Analytical study of time series generation by feed-forward networks

I. Kanter, D. A. Kessler, A. Priel and E. Eisenstein
Minerva Center and Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel

Generation of time series is studied analytically for a generalization of the Bit-Generator to continuous activation functions and multi-layer architectures. The network exhibits at asymptotically large times the following characteristic features: (a) flows can be periodic or quasi-periodic depending on the phase of the weights, (b) the dimension of the attractor is a function of the gain of the activation function and the number of hidden units, (c) a phase shift in the weights results in a frequency shift in the output, so that the system operates as a phase detector.

PACS numbers: 05.20.-y, 87.10.+e

The construction of neural networks with complex, as opposed to chaotic, time dynamics poses an important and interesting challenge. Progress in this field is crucial for understanding real-life intelligence and information processing. [1] Intelligence presupposes the ability to generate chains of highly correlated temporal activity, with the future activity of the network determined by nontrivial feedback from its current state. Comprehensive studies of such networks have begun only recently and very little knowledge of an analytic nature has at yet been achieved. [1]

In the well-studied cases of symmetric and feed-forward networks, there is no non-trivial dynamical behavior. [2,3] The central issue is rather the dependence of the output (or equilibrium) state on a set of random inputs. Here the input to the network is constantly changing in response to the current state of the network and so the issue of the dynamical behavior of the network rises to the fore. Hence a new repertoire of analytical techniques and concepts is required to begin to explore the implications of different architectures and feedback schemes for the behavior of the network.

In this paper we shall present an analytic study of a particular class of network, the Sequence Generator (SGen). This network is a natural generalization of the Bit Generator (BG), [4] which was shown in simulation studies to possess interesting and nontrivial time-dynamics of the kind described above. The SGen is in essence a perceptron [5] with a single continuous output, whose input at each successive time step is chosen as follows: the inputs from the previous time step are shifted one unit to the right with the state of the leftmost input unit set equal to the state of the output unit in the previous time step. Mathematically, if we denote the N inputs at time μ by the vector $\boldsymbol{S}^\mu = (S_1^\mu, \ldots, S_N^\mu)$, the process can be summarized by

$$S_1^{\mu+1} = o^\mu \;\; ; \;\; S_i^{\mu+1} = S_{i-1}^\mu \quad i = 2, \ ..., \ N. \tag{1}$$

Here, o^μ, the output at each time step is determined from the inputs at that time and the fixed weight vector $\boldsymbol{W}$ via the nonlinear activation function

$$o^\mu = \tanh[\beta \boldsymbol{W} \cdot \boldsymbol{S}^\mu] \;\; . \tag{2}$$

In the limit where the coupling $\beta (> 0)$ goes to ∞, this network reduces to the case of the BG, with discrete inputs and outputs, ± 1. [4] It should be pointed out that the continuous nature of the outputs of the SGen make it more useful for application to the realm of time-series. We will also consider further on the extension of this SGen to other activation functions and to multi-layer networks.

Equivalently, we may think of the network as generating an infinite sequence of output numbers, s_l, $l = 0, -1, \ldots, -\infty$, starting from the initial inputs, $s_i = S_i^0$ via the dynamical equation

$$s_l = \tanh[\beta \sum_{j=1}^{N} W_j s_{j+l}] \tag{3}$$

This is the origin for the name 'sequence generator' given to this machine.

Simulation of the model, as well as previous studies [4] of the BG, show that the output is primarily determined by the large amplitude Fourier components of the weight vector, $\boldsymbol{W}$. Thus it is natural to consider first the case where the weight vector consists of a single Fourier component. In particular, let

$$W_j = \cos[\frac{2\pi}{N} kj], \tag{4}$$

the integer $k \neq 0$ denoting the wave-number. Then, a dynamical solution of eqs. (3-4) is given by

$$s_j = \tanh[A \cos(\; \frac{2\pi}{N} kj + \delta \;)] \tag{5}$$

where δ can take the values $2\pi k/N, 4\pi k/N, \ldots, \pi k(N-1)/N$, corresponding to a trivial time translation, and π, corresponding to a global inversion symmetry. We will not explicitly discuss this symmetry-induced degeneracy in the rest of the analysis. This solution can be verified by evaluating the sum over j in eq. (3), expanding s_{j+l} in a power series in A. The constant A depends on β through the equation

