

- Bryan, J.K., Dwyer, S.J., and Lago, G.V. (1969). "Non-Parametric Decision Schemes for EEG Classification". Sixth Annual Rocky Mountain Bio-Engineering Symposium Conference Record, 56-61, University of Wyoming, Laromie, Wyoming.
- Cutteridge, O.P.D. (1974). "Powerful 2-Part Program for Solution of Nonlinear Simultaneous Equations". Electronics Letters, Vol. 10, No. 10, 182-184.
- Fletcher, R. and Powell, M.J.D. (1963). "A Rapidly Convergent Descent Method for Minimisation". The Computer Journal, 6, 163-168.
- Hooke, R. and Jeeves, T.A. (1969). "Direct Search Solution of Numerical and Statistical Problems". Journal of the Association for Computing Machinery, 8, 212-229.
- Nelder, J.A. and Mead, R. (1965). "A Simplex Method for Function Minimisation". The Computer Journal, 7, 308-313.
- Rosenbrock, H.H. (1960). "An Automatic Method for Finding the Greatest or Least Value of a Function". The Computer Journal, 3, 175-184.

TOWARDS GLOBAL OPTIMISATION 2
L.C.W. Dixon and G.P. Szegö (eds.)
© North-Holland Publishing Company (1978)

A STOCHASTIC METHOD FOR GLOBAL OPTIMIZATION:
ITS STRUCTURE AND NUMERICAL PERFORMANCE(*)

L. DE BIASE, ISTITUTO DI MATEMATICA
UNIVERSITÀ DEGLI STUDI DI MILANO
MILANO, ITALY
F. FRONTINI, C.I.S.E. SEGRATE, ITALY

In this paper a method is developed for the minimization of a function $f: \mathbb{S} \subset \mathbb{R}^N \rightarrow \mathbb{R}$. The root of a suitable approximation $A(\xi)$ of the level set measure yields an approximation to f^* . $A(\xi)$ is given by a recursive spline technique, smoothing the data obtained by a sequential uniform sampling both on \mathbb{S} and on the expected range of function values.

INTRODUCTION

Let $f(x)$ be a continuous real valued function defined on a compact set $\mathbb{S} \subset \mathbb{R}^N$; our problem is to single out a point x^* such that

$$f^* = f(x^*) \leq f(x), \quad x \in \mathbb{S}.$$

For particular classes of functions many satisfactory algorithms have been exhibited, but in the general case, mainly when the function to be minimized is multiextremal, usual optimization techniques cannot be applied. An interesting stochastic strategy relies upon a one variable function $\psi(\xi)$ defined as the normalized Lebesgue measure of the set

$$E(\xi) = \{x \in \mathbb{S}: f(x) \leq \xi\}.$$

If f^* is the essential infimum of $f(x)$ and x^* is an isolated minimum, we have obviously: $\psi(\xi) = 0$ for $\xi < f^*$ and $\psi(\xi) > 0$ for $\xi > f^*$. In order to approximate $\psi(\xi)$, $f(x)$ is evaluated at random points out of a uniform distribution in \mathbb{S} , yielding a "noisy" pointwise approximation of $\psi(\xi)$. The sampling and the approximation processes are tuned together sequentially in order to get a convergent estimate of f^* at the root

(*) This work was supported in part by G.N.I.M. (Gruppo Nazionale Informatica Matematica) of C.N.R. (Italian National Research Council)

of the approximating function and to control the sample size.

ANALYSIS OF THE METHOD

Let $f(x)$ be a real valued continuous function, defined on a compact set $S \subset \mathbb{R}^N$. Let $E(\xi)$ be the set

$$E(\xi) = \{x \in S : f(x) \leq \xi\} \quad \xi \in \mathbb{R}^1.$$

Let $m(\cdot)$ be a measure proportional to the Lebesgue measure on S , such that $m(S)=1$. We define a function

$$\psi(\xi) = m(E(\xi)) \quad (1.1)$$

mapping \mathbb{R} into \mathbb{R}^+ .

As S is a compact set, $E(\xi)$ belongs to the σ -algebra of Lebesgue sets in S ; thus $\psi(\xi)$ is properly defined and $0 \leq \psi(\xi) \leq 1$.

Let f^* be the essential infimum of $f(x)$ in S . If $\xi < f^*$, $\psi(\xi) = 0$; $m(f^{-1}(f^*)) = 0$ implies $\psi(\xi) = 0$ iff $\xi \leq f^*$.

The function $\psi(\xi)$ has some regularity properties:

Theorem: a) $\psi(\xi)$ is non decreasing in \mathbb{R}^1 ;

b) $\psi(\xi)$ is almost everywhere differentiable;

c) $\psi(\xi)$ is continuous in \mathbb{R}^1 provided that no set $H \subset S$ exists such that $f(x) = \text{const.}$ in H and $m(H) > 0$.

