Controlled Lyapunov-exponents with application in optimization, finance and biology

László Gerencsér, Miklós Rásonyi, Zsuzsanna Vágó

Abstract—Let $X = (X_n)$ be a stationary process of $k \times k$ real-valued matrices, depending on some vector-valued parameter $\theta \epsilon D \subset \mathbb{R}^p$, satisfying

$$\mathrm{E} \log^+ \|X_0(\theta)\| < \infty,$$

for all $\theta \epsilon D$. The top-Lyapunov exponent of X is defined as

$$\lambda(\theta) = \lim_n \frac{1}{n} \mathrm{E} \log \|X_n \cdot X_{n-1} \ldots \cdot X_0\|.$$

We present an iterative scheme which converges to the parameter value optimizing $\lambda(\theta)$. Three potential applications are given: optimizing the convergence rate of a randomized optimization procedure, SPSA; the optimization of the growth rate of a portfolio's value; and the optimization of the growth rate of a biomass.

I. Random matrix-products

Let $X = (X_n), n = 0, 1, \ldots$ be a stationary process of $k \times k$ real-valued matrices over some probability space $(\Omega, \mathcal{F}, \mathcal{P})$, satisfying

$$\mathrm{E} \log^+ \|X_0\| < \infty, \tag{1}$$

where $\log^+ x$ denotes the positive part of $\log x$. It is well-known (see [1]) that under the above condition

$$\lambda = \lim_n \frac{1}{n} \mathrm{E} \log \|X_n \cdot X_{n-1} \cdots X_0\| \tag{2}$$

exists. Here $\lambda = -\infty$ is allowed. The following result is fundamental in multiplicative ergodic theory, see [1]:

Theorem 1: Assume that the process $X = (X_n)$ described above satisfies (1) and it is ergodic. Then P-almost surely

$$\lambda = \lim_n \frac{1}{n} \log \|X_n \cdot X_{n-1} \cdots X_0\|. \tag{3}$$

The number λ is the exponential rate of growth of the norm of the product $\|X_n \cdot X_{n-1} \ldots \cdot X_0\|$, and is called the top Lyapunov-exponent of the process $X = (X_n)$ for reasons that will become clear later.

We also recall a part of Oseledec's theorem (see [2] and [3]) which describes what happens if we apply the above random matrix products to a fixed vector.

Theorem 2: There exists a subset $\Omega' \subset \Omega$ of probability 1 such that for all $\omega \epsilon \Omega'$ there is a proper subspace $H(\omega) \subset \mathbb{R}^k$ of fixed dimension such that for all $v \epsilon \mathbb{R}^k \setminus H(\omega)$

$$\lim_n \frac{1}{n} \log \|X_n(\omega) X_{n-1}(\omega) \cdots X_0(\omega) v\| = \lambda.$$

Assume now that the matrices $X_n, n = 0, 1, \ldots$ depend on a common parameter, say θ, where $\theta \epsilon D \subset \mathbb{R}^p$, and D is an open domain. θ is considered as a control-parameter that we can set freely. Thus the top Lyapunov-exponent $\lambda = \lambda(\theta)$ will be a function of θ, and will be called a controlled Lyapunov-exponent. The problem that we consider in this paper is:

$$\min_{\theta \epsilon D} \lambda(\theta). \tag{4}$$

The maximization of $\lambda(\theta)$ is a completely analogous problem.

II. Minimization of the top-Lyapunov exponent

In developing an iterative procedure for solving the above minimization problem an alternative expression for $\lambda = \lambda(\theta)$ will play a key role. Let us define a $k \times k$ matrix-valued process $Z = (Z_n), n = 0, 1, \ldots$ as follows:

$$Z_n = X_n \cdot X_{n-1} \cdots X_0 / \|X_n \cdot X_{n-1} \cdots X_0\| \tag{5}$$

assuming that the denominator is not zero. In the latter case we write $Z_n = 0$. Obviously, $Z = (Z_n)$ can be defined recursively as follows:

$$Z_{n+1} = X_{n+1} Z_n / \|X_{n+1} Z_n\| \tag{6}$$

with initial condition $Z_0 = X_0 / \|X_0\|$, and the convention that $0/0 = 0$. It is easily seen that Theorem 1 implies

$$\lambda = \lim_n \frac{1}{n} \sum_{k=0}^{n-1} \log \|X_{k+1} Z_k\| \tag{7}$$

P-almost surely.

