Products of random matrices and investment strategies

Stefano Galluccio and Yi-Cheng Zhang
Institut de Physique Théorique, Université de Fribourg, CH-1700, Switzerland
(Received 13 August 1996)

A simple stochastic model of investment based on communication theory is introduced and analyzed in detail. We solve it exactly in a simple case and we use a weak disorder expansion to deal with the small fluctuations of the capital between two consecutive trading periods. Some possible generalizations are also discussed. [S1063-651X(96)52011-8]

PACS number(s): 02.50.Le, 05.20.−y

Let us consider an investor, whose aim is to increase a given capital $\mathcal{Z}$ by investing it in several stocks i $(i=1,\ldots,M)$ with a given strategy. The total amount of capital is the sum $\mathcal{Z}=\sum_{i=1}^{M}\mathcal{Z}(i)+\mathcal{Z}(0)$ where $\mathcal{Z}(i)$ is the fraction of the capital invested in the stock i and $\mathcal{Z}(0)$ is the capital left at the bank. The capital $\mathcal{Z}(i)$ is multiplied, at each trading period, by a factor a, a being a stochastic variable whose distribution is, in general, unknown. For our simple model we assume uncorrelated distributions; however, in reality, it is well known that such data can be correlated in time. We believe the following analysis can be readily generalized to the correlated case as well.

The investment strategy is the following: at each ''time'' n $(n=1,\ldots,N)$, one takes a fraction $\tau_i\in[0,1]$ of the capital $\mathcal{Z}(i)$ and transfers it to the cash bank; at the same time, a fraction σ/M of the bank capital is taken from the bank itself and then added to $\mathcal{Z}(i)$. We can then write our system as

$$\mathcal{Z}_{n+1}(0)=(1-\sigma)\mathcal{Z}_n(0)+\sum_{i=1}^{M}\tau_i\mathcal{Z}_n(i),$$

$$\mathcal{Z}_{n+1}(i)=a_i\frac{\sigma}{M}\mathcal{Z}_n(0)+a_i(1-\tau_i)\mathcal{Z}_n(i), \tag{1}$$

where, as above defined, we suppose that a_i are independent random variables with common distribution given by $\rho(a)$. The system is completely specified by a $(M+1)\times(M+1)$ random transfer matrix $\mathbf{A}_n$ since we can write $\vec{\mathcal{Z}}_{n+1}=\mathbf{A}_n\vec{\mathcal{Z}}_n$. Note that the system is normalized: if $a_i=1$ for all i, i.e., if there is no gain or loss at each time step, then the total capital is conserved: $\mathcal{Z}_{n+1}=\mathcal{Z}_n$. The first equation implies that after a trading period the amount deposited at the bank is the sum of two contributions: the part which has not been transferred and the part coming from $\mathcal{Z}(i)$. On the other hand, for each stock i, we have that $\mathcal{Z}(i)$ is incremented or diminished depending on the value of the stochastic variable a_i. Note that the choice of (1) is not the only one possible: one may define $\mathcal{Z}_{n+1}(i)=\sigma/M\,\mathcal{Z}_n(0)+a_i(1-\tau_i)\mathcal{Z}_n(i)$, e.g., we may decide not to bet the fraction of the capital transferred from the bank. This choice, however, has the disadvantage of rendering calculations much harder and leaving the qualitative behavior unchanged.

The goal is to optimize the capital gain by varying the set $\{\xi\}=\{\sigma,\tau_1,\ldots,\tau_M\}$ [that is to have the largest $\mathcal{Z}=\sum_{i=0}^{M}\mathcal{Z}(i)$ after N trading periods, in the limit $N\to\infty$]. How is it possible to decide the best strategy once we know $\rho(a)$, and with fixed M? We find that a special case of this problem can be mapped to a classical one in communication theory. Indeed in the 1950s Kelly considered this type optimization, albeit using information theory [1]. His model is based on the concept of the so-called *rate of transmission* in a communication channel, as introduced by Shannon [2]. Suppose one considers a channel and uses it to transmit a set of data between two distant units, e.g., the results of a change situation before they become common knowledge. A given gambler would like to use this source of information to put bets in order to increase his money. The question is: how much should he bet each time in order to maximize his gain? The answer, of course, relies on the ''quality'' of the signal he receives. If the channel is noiseless, i.e., if the gambler has exact information, then the best strategy is obviously to bet the whole amount of money he has. For example, if one is supposed to gain a fraction K of the capital after each good bet, then, after N steps he will have K^N times the original bankroll. The situation is more complicated if one considers a noisy channel, that is the situation in which the gambler has a given probability to get the correct information through the channel (as a result of noise and disturbances in the transmission). In that case if the gambler would bet at each time n his entire capital $\mathcal{Z}_n$, in order to maximize the expected value $\langle\mathcal{Z}_n\rangle$, then the result would be that if N is very large he will lose the whole capital. In fact in [1] it is shown that the correct strategy consists of maximizing the *exponential rate of growth of the gambler capital*, defined as

