A generalised Golden-Section algorithm for line-search*

Luc Pronzato†, Henry P. Wynn‡and Anatoly A. Zhigljavsky§

Abstract—This paper aims at promoting a generalisation of the Golden-Section line-search algorithm, with better performances for functions locally symmetric around their optimum. Any line-search algorithm can be represented as a non-convergent dynamic system when a suitable renormalisation of the uncertainty interval is performed at each iteration. This allows a detailed study of the finite sample and asymptotic behaviour of the algorithm through a Markov-chain representation. We show that an expansion of the initial uncertainty interval improves this behaviour. Some asymptotic characteristics based on the evolution of the length of the uncertainty interval are shown to be related to classical ergodic characteristics of dynamic systems, such as Lyapunov exponent and Kolmogorov entropy. However, other ergodic characteristics, related to the Renyi entropy, are suggested as being more suitable in this context.

Abbreviated title—A generalised Golden-Section algorithm

Key words—Golden Section, Fibonacci, line search, dynamic system, Markov chain, entropy

AMS subject classification—65-K05, 68-Q25, 28-D20

*This work was supported by a French-British Alliance grant (n94002), by the UK Science and Engineering Research Council grant for the second author, and by a grant of the French Ministère de l'Education Nationale (invitation triennale, procédure PAST) for the third author.

†Laboratoire I3S, bât. 4, 250 rue A. Eistein, CNRS-URA 1376, Sophia Antipolis, 06560 Valbonne, France. Author for correspondence. Tel: (33) 92 94 27 08, Fax: (33) 92 94 28 98, e-mail: pronzato@unice.fr

‡Dept of Statistics, University of Warwick, Coventry, CV4 7AL, UK

§Dept of Mathematics, St.Petersburg University, Bibliotechnaya sq. 2, 198904, Russia

1 Introduction

We consider the minimisation of a uniextremal function $f(.)$ on a given interval $[A, B]$ using a "second-order" algorithm, as defined by Kiefer (1957). Let x^* be the unknown point at which $f(.)$ is minimum, with $f(.)$ monotonously increasing (resp. decreasing) for $x > x^*$ (resp. $x < x^*$). At iteration n we compare the values of $f(.)$ at two points U_n and V_n in the current uncertainty interval $[A_n, B_n]$, with $U_n < V_n$. Then, if $f(U_n) \geq f(V_n)$ we delete $[A_n, U_n)$, otherwise we delete $(V_n, B_n]$. Note that, in a practical implementation of the algorithm, both $[A_n, U_n)$ and $(V_n, B_n]$ can be deleted in the case where $f(U_n) = f(V_n)$ (but the algorithm should then be reinitialised). This will not be considered here because it has no effect on the performance characteristics that are considered. The remaining part of the interval defines the uncertainty interval $[A_{n+1}, B_{n+1}]$ for the next iteration, see Figure 1. On this figure, (R) and (L) stand respectively for Right and Left deletion. In each case, one of the two points U_n, V_n is carried forward to $[A_{n+1}, B_{n+1}]$. Let E_{n+1} denote this point. At iteration $n+1$ we thus only need to compare $f(E_{n+1})$ to the value of $f(.)$ at a new point E'_{n+1}. Notice that the actual values of the function $f(.)$ are not used, which makes it possible to optimize a function defined e.g. by a convergent series, for which arbitrary precise bounds can be constructed. A second-order line-search algorithm is therefore defined by the choice of the:

(i) initial uncertainty interval $[A_1, B_1] \supseteq [A, B]$,

(ii) initial test-point $E_1 \in [A_1, B_1]$,

(iii) selection rule for E'_{n+1}, $n \geq 0$.

POSSIBLE LOCATION OF FIGURE 1

The most famous algorithm in this class is the Golden-Section (GS) algorithm, for which

$$[A_1, B_1] = [A, B]\,, \quad E_1 = A_1 + \lambda L_1\,,$$
$$E'_n = \begin{cases} A_n + \lambda L_n & \text{if} \quad E_n = A_n + (1-\lambda)L_n\,, \\ A_n + (1-\lambda)L_n & \text{if} \quad E_n = A_n + \lambda L_n\,, \end{cases}$$

where $L_n = B_n - A_n$ and where $\lambda = \frac{\sqrt{5}-1}{2} \simeq 0.61804$ is the solution of $\lambda^2 + \lambda = 1$ and is called the Golden-Section ratio. The key property of the algorithm is that E_{n+1} satisfies:

$$\frac{E_{n+1} - A_{n+1}}{L_{n+1}} \in \{1-\lambda, \lambda\}\,. \tag{1}$$

This algorithm is known to be asymptotically worst-case optimal in the class of all uniextremal functions, see [5] and [4], Theorem 9.2.2, p. 181. The reduction, or convergence, rate at iteration n is defined as

$$r_n = \frac{L_n}{L_{n+1}}\,, \tag{2}$$

so that

$$L_n = L_0 \prod_{i=0}^{n-1} r_i \,,$$

with $L_0 = B - A$. For the GS algorithm, $r_0 = 1$ and $r_n = \lambda$, $\forall n \geq 1$, so that $L_n = L_0 \lambda^{n-1}$. When the number of function evaluations is fixed, say equal to N, the worst-case optimal algorithm in the sense of L_N is the Fibonacci algorithm [5], for which

$$L_N = \frac{L_0}{F_{N+1}} \,, \tag{3}$$

where $\{F_i\}_{i=1}^{\infty}$ is the Fibonacci sequence, defined by $F_1 = F_2 = 1$ and $F_n = F_{n-1} + F_{n-2}, n > 2$.

This paper aims at promoting a generalisation of the GS algorithm, with four possible values for $\frac{E_{n+1}-A_{n+1}}{L_{n+1}}$, see (4). This algorithm, which we call GS4, has better asymptotic and finite sample performances than the GS algorithm for functions locally symmetric around their minimum point x^*, and is the best within a class presented in [7, 8].

1.1 Outline of the paper

The algorithm GS4 is presented in Section 2. A key idea for the study of its behaviour is renormalisation, which is explained in Section 3, where we also introduce a dynamic system and Markov chain associated with the algorithm. Section 4 is devoted to the study of the finite sample behaviour of the algorithm GS4. Section 4.1 introduces some performance characteristics for line-search algorithms. The case of functions symmetric around x^* is considered in Section 4.2. Robustness of the performance characteristics with respect to non-symmetry is investigated in Section 4.3. The study of the asymptotic behaviour of the algorithm GS4 is considered in Section 5. The invariant measure for the dynamic system associated with the algorithm is presented in Section 5.1. The ergodic properties of the algorithm are studied in Section 5.2. In Section 5.3 we establish relations between limits of the performance characteristics of Section 4.1 and various entropies of the dynamic system.

Although we only consider here the particular case of the GS4 algorithm, the methodology which is used for studying convergence can be applied to some other search algorithms, see, e.g. [7]. Among the results that will be presented we want to draw special attention on the following ones: (i) expansion of the initial uncertainty interval can improve finite sample and asymptotic performances, see Sections 4.2 and 5.2; (ii) classical ergodic characteristics of dynamic systems, such as Lyapunov exponent or Kolmogorov entropy may not be suitable as performance characteristics for search algorithms, see Section 5.2 and 5.3.

2 The GS4 algorithm

We generalise the property (1) of the GS algorithm to

$$\frac{E_{n+1} - A_{n+1}}{L_{n+1}} \in \{a, b, c, d\}\,, \tag{4}$$

with a, b, c and d in $[0, 1]$. Such an algorithm is called 4-point cycle in [8]. As pointed out in this paper, there are infinitely many possible choices for (a, b, c, d). For the choice proposed hereafter, the ergodic characteristics of Section 5.2 are close to their best achievable values, and, moreover, the finite sample characteristics of Section 4.1 are close to their limits.

Define a as the smallest real positive root of

$$2t^4 - 8t^3 + 11t^2 - 7t + 1 = 0\,, \tag{5}$$

and

$$a' = 2a - a^2\,,\quad b = 2a^3 - 4a^2 + 3a\,,\quad c = 1 - b\,,\quad d = 1 - a\,, \tag{6}$$

so that

$$a \simeq 0.19412\,,\quad a' \simeq 0.35055\,,\quad b \simeq 0.44625\,,\quad c \simeq 0.55375\,,\quad d \simeq 0.80588\,.$$

The algorithm GS4 is then defined as follows.

Step (0): Let $f(.)$ be the function to be minimized on $[A, B]$. Choose N and/or δ to define the stopping rule of Step (iii).

Step (i): Compute $A_1 = A - \epsilon(B - A)$, $B_1 = B + \epsilon(B - A)$, with $\epsilon = \frac{1-a}{2} \simeq 0.43008$. Take $E_1 = A_1 + b(B_1 - A_1)$, compute $f(E_1)$ and set $n = 1$.

Step (ii): Compute E'_n according to the following rule :

$$E'_n = \begin{cases} A_n + a'L_n & \text{if} \quad E_n = A_n + aL_n\,, \\ A_n + cL_n & \text{if} \quad E_n = A_n + bL_n\,, \\ A_n + bL_n & \text{if} \quad E_n = A_n + cL_n\,, \\ A_n + (1 - a')L_n & \text{if} \quad E_n = A_n + dL_n\,. \end{cases} \tag{7}$$

Compute $f(E'_n)$.

$$\begin{array}{lll} \text{If } E_n < E'_n & \text{set} & U = E_n\,,\ V = E'_n\,,\ f_U = f(E_n)\,,\ f_V = f(E'_n)\,, \\ \text{otherwise} & \text{set} & U = E'_n\,,\ V = E_n\,,\ f_U = f(E'_n)\,,\ f_V = f(E_n)\,. \end{array}$$

Step (iii):

$$\begin{array}{lll} \text{If } f_U < f_V & \text{set} & [A_{n+1}, B_{n+1}] = [A_n, V]\,, \\ \text{otherwise} & \text{set} & [A_{n+1}, B_{n+1}] = [U, B_n]\,. \end{array}$$

If $n + 1 \geq N$ and/or $L_{n+1} \leq \delta$ stop; otherwise set $n + 1 \to n$ and go to Step (ii).

A key point in the structure of the algorithm is that for any (a', b, c, d) satisfying (6), E_{n+1} satisfies (4) whatever the value of a in $(0, 1-\lambda)$, where λ is the Golden Section. This can easily be checked. The values of a and ϵ could be chosen optimally for each N and each performance criterion. Here, the choice indicated for a is made as follows. We first optimise numerically the value of the ergodic characteristic $\lim_{N\to\infty} \frac{1}{N} \log EL_N$, see Section 5.2, obtained from simulations. This defines a small range of suitable values for a. We then select in this range a value of a according to some special criterion to be described in Section 3.3. This gives equation (5). The value $\epsilon = \frac{1-a}{2}$ makes the interval $[\frac{A-A_1}{L_1}, \frac{B-A_1}{L_1}]$ coincide with the support of the invariant density of the dynamic system (x_n, e_n) (9) of Section 3.2, conditional on $e_n = b$. Also, this value happens to be close to the optimal values obtained numerically for different N and different performance criteria. The fact that the choice of $\epsilon \geq 0$, which corresponds to an expansion of the initial uncertainty interval, only affects the initialisation of the dynamic system presented in Section 3.2, has strong influence on some asymptotic performance characteristics will be enhanced in Section 5.2. For that reason, we also detail the behaviour of the algorithm with $\epsilon = 0$.

3 Renormalisation and dynamic system

3.1 Renormalisation

The cornerstone of the study of the behaviour of the algorithm is renormalisation. After (L) or (R) deletion, we renormalise each uncertainty interval $[A_n, B_n]$ to $[0, 1]$. Thus introduce normalised variables in $[0, 1]$:

$$x_n = \frac{x^* - A_n}{L_n}\,, \quad e_n = \frac{E_n - A_n}{L_n}\,, \quad e'_n = \frac{E'_n - A_n}{L_n}\,,$$

and $u_n = \min(e_n, e'_n)$, $v_n = \max(e_n, e'_n)$. Straightforward calculations then show that right and left deletions respectively give:

$$x_{n+1} = \begin{cases} \frac{x_n}{v_n} & (R) \\ \frac{x_n - u_n}{1-u_n} & (L) \end{cases} \tag{8}$$

Moreover, from the definition of E_{n+1}, we obtain

$$e_{n+1} = \begin{cases} \frac{u_n}{v_n} & (R) \\ \frac{v_n - u_n}{1-u_n} & (L) \end{cases}$$

In the case of the GS algorithm, (1) implies that $e_n \in \{1-\lambda, \lambda\}\,,\ n \geq 1$, and the algorithm can be summarized as in Table 1.

