A. Proof of Theorem 1

Recall that $\mathbf{E}(t)$ is defined as in Eq. (7), and consider the representation Eq. (18). We begin by decomposing the first term $e^{-s\mathbf{Q}}\mathbf{J}e^{s\mathbf{Q}}$ inside of the integral in Eq. (18) as

$$\begin{aligned} &e^{-s\mathbf{Q}}\mathbf{J}e^{s\mathbf{Q}} \\ &\quad= e^{-s\mathbf{Q}}\mathbf{Q}^+\mathbf{Q}\mathbf{J}\mathbf{Q}\mathbf{Q}^+e^{s\mathbf{Q}} + e^{-s\mathbf{Q}}(\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}\mathbf{Q}^+e^{s\mathbf{Q}} \\ &\qquad + e^{-s\mathbf{Q}}\mathbf{Q}^+\mathbf{Q}\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+)e^{s\mathbf{Q}} \\ &\qquad + e^{-s\mathbf{Q}}(\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+)e^{s\mathbf{Q}} \\ &\quad= e^{-s\mathbf{Q}}\mathbf{Q}^+\mathbf{Q}\mathbf{J}\mathbf{Q}\mathbf{Q}^+e^{s\mathbf{Q}} + (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}\mathbf{Q}^+e^{s\mathbf{Q}} \\ &\qquad + e^{-s\mathbf{Q}}\mathbf{Q}^+\mathbf{Q}\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) + (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+), \end{aligned} \quad [52]$$

where we used Eq. (21) for the second identity. Regarding the middle two terms, with our commutativity Eq. (22) assumption, we observe that

$$\begin{aligned} &(\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}\mathbf{Q}^+e^{s\mathbf{Q}} + e^{-s\mathbf{Q}}\mathbf{Q}^+\mathbf{Q}\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) \\ &\quad= (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}^+\mathbf{Q}e^{s\mathbf{Q}} + e^{-s\mathbf{Q}}\mathbf{Q}\mathbf{Q}^+\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) \\ &\quad= \frac{d}{ds}\left((\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}^+e^{s\mathbf{Q}} - e^{-s\mathbf{Q}}\mathbf{Q}^+\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+)\right). \end{aligned} \quad [53]$$

Hence, integrating in time, and then using Eq. (20), Eq. (22) for the first and last term terms on the right hand side, we have

$$\begin{aligned} &\int_0^t \left\{(\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}\mathbf{Q}^+e^{s\mathbf{Q}} + e^{-s\mathbf{Q}}\mathbf{Q}^+\mathbf{Q}\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+)\right\}ds \\ &\quad= (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}^+e^{t\mathbf{Q}} - e^{-t\mathbf{Q}}\mathbf{Q}^+\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) \\ &\qquad - (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}^+ + \mathbf{Q}^+\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) \\ &\quad= (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}(\mathbf{Q}^+)^2\mathbf{Q}e^{t\mathbf{Q}} - e^{-t\mathbf{Q}}\mathbf{Q}^+\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) \\ &\qquad - (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}^+ + \mathbf{Q}(\mathbf{Q}^+)^2\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) \end{aligned} \quad [54]$$

Finally let us note, regarding the second term on the right hand side of Eq. (18), using again Eq. (21) and Eq. (20), Eq. (22)

$$-te^{t\mathbf{Q}}\mathbf{J} = -te^{t\mathbf{Q}}\mathbf{Q}\mathbf{Q}^+\mathbf{J} - t(\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}. \quad [55]$$

By combining Eq. (52)–Eq. (55) and comparing with Eq. (18), we now find that

$$\begin{aligned} \mathbf{E}(t) = &\int_0^t e^{(t-s)\mathbf{Q}}\mathbf{Q}^+\mathbf{Q}\mathbf{J}\mathbf{Q}\mathbf{Q}^+e^{s\mathbf{Q}}ds + (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}(\mathbf{Q}^+)^2\mathbf{Q}e^{t\mathbf{Q}} \\ &+ e^{t\mathbf{Q}}\mathbf{Q}(\mathbf{Q}^+)^2\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) - te^{t\mathbf{Q}}\mathbf{Q}\mathbf{Q}^+\mathbf{J} \\ &- \mathbf{Q}^+\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) - (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}^+ \\ &- t(\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}\mathbf{Q}^+. \end{aligned} \quad [56]$$

Rearranging Eq. (56), we have

$$\begin{aligned} &\mathbf{Q}^+\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) + (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}^+ + t(\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}\mathbf{Q}^+ \\ &\qquad + \nabla_{\mathbf{J}}e^{t\mathbf{Q}} - te^{t\mathbf{Q}}\mathbf{J} = \mathbf{T}_1 + \mathbf{T}_2, \end{aligned}$$

where

$$\begin{aligned} \mathbf{T}_1 :=& \int_0^t e^{(t-s)\mathbf{Q}}\mathbf{Q}^+\mathbf{Q}\mathbf{J}\mathbf{Q}\mathbf{Q}^+e^{s\mathbf{Q}}ds, \\ \mathbf{T}_2 :=& (\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}(\mathbf{Q}^+)^2\mathbf{Q}e^{t\mathbf{Q}} \\ &+ e^{t\mathbf{Q}}\mathbf{Q}(\mathbf{Q}^+)^2\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+) - te^{t\mathbf{Q}}\mathbf{Q}\mathbf{Q}^+\mathbf{J}. \end{aligned}$$

We estimate each of $\mathbf{T}_1$ and $\mathbf{T}_2$ in turn. Regarding $\mathbf{T}_1$, using once more Eq. (22), as well as Eq. (23), we have

$$\begin{aligned} \|\mathbf{T}_1\| \leq& \int_0^t \|\mathbf{Q}e^{(t-s)\mathbf{Q}}\|\|\mathbf{Q}^+\mathbf{J}\mathbf{Q}^+\|\|\mathbf{Q}e^{s\mathbf{Q}}\|ds \\ \leq& C_0^2\int_0^t e^{-(t-s)\kappa}\|\mathbf{Q}^+\mathbf{J}\mathbf{Q}^+\|e^{-s\kappa}ds \\ =& C_0^2te^{-t\kappa}\|\mathbf{Q}^+\mathbf{J}\mathbf{Q}^+\|. \end{aligned}$$

Turning to $\mathbf{T}_2$, by Eq. (23) we have

$$\begin{aligned} \|\mathbf{T}_2\| \leq C_0(&\|(\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}(\mathbf{Q}^+)^2\| + \|(\mathbf{Q}^+)^2\mathbf{J}(\mathbf{I}-\mathbf{Q}\mathbf{Q}^+)\| \\ &+ t\|\mathbf{Q}^+\mathbf{J}\|)e^{-t\kappa}. \end{aligned}$$

Combining these two bound completes the proof.