$$N\beta = \left[\sum_{\rho=1}^{\infty} \frac{A^{2\rho-2}}{(\rho!)^2} (2^{2\rho} - 1) B_{2\rho} \right]^{-1} \tag{6}$$

where B_ρ are the Bernoulli numbers. [6] This solution is exact for any system size N and wave-number k. Note that eq. (6) is independent of k and that for $A > \pi/2$ the sum in eq. (6) diverges and the summation should be done via a continued fraction expansion. As the right hand side of eq. (6) behaves as $2+O(A^2)$ for small A and increases monotonically with A, we find that in the small coupling phase, $\beta < \beta_c = 2/N$, the only solution is the trivial fixed point $s_j \equiv 0$. At β_c this solution becomes unstable and the system undergoes a Hopf bifurcation to a periodic orbit of length N (for k and N relatively prime) characterized by a non-zero A, where just above the transition $A \sim (\beta_c - \beta)^{0.5}$. As β is increased, A increases and in the large β limit A diverges and a binary sequence with periodicity N is achieved. Results of the asymptotic long-time behavior of simulations on a system with $N = 100$ are presented in Fig. 1, and reproduce the analytic solution eq. (5). Thus this solution is indeed an attractor for the dynamics. We have not found initial conditions which do not lead to this attractor.

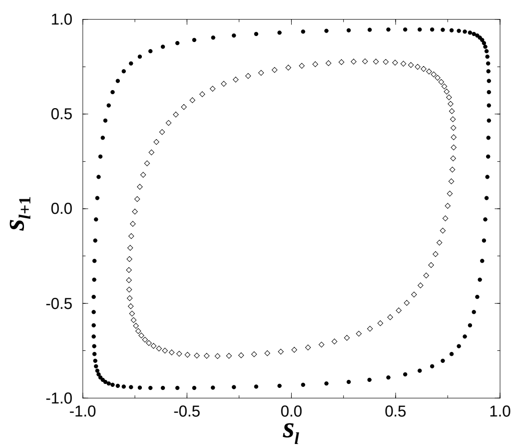

FIG. 1. Result of simulations for s_{l+1} vs. s_l for SGen, eqs. (3)-(4), with $N = 100$, $k = 11$ and $\beta = 1/30$ (•), $1/40$ (⋄). The analytical solution, eq. (5), gives exactly the same points.

In the more general case where the weights consist of a single Fourier mode with general phase, the weights are

$$W_j = R\cos[\frac{2\pi}{N}kj - \pi\phi] \qquad (7)$$

where R is an amplitude and $\phi \in [-1:1]$. This phase shift in the weights has a surprisingly dramatic effect on the dynamics. Proceeding as above, we find that in the limit $N \to \infty$ and $k \gg 1$, the translating solution is

$$s_j = \tanh[A\cos(\ \frac{2\pi}{N}(k-\phi)j\ )] \qquad (8)$$

with higher harmonic corrections of $O(1/k)$ and A now related to β through

$$N\beta = \frac{\pi\phi}{R\sin\pi\phi}\left[\sum_{\rho=1}^{\infty}\frac{A^{2\rho-2}}{(\rho!)^2}(2^{2\rho}-1)B_{2\rho}\right]^{-1}. \qquad (9)$$

The system undergoes a Hopf bifurcation as before but the critical coupling now depends on both the amplitude R and the phase ϕ

$$\beta_c = \frac{2}{RN}\frac{\pi\phi}{\sin(\pi\phi)} \qquad (10)$$

so that β increases with ϕ. For $\phi = 1$ one can show by taking into account higher order terms that $\beta_c = 1$. [7]

From the generic asymptotic dynamical behavior of a SGen with one component in the power spectrum of $\mathbf{W}$, eqs. (7-8), one can deduce the following results:

(a) **Phase detector**: A phase shift in the weights of the SGen results in a phase shift in the power spectrum, $k \to k - \phi$. The sensitivity of the SGen to the phase of the weights is a surprising result, since in the realm of classical spin systems the phase is almost the first piece of information which is lost. Furthermore, the piecewise linear interplay between the phase of the weights and the frequency shift is far from intuitive. For the finite N case, the dynamics can be solved perturbatively near the transition, and a more complicated dependence of the frequency shift on the phase ϕ is obtained. [7]

(b) **Transition as a function of** β: The critical β is a function of both the amplitude and the phase of the weights (though not the frequency). The interplay between these two parameters plays an important role when the case of weights with a more complex power spectrum is discussed below.

