IMA Journal of Mathematical Control and Information (2001) **18**, 281–296

Analysis of performance of symmetric second-order line search algorithms through continued fractions

LUC PRONZATO†
Laboratoire I3S, CNRS–UNSA, bât. Euclide, 'Les Algorithmes', 2000 route des Lucioles, Sophia Antipolis, 06410 Biot, France

HENRY P. WYNN
Department of Statistics, University of Warwick, Coventry CV4 7AL, UK

AND

ANATOLY A. ZHIGLJAVSKY
School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF2 4YH, UK

[Received on 22 January 2000; revised on 13 April 2000]

The paper makes a connection between classical dynamical systems, namely the Gauss map and the associated Farey map, and symmetric second-order line search, of which the classical Golden-Section and Fibonacci search methods are particular cases. The main result is an expression for the actual finite sample rate for symmetric line-search algorithms in terms of the continued fraction expansion of the first observation point. This yields formulae for asymptotic rates in the almost sure ergodic case and exact rates for the two special cases, (i) the rational case when the algorithm stops and (ii) the cyclic case which corresponds to quadratic irrationality. The asymptotic rate is sub-exponential for almost all starting points. Moreover, the dynamical system suffers so-called chaos with intermittency, typical of systems which have fairly normal (exponential) expansion punctuated by periods of very slow expansion. The property follows immediately through the link with the Farey map which has slow growth close to zero.

Keywords: dynamical system; farey map; Fibonacci; Golden Section; Gauss map; line search.

1. Introduction

This paper is a contribution to a series studying dynamical systems embedded within certain search algorithms (Pronzato *et al.*, 1997, 1998, 1999). This is a new area in that it describes kind of stochasticity different from what is commonly considered as a random algorithm, for example arising in graph theory or in Markovian algorithms such as simulated annealing or genetic algorithms.

The paper makes a surprising connection between classical dynamical systems, namely the Gauss map and the associated Farey map, and a special kind of line search. The latter shares a symmetry property with the classical Golden-Section and Fibonacci search methods. Unlike previous work by the authors, which concentrates on the path of the

†Author for correspondence.

optimum after a renormalization, this paper covers the path of the observation points. Some progress towards the ideas of this paper appeared in Wynn & Zhigljavsky (1993a). The main result is to give a formula for the actual finite sample rate for symmetric line search in terms of the continued fraction expansion of the first observation point. This leads to asymptotic rates in the almost sure ergodic case (based on the Gauss map's invariant measure) and exact rates for the two special cases, (i) the rational case when the algorithm stops and (ii) the cyclic case which corresponds to quadratic irrationality.

This represents a more or less complete analysis of the rates of convergence for a symmetric line-search algorithm. Except for particular cases (zero Lebesgue measure for the starting point) such as the Golden-Section case, the other asymptotic rates are sub-exponential. Moreover, they suffer so-called chaos with intermittency, typical of systems which have fairly normal (exponential) expansion punctuated by periods of very slow expansion. Such behaviour is commented upon in various work in optimization where one is encouraged to take care in the initialization of a symmetric algorithm, even the Golden Section, after each iterate. The property follows immediately through the link with the Farey map which has slow growth close to zero.

The paper is organized as follows. Section 2 briefly introduces second-order line search, with the Golden-Section and Fibonacci algorithms as particular symmetric cases. A renormalization of variables in [0, 1] is used in Section 3, and we show in Section 4 how the dynamical system we associate with a symmetric algorithm is related to the Farey map. The relation between symmetric line search algorithms and continued fractions is investigated in Section 5.

2. Optimization based on comparisons of function values

2.1 *Second-order line search*

We consider the minimization of a uniextremal function $f(\cdot)$ on a given interval $[A_1, B_1]$ using a *second-order* algorithm, as defined by Kiefer (1957). Let x^* be the unknown point at which $f(\cdot)$ is minimum. We assume that $f(\cdot)$ is decreasing for $x \leqslant x^*$ and non-decreasing for $x > x^*$ (or non-increasing for $x \leqslant x^*$ and increasing for $x > x^*$). If these conditions on $f(\cdot)$ are not satisfied, then convergence to a local minimum of $f(\cdot)$ in $[A_1, B_1]$ may occur.

At iteration n we compare the values of $f(\cdot)$ at two points U_n and V_n in the current uncertainty interval $[A_n, B_n]$, with $U_n < V_n$. Then, after this iteration if $f(U_n) \geqslant f(V_n)$ 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 re-initialized. (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 Fig. 1 in which (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 compare $f(E_{n+1})$ to the value of $f(\cdot)$ at a new point E'_{n+1}.

A second-order line-search algorithm is therefore defined by the choice of the

(i) initial uncertainty interval $[A_1, B_1]$,

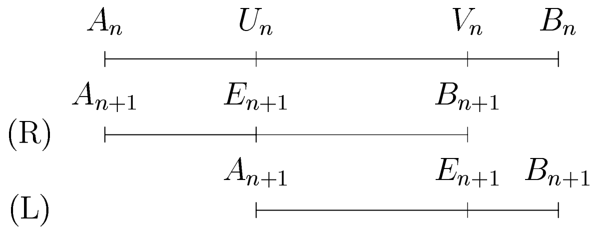

FIG. 1. One iteration in second-order line search.

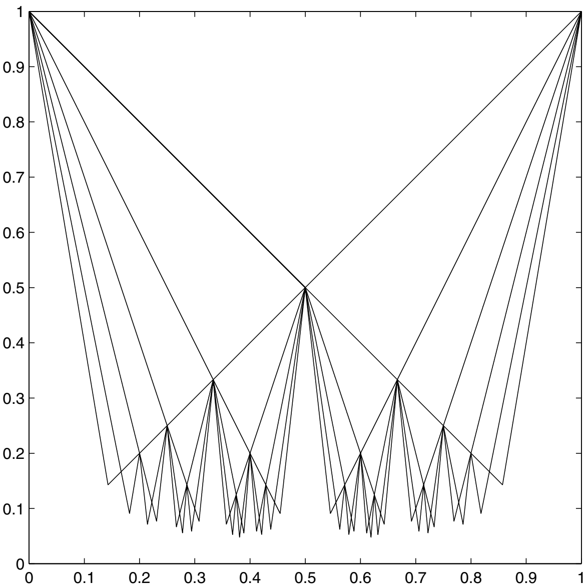

FIG. 2. $L_n(E_1)/L_1$ as a function of E_1, $n = 2, \ldots, 7$, for a symmetric algorithm.