$$\lambda(\{\xi\})=\lim_{N\to\infty}\frac{1}{N}\left\langle\ln\frac{\mathcal{Z}_N}{\mathcal{Z}_0}\right\rangle. \tag{2}$$

This quantity has a great importance: in the mathematical theory of communication it represents the *capacity* of a discrete channel [2], while in theoretical physics corresponds to the *Lyapunov characteristic exponent*. It has important applications in chaotic dynamical systems, statistical mechanics of disordered systems, localization problems and turbulence. Without entering into any mathematical definition (the interested reader may see [3]) we point out that the problem of maximizing (2), instead of the usual mean, comes from the need of the *quenched* average in disordered systems. In fact, as is well known, if one considers the product $P_N=x_1x_2\cdots x_N$ of N independent stochastic variables, then the typical value will be given by $\exp(N\langle\ln x\rangle)$ and not by the *annealed* average $\langle x\rangle^N$, since only the fluctuations around the first quantity vanish for $N\to\infty$. In our model the product P_N is obviously given by $P_N=\Pi_{i=0}^N \mathcal{Z}_{i+1}/\mathcal{Z}_i$. Physically it means that if one tries to increase his capital by maximizing the annealed average $\langle \mathcal{Z}_N\rangle$ he could finally be broke, since the fluctuations around that quantity increase with N.

The noisy channel corresponds to the stochastic fluctuations of stock prices and a bet of the gambler to the investment in a given stock in one trading period.

Using our present notation, instead of (1), Kelly's model is specified by the following stochastic evolution equations (in his paper Kelly considered only the case $M=1$):

$$\mathcal{Z}_{n+1}(0)=(1-\sigma)\mathcal{Z}_n(0)+(1-\sigma)\mathcal{Z}_n(1),$$

$$\mathcal{Z}_{n+1}(1)=a\sigma\mathcal{Z}_n(0)+a\sigma\mathcal{Z}_n(1). \tag{3}$$

The physical meaning is that at each time step, after collecting the whole amount of money, one decides to preserve a fraction $1-\sigma$ and to bet a fraction σ. In other words, this can be considered as a ''nonlocal'' version of our model, since in (1) we have the freedom to decide how much of the capital $\mathcal{Z}(i)$ to bet, independently from others $\mathcal{Z}(j)$. In Kelly's model the random transfer matrix defined by (3) has vanishing determinant. This means that the two equations are not independent, and in fact it is simple to write the above system by means of a single stochastic equation for the global capital $\mathcal{Z}(0)+\mathcal{Z}(1)$. Therefore the problem reduces to a product of random numbers, rather than matrices, and it presents no mathematical troubles since the calculations of λ is in this case straightforward [3]. The main point is that matrices do not commute in general: Kelly's model was far easier to solve since numbers allow one to use a special order and then large number law applies.

Let us now turn back to the model defined in (1): in order to calculate the exponential rate of growth λ, we should consider the quantity (sometimes called *response function*) $R_n=|\vec{\mathcal{Z}}_{n+1}|/|\vec{\mathcal{Z}}_n|$, where $|\vec{\mathcal{Z}}_n|$ is, as usual, the norm of $\vec{\mathcal{Z}}_n$. The Oseledec theorem [4] ensures that the Lyapunov characteristic exponent is a self-averaging quantity, that is in (1) we can disregard the average $\langle\ \rangle$ and calculate λ from a single disorder realization. Equivalently, we can calculate λ from

