

A GENERAL-PURPOSE GLOBAL OPTIMIZER: IMPLEMENTATION AND APPLICATIONS

Luc PRONZATO and Eric WALTER

Laboratoire des Signaux et Systèmes, C.N.R.S.-E.S.E., 91190 Gif sur Yvette, France

Alain VENOT and Jean-Francois LEBRUCHEC

F.R. Synthèse et Analyse d'Images Médicales, Hôpital Cochin, 75674 Paris Cedex 14, France, and INSERM 4194. UER Pitie Salpêtrière, Paris, France

This paper, written from a user stand-point, advocates the Adaptive Random Search strategy as an efficient tool for global optimization. First is presented a brief overview of the various types of methods available in the literature for global optimization, and practical advantages of the random search approach are advanced. Some modifications, which were found to improve the efficiency and versatility of the method, and a detailed description of the practical implementation of the resulting algorithm are presented. The routine is used first to treat seven test-cases from the literature for comparison purposes. Then two examples are treated related to automatic control theory. The first one is a parameter estimation problem, the second one a control problem. Finally a practical application of the method to automated registration in medical nuclear imagery is presented.

1. Introduction

Optimization problems occur in an ever increasing number of fields. In most cases one is interested in finding the global extremum of a given criterion with respect to several parameters. Many methods have been proposed for this purpose in the literature (see e.g. [7,8]). Their efficiency is usually justified by examples only, because of a definite lack of theoretical results concerning their rates of convergence. Since authors most often limit themselves to presenting results of their own methods, one is often at a loss when one has to choose a method to use for practical applications.

In this paper, written from a user's stand-point, we advocate an improved version of a random search method proposed by Bekey et al. [1,2,18], which we have found to be simple, reliable and powerful on a number of practical and test-case problems. The paper is organized as follows. Section 2 presents a brief overview of the various types of methods available in the literature for global optimization and explains the practical ad-

vantages of the random search approach [21]. Section 3 is devoted to the strategy proposed by Bekey et al. and to several ideas that we have found on our examples to improve its performance. Enough details are given for an easy implementation of this method on a computer. Section 4 first describes the behavior of the resulting algorithm when applied to several test-cases from the literature. This allows a comparison with other algorithms to be made. Then two examples are treated relating to the field of automatic control theory.

Finally Section 5 reports a practical application of the method to automated registration of medical images, where it has proved extremely valuable.

2. Overview of global optimization methods

The most commonly used algorithms for minimizing a criterion J with respect to a vector θ of parameters proceed by iterative refinements. If the criterion is sufficiently regular, its expansion in

Taylor series can be used to produce efficient algorithms which converge to a local minimum of the criterion. (This is the case for gradient or Newton methods for instance.) When the criterion is unimodal the located minimum is global and the procedure is satisfactory. In the case of a multimodal criterion, trajectory-description methods [12,15,28] can be used to generate trajectories in the parameter space which (hopefully) pass through all the minima of the criterion and thus succeed in locating the global minimum. The most commonly used trajectory-description method is due to Branin [10] and consists of changing the sign of the displacement in Newton's algorithm to make it diverge each time a local minimum has been found. The practical implementation of Branin's method raises many numerical and theoretical difficulties (especially when the Hessian matrix tends to become singular or when the trajectory of the parameter vector leaves the admissible domain in the parameter space) and it appears totally unworkable for optimization with more than three parameters. Moreover the method applies only to sufficiently regular criteria, since j must be at least twice continuously differentiable with respect to θ . For these reasons trajectory methods do not qualify for being used as general-purpose global optimizers.

Another idea, very commonly used because of its simplicity, is to divide the initial search domain into several subdomains and then to perform a local minimization in every one of them. But again this approach proves unworkable in the case of a large parameter vector (for a six-parameter problem, a division of the interval of variation of each parameter into merely five subintervals would already require 15 625 local minimizations [8]). Several methods have been proposed to limit the computational burden of this method by determining the domains of attraction of the local minima [4,11,20,27]. These methods try to cluster all the points in the parameter space which seem to lead to the same minimum so as to avoid unnecessary local minimizations. Their practical implementation appears cumbersome, and the results obtained depend heavily on the initial points and clustering policy chosen. All reported uses of these methods concern problems with a limited number of parameters.

Interval analysis can also be used to locate the global minimum of a multivariable function [13,14,16]. The original admissible domain in the parameter space is successively divided into several subdomains. On each of these subdomains upper and lower bounds of the function on the interval are computed. Any subdomain to which the global minimum cannot belong is then discarded. This approach requires an analytic expression of the criterion to be available—an assumption which is not commonly satisfied by real-life problems—and the implied computational burden can be considerable for problems with a large number of parameters.