(ii) initial test-point $E_1 \in (A_1, B_1)$,
(iii) selection rule for E'_{n+1}, $n \geqslant 0$.

The reduction rate at iteration n is defined as $r_n = L_{n+1}/L_n$, where $L_k = B_k - A_k$ is the length of the uncertainty interval at iteration k, so that

$$L_n = L_n(E_1) = L_1 \prod_{i=1}^{n-1} r_i, \; n = 2, \ldots.$$

A *symmetric algorithm* corresponds to the case where E'_n is selected according to the rule $E'_n = A_n + B_n - E_n$. In that case, the length L_n does not depend on the sequence of (R) and (L) deletions and is thus independent of the objective function $f(\cdot)$. It depends only on E_1. Figure 2 presents L_n/L_1 as a function of E_1 for $n = 2, \ldots, 7$.

The most famous second-order line-search algorithms are the Fibonacci and the Golden-Section (GS) algorithms. Both are symmetric.

2.2 *GS and Fibonacci algorithms*

The well-known GS algorithm is defined by

$$E_1 = A_1 + \varphi L_1,$$
$$E'_n = \begin{cases} A_n + \varphi L_n & \text{if} \quad E_n = A_n + (1-\varphi)L_n \\ A_n + (1-\varphi)L_n & \text{if} \quad E_n = A_n + \varphi L_n \end{cases} \tag{1}$$

where $L_n = B_n - A_n$ and where $\varphi = (\sqrt{5}-1)/2 \simeq 0{\cdot}618\,04$ is the largest root of $\varphi^2 + \varphi - 1 = 0$ and is called the GS ratio. This algorithm is known to be asymptotically worst-case optimal in the class of all uniextremal functions (see Kiefer (1957) and Du & Hwang (1993) Theorem 9.2.2, p. 181). We come back to this point in Section 5.3. The convergence rate at iteration n satisfies $r_n = \varphi$, $\forall n \geqslant 1$, so that $L_n = L_1\varphi^{n-1}$. When the number of function evaluations is fixed *a priori*, say equal to N, the worst-case optimal algorithm, in the sense of L_N, is the Fibonacci algorithm (Kiefer, 1957), for which

$$E_1 = A_1 + \frac{F_N}{F_{N+1}} L_1,$$
$$E'_n = A_n + B_n - E_n,$$

where $(F_i)_{i=1}^{\infty} = \{1, 1, 2, 3, 5, 8, 13, \dots\}$ is the Fibonacci sequence, defined by $F_1 = F_2 = 1$ and $F_n = F_{n-1} + F_{n-2}$, $n > 2$. The algorithm satisfies

$$r_n = \frac{F_{N+1-n}}{F_{N+2-n}}, \qquad 1 \leqslant n < N,$$
$$L_n = L_1 \frac{F_{N+2-n}}{F_{N+1}}, \qquad 1 \leqslant n \leqslant N.$$

(see Fig. 3). The algorithm stops at $n = N$, so that one can define $r_n = 1$ and $L_n = L_N$ for $n \geqslant N$. Note that in the last iteration one assumes that the two test points coincide. In practice they should be chosen as close as possible. From the recurrence defining F_n, we obtain

$$F_n = \frac{(1+\varphi)^n - (-\varphi)^n}{\sqrt{5}}.$$

When N tends to infinity, the ratio of L_N for the GS algorithm to L_N for the Fibonacci algorithm tends to $(2+\varphi)/\sqrt{5} \simeq 1{\cdot}170\,82$.

3. Renormalization

Consider the line-search algorithms of the previous section. After left or right deletion, we renormalize each uncertainty interval $[A_n, B_n]$ to [0, 1]. Thus introduce normalized variables in [0, 1]

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

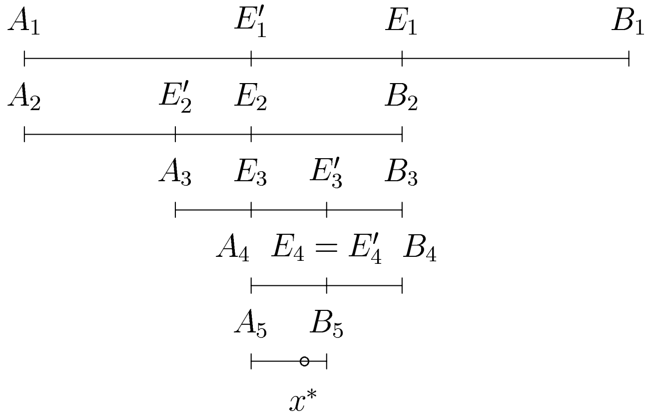

FIG. 3. Five iterations of the Fibonacci algorithm.

and

$$u_n = \min(e_n, e_n'), \; v_n = \max(e_n, e_n').$$

The deletion rule is

$$\begin{cases} \text{(R):} \quad \text{if } f_n(u_n) < f_n(v_n) \text{ delete } (v_n, 1] \\ \text{(L):} \quad \text{if } f_n(u_n) \geqslant f_n(v_n) \text{ delete } [0, u_n) \end{cases}$$

with

$$f_n(x) = f(L_n x + A_n).$$

The remaining interval is then renormalized to $[0, 1]$. The successive lengthes of uncertainty intervals satisfy

$$L_{n+1} = \begin{cases} L_n v_n & \text{(R)} \\ L_n(1 - u_n) & \text{(L)} \end{cases}$$

and therefore

$$r_n = \begin{cases} v_n & \text{(R)} \\ 1 - u_n & \text{(L).} \end{cases}$$

Straightforward calculation shows that right and left deletions, respectively, give

$$x_{n+1} = \begin{cases} \dfrac{x_n}{v_n} & \text{(R)} \\ \dfrac{x_n - u_n}{1 - u_n} & \text{(L).} \end{cases}$$