$$\lambda(\{\xi\})=\lim_{N\to\infty}\frac{1}{N}\ln\|\mathbf{P}_N\|, \tag{4}$$

where $\mathbf{P}_N=\Pi_{n=1}^N\mathbf{A}_n$. Usually the norm $\|\ \|$ that appears in the above formula is the maximum of the eigenvalues of the spectrum of $\mathbf{P}_N$. Our problem is then reduced to the calculation of the Lyapunov characteristic exponent of a product of independent random matrices with a given distribution. In general, there is no hope to accomplish this goal analytically, apart from very special situations, for example the case in which all $\mathbf{A}_n$ commute with each other and one can simply apply the large numbers law to get the result that $\lambda=\max\{\langle\ln|\gamma_k|\rangle\}$ where γ_k are the eigenvalues of $\mathbf{A}$ [3]. This is, for instance, the case of Kelly's model.

At this point we are then forced to introduce some simplifications or rely on numerical simulations. The simplest nontrivial model is defined by a 2×2 transfer matrix; this implies investment in only one stock and a cash bank:

$$\mathbf{A}=\frac{1}{1+s}\begin{pmatrix}1 & Kt\\ sa & Ka\end{pmatrix}=\frac{1}{1+s}\mathbf{A}', \tag{5}$$

with $s=\sigma/(1+\sigma)$, $t=\tau/(1+\tau)$, and $K=(1+s)/(1+t)$. These new parameters are simply introduced for future convenience. We finally find that the rate of growth is given by

$$\lambda(s,t)=-\ln(1+s)+\lim_{N\to\infty}\frac{1}{N}\ln\,\mathrm{Tr}\left(\prod_{n=1}^{N}\mathbf{A}'_n\right). \tag{6}$$

It is easy to show that our system (1), for $M=1$, is deeply related to the one-dimensional (1D) random field Ising model (RFIM) since $\mathbf{A}'$ is mathematically equivalent to the transfer matrix of that system. For a general distribution $\rho(a)$ no analytical solutions are available even for this simplified model [5]. Derrida and Hilhorst showed that a naive Taylor expansion around the low-temperature phase $(t,s\to0)$ does not converge if $\langle a\rangle>1$.

In one nontrivial case, however, the full solution can be found, and it corresponds to the situation in which at each trading period, with probability p the fraction of capital $\mathcal{Z}(i)$ is increased by a factor $\alpha>1$, and with probability $1-p$ it is completely lost. In other words, we define a product of Bernoulli random matrices with density $\rho(a)=p\,\delta(a-\alpha)+(1-p)\,\delta(a)$. We outline the main steps of the calculation. Due to fact that $\mathcal{Z}(0)/\mathcal{Z}(1)$ has upper and lower bounds not depending on N, the response function can be expressed as $R_n=\mathcal{Z}_{n+1}(0)/\mathcal{Z}_n(0)=\mathcal{Z}_{n+1}(1)/\mathcal{Z}_n(1)$ and it satisfies the following nonlinear difference equation:

$$R_{n+1}=1+aK+aK\frac{st-1}{R_n}=F(R_n), \tag{7}$$

with initial conditions given by $R_0=R_1=1$. The solution of this equation is given by

$$R_n=\frac{f_1\zeta_1^{n+1}+f_2\zeta_2^{n+1}}{f_1\zeta_s^n+f_2\zeta_s^n},\quad f_{1,2}=\frac{\zeta_{1,2}-aK}{\zeta_{1,2}-\zeta_{2,1}},$$

$$\zeta_{1,2}=\frac{1+aK}{2}\pm[(1+aK)^2+4aK(st-1)]^{1/2}. \tag{8}$$

In general, the Lyapunov characteristic exponent λ can be calculated once the stationary distribution of the response function $\mathcal{P}(R)$ is known, since, supposing the system ergodic, $\lambda=\int\mathcal{P}(R)\ln(R)dR$. As we know the evolution of the response function, given by (7) and (8), one is tempted to write down an integral equation for $\mathcal{P}(R)$ by using the transformation laws between two stochastic variables R_{n+1} and R_n:

$$\mathcal{P}(R_{n+1})=\int da\,\rho(a)\int dR_n\mathcal{P}(R_n)\,\delta(R_{n+1}-F(R_n)). \tag{9}$$