Each ordered pair $(e, e') \in \{(1-\lambda, \lambda), (\lambda, 1-\lambda)\}$ can be interpreted as a state of the algorithm. Similarly, for the GS4 algorithm we obtain Table 2, where a, b, c, d and a' are defined by (5-6).

e_n	e'_n	e_{n+1}	
		(L)	(R)
$1-\lambda$	λ	$1-\lambda$	λ
λ	$1-\lambda$	$1-\lambda$	λ

Table 1: The GS algorithm.

e_n	e'_n	e_{n+1}	
		(L)	(R)
a	a'	a	c
b	c	a	d
c	b	a	d
d	$1-a'$	b	d

Table 2: The GS4 algorithm.

3.2 Dynamic system representation

When the function $f(.)$ is symmetric with respect to x^*, the decision to left or right delete only depends on the position of x^* with respect to $\frac{E_n+E'_n}{2}$. In the renormalised form we thus obtain

$$\begin{cases} (R) & \text{if } x_n < \frac{e_n+e'_n}{2}\,, \\ (L) & \text{if } x_n \geq \frac{e_n+e'_n}{2}\,. \end{cases}$$

For the GS algorithm, the updating rule (8) simply becomes

$$x_{n+1} = \begin{cases} (1+\lambda)x_n & \text{if } x_n < \frac{1}{2}\,, \\ (1+\lambda)x_n - \lambda & \text{if } x_n \geq \frac{1}{2}\,, \end{cases}$$

which defines a dynamic system. Its ergodic behaviour was studied in [10], and it was shown in [6] that this ergodic behaviour is the same when $f(.)$ is only locally symmetric at x^*.

For GS4, the updating rule is

$$(x_{n+1}, e_{n+1}) = \begin{cases} (\frac{x_n}{a'}, c) & \text{if } e_n = a \text{ and } x_n < \frac{a+a'}{2}\,, \\ (\frac{x_n-a}{1-a}, a) & \text{if } e_n = a \text{ and } x_n \geq \frac{a+a'}{2}\,, \\ (\frac{x_n}{c}, d) & \text{if } (e_n = b \text{ or } e_n = c) \text{ and } x_n < \frac{b+c}{2}\,, \\ (\frac{x_n-b}{c}, a) & \text{if } (e_n = b \text{ or } e_n = c) \text{ and } x_n \geq \frac{b+c}{2}\,, \\ (\frac{x_n}{d}, d) & \text{if } e_n = d \text{ and } x_n < 1 - \frac{a+a'}{2}\,, \\ (\frac{x_n-(1-a')}{a'}, b) & \text{if } e_n = d \text{ and } x_n \geq 1 - \frac{a+a'}{2}\,. \end{cases} \tag{9}$$

Due to the symmetry of the possible values of e_n in $[0,1]$ and the symmetry of the possible choices of e'_n, see Table 2 and (6), the dynamic system can be simplified as

follows. Define for $n \geq 2$

$$(y_n, g_n) = \begin{cases} (x_n, e_n) & \text{if } e_n = a \text{ or } e_n = b\,, \\ (1 - x_n, 1 - e_n) & \text{if } e_n = c \text{ or } e_n = d\,. \end{cases}$$

Since the rate of convergence of the algorithm is the same for the initial values x_1 and $1 - x_1$, we define

$$(y_1, g_1) = \begin{cases} (x_1, b) & \text{if } x_1 < \frac{1}{2}\,, \\ (1 - x_1, b) & \text{if } x_1 \geq \frac{1}{2}\,, \end{cases}$$

so that $y_1 \leq \frac{1}{2}$. Then the new system (y_n, g_n) obeys the simplified updating rule:

$$(y_{n+1}, g_{n+1}) = \begin{cases} (1 - \frac{y_n}{a'}, b) & \text{if } g_n = a \text{ and } y_n < \frac{a+a'}{2}\,, \\ (\frac{y_n - a}{1-a}, a) & \text{if } g_n = a \text{ and } y_n \geq \frac{a+a'}{2}\,, \\ (1 - \frac{y_n}{c}, a) & \text{if } g_n = b \text{ and } y_n < \frac{1}{2}\,, \\ (\frac{y_n - b}{c}, a) & \text{if } g_n = b \text{ and } y_n \geq \frac{1}{2}\,. \end{cases} \tag{10}$$

The price for this simplification is that knowing (y_n, g_n) we do not know whether (x_n, e_n) equals $(y_n, g_n), (1 - y_n, g_n), (y_n, 1 - g_n)$ or $(1 - y_n, 1 - g_n)$. However, this has no consequence on the calculation of the performance characteristics presented in Section 4.1. In order to obtain a one-dimensional dynamic system on $[0, 1]$, we now introduce

$$\begin{cases} z_n = \frac{y_n}{2} & \text{if } g_n = a\,, \\ z_n = \frac{1+y_n}{2} & \text{if } g_n = b\,, \end{cases}$$

which gives

$$z_{n+1} = T(z_n) = \begin{cases} 1 - \frac{z_n}{a'} & \text{if } z_n < \frac{a+a'}{4}\,, \\ \frac{2z_n - a}{2(1-a)} & \text{if } \frac{a+a'}{4} \leq z_n < \frac{1}{2}\,, \\ \frac{1}{2} - \frac{2z_n - 1}{2c} & \text{if } \frac{1}{2} \leq z_n < \frac{3}{4}\,, \\ \frac{2z_n - 1 - b}{2c} & \text{if } \frac{3}{4} \leq z_n\,. \end{cases} \tag{11}$$

The transformation $T(.)$ is presented on Figure 2. Note that the convention used to define the mapping $T(.)$ at $1/2$ is arbitrary since $1/2$ can be reached only if the process is initialised at $x^* = A_1$ or B_1, which is impossible if $\epsilon > 0$.

POSSIBLE LOCATION OF FIGURE 2

In order to be able to base our study of the performances the GS4 algorithm on that of the dynamic system (11), we need to know the convergence rate (2) obtained at each iteration of (11). This is given by:

$$r_n = \begin{cases} a' & \text{if } z_n < \frac{a+a'}{4}\,, \\ 1 - a & \text{if } \frac{a+a'}{4} \leq z_n < \frac{1}{2}\,, \\ c & \text{if } \frac{1}{2} \leq z_n\,. \end{cases} \tag{12}$$

Note that $r_n = \frac{1}{|T'(z_n)|}$, that is r_n corresponds to the inverse of the modulus of the slopes of the piecewise linear mapping $T(.)$ (a term which appears in the Perron-Frobenius equation in Section 5.1).

3.3 Markov-chain representation

Consider the collection $\mathcal{A}$ of all points mentioned in the right-hand side of (11), that is $\mathcal{A} = \{0, \frac{a+a'}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ and define $\mathcal{S}^\infty = \cup_{n=0}^\infty T^n(\mathcal{A})$, where $T^n(\mathcal{A}) = \{T^n(z), z \in \mathcal{A}\}$. Straightforward calculations, using computer algebra, show that when a is chosen as indicated in Section 2, $\mathcal{S}^\infty$ is finite and given by $\mathcal{S}^\infty = \{\alpha_0, \alpha_1, \ldots, \alpha_{12}\}$, with

$$\begin{aligned}
\alpha_0 &= 0\,, \quad \alpha_1 = \frac{a}{4} \simeq 0.04853\,, \quad \alpha_2 = \frac{a^3}{2} - 2a^2 + \frac{9}{4}a - \frac{1}{4} \simeq 0.11506\,, \\
\alpha_3 &= -\frac{a^2}{4} + \frac{3}{4}a \simeq 0.13617\,, \quad \alpha_4 = \frac{a^3}{2} - \frac{3}{2}a^2 + \frac{5}{4}a \simeq 0.18978\,, \quad \alpha_5 = \frac{1}{4}\,, \\
\alpha_6 &= \frac{1}{4} + \frac{a}{4} \simeq 0.29853\,, \quad \alpha_7 = \frac{1}{2}\,, \quad \alpha_8 = \frac{a^3}{2} - a^2 + \frac{3}{4}a + \frac{1}{2} \simeq 0.61156\,, \\
\alpha_9 &= \frac{a^2}{2} - \frac{a}{2} + \frac{3}{4} \simeq 0.67178\,, \quad \alpha_{10} = \frac{3}{4}\,, \quad \alpha_{11} = \frac{a^3}{2} - a^2 + \frac{3}{4}a + \frac{3}{4} \simeq 0.86156\,, \quad \alpha_{12} = 1\,.
\end{aligned}$$

The set $\mathcal{S}^\infty$ defines a partition of $[0,1]$ into 12 intervals $\mathcal{I}_i = [\alpha_{i-1}, \alpha_i], i = 1, \ldots, 12$. The points α_i are presented in Figure 2.

Let us take a Bayesian viewpoint, and assume that z_1 has a probability density $\phi_z^1(.)$ on $[0,1]$, such that $\phi_z^1(z)$ is constant on each interval $\mathcal{I}_i, i = 1, \ldots, 12$. In particular, the GS4 algorithm presented in Section 2 gives, when x^* is uniformly distributed on $[A, B]$:

$$\begin{aligned}
e_1 = b\,, \quad &\text{so that} \quad \phi_z^1(z) = 0\,,\ z \in [0, \frac{1}{2}] = \cup_{i=1}^7 \mathcal{I}_i\,, \\
\epsilon = \frac{1-a}{2}\,, \quad &\text{so that} \quad \phi_z^1(z) = 0\,,\ z \in \mathcal{I}_8\,.
\end{aligned}$$

Moreover, $\phi_z^1(z) = 0$, $z \in \mathcal{I}_{11} \cup \mathcal{I}_{12}$ since $y_1 \leq \frac{1}{2}$ by definition. The density induced on $\mathcal{I}_9$ and $\mathcal{I}_{10}$ is thus

$$\phi_z^1(z) = \frac{1}{\alpha_{10} - \alpha_8} = 8 - 4a \simeq 7.22353\,, \quad z \in \mathcal{I}_9 \cup \mathcal{I}_{10}\,.$$

Similarly, one obtains for $\epsilon = 0$

$$\phi_z^1(z) = 4\,, \quad z \in \left[\frac{1}{2}, \frac{3}{4}\right] = \mathcal{I}_8 \cup \mathcal{I}_9 \cup \mathcal{I}_{10}\,.$$

The density of z_n then remains constant on each interval $\mathcal{I}_i$. Let $\pi_i^{(n)}$ denote the probability $Pr(z_n \in \mathcal{I}_i)$. The interval $\mathcal{I}_i, i = 1, \ldots, 12$, can be interpreted as the ith state S_i of a finite Markov chain, with $\pi_i^{(n)}$ the probability to be in state S_i at iteration n. The transition probability π_{ij} from state S_i to state S_j is simply given by:

$$\pi_{ij} = \begin{cases} \frac{|\mathcal{I}_j|}{\sum_{k \in \Omega_i} |\mathcal{I}_k|} & \text{if } j \in \Omega_i\,, \\ 0 & \text{otherwise}\ , \end{cases} \tag{13}$$

where $|\mathcal{I}_k|$ is the length of the interval $\mathcal{I}_k$ and

$$\Omega_i = \{j | \mathcal{I}_j \subseteq T(\mathcal{I}_i)\}\,. \tag{14}$$

Let $\boldsymbol{P}$ denote the transition matrix with elements $\pi_{ij}, i, j = 1, \ldots, 12$, it is given by

$$\begin{aligned}
\pi_{1,12} &= 1\,,\ \ \pi_{2,10} = -\frac{2}{5}a^3 + \frac{2}{5}a^2 + \frac{2}{5}\,,\ \ \pi_{2,11} = \frac{2}{5}a^3 - \frac{2}{5}a^2 + \frac{3}{5}\,,\ \ \pi_{3,9} = 1\,,\\
\pi_{4,2} &= 1\,,\ \ \pi_{5,3} = a^3 - 3a^2 + 2a\,,\ \ \pi_{5,4} = -a^3 + 3a^2 - 2a + 1\,,\ \ \pi_{6,5} = 1\,,\\
\pi_{7,6} &= a\,,\ \ \pi_{7,7} = 1 - a\,,\ \ \pi_{8,7} = 1\,,\ \ \pi_{9,5} = -2a^3 + 4a^2 - 3a + 1\,,\\
\pi_{9,6} &= 2a^3 - 4a^2 + 3a\,,\ \ \pi_{10,2} = -a^3 + 2a^2 - \frac{a}{2} + \frac{1}{2}\,,\ \ \pi_{10,3} = a^2 - 2a + \frac{1}{2}\,,\\
\pi_{10,4} &= a^3 - 3a^2 + \frac{5}{2}a\,,\ \ \pi_{11,2} = -2a^3 + 4a^2 + a\,,\ \ \pi_{11,3} = 2a^2 - 5a + 1\,,\\
\pi_{11,4} &= 2a^3 - 8a^2 + 8a - 1\,,\ \ \pi_{11,5} = 2a^2 - 4a + 1\,,\ \ \pi_{12,6} = a\,,\ \ \pi_{12,7} = 1 - a\,,
\end{aligned}$$

the other elements being equal to 0. This gives

$$\boldsymbol{P} \simeq \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.412 & 0.588 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0.283 & 0.717 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0.194 & 0.806 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0.554 & 0.446 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0.471 & 0.149 & 0.380 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0.330 & 0.105 & 0.266 & 0.299 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0.194 & 0.806 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}. \tag{15}$$