To compute the gradient of λ with respect to θ consider first the expression $\|XZ\|$ with $X \epsilon \mathbb{R}^{k \times k}$ fixed. Let $(Z(t)), t \geq 0$ be a smooth curve in $\mathbb{R}^{k \times k}$ with $Z(0) = Z, \dot{Z}(0) = \dot{Z}$ such that $XZ \neq 0$. Then simple calculus gives that at $t = 0$ we have

$$\frac{d}{dt} \|XZ(t)\| = \frac{1}{\|XZ\|} \mathrm{tr}\ (\dot{Z} Z^T X^T X). \tag{8}$$

A similar result is obtained if the roles of X and Z are interchanged. Thus we finally arrive at the following result:

Lemma 1: Let $X(t), Z(t), t \geq 0$ be smooth curves in $\mathbb{R}^{k \times k}$, with $X(0) = X, Z(0) = Z, \dot{X}(0) = \dot{X}, \dot{Z}(0) = \dot{Z}$, such that $XZ \neq 0$. Then at $t = 0$ we have that $(d/dt)\|X(t)Z(t)\|$ is equal to

$$\frac{1}{\|XZ\|} \mathrm{tr}\ (\dot{Z} Z^T X^T X + \dot{X} Z Z^T X^T). \tag{9}$$

Let us now consider the case where $X_n = X_n(\theta)$ is a smooth function of θ, as above, i.e. $\theta \epsilon D \subset \mathbb{R}^p$, and D is an open domain. Assume that $X_n(\theta)$ is non-singular for all n and all $\theta \epsilon D$. Thus we get a well-defined sequence $(Z_n) = (Z_n(\theta))$, and for all n $Z_n(\theta)$ is a smooth function of θ. Let θ_i for some $i = 1, ..., p$ be a fixed coordinate direction and let us introduce the notations

$$X_{\theta_i,n} = \frac{\partial}{\partial \theta_i} X_n(\theta) \quad Z_{\theta_i,n} = \frac{\partial}{\partial \theta_i} Z_n(\theta).$$

Formally differentiating (7) with respect to θ_i and using Lemma 1 we get that $\lambda_{\theta_i} = (\partial/\partial\theta_i)\lambda(\theta)$ equals

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \frac{1}{\|X_{k+1} Z_k\|^2} \mathrm{tr}(U_k + V_k), \tag{10}$$

where

$$\begin{aligned} U_k &= (Z_{\theta_i,k} Z_k^T X_{k+1}^T X_{k+1}, \\ V_k &= X_{\theta_i,k+1} Z_k Z_k^T X_{k+1}^T). \end{aligned}$$

Introduce the notation $\dot{H}(X, \dot{X}, Z, \dot{Z})$ for

$$\frac{1}{\|XZ\|^2} \mathrm{tr}\ (\dot{Z} Z^T X^T X + \dot{X} Z Z^T X^T)$$

and set

$$H_i(X, X_{\theta_i}, Z, Z_{\theta_i}) = \dot{H}(X, X_{\theta_i}, Z, Z_{\theta_i}).$$

Finally let $H(X, X_\theta, Z, Z_\theta)$ denote the vector

$$(H_1(X, X_{\theta_1}, Z, Z_{\theta_1}), \ldots, H_p(X, X_{\theta_p}, Z, Z_{\theta_p})).$$

It is assumed that the partial derivatives $X_{\theta_i,k+1}$ are computable explicitly. On the other hand the partial derivatives $Z_{\theta_i,k}$ will be computed recursively, taking into account the recursive definition of Z_n given in (6). For this purpose define the mapping of $\mathbb{R}^{k\times k} \times \mathbb{R}^{k\times k}$ into $\mathbb{R}^{k\times k}$ by

$$f(X, Z) = XZ/\|XZ\| \tag{11}$$

assuming that $XZ \neq 0$. To obtain the derivative of f with respect to Z let $X \epsilon \mathbb{R}^{k\times k}$ be fixed and let $(Z(t)), t \geq 0$ be a smooth curve in $\mathbb{R}^{k\times k}$ with $Z(0) = Z, \dot{Z}(0) = \dot{Z}$. Then at $t = 0$ we have

$$\frac{d}{dt} f(X, Z(t)) = \frac{X\dot{Z}}{\|XZ\|} - XZ \frac{1}{\|XZ\|^2} \frac{d}{dt} \|XZ(t)\|.$$