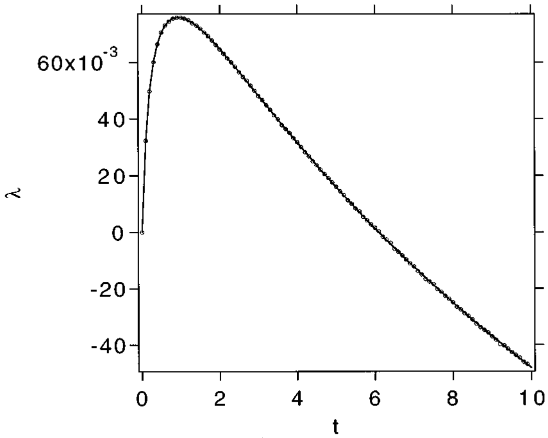

FIG. 1. The Lyapunov exponent plotted versus t (here $s=t$) for the case discussed in the text: $p=0.7$, $\alpha=2$. Solid line: exact solution [Eq. (10) of the text]. Circles: numerical simulation. The maximum of the capital gain is at $t\simeq 0.9356$.

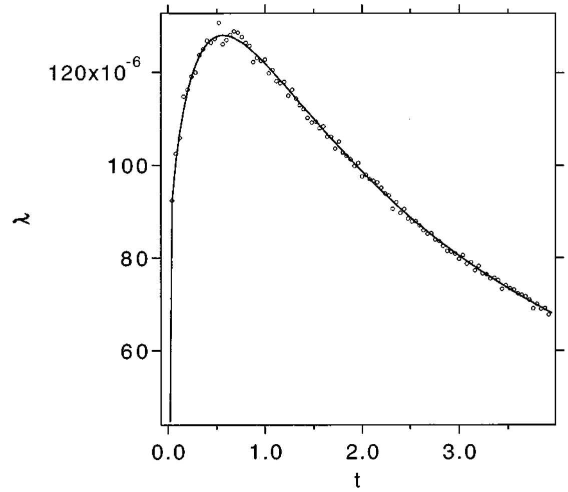

FIG. 2. The same as in Fig. 1 for the weak disorder case. We have chosen $\alpha=1$, $\varepsilon=0.05$, $p=0.51$. Solid line: analytical result from the perturbation expansion truncated at $O(\varepsilon^3)$. Circles: numerical simulation. The maximum is found at $t\simeq 0.564$.

The stationary distribution is the fixed point of the above integral equation (if it exists). Unfortunately, once the full expression for $F(R_n)$ is introduced, and the δ function is expressed in terms of the integration variable, the resulting integral equation is too difficult to be solved, in general. With our assumption on $\rho(a)$, however, it is simple to show that the solution must have the form

$$\mathcal{P}(R)=(1-p)\sum_{n=0}^{\infty} p^n \delta(R-R_n),$$

$$\lambda(s,t)=(1-p)\sum_{n=0}^{\infty} p^n[\ln(f_1\zeta_1^{n+1}+f_2\zeta_2^{n+1}) -\ln(f_1\zeta_1^{n}+f_2\zeta_2^{n})]-\ln(1+s), \tag{10}$$

with $a=\alpha$ in the preceding expression. In Fig. 1 we compare the exact solution (10) with the numerical calculation of λ for the particular case $s=t$, $\alpha=2$, $p=0.7$. We see that, as one may expect, there is a maximum of the exponential rate indicating which is the optimizing fraction of capital one has to invest to get the best profit. Due to the infinite sum appearing in (10) it is impossible to exactly calculate the maximum of λ. By taking into account the first 40 terms only [equivalent to an expansion of order $O(p^{40})$] we can numerically find the maximum. We obtain that $t_{\rm opt}\simeq 0.9356$, or $\tau_{\rm opt}\simeq 0.4834$, that is, the better strategy for that given set of parameters consists in transferring, at each step, a bit less than 50% of the capital.

For $s\neq t$ the situation is more interesting, but calculations become much more complicated, since now one has to find the maximum of a very complicated function of two variables. With the same truncated expression and the same set of parameters we finally obtained that $\sigma_{\rm opt}\simeq 0.652$ and $\tau_{\rm opt}\simeq 0.528$ in perfect agreement with the numerical solution.

