Proc. Estonian Acad. Sci. Phys. Math., 2007, **56**, 4, 299–309

Evolution strategies in optimization problems

Pedro A. F. Cruz and Delfim F. M. Torres

Department of Mathematics, University of Aveiro, 3810-193 Aveiro, Portugal; {pedrocruz,delfim}@ua.pt

Received 23 February 2007, in revised form 2 July 2007

Abstract. Evolution strategies are inspired in biology and form part of a larger research field known as evolutionary algorithms. Those strategies perform a random search in the space of admissible functions, aiming to optimize some given objective function. We show that simple evolution strategies are a useful tool in optimal control, permitting one to obtain, in an efficient way, good approximations to the solutions of some recent and challenging optimal control problems.

Key words: random search, Monte Carlo method, evolution strategies, optimal control, discretization.

1. INTRODUCTION

Evolution strategies (ES) are algorithms inspired in biology, with publications dating back to 1965 by separate authors H. P. Schwefel and I. Rechenberg (cf. [1]). Evolution strategies form part of a larger area called evolutionary algorithms that perform a random search in the space of solutions aiming to optimize some objective function. It is common to use biological terms to describe these algorithms. Here we make use of a simple ES algorithm known as the (μ, λ)-ES method [1], where μ is the number of *progenitors* and λ is the number of generated approximations, called *offsprings*. Progenitors are *recombined* and *mutated* to produce, at each generation, λ *offsprings* with innovations sampled from a multivariate normal distribution. The variance can also be subject to mutation, meaning that it is part of the *genetic code* of the population. Every solution is evaluated by the objective function and one or some of them are selected to be

Partially presented at the 5th Junior European Meeting on Control and Information Technology (JEM'06), September 20–22, 2006, Tallinn, Estonia.

the next *progenitors*, allowing the search to go on, stopping when some criteria are met. In this paper we use a recent convergence result proved by A. Auger in 2005 [2]. The log-linear convergence is achieved for the optimization problems we investigate here, and depends on the number λ of search points.

Usually optimal control problems are approximately solved by means of numerical techniques based on the gradient vector or the Hessian matrix [3]. Compared with these techniques, ES provide easier computer coding because they only use measures from a discretized objective function. The first work combining these two research fields (ES and optimal control) was published by D. S. Szarkowicz in 1995 [4], where the Monte Carlo method (an algorithm with the same principle as ES) is used to find an approximation to the classical brachistochrone problem. In the late 1990s, B. Porter and his collaborators showed how ES can be used to synthesize optimal control policies that minimize manufacturing costs while meeting production schedules [5]. The use of ES in control has grown during the last ten years, and is today an active and promising research area. Recent results, showing the power of ES in control, include Hamiltonian synthesis [6], robust stabilization [7], and optimization [8]. Very recently it has also been shown that the theory of optimal control provides insights that permit developing more effective ES algorithms [9].

In this work we are interested in two classical problems of the calculus of variations: the 1696 brachistochrone problem and the 1687 Newton's aerodynamical problem of minimal resistance (see, e.g., [10]). These two problems, although classical, are a source of strong current research on optimal control and provide many interesting and challenging open questions [11,12]. We focus our study on the brachistochrone problem with restrictions proposed by A. G. Ramm in 1999 [13], for which some questions still remain open (see some conjectures in [13]); and on a generalized aerodynamical minimum resistance problem with non-parallel flux of particles, recently studied by Plakhov and Torres [11,14]. Our results show the effectiveness of ES algorithms for this class of problems and motivate further work in this direction in order to find the (yet) unknown solutions to some related problems, as the ones formulated in [15].

2. PROBLEMS AND SOLUTIONS

All the problems we are interested in share the same formulation:

$$\min T[y(\cdot)] = \int_{x_0}^{x_f} L(x, y(x), y'(x))\mathrm{d}x$$

on some specified class of functions, where $y(\cdot)$ must satisfy some given boundary conditions (x_0, y_0) and (x_f, y_f).