On the other hand, probabilistic methods appear particularly convenient when one is concerned with the global optimization of a criterion which depends on a large parameter vector. Their main idea is to generate, at each iteration k , new parameter vectors θ until $j(\theta) < j(\theta^k)$, where θ^k is the estimate of the optimal θ at the beginning of iteration k . The last value of θ is then accepted as θ^{k+1} . Such random search methods have only been proved to converge in probability, which may explain their lack of acceptance. However their implementation is extraordinarily simple, and no special analytic property of the criterion has to be assumed. Moreover the volume of computation required is far smaller than with deterministic methods. It has been claimed that the computation time increases exponentially with the number N of parameters for deterministic methods [4], and that with random search methods the increase will be linear only [22]. Solis et al. [26] present experimental evidence supporting this statement and propose an explanation. Masri et al. report [18] that the computation time required by their adaptive random search (A.R.S.) method increases approximately as the square root of N , which makes it especially attractive. To the best of our knowledge, no theoretical study of the convergence rates of these efficient algorithms has been presented up to now, and the only rigorous mathematical developments which have appeared have dealt with elementary algorithms applied to very simple examples [6,24,26].

From the above discussion, the A.R.S. method appears extremely promising. However actual per-

formance of algorithms may depend heavily on their practical implementation. The next section describes how we have implemented the A.R.S. strategy and the various modifications that we have found to improve its behavior.

3. Implementation of the A.R.S. strategy

If $\theta^k = (\theta_1^k, \theta_2^k, \dots, \theta_N^k)^T$ is the parameter vector at iteration k and $j^k = j(\theta^k)$ is the corresponding value of the criterion, the basic structure of almost any random search algorithm can be summarized as follows:

Step 0: Choose θ^0 , set $k = 0$.

Step 1: Generate one (or several) new trial point $y^k = \theta^k + r^k$, where r^k is a suitably distributed random vector. (Any non admissible y^k has to be discarded and replaced by an admissible one.)

Step 2: Generate θ^{k+1} according to some rule (e.g. $\theta^{k+1} = y^k$ if $j(y^k) < j(\theta^k)$ and $\theta^{k+1} = \theta^k$ otherwise).

Step 3: Increase k by one and go to Step 1.

Most random search algorithms only differ by the probability density function chosen for r^k and by the selection rule for θ^{k+1} .

When θ^k is far from the optimum value θ^* , best results will be achieved with r^k having a large variance, whereas a smaller variance is to be preferred when θ^k is close to θ^* . Since one never knows how far θ^k is from θ^* , the A.R.S. strategy [1,2,18] consists in repeatedly alternating two phases. During the first one the variance of r^k yielding the best value of the criterion is selected among a set of predefined variances. During the second phase this selected variance is used for a fixed number of iterations.

The implementation of this procedure requires choosing the actual value of many parameters, such as the number and values of the optional variances, the number of trials during the variance-selection phase and the total number of computations of the criterion allowed. In order to obtain an all-purpose optimizer one must avoid having to tune these parameters before each opti-

mization and the algorithm must be as problem-independent as possible. Because of the adaptive features of the algorithm, it is hoped that a single policy may be used for any optimization problem, contrary to what is done in [1], where each example is treated with a different tuning of the algorithm. The examples to be presented in Sections 4 and 5 support this hope.

We shall now describe in some detail our implementation of A.R.S. strategy so as to make it possible for any interested reader to program it easily.

The user is supposed to specify the criterion to be minimized and the admissible range for each parameter θ_i ,

$$\theta_{i,\min} \leq \theta_i \leq \theta_{i,\max}.$$

No other information has to be supplied. The routine chooses the initial point according to the rule

$$\theta_i^0 = \frac{\theta_{i,\max} + \theta_{i,\min}}{2}, \quad i = 1, 2, \dots, N.$$

This rule can be overridden at the user's will.

The random displacement vector r^k is generated according to a normal distribution with zero mean and a covariance Σ satisfying

$$\Sigma(\sigma) := \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_N] \quad \text{where}$$

$$\sigma = [\sigma_1, \sigma_2, \dots, \sigma_N]^T.$$

During the variance selection phase σ is taken in the set $\{\sigma^1, \sigma^2, \dots, \sigma^{f_1}\}$, where by convention

$$\sigma^{i+1} \sigma_j < \sigma^i \sigma_j, \quad j = 1, 2, \dots, N.$$

The range of the possible values for σ must be large enough to allow an easy exploration of the whole admissible parameter set as well as a precise localization of the optimal parameter vector. For these reasons we have chosen the following tactics:

$$\sigma^1 = \theta_{\max} - \theta_{\min},$$

where θ_{\max} and θ_{\min} are the upper and lower bounds defining the admissible parameter domain Θ . The other standard deviations are given by

$$\sigma^i = 0.1^{(i-1)} \sigma^1, \quad i = 2, \dots, f_1.$$

(Typical value $f_1 = 5$.)

Each possible value ' σ ' is used $f_2(i)$ times. Experimentally, it was found that for the small values of ' σ ', corresponding to a local search, it was not necessary to perform as many iterations as for the larger values of ' σ ', which are designed to escape from local minima. This is why we have chosen the number of trials with the i th variance to satisfy

$$f_2(i) = \frac{f_3}{i},$$

where the typical value for f_3 is 100.

The most successful variance ' σ^2 ' is then used f_4 times during the second phase of the algorithm (typical value $f_4 = 100$) before resuming a variance selection phase, unless the decision to stop is taken. In [18] the algorithm is stopped when a maximum number of iterations is reached or when the criterion has reached a predefined level \tilde{j} . It is often difficult, however, to specify \tilde{j} , and we have added to the maximum number of iterations the following rule: if $f_1\sigma$ has been selected more than f_5 times consecutively then we decide that the global minimum has been located and stop the algorithm (typically $f_5 = 5$).