Despite the complexity of this calculation, we need a more realistic situation since the hypothesis on $\rho(a)$ is still too strong. What actually happens in a given trading period (if it is short enough) is that the stock prices change by a small amount with respect to the step before, e.g., a few percent. This means, for example, that one can imagine defining a random matrix with distribution $\rho(a)=p\,\delta(a-(1+\alpha\varepsilon))+(1-p)\,\delta(a-(1-\alpha\varepsilon))$ with $\varepsilon\ll 1$. In the RFIM, that corresponds to a system to which a small external random field with opposite signs is applied. In our economic context this assumption means that at each step we are allowed to gain or lose only a small fraction of the total amount of money we have invested in a given stock. From the mathematical point of view we can use a weak disorder expansion in ε, that is, we consider our matrix $\mathbf{A}'$ as the sum of a diagonalizable matrix $\mathbf{B}$ and a random matrix $\mathbf{C}$ times a small expansion parameter: $\mathbf{A}'=\mathbf{B}+\varepsilon\mathbf{C}$. Here the fundamental hypothesis is the nondegeneration of the eigenvalues of $\mathbf{B}$ [6], since in the very general situation it is not possible to get a correct perturbation expansion [7]. If, however, the above hypothesis is fully satisfied, we have (see also [6])

$$\lambda=\ln\frac{\gamma_1}{1+t}-\frac{\varepsilon^2}{2}\frac{\langle C_{11}^2\rangle}{\gamma_1^2}+\frac{\varepsilon^3}{3}\frac{\langle C_{11}^3\rangle}{\gamma_1^3}+O(\varepsilon^4), \tag{11}$$

where now γ_1 is the maximum of the eigenvalues of $\mathbf{B}$, and C_{11} is the diagonal element of $\mathbf{C}$ (corresponding to γ_1) in the base in which $\mathbf{B}$ is diagonal. The average $\langle\ \rangle$ is made with the density $\rho(a)$. Note that there are no first order corrections to the pure (nonrandom) case. In our 2×2 case, from (5), and after some algebra, we get (for simplicity we have here considered the symmetric case $s=t$)

$$\gamma_1=\frac{1+u+v}{2},\quad \langle C_{11}^2\rangle=4p(1-p)\alpha^2\left(\frac{u+v+2t^2-1}{2v}\right)^2, \tag{12}$$

with $u=1+(2p-1)\alpha$ and $v=\sqrt{(u-1)^2+4ut^2}$. The result (up to third order in ε) is then compared in Fig. 2 with the numerical calculation, in the case $\alpha=1$, $\varepsilon=0.05$, and $p=0.51$. From Eqs. (11) and (12) it is a simple matter to analytically find the value of t at which λ attains its maximum: for the above case we find that $t_{\rm opt}\simeq 0.564$ or $\tau_{\rm opt}\simeq 0.353$.

The most interesting aspect of this approximation is that it can be employed, under certain hypotheses, for the general case in which one invests his capital in more than one stock ($M>1$). Due to the symmetry of the matrix, we can still perform the calculations, as we will show in a longer paper [8].

Another interesting situation is represented by the limit in which we take M very large. We could then introduce a trial mean-field approximation by considering $Z=1/M\Sigma_{i=1}^{M}\mathcal{Z}(i)\simeq\langle\mathcal{Z}\rangle$, and by defining, from (1), a new problem for the averaged capital $\langle\mathcal{Z}\rangle$ by means of the mean-field matrix

$$\mathbf{A}_{MF}=\frac{1}{1+t}\begin{pmatrix}1 & t\\ t\overline{a} & \overline{a}\end{pmatrix}, \tag{13}$$

where $\overline{a}=1/M\Sigma_{i=1}^{M}a_i$ and we have considered the symmetric case with all τ_i equal. If the a_i are, by hypothesis, independent stochastic variables with density $\rho(a)=p\,\delta(a-\alpha)+q\,\delta(a-\beta)$ $(p+q=1)$, one should expect, if M is large enough, that the distribution of $\overline{a}$ is Gaussian with mean equal to $\mu=p\alpha+q\beta$ and variance $[(p-p^2)\alpha^2+(q-q^2)\beta^2-2pq\alpha\beta]/M$. Unfortunately, even in the $M\rightarrow\infty$ limit, this is not a good approximation as one can easily see, for instance, by considering what happens in the limit $\tau\rightarrow 0$ in which all matrices commute and we can easily find the Lyapunov exponent. We have that $\lambda=\max\{0,p\ \ln\alpha+q\ \ln\beta\}$ in the true case, while from the mean-field approximation we obtain $\lambda=\ln(p\alpha+q\beta)$ (in the limit $M\rightarrow\infty$). This discrepancy is due to the fact that for $t\ll 1$ the limits $M\rightarrow\infty$ and $N\rightarrow\infty$ do not commute. This situation corresponds to the low-temperature phase of the 1D RFIM [5], in which the pure system attains a phase transition and then fluctuations are so strong that the mean-field picture completely fails (see Fig. 3).