The initial distribution, given by $\pi_i^{(1)}, i = 1, \ldots, 12$, is that induced on the states by the prior distribution of x^*, and depends on the choice of ϵ. For $\epsilon = \frac{1-a}{2}$, it is

$$\begin{aligned}
\boldsymbol{\pi}^{(1)} &= (0, 0, 0, 0, 0, 0, 0, 0, -2a^3 + 6a^2 - 4a + 1, 2a^3 - 6a^2 + 4a, 0, 0)^T\\
&\simeq (0, 0, 0, 0, 0, 0, 0, 0, 0.4350, 0.5650, 0, 0)^T\,,
\end{aligned}$$

while for $\epsilon = 0$ it is

$$\begin{aligned}
\boldsymbol{\pi}^{(1)} &= (0, 0, 0, 0, 0, 0, 0, 2a^3 - 4a^2 + 3a, -2a^3 + 6a^2 - 5a + 1, -2a^2 + 2a, 0, 0)^T\\
&\simeq (0, 0, 0, 0, 0, 0, 0, 0.4462, 0.2409, 0.3129, 0, 0)^T\,.
\end{aligned}$$

This Markov-chain representation will be used in Sections 5.2 and 5.3 in the ergodic analysis, and in Section 4.2 to obtain closed-form expressions for certain finite sample performance characteristics of the algorithm.

4 Finite sample behaviour of the GS4 algorithm

When the function $f(.)$ is symmetric with respect to x^*, the behaviour of the algorithm only depends on the location of x^* in $[A, B]$. When $f(.)$ is not symmetric around x^*, the behaviour also depends on the shape of $f(.)$. However, for any fixed $f(.)$, we can write $f(x) = \tilde{f}(x - x^*)$, and consider x^* as a location parameter. The behaviour of the algorithm for a fixed $\tilde{f}(.)$ then only depends on x^*. Various performance characteristics are introduced in Section 4.1. Section 4.2 is devoted to functions symmetric around x^*, whereas in Section 4.3 robustness with respect to non-symmetry is considered through the construction of worst-case functions $\tilde{f}(.)$.

4.1 Performance characteristics

We shall consider the following performance criteria, which are functions of ϵ and N:

$$\begin{aligned} lEL_N &= \log EL_N = \log E_{x^*}\{L_N(x^*)\}\,, \\ lL_N^{1-\alpha} &= \log L_N^{1-\alpha} = \log \inf\{t | Pr(L_N(x^*) \geq t) < \alpha\}\,, \\ P_N^{GS} &= Pr\{L_N(x^*) < L_0 \lambda^{N-1}\}\,, \;\; P_N^F = Pr\{L_N(x^*) < \frac{L_0}{F_{N+1}}\}\,, \end{aligned}$$

where x^* is assumed uniformly distributed on $[A, B]$. The criteria P_N^{GS} and P_N^F respectively correspond to the probability that the GS4 algorithm has a faster convergence (in terms of $L_N(x^*)$) than the GS and the Fibonacci algorithms. We shall also consider the worst-case performance criterion

$$lML_N = \log ML_N = \log \max_{x^* \in [A,B]} \{L_N(x^*)\}\,,$$

and the following variants of lEL_N:

$$\begin{aligned} lEL_N^{\gamma} &= \log E_{x^*}\{L_n^{\gamma}(x^*)\}\,, \;\; \gamma > 0\,, \\ ElL_N &= E_{x^*}\{\log L_N(x^*)\}\,. \end{aligned} \tag{16}$$

Note that EL_N and ML_N are more classical criteria than ElL_N for characterizing the precision of the localisation of x^*.

From (12), a reduction rate R_i is associated with each state S_i. Now, R_i is the inverse of the modulus of the slope of the transformation $T(.)$ on the interval $\mathcal{I}_i$,

$$R_i = \frac{|\mathcal{I}_i|}{\sum_{j \in \Omega_i} |\mathcal{I}_j|}\,, \tag{17}$$

with Ω_i defined by (14). The vector R of rates associated with the twelve states is thus

$$R = (a', a', a', 1-a, 1-a, 1-a, 1-a, c, c, c, c, c)^T\,. \tag{18}$$

4.2 Finite sample behaviour of GS4 when $f(.)$ is symmetric

From the values of the rates R_i (18), we can easily compute the analytical expressions of ElL_N and lEL_N^γ for any $\gamma > 0$.

Theorem 1 *For the GS4 algorithm with initial distribution $\boldsymbol{\pi}^{(1)} = (\pi_1^{(1)}, \ldots, \pi_{12}^{(1)})^T$ for the states $S_i, i = 1, \ldots, 12$, we have when $f(.)$ is symmetric with respect to x^*:*

$$\forall \gamma > 0\,, \;\; lEL_N^\gamma = \gamma \log L_0 + \log \boldsymbol{p}_\gamma^T \boldsymbol{R}_\gamma^{N-2} \boldsymbol{q}_\gamma\,, \tag{19}$$

with

$$(\boldsymbol{p}_\gamma)_i = \pi_i^{(1)} |\mathcal{I}_i|^\gamma\,, \;\; (\boldsymbol{q}_\gamma)_i = \frac{R_i^\gamma}{|\mathcal{I}_i|^\gamma}\,, \;\; (\boldsymbol{R}_\gamma)_{ij} = (\pi_{ij})^{1+\gamma}\,, \tag{20}$$

and

$$ElL_N = \log L_0 + {\boldsymbol{\pi}^{(1)}}^T \boldsymbol{Q}_{N-2} \boldsymbol{l}\,, \tag{21}$$

with

$$\boldsymbol{Q}_{N-2} = \sum_{k=0}^{N-2} \boldsymbol{P}^k\,, \;\; \boldsymbol{l} = (\log R_1, \ldots, \log R_{12})^T\,. \tag{22}$$

Proof.

First note that $L_N(x^*) = L_0 R_{i_1} R_{i_2} \ldots R_{i_{N-1}}$, where $i_1, \ldots, i_{N-1}$ denotes the sequence of states visited by $z_1, \ldots, z_{N-1}$, which is function of x^*. The probability of visiting this particular sequence of states is

$$Pr(z_1 \in \mathcal{I}_{i_1}, z_2 \in \mathcal{I}_{i_2}, \ldots, z_{N-1} \in \mathcal{I}_{i_{N-1}}) = \pi_{i_1}^{(1)} \pi_{i_1 i_2} \ldots \pi_{i_{N-2} i_{N-1}}\,.$$

We thus have

$$lEL_N^\gamma = \log \left(L_0^\gamma \sum_{i_1, \ldots, i_{N-1}} \pi_{i_1}^{(1)} \pi_{i_1 i_2} \ldots \pi_{i_{N-2} i_{N-1}} (R_{i_1} R_{i_2} \ldots R_{i_{N-1}})^\gamma \right)\,. \tag{23}$$

From (13) and (17), we have

$$R_{i_k} = \frac{\pi_{i_k i_{k+1}} |\mathcal{I}_{i_k}|}{|\mathcal{I}_{i_{k+1}}|}\,, \forall i_{k+1} \in \Omega_{i_k}\,, \tag{24}$$

and substituting back in (23) from $k = 1$ to $k = N - 2$ we obtain

$$lEL_N^\gamma = \log \left(L_0^\gamma \sum_{i_1, \ldots, i_{N-1}} \pi_{i_1}^{(1)} |\mathcal{I}_{i_1}|^\gamma (\pi_{i_1 i_2} \ldots \pi_{i_{N-2} i_{N-1}})^{1+\gamma} \frac{R_{i_{N-1}}^\gamma}{|\mathcal{I}_{i_{N-1}}|^\gamma} \right)\,,$$

which can be written as (19).

Using similar arguments, $ElL_N = E_{x^*}\{\log L_N(x^*)\}$ can be written as

$$\begin{aligned} ElL_N &= \sum_{i_1,\ldots,i_{N-1}} \pi^{(1)}_{i_1}\pi_{i_1 i_2}\ldots\pi_{i_{N-2}i_{N-1}}\log(L_0 R_{i_1}R_{i_2}\ldots R_{i_{N-1}}) \qquad (25)\\ &= \log L_0 + \sum_{i_1}\pi^{(1)}_{i_1}\log R_{i_1} + \sum_{k=2}^{N-1}\sum_{i_1,\ldots,i_k}\pi^{(1)}_{i_1}\pi_{i_1 i_2}\ldots\pi_{i_{k-1}i_k}\log R_{i_k}\\ &= \log L_0 + {\boldsymbol{\pi}^{(1)}}^T\boldsymbol{l} + \sum_{k=2}^{N-1}{\boldsymbol{\pi}^{(1)}}^T\boldsymbol{P}^{k-1}\boldsymbol{l}\,, \end{aligned}$$

which gives the result (21).

■

The expression for lML_N can be computed from the determination of the worst cases. The two choices $\epsilon = \frac{1-a}{2}$ and $\epsilon = 0$ have to be considered separately.

Case a: $\epsilon = \frac{1-a}{2}$ The initial probabilities $\pi^{(1)}_1, \pi^{(1)}_7, \pi^{(1)}_8$ and $\pi^{(1)}_{12}$ are equal to zero, so that, from the structure of the Markov chain,

$$\pi^{(k)}_1 = \pi^{(k)}_7 = \pi^{(k)}_8 = \pi^{(k)}_{12} = 0\,,\ \ \forall k \geq 1\,.$$

We can thus consider a reduced Markov chain with eight states,

$$S'_1 = S_2\,,\ S'_2 = S_3\,,\ S'_3 = S_4\,,\ S'_4 = S_5\,,\ S'_5 = S_6\,,\ S'_6 = S_9\,,\ S'_7 = S_{10}\,,\ S'_8 = S_{11}\,, \quad (26)$$

and the associated eight intervals $\mathcal{I}'_i$. Its transition probability matrix $\boldsymbol{P}'$ is given by the corresponding submatrix of $\boldsymbol{P}$. We also define the adjacency matrix for the states $S'_i, i = 1,\ldots,8$ as

$$\tilde{\boldsymbol{P}}'_{ij} = \tilde{\pi}'_{ij} = \begin{cases} 1 & \text{if } \boldsymbol{P}'_{ij} > 0\,,\\ 0 & \text{otherwise}\,, \end{cases} \qquad (27)$$

that is $\tilde{\boldsymbol{P}}'_{ij} = 1$ if and only if state S'_j is reachable in one step from state S'_i. The corresponding transition graph is presented on Figure 3.

POSSIBLE LOCATION OF FIGURE 3

Theorem 2 *For the GS4 algorithm with $\epsilon = \frac{1-a}{2}$ we have when $f(.)$ is symmetric with respect to x^*:*

$$\forall N \geq 3\,,\ \ lML_N = (2m+1)\log(1-a) + m\log a' + m\log c + l_k\,,$$

where $k = (N-3)[mod\ 4]$, $m = \frac{N-3-k}{4}$ and

$$l_k = \begin{cases} 0 & \text{if } k = 0\,,\\ \log(1-a) & \text{if } k = 1\,,\\ 2\log(1-a) & \text{if } k = 2\,,\\ 2\log(1-a) + \log a' & \text{if } k = 3\,. \end{cases}$$

Proof.

One can easily check that there are just two cycles which give the same asymptotic worst rate, namely:

$$S'_4 \longrightarrow S'_3 \longrightarrow S'_1 \longrightarrow S'_8 \longrightarrow S'_4 \longrightarrow \dots$$
$$S'_4 \longrightarrow S'_2 \longrightarrow S'_6 \longrightarrow S'_5 \longrightarrow S'_4 \longrightarrow \dots$$

with rate $a'c(1-a)^2$ for 4 iterations.

