Evolutionary behaviour of 'inflating' random real matrices for economy or biology: stasis statistics of vector iterations upon growth

Henri Benisty

Journal of Physics: Complexity

PAPER • OPEN ACCESS

Evolutionary behaviour of 'inflating' random real matrices for economy or biology: stasis statistics of vector iterations upon growth

To cite this article: Henri Benisty 2022 *J. Phys. Complex.* **3** 025006

IOP Publishing J.Phys.Complex. 3 (2022) 025006 (20pp) https://doi.org/10.1088/2632-072X/ac718f

Journal of Physics: Complexity

OPEN ACCESS

RECEIVED
17 January 2022

REVISED
6 May 2022

ACCEPTED FOR PUBLICATION
19 May 2022

PUBLISHED
6 June 2022

PAPER

Evolutionary behaviour of 'inflating' random real matrices for economy or biology: stasis statistics of vector iterations upon growth

Henri Benisty*

Laboratoire Charles Fabry, Institut d'Optique Graduate School, Université Paris-Saclay, Palaiseau, France
Laboratoire Interdisciplinaire des Energies de Demain (LIED), Université de Paris, CNRS UMR 8236, Paris, France
* Author to whom any correspondence should be addressed.

E-mail: henri.benisty@institutoptique.fr

Keywords: random matrix theory, eigenvalues and eigenvectors, punctuated equilibria, matrix inflation, stasis, Schumpeter, q-exponential

Abstract

A scheme is proposed for describing stasis and transitions in evolutionary contexts defined by a growing number of interrelated items. These items could be genes/species in biology, or tools/products in economy. The target is a frame to describe the advent of stasis marked by dominant species or dominant objects (car, smartphone) between sharp transitions (quakes). The tool of random matrices is adapted to add an explicit varietal growth, through an 'inflation' of a real random matrix (Ginibre set), by regularly adding a line and a column, such a matrix operating at each unit time step on the evolving vector, $U(t+1) = M(t)U(t)$. In this view, $U(t) \propto \log(C(t))$ with $C(t)$ the vector of abundances of genes in a gene pool, or of abundance of tools in a multi-sector production economy (as in Leontieff matrices of sector-wise productions). $U(t)$ is trending toward the eigenvector with the largest-modulus eigenvalue (ev) $U^{(N)}$ for the current $N(t)$. Most times, the next such eigenvector $U^{(N+1)}$ of the inflated matrix is close and mostly colinear to $U^{(N)}$. But, as time goes and N also grows, a wholly unrelated eigenvector may acquire a larger modulus ev and become the new attractor. Thus, there are slowly-moving stases punctuated by quakes. The leading-ev-modulus competition is elucidated, as well as the main features of the stasis duration distribution it entails, akin to a q-exponential law. This is done by means of a two-particles model of Brownian motion taking place with an N-dependent drift and diffusion. To minimally flesh the issue, a set of bibliographic data (yearly number of papers among all journals of a domain) is used, lending support to the vector-colinearity-based methods used for the detection of stases vs quakes. Hints are given for various developments tackling the appropriateness of the model to various growth contexts, e.g. with sparse network or with death and not only novelty/birth.

1. Introduction: novelty, growth and random matrices

1.1. Growth and novelty in life and in economy

Life on earth, as well as human economy at large, both entangle growth and evolution. Both are acknowledged to expand, overall, in terms of diversity and quantity, but through a process far from linear or regular. Punctuated equilibria, long stasis separated by short and much quicker changes or 'quakes', embody one such recognized way to describe life evolution beyond the 'modern synthesis' of Darwinian theory. As for economy, its many crisis, exogenous or endogenous, financial or industrial, while feeding debates among various schools of thoughts, are undoubtedly key phenomena accruing rapid changes, Schumpeter remaining famous for viewing their role as 'destructive creation' (references will appear in the next subsection).

Novelty is hard to properly define in both cases, being neither purely instantaneous nor purely gradual. But it plays an essential role in economic drifts, technical innovation, new energy sources and information management capabilities, all feeding huge socio-economic changes, e.g. the smartphone recently. In life description, novelty is even harder to spot *per se*. Common mutations could be thought to produce new genes and 'new proteins', but many further stages intervene before *bona fide* speciation (the separation of a new species unable to reproduce with the parent one) and yet other stages before a large shift in the gene pool composition occurs. The gene pool may be used to conveniently keep track of evolutionary events (of its 'accounting' part, not operating selection directly, as Gould, the promoter of punctuated equilibria in evolution, insists). The debate on biological evolutionary hierarchies (see e.g. the holobiont concept) is still going on. Nevertheless, one can think of key new metabolic capabilities as novelties that shift the fitness in the framework of speciation (say, among carbohydrates, among nitrogen-rich nutrients), or of key thermal, mechanical and transport capabilities (feathers and wings), or yet of key sensory capabilities (eyes).

We will give in the next subsection some views on the literature that point out the gap that we want to address in this paper relating growth and the stasis/quick evolution succession (punctuated equilibria) with the tools of linear algebra and basic spectral (eigenvalues (ev's)) theory.

1.2. Growth and evolutionary models, a partial view

The literature on the understanding of either life evolution or economy evolution is huge. Chronologically, life and Earth evolution was believed to be due, before Lyell and Darwin in the 19th century, to a sequence of catastrophic events. As life history started to become clear and to be millions years long, gradualism gave a number of keys to grasp main features, and last but not least in life, explain the variety of species around phylogeny and a phylogenetic tree. In economy, equilibrium at the meeting point of consumers and producers 'utilities' was also a tenet of explanations around 1900, with crisis being seen as exogenous (e.g., caused by climate exceptional events or by variable gold and silver supplies). Schumpeter's famous proposal of destructive creation together with Keynes period and its broader views to economical agents motivations, e.g. their psychology at 'Minsky moments' of tidal reversal, was a sign that stability of (growing) complex systems was intrinsically not granted, especially with a large financial sector. The reader can turn here to textbooks: 'Darwinism Evolving' by Depew and Weber [1] and 'Debt, Innovations, and Deflation' by Raines and Leathers [2].

In the period 1970–1990, physics and biology also debated about growth with steps or stasis. In physics, self-organized criticality by Bak and Sneppen, became a new paradigm, readily considered for punctuated equilibrium [3]. In biology, the punctuated equilibria by Gould and Eldredge (detailed in reference [4]) had already deeply questioned the gradualism of Darwinian synthesis, which had to be critically revised [1]. At the same period, May famously proposed an analysis of complex systems based on random matrices and their ev's [5] (basics of spectral theory can be found in textbooks [6]), evidencing the May–Wigner transition. Linearity was here obtained by the study of the Jacobian at a stationary point, while the exact overall ecosystem dynamics would rather obey some generalized flavour of Lotka–Volterra equations, the reference prey–predator model.

To our own surprise, few papers, comparatively, have investigated the use of random matrices to produce a punctuated evolutionary growth. At this stage, it may be good to recall the breadth of the topic that have been addressed by random matrix theory (RMT) since, say, Dyson's enlightening contributions [7]. On the mathematical side, there are deep relations with equations of fundamental interest such as the Burgers equation [8]. The dynamics of random matrices following a matrix-level random walk is a topic of interest at a high level [9] with ramifications in another topic coined by the term 'growth': the growth of polymer or of surface shapes made of crystal steps, see [10, 11]. We are not able, here, to connect our views to growth phenomena of this kind.

We note that a nonlinear analogue of the May–Wigner transition was only proposed a few years ago [11]. A review of physical approaches to complex systems such as reference [12] could also help the reader to grasp issues tackled in recent decades.

Now if we come back to economy to assess the impact of random matrices, Allesina and Tang contribution [13] is a synthetic description of the research advances and status. There is also a large piece of economic literature that deals with innovation, a prime engine of growth, focussing on the inner workings of the disruption adoption [14–17]. When the emphasis is on the distribution of innovations, most works deal with novelty in a mean-field spirit, with a 'rain of innovation' on a continuous distribution of actions that spread across some market risk spectrum [18–21]. In bold works of 2010 on evolution models [22, 23], the issue of growth with punctuated equilibria is similarly tackled in economy and in life evolution, but still rather with local rules that modify graphs, not with a genuinely 'agnostic' approach.

As for biology, there are evolutionary models that have successfully produced so-called q-ESS: *quasi evolutionary stable strategies* with stable populations and species, separated by more hectic evolution phases. The tangled nature (TaNa) model [24] is a prominent such example (although its 'noisy' hectic phases can represent relatively long time lapses rather than shorter 'quakes'). Subsequent works by Murase *et al* have provided more insight [25–27]. In reference [28], the relationship to analytically tractable stability criteria was proposed. One of the goals was to test the results against the fossil record distribution of [29], fitted to a 'q-exponential' law. These models are as agnostic as they could be, but they understandably wanted to stick to established paradigms of evolutionary biology such as assessing the role of the genome itself and its intrinsically discrete mutation rules vs other evolutionary causes such as migrations in the kind of dynamics produced by the model. It is thus logical that the models in this rich branch of investigation still have to include some assumptions on reproduction, mutation, and survival, and that they thus make rather marginal use of ev's and eigenvectors. It is interesting to note that another larger series of work has indeed turned to the issue opposite to growth, i.e. massive extinctions or collapse, addressing them in several branches of ecology (and possibly in economy) [30–35], unlike the relative scarcity of descriptions that regard intertwined growth and disruption. There are some descriptions of evolution e.g. following new territory invasions by new people or new species, which can be more understood as sequels of a single singularity rather than a statistical sequence of disruptions.

Hence we found none of such works that would amount, even remotely, to a direct use of random matrices of increasing size, virtually free of explicit rules ('agnostic'), that we will present in the next section. It is difficult to provide a single reason for the lack of attempt to use the powerful universality of random matrix tools to describe growth and disruptions. Maybe the first impression is that matrix randomness is a 'double randomness' that sheds light on stability or through subtle mathematics on a set of specialized dynamics (Burgers equation, KPZ instability (KPZ standing for Kardar Parisi Zhang), advanced gauge theories). The author own impression is that a better understanding of eigenvector dynamics [36–38] will anyway help going in the direction that we propose, filling a similar gap in an alternative manner.

A related point is that matrices that are closer to sensible depictions of the reality rarely operate as endomorphisms, i.e. operators that directly relate ('evolve') vectors in the same space. Instead, in biology, there are matrices that relate genotypes to phenotypes and traits (pleiotropic effects possibly modeled by random matrices [39]), or in economy there are Leontief input–output matrices that relate input and output by sectors [40, 41], and the different nature of in and out vectors, made visually obvious by the rectangular form of many such matrices, does result in interesting applications of random matrices, but mostly in terms of the correlations: this is the field of Wishart matrices, for which the Marçenko–Pastur limit distribution applies, but whose ev's do not translate directly in a view on the stability or the dominant mode of the studied system (however, in various task-solving problems involving artificial neural networks, rectangular matrices are very useful in shrinking the dimension of a problem, say from 10^6 to 10^3 in an agnostic way, or also to inflate dimensionality so as to better tackle nonseparable problems of mesoscopic size). The topic of random matrices has indeed become broad enough for specialists of various disciplines to use it without being easily able to see the correspondences, so it is a victim of its success.

It appears that Cui *et al* recent contributions to this field [42] are helpful in several respects, suggesting that the perceived gaps on application of random matrices are starting to be addressed, thus making the present proposal a timely one.

A last word is in order in this introduction to justify the focus on inflation only and not on disappearance of tools/genes that could of course become extinct. It is likely that extinction will have its own interest in later applications, but that for the time being, it would introduce less relevant complications. Anyway, some components will become silent for very long periods, without the need to zero them explicitly. In economy, to exploit an anecdotal fact, goods as simple as wax have been used for the Latin 'tabula' writing support, which revives today as our modern pads, while wax found many other uses.

