CONTROLLED LYAPUNOV-EXPONENTS IN OPTIMIZATION AND FINANCE

L. Gerencsér*, **M. Rásonyi,*** **Zs. Vágó*†**

* Computer and Automation Institute of the Hungarian Academy of Sciences
H-1111 Budapest, Kende 13-17, Hungary.
email: gerencser@sztaki.hu, rasonyi@sztaki.hu
† PPCU Department of Information Technology
H-1052 Budapest Piarista kz 1.
email: vago@sztaki.hu

Keywords: Random products, Lyapunov exponents, SPSA, growth rate, recursive estimation

Abstract

Let $X = (X_n)$ be a stationary process of $k \times k$ real-valued matrices, depending on some vector-valued parameter $\theta \epsilon \mathbb{R}^p$, satisfying $\mathrm{E} \log^+ ||X_0(\theta)|| < \infty$ for all θ. The top-Lyapunov exponent of X is defined as

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

Top-Lyapunov exponents play a prominent role in randomization procedures for optimization, such as SPSA, and in finance, giving the growth-rate of a self-financing currency-portfolio with a fixed strategy. We develop a convergent iterative procedure for the optimization of $\lambda(\theta)$. In the case when X is a Markov-process, the proposed procedure is formally within the class defined in [1], however the general case requires fundamentally different techniques.

1 Random matrix-products

Let $X = (X_n), n = 0, 1, ...$ 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 [2]) that under the above condition

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

exists. Here $\lambda = -\infty$ is allowed. The following result is fundamental in multiplicative ergodic theory (see [2]):

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

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

The number λ, the exponential growth rate of the product $||X_n \cdot X_{n-1} ... \cdot X_0||$, 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 [8] and [6]) which describes what happens if we apply the above random matrix products to a fixed vector.

Theorem 2 *Under the conditions of Theorem 1 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 $X_n, n = 0, 1...$ 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:

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

A theoretical expression for λ can be obtained as follows (cf. [2]). Let $Y_k = X_k ... X_1$ and define the normalized products $Z_k = Y_k / ||Y_k||$. Then it can be shown that the process (Z_k, X_{k+1}) is asymptotically stationary. Let μ denote the stationary distribution of (X_2, Z_1). Then we have

$$\lambda = \int \log ||X_2 Z_1|| d\mu. \tag{5}$$

Obviously this expression is not very useful for practical computations.

2 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 $Z = (Z_n), n = 0, 1, ...$ as follows:

$$Z_n = X_n \cdot X_{n-1} ... \cdot X_0 / ||X_n \cdot X_{n-1} ... \cdot X_0|| \tag{6}$$

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:

$$Z_{n+1} = X_{n+1} Z_n / ||X_{n+1} Z_n|| \tag{7}$$

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

$$\log ||X_n \cdot X_{n-1} ... \cdot X_0|| = \sum_{k=0}^{n-1} \log ||X_{k+1} Z_k|| + \log ||X_0||.$$

Thus Theorem 1 implies

$$\lambda = \lim_n \frac{1}{n} \sum_{k=0}^{n-1} \log ||X_{k+1} Z_k|| \tag{8}$$

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 at $t = 0$ we have, using $||XZ(t)|| =$ tr $(XZZ^T X^T)^{1/2}$, that $\frac{d}{dt}||XZ(t)||$ equals

$$\frac{1}{2}\text{tr } (XZZ^T X^T)^{-1/2} \cdot \text{tr } (X\dot{Z}Z^T X^T + XZ\dot{Z}^T X^T).$$

Using the identities $\text{tr}A = \text{tr}A^T$ and $\text{tr}AB = \text{tr}BA$, we get

$$\frac{d}{dt}||XZ(t)|| = \frac{1}{||XZ||}\text{tr } (\dot{Z}Z^T X^T X). \tag{9}$$

Now let the role of X and Z be interchanged: let $Z \epsilon \mathbb{R}^{k\times k}$ be fixed and let $X(t)$ be a smooth curve in $\mathbb{R}^{k\times k}$ with $X0) = X$ such that $XZ \neq 0$. Proceed as above, and note that, in analogy with (9) we have

$$\frac{d}{dt}||X(t)Z|| = \frac{1}{||XZ||}\text{tr } (\dot{X}ZZ^T X^T). \tag{10}$$