The proof of this theorem is in Archetti-Betrò (1976); we give here some hint of it:

a) follows immediately from properties of a Lebesgue measure; to prove

b) $m(f^{-1}(\cdot))$ is shown to be a finite Borel measure. After known theorems (Rudin (1970)) this implies that $m(f^{-1}(-\infty, \xi)) = \psi(\xi)$ is almost everywhere differentiable; using properties of real monotone sequences right and left continuity are stated and c) is proved.

The behaviour of $\psi(\xi)$ is shown in figure 1 for an $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ (figure 1).

By the properties of $\psi(\xi)$ an approximation to a value β^* such that $\psi(\xi) = 0$ for $\xi \leq \beta^*$ and $\psi(\xi) > 0$ for $\xi > \beta^*$ may be instrumental in controlling the optimization process: thus we look for an approximation of $\psi(\xi)$ over the interval $[f^*, \|f\|_\infty]$, where $\|f\|_\infty$ is the supremum in S of $f(x)$.

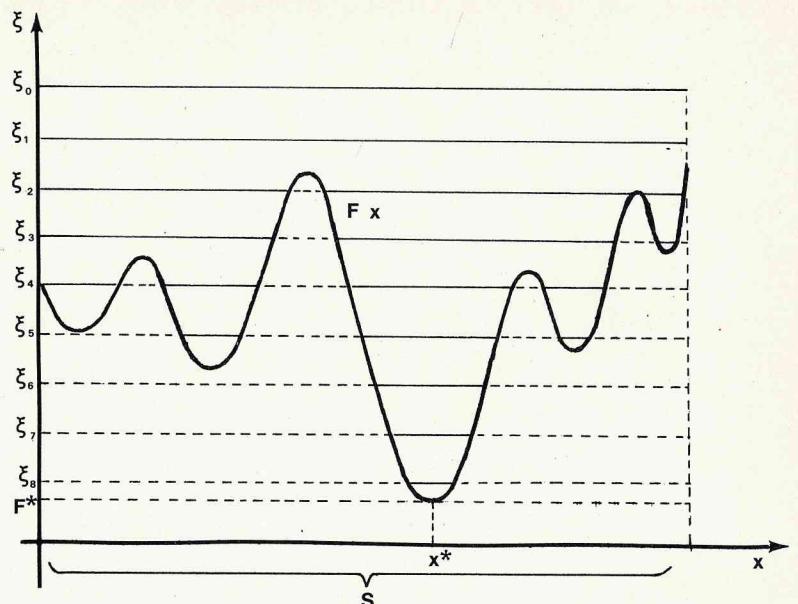


fig. 1

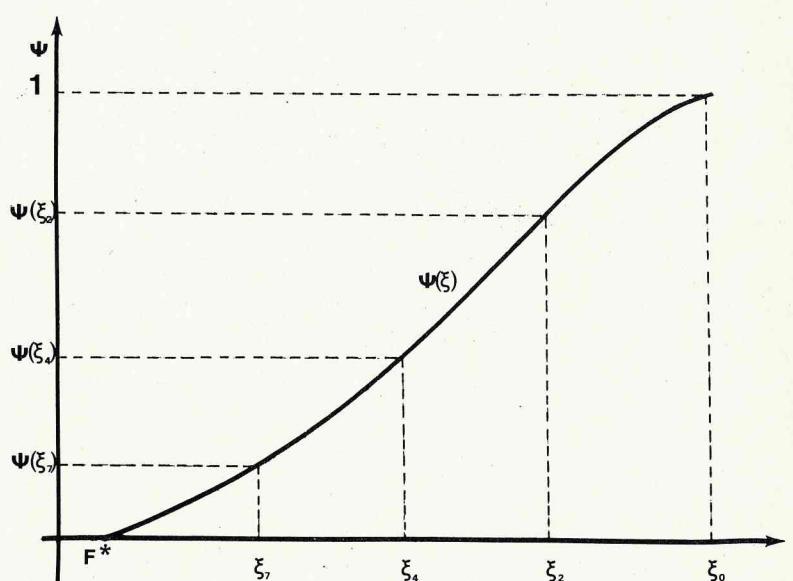


fig. 2

Except for trivial cases, an analytic expression of $\psi(\xi)$ is not available; it is then necessary to set up a stochastic sampling technique to provide an approximation of it in analytical form.

Let $w \in \mathbb{R}^N$ be a vector uniformly distributed in S . Let's define

$$P(\xi) = \text{prob } \{ f(w) \leq \xi \}$$

as the probability of hitting the region $E(\xi)$.

