

Surviving species in a Large Lotka-Volterra system of differential equations

Maxime CLÉNET¹, François MASSOL², Jamal NAJIM^{1*}

¹CNRS and Université Gustave Eiffel, France
5 Bd Descartes, 77420 Champs-sur-Marne

²CNRS, Université de Lille, INSERM, CHU, Institut Pasteur Lille
1 Rue du Professeur Calmette, 59000 Lille

maxime.clenet@univ-eiffel.fr, francois.massol@univ-lille.fr,
jamal.najim@univ-eiffel.fr

Résumé – Les équations différentielles de Lotka-Volterra jouent un rôle clé pour la modélisation mathématique de systèmes dynamiques complexes d’agents en interaction. Quand le nombre d’agents (on parlera d’espèces dans un contexte biologique ou écologique) devient grand, des questions simples telles que le nombre d’espèces survivantes n’ont pas de réponses théoriques claires. Dans cet article, nous considérons un grand système d’équations de Lotka-Volterra où les interactions entre les différentes espèces sont aléatoires, et présentons une heuristique pour calculer le nombre d’espèces survivantes. Cette heuristique combine des arguments de la théorie des grandes matrices aléatoires, d’optimisation mathématique et de théorie des valeurs extrêmes. Des simulations numériques illustrent la précision et la portée des résultats présentés.

Abstract – Lotka-Volterra differential equations play a key role in the mathematical modeling of complex dynamical systems of interacting components. When the number of components (we shall refer to species in a biological or ecological context) becomes large, basic but fundamental questions such as computing the number of surviving species still lack theoretical answers. In this paper, we consider a large system of Lotka-Volterra equations where the interactions between the various species are random, and present a heuristics to compute the number of surviving species. This heuristics combines arguments from Random Matrix Theory, mathematical optimization (LCP), and standard extreme value theory. Numerical simulations illustrate the accuracy and scope of the results.

1 Introduction

Model and assumptions. Large Lotka-Volterra (LV) systems of differential equations are widely used in various scientific fields involving complex dynamical systems with interacting components, such as biology, ecology, chemistry, etc. [1, 2]. A LV system represents a good trade-off between a fairly realistic model and a mathematically tractable one, which has yielded insights on issues such as the link between the productivity and stability of ecosystems [3] or the resistance of species communities against invasions [4]. In the sequel, we use the ecological terminology and refer to the interacting components as *species*.

A large LV system is a system of differential equations :

$$\frac{dx_k(t)}{dt} = x_k(t) \left(r_k - \theta x_k(t) + \frac{1}{\alpha\sqrt{n}} \sum_{\ell \in [n]} A_{k\ell} x_\ell(t) \right), \quad (1)$$

where $k \in [n] = \{1, \dots, n\}$.

The number n represents the number of species within the system, the unknown vector $\mathbf{x} = (x_k)_{k \in [n]}$ is the vector of abundances of the various species in the foodweb and evolves with time $t > 0$ according to the dynamics (1). Quantity $x_k(t)$

represents the abundance (a value representing the population size) of species k at time t .

In Eq. (1), r_k represents the intrinsic growth rate of species k , θ is an intraspecific competition coefficient, and $A_{k\ell}$ is the per capita effect of species ℓ on species k (interactions).

Hereafter, we focus on the idealized model $r_k = \theta = 1$:

$$\frac{dx_k}{dt} = x_k \left(1 - x_k + \frac{(A\mathbf{x})_k}{\alpha\sqrt{n}} \right), \quad (2)$$

In an ecological or biological context (think of animal species interacting in a lake or a remote valley, or the human microbiome), it is often extremely difficult and/or expensive to estimate precisely each interaction strength $A_{k\ell}$. In the absence of any prior information, these interactions can be modeled as random (see for instance [5]), which we assume in the sequel : matrix $(A_{k\ell})_{k, \ell \in [n]}$ is a $n \times n$ matrix of independent and identically distributed (i.i.d.) standard Gaussian $\mathcal{N}(0, 1)$ random variables (RV). Notice that each $A_{k\ell}$ variable has a $(\alpha\sqrt{n})^{-1}$ normalizing factor. The positive number α is an extra parameter reflecting the interaction strength.

It is our belief that understanding such a simplified model will provide key information and guidance to analyze more sophisticated and realistic random models.