We consider a simplified (μ, λ)-ES algorithm where we put $\mu = 1$, meaning that on each generation we keep only one *progenitor* to generate other candidate solutions, and set $\lambda = 10$ meaning we generate 10 candidate solutions called *offsprings* (this value appears as a reference value in the literature). Also, the algorithm uses an individual and constant σ^2 variance on each coordinate, which is fixed to a small value related to the desired precision. The number of iterations was $100\,000$ and σ^2 was tuned for each problem. We got convergence in useful time. The simplified $(1, 10)$-ES algorithm goes as follows:

1. set an equal-spaced sequence of n points $\{x_0, \ldots, x_i, \ldots, x_f\}$, where $i = 1, \ldots, n-2$; x_0 and x_f are kept fixed (given boundary conditions);
2. generate a randomly piecewise linear function $y(\cdot)$ that approximates the solution, defined by a vector $y = \{y_0, \ldots, y_i, \ldots, y_f\}$, $i = 1, \ldots, n-2$; transform y in order to satisfy the boundary conditions y_0 and y_f and the specific problem restrictions on y, y' or y'';
3. perform the following steps a fixed number N of times:
   (a) based on y find λ new candidate solutions Y^c, $c = 1, \ldots, \lambda$, where each new candidate is produced by $Y^c = y + \mathrm{N}(0, \sigma^2)$, where $\mathrm{N}(0, \sigma^2)$ is a vector of random perturbations from a normal distribution; transform each Y^c to obey boundary conditions y_0 and y_f and other problem restrictions on y, y' or y'';
   (b) determine $T^c := T[Y^c]$, $c = 1, \ldots, \lambda$, and choose the new $y := Y^c$ as the one with minimum T^c.

In each iteration the best solution must be kept because (μ, λ)-ES algorithms do not keep the best solution from iteration to iteration.

The next subsections contain a description of the studied problems, respective solutions, and the approximations found by the described algorithm.

2.1. The classical brachistochrone problem, 1696

Problem statement. The brachistochrone problem consists in determining the curve of minimum time when a particle starting at a point $A = (x_0, y_0)$ of a vertical plane goes to a point $B = (x_1, y_1)$ in the same plane under the action of the gravity force and with no initial velocity. According to the energy conservation law $\frac{1}{2}mv^2 + mgy = mgy_0$, one easily deduces that the time a particle needs to reach B starting from the point A along curve $y(\cdot)$ is given by

$$T[y(\cdot)] = \frac{1}{\sqrt{2g}} \int_{x_0}^{x_1} \sqrt{\frac{1 + (y')^2}{y_0 - y}} \mathrm{d}x, \tag{1}$$

where $y(x_0) = y_0$, $y(x_1) = y_1$, and $y \in C^2(x_0, x_1)$. The minimum to (1) is given by the famous Cycloid

$$\gamma : \begin{cases} x = x_0 + \dfrac{a}{2}(\theta - \sin\theta), \\ y = y_0 - \dfrac{a}{2}(1 - \cos\theta), \end{cases}$$