Since the distribution is uniform and $m(S)=1$, it follows that

$$\psi(\xi) = m(E(\xi)) = P(\xi).$$

We perform a sample θ_q and evaluate $f(x)$ at every random point w_j of θ_q ($j=1, 2, \dots, q$).

Let:

$$\theta_x = \min_{j=1, 2, \dots, q} f(w_j) \quad \text{and} \quad (1.2)$$

$$\theta_u = \max_{j=1, 2, \dots, q} f(w_j),$$

and let's choose ξ out of a uniform distribution in $[\theta_x, \theta_u]$.

Let p be the number of trial points hitting $E(\xi)$; $\psi(\xi)$ is approximated by $\tilde{\psi}(\xi) = p/q$.

The random variable $\tilde{\psi}(\xi)$ is an unbiased estimator of $\psi(\xi)$.

For the law of large numbers, we have:

$$\text{Prob}\{\lim_{q \rightarrow \infty} p/q = \psi(\xi)\} = 1.$$

The random variable $\tilde{\psi}(\xi)$ follows a binomial distribution with parameter $\psi(\xi)$; the error $|p/q - \psi(\xi)|$ is of the order of $1/\sqrt{q}$; its expected value is 0 and its standard deviation is $\{\psi(\xi)(1-\psi(\xi))/q\}^{1/2}$; thus $\psi(\xi)$ turns out to be an unknown regression function and we look for its root ξ^* , such that $\psi(\xi^*)=0$, $\psi(\xi)>0$ for $\xi>\xi^*$.

Stochastic optimization techniques are likely to fail in this problem because the root we are looking for is located just outside the actually sampled range in the ξ variable: we are thus faced with stochastic extrapolation.

To surmount this difficulty we provide a least squares approximation to $\psi(\xi)$, so that its root may be analytically evaluated.

Let $\phi_\ell(\xi)$ ($\ell=1, 2, \dots, k$) be our basic functions and $\sum_1^k \lambda_\ell \phi_\ell(\xi)$ the approximation. To calculate the optimal values of the coefficients λ_ℓ ($\ell=1, 2, \dots, k$), we consider the following error functional

$$D(\lambda) = \int_{f^*}^{\|f\|_\infty} [f(\xi) - \sum_{\ell=1}^k \lambda_\ell \phi_\ell(\xi)]^2 d\xi \quad (1.3)$$

and minimize it.

Since $\psi(\xi)$, $\|f\|_\infty$, f^* are not available, $D(\lambda)$ is replaced by the empirical risk functional

$$J^r(\lambda) = \left\{ \sum_{i=1}^r \tilde{\psi}_i(\xi_i) - \sum_{\ell=1}^k \lambda_\ell \phi_\ell(\xi_i) \right\}^2 \quad (1.4)$$

where r , $\tilde{\psi}_i$ and ξ_i are defined below:

let $\theta_{q_i}^r = \{f(w_{j_i}^i), j_i=1, 2, \dots, q_i; i=1, 2, \dots, r\}$ be the i -th sample performed, where $w_{j_i}^i$, $i=1, 2, \dots, r$ and $j_i=1, \dots, q_i$ have already been described. We define

$$w_\ell^r = \min_{i} \min_{j_i} \{f(w_{j_i}^i)\} \quad \text{and} \quad (1.5)$$

$$w_u^r = \max_{i} \max_{j_i} \{f(w_{j_i}^i)\}.$$

Consider now a random variable ξ_i uniformly distributed in $[w_\ell^r - \epsilon, w_u^r + \delta]$, with $\epsilon, \delta > 0$; the following theorem holds (Archetti-Betrò (1976)):

Theorem. Let $m_i = \min\{f(x_j), j=1, 2, \dots, i\}$ and $M_i = \max\{f(x_j), j=1, 2, \dots, i\}$ where x_j are independent random points uniformly distributed in S . Then $m_i \rightarrow f^*$ and $M_i \rightarrow \|f\|_\infty$ with probability 1 as $i \rightarrow \infty$.

This theorem ensures that for r sufficiently large

$$[w_\ell^r - \epsilon, w_u^r + \delta] \ni [f^*, \|f\|_\infty]$$

with probability 1.

Now we can define an unbiased estimator of $\psi(\xi)$ at every level ξ_i

$$\tilde{\psi}_i = \tilde{\psi}(\xi_i) = p_i / q_i, \quad (1.6)$$

where p_i is the number of trial points hitting the region $E(\xi_i)$ in $\theta_{q_i}^r$. Moreover, as the samples $\theta_{q_i}^r$ are independent, we have in hand an independent sample $\{(\xi_i, \tilde{\psi}_i), i=1, 2, \dots, r\}$ for the construction of $\psi(\xi)$.