*Supported by CNRS Project 80 Prime — KARATE.

Remark 1. Although such a matrix A is a complex random object, Large Random Matrix Theory (RMT) provides a number of valuable results, among which the almost sure (a.s.) convergence of the spectral radius and the spectral norm

$$\rho(A/\sqrt{n}) \xrightarrow[n \rightarrow \infty]{a.s.} 1 \quad \text{and} \quad \|A/\sqrt{n}\| \xrightarrow[n \rightarrow \infty]{a.s.} 2,$$

and also the a.s. convergence in distribution (\mathcal{D}) of the spectral measure to the circular law (see for instance [6]):

$$(a.s.) \quad \frac{1}{n} \sum_{k \in [n]} \delta_{\lambda_k(A/\sqrt{n})} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \frac{\mathbf{1}_{\{x^2+y^2 \leq 1\}}}{\pi} dx dy.$$

As a consequence, the normalized interaction matrix A/\sqrt{n} has a macroscopic effect, non vanishing asymptotically on the dynamics of system (1).

Understanding the dependence between factor $\alpha > 0$ and the number of surviving species is the key issue addressed in this paper. A complementary result addressing the elliptical random matrix model by means of theoretical physics methods can be found in [7] and [8].

Notice that in [9], Bizeul *et al.* have described an appropriate scaling for α , namely $\alpha \sim \sqrt{2 \log(n)}$, to guarantee the survival of all species. For fixed $\alpha > 0$, which we will assume in the sequel, a consequence of Dougoud *et al.* [10] is that some species will go to extinction.

2 Equilibrium and stability results

A key element to understand the dynamics of (1) is the existence of an equilibrium $\mathbf{x}^* = (x_k^*)_{k \in [n]}$ such that

$$x_k^* \left(1 - x_k^* + \frac{(A\mathbf{x}^*)_k}{\alpha\sqrt{n}} \right) = 0 \quad \forall k \in [n], \quad (3)$$

and the study of its stability, that is the convergence of a solution \mathbf{x} to the equilibrium \mathbf{x}^* : $\mathbf{x}(t) \xrightarrow[t \rightarrow \infty]{} \mathbf{x}^*$ if $\mathbf{x}(0)$ is sufficiently close to \mathbf{x}^* .

It is well known that for LV equations, the fact that $\mathbf{x}(0) > 0$ (componentwise) implies that $\mathbf{x}(t) > 0$ for every $t > 0$, but one can have some components $x_k(t)$ of $\mathbf{x}(t)$ vanishing to zero. We hence only consider non-negative equilibria $\mathbf{x}^* \geq 0$.

Notice that the situation differs substantially whether $\mathbf{x}^* > 0$ or \mathbf{x}^* has vanishing component(s). In the former case, the equilibrium set of equations becomes a linear equation:

$$\mathbf{x}^* = \mathbf{1} + \frac{A\mathbf{x}^*}{\alpha\sqrt{n}}$$

whose positive solution has been studied in [9] (notice that the existence of such a solution requires $\alpha \gg \sqrt{2 \log(n)}$).

In the latter case, either $x_k^* = 0$ or $1 - x_k^* + \frac{(A\mathbf{x}^*)_k}{\alpha\sqrt{n}} = 0$. The equilibrium equations become ill-posed as there might be many equilibria. In any case, we no longer remain in the comfort zone of a linear matrix equation. However, relying on standard properties of dynamical systems, see for instance [11,

Theorem 3.2.5], a necessary condition for the equilibrium \mathbf{x}^* to be stable is that

$$1 - x_k^* + \frac{(A\mathbf{x}^*)_k}{\alpha\sqrt{n}} \leq 0.$$

This casts the problem of finding a non negative equilibrium into the class of Linear Complementarity Problems (LCP).

Linear Complementarity Problem (LCP). LCP is a class of problems from mathematical optimization which in particular encompasses linear and quadratic programs; standard references are [12, 13]. Given a $n \times n$ matrix M and a $n \times 1$ vector \mathbf{q} , the associated LCP denoted by $LCP(M, \mathbf{q})$ consists in finding two $n \times 1$ vectors \mathbf{z}, \mathbf{w} satisfying the constraints:

$$\begin{cases} \mathbf{z} \geq 0, \\ \mathbf{w} = M\mathbf{z} + \mathbf{q} \geq 0, \\ \mathbf{w}^T \mathbf{z} = 0 \quad (\Leftrightarrow \quad w_k z_k = 0 \quad \forall k \in [n]). \end{cases} \quad (4)$$

Since \mathbf{w} can be inferred from \mathbf{z} , we denote $\mathbf{z} \in LCP(M, \mathbf{q})$ if (\mathbf{w}, \mathbf{z}) is a solution of (4).