1.3. Outline of the paper

In the next section (section 2), the basic ingredients of the model are presented. Its basis is the iteration of an evolving matrix operating on a vector that represents a pool of items. This pool can be a gene pool in a life evolution account or a 'pool of tools' in economy. The matrix at any time is real non-symmetric, thus non Hermitian, and random. It can be thought as a Ginibre set, but not rigorously because the size evolution puts the model partly at odds with RMT, as the probabilistic core of these theories are built on sets of given size N, with the asymptotic large-N quests being undertaken at the next step. The parameters used in the model are very few, unlike more explicit models of growth that use explicit thresholds as triggers to novelty, or some other explicit mechanism (in other words we exploit the agnostic aspect of random matrices).

The two connections of the model are described: how it relates to the growth of a gene pool or a pool of tools on the one hand through its vector and the inflation of a random matrix, and how the ev's mathematically evolve for a single inflation step $N \to N + 1$.

In the subsequent section (section 3), typical growth sequences for a model size N growing up to 150 exemplify its operation, notably the quakes and stasis and the role of gradual renewal. The tracking of the quakes can then be implemented and the stasis duration statistics obtained. A drift-diffusion model is discussed in this section, which provides a simplified account of the top ev's dynamics and their competition. It shall help bridging the accurate numerical approach to almost equally accurate analytical ones. This section also relates the resulting stasis distribution to a q-exponential law of parameter $q \simeq 1.24$ close to that of reference [29].

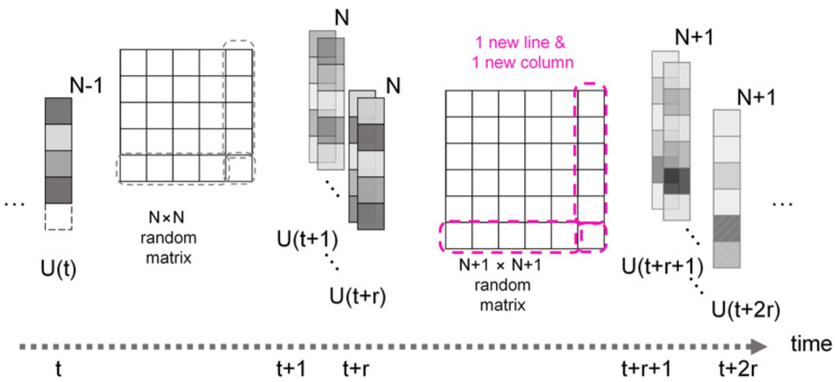

Figure 1. Sketch of vector U and matrices used to form $U(t)$ by inflation, here from $N-1$ at t to $N+1$ till $t+2r$ with r the number of repeat sequences at given N. During a sequence at given N, the convergence to the eigenvector $U^{(1)}$ with the largest $|\lambda|$ is akin to a fitness improvement of the whole system, as the less successful ev's and eigenvectors are filtered out.

In the fourth section (section 4), an analysis of selected bibliometry data through the last 6–7 decades is proposed, in order to minimally 'flesh' the model, especially in terms of eigenvector evolution.

In the last section (section 5), the perspectives and application potential to evolutionary biology and economy are discussed in terms of the kind of data that would grant a fair applicability test. Also, important variants to the basic model, such as the sparsity, the non-Hermiticity and the possible use of non-linear ingredients, which correspond by and large to species disappearance in biology, are considered. The conclusion follows.

2. Model basis

We describe the core ingredients of our proposed model, temporarily evading some complexities of twice-degenerate ev's of non-Hermitian matrices for clarity. We use figure 1 as a support. We first consider iteration at constant N, and next the 'true' evolution, whereby iteration and growing N both intervene.

2.1. Constant size N and stability of largest eigenvalue/eigenvector

We start with a model without any change in its pool basis, of given size N.

- We have a discrete time with integer times steps $t = 1, 2, \ldots$.
- The (eco)system pool is described by a state-vector $U = U(t)$ of N elements detailed below. The evolution in time $t \to t+1$ is described by a real-coefficient matrix M that accounts for the aggregated interaction between elements:

$$U(t+1) = MU(t). \tag{1}$$

- U is real-valued, positive or negative: U_i's will be argued to represent 'chemical activities' of the basic building-blocks of the pool having physical concentration $c_i > 0$ according to the relation

$$U_i = K \log(c_i/c_{i,\mathrm{ref}}). \tag{2}$$

Thus negative U_i's, for instance, represent low concentrations. This could justify deleting those elements that are so low that they have no window for physical interaction in the modeled system, and get decoupled. Furthermore, the fact that interactions are thus linearly scaling with an 'activity' could be further consolidated from out-of-equilibrium thermodynamic arguments. We use here the term 'concentration' also in order to connect to the chemists and non-equilibrium physics community. It is clear that for more usual evolutionary biological models or economy models, we could use the total quantity (the extensive one) rather than the intensive one c_i. For very ubiquitous objects in economy (wheels, batteries, ...), one would recourse to quantities per capita for instance. Hence the approach of the concentration is partly a metaphor. The interest of going to a log when building up the vector U_i is to give relevance to familiar random matrices whose entries are a-dimensioned and whose standard deviation remains not too far from unity. The elements c_i labelled by $\{i \in 1, \ldots, N\}$ could be, for life evolution modelling, genes (or alleles) of a *gene pool*, whatever the species carrying them, thus in a more abstract way than the TaNa model [24]. For economy modelling, they could be different tools of the *tool pool* used by different industries and other socio-economical units, in order to produce {tools + consumer products + services}. The

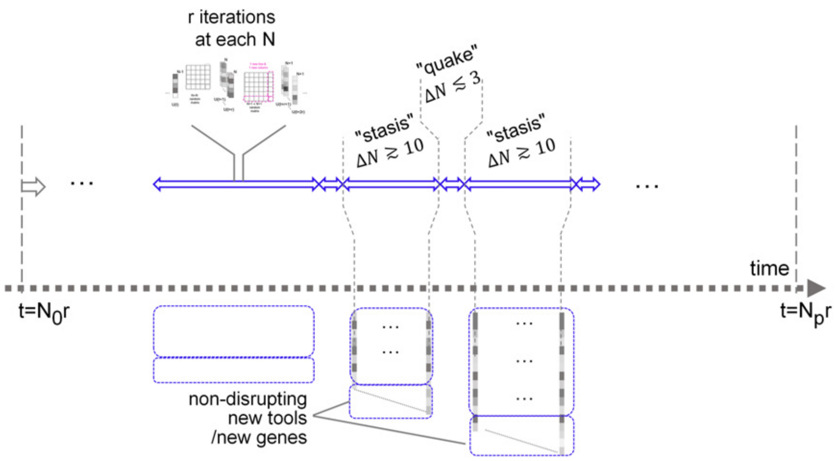

Figure 2. Longer time sketch of evolution, outlining the different possibilities vs growth of N. There are long period during which U does not change, essentially because it tracks an eigenvector that undergoes little changes in spite of the addition of new lines and columns (non-orthogonal contributions exist). These are evolutionary stasis. There are also a few short periods ('quakes') whereby the addition of very few elements (arbitrarily we propose less than three in the figure, but this depends on the renewal rate, among other factors) fully changes the structure of U, fundamentally because the new largest ev is based on a different rather orthogonal eigenvector.

more actual units (species/cells or workshops) share such common elements, the more our $N \gg 1$ random matrix approach is justified. Of course, sparser matrices and with them a graph approach could also be devised.

- In the case where U has a given *fixed* size N, $N \gg 1$, and M is $N \times N$, the first idea is that U represents the limit stable situation only if $U = U^{(1)}$ is the eigenvector of M with the largest modulus $|\lambda|$ among the set of ev's $\{\lambda\}$. Apart the issue of degeneracy, we may order the ev's in descending modulus order, $|\lambda^{(1)}| \geqslant |\lambda^{(2)}| \geqslant \ldots$, so that

$$MU^{(1)} = \lambda^{(1)} U^{(1)}. \tag{3}$$

- We deal with matrices M with randomly chosen entries, and fixed standard deviation $\sigma_M = \frac{1}{N}$, which means that almost surely the spectral radius $\rho(M) = |\lambda^{(1)}|$ is around unity, with documented distributions. The Ginibre ensemble (real non symmetric matrices) that we shall use is however rather the least tractable in this respect.
- It is trivial that the largest eigenvector is an attractor of the process equation (1) at constant N. If $U(t) = \sum a_k(t) U^{(k)}$ is a decomposition on eigenvectors ($k = 1, 2 \ldots$) with $a^{(1)} \neq 0$, its convergence to $U^{(1)}$ is controlled by the difference $|\lambda^{(1)}| - |\lambda^{(2)}|$. This is because

$$MU(t) = \sum \lambda^{(k)} a_k(t) U^{(k)} = \sum a_k(t+1) U^{(k)}, \tag{4}$$

with k spanning the set of λ, hence the evolution of the ratio $|a_2(t+1)/a_1(t)|$ (the longest-lasting components relative weight) can be cast by a per-cycle-rate (log-decrement) α_{12} such that $|a_2(t+1)/a_1(t)| = |\lambda^{(2)}/\lambda^{(1)}| = \exp(-\alpha_{12})$. It typically demands a duration of $T_{12} = \alpha_{12}^{-1}$ time steps to purify U by a factor e vs the second longest-lasting component.
- What we thus implement, especially for the limit case, is the idea that (in bio language) 'gene i in the pool with activity $U_i = K \log(c_i/c_{i,\mathrm{ref}})$ contributes the activity of gene j with a $M_{ji} U_i$ term', and that the stability of the gene pool is nothing but the attainment of the largest ev. The fact that $|\lambda^{(1)}|$ typically exceeds a little unity in our M choice is just compensated by normalizing U at every time step t (using the classical two-norm), noting that this has no impact on the relative values of the U_i's (activity ratio).

Let us eventually note that apart the size N itself, there is no free parameter at this stage. As for the issue of degeneracy, real non-Hermitian matrices have complex conjugate ev's, and nondegenerate real ones (almost surely, i.e. unless accidental degeneracies occur). Because the vectors we use are real, we always use, for nonreal ev's, a combination of two associated ev's and eigenvectors analogous to the cosine representation as $(e^{i\theta} + e^{-i\theta})/2$ with some *common* amplitude ρ, hence the pairing of these non-real ev's is mostly a detail vs the existence of an ordered series of their modulus (bounded by the spectral radius) once adequately paired.

2.2. Ecosystem with growing pool size $N(t)$

The next point is to operate the same algorithm, but to 'inflate the pool'. Here we do this step-wise, at a constant rate $g = 1/r$ per unit time with $r \geqslant 1$ an integer, and we assume that the new element added to the pool readily comes with its cross-capabilities mature. This means that every r iterations (r thus stands for 'repeat'), we increase the size of the matrix M by one line and one column, *taken with the same distribution as the previous additions at former inflation steps*. The only provision is to take care of the $\sigma_M = \frac{1}{N}$ normalization to properly track ev's in later treatments, but note also that the normalization of U removes a kind of global consequence of the choice. The very inception of the new gene/tool (which interacts through the new line + new column) is not our concern (e.g., with which mutation rate, dependent on the gene pool itself and which mechanisms act on it; or for a new tool, what triggered an extra adoptable invention from the previous tools' uses and related social, energetic, or other determinations).