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

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

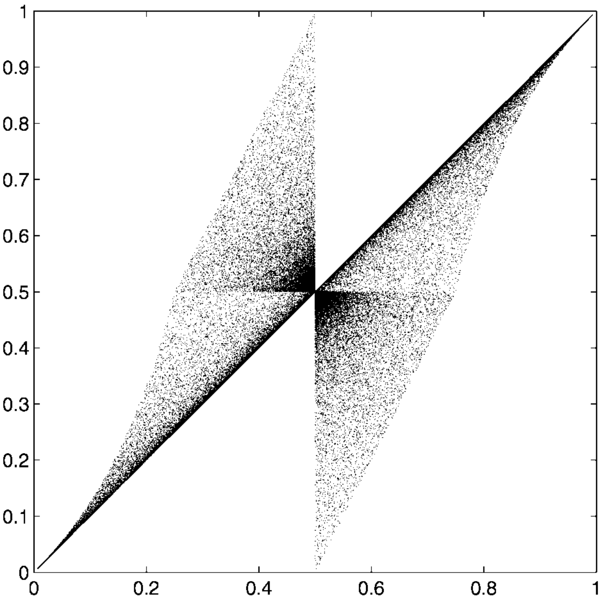

FIG. 4. Typical chaotic sequence of iterates (x_n, x_{n+1}), $n = 1, \dots, 100\,000$, in a symmetric algorithm.

Figure 4 presents a plot of a typical sequence of iterates (x_n, x_{n+1}), $n = 1, \dots, 100\,000$ for a non-periodic symmetric algorithm, with x_n the renormalized location of x^*.

The GS algorithm corresponds to

$$v_n = 1 - u_n = \varphi,$$

which gives $x_{n+1} = h_n(x_n)$, with

$$h_n(x_n) = \begin{cases} x_n(1+\varphi) & \text{if } f_n(1-\varphi) < f_n(\varphi) \quad \text{(R)} \\ x_n(1+\varphi) - \varphi & \text{if } f_n(1-\varphi) \geqslant f_n(\varphi) \quad \text{(L).} \end{cases}$$

The ergodic behaviour of this dynamical system is studied in Wynn & Zhigljavsky (1993b) in the case where $f(\cdot)$ is symmetric with respect to x^*. We show in Pronzato *et al.* (1995) that this ergodic behaviour is the same for functions that are only locally symmetric.

4. Relation with Farey map

In the case of a symmetric algorithm one has for every n

$$e'_n = 1 - e_n.$$

This last condition implies $u_n = 1 - v_n$, which gives both for the (R) and (L) cases $r_n = v_n$ and

$$r_{n+1} = \begin{cases} \dfrac{1-r_n}{r_n} & \text{if } \frac{1}{2} \leqslant r_n < \frac{2}{3} \\ 2 - \dfrac{1}{r_n} & \text{if } \frac{2}{3} \leqslant r_n \leqslant 1 \end{cases}$$

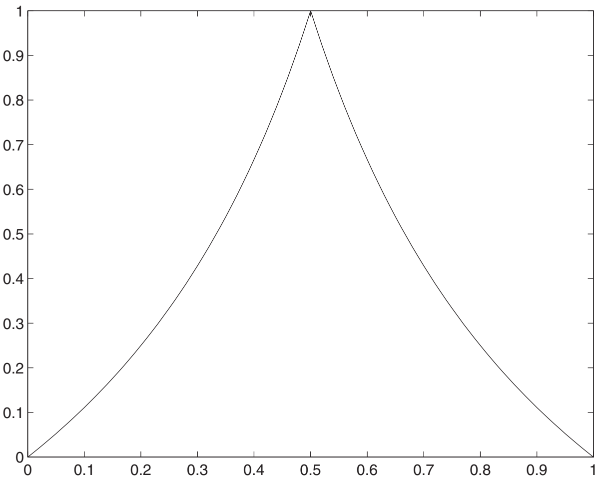

FIG. 5. The Farey map.

with the sequence (r_n) living in $[\frac{1}{2}, 1]$. Note that the updating formula for r_n implies by induction that, for fixed n, the function $L_n(E_1)/L_1$ in Fig. 2 is piecewise linear.

The variable z_n obtained by the simple transformation $z_n = (1/r_n) - 1$ follows the dynamical system $z_{n+1} = T(z_n)$, with $T(\cdot)$ given by (7) below, that is the Farey map. However, we shall not pursue this connection but use a different one.

Neither the shape of the objective function, nor the location of x^* has any effect on the behaviour of (r_n). We can thus assume for simplicity that the objective function monotonously increases, so that the rule (R) always applies. Then, the evolution of (e_n) becomes

$$e_{n+1} = \frac{u_n}{v_n} = \begin{cases} \dfrac{e_n}{1-e_n} & \text{if} \quad 0 \leqslant e_n < \frac{1}{2} \\ \dfrac{1-e_n}{e_n} & \text{if} \quad \frac{1}{2} \leqslant e_n \leqslant 1 \end{cases} \tag{2}$$

which is exactly the Farey map (see Bowen, 1979), presented in Fig. 5. The associated sequence of rates is

$$r_n = \begin{cases} 1 - e_n & \text{if} \quad 0 \leqslant e_n < \frac{1}{2} \\ e_n & \text{if} \quad \frac{1}{2} \leqslant e_n \leqslant 1. \end{cases}$$

The Frobenius–Perron equation for the dynamical system $e_n \mapsto e_{n+1} = T(e_n)$ is given by

$$\phi(e) = \frac{1}{(e+1)^2}\left(\phi\left(\frac{e}{e+1}\right) + \phi\left(\frac{1}{e+1}\right)\right),$$

which has solution

$$\phi(e) = \frac{1}{e}, \quad 0 < e < 1. \tag{3}$$

The fact that the invariant measure above is not finite is related to the slope of the mapping $T(\cdot)$ being unity at zero, a fixed point of the mapping, which causes non-exponential divergence of the trajectories. Such a map is called *almost expanding*, which relates to the phenomenon of *chaos with intermittency*. This term is used to describe the occasional regular behaviour of the trajectories which occurs in the present case near zero.