The initial distribution is concentrated on $\mathcal{I}'_6$ and $\mathcal{I}'_7$, i.e. the initial state is S'_6 or S'_7. Starting at S'_6, where the rate is c, we go in one iteration to S'_4 or S'_5, where the rate is $1-a$. In the next iteration we then respectively enter one of the worst cycles described above, or go first from S'_5 to S'_4 and then enter one of the worst cycles. The later case gives the worst case.

Starting at S'_7, where the rate is c, we first go either to S'_1 or S'_2, which belong to the cycles above and have rate a'.

The worst path thus starts at S'_6:

$$S'_6 \longrightarrow S'_5 \longrightarrow S'_4 \longrightarrow \dots S'_4 \longrightarrow S'_3 \longrightarrow S'_1 \longrightarrow S'_8\,,$$

and stops at S'_4, S'_3, S'_7 or S'_8 depending on the value of N. Taking into account that the rate r_1 corresponding to the initial expansion of the interval is $\frac{1}{c}$, we have the stated result.

■

Corollary 1 *For the GS4 algorithm with $\epsilon = \frac{1-a}{2}$ and for functions symmetric with respect to x^*, we have:*

$$\forall N \geq 38\,,\quad \forall x^* \in [A,B]\,,\quad L_N(x^*) < L_0\lambda^{N-1}$$
$$\forall N \geq 42\,,\quad \forall x^* \in [A,B]\,,\quad L_N(x^*) < \frac{L_0}{F_{N+1}}$$

that is the GS4 algorithm performs better than the GS (respectively Fibonacci) algorithm for any $N \geq 38$ (respectively 42) and any x^* in $[A,B]$.

There are only two values for x^* that make the algorithm stay in the worst-case path forever. They can be determined as follows. One can check that the point

$$\hat{z} = -\frac{2}{31}a^3 - \frac{3}{31}a^2 + \frac{19}{31}a + \frac{3}{31} \simeq 0.21163$$

is the only point in $\mathcal{I}'_4$ such that $T^4(\hat{z}) = \hat{z}$ and $T(\hat{z}) \in \mathcal{I}'_3$, $T^2(\hat{z}) \in \mathcal{I}'_1$ and $T^3(\hat{z}) \in \mathcal{I}'_8$. With $\hat{z}$ is associated a unique point z_1 in $\mathcal{I}'_6$ such that $T(z_1) \in \mathcal{I}'_5$ and $T^2(z_1) = \hat{z}$. It is given by

$$z_1 = \frac{7}{31}a^3 - \frac{5}{31}a^2 - \frac{9}{62}a + \frac{41}{62} \simeq 0.62869\,.$$

To z_1 correspond two possible values for x^*, given by $x^* = x_1L_1 + A_1$, with $L_1 = (2-a)(B-A), A_1 = A - \frac{1-a}{2}(B-A)$ and $x_1 = 2z_1 - 1$ or $x_1 = 2 - 2z_1$, that is

$$x^* \simeq A + 0.061843(B-A)\,,\ \text{or}\ x^* \simeq A + 0.938157(B-A)\,.$$

Case b: $\epsilon = 0$ When $\epsilon = 0$, we need to consider all states $S_i, i = 1, \ldots, 12$. The worst cycle, corresponding to $x^* = A$ or $x^* = B$, is then

$$S_7 \longrightarrow S_7 \ldots$$

with rate $1 - a$ for 1 iteration. We then obtain the following property.

Theorem 3 *For the GS4 algorithm with $\epsilon = 0$ and for functions symmetric with respect to x^*, we have:*

$$\forall N \geq 2\,, \;\; lML_N = \log c + (N-2)\log(1-a)\,.$$

Proof.

The algorithm is initialised in S_8 or S_9 or S_{10}. The worst path is

$$S_8 \longrightarrow S_7 \longrightarrow \ldots \longrightarrow S_7$$

which gives the result.

■

The criteria $lL_N^{1-\alpha}, P_N^{GS}$ and P_N^F are difficult to compute analytically, but can be evaluated with any arbitrary precision for any reasonable N. We simply need to compute the value of L_N and the probability $\pi_{i_1}^{(1)} \pi_{i_1 i_2} \ldots \pi_{i_{N-2} i_{N-1}}$ associated with any sequence of states $S_{i_1}, S_{i_2}, \ldots, S_{i_{N-1}}$. Note that from Corollary 1, $P_N^{GS} = 1, N \geq 38$ and $P_N^F = 1, N \geq 42$ when $\epsilon = \frac{1-a}{2}$.

Table 3 presents the numerical performances of the GS4 algorithm with $\epsilon = \frac{1-a}{2}$ with respect to all criteria above for $1 \leq N \leq 30$. Note the expansion of the initial interval, that is $L_1 > L_0$.

Tables 4 presents the performances achieved when $\epsilon = 0$. The comparison with Table 3 stresses the importance of expanding the initial interval for the finite sample behaviour. The fact that this expansion is also important asymptotically will be demonstrated in Section 5.3.

Table 5 presents the value of N required for the corresponding characteristic to reach the precision indicated. For instance, the Fibonacci algorithm requires 30 function evaluations to reduce the length of the initial interval by a factor 10^6, while GS4 with $\epsilon = \frac{1-a}{2}$ requires respectively 25 and 28 evaluations to achieve the same precision, on the average and with probability 0.99.

Figure 4 gives the evolution of some performance characteristics as function of N for GS4.

POSSIBLE LOCATION OF FIGURE 4

4.3 Robustness with respect to non symmetry

We consider the case where $f(.)$ is only locally symmetric around x^*, that is

$$C|x-x^*|^\beta - D|x-x^*|^{\beta+\omega} \leq |f(x)-f(x^*)| \leq C|x-x^*|^\beta + D|x-x^*|^{\beta+\omega}\,, \qquad (28)$$

with $\beta > 0, \omega > 0, C > 0, D \geq 0$. Since the algorithm only uses function comparisons, $f(.)$ can be scaled down so that we can assume that $C = 1$.

Figure 5 presents the graph of the functions $f_{0.5}(x-x^*)$ and $f_2(x-x^*)$, with

$$f_D(z) = \begin{cases} \frac{4}{27D^2} & \text{if } z \leq -\frac{2}{3D} \\ z^2 + Dz^3 & \text{if } -\frac{2}{3D} < z\,, \end{cases}$$

which can be considered as the worst uniextremal function in the class above with respect to the symmetry condition, for a given value D, when $\beta = 2, \omega = 1$. The functions $f_D(.)$ are constant for $x \leq \bar{x} = -\frac{2}{3D}$, however, since we delete $[A_n, U_n]$ when $f(U_n) \geq f(V_n)$ this has no effect on the behaviour of the algorithm.

POSSIBLE LOCATION OF FIGURE 5

Numerical simulations with x^* uniformly distributed in $[A, B]$ were carried out, with the non-symmetry parameter D varying from 0 to 10. Figure 6 presents the evolution of the empirical values of the performance characteristics $EL_{30} = E_{x^*}\{L_{30}(x^*)\}$ and $ML_{30} = \max_{x^*}\{L_{30}(x^*)\}$ as functions of D for the GS4 algorithm with $\epsilon = \frac{1-a}{2}$. Note that EL_{30} remains much smaller than $\frac{1}{F_{31}}$ even for functions quite non-symmetric with respect to x^*. Note in particular from Figure 5 that $f_D(.)$ is already very non-symmetric when $D = 2$. The value of ML_{30} jumps from $ML_{30}^{(1)} \simeq 7.366\ 10^{-7}$ to $ML_{30}^{(2)} = \frac{1-a}{a'} ML_{30}^{(1)} \simeq 1.693\ 10^{-6}$ at $D \simeq 1.35$. This jump corresponds to replacing one transition with rate a' by another one with rate $1-a$ in the worst path through the graph presented in Figure 3.

Figure 7 presents the evolution of the empirical probability P_N^F as a function of D and N. Again, the performances of the GS4 algorithm remain fairly stable while non symmetry increases.

POSSIBLE LOCATION OF FIGURE 6

POSSIBLE LOCATION OF FIGURE 7

Throughout the rest of the paper the function $f(.)$ will be assumed symmetric with respect to x^*.

5 Asymptotic behaviour of GS4

5.1 Invariant measure

The mapping $T(.)$ (11) is expanding, see Figure 2, which implies the existence of a T-invariant measure $\mu_z(.)$, absolutely continuous with respect to the Lebesgue measure

(see [1], p. 210). Its density, denoted in what follows by $\phi_z(.)$, can be normalised and considered as a probability density, that is $\int_0^1 \phi_z(z)dz = 1$. It satisfies the Perron-Frobenius equation:

$$\phi_z(z) = \sum_{y_j|T(y_j)=z} \frac{1}{|T'(y_j)|}\phi_z(y_j)\,. \tag{29}$$

Since $T(.)$ is piecewise linear, and in view of the definition of $\mathcal{S}^\infty$, $\phi_z(.)$ is piecewise constant on the intervals $\mathcal{I}_i$: $\phi_z(z) = \bar{\phi}_i, z \in \mathcal{I}_i, i = 1, \ldots, 12$. Equation (29) then reduces to a linear equation $\bar{\boldsymbol{\phi}} = \boldsymbol{M}\bar{\boldsymbol{\phi}}$, with $\bar{\boldsymbol{\phi}} = (\bar{\phi}_1, \ldots, \bar{\phi}_{12})^T$, and

$$M_{ij} = \pi_{ji}\frac{|\mathcal{I}_j|}{|\mathcal{I}_i|}\,,$$

which is the reduction rate from state S_j to state S_i, see (24). The equation $\bar{\boldsymbol{\phi}} = \boldsymbol{M}\bar{\boldsymbol{\phi}}$ thus gives

$$\bar{\phi}_i|\mathcal{I}_i| = \sum_{j=1}^{12} \pi_{ji}\bar{\phi}_j|\mathcal{I}_j|\,,$$

or equivalently

$$\bar{\boldsymbol{\pi}} = \boldsymbol{P}^T\bar{\boldsymbol{\pi}}\,,$$

with $\bar{\pi}_i = \bar{\phi}_i|\mathcal{I}_i|$. This means that $\bar{\boldsymbol{\pi}}$ is the invariant distribution for the Markov chain with matrix $\boldsymbol{P}$, given by the normalised eigenvector associated with the eigenvalue 1 for the matrix $\boldsymbol{P}^T$. Ergodicity of $T(.)$ follows from the existence of a single absorbing class for the Markov chain consisting of states $\{2, 3, 4, 5, 6, 9, 10, 11\}$. This distribution $\bar{\boldsymbol{\pi}}$ is given in Table 6.

5.2 Asymptotic performance characteristics

We shall compute now the asymptotic values of the performance characteristics of Section 4.1. In particular, we shall see that $\lim_{N\to\infty} \frac{1}{N} ElL_N$ is not sensitive to the choice of ϵ in the GS4 algorithm, whereas $\lim_{N\to\infty} \frac{1}{N} lEL_N$ and $\lim_{N\to\infty} \frac{1}{N} lML_N$ are. This reinforces our view that ElL_N is not a suitable criterion for evaluating the performances of the algorithm.

5.2.1 Lyapunov exponent and ergodic log-rate

The Lyapunov exponent of the dynamic system (11) is defined by

$$\Lambda = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \log|T'(z_k)|\,, \tag{30}$$

if this limit exists and is the same for almost all z_1. Birkhoff's ergodic Theorem, see [2] p. 44, implies that Λ exists and is given by

$$\Lambda = -\sum_{i=1}^{12} \bar{\pi}_i \log R_i \simeq 0.63006\,, \tag{31}$$

where $\bar{\pi}_i$ and R_i respectively correspond to the invariant probability and reduction rate for state S_i, given by Table 6 and (18). Since $|T'(z_k)| = \frac{1}{R_i}$ for $z_k \in \mathcal{I}_i$, the expression for Λ can also be written as

$$\Lambda = -\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log r_k = -\frac{1}{n} \lim_{n\to\infty} \log L_n(x^*)\,.$$

This last limit is called the ergodic log-rate in [7], and it is the same for almost all x^* and any fixed $\epsilon \geq 0$. This is a consequence of the piecewise linearity of the mapping $T(.)$, see [7]. The next theorem relates Λ and the quantity

$$H = -\lim_{N\to\infty} \frac{1}{N} E_{x^*}\{\log L_N(x^*)\}\,. \tag{32}$$

Theorem 4 *Assume that x^* has a prior density $p(x^*)$ on $[A, B]$ absolutely continuous with respect to the Lebesgue measure. Then for the GS4 algorithm $\Lambda = H \simeq 0.63006$.*

Proof.