B. Proof of Theorem 2

This result is established from Theorem 1 and some basic properties $\mathbf{Q}$ under the given assumptions. First, note that, cf. Eq. (32), (8)

$$|\lambda_d(\mathbf{Q})| \leq \max_{j=1,\dots,d_-} |\lambda_j(\mathbf{Q})| \leq |\lambda_1(\mathbf{Q})|. \quad [57]$$

Referring back to Eq. (10), by Eq. (57) we have the estimate

$$\begin{aligned} \|\mathbf{Q}e^{t\mathbf{Q}}\|_F =& \left(\sum_{j=1}^{d_-}|\lambda_j(\mathbf{Q})e^{t\lambda_j(\mathbf{Q})}|^2\right)^{1/2} \\ \leq& \sqrt{d_-}|\lambda_1(\mathbf{Q})|e^{-t|\lambda_{d_-}(\mathbf{Q})|}. \end{aligned} \quad [58]$$

Similarly, from Eq. (11) and Eq. (57), we have that

$$\begin{aligned} \|\mathbf{Q}e^{t\mathbf{Q}}\|_{\mathrm{op}} \leq& \max_{j=1,\dots,d_-}|\lambda_j(\mathbf{Q})|e^{-t|\lambda_j(\mathbf{Q})|} \\ \leq& |\lambda_1(\mathbf{Q})|e^{-t|\lambda_{d_-}(\mathbf{Q})|}. \end{aligned} \quad [59]$$

Likewise, we observe that

$$\|\mathbf{Q}\|_F^2 = \sum_{k=1}^{d}\lambda_k(\mathbf{Q})^2, \quad \|\mathbf{Q}^+\|_F^2 = \sum_{k=1}^{d_-}\frac{1}{\lambda_k(\mathbf{Q})^2}, \quad [60]$$

and that

$$\|\mathbf{Q}\|_{\mathrm{op}}^2 = \lambda_1(\mathbf{Q})^2, \quad \|\mathbf{Q}^+\|_{\mathrm{op}}^2 = \frac{1}{\lambda_{d_-}(\mathbf{Q})^2}. \quad [61]$$

Let $\mathbf{U}$ be as in Eq. (33). Noting, furthermore, that $\mathbf{I}-\mathbf{Q}\mathbf{Q}^+ = \mathbf{U}(\mathbf{I}-\mathbf{I}_{d_-})\mathbf{U}^*$, where $\mathbf{I}_{d_-}$ is the matrix with 1's along the first d_- diagonal elements and zero otherwise, we have the bounds

$$\|\mathbf{I}-\mathbf{Q}\mathbf{Q}^+\|_F = \sqrt{d-d_-}, \quad \|\mathbf{I}-\mathbf{Q}\mathbf{Q}^+\|_{\mathrm{op}} = 1. \quad [62]$$

Now, for any matrix norm $\|\cdot\|$, Eq. (24) implies that

$$\begin{aligned} &\|t(\mathbf{I}-\mathbf{Q}^+\mathbf{Q})\mathbf{J}\mathbf{Q}\mathbf{Q}^+ + \nabla_{\mathbf{J}}e^{t\mathbf{Q}} - te^{t\mathbf{Q}}\mathbf{J}\| \\ &\qquad \leq C(1+t)e^{-\kappa t} + 2\|\mathbf{Q}^+\|\|\mathbf{J}\|\|\mathbf{I}-\mathbf{Q}\mathbf{Q}^+\| \end{aligned} \quad [63]$$

for any $t \geq 0$, where $C > 0$ is given by Eq. (25). Thus, for the choice of norm $\|\cdot\| = \|\cdot\|_F$, by Eq. (63), Eq. (58), Eq. (61) and Eq. (62), we obtain Eq. (35). Similarly, for the choice of norm $\|\cdot\| = \|\cdot\|_{\mathrm{op}}$, Eq. (63), in conjunction with Eq. (59), Eq. (61) and Eq. (62), yields the bound Eq. (36). Since Eq. (38) follows immediately from Eq. (35), and similarly between Eq. (39) and Eq. (36), the proof is now complete.

C. Proof of Theorem 3

For use in this section, we introduce a probabilistic version of the typical O and o asymptotic notations. Given collections of random variable $\{X_d\}_{d\in\mathbb{N}}$, $\{Y_d\}_{d\in\mathbb{N}}$, we write

$$X_d = O_{\mathrm{a.s.}}(f(d)), \quad Y_d = o_{\mathrm{a.s.}}(f(d)), \quad [64]$$

for some $f : \mathbb{N} \to \mathbb{R}^+$ to mean, respectively, that there exists a random variable $C = C(\omega)$, not dependent on d, such that $|X_d|/f(d) \leq C$ for all $d \in \mathbb{N}$ a.s., and also that $\lim_{d\to\infty} Y_d/f(d) = 0$ a.s.

In everything that follows we will make use of the so-called *Weyl's inequalities* (see e.g. (30, Theorem 4.3.1, p. 239)). Namely, let $\mathbf{A}, \mathbf{B} \in \mathcal{S}(d, \mathbb{C})$, each with its own eigenvalues listed in increasing order as in our convention Eq. (8). Then, for $i = 1, \dots, d$,

$$\lambda_i(\mathbf{A}+\mathbf{B}) \leq \lambda_{i+j}(\mathbf{A}) + \lambda_{d-j}(\mathbf{B}), \quad j = 0, 1, \dots, d-i, \quad [65]$$

and

$$\lambda_{i-j+1}(\mathbf{A}) + \lambda_j(\mathbf{B}) \leq \lambda_i(\mathbf{A}+\mathbf{B}), \quad j = 1, \dots, i. \quad [66]$$

In what follows, we also make use of the fact that, for $\mathbf{A} \in \mathcal{S}(d, \mathbb{C})$,

$$\lambda_\ell(-\mathbf{A}) = -\lambda_{d-\ell+1}(\mathbf{A}), \quad \ell = 1, \ldots, d. \tag{67}$$

With these preliminaries now in hand, we turn to the proof of Theorem 3. This result is established with the aide of four auxiliary results, Lemma 1, Lemma 2, Lemma 3 and Lemma 4, whose precise statements and proof are provided immediately afterward.