THE APPROXIMATION OF $\psi(\xi)$

$$\text{Let } s_k^r(\xi) = \sum_{l=1}^k \lambda_l^{*r} \phi_l(\xi) \quad (2.1)$$

where λ_l^{*r} is such that $J(\lambda^{*r}) \leq J(\lambda)$ for every λ .

As an approximation of f^* we shall consider the value β_k^r , such that

$$s_k^r(\beta_k^r) = 0, \quad s_k^r(\xi) > 0 \text{ for } \xi > \beta_k^r.$$

We remark that such a β_k^r may fail to exist. In order to ensure its existence (as can easily be shown) we generalize its definition as follows:

$$s_k^r(\beta_k^r) = \varepsilon_k, \quad s_k^r(\xi) > \varepsilon_k \text{ for } \xi > \beta_k^r \quad (2.2)$$

with $\varepsilon_k \rightarrow 0^+$ for $k \rightarrow +\infty$.

We must of course be sure that our approximation ensures $\beta_k^r \rightarrow f^*$ as $r, k \rightarrow \infty$, i.e. that every piece of information gained during sampling improves the approximation to $\psi(\xi)$ as well as to f^* . This behaviour is ensured if s_k^r converges uniformly to the regression function $\psi(\xi)$ as $r, k \rightarrow \infty$, and this happens under mild conditions if $s_k^r(\xi)$ is a spline function of odd degree, as stated in Betrò-De Biase (1976) (*). We now set $k=N+2m$ (the reason will be clear later).

Let $s_k^r(\xi)$ be expressed in terms of spline functions of order $2m-1$, with N equidistributed knots z_j , $j=1, 2, \dots, N$ in $[f^*, \|f\|_\infty]$.

To find out the $N+2m$ parameters $(\lambda_1, \lambda_2, \dots, \lambda_{N+2m})$ of our spline function $s_k^r(\xi)$, we minimize the functional

$$J^r(s_k^r) = \sum_{i=1}^r [\psi_i - s_k^r(\xi_i)]^2. \quad (2.3)$$

Uniform convergence of the spline functions sequence to $\psi(\xi)$ is ensured by a result of Mikhalkin's (1974):

(*) We recall that a spline function $s_N(\xi)$ of order $2m-1$ with N equidistributed knots z_i , $i=1, 2, \dots, N$ over a closed interval $[a, b]$ is a piecewise polynomial function of $C^{2m-2}[a, b]$, i.e. it is a function such that $s_N(\xi)$ is a polynomial of degree at most $2m-1$ on every interval $(-\infty, z_1], (z_N, +\infty), (z_i, z_{i+1})$, $i=1, 2, \dots, N-1$, with derivatives joining at every knot, up to the $(2m-2)$ -th order.

Theorem: Let $s_{N(r)}^r(\xi)$ be a spline function for which the number $N(r)$ of knots, equidistributed in a closed interval $[a, b]$ increases as $r \rightarrow \infty$. Let $f(x)$ be a continuous function defined on $[a, b]$ and $\{(x_i, f_i)\}$, $i=1, 2, \dots, r$ an independent sample such that f_i is an unbiased estimator of $f(x_i)$; if the sequence $s_{N(r)}^r$ minimizes the empirical risk, then it converges uniformly to $\psi(\xi)$ provided that the following condition holds:

$$\lim_{r \rightarrow \infty} \frac{N^2(r) \ln r}{r} = 0.$$

Relying upon all the considerations quoted above about the properties of $\psi(\xi)$ and about the sequential sampling procedure, we can apply this result to the approximation of $\psi(\xi)$.

From now on we shall write s_N^r for the approximation, to stress the number of knots.

SPLINE APPROXIMATION

It is shown (Betrò-De Biase (1976)) that a necessary and sufficient condition for the existence of a spline function $s_N^*(\xi)$ of degree $2m-1$ minimizing the empirical risk is the existence of a subset P of the set $\{\xi_i\}$, $i=1, 2, \dots, r$ with $r > N+2m$, such that

$$P = \{\xi_{j_i}\}, \quad i=1, 2, \dots, N+2m \text{ with } \xi_{j_1} < \xi_{j_2} < \dots < \xi_{j_{N+2m}} \text{ and}$$

$$\xi_{j_i} < z_i < \xi_{j_{i+2m}}. \quad (3.1)$$

The proof of this theorem relies upon two theorems which provide conditions to ensure that the determinant $|(d_i - b_j)|$ (*) is positive given any $n > 0$ and any numbers $\{d_i\}$, $i=1, 2, \dots, r\}$, $\{b_j\}$, $j=1, 2, \dots, r\}$ and that a system $s_N(x_i) = y_i$, $i=1, 2, \dots, N+2m$, and $x_1 < x_2 < \dots < x_{N+2m}$ has one and only one solution.