A theorem by Murty [14] states that the $LCP(M, \mathbf{q})$ has a unique solution (\mathbf{w}, \mathbf{z}) iff M is a P -matrix, that is:

$$\det(M_{\mathcal{I}}) > 0, \quad \forall \mathcal{I} \subset [n], \quad M_{\mathcal{I}} = (M_{k\ell})_{k, \ell \in \mathcal{I}}.$$

Let us denote $\check{A} = \frac{A}{\alpha\sqrt{n}}$. Gathering the constraints of the equilibrium \mathbf{x}^* defined in (3), we get:

$$\begin{cases} \mathbf{x}^* \geq 0, \\ 1 - x_k^* + (\check{A}\mathbf{x}^*)_k \leq 0, \\ x_k^* (1 - x_k^* + (\check{A}\mathbf{x}^*)_k) = 0. \end{cases}$$

Otherwise stated, $\mathbf{x}^* \in LCP(I - \check{A}, -\mathbf{1})$.

The equilibrium \mathbf{x}^* and its stability. For a generic LV system

$$\frac{dy_k(t)}{dt} = y_k(r_k + (B\mathbf{y})_k), \quad k \in [n], \quad (5)$$

Takeuchi and Adachi provide a criterion for the existence of a unique equilibrium \mathbf{y}^* and the global stability of LV systems, see Theorem 3.2.1 in [11].

Theorem 1 (Takeuchi and Adachi, see [11]). *If there exists a positive diagonal matrix Δ such that $\Delta B + B^T \Delta$ is negative definite, then $LCP(-B, \mathbf{r})$ admits a unique solution and there is a unique equilibrium \mathbf{y}^* to (5), which is globally stable:*

$$\forall \mathbf{y}_0 > 0, \quad \begin{cases} \mathbf{y}(0) = \mathbf{y}_0 \\ \mathbf{y}(t) \text{ satisfies (5)} \end{cases}, \quad \mathbf{y}(t) \xrightarrow[t \rightarrow \infty]{} \mathbf{y}^*.$$

Combining this result (with $I - \check{A} = -B$) with results from RMT, we can guarantee the existence of a globally stable equilibrium \mathbf{x}^* of (1) for a wide range of α 's.

Proposition 2. *Let $\alpha > \sqrt{2}$, then almost surely, matrix*

$$(I - \check{A}) + (I - \check{A})^T$$

is eventually positive definite : with probability one, for a given realization of the matrix \check{A}_ω , there exists $N(\omega)$ such that for $n \geq N(\omega)$, $(I - \check{A}_\omega) + (I - \check{A}_\omega)^T$ is positive definite. In particular, there exists a unique globally stable equilibrium $\mathbf{x}^* \in \text{LCP}(I - \check{A}_\omega, -1)$.

Proof. We have

$$I - \check{A} + I - \check{A}^T = 2I - \frac{\sqrt{2}}{\alpha} \frac{1}{\sqrt{n}} \left(\frac{A + A^T}{\sqrt{2}} \right).$$

Notice that $(A + A^T)/\sqrt{2}$ is a symmetric matrix with independent $\mathcal{N}(0, 1)$ entries above the diagonal (the distribution of the diagonal entries is different from the off-diagonal entries, with no asymptotic effect). In this case, the largest eigenvalue of the normalized matrix (or equivalently its spectral norm since the matrix is symmetric) a.s. converges to the right edge of the support of the semi-circle law (see [15, Theorem 5.2]) :

$$\lambda_{\max} \left(\frac{A + A^T}{\sqrt{2n}} \right) \xrightarrow[n \rightarrow \infty]{a.s.} 2.$$

Now if $\alpha > \sqrt{2}$, we can conclude that (a.s.) eventually the smallest eigenvalue of $2I - (\check{A} + \check{A}^T)$ is positive. We can then rely on Theorem 1 to conclude. \square

3 Surviving species : A heuristics

In Section 2, we have presented a theoretical guarantee, condition $\alpha > \sqrt{2}$, for the existence of a globally stable equilibrium \mathbf{x}^* to (1). As \mathbf{x}^* depends on the realization of matrix A , it is a random vector that will feature vanishing components as $\alpha > 0$ is fixed and does not depend on n .