Proof. To demonstrate our desired result, Eq. (44), it suffices to establish the lower bound

$$\sigma_2^2(\mathbf{Q}) \geq (d-1)^2 \left\{ \left(\frac{\mu\, d}{d-1} \right)^2 + O_{\text{a.s.}}\left(\sqrt{\frac{\log d}{d}} \right) \right\} \tag{68}$$

as well as the upper bound

$$\sigma_d^2(\mathbf{Q}) \leq (d-1)^2 \left\{ \left(\frac{\mu\, d}{d-1} \right)^2 + O_{\text{a.s.}}\left(\sqrt{\frac{\log d}{d}} \right) \right\}. \tag{69}$$

With this in mind, we now decompose $\mathbf{Q}$ as follows, starting with the case Eq. (42). Fix $d \in \mathbb{N}\backslash\{1\}$. Taking F_X to be the distribution defining the elements in Eq. (42), draw

$$\{\widetilde{q}_{ii}\}_{i=1,\ldots,d} \overset{\text{iid}}{\sim} F_X \text{ independently of } \{q_{ij}\}_{i,j=1,\ldots,d}. \tag{70}$$

Here note carefully that $\widetilde{q}_{ii} \overset{d}{=} q_{ij}$ and $\widetilde{q}_{ii} \overset{d}{\neq} q_{ii}$. Now recast $\mathbf{Q} = \{q_{ij}\}_{i,j}$ as

$$\begin{pmatrix} 0 & q_{12}-\mu & \ldots & q_{1d}-\mu \\ q_{21}-\mu & 0 & \ldots & q_{2d}-\mu \\ \vdots & \vdots & \ddots & \vdots \\ q_{d1}-\mu & q_{d2}-\mu & \ldots & 0 \end{pmatrix} + \begin{pmatrix} q_{11} & \mu & \ldots & \mu \\ \mu & q_{22} & \ldots & \mu \\ \vdots & \vdots & \ddots & \vdots \\ \mu & \mu & \ldots & q_{dd} \end{pmatrix}$$

$$= \left\{ \begin{pmatrix} \widetilde{q}_{11}-\mu & q_{12}-\mu & \ldots & q_{1d}-\mu \\ q_{21}-\mu & \widetilde{q}_{22}-\mu & \ldots & q_{2d}-\mu \\ \vdots & \vdots & \ddots & \vdots \\ q_{d1}-\mu & q_{d2}-\mu & \ldots & \widetilde{q}_{dd}-\mu \end{pmatrix} \right. \tag{71}$$

$$\left. + \begin{pmatrix} -\widetilde{q}_{11}+\mu & 0 & \ldots & 0 \\ 0 & -\widetilde{q}_{22}+\mu & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & -\widetilde{q}_{dd}+\mu \end{pmatrix} \right\}$$

$$+ (1-d) \left\{ \begin{pmatrix} \frac{-q_{11}-(d-1)\mu}{d-1} & 0 & \ldots & 0 \\ 0 & \frac{-q_{22}-(d-1)\mu}{d-1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{-q_{dd}-(d-1)\mu}{d-1} \end{pmatrix} \right.$$

$$\left. + \begin{pmatrix} \mu & \frac{-\mu}{d-1} & \ldots & \frac{-\mu}{d-1} \\ \frac{-\mu}{d-1} & \mu & \ldots & \frac{-\mu}{d-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-\mu}{d-1} & \frac{-\mu}{d-1} & \ldots & \mu \end{pmatrix} \right\} \tag{72}$$

$$=: \mathbf{Q}_1 + \mathbf{Q}_2 + (1-d)\left\{ \mathbf{Q}_3 + \mathbf{Q}_4 \right\} =: \mathbf{R} + (1-d)\mathbf{Q}_4 \tag{73}$$

Working from Eq. (73) we start by establishing the first bound Eq. (68) under Eq. (42). Noting from Lemma 4 that $\mathbf{Q}_4$ is symmetric and positive, making use of Eq. (66), first with $i = 2, j = 1$, then with $i = 1, j = 1$, invoking Eq. (67) and finally using that $\mathbf{R}\mathbf{R}^*$ is symmetric and non-negative, we find

$$\begin{aligned} &\sigma_2^2(\mathbf{Q}) = \lambda_2(\mathbf{Q}\mathbf{Q}^*) \\ &\geq (d-1)^2 \left\{ \lambda_2(\mathbf{Q}_4^2) + \lambda_1\left(\frac{\mathbf{R}\mathbf{R}^*}{(d-1)^2} - \frac{\mathbf{Q}_4\mathbf{R}^* + \mathbf{R}\mathbf{Q}_4^*}{d-1} \right) \right\} \\ &\geq (d-1)^2 \left\{ \lambda_2(\mathbf{Q}_4^2) + \lambda_1\left(\frac{\mathbf{R}\mathbf{R}^*}{(d-1)^2} \right) + \lambda_1\left(\frac{\mathbf{Q}_4\mathbf{R}^* + \mathbf{R}\mathbf{Q}_4^*}{1-d} \right) \right\} \\ &\geq (d-1)^2 \left\{ \lambda_2(\mathbf{Q}_4^2) - \lambda_d\left(\frac{\mathbf{Q}_4\mathbf{R}^* + \mathbf{R}\mathbf{Q}_4^*}{d-1} \right) \right\}. \end{aligned} \tag{74}$$