After these theorems it can be shown that the system matrix is positive definite iff a subset like P exists, because the proof of positive definiteness reduces to looking for the kernel of a system with r equations and $N+2m$ variables.

In our algorithm a more restrictive condition than (3.1) is needed,

(*) For the definition of $|(d_i - b_j)|$ see (3.3).

i. e. at least one point occurs between every two knots. If this is verified the algorithm proceeds, otherwise other points ξ_i are generated until our requirement is fulfilled.

The number of parameters to be evaluated to determine the spline function is $N+2m$. Indeed, a polynomial of degree $2m-1$ has $2m$ coefficients and we have $N+1$ polynomials (we do not consider the bounds of the interval as knots); so we have $(N+1)2m$ parameters. We have also $(2m-1)N$ conditions, since we require our polynomials to join at every knot together with equal derivatives up to order $2m-2$. So the number of free parameters is $N+2m$.

Setting $s_N^{2m-2}(\xi) = \text{const.}$ on every interval between two knots and integrating with respect to ξ , it is easy to show that every spline function has the following representation:

$$s_N(\xi) = \sum_{i=0}^{2m-1} a_i \xi^i + \sum_{j=1}^N c_j (\xi - z_j)_+^{2m-1} \quad (3.2)$$

where, for every function Γ we define

$$\Gamma_+ = \begin{cases} \Gamma & \text{if } \Gamma \geq 0 \\ 0 & \text{if } \Gamma < 0. \end{cases} \quad (3.3)$$

We remark that this representation has

$$1, \xi, \xi^2, \dots, \xi^{2m-1}, \{(\xi - z_j)_+^{2m-1}, j=1, 2, \dots, N\} \quad (3.4)$$

as basic functions, i.e. exactly $N+2m$ parameters are needed.

A severe ill-conditioning of the matrix C may arise in the system $Cx=b$, where C and b are respectively the matrix and the vector obtained when minimizing $J(s_N)$.

In order to avoid ill-conditioning, a proper basis must be chosen. A very good one is composed by the $(2m)$ -th divided differences of the functions

$$(\xi - z_j)_+^{2m-1}, \quad j=1, 2, \dots, N+2m. \quad (3.5)$$

This basis was introduced by Schumaker (1969) for spline

interpolation problems and in this case it yielded tridiagonal matrices. When dealing with least squares approximations as in our problem, a matrix is obtained which is no longer tridiagonal, but retains diagonal dominance. Therefore this basis is particularly effective in our problem.

ANALYSIS OF THE ALGORITHM

Let's now outline the algorithm step by step.

Step 1 Sampling range.

We must single out an interval for the sampling of ξ_i , $i=1, 2, \dots, r$. As discussed previously, θ_l and θ_u are used as an approximation respectively of f^* and $\|f\|_\infty$.

We perform a sampling of q points in S and evaluate $f(x)$ at each of them. An arithmetic mean of function values obtained in this way will be assumed as an approximation of θ_u .

It is obvious that, when looking for a global minimum, it is not important to extend the range of $f(x)$ upwards, and, on the contrary, in our case, we prefer a higher density of points neighbouring f^* . A more delicate problem is the choice of θ_l .

We proceed as follows: the amplitude of the interval between θ_u and the minimum value obtained for $f(x)$ is found out; it is multiplied by a proper parameter greater than 1 (depending on the particular function $f(x)$ one is minimizing) and this is assumed as the amplitude of $[\theta_l, \theta_u]$.

Step 2 Sampling strategy.

At every sample a random value $\xi_i \in [\theta_l, \theta_u]$ and q random points (q n -tuples of random values) in S are generated; for every ξ_i we evaluate the frequency $\psi(\xi_i)$ (i.e. $m(E(\xi_i))$) by means of $\tilde{\psi}_i = p_i/q$.

Step 3 Updating of frequencies.

To prevent an anomalous sample from influencing the approximation too strongly, the frequencies obtained up to now are updated by the rule:

$$\tilde{\psi}_j^i = \frac{\tilde{\psi}_{j-1}^{i-1}(i-1) + p_j}{i}, \quad i \geq 2, \quad (4.1)$$

where $\tilde{\psi}_j^i$ is the approximation obtained at sample i for $\psi(\xi_j)$.

Step 4 Spline approximation at the i -th sample.

At the i -th sample we have:

$$s_N^i(\xi) = \sum_{j=1}^{N+2m} \lambda_j \phi_j(\xi) \quad (4.2)$$

where $\phi_j(\xi)$, $j=1, 2, \dots, N+2m$, are the basic functions discussed in previous section.