In this section, we study the proportion of \mathbf{x}^* 's non-vanishing components, referred to as surviving species in an ecological context. The vanishing components correspond to the species going to extinction : if $x_k^* = 0$ then $x_k(t) \xrightarrow[t \rightarrow \infty]{} 0$. Associated to \mathbf{x}^* , we introduce :

$$\mathcal{S} = \{i \in [n], x_i^* > 0\}, \quad \hat{p} = \frac{|\mathcal{S}|}{n}, \quad \hat{\sigma}^2 = \frac{1}{|\mathcal{S}|} \sum_{i \in [n]} (x_i^*)^2.$$

Denote by $Z \sim \mathcal{N}(0, 1)$ a standard Gaussian random variable and by Φ the cumulative Gaussian distribution function :

$$\Phi(x) = \int_{-\infty}^x \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du. \quad (6)$$

Heuristics 1. Let $\alpha > \sqrt{2}$. The following system of two equations and two unknowns (p, σ)

$$\sigma \sqrt{p} \Phi^{-1}(1-p) + \alpha = 0, \quad (7)$$

and

$$1 + \frac{2\sigma\sqrt{p}}{\alpha} \mathbb{E}(Z | Z > -\delta) + \frac{\sigma^2 p}{\alpha^2} \mathbb{E}(Z^2 | Z > -\delta) = \sigma^2 \quad (8)$$

where $\delta = \frac{\alpha}{\sigma\sqrt{p}}$, admits a unique solution (p^*, σ^*) and

$$\hat{p} \xrightarrow[n \rightarrow \infty]{a.s.} p^* \quad \text{and} \quad \hat{\sigma} \xrightarrow[n \rightarrow \infty]{a.s.} \sigma^*.$$

Simulations. We fix $n = 500$, draw L independent realizations of matrices $A^{(i)}$, compute corresponding equilibria $\mathbf{x}^{*(i)}(\alpha)$ and related quantities $(\hat{p}^{(i)}(\alpha), \hat{\sigma}^{(i)}(\alpha))$ for fixed $\alpha > 0$. We then compare the empirical Monte Carlo (MC) averages

$$\hat{p}_L(\alpha) = \frac{1}{L} \sum_{i=1}^L \hat{p}^{(i)}(\alpha) \quad \text{and} \quad \hat{\sigma}_L(\alpha) = \frac{1}{L} \sum_{i=1}^L \hat{\sigma}^{(i)}(\alpha)$$

to their theoretical counterparts $p^*(\alpha), \sigma^*(\alpha)$ with $L = 400$. As shown in Figure 1, the matching is remarkable.

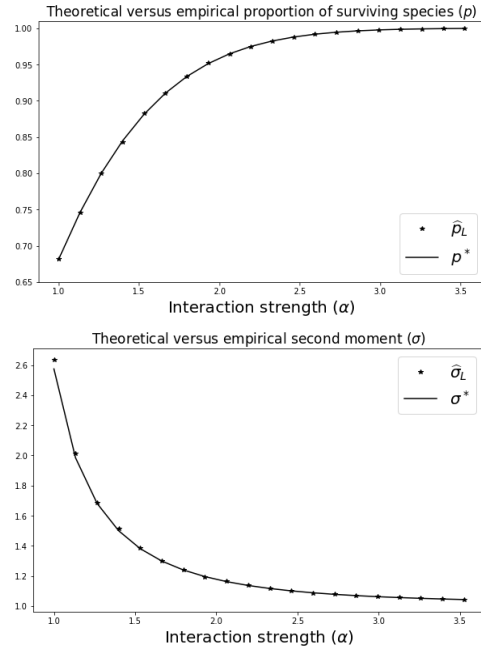


FIGURE 1 – The plot represents a comparison between the theoretical proportion of surviving species $p^*(\alpha)$ (up) and second moment (down) computed as solutions of (7)-(8), and their empirical MC counterpart $(\hat{p}_L(\alpha), \hat{\sigma}_L(\alpha))$. The parameter α on the x-axis ranges from 1 to $\sqrt{2} \log(n) \simeq 3.53$. The threshold $\alpha > \sqrt{2}$ represents the theoretical guarantee to have a stable equilibrium ; $\alpha = \sqrt{2} \log(n)$ is the upper-limit above which we have no extinction ($p^* = 1$). Notice that for $\alpha \in [1, \sqrt{2}]$, the heuristics shows a remarkable matching with the empirical data despite no theoretical guarantees.