From the definition of the algorithm:

$$L_0(a')^{N-1} \leq L_N(x^*) \leq L_0(1-a)^{N-1}\,,\ \ \forall N \geq 2\,,\ \ \forall x^* \in [A, B]\,,$$

where a and a' are defined in Section 2. This gives

$$-\frac{\log L_0}{N-1} - \log(1-a) \leq -\frac{\log L_N(x^*)}{N-1} \leq -\frac{\log L_0}{N-1} - \log a'\,.$$

Lebesgue's Theorem on dominated convergence then implies

$$\begin{aligned} -\lim_{N\to\infty} \frac{1}{N} E_{x^*}\{\log L_N(x^*)\} &= -\lim_{N\to\infty} \int_A^B \frac{\log L_N(x^*)}{N} p(x^*)dx^* \\ &= \int_A^B -\lim_{N\to\infty} \left(\frac{\log L_N(x^*)}{N}\right) p(x^*)dx^* = \Lambda\,. \end{aligned}$$

■

In the case where x^* has a prior distribution uniform on $[A, B]$, the equality $H = \Lambda$ could also have been obtained through the expression given in Theorem 1 for ElL_N. Indeed,

$$\frac{ElL_N}{N} = \frac{\log L_0}{N} + \frac{(\boldsymbol{\pi}^{(1)})^T \boldsymbol{Q}_{N-2} \boldsymbol{l}}{N}\,,$$

with $\boldsymbol{Q}_{N-2}$ and $\boldsymbol{l}$ given by (22). Now, $(\boldsymbol{\pi}^{(1)})^T \boldsymbol{P}^k \boldsymbol{l}$ tends to $\bar{\boldsymbol{\pi}}^T \boldsymbol{l}$ when k tends to infinity, so that

$$H = -\lim_{N\to\infty} \frac{ElL_N}{N} = -\bar{\boldsymbol{\pi}}^T \boldsymbol{l} = \Lambda\,.$$

The next result shows that the quantiles $lL_N^{1-\alpha}$ behave asymptotically as ElL_N.

Theorem 5 *Assume that x^* has a prior density $p(x^*)$ on $[A, B]$ absolutely continuous with respect to the Lebesgue measure. Then for any fixed α, $0 < \alpha < 1$, one has*

$$\lim_{N\to\infty} \frac{1}{N} l L_N^{1-\alpha} = -H \,, \tag{33}$$

where H is defined by (32).

Proof.

The asymptotic relation (30) implies that $\forall \delta, \gamma > 0, \exists N_0(\delta, \gamma)$ such that

$$\forall N \geq N_0(\delta, \gamma) \,, \quad Pr(|\frac{1}{N} \log L_N(x^*) + \Lambda| \geq \delta) < \gamma \,. \tag{34}$$

Let t' be any number larger than $-\Lambda$. We can show that $\frac{1}{N} l L_N^{1-\alpha} \leq t'$ for N large enough. Indeed, take $\gamma = \alpha$ and $\delta = t' + \Lambda > 0$. Then (34) implies

$$\forall N \geq N_0(\delta, \gamma) \,, \quad Pr(\frac{1}{N} \log L_N(x^*) \geq t') < \alpha \,,$$

and thus

$$\forall N \geq N_0(\delta, \gamma) \,, \quad \frac{1}{N} l L_N^{1-\alpha} = \inf\{t | Pr(\frac{1}{N} \log L_N(x^*) \geq t) < \alpha\} \leq t' \,.$$

Similar arguments show that $\forall t' < -\Lambda$, $\frac{1}{N} l L_N^{1-\alpha} \geq t'$ for N large enough. We thus obtain for any $\delta > 0$:

$$\limsup_{N\to\infty} \frac{1}{N} l L_N^{1-\alpha} \leq -\Lambda + \delta \,,$$

and

$$\liminf_{N\to\infty} \frac{1}{N} l L_N^{1-\alpha} \geq -\Lambda - \delta \,.$$

This implies

$$\lim_{N\to\infty} \frac{1}{N} l L_N^{1-\alpha} = -\Lambda \,,$$

which together with Theorem 4 gives the result (33).

■

Corollary 2 *Assume that x^* has a prior density $p(x^*)$ on $[A, B]$ absolutely continuous with respect to the Lebesgue measure. Then*

$$\lim_{N\to\infty} P_N^{GS} = \lim_{N\to\infty} P_N^F = 1 \,.$$

Proof.

We showed in the proof of Theorem 5 that $\forall \delta, \gamma > 0, \exists N_0(\delta, \gamma)$ such that

$$\forall N > N_0(\delta, \gamma) \,, \quad Pr(\frac{1}{N} \log L_N(x^*) \geq -\Lambda + \delta) < \gamma \,. \tag{35}$$

Therefore, taking $\delta = -\Lambda + \log\lambda > 0.148$ one obtains

$$\begin{aligned} \forall N > N_0(\delta,\gamma)\,,\ \forall\gamma>0\,,\ 1-P_N^{GS} &= Pr(\frac{1}{N}\log L_N(x^*) > \frac{N-1}{N}\log\lambda) \\ &< Pr(\frac{1}{N}\log L_N(x^*) > \log\lambda) < \gamma\,, \end{aligned}$$

so that $\lim_{N\to\infty} P_N^{GS} = 1$.

Now, since $\left(\frac{1}{F_{N+1}}\right)^{1/N}$ converges to λ [5], we have

$$\forall\omega\,,\ 0<\omega<\lambda-\exp(\Lambda)\,,\ \exists N_1 \text{ such that } \forall N\geq N_1\,,\ \left(\frac{L_0}{F_{N+1}}\right)^{1/N} > \lambda-\omega\,.$$

One then obtains for $N \geq N_1$:

$$\begin{aligned} 1-P_N^F &= Pr(\frac{1}{N}\log L_N(x^*) > \frac{1}{N}\log\left(\frac{L_0}{F_{N+1}}\right)) \\ &\leq Pr(\frac{1}{N}\log L_N(x^*) > \log(\lambda-\omega))\,. \end{aligned}$$

Taking $\delta = -\Lambda + \log(\lambda-\omega) > 0$ in (35) then gives the result $\lim_{N\to\infty} P_N^F = 1$.

■

5.2.2 Asymptotic behaviour of the expected length

As we shall see below, the asymptotic expression $\lim_{N\to\infty}\frac{1}{N}lEL_N^\gamma$, with lEL_N^γ given by (16) depends on the choice of ϵ, whereas it was not the case for the criterion $\lim_{N\to\infty}\frac{1}{N}ElL_N$. For that reason, we consider the two cases $\epsilon = \frac{1-a}{2}$ and $\epsilon = 0$ separately.

Case a: $\epsilon = \frac{1-a}{2}$ We have the following property.

Theorem 6 *Assume that x^* has a uniform prior distribution on $[A,B]$ and that $\epsilon = \frac{1-a}{2}$. Then for any $\gamma > 0$, the algorithm GS4 is such that*

$$\lim_{N\to\infty}\frac{1}{N}lEL_N^\gamma = \log\lambda_{\max}(\boldsymbol{R}'_\gamma)\,, \tag{36}$$

where $\lambda_{\max}(\boldsymbol{M})$ denotes the maximal eigenvalue of the matrix $\boldsymbol{M}$ and where

$$(\boldsymbol{R}'_\gamma)_{ij} = (\boldsymbol{P}'_{ij})^{1+\gamma}\,, \tag{37}$$

with $\boldsymbol{P}'$ the submatrix of $\boldsymbol{P}$ (15) associated with the eight states S'_i (26). In particular for $\gamma = 1$

$$\lim_{N\to\infty}\frac{1}{N}\log E_{x^*}\{L_N(x^*)\} = \log\mu \simeq -0.61273\,, \tag{38}$$

where $\mu \simeq 0.54187$ is the largest positive root of the equation

$$\begin{aligned} 4t^6 - 8a^2t^4 + (-24a^3+54a^2-42a+6)t^3 + (-12a^3+18a^2-14a+2)t^2 \\ +(52a^3-102a^2+90a-14)t + 68a^3-125a^2+99a-15 = 0\,. \end{aligned}$$

Proof.

From Theorem 1, we have

$$lEL_N^\gamma = \gamma \log L_0 + \log(\boldsymbol{p}_\gamma^T \boldsymbol{R}_\gamma^{N-2} \boldsymbol{q}_\gamma)\,,$$

with $\boldsymbol{p}_\gamma, \boldsymbol{q}_\gamma$ and $\boldsymbol{R}_\gamma$ given by (20). Since $\epsilon = \frac{1-a}{2}$, $(\boldsymbol{p}_\gamma)_1 = (\boldsymbol{p}_\gamma)_7 = (\boldsymbol{p}_\gamma)_8 = (\boldsymbol{p}_\gamma)_{12} = 0$, and from the structure of the matrix $\boldsymbol{P}$ we have

$$lEL_N^\gamma = \gamma \log L_0 + \log((\boldsymbol{p}'_\gamma)^T (\boldsymbol{R}'_\gamma)^{N-2} \boldsymbol{q}'_\gamma)\,,$$

where

$$\begin{aligned} \boldsymbol{p}'_\gamma &= \Big((\boldsymbol{p}_\gamma)_2, (\boldsymbol{p}_\gamma)_3, (\boldsymbol{p}_\gamma)_4, (\boldsymbol{p}_\gamma)_5, (\boldsymbol{p}_\gamma)_6, (\boldsymbol{p}_\gamma)_9, (\boldsymbol{p}_\gamma)_{10}, (\boldsymbol{p}_\gamma)_{11}\Big)^T\,, \\ \boldsymbol{q}'_\gamma &= \Big((\boldsymbol{q}_\gamma)_2, (\boldsymbol{q}_\gamma)_3, (\boldsymbol{q}_\gamma)_4, (\boldsymbol{q}_\gamma)_5, (\boldsymbol{q}_\gamma)_6, (\boldsymbol{q}_\gamma)_9, (\boldsymbol{q}_\gamma)_{10}, (\boldsymbol{q}_\gamma)_{11}\Big)^T\,, \end{aligned}$$

and $\boldsymbol{R}'_\gamma$ is given by (37). The matrix $\boldsymbol{R}'_\gamma$ satisfies the strong mixing condition, that is

$$((\boldsymbol{R}'_\gamma)^n)_{ij} > 0\,, \quad \forall i, j = 1, \ldots, 8\,,$$

for any $n \geq 8$. Perron Frobenius Theorem, see [2], Theorem 7.25 p.205, gives

$$\lim_{N\to\infty} \lambda_{\max}^{-(N-2)} (\boldsymbol{p}'_\gamma)^T (\boldsymbol{R}'_\gamma)^{N-2} \boldsymbol{q}'_\gamma = ((\boldsymbol{p}'_\gamma)^T \boldsymbol{u}_\gamma)((\boldsymbol{q}'_\gamma)^T \boldsymbol{v}_\gamma)\,,$$

where $\lambda_{\max}$ is the (simple) maximum eigenvalue of $\boldsymbol{R}'_\gamma$ and $\boldsymbol{u}_\gamma$ is the associated eigenvector, $\boldsymbol{v}_\gamma$ is the eigenvector of $(\boldsymbol{R}'_\gamma)^T$ associated with the same eigenvalue $\lambda_{\max}$. Moreover, $\boldsymbol{u}_\gamma$ and $\boldsymbol{v}_\gamma$ have strictly positive elements and $\boldsymbol{u}_\gamma^T \boldsymbol{v}_\gamma = 1$. This implies

$$\lim_{N\to\infty} \frac{1}{N} \log\Big(\lambda_{\max}^{-(N-2)} (\boldsymbol{p}'_\gamma)^T (\boldsymbol{R}'_\gamma)^{N-2} \boldsymbol{q}'_\gamma\Big) = 0\,,$$

so that

$$\lim_{N\to\infty} \frac{1}{N} \log\Big((\boldsymbol{p}'_\gamma)^T (\boldsymbol{R}'_\gamma)^{N-2} \boldsymbol{q}'_\gamma\Big) = \log \lambda_{\max}\,,$$

which gives (36).

In particular when $\gamma = 1$, elementary but tedious algebraic calculations give (38).

