finally,

$$
2\Lambda L_2 n^{-1}\|\Sigma^{-1/2}\|\,|X-nc|^2\,|X-nc|_\Sigma \ \leq\ \tfrac{1}{2}\alpha_1|X-nc|^2_\Sigma \qquad (2.13)
$$

if $|X-nc|_\Sigma \leq n\,\min\{\delta_{2.2},\delta'_{2.2}(d)\}$, where

$$
\delta_{2.2} \ :=\ \frac{\delta_0}{\sqrt{\lambda_{\max}(\Sigma)}}; \quad \delta'_{2.2}(d) \ :=\ \frac{1}{d}\,\frac{\alpha_1\sqrt{\lambda_{\min}(\Sigma)}}{4\Lambda L_2\lambda_{\max}(\Sigma)}. \qquad (2.14)
$$

This proves the first part of the lemma for all X such that $|X-nc|_\Sigma < n\delta - J^\Sigma_{\max}$.

If $n\delta - J^\Sigma_{\max} \leq |X-nc|_\Sigma \leq n\delta$, we may have $g^J(n^{-1}X) > g^J_\delta(n^{-1}X) = 0$ for some J. However, from the definition of h_0, these J represent transitions for which $h_0(X+J)-h_0(X) > 0$, and replacing $g^J(n^{-1}X)$ by zero makes the value of $\mathcal{A}^\delta_n h_0(X)$ even smaller than that given in (2.7), and hence preserves the inequality (2.9).

For the second part, taking $\delta \leq \delta_{2.2}$, we note that $e^x - 1 \leq x + x^2$ in $x \leq 1$. Now, for $J^\Sigma_{\max} \leq n\delta_{2.2}$ and $|X-nc|_\Sigma \leq n\delta_{2.2}$, we have

$$
\frac{\theta}{n}|h_0(X+J)-h_0(X)| \ \leq\ \frac{\theta}{n}\{2J^\Sigma_{\max}\,|X-nc|_\Sigma + (J^\Sigma_{\max})^2\} \ \leq\ 3\theta J^\Sigma_{\max}\delta_{2.2},
$$

and $J^\Sigma_{\max} \leq n\delta_{2.2}$ if $n \geq (d^{-1}J^\Sigma_{\max}/\delta_{2.2})^{4/3} =: n_{2.2}$, because $n \geq d^4$. Hence it follows that $n^{-1}\theta|h_0(X+J)-h_0(X)| \leq 1$ for all $X \in \widetilde{B}_{n,\delta}(c)$, if $\theta \leq \theta_1$, $n \geq n_{2.2}$ and

$$
\theta_1 J^\Sigma_{\max}\delta_{2.2} \ \leq\ 1/3; \qquad (2.15)
$$

note that then $d\theta_1 \in \mathcal{K}$. Then, for X such that $|X-nc|_\Sigma < n\delta - J^\Sigma_{\max}$, and with $x := n^{-1}X$,

$$
\mathcal{A}^\delta_n h_\theta(X) \ =\ nh_\theta(X)\sum_{J\in\mathcal{J}} g^J(x)\big\{e^{n^{-1}\theta(h_0(X+J)-h_0(X))} - 1\big\}.
$$

Hence, if $|X - nc|_\Sigma < n\delta - J^\Sigma_{\max}$, we have

$$\begin{aligned} n \sum_{J \in \mathcal{J}} g^J(x) \big\{ e^{n^{-1}\theta(h_0(X+J) - h_0(X))} - 1 \big\} & \\ \le\ n^{-1}\theta \mathcal{A}^\delta_n h_0(X) + n \sum_{J \in \mathcal{J}} g^J(x) n^{-2} \theta^2 |h_0(X+J) - h_0(X)|^2. & \end{aligned}$$

Since

$$\begin{aligned} |h_0(X+J) - h_0(X)|^2 &\le \{2|X - nc|_\Sigma |J|_\Sigma + |J|^2_\Sigma\}^2 \\ &\le |J|^2_\Sigma (8|X - nc|^2_\Sigma + 2(J^\Sigma_{\max})^2), \end{aligned}$$

it follows in turn that, if $\delta \le \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, then

$$\begin{aligned} n \sum_{J \in \mathcal{J}} g^J(x) \big\{ e^{n^{-1}\theta(h_0(X+J) - h_0(X))} - 1 \big\} & \\ \le\ -n^{-1}\theta\alpha_1 h_0(X) + 2n^{-1}\theta^2 L_0 \mathrm{Tr}\,(\sigma^2_\Sigma)\{4h_0(X) + (J^\Sigma_{\max})^2\}, & \end{aligned}$$

if $\theta \le \theta_1$. But now, if θ_1 is also chosen so that

$$8d\theta_1 L_0 \lambda_{\max}(\sigma^2_\Sigma) \ \le\ \tfrac{1}{4}\alpha_1 \ =\ \tfrac{1}{8}\lambda_{\min}(\sigma^2_\Sigma), \tag{2.16}$$

we have $8\theta^2 L_0 \mathrm{Tr}\,(\sigma^2_\Sigma) h_0(X) \le \frac{1}{4}\theta\alpha_1 h_0(X)$, and if

$$2d\theta_1 L_0 \lambda_{\max}(\sigma^2_\Sigma)\,(J^\Sigma_{\max})^2 \ \le\ \tfrac{1}{4}\alpha_1\, dK^2_{2.2}, \tag{2.17}$$

and $|X - nc|_\Sigma \ge K_{2.2}\sqrt{nd}$, we have $2\theta^2 L_0 \mathrm{Tr}\,(\sigma^2_\Sigma)(J^\Sigma_{\max})^2 \le \frac{1}{4}\theta\alpha_1 h_0(X)$ also, so that then

$$n \sum_{J \in \mathcal{J}} g^J(x) \big\{ e^{n^{-1}\theta(h_0(X+J) - h_0(X))} - 1 \big\} \ \le\ -\tfrac{1}{2} n^{-1}\alpha_1 \theta h_0(X). \tag{2.18}$$

Note that (2.15), (2.16) and (2.17) are satisfied by choosing

$$d\theta_1 \ =\ \min\{1/(3d^{-1}J^\Sigma_{\max}\delta_{2.2}), 1/(64L_0\rho(\sigma^2)\rho(\Sigma)), 1/4(d^{-1}J^\Sigma_{\max})^2\} \ \in\ \mathcal{K}, \tag{2.19}$$

since we assume that $n \ge d^4$. As for the first part, if $n\delta - J^\Sigma_{\max} \le |X - nc|_\Sigma \le n\delta$, the inequality (2.18) is still true, completing the proof of the second statement of the lemma. □

Remark 2.3. If the functions g^J are *linear* within $B_{\delta_0,\Sigma}$, then $L_2 = 0$, and we can take $\min\{\delta_{2.2}, \delta'_{2.2}(d)\} = \delta_{2.2} = \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$.

The first of the drift inequalities in Lemma 2.2 is now used to show that X^δ_n quickly reaches even small balls around nc, if $\delta \le \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$.

Lemma 2.4. *Let X_n be a sequence of Markov jump processes, whose transition rates are given in (1.1), and such that Assumptions G0–G4 are satisfied. Let α_1 be as in (2.4) and $K_{2.2}$, $\delta_{2.2}$ and $\delta'_{2.2}(d)$ as in Lemma 2.2. Then, if $\delta \le \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$ and $\eta > \max\{K_{2.2}\sqrt{d/n}, 2n^{-1}J^{\Sigma}_{\max}\}$, we have*

$$\mathbb{P}[\tilde{\tau}^{\delta}_n(\eta) > t \,|\, X^{\delta}_n(0) = X_0] \;\le\; 4(n\eta)^{-2}|X_0 - nc|^2_{\Sigma}\, e^{-\alpha_1 t}.$$

Proof. As before, let $h_0(X) := |X - nc|^2_{\Sigma}$, and define $M_0(t) := h_0(X^{\delta}_n(t))e^{\alpha_1 t}$. Then it follows from the first part of Lemma 2.2, by a standard argument, that $M_0(t \wedge \tilde{\tau}^{\delta}_n(K_{2.2}\sqrt{d/n}))$, $t \ge 0$, is a non-negative supermartingale with respect to the filtration $\mathcal{F}^{X^{\delta}_n} :=(\mathcal{F}^{X^{\delta}_n}_t,\, t \ge 0)$ generated by X^{δ}_n. This implies that

$$\begin{aligned}(n\eta - J^{\Sigma}_{\max})^2 \mathbb{E}\big\{e^{\alpha_1 \tilde{\tau}^{\delta}_n(\eta)} \mathbb{1}\{\tilde{\tau}^{\delta}_n(\eta) \le t\} \,|\, X^{\delta}_n(0) = X_0\big\} \\ \le\; \mathbb{E}\{M_0(t \wedge \tilde{\tau}^{\delta}_n(\eta)) \,|\, X^{\delta}_n(0) = X_0\} \;\le\; h_0(X_0),\end{aligned}$$

since $h_0(X^{\delta}_n(\tilde{\tau}^{\delta}_n(\eta))) \ge (n\eta - J^{\Sigma}_{\max})^2$, because the jumps of X^{δ}_n are bounded in Σ-norm by $J^{\Sigma}_{\max}$. Letting $t \to \infty$, we have

$$\mathbb{E}\big\{e^{\alpha_1 \tilde{\tau}^{\delta}_n(\eta)} \,|\, X^{\delta}_n(0) = X_0\big\} \;\le\; \left\{\frac{|X_0 - nc|_{\Sigma}}{n\eta - J^{\Sigma}_{\max}}\right\}^2.$$

The lemma now follows immediately. □

The second drift inequality in Lemma 2.2 implies that the process X^{δ}_n takes a long time to get far away from neighbourhoods of nc. For use in what follows, we define

$$\psi(n) \;:=\; 4\sqrt{\frac{\log n}{(d\theta_1)n^{3/4}}} \quad \text{and} \quad \psi^{-1}(\eta) \;:=\; \min\{n \ge 4\colon\, \psi(n) \le \eta\}. \qquad (2.20)$$

Lemma 2.5. *Let X_n be a sequence of Markov jump processes, whose transition rates are given in (1.1), and such that Assumptions G0–G4 are satisfied. Then there exists $K_{2.5} \in \mathcal{K}$ such that, for all $\eta \le \delta \le \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$ and for θ_1 as in Lemma 2.2, we have*

$$\mathbb{P}[\tau^{\delta}_n(\eta) \le t \,|\, X^{\delta}_n(0) = X_0] \;\le\; (nK_{2.5}\Lambda t + \exp\{n^{-1}\theta_1|X_0 - nc|^2_{\Sigma}\})e^{-n\theta_1\eta^2},$$

if $n \ge n_{2.2}$. In particular, for any $\delta \le \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, for any $\eta \le \delta$, and for any $T > 0$, there exists $n_{2.5}(T) \in \mathcal{K}(\overline{\Lambda}T)$ such that, for all $|X_0 - nc|_{\Sigma} \le n\eta/2$ and $t \le T$, we have

$$\mathbb{P}[\tau^{\delta}_n(3\eta/4) \le t \,|\, X^{\delta}_n(0) = X_0] \;\le\; 2n^{-4},$$

as long as $n \ge \max\{n_{2.5}(T), \psi^{-1}(\eta)\}$. The quantities $K_{2.5}$ and $n_{2.5}(T)$ are defined in (2.22) and (2.23), respectively.

Proof. It follows from the second part of Lemma 2.2 that, for $0 \leq \theta \leq \theta_1$,

$$M_\theta(t) \;:=\; h_\theta(X_n^\delta(t)) - H_\theta \int_0^t \mathbb{1}\{|X_n^\delta(s) - nc|_\Sigma \leq K_{2.2}\sqrt{nd}\}\, ds$$

is an $\mathcal{F}^{X_n^\delta}$-supermartingale, where

$$H_\theta \;:=\; \max_{X \in \mathbb{Z}^d \colon |X - nc|_\Sigma \leq K_{2.2}\sqrt{nd}} \mathcal{A}_n^\delta h_\theta(X).$$

Clearly, recalling $n \geq d^4$, H_θ is bounded by

$$n \sum_{J \in \mathcal{J}} \|g^J\|_{\delta_0} \exp\{n^{-1}\theta[K_{2.2}\sqrt{nd} + J^\Sigma_{\max}]^2\} \;\leq\; n\Lambda K_{2.5}, \tag{2.21}$$

for

$$K_{2.5} \;:=\; L_0 \exp\{\theta_1[K_{2.2} + d^{-1}J^\Sigma_{\max}]^2\} \in \mathcal{K}. \tag{2.22}$$

By the optional stopping theorem, applied to $M_\theta(\min\{t, \tau_n^\delta(\eta)\})$, it thus follows that

$$e^{n\theta\eta^2}\mathbb{P}[\tau_n^\delta(\eta) \leq t \,|\, X_n^\delta(0) = X_0] - n\Lambda K_{2.5}t \;\leq\; \exp\{n^{-1}\theta|X_0 - nc|_\Sigma^2\},$$

proving the first claim. The second follows for $n \geq \max\{n_{2.5}(T), \psi^{-1}(\eta)\}$, where

$$n_{2.5}(T) \;:=\; \max\{K_{2.5}\overline{\Lambda}T, n_{2.2}\}, \tag{2.23}$$

since, for such choices of n,

$$nK_{2.5}\Lambda T \;\leq\; n^{9/4} \;\leq\; n^4 \;\leq\; e^{n\theta_1\eta^2/4}, \quad \text{and thus} \quad e^{-5n\theta_1\eta^2/16} \;\leq\; n^{-4}. \qquad \square$$

3 The analysis of X_n^δ: elementary processes

In this section, we conduct a more detailed analysis of the Markov jump processes X_n^δ. The results that follow are used to bound the solution to the Stein equation (1.15) and its differences, using the representation given in (1.17); this is an essential step in proving our approximation theorem. In order to find Markov jump processes that yield a given pair A, σ^2, we only need to consider ones whose transition rates satisfy more restrictive conditions than Assumptions G0–G4; we refer to them as *elementary* (sequences of) processes. Since this simplifies some of the coming arguments, we conduct them within the context of elementary processes, though analogous results hold under the previous assumptions: see Remark 6.4. We retain Assumptions G0 and G1, replacing the remainder with the Assumptions S2–S4 below.

Assumption S2. The set $\mathcal{J}$ contains the vectors $\{\pm e^{(j)},\, 1 \le j \le d\}$.

Assumption S3. The transition rates $g^J(x)$ are constant in $B_{\delta_0}(c)$, for all $J \in \mathcal{J} \setminus \{e^{(j)},\, 1 \le j \le d\}$.

Assumption S4. For $1 \le j \le d$, $g^{e^{(j)}}(x)$ is linear and satisfies $g^{e^{(j)}}(x) \ge \frac{1}{2} g^{e^{(j)}}(c)$ in $x \in B_{\delta_0}(c)$.

Defining $I^{(j)} := \{i \colon 1 \le i \le d, A_{ij} \ne 0\}$, $1 \le j \le d$, we write

$$g^{(j)} \;:=\; g^{-e^{(j)}}(c), \quad G^{(j)} \;:=\; \sum_{i \in I^{(j)}} g^{e^{(i)}}(c), \quad g_* \;:=\; \min_{1 \le j \le d} (g^{(j)} \wedge G^{(j)}), \tag{3.1}$$

observing that $G^{(j)} \le \Lambda$, $1 \le j \le d$. We retain the definitions (2.3), noting that, for elementary processes, $L_2 = 0$ and that $L_0 \le 3/2$, and that ε_0 as defined in Assumption G3 can be taken to be $1/2$. As observed in Remark 2.3, since $L_2 = 0$, we have

$$\min\{\delta_{2.2}, \delta'_{2.2}(d)\} \;=\; \delta_{2.2} \;=\; \delta_0 / \sqrt{\lambda_{\max}(\Sigma)}$$

for the upper bound on δ in Lemma 2.2. We also define

$$n_{(3.2)} \;:=\; \max\Bigl\{(5(d^{-1} J^{\Sigma}_{\max}) \max\{1, \sqrt{d\theta_1}\})^{8/3}, n_{2.5}(1/g_*)\Bigr\} \;\in\; \mathcal{K}. \tag{3.2}$$

After some work, it follows from the definitions of ψ and $n_{(3.2)}$, and because $d^4 \le n$, that $n \ge \max\{n_{(3.2)}, \psi^{-1}(\delta)\}$ implies that

$$\delta \;\ge\; 20 n^{-3/4} (d^{-1} J^{\Sigma}_{\max}) \;\ge\; 20 J^{\Sigma}_{\max}/n; \tag{3.3}$$

these inequalities are used later.

3.1 Any c, A and σ^2 can be associated with an elementary process

In this section, we relate the generator $\widetilde{\mathcal{A}}_n$, defined using an arbitrary choice of c, A and σ^2, to the generator $\mathcal{A}^{\delta}_n$ of an elementary process. The main difficulty is to match σ^2, overcome by using Tropp (2015, Theorem 1.1).

Lemma 3.1. *Let σ^2 be any $d \times d$ covariance matrix with positive eigenvalues $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_d > 0$. Then σ^2 can be represented in the form*

$$\sigma^2 \;=\; \sum_{J \in \mathcal{J}} \tilde{g}(J) J J^T,$$

for a finite set $\mathcal{J} \in \mathbb{Z}^d$ such that $e^{(i)} \in \mathcal{J}$, $1 \le i \le d$, such that $J \in \mathcal{J}$ implies that $-J \in \mathcal{J}$, with $\tilde{g}(-J) = \tilde{g}(J)$, and such that

$$\max_{J \in \mathcal{J}} \max_{1 \le i \le d} |J_i| \;\le\; 1 + \tfrac{1}{2}\sqrt{2(d-1)\rho(\sigma^2)}.$$

Furthermore, $\tilde{g}(e^{(i)}) \ge \frac{1}{4}\lambda_d$ for each $1 \le i \le d$.

Proof. Write $\lambda_0 := \frac{1}{2}\lambda_d = \frac{1}{2}\lambda_{\min}(\sigma^2)$, so that $\sigma^2 - \lambda_0 I$ is positive definite, and has condition number $\rho(\sigma^2 - \lambda_0 I) \le 2\rho(\sigma^2)$. By Theorem 1.1 of Tropp (2015), we can write

$$\sigma^2 - \lambda_0 I \;=\; \sum_{J \in \mathcal{J}_1} \gamma(J) J J^T,$$

where the set $\mathcal{J}_1$ is finite, $\gamma(J) > 0$ for each $J \in \mathcal{J}_1$, and the vectors J have integer coordinates with $|J_i| \le 1 + \frac{1}{2}\sqrt{(d-1)\rho(\sigma^2 - \lambda_0 I)}$. Note that the same covariance matrix is obtained if $\gamma(J)JJ^T$ is replaced by $\frac{1}{2}\gamma(J)\{JJ^T + (-J)(-J)^T\}$, which we do, expanding the set $\mathcal{J}_1$ if necessary. Writing $\lambda_0 I = \sum_{i=1}^d \frac{1}{2}\lambda_0\{e^{(i)}(e^{(i)})^T + (-e^{(i)})(-e^{(i)})^T\}$, and taking $\mathcal{J} = \mathcal{J}_1 \cup \{\pm e^{(i)},\, 1 \le i \le d\}$, the lemma follows. □

Fitting A and c as well, in such a way that Assumptions G0–G1 and S2–S4 are all satisfied, is now easy.

Theorem 3.2. *For any $c \in \mathbb{R}^d$, A whose eigenvalues all have negative real parts, and positive definite σ^2, there exists a sequence of elementary processes having $F(c) = 0$, $DF(c) = A$ and σ^2 given by (1.4). For these processes, defining $\delta_0 := \lambda_{\min}(\sigma^2)/(8\|A\|)$ and $\overline{\Lambda} := \bar{\lambda}(\sigma^2)$, we have $\varepsilon_0 \ge 1/2$ in Assumption S4, and the quantities in $\mathcal{K}$ are all bounded by continuous functions of $\|A\|/\overline{\Lambda}$ and the elements of $\mathrm{Sp}'(\sigma^2/\overline{\Lambda})$ and $\mathrm{Sp}'(\Sigma)$.*

Proof. Represent σ^2 as in Lemma 3.1. For $J \in \mathcal{J}$, define

$$g^J(x) \;:=\; \begin{cases} \tilde{g}(J), & \text{if } J \in \mathcal{J} \setminus \{e^{(i)},\, 1 \le i \le d\}; \\ \tilde{g}(J) + (A(x-c))_i, & \text{for } J = e^{(i)},\, 1 \le i \le d. \end{cases}$$

With these functions g^J, we have $\sigma^2 \;=\; \sum_{J \in \mathcal{J}} g^J(c) J J^T$ and, writing $F(x) := \sum_{J \in \mathcal{J}} J g^J(x)$, we also have $F(c) = 0$ and $DF(c) = A$; define $\bar{\gamma}(\sigma^2) := d^{-3/2} \sum_{J \in \mathcal{J}} g^J(c)|J|^3$.