4 Construction of the heuristics

We successively establish (7) and (8).

Equation (7). We first recall a result on order statistics of a Gaussian sample (see [16]). Consider a family $(Z_k)_{k \in [n]}$ of i.i.d. random variables $\mathcal{N}(0, 1)$ and their order statistics

$$Z_1^* \leq Z_2^* \leq \dots \leq Z_n^*.$$

Consider index $[n\alpha] \in [n]$ where $\alpha \in (0, 1)$ is fixed, then the typical location of $Z_{[n\alpha]}^*$ is $\Phi^{-1}(\alpha)$, where Φ is defined in (6) :

$$Z_{[n\alpha]}^* \simeq \Phi^{-1}(\alpha) \quad \text{as} \quad n \rightarrow \infty. \quad (9)$$

Intuitively, (9) follows from the empirical approximation :

$$\Phi(Z_{[\alpha n]}^*) \simeq \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, Z_i]}(Z_{[\alpha n]}^*) = \frac{[\alpha n]}{n}.$$

Let \mathbf{x}^* be the equilibrium of (1) and consider the RV :

$$\check{Z}_k = \frac{1}{\alpha\sqrt{n}} \sum_{i \in \mathcal{S}} A_{ki} x_i^*.$$

Assume that asymptotically, the x_i^* 's are i.i.d. and independent from the A_{ki} 's. A natural consequence is that \check{Z}_k is asymptotically Gaussian and that the \check{Z}_k 's are uncorrelated. Suppose also that \mathcal{S} is deterministic (which is clearly a limitation of this computation) ; denote by $\sigma^2 = \mathbb{E}(x_1^*)^2$.

$$\mathbb{E}\check{Z}_k = 0 \quad \text{and} \quad \mathbb{E}\check{Z}_k^2 = \frac{1}{\alpha^2 n} \sum_{i \in \mathcal{S}} \mathbb{E}A_{ki}^2 \mathbb{E}(x_i^*)^2 = \frac{p^*}{\alpha^2} \sigma^2.$$

We now introduce the $\mathcal{N}(0, 1)$ random variables $(Z_k)_{k \in [n]}$ such that $Z_k = \alpha/(\sigma\sqrt{p^*})\check{Z}_k$. Consider the equilibrium $\mathbf{x}^* = (x_k^*)_{i \in [n]}$. If $k \in \mathcal{S}$, that is $x_k^* > 0$, we have

$$1 - x_k^* + \frac{(A\mathbf{x}^*)_k}{\alpha\sqrt{n}} = 0 \Rightarrow 1 + \frac{(A\mathbf{x}^*)_k}{\alpha\sqrt{n}} = 1 + \frac{\sigma\sqrt{p^*}}{\alpha} Z_k > 0.$$

If $k \notin \mathcal{S}$ then

$$1 + \frac{(A\mathbf{x}^*)_k}{\alpha\sqrt{n}} = 1 + \frac{\sigma\sqrt{p^*}}{\alpha} Z_k \leq 0.$$

Otherwise stated,

$$\begin{cases} Z_k \leq -\alpha/(\sigma\sqrt{p^*}) & \text{if } k \notin \mathcal{S}, \\ Z_k \geq -\alpha/(\sigma\sqrt{p^*}) & \text{if } k \in \mathcal{S}. \end{cases}$$

Considering the order statistics of the Z_k 's we obtain :

$$Z_1^* \leq \dots \leq Z_i^* \leq -\frac{\alpha}{\sigma\sqrt{p^*}} \leq Z_{i+1}^* \leq \dots \leq Z_n^*.$$

Now, there are exactly $n - |\mathcal{S}| = n(1 - p^*)$ indices before the threshold so $i = n(1 - p^*)$ and $Z_i^* \simeq -\frac{\alpha}{\sigma\sqrt{p^*}}$. Relying on (9) and replacing p^* by its (supposedly existing) limit p , we finally obtain $\Phi^{-1}(1 - p) = -\frac{\alpha}{\sigma\sqrt{p}}$, which is exactly (7).