■

Case b: $\epsilon = 0$ In that case we can only obtain a bound on $\liminf_{N\to\infty} \frac{1}{N} \log E_{x^*}\{L_N(x^*)\}$.

Theorem 7 *Assume that x^* has a uniform distribution on $[A, B]$ and let $\epsilon = 0$. Then, for any $\gamma > 0$ and $N > 3$ the GS4 algorithm is such that*

$$\frac{lEL_N^\gamma}{N-2} \geq (1+\gamma)\log(1-a) + \frac{\gamma \log L_0}{N-2} + \frac{1}{N-2} \log \frac{(1-c)^{1+\gamma}}{(1-a)^2}\,, \tag{39}$$

and therefore

$$\liminf_{N\to\infty} \frac{1}{N} lEL_N^\gamma \geq (1+\gamma)\log(1-a)\,.$$

Proof.

From Theorem 1, we have

$$lEL_N^\gamma = \gamma \log L_0 + \log(\boldsymbol{p}_\gamma^T \boldsymbol{R}_\gamma^{N-2} \boldsymbol{q}_\gamma)\,,$$

with $\boldsymbol{p}_\gamma, \boldsymbol{q}_\gamma$ and $\boldsymbol{R}_\gamma$ given by (20). When $\epsilon = 0$, $(\boldsymbol{p}_\gamma)_1 = (\boldsymbol{p}_\gamma)_{12} = 0$, and due to the structure of the matrix $\boldsymbol{P}$ we have

$$lEL_N^\gamma = \gamma \log L_0 + \log((\boldsymbol{p}''_\gamma)^T (\boldsymbol{R}''_\gamma)^{N-2} \boldsymbol{q}''_\gamma)\,,$$

where

$$\boldsymbol{p}''_\gamma = \Big((\boldsymbol{p}_\gamma)_2, (\boldsymbol{p}_\gamma)_3, \ldots, (\boldsymbol{p}_\gamma)_{11}\Big)^T\,, \quad \boldsymbol{q}''_\gamma = \Big((\boldsymbol{q}_\gamma)_2, (\boldsymbol{q}_\gamma)_3, \ldots, (\boldsymbol{q}_\gamma)_{11}\Big)^T\,,$$

and $\boldsymbol{R}''_\gamma$ is the submatrix of $\boldsymbol{R}_\gamma$ corresponding to the states $S_2, \ldots, S_{11}$. We can reorder the states and rearrange S_7 and S_8 to the last positions, so as to obtain the following representation for $\boldsymbol{R}''_\gamma$:

$$\boldsymbol{R}''_\gamma = \begin{pmatrix} \boldsymbol{R}'_\gamma & 0 \\ \boldsymbol{G} & \boldsymbol{F} \end{pmatrix}\,,$$

with

$$\boldsymbol{F} = \begin{pmatrix} (1-a)^{1+\gamma} & 0 \\ 1 & 0 \end{pmatrix}\,, \quad \boldsymbol{G} = \begin{pmatrix} 0 & 0 & 0 & 0 & a^{1+\gamma} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}\,,$$

which corresponds to transition probabilities (to the power $1+\gamma$) from states S_7 and S_8. Then $(\boldsymbol{R}''_\gamma)^{N-2}$ takes the form

$$(\boldsymbol{R}''_\gamma)^{N-2} = \begin{pmatrix} (\boldsymbol{R}'_\gamma)^{N-2} & 0 \\ \boldsymbol{G}_{N-2} & \boldsymbol{F}^{N-2} \end{pmatrix}\,,$$

where $\boldsymbol{G}_{N-2}$ is a matrix with non-negative elements. With the same rearrangement of the states we define

$$\begin{aligned} (\boldsymbol{p}_\gamma)_{7,8} &= ((\boldsymbol{p}_\gamma)_7, (\boldsymbol{p}_\gamma)_8)^T = (0, \frac{(1-c)^{1+\gamma}}{4^\gamma})^T\,, \\ (\boldsymbol{q}_\gamma)_{7,8} &= ((\boldsymbol{q}_\gamma)_7, (\boldsymbol{q}_\gamma)_8)^T = (4^\gamma (1-a)^{\gamma-1}, 4^\gamma (2 - 5a + 6a^2 - 2a^3)^\gamma)^T\,. \end{aligned}$$

Then, since all components involved are non-negative,

$$lEL_N^\gamma \geq \gamma \log L_0 + \log((\boldsymbol{p}_\gamma)_{7,8}^T \boldsymbol{F}^{N-2} (\boldsymbol{q}_\gamma)_{7,8})\,.$$

The matrix $\boldsymbol{F}$ can be decomposed into

$$\boldsymbol{F} = \begin{pmatrix} 1 & (1-a)^{1+\gamma} \\ 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 0 \\ 0 & (1-a)^{1+\gamma} \end{pmatrix} \begin{pmatrix} 1 & (1-a)^{1+\gamma} \\ 1 & 0 \end{pmatrix}\,,$$

which gives (39) after elementary calculations.

■

When $\gamma = 1$ we thus have the following:

Corollary 3 *For the GS4 algorithm with $\epsilon = 0$, one has when x^* has a prior distribution uniform on $[A, B]$:*

$$\liminf_{N\to\infty} \frac{1}{N} \log E_{x^*}\{L_N(x^*)\} \geq 2\log(1-a) \simeq -0.43163\,.$$

Note that the bound $2\log(1-a)$ of Corollary 3 implies that the performances in terms of expected length of the uncertainty interval are significantly worse for the GS4 algorithm with $\epsilon = 0$ than for the GS algorithm, for which $\frac{lEL_N}{N} = \frac{N-1}{N}\log\lambda$, with $\log\lambda \simeq -0.48121$.

5.2.3 Asymptotic behaviour of worst-case performances

We consider now the asymptotic worst-case characteristic $\lim_{N\to\infty} \frac{1}{N} lML_N$. Again, the two cases $\epsilon = \frac{1-a}{2}$ and $\epsilon = 0$ must be treated separately.

Case a: $\epsilon = \frac{1-a}{2}$ From Theorem 2, we obtain the asymptotic expression

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} lML_N &= \frac{1}{4}\log(a(1-a)^2)\,, \\ &\simeq -0.51773\,. \end{aligned}$$

A crucial point here is that this value is less than $\log\lambda \simeq -0.48121$, which corresponds to the performance of the GS algorithm.

Case b: $\epsilon = 0$ Theorem 3 now gives the asymptotic expression

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} lML_N &= \log(1-a)\,, \\ &\simeq -0.21582 \end{aligned}$$

This value is now much larger that $\log\lambda$. However, we still have convergence to 1 for P_N^{GS} and P_N^F, see Corollary 2.

5.3 Partitions and entropies

Some asymptotic performance characteristics of the GS4 algorithm are closely connected to various entropies of the associated dynamic system. The link is a consequence of the connections between the transition probabilities of the Markov chain, the lengths of the intervals of the partition S^∞ and the rates R_i. These calculations are really of a general nature, but are presented here specialised to the GS4 algorithm. An important consequence is the contention that the second-order Renyi entropy is to be preferred to Shannon entropy for a wide class of search algorithms.

Consider the partition $\mathcal{P}_1$ of the interval $[0,1]$ defined by the intervals $\mathcal{I}_i, i = 1,\ldots,12$. When applying the mapping $T(.)$ n times we obtain another partition of $[0,1]$:

$$\mathcal{P}_n = T^{-n}(\mathcal{P}_1) = \cup_{i=1}^{12}\{z_1 \in [0,1] | z_n = T^n(z_1) \in \mathcal{I}_i\}\,.$$

The join of partitions $\mathcal{P}_1,\ldots,\mathcal{P}_n$ is then defined as

$$\begin{aligned}\mathcal{Q}_n(\mathcal{P}_1) &= \mathcal{P}_1 \vee \mathcal{P}_2 \vee \ldots \vee \mathcal{P}_n \\ &= \{\cap_{i_1,\ldots,i_n}\mathcal{I}_{i_1,\ldots,i_n}\,,\ \ i_k \in \{1,\ldots,12\}\,,\ \ k = 1,\ldots,n\}\ ,\end{aligned}$$

which contains all intersections of elements of partitions $\mathcal{P}_1,\ldots,\mathcal{P}_n$, that is

$$\mathcal{I}_{i_1,\ldots,i_n} = \{z_1 \in [0,1] | z_1 \in \mathcal{I}_{i_1},\ldots,z_n \in \mathcal{I}_{i_n}\}\,.$$

Let $\mu(.)$ be a probability measure on Borel sets of $[0,1]$, $\mathcal{P}$ be a partition of $[0,1]$ and $\Phi_0(\mathcal{Q}_n(\mathcal{P}),\mu)$ be the Shannon entropy of the partition $\mathcal{Q}_n(\mathcal{P}) = \mathcal{P} \vee T^{-1}(\mathcal{P}) \vee \ldots \vee T^{-n+1}(\mathcal{P})$. If $\mathcal{P} = \mathcal{P}_1$ then

$$\Phi_0(\mathcal{Q}_n(\mathcal{P}_1),\mu) = -\sum_{i_1,\ldots,i_n} \mu(\mathcal{I}_{i_1,\ldots,i_n})\log(\mu(\mathcal{I}_{i_1,\ldots,i_n}))\,,$$

Define

$$h_0(T,\mathcal{P}_1,\mu) = \lim_{n\to\infty}\frac{1}{n}\Phi_0(\mathcal{Q}_n(\mathcal{P}_1),\mu)\,,$$

if the limit exists. The limit certainly exists if $\mu(\cdot)$ is the invariant measure $\mu_z(\cdot)$ for the mapping $T(\cdot)$. The Kolmogorov, or metric, entropy of the dynamic system with respect to the measure $\mu_z(\cdot)$ is defined as:

$$h_0(T) = h_0(T,\mu_z) = \sup_{\mathcal{P}} h_0(T,\mathcal{P},\mu_z)\,.$$

Since for any pair of points $z_1 \neq z_1'$ there exists n such that $z_n = T^n(z_1) \in \mathcal{I}_{i_n}$ and $z_n' = T^n(z_1') \notin \mathcal{I}_{i_n}$, then $\mathcal{Q}_\infty(\mathcal{P}_1) = \vee_{i=1}^{\infty}\mathcal{P}_i$ consists only of singletons. Thus, from [1], Theorem 4.6 p. 215, $\mathcal{P}_1$ is a generating partition and $h_0(T) = h_0(T,\mathcal{P}_1,\mu_z)$.

From the Markov-chain representation of Section 3.3, the measure $\mu_z(\mathcal{I}_{i_1,\ldots,i_n})$ can be written as $\mu_z(\mathcal{I}_{i_1,\ldots,i_n}) = \bar{\pi}_{i_1}\pi_{i_1 i_2}\ldots\pi_{i_{n-1}i_n}$, which, using (13) and (17), gives the following expression for $h_0(T)$:

$$\begin{aligned}h_0(T) &= \lim_{n\to\infty} -\frac{1}{n}\sum_{i_1,\ldots,i_n} \bar{\pi}_{i_1}\pi_{i_1 i_2}\ldots\pi_{i_{n-1}i_n}\log(\bar{\pi}_{i_1}\pi_{i_1 i_2}\ldots\pi_{i_{n-1}i_n}) \\ &= \lim_{n\to\infty} -\frac{1}{n}\sum_{i_1,\ldots,i_n} \bar{\pi}_{i_1}\pi_{i_1 i_2}\ldots\pi_{i_{n-1}i_n}\log(\bar{\pi}_{i_1}R_{i_1}R_{i_2}\ldots R_{i_{n-1}}\frac{|\mathcal{I}_{i_n}|}{|\mathcal{I}_{i_1}|}) \\ &= \lim_{n\to\infty} -\frac{1}{n}\sum_{i_1,\ldots,i_n} \bar{\pi}_{i_1}\pi_{i_1 i_2}\ldots\pi_{i_{n-1}i_n}\log(L_0 R_{i_1}R_{i_2}\ldots R_{i_{n-1}})\,.\end{aligned}$$

The only difference with $H = \lim_{N\to\infty} -\frac{1}{N}ElL_N$, where ElL_N is given by (25) for x^* uniformly distributed on $[A,B]$, comes from the use of the invariant distribution

in the definition of $h_0(T)$. However, when $\epsilon = \frac{1-a}{2}$, for any $\delta > 0$, there exists N_0 such that $|\pi_i^{(N_0)} - \bar{\pi}_i| < \delta\bar{\pi}_i, i = 1, \ldots, 12$, and we can then easily prove that $h_0(T) = \lim_{N\to\infty} -\frac{1}{N} E_{x^*}\{\log L_N(x^*)\}$. Note that, under the less restrictive assumption that x^* has a prior density on $[A, B]$ absolutely continuous with respect to the Lebesgue measure, this also follows from Theorem 2 and the Shannon-McMillan-Breiman Theorem, see [1] p. 214, which implies that $h_0(T) = \Lambda$.