- We separate a stage of r repeats of equation (1), done at a given size $N(t) = N(t+1) = \ldots = N(t+r)$, from the size-change step (hence the $g = 1/r$ rate of the latter). This is intended to leave time for the system to relax toward the next stable position. However, we do not tune r as a function of α_{12}, we leave it fixed for simplicity, typically $r = 5$ to 50 to set ideas. Hence we end up with inflation steps at times $t(N \rightarrow N+1) = Nr + t_0$: the chronology of novelty is r times slower than the basic 'regeneration' process.
- The motion of relaxation to the largest ev/eigenvector can be seen as the system moving to a fitness optimum. The Rayleigh quotient U^+MU/U^+U (the observable value of textbook quantum mechanics), for instance, will gently shift toward the highest ev. The so-called 'Malthusian fitness' can, thus, be related to these quantities as they dictate differential ratio of evolution, with the largest ev the fittest. As $U^{(1)}$ takes over $U^{(2)}$, it has the differential reproduction advantage due to $|\lambda^{(2)}/\lambda^{(1)}| = \exp(-\alpha_{12}) < 1$. However it would be a *fitness of the eigenvector* U, i.e. of an ecosystem configuration, not that of any given species, and furthermore the U are activities, not direct concentrations or number of species members.
- In the algorithm, rather than technically drawing the new column and lines of M at the precise inflation times, we draw a larger M that has the final calculated size N_{p} (the 'present' size if history $1 \rightarrow N$ is in the past) and impose its truncation to $N(t) \leqslant N_{\mathrm{p}}$ (and apply the factor for adapting $\sigma_M(N)$) by extracting the proper subblock. We denote by $M_{N(t)}$ this restricted matrix (the alternative option is 'zero padding' and use of trivial subblock-diagonal matrices to achieve the same result).
- The chronology thus has a period r between growth by {one line and column} (figure 1). Specifically, we first do the repeats and then change the size of M by one unit to let the $N(t)+1$ line and column operate once. Hence, the equation of a sequence of steps that leads from $N-1$ to $N+1$ reads:

$$U(t+r+1) = M_{N(t)+1}U(t+r) = M_{N(t)+1}(M_{N(t)})^r U(t). \tag{5}$$

  In this formulation covering in total $r+2$ steps, we can track the appearance of new components in the vector. At the start, $U(t)$ had $N(t)-1$ nonzero entries. Then $U(t+1) = M_{N(t)}U(t)$ provides, at time $t+1$, the first occurrence of a vector whose $N(t)$th component is the nonzero term $U_{N(t)}(t+1)$ (due to the new matrix line and column introduced after time t). Next come the r repeats. And the next step $t+r+1$ was written explicitly to get the first occurrence of a nonzero $U_{N(t)+1}(t+r+1)$ component.
- We note that it is possible to operate deletion of elements 'i' (in biology, the species carrying the gene *all* become extinct, in industry, the tool disappears in *all* workshops). First think of running the process backward as regards $N(t)$, turning apparitions into deletions. But of course we can delete in the forward time direction. The effective matrix size becomes less than scheduled for pure growth. This is evoked in the final discussion section.
- In a nutshell for what regards the issue of stasis and quakes, the large majority of growth processes ($N \rightarrow N+1$), figure 2, shall not trigger a large change in the eigenvector and will correspond to a stasis. Most growths only shift the iterations close to the earlier ones, accommodating the new member with an eigenvector almost colinear to the previous one. *Some growths, however, disrupt* and introduce a clear 'quake' in the vector, when the maximum ev shifts to a 'different' one, i.e., one with a rather orthogonal eigenvector. This is reminiscent of punctuated growth (stasis and quakes) in evolution, and reminiscent of disruptions/growth crisis in economy (due to 'game-changing inventions').
- Tracking the walk (constrained and yet partly random) of the largest ev's will be helpful to argue of the interest and tractability of our model.

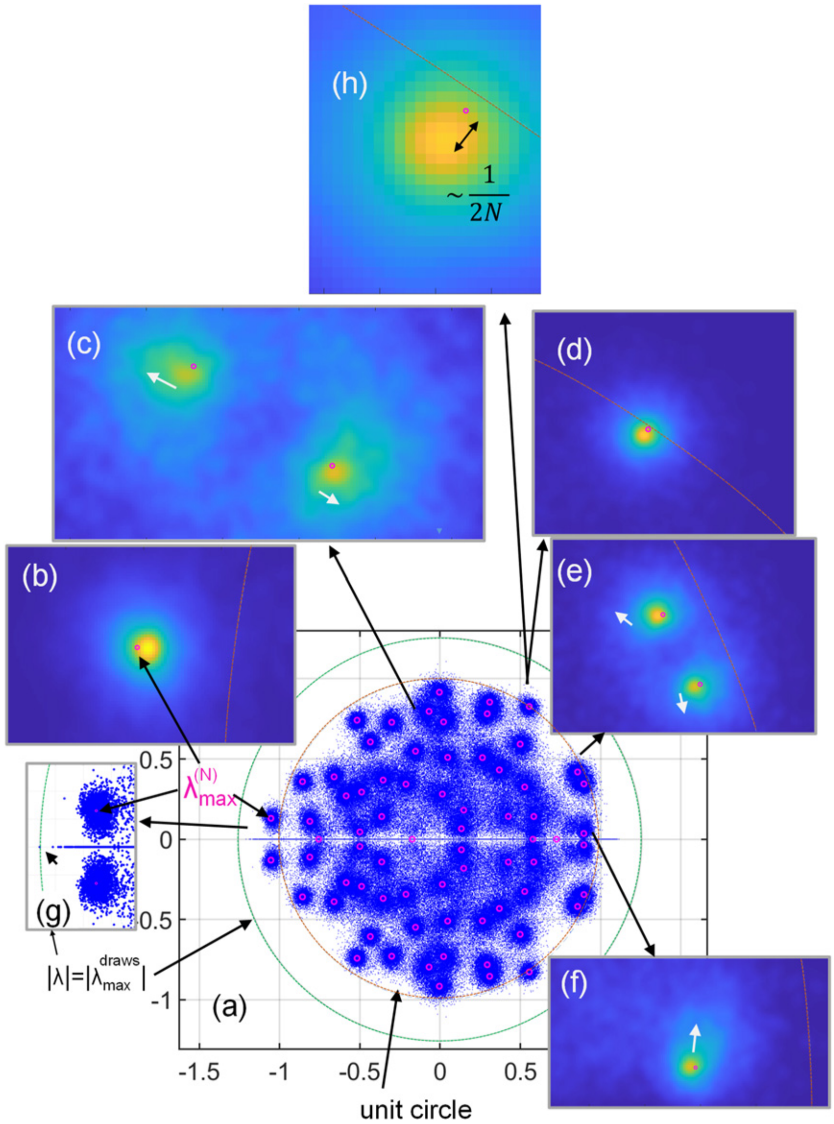

Figure 3. Evolution of ev's in a case study $N = 70, N + 1 = 71, 1/\sqrt{N} \simeq 0.120, 1/2N \simeq 0.007$. (a) Ev's for M at $N = 70$ (magenta circles) and of inflated matrix, $N = 71$, for 6000 random draws of last line and column (blue dots); the unit circle is brownish and the largest spectral radius attained is shown as the green circle. (b–f) Zoom on the density of ev's (i.e. blue dots of (a)) as a guide to the eye, using a squared-Lorentzian convolution with suited colormap. The central yellow spot is generally inward, but the outliers show an oriented pattern in several cases, suggested by white arrow; (g) real axis where the highest draws are obtained (small arrow: largest, on real axis); (h) for a complex ev obeying $|\lambda| \approx 1$ and reasonably isolated, the eigenvector is rather unchanged and the inward shift is $\approx 1/2N$. Hence a future largest ev tends to emerge from the pre-inflation matrix thanks to outlier trends.

2.3. Specifics of eigenvalue/eigenvector evolution upon inflation

The main feature that helps us tracking the stasis and quakes is the way ev's and eigenvectors evolve. This is pictured in figure 3 in a chosen example. The magenta circles are the ev of a $N = 70$ real random matrix with standard deviation $\sigma_M(N) = 1/N$ from a numerical generator for its N^2 Gaussian entries (think of the Ginibre set). As is well known, they fill the unit circle, with a sizable fraction clustering on the real axis and a small exclusion effect around (so-called 'Saturn effect'). The blue fine dots are the ev's of 6000 draws of a 71 × 71 matrix with the same $N = 70$ main part and one random extra line and column, only normalized with the $\sqrt{N}/\sqrt{N+1} \simeq 1 - \frac{1}{2N}$ relative factor.

As can be seen from the main figure, the blue dots for the +1 inflated matrix seem, at first sight, to gently shift isotropically around the $N = 70$ ev's in comparable amounts everywhere, reminiscent of Brownian motion of ev's (a tenet of RMT). However, the density of blue dots is large and saturates (our target in doing so was to make also the extreme shifts visible). To visualize the detail, we convert dots into a continuous density by performing a convolution with a squared-Lorentzian kernel. Most of the cases shown (b)–(f) clearly indicate a first-order inward effect of drift, of the expected $-\frac{1}{2N}$ magnitude. The cases (d) and (h) with an initial ev $\simeq 1$ show, using suitable colormaps and magnifications, that this drift is similar to the extent of the diffusion. However closer examination of various cases show that outliers are not isotropic (see (c) and (e)) to the point where the inward motion is largely compounded by a lateral motion (at constant radius). Also, for the largest starting ev's and more generally those close to the real axis, anisotropies clearly arise. Such effects are part of the organisation of ev's of the Ginibre set, with those ev's closer to the real axis having, qualitatively, more Hermitian repelling features in their interaction pattern. We have done various checks of the statistics of the $N \to N+1$ process to assess that, whatever the exact relative weight of these elements, we had reached in the following a robust overall view of the evolution of ev's. It is thus argued that the effects appearing pictorially in figure 3 are correctly embedded in our further treatments. Notably, we checked that the distribution of the Kth largest ev modulus evolution $|\lambda_{N+1}^{K}| - |\lambda_{N}^{K}|$ had stable typical characteristics of their probability distribution functions (pdf) that are consistent with a 'competition picture' of the ev's. Furthermore, in figure 10 of the appendix, a picture is given of the pdf and cumulative distribution functions (cdf) of the scalar products of eigenvectors associated to the ev of a given K rank ($K = 1$ to 6) before and after an inflation step, that lends support to the above view.

These features taken altogether suggest how the largest ev $\lambda_{\max}^{(N)}$ shall evolve with N. Most ev's shall shift inward. But a few ones can go outward. Among them will, sometimes, be the new largest $|\lambda|$. The important point is that there is no specific correlation of eigenvectors between the present one of largest-ev-modulus $U^{(N)}$ (at step N) and the future one $U^{(N+1)}$ (at step $N+1$) if/once these U's are not 'the same', on account of the different ev's. Consider for instance a ring of radial extent $r \in \left[1 - \frac{k}{N}, 1\right]$, with k an integer. Under the simplified assumption of a constant areal densities of ev's in the complex plane, since this ring represents an area $\simeq 2k\pi/N$ it contains $N \times \frac{2k\pi/N}{\pi} = 2k$ ev's, thus $\sim k$ non-conjugate. One of them can statistically diffuse outward (to counter the inward trend) by more than $1/2N$ with a nonvanishing probability p (at given N). (This is reminiscent of Fisher's geometric model [43] in 2D version for fitness and mutation, although it is coincidental here, but clearly we expect $p \ll 1$ as both the deviation magnitude *and* the direction are at stake). We can guess more about this by multiple draws for a given ($N \to N+1$) inflation step, as we shall do. However, the more interesting point of stasis and quake is addressed by following a whole string of inflations ($N \to N + 1 \ldots \to \ldots N + \Delta N$) till such an event shows up. This must happen, typically once ΔN satisfies $p\Delta N \simeq 1$: it is assumed that p is constant enough, as is plausible if $\Delta N \ll N$. The efforts in the paper are towards justifying the pdf of stasis duration ΔN generated in a given sequence ($r\Delta N$ time-wise), and by the same token to check this pdf independence against the starting and end points, making it as universal as possible at this stage. It is found that the distribution of stasis in the basic process exposed above has a roughly Poissonian 'body', but also features a small-weight longer tail. Thus a specific subsection will be devoted to a positive comparison with a q-exponential law as in [29].