Now all the transition rates $g^J(x)$ are constant in x, except for $J = e^{(i)}$, $1 \le i \le d$, when they are linear. For $g^{e^{(i)}}$, we have

$$\frac{g^{e^{(i)}}(x)}{g^{e^{(i)}}(c)} \;=\; \frac{\tilde{g}(e^{(i)}) + (A(x-c))_i}{\tilde{g}(e^{(i)})},$$

and this is at least 1/2 if

$$|x-c|\,\|A\| \;\le\; \tfrac{1}{8}\lambda_{\min}(\sigma^2) \;\le\; \tfrac{1}{2}\tilde g(e^{(i)}),$$

which is in turn true if $|x-c| \le \delta_0$, so that we can take $\varepsilon_0 = 1/2$. The same calculation shows that $L_0 \le 3/2$, and it is also immediate, from Lemma 3.1, that

$$\begin{aligned} L_1 &\le 2\|A\|/\min_{1\le i\le d}\tilde g(e^{(i)}) \;\le\; 4\|A\|/\lambda_{\min}(\sigma^2); \\ \overline{\Lambda}/g_* &\le \bar\lambda(\sigma^2)/\min_{1\le i\le d}\tilde g(e^{(i)}) \;\le\; 4\rho(\sigma^2/\bar\lambda(\sigma^2)); \quad (3.4) \\ d^{-1/2}\bar\gamma(\sigma^2)/\overline{\Lambda} &\le \sqrt{1+\rho(\sigma^2/\bar\lambda(\sigma^2))}. \quad (3.5) \end{aligned}$$

Finally, again from Lemma 3.1,

$$d^{-1}J_{\max} \;\le\; d^{-1}\bigl\{d\bigl(1+\tfrac{1}{2}\sqrt{2(d-1)\rho(\sigma^2)}\bigr)^2\bigr\}^{1/2} \;\le\; 1+\sqrt{\tfrac{1}{2}\rho(\sigma^2/\bar\lambda(\sigma^2))}. \tag{3.6}$$

Hence, for this choice of δ_0, the quantities in $\mathcal{K}$ are all bounded by continuous functions of $\|A\|/\overline{\Lambda}$ and the elements of $\mathrm{Sp}'(\sigma^2/\overline{\Lambda})$ and $\mathrm{Sp}'(\Sigma)$. □

3.2 The dependence of $\mathcal{L}(X_n^\delta(U))$ on $X_n^\delta(0)$

We first show that the distribution $\mathcal{L}(X_n^\delta(U)\,|\,X_n^\delta(0)=X)$ does not change too much if the initial condition is slightly altered. The argument is based on that for one-dimensional processes given in Socoll & Barbour (2010). We begin by bounding differences of the form

$$\mathbb{E}\{f(X_n^\delta(U))\,|\,X_n^\delta(0)=X-e^{(j)}\} - \mathbb{E}\{f(X_n^\delta(U))\,|\,X_n^\delta(0)=X\},$$

and then prove a sharper bound on second differences.

Theorem 3.3. *Let X_n be a sequence of elementary processes. Fix any $\delta < \delta_{2.2}$. Then there are constants $K_{3.3}^j$, $1\le j\le d$, in $\mathcal{K}$, such that, for all $n \ge \max\{n_{(3.2)}, \psi^{-1}(\delta)\}$ as in (3.2),*

$$\begin{aligned} \sup_{f:\,\|f\|_\infty=1} &|\mathbb{E}\{f(X_n^\delta(U))\,|\,X_n^\delta(0)=X-e^{(j)}\} - \mathbb{E}\{f(X_n^\delta(U))\,|\,X_n^\delta(0)=X\}| \\ &\le K_{3.3}^j n^{-1/2}\left(\frac{G^{(j)}}{g^{(j)}}\right)^{1/4}\max\left\{1, \frac{1}{(g^{(j)}G^{(j)})^{1/4}\sqrt{U}}\right\}, \quad (3.7) \end{aligned}$$

uniformly for all $U>0$ and $|X-nc|_\Sigma \le n\delta/2$.

Proof. For any $x \in \ell_1$ and any stochastic matrix P, we have $|x^T P|_1 \leq |x|_1$. Hence the quantity being bounded in (3.7) is non-increasing in U. We can thus take $U \leq U^{(j)} := 1/\sqrt{G^{(j)} g^{(j)}}$ in what follows, and use the bound obtained for $U = U^{(j)}$ as a bound for all larger values of U. Note that $U^{(j)} \leq 1/g_*$.

We begin by realizing the chain X_n^δ with $X_n^\delta(0) = X_0$ in the form $X_n^\delta(u) := X_0 - e^{(j)} N_n^\delta(u) + W_n^\delta(u)$, where the bivariate chain (N_n^δ, W_n^δ) with state space $\mathbb{Z}_+ \times \mathbb{Z}^d$ starts at $(0,0)$, and, at times u such that $|X_n^\delta(u) - nc|_\Sigma \leq n\delta - J_{\max}^\Sigma$, has transition rates given by

$$
\begin{array}{lll}
(l,W) \to (l+1,W) & \text{at rate } ng^{(j)}; & \\
(l,W) \to (l, W+J) & \text{at rate } ng^J\big((X_0 - le^{(j)} + W)/n\big), & J \neq -e^{(j)} \in \mathcal{J};
\end{array}
\tag{3.8}
$$

note that the first of these transitions *reduces* the j-coordinate of X_n^δ by 1. At other values of X, it may be that $g_\delta^J(n^{-1}X)$ does not agree with $g^J(n^{-1}X)$, and so the transition rates of (N_n^δ, W_n^δ) may be different from those given in (3.8). For this reason, if the time interval $[0,U]$ is of interest, we treat any paths of X_n^δ for which $\sup_{0\leq u\leq U} |X_n^\delta(u) - nc|_\Sigma > n\delta - 3J_{\max}^\Sigma$ separately; the factor 3 ensures that shifting a path by a vector $J' + J''$, for any $J', J'' \in \mathcal{J}$, still leaves it entirely within $\{X \colon |X - nc|_\Sigma \leq n\delta - J_{\max}\}$ over $[0,U]$.

Using the bivariate process, we deduce that

$$
\begin{aligned}
&d_{TV}\{\mathcal{L}_{X_0}(X_n^\delta(U)), \mathcal{L}_{X_0 - e^{(j)}}(X_n^\delta(U))\} \\
&\quad= \frac{1}{2} \sum_{X\in\mathbb{Z}^d} |\mathbb{P}_{X_0}[X_n^\delta(U) = X + X_0] - \mathbb{P}_{X_0 - e^{(j)}}[X_n^\delta(U) = X + X_0]| \\
&\quad= \frac{1}{2} \sum_{X\in\mathbb{Z}^d} \Big| \sum_{l\geq 0} \mathbb{P}_{X_0}[N_n^\delta(U) = l] \mathbb{P}_{X_0}[W_n^\delta(U) = X + le^{(j)} \,|\, N_n^\delta(U) = l] \\
&\qquad\qquad - \sum_{l\geq 1} \mathbb{P}_{X_0}[N_n^\delta(U) = l-1] \mathbb{P}_{X_0 - e^{(j)}}[W_n^\delta(U) = X + le^{(j)} \,|\, N_n^\delta(U) = l-1] \Big| \\
&\quad\leq \frac{1}{2} \sum_{X\in\mathbb{Z}^d} \sum_{l\geq 0} |\mathbb{P}_{X_0}[N_n^\delta(U) = l] - \mathbb{P}_{X_0}[N_n^\delta(U) = l-1]| q^U_{l-1, X_0 - e^{(j)}}(X + le^{(j)}) \\
&\qquad\qquad + \frac{1}{2} \sum_{X\in\mathbb{Z}^d} \sum_{l\geq 1} \mathbb{P}_{X_0}[N_n^\delta(U) = l] | q^U_{l,X_0}(X + le^{(j)}) - q^U_{l-1, X_0 - e^{(j)}}(X + le^{(j)})|,
\end{aligned}
\tag{3.9}
$$

where

$$
q^U_{l,X}(W) := \mathbb{P}[W_n^\delta(U) = W \,|\, N_n^\delta(U) = l, X_n^\delta(0) = X]. \tag{3.10}
$$

Now, from Barbour, Holst & Janson (1992, Proposition A.2.7),

$$\sum_{l \geq 0} |\mathrm{Po}\,(\lambda)\{l\} - \mathrm{Po}\,(\lambda)\{l-1\}| \;=\; 2 \max_{l \geq 0} \mathrm{Po}\,(\lambda)\{l\} \;\leq\; \frac{1}{\sqrt{\lambda}}. \qquad (3.11)$$

Hence, since N_n^δ is a Poisson process of rate $ng^{(j)}$ until the time

$$\hat{\tau}_n^\delta \;:=\; \tau_n^\delta(\delta - 3n^{-1}J_{\max}^\Sigma), \qquad (3.12)$$

where $\tau_n^\delta(\eta)$ is as defined in (2.6), it follows that the first term in (3.9) is bounded by

$$\mathbb{P}_{X_0}[\hat{\tau}_n^\delta \leq U] + \tfrac{1}{2}\{ng^{(j)}U\}^{-1/2}. \qquad (3.13)$$

Recall that $n \geq \max\{n_{(3.2)}, \psi^{-1}(\delta)\}$, so that, from (3.3), $\delta - 3n^{-1}J_{\max}^\Sigma > 3\delta/4$. Hence, for any $U \leq U^{(j)} \leq 1/g_*$, we can use Lemma 2.5 and the definition of $\hat{\tau}_n^\delta$ to give

$$\mathbb{P}_X[\hat{\tau}_n^\delta \leq U] \;\leq\; \mathbb{P}_X[\tau_n^\delta(3\delta/4) \leq U^{(j)}] \;\leq\; 2n^{-4}, \qquad (3.14)$$

uniformly in $|X - nc|_\Sigma \leq n\delta/2$. Putting this into (3.13), for $U \leq U^{(j)}$, gives a contribution to $d_{TV}\{\mathcal{L}_{X_0}(X_n^\delta(U)), \mathcal{L}_{X_0-e^{(j)}}(X_n^\delta(U))\}$ from the first part of (3.9) of at most

$$2n^{-4} + \tfrac{1}{2}\{ng^{(j)}U\}^{-1/2}. \qquad (3.15)$$

It thus remains only to control the differences between the conditional probabilities $q_{l,X}^U(W)$ and $q_{l-1,X-e^{(j)}}^U(W)$.

To make the comparison between $q_{l,X}^U(W)$ and $q_{l-1,X-e^{(j)}}^U(W)$ for $l \geq 1$, we first condition on the whole paths of N_n^δ leading to the events $\{N_n^\delta(U) = l\}$ and $\{N_n^\delta(U) = l-1\}$, respectively, chosen to be suitably matched; we write

$$\begin{aligned} q_{l,X}^U(W) \;&=\; \frac{1}{U^l}\int_{[0,U]^l} ds_1 \,\ldots\, ds_{l-1}\, ds^* \\ &\qquad \mathbb{P}_X[W_n^\delta(U) = W \mid (N_n^\delta)^U = \nu_l(\cdot\,; s_1, \ldots, s_{l-1}, s^*)]; \\ q_{l-1,X-e^{(j)}}^U(W) \;&=\; \frac{1}{U^l}\int_{[0,U]^l} ds_1 \ldots ds_{l-1} ds^* \\ &\qquad \mathbb{P}_{X-e^{(j)}}[W_n^\delta(U) = W \mid (N_n^\delta)^U = \nu_{l-1}(\cdot\,; s_1, \ldots, s_{l-1})], \end{aligned} \qquad (3.16)$$

where

$$\nu_r(u; t_1, \ldots, t_r) \;:=\; \sum_{i=1}^r \mathbb{1}_{[0,u]}(t_i), \qquad (3.17)$$

and, for a function Y on $\mathbb{R}_+$, Y^u is used to denote $(Y(s),\, 0 \leq s \leq u)$. Fixing $\mathbf{s}_{l-1} := (s_1, s_2, \ldots, s_{l-1})$, let $\mathbb{P}_{(\mathbf{s}_{l-1}, s_*), X}^U$ denote the distribution of $(W_n^\delta)^U$, conditional on $(N_n^\delta)^U = \nu_l(\cdot\,;\mathbf{s}_{l-1},s^*)$ and $X_n^\delta(0)=X$, and let $\mathbb{P}^U_{\mathbf{s}_{l-1},X}$ denote the distribution conditional on $(N_n^\delta)^U = \nu_{l-1}(\cdot\,;\mathbf{s}_{l-1})$ and $X_n^\delta(0)=X$. Write $\hat R^U_{(\mathbf{s}_{l-1},s_*),j,X}(u,w^u)$ to denote the Radon–Nikodym derivative $d\mathbb{P}^U_{\mathbf{s}_{l-1},X-e^{(j)}}/d\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),X}$ evaluated at the path w^u, for any $0\le u\le U$. Then

$$\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),X}[W_n^\delta(U)=W] \;=\; \int_{\{w^U\,:\,w(U)=W\}} \hat R^U_{(\mathbf{s}_{l-1},s_*),j,X}(U,w^U)\,d\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),X}(w^U),$$

and hence

$$\begin{aligned}&\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),X}[W_n^\delta(U)=W] - \mathbb{P}^U_{\mathbf{s}_{l-1},X-e^{(j)}}[W_n^\delta(U)=W]\\ &\quad= \int 1\!\!1_{\{W\}}(w(U))\{1-\hat R^U_{(\mathbf{s}_{l-1},s_*),j,X}(U,w^U)\}\,d\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),X}(w^U). \qquad (3.18)\end{aligned}$$

Thus

$$\begin{aligned}&\sum_{W\in\mathbb{Z}^d} |q^U_{l,X}(W) - q^U_{l-1,X-e^{(j)}}(W)|\\ &\quad\le \frac{1}{U^l}\int_{[0,U]^l} ds_1\,\ldots\,ds_{l-1}\,ds^*\\ &\qquad \sum_{W\in\mathbb{Z}^d}\mathbb{E}^U_{(\mathbf{s}_{l-1},s_*),X}\Big\{|\hat R^U_{(\mathbf{s}_{l-1},s_*),j,X}(U,(W_n^\delta)^U)-1|1\!\!1_{\{W\}}(W_n^\delta(U))\Big\} \qquad (3.19)\\ &\quad\le \frac{2}{U^l}\int_{[0,U]^l} ds_1\,\ldots\,ds_{l-1}\,ds^*\,\mathbb{E}^U_{(\mathbf{s}_{l-1},s_*),X}\Big\{[1-\hat R^U_{(\mathbf{s}_{l-1},s_*),j,X}(U,(W_n^\delta)^U)]_+\Big\}.\end{aligned}$$

To evaluate the expectation, note that $\hat R^U_{(\mathbf{s}_{l-1},s_*),j,X}(u,(W_n^\delta)^u)$, $u\ge 0$, is a $\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),X}$-martingale with respect to the filtration $\mathcal{F}^{X_n^\delta}$, with expectation 1. Now, if the path w^U has r jumps of vectors $J_1,\ldots,J_r$ at times $t_1<\cdots<t_r$, write

$$x_Y(v) \;:=\; n^{-1}(w(v) - e^{(j)}\nu_{l-1}(v\,;s_1,\ldots,s_{l-1}) + Y), \qquad (3.20)$$

and define

$$\hat g^{J'}(\cdot) \;:=\; g^{J'}(\cdot),\; J'\ne -e^{(j)}; \qquad \hat g^{-e^{(j)}}(\cdot) \;:=\; 0; \qquad \hat g(\cdot) \;:=\; \sum_{J'\in\mathcal{J}} \hat g^{J'}(\cdot). \qquad (3.21)$$

Then, for $u \leq \hat{\tau}_n^\delta$, we have

$$\hat{R}^U_{(\mathbf{s}_{l-1},s_*),j,X}(u, w^u) \tag{3.22}$$

$$= \begin{cases} \exp\left(n\int_0^u \{\hat{g}(x_{X-e^{(j)}}(v)) - \hat{g}(x_{X-e^{(j)}}(v) - e^{(j)}n^{-1})\}\,dv\right) \\ \qquad \prod_{\{k:\,0\leq t_k\leq u\}} \left\{\hat{g}^{J_k}(x_{X-e^{(j)}}(t_k-) - e^{(j)}n^{-1})/\hat{g}^{J_k}(x_{X-e^{(j)}}(t_k-))\right\} \\ \qquad\qquad \text{if } u < s^*; \\ \exp\left(n\int_0^{s^*} \{\hat{g}(x_{X-e^{(j)}}(v)) - \hat{g}(x_{X-e^{(j)}}(v) - e^{(j)}n^{-1})\}\,dv\right) \\ \qquad \prod_{\{k:\,0\leq t_k\leq s^*\}} \left\{\hat{g}^{J_k}(x_{X-e^{(j)}}(t_k-) - e^{(j)}n^{-1})/\hat{g}^{J_k}(x_{X-e^{(j)}}(t_k-))\right\} \\ \qquad\qquad \text{if } u \geq s^*; \end{cases}$$

after the 'extra jump' at s^*, the chains have come together. Note that $\hat{R}^U_{(\mathbf{s}_{l-1},s_*),j,X}(u, w^u)$ is absolutely continuous except for jumps at the times t_k. Then also, from Assumptions S3 and S4,

$$\frac{\hat{g}^{J'}(x - e^{(j)}n^{-1})}{\hat{g}^{J'}(x)} \;=\; 1, \qquad J' \notin \{e^{(i)},\, i \in I^{(j)}\},$$

and

$$\left| \frac{\hat{g}^{e^{(i)}}(x - e^{(j)}n^{-1})}{\hat{g}^{e^{(i)}}(x)} - 1 \right| \;\leq\; \frac{2\|Dg^{e^{(i)}}\|_{\delta_0}}{ng^{e^{(i)}}(c)} \;\leq\; 2L_1/n, \quad i \in I^{(j)}, \tag{3.23}$$

uniformly in $|x - c| \leq \delta_0$. Hence, if we define the stopping time

$$\hat{\varphi}_n \;:=\; \inf\{u \geq 0\colon\; \hat{R}^U_{(\mathbf{s}_{l-1},s_*),j,X_0}(u, (W_n^\delta)^u) \geq 2\}, \tag{3.24}$$

the jumps of the martingale $\hat{R}^U_{(\mathbf{s}_{l-1},s_*),j,X_0}(u, (W_n^\delta)^u)$, stopped at the time $\min(U, \hat{\tau}_n^\delta, \hat{\varphi}_n)$, are of size at most $4L_1/n$. Hence, recalling that $L_0 \leq 3/2$, the stopped martingale has expected quadratic variation up to time u of at most

$$\int_0^u \left(\frac{4L_1}{n}\right)^2 n \sum_{i\in I^{(j)}} \|g^{J'}\|_{\delta_0}\,dv \;\leq\; n^{-1}K_{(3.25)}G^{(j)}u, \tag{3.25}$$

where $K_{(3.25)} := 24L_1^2 \in \mathcal{K}$. This in turn also implies that, for $0 < u \leq U$,

$$\mathbb{E}^U_{(\mathbf{s}_{l-1},s_*),X_0}\{(\hat{R}^U_{(\mathbf{s}_{l-1},s_*),j,X_0}(u\wedge\hat{\tau}_n^\delta\wedge\hat{\varphi}_n, (W_n^\delta)^{u\wedge\hat{\tau}_n^\delta\wedge\hat{\varphi}_n}) - 1)^2\} \;\leq\; n^{-1}K_{(3.25)}G^{(j)}u. \tag{3.26}$$

Clearly, from (3.26) and from Kolmogorov's inequality, once again taking $U = U^{(j)}$,

$$\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),X_0}[\hat{\varphi}_n < \min\{U, \hat{\tau}_n^\delta\}] \;\leq\; n^{-1}K_{(3.25)}G^{(j)}U^{(j)} = n^{-1}K_{(3.25)}(G^{(j)}/g^{(j)})^{1/2}. \tag{3.27}$$

Hence, for this choice of U, from (3.26) and (3.27),

$$\mathbb{E}^U_{(\mathbf{s}_{l-1},s_*),X_0}\Big\{[1-\hat{R}^U_{(\mathbf{s}_{l-1},s_*),j,X_0}(U,(W^\delta_n)^U)]_+\Big\}$$
$$\leq\ \min\big\{1, 2n^{-1/2}K_{(3.25)}(G^{(j)}/g^{(j)})^{1/4}+\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),X_0}[\hat{\tau}^\delta_n<U]\big\}. \quad (3.28)$$

In view of Lemma 2.5, the expectation of the term $\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),X_0}[\hat{\tau}^\delta_n<U]$ is bounded by $2n^{-4}$, uniformly in $|X_0-nc|_\Sigma\leq n\delta/2$, because $n\geq\max\{n_{(3.2)},\psi^{-1}(\delta)\}$. Substituting this into (3.9), and using (3.19), it follows that

$$\sum_{l\geq 1}\mathbb{P}[N^\delta_n(U)=l-1]\sum_{W\in\mathbb{Z}^d}|q^U_{l,X_0}(W+le^{(j)})-q^U_{l-1,X_0-e^{(j)}}(W+le^{(j)})|$$
$$\leq\ 2\{2n^{-1/2}K_{(3.25)}(G^{(j)}/g^{(j)})^{1/4}+2\mathbb{P}_{X_0}[\hat{\tau}^\delta_n<U]\}$$
$$\leq\ 2\{2n^{-1/2}K_{(3.25)}(G^{(j)}/g^{(j)})^{1/4}+4n^{-4}\}, \quad (3.29)$$

uniformly for X_0 such that $|X_0-nc|_\Sigma\leq n\delta/2$, and for $n\geq\max\{n_{(3.2)},\psi^{-1}(\delta)\}$. Thus the contribution to $d_{TV}\{\mathcal{L}_{X_0}(X^\delta_n(U)),\mathcal{L}_{X_0-e^{(j)}}(X^\delta_n(U))\}$ from the second part of (3.9) is at most

$$2K_{(3.25)}(G^{(j)}/g^{(j)})^{1/4}n^{-1/2}+4n^{-4}, \quad (3.30)$$

and this, with (3.15), proves the theorem. □

Remark 3.4. As observed after (3.1), we always have $G^{(j)}\leq d\overline{\Lambda}$; however, if $A=-\lambda I$ and (X_n) is as in Theorem 3.2, $G^{(j)}/g^{(j)}=1$ does not grow with d.