To calculate coefficients λ_j^i we solve the linear system $C^i \lambda = b^i$ obtained when minimizing

$$J^R(\lambda) = \sum_{\ell=1}^i \{ \tilde{\psi}_\ell - \sum_{j=1}^{N+2m} \lambda_j \phi_j(\xi_\ell) \}^2. \quad (4.3)$$

Indeed, we must have:

$$\frac{\partial J(\lambda)}{\partial \lambda_k} = 0, \quad k=1, 2, \dots, N+2m.$$

And this means

$$2 \sum_{\ell=1}^i \{ \tilde{\psi}_\ell - \sum_{j=1}^{N+2m} \lambda_j \phi_j(\xi_\ell) \} \phi_k(\xi_\ell) = 0, \quad k=1, 2, \dots, N+2m. \quad (4.4)$$

This yields:

$$\sum_{\ell=1}^i \tilde{\psi}_\ell \phi_k(\xi_\ell) = \sum_{j=1}^{N+2m} \lambda_j \sum_{\ell=1}^i \phi_j(\xi_\ell) \phi_k(\xi_\ell), \quad k=1, 2, \dots, N+2m \quad (4.5)$$

Then we have $C^i = \{a_{pq}^i\}$ and $b^i = \{b_p^i\}$, where

$$a_{pq}^i = \sum_{\ell=1}^i \phi_p(\xi_\ell) \phi_q(\xi_\ell); \quad b_p^i = \sum_{\ell=1}^i \phi_p(\xi_\ell) \tilde{\psi}_\ell. \quad (4.6)$$

We remark that updating of frequencies $\tilde{\psi}_i$ modifies every component of b^i at each sample, but it does not anyway affect C^i .

This is a very important fact because it allows us to reduce computation. Indeed, we do not need inversion of C^i at every sample to evaluate coefficients λ_j , $j=1, 2, \dots, N+2m$, until the number $N(i)$ of knots is increased.

As a matter of fact, we have

$$C_N^{i+1} = C_N^i + \Phi(\xi_{i+1}) \Phi^T(\xi_{i+1}), \quad (4.7)$$

where

$$\Phi(\xi) = \{\phi_1(\xi), \phi_2(\xi), \dots, \phi_{N+2m}(\xi)\},$$

and this allows a recursive technique to update $(C_N^{i+1})^{-1}$ on the basis of $(C_N^i)^{-1}$ by means of the formula:

$$(C_N^{i+1})^{-1} = (C_N^i)^{-1} - \frac{(C_N^i)^{-1} \Phi(\xi_{i+1}) \Phi^T(\xi_{i+1}) (C_N^i)^{-1}}{1 + \Phi^T(\xi_{i+1}) (C_N^i)^{-1} \Phi(\xi_{i+1})} \quad (4.8)$$

With this device we solve our system with a number of operations of the order of $2(N+2m)^2$ against $(N+2m)^3/3$ of usual solution.

Step 5 Termination criteria.

A crucial point in the algorithm is the choice of proper termination criteria. A first test is made on the convergence of the coefficients λ_j , $j=1, 2, \dots, N+2m$, of the basic spline functions, in order to verify whether, with a fixed number of knots, information deriving from each new sample is completely exploited.

This means that, given a precision η , before changing the number of knots, the condition

$$\frac{|\lambda_j^i - \lambda_j^{i-1}|}{|\lambda_j^{i-1}|} \leq \eta \quad j=1, 2, \dots, N+2m \quad (4.9)$$

must be verified, where λ_j^i denotes the j -th component of the vector λ at the i -th sample.

The roots (in the sense of (2.2)) of the approximation to $\psi(\xi)$ corresponding to every reached convergence of the vectors λ^i are composed in their mean value λ_{mean} by the law:

$$\beta_{\text{mean}}^{i+1} = \frac{1}{i+1} \{ \beta_{\text{mean}}^i \cdot i + \beta^{i+1} \} \quad (4.10)$$

If we meet for a certain fixed number of times a β^{i+1} which is "far" from β_{mean}^i i.e. if

$$\frac{|\beta^{i+1} - \beta_{\text{mean}}^i|}{|\beta^{i+1}|} \geq 1 \quad (4.11)$$

we give less relevance to β_{mean}^i by a new updating formula:

$$\beta_{\text{mean}}^{i+1} = \frac{\beta_{\text{mean}}^i \cdot i' + \beta^{i+1}}{i'+1}$$

where the first value for i' is 1.

The reason for doing this only after several discrepancies between the new nihilating point β^{i+1} and β_{mean}^i is that, if it happens only once, possibly the anomalous result is β^{i+1} .