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*

$$\frac{d}{dt}||X(t)Z(t)|| = \frac{1}{||XZ||}\text{tr } (\dot{Z}Z^T X^T X + \dot{X}ZZ^T X^T). \tag{11}$$

Let us now consider the case where 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

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

Differentiating (8), and using Lemma 1 with $X = X_{k+1}(\theta)$, $Z = Z_k(\theta)$ we get, after formal derivation, the following expression for $\lambda_{\theta_i} = (\partial/\partial\theta_i)\lambda(\theta)$:

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \frac{\text{tr } (Z_{\theta_i,k} Z_k^T X_{k+1}^T X_{k+1} + X_{\theta_i,k+1} Z_k Z_k^T X_{k+1}^T)}{||X_{k+1} Z_k||^2}. \tag{12}$$

Introduce the notations:

$$\begin{aligned} \dot{H}(X, \dot{X}, Z, \dot{Z}) &= \frac{\text{tr } (\dot{Z}Z^T X^T X + \dot{X}ZZ^T X^T)}{||XZ||^2} \\ H_i(X, X_{\theta_i}, Z, Z_{\theta_i}) &= \dot{H}(X, X_{\theta_i}, Z, Z_{\theta_i}) \\ H(X, X_\theta, Z, Z_\theta) &= (H_1(\ldots), ..., H_p(\ldots)). \end{aligned} \tag{13}$$

It is assumed that the partial derivatives $X_{\theta_i,k+1}$ are available *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 (7). For this purpose consider the mapping of $\mathbb{R}^{k\times k} \times \mathbb{R}^{k\times k}$ into $\mathbb{R}^{k\times k}$ defined by

$$f(X, Z) = XZ/||XZ|| \tag{14}$$

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 (9) we get

$$\frac{d}{dt} f(X, Z(t)) = \frac{X\dot{Z}}{||XZ||} - \frac{XZ}{||XZ||^3}\text{tr } (\dot{Z}Z^T X^T X). \tag{15}$$

Now interchanging the role of X and Z we get

$$\frac{d}{dt} f(X(t), Z) = \frac{\dot{X}Z}{||XZ||} - \frac{XZ}{||XZ||^3}\text{tr } (\dot{X}ZZ^T X^T). \tag{16}$$

Thus we arrive at the following result:

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*

$$\begin{aligned} \frac{d}{dt} XZ/||XZ|| &= \frac{X\dot{Z}}{||XZ||} + \frac{\dot{X}Z}{||XZ||} - \\ &- \frac{XZ}{||XZ||^3}\left(\text{tr } (\dot{Z}Z^T X^T X) + \text{tr } (\dot{X}ZZ^T X^T)\right) = \\ &= g(X, Z, \dot{X}, \dot{Z}). \end{aligned} \tag{17}$$

Thus we can write

$$\frac{d}{dt} f(X(t), Z(t)) = g(X, Z, \dot{X}, \dot{Z}). \tag{18}$$

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

$$Z_{\theta_i,n+1} = g(X_{n+1}, Z_n, X_{\theta_i,n+1}, Z_{\theta_i,n}). \tag{19}$$

The iterative scheme. Assume, that 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 $X_{\theta,n+1} = X_{\theta,n+1}(\theta_n)$. Then set

$$\begin{aligned} Z_{n+1} &= X_{n+1}Z_n/||X_{n+1}Z_n|| \\ Z_{\theta_i,n+1} &= g(X_{n+1}, Z_n, X_{\theta_i,n+1}, Z_{\theta_i,n}) \\ H_n &= H(X_{n+1}, X_{\theta,n+1}, Z_n, Z_{\theta,n}) \\ \theta_{n+1} &= \theta_n - \frac{1}{n}H_n \end{aligned} \quad (20)$$

An important technical tool is enforced boundedness which is achieved by resetting: if θ_{n+1} would leave a fixed compact domain then we reset to its initial value.

The algorithm formally falls within the class of recursive estimation methods described in [1] if X is a Markov-process, but the application of the results of [1] is not straightforward. In particular, [1] does not consider the effect of resetting. The convergence analysis requires completely different tools if X is not Markov. 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).