Note that any statement on the asymptotic behaviour of (e_n) deriving from the Farey map can be reformulated in terms of the behaviour of (r_n). For example, the invariant density for (r_n) is

$$\phi(r) = \frac{1}{r(1-r)}, \quad \tfrac{1}{2} \leqslant r < 1. \tag{4}$$

The behaviour (and the convergence rate) of the symmetric algorithm is completely determined by e_1. We show later that the best asymptotic convergence rate is for the GS algorithm, that is, when $e_1 = \varphi$, the Golden Section. The relation between the rate sequence (r_n) and the Farey map will be exploited further in Section 5.2.

5. Relation with continued fractions

5.1 *Continued fraction expansion*

Any irrational number α in [0, 1) has a unique continued fraction expansion of the form

$$\alpha = [a_1, a_2, \dots] = \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}},$$

where the *partial quotients* $a_1, a_2, \dots$ are positive integers. For any n, there exists a unique $\epsilon_{n+1} \in (0, 1)$, such that

$$\alpha = [a_1, a_2, \dots] = [a_1, a_2, \dots, a_n + \epsilon_{n+1}] = \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots \frac{1}{a_n + \epsilon_{n+1}}}}.$$

The sequence (ϵ_n) follows the dynamical system

$$\begin{aligned}
\epsilon_1 &= \alpha \\
\epsilon_2 &= \left\{\frac{1}{\epsilon_1}\right\} = \frac{1}{\epsilon_1} - \left\lfloor \frac{1}{\epsilon_1} \right\rfloor \\
&\ \ \vdots \\
\epsilon_{n+1} &= \left\{\frac{1}{\epsilon_n}\right\} = \frac{1}{\epsilon_n} - \left\lfloor \frac{1}{\epsilon_n} \right\rfloor
\end{aligned} \tag{5}$$

where $\{x\}$ is the fractional part and $\lfloor x \rfloor$ the integer part of x. For each n, the partial quotient a_n is $a_n = \lfloor 1/\epsilon_n \rfloor$.

The continued fraction expansion for a rational number α is finite and non-unique:

$$\alpha = [a_1, a_2, \ldots, a_n] = \begin{cases} [a_1, a_2, \ldots, a_n - 1, 1] & \text{if } a_n > 1, \\ [a_1, a_2, \ldots, a_{n-1} + 1] & \text{if } a_n = 1. \end{cases}$$

Note that one can also write (see Schoisengeier, 1990)

$$\alpha = [a_1, a_2, \ldots, a_n] = [a_1, a_2, \ldots, a_n, \infty]. \tag{6}$$

The dynamical system $\epsilon_n \mapsto \epsilon_{n+1} = T(\epsilon_n) = \{1/\epsilon_n\}$, defined by (5) is called the *Gauss map*. Its invariant density is

$$\phi(x) = \frac{1}{(1+x)\log 2}, \quad x \in [0, 1).$$

It describes the asymptotic behaviour of ϵ_n for almost all α in $[0, 1)$.

Continued fractions can be related to other dynamical systems in various ways, as a consequence of the relationship to the Gauss map. First consider the map

$$T_1(x) = \begin{cases} 1/x & \text{if } 0 < x \leqslant 1 \\ x - 1 & \text{if } 1 < x \end{cases}$$

defined on $(0, \infty)$. Starting with $\alpha \in (0, 1)$ and iterating, we immediately jump to the second branch and remain there until switching back to the first branch with the value $\{1/\alpha\} = 1/\alpha - \lfloor 1/\alpha \rfloor$. This is repeated so that at the nth switch back to the first branch we have the same ϵ_n as in the Gauss map. Moreover, the number of iterations spent in the second branch is precisely $a_n = \lfloor 1/\epsilon_n \rfloor$.

For the iteration just before coming to the first branch, the process is in $(1, 2]$. Consider a reduced mapping $T_2(\cdot)$, obtained by joining two iterations of $T_1(\cdot)$ when $x \in (1, 2]$. We have

$$T_2(x) = \begin{cases} T_1[T_1(x)] = 1/(x-1) & \text{if } 1 < x \leqslant 2 \\ T_1(x) = x - 1 & \text{if } 2 < x. \end{cases}$$

Now, the change of variable $e = 1/x$ in $T_2(\cdot)$ gives the Farey map presented in Fig. 5, (see (2))

$$T(e) = \begin{cases} e/(1-e) & \text{if } 0 < e < \frac{1}{2} \\ (1-e)/e & \text{if } \frac{1}{2} \leqslant e < 1. \end{cases} \tag{7}$$

Note that the number of iterations spent in the second branch of $T_2(\cdot)$ before reaching the first, and in the first branch of $T(\cdot)$ before reaching the second, is $a_n - 1 = \lfloor 1/\epsilon_n \rfloor - 1$.

5.2 *L_n as a function of e_1*

The following theorem gives the expression of the length of the uncertainty interval $[A_n, B_n]$ for a symmetric algorithm as a function of the continued fraction expansion of e_1. Note that the expression is valid for any objective function on which the line-search algorithm is applied.