The number β_{mean} is used for a test of convergence of β^i to f^* . We check whether

$$\frac{|\beta_{\text{mean}} - \theta_\ell|}{|\theta_u - \theta_\ell|} < \epsilon \quad (4.12)$$

where $\epsilon > 0$ is a fixed precision.

The reason to perform this kind of test (and to wait until it is verified more than once) is that it, somehow, ensures β_{mean} , which is an extrapolation result, to stay within a "small" (with respect to $\theta_u - \theta_\ell$) neighborhood of the sampled minimum.

Another obvious criterion which leads to termination is verified when the function $J(\lambda)$ is less than a prefixed precision.

When at least one of the last two exposed criteria is fulfilled, the algorithm performs some local searches, starting from a prefixed number of points (trial number), selected among the best points obtained along sampling, to the extent of calculating the actual value F_c and coordinates x^* of the global optimum.

If we have

$$\frac{|F_c - \beta^*|}{|\theta_u - F_c|} \leq \alpha \quad (4.13)$$

where β^* is the last β_{mean} and α a fixed precision, the algorithm terminates and (x_c, F_c) (where x_c is the actually calculated x^*) is accepted as final output.

Otherwise the strategy is started again from step 2 within a fixed number of function evaluations.

COMPUTATIONAL RESULTS

This algorithm was tried on some test functions from the literature and on some technological problems (Archetti-Frontini (1977)).

1) "six hump camel back function" (Branin (1972)):

$$f(x) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4.$$

This function has six minima, two maxima and seven saddle points. There are two global minima in which $f(x) = -1.0316$, and they are attained at $(0.0898, -0.7126)$ and $(0.7126, -0.0898)$.

Our minimization is performed over the compact set $-2.5 \leq x_1 \leq 2.5$, $-1.5 \leq x_2 \leq 1.5$.

The results we obtained are shown in Table A.

2) Branin (RCOS) (1972):

$$f(x) = (x_2 - \frac{5.1}{4\pi^2} x_1^2 + \frac{5}{\pi} x_1 - 6)^2 + 10(1 - \frac{1}{8\pi}) \cos x_1 + 10.$$

This function has three global minima in the region $-5 \leq x_1 \leq 10$, $0 \leq x_2 \leq 15$.

3) Goldstein&Price (GOLDPR) (1972):

$$f(x) = [1 + (x_1 + x_2 + 1)^2 \cdot (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] \cdot [30 + (2x_1 - 3x_2)^2 \cdot (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)].$$

This function has 4 local minima in the region

$$-2 \leq x_1 \leq 2, -2 \leq x_2 \leq 2.$$

The global minimum is $f(x)=3$, at $x_1=0, x_2=-1$.

A direct comparison to evaluate effectiveness of this procedure can be made by Hartman's method (1973), using some of his test problems. Hartman's method has no global convergence test; this algorithm, on the contrary, spends a number of function evaluations to satisfy internal and external tests even if the problem has already been solved. The objective function is:

$$4) \quad f(x) = - \sum_{i=1}^m c_i \exp\{(x-p_i)^T A(x-p_i)\},$$

where c_i , $i=1, 2, \dots, m$ are the levels of the m minima and p_i , $i=1, 2, \dots, m$, their position vectors.

Shape and amplitude of attraction regions are determined by matrix A (n rows and n columns, n being the dimension of the problem), which has to be negative definite.

Also for this test family some trials were performed, with results shown in table A.

Another classical family of test functions was proposed by Shekel (1973), with the form:

$$5) \quad f(x) = - \sum_{i=1}^m \frac{1}{\|x-a^i\|^2 + c_i}.$$

This function has m minima in positions a^i ($i=1, 2, \dots, m$) with levels c_i .

Note: interaction among minima in Hartman's family acts heavily on the levels of minima but does not affect their positions, while in Shekel's family positions differ a lot from our a^i , $i=1, 2, \dots, m$, but levels are sufficiently preserved.

Regarding this method an important feature becomes evident from the trials performed: the number of function evaluations needed is almost insensitive to the dimension of the problem and to the number of minima.

| Function | N | r | q | F.E. | Total F.E. | β^* | F_C | $\left \frac{F_C - \beta^*}{\theta_u - F_C} \right $ | time (prime minutes) |
|--------------------|---|----|----|------|------------|-----------|-----------|---|----------------------|
| S.H.C.B.F. | 2 | 67 | 10 | 670 | 717 | -1.13418 | -1.031629 | 0.024 | 2.09 |
| Branin RCOS | 2 | 37 | 15 | 555 | 597 | 0.9433 | 1.250 | 0.0061 | 1.37 |
| Goldstein&Price | 2 | 21 | 15 | 315 | 378 | 4.6044 | 2.9997 | 0.5×10^{-5} | 1.51 |
| Hartman (4 minima) | 3 | 45 | 15 | 675 | 732 | -3.8602 | -3.8627 | 0.0008 | 1.58 |
| Hartman (4 minima) | 6 | 47 | 15 | 705 | 807 | -3.2013 | -3.3223 | 0.04 | 2.08 |
| Shekel (5 minima) | 4 | 36 | 15 | 540 | 620 | -1.534 | -10.14 | 0.9 (*) | 2.35 |
| Shekel (7 minima) | 4 | 47 | 15 | 705 | 788 | -1.53 | -10.399 | 0.8 (*) | 2.07 |
| Shekel (10 minima) | 4 | 54 | 20 | 1080 | 1160 | -1.67 | -10.43 | 0.87 (*) | 3.03 |