Now observe

$$|\lambda_j(\mathbf{A})| \leq \sigma_d(\mathbf{A}) = \|\mathbf{A}\|_{op} \quad \text{for any } j = 1, \ldots, d, \tag{75}$$

whenever $\mathbf{A}$ is symmetric. Furthermore,

$$\lambda_2(\mathbf{Q}_4^2) = \lambda_2^2(\mathbf{Q}_4) = \frac{\mu^2 d^2}{(d-1)^2}, \tag{76}$$

where the equality follows from Eq. (90) in Lemma 4. Combining Eq. (74), Eq. (75) and Eq. (76), we therefore infer

$$\sigma_2^2(\mathbf{Q}) \geq (d-1)^2 \left\{ \left(\frac{\mu\, d}{d-1} \right)^2 - \frac{1}{d-1} \|\mathbf{Q}_4\mathbf{R}^* + \mathbf{R}\mathbf{Q}_4^*\|_{\text{op}} \right\}. \tag{77}$$

However, by Eq. (80) in Lemma 1, Eq. (83) in Lemma 2 and Eq. (85) from Lemma 3 we have

$$\begin{aligned} \frac{d}{d-1} \frac{\|\mathbf{R}\|_{\text{op}}}{d} &\leq \frac{d}{d-1} \left(\frac{\|\mathbf{Q}_1\|_{\text{op}}}{d} + \frac{\|\mathbf{Q}_2\|_{\text{op}}}{d} \right) + \|\mathbf{Q}_3\|_{\text{op}} \\ &= O_{\text{a.s.}}\left(\sqrt{\frac{\log d}{d}} \right). \end{aligned} \tag{78}$$

Hence, by Eq. (78) and Eq. (90) in Lemma 4,

$$\begin{aligned} \frac{1}{d-1} \|\mathbf{Q}_4\mathbf{R}^* + \mathbf{R}\mathbf{Q}_4^*\|_{\text{op}} &\leq \frac{2d}{d-1} \|\mathbf{Q}_4\|_{\text{op}} \left\| \frac{\mathbf{R}}{d} \right\|_{\text{op}} \\ &= O(1) \cdot O_{\text{a.s.}}\left(\sqrt{\frac{\log d}{d}} \right) = O_{\text{a.s.}}\left(\sqrt{\frac{\log d}{d}} \right). \end{aligned} \tag{79}$$

Thus, by Eq. (77), Eq. (78) and Eq. (79), the lower bound Eq. (68) holds.

We now turn to verify Eq. (69). Here, by Eq. (73), Eq. (65) with $j = 0, i = d$, Eq. (76) and finally Eq. (75),

$$\begin{aligned} &\sigma_d^2(\mathbf{Q}) = \lambda_d(\mathbf{Q}\mathbf{Q}^*) \\ &\leq (d-1)^2 \left\{ \lambda_d(\mathbf{Q}_4^2) + \lambda_d\left(\frac{\mathbf{R}\mathbf{R}^*}{(d-1)^2} \right) + \lambda_d\left(\frac{\mathbf{Q}_4\mathbf{R}^* + \mathbf{R}\mathbf{Q}_4^*}{1-d} \right) \right\} \\ &\leq (d-1)^2 \left\{ \left(\frac{\mu\, d}{d-1} \right)^2 + \left\| \frac{\mathbf{R}}{d-1} \right\|_{\text{op}}^2 + \frac{1}{d-1} \|\mathbf{Q}_4\mathbf{R}^* + \mathbf{R}\mathbf{Q}_4^*\|_{\text{op}} \right\}. \end{aligned}$$

The upper bound Eq. (69) now follows from Eq. (78) and Eq. (79). Thus, as a consequence of Eq. (68) and Eq. (69), Eq. (44) holds under condition Eq. (42), as claimed. We have established the desired result in the first case Eq. (42).

Now suppose that condition Eq. (43) holds. In order to establish Eq. (44), it suffices to replace q_{ij}, $i < j$, with q_{ji} in expression Eq. (71) and in the definition of $\mathbf{Q}_3$. We then follow the rest of the argument for proving Eq. (44) under condition Eq. (42) noting that Lemma 1, Lemma 3 also hold in the symmetric case. This concludes the proof. □

We turn now to the Lemmata 1, 2, 3 and 4, which are used in the proof of Theorem 3. These results correspond to the bounds we use for each of elements in the decomposition Eq. (73). We start off with Lemma 1 which essentially packages results found in (39), (21).

Lemma 1. *Let $\mathbf{Q}_1 = \mathbf{Q}_1(d)$, $d \in \mathbb{N} \setminus \{1\}$, be the sequence of random matrices defined as in* Eq. (71). *Here we suppose that the off diagonal elements q_{ij} defining $\mathbf{Q}_1$ are either as in* Eq. (42) *or as in* Eq. (43) *and that the diagonal elements $\widetilde{q}_{ii}$ are drawn as in* Eq. (70). *Then, in either case, we have*

$$\frac{\|\mathbf{Q}_1\|_{\text{op}}}{d} = O_{\text{a.s.}}\left(\frac{1}{\sqrt{d}} \right). \tag{80}$$

Proof. To establish Eq. (80) in the first case, Eq. (42), we note that $d^{-1}\mathbf{Q}_1\mathbf{Q}_1^*$ forms a sample covariance matrix, where each entry of $\mathbf{Q}_1$ is centered and has infinitely many moments; cf. Remark 2. Thus, by (39, Theorem 3.1, p. 517),

$$\|\mathbf{Q}_1\|_{\text{op}} = \sqrt{d} \cdot \sqrt{\lambda_d\left(\frac{\mathbf{Q}_1\mathbf{Q}_1^*}{d} \right)} = \sqrt{d} \cdot \sqrt{4\sigma^2 + o_{\text{a.s.}}(1)}, \tag{81}$$

which immediately yields Eq. (80).

In the second case, where all the off-diagonal elements are determined starting from Eq. (43), $\mathbf{Q}_1$ is a centered (mean-zero) Wigner matrix in the typical nomenclature followed by (21). Thus, noting Eq. (11) and that, as in the previous case, $\mathbf{Q}_1$ has infinitely many moments Eq. (80) is thus a direct consequence of (21, Theorem 2.12, p. 630). $\square$

We next state and establish Lemma 2. Note that the assumption of independence among the random variables is not needed for this result. Here and below in what follows we adopt the useful notational convention that $o_{\text{a.s.}}(1)$ denotes a random variable that vanishes almost surely, as $d \to \infty$ (cf. Eq. (64)).