3 Noise-free SPSA

We consider the following problem:

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

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

$$G(\theta) = L_\theta(\theta). \quad (21)$$

Following [7] 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 some probability space $(\Omega, \mathcal{F}, \mathcal{P})$ satisfying certain weak technical conditions given in [7]. 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

$$\widehat{\theta}_{k+1} = \widehat{\theta}_k - aH(k+1, \widehat{\theta}_k) \quad (22)$$

with $a > 0$ fixed. The peculiarity 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 [4] 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 [5]. The convergence properties of the proposed fixed gain SPSA method can be easily established for quadratic functions. We have the following result:

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

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

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.

Sketch the proof: first, it is easy to see that for quadratic functions

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

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

$$\delta\theta_{k+1} = (I - a\Delta_k^{-1}\Delta_k^T A)\delta\theta_k. \quad (23)$$

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

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

is stationary and ergodic. Applying Oseledec's multiplicative ergodic theorem (cf. [8, 6]) the claim of the theorem follows immediately with some deterministic, not necessarily negative λ. To show that $\lambda < 0$ for small a we use the result of [3].

Simple adaptive procedures for noise-free SPSA have been considered in earlier works. A simple 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).

4 Growth rate of wealth-processes

Let us consider a currency portfolio $\phi = (\phi_n)$ consisting of k currencies. Thus $\phi_n = (\phi_{i,n}),\ i = 1,..,k$, where $\phi_{i,n}$ denotes the absolute size of the portfolio held in the i-th currency at time n. At any time n the exchange rates are collected in a $k \times k$ matrix β_n. Obviously, β_n is random. (β_n) will be assumed to be a strictly stationary process. Based on past and present values of β_n a rebalancing of the portfolio will take place, so that a certain fixed percentage of dollar will be converted into Euro or the other way round. This rebalancing can be described by a linear transformation:

$$\phi_{n+1} = X_n \phi_n, \tag{24}$$

where (X_n) is a strictly stationary sequence of $k \times k$ random matrices, describing the strategy of the investor. Let us focus on a parametric set of strategies $X = X(\theta)$. There is no reason to assume that (X_n) is a Markov-process. The wealth or the value of the portfolio in say Euros will be obtained from a scalar product of the form

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

where γ_n is an appropriate row of the random matrix β_n. Then the growth rate of the wealth 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). Its the maximization can be carried out by the procedure proposed in Section 1.

5 Simulation results

Our first experiments show the dependence of the top-Lyapunov exponent λ on the stepsize a in the fixed gain SPSA method. As a benchmark example, we considered the problem of minimizing a quadratic function of the form

$$L(\theta) = (\theta - \theta^*)^T A(\theta - \theta^*).$$

The minimizing point θ^* was generated uniformly within the unit cube. The matrix A was also generated randomly in the following way: first we generated the eigenvalues of A, λ_j, according to exponential distribution with parameter $\mu = 0.5$, and considered the matrix $\widetilde{A} = \text{diag}(\lambda_j)$. Then we applied randomly chosen rotations, and considered $A = T\widetilde{A}T^{-1}$, where T is the product of p random rotations.

In Figure 1 we plotted the estimation of the top-Lyapunov exponent λ as the function of the stepsize a.

In Figure 2 we plotted the estimation of the gradient of the top Lyapunov exponent, as it is computed in Section 2. It is seen, that the gradient vanishes around the minimizing point, $a \approx 0.05$.

In Figure 3 we plotted the result of the proposed iterative scheme to find the optimal control Lyapunov exponents in 3000 iterations.

Acknowledgement

The author expresses his 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] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*. Springer-Verlag, Berlin, 1990.

[2] H. Furstenberg and H.Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960.

[3] L. Gerencsér. Almost sure exponential stability of random linear differential equations. *Stochastics*, 36:411–416, 1991.

[4] L. Gerencsér, S. D. Hill, and Zs. Vágó. Optimization over discrete sets via SPSA. In *Proceedings of the 38-th Conference on Decision and Control, CDC'99*, pages 1791–1794. IEEE, 1999.

[5] L. Gerencsér and Zs. Vágó. A general framework for noise-free SPSA. In *Proceedings of the 40-th IEEE Conference on Decision and Control, CDC'01*, page submitted., 2001.