Note that the f_i 's are parameters which can always be chosen according to the particular problem to be treated. However, it will be seen in Section 4 that the typical values proposed permit to solve the various problems presented efficiently, and thus confer to the routine some problem independence.

Since the smallest value of σ corresponds to a local exploration of the parameter space we thought it would be worthwhile to offer the possibility of performing a local minimization when $f_1\sigma$ is selected. In order to avoid unnecessary computations, each local minimization must be followed by a random search leading to a choice of σ different from $f_1\sigma$, before a new local minimization is authorised. As we do not suppose the gradient of the criterion to be provided by the user, we have chosen a quasi-Newton method which evaluates the required derivatives itself and has been implemented by IMSL under the name ZXMIN [17]. The use of this routine is made optional for one may wish to solve problems with non-differentiable criteria.

4. Test-cases

The first seven test-cases considered are given in Table 1. All of them are classical examples of the literature. The admissible domain Θ has been chosen to allow comparison with other global minimization methods.

In order to illustrate the ability of the A.R.S. approach to handle engineering problems we have added two test-cases related to the automatic control field.

TC 8 is a parameter estimation problem. The criterion to be minimized is

$$j(\theta) = \sum_{i=1}^{20} [y_p(10i) - y_m(10i, \theta)]^2.$$

The 'process' output is given by

$$y_p(t) = 5e^{-t/25} - 10e^{-t/12.5} + 5e^{-t/50} + \epsilon(t),$$

where $\epsilon(t)$ is an additive white noise normally distributed with zero mean and a standard deviation of 0.179.

The model output satisfies

$$y_m(t, \theta) = \theta_1 e^{-t/\theta_2} + \theta_3 e^{-t/\theta_4} - (\theta_1 + \theta_3) e^{-t/\theta_5}.$$

The (rather large) admissible domain is given by

$$\Theta = [-100, 100] \otimes [0, 100] \otimes [-100, 100] \\ \otimes [0, 100] \otimes [0, 100].$$

TC 9 is a control problem. The criterion to be minimized is

$$j(\theta) = \sum_{k=0}^{151} [y_r(k) - y_p(k, u(k, \theta))]^2,$$

where y_p is the process output, related to the control input u by the nonlinear state equation

$$x_1(k+1) = 2x_1(k) + 0.1x_1(k)x_2(k) + u(k, \theta).$$

$$x_1(0) = 0,$$

$$x_2(k+1) = x_1(k), \quad x_2(0) = 0,$$

$$y_p(k) = x_2(k),$$

and where y_r is the reference output to be followed by the process, given by the linear recur-

Table 1
Description of the first seven test-cases

Test-cases	Criterion	Admissible domain Θ	Minima ^{a)} $j(\theta^*), \theta^{*T}$
TC ₁ Rosenbrock [15,16,23,25]	$j(\theta): 100 \cdot (\theta_2 - \theta_1^2)^2 + (1 - \theta_1)^2$	$[-5, 5]^{\otimes 2}$	G: 0, (1, 1)
TC ₂ Beale [25]	$j(\theta) = \sum_{i=1}^3 (C_i - \theta_1(1 - \theta_2^i))^2$ where $C_1 = 1.5, C_2 = 2.25, C_3 = 2.625$	$[-10, 10]^{\otimes 2}$	G: 0, (3, 0.5)
TC ₃ Powell [25]	$j(\theta) = (\theta_1 + 10\theta_2)^2 + 5(\theta_3 - \theta_4)^2 + (\theta_2 - 2\theta_3)^2 + (10\theta_1 - \theta_4)^4$	$[-20, 20]^{\otimes 4}$	G: 0, (0, 0, 0, 0)
TC ₄ Colville [25]	$j(\theta) = 100 \cdot (\theta_1^2 - \theta_2)^2 + (1 - \theta_1)^2 + 10 \cdot (\theta_4 - \theta_3^2)^2 + (1 - \theta_3)^2 + 10.1(\theta_2 - 1)^2 + (\theta_4 - 1)^2 + 19.8(\theta_2 - 1)^2(\theta_4 - 1)^2$	$[-10, 10]^{\otimes 4}$	G: 0, (1, 1, 1, 1)
TC ₅ Hosaki [3]	$j(\theta) = (1 - 8\theta_1 + 7\theta_1^2 - \frac{7}{3}\theta_1^3 + \frac{1}{4}\theta_1^4)\theta_2^2 e^{-\theta_2}$	$[0, 5] \otimes [0, 6]$	G: -2.345, (4, 2) L: -1.127, (1, 2)
TC ₆ Goldstein and Price [8,15]	$j(\theta) = (1 + (\theta_1 + \theta_2 + 1)^2 \cdot f_1(\theta)) \cdot (30 + (2\theta_1 - 3\theta_2)^2 \cdot f_2(\theta))$ where $f_1(\theta) = 19 - 14\theta_1 + 3\theta_1^2 - 14\theta_2 + 6\theta_1\theta_2 + 3\theta_2^2$, $f_2(\theta) = 18 - 32\theta_1 + 12\theta_1^2 + 48\theta_2 - 36\theta_1\theta_2 + 27\theta_2^2$	$[-2, 2]^{\otimes 2}$	G: 3, (0, -1) 4 local minima in Θ
TC ₇ 3 Hump camel-back [14,15,16]	$j(\theta) = 2\theta_1^2 - 1.05\theta_1^4 + \frac{1}{6}\theta_1^6 - \theta_1\theta_2 + \theta_2^2$	$[-3, 3] \otimes [-1.5, 1.5]$	G: 0, (0, 0) L: (1.7455, -0.87377) L: (-1.7455, 0.87377)