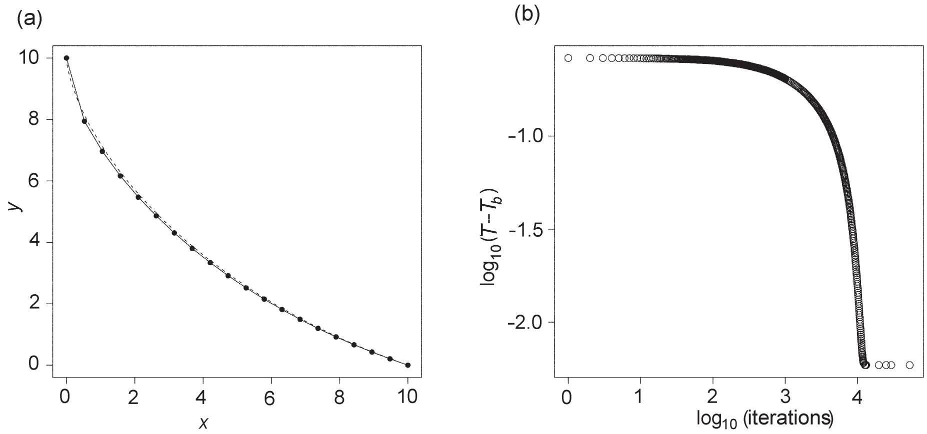

Fig. 1. The brachistochrone problem and approximate solution. (a) The continuous line with dots is the piecewise approximate solution, the dashed line the optimal solution. (b) Logarithm of iterations vs. logarithmic distance to the minimum value of (1).

with $\theta_0 \le \theta \le \theta_1$, θ_0 and θ_1 being the values of θ in the starting and ending points (x_0, y_0) and (x_1, y_1). The minimum time is given by $T = \sqrt{a/(2g)}\theta_1$, where parameters a and θ_1 can be determined numerically from boundary conditions.

Results and implementation details. Consider the following three curves and the time a particle needs to go from A to B through them:

T_b: the brachistochrone for the problem with $(x_0, x_1) = (0, 10)$, $(y_0, y_1) = (10, 0)$ has parameters $a \simeq 5.72917$ and $\theta_1 = 2.41201$; the time is $T_b \simeq 1.84421$;

T_{es}: a piecewise linear function with 20 segments shown in Fig. 1a was found by ES; the time is $T_{es} = 1.85013$;

T_o: a piecewise linear function with 20 segments defined over the brachistochrone; the time is $T_o = 1.85075$.

From Fig. 1a one can see that the piecewise linear solution is made of points that are not over the brachistochrone, because that is not the best solution for piecewise functions. We use $\sigma = 0.01$ (see Appendix for cpu-times). Figure 1b shows that a little more than 10 000 iterations are needed to reach a good solution for the 20 line segment problem.

2.2. The brachistochrone problem with restrictions, 1999

Problem statement. Ramm [13] presents a conjecture about a brachistochrone problem over the set S of convex functions y (with $y''(x) \ge 0$ a.e.) and $0 \le y(x) \le y_0(x)$, where y_0 is a straight line between $A = (0, 1)$ and $B = (b, 0)$, $b > 0$. Up to a constant, the functional to be minimized is formulated as in (1):

$$T[y(\cdot)] = \int_0^b \frac{\sqrt{1 + (y')^2}}{\sqrt{1 - y}} \mathrm{d}x \, .$$

Let P be the line connecting AO and OB, where $O = (0,0)$; P_{br} be the polygonal line connecting AC and CB, $C = (\pi/2, 0)$. Then, $T_0 := T(y_0) = 2\sqrt{1+b^2}$, $T_P := T(P) = 2 + b$, $T(P_{br}) = \sqrt{4+\pi^2} + b - \pi/2$. Let the brachistochrone be y_{br}. The following inequalities, for each $y \in S$, hold [13]:

1. if $0 < b < 4/3$, then $T(y_{br}) \leq T(y) < T_P$;
2. if $4/3 \leq b \leq \pi/2$, then $T(y_{br}) \leq T(y) \leq T_0$;
3. if $b > \pi/2$, then $T(P_{br}) < T(y) \leq T_0$.

The classical brachistochrone solution holds for cases 1 and 2 only. For the third case, Ramm has conjectured that the minimum time curve is composed by the brachistochrone between $(0,1)$ and $(\pi/2, 0)$ and then by the horizontal segment between $(\pi/2, 0)$ and $(x_f, 0)$.