THEOREM 1 Consider a symmetric second-order line-search algorithm, initialized at e_1 with continued fraction expansion $e_1 = [a_1, a_2, \dots]$. Define

$$n_0 = 0,\ n_j = \sum_{i=1}^{j} a_i,\ j \geqslant 1. \tag{8}$$

Then, for any N such that $n_j \leqslant N < n_{j+1}$, $j \geqslant 0$,

$$L_{N+1} = L_1 \times \epsilon_0 \times \cdots \times \epsilon_j \times [1 - (N - n_j)\epsilon_{j+1}], \tag{9}$$

where $\epsilon_0 = 1$, $\epsilon_1 = e_1$ and $\epsilon_{j+1} = \{1/\epsilon_j\}$, $j \geqslant 1$.

Proof. We assume that $e_1 < \frac{1}{2}$. The case $e_1 > \frac{1}{2}$ could be treated similarly. Recalling the relation between the Farey and continued fraction maps, we notice that between two successive visits to the interval $[\frac{1}{2}, 1]$, say the jth and $(j+1)$th, the sequence (2) spends $a_{j+1} - 1$ iterations in $[0, \frac{1}{2})$. The subsequence (e_{n_j}), with n_j given by (8), then contains all terms of the sequence (e_n) that belong to $[\frac{1}{2}, 1)$. Moreover, for $j \geqslant 0$ e_{n_j+1} exactly coincides with ϵ_{j+1} in the Gauss-map sequence (ϵ_j), defined by $\epsilon_{j+1} = \{1/\epsilon_j\}$ and $\epsilon_1 = e_1$. Also, if we consider $N = n_j$ iterations of the algorithm (2), then we arrive at the point $e_{n_j+1} = \epsilon_{j+1}$ and the length of the unnormalized uncertainty interval after these N iterations is equal to

$$L_{n_j+1} = L_1 \times r_1 \times \cdots \times r_{n_j} = L_1 \times \epsilon_1 \times \cdots \times \epsilon_j. \tag{10}$$

This can be proved as follows.

Let $j = 0$, and compute the reduction rate in n_1 iterations, from $e_{n_0+1} = e_1 = \epsilon_1$ to $e_{n_1+1} = \epsilon_2$. The length of the uncertainty interval then becomes

$$L_{n_1+1} = L_{n_0+1} \times r_{n_0+1} \times \cdots \times r_{n_1}.$$

As

$$\begin{cases} e_i < \frac{1}{2} & \text{and} \quad r_i = 1 - e_i,\ i = n_0 + 1, \dots, n_1 - 1, \\ \frac{1}{2} \leqslant e_{n_1} & \text{and} \quad r_{n_1} = e_{n_1}, \end{cases}$$

one obtains by induction

$$e_i = \frac{e_1}{1 - (i-1)e_1},\ r_i = \frac{1 - ie_1}{1 - (i-1)e_1},\ i = n_0 + 1, \dots, n_1 - 1, \tag{11}$$

and $r_{n_1} = e_{n_1} = e_{n_0+1}/[1 - (n_1 - 1)e_{n_0+1}]$. Therefore,

$$L_{n_1+1} = L_{n_0+1} e_1 = L_{n_0+1} \epsilon_1.$$

The same arguments generalize to arbitrary $j \geqslant 1$, and $L_{n_j+1} = L_{n_{j-1}+1}\epsilon_j$, which gives (10). Finally, using (11), one can easily find the value of L_{N+1} for arbitrary N, that is equation (9) for $n_j \leqslant N < n_{j+1}$. □

Theorem 1 implies in particular that, when e_1 is a rational in [0, 1), the length of the uncertainty interval remains constant after some iteration. Indeed, one can write $e_1 = [a_1, \ldots, a_m, \infty]$ (see equation (6)). This gives $\epsilon_{m+1} = 0$, $n_{m+1} = \infty$, and thus from equation (9)

$$L_{N+1} = L_1 \times \epsilon_1 \times \cdots \times \epsilon_m$$

for any $N \geqslant N_* = a_1 + \cdots + a_m$. Thus, for rational e_1, the algorithm terminates (degenerates) at a certain iteration N_*, where $N_* = a_1 + \cdots + a_m$ could also be defined as the first n such that $e_n = 1$. (Recall that we assume that at the last iteration we halve the uncertainty interval by making the two last observations at the same point, the midpoint of the interval.)

Because the termination point N_* is a function of e_1 we need to be careful in the specification of the sample size of the algorithm. If we set $N > N_*$ then there is no further improvement after N_* so that the per iteration rate will have declined. However, we have the option of deciding in advance that the number of iterations is $N = N_*$ and then controlling e_1 to produce $\min_{e_1} L_N(e_1)$, the minimal length of the uncertainty interval in N iterations. Fortunately, the case $N = N_*$ gives a nice expression for the value of $L_N(e_1)$ in terms of the so-called Farey tree (see definitions below and Corollary 1) and the value of e_1 where the minimum $\min_{e_1} L_N(e_1)$ is achieved (see Corollary 2).

Consider the mediant $\mathrm{med}(p/q,\ p'/q')$ of two fractions p/q and p'/q', $\mathrm{med}(p/q,\ p'/q') = (p + p')/(q + q')$, where p, p', q and q' are positive integers. One can easily check that for $p/q < p'/q'$, $\mathrm{med}(p/q,\ p'/q')$ always belongs to the interval $(p/q,\ p'/q')$.