^{a)} G = global, L = local.

rence equation

$$y_r(k+2) = y_r(k+1) - 0.5y_r(k) + 0.5,$$

$$y_r(0) = y_r(1) = 0.$$

The control input is supposed to satisfy

$$u(k, \theta) = \theta_1 y_p(k) + \theta_2 y_p(k-1) + \theta_3 y_p(k) y_p(k-1) + \theta_4 y_p^2(k) + \theta_5,$$

and the chosen admissible parameter domain is

$$\Theta = [-5, 5]^{\otimes 5}.$$

When treating these examples we address the following questions:

Q1: Is it worthwhile spending time to select the variance of the random search?

Q2: Is it useful to combine a global random search algorithm with a deterministic local minimization?

Q3: Is it possible to treat various problems while keeping the same value for all the optional parameters f of the algorithm?

Q4: Does the resulting algorithm compare reasonably with classical global search algorithms of the literature?

4.1. Interest of variance selection

Two random search algorithms are compared. The first one follows the A.R.S. strategy (without

Table 2
Effect of variance selection

Test-case	Algorithm	Initial point θ^0	Initial criterion value	Final criterion value ^{a)}	Number of Evaluations of the criterion ^{a)}
TC ₃	A.R.S. strategy	$(3, -1, 0, 1)^T$	$7.073 \cdot 10^5$	$3.105 \cdot 10^{-4}$	1969
	Unique variance value strategy	$(3, -1, 0, 1)^T$	$7.073 \cdot 10^5$	1.255	4055

^{a)} Averaged result over 25 successive runs.

Table 3
Usefulness of deterministic local minimization

Test-case	Algorithm	Initial point θ^0	Initial criterion value	Final criterion value ^{a)}	Number of evaluations of the criterion ^{a)}
TC ₃	A.R.S. with local minimization	$(3, -1, 0, 1)^T$	$7.073 \cdot 10^5$	$3.102 \cdot 10^{-13}$	1202
	A.R.S. without local minimization	$(3, -1, 0, 1)^T$	$7.073 \cdot 10^5$	$3.105 \cdot 10^{-4}$	1969

^{a)} Averaged result over 25 successive runs.

Table 4
Problem independence

Test-cases	Initial parameter vector	Initial criterion value	Final parameter vector	Final criterion value	Evaluations of the criterion
TC ₁	$(-1.2, 1)^T$	$7.490 \cdot 10^2$	$(1.000, 0.9999)^T$	$1.958 \cdot 10^{-9}$	796
TC ₂	$(0, 0)^T$	$1.420 \cdot 10^1$	$(3.000, 0.5000)^T$	$1.421 \cdot 10^{-14}$	783
TC ₃	$(3, -1, 0, 1)^T$	$7.073 \cdot 10^5$	$(4.983 \cdot 10^{-6}, -4.983 \cdot 10^{-7}, -7.575 \cdot 10^{-5}, -7.575 \cdot 10^{-5})^T$	$7.821 \cdot 10^{-16}$	1129
TC ₄	$(-3, -1, -3, -1)^T$	$1.139 \cdot 10^4$	$(1.000, 1.000, 1.000, 1.000)^T$	$1.829 \cdot 10^{-12}$	839
TC ₅	$(1, 4.5)^T$	-0.4687	$(4.001, 1.999)^T$	-2.346	830
TC ₆	$(1, 1)^T$	$1.876 \cdot 10^3$	$(3.055 \cdot 10^{-4}, -0.9995)^T$	3.000 ^{a)}	657
TC ₇	$(1.74755, -0.87377)^T$ local minimum	0.29863	$(2.962 \cdot 10^{-13}, -1.444 \cdot 10^{-11})^T$	$2.687 \cdot 10^{-24}$	838

^{a)} Result achieved without performing any local minimization.

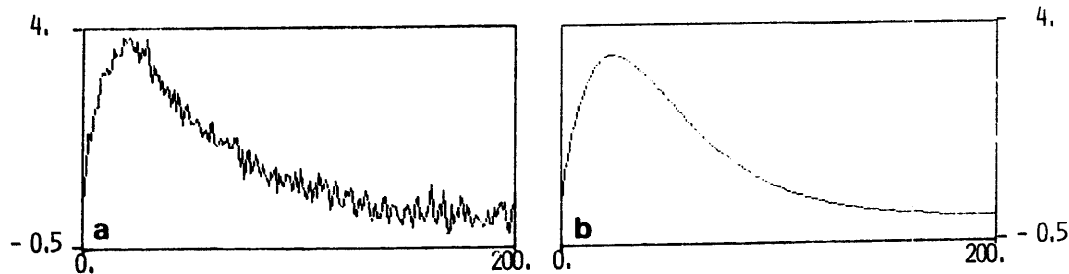


Fig. 1. (a) Process output; (b) optimized model output.