3. Drift-diffusion model for largest eigenvalue competition

3.1. Numerical example of growth with stasis

Let us start by an illustrative example of an eigenvector evolution in the proposed matrix inflation process, obtained numerically with a standard random generator for M. We illustrate on figure 4 the eigenvectors by a map of their components modulus $|U_k|$ occupying the lower right triangle with $N = 1$ to $N = 150$. We chose $r = 25$ repeats at each stage, so there are $25 \times 150 = 3750$ basic iterations. The abrupt changes in U composition cannot be missed. There are also oscillations, this is expected because we start from a random but real U at the process initialisation. Sets of conjugate ev's (real matrix) are then associated with two conjugate eigenvectors (a cosine from two exponentials as said in section 2 above) so as to build up a modulus oscillation. The abruptness can be smoothed by tapering the renewal (U is only partially replaced by the new one) and oscillations can either be made to look like amplitude and phase with standard signal processing techniques (Fourier transform based), or tamed with minor tweaks, such as zeroing the ev's that lie on the negative axis. This was done in figure 4(b) (renewal of $\alpha = 0.2$ instead of full one $\alpha = 1$, and 8.9% of ev's zeroed in the diagonal form of the same 150×150 matrix as in case (a)). This can sometimes lead to a different occurrence of some of the stasis and quakes, as the slow renewal gives the possibility to an eigenvector of surviving through the momentary lead of another eigenvector's ev. The aim of this empirical-numerical basic exploration is to broaden the possible analogies with economy or evolution and not being annoyed by these oscillations. Oscillations are indeed often part of the picture when linear algebra describes a vector evolution, but the most regular oscillations (think of Lotka–Volterra dynamics) are factored in various approaches familiar to audiences of concerned domains.

What appears from a set of examples like this one is that there is no strong N scaling of the stasis, the stable periods have similar sizes on the left and right sides in spite of a factor $\sim$10 in relevant N values. So this must relate to the universal aspects of ev distributions in random matrices. While we are not able at the present stage to propose an *ab initio* mathematical analysis, it seems instructive and useful for further modelling to cast the phenomenon as a competition between the few top ev's that randomly drift in a common effective potential.

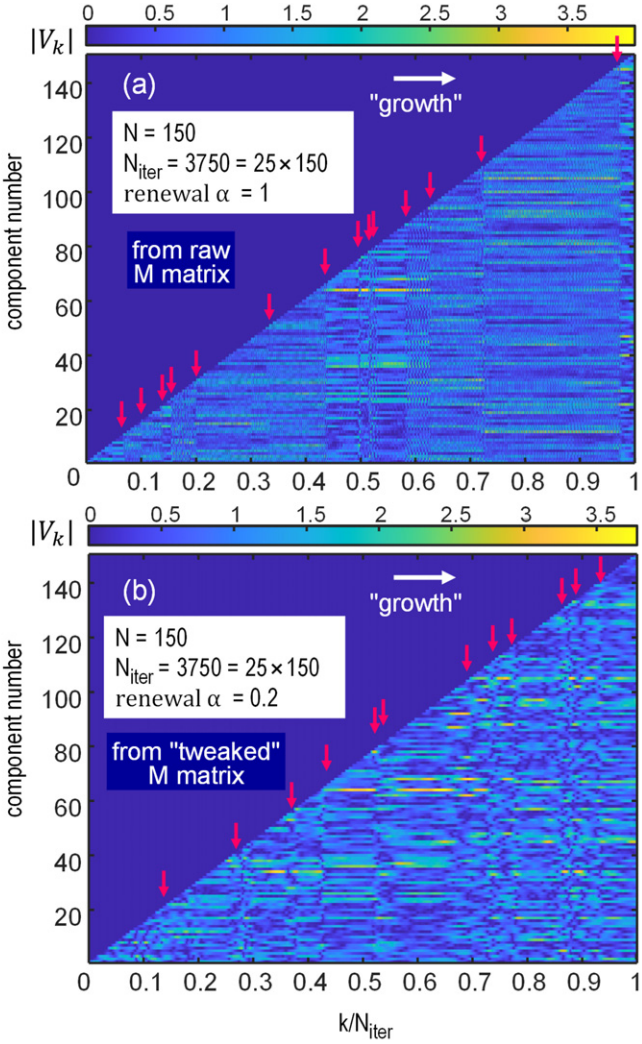

Figure 4. Typical sequences of eigenvectors (given as component modulus $|V_k|$) from numerical examples of growth over $N = 150$ inflation steps (arranged as a upper diagonal matrix and upward ordinate) (a) presents the result of 3750 iteration on a raw random matrix with full renewal at each of the 25 iterations among inflation steps ('growth' direction) with stasis and quakes (magenta arrows) as well as oscillations; (b) presents a similar map but with two 'tweaks': the ev's on the negative axis (<10%) have been made to nearly vanish, and the renewal rate was set at 0.2 only. As a result, oscillations are almost suppressed, and the transitions are more smooth.

The fate of two distinct ev's modulus suffices to get the main trends, requiring however some approximations to sweep all other ev's under the proverbial rug. Here, a steep reflecting mechanism operating only at the bottom of the studied range of ev's appears to provide an adequate model.

3.2. Diffusion-drift of the top eigenvalues

To establish a simple model of ev evolution and competition, it is attempted to analyze the statistics of $|\lambda|$ evolution for inflation at a given rank N, in order to cast a long series evolution in terms of a diffusion-drift model incorporating its local N-dependence (which does not preclude an overall N-independence of the stasis duration pdf, see section 3.4). Thus, the studied quantity is the change of the few top ev's of a randomly inflated matrix in a $N \rightarrow N + 1$ process. Importantly, it is critical to follow ev's based on eigenvectors and not on rank as done by usual software packages. Specifically, one has to track which pair of eigenvectors before and after the inflation yields a large (near unity) normalized scalar product. In this way, one can establish the pdf of the ev motion from any starting point and toward any other nearby value.

This could be seen as a pair distribution function, but it is rather analyzed here as a drift-diffusion model: from any point, the motion has a first statistical momentum that can be seen as a result of a force. It also has a second momentum (spread), that reflects a diffusion process. The details are largely omitted as they are somewhat cumbersome, and the reader is first referred to the numerical result in appendix. It is shown in figure 11 as various surfaces on a $(N, |\lambda|)$ semilog-x basal plane, with some added contours. Minimal technicalities are as follows: the drift and diffusion were calculated for several values of N in a log-spaced series with six values per octave, spanning over 5 octaves (~2.5 decades). This discrete grid of 30 N values is fine enough to provide

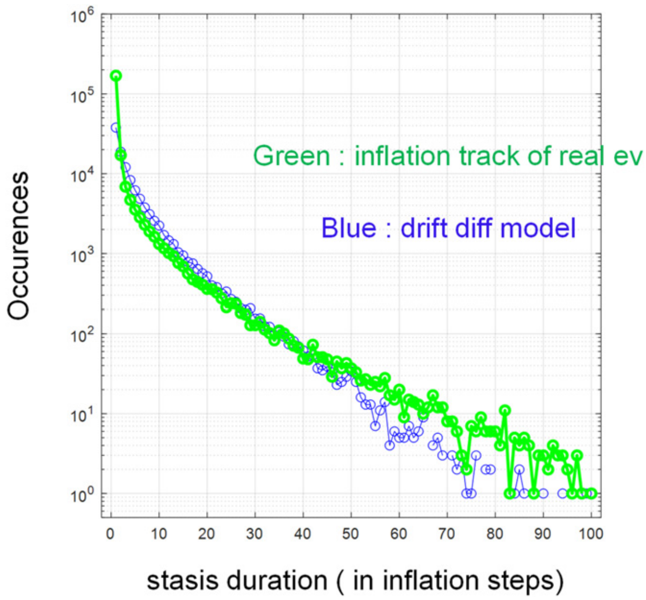

Figure 5. Comparison of stasis statistics from direct tracking of ev's ('along eigenvectors') and from the drift-diffusion two-ev simplified model.

a decent continuous approximation of an N-dependent drift-diffusion statistics. It is hypothesized that this map can be retrieved, asymptotically at least, from an analytical expression, much as many of the universal trends of RMT.

It is further acknowledged that the present version does not target high accuracy, but rather heuristically illustrates the way to get a diffusion-drift model in the absence of theoretical results on ev evolution for inflating matrices. A last noticeable point is that, for each N, a simple interpolated affine dependence on the ev modulus $|\lambda|$ is retained. Since the simplicity of this coarse approach gives acceptably accurate results, the core ingredients of the issue have been properly captured.

3.3. Comparison of direct eigenvalue tracking models and drift-diffusion

It is now possible to tackle the issue of quakes and stasis in terms of ev's modulus competition during an inflation process. The scope of the following paragraphs is to explain the two numerically obtained pdf of figure 5. These pdf describe the stasis duration. They are clearly central either in terms of understanding the underlying processes, or in terms of evaluating whether any application to/from real world is meaningful (in our case, economy or biology): the similarity of the pdf would be a prerequisite.

In the rest of this subsection, a text-based account of the two processes is given. In the following subsection 3.4, figure 6 will be exploited to evaluate the result vs a q-exponential law and also to give supporting elements for the asymptotic N-independence of the observed stasis duration distribution. In the next subsection 3.5, a graphical account, figure 7, is given for both processes. This two-step account avoids simultaneously reporting on the general process and on a particular realisation.

The two curves derive from the two approaches that we first describe below. The first, from which the green curve is derived, is an inflation of actual matrices and a track of their actual eigenvectors to detect if they show quakes at inflation steps. The second, from which the blue curve is derived, makes use of a Brownian motion analogy of the ev moduli on the basis of their approximate drift and diffusion to deliver the duration of stasis based on the crossing of such Brownian motions, since those crossing must represent the quakes bounding the stasis.