Lemma 2. *Let $\{Y_d\}_{d\in\mathbb{N}}$ be an identically distributed sequence of sub-exponential variables with mean μ; cf. Definition 1. Let*

$$Z_d = \max_{k=1,\ldots,d} |Y_k - \mu|, d \in \mathbb{N}. \qquad [82]$$

Then, for any fixed $\varepsilon_0 > 0$,

$$\frac{Z_d}{d^{\varepsilon_0}} = o_{\text{a.s.}}(1). \qquad [83]$$

In particular, for the sequence of diagonal matrices $\mathbf{Q}_2 \equiv \mathbf{Q}_2(d)$ given as in Eq. (73)*, and any $\varepsilon_0 > 0$,*

$$\frac{\|\mathbf{Q}_2\|_{op}}{d^{\varepsilon_0}} = o_{\text{a.s.}}(1). \qquad [84]$$

Proof. For any $M > 0$,

$$\mathbb{P}\Big(\frac{Z_d}{d^{\varepsilon_0}} \geq \frac{M}{d^{\varepsilon_0/2}}\Big) = \mathbb{P}\Big(\bigcup_{k=1}^{d}\Big\{\frac{|Y_k-\mu|}{d^{\varepsilon_0}} \geq \frac{M}{d^{\varepsilon_0/2}}\Big\}\Big)$$
$$\leq d\cdot\mathbb{P}\Big(|Y_1-\mu| \geq d^{\varepsilon_0/2}M\Big) \leq 2d\cdot e^{-\frac{d^{\varepsilon_0/2}M}{K}}.$$

In the first and second inequalities, respectively, we use the assumption that the random variables $\{Y_d\}_{d\in\mathbb{N}}$ are identically distributed as well as relation Eq. (40) for some suitable value $K > 0$. Therefore, $\sum_{d=1}^{\infty}\mathbb{P}(Z_d/d^{\varepsilon_0/2} \geq M/d^{\varepsilon_0/2}) < \infty$. Since $\lim_{d\to\infty} M/d^{\varepsilon_0/2} = 0$, Eq. (83) now follows as a consequence of the Borel-Cantelli lemma.

Regarding the second claim observe that, cf. Eq. (70), $\mathbf{Q}_2$ is a diagonal matrix containing identically distributed sub-exponential random variables. This mean that, taking $Y_k := \tilde{q}_{kk}$, $\|\mathbf{Q}_2\|$ is of the form Eq. (82). As such the second claim Eq. (84) now follows from the first Eq. (83), completing the proof. $\square$

Turning to our bounds on $\|\mathbf{Q}_3\|_{op}$, we have the following lemma.

Lemma 3. *Let $\mathbf{Q}_3 \equiv \mathbf{Q}_3(d)$, $d \in \mathbb{N}$, be a sequence of random matrices as in* Eq. (73)*. Here we assume that the q_{ii} are determined either according to* Eq. (42) *or according to* Eq. (43)*. Then, in either of these cases,*

$$\|\mathbf{Q}_3\|_{\text{op}} = O_{\text{a.s.}}\Big(\sqrt{\frac{\log d}{d}}\Big). \qquad [85]$$

Proof. Under either Eq. (42) or Eq. (43), each of the entries along the main diagonal is a renormalized sums of iid sub-exponential random variables i.e.

$$\frac{-q_{ii}-\mu(d-1)}{d-1} = \frac{1}{d-1}\sum_{j\neq i}^{d}(q_{ij}-\mu) \qquad [86]$$

While these diagonal elements are not independent under Eq. (43) note carefully that we do not use the independence of the rows of $\mathbf{Q}_3$ in the arguments that follow.

Now, by Bernstein's inequality, (31, p. 29, Theorem 2.8.1), there exists a constant $C > 0$ such that, for any $\varepsilon > 0$,

$$\mathbb{P}\Big(\Big|-q_{ii}-\mu(d-1)\Big| > \varepsilon\Big)$$
$$\leq 2\cdot\exp\Big\{-C\min\Big\{\frac{\varepsilon^2}{(d-1)\|X_\mu\|_{\psi_1}^2}, \frac{\varepsilon}{\|X_\mu\|_{\psi_1}}\Big\}\Big\}, \qquad [87]$$

for $i = 1,\ldots,d$ where $X_\mu := X - \mu$ for $X \sim F_X$ and the right-hand side of Eq. (87) involves the sub-exponential norm Eq. (41). Fix $\delta > 0$, and let $\eta := (2+\delta)/C$ where C is the constant in the upper bound in Eq. (87). Hence, for all $d \in \mathbb{N}\backslash\{1\}$,

$$\mathbb{P}\Big(\sqrt{\frac{d-1}{\eta\log d}}\,\|\mathbf{Q}_3(d)\|_{\text{op}} > \|X_\mu\|_{\psi_1}\Big)$$
$$= \mathbb{P}\Big(\max_{i=1,\ldots,d}\Big|\frac{-q_{ii}-\mu(d-1)}{d-1}\Big| > \|X_\mu\|_{\psi_1}\sqrt{\frac{\eta\log d}{d-1}}\Big)$$
$$= \mathbb{P}\Big(\bigcup_{i=1}^{d}\Big\{\Big|-q_{ii}-\mu(d-1)\Big| > \|X_\mu\|_{\psi_1}\sqrt{\eta(d-1)\log d}\Big\}\Big)$$
$$\leq \sum_{i=1}^{d}\mathbb{P}\Big(\Big|-q_{ii}-\mu(d-1)\Big| > \|X_\mu\|_{\psi_1}\sqrt{\eta(d-1)\log d}\Big)$$
$$\leq 2d\cdot\exp\Big\{-C\min\Big\{\eta\log d, \sqrt{\eta(d-1)\log d}\Big\}\Big\}. \qquad [88]$$

In the second inequality in Eq. (88), we use Eq. (87) with $\varepsilon := \|X_\mu\|_{\psi_1}\sqrt{\eta(d-1)\log d}$ and the fact that $\{q_{ii}\}_{i=1,\ldots,d}$ are identically (but not necessarily independently) distributed. Therefore, for every d sufficiently large,

$$\mathbb{P}\Big(\sqrt{\frac{d-1}{\eta\log d}}\|\mathbf{Q}_3(d)\|_{\text{op}} > \|X_\mu\|_{\psi_1}\Big) \leq \frac{2}{d^{1+\delta}},$$

and as a consequence,

$$\sum_{d=1}^{\infty}\mathbb{P}\Big(\|\mathbf{Q}_3(d)\|_{\text{op}} > \|X_\mu\|_{\psi_1}\sqrt{\frac{\eta\log d}{d-1}}\Big) < \infty.$$

Hence, by the Borel-Cantelli lemma,

$$\mathbb{P}\Big(\|\mathbf{Q}_3(d)\|_{\text{op}} > \|X_\mu\|_{\psi_1}\sqrt{\frac{\eta\log d}{d-1}}\ \text{i.o.}\Big) = 0,$$

so that the relation Eq. (85) holds. The proof is complete. $\square$

Finally we conclude with Lemma 4 as follows.