Theorem 3.3 bounds differences of the form

$$\mathbb{E}\{f(X^\delta_n(U)\,|\,X^\delta_n(0)=X_0-e^{(j)}\}-\mathbb{E}\{f(X^\delta_n(U)\,|\,X^\delta_n(0)=X_0\},$$

showing that they are of order $O(n^{-1/2})$ uniformly in $U\geq 0$, for f such that $\|f\|_\infty\leq 1$. We now show that the corresponding second differences are of order $O(n^{-1})$.

Theorem 3.5. *Let X_n be a sequence of elementary processes. Fix any $\delta<\delta_{2.2}$. Then there are constants $(K^{ji}_{3.5},\ 1\leq j,i\leq d)$ in $\mathcal{K}$ such that, for any function f with $\|f\|_\infty\leq 1$,*

$$\big|\mathbb{E}\{f(X^\delta_n(U))\,|\,X^\delta_n(0)=X_0-e^{(j)}-e^{(i)}\}-\mathbb{E}\{f(X^\delta_n(U))\,|\,X^\delta_n(0)=X_0-e^{(j)}\}$$
$$-\,\mathbb{E}\{f(X^\delta_n(U))\,|\,X^\delta_n(0)=X_0-e^{(i)}\}+\mathbb{E}\{f(X^\delta_n(U))\,|\,X^\delta_n(0)=X_0\}\big|$$
$$\leq\ K^{ji}_{3.5}\,n^{-1}\left(\frac{G^+_{ij}}{g^-_{ij}}\right)^{1/2}\max\left\{1,\frac{1}{U\sqrt{G^+_{ij}g^+_{ij}}}\right\}, \quad (3.31)$$

uniformly for all $U > 0$, *for* $|X_0 - nc|_\Sigma \le n\delta/4$, *and for* $n \ge \max\{n_{(3.2)}, \psi^{-1}(\delta)\}$, *where*

$$g^+_{ij} \ := \ \max\{g^{(i)}, g^{(j)}\}; \quad g^-_{ij} \ := \ \min\{g^{(i)}, g^{(j)}\}; \quad G^+_{ij} \ := \ \max\{g^+_{ij}, (G^{(i)} + G^{(j)})\}.$$

Proof. As in the previous theorem, the supremum over f of the quantity being bounded in (3.31) is non-increasing in U, so that we can argue for $U \le U^{(i,j)} := (G^+_{ij} g^+_{ij})^{-1/2} \le 1/g^+_{ij}$, and then use the bound for $U = U^{(i,j)}$ for all larger values of U. We give the detailed argument for j and i distinct; it is almost identical if they are the same.

Much as for (3.9), we split off Poisson processes of $-e^{(j)}$ and $-e^{(i)}$ jumps. We write $X^\delta_n(u) := X_0 - e^{(j)} N^\delta_n(u) - e^{(i)} (N')^\delta_n(u) + W^\delta_n(u)$, where the trivariate chain $(N^\delta_n, (N')^\delta_n, W^\delta_n)$ with state space $\mathbb{Z}^2_+ \times \mathbb{Z}^d$ has transition rates

$$\begin{array}{lll} (l, l', W) \to (l+1, l', W) & \text{at rate} & ng^{(j)}; \\ (l, l', W) \to (l, l'+1, W) & \text{at rate} & ng^{(i)}; \\ (l, l', W) \to (l, l', W+J) & \text{at rate} & ng^J\big((X_0 - le^{(j)} - l'e^{(i)} + W)/n\big), \\ & & \qquad J \notin \{-e^{(j)}, -e^{(i)}\}, \end{array} \tag{3.32}$$

up to the time $\hat\tau^\delta_n$, and starts at $(0,0,0)$. Defining

$$\begin{aligned} q^u_{l,l',X}(W) \ &:= \ \mathbb{P}_X[W^\delta_n(u) = W \,|\, N^\delta_n(u) = l, (N')^\delta_n(u) = l']; \\ p_X(l, l', u) \ &:= \ \mathbb{P}_X[N^\delta_n(u) = l, (N')^\delta_n(u) = l'], \end{aligned}$$

this allows us to deduce that

$$\begin{aligned} &\mathbb{E}\{f(X^\delta_n(U)) \,|\, X^\delta_n(0) = X_0 - e^{(j)} - e^{(i)}\} - \mathbb{E}\{f(X^\delta_n(U)) \,|\, X^\delta_n(0) = X_0 - e^{(j)}\} \\ &\qquad - \mathbb{E}\{f(X^\delta_n(U)) \,|\, X^\delta_n(0) = X_0 - e^{(i)}\} + \mathbb{E}\{f(X^\delta_n(U)) \,|\, X^\delta_n(0) = X_0\} \\ &= \ \sum_{X \in \mathbb{Z}^d} f(X) \sum_{l \ge 0} \sum_{l' \ge 0} \Big\{ p_{X_0}(l-1, l'-1, U) q^U_{l-1,l'-1,X_0 - e^{(j)} - e^{(i)}}(X + le^{(j)} + l'e^{(i)}) \\ &\qquad\qquad - p_{X_0}(l-1, l', U) q^U_{l-1,l',X_0 - e^{(j)}}(X + le^{(j)} + l'e^{(i)}) \\ &\qquad\qquad\quad - p_{X_0}(l, l'-1, U) q^U_{l,l'-1,X_0 - e^{(i)}}(X + le^{(j)} + l'e^{(i)}) \\ &\qquad\qquad\qquad + p_{X_0}(l, l', U) q^U_{l,l',X_0}(X + le^{(j)} + l'e^{(i)}) \Big\}. \end{aligned} \tag{3.33}$$

Write $r_{jk,X}(l, l', u) := p_X(l-j, l'-k, u)/p_X(l, l', u)$ for $j, k \in \{0, 1\}$, and

$$R^u_{j,k,Y;l,l',X}(W) \ := \ q^u_{l-j,l'-k,X+Y}(W)/q^u_{l,l',X}(W).$$

Then the right hand side of (3.33) can be expressed as

$$\sum_{l\ge 0}\sum_{l'\ge 0} p_{X_0}(l,l',U) \sum_{w\in\mathbb{Z}^d} f(w - le^{(j)} - l'e^{(i)}) q^U_{l,l',X_0}(w)$$
$$\Big\{ r_{11,X_0}(l,l',U) R^U_{1,1,-e^{(j)}-e^{(i)};l,l',X_0}(w) - r_{10,X_0}(l,l',U) R^U_{1,0,-e^{(j)};l,l',X_0}(w)$$
$$- r_{01,X_0}(l,l',U) R^U_{0,1,-e^{(i)};l,l',X_0}(w) + 1 \Big\}. \qquad (3.34)$$

We now use the decomposition

$$rR \;=\; (r-1)(R-1) + (r-1) + (R-1) + 1$$

in each term of (3.34). The sum corresponding to taking 1 yields nothing. Then, for the sum corresponding to taking $(r-1)$ alone, summing over w first and using $\|f\|_\infty \le 1$, we have

$$\sum_{l,l'\ge 0} p_{X_0}(l,l',U) \sum_{w\in\mathbb{Z}^d} |f(w - le^{(j)} - l'e^{(i)})|\, q^U_{l,l',X_0}(w)$$
$$|r_{11,X_0}(l,l',U) - r_{10,X_0}(l,l',U) - r_{01,X_0}(l,l',U) + 1| \qquad (3.35)$$
$$\le\; \sum_{l\ge 0}\sum_{l'\ge 0} p_{X_0}(l,l',U) |r_{11,X_0}(l,l',U)) - r_{10,X_0}(l,l',U) - r_{01,X_0}(l,l',U) + 1|.$$

As for (3.9) and (3.15), the processes $(N^\delta_n, (N')^\delta_n)$ can be coupled to independent Poisson processes with rates $ng^{(j)}$ and $ng^{(i)}$ respectively on the interval $[0,U]$, with failure probability at most $\mathbb{P}_{X_0}[\hat\tau^\delta_n < U]$. Hence, using $\pi^{(j)}$ to denote $\mathrm{Po}\,(nUg^{(j)})$, (3.35) gives a contribution to (3.34) of at most

$$\sum_{l\ge 0}\sum_{l'\ge 0} |\pi^{(j)}\{l\} - \pi^{(j)}\{l-1\}|\, |\pi^{(i)}\{l'\} - \pi^{(i)}\{l'-1\}| + 4\mathbb{P}_{X_0}[\hat\tau^\delta_n < U]$$
$$=\; 4 d_{\mathrm{TV}}\big(\pi^{(j)}, \pi^{(j)} * \varepsilon_1\big)\, d_{\mathrm{TV}}\big(\pi^{(i)}, \pi^{(i)} * \varepsilon_1\big) + 4\mathbb{P}_{X_0}[\hat\tau^\delta_n < U]$$
$$\le\; \frac{4}{\sqrt{g^{(j)}g^{(i)}}}\,\frac{1}{nU} + 8n^{-4}, \qquad (3.36)$$

for $n \ge \max\{n_{(3.2)}, \psi^{-1}(\delta)\}$, uniformly in $|X_0 - nc|_\Sigma \le n\delta/4$.

We separate the sum corresponding to $(r-1)(R-1)$ in (3.34) into three pieces, corresponding to the subscripts $(1,1)$, $(1,0)$ and $(0,1)$, and use $\|f\|_\infty \le 1$. We then use an argument similar to that leading to (3.29); we sketch it for the $(1,1)$ case. First, by conditioning on the paths of N^δ_n and $(N')^\delta_n$ and using (3.46) below, it follows, much as for (3.29) and for (3.28), that, for each $l, l' \geq 0$,

$$\sum_{w \in \mathbb{Z}^d} q^U_{l,l',X_0}(w) |1 - R^U_{1,1,-e^{(j)}-e^{(i)};l,l',X_0}(w)|$$
$$\leq \min\{2, 2n^{-1/2}\sqrt{K_{(3.25)}(G^{(i)} + G^{(j)})U} + 4n^{-1}K_{(3.25)}(G^{(i)} + G^{(j)})U$$
$$+ 2\mathbb{P}_{X_0}[\hat{\tau}^\delta_n < U \mid N^\delta_n(u) = l, (N')^\delta_n(u) = l']\} \quad (3.37)$$
$$\leq 4n^{-1/2}\sqrt{K_{(3.25)}(G^{(i)} + G^{(j)})U} + 2\mathbb{P}_{X_0}[\hat{\tau}^\delta_n < U \mid N^\delta_n(u) = l, (N')^\delta_n(u) = l'].$$

Then, as in treating (3.35), and using Lemma 2.5, we have

$$\sum_{l,l' \geq 0} p_{X_0}(l, l', U) |r_{11,X_0}(l, l', U) - 1|$$
$$\leq 2\{d_{\mathrm{TV}}(\pi^{(j)}, \pi^{(j)} * \varepsilon_1) + d_{\mathrm{TV}}(\pi^{(i)}, \pi^{(i)} * \varepsilon_1)\} + 4\mathbb{P}_{X_0}[\hat{\tau}^\delta_n < U]$$
$$\leq \frac{2}{\sqrt{nUg^{(j)}}} + \frac{2}{\sqrt{nUg^{(i)}}} + 8n^{-4} \leq \frac{4}{\sqrt{nUg^-_{ij}}} + 8n^{-4}, \quad (3.38)$$

for $n \geq \max\{n_{(3.2)}, \psi^{-1}(\delta)\}$, uniformly in $|X_0 - nc|_\Sigma \leq n\delta/4$.

Combining the first part of (3.37) with (3.38) gives a contribution to (3.34) bounded by

$$Kn^{-1}d^{1/2}\big((G^{(i)} + G^{(j)})/g^-_{ij}\big)^{1/2} + 12n^{-2}, \quad (3.39)$$

uniformly for $U \leq U^{(i,j)}$ and $|X_0 - nc|_\Sigma \leq n\delta/4$, for $K := 4\sqrt{K_{(3.25)}} \in \mathcal{K}$. Taking the second part of (3.37) with (3.38), it is immediate that

$$2\sum_{l,l' \geq 0} p_{X_0}(l, l', U) |r_{11,X_0}(l, l', U) - 1| \mathbb{P}_{X_0}[\hat{\tau}^\delta_n < U \mid N^\delta_n(u) = l, (N')^\delta_n(u) = l']$$
$$\mathbb{1}\{r_{11,X_0}(l, l', U) \leq n^2\}$$
$$\leq 2n^2\mathbb{P}_{X_0}[\hat{\tau}^\delta_n < U] \leq 4n^{-2},$$

by Lemma 2.5, since $n \geq \max\{n_{(3.2)}, \psi^{-1}(\delta)\}$. For the remainder, we have at most

$$2\sum_{l,l' \geq 0} p_{X_0}(l, l', U)\mathbb{1}\{r_{11,X_0}(l+1, l'+1, U) > n^2\} \quad (3.40)$$
$$\leq 2\mathbb{P}_{X_0}[\hat{\tau}^\delta_n < U] + 2\sum_{l,l' \geq 0} \pi^{(j)}\{l\}\pi^{(i)}\{l'\}\mathbb{1}\{r_{11,X_0}(l+1, l'+1, U) > n^2\}.$$

Now

$$|p_{X_0}(l, l', U) - \pi^{(j)}\{l\}\pi^{(i)}\{l'\}| \leq \mathbb{P}_{X_0}[\hat{\tau}^\delta_n < U] \leq 2n^{-4}.$$

This implies that, if

$$\min(\pi^{(j)}\{l\},\pi^{(i)}\{l'\},\pi^{(j)}\{l+1\},\pi^{(i)}\{l'+1\}) \;\ge\; 2n^{-2}, \tag{3.41}$$

then $r_{11,X_0}(l+1,l'+1,U) \le n^2$, giving no contribution to the sum in (3.40). This is because

$$r_{11,X_0}(l+1,l'+1,U) \;\le\; 3\frac{\pi^{(j)}\{l\}\pi^{(i)}\{l'\}}{\pi^{(j)}\{l+1\}\pi^{(i)}\{l'+1\}} \;\le\; 3\frac{(l+1)(l'+1)}{n^2U^2g_{ij}^-g_{ij}^+}\,;$$

by Proposition A.2.3 (i) of Barbour, Holst & Janson (1992), if (3.41) holds,

$$3\frac{(l+1)(l'+1)}{n^2U^2g_{ij}^-g_{ij}^+} \;\le\; 100(\log n)^2 \;<\; n^2,$$

for all $n \ge 40$. In proving the first inequality, we assume that $nUg_{ij}^- \ge 1$, since the inequality in the statement of the theorem is immediate for smaller nU. This leaves only a contribution to the sum in (3.40) from l, l' for which (3.41) does not hold, and this is at most

$$2\sum_{l\ge 0}\bigl\{\pi^{(j)}\{l\}\mathbb{1}\{\pi^{(j)}\{l\}\le 2n^{-2}\}+\pi^{(i)}\{l\}\mathbb{1}\{\pi^{(i)}\{l\}\le 2n^{-2}\}\bigr\} \;\le\; 8n^{-3/2},$$

by Proposition A.2.3 (ii), (iii) and (iv) of Barbour, Holst & Janson (1992), if $n \ge 10$, because we also have $nUg_{ij}^+ \le n$ in $U \le U^{(i,j)}$.

The trickiest sum is that corresponding to $(R-1)$ alone. Using $\|f\|_\infty \le 1$, we need first to examine the quantity

$$\sum_{w\in\mathbb{Z}^d} q^U_{l,l',X_0}(w)\Bigl|R^U_{1,1,-e^{(j)}-e^{(i)};l,l',X_0}(w)-R^U_{1,0,-e^{(j)};l,l',X_0}(w)-R^U_{0,1,-e^{(i)};l,l',X_0}(w)+1\Bigr|. \tag{3.42}$$

We treat it, after conditioning on realizations of the underlying Poisson processes N_n^δ and $(N')_n^\delta$, as the expectation of the absolute value at time U of an $\mathcal{F}^{X_n^\delta}$-martingale $M^{(2)}(W_n^\delta)$, defined in (3.43) below. Let $W^u := (W(t),\, 0 \le t \le u)$ denote the restriction of a function W on $\mathbb{R}_+$ to $[0,u]$. Write $\mathbf{s}_l := (s_1,\ldots,s_l)$, $\mathbf{s}'_{l'} := (s'_1,\ldots,s'_{l'})$. If realizations of N_n^δ and $(N')_n^\delta$, having l and l' points respectively in $[0,U]$, are denoted by $\nu_l(\cdot;\mathbf{s}_l)$ and $\nu'_{l'}(\cdot;\mathbf{s}'_{l'})$, as in (3.17), we then denote conditional probability and expectation, given $(N_n^\delta)^U = \nu_l(\cdot;\mathbf{s}_l)$, $((N')_n^\delta)^U = \nu'_{l'}(\cdot;\mathbf{s}'_{l'})$ and $X_n^\delta(0) = X$, by $\mathbb{P}^U_{\mathbf{s}_l,\mathbf{s}'_{l'},X}$ and $\mathbb{E}^U_{\mathbf{s}_l,\mathbf{s}'_{l'},X}$, and we denote the corresponding conditional density of $(W_n^\delta)^u$ at the path segment W^u, with respect to some suitable reference measure, by

$$q^U(u,W^u;\,\mathbf{s}_l,\mathbf{s}'_{l'},X).$$

We then define the Radon–Nikodym derivatives

$$\begin{aligned}
R^U_{11}(u,W^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0) &:= \frac{q^U(u,W^u;\ \mathbf{s}_{l-1},\mathbf{s}'_{l'-1},X_0-e^{(j)}-e^{(i)})}{q^U(u,W^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)};\\
R^U_{10}(u,W^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0) &:= \frac{q^U(u,W^u;\ \mathbf{s}_{l-1},(\mathbf{s}'_{l'-1},s'_*),X_0-e^{(j)})}{q^U(u,W^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)};\\
R^U_{01}(u,W^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0) &:= \frac{q^U(u,W^u;\ (\mathbf{s}_{l-1},s_*),\mathbf{s}'_{l'-1},X_0-e^{(i)})}{q^U(u,W^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)};
\end{aligned}$$

these have explicit formulae analogous to (3.22). We use them to formulate the analogue of the argument used in the proof of Theorem 3.3. For example, we can write

$$\begin{aligned}
&\sum_{w\in\mathbb{Z}^d} q^U_{l,l',X_0}(w)R^U_{1,1,-e^{(j)}-e^{(i)};l,l',X_0}(w)\\
&= \frac{1}{U^{l+l'}}\int_{[0,U]^{l+l'}} ds_1\ldots ds_{l-1}ds_*\,ds'_1\ldots ds'_{l'-1}ds'_*\\
&\qquad\qquad \sum_{w\in\mathbb{Z}^d}\mathbb{P}^U_{\mathbf{s}_{l-1},\mathbf{s}'_{l'-1},X_0-e^{(j)}-e^{(i)}}[W(U)=w]\\
&= \frac{1}{U^{l+l'}}\int_{[0,U]^{l+l'}} ds_1\ldots ds_{l-1}ds_*\,ds'_1\ldots ds'_{l'-1}ds'_*\\
&\qquad \sum_{w\in\mathbb{Z}^d}\mathbb{E}^U_{(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0}\{R^U_{11}(U,W^U;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)I[W(U)=w]\}.
\end{aligned}$$

The mean zero martingale $M^{(2)}(W^\delta_n)$ of main interest to us can then be expressed as

$$\begin{aligned}
M^{(2)}(W^\delta_n)(u) \;:=\; & R^U_{11}(u,(W^\delta_n)^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)\\
&- R^U_{10}(u,(W^\delta_n)^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)\\
&- R^U_{01}(u,(W^\delta_n)^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)+1,
\end{aligned}\tag{3.43}$$

with $(W^\delta_n)^U$ a random element with distribution $\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0}$. We also define the $\mathcal{F}^{X^\delta_n}$-martingale

$$M^{(1)}(W^\delta_n)(u) \;:=\; R^U_{11}(u,(W^\delta_n)^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)-1,$$

for use in the proof below, as well as for the proof of the estimate of the $(1,1)$ term in (3.37) above.