Dealing, thus, with the approach of the inflation of actual matrices, with the help gained from the ev and eigenvector visualizations, it is clear that the competition of the top few ev's modulus directly determines these stasis and quakes. The eigenvector has no large change (i.e., a stasis) as long as 'its' ev is not beaten by an emerging larger one from another eigenvector (the eigenvector that wins the iteration competition, supposed to account for the evolutionary processes mentioned in section 2 in economy or evolutionary biology). When it is beaten, the change is large, akin to a disruption or a 'quake'. Thus, for numerical purpose, by simply running an inflation process, then getting at each N value the few first eigenvectors and ev's, and by finally tracking the modulus of successive scalar products of the distinct-modulus eigenvectors $|U_1^N U_1^{N+1}|$ associated to the top distinct ev (hence the subscript 1), a sequence of numbers is obtained that are often close to unity, but occasionally, at quakes, much smaller than unity. Given the marked bi-modal nature of the distribution of

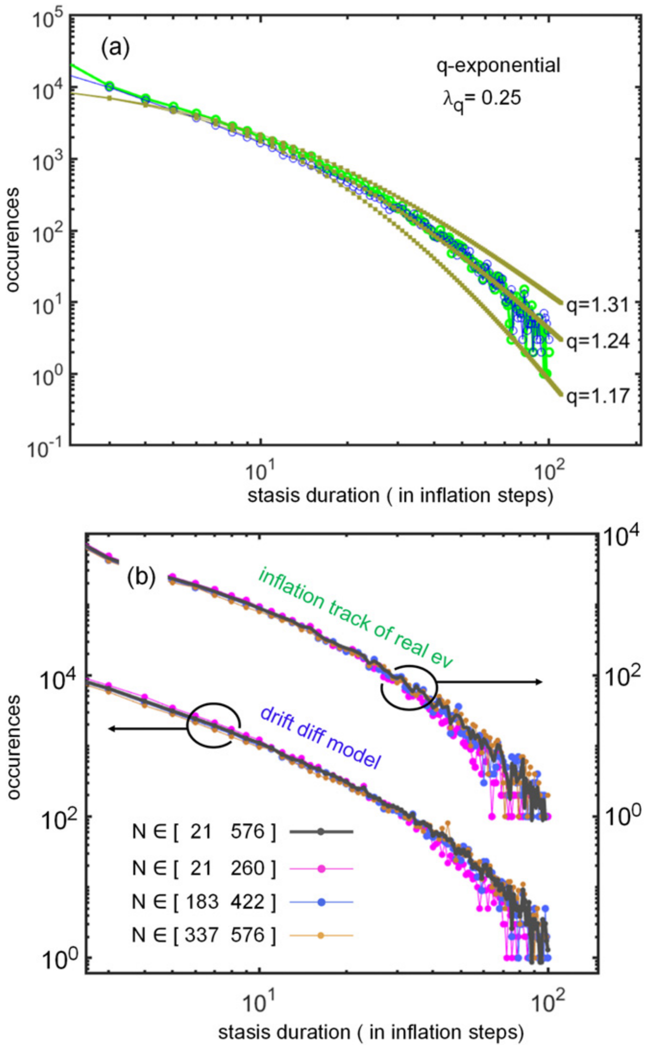

Figure 6. (a) Same statistics as in figure 5, but on a log–log plot (blue: 'along eigenvectors', green: 'drift-diffusion model'), in brownish color, comparison with q-exponential laws for the indicated q values and for the q-exponential parameter $\lambda_q = 0.25$; (b) Comparison on log–log plots of the distribution for three identical ranges of N ($N_{\text{final}} - N_{\text{start}} + 1 = 240$) taken at the three intervals dispatched, as indicated, at the start, the middle and the end of the full N sequence. The top set and left axis is for the series 'along eigenvectors', thus shifted for clarity from the bottom set and right axis, obtained from the drift-diffusion model. Only the magenta set of small N shows a deviation at the largest stasis duration (stasis duration $\gtrsim$40).

such 'adjacent-N eigenvectors scalar products', the resulting stasis duration statistics (its pdf) shows very little sensitivity to the exact threshold chosen to discriminate quakes.

As for the second approach, the knowledge of local drift and diffusion of ev's can be exploited to generate a kind of Brownian-motion model of ev's modulus, that we term $x(N)$ in this subsection to stress the analogy with a generic random walk.

Using a basic normal distribution random generator (MATLAB's in our case), a drift and a diffusion are drawn at each N value, dictated by the local ev position $x \equiv |\lambda|$ and by the interpolated mean drift and mean diffusion as had been numerically obtained (in a process whose details are in appendix). A specific correction consists of modifying this Brownian motion at the bottom, around a limit situated at about one standard deviation below the mean, where a stronger restoring force is applied. This correction is necessary because the drift-diffusion data were obtained by tracking the modulus, but this track does not account for the 'loss to the bottom' of ev's, whenever the rank after the inflation step becomes the third or less in modulus, and it does not either account for the reciprocal 'replenishment' to the top ev's from the reservoir of all other ev's, e.g., when the previously third etc ev becomes the second one. Hence the downward motion was too relaxed to the bottom. Logically, it can be thought that the replenishment process locally stiffens the apparent motion in this bottom range where exchange among second, third, etc is most likely. Similar issues around crossing and no-crossing random walks are generically known in statistical physics.

An exact treatment would be quite heavy, hence a simple linear model of restoring force (barrier) applied at an adjusted limit around $\bar{\lambda}_{\text{int}} - \sigma$ was taken (with $\bar{\lambda}_{\text{int}} = \bar{\lambda}_{\text{int}}(N)$ the mean of the ev modulus as interpolated

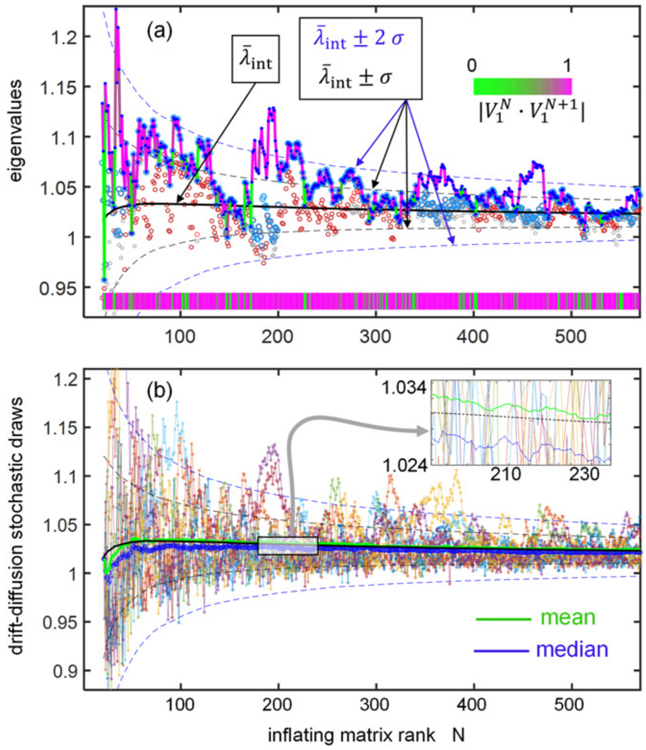

Figure 7. Largest ev's behavior during inflation, for N covering 20 to 570; (a) plot of the four largest ev's modulus of an actual random matrix inflation sequence. Four ev's are needed to always get two or more distinct modulus values. Their symbols are (1) blue dots, (2) blueish circles, (3) red small circles, (4) grey tiny circles. The top (#1) ev's (blue dots) are connected by lines whose color code depend on the magnitude of the scalar product of corresponding eigenvectors ($|U_1^N \cdot U_1^{N+1}|$), with the magenta/green color map indicated. The thick ribbon at the bottom gives a homogeneous view of the sequence of these scalar products. There are few green 'quakes' and large magenta stasis, and their duration can be extracted by appropriate treatment of such data; (b) for the drift diffusion model, the plot represent ten draws of pseudo-ev's modulus trajectories, with consistently modified lower bound as explained in text. In both (a) and (b), a superimposed set of curves is the statistical account of actual top ev's modulus obtained from a large set of 720 000 matrices, logarithmically sampling 30 values of N from 20 to 576, and interpolated (along N) in between: the black line is the N-dependent mean denoted $\bar{\lambda}$ of a large set (black line) and $\bar{\lambda} \pm \sigma$ and $\bar{\lambda} \pm 2\sigma$ (thin dashed and blue black lines). They delineate the spread of the top ev's (see text) with σ the standard deviation of the set. The inset shows that the mean (green) and the median (blue) of a set of 1060 draws are close to the correct value, justifying that the lower bound correction is consistent and thus neutral.

from the 30 logarithmically spaced N values to the whole 20 ... 576 set and $\sigma = \sigma(N)$ its standard deviation), with a strength (stiffness) self-consistently chosen so that the Brownian motion's ensemble average $\bar{x}$ sticks accurately enough to the average trajectory $\bar{\lambda}_{\text{int}}(N)$.

Equipped with a collection of many trajectories $x_k(N)$ of 556 points ($N = 20 \ldots 576$ and typically $k = 1$ to 1060 trajectories for the choices made here), the comparison of two series, for instance x_k and x_{k+1}, is performed to detect their crossing (there is statistical independence if we use each walk x_k only once with its upper neighbour). Each crossing of the two series x_k and x_{k+1} is assumed to be a quake, and so the duration of the stasis in between can be deduced from a basic treatment. From the ensemble of 1059 trajectory comparisons, the empirical pdf of stasis is deduced as well.

The fact that the two pdf of figure 5 are quite similar (the green somehow more curved than the blue, but the similarity is quite good over almost 3 decades) is an important achievement of this work. It suggests that the proposed descriptions are indeed consistent far enough in the long tails, spotting the longer stasis, which are the most remarkable features in the real data that inspire this work in evolutionary biology or in economy. Thus, many ingredients facilitating the desirable emergence of an analytical solution are validated from this result. It also indicates that the amount of empirical data needed to check the validity of the inflation model for a given targeted set of real-world statistical data is rather large: it must plausibly provide the empirical statistics of quakes over 2 to 3 orders of magnitudes to attain a meaningful diagnosis.

3.4. Correspondence with q-exponential law and trend of N-independence

In figure 6(a), the two pdf of figure 5 are reproduced on a log–log plot. The curvature is opposite of the semilog plot. Attempting to fit with a q-exponential law as those put forward in fossil records [29] (namely $A_q[1 + (1-q)(-\lambda_q x)]^{1/(1-q)}$ with adapted normalization A_q and x the stasis duration), it is first found that the range imposes the parameter λ_q to be 0.25 (thus a 'time scale' of $4 = \lambda_q^{-1}$ steps). Next, by comparison to the range of values $q = 1.1$ to 1.35, it is seen on figure 6(a) that $q = 1.24$ approaches the curve very well, and that the accuracy of a fit on such data is safely $\delta q \lesssim 0.01$. These values turn out to be very similar to those of [29] ($q = 1.2344 \pm 0.033$). This author is not aware of the systematic use of the q-exponential law in economy, although train delays have indeed been considered [44], and this issue is deferred to future work. Both approaches of figure 5 seem, without extra in-depth investigations, equally compatible vs the q-exponential law.

The log–log plots are also a good opportunity to test the degree of N-independence of the stasis distribution. This test is conducted by comparing the distribution obtained in three segments (short-N, mid-N and long-N subsets) of the overall inflation used for figure 5, based on the very same set of random realizations. Segments covering 240 steps were chosen, out of the total of 556, so as to still sample the longest $\Delta N = 100$ duration (beyond which scarcity strikes). Thus the N-ranges are those indicated inside figure 6(b), partly overlapping. The comparison between the subsets with the two largest N ranges indicates a difference at most of the order of the statistical noise (brown and blueish data, the full set being the dark grey data), for either models (top and bottom curves, right and left scales), with a residual dearth of long stasis occurrences for the mid-N range vs the long-N range. For the smaller subset, covering $N = 21$ to 260 (magenta), the range of duration 4–40 still shows no difference. But a clear lack of occurrences is seen for stasis duration $\gtrsim$40, reaching about a factor of 2 at the largest sampled durations of 90–100. Overall, the picture is thus an N-independence in asymptotic terms, with a modification at small N, typically for stasis duration smaller than N or $N/2$: smaller matrices do not offer enough 'inertia' of their eigenvectors to sustain competition when the matrix size is doubled or so. Such considerations are reasonable intuitive drivers of the low-N deviation indicated by these results. Conversely, the validity of the N-independence at large N could be logical with the statistics of Brownian motions of the largest ev's, within the elements discussed in section 2.3. Even the less obvious behaviour of ev's close the real axis (e.g. in the depleted area around the 'Saturn Ring' of the Ginibre set, figure 3) will eventually cause negligible deviations with respect to the main process.

3.5. Graphical illustrations of direct eigenvalue tracking and drift-diffusion models

Graphical illustrations of the two processes, the actual inflation and the Brownian motion, are given to support the above description and weigh its various aspects.