The Farey tree (also called Brocot sequence; see Lagarias, 1992) $\mathcal{F}_n$ of order n is defined inductively as follows. $\mathcal{F}_0$ consists of two elements, zero and unity written as $0/1$ and $1/1$. Then at iteration n, for every pair $\{p/q,\ p'/q'\}$ of adjacent fractions in $\mathcal{F}_{n-1}$, their mediant $\mathrm{med}(p/q,\ p'/q')$ is added to the elements of $\mathcal{F}_{n-1}$. Thus,

$$\mathcal{F}_n = \mathcal{F}_{n-1} \bigcup \mathrm{med}\,(p/q,\ p'/q'),$$

where the union is taken over all adjacent pairs $\{p/q,\ p'/q'\}$ in $\mathcal{F}_{n-1}$. For example,

$$\mathcal{F}_1 = \left\{0, \tfrac{1}{2}, 1\right\}, \quad \mathcal{F}_2 = \left\{0, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, 1\right\}, \quad \mathcal{F}_3 = \left\{0, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{2}{5}, \tfrac{1}{2}, \tfrac{3}{5}, \tfrac{2}{3}, \tfrac{3}{4}, 1\right\}.$$

One can easily check (see Schroeder, 1991, p. 337 and Cornfeld *et al.*, 1982) that there are exactly 2^{n-1} Farey fractions of level n, i.e. elements in $\mathcal{F}_n \setminus \mathcal{F}_{n-1}$, and all of them have a finite continued fraction representation $p/q = [a_1, a_2, \ldots, a_k]$ satisfying $a_1 + \cdots + a_k = n$. Also, the number of elements in $\mathcal{F}_n$ is $|\mathcal{F}_n| = 2^n + 1$ and these elements, apart from zero and unity, have a continued fraction representation $[a_1, a_2, \ldots, a_k]$ satisfying $a_1 + \cdots + a_k \leqslant n$.

This gives a characterization of rational starting points e_1 such that the corresponding symmetric line-search algorithm stops at iteration n. Specifically, these points are exactly the points with a finite continued fraction representation $e_1 = [a_1, a_2, \ldots, a_k]$ with $a_1 + \cdots + a_k = n$ and therefore are exactly the Farey fractions of level n.

If we return to Fig. 2 and consider the function $L_n(e_1)$ for a given n, we observe that it is piecewise linear, with 2^{n-1} local minima at the Farey fractions of level n and $2^{n-1} + 1$ local maxima at 0, 1 and the Farey fractions of all levels smaller than n.

The following property gives the value of L_N when the algorithm degenerates.

COROLLARY 1 Let $[A_1, B_1] = [0, 1]$ and $e_1 \in [0, 1]$ be a rational number $e_1 = p/q$ with $\gcd(p, q) = 1$ and continued fraction expansion $e_1 = [a_1, a_2, \ldots, a_k]$, $\sum_{i=1}^{k} a_i = N_*$. Then $L_N(e_1) = 1/q$ for all $N \geqslant N_*$.

Proof. Consider the sequence $\epsilon_1, \epsilon_2, \ldots$ with $\epsilon_1 = e_1 = p/q$. Then

$$\epsilon_2 = \frac{1}{\varepsilon_1} - a_1 = \frac{q}{p} - a_1 = \frac{q - pa_1}{p} = \frac{p_2}{p_1},$$

where $p_1 = p$ and $p_2 = q - pa_1$ is some integer such that $p_2 < p_1$ and $\gcd(p_1, p_2) = 1$. (The fact that $p_2 < p_1$ is equivalent to the fact that $\varepsilon_2 < 1$.)

Analogously, for all $i \leqslant k$, $\epsilon_{i-1} = p_i/p_{i-1}$ implies $\epsilon_i = p_{i+1}/p_i$ with $p_{i+1} < p_i$ and $\gcd(p_{i+1}, p_i) = 1$. When $i = k-1$, $\epsilon_{k-1} = 1/a_k$ so that $p_k = 1$ and $p_{k-1} = a_k$.

According to (9), ϵ_{k-1} is the last term we need in the product $\prod_i \varepsilon_i$ for all $N \geqslant N_*$ (because $a_{k+1} = \infty$ and $\epsilon_k = 0$). The product itself equals

$$L_{N+1}(e_1) = \prod_{i=1}^{k-1} \varepsilon_i = \frac{p_1}{q}\frac{p_2}{p_1}\frac{p_3}{p_2}\cdots\frac{p_{k-1}}{p_{k-2}}\frac{1}{p_{k-1}} = \frac{1}{q}, \quad N \geqslant N_*.$$

□

The Fibonacci algorithm can easily be shown to be optimal among symmetric algorithms.

COROLLARY 2 (Optimality of the Fibonacci search)
For a fixed N, the minimum

$$\min_{e_1} L_N(e_1)$$

equals $1/F_{N+1}$ and is achieved at $e_1 = F_N/F_{N+1}$ or $e_1 = F_{N-1}/F_{N+1}$, where $F_1, F_2, \ldots$ is the Fibonacci sequence.

Proof. Without any loss of generality we restrict our attention to the case $e_1 > \frac{1}{2}$. As noted in Section 4, the function $L_N(e_1)$ is piecewise linear on $[\frac{1}{2}, 1]$ (see Fig. 2). It takes its minimum and maximum values at the rational $e_1 = p/q$ with $\gcd(p, q) = 1$ and continued fraction expansion $e_1 = [a_1, a_2, \ldots, a_k]$ such that $\sum_{i=1}^{k} a_i \leqslant N$.

According to Corollary 1, $L_N(e_1) = 1/q$ at these points. However, one of the properties of the Farey tree is that the maximum value of the denominator q among all the points p/q with $\gcd(p, q) = 1$ and continued fraction expansion $p/q = [a_1, a_2, \ldots, a_k]$ such that $\sum_{i=1}^{k} a_i \leqslant N$, is achieved when $p/q = F_N/F_{N+1}$. This property can easily be proved by induction (see also Schroeder, 1991, p. 339). □

5.3 *Asymptotic performance*

Using (9), we can easily evaluate the performance of any symmetric algorithm in terms of the asymptotic convergence rate R defined as follows:

$$R(e_1) = \limsup_{N\to\infty} [L_{N+1}(e_1)]^{\frac{1}{N}} = \limsup_{N\to\infty} \left[L_1 \prod_{n=1}^{N} r_n(e_1) \right]^{\frac{1}{N}}. \tag{12}$$

Because $L_1 < \infty$ this becomes

$$R(e_1) = \limsup_{N\to\infty} \left[\prod_{n=1}^{N} r_n(e_1) \right]^{\frac{1}{N}} .$$

In what follows we assume, without any loss of generality, that $L_1 = 1$.