Lemma 4. *Let $\mathbf{Q}_4 \equiv \mathbf{Q}_4(d)$, $d \in \mathbb{N}$, be the sequence of symmetric matrices defined in* Eq. (73)*. Then, for any $d \in \mathbb{N}\backslash\{1\}$,*

$$\lambda_1(\mathbf{Q}_4) = 0, \qquad [89]$$

and

$$\lambda_\ell(\mathbf{Q}_4) = \frac{\mu\, d}{d-1}, \quad \text{for } \ell = 2,\ldots,d. \qquad [90]$$

Proof. The statement Eq. (89) is a consequence of the fact that $\mathbf{Q}_4\,(1,\ldots,1)^\top = \mathbf{0} \in \mathbb{R}^n$. To establish Eq. (90), it suffices to prove that

$$\frac{d}{d-1} \leq \frac{\lambda_2(\mathbf{Q}_4)}{\mu} \leq \frac{\lambda_d(\mathbf{Q}_4)}{\mu} \leq \frac{d}{d-1}, \quad d \in \mathbb{N}\backslash\{1\}. \qquad [91]$$

Let $\mathbf{1} \in \mathcal{S}(d,\mathbb{R})$ be a matrix of ones and recast

$$\frac{1}{\mu}\mathbf{Q}_4 = \mathbf{I} + \frac{1}{d-1}(\mathbf{I}-\mathbf{1}). \qquad [92]$$

By Eq. (92), Eq. (66) with $i = 2, j = 1$ and Eq. (67),

$$\frac{\lambda_2(\mathbf{Q}_4)}{\mu} \geq \lambda_1(\mathbf{I}) + \lambda_2\Big(\frac{1}{d-1}(\mathbf{I}-\mathbf{1})\Big)$$
$$= \lambda_1(\mathbf{I}) - \frac{1}{d-1}\lambda_{d-1}(\mathbf{1}-\mathbf{I}). \qquad [93]$$

However, by Eq. (65), in the case $i = d-1, j = 1$,

$$\lambda_{d-1}(\mathbf{1}-\mathbf{I}) \leq \lambda_{d-1}(\mathbf{1}) + \lambda_d(-\mathbf{I}) = -1, \qquad [94]$$

where we used that

$$\lambda_1(\mathbf{1}) = \cdots = \lambda_{d-1}(\mathbf{1}) = 0. \qquad [95]$$

This latter claim follows immediately from the fact that $\text{rank}(\mathbf{1}) = 1$ and $\mathbf{1}\,(1,\ldots,1)^\top = d(1,\ldots,1)^\top$. The first inequality in Eq. (91) is now a consequence of Eq. (93) and Eq. (94).

On the other hand, again by Eq. (92), Eq. (65), this time with $i = d, j = 0$, and Eq. (67),

$$\frac{\lambda_d(\mathbf{Q}_4)}{\mu} \leq \lambda_d(\mathbf{I}) + \lambda_d\left(\frac{1}{d-1}(\mathbf{I}-\mathbf{1})\right)$$
$$= \lambda_d(\mathbf{I}) - \lambda_1\left(\frac{1}{d-1}(\mathbf{1}-\mathbf{I})\right). \quad [96]$$

However, by Eq. (66) and Eq. (95),

$$\lambda_1(\mathbf{1}-\mathbf{I}) \geq \lambda_1(\mathbf{1}) + \lambda_1(-\mathbf{I}) = -1. \quad [97]$$

The third inequality in Eq. (91) is now a consequence of Eq. (96) and Eq. (97). This establishes Eq. (91) and, hence, Eq. (90), completing the proof. □

D. Surrogate-trajectory Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) (40, 41) is an advanced MCMC procedure that uses numerical approximations to Hamiltonian trajectories to generate Metropolis-type (16) proposals far away from the current Markov chain state. On the one hand, this approach to proposal generation helps reduce autocorrelation between chain states and is particularly helpful within higher-dimensional state spaces. On the other hand, numerical integration of Hamilton's equations requires repeated evaluation of the Hamiltonian potential energy's gradient, and these repeated floating-point operations may become computationally burdensome.

Lets briefly recall this HMC approach to resolve a given density $\pi(\cdot)$ of a 'target' probability measure of interest.* We proceed by considering a potential energy of the form $U(\boldsymbol{\theta}) := -\log \pi(\boldsymbol{\theta})$. We then select an associated kinetic energy $V(\boldsymbol{\xi}) := |\mathbf{G}^{-1/2}\boldsymbol{\xi}|^2$ for an appropriately chosen symmetric-positive-definite mass matrix $\mathbf{G}$ (which is often taken as the identity for simplicity). One then observes that the Gibbs measure, proportional to $e^{-H(\boldsymbol{\theta},\boldsymbol{\xi})}$ where $H = U + V$, is invariant under the associated Hamiltonian dynamic. Here note that the $\boldsymbol{\theta}$ marginal of $e^{-H(\boldsymbol{\theta},\boldsymbol{\xi})}$ coincides with π while the $\boldsymbol{\xi}$ marginal is normally distributed as $\mathcal{N}(\mathbf{0}, \mathbf{G})$.