Taking into account (8) we get for the derivative $(d/dt) f(X, Z(t))$

$$\frac{X\dot{Z}}{\|XZ\|} - \frac{XZ}{\|XZ\|^3} \mathrm{tr}\ (\dot{Z} Z^T X^T X). \tag{12}$$

Now interchanging the role of X and Z we finally get:

Lemma 2: Let $X(t), Z(t), t \geq 0$ be smooth curves in $\mathbb{R}^{k\times k}$ with $X(0) = X, Z(0) = Z, \dot{X}(0) = \dot{X}, \dot{Z}(0) = \dot{Z}$ such that $XZ \neq 0$. Then at $t = 0$ we have for $(d/dt)\dfrac{XZ}{\|XZ\|}$

$$\frac{X\dot{Z}}{\|XZ\|} - \frac{XZ}{\|XZ\|^3} \mathrm{tr}\ (\dot{Z} Z^T X^T X) + \frac{\dot{X} Z}{\|XZ\|} - \frac{XZ}{\|XZ\|^3} \mathrm{tr}\ (\dot{X} Z Z^T X^T).$$

Thus we can write in short

$$\frac{d}{dt} f(X(t), Z(t)) = g(X, Z, \dot{X}, \dot{Z}), \tag{13}$$

with $g(X, Z, \dot{X}, \dot{Z})$ given above.

Applying the above notations we can express the derivatives $Z_{\theta_i,n}(\theta)$ in a recursive manner for any θ, and we have

$$Z_{\theta_i,n+1} = g(X_{n+1}, Z_n, X_{\theta_i,n+1}, Z_{\theta_i,n}). \tag{14}$$

The iterative scheme. The proposed iterative scheme for minimizing $\lambda(\theta)$ is as follows: at time n we have at our disposal the latest estimator θ_n and the matrices $X_n, X_{\theta,n}, Z_n, Z_{\theta,n}$. Observe $X_{n+1} = X_{n+1}(\theta_n)$ and compute $X_{\theta,n+1} = X_{\theta,n+1}(\theta_n)$. Then set

$$\begin{aligned} Z_{n+1} &= X_{n+1} Z_n / \|X_{n+1} Z_n\|, \\ Z_{\theta_i,n+1} &= g(X_{n+1}, Z_n, X_{\theta_i,n+1}, Z_{\theta_i,n}), \\ H_n &= H(X_{n+1}, X_{\theta,n+1}, Z_n, Z_{\theta,n}), \\ \theta_{n+1} &= \theta_n - \frac{1}{n} H_n \end{aligned} \tag{15}$$

An important technical tool is enforced boundedness which is achieved by resetting (cf. [4]): if θ_{n+1} would leave a compact domain then we reset to its initial value. Modulo some technical conditions, it is possible to show that the above iteration scheme converges to the optimal θ almost surely.

The algorithm formally falls within the class of recursive estimation methods described in [5] if X is a Markov-process, but the application of the results of [5] is not straightforward. In particular, [5] does not consider the effect of resetting. The convergence analysis requires completely different tools if X is non-Markovian. The first step is relatively easy: the extension of the ODE-method to recursive estimation processes with resetting, when the correction term is strictly stationary (asymptotically) for each fixed θ. The hard part is to establish uniform laws of large numbers with respect to θ for sums defined in terms of the process (X_{n+1}, Z_n).

The above arguments are applicable with minor changes to normalized processes of the form

$$z_n = X_n X_{n-1} \cdots X_0 v / \|X_n X_{n-1} \cdots X_0 v\|,$$

where v is a non-zero vector in $\mathbb{R}^k$.

We will also study simulation results so as to see the rate of convergence of the estimators θ_n.

III. Noise-free SPSA

Consider the following problem:

$$\min\ L(\theta),$$

where $L(\theta)$ is defined for $\theta \epsilon \mathbb{R}^p$. Assume that the computation of $L(.)$ is expensive and the gradient of $L(.)$ is not computable at all, and therefore, we need a numerical procedure to estimate the gradient of $L(.)$ denoted by

$$G(\theta) = L_\theta(\theta). \tag{16}$$

Following [6] we consider random perturbations of the components of θ. For this we first consider a sequence of independent, identically distributed (i.i.d.) random variables $\Delta_{ki},\ k=1,...,\ i=1,...,p$ defined over a probability space $(\Omega,\mathcal{F},\mathcal{P})$ satisfying certain weak technical conditions given in [6]. E.g. they may be chosen Bernoulli with