[6] M. S. Ragunathan. A proof of oseledec's multiplicative ergodic theorem. *Israel Journal of Mathematics*, 42:356–362, 1979.

[7] J.C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Automat. Contr.*, 37:332–341, 1992.

[8] V.I.Oseledec. A multiplicative ergodic theorem. Lyapunov charasteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, 19:197–231, 1968.

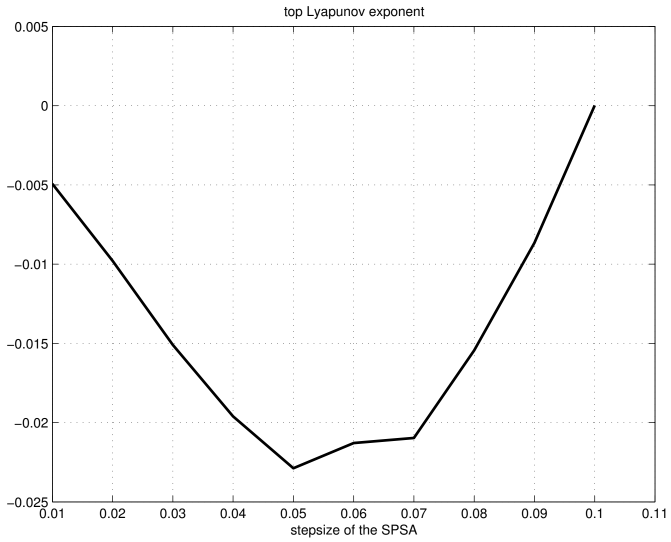

Figure 1: Top Ljapunov exponent as the function of the step-size.

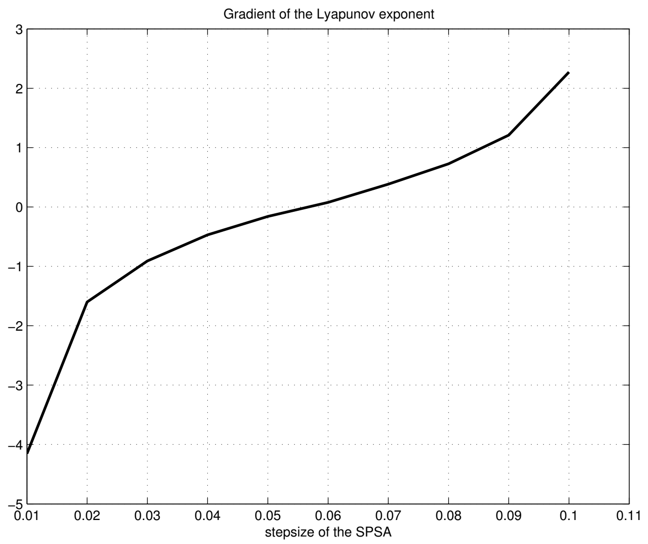

Figure 2: The gradient of the top-Lyapunov exponent.

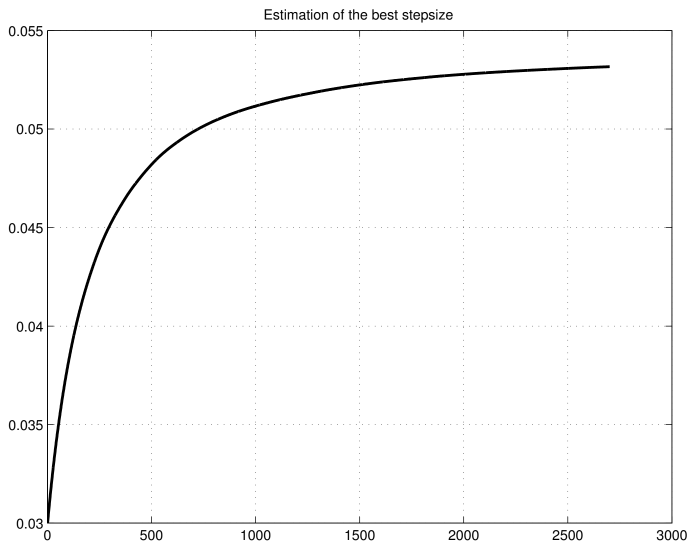

Figure 3: The result of the iterative scheme.