The 'actual inflation with eigenvector tracking' process is exemplified in figure 7(a). Here, a typical single run is shown. At each N, four ev's modulus are shown as circle symbols (the four top ones, distinct or not), with blue and blueish colors for the top two, other colours and smaller symbol sizes for the third and fourth. A solid line joins the top ev's along the inflation (along N), but the line coloring was made to depend on the adjacent scalar product $S = |U_1^{N-1} U_1^N|$: it is magenta if S is close to unity and green otherwise. The corresponding quake/stasis sequence is indicated at the bottom of figure 7(a) as a linear fresco ribbon, predominantly magenta with green stripes. There are a few points on this fresco that are in the brownish hue region of S, i.e. hard to decide between similar or different eigenvector. However they account for a negligible fraction of the candidate quake and shifting the quake criterion does not notably affect the statistical pdf of figure 5.

As for the impression given by the series themselves, long stasis are associated with large excursions of a top ev. So there is no mysterious mechanism that would for instance maintain a tiny but fixed spacing between the two top ev's, a mechanism that would poorly be accounted by the drift-diffusion modelling of two random motions crossing each other. Given the statistical limits to drift and diffusion, once a large gap with the subsequent smaller moduli has opened, it cannot generally be suppressed in one inflation step. Hence if enough steps are made upwards (vs the next larger modulus) at a given moment, a long stasis is rather granted. This is, at least informally, reminiscent of polymer dynamics and how they return to a reference line, a topic that is among the major application successes of RMT. The mean modulus $\bar{\lambda}_{\text{int}}$ and mean $\pm$ one and two standard-deviations σ of the interpolated series used for the other model are shown and it is seen that the range actually sampled with respect to this statistics is almost always above the $\bar{\lambda}_{\text{int}} - \sigma$ limit where the added barrier of the Brownian motion, commented below, has been implemented.

Next, the Brownian motion process is exemplified in figure 7(b). A series of ten $x_k(N)$ series is drawn. Some cases of top ev's drifting up for a rather long while (say $\Delta N >\sim 20$) are easily seen. These must correspond to stasis, if we choose here to compare any two such walks (as said, to establish the statistics, we considered two consecutive draws, thus using each walk only once on a given side of a comparison). Along the same picture, when a different line color emerges as the top one, a quake has likely taken place. On the figure is plotted the mean and median of the 1060 Brownian motions launched. The mean $\bar{\lambda}_{\text{int}}$ and mean $\pm$ one and two standard-deviations σ of the interpolated series are shown. The effect of the correction, the extra restoring force exerted at the bottom below $\bar{\lambda}_{\text{int}} - \sigma$, is very apparent, with a quick decrease of the abundance of the random walk passages $x(N)$ below this line. The adjustment of the strength of this correction was made to align the mean $\bar{x}$ close to the black curve, giving a kind of self-consistency to the approach. The inset shows that the residual error is in the few 10^{-3} at most, thus amounting to less than $\sim$5% of a standard deviation from the mean

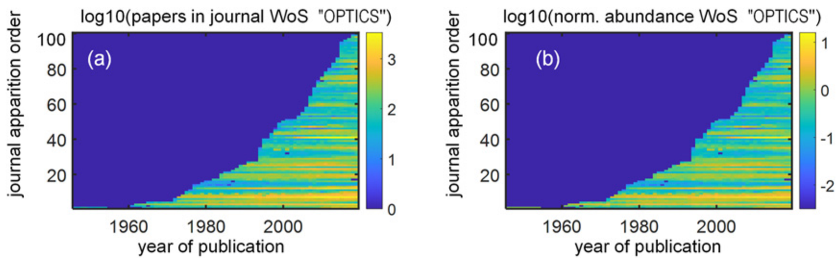

Figure 8. Statistics of papers in OPTICS category of Web of Science, curated; (a) colormap of $\log_{10}$(number of paper) with year in abscissa and journal apparition order in ordinate (b) same data after normalisation within each year to make it similar to normalized eigenvectors.

(green curve). This latter is, as expected, a bit larger than the median. The further exploration of the detail of this correction is believed to be of secondary importance, a better mathematical treatment of these random walks with a more rigorous account of the exchange with the reservoir of $N-2$ unaccounted ev's (those at ranks $3\ldots N$) being a more important aim.

4. Example of real-world data: bibliometric records

Getting meaningful real-world data to account for quakes and stasis in a statistical sense is a task beyond what can be proposed here. It entails many large series with a long sampling time, providing a thousand time points or so to address the tail of the distributions in figure 7.

Nevertheless, without the embryo of an example, the message could be missed. Therefore, a minimal example was explored in the area of scientific bibliometry, so that the audience is familiar with the data. Curation of data in biology and economy (see for example the data on world GDP and world energy in [45]) is a very demanding task, so it is likely that several proxies have to be used to robustly test the approach proposed here.

The chosen example is a subcategory of the Web of Science (WoS) currently operated by Clarivate, using the topic 'OPTICS' whereby this author could easily perform a sensible curation (discarding journals that disappear a couple of years because of contingent editorial issues in the 60s and so on). We take the set of journals as the vector-space basis, the size N of this set grows from a few units in the early 60s to $\sim$100 journals around 2020. The vector *components* are based on the yearly abundance of articles $A_N(j)$ in each journal j in year N counted from a convenient 'year 1', say 1950 (by and large, their market share in optics in year N). As explained in section 2, it is more sensible to use the logarithm $U_j = \log(A(j))$ for the components of the vector (and the zeroes of $A(j)$ are represented as the most negative color map hue, they are not involved in determining relevant quantities whatsoever). With these specific choices, figure 8(a) shows this logarithmic abundance. The yellow stripe appearing at $j = 41$ is the journal 'Optics Express', that did manage to capture a large share of the optics submissions/publications, thanks to its open access model, somehow a pioneer of this choice in optics in the 90s. A modification of this map is shown in figure 8(b). We further normalize in each year the vectors by using the mean A_N over the active journals in the year in question. This gives vectors with balanced positive and negative components, an important point to work out correlation and stasis easily (unlike positive series).

Formally, once a series of vector is given, it is possible to infer the matrix M that, under the inflation process, solves for the 'inflation equation' (equation (1) or equation (5)), with the unknowns located in the last row (or column). This approach would be interesting to follow, with the appropriate tools of linear algebra for the under/over determination of the various elements (singular value decomposition, pseudo-inverse matrix etc). The properties of the matrix M thus deduced would of course be of interest (the topic is reminiscent of so-called Leslie matrices widely used in demographic studies, as they obey the same iteration rules as the inflation process discussed here). It is not attempted to go further in this direction here.

Rather, the vectors can be used themselves to detect the quakes and the stasis. Graphically, when many new journals appear in a short duration, it could look like a quake. However, journals with few papers do not count as much as those with many papers. With the normalized version of the logarithmically represented data, figure 8(b), it becomes sensible to identify quakes thanks to the same product of adjacent vectors that was used to track eigenvectors, $S = |U_1^{N-1} U_1^N|$. This exercise is done for the 'OPTICS' category in figure 9(a). The same map as in the mathematical study is used, except that we apply it to S^2 (blue line) to get a more similar visual contrast (it might also be that disruption in journal shares cannot be as quick as market disruptions in economy or species invasion in biology, but this is not obvious). The sturdy fate of legacy journals probably imposes a

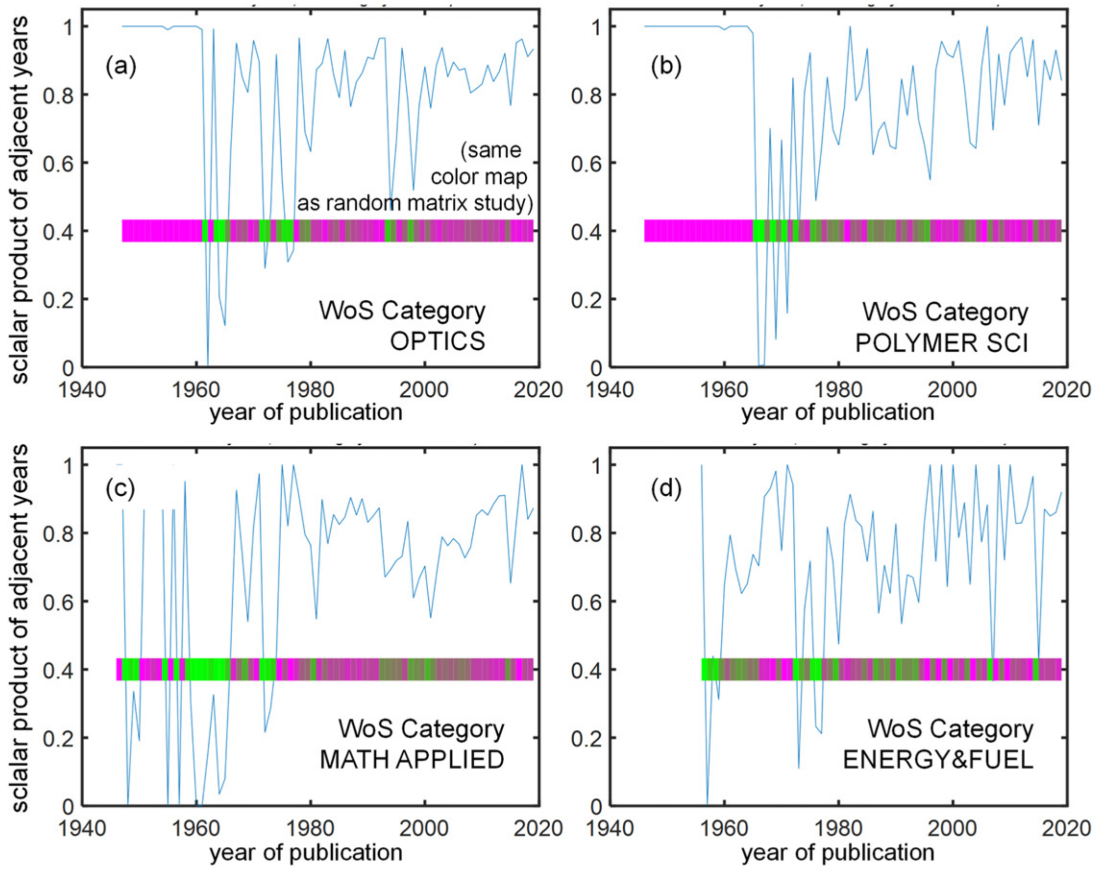

Figure 9. Analysis of journal data based on scalar product between adjacent year, so as to detect quakes vs stasis. The curves represent the successive squared scalar products of the normalized $\log_{10}$(number of paper), while the magenta/brown/green colormap is the same as used above for punctuated growth analysis. Cases (a)–(d) are for the four following categories of WoS: Optics, Polymer Science, Applied Mathematics and Energy-and-Fuel.

strong continuous component here. Knowledge may remain a nonrival good but still with some hierarchical rules (from its social role and from epistemology), making it plausibly a more conservative example than the other aimed targets of the present approach.

Nevertheless, a sequence of stasis and quakes emerges in this way, which very much differs from a white noise or from a pink noise. The same exercise was expanded to three other WoS bibliography categories shown in figures 9(b)–(d). Curation, in small amounts, was done based on the teaching on the optics categories, but it cannot be claimed to be as safe as in the OPTICS case, as the author has little direct knowledge of those domains. The pattern for these three domains is still very much indicative of a stasis + quake appearance, maybe slightly less because brownish areas can acquire some extent, for example in (c) (MATH APPLIED) during a decade around 2000. All in all, these examples give a favourable weight to the assumption that actual growth series from the real world, if sufficiently documented to reveal disruptions in a broad time vector evolution $A_N(j)$, can be cast into a model of quakes and stasis, and further studied against a random matrix ansatz much as the one substantiated here.