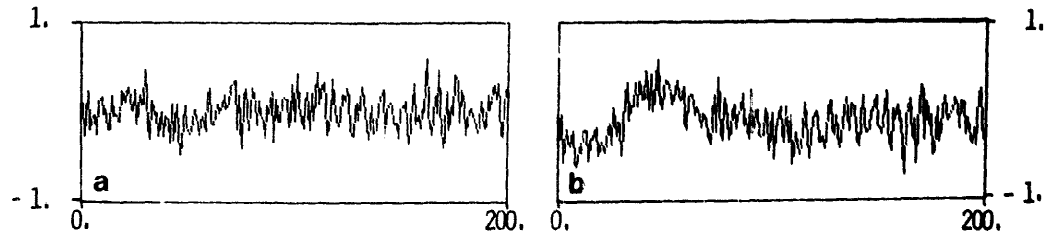


Fig. 2. (a) Additive noise; (b) output error.

local minimization), while the second one uses a unique variance value σ , given by

$$\sigma = {}^2\sigma.$$

A representative sample of available results is reported in Table 2. Similar results have been obtained for each problem treated, A.R.S. strategy always leads to a far higher efficiency compared with classical random search using a unique variance value, independently of the chosen value.

4.2. Usefulness of deterministic local minimizations

Deterministic local minimization has been found almost always to increase the efficiency of A.R.S. strategies. Table 3 presents a typical example of the improvement that can be obtained.

4.3. Problem independence

All the different test-cases considered have been treated with the same value of the optional param-

eters, equal to the typical value indicated in Section 3. Results for the first seven test-cases are summarized in Table 4. Results for the automatic control test-cases are as follows:

TC8: The initial parameter vector is $(0, 50, 0, 50, 50)^T$. The 'process' output and the optimized model output are given in Fig. 1. Figure 2 makes it possible to compare the additive noise, which is available in this fictitious example, to the error between 'process' and model outputs.

The criterion has been computed 965 times. Its final value is 0.7779 (lower than the contribution of the noise 0.9362).

TC9: The initial parameter vector is $(0, 0, 0, 0, 0)^T$. The reference output and the process response to a unit step function are given in Fig. 3. Figure 4 shows the output of the process when controlled by the computed control law and the corresponding output error.

The criterion has been evaluated 1672 times, its final value is $0.5332 \cdot 10^{-2}$.

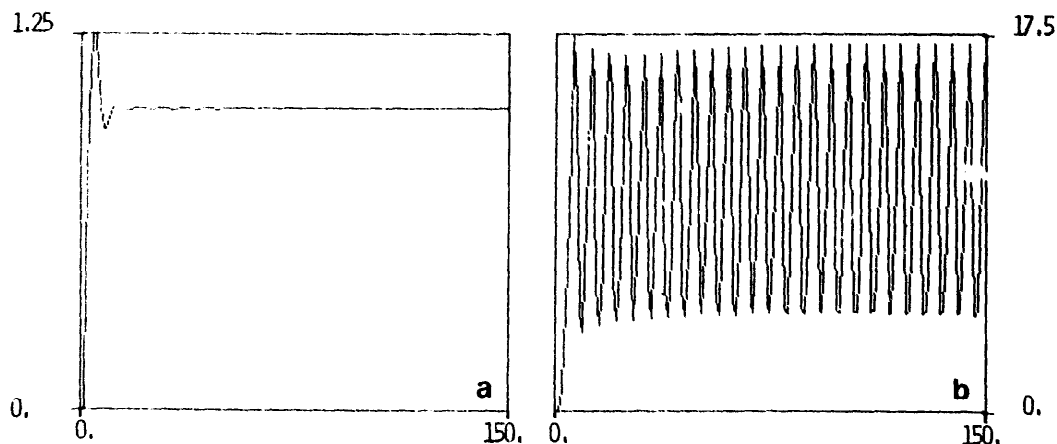


Fig. 3. (a) Reference output; (b) process response to a unit step function.