One operationalizes these observations as an algorithmic sampling procedure as follows. At each step, given a current sample $\boldsymbol{\theta}^{(n)}$, one draws $\boldsymbol{\xi}^{(n)} \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$. From this initial state $(\boldsymbol{\theta}(0), \boldsymbol{\xi}(0)) := (\boldsymbol{\theta}^{(n)}, \boldsymbol{\xi}^{(n)})$ one then numerically approximates the associated Hamiltonian dynamics using a Störmer-Verlet (velocity Verlet) or leapfrog integrator up to a total integration time $\tau > 0$ and using integration step size $\epsilon > 0$. In this context, note that a single iteration of this integrator takes the form, (42),

$$\begin{aligned} \boldsymbol{\xi}\left(s + \frac{\epsilon}{2}\right) &:= \boldsymbol{\xi}(s) + \frac{\epsilon}{2}\nabla \log \pi(\boldsymbol{\theta}(s)), \\ \boldsymbol{\theta}(s+\epsilon) &:= \boldsymbol{\theta}(s) + \epsilon\,\mathbf{G}^{-1}\boldsymbol{\xi}(s + \frac{\epsilon}{2}), \\ \boldsymbol{\xi}(s+\epsilon) &:= \boldsymbol{\xi}\left(s + \frac{\epsilon}{2}\right) + \frac{\epsilon}{2}\nabla \log \pi(\boldsymbol{\theta}(s+\epsilon))\,. \end{aligned} \quad [98]$$

In this fashion the proposed new state is given through Eq. (98) by $\boldsymbol{\theta}(\tau)$. To remove bias this procedure can be augmented with an acceptance probability of the form $\alpha^{(n)} := \exp(H(\boldsymbol{\theta}(0), \boldsymbol{\xi}(0)) - H(\boldsymbol{\theta}(\tau), \boldsymbol{\xi}(\tau)) \wedge 1$.

Different strategies aim to speed up the leapfrog integrator's many log-posterior gradient evaluations $\nabla \log \pi(\boldsymbol{\theta})$, as these numerical routines often represent the algorithm's computational bottleneck. In model-specific contexts, (3, 4, 43) yield parallelization strategies, and (35) develops dynamic programming techniques, to accelerate the evaluation of $\nabla \log \pi(\boldsymbol{\theta})$.

A small body of work considers another approach by replacing $\nabla \log \pi(\boldsymbol{\theta})$ with a suitable approximation $\widetilde{\nabla} \log \pi(\boldsymbol{\theta})$ and recognizing that the modified Eq. (98) continues to satisfy path reversibility and volume preservation, two essential ingredients for well-specified HMC. It follows immediately that the resulting 'surrogate trajectory HMC' continues to sample the correct target distribution $\pi(\cdot)$; see further details (33, 41). Nonetheless, the acceptance rates and overall efficiency of such samplers may suffer when approximations are poor.

*In the Bayesian inference context, we are typically considering the posterior density function $\pi(\boldsymbol{\theta}) := p(\boldsymbol{\theta}|\mathbf{Y})$ for the parameter $\boldsymbol{\theta} \in \mathbb{R}^K$ given observed data $\mathbf{Y}$.

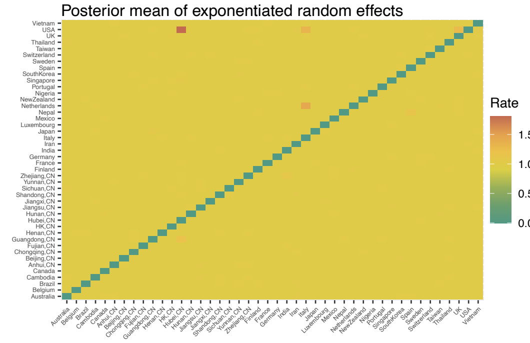

Fig. S1. Posterior means for exponentiated random effects convey expected multiplicative deviations from the portion of the rate attributable to fixed effects for each corresponding element of the generator matrix. Notably, we infer a roughly 1.81-fold posterior mean increase in the rate of transitions from the US to Hubei, CN, beyond that portion of the rate which may be explained by fixed effects. Less pronounced are posterior mean multiplicative increases of 1.34 (US to Italy), 1.27 (US to UK), 1.44 (Netherlands to Italy) and 1.21 (Guangdong, CN, to Hubei, CN).

The majority of surrogate HMC methods first obtain a small sample of exact gradient evaluations and then use some model to interpolate: (44) assumes the approximate gradient follows a piecewise constant form across a grid; (45, 46) construct approximations using Gaussian processes; and (47, 48) do the same using neural networks. But an even simpler approach to surrogate HMC may be appropriate when gradients have series representations as we can leverage here in Eq. (3).

E. Visualizing the posterior mean random effects

Figure S1 displays posterior means of the exponentiated random effects, which one may interpret as multiplicative deviations from the fixed effects' contributions to the generator matrix. Whereas the vast majority of exponentiated random effects have posterior means close to 1, indicating no deviation from the fixed-effect model, a few exhibit posterior means that are significantly greater than 1. In particular, the rate element corresponding to transfer from the US to Hubei, CN, exhibits a 1.81-fold random-effect derived increase to the fixed-effect component. This large multiplicative increase agrees with, but goes beyond, the influence that the Hubei asymmetry holds for the entire generator matrix model.

1. P Lemey, A Rambaut, AJ Drummond, MA Suchard, Bayesian phylogeography finds its roots. *PLoS computational biology* **5**, e1000520 (2009).
2. P Lemey, et al., Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza h3n2. *PLoS pathogens* **10**, e1003932 (2014).
3. AJ Holbrook, et al., Massive parallelization boosts big bayesian multidimensional scaling. *J. Comput. Graph. Stat.* **30**, 11–24 (2021).
4. AJ Holbrook, X Ji, MA Suchard, From viral evolution to spatial contagion: a biologically modulated hawkes model. *Bioinformatics* **38**, 1846–1856 (2022).
5. JR Norris, *Continuous-time Markov chains I*, Cambridge Series in Statistical and Probabilistic Mathematics. (Cambridge University Press), p. 60–107 (1997).
6. C Moler, C Van Loan, Nineteen dubious ways to compute the exponential of a matrix. *SIAM review* **20**, 801–836 (1978).
7. C Moler, C Van Loan, Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review* **45**, 3–49 (2003).
8. J Felsenstein, Evolutionary trees from dna sequences: a maximum likelihood approach. *J. molecular evolution* **17**, 368–376 (1981).
9. JS Sinsheimer, JA Lake, RJ Little, Bayesian hypothesis testing of four-taxon topologies using molecular sequence data. *Biometrics* pp. 193–210 (1996).
10. Z Yang, B Rannala, Bayesian phylogenetic inference using dna sequences: a markov chain monte carlo method. *Mol. biology evolution* **14**, 717–724 (1997).
11. B Mau, MA Newton, B Larget, Bayesian phylogenetic inference via markov chain monte carlo methods. *Biometrics* **55**, 1–12 (1999).
12. MA Suchard, RE Weiss, JS Sinsheimer, Bayesian selection of continuous-time markov chain evolutionary models. *Mol. biology evolution* **18**, 1001–1013 (2001).
13. TH Jukes, CR Cantor, , et al., Evolution of protein molecules. *Mammalian protein metabolism* **3**, 21–132 (1969).
14. M Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. molecular evolution* **16**, 111–120 (1980).
15. M Hasegawa, H Kishino, Ta Yano, Dating of the human-ape splitting by a molecular clock of mitochondrial dna. *J. molecular evolution* **22**, 160–174 (1985).
16. N Metropolis, AW Rosenbluth, MN Rosenbluth, AH Teller, E Teller, Equation of state calculations by fast computing machines. *The journal chemical physics* **21**, 1087–1092 (1953).