Some approximation methods for calculating Lyapunov exponents are well known in the literature [3], but unfortunately they are not applicable to our case or they give very bad results. This is the case of the so-called *microcanonical approximation*, introduced by Deutsch and Paladin [9], which consists in replacing the quenched average in (2) with an annealed average and by imposing the constraint that the average is taken *only* on those configurations that satisfy the large number law. Even though it is usually believed that this approximation is a very satisfactory one (it gives a rigorous

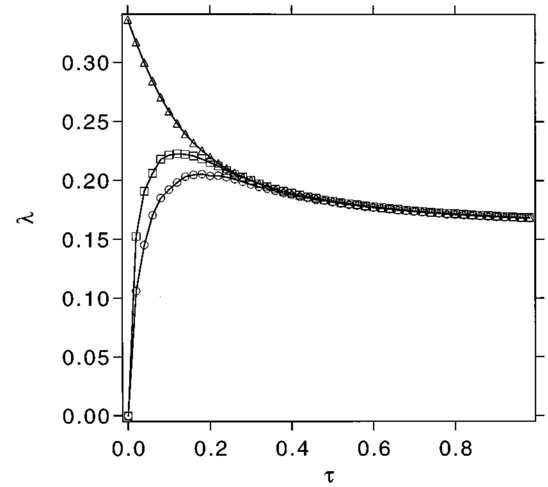

FIG. 3. Comparison between the mean-field solution for $s=t$ (△) and two numerical solutions in the large M limit. In particular, $M=10^3$ (○) and $M=10^4$ (□). Solid lines are guides for the eyes. Note the failure of the mean-field approximation in the ''low-temperature'' phase at $t\rightarrow 0$.

upper bound of the Lyapunov exponent), one can prove that in general this is not the case [8], since the fluctuations due to the noncommutative nature of the matrices are very strong. One should then consider the same annealed average but with more that one constraint [10]. In fact, this method seems to give very good approximations for the statistical mechanics of disordered magnetic systems. We will show our results based on the so-called *constrained annealing* approximation in our longer work.

In this paper we have analyzed a simple investment model. In traditional portfolio theory one is only interested in a single time period optimization. We show that for a very simple 2×2 matrix model the long time limit result can be obtained. In fact, the capital growth rate per unit time step can be interpreted as a Lyapunov exponent, which is a familiar concept in theoretical physics. For more general modes ($M>1$), we show, although exact calculations are hard to perform, meaningful approximations can yield interesting results. With the availability of more realistic data and distributions, doubtlessly the present study can be extended.

S.G. would like to thank Matteo Marsili for useful comments. This work has been supported by the Swiss National Foundation for Scientific Research.

[1] J. L. Kelly, Jr., Bell Syst. Tech. J. **35**, 917 (1956).
[2] C. Shannon, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).
[3] A. Crisanti, G. Palandin, and A. Vulpiani, *Products of Random Matrices in Statistical Physics*, Series in Solid States Sciences Vol. 104 (Springer-Verlag, Berlin, 1993).
[4] V. I. Oseledec, Trans. Moscow Math. Soc. **19**, 197 (1968).
[5] B. Derrida and H. J. Hilhorst, J. Phys. A **16**, 2641 (1983).
[6] B. Derrida, K. Mecheri, and J. L. Pichard, J. Phys. (France) **48**, 733 (1987).
[7] N. Zanon and B. Derrida, J. Stat. Phys. **50**, 509 (1988).
[8] S. Galluccio and Y.-C. Zhang (unpublished).
[9] J. M. Deutsch and G. Paladin, Phys. Rev. Lett. **62**, 695 (1989).
[10] M. Serva and G. Paladin, Phys. Rev. Lett. **70**, 105 (1993).