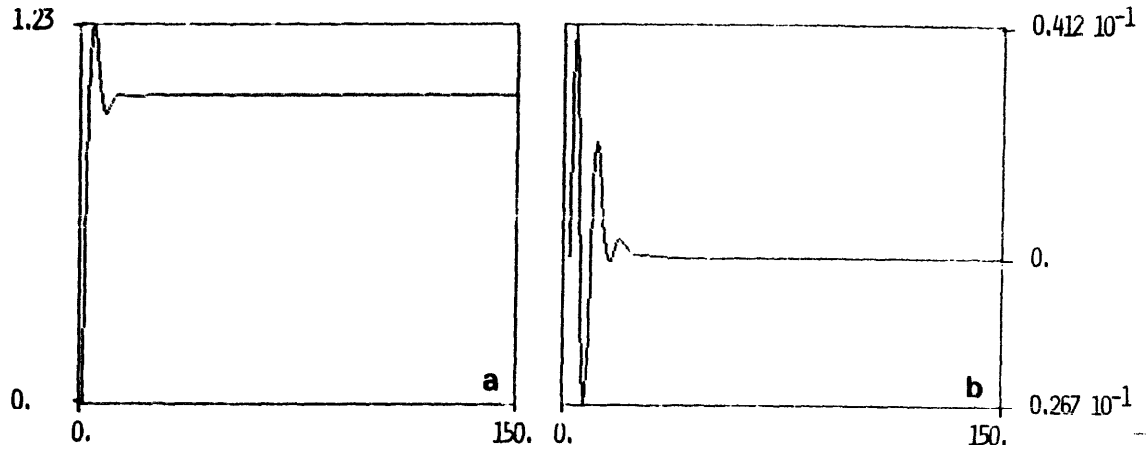


Fig. 4. (a) Controlled process output; (b) output error.

Table 5
Results found in literature

Test-cases	Method	Initial parameter vector	Final parameter vector	Final criterion value	Evaluations of the criterion
TC ₁	Scheela [25]	$(-1.2, 1.0)^T$	$(1.000, 1.001)^T$	$0.657 \cdot 10^{-6}$	1941
	Hardy [15]	(*)	G(*)	(*)	1259
TC ₂	Scheela [25]	$(0, 0)^T$	$(2.978, 0.494)^T$	$0.737 \cdot 10^{-4}$	988
TC ₃	Scheela [25]	$(3, -1, 0, 1)^T$	$(-0.0211, 0.0019, -0.045, -0.045)^T$	$0.83 \cdot 10^{-3}$	4006
TC ₄	Scheela [25]	$(-3, -1, -3, -1)^T$	$(1.01, 1.02, 0.98, 0.97)^T$	$0.98 \cdot 10^{-3}$	97813
TC ₅	Bekey/Ung [3]	$(1, 4.5)^T$	$(3.99093, 2.01405)^T$ (one failure recorded out of 27 runs)	-2.34555	500
TC ₆	Hardy [15]	(*)	G(*)	(*)	4242
	Bremmermann [8]	(*)	L(*)	(*)	210
	Mod. Bremmermann [8]	(*)	G(*)	(*)	300
	Gomulka/VM [11]	-	G. Accuracy 0.005	(*)	1495
	Törn [27]	-	$(0.202 \cdot 10^{-3}, -1.00007)^T$	3.00001	2499
	Price [20]	-	$(0.00000, -1.00000)^T$	3.00000	2500
	Fagioli [9]	-	$(0.149 \cdot 10^{-8}, -1)^T$	3.0	158
					(1096 to achieve converg. in probab. 0.90)
	De Biase Frontini [5]	-	G(*)	2.9997	378
	Mockus [19]	-	G. $\ \text{gradient}(j)\ < 5 \cdot 10^{-4}$	(*)	362
Zilinskas [19]	(*)	$(0.00001, -1.00000)^T$	3.0	520	
TC ₇	Hardy [15]	(*)	G(*)	(*)	1136

(*) Unavailable result. G: Global minimum located. L: Local minimum located.

4.4. Comparison with other algorithms

Some available results of the literature concerning the first seven test-cases are presented in Table 5. Together with Table 4 they make it possible to compare the efficiencies of various methods with that of our implementation of A.R.S. strategy. On the first five examples our routine performs better than or comparably with the methods of the literature. For the two remaining test-cases, information was not always available on the final value of the parameters and of the criterion, so that a comparison is more difficult, since the number of iterations depends heavily on the required precision. In any case, our routine has successfully and precisely located the minima in a reasonable number of iterations, without needing any problem-dependent tuning.

5. Application to automated registration of medical images

Our implementation of A.R.S. strategy is used to solve an original optimization problem arising in the field of medical nuclear imagery. The comparison of two scintigraphic images of the same organ explored under varying conditions is a routine problem. To automate this comparison it is necessary first to carry out the registration of the images (e.g. alignment, normalization of the gray levels), and second to analyse the images pixel by pixel.

The registration step is performed by optimizing a similarity measure with respect to the registration parameters. Let $F_1(i, j)$ and $F_2(i, j)$ be two digitized images of a same organ. The registration procedure consists in transforming F_2 according to the rule

$$F_{1,m}(i, j) = a \cdot [\mathcal{T} \circ \mathcal{R} \{ F_2 \}] + b,$$

so as to maximize some measure of the similarity between F_1 and $F_{1,m}$. \mathcal{T} is a translation characterized by Δ_x and Δ_y , the shifts along the horizontal and vertical axis, \mathcal{R} is a rotation of an angle α , a (resp. b) is a multiplicative (resp. additive) normalization factor of the gray levels. The selected

similarity measure is the S.S.C. criterion [29,30] defined as the number of sign changes in the pixels of the subtraction image $F_1 - F_{1,m}$ scanned line by line. This criterion has to be maximized with respect to five parameters $\theta = (\Delta_x, \Delta_y, \alpha, a, b)^T$. $(\Delta_x, \Delta_y, \alpha, a, b)^T$.