5. Perspectives and discussion

The idea of matrix inflation is to capitalise on the ‘agnostic’ aspect of random matrices to describe systems whose states evolve by stasis and quakes. Such systems are notorious in biology (justifying the emergence of the punctuated equilibria concept in evolution) and in economy (radical Schumpeterian changes in technologies). The frame to implement the independence of the ‘old’ (pre-inflation) and the ‘new’ (post-inflation) eigenvector is thus chosen as that of real non Hermitian matrices, the Ginibre set. To retain real signed vectors and real matrix, the use of the logarithm of the actual quantities (akin to activities, notably in biochemistry) could be logical. The outcome is the statistics of stasis and quakes, whose long tail is of most interest as it leads to the most recognizable long stasis patterns. Of course, knowledge on very long tails is inherently affected by poorer determination of scaling laws, risks of cherry-picking of scarce event in long series by epistemic biases, and so on.

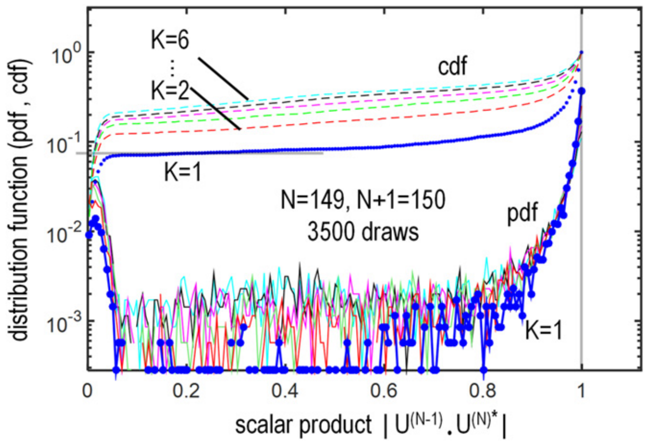

Figure 10. Statistics of scalar product $|c^{(K)}| = |U^{N,(K)} \cdot U^{N+1,(K)*}|$ of the eigenvectors associated to the Kth largest ev, based on 3500 draws in the particular case $N = 149 \to N + 1 = 150$, in the form of cdf and pdf. The blue lines with the dots are for $K = 1$. The other coloured lines are for the K values indicated, the thin dashed lines for the cdf (on top) and the thin solid lines for the pdf.

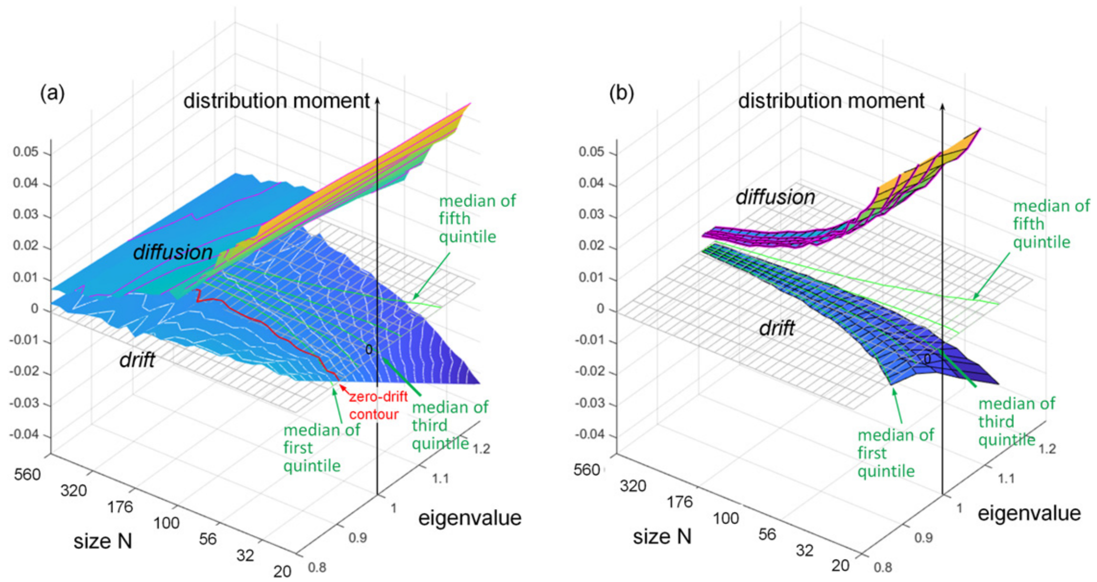

Figure 11. 3D maps of moments of the ev distribution change discussed in the main text, as a function of ev and of matrix size. These maps are the basic ingredient of the simplified model whereby two ev's are left to drift and diffuse during growth of size N. In (a), the interpolated maps for drift (first moment) and diffusion (second moment) are represented. The interpolation is a linear one along the ev axis based on the five points obtained as means for each of the five (ranked) quintiles of the distribution, for both quantities (see text). The drift map crosses the zero on the red line. In the zero moment horizontal plane, we have reported the ev statistics (median of each of the five quintiles) as green lines. In (b), we have only represented the momentum maps not interpolated, i.e. the means taken at the values of the green lines, which are the median of the five quintiles.

There are however several ways to revisit (and expand) this choice, that we attempt to itemize. We do not comment again on the role of renewal rate evidenced in figure 4, but we hypothesize that it echoes the studies of pseudospectra and the underlying non-orthogonal vector combinations that feature distinct evolutions compared to orthogonal cases [38].

Firstly, the generalized interaction of all elements of a full random matrix can be replaced by a sparse random matrix. This was considered already in May's seminal work, as the matrix sparsity and the ecological network connectivity are essentially the same thing. There is no special difficulty in running the basic algorithm that gives figure 4 with a sparse matrix. Our first observation is that the statistics of quakes and stasis are similar, but that the local shifts in ev's tend to be larger, as the averaging is less. Extending the present study in this direction is surely of interest. The more general issue of network topology (tackled in a different 'ev context' in [46], but much more general in graph theory, the small-world approach, etc) could also fuel further investigations.

We also note that Rivkold [28] has given important basic trends for interaction mediated by structured matrices: completely random matrices tend to evolve into strongly connected, mutualistic communities, while antisymmetric matrices mimicking predator–prey dynamics lead to tree-shaped communities resembling food webs. This certainly suggests a relevant path for further investigations.

Next, the non-Hermitian aspect could be modified. A Hermitian matrix has ev on the real axis that must lead to anticrossing and gradual exchange of eigenvectors when the ev's are on a crossing trajectory. This means that abrupt change of dominant eigenvector does not take place. The eigenvector evolution from an inflation process can still have some salient features, but mainly, the quakes are washed out. Playing with the degree of non-Hermiticity essentially amounts to make the matrix M closer to a symmetric one (decomposing it into symmetric and antisymmetric part and diminishing the anti-symmetric contribution). It would be interesting if data exist for such cases. But within our application targets, symmetry of M has quite few reasons to emerge. In economy, if a product j depends on another product k for its production, there is generally no reciprocity (tyres are requested on all cars and trucks, but the tyre industry's overall operation uses few cars, certainly it uses more trucks). In ecological networks, the same remark holds. Of course, the well-known predator–prey model suggests some reciprocity, but it lies in the nonlinear aspect. In terms of the direct dependence, traced along a precise set of trophic chains, the symmetry is the exception (or it evolves in symbiosis and germane cases). But mathematical theories of inflation could be formulated that take the degree of symmetry into account, to assess all aspects of such theories and consolidate them.

Other modifications can be devised based on the known facts of the targeted domains. Some species become extinct, some products stop existing not only out of obsolescence but because the skills are not worth exploiting. Wooden wheels with steel rims, not obvious to assemble to say the least, convey the idea of such a unworthy skill situation. This means that some components of vector U could become extinct from some point on in time (during the inflation). Pushing components to zero or, conversely, to a near saturation, both tick the category of nonlinear behaviours. Certainly, nonlinear aspects are of utmost importance in the actual dynamics of innovation or of biological evolution. Some preliminary investigations have been done by the author along this line, rather conclusively. We intend to report them elsewhere. The interesting point is that they modify the pdf of quakes and stasis, influencing the shape of the whole distribution and especially the long tail in the stasis duration distribution in figure 7. Still, the availability of large data would help assessing the relevance of this nonlinearity. It also remains to be understood whether the known dynamics of ecosystems that generalize the Lotka–Volterra prey–predator oscillations must be related to the natural oscillation of real eigenvectors due to the conjugated cosine-type ev evolution that was mentioned in the comments of figure 4.

6. Conclusion

We have proposed a model of matrix inflation, dealt with in the spirit of RMT, in order to account for processes where novelty, i.e., growth in the variety and population of a set composition, is accompanied by quakes and stasis. Such marked behaviour is recognized to have a large role in evolutionary biology and in economy. In evolutionary biology, even though there is no unanimous consensus on punctuated equilibria, it is one of the main defining paradigms in the domain. In economy, Schumpeterian destructive creation epitomizes the innovation-triggered quakes of the economy, while the stasis are often related to a lack of technological progress and a use of less qualified abundant labour.

The creation of new species or of new tools can be thought as a permanent, never steady process that has most of the time little impact on the overall ecosystem architecture, or the economy organisation. But sometimes a new product or a new species disrupts these slow-evolving states.

In the model that was proposed, the instantaneous situation is described by an eigenvector whose basic elements must describe as well as possible a generic interdependence. Our guess is that in biology, it is a pool of gene and in economy a pool of tools and products.

Then, it is posited that the evolution by a given matrix M, which embodies the interrelations of these elements, 'lands' on the dominant eigenvector of M, using random matrices in a way that is as 'agnostic' as May's seminal proposal on ecological networks, but distinctly different in purpose.

The essential new point is to introduce growth and innovation in economy or in the gene pool as a matrix inflation process, rank by rank. The new line and column dictate what happens to the eigenvectors and ev's. A Ginibre set framework is chosen so that the crossings of ev modulus do not entail an exchange of eigenvectors, and thus make it possible to drastically change the eigenvector when an innovation displaces an 'incumbent' eigenvector in favour of a new different one.

The paper has given some mathematical clues about the distribution of quakes and stasis that arises within the most basic case that could be devised. There is a pdf tail of stasis, which is not exponential, and which tends to be fatter than a Poissonian law. It could be tracked over about three decades. A q-exponential law was found to agree well with these results, with $q \approx 1.24$. Notably, the analysis of the motion of top ev's in a drift-diffusion framework taking place over successive innovations appears to be able to describe fairly faithfully the pdf of stasis and quakes that can be seen in a numerical model. The heuristic value of this approach to better understand the operation of matrix inflation is hoped to trigger further studies in the domain. The elements in favour of an N-independence of the statistics of stasis suggest that its signature might be recognized across very different contexts. Various alternate ways to operate the inflation and to constrain the matrix, e.g. through its sparsity, or through the degree of its non-Hermitian character, have also been suggested. Several perspectives have been discussed accordingly. It is hoped notably that this study can later contribute to analyse the 'noise' of world-scale innovation and energy used, as recently described in [45], so as to better deal with the tightly related 'signal' of the primary energy consumption and of its environmental and societal impacts.

The mathematical tools that could be employed to further explore this problem are certainly already existing in the large body of RMT related works. We expect that the use of these tools would naturally enrich the ideas and the trends that we have tracked mostly numerically in this first version of matrix inflation.

Acknowledgments

HB is grateful to Hervé Bercegol for support of an early version of these ideas. He also thanks Guillaume Martin, François Massol, Nicolas Pouyanne, Julien Randon-Furling and Thierry Huillet for several useful discussions in mathematics and evolutionary biology.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Appendix

In this appendix we first depict the typical distribution of eigenvector scalar products that sets the frame of the competition picture exploited in the main text. Next are presented the main ingredients that were elaborated for implementing a drift-diffusion model of ev evolution as caused by the matrix inflation process.