We now set $x^\delta_n(u):=n^{-1}(W^\delta_n(u)+X_0-e^{(j)}\nu_{l-1}(u;\mathbf{s}_{l-1})-e^{(i)}\nu'_{l'-1}(u;\mathbf{s}'_{l'-1}))$ for $u<\min\{s_*,s'_*\}$. If, for $u<\min\{s_*,s'_*\}$ and $|x^\delta_n(u)-c|_\Sigma\le\delta-3n^{-1}J^\Sigma_{\max}$, there is a jump of $e^{(r)}$ in W_n^δ at time u, for some $1 \le r \le d$, this gives rise to a jump in the martingale $M^{(2)}(W_n^\delta)$ at u of

$$
\begin{aligned}
&R_{11}^U(u-, (W_n^\delta)^{u-}; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0)\Big(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u-) - n^{-1}(e^{(j)} + e^{(i)}))}{\hat{g}^{e^{(r)}}(x_n^\delta(u-))} - 1\Big) \\
&\quad - \quad R_{10}^U(u-, (W_n^\delta)^{u-}; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0)\Big(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u-) - n^{-1}e^{(j)})}{\hat{g}^{e^{(r)}}(x_n^\delta(u-))} - 1\Big) \\
&\quad - \quad R_{01}^U(u-, (W_n^\delta)^{u-}; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0)\Big(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u-) - n^{-1}e^{(i)})}{\hat{g}^{e^{(r)}}(x_n^\delta(u-))} - 1\Big).
\end{aligned}
$$

If $s_* < u < s'_*$, the elements $-n^{-1}e^{(j)}$ are removed from the arguments of $\hat{g}^{e^{(r)}}$, simplifying the considerations, but then $x_n^\delta(u)$ is replaced by $x_n^\delta(u) - n^{-1}e^{(j)}$; the elements $-n^{-1}e^{(i)}$ are removed if $s'_* < u < s_*$, and then $x_n^\delta(u)$ is replaced by $x_n^\delta(u) - n^{-1}e^{(i)}$; if $u > \max\{s_*, s'_*\}$, both elements $-n^{-1}e^{(j)}$ and $-n^{-1}e^{(i)}$ are removed, and so there is no jump. Now, because the transition rate $g^{e^{(r)}}(x)$ is linear in x,

$$
\begin{aligned}
\Big(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u) - n^{-1}(e^{(j)} + e^{(i)}))}{\hat{g}^{e^{(r)}}(x_n^\delta(u))} - 1\Big) - \Big(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u) - n^{-1}e^{(j)})}{\hat{g}^{e^{(r)}}(x_n^\delta(u))} - 1\Big) \\
- \Big(\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u) - n^{-1}e^{(i)})}{\hat{g}^{e^{(r)}}(x_n^\delta(u))} - 1\Big) \;=\; 0,
\end{aligned}
$$

and so R^U can be replaced by $|R^U - 1|$ when bounding the sizes of the jumps, irrespective of the relative positions of s_*, s'_* and u. Since also, from (2.3) and Assumption S4,

$$
\left|\frac{\hat{g}^{e^{(r)}}(x_n^\delta(u) + n^{-1}Y)}{\hat{g}^{e^{(r)}}(x_n^\delta(u))} - 1\right| \;\le\; 2n^{-1}|Y|L_1, \tag{3.44}
$$

the remaining contributions to the jump in $M^{(2)}(W_n^\delta)$ are at most

$$
\begin{aligned}
\frac{4L_1}{n}\{&|R_{11}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) - 1| \\
&+ |R_{10}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) - 1| \\
&\quad + |R_{01}^U(u, (W_n^\delta)^u; (\mathbf{s}_{l-1}, s_*), (\mathbf{s}'_{l'-1}, s'_*), X_0) - 1|\}.
\end{aligned} \tag{3.45}
$$

We can now bound the quadratic variation arising from each of the three terms individually, by the argument leading to (3.26). Defining

$$
\tilde{\varphi}_n \;:=\; \inf\{u \ge 0\colon\; \tilde{m}(u) \ge 2\},
$$

where

$$\begin{aligned}\tilde{m}(u) \;:=\; \max\bigl\{ & R_{11}^U(u,(W_n^\delta)^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0), \\ & R_{10}^U(u,(W_n^\delta)^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0), \\ & R_{01}^U(u,(W_n^\delta)^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)\bigr\},\end{aligned}$$

we use the martingale $M^{(1)}(W_n^\delta)$ and (3.44) with the argument leading to (3.25) to give

$$\begin{aligned}&\mathbb{E}^U_{(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0}\{[R_{11}^U(u\wedge\hat{\tau}_n^\delta\wedge\tilde{\varphi}_n,(W_n^\delta)^{u\wedge\hat{\tau}_n^\delta\wedge\tilde{\varphi}_n};(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)-1]^2\} \\ &\quad\le\; n^{-1}4K_{(3.25)}(G^{(i)}+G^{(j)})u; \qquad (3.46)\end{aligned}$$

the same bound holds for R_{10}^U and R_{01}^U also, but with $4(G^{(i)}+G^{(j)})$ replaced by $G^{(j)}$ and $G^{(i)}$ respectively. Hence the expected quadratic variation of the martingale $M^{(2)}(W_n^\delta)$ stopped at $u\wedge\hat{\tau}_n^\delta\wedge\tilde{\varphi}_n$ is at most

$$\begin{aligned}&n(G^{(i)}+G^{(j)})\int_0^u\left(\frac{12L_1}{n}\right)^2\left(\frac{4K_{(3.25)}(G^{(i)}+G^{(j)})v}{n}\right)dv \\ &\qquad\le\; 2n^{-2}\bigl((G^{(i)}+G^{(j)})u\bigr)^2(12L_1)^2K_{(3.25)} \;\le\; n^{-2}K_8\bigl((G^{(i)}+G^{(j)})u\bigr)^2,\end{aligned}$$

uniformly in $|X_0-nc|_\Sigma\le n\delta$, and in l, l', $\mathbf{s}_{l-1}$, $\mathbf{s}'_{l'-1}$, s_* and s'_*, for $K_8 := 2(12L_1)^2K_{(3.25)}\in\mathcal{K}$. This gives a contribution of at most $n^{-1}\sqrt{K_8}\,(G^{(i)}+G^{(j)})U$ to (3.42), and hence to (3.34), from the expectation of $|M^{(2)}-1|$, stopped at $U\wedge\hat{\tau}_n^\delta\wedge\tilde{\varphi}_n$.

Because the martingale $M^{(2)}(W_n^\delta)$ is not uniformly bounded from below, we can no longer use an argument as for (3.28) to bound the contributions to (3.34) from the events $\hat{\tau}_n^\delta<U$ and $\tilde{\varphi}_n<U$. Instead, we consider their contributions for each element of $M^{(2)}(W_n^\delta)$ separately. For example, writing

$$\begin{aligned}\widetilde{R}_*^U &:= R_{11}^U((W_n^\delta)^U;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0); \\ \mathbb{E}_*^U &:= \mathbb{E}^U_{(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0}; \quad \mathbb{P}_*^U := \mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0},\end{aligned}$$

we have

$$
\begin{aligned}
&\sum_{w\in\mathbb{Z}^d} f(w-le^{(j)}-l'e^{(i)})\mathbb{E}_*^U\{\widetilde{R}_*^U I[W_n^\delta(U)=w, \hat\tau_n^\delta<U]\}\\
&\quad\le\ \mathbb{E}_*^U\Big\{\widetilde{R}_*^U I\Big[\sup_{0\le u\le U}|X_0+W_n^\delta(u)-e^{(j)}\nu(u;\mathbf{s}_{l-1},s_*)\\
&\qquad\qquad\qquad -e^{(i)}\nu'(u;\mathbf{s}'_{l'-1},s'_*)-nc|\ge n\delta-3J^\Sigma_{\max}\Big]\Big\}\\
&\quad=\ \mathbb{E}^U_{\mathbf{s}_{l-1},\mathbf{s}'_{l'-1},X_0-e^{(j)}-e^{(i)}}\Big\{I\Big[\sup_{0\le u\le U}|X_0+W_n^\delta(u)-e^{(j)}\nu(u;\mathbf{s}_{l-1})\\
&\qquad -e^{(i)}\nu'(u;\mathbf{s}'_{l'-1})-e^{(j)}\mathbb{1}_{[s_*,U]}(u)-e^{(i)}\mathbb{1}_{[s'_*,U]}(u)-nc|\ge n\delta-3J^\Sigma_{\max}\Big]\Big\}\\
&\quad\le\ \mathbb{P}^U_{\mathbf{s}_{l-1},\mathbf{s}'_{l'-1},X_0-e^{(j)}-e^{(i)}}[\tau_n^\delta(\delta-3n^{-1}J^\Sigma_{\max}-2n^{-1}/\sqrt{\lambda_{\min}(\Sigma)})<U].
\end{aligned}
$$

Now both the inequalities

$$|X_0-e^{(j)}-e^{(i)}-nc|_\Sigma\ \le\ n\delta/2$$

and

$$\tau_n^\delta\big(\delta-3n^{-1}J^\Sigma_{\max}-2n^{-1}/\sqrt{\lambda_{\min}(\Sigma)}\big)\ \ge\ \tau_n^\delta(3\delta/4)$$

are satisfied if $n^{3/4}\delta>20d^{-1}J^\Sigma_{\max}$ and $|X_0-nc|_\Sigma\le n\delta/4$. Taking expectations over the realizations of $(N_n^\delta)^U$ and $((N')_n^\delta)^U$ and invoking Lemma 2.5 thus gives a contribution to (3.34) of at most $2n^{-4}$, uniformly in $|X_0-nc|_\Sigma\le n\delta/4$, for $n\ge\max\big\{n_{2.5}(1/g_*),(20d^{-1}J^\Sigma_{\max}/\delta)^{4/3},\psi^{-1}(\delta)\big\}$; this inequality is satisfied if $n\ge\max\{n_{(3.2)},\psi^{-1}(\delta)\}$. Then

$$
\begin{aligned}
&\sum_{w\in\mathbb{Z}^d} f(w-le^{(j)}-l'e^{(i)})\mathbb{E}_*^U\Big\{\widetilde{R}_*^U I[W_n^\delta(U)=w,\tilde\varphi_n<\min\{U,\hat\tau_n^\delta\}]\Big\}\\
&\quad\le\ \mathbb{E}_*^U\{\widetilde{R}_*^U I[\tilde\varphi_n<\min\{U,\hat\tau_n^\delta\}]\}\\
&\quad\le\ 2\mathbb{P}_*^U[\tilde\varphi_n<\min\{U,\hat\tau_n^\delta\}]+\mathbb{E}_*^U\{\widetilde{R}_*^U I[\tilde\varphi_n^{11}<\min\{U,\hat\tau_n^\delta\}]\},
\end{aligned}
$$

where

$$\tilde\varphi_n^{11}\ :=\ \inf\big\{u\ge0\colon\ R^U_{11}(u,(W_n^\delta)^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0)\ge2\big\}.$$

The first of these terms is at most $5n^{-1}K_{(3.25)}(G^{(i)}+G^{(j)})U$, using (3.46) and its analogues for the quantities R^U_{11}, R^U_{10} and R^U_{01} appearing in the definition of $\tilde m(U)$, and then applying Kolmogorov's inequality; the argument is much as for (3.27). The second is no larger than

$$2\mathbb{P}^U_{\mathbf{s}_{l-1},\mathbf{s}'_{l'-1},X_0-e^{(j)}-e^{(i)}}[\tilde\varphi_n^{11}<\min\{U,\hat\tau_n^\delta\}]. \tag{3.47}$$

However, under $\mathbb{P}^U_{\mathbf{s}_{l-1},\mathbf{s}'_{l'-1},X_0-e^{(j)}-e^{(i)}}$, the process

$$M' := \{1/R^U_{11}(u,(W^\delta_n)^u;(\mathbf{s}_{l-1},s_*),(\mathbf{s}'_{l'-1},s'_*),X_0),\, u \geq 0\}$$

is an $\mathcal{F}^{X^\delta_n}$-martingale with mean 1. Arguing much as for (3.25), its expected quadratic variation up to the time $\min\{U,\hat{\tau}^\delta_n,(\tilde{\varphi}'_n)^{11}\}$ can be shown to be at most $4n^{-1}K_{(3.25)}(G^{(i)}+G^{(j)})U$; here, $(\tilde{\varphi}'_n)^{11} := \inf\{u \geq 0\colon\ M'(u) \geq 2\}$. Using an argument much as that for (3.27), Kolmogorov's inequality now shows that the quantity in (3.47) is itself at most $8n^{-1}K_{(3.25)}(G^{(i)}+G^{(j)})U$, giving a contribution to (3.34) of order $O(n^{-1}(G^{(i)}+G^{(j)})U)$. Combining these considerations with (3.36) and (3.39), the inequality of the theorem follows for $n \geq \max\{n_{(3.2)},\psi^{-1}(\delta)\}$. □

3.3 Coupling copies of X^δ_n

In this section, we show that copies of X^δ_n with different initial states can be defined on the same probability space, in such a way that they coincide rather quickly. As a consequence, the total variation distance between their distributions becomes small as time increases. Our arguments are reminiscient of those in Roberts & Rosenthal (1996).

The basic coupling that we use relies mainly on the drift towards nc to achieve this. We define the process $(X^\delta_{n,1}(t),X^\delta_{n,2}(t))$ on $\widetilde{B}_{n,\delta}(c)\times\widetilde{B}_{n,\delta}(c)$ to have the transition rates

$$\begin{array}{llll}
(X_1,X_2) & \to & (X_1+J,X_2+J) & \text{at rate} \quad n\{g^J_\delta(n^{-1}X_1)\wedge g^J_\delta(n^{-1}X_2)\}; \\
(X_1,X_2) & \to & (X_1,X_2+J) & \text{at rate} \quad n\{g^J_\delta(n^{-1}X_2)-g^J_\delta(n^{-1}X_1)\}_+; \\
(X_1,X_2) & \to & (X_1+J,X_2) & \text{at rate} \quad n\{g^J_\delta(n^{-1}X_1)-g^J_\delta(n^{-1}X_2)\}_+,
\end{array}$$

for each $J \in \mathcal{J}$. Let its generator be denoted by $\widetilde{\mathcal{A}}^\delta_n$. Our coupling argument begins with a drift inequality.

Lemma 3.6. *Let X_n be a sequence of elementary processes. Define $h_1(X_1,X_2) := |X_1-X_2|^2_\Sigma$; let α_1 be as in (2.4) and $\delta_1 := \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$. Then, for $\delta \leq \delta_1/3$, there exists $K_{3.6} \in \mathcal{K}$, defined in (3.48), such that, for all (X_1,X_2) with* $\max\{|X_1-nc|_\Sigma,|X_2-nc|_\Sigma\} \leq n\delta - J^\Sigma_{\max}$ *and $|X_1-X_2|_\Sigma \geq dK_{3.6}$, we have*

$$\widetilde{\mathcal{A}}^\delta_n h_2(X_1,X_2) \;\leq\; -\tfrac{1}{2}\alpha_1 h_2(X_1,X_2),$$

where $h_2(X_1,X_2) := h_1(X_1,X_2)+d^2K^2_{3.6}$.

Proof. For any $\xi > 0$, write $h^{(\xi)}(X_1,X_2) := h_1(X_1,X_2)+\xi$. By the definition of h_1, the transitions where both components of (X_1,X_2) make the same jump make no contribution to $\widetilde{\mathcal{A}}_n^\delta h^{(\xi)}(X_1, X_2)$. Hence, for (X_1, X_2) with $\max\{|X_1 - nc|_\Sigma, |X_2 - nc|_\Sigma\} \le n\delta - J^\Sigma_{\max}$, and writing $x_i := n^{-1}X_i$, $i = 1, 2$, we have

$$\begin{aligned}\widetilde{\mathcal{A}}_n^\delta h^{(\xi)}(X_1, X_2) \;&=\; n \sum_{J\in\mathcal{J}} \Big\{ (g^J(x_1) - g^J(x_2))_+ \{2J^T\Sigma^{-1}(X_1 - X_2) + J^T\Sigma^{-1}J\} \\ &\qquad + (g^J(x_2) - g^J(x_1))_+ \{-2J^T\Sigma^{-1}(X_1 - X_2) + J^T\Sigma^{-1}J\} \Big\} \\ &= 2n(F(x_1) - F(x_2))^T\Sigma^{-1}(X_1 - X_2) + n \sum_{J\in\mathcal{J}} |g^J(x_1) - g^J(x_2)|\, |J|^2_\Sigma,\end{aligned}$$

since, for such (X_1, X_2), $g^J_\delta(x_1) = g^J(x_1)$ and $g^J_\delta(x_2) = g^J(x_2)$. Now, since the transition rates $g^J(x)$ are all linear in x, we have

$$\begin{aligned}2n(F(x_1) - F(x_2))^T\Sigma^{-1}(X_1 - X_2) \;&=\; -(X_1 - X_2)^T\Sigma^{-1}\sigma^2\Sigma^{-1}(X_1 - X_2) \\ \le\; -\lambda_{\min}(\sigma^2_\Sigma)|X_1 - X_2|^2_\Sigma \;&=\; -2\alpha_1 h_1(X_1, X_2).\end{aligned}$$

Then

$$\begin{aligned}n \sum_{J\in\mathcal{J}} |g^J(x_1) - g^J(x_2)|\, |J|^2_\Sigma \;&\le\; L_1(\Lambda/\lambda_{\min}(\Sigma))|X_1 - X_2|_\Sigma \sqrt{\lambda_{\max}(\Sigma)} \\ &\le\; \alpha_1 |X_1 - X_2|^2_\Sigma \;=\; \alpha_1 h_1(X_1, X_2),\end{aligned}$$

if $|X_1 - X_2|_\Sigma \ge dK_{3.6}$, where

$$K_{3.6} \;:=\; \max\{1, L_1(\overline{\Lambda}/\alpha_1)\sqrt{\rho(\Sigma)/\lambda_{\min}(\Sigma)}\} \;\in\; \mathcal{K}. \tag{3.48}$$

From this, it follows that

$$\widetilde{\mathcal{A}}_n^\delta h^{(\xi)}(X_1, X_2) \;\le\; -\alpha_1 h_1(X_1, X_2) \;\le\; -\tfrac{1}{2}\alpha_1(h_1(X_1, X_2) + d^2K^2_{3.6}),$$

for $\delta \le \delta_1/3$ and for $|X_1 - X_2|_\Sigma \ge dK_{3.6}$. Taking $\xi = d^2K^2_{3.6}$ proves the lemma. □

We now convert the drift inequality into a bound on the distribution of the coupling time

$$\tau_C \;:=\; \inf\{t \ge 0\colon X^\delta_{n,1}(t) = X^\delta_{n,2}(t)\},$$

for arbitrary values of $(X^\delta_{n,1}(0), X^\delta_{n,2}(0))$. Our broad strategy is as follows. If $\max\{|X^\delta_{n,1}(0) - nc|_\Sigma, |X^\delta_{n,2}(0) - nc|_\Sigma\} > 3n\delta/8$, we run both processes *independently* for a fixed time interval t_1, chosen in such a way that we have $\max\{|X^\delta_{n,1}(t_1) - nc|_\Sigma, |X^\delta_{n,2}(t_1) - nc|_\Sigma\} \le 3n\delta/8$, with probability at least $1/16$. If not, we continue to repeat the procedure, over intervals of length t_1, until both $X^\delta_{n,1}$ and $X^\delta_{n,2}$ are within $3n\delta/8$ of nc in the $|\cdot|_\Sigma$-norm. We then couple the processes $X^\delta_{n,1}$ and $X^\delta_{n,2}$ as for Lemma 3.6, and run them until the minimum $(\tau_3 \wedge \tau_0)$ of the time τ_3, at which $|X^\delta_{n,1}(t) - X^\delta_{n,2}(t)|_\Sigma$ first falls below the value $dK_{3.6}$, and the time τ_0, at which first $\max\{|X^\delta_{n,1}(t) - nc|_\Sigma, |X^\delta_{n,2}(t) - nc|_\Sigma\} > n\delta/2$. We call these two stages together a 'drift phase'.

If τ_0 is the first to occur, we begin another drift phase. If not, we enter a 'trial phase', of length $t_3 = 1/\alpha_1$. By Theorem 3.3 and Assumption S2, and because $|X^\delta_{n,1}(t) - X^\delta_{n,2}(t)|_\Sigma \le dK_{3.6}$ implies that $|X^\delta_{n,1}(t) - X^\delta_{n,2}(t)|_1 \le d^{3/2} K_{3.6} \sqrt{\lambda_{\max}(\Sigma)}$, we have

$$d_{\rm TV}\bigl(\mathcal{L}(X^\delta_{n,1}(\tau_3 + t_3)), \mathcal{L}(X^\delta_{n,2}(\tau_3 + t_3))\bigr) \;\le\; d^{3/2} C_* n^{-1/2} \left(\frac{d\overline{\Lambda}}{g_*}\right)^{1/4}, \quad (3.49)$$

for $n \ge n_{(3.2)}$, where we use $G^{(j)} \le d\overline{\Lambda}$, and where

$$C_* \;:=\; K_{3.6} \sqrt{\lambda_{\max}(\Sigma)} \max_{1 \le j \le d} \left\{ K^j_{3.3} \max\left(1, \frac{\sqrt{\alpha_1}}{(dg^{(j)}\overline{\Lambda})^{1/4}}\right)\right\} \;\in\; \mathcal{K}. \quad (3.50)$$

Hence the two processes can be coupled in such a way that $X^\delta_{n,1}(\tau_3 + t_3) = X^\delta_{n,2}(\tau_3 + t_3)$, except on an event of probability at most $d^{3/2} C_* n^{-1/2} (\Lambda/g_*)^{1/4}$, and this is the coupling that we use. On the event that the values of the two processes are equal at time $\tau_3 + t_3$, the coupling is said to have been successful. If not, a new drift phase begins, (or another trial phase, if $|X^\delta_{n,1}(\tau_3 + t_3) - X^\delta_{n,2}(\tau_3 + t_3)|_\Sigma \le dK_{3.6}$ and $\max\{|X^\delta_{n,1}(\tau_3 + t_3) - nc|_\Sigma, |X^\delta_{n,2}(\tau_3 + t_3) - nc|_\Sigma\} \le n\delta/2$). This sequence of steps is repeated until coupling is achieved.