This problem presents two specific difficulties: First the criterion has an integer value and therefore is not differentiable, second the three parameters Δ_x , Δ_y and b are integer.

The routine coded in FORTRAN is implemented on a mini computer¹ connected with an array processor² which is used for the computation of the S.S.C. criterion.

An example corresponding to two bone scintigraphic images (128×128) of a pelvis is given in Fig. 5. The admissible domain for the parameter vector is

$$\Theta = [-30, 30] \otimes [-30, 30] \otimes [-45, 45] \\ \otimes [0.1, 10] \otimes [-20, 20].$$

Figure 5(c) presents the result of the registration procedure obtained in three minutes and 2000 computations of the criterion, without using the local minimization option.

6. Conclusion

Among the many methods available for global optimization, the Adaptive Random Search strategy has been selected owing to its simplicity and efficiency. The various test-cases and the practical application treated in this paper support the idea that our implementation of A.R.S. strategy is a useful general-purpose tool for optimizing criteria which are not necessarily convex functions of the parameters. It is now in routine use in our laboratories. The only information that the user has to supply is a rule for computing the criterion and an admissible domain for the parameter. No tuning is required for each specific problem and the possibility of combining global random search and

¹ IMAC 7300, CGR Medecine Nucleaire (French version of ADAC system I).

² AB120B, Floating Point System.

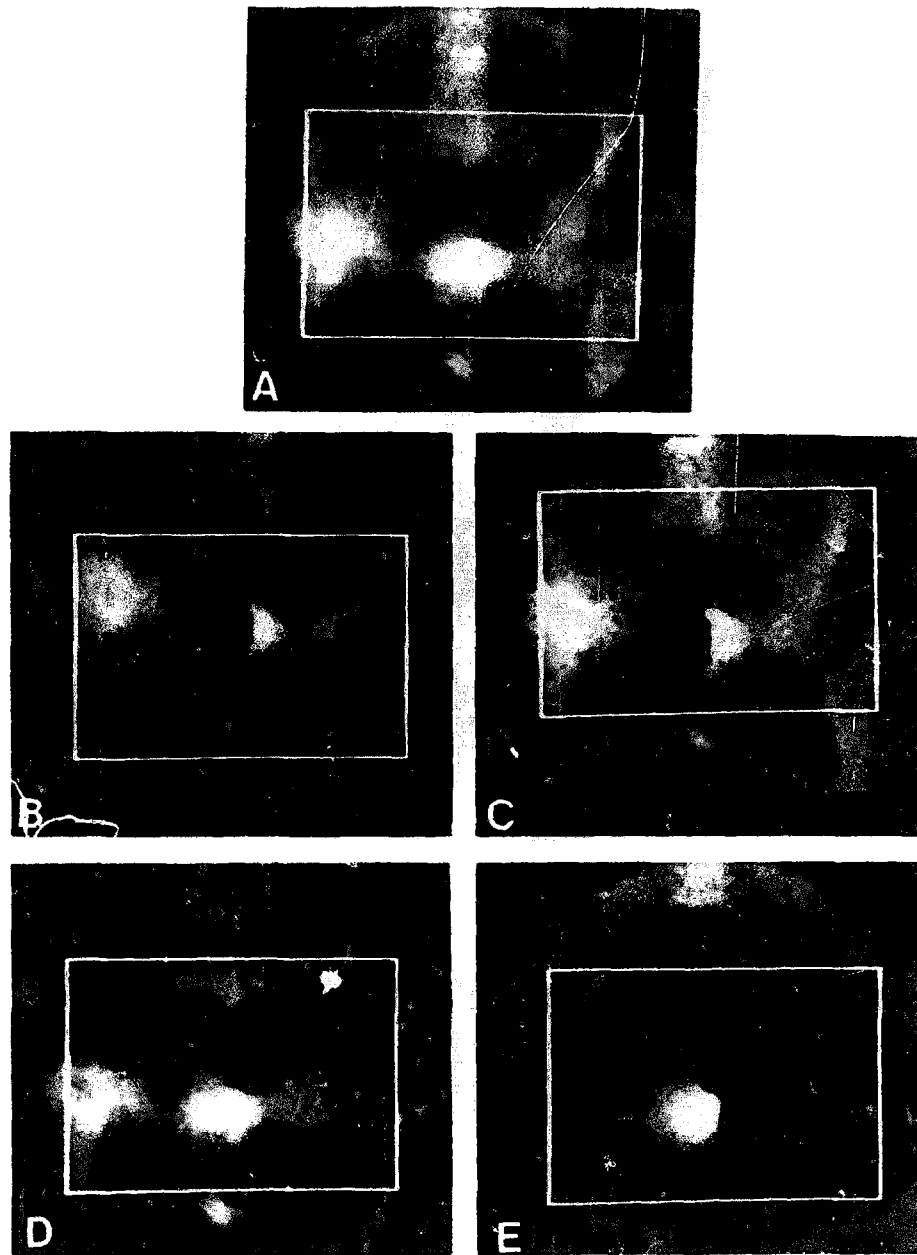


Fig. 5. Example of the automated registration procedure applied to scintigraphic images of a pelvis, the considered part is limited by the white rectangle: (A) original image; (B) same image of the pelvis after translation, rotation, modification of the image and of the gray level intensity; (C) registered version of image (B) after application of the A.R.S. method; (D) subtraction image (A) - (B); (E) subtraction image (A) - (C).

deterministic local search has been shown to improve the performance of the algorithm considerably.