Results and implementation details. We study the problem with $b = 2$. Our results give force to Ramm's conjecture mentioned above for case 3. We compare three descendant times:

T_{br}: the conjectured solution in continuous time takes $T_{br} = \sqrt{\alpha/9.8}\theta_f + (b - \pi/2)/\sqrt{2 * 9.8} = 0.8066$;

T_{es}: the 20 segment piecewise linear solution found by ES needs $T_{es} = 0.8107$;

T_o: the 20 segment piecewise linear solution with points over the conjectured solution needs $T_o = 0.8111$.

Previous values and Fig. 2 permit us to make conclusions similar to the ones obtained for the pure brachistochrone problem (§2.1). We use $\sigma = 0.001$ (see Appendix for cpu-times). Figure 2b shows that less than 10 000 iterations are needed to reach a good solution.

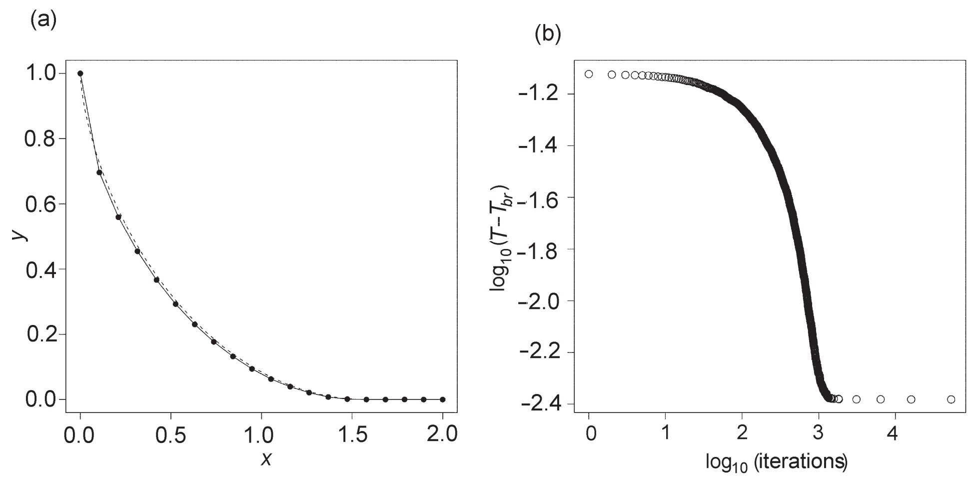

Fig. 2. The solution and approximate solution conjectured by Ramm [13]. (a) The continuous line with dots is the obtained approximated solution, the dashed line Ramm's conjectured solution. (b) Logarithm of iterations vs. logarithmic distance to minimum integral value.

2.3. Newton's minimum resistance, 1687

Problem statement. Newton's aerodynamical problem consists in determining the minimum resistance profile of a body of revolution moving at constant speed in a rare medium of equally spaced particles that do not interact with each other. Collisions with the body are assumed to be perfectly elastic. Formulation of this problem is: minimize

$$R[y(\cdot)] = \int_0^r \frac{x}{1 + \dot{y}(x)^2} \mathrm{d}x,$$

where $0 \le x \le r$, $y(0) = 0$, $y(r) = H$, and $y'(x) \ge 0$. The solution is given in parametric form:

$$x(u) = 2\lambda u, \quad y(u) = 0, \quad \text{for } u \in [0, 1];$$

$$x(u) = \frac{\lambda}{2}\left(\frac{1}{u} + 2u + u^3\right), \quad y(u) = \frac{\lambda}{2}\left(-\log u + u^2 + \frac{3}{4}u^4\right) - \frac{7\lambda}{8},$$
$$\text{for } u \in [1, u_{\max}].$$

Parameters λ and $u_{\max}$ are obtained by solving $x(u_{\max}) = r$ and $y(u_{\max}) = H$.