(c) **Quasi-periodicity**: In the generic case, ϕ is irrational and the flow is quasi-periodic. A typical result for s_{l+1} versus s_l is presented in Fig.2. Note that in contrast to the case $\phi = 0$ where the graph consists of only N distinct points (Fig.1), for ϕ irrational, the attractor densely fills the orbit.

(d) **Attractor dimension (AD)**: We have seen that for ϕ rational, the orbit consists of a finite set of points, so that the attractor dimension , AD, equals zero. However, for the generic case of ϕ irrational, the flow is quasi-periodic, eq. (8), and hence, the AD is one (see Fig. 2). In the following, we examine the robustness of this result to weights with more than one Fourier component.

(e) **Right SGen**: One can reverse the direction of the SGen, and generate an infinite sequence now to the right of an initial set of inputs, $s_{N-l+1} = \tanh[\beta\sum_{j=1}^{N} W_j s_{j-l}]$. All our results hold also for the right SGen with only a modification to the phase of the weights, $\phi_{right} = -\phi + 2k/N$. It is clear that one can find a SGen and a right SGen with the same asymptotic quasi-periodic structure and this two-fold degeneracy cannot be lifted without any further information about the direction of the SGen.

(f) **Uniqueness of the solution**: Assume that the asymptotic quasi-periodic flow of a SGen is given. A crucial question is whether this information is enough to determine uniquely the weights of the SGen (assuming a single Fourier component in the power spectrum).

From the discussion above it is clear that the weights are uniquely determined up to the following two-fold degeneracy, $\cos(2\pi k/N - \pi\phi)$ and $\cos(2\pi(k-1)/N + \pi(1-\phi))$.

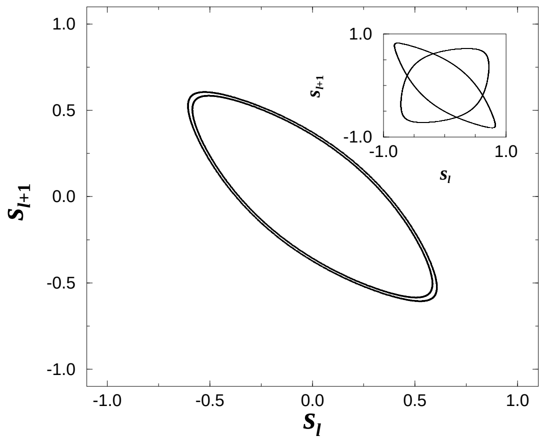

FIG. 2. Result of simulations for s_{l+1} vs. s_l for SGen, eqs. (3) and (7), with $N = 100$, $k = 41$, $\phi = 0.411$ and $\beta = 1/33$, together with the analytical solution, (outer curve) eqs. (8)-(9), are presented. For larger N, k the corrections are invisible. Inset: The two attractors for the two component SGen, eq. (12), with $R_i = \{1,\ 0.8\}$, $k_i = \{211,\ 91\}$, $\phi_i = \{0.7,\ 0.67\}$ and $N = 500$, $\beta = 1/70$.

(g)**State characterization**: From the graph of Fig. (2) it is clear that at asymptotically large times the two values $(s_l,\ s_{l+1})$ are sufficient to uniquely characterize the state of the system and hence its further time development. This is of course consistent with the general result [8] that at most $2(AD)+1$ values are sufficient to characterize a system state.

The above mentioned results can easily be generalized to other non-linear activation functions beyond the tanh. In particular, for any odd activation function, O, and weights $\cos(2\pi kj/N - \pi\phi)$, one can verify that the translating solution in the leading order is

$$s_j = O[A\cos(\frac{2\pi}{N}(k-\phi)j)] \qquad (11)$$

where A is determined by the explicit form of the activation function O. For the case of activation function sin and $\phi = 0$, for instance, one can verify that A is determined by the equation $A = \beta N J_1(A)$, where J_1 is the Bessel function. The result eq. (11) indicates the following important rule. The solution is essentially given by the weight function passed through the non-linear activation function, O, just as the inputs themselves are passed through the weight function and then acted up by O. This rule will be generalized when we next analyze the case of Multi-Layer feed-forward Networks (MLN).