References

- [1] A.M. Andronikou, G.A. Bekey and S.F. Masri, Identification of nonlinear hysteretic systems using random search.

- Proc. 6th IFAC Symposium on Identification and System Parameter Estimation*, Vol. 1 (1982) 263–268.
- [2] G.A. Bekey and S.F. Masri, Random search techniques for optimization of nonlinear systems with many parameters, *Math. Comput. Simulation* 25 (3) (1983) 210–213.
- [3] G.A. Bekey and M.T. Ung, A comparative evaluation of two global search algorithms, *IEEE Trans. Systems, Man Cybernet.* 4 (1) (1974) 112–116.
- [4] C.G.E. Boender, A.H.G. Rinnooy Kan, G.T. Timmer and L. Stougie, A stochastic method for global optimization, *Math. Programming* 22 (1982) 125–140.
- [5] L. De Biase and F. Frontini, A stochastic method for global optimization: Its structure and numerical performance, in: [8], pp. 85–102.
- [6] L.P. Devroye, Progressive global random search of continuous functions, *Math. Programming* 15 (1978) 330–342.
- [7] L.C.W. Dixon and G.P. Szego (Eds.), *Towards Global Optimisation* (North-Holland, Amsterdam, 1975).
- [8] L.C.W. Dixon and G.P. Szego (Eds.), *Towards Global Optimisation 2* (North-Holland, Amsterdam, 1978).
- [9] E. Faggioli, P. Pianca and M. Zecchin, A mixed stochastic-deterministic technique for global optimisation, in: [8], pp. 103–116.
- [10] J. Gomulka, Remarks on Branin's method for solving nonlinear equations, in: [7], pp. 96–106.
- [11] J. Gomulka, A users experience with Torn's clustering algorithm, in: [8], pp. 63–70.
- [12] A.O. Griewank, Generalized descent for global optimization, *J. Optimization Theory Appl.* 34 (1) (1981) 11–39.
- [13] E.R. Hansen, Global optimization using interval analysis: the one-dimensional case, *J. Optimization Theory Appl.* 29 (3) (1979) 331–344.
- [14] E.R. Hansen, Global optimization using interval analysis — The multi-dimensional case, *Numer. Math.* 34 (1980) 247–270.
- [15] J.W. Hardy, An implemented extension of Branin's method, in: [7], pp. 119–142.
- [16] K. Ichida and Y. Fujii, An interval arithmetic method for global optimization, *Computing* 23 (1979) 85–97.
- [17] IMSL, Library Reference Manual, Edition 8, Houston Texas (1980).
- [18] S.F. Masri, G.A. Bekey and F.B. Safford, An adaptive random search method for identification of large-scale nonlinear systems, *Proc. 4th IFAC Symposium on Identification and System Parameter Estimation* (1976) 246–255.
- [19] J. Mockus, V. Tiesis and A. Zilinskas, The application of Bayesian methods for seeking the extremum, in: [8], pp. 117–129.
- [20] W.L. Price, A controlled random search procedure for global optimisation, in: [8], pp. 71–84.
- [21] L.A. Rastrigin, The convergence of the random search method in the extremal control of a many-parameter system, *Automation and Remote Control* 24 (11) (1963) 1337–1342.
- [22] W.E. Red, A method for determining the equilibrium states of dynamics systems, *J. Optimization Theory Appl.* 21 (3) (1977) 299–317.
- [23] H.H. Rosenbrock, An automatic method for finding the greatest or least value of a function, *Comput. J.* 3 (1960) 175–184.
- [24] Y.R. Rubinstein and G. Samorodnitsky, Efficiency of the random search method, *Math. Comput. Simulation* 23 (1982) 257–268.
- [25] B.V. Sheela, An optimized step-size random search (OS-SRS), *Comput. Methods Appl. Mech. Engrg.* 19 (1979) 99–106.
- [26] F.J. Solis and R.J.B. Wets, Minimization by random search techniques, *Math. Operations Res.* 6 (1) (1981) 19–30.
- [27] A.A. Thurn, A search-clustering approach to global optimization, in: [8], pp. 49–62.
- [28] G. Treccani, A new strategy for global minimisation, in: [7], pp. 143–148.
- [29] A. Venot, J.F. Lebruchec, J.L. Golmard and J.C. Roucayrol, An automated method for the normalization of scintigraphic images, *J. Nucl. Med.* 24 (1983) 529–534.
- [30] A. Venot, J.L. Golmard, J.F. Lebruchec, L. Pronzato, E. Walter, G. Fria and J.C. Roucayrol, Digital methods for change detection in medical images, in: *Information Processing in Medical Images* (Martinus Nijhoff, The Hague) to appear.