Results and implementations details. For $H = 2$ we have:

R_{newton}: the exact solution has resistance $R_{newton} = 0.0802$;

R_{es}: the 20 segment piecewise linear solution found by ES has $R_{es} = 0.0809$;

R_o: the 20 segment piecewise linear solution with points over the exact solution leads to $R_o = 0.0808$.

Newton's problem turns out to be more complex than previously studied brachistochrone problems. The trial-and-error method was needed in order to find a useful σ^2 value. For example, using $\sigma = 0.001$, our algorithm seems to stop in

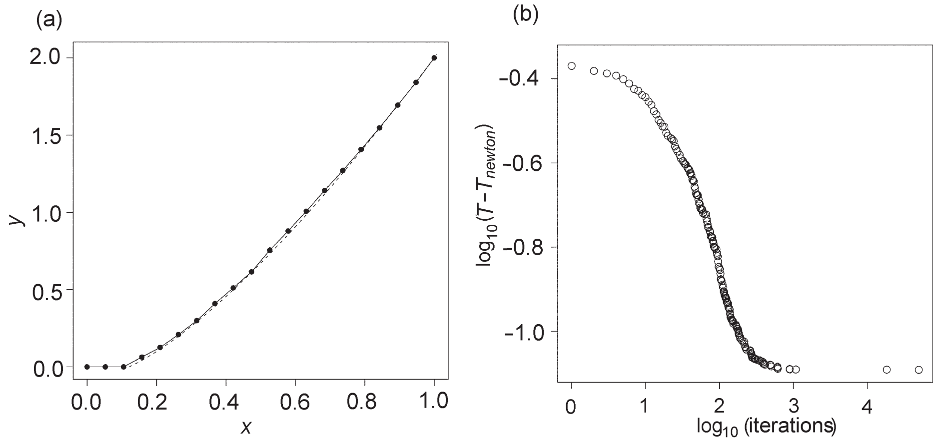

Fig. 3. Optimal solution to Newton's problem and approximation. (a) The continuous line with dots is the obtained approximation, the dashed line the optimal solution. (b) Logarithm of iterations vs. logarithmic distance to minimum integral value.

some local minimum. In Fig. 3 an approximate solution with $\sigma = 0.01$ is shown. We have also observed that changing the starting point causes minor differences in the approximate solution. The achieved ES solution should be better since R_o is better than R_{es}. One possible explanation for this fact is that we are using 20 x_i fixed points and the optimal solution has a break point at $x = 2\lambda$. We use $\sigma = 0.01$ (see Appendix for cpu-times). Figure 3b shows that less than 1000 iterations are needed to reach a good solution.

2.4. Newton's problem with temperature, 2005

Problem statement. The problem consists in determining the body of minimum resistance, moving with constant velocity in a rarefied medium of chaotically moving particles with velocity distributions assumed to be radially symmetric in the Euclidean space $\mathbb{R}^d$. This problem was posed and solved in 2005–2006 by Plakhov and Torres [11,14]. It turns out that the two-dimensional problem ($d = 2$) is richer than the three-dimensional one, having five possible types of solutions when the velocity of the moving body is not "too slow" or "too fast" compared with the velocity of particles.

The pressure at the body surface is described by two functions: in the front of the body the flux of particles causes resistance, in the back the flux causes acceleration. We consider functions found in [14], where the two flux functions p_+ and p_- are given by $p_+(u) = \frac{1}{1+u^2} + 0.5$ and $p_-(u) = \frac{0.5}{1+u^2} - 0.5$. We also consider a body of fixed radius 1. The optimal solution depends on the body height h: the front solution is denoted by f_{h_+}, which depends on some appropriate front height h_+; and the solution for the rear is denoted by f_{h_-}, depending on some appropriate height h_-. Optimal solutions f_{h_+} and f_{h_-} are obtained:

$$f_{h_+} = \min_{f_h} \quad \mathbb{R}_+(f_h) = \int_0^1 p_+(f_h'(t))\mathrm{d}t$$

and

$$f_{h_-} = \min_{f_h} \mathbb{R}_-(f_h) = \int_0^1 p_-(f_h'(t))\mathrm{d}t.$$

Then, the body shape is determined by minimizing

$$R(h) = \min_{h_+ + h_- = h} \left(\mathbb{R}_+(f_{h_+}') + \mathbb{R}_-(f_{h_-}')\right).$$