Although the solution of a SGen with a single Fourier component weight vector reveals many interesting results, it remains to be seen whether its features are common also for more general weight vectors. In particular it is interesting to verify the interplay, if any, between the structure of the power spectrum of the weights and the AD of the translating solution of the dynamics. In the following the analytical results for one Fourier component weights are extended to the case of two components

$$W_j = \sum_{m=1}^{2} R_m \cos[\frac{2\pi}{N}k_m j - \pi\phi_m] \qquad (12)$$

Note that the case $m > 2$ can be analyzed in a similar way. Following eq. (8), the general form of the solution of the output for the two component case is given by

$$s_j = \tanh[\sum_{m=1}^{2} A_m \cos(\ \frac{2\pi}{N}(k_m - \phi_m)j\ )] \quad . \qquad (13)$$

Inserting this solution into eq. (3) leads, for k_m, $k_{m'}$ relatively prime, to the following self-consistent coupled equations for the coefficients A_i as a function of β

$$A_m = \beta\tilde{A}_m N \sum_{\rho=1}^{\infty} C(\rho) \sum_{t=0}^{\rho-1} \frac{(A_m)^{2t}(A_{m'})^{2\rho-2t-2}}{(t+1)!t!(\rho-t-1)!^2} \qquad (14)$$

where $\tilde{A}_m \equiv A_m R_m \sin(\pi\phi_m)/\pi\phi_m$, $C(\rho) \equiv (2^{2\rho}-1)B_{2\rho}/\rho$, and $m' = 2$ for $m = 1$ and vice-versa. The solution of these two coupled equations reveals the following three regimes:

(a) In the small coupling phase, $\beta < \beta_1 \equiv min[\beta_{c(1)},\ \beta_{c(2)}]$, the only solution is the fixed point $s_j = 0$. Hence $\beta_{c(m)}$ stands for the critical coupling strength where the transition occurs in the corresponding one component case, eq. (10).

(b) At β_1, the SGen undergoes a supercritical bifurcation to a phase with only one non-zero component A, the one corresponding to the lower β_c. Since β_c is a function of both R and ϕ, this is not always the one with the largest amplitude. This observation was quantitatively confirmed in simulations [7] and also explains the result that for a BG with random weights the power spectrum of the output does not always follow the dominant component of the power spectrum of the weights. [4]

(c) For $\beta_2 > max[\beta_{c(1)},\ \beta_{c(2)}]$, there are two basins of attraction for the dynamics, each one simply the corresponding one component solution, see the inset of Fig. 2.

An important conclusion is that even in the two-component case, in the strong coupling phase the AD is generically equal to one. For $m > 2$, simulations indicates similar results, namely, that each translating solution has only one non-zero component and that the number of different translating solutions increases with β. Hence, the results with two components capture the behavior of a generic SGen and in particular the feature that the AD of a generic perceptron functioning as a SGen is one.

It is interesting to note that the model, eq. (3), has been extensively examined for the case $N = 2,\ 3$ in the context of an ANNNI model on a Cayley tree. [9] The introduction of bias (i.e. $k \neq 0$) in their studies, allows for, beside an additional "ferromagnetic" fixed point, the appearance of strange attractors, at least for $N = 2$. The fact that almost all orbits, for $N = 2$, are periodic, not quasiperiodic, is due to the absence of a meaningful phase, ϕ, in this case. For $N = 3$, the generic case for bias 0 is quasiperiodic, in agreement with our analysis.

Since the generic flows of SGens are restricted to an AD of one, independent of the complexity of the power spectrum of the weights, it is interesting to examine the generalization of the SGen to the case of a MLN. In this case the dynamics is the same but with the state of the leftmost input unit set equal to the output unit of the MLN in the previous time step. In particular it will be interesting to examine the interplay, if any, between the AD and the number of hidden units. Here we present results for only the simplest case of a MLN consisting of N input units, m hidden units and one output unit. The activation function of each one of the hidden units is tanh as given by eq. (3), the output is the linear summation of the hidden units and the weights of each one of the hidden units consists of only one Fourier component $W_j^q = R_q \cos[(2\pi/N)k_q j - \pi\phi_q]$. For this generalized SGen one can show [7] that the solution is of the type

$$s_j = \sum_{q=1}^{m} \tanh[A_q \cos(\frac{2\pi}{N}(k_q - \phi_q)j)] \qquad (15)$$