Theorem 3.7. *Let X_n be a sequence of elementary processes. Let α_1 be as in (2.4), and let $\delta_1 := \min\{3, \delta_0/\sqrt{\lambda_{\max}(\Sigma)}\}$. Then, for any $\delta \le \delta_1/3$, there is a constant $n_{3.7}$ in $\mathcal{K}$ such that, whatever the values of $X_1, X_2 \in \widetilde{B}_{n,\delta}(c) := \{X \in \mathbb{Z}^d \colon |X - nc|_\Sigma \le n\delta\}$ and $t \ge 0$,*

$$d_{\rm TV}\bigl(\mathcal{L}(X^\delta_n(t) \,|\, X^\delta_n(0) = X_1), \mathcal{L}(X^\delta_n(t) \,|\, X^\delta_n(0) = X_2)\bigr) \;\le\; 9\,(2n)^{1/16} e^{-\alpha_2 t},$$

for all $n \ge \max\{d^4, n_{3.7}, \psi^{-1}(\delta/2)\}$, where $\alpha_2 := \alpha_1/128$. The quantity $n_{3.7}$ is defined in (3.63).

Proof. Recalling the definition (1.13) of $\widetilde{B}_{n,\delta}(c)$, we begin by writing

$$\begin{aligned}
B_0 &:= \widetilde{B}_{n,\delta}(c) \times \widetilde{B}_{n,\delta}(c);\\
B_1 &:= \{(X_1, X_2) \in B_0 \colon \max_{i=1,2} |X_i - nc|_\Sigma \le n\delta/2\};\\
B_2 &:= \{(X_1, X_2) \in B_0 \colon \max_{i=1,2} |X_i - nc|_\Sigma \le 3n\delta/8\};\\
B_3 &:= \{(X_1, X_2) \in B_1 \colon |X_1 - X_2|_\Sigma \le dK_{3.6}\}.
\end{aligned}$$

Clearly, $B_3 \subset B_1 \subset B_0$ and $B_2 \subset B_1$. Let

$$\begin{aligned}
\tau_0 &:= \inf\{t \geq 0\colon (X^\delta_{n,1}(t), X^\delta_{n,2}(t)) \notin B_1\};\\
\tau_2 &:= \inf\{t \geq 0\colon (X^\delta_{n,1}(t), X^\delta_{n,2}(t)) \in B_2\};\\
\tau_3 &:= \inf\{t \geq 0\colon (X^\delta_{n,1}(t), X^\delta_{n,2}(t)) \in B_3\}.
\end{aligned}$$

Then, for any $s, \beta > 0$, and for $0 \leq i \leq 4$, we define

$$\varphi_i(\beta, s) := \max_{(X_1,X_2)\in B_i} \mathbb{E}_{X_1,X_2}\{e^{\beta(\tau_C \wedge s)}\}, \tag{3.51}$$

where the coupling time τ_C is defined for copies of the processes $X^\delta_{n,1}$ and $X^\delta_{n,2}$ specified using drift and trial phases as above, and where $\mathbb{E}_{X_1,X_2}$ and $\mathbb{P}_{X_1,X_2}$ refer to the distribution conditional on $(X^\delta_{n,1}(0), X^\delta_{n,2}(0)) = (X_1, X_2)$. We shall establish that, for n large enough,

$$\mathbb{P}_{X_1,X_2}[\tau_C > t] \leq 9(2n)^{1/16} e^{-\alpha_1 t/128}.$$

Fix t_1' such that $e^{t_1'} = 128$, and write $t_1 := \alpha_1^{-1} t_1'$. Then, for any $(X_1, X_2) \in B_0$, it follows from Lemma 2.4 that

$$\mathbb{P}_{X_1,X_2}[\inf\{t > 0\colon |X^\delta_{n,1}(t)|_\Sigma \leq n\delta/4\} > t_1] \leq 1/2,$$

if also $n\delta \geq 8J^\Sigma_{\max}$; this is true, from (3.3), for $n \geq \max\{n_{(3.2)}, \psi^{-1}(\delta/2)\}$. But now, taking $\eta = 3\delta/8$ and $|X_0 - nc|_\Sigma \leq \delta/4$ in Lemma 2.5, it follows that

$$\mathbb{P}_{X_1,X_2}[|X^\delta_{n,1}(t_1)|_\Sigma \leq 3\delta/8] \geq 1/4,$$

provided that $n \geq \max\{n_{2.2}, (t_1' K_{2.5}(\overline{\Lambda}/\alpha_1))^{1/2}\}$, and that $2n^{-4} \leq 1/2$. Hence, for any choice of $(X_1, X_2) \notin B_2$ and for $n \geq \max\{n_1, \psi^{-1}(\delta/2)\}$, where

$$n_1 := \max\{d^4, n_{(3.2)}\} \in \mathcal{K},$$

it follows that $\mathbb{P}_{X_1,X_2}[(X^\delta_{n,1}(t_1), X^\delta_{n,2}(t_1)) \in B_2] \geq 1/16$, if $X^\delta_{n,1}$ and $X^\delta_{n,2}$ are run independently over the interval $[0, t_1]$. Thus, defining

$$\varphi(\beta, s) := \max_{(X_1,X_2)\in B_0} \mathbb{E}_{X_1,X_2}\big\{e^{\beta(\tau_2 \wedge s)}\big\},$$

the Markov property yields

$$\varphi(\beta, s) \leq e^{\beta t_1}\{1 + 15\varphi(\beta, s)\}/16.$$

Choosing $u_0 = 1/32$, so that, in particular, $15e^{u_0 t_1'}/16 = 15(128)^{u_0}/16 < 31/32$, and then $\beta = u_0\alpha_1$, it follows for any $s > 0$ that

$$\varphi(\beta, s) \leq 2(128)^{u_0} \leq 31/15.$$

Considering the possibilities if $\tau_C \leq \tau_2$ or if $\tau_C > \tau_2$, it now follows by the strong Markov property that

$$\varphi_0(u_0\alpha_1, s) \; \leq \; 31\varphi_2(u_0\alpha_1, s)/15. \tag{3.52}$$

We next consider what happens if the process starts in B_2. If $dK_{3.6} \geq 3n\delta/4$, then $B_2 \subset B_3$, and so $\varphi_2(u_0\alpha_1, s) \leq \varphi_3(u_0\alpha_1, s)$. If not, note that

$$\begin{aligned}
&\mathbb{E}_{X_1,X_2}\{e^{\beta(\tau_C \wedge s)}\} \\
&\leq \; \mathbb{E}_{X_1,X_2}\big\{e^{\beta(\tau_C\wedge s)}1\!\!1\{\tau_3 \leq (\tau_0 \wedge s)\}\big\} + \mathbb{E}_{X_1,X_2}\big\{e^{\beta(\tau_C\wedge s)}1\!\!1\{\tau_0 \leq (\tau_3 \wedge s)\}\big\} \\
&\qquad + \mathbb{E}_{X_1,X_2}\big\{e^{\beta(\tau_C\wedge s)}1\!\!1\{(\tau_0 \wedge \tau_3) > s\}\big\} \\
&\leq \; \mathbb{E}_{X_1,X_2}\big\{e^{\beta\tau_3}1\!\!1\{\tau_3 \leq (\tau_0 \wedge s)\}\big\}\varphi_3(\beta, s) \\
&\qquad + e^{\beta s}\{\mathbb{P}_{X_1,X_2}[\tau_0 \leq (\tau_3 \wedge s)] + \mathbb{P}_{X_1,X_2}[(\tau_0 \wedge \tau_3) > s]\}\varphi_0(\beta, s),
\end{aligned} \tag{3.53}$$

with the last inequality following from the strong Markov property. Now, in view of Lemma 3.6, if we define

$$M_1(t) \; := \; e^{\alpha_1 t/2} h_2(X^\delta_{n,1}(t), X^\delta_{n,2}(t)), \tag{3.54}$$

then $M_1(t \wedge \tau_3 \wedge \tau_0)$ is an $\mathcal{F}^{X^\delta_{n,1},X^\delta_{n,2}}$-supermartingale. This implies that

$$d^2K^2_{3.6}\,\mathbb{E}_{X_1,X_2}\big\{e^{\alpha_1\tau_3/2}1\!\!1\{\tau_3 \leq (\tau_0 \wedge s)\}\big\} \; \leq \; h_2(X_1, X_2), \tag{3.55}$$

and also that

$$d^2K^2_{3.6}\,e^{\alpha_1 s/2}\mathbb{P}_{X_1,X_2}[(\tau_3 \wedge \tau_0) > s] \; \leq \; h_2(X_1, X_2). \tag{3.56}$$

Thence, by Jensen's inequality and (3.55), for any $0 \leq u \leq 1$, we also have

$$\mathbb{E}_{X_1,X_2}\big\{e^{u\alpha_1\tau_3/2}1\!\!1\{\tau_3 \leq (\tau_0 \wedge s)\}\big\} \; \leq \; \{h_2(X_1, X_2)/d^2K^2_{3.6}\}^u. \tag{3.57}$$

Finally, from Lemma 2.5 with $\eta = \delta/2$, for $(X_1, X_2) \in B_2$ and for θ_1 as in Lemma 2.2,

$$\mathbb{P}_{X_1,X_2}[\tau_0 \leq (\tau_3 \wedge s)] \; \leq \; \mathbb{P}_{X_1,X_2}[\tau_0 \leq s] \; \leq \; 2\exp\{-7n\theta_1\delta^2/64\}, \tag{3.58}$$

if $s \leq n/\alpha_1$ and $n \geq \max\{n_2, \psi^{-1}(\delta/2)\}$, where

$$n_2 \; := \; \max\left\{n_{(3.2)}, \frac{K_{2.5}\overline{\Lambda}}{\alpha_1}\right\} \; \in \; \mathcal{K};$$

this follows because $n_{2.2} \leq n_{(3.2)}$, and because, for n and s chosen in this way, $\exp\{n^{-1}\theta_1(3n\delta/8)^2\} \geq n^9 \geq nK_{2.5}\Lambda s$, because $\delta \geq 2\psi(n)$ and $n \geq d^4$.

Hence, taking $\beta = u_0\alpha_1/2$, with $u_0 = 1/32$ as above, substituting (3.56), (3.57) and (3.58) into (3.53), and then using (3.52), we have

$$\begin{aligned}\varphi_2(u_0\alpha_1/2, s) &\le (2n\delta/dK_{3.6})^{2u_0}\varphi_3(u_0\alpha_1/2, s) + P(u_0, s, n)\varphi_0(u_0\alpha_1/2, s)\\ &\le (2n\delta/dK_{3.6})^{2u_0}\varphi_3(u_0\alpha_1/2, s) + 31P(u_0, s, n)\varphi_2(u_0\alpha_1/2, s)/15,\end{aligned}$$

where

$$\begin{aligned}P(u, s, n) &:= e^{u\alpha_1 s/2}\{\mathbb{P}_{X_1,X_2}[\tau_0 \le (\tau_3 \wedge s)] + \mathbb{P}_{X_1,X_2}[(\tau_3 \wedge \tau_0) > s]\}\\ &\le e^{u\alpha_1 s/2}\{2\exp\{-7n^{3/4}d\theta_1\delta^2/64\} + (2n\delta/dK_{3.6})^{2u_0}e^{-\alpha_1 s/2}\},\end{aligned}$$

recalling $n \ge d^4$ for the final inequality. It thus follows that $P(u_0, s_n, n) \le 15/62$ for all $n \ge \max\{n_2, d^4, \psi^{-1}(\delta/2)\}$ such that

$$\left(\frac{2n\delta}{dK_{3.6}}\right)^{2u_0} e^{-\alpha_1(1-u_0)s_n/2} \le \frac{7}{62} \quad \text{and} \quad e^{-(7n^{3/4}d\theta_1\delta^2/64 - u_0\alpha_1 s_n/2)} \le \frac{2}{31}. \tag{3.59}$$

Picking $s = s_n := 64\log n/\alpha_1$, and recalling that $u_0 = 1/32$ and $2\psi(n) \le \delta \le \delta_1/3 \le 1 \le K_{3.6}$, it is enough that $n \ge n_3$, where

$$n_3 := \max\left\{\left[2^{1/16}\left(\frac{62}{7}\right)\right]^{1/30}, \left(\frac{31}{2}\right)^{1/6}\right\} \in \mathcal{K}.$$

Hence, for $n \ge \max\{n_2, n_3, \psi^{-1}(\delta/2)\}$,

$$\varphi_2(u_0\alpha_1/2, s_n) \le 2(2n\delta/dK_{3.6})^{2u_0}\varphi_3(u_0\alpha_1/2, s_n). \tag{3.60}$$

For $(X_1, X_2) \in B_3$, we take $t_3 = 1/\alpha_1$, and use (3.49) to conclude that

$$\varphi_3(u\alpha_1/2, s_n) \le e^{u/2}\{1 + d^{3/2}C_* n^{-1/2}(\Lambda/g_*)^{1/4}\,\varphi_0(u\alpha_1/2, s_n)\}, \tag{3.61}$$

for C_* as defined in (3.50), if $n \ge n_{(3.2)}$. From (3.52) and (3.60), we have

$$\varphi_0(u_0\alpha_1/2, s_n) \le 62(2n\delta/dK_{3.6})^{2u_0}\varphi_3(u_0\alpha_1/2, s_n)/15, \tag{3.62}$$

for all $n \ge \max\{d^4, n_2, n_3, \psi^{-1}(\delta/2)\}$. Taking $u = u_0 = 1/32$ in (3.61) and using (3.62), the coefficient of $\varphi_3(u_0\alpha_1/2, s_n)$ on the right hand side of (3.61) is at most

$$e^{1/64}d^{3/2}C_* n^{-1/2}(\Lambda/g_*)^{1/4}62(2n\delta/dK_{3.6})^{1/16}/15 \le 1/2$$

if, using $n \ge d^4$ and $\delta \le \delta_1/3 \le 1 \le K_{3.6}$,

$$n \ge n_4 := e\,2^4\left(\frac{124C_*}{15}\right)^{64}\left(\frac{\overline{\Lambda}}{g_*}\right)^{16} \in \mathcal{K}.$$

Hence, from (3.61) and (3.62), for $n \geq \max\{\max_{2\leq l\leq 4} n_l, \psi^{-1}(\delta/2)\}$, we have

$$\varphi_3(u_0\alpha_1/2, s_n) \;\leq\; 2e^{1/64}.$$

Combining this with (3.62) and the definition (3.51) of φ_0, it follows that, for all $(X_1, X_2) \in B_0$, we have

$$\mathbb{P}_{X_1,X_2}[\tau_C > t] \;\leq\; 9(2n)^{2u_0} e^{-u_0\alpha_1 t/2}, \quad 0 \leq t \leq s_n,$$

for $n \geq \max\{n_{(3.2)}, \max_{1\leq l\leq 4} n_l, \psi^{-1}(\delta/2)\}$. In particular,

$$\mathbb{P}_{X_1,X_2}[\tau_C > s_n] \;\leq\; 9(2n)^{2u_0} e^{-u_0\alpha_1 s_n/2}.$$

However, by the strong Markov property, for $t > s_n$,

$$\mathbb{P}_{X_1,X_2}[\tau_C > t] \;\leq\; \mathbb{P}_{X_1,X_2}[\tau_C > s_n] \max_{(X_1,X_2)\in B_0} \mathbb{P}_{X_1,X_2}[\tau_C > t - s_n].$$

Arguing inductively, it follows that, for $r \in \mathbb{Z}_+$ and $0 \leq v < s_n$,

$$\mathbb{P}_{X_1,X_2}[\tau_C > rs_n + v] \;\leq\; \bigl(9(2n)^{2u_0} e^{-u_0\alpha_1 s_n/2}\bigr)^r 9(2n)^{2u_0} e^{-u_0\alpha_1 v/2}.$$

Take n so large that $9(2n)^{2u_0} e^{-u_0\alpha_1 s_n/4} \leq 1$; this is true, for $u_0 = 1/32$ and $s_n = 64\log n/\alpha_1$, if $n \geq n_5 := 9^4 2^{1/4}$. Then, for all $(X_1, X_2) \in B_0$, we have

$$\mathbb{P}_{X_1,X_2}[\tau_C > rs_n + v] \;\leq\; 9(2n)^{2u_0} e^{-u_0\alpha_1(rs_n+v)/4}.$$

Since $u_0 = 1/32$, the inequality in the theorem is thus proved for $n \geq \max\{n_{3.7}, \psi^{-1}(\delta/2)\}$, where

$$n_{3.7} \;:=\; \max_{1\leq l\leq 5} n_l \in \mathcal{K}, \tag{3.63}$$

with $\alpha_1/128$ for α_2. □

4 Stein's method based on X_n^δ

4.1 Bounding the solutions of the Stein equation

We now use the results of the previous section to bound the first and second differences of the solutions $h_B := h^\delta_{B,n}$ of the Stein equation corresponding to the generator $\mathcal{A}^\delta_n$ defined in (1.14), for the elementary processes satisfying Assumptions G0, G1 and S2–S4. We recall from the introduction the definitions

$$\|\Delta f(X)\|_\infty \;:=\; \max_{1\leq j\leq d} |\Delta_j f(X)|; \quad \|\Delta^2 f(X)\|_\infty \;:=\; \max_{1\leq j,k\leq d} |\Delta_{jk} f(X)|, \tag{4.1}$$

where Δ_j and Δ_{jk}, as defined in (1.8), denote the components of the first and second difference operators Δ and Δ^2, respectively.

Theorem 4.1. *Let X_n be an elementary process, satisfying Assumptions G0, G1 and S2–S4 for some $\delta_0 > 0$. Let $\delta_1 := \min\{3, \delta_0/\sqrt{\lambda_{\max}(\Sigma)}\}$. Then there are constants $\kappa_0, \kappa_1, \kappa_2 \in \mathcal{K}$ such that, for any $B \subset \mathbb{Z}^d$ and any $\delta \le \delta_1/3$, the solution $h_B := h^\delta_{B,n}$ of the Stein equation*

$$\mathcal{A}^\delta_n h_B(X) \;=\; \mathbb{1}_B(X) - \Pi^\delta_n\{B\}$$

satisfies

$$\begin{aligned} |h_B(X)| \;&\le\; \alpha_1^{-1}\kappa_0 \log n; \quad \|\Delta h_B(X)\|_\infty \;\le\; \alpha_1^{-1}\kappa_1 d^{1/4} n^{-1/2}\log n; \\ \|\Delta^2 h_B(X)\|_\infty \;&\le\; \alpha_1^{-1}\kappa_2 d^{1/2} n^{-1}\log n, \end{aligned}$$

for all $|X - nc|_\Sigma \le n\delta/4$ and $n \ge \max\{n_{3.7}, \psi^{-1}(\delta/2)\}$. The constants $\kappa_0, \kappa_1, \kappa_2$ are given in (4.3), (4.5) and (4.6), respectively, and α_1 is as in (2.4) .