The solution can be of five types ($d = 2$). From functions p_+ and p_- one can determine constants u_+^0, u_*, u_-^0, and h_-. Then, depending on the choice of the height h, the theory developed in [11,14] asserts that the minimum resistance body is:

1. a trapezium if $0 < h < u_+^0$;
2. an isosceles triangle if $u_+^0 \leq h \leq u_*$;

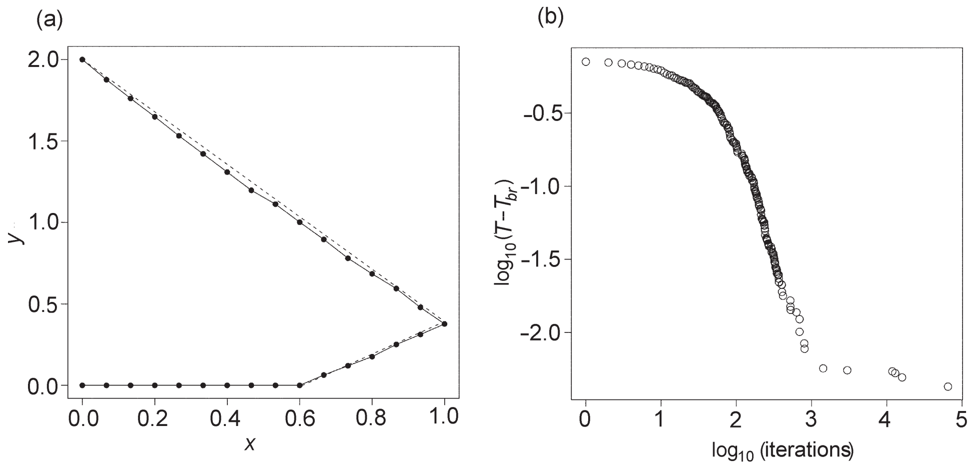

Fig. 4. 2D Newton-type problem with temperature. (a) The continuous line with dots is the obtained approximation, the dashed line the optimum. (b) Logarithm of iterations vs. logarithmic distance to minimum integral value.

3. the union of a triangle and a trapezium if $u_* < h < u_* + u_-^0$;
4. if $h \geq u_* + u_-^0$, the solution depends on h_- and can be a union of two isosceles triangles with a common base with heights h_+ and h_- or the union of two isosceles triangles and a trapezium;
5. a combination of a triangle, trapezium, and other triangle, depending on some other particular conditions (cf. [11]).

Results and implementation details. We illustrate the use of ES algorithms for $h = 2$. Following section 4.1 of [11], we have $u_* \simeq 1.60847$ and $u_-^0 = 1$, so this is case 3 above: $u_* < h < u_* + u_-^0$. The resistance values are:
R_{pd}: the exact solution has resistance $R_{pd} = 0.681$;
R_{es}: the 31 segment piecewise linear solution found by ES has $R_{es} = 0.685$.

Similar to the classical problem of Newton (§2.3), some hand search for the parameter σ^2 was needed. We use $\sigma = 0.01$ and piecewise approximation with 31 equal-spaced segments in xx (see Appendix for cpu-times). Figure 4b shows that only a little more than 1000 iterations are needed to reach a good solution.