$$P(\Delta_{ki}=+1)=1/2 \qquad P(\Delta_{ki}=-1)=1/2.$$

Now let $c_k>0$ be a fixed sequence of numbers. For any $\theta\epsilon\mathbb{R}^p$ we evaluate $L(.)$ at two randomly and symmetrically chosen points $\theta+c_k\Delta_k$ and $\theta-c_k\Delta_k$, respectively. Define the random vector

$$\Delta_k^{-1}=\left[\Delta_{k1}^{-1},\ldots,\Delta_{kp}^{-1}\right]^T.$$

Then the estimator of the gradient is defined as

$$H(k,\theta)=\Delta_k^{-1}\frac{1}{2c_k}\Big(L(\theta+c_k\Delta_k)-L(\theta-c_k\Delta_k)\Big).$$

The fixed gain SPSA (simultaneous perturbation stochastic approximation) procedure is then defined by

$$\widehat{\theta}_{k+1}=\widehat{\theta}_k-aH(k+1,\widehat{\theta}_k) \tag{17}$$

with $a>0$ fixed.

The pecularity of the procedure is, that for $\theta=\theta^*$ and $c_k\to 0$ the correction term $H(k,\theta^*)$ vanishes asymptotically. Fixed gain SPSA methods have been first considered in [7] in connection with discrete optimization.

A main result is that fixed gain SPSA applied to noise-free optimization yields geometric rate of convergence almost surely, just like deterministic gradient methods under appropriate conditions, see [8]. The convergence properties of the proposed fixed gain SPSA method can be easily established for quadratic functions.

First, it is easy to see that for quadratic functions

$$H(k,\theta)=\Delta_k^{-1}\Delta_k^TG(\theta).$$

Since $G(\theta)=A(\theta-\theta^*)$, say, we get the following recursion for $\delta\theta_k=\theta-\theta^*$:

$$\delta\theta_{k+1}=(I-a\Delta_k^{-1}\Delta_k^TA)\delta\theta_k. \tag{18}$$

Now the sequence Δ_k is i.i.d., hence the matrix-valued process

$$A_k=(I-a\Delta_k^{-1}\Delta_k^TA)$$

is stationary and ergodic. Applying Oseledec's multiplicative ergodic theorem (cf. [2], [3]) and a discrete time version of the results in [9] we get the following result:

Theorem 3: Let L be a positive definite quadratic function,

$$L(\theta)=\frac{1}{2}(\theta-\theta^*)^TA(\theta-\theta^*),$$

and let $c_k=c$ be fixed. Then, for sufficiently small a there is a deterministic constant $\lambda<0$, depending on a, such that for any initial condition θ_0 outside of a set of Lebesgue-measure zero we have

$$\lim_{k\to\infty}\frac{1}{k}\log|\widehat{\theta}_k-\theta^*|=\lambda$$

with probability 1.

Simple adaptive noise-free SPSA procedures have been considered in earlier works. One procedure is to use two gains and choose the one in each step that gives smaller function value. To our knowledge the best switching strategy, minimizing the top-Lyapunov exponent is not known. The problem is hard even for two fixed matrices, and has been solved only recently by V. Blondel (yet unpublished).

IV. Growth rate of wealth processes

Let us consider portfolios consisting of k financial assets (investments) such that ϕ_n^i denotes the amount of money we have in asset $i=1,\ldots,k$ at time n. We refer to Chapter 1 of [10] for basic notions of discrete-time financial market models.

For the sake of simplicity, we deal with the case where $k=2$. Thus let us consider two stocks whose price at time n is given by

$$S_n^i:=\exp\left\{\mu_in+\sum_{j=1}^nW_j^i\right\},\ i=1,2,$$

where the μ_i are real constants, (W_n^1,W_n^2) is a stationary ergodic stochastic process, see p. 167 of [10]. Then the process S_{n+1}/S_n is also a stationary ergodic stochastic process.

Consider and investor pursuing the following strategy: if the present growth rate S_n^1/S_{n-1}^1 of the first stock is greater then that of the second stock, i.e. S_n^2/S_{n-1}^2, then he sells a proportion $\beta\epsilon(0,1)$ of the second stock and invests the money in the first. If the opposite situation occurs then he transfers a proportion α of his wealth in the first stock into the second. The parameters α,β are fixed by the investor.