Proof. The argument starts from the explicit representation of h_B given in (1.17), which immediately yields

$$\begin{aligned} h_B(X) \;&=\; -\int_0^\infty (\mathbb{P}[X^\delta_n(t) \in B \mid X^\delta_n(0) = X] - \Pi^\delta_n\{B\})\,dt \\ &=\; -\sum_{Y \in \widetilde{B}_{n,\delta}(c)} \int_0^\infty (\mathbb{P}[X^\delta_n(t) \in B \mid X^\delta_n(0) = X] \\ &\qquad\qquad - \mathbb{P}[X^\delta_n(t) \in B \mid X^\delta_n(0) = Y])\Pi^\delta_n(Y)\,dt. \end{aligned} \tag{4.2}$$

Because $\delta \le \delta_1/3$ and $9 \cdot 2^{1/16} \le 10$, we can use Theorem 3.7 for $n \ge \max\{n_{3.7}, \psi^{-1}(\delta/2)\}$ to give

$$\begin{aligned} |h_B(X)| \;&\le\; 2\alpha_2^{-1}\log n + \int_{2\alpha_2^{-1}\log n}^\infty 10 n^{1/16} e^{-\alpha_2 t}\,dt \\ &\le\; 2\alpha_2^{-1}\log n + 10\alpha_2^{-1} n^{-1}, \end{aligned}$$

with $\alpha_2 = \alpha_1/128$, proving the bound on $|h_B(X)|$, with

$$\kappa_0 \;:=\; 1536. \tag{4.3}$$

Next, from (4.2), we have

$$\begin{aligned} &h_B(X - e^{(j)}) - h_B(X) \\ &\quad= \int_0^\infty (\mathbb{P}[X^\delta_n(t) \in B \mid X^\delta_n(0) = X] - \mathbb{P}[X^\delta_n(t) \in B \mid X^\delta_n(0) = X - e^{(j)}])\,dt. \end{aligned} \tag{4.4}$$

Taking $f(X) := \mathbb{1}_B(X)$, Theorem 3.3, with $G^{(j)}$ bounded by Λ, implies that

$$\int_0^{2\alpha_2^{-1}\log n} |\mathbb{P}[X_n^\delta(t) \in B \,|\, X_n^\delta(0) = X] - \mathbb{P}[X_n^\delta(t) \in B \,|\, X_n^\delta(0) = X - e^{(j)}]|\, dt$$
$$\leq\; K_{3.3}^j \alpha_2^{-1} n^{-1/2}\{2\log n + 2\alpha_2(\Lambda g^{(j)})^{-1/2}\} \left(\frac{\Lambda}{g^{(j)}}\right)^{1/4},$$

if $|X - nc|_\Sigma \leq n\delta/4$ and $n \geq \max\{n_{(3.2)}, \psi^{-1}(\delta/2)\}$; in performing the integration, the range is split at $t = (\Lambda g^{(j)})^{-1/2}$. The remainder of the integral is bounded by $10\alpha_2^{-1}n^{-1}$, as above, if also $n \geq \max\{n_{3.7}, \psi^{-1}(\delta/2)\}$, since $n_{3.7} \geq n_{(3.2)}$, completing the bound on $|h(X - e^{(j)}) - h(X)|$, and thence on $\|\Delta h(X)\|_\infty$, with

$$\kappa_1 \;:=\; 128\left\{10 + K_*^{(1)}\left(2 + \frac{2\alpha_2}{\sqrt{\Lambda g_*}}\right)\right\}\left(\frac{\Lambda}{g_*}\right)^{1/4} \;\in\; \mathcal{K}, \tag{4.5}$$

where $K_*^{(1)} := \max_{1\leq j\leq d} K_{3.3}^j$.

For the second differences, the argument is entirely similar, using Theorem 3.5 for the bulk of the estimate, and bounding the integrand by 2 for $0 \leq t \leq n^{-1}\big(\Lambda g_{ij}^+\big)^{-1/2}$. This gives the bound on $\|\Delta^2 h(X)\|_\infty$, with

$$\kappa_2 \;:=\; 128\left\{10 + \frac{2\alpha_2}{\sqrt{\Lambda g_*}} + K_*^{(2)}\left(\frac{\Lambda}{g_*}\right)^{1/2}\left(2 + \frac{\alpha_2}{\sqrt{\Lambda g_*}}\right)\right\} \;\in\; \mathcal{K}, \tag{4.6}$$

where $K_*^{(2)} := \max_{1\leq i,j\leq d} K_{3.5}^{ji}$. □

Remark 4.2. As in Remark 3.4, the dependence on d, appearing through the factors $d^{1/4}$ and $d^{1/2}$ in the bounds on $\|\Delta h_B(X)\|_\infty$ and $\|\Delta^2 h_B(X)\|_\infty$, respectively, is not needed if $g_*^{-1}\max_{1\leq j\leq d}\{G^{(j)} + g^{(j)}\}$ remains bounded as $n \to \infty$. This is the case if $A = -\lambda I$, for some $\lambda > 0$, and if (X_n) is as in Theorem 3.2.

4.2 Reducing the generator

In this section, we show that we can work with the simpler operator $\widetilde{\mathcal{A}}_n$, defined in (1.7), in place of the generator $\mathcal{A}_n^\delta$. As a first step in the reduction, we use two technical lemmas to bound the expectation of the Newton remainder

$$e_2(W, J, h) \;:=\; h(W + J) - h(W) - \Delta h(W)^T J - \tfrac{1}{2} J^T \Delta^2 h(W) J, \tag{4.7}$$

for W a random vector. For $\eta > 0$ and $f\colon \mathbb{Z}^d \to \mathbb{R}$, we use the notation

$$\|f\|^{\Sigma}_{n\eta,\infty} \;:=\; \max_{|X-nc|_\Sigma \le n\eta} |f(X)|, \tag{4.8}$$

analogous to that of (2.2), but using $|\cdot|_\Sigma$-balls, with nc implicit. Similarly, we write $\|\Delta h\|^{\Sigma}_{n\eta,\infty}$ and $\|\Delta^2 h\|^{\Sigma}_{n\eta,\infty}$ for $\|f\|^{\Sigma}_{n\eta,\infty}$, when $f(X) = \|\Delta h(X)\|_\infty$ and $f(X) = \|\Delta^2 h(X)\|_\infty$, respectively, so that the conclusion of Theorem 4.1 can be expressed as

$$\begin{aligned} \|h_B\|^{\Sigma}_{n\delta/4,\infty} \;&\le\; \alpha_1^{-1}\kappa_0 \log n; \quad \|\Delta h_B\|^{\Sigma}_{n\delta/4,\infty} \;\le\; \alpha_1^{-1}\kappa_1 d^{1/4} n^{-1/2}\log n; \\ \|\Delta^2 h_B\|^{\Sigma}_{n\delta/4,\infty} \;&\le\; \alpha_1^{-1}\kappa_2 d^{1/2} n^{-1}\log n, \end{aligned} \tag{4.9}$$

for $n \ge \max\{n_{3.7}, \psi^{-1}(\delta/2)\}$ and $\delta \le \delta_1$.

Our control over the differences of the functions h_B is only such that we can bound their $\|\cdot\|^{\Sigma}_{n\eta,\infty}$ norms for suitable η, so these are the quantities that we need in our estimates. For instance, if W is a random vector in $\mathbb{Z}^d$ such that $d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \;\le\; \varepsilon$, we immediately have the bound

$$|\mathbb{E}\{f(W + e^{(j)}) - f(W)\}| \;\le\; 2\varepsilon\|f\|_\infty.$$

Because we often only have control within certain (large) balls, we are led instead to bounding a truncated quantity

$$|\mathbb{E}\{(f(W + e^{(j)}) - f(W)) I[|W - nc|_\Sigma \le n\eta_1]\}|$$

in terms of $\|f\|^{\Sigma}_{n\eta_2,\infty}$, for suitable choices of η_1 and η_2. The following lemma, proved in Section 6.1, provides what we need.

Lemma 4.3. *Suppose that W is a random vector in $\mathbb{Z}^d$ such that*

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \;\le\; \varepsilon_1,\; 1 \le j \le d; \quad \mathbb{P}[|W - nc|_\Sigma > n\delta/4] \;\le\; \varepsilon_2.$$

Then, for $f\colon \mathbb{Z}^d \to \mathbb{R}$ and $1 \le j \le d$, and for any $U, X \in \mathbb{Z}^d$ such that $\max\{|X|_\Sigma, |X + U|_\Sigma\} \le n\delta/6$ *and* $n\delta \ge 12|U|_\Sigma$,

$$\begin{aligned} &|\mathbb{E}\{(f(W + X + U) - f(W + X)) I[|W - nc|_\Sigma \le n\delta/3]\}| \\ &\quad\le\; (\varepsilon_1|U|_1 + \varepsilon_2)\|f\|^{\Sigma}_{n\delta/2,\infty}. \end{aligned}$$

We use Lemma 4.3 to bound the Newton remainder defined in (4.7). Instead of bounding $e_2(W, J, h)$ directly, we bound a perturbation of it,

$$E_2(W, J, h) \;:=\; e_2(W, J, h) + \frac{1}{2}\sum_{j=1}^{d} J_j \Delta_{jj} h(W), \tag{4.10}$$

which can be represented as a sum of third differences of h. The result, proved in Section 6.2, is as follows.

Lemma 4.4. *If W is a random vector in $\mathbb{Z}^d$ such that*

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \;\le\; \varepsilon_1,\; 1 \le j \le d; \quad \mathbb{P}[|W - nc|_\Sigma \ge n\delta/4] \;\le\; \varepsilon_2,$$

then, for any $J \in \mathbb{Z}^d$ such that $|J|_\Sigma \le n\delta/12$,

$$(i) \qquad \Big|\mathbb{E}\Big\{E_2(W, J, h) I[|W - nc|_\Sigma \le n\delta/3]\Big\}\Big| \\ \le\; \|\Delta^2 h\|^{\Sigma}_{n\delta/2,\infty}(C^{(1)}_{4.4}(J)\varepsilon_1 + C^{(2)}_{4.4}(J)\varepsilon_2),$$

where

$$C^{(1)}_{4.4}(J) \;:=\; \tfrac{1}{6}|J|_1(|J|_1 + 1)(|J|_1 + 2); \quad C^{(2)}_{4.4}(J) \;:=\; \tfrac{1}{2}|J|_1(|J|_1 + 1). \quad (4.11)$$

If the conditions are replaced by $\mathbb{P}[|W - nc| \ge n\delta/4] \;\le\; \varepsilon^E_2$ and $|J| \le n\delta/12$, then

$$(ii) \qquad \Big|\mathbb{E}\Big\{E_2(W, J, h) I[|W - nc| \le n\delta/3]\Big\}\Big| \\ \le\; \|\Delta^2 h\|_{n\delta/2,\infty}(C^{(1)}_{4.4}(J)\varepsilon_1 + C^{(2)}_{4.4}(J)\varepsilon^E_2),$$

where $\|\Delta^2 h\|_{n\delta/2,\infty}$ is as defined in (2.2).

Remark 4.5. Note that, because $0 \ne J \in \mathbb{Z}^d$, we have

$$C^{(1)}_{4.4}(J) \;\le\; |J|^3_1 \;\le\; d^{3/2}|J|^3; \quad C^{(2)}_{4.4}(J) \;\le\; 2d|J|^2.$$

Lemma 4.4 allows us to prove the following reduction theorem, useful for approximating the generator of any Markov jump process satisfying our general assumptions.

Theorem 4.6. *Suppose that $(g^J,\ J \in \mathcal{J})$, c, A, σ^2, γ, δ_0, $\mathcal{A}^\delta_n$ and $\widetilde{\mathcal{A}}_n$ are as in Sections 1 and 2.1, and that Assumptions G0–G4 are satisfied. Suppose that W is a random vector in $\mathbb{Z}^d$, such that, for some $\varepsilon, v > 0$,*

$$\begin{aligned} &\text{(i)} \quad \mathbb{E}|W - nc|^2_\Sigma \;\le\; dvn; \\ &\text{(ii)} \quad d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \;\le\; \varepsilon, \text{ for each } 1 \le j \le d. \end{aligned} \qquad (4.12)$$

Then, for any $\delta \le \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$ and for any $0 < \delta' \le \delta/2$, and for $n \ge \max\{n_{(3.2)}, \psi^{-1}(\delta/2)\}$,

$$|\mathbb{E}\{(\mathcal{A}^\delta_n h(W) - \widetilde{\mathcal{A}}_n h(W)) I[|W - nc|_\Sigma \le n\delta'/3]\}| \\ \le\; d^{5/2}\overline{\Lambda}\Big(\tfrac{1}{2}L_2\lambda_{\max}(\Sigma)v\|\Delta h\|^{\Sigma}_{n\delta/4,\infty} \qquad (4.13) \\ + n\|\Delta^2 h\|^{\Sigma}_{n\delta/4,\infty}\big\{L_1\sqrt{v\lambda_{\max}(\Sigma)}n^{-1/2} + d^{1/2}(\bar{\gamma}/\overline{\Lambda})\varepsilon + 32d^{1/2}v/\{n(\delta')^2\}\big\}\Big),$$

where $\|\Delta h\|^{\Sigma}_{n\eta,\infty}$ and $\|\Delta^2 h\|^{\Sigma}_{n\eta,\infty}$ are bounded in (4.9).

Proof. Consider

$$\begin{aligned}
\mathcal{A}_n^\delta h(X) &:= n\sum_{J\in\mathcal{J}} g_\delta^J(X/n)\{h(X+J)-h(X)\} \qquad (4.14)\\
&= n\sum_{J\in\mathcal{J}}\{g^J(c)+Dg^J(c)^T n^{-1}(X-nc)+e_1(X,J,g_\delta^J)\}\\
&\qquad\times\{\Delta h(X)^T J+\tfrac{1}{2}J^T\Delta^2 h(X)J+e_2(X,J,h)\},
\end{aligned}$$

where

$$e_1(X,J,g_\delta^J) := g_\delta^J(X/n)-g^J(c)-n^{-1}Dg^J(c)^T(X-nc),$$

and e_2 is as in (4.7). Observing that $\sum_{J\in\mathcal{J}} g^J(c)\Delta h(X)^T J=\Delta h(X)^T F(c)=0$, because $F(c)=0$, and that

$$\begin{aligned}
\sum_{J\in\mathcal{J}} g^J(c)J^T\Delta^2 h(X)J &= \mathrm{Tr}\,\{\sigma^2\Delta^2 h(X)\};\\
\sum_{J\in\mathcal{J}} Dg^J(c)^T n^{-1}(X-nc)\Delta h(X)^T J &= n^{-1}\mathrm{Tr}\,\{A(X-nc)\Delta h(X)^T\},
\end{aligned}$$

it follows, once again writing $I_n^\eta(W):=I[|W-nc|_\Sigma\le n\eta/3]$, that

$$\begin{aligned}
&|\mathbb{E}\{(\mathcal{A}_n^\delta h(W)-\widetilde{\mathcal{A}}_n h(W))I_n^{\delta'}(W)\}|\\
&\quad\le n\mathbb{E}\left\{\sum_{J\in\mathcal{J}}|e_1(W,J,g_\delta^J)|\,|h(W+J)-h(W)|I_n^{\delta'}(W)\right\}\\
&\qquad+\mathbb{E}\left\{\sum_{J\in\mathcal{J}}|Dg^J(c)^T(W-nc)|\,|h(W+J)-h(W)-\Delta h(W)^T J|I_n^{\delta'}(W)\right\}\\
&\qquad+n\Big|\mathbb{E}\Big\{\sum_{J\in\mathcal{J}} g^J(c)e_2(W,J,h)I_n^{\delta'}(W)\Big\}\Big|. \qquad (4.15)
\end{aligned}$$

Now, from (2.3), and recalling (1.11), it follows that, for $X\in\mathcal{X}_n^\delta(J)$,

$$\begin{aligned}
|e_1(X,J,g_\delta^J)| &= |g_\delta^J(X/n)-g^J(c)-Dg^J(c)^T n^{-1}(X-nc)|\\
&\le \tfrac{1}{2}n^{-2}|X-nc|^2 L_2 g^J(c), \qquad (4.16)
\end{aligned}$$

provided that $\delta\sqrt{\lambda_{\max}(\Sigma)}\le\delta_0$. Since, for $|X-nc|_\Sigma\le n\delta'/3\le n\delta/6$ and $n\delta/12>J_{\max}^\Sigma$, we have

$$|X+J-nc|_\Sigma \le J_{\max}^\Sigma+|X-nc|_\Sigma \le n\delta/4,$$

it follows, for such X and n, that $|h(X+J)-h(X)| \le |J|_1 \,\|\Delta h\|^{\Sigma}_{n\delta/4,\infty}$ and that (4.16) is satisfied. Hence, since $\mathbb{E}|W-nc|^2_{\Sigma} \le dvn$, we have

$$\begin{aligned} & n\mathbb{E}\left\{ \sum_{J\in\mathcal{J}} |e_1(W,J,g^J_\delta)|\,|h(W+J)-h(W)|\,I^{\delta'}_n(W) \right\} \\ & \le \tfrac12 L_2 \sum_{J\in\mathcal{J}} g^J(c)|J|_1 \,\|\Delta h\|^{\Sigma}_{n\delta/4,\infty} n^{-1}\mathbb{E}|W-nc|^2 \\ & \le \tfrac12 L_2 \sum_{J\in\mathcal{J}} g^J(c)|J|_1 \,\|\Delta h\|^{\Sigma}_{n\delta/4,\infty} dv\lambda_{\max}(\Sigma) \\ & \le \tfrac12 L_2 v d^{5/2}\overline{\Lambda}\lambda_{\max}(\Sigma)\,\|\Delta h\|^{\Sigma}_{n\delta/4,\infty}, \end{aligned}$$

if $n\delta/12 > J^{\Sigma}_{\max}$, and this condition is satisfied for $n \ge \max\{n_{(3.2)}, \psi^{-1}(\delta/2)\}$.

Then, from (6.20) and (2.3), if $n \ge n_{(3.2)}$ and $|X-nc|_{\Sigma} \le n\delta'/3$,

$$\begin{aligned} & \sum_{J\in\mathcal{J}} |Dg^J(c)^T(X-nc)|\,|h(X+J)-h(X)-\Delta h(X)^T J|\,I^{\delta'}_n(X) \\ & \le \tfrac12 L_1|X-nc|\,\|\Delta^2 h\|^{\Sigma}_{n\delta/4,\infty} \sum_{J\in\mathcal{J}} g^J(c)|J|_1(|J|_1+1) \\ & \le d^2\overline{\Lambda}L_1|X-nc|\,\|\Delta^2 h\|^{\Sigma}_{n\delta/4,\infty}; \end{aligned}$$

hence, since $\mathbb{E}|W-nc|_{\Sigma} \le \sqrt{dvn}$,

$$\begin{aligned} & \mathbb{E}\left\{ \sum_{J\in\mathcal{J}} |Dg^J(c)^T(W-nc)|\,|h(W+J)-h(W)-\Delta h(W)^T J|\,I^{\delta'}_n(W) \right\} \\ & \le d^{5/2}\overline{\Lambda}L_1\sqrt{v\lambda_{\max}(\Sigma)n}\,\|\Delta^2 h\|^{\Sigma}_{n\delta/4,\infty}. \end{aligned}$$

Finally, from Lemma 4.4 and Chebyshev's inequality,

$$\begin{aligned} & n\left|\mathbb{E}\left\{ \sum_{J\in\mathcal{J}} g^J(c)\left(e_2(W,J,h) + \tfrac12 \sum_{j=1}^d J_j\Delta_{jj}h(W) \right) I^{\delta'}_n(W) \right\}\right| \\ & \le \sum_{J\in\mathcal{J}} g^J(c) n\|\Delta^2 h\|^{\Sigma}_{n\delta/4,\infty}(C^{(1)}_{4.4}(J)\varepsilon + 16C^{(2)}_{4.4}(J)dv/\{n(\delta')^2\}) \\ & \le n\|\Delta^2 h\|^{\Sigma}_{n\delta/4,\infty}(d^3\bar{\gamma}\varepsilon + 32d^3\overline{\Lambda}v/\{n(\delta')^2\}), \end{aligned}$$

and, for each j, $\sum_{J\in\mathcal{J}} J_j g^J(c) = 0$, because $F(c)=0$. This completes the proof of the theorem. □

Remark 4.7. Suppose that processes $(X_n,\, n \ge 0)$ are *almost* density dependent, in that they have rates as in (1.1), but with $g^J(x)$ being replaced by $g^J(x,n)$, where $\lim_{n\to\infty} n^{1/2}\varepsilon_n^\delta = 0$ for all $\delta > 0$ small enough, where $\varepsilon_n^\delta := \sup_{|x-c|\le\delta} \sum_{J\in\mathcal{J}} |g^J(x,n) - g^J(x)|$. Then it is immediate from (4.14) that, if Assumptions G0–G4 hold for the transition rates $g^J(x)$, then the conclusion of Theorem 4.6 continues to hold, with the limiting definitions of c, A and σ^2, provided only that an asymptotically small term

$$d\overline{\Lambda}\{n^{1/2}\varepsilon_n^\delta\}\, n^{1/2}\|\Delta h\|^{\Sigma}_{n\delta/4,\infty}$$

is added to the bound given in (4.13). In particular, if $n\varepsilon_n^\delta$ is bounded as $n \to \infty$, the asymptotic order of the error bound is not increased.

4.3 Total variation approximation

We are now in a position to prove Theorem 4.8, which gives a measure of the error in the approximation of the distribution of a random vector W in $\mathbb{Z}^d$ by the distribution Π_n^δ, if the process X_n^δ satisfies the special assumptions of Section 3. The statement of Theorem 4.8 is to some extent complicated by the presence of the indicator truncating the range of W in the main condition (iii). The truncation is necessary, because Theorem 4.1 only enables one to bound the differences of the functions h_B solving the Stein equation (1.15) in balls of radius $n\delta/4$, for any $\delta \le \delta_0/3\sqrt{\lambda_{\max}(\Sigma)}$.