On figure 10, we show on a semilog scale the pdf and cdf of the six largest eigenvector scalar products, before and after an $N \to N+1$ inflation step, $|c^{(K)}| = |U^{N,(K)} \cdot U^{N+1,(K)*}|$. The pdf are seen to be markedly bimodal, with a small bump at near-zero scalar product (for random vectors, there is a $|c| \sim 1/\sqrt{N}$ scaling) and a high abundance near unity. The cdf for $K = 1$ has a plateau around 0.075 on the right of the initial pdf bump, and has most of its pdf contributions gained near $|c| \simeq 1$. This means that most of the times, an eigenvector of a given rank K remains after the inflation close to its parent before the inflation, with little motion between them, whereas in a few cases, a completely different eigenvector replaces it. In the case $K = 1$, it becomes the leader, as the vector in question is bound to win the iteration process, which is in essence, at a given N, a competition among ev's won at a speed dictated by the ratio to the next smaller $|\lambda|$ one.

As for the ingredients of the drift-diffusion model, on figure 11 are shown the results of the analysis of a large set of matrices done at sizes $N \in \{20, 22, 25, 28, 32, 36\}$ and at 4 larger octave-shifted homothetic sets starting at $N = 40, 80, 160, 320$, hence up to $N = 576$.

In figure 11(a), the two relevant surfaces, depicting the drift and the diffusion of ev's, are shown. Magenta contours have been overlaid. The evolution paths of ev's are tracked here for a 'given eigenvector', i.e. one follows ev's whose scalar products of eigenvectors ($c = U^N \cdot U^{N+1}$) before and after $N \to N+1$ inflation has a modulus $|c|$ near unity. The scalar product modulus of the few first eigenvectors (typically 16 products $|U_j^N \cdot U_k^{N+1}|$ if one tracks the four top ev's, i.e. $j, k \in \{1, 2, 3, 4\}$ in terms of usual rank, as delivered by numerical packages) are distributed very unevenly, making this assignment almost always unambiguous.

Furthermore, once one has the collection of individual changes at a given N, one would ideally finely 'bin' the ev's modulus in as many bins as possible to get the data of drift and diffusion around given values $|\lambda|$ (the short notation in the following for ev's modulus). Because the tails ($|\lambda|$ furthest from unity, essentially), that are important here, would be too noisy for fine bins derived from still tractable pseudo-random numerical sets, a simpler approximate option has been preferred: the merging in a single analysis of the two top distinct ev's, further splitting the resulting distributions of $|\lambda|$'s shift and $|\lambda|$'s diffusion along the five quintiles of the $|\lambda|$ distribution (five slices of equal population, taken in the $|\lambda|$ distribution, and not in the drift or diffusion sets), i.e. a coarse view in terms of bins. Noting $\bar{\lambda}$ the mean and σ the standard deviation of the whole aggregated set of collected $|\lambda|$ distinct pairs $(|\lambda_1|, |\lambda_2|)$, the four edges of these quintiles are analysed for consistency. They are found to be typically arranged in a staggered fashion within a 'knowledgeable' reference set such as $\{\bar{\lambda} - 2\sigma, \bar{\lambda} - \sigma, \bar{\lambda}, \bar{\lambda} + \sigma, \bar{\lambda} + 2\sigma\}$.

For each of the five quintiles $q = 1, \ldots, 5$, one calculates the mean drift $\bar{\Delta}_q$ and the mean standard deviation d_q (diffusion) of the ev modulus. One thus obtains, for each of the two quantities, five numbers. They are represented as two surfaces on figure 11(b). The axis of size, N, is a logarithmic one, but that of ev modulus $|\lambda|$ is linear. The momentum (drift and diffusion) are on a vertical linear axis.

It was then simply chosen to apply a linear interpolation, thus reducing the five numbers to just two, with the benefit of getting an approximate value of drift and diffusion from any starting point $|\lambda|$. It was tried to use quadratic or higher order interpolation but the results was observe to diverge too much in the tails in large swaths of the ranges of interest. The consistency across the different N values of the set (30 N values log-spaced) gives a fair confidence in the relevance of the results. The locations of each quintile's median ev's modulus are the five green lines in figures 11(a) and (b). Figure 11(a) gives the interpolated map calculated from the linear interpolation just mentioned, that was thus used in further models to grasp the behaviour of ev's during a whole inflation series over large spans of N.

ORCID iDs

Henri Benisty https://orcid.org/0000-0001-7428-7196

References

[1] Depew D J and Weber B H 1994 *Darwinism Evolving: Systems Dynamics and the Genealogy of Natural Selection* (Cambridge, MA: MIT Press)

[2] Patrick Raines J and Leathers C G 2008 *Debt, Innovations, and Deflation: The Theories of Veblen, Fisher, Schumpeter, and Minsky* (Cheltenham: Edward Elgar)

[3] Bak P and Sneppen K 1993 Punctuated equilibrium and criticality in a simple model of evolution *Phys. Rev. Lett.* **71** 4083

[4] Gould S J 2002 *The Structure of Evolutionary Theory* (Cambridge, MA: Harvard University Press)

[5] May R M 1972 Will a large complex system be stable? *Nature* **238** 413

[6] Chatelin F 1988 *Valeurs propres de matrices* (Paris: Masson)

[7] Dyson F J 1962 Statistical theory of the energy levels of complex systems. I *J. Math. Phys.* **3** 140 and references therein

[8] Menon G 2012 Lesser known miracles of Burgers equation *Acta Math. Sci.* **32** 281

[9] Burda Z, Grela J, Nowak M A, Tarnowski W and Warchoł P 2014 Dysonian dynamics of the Ginibre ensemble *Phys. Rev. Lett.* **113** 104102

[10] Schehr G 2012 Extremes of *N* Vicious walkers for large *N*: application to the directed polymer and KPZ interfaces *J. Stat. Phys.* **149** 385

[11] Fyodorov Y V and Khoruzhenko B A 2016 Nonlinear analogue of the May–Wigner instability transition *Proc. Natl Acad. Sci. USA* **113** 6827–32

[12] Kwapień J and Drożdż S 2012 Physical approach to complex systems *Phys. Rep.* **515** 115

[13] Allesina S and Tang S 2015 The stability-complexity relationship at age 40: a random matrix perspective *Popul. Ecol.* **57** 63

[14] Watts D J 2002 A simple model of global cascades on random networks *Proc. Natl Acad. Sci. USA* **99** 5766

[15] Maih J 2015 Efficient perturbation methods for solving regime-switching DSGE models *SSRN Electron. J.* volume number **1** 1

[16] McCullen N J, Rucklidge A M, Bale C S E, Foxon T J and Gale W F 2013 Multiparameter models of innovation diffusion on complex networks *SIAM J. Appl. Dyn. Syst.* **12** 515

[17] Mehrotra R, Soni V and Jain S 2008 Diversity sustains an evolving network *J. R. Soc. Interface* **6** 793

[18] Benhabib J, Perla J and Tonetti C 2021 Reconciling models of diffusion and innovation: a theory of the productivity distribution and technology frontier *Econometrica* **89** 2261–301

[19] Silverberg G and Verspagen B 2005 A percolation model of innovation in complex technology spaces *J. Econ. Dyn. Control* **29** 225

[20] Napolitano L, Evangelou E, Pugliese E, Zeppini P and Room G 2018 Technology networks: the autocatalytic origins of innovation *R. Soc. Open Sci.* **5** 172445

[21] König M D, Battiston S and Schweitzer F 2009 Modeling evolving innovation networks *Innovation Networks* (*Understanding Complex Systems*) ed A Pyka and A Scharnhorst (Berlin: Springer) pp 187–267

[22] Thurner S, Klimek P and Hanel R 2010 Schumpeterian economic dynamics as a quantifiable model of evolution *New J. Phys.* **12** 075029

[23] Thurner S, Hanel R and Klimek P 2010 Physics of evolution: selection without fitness *Physica* A **389** 747

[24] Christensen K, Di Collobiano S A, Hall M and Jensen H J 2002 Tangled nature: a model of evolutionary ecology *J. Theor. Biol.* **216** 73–84

[25] Murase Y, Shimada T, Ito N and Rikvold P A 2010 Random walk in genome space: a key ingredient of intermittent dynamics of community assembly on evolutionary time scales *J. Theor. Biol.* **264** 663–72

[26] Murase Y, Shimada T and Ito N 2010 A simple model for skewed species-lifetime distributions *New J. Phys.* **12** 063021

[27] Murase Y, Shimada T and Ito N 2009 Phase diagram and stability of ecosystems *Artif. Life Robot.* **13** 460–3

[28] Rikvold P A 2007 Self-optimization, community stability, and fluctuations in two individual-based models of biological coevolution *J. Math. Biol.* **55** 653–77

[29] Shimada T, Yukawa S and Ito N 2003 Life-span of families in fossil data forms *q*-exponential distribution *Int. J. Mod. Phys.* C **14** 1267–71

[30] Tokita K and Yasutomi A 1999 Mass extinction in a dynamical system of evolution with variable dimension *Phys. Rev.* E **60** 842

[31] Aguirre J and Manrubia S 2015 Tipping points and early warning signals in the genomic composition of populations induced by environmental changes *Sci. Rep.* **5** 9664

[32] Solé R V, Alonso D and McKane A 2002 Self-organized instability in complex ecosystems *Phil. Trans. R. Soc.* B **357** 667

[33] Schweitzer F, Fagiolo G, Sornette D, Vega-Redondo F and White D R 2009 Economic networks: what do we know and what do we need to know? *Adv. Complex Syst.* **12** 407

[34] Shimada T 2014 A universal transition in the robustness of evolving open systems *Sci. Rep.* **4** 4082

[35] Sandoval L and Franca I D P 2012 Correlation of financial markets in times of crisis *Physica* A **391** 187

[36] Allez R and Bouchaud J-P 2012 Eigenvector dynamics: general theory and some applications *Phys. Rev.* E **86** 046202

[37] Burda Z, Grela J, Nowak M A, Tarnowski W and Warchoł P 2015 Unveiling the significance of eigenvectors in diffusing non-Hermitian matrices by identifying the underlying Burgers dynamics *Nucl. Phys.* B **897** 421

[38] Trefethen L N 1992 Pseudospectra of matrices *Numerical Analysis* ed D F Griffiths and G A Watson (Harlow: Longman Scientific and Technical) pp 234–65

[39] Chevin L-M, Martin G and Lenormand T 2010 Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution *Evolution* **64** 3213–31

[40] Dietzenbacher E 1992 The measurement of interindustry linkages: key sectors in The Netherlands *Econ. Modelling* **9** 419

[41] Cerina F, Zhu Z, Chessa A and Riccaboni M 2015 World input–output network *PLoS One* **10** e0134025

[42] Cui W, Marsland R III and Mehta P 2021 Diverse communities behave like typical random ecosystems *Phys. Rev.* E **104** 034416

[43] Orr H A 2005 The genetic theory of adaptation: a brief history *Nat. Rev. Genet.* **6** 119–27

[44] Briggs K and Beck C 2007 Modelling train delays with q-exponential functions *Physica* A **378** 498–504

[45] Bercegol H and Benisty H 2022 An energy-based macroeconomic model validated by global historical series since 1820 *Ecol. Econ.* **192** 107253

[46] Benisty H and Goupil C 2020 Configuration barrier towards parity-time symmetry in randomly connected mesoscopic sets on a graph *Eur. Phys. J.* B **93** 192