3. CONCLUSIONS AND FUTURE DIRECTIONS

Our main conclusion is that a simple ES algorithm can be effectively used as a tool to find approximate solutions to some optimization problems. In the present work we report simulations that motivate the use of ES algorithms to find good approximate solutions to brachistochrone-type and Newton-type problems. We illustrate our approach with the classical problems and with some recent and still challenging problems. More precisely, we considered the 1696 brachistochrone problem (B); the 1687 Newton's aerodynamical problem of minimal resistance (N); a recent brachistochrone problem with restrictions (R) studied by Ramm in 1999, and where some open questions still remain [13]; and finally a generalized aerodynamical minimum resistance problem with non-parallel flux of particles (P), recently studied by Plakhov and Torres [11,14] and which gives rise to other interesting questions [15].

We argue that the approximated solutions we have found by the ES algorithm are of good quality. We give two reasons. First, for the brachistochrone and Ramm problems the functional value for the ES approximation was better than the linear interpolation over the exact solution, showing that the ES algorithm is capable of good precision. The second reason is the low relative error $r(T_Y, T_y)$ between the functional over the exact solution T_y and the approximate solution T_Y, as shown in Table 1.

In Table 1, y_k are points over the exact solution of the problem and Y_k are points from the piecewise approximation. We note that $\max|Y_k - y_k|$ need not be zero because the best continuous solution and the best linear solution cannot be superposed.

Evolution strategies algorithms use computers in an intensive way. For brachistochrone-type and Newton-type problems, and nowadays computing power, few minutes of simulation (or less) were enough on an interpreted language (see Appendix).

More research is needed to tune algorithms of this kind and obtain more accurate solutions. Special attention must be paid to qualifying an obtained ES approximation: Is it a minimum of the energy function? Is it local or global? Another question is computer efficiency. Waiting for few minutes in recent computers is not bad, but can we improve the running times?

Concerning the accuracy, several new ES algorithms have been proposed. These algorithms can tune σ values and use generated second-order information that can influence the precision and time needed. Also the use of random xx points (besides y piecewise linear solutions) should be investigated.

We believe that the simplicity of the technique considered in the present work can help in the search of solutions to some open problems in optimal control. This is under investigation and will be addressed elsewhere.

Table 1. Performance achieved for problems (B), (R), (N), and (P)

Problem	$\max\|Y_k - y_k\|$	$r(T_Y, T_y)$	Problem	$\max\|Y_k - y_k\|$	$r(T_Y, T_y)$
(B)	0.15	0.001	(N)	0.08	0.01
(R)	0.09	0.003	(P)	0.07	0.001

ACKNOWLEDGEMENTS

The authors were supported by the R&D unit *Centre for Research on Optimization and Control* (CEOC) from the Portuguese Foundation for Science and Technology, cofinanced by the European Community fund FEDER/POCI 2010.

APPENDIX

HARDWARE AND SOFTWARE

The code developed for this work can be freely obtained from the first author's web page, at `http://www.mat.ua.pt/jpedro/evolution/`.

In most of our investigations few minutes were sufficient for getting a good approximation for all the considered problems, even using a code style prone to humans rather than machines (code was done concerning clearness of concepts rather than the execution speed). Our simulations used a Pentium 4 CPU 3 GHz, running Debian Linux `http://www.debian.org`. The language was R [[16]], chosen because it is a fast interpreted language, numerically oriented to statistics and freely available.

The cpu-times in Table 2 were obtained with the command

```
time R CMD BATCH problem.R
```

where `time` keeps track of the cpu used and `R` calls the interpreter. The times are rounded and the last column estimates the time for the first good solution.

We note that the per iteration "step" was $\sigma = 0.001$ in the brachistochrone (-type) problems and $\sigma = 0.01$ for the Newton(-type) problems. Using a compiled language like C, one can certainly improve times by several orders of magnitude.