Equation (8). Let $k \in \mathcal{S}$, then

$$1 - x_k^* + \frac{(A\mathbf{x}^*)_k}{\alpha\sqrt{n}} = 0 \Leftrightarrow x_k^* = 1 + \frac{\sigma\sqrt{p^*}}{\alpha} Z_k, \quad Z_k \sim \mathcal{N}(0, 1)$$

provided that $Z_k > -\delta := -\frac{\alpha}{\sigma\sqrt{p^*}}$. Taking the square and the expectation conditionnally to $Z_k > -\delta$, we end up with

$$\mathbb{E}(x_k^*)^2 = 1 + \frac{2\sigma\sqrt{p^*}}{\alpha} \mathbb{E}(Z \mid Z > -\delta) + \frac{\sigma^2 p^*}{\alpha^2} \mathbb{E}(Z^2 \mid Z > -\delta),$$

which is the desired result, once replacing p^* by p .

5 Concluding remarks

In this paper, we provide a heuristics to evaluate the proportion of surviving species in a large LV system. Simulations show a striking matching with the empirical data.

Two questions naturally arise. First, is it possible to prove mathematically this heuristics? The dependence induced by the LCP procedure seems a priori difficult to handle. Second, would it be possible to extend this heuristics to non-centered elliptical matrix models as in [7]? We are confident that this should be possible, at least in the non-centered i.i.d. case.

Références

- [1] J. Hofbauer and K. Sigmund, Evolutionary games and population dynamics, *Cambridge University Press*, 1998.
- [2] K. Z. Coyte, J. Schluter, and K. R. Foster “The ecology of the microbiome : Networks, competition, and stability,” *Science*, vol. 350, pp. 663–666, 2015.
- [3] J. C. Moore, P. C. de Ruiter, and H. W. Hunt. 1993. “Influence of productivity on the stability of real and model ecosystems,” *Science*, vol. 261, pp. 906–908, 1993.
- [4] T. J. Case “Invasion resistance arises in strongly interacting species-rich model competition communities,” *Proc. Natl. Acad. Sci. USA*, vol. 87, pp. 9610–9614, 1990.
- [5] S. Allesina and S. Tang, “Stability criteria for complex ecosystems,” *Nature*, vol. 483, no. 7388, pp. 205, 2012.
- [6] C. Bordenave and D. Chafaï, “Around the circular law,” *Probab. Surv.*, vol. 9, pp. 1–89, 2012.
- [7] G. Bunin, “Ecological communities with lotka-volterra dynamics,” *Physical Review E*, vol. 95, no. 4, pp. 042414, 2017.
- [8] T. Galla, “Dynamically evolved community size and stability of random Lotka-Volterra ecosystems,” *EPL (Europhysics Letters)*, vol. 123, no. 4, 2018.
- [9] P. Bizeul and J. Najim, “Positive solutions for large random linear systems,” *arXiv :1904.04559, to be published in Proceedings of the A.M.S.*, 2021.
- [10] M. Dougoud, L. Vinckenbosch, R. P Rohr, L-F. Bersier, and C. Mazza, “The feasibility of equilibria in large ecosystems : A primary but neglected concept in the complexity-stability debate,” *PLoS computational biology*, vol. 14, no. 2, pp. e1005988, 2018.
- [11] Y. Takeuchi, *Global dynamical properties of Lotka-Volterra systems*, World Scientific, 1996.
- [12] K. G. Murty and F-T. Yu, *Linear complementarity, linear and nonlinear programming*, vol. 3, Citeseer, 1988.
- [13] R. W Cottle, J-S. Pang, and R.E. Stone, *The linear complementarity problem*, SIAM, 2009.
- [14] K. G. Murty, “On the number of solutions to the complementarity problem and spanning properties of complementary cones,” *Linear Algebra and its Applications*, vol. 5, no. 1, pp. 65–108, 1972.
- [15] Z. D. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, Springer Series in Statistics. Springer, New York, second edition, 2010.
- [16] A. Balkema and L. De Haan, “Limit distributions for order statistics. i,” *Theory of Probability & Its Applications*, vol. 23, no. 1, pp. 77–92, 1978.