where the A_q's are given by the same equation as for the SGen, eq. (9). As the strength of the coupling, β, increases one finds that the number of non-zero A_q increases following eq. (10). Hence, one can conclude that the AD is a function of the coupling strength and is bounded from above by the number of hidden units, m. The AD of the generalized SGen is equal to the number of non-zero A_q. The detailed proof of this statement for the general case will be given elsewhere. [7] For example, where there are only two non-zero A_q, s_l can be written in the following way $s_l = \tanh(A_1 x) + \tanh(A_2 y)$ where x and y are two essentially uncorrelated numbers which can take values $\in [-1:1]$ with equal probability. Since there is a one dimensional continuum of (x, y) pairs with the same s_l, the possible s_{l+1} also form a one dimension continuum. Therefore the AD is two. A typical solution for $m = 2$ is presented in Fig.3.

The solution of the generalized SGen indicates that there is an interplay between the architecture of the MLN and the AD, the AD being a function of the coupling and the number of hidden units. This feature quantitatively distinguishes between the computational ability of MLN with different number of hidden units. Adding additional hidden units vastly enlarges the set of sequences generable with the network.

Finally we note that in the discussed MLN with tanh activation function both in the hidden units (with gain β) and in the output unit (with gain β'), the solution is of the form $s_l = \tanh[\beta'[\sum_q \tanh(A_q \cos(2\pi/N(k_q - \phi_q)j))]]$, which suggests that in the general case of a MLN , the translating solution schematically mirrors the architecture of the network itself, with the weight function on the lowest level acted upon by the activation functions of the hidden units and then the 2nd level weights and the top activation function. Details and further examples will be given elsewhere. [7]

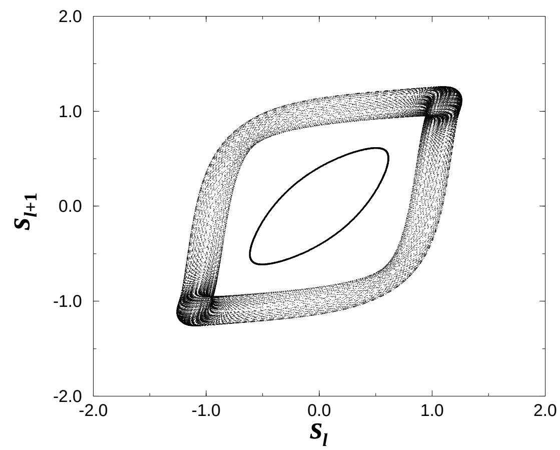

FIG. 3. Result of simulations for s_{l+1} vs. s_l for the generalized SGen consisting of two hidden units with $N = 100$, $k_i = \{11, 19\}$, $\phi_i = \{1/\sqrt{5}, 1/\sqrt{3}\}$ and $R_i = \{1, 0.7\}$. The orbit is for $\beta = .03$ and the surface for $\beta = .06$ If more points on the attractor are plotted, the surface fills densely.

IK thanks E. Domany and W. Kinzel for fruitful discussions. We thank W. Kinzel for bringing Ref. [9] to our attention. IK and DAK acknowledge the support of the Israel Academy of Sciences.

[1] A. S. Weigand and N. A. Gershenfeld (eds.), "Time Series Prediction", Santa Fe Proceedings, Addison Wesley, 1994.

[2] T. L. H. Watkin, A. Rau and M. Biehl, Rev. Mod. Phys. **65**, 499 (1993).

[3] J. A. Hertz, A. Krogh and R. G. Palmer, *Introduction to the Theory of Neural Computation*, (Addison-Wesley, CA (1991)).

[4] E. Eisenstein, I. Kanter, D. A. Kessler and W. Kinzel, Phys. Rev. Lett. **74**, 6 (1995).

[5] M. L. Minsky and S. Papert, " Perceptron" (MIT Press, Cambridge, 1969).

[6] M. Abramowitz and I. A. Stegun, "Handbook of Mathematical Functions" pg. 804 (Dover Publications, New York, 1970).

[7] I. Kanter, D. A. Kessler, A. Priel and E. Eisenstein (unpublished).

[8] T. Sauer, A. Yorke and M. Casdagli, J. Stat. Phys. **65**, 579 (1991).

[9] C. S. O. Yokoi, M. J. Oliviera and S. R. Salinas, Phys. Rev. Lett **54**, 163 (1985); J. G. Moreira and S. R. Salinas, J. Phys. A **20**, 1621 (1987).