In what follows we consider entropies of the dynamic system (11) based on the following functionals on partitions:

$$\begin{aligned} \Phi_\gamma(\mathcal{Q}_n(\mathcal{P}_1), \mu) &= -\frac{1}{\gamma} \log \left(\sum_{i_1,\ldots,i_n} \mu^{1+\gamma}(\mathcal{I}_{i_1,\ldots,i_n}) \right) , \quad \gamma > 0 \,, \\ \Phi_*(\mathcal{Q}_n(\mathcal{P}_1)) &= \log |\mathcal{Q}_n(\mathcal{P}_1)| \,, \end{aligned} \tag{40}$$

where $|\mathcal{Q}_n|$ is the number of elements in the partition $\mathcal{Q}_n$ and $\mu(.)$ is a fixed measure on $[0, 1]$. $\Phi_\gamma(\mathcal{Q}_n, \mu)$ belongs to the class of Renyi entropies for the partition $\mathcal{Q}_n$ [9].

The topological entropy for the dynamic system $T(.)$ with respect to $\mathcal{P}_1$, which is a measure of the complexity of the algorithm, is defined by

$$h_*(T, \mathcal{P}_1) = \lim_{n\to\infty} \frac{1}{n} \Phi_*(\mathcal{Q}_n(\mathcal{P}_1)) \,, \tag{41}$$

if the limit exists, and we define the expected-cost entropy of degree γ with respect to $\mathcal{P}_1$ and $\mu(.)$ as

$$h_\gamma(T, \mathcal{P}_1, \mu) = \lim_{n\to\infty} \frac{1}{n} \Phi_\gamma(\mathcal{Q}_n(\mathcal{P}_1), \mu) \,, \tag{42}$$

if the limit exists. When $\gamma = 1$, we simply call $h_1(T, \mathcal{P}_1, \mu)$ the expected-cost entropy, which can be justified from the following property.

Theorem 8 *Let $\mu(.)$ be any measure with a density constant on the intervals $\mathcal{I}'_i$ associated with the eight states S'_i (26), and $\mathcal{P}_1$ be the partition defined by the intervals $\mathcal{I}_i, i = 1, \ldots, 12$. Then the limit (42) exists and equals*

$$h_\gamma(T, \mathcal{P}_1, \mu) = -\frac{1}{\gamma} \log \lambda_{\max}(\boldsymbol{R}'_\gamma) \,, \tag{43}$$

independently of μ, where $\boldsymbol{R}'_\gamma$ is defined by (37).

Proof.

Let $\pi'_i = \mu(\mathcal{I}'_i), i = 1, \ldots, 8$. Then $\mu(\mathcal{I}_{i_1,\ldots,i_n}) = \pi'_{i_1}\pi'_{i_1 i_2} \ldots \pi'_{i_{n-1} i_n}$, with π'_{ij} the element i, j of the submatrix $\boldsymbol{P}'$ of the matrix $\boldsymbol{P}$ associated with the eight states S'_i (26). From (40) and (42), this gives the following expression for $h_\gamma(T, \mathcal{P}_1, \mu)$:

$$\begin{aligned} h_\gamma(T, \mathcal{P}_1, \mu) &= \lim_{n\to\infty} -\frac{1}{n\gamma} \log \left(\sum_{i_1,\ldots,i_n} (\pi'_{i_1}\pi'_{i_1 i_2} \ldots \pi'_{i_{n-1} i_n})^{1+\gamma} \right) , \\ &= \lim_{n\to\infty} -\frac{1}{n\gamma} \log((\boldsymbol{\pi}'_\gamma)^T (\boldsymbol{R}'_\gamma)^{n-1} \mathbf{1}) \,, \end{aligned}$$

where $\mathbf{1} = (1,1,1,1,1,1,1,1)^T$ and $\boldsymbol{\pi}'_\gamma = ((\pi'_1)^{1+\gamma}, \ldots, (\pi'_8)^{1+\gamma})^T$. Since $\boldsymbol{R}'_\gamma$ is strong mixing, see the proof of Theorem 4, we obtain the result (43). That the existence of the limit for one measure implies the same limit for equivalent measures can also be established from ergodic considerations.

■

Corollary 4 *Take μ and $\mathcal{P}_1$ as in Theorem 8, and let x^* be uniformly distributed on $[A,B]$, then for the* GS4 *algorithm with $\epsilon = \frac{1-a}{2}$ we have for any $\gamma > 0$:*

$$h_\gamma(T, \mathcal{P}_1, \mu) = -\frac{1}{\gamma} \lim_{N\to\infty} \frac{lEL_N^\gamma}{N} .$$

In particular, when $\gamma = 1$

$$h_1(T, \mathcal{P}_1, \mu) = -\lim_{N\to\infty} \frac{1}{N} \log E_{x^*}\{L_N(x^*)\} .$$

Note that Theorem 8 may not be valid when the support of $\mu(.)$ is not restricted to the intervals $\mathcal{I}'_i, i = 1, \ldots, 8$ (e.g. when $\epsilon = 0$ in the GS4 algorithm, see Theorem 7 for a lower bound on lEL_N^γ).

It should be stressed that although h_1 has more practical interpretation than h_0 in terms of performance characteristics, it does not seem to have been considered in the literature on dynamic systems.

Similarly to the definition of Kolmogorov entropy we can define

$$h_\gamma(T, \mu) = \sup_{\mathcal{P}} h_\gamma(T, \mathcal{P}, \mu) ,$$

if $\lim_{n\to\infty} \frac{1}{n}\Phi_\gamma(\mathcal{Q}_n(\mathcal{P}), \mu)$ exists. Since $\mathcal{P}_1$ is a generating partition, we believe that $h_\gamma(T, \mu) = h_\gamma(T, \mathcal{P}_1, \mu)$, even if this point requires further investigations.

Analogous considerations for the topological entropy yield the following property.

Theorem 9 *Let $\mathcal{P}_1$ be the partition defined by the intervals $\mathcal{I}_i, i = 1, \ldots, 12$. Then the limit (41) exists for the dynamic system associated with the GS4 algorithm with $\epsilon = \frac{1-a}{2}$ and equals*

$$h_*(T, \mathcal{P}_1) = \log \lambda_{\max}(\tilde{\boldsymbol{P}}') \simeq 0.65103 , \tag{44}$$

where $\tilde{\boldsymbol{P}}'$ is defined by (27) and where $\lambda_{\max}(\tilde{\boldsymbol{P}}') \simeq 1.9175$ is the maximal root of $t^5 - t^4 - t^3 - 2t^2 + 2 = 0$.

Proof.

The topological entropy counts the growth in the number of trajectories of the algorithm, and is given by

$$h_*(T, \mathcal{P}_1) = \lim_{n\to\infty} \frac{1}{n} \log(\text{number of } i_1, \ldots, i_n | z_{i_1} \in \mathcal{I}_{i_1}, \ldots, z_{i_n} \in \mathcal{I}_{i_n}) .$$

Since $\epsilon = \frac{1-a}{2}$, the intervals $\mathcal{I}_1, \mathcal{I}_7, \mathcal{I}_8$ and $\mathcal{I}_{12}$ are never visited and $h_*(T, \mathcal{P}_1)$ can be written as:

$$h_*(T, \mathcal{P}_1) = \lim_{n\to\infty} \frac{1}{n} \log \left(\sum_{i_1,\ldots,i_n} \tilde{\pi}'_{i_1 i_2} \ldots \tilde{\pi}'_{i_{n-1} i_n} \right) ,$$

with $\tilde{\pi}'_{ij}$ defined by (27), $i, j = 1, \ldots, 8$. Similarly to the proof of Theorem 6, we obtain

$$h_*(T, \mathcal{P}_1) = \lim_{n\to\infty} \frac{1}{n} \log \mathbf{1}^T (\tilde{\boldsymbol{P}}')^{n-1} \mathbf{1} ,$$

which give the result (44).

■

Note that (44) gives a lower bound for the topological entropy of the dynamic system $T(.)$ (11) associated with the GS4 algorithm with $\epsilon = 0$.

6 Discussion and conclusions

The asymptotic performances of the GS and GS4 algorithms are summarized in Table 7.

The performances in terms of $\lim_{N\to\infty} \frac{1}{N} lEL_N^{1-\alpha}$ are not indicated in the table since this characteristic coincides with $\lim_{N\to\infty} \frac{1}{N} ElL_N$, see Theorem 5. The fact that $-\log 2$ is a lower bound for $\lim_{N\to\infty} \frac{1}{N} ElL_N$ is proved in [7], and an algorithm that achieves this lower bound for locally symmetric functions satisfying (28) is detailed in [11]. The fact that $-\log 2$ is a lower bound for $\lim_{N\to\infty} \frac{1}{N} lEL_N$ then follows from Jensen's inequality:

$$\log E_{x^*}\{L_N(x^*)\} \geq E_{x^*}\{\log L_N(x^*)\} .$$

A family of algorithms with performances $\lim_{N\to\infty} \frac{1}{N} lEL_N$ arbitrarily close to the bound $-\log 2$ for functions symmetric with respect to x^* is presented in [8]. However, these algorithms are mainly of theoretical interest since their finite sample behaviour is inferior. The bound $-\log 2$ for $\lim_{N\to\infty} \frac{1}{N} lML_N$ simply follows from

$$\max_{x^*\in[A,B]} \{\log L_N(x^*)\} \geq E_{x^*}\{\log L_N(x^*)\} .$$

The existence of second-order algorithms achieving this bound for functions symmetric with respect to x^* remains an open question.

Table 7 shows that the GS4 algorithm with $\epsilon = \frac{1-a}{2}$ has much better asymptotic performances than the GS algorithm. This alone is not enough, however, to give GS4 some practical interest. One should also consider Tables 4 and 5 which illustrate the superiority of GS4 over GS for finite N. The robustness of the performances with respect to non-symmetry of the function is illustrated by Figures 6 and 7. All these results make GS4 a promising alternative to Golden-Section and Fibonacci algorithms.

References

[1] O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall, eds., *Networks and Chaos – Statistical and Probabilistic Aspects*, Chapman & Hall, London, 1993.

[2] T. Bedford, M. Keane, and C. Series, eds., *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*, Oxford University Press, New York, 1991.

[3] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Math., Springer Verlag, Heidelberg, 1975.

[4] D.-Z. Du and F.K. Hwang, *Combinatorial Group Testing and its Applications*, World Scientific, Singapore, 1993.

[5] J. Kiefer, *Optimum sequential search and approximation methods under minimum regularity assumptions*, J. Soc. Indust. Appl. Math., 5(3), 1957, pp. 105–136.

[6] L. Pronzato, H.P. Wynn and A.A. Zhigljavsky, *Improving on Golden-Section optimisation for locally symmetric functions*, Proceedings MODA 4–Advances in Model-Oriented Data Analysis, Ch.P. Kitsos and W.G. Müller, eds., Physica Verlag, Heidelberg, 1995, pp. 285–297.

[7] L. Pronzato, H.P. Wynn and A.A. Zhigljavsky, *Dynamic systems in search and optimisation*, CWI Quarterly, 1995, to appear.

[8] L. Pronzato, H.P. Wynn and A.A. Zhigljavsky. *Stochastic analysis of convergence via dynamic representation for a class of line-search algorithms*, Technical Report 95-54, Laboratoire I3S, CNRS-URA 1376, Sophia-Antipolis, 06560 Valbonne, France, submitted to *Combinatorics, Probability and Computing*, 1995.

[9] A. Rényi, *On the foundations of information theory*, Rev. Inst. Internat. Stat., 33, 1965, pp. 1–14.

[10] H.P. Wynn and A.A. Zhigljavsky, *Chaotic behaviour of search algorithms*, Acta Applicandae Mathematicae, 32, 1993, pp. 123–156.

[11] H.P. Wynn and A.A. Zhigljavsky, *Achieving the ergodically optimal convergence rate for a one-dimensional minimization problem*, J. of Complexity, 11, 1995, pp. 196–226.