Table 2. Cpu-times obtained

Problem	Section	100 000 iterations	"Good solution" at
Brachistochrone	§2.1	10 min	1 min
Ramm conjecture	§2.2	10 min	1 min
Newton	§2.3	9 min	10 s
Plakhov and Torres	§2.4	14 min	10 s

REFERENCES

1. Beyer, H.-G. and Schwefel, H.-P. Evolution strategies – a comprehensive introduction. *Nat. Comput.*, 2002, **1**, 3–52.
2. Auger, A. Convergence results for the $(1, \lambda)$-SA-ES using the theory of ϕ-irreducible Markov chains. *Theor. Comput. Sci.*, 2005, **334**, 35–69.
3. Arnăutu, V. and Neittaanmäki, P. *Optimal Control from Theory to Computer Programs*. Kluwer Acad. Publ., Dordrecht, 2003.
4. Szarkowicz, D. S. Investigating the brachistochrone with multistage Monte Carlo method. *Int. J. Systems Sci.*, 1995, **26**, 233–243.
5. Porter, B. and Merzougui, T. Evolutionary synthesis of optimal control policies for manufacturing systems. In *Emerging Technologies and Factory Automation Proceedings, 1997. ETFA '97, 6th International Conference on 9–12 Sept*. 1997, 304–309.
6. Raimúndez Álvarez, J. C. Port controller Hamiltonian synthesis using evolution strategies. In *Dynamics, Bifurcations, and Control (Kloster Irsee, 2001)*. Lecture Notes in Control and Inform. Sci., 2002, 273, 159–172.
7. Raimúndez, C. Robust stabilization for the nonlinear benchmark problem (TORA) using neural nets and evolution strategies. In *Nonlinear Control in the Year 2000, Vol. 2 (Paris)*. Lecture Notes in Control and Inform. Sci., 2001, 259, 301–313.
8. Beielstein, T., Ewald, C.-P. and Markon, S. Optimal elevator group control by evolution strategies. *Lecture Notes Comput. Sci.*, 2003, 1963–1974.
9. Aktan, B., Greenwood, G. W. and Shor, M. H. Using optimal control principles to adapt evolution strategies. In *2006 IEEE Congress on Evolutionary Computation, Vancouver, Canada, July 16–21*. 2006, 291–296.
10. Tikhomirov, V. M. *Stories About Maxima and Minima*. Amer. Math. Soc., Providence, RI, 1990 (translated from the 1986 Russian original by Abe Shenitzer).
11. Plakhov, A. Yu. and Torres, D. F. M. Newton's aerodynamic problem in media of chaotically moving particles. *Sb. Math.*, 2005, **196**, 885–933.
12. Sussmann, H. J. and Willems, J. C. The brachistochrone problem and modern control theory. In *Contemporary Trends in Nonlinear Geometric Control Theory and Its Applications (México City, 2000)*. World Sci. Publ., River Edge, NJ, 2002, 113–166.
13. Ramm, A. G. Inequalities for brachistochrone. *Math. Inequal. Appl.*, 1999, **2**, 135–140.
14. Torres, D. F. M. and Plakhov, A. Yu. Optimal control of Newton-type problems of minimal resistance. *Rend. Semin. Mat. Univ. Politec. Torino*, 2006, **64**, 79–95.
15. Plakhov, A. Yu. Billiards in unbounded domains that reverse the direction of motion of the particles. *Uspekhi Mat. Nauk*, 2006, **61**, 183–184.
16. R (Development Core Team). *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0, http://www.R-project.org.

Evolutsioonistrateegiad optimeerimisprobleemides

Pedro A. F. Cruz ja Delfim F. M. Torres

Evolutsioonistrateegiad on osa laiemast uurimisvaldkonnast, mis on tuntud evolutsiooniliste algoritmide nime all. Nimetatud strateegiad teostavad juhuslikku otsingut lubatavate funktsioonide ruumis eesmärgiga optimeerida mõnd etteantud sihifunktsiooni. On näidatud, et lihtsad evolutsioonistrateegiad on kasulikud mitmetes optimaaljuhtimise keerukates probleemides, võimaldades efektiivselt leida häid lahendite aproksimatsioone.