17. WK Hastings, Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970).
18. AF Magee, et al., Random-effects substitution models for phylogenetics via scalable gradient approximations. *arXiv preprint arXiv:2303.13642* **1** (2023).
19. I Najfeld, TF Havel, Derivatives of the matrix exponential and their computation. *Adv. applied mathematics* **16**, 321–375 (1995).
20. Y Takahashi, Markov chains with random transition matrices in *Kodai Mathematical Seminar Reports.* Vol. 21, pp. 426–447 (1969).
21. ZD Bai, Methodologies in spectral analysis of large dimensional random matrices, a review. *Stat. Sinica* **9**, 611–677 (1999).
22. D Chafaï, The Dirichlet Markov Ensemble. *J. Multivar. Analysis* **101**, 555–567 (2010).
23. C Bordenave, P Caputo, D Chafaï, Circular law theorem for random markov matrices. *Probab. Theory Relat. Fields* **152**, 751–779 (2012).
24. A Chatterjee, RS Hazra, Spectral properties for the Laplacian of a generalized Wigner matrix. *Random Matrices: Theory Appl.* **11**, 2250026 (2022).
25. G Nakerst, S Denisov, M Haque, Random sparse generators of markovian evolution and their spectral properties. *Phys. Rev. E* **108**, 014102 (2023).
26. W Bryc, A Dembo, T Jiang, Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Annals Probab.* **34**, 1–38 (2006).
27. X Ding, T Jiang, Spectral distributions of adjacency and laplacian matrices of random graphs. *Annals Appl. Probab.* pp. 2086–2117 (2010).
28. C Bordenave, P Caputo, D Chafaï, Spectrum of markov generators on sparse random graphs. *Commun. on Pure Appl. Math.* **67**, 621–669 (2014).
29. LN Trefethen, D Bau, *Numerical linear algebra.* (Siam) Vol. 181, (2022).
30. RA Horn, CR Johnson, *Matrix Analysis.* (Cambridge University Press), (2012).
31. R Vershynin, *High-Dimensional Probability: an Introduction with Applications in Data Science.* (Cambridge University Press) Vol. 47, (2018).
32. CM Goldie, C Klüppelberg, Subexponential distributions. *A practical guide to heavy tails: statistical techniques applications* pp. 435–459 (1998).
33. NE Glatt-Holtz, JA Krometis, CF Mondaini, On the accept-reject mechanism for metropolis-hastings algorithms. *Annals Appl. Probab.* (2023) (to appear).
34. NG Polson, JG Scott, J Windle, The Bayesian bridge. *J. Royal Stat. Soc. Ser. B: Stat. Methodol.* **76**, 713–733 (2014).
35. X Ji, et al., Gradients do grow on trees: a linear-time o (n)-dimensional gradient for statistical phylogenetics. *Mol. biology evolution* **37**, 3047–3060 (2020).
36. M Worobey, et al., The emergence of sars-cov-2 in europe and north america. *Science* **370**, 564–570 (2020).
37. A Nishimura, MA Suchard, Shrinkage with shrunken shoulders: Gibbs sampling shrinkage model posteriors with guaranteed convergence rates. *Bayesian Analysis* **1**, 1–24 (2022).
38. P Lemey, et al., Accommodating individual travel history and unsampled diversity in bayesian phylogeographic inference of sars-cov-2. *Nat. communications* **11**, 5110 (2020).
39. YQ Yin, ZD Bai, PR Krishnaiah, On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Theory Relat. Fields* **78**, 509–521 (1988).
40. S Duane, AD Kennedy, BJ Pendleton, D Roweth, Hybrid monte carlo. *Phys. letters B* **195**, 216–222 (1987).
41. R Neal, Mcmc using hamiltonian dynamics. *Handb. Markov Chain Monte Carlo* **1**, 113–162 (2011).
42. B Leimkuhler, S Reich, *Simulating hamiltonian dynamics.* (Cambridge university press) No. 14, (2004).
43. AJ Holbrook, X Ji, MA Suchard, Bayesian mitigation of spatial coarsening for a hawkes model applied to gunfire, wildfire and viral contagion. *The Annals Appl. Stat.* **16**, 573–595 (2022).
44. C Zhang, B Shahbaba, H Zhao, Precomputing strategy for hamiltonian monte carlo method based on regularity in parameter space. *Comput. Stat.* **32**, 253–279 (2017).
45. CE Rasmussen, Gaussian processes to speed up hybrid monte carlo for expensive bayesian integrals in *Seventh Valencia international meeting, dedicated to Dennis V. Lindley.* (Oxford University Press), pp. 651–659 (2003).
46. S Lan, T Bui-Thanh, M Christie, M Girolami, Emulation of higher-order tensors in manifold monte carlo methods for bayesian inverse problems. *J. Comput. Phys.* **308**, 81–101 (2016).
47. C Zhang, B Shahbaba, H Zhao, Hamiltonian monte carlo acceleration using surrogate functions with random bases. *Stat. computing* **27**, 1473–1490 (2017).
48. L Li, A Holbrook, B Shahbaba, P Baldi, Neural network gradient hamiltonian monte carlo. *Comput. statistics* **34**, 281–299 (2019).