A_n U_n V_n B_n

A_{n+1} E_{n+1} B_{n+1}

(R)

A_{n+1} E_{n+1} B_{n+1}

(L)

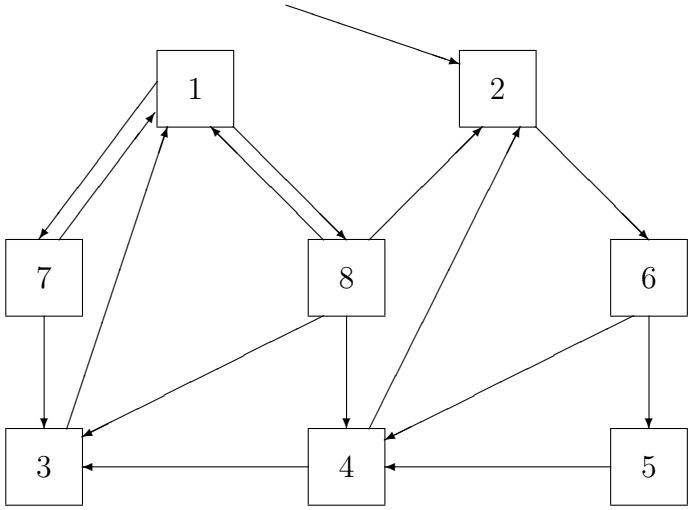
1
2
7
8
6
3
4
5

FIGURE CAPTIONS

Figure 1: One iteration in a second-order line search algorithm.

Figure 2: Graph of the Mapping $T(.)$ (11). The α_i's are indicated by stars.

Figure 3 : Graph of transition for states $S'_i, i = 1, \ldots, 8$.

Figure 4: Decimal logarithm of various performance characteristics ($L_0 = 1$).

Figure 5: Graphs of $f_{0.5}(.)$ and $f_2(.)$.

Figure 6: Evolution of EL_{30} and ML_{30} as functions of D ($L_0 = 1$).

Figure 7: Evolution of P_N^F as a function of D and N.

$z \qquad T(z)$

$N \qquad \log_{10}(1/F_{N+1}) \qquad \log_{10}(ML_N)$

$\log_{10}(L_N^{0.99}) \qquad \log_{10}(EL_N)$

$x - x^* \qquad f_{0.5}(x - x^*) \qquad f_2(x - x^*)$

$D \qquad EL_{30}$

$D \qquad ML_{30} \qquad \lambda^{29} \qquad 1/F_{31}$

$D \qquad N \qquad P_N^F$

N	λ^{N-1}	$\frac{1}{F_{N+1}}$	EL_N	ML_N	P_N^{GS}	P_N^F	$L_N^{0.99}$
1	1.0000	1.0000	1.8059	1.8059	0.0000	0.0000	1.8059
2	0.6180	0.5000	1.0000	1.0000	0.0000	0.0000	1.0000
3	0.3820	0.3333	0.6463	0.8059	0.3506	0.0000	0.8059
4	0.2361	0.2000	0.3862	0.6494	0.3506	0.3506	0.6494
5	0.1459	0.1250	0.2117	0.5234	0.1361	0.1361	0.5234
6	$9.017\ 10^{-2}$	$7.692\ 10^{-2}$	0.1064	0.1835	0.3555	3.555	0.1835
7	$5.573\ 10^{-2}$	$4.762\ 10^{-2}$	$6.076\ 10^{-2}$	0.1016	0.6100	0.6100	0.1016
8	$3.444\ 10^{-2}$	$2.941\ 10^{-2}$	$3.345\ 10^{-2}$	$8.188\ 10^{-2}$	0.6100	0.6100	$8.188\ 10^{-2}$
9	$2.129\ 10^{-2}$	$1.818\ 10^{-2}$	$1.764\ 10^{-2}$	$6.598\ 10^{-2}$	0.8421	0.3823	$6.598\ 10^{-2}$
10	$1.316\ 10^{-2}$	$1.124\ 10^{-2}$	$9.511\ 10^{-3}$	$2.313\ 10^{-2}$	0.6196	0.6196	$2.313\ 10^{-2}$
11	$8.131\ 10^{-3}$	$6.944\ 10^{-3}$	$5.266\ 10^{-3}$	$1.281\ 10^{-2}$	0.8207	0.8207	$1.281\ 10^{-2}$
12	$5.025\ 10^{-3}$	$4.292\ 10^{-3}$	$2.835\ 10^{-3}$	$1.032\ 10^{-2}$	0.9437	0.8207	$1.032\ 10^{-2}$
13	$3.106\ 10^{-3}$	$2.653\ 10^{-3}$	$1.521\ 10^{-3}$	$8.318\ 10^{-3}$	0.9437	0.9437	$3.618\ 10^{-3}$
14	$1.919\ 10^{-3}$	$1.639\ 10^{-3}$	$8.300\ 10^{-4}$	$2.916\ 10^{-3}$	0.8080	0.8080	$2.004\ 10^{-3}$
15	$1.186\ 10^{-3}$	$1.013\ 10^{-3}$	$4.515\ 10^{-4}$	$1.615\ 10^{-3}$	0.9267	0.9267	$1.615\ 10^{-3}$
16	$7.331\ 10^{-4}$	$6.261\ 10^{-4}$	$2.433\ 10^{-4}$	$1.301\ 10^{-3}$	0.9817	0.9817	$1.301\ 10^{-3}$
17	$4.531\ 10^{-4}$	$3.870\ 10^{-4}$	$1.318\ 10^{-4}$	$1.049\ 10^{-3}$	0.9817	0.9817	$4.562\ 10^{-4}$
18	$2.800\ 10^{-4}$	$2.392\ 10^{-4}$	$7.166\ 10^{-5}$	$3.676\ 10^{-4}$	0.9978	0.9134	$2.526\ 10^{-4}$
19	$1.731\ 10^{-4}$	$1.478\ 10^{-4}$	$3.879\ 10^{-5}$	$2.036\ 10^{-4}$	0.9725	0.9725	$2.036\ 10^{-4}$
20	$1.070\ 10^{-4}$	$9.136\ 10^{-5}$	$2.098\ 10^{-5}$	$1.640\ 10^{-4}$	0.9943	0.9943	$7.136\ 10^{-5}$
21	$6.611\ 10^{-5}$	$5.646\ 10^{-5}$	$1.138\ 10^{-5}$	$1.322\ 10^{-4}$	0.9995	0.9943	$3.952\ 10^{-5}$
22	$4.086\ 10^{-5}$	$3.490\ 10^{-5}$	$6.173\ 10^{-6}$	$4.634\ 10^{-5}$	0.9994	0.9994	$3.184\ 10^{-5}$
23	$2.525\ 10^{-5}$	$2.157\ 10^{-5}$	$3.341\ 10^{-6}$	$2.566\ 10^{-5}$	0.9903	0.9903	$1.116\ 10^{-5}$
24	$1.561\ 10^{-5}$	$1.333\ 10^{-5}$	$1.811\ 10^{-6}$	$2.068\ 10^{-5}$	0.9983	0.9983	$6.182\ 10^{-6}$
25	$9.645\ 10^{-6}$	$8.238\ 10^{-6}$	$9.817\ 10^{-7}$	$1.667\ 10^{-5}$	0.9999	0.9999	$4.982\ 10^{-6}$
26	$5.961\ 10^{-6}$	$5.091\ 10^{-6}$	$5.318\ 10^{-7}$	$5.843\ 10^{-6}$	1.0000	0.9999	$4.015\ 10^{-6}$
27	$3.684\ 10^{-6}$	$3.147\ 10^{-6}$	$2.881\ 10^{-7}$	$3.235\ 10^{-6}$	1.0000	0.9967	$1.407\ 10^{-6}$
28	$2.277\ 10^{-6}$	$1.945\ 10^{-6}$	$1.561\ 10^{-7}$	$2.607\ 10^{-6}$	0.9995	0.9995	$7.793\ 10^{-7}$
29	$1.407\ 10^{-6}$	$1.202\ 10^{-6}$	$8.462\ 10^{-8}$	$2.101\ 10^{-6}$	1.0000	1.0000	$6.280\ 10^{-7}$
30	$8.697\ 10^{-7}$	$7.428\ 10^{-7}$	$4.584\ 10^{-8}$	$7.366\ 10^{-7}$	1.0000	1.0000	$2.202\ 10^{-7}$

Table 3: Performances of the GS4 algorithm with $\epsilon = \frac{1-a}{2}$ ($L_0 = 1$).

N	EL_N	ML_N	P_N^{GS}	P_N^F	$L_N^{0.99}$
1	1	1	0	0	1
2	0.5537	0.5537	0	0	0.5537
3	0.3973	0.4463	0.1941	0.1941	0.4463
4	0.2789	0.3596	0.3506	0.3506	0.3596
5	0.1943	0.2898	0.4766	0.3506	0.2898
6	0.1336	0.2336	0.4766	0.4766	0.2336
7	$9.069\ 10^{-2}$	0.1882	0.3378	0.3378	0.1882
8	$6.098\ 10^{-2}$	0.1517	0.4663	0.4663	0.1517
9	$4.078\ 10^{-2}$	0.1222	0.5699	0.5699	0.1222
10	$2.712\ 10^{-2}$	$9.851\ 10^{-2}$	0.6734	0.5699	$9.851\ 10^{-2}$
15	$3.350\ 10^{-3}$	$3.348\ 10^{-2}$	0.7209	0.7209	$3.348\ 10^{-2}$
20	$3.972\ 10^{-4}$	$1.138\ 10^{-2}$	0.8173	0.8173	$1.138\ 10^{-2}$
25	$4.637\ 10^{-5}$	$3.869\ 10^{-3}$	0.8830	0.8205	$1.156\ 10^{-3}$
30	$5.381\ 10^{-6}$	$1.315\ 10^{-3}$	0.8838	0.8838	$5.110\ 10^{-5}$

Table 4: Performances of the GS4 algorithm with $\epsilon = 0$ ($L_0 = 1$).

precision	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
λ^{N-1}	6	11	16	21	25	30
$\frac{1}{F_{N+1}}$	6	11	16	20	25	30
EL_N	7	10	14	18	22	25
ML_N	8	13	18	22	26	30
$L_N^{0.99}$	8	13	17	20	24	28

Table 5: Number of functions evaluations required to achieve a given precision ($L_0 = 1$).

$\bar{\pi}_1$	0	
$\bar{\pi}_2$	$\frac{1}{1763}(112a^3 + 4a^2 - 879a + 657)$	$\simeq 0.2764$
$\bar{\pi}_3$	$\frac{1}{1763}(100a^3 - 752a^2 + 1293a - 106)$	$\simeq 0.0666$
$\bar{\pi}_4$	$\frac{1}{1763}(-792a^3 + 2994a^2 - 2977a + 769)$	$\simeq 0.1691$
$\bar{\pi}_5$	$\frac{1}{1763}(-276a^3 + 242a^2 + 592a + 81)$	$\simeq 0.1152$
$\bar{\pi}_6$	$\frac{1}{1763}(644a^3 - 1740a^2 + 1557a - 189)$	$\simeq 0.0297$
$\bar{\pi}_7$	0	
$\bar{\pi}_8$	0	
$\bar{\pi}_9$	$\frac{1}{1763}(100a^3 - 752a^2 + 1293a - 106)$	$\simeq 0.0666$
$\bar{\pi}_{10}$	$\frac{1}{1763}(244a^3 - 495a^2 - 89a + 235)$	$\simeq 0.1139$
$\bar{\pi}_{11}$	$\frac{1}{1763}(-132a^3 + 499a^2 - 790a + 422)$	$\simeq 0.1625$
$\bar{\pi}_{12}$	0	

Table 6: Invariant distribution for the Markov chain.

	GS	GS4 ($\epsilon = 0$)	GS4 ($\epsilon = \frac{1-a}{2}$)	lower bound
$\lim_{N\to\infty} \frac{1}{N} ElL_N$	$\simeq -0.48121$	$\simeq -0.63006$	$\simeq -0.63006$	$-\log 2 \simeq -0.6930$
$\lim_{N\to\infty} \frac{1}{N} lEL_N$	$\simeq -0.48121$	$> -0.43163^*$	$\simeq -0.61273$	$-\log 2 \simeq -0.6930$
$\lim_{N\to\infty} \frac{1}{N} lML_N$	$\simeq -0.48121$	$\simeq -0.21582$	$\simeq -0.51773$	$-\log 2 \simeq -0.6930$

Table 7: Asymptotic performance characteristics of the GS and GS4 algorithms. The sign * indicates that the bound concerns $\liminf_{N\to\infty} \frac{1}{N} lEL_N$.