The algorithm has *exponential convergence* if $R(e_1) < 1 - \alpha$ for some $\alpha > 0$. Corollary 3 shows how the asymptotic rate of convergence of the algorithm is related to the choice of e_1.

COROLLARY 3 The convergence of the asymmetric algorithm initialized at $e_1 \in (0, 1)$ is,

(i) sub-exponential if e_1 is a rational number (in this case, the algorithm is such that the two test points E_n, E'_n coincide at some iteration n).
(ii) exponential if e_1 is a badly approximable number, that is a number with bounded partial quotients,
(iii) sub-exponential for almost all values of e_1.

Proof.

(i) We have seen in Section 5.2 that, when e_1 is rational, $L_n(e_1)$ remains constant for n larger than some N, and therefore $R(e_1) = 1$.
(ii) When e_1 is a badly approximable number, one has $e_1 = [1, a_1, a_2, \dots]$ with $a_i < A$ for all i and some $A < \infty$. (Note that unlike the set of quadratic irrationals, the set of badly approximable numbers is uncountable. However, it still has zero Lebesgue measure.) From (10), we obtain

$$R(e_1) = \limsup_{j\to\infty} (\epsilon_1 \times \cdots \times \epsilon_j)^{1/n_j},$$

with n_j defined by (8). Therefore,

$$R(e_1) \leqslant \limsup_{j\to\infty} (\epsilon_1 \times \cdots \times \epsilon_j)^{1/(jA)}.$$

From the recurrence $\epsilon_{i+1} = \{1/\epsilon_i\}$, we obtain $\epsilon_i\epsilon_{i+1} < \varphi$. Indeed, the result is obvious for $\epsilon_i \leqslant \varphi$, and $\epsilon_i > \varphi$ implies $\epsilon_{i+1} = 1/\epsilon_i - 1 < \varphi$. Therefore,

$$R(e_1) \leqslant \varphi^{1/(2A)} < 1,$$

and the convergence is exponential.
(iii) Ergodic arguments can be used to show that for almost all values of e_1 the convergence is sub-exponential. Indeed, for almost all e_1, the dynamical system (2) has the invariant density of the Farey map given by (3), and the density of the sequence (r_n) is (4). These densities are not normalized; however, they can be used to construct the proportion of points falling in different intervals. Consider in particular the interval $\mathcal{I}_\delta = [1 - \delta, 1)$, $0 < \delta < \frac{1}{2}$, and let n_δ denote the number of points r_i, $i = 1 \dots, n$, in $\mathcal{I}_\delta$. Then n_δ/n tends to unity for any δ. This gives

$$\begin{aligned} R(e_1) &= \limsup_{n\to\infty} (r_1 \times \cdots \times r_n)^{1/n} \\ &> \limsup_{n\to\infty} [(1-\delta)^{n_\delta} (\tfrac{1}{2})^{n-n_\delta}]^{1/n} = 1 - \delta \end{aligned}$$

for any δ. Therefore $R(e_1) = 1$ for almost all e_1.

□

As a complement to the case (iii) above, notice that ergodic properties of the Gauss map imply (see Cornfeld *et al.*, 1982) that the limit

$$\rho'(e_1) = \lim_{j\to\infty} \frac{1}{j} \sum_{i=1}^{j} \log \epsilon_i$$

exists for almost all e_1 in (0, 1) and equals

$$\rho' = \rho'(e_1) = \frac{\pi^2}{12 \log 2} \simeq 1{\cdot}186\,569,$$

which is the Lyapunov exponent for the Gauss map. A consequence of this relation is that for almost all e_1 in (0, 1) for large N

$$\prod_{i=1}^{j} \epsilon_i \sim \exp(j\rho')$$

which, together with (9), implies $L_N \sim \exp(j\rho')$ where j is such that $n_j \leqslant N < n_{j+1}$.

Corollary 3(i) does not mean that symmetric algorithms have necessarily poor performance for finite N (see, for instance, the Fibonacci algorithm, for which $e_1 = F_N/F_{N+1} = [1, \ldots, 1]$ with N unities, where (F_N) is the Fibonacci sequence).

A particular case of Corollary 3(ii) is when e_1 is a quadratic irrational, i.e. solution of a quadratic equation with integer coefficients. Its continued fraction expansion is then periodic, starting with some n (and only quadratic irrationals have this property; see Rockett & Szüsz, 1992). Let the period be $b_1, \ldots, b_k$, that is

$$e_1 = [1, a_1, \ldots, a_n, \underbrace{b_1, \ldots, b_k}, \underbrace{b_1, \ldots, b_k}, \ldots].$$

Then, (10) implies that the ergodic rate of convergence of the corresponding algorithm is

$$R(e_1) = (\epsilon_1' \times \ldots \epsilon_k')^{1/(b_1+\cdots+b_k)},$$

where

$$\begin{aligned}
\epsilon_1' &= [\underbrace{b_1, \ldots, b_k}, \underbrace{b_1, \ldots, b_k}, \ldots] \\
\epsilon_2' &= [\underbrace{b_2, \ldots, b_k, b_1}, \underbrace{b_2, \ldots, b_k, b_1}, \ldots] \\
&\vdots \\
\epsilon_k' &= [\underbrace{b_k, b_1, \ldots, b_{k-1}}, \underbrace{b_k, b_1, \ldots, b_{k-1}}, \ldots].
\end{aligned}$$

All the ϵ_i' values are smaller than unity and the convergence is therefore exponential.