It is easy to see that the portfolio dynamics is described by the following equations:

$$\begin{aligned}
\phi_{n+1}^1 &= \frac{S_{n+1}^1}{S_n^1}\Big[\phi_n^1+I_{\{S_n^1/S_{n-1}^1>S_n^2/S_{n-1}^2\}}\beta\phi_n^2\\
&\quad - I_{\{S_n^1/S_{n-1}^1<S_n^2/S_{n-1}^2\}}\alpha\phi_n^1\Big],\\
\phi_{n+1}^2 &= \frac{S_{n+1}^2}{S_n^2}\Big[\phi_n^2+I_{\{S_n^1/S_{n-1}^1<S_n^2/S_{n-1}^2\}}\alpha\phi_n^1\\
&\quad - I_{\{S_n^1/S_{n-1}^1>S_n^2/S_{n-1}^2\}}\beta\phi_n^2\Big].
\end{aligned} \tag{19}$$

It is obvious that the sequence ϕ_n is the image of ϕ_0 under the linear mapping that is represented by an n-fold random product of members of a stationary ergodic matrix-valued sequence. This is true even in the case when α and β depend on the price-process.

The total value of the portfolio is of the form

$$V_n = e^T \phi_n,$$

where all the components of the vector e are equal to 1. Then the growth rate of the total value will be

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \log V_n$$

which, under reasonable conditions, is equal to the top-Lyapunov exponent of (X_n). Thus we can apply the optimizing procedure proposed in section 2 and find the value of α, β which maximize the growth rate of the total value.

Another, even simpler example is where the investor keeps a fixed proportion $\alpha \epsilon (0, 1)$ of his wealth in asset 1 and the rest in asset 2 and rebalances his portfolio at each time step. Again, the scheme of section 2 provides the optimal α.

V. Population growth

Let us consider a population with k types of individuals. At each reproducing time $n = 0, 1, \ldots$ one individual of type i breeds X_n^{ji} individuals of type j, where $1 \leq i, j \leq k$. Let $X_n = (X_n^{ji})$. If the number of type i individuals at time n is ϕ_n^i, then

$$\phi_{n+1} = X_n \phi_n,$$

where $\phi_n = (\phi_n^1, ..., \phi_n^k)$.

It is reasonable to suppose that the environment varies in a random, but stationary way, and hence the rate of reproduction expressed by X is a matrix-valued stationary stochastic process. Consider now the example of a biomass where e.g. temperature is in our control. Then maximizing the population size can be performed by maximizing the top-Lyapunov exponent of the process $X(\theta)$.

Acknowledgement

The authors express their thanks to James C. Spall and John L. Maryak of the Applied Physics Laboratory of Johns Hopkins University for cooperating in this research. This research has been supported by the National Research Foundation (OTKA) of Hungary under grant no T 032932.

References

[1] H. Furstenberg and H. Kesten, "Products of random matrices", Ann. Math. Statist., vol. 31, pp. 457–469, 1960.

[2] V. I. Oseledec, "A multiplicative ergodic theorem. Lyapunov charasteristic numbers for dynamical systems", Trans. Moscow Math. Soc., vol. 19, pp. 197-231, 1968.

[3] M. S. Ragunathan, "A proof of Oseledec's multiplicative ergodic theorem", Israel Journal of Mathematics, vol. 42, pp. 356–362, 1979.

[4] L. Gerencsér, "Rate of convergence of recursive estimators", SIAM J. Control and Optimization, vol. 30, pp. 1200–1227, 1992.

[5] A. Benveniste, M. Métivier and P. Priouret, "Adaptive algorithms and stochastic approximations", Springer-Verlag, Berlin, 1990.

[6] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation", "IEEE Trans. Automat. Contr.", vol. 37, pp. 332–341, 1992.

[7] L. Gerencsér, S. D. Hill and Zs. Vágó, "Optimization over discrete sets via SPSA", Proceedings of the 38-th Conference on Decision and Control, CDC'99, Phoenix, pp. 1791–1794, 1999.

[8] L. Gerencsér and Zs. Vágó, "The mathematics of noise-free SPSA", Proceedings of the 40-th IEEE Conference on Decision and Control, CDC'01, Orlando, pp. 4400–4405, 2001.

[9] L. Gerencsér, "Almost sure exponential stability of random linear differential equations', Stochastics, vol. 36, 411–416, 1991.

[10] D. Lamberton and B. Lapeyre, "Introduction to Stochastic Calculus", Chapman & Hall, 1996.