Theorem 4.8. *Given any $c \in \mathbb{R}^d$ and $d\times d$ matrices A and σ^2, with A having eigenvalues all with negative real parts, and σ^2 being positive definite, there exists a sequence of elementary processes $(X_n,\, n \ge 1)$, given in Theorem 3.2, satisfying Assumptions G0, G1 and S2–S4 for $\delta_0 = \lambda_{\min}(\sigma^2)/(8\|A\|) > 0$, having $F(c) = 0$, $DF(c) = A$ and σ^2 given by (1.4). Let Σ be as in (1.9). Define $\tilde\delta_0 := \min\{1, \delta_0/3\sqrt{\lambda_{\max}(\Sigma)}\}$ and $\overline{\Lambda} := \bar\lambda(\sigma^2)$, and suppose that $\delta' \le \tilde\delta_0/2$. Then, for any $v > 0$, there exists a constant $C_{4.8}(v,\delta')$, which is a function of v, δ', $\|A\|/\overline{\Lambda}$ and the elements of $\mathrm{Sp}'(\sigma^2/\overline{\Lambda})$ and $\mathrm{Sp}'(\Sigma)$, with the following property: if W is any random vector in $\mathbb{Z}^d$ such that, for some $n \ge \max\{n_{3.7}, \psi^{-1}(\delta')\}$ and for some $\varepsilon_1, \varepsilon_{20}, \varepsilon_{21}, \varepsilon_{22} > 0$,*

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

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

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

where $\widetilde{\mathcal{A}}_n$ is as defined in (1.7), then, for any δ such that $2\delta' \le \delta \le \tilde\delta_0$,

$$d_{\mathrm{TV}}(\mathcal{L}(W), \Pi_n^\delta) \;\le\; C_{4.8}(v,\delta')(d^3 n^{-1/2} + d^4\varepsilon_1 + \varepsilon_{20} + d^{1/4}\varepsilon_{21} + d^{1/2}\varepsilon_{22})\log n.$$

Proof. From (1.16), we have

$$d_{\rm TV}(\mathcal{L}(W),\Pi_n^\delta) \\ \le \sup_{B\subset \widetilde B_{n,\delta}(c)} |\mathbb{E}\{\mathcal{A}_n^\delta h_B(W) I[W\in \widetilde B_{n,\delta'/3}(c)]\}| + \mathbb{P}[W\notin \widetilde B_{n,\delta'/3}(c)],$$

where $\widetilde B_{n,\eta}(c) := \mathbb{Z}^d \cap B_{n\eta,\Sigma}(nc)$ and $h_B := h^\delta_{B,n}$ is as for (1.15). The probability in the second term is at most $9dv/\{n(\delta')^2\}$, by (i) and Chebyshev's inequality. Then, for $2\delta' \le \delta \le \tilde\delta_0$ and for $n \ge \max\{n_{3.7}, \psi^{-1}(\delta')\}$, we can use (4.9) in (iii), giving

$$|\mathbb{E}\{\widetilde{\mathcal{A}}_n h_B(W)\} I[|W-nc|_\Sigma \le n\delta'/3]| \\ \le \bar\lambda(\sigma^2)(\varepsilon_{20}\|h_B\|^\Sigma_{n\tilde\delta_0/4,\infty} + \varepsilon_{21} n^{1/2}\|\Delta h_B\|^\Sigma_{n\tilde\delta_0/4,\infty} + \varepsilon_{22} n \|\Delta^2 h_B\|^\Sigma_{n\tilde\delta_0/4,\infty}) \\ \le (\bar\lambda(\sigma^2)/\alpha_1)(\varepsilon_{20}\kappa_0 + \varepsilon_{21}\kappa_1 d^{1/4} + \varepsilon_{22}\kappa_2 d^{1/2})\log n.$$

Finally, from Theorem 4.6 and (4.9) and recalling that $L_2 = 0$ for elementary processes, for $n \ge \max\{n_{3.7}, \psi^{-1}(\delta')\}$ and $2\delta' \le \delta \le \tilde\delta_0$, we have

$$|\mathbb{E}\{(\mathcal{A}_n^\delta h_B(W) - \widetilde{\mathcal{A}}_n h_B(W)) I[W\in \widetilde B_{n,\delta'/3}(c)]\}| \\ \le d^3 \log n\,\frac{\bar\lambda(\sigma^2)}{\alpha_1}\left\{\frac{\kappa_2 L_1\sqrt{v\lambda_{\max}(\Sigma)}}{\sqrt n} + \frac{32\kappa_2 d^{1/2} v}{n(\delta')^2} + \kappa_2 d^{1/2}(\bar\gamma/\overline\Lambda)\,\varepsilon_1\right\},$$

and noting that $\bar\gamma \le d^{-1/2}\overline\Lambda J_{\max}$ completes the proof of the theorem. □

5 Application: approximating a Markov jump process

Suppose that $(X_n,\, n\ge 1)$ is a fixed sequence of Markov jump processes with $X_n(\cdot) \in n^{-1}\mathbb{Z}^d$ for some fixed d, and with transition rates determined by the fixed collection of functions $(g^J : \mathbb{R}^d \to \mathbb{R}_+,\ J\in\mathcal{J})$, satisfying Assumptions G0–G4 for some c and δ_0. Then, for large n, X_n has a quasi-equilibrium behaviour near nc, in the sense that the process, if started near nc, remains within any ball $\widetilde B_{n,\delta}(c)$ for a length of time whose expectation, for fixed $\delta > 0$, grows exponentially with n. During this time, its behaviour is asymptotically extremely close to that of X_n^δ, the choice of $\delta < \delta_0$ having almost no effect: see Barbour & Pollett (2012, Section 4). Thus it has a quasi-equilibrium distribution that, as $n\to\infty$, is asymptotically extremely close to Π_n^δ, for any $0<\delta<\delta_0$. Theorem 5.3 below shows that Π_n^δ, in turn, can be closely approximated by the equilibrium distribution $\widehat\Pi_n^\delta$ of an elementary process.

We begin by noting that the variance of Π_n^δ is of the correct order, satisfying Condition (i) of Theorem 4.8, and that Condition (ii) is also satisfied, with $\varepsilon_1 = O(n^{-1/2})$. The proofs of these two results are in Section 6.3. Since, for this application, all the data of the problem, apart from n, are fixed, we can simplify many statements to order expressions as $n \to \infty$.

Lemma 5.1. *Let X_n be a Markov jump process whose transition rates are given in (1.1), satisfying Assumptions G0–G4 for some $\delta_0 > 0$, and let $\delta_{2.2}$ and $\delta'_{2.2}(d)$ be as in Lemma 2.2. Then, for any $0 < \delta \le \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, if $X_n^\delta \sim \Pi_n^\delta$, we have*

$$\mathbb{E}|X_n^\delta - nc|_\Sigma^2 \;=\; O(n).$$

Proposition 5.2. *Under Assumptions G0–G4, if $X_n^\delta \sim \Pi_n^\delta$ for some fixed $\delta \le \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, then, for each $1 \le j \le d$,*

$$d_{\mathrm{TV}}(\Pi_n^\delta, \Pi_n^\delta * \varepsilon_{e(j)}) \;=\; O(n^{-1/2}),$$

where ε_J denotes the point mass at J and $*$ denotes convolution.

We now give the approximation theorem.

Theorem 5.3. *Under the above assumptions on X_n, there is a sequence of elementary processes $\widehat{X}_n$ such that, for any fixed $0 < \delta < \min\{\delta_0/\sqrt{\lambda_{\max}(\Sigma)}, \tilde{\delta}_0\}$, $\widehat{X}_n^\delta$ has equilibrium distribution $\widehat{\Pi}_n^\delta$ satisfying*

$$d_{\mathrm{TV}}(\Pi_n^\delta, \widehat{\Pi}_n^\delta) \;=\; O(n^{-1/2}\log n),$$

as $n \to \infty$, where δ_0 is as in Assumptions G0–G4 for X_n, and $\tilde{\delta}_0$ is as in Theorem 4.8 for $\widehat{X}_n$.

Proof. We apply Theorem 4.8, using the elementary process $\widehat{X}_n$, given in Theorem 3.2, that shares the same c, A and σ^2, and hence the same $\widetilde{\mathcal{A}}_n$, as X_n. We show that, for $W \sim \Pi_n^\delta$, Conditions (i)–(iii) of Theorem 4.8 are satisfied, with suitable choices of v, ε_1 and ε_{2l}, $0 \le l \le 2$.

Condition (i) follows immediately for some $v > 0$ from Lemma 5.1, and Condition (ii) is implied by Proposition 5.2, with $\varepsilon_1 = O(n^{-1/2})$. It then follows from Theorem 4.6 with $\delta' = \delta/2$ that, for any function h, we have

$$\begin{aligned}|\mathbb{E}\{(\mathcal{A}_n^\delta h(W) - \widetilde{\mathcal{A}}_n h(W))I[|W - nc|_\Sigma \le n\delta/6]\}| \\ \;=\; O\big(n^{-1/2}(n^{1/2}\|\Delta h\|_{n\delta/4,\infty}^\Sigma + n\|\Delta^2 h\|_{n\delta/4,\infty}^\Sigma)\big).\end{aligned} \tag{5.17}$$

However, since Π_n^δ is the equilibrium distribution of X_n^δ, it follows that $\mathbb{E}\{\mathcal{A}_n^\delta h(W)\} = 0$. Then, since $|W - nc|_\Sigma \le n\delta$ implies that $|W - nc| \le \delta_0$, because $\delta < \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$, it follows that

$$
\begin{aligned}
&|\mathbb{E}\{\mathcal{A}_n^\delta h(W) I[|W - nc|_\Sigma > n\delta/6]\}| \\
&\quad \le\; n \sum_{J \in \mathcal{J}} |g^J|_{\delta_0} \mathbb{E}\{|h(W+J) - h(W)| I[|W - nc|_\Sigma > n\delta/6]\} \\
&\quad \le\; 2nL_0 \Lambda \|h\|_{n\delta_0,\infty} \mathbb{P}[|W - nc|_\Sigma > n\delta/6], \qquad (5.18)
\end{aligned}
$$

where L_0, Λ and $|\cdot|_\delta$ are as in Section 2.1. From Lemma 6.1, with $r = 2$, $\mathbb{P}[|W - nc|_\Sigma > n\delta/6] = O(n^{-2})$ as $n \to \infty$. Then, since the left hand side of (5.17) is unchanged if we set $h(X) = 0$ for $|X - nc|_\Sigma > n\delta/4$, provided that $J^\Sigma_{\max} \le n\delta/12$, we can replace $\|h\|_{n\delta_0,\infty}$ by $\|h\|^\Sigma_{n\delta/4,\infty}$ in (5.18) for all n sufficiently large. These two observations imply, with (5.18), that

$$
|\mathbb{E}\{\mathcal{A}_n^\delta h(W) I[|W - nc|_\Sigma \le n\delta/6]\}| \;=\; O(n^{-1}\|h\|^\Sigma_{n\delta/4,\infty}).
$$

Combining this with (5.17), it follows that

$$
\begin{aligned}
&|\mathbb{E}\{\widetilde{\mathcal{A}}_n h(W) I[|W - nc|_\Sigma \le n\delta/6]\}| \\
&\quad =\; O\bigl(n^{-1/2}(\|h\|^\Sigma_{n\delta/4,\infty} + n^{1/2}\|\Delta h\|^\Sigma_{n\delta/4,\infty} + n\|\Delta^2 h\|^\Sigma_{n\delta/4,\infty})\bigr),
\end{aligned}
$$

which in turn implies that Condition (iii) of Theorem 4.8 is satisfied, with $\varepsilon_{20}, \varepsilon_{21}$ and ε_{22} all of order $O(n^{-1/2})$, proving the result. □

It is shown in Part II, Theorem 2.3, that the equilibrium distributions of elementary processes are at distance $O(n^{-1/2}\log n)$ in total variation from discrete normal distributions. The theorem above thus extends this to the quasi-equilibrium distributions of very general Markov jump processes. However, other equally explicit approximations may be available. For example, consider the bivariate immigration–death process X_n with immigration rates $n\alpha_1$ for a single type 1 individual, $n\alpha_2$ for a single type 2 individual, and $n\alpha_{12}$ for a pair with one of each type. Assume that individuals have independent exponentially distributed lifetimes, with mean $1/\mu_i$ for type i. This process has equilibrium distribution $\Pi_n := \mathcal{L}(N_1 + N_3, N_2 + N_3)$, where N_1, N_2 and N_3 are independent Poisson random variables with means

$$
\mathbb{E}N_1 \;=\; \frac{n}{\mu_1}\left(\alpha_1 + \frac{\alpha_{12}\mu_2}{\mu_1+\mu_2}\right); \quad \mathbb{E}N_2 \;=\; \frac{n}{\mu_2}\left(\alpha_2 + \frac{\alpha_{12}\mu_1}{\mu_1+\mu_2}\right); \quad \mathbb{E}N_3 \;=\; \frac{n\alpha_{12}}{\mu_1+\mu_2}.
$$

It is then easy to show that, for any $\delta > 0$, $d_{\rm TV}(\Pi_n, \Pi_n(\delta)) = O(n^{-1})$, where $\Pi_n(\delta)$ denotes the equilibrium distribution of the restriction of X_n to an $n\delta$-ball around its mean $n\hat{c}$, given by $\hat{c} := \bigl(\mu_1^{-1}(\alpha_1 + \alpha_{12}), \mu_2^{-1}(\alpha_2 + \alpha_{12})\bigr)^T$. Taking any $a := (a_1, a_2)^T > -\hat{c}$, and then translating X_n by $(\lfloor na_1 \rfloor, \lfloor na_2 \rfloor)^T$

(the integer parts are needed, to stay in $\mathbb{Z}^2$), we obtain an almost density dependent Markov jump process $\widetilde{X}_n$ satisfying Assumptions G0–G4, with $\mathcal{J} := \{(1,0),(0,1),(1,1),(-1,0),(0,-1)\}$ and

$$\begin{aligned} g^{(1,0)}(x,n) &= \alpha_1; \quad g^{(0,1)}(x,n) \;=\; \alpha_2; \quad g^{(1,1)}(x,n) \;=\; \alpha_{12}; \\ g^{(-1,0)}(x,n) &= \mu_1(x_1 - n^{-1}\lfloor na_1 \rfloor); \quad g^{(0,-1)}(x,n) \;=\; \mu_2(x_2 - n^{-1}\lfloor na_2 \rfloor), \end{aligned}$$

having equilibrium distribution $\widetilde{\Pi}_n := \Pi_n * \varepsilon_{(\lfloor na_1 \rfloor, \lfloor na_2 \rfloor)}$.

By Remark 4.7 and Theorem 4.8, $d_{\mathrm{TV}}(\widetilde{\Pi}_n^\delta, \widehat{\Pi}_n^\delta) = O(n^{-1/2}\log n)$, where $\widehat{X}_n$ is the elementary process from Theorem 3.2, having

$$c \;:=\; a + \hat{c}, \quad A \;:=\; \begin{pmatrix} -\mu_1 & 0 \\ 0 & -\mu_2 \end{pmatrix} \quad \text{and} \quad \sigma^2 \;:=\; \begin{pmatrix} 2(\alpha_1 + \alpha_{12}) & \alpha_{12} \\ \alpha_{12} & 2(\alpha_2 + \alpha_{12}) \end{pmatrix}.$$

Using Theorem 4.8 again, it follows that any other Markov jump process X_n' satisfying Assumptions G0–G4 with the same c, A and σ^2, restricted to any $n\delta$-ball around nc, has an equilibrium distribution at distance of order $O(n^{-1/2}\log n)$ from $\widetilde{\Pi}_n^\delta$, and hence also from $\widetilde{\Pi}_n$. This covers a wide range of processes, but by no means all. For instance, it is easy to check that $0 \le \mathrm{Corr}\,(X_n^1, X_n^2) \le 1/2$ for all positive choices of the birth and death rates.

6 Technicalities

6.1 Proof of Lemma 4.3

In order to bound $|\mathbb{E}\{(f(W+X+U) - f(W+X))I[|W-nc|_\Sigma \le n\delta/3]\}|$, we write

$$\begin{aligned} &(f(W+X+U) - f(W+X))I[|W-nc|_\Sigma \le n\delta/3] \\ &\quad = \tilde{f}_X(W+U) - \tilde{f}_X(W) \\ &\qquad + f(W+X+U)\{I[|W-nc|_\Sigma \le n\delta/3] - I[|W+U-nc|_\Sigma \le n\delta/3]\}, \end{aligned}$$

where $\tilde{f}_X(Y) := f(Y+X)I[|Y-nc|_\Sigma \le n\delta/3]$. Since $|\tilde{f}_X(Y)| \le \|f\|_{n\delta/2,\infty}^\Sigma$ if $|X|_\Sigma \le n\delta/6$, it is immediate that

$$|\mathbb{E}\{\tilde{f}_X(W+U) - \tilde{f}_X(W)\}| \;\le\; \varepsilon_1 |U|_1 \|f\|_{n\delta/2,\infty}^\Sigma.$$

Then, on the set

$$\{|Y-nc|_\Sigma \le n\delta/3 < |Y+U-nc|_\Sigma\} \cup \{|Y+U-nc|_\Sigma \le n\delta/3 < |Y-nc|_\Sigma\},$$

it follows that $|Y - nc|_\Sigma > n\delta/3 - |U|_\Sigma \geq n\delta/4$ if $n\delta \geq 12|U|_\Sigma$, and that

$$|Y + X + U - nc|_\Sigma \;\leq\; n\delta/3 + \max\{|X|_\Sigma, |X+U|_\Sigma\} \;\leq\; n\delta/2,$$

this last by assumption. Hence

$$\begin{aligned}&|\mathbb{E}\{f(W+X+U)(I[|W-nc|_\Sigma \leq n\delta/3] - I[|W+U-nc|_\Sigma \leq n\delta/3])\}| \\ &\quad\leq\; \varepsilon_2 \|f\|^{\Sigma}_{n\delta/2,\infty},\end{aligned}$$

and Lemma 4.3 is proved. □

6.2 Proof of Lemma 4.4

We prove only part (i), showing that, if $E_2(W, J, h)$ is as in (4.10), then

$$\begin{aligned}&\Big|\mathbb{E}\Big\{E_2(W,J,h)I[|W-nc|_\Sigma \leq n\delta/3]\Big\}\Big| \\ &\quad\leq\; \|\Delta^2 h\|^{\Sigma}_{n\delta/2,\infty}(C^{(1)}_{4.4}(J)\varepsilon_1 + C^{(2)}_{4.4}(J)\varepsilon_2),\end{aligned}$$

for constants $C^{(1)}_{4.4}(J), C^{(2)}_{4.4}(J)$ given in (4.11), whenever

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathcal{L}(W + e^{(j)})) \;\leq\; \varepsilon_1,\; 1 \leq j \leq d; \quad \mathbb{P}[|W - nc|_\Sigma \geq n\delta/4] \;\leq\; \varepsilon_2;$$

the proof of part (ii) is entirely similar. We begin by taking any function $f\colon \mathbb{Z}^d \to \mathbb{R}$ and any $k \in \mathbb{Z}$ and $1 \leq j \leq d$. First, we note that, for $X \in \mathbb{Z}^d$,

$$f(X+ke^{(j)}) - f(X) \;=\; \begin{cases} \sum_{l=1}^{k} \Delta_j f(X + (l-1)e^{(j)}) & \text{if } k \geq 1; \\ -\sum_{l=1}^{|k|} \Delta_j f(X - le^{(j)}) & \text{if } k \leq -1. \end{cases} \tag{6.19}$$

Hence, by considering positive and negative k separately, we find that

$$|f(X + ke^{(j)}) - f(X) - k\Delta_j f(X)| \;\leq\; \tfrac{1}{2}|k|(|k|+1)\|\Delta^2 f\|_\infty.$$

For more general increments $J \in \mathbb{Z}^d$, we define

$$J^{(s)} \;:=\; (J_1, J_2, \ldots, J_s, 0, 0, \ldots, 0), \quad s \geq 1; \qquad J^{(0)} \;:=\; (0, \ldots, 0).$$

Then, from the inequalities above, we have

$$|f(X+J^{(s)}) - f(X+J^{(s-1)}) - J_s\Delta_s f(X + J^{(s-1)})| \;\leq\; \tfrac{1}{2}|J_s|(|J_s|+1)\|\Delta^2 f\|_\infty,$$

and

$$|\Delta_s f(X + J^{(s-1)}) - \Delta_s f(X)| \;\leq\; |J^{(s-1)}|_1 \|\Delta^2 f\|_\infty.$$

Hence it follows that

$$|f(X+J^{(s)}) - f(X+J^{(s-1)}) - J_s\Delta_s f(X)| \\ \le \left\{\tfrac{1}{2}|J_s|(|J_s|+1) + |J_s||J^{(s-1)}|_1\right\} \|\Delta^2 f\|_\infty.$$

Adding over $1 \le s \le d$, this gives

$$|f(X+J) - f(X) - Df(X)^T J| \le \tfrac{1}{2}|J|_1(|J|_1+1)\|\Delta^2 f\|_\infty.$$

The same argument also shows that

$$|f(X+J) - f(X) - Df(X)^T J| \le \tfrac{1}{2}|J|_1(|J|_1+1)\|\Delta^2 f\|^{\Sigma}_{n\delta/2,\infty}, \quad (6.20)$$

if $|X-nc|_\Sigma \le n\delta/3$ and $|J|_\Sigma \le n\delta/6$.