TABLE 1

Values of $R(e_1)$ for different choices of quadratic irrationals e_1

e_1	$R(e_1)$
$\varphi=[1,1,\ldots]$	$\varphi\simeq 0{\cdot}61803$
$2-\sqrt{2}=[1,1,2,2,2,\ldots]$	$\sqrt{\sqrt{2}-1}\simeq 0{\cdot}64359$
$\sqrt{2}/2=[1,2,2,2,\ldots]$	$\sqrt{\sqrt{2}-1}\simeq 0{\cdot}64359$
$\sqrt{3}-1=[\underbrace{1,2},\underbrace{1,2},\ldots]$	$(\sqrt{3}-2)^{1/3}\simeq 0{\cdot}64469$
$\sqrt{3}/3=[1,\underbrace{1,2},\underbrace{1,2},\ldots]$	$(\sqrt{3}-2)^{1/3}\simeq 0{\cdot}64469$
$\sqrt{10}/2-1=[\underbrace{1,1,2},\underbrace{1,1,2},\ldots]$	$(\sqrt{10}-3)^{1/4}\simeq 0{\cdot}63469$
$(\sqrt{10}-1)/3=[\underbrace{1,2,1},\underbrace{1,2,1},\ldots]$	$(\sqrt{10}-3)^{1/4}\simeq 0{\cdot}63469$

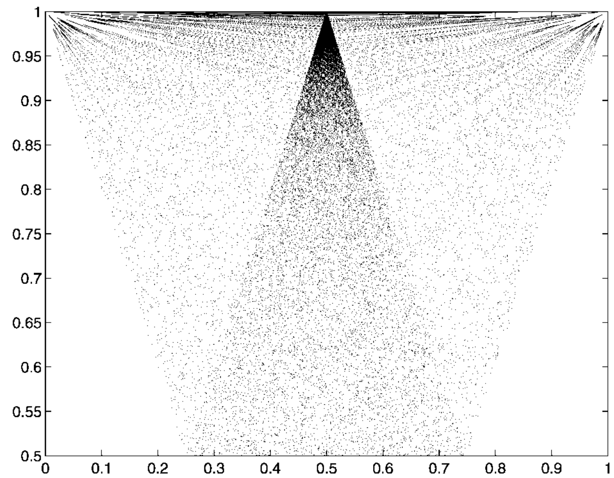

FIG. 6. Typical chaotic sequence of iterates (x_n, r_n), $n=1,\ldots,100\,000$, in a symmetric algorithm.

A famous example is the GS algorithm, for which $e_1=\varphi=[1,1,\ldots]$, which gives $R(e_1)=\varphi\simeq 0{\cdot}618\,03$. More generally, the same value for R is obtained when e_1 is a so-called noble number, that is, when e_1 has a continued fraction expansion $e_1=[a_1,\ldots,a_n,1,1,\ldots]$ all ending in unity (see Schroeder, 1991, p. 392). Some other examples are given in Table 1.

Figure 6 presents a plot of a typical sequence of iterates (x_n, r_n), $n=1,\ldots,100\,000$ for a non-periodic symmetric algorithm, corresponding to Corollary 3(iii). This propery has the important consequence that a direct implementation of a symmetric algorithm, based on the application of the rule $E'_n=A_n+B_n-E_n$ yields sub-exponential convergence ($R=1$) due to numerical inaccuracies. In particular, this is the case for the GS algorithm; hence the usual recommendation to use the implementation (1). Note that among symmetric algorithms, $R(e_1)$ is minimum when $e_1=\varphi$, the Golden Section. This follows from

Corollary 2 and the fact that

$$\frac{L_n(\varphi)}{L_n(F_n/F_{n+1})} \to \frac{2+\varphi}{\sqrt{5}}, \quad n \to \infty,$$

(see Section 2.2).

REFERENCES

BOWEN, R. 1979 Invariant measures for Markov maps of the interval. *Commun. Math. Phys.*, **69**, 1–17.

CORNFELD, I., FOMIN, S. V., & SINAI, YA. G. 1982 *Ergodic Theory*. New York: Springer.

DU, D. Z. & HWANG, F. K. 1993 *Combinatorial Group Testing and its Applications*. Singapore: World Scientific.

KIEFER, J. 1957 Optimum sequential search and approximation methods under minimum regularity assumptions. *J. Soc. Indust. Appl. Math.*, **5**, 105–136.

LAGARIAS, J. C. 1992 Number theory and dynamical systems. *Proc. Symp. Appl. Math.*, **46**, 35–72.

PRONZATO, L., WYNN, H. P., & ZHIGLJAVSKY, A. A. 1995 Improving on golden-section optimisation for locally symmetric functions. *MODA4—Advances in Model-Oriented Data Analysis, Spetses (Greece), June 1995*. (C. P. Kitsos & W. G. Müller eds). Heidelberg: Physica, pp. 285–297.

PRONZATO, L., WYNN, H. P., & ZHIGLJAVSKY, A. A. 1997 Stochastic analysis of convergence via dynamic representation for a class of line-search algorithms. *Comb. Probab. Comput.*, **6**, 205–229.

PRONZATO, L., WYNN, H. P., & ZHIGLJAVSKY, A. A. 1998 A generalised golden-section algorithm for line-search. *IMA J. Math. Control Inform.*, **15**, 185–214.

PRONZATO, L., WYNN, H. P., & ZHIGLJAVSKY, A. A. 1999 Finite sample behaviour of an ergodically fast line-search algorithm. *Comput. Optim. Appl.*, **14**, 75–86.

ROCKETT, A. M. & SZÜSZ, P. 1992 *Continued Fractions*. Singapore: World Scientific.

SCHOISENGEIER, J. 1990 An asymptotic expansion for $\sum_{n \leqslant N}\{n\alpha + \beta\}$. *Number-Theoretic Analysis*. (H. Hlawka & R. F. Tichy eds). Heidelberg: Springer, pp. 199–205.

SCHROEDER, M. 1991 *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. New York: Freeman and Company.

WYNN, H. P. & ZHIGLJAVSKY, A. A. 1993a Chaotic behaviour of search algorithms. *Acta Applicandae Mathematicae*, **32**, 123–156.

WYNN, H. P. & ZHIGLJAVSKY, A. A. 1993b Chaotic behaviour of search algorithms: introduction. *Model-Oriented Data Analysis III, Proceedings MODA3, St Petersburg, May 1992*. (W. G. Müller, H. P. Wynn & A. A. Zhigljavsky eds). Heidelberg: Physica, pp. 199–211.