We now prove Part (i) of the lemma by induction on the number r of non-zero components of J. We write $I^\eta_n(X)$ as shorthand for $I[|X-nc|_\Sigma \le n\eta/3]$, for any $\eta > 0$. Starting with $r=1$, we consider three cases. For $J = ke^{(j)}$ and $k \ge 1$, we have

$$\begin{aligned} & h(X+ke^{(j)}) - h(X) - k\Delta_j h(X) - \tfrac{1}{2}k(k-1)\Delta_{jj}h(X) \\ &= \sum_{l=1}^{k-1}\{\Delta_j h(X+le^{(j)}) - \Delta_j h(X)\} - \tfrac{1}{2}k(k-1)\Delta_{jj}h(X) \\ &= \sum_{l=1}^{k-1}\sum_{r=1}^{l-1}\{\Delta_{jj}h(X+re^{(j)}) - \Delta_{jj}h(X)\}. \end{aligned} \quad (6.21)$$

From Lemma 4.3, with $X=0$ and $U = re^{(j)}$, it follows that

$$|\mathbb{E}\{(\Delta_{jj}h(W+re^{(j)}) - \Delta_{jj}h(W))\, I^\delta_n(W)\}| \le (r\varepsilon_1+\varepsilon_2)\|\Delta^2 h\|^{\Sigma}_{n\delta/2,\infty} \quad (6.22)$$

for $r \le k-2$, if $|ke^{(j)}|_\Sigma = |J|_1|e^{(j)}|_\Sigma \le n\delta/12$. Multiplying (6.21) by $I^\delta_n(X)$, replacing X by W, then taking expectations, invoking (6.22), and adding, this yields the claim for $J = ke^{(j)}$ and $k \ge 1$, with the upper bounds $C^{(1)}_{4.4}(ke^{(j)}) = \frac{1}{6}(k-2)(k-1)k$ and $C^{(2)}_{4.4}(ke^{(j)}) = \frac{1}{2}(k-1)k$. If $J = ke^{(j)}$ and $k=0$, there is nothing to prove. For $J = -ke^{(j)}$ and $k \ge 1$, we have

$$\begin{aligned} & h(X-ke^{(j)}) - h(X) - (-k)\Delta_j h(X) - \tfrac{1}{2}(-k)(-k-1)\Delta_{jj}h(X) \\ &= \sum_{l=1}^{k}\{\Delta_j h(X) - \Delta_j h(X-le^{(j)})\} - \tfrac{1}{2}k(k+1)\Delta_{jj}h(X) \\ &= \sum_{l=1}^{k}\sum_{r=1}^{l}\{\Delta_{jj}h(X-re^{(j)}) - \Delta_{jj}h(X)\}. \end{aligned} \quad (6.23)$$

Arguing as before yields the claim for $J = -ke^{(j)}$ and $k \geq 1$, with

$$C_{4.4}^{(1)}(-ke^{(j)}) \;=\; \tfrac{1}{6}k(k+1)(k+2); \qquad C_{4.4}^{(2)}(-ke^{(j)}) \;=\; \tfrac{1}{2}k(k+1),$$

again if $|ke^{(j)}|_\Sigma = |J|_1|e^{(j)}|_\Sigma \leq n\delta/12$. This establishes that the inequality (i) is true for $r = 1$, when J has just one non-zero component.

Now, for any $2 \leq r \leq d$, we assume that (i) is true for all J with at most $r-1$ non-zero components, and show that this implies that (i) is also true for all J with at most r non-zero components. Without loss of generality, we consider any J with $J_j = 0$ for $r < j \leq d$. First, we write

$$h(X+J) - h(X) \;=\; \{h(X+J) - h(X+J^{(r-1)})\} + \{h(X+J^{(r-1)}) - h(X)\}.$$

The induction hypothesis gives

$$\begin{aligned}&\Big|\mathbb{E}\Big\{\Big(e_2(W, J^{(r-1)}, h) + \tfrac{1}{2}\sum_{j=1}^{r-1} J_j \Delta_{jj} h(W)\Big)\, I_n^\delta(W)\Big\}\Big| \\ &\quad\leq\; \|\Delta^2 h\|_{n\delta/2,\infty}^\Sigma (C_{4.4}^{(1)}(J^{(r-1)})\varepsilon_1 + C_{4.4}^{(2)}(J^{(r-1)})\varepsilon_2).\end{aligned}$$

Thus it remains only to consider the expectation of the quantity

$$\begin{aligned}&\Big(h(W+J) - h(W+J^{(r-1)}) - J_r\Delta_r h(W) \\ &\quad - \tfrac{1}{2}J_r^2\Delta_{rr}h(W) - J_r\sum_{j=1}^{r-1} J_j\Delta_{rj}h(W) + \tfrac{1}{2}J_r\Delta_{rr}h(W)\Big)\, I_n^\delta(W).\end{aligned}$$

The one dimensional result gives

$$\begin{aligned}&\big|\mathbb{E}\big\{\big(h(W+J) - h(W+J^{(r-1)}) - J_r\Delta_r h(W+J^{(r-1)}) \\ &\qquad - \tfrac{1}{2}J_r(J_r-1)\Delta_{rr}h(W+J^{(r-1)})\big)\, I_n^\delta(W)\big\}\big| \\ &\quad\leq\; \tfrac{1}{6}|J_r|(|J_r|+1)\|\Delta^2 h\|_{n\delta/2,\infty}^\Sigma((|J_r|+2)\varepsilon_1 + 3\varepsilon_2),\end{aligned}$$

leaving an expectation involving the expression

$$\begin{aligned}&J_r\left\{\Delta_r h(X+J^{(r-1)}) - \Delta_r h(X) - \sum_{j=1}^{r-1} J_j\Delta_{rj}h(X)\right\} \\ &\qquad + \tfrac{1}{2}J_r(J_r-1)\{\Delta_{rr}h(X+J^{(r-1)}) - \Delta_{rr}h(X)\}. \qquad (6.24)\end{aligned}$$

The first line in (6.24) can be expressed as

$$J_r\sum_{s=1}^{r-1}\{\Delta_r h(X+J^{(s)}) - \Delta_r(X+J^{(s-1)}) - J_s\Delta_{rs}h(X)\}. \qquad (6.25)$$

From (6.19) with $f := \Delta_d h$, we have

$$\Delta_r h(X + J^{(s)}) - \Delta_r h(X + J^{(s-1)}) - J_s \Delta_{rs} h(X)$$
$$= \begin{cases} \sum_{l=1}^{J_s} \{\Delta_{rs} h(X + J^{(s-1)} + (l-1)e^{(s)}) - \Delta_{rs} h(X)\}, & \text{if } J_s \geq 1; \\ 0, & \text{if } J_s = 0; \\ -\sum_{l=1}^{|J_s|} \{\Delta_{rs} h(X + J^{(s-1)} - le^{(s)}) - \Delta_{rs} h(X)\}, & \text{if } J_s \leq -1. \end{cases}$$

Hence, multiplying by $I_n^\delta(X)$, taking expectations with W in place of X and using Lemma 4.3, it follows for the first line in (6.24) that

$$\begin{aligned} &|J_r| \left| \mathbb{E}\left\{ \sum_{s=1}^{r-1} \{\Delta_r h(W + J^{(s)}) - \Delta_r (W + J^{(s-1)}) - J_s \Delta_{rs} h(W)\} \, I_n^\delta(W) \right\} \right| \\ &\leq \; |J_r| \sum_{s=1}^{r-1} |J_s| \|\Delta^2 h\|_{n\delta/2,\infty}^\Sigma \{(|J^{(s-1)}|_1 + \tfrac{1}{2}|J_s|)\varepsilon_1 + \varepsilon_2\} \\ &\leq \; |J_r| |J^{(r-1)}|_1 \|\Delta^2 h\|_{n\delta/2,\infty}^\Sigma (\tfrac{1}{2}|J^{(r-1)}|_1 \varepsilon_1 + \varepsilon_2). \end{aligned}$$

The second line in (6.24) is directly bounded using Lemma 4.3, giving

$$\begin{aligned} &\tfrac{1}{2}|J_r|(|J_r|+1) \big| \mathbb{E}\{(\Delta_{rr} h(W + J^{(r-1)}) - \Delta_{rr} h(W)) \, I_n^\delta(W)\} \big| \\ &\quad \leq \; \tfrac{1}{2}|J_r|(|J_r|+1) \|\Delta^2 h\|_{n\delta/2,\infty}^\Sigma (|J^{(r-1)}|_1 \varepsilon_1 + \varepsilon_2). \end{aligned}$$

This establishes the inequality (i) for J with $J_j = 0$ for $r < j \leq d$, since it is easily checked that

$$C_{4.4}^{(1)}(J) \;\geq\; C_{4.4}^{(1)}(J^{(r-1)}) + \tfrac{1}{2}|J_r| \, |J^{(r-1)}|_1^2 + \tfrac{1}{2}|J_r|(|J_r|+1) \, |J^{(r-1)}|_1,$$

and that

$$C_{4.4}^{(2)}(J) \;\geq\; C_{4.4}^{(2)}(J^{(r-1)}) + |J_r| \, |J^{(r-1)}|_1 + \tfrac{1}{2}|J_r|(|J_r|+1).$$

The lemma now follows by induction. □

6.3 Proofs of Lemma 5.1 and Proposition 5.2

To prove Lemma 5.1, we need to show that, if $X_n^\delta \sim \Pi_n^\delta$ for some $\delta \leq \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$, then $\mathbb{E}\{|X_n^\delta - nc|_\Sigma^2\} = O(n)$. Now, by Dynkin's formula, we have $\mathbb{E}\{\mathcal{A}_n^\delta h(X_n^\delta)\} = 0$ for any choice of h. Take $h(X) = h_0(X) = |X - nc|_\Sigma^2$ as in Lemma 2.2, for which $\mathcal{A}_n^\delta h_0(X) \leq -\alpha_1 h_0(X)$ in $|X - nc|_\Sigma \geq K_{2.2}\sqrt{nd}$.

Then, from (2.9), for all $|X - nc|_\Sigma \le n\,\min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, $\mathcal{A}_n^\delta h_0(X) \le ndK'_1$, where, from (2.11), and (2.13),

$$K'_1 \;:=\; L_0 d^{-1}\mathrm{Tr}\,(\sigma_\Sigma^2) \le L_0\lambda_{\max}(\sigma_\Sigma^2) \;\le\; L_0\alpha_1\rho(\sigma^2)\rho(\Sigma).$$

This implies that

$$0 \;=\; \mathbb{E}\{\mathcal{A}_n^\delta h_0(X_n^\delta)\} \;\le\; -\alpha_1\mathbb{E}\{|X_n^\delta - nc|_\Sigma^2 I[|X - nc|_\Sigma \ge K_{2.2}\sqrt{nd}]\} + ndK'_1.$$

Since also, using (2.12),

$$\mathbb{E}\{|X_n^\delta - nc|_\Sigma^2 I[|X - nc|_\Sigma < K_{2.2}\sqrt{nd}]\} \;\le\; ndK_{2.2}^2 \;\le\; 4ndL_0\rho(\sigma^2)\rho(\Sigma),$$

the claim is proved. □

To prove Proposition 5.2, we start with a concentration bound, used to handle the truncation.

Lemma 6.1. *Define $K_\Sigma := 2\overline{\Lambda}K_{2.5}/(d\theta_1\alpha_1) \in \mathcal{K}$. Under Assumptions G0–G4, for any $0 < \delta \le \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, $n \ge n_{2.2}$ and $\eta > K_{2.2}\sqrt{d/n}$,*

$$\Pi_n^\delta\{|X - nc|_\Sigma > n\eta\} \;\le\; \eta^{-2}d^2K_\Sigma e^{-n\theta_1\eta^2}.$$

In particular, for any fixed $\eta > 0$, $\Pi_n^\delta\{|X - nc|_\Sigma > n\eta\} = O(n^{-r})$ as $n \to \infty$, for any $r \ge 1$.

Proof. Again, for $\delta \le \delta_0/\sqrt{\lambda_{\max}(\Sigma)}$ and $X_n^\delta \sim \Pi_n^\delta$, we have $\mathbb{E}\{\mathcal{A}_n^\delta h(X_n^\delta)\} = 0$ for any choice of h, by Dynkin's formula. Take $h(X) = h_{\theta_1}(X)$ as in Lemma 2.2, for which, from (2.18) and for $n \ge n_{2.2}$,

$$\mathcal{A}_n^\delta h_{\theta_1}(X) \;\le\; -\tfrac{1}{2}n^{-1}\alpha_1\theta_1 h_0(X)h_{\theta_1}(X) \quad \text{for} \quad |X - nc|_\Sigma \ge K_{2.2}\sqrt{nd}.$$

This, with (2.21), implies that

$$\tfrac{1}{2}n^{-1}\alpha_1\theta_1\mathbb{E}\big\{h_0(X)h_{\theta_1}(X)I\big[|X - nc|_\Sigma \ge K_{2.2}\sqrt{nd}\big]\big\} \;\le\; n\Lambda K_{2.5}.$$

Hence, for $\eta > K_{2.2}\sqrt{d/n}$, it follows that

$$\tfrac{1}{2}n^{-1}\alpha_1\theta_1\,(n\eta)^2 e^{n^{-1}\theta_1(n\eta)^2}\Pi_n^\delta\{|X - nc|_\Sigma > n\eta\} \;\le\; n\Lambda K_{2.5},$$

proving the first part. The second is then immediate. □

The proof of the next lemma is rather close to that of Theorem 3.3, so we only give a quick sketch.

Lemma 6.2. *Under Assumptions G0–G3, for any fixed* $\delta \le \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$, *and for any* $J \in \mathcal{J}$,

$$d_{TV}\{\Pi_n^\delta, \Pi_n^\delta * \varepsilon_J\} \;=\; O(n^{-1/2})$$

as $n \to \infty$, *where* ε_J *denotes the point mass at* J *and* $*$ *denotes convolution.*

Remark 6.3. Note that we cannot directly replace J by $e^{(j)}$ here, to obtain Proposition 5.2, because $e^{(j)}$ may not belong to $\mathcal{J}$. Under Assumption G4, we can do so: see (6.30) below.

Proof. Fix any $U > 0$, and use the stationarity of Π_n^δ to give the inequality

$$d_{TV}\{\Pi_n^\delta, \Pi_n^\delta * \varepsilon_J\} \;\le\; \sum_{X \in \mathbb{Z}^d} \Pi_n^\delta(X)\, d_{TV}\{\mathcal{L}_X(X_n^\delta(U)), \mathcal{L}_X(X_n^\delta(U) + J)\}, \tag{6.26}$$

where $\mathcal{L}_X$ denotes distribution conditional on $\{X_n^\delta(0) = X\}$. By Lemma 5.1, it then follows that, for any $\delta \le \min\{\delta_{2.2}, \delta'_{2.2}(d)\}$,

$$d_{TV}\{\Pi_n^\delta, \Pi_n^\delta * \varepsilon_J\} \;=\; D_{Jn}(\delta/2) + O(n^{-1}), \tag{6.27}$$

where

$$D_{Jn}(\delta') \;:=\; \sum_{X:\, |X - nc|_\Sigma \le n\delta'} \Pi_n^\delta(X)\, d_{TV}\{\mathcal{L}_X(X_n^\delta(U)), \mathcal{L}_X(X_n^\delta(U) + J)\}.$$

This alters our problem to one of finding a bound of similar form, but now involving the transition probabilities of the process X_n^δ over a finite time U, started in any state X which is reasonably close to nc.

By Assumption G3, J-jumps occur in X_n^δ with rate at least $n\mu_0^J$, whenever it is in the set $\mathcal{X}_n^\delta(J)$. Thus, by analogy with (3.8), we can realize the chain X_n^δ with $X_n^\delta(0) = X_0$ in the form $X_n^\delta(u) := X_0 + JN_n^\delta(u) + W_n^\delta(u)$, where the transition $(l, W) \to (l+1, W)$ occurs at rate $n\mu_0^J$. This leads to a decomposition

$$\begin{aligned}
&d_{TV}\{\mathcal{L}_X(X_n^\delta(U)), \mathcal{L}_X(X_n^\delta(U) + J)\} \\
&\quad\le\; \frac{1}{2} \sum_{l \ge 0} |\mathbb{P}_{X_0}[N_n^\delta(U) = l] - \mathbb{P}_{X_0}[N_n^\delta(U) = l-1]| \qquad (6.28) \\
&\qquad + \frac{1}{2} \sum_{X \in \mathbb{Z}^d} \sum_{l \ge 1} \mathbb{P}_{X_0}[N_n^\delta(U) = l-1] |q_{l,X_0}^U(X - lJ) - q_{l-1,X_0}^U(X - lJ)|,
\end{aligned}$$

where $q_{l,X}^U(W)$ is as defined in (3.10).

Much as for (3.13), and using Lemma 2.5, the first sum in (6.28) is bounded by

$$\mathbb{P}_{X_0}[\hat\tau_n^\delta \le U] + \{n\mu_0^J U\}^{-1/2} \;=\; O(n^{-1/2}), \qquad (6.29)$$

if we choose $U = 1/\sqrt{\Lambda\mu_0^J}$, where $\hat\tau_n^\delta$ is as defined in (3.12). For the second part of (6.28), we argue as in the proof of Theorem 3.3, using the Radon–Nikodym derivative $R^U_{(\mathbf{s}_{l-1},s_*),X}(u,w^u) := d\mathbb{P}^U_{(\mathbf{s}_{l-1},s_*),X}/d\mathbb{P}^U_{\mathbf{s}_{l-1},X}(w^u)$. As long as $R^U_{(\mathbf{s}_{l-1},s_*),X}(u,w^u) \le 2$ and $u \le \hat\tau_\delta$, the $\mathbb{P}^U_{\mathbf{s}_{l-1},X}$-martingale $R^U_{(\mathbf{s}_{l-1},s_*),X}(u,w^u)$ makes jumps of size at most $2|J|L_1/(n\varepsilon_0)$, and this enables the quadratic variation of the stopped martingale to be controlled by $n^{-1}|J|^2 4L_0(L_1/\varepsilon_0)^2\Lambda u$, as in (3.25). Choosing $U = 1/\sqrt{\Lambda\mu_0^J}$ once more, and arguing as for (3.27) and (3.29), the second part in (6.28) is also shown to be of order $O(n^{-1/2})$. Combining these observations with (6.27), the lemma follows. □

To deduce Proposition 5.2 from Lemma 6.2, take any $1 \le j \le d$. Then it is immediate from the triangle inequality that, because $\sum_{l=1}^{r(j)} J_l^{(j)} = e^{(j)}$ for $J_1^{(j)},\ldots,J_{r(j)}^{(j)}$ as given in Assumption G4, we also have

$$d_{\mathrm{TV}}(\Pi_n^\delta, \Pi_n^\delta * \varepsilon_{e^{(j)}}) \;\le\; \sum_{l=1}^{r(j)} d_{\mathrm{TV}}\bigl(\Pi_n^\delta, \Pi_n^\delta * \varepsilon_{J_l^{(j)}}\bigr) \;=\; O(n^{-1/2}). \qquad (6.30)$$

Remark 6.4. Replacing $e^{(j)}$ by any $J \in \mathcal{J}$ in the statement of Theorem 3.3, the corresponding conclusion can be established, by adapting the proof much as for Lemma 6.2 above, for sequences of Markov jump processes satisfying Assumptions G0–G4. In the bounds, $K_{3.3}^J$ depends on J through a factor of $|J|$, and $G^{(j)}$ is replaced by Λ.

References

[1] A. D. Barbour (1988) Stein's method and Poisson process convergence. *J. Appl. Probab.* **25(A)**, 175–184.

[2] A. D. Barbour, L. Holst & S. Janson (1992) *Poisson Approximation.* Oxford Univ. Press.

[3] A.D. Barbour, M. J. Luczak & A. Xia (2016) Multivariate approximation in total variation, II: discrete normal approximation. `arXiv:***.***`

[4] A. D. Barbour & P. K. Pollett (2012) Total variation approximation for quasi-equilibrium distributions, II. *Stoch. Procs. Applics* **122**, 3740–3756.

[5] A. D. Barbour & A. Xia (1999) Poisson perturbations. *ESAIM: P&S* **3**, 131–150.

[6] L. H. Y. Chen (1975) Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–545.

[7] X. Fang (2014) Discretized normal approximation by Stein's method. *Bernoulli* **20**, 1404–1431.

[8] J. G. Kemeny & J. L. Snell (1960) *Finite Markov chains.* van Nostrand, Princeton NJ.

[9] J. G. Kemeny & J. L. Snell (1961) Finite continuous time Markov chains. *Teor. Verojatnost. i Primenen.* **6**, 101–105.

[10] H. K. Khalil (2002) *Nonlinear systems.* 3rd Edn, Prentice Hall, New Jersey.

[11] E. L. Presman (1983) On the approximation of binomial distributions by means of infinitely divisible ones. *Teor. Verojatnost. i Primenen.* **28**, 393–403.

[12] G. O. Roberts & J. S. Rosenthal (1996) Quantitative bounds for convergence rates of continuous time Markov processes. *Elec. J. Probab.* **1**, Paper no. 9, 21pp.

[13] A. Röllin (2005) Approximation of sums of conditionally independent random variables by the translated Poisson distribution. *Bernoulli* **11**, 1115–1128.

[14] A. Röllin (2007) Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.* **17**, 1596–1614.

[15] S. N. Socoll & A. D. Barbour (2010) Translated Poisson approximation to equilibrium distributions of Markov population processes. *Meth. Comp. Appl. Probab.* **12**, 567–586.

[16] J. A. Tropp (2015) Integer factorization of a positive-definite matrix. *SIAM J. Discrete Math.* **29**, 1783–1791.