Table A

(*) Termination reached by the test $J^r(\lambda) \leq 10^{-6}$.

The time required for 1000 Shekel 5 evaluations at (4., 4., 4., 4.) was 0.09 prime minutes.

We remark that in table A we quoted two different numbers of function evaluations: the first is the number of function evaluations required by our method without local searches, the second is the total number of them.

In table B we show the behaviour of β_{mean} as r increases, for all the functions described in table A.

| r | S.H.C.B.F. | Branin | RCOS | Goldst. & Price | Hartman 4;3 | Hartman 4;6 | Shekel 5;4 | Shekel 7;4 | Shekel 10;4 |
|-----|------------|--------|--------|-----------------|-------------|-------------|------------|------------|-------------|
| 15 | -1.1622 | 1.3593 | 4.6050 | -3.6717 | -2.6210 | -1.393 | -1.422 | -1.098 | |
| 20 | -1.1622 | 1.3593 | 4.6044 | -3.6717 | -2.6463 | -1.393 | -1.422 | -1.098 | |
| 25 | -1.1622 | 1.3597 | | -3.6720 | -2.6453 | -1.393 | -1.490 | -1.773 | |
| 30 | -1.1622 | 0.5197 | | -3.6945 | -2.6453 | -1.393 | -1.448 | -1.792 | |
| 35 | -1.1622 | 0.9433 | | -3.6945 | -2.6508 | -1.534 | -1.465 | -1.776 | |
| 40 | -1.1622 | | | -3.8602 | -2.9810 | | -1.456 | -1.776 | |
| 45 | -1.1622 | | | -3.8602 | -3.1512 | | -1.528 | -1.721 | |
| 50 | -1.1622 | | | | | | -1.690 | | |
| 55 | -1.1627 | | | | | | | -1.668 | |
| 60 | -1.1627 | | | | | | | | |
| 65 | -1.1341 | | | | | | | | |

Table B

It seems on the contrary dependent on the shape of the function. Indeed, if the function has very sharp minima the number of function evaluations increases noticeably; if it is very smooth very few function evaluations are needed, even for high dimension problems. Comparing our algorithm with Hartman's method A3, we can notice that for cases A, B, F, H our number of function evaluations is of the order of 50% of the average number needed by Hartman's method to reach the global minimum.

We remark that even if it is not explicitly stressed in table A, all trials reached the position of the global minimum.

We point out, however, that our main interest in testing this method was in obtaining a good prediction of the level of global optima; this will be clear from the close checks introduced into the algorithm described at the end of the previous section.

If our goal were a dependable answer after a small number of function evaluations, this total number could significantly be reduced.

It is worth noticing that this control on the goodness of the prevision was violated an insignificant number of times during all the trials we performed.

ACKNOWLEDGEMENTS

We thank Professor M. Cugiani for reading the draft of this paper. We are also grateful to Dr. J. Gomulka and Dr. L.C.W. Dixon for sending test functions and for fruitful discussions.

REFERENCES

- Archetti, F. (1975). A sampling technique for global optimization; in: Towards Global Optimization, Szego & Dixon eds., North-Holland Publishing Company.
- Archetti, F. and Betrò, B. (1976). Recursive stochastic evaluation of the level set measure in global optimization problems; quaderni

- del Dipartimento di Ricerca Operativa e Scienze Statistiche, A21, Università di Pisa.
- Archetti, F. and Frontini, F. (1977). A global optimization method and its application to technological problems. This book.
- Betrò, B. and De Biase, L. (1976). A recursive spline technique for uniform approximation of sampled data; quaderni del Dipartimento di Ricerca Operativa e Scienze Statistiche, A31, Università di Pisa.
- Branin, F.H. (1972). A widely convergent method for finding multiple solutions of simultaneous equations; Tech. Rept. 21466, I.B.M. S.D. D.L., I.B.M. Journal on Research & Development.
- Christensen, T. and Hartman, K. (1974). A strategy of optimization based on random search; Proceedings of I.F.A.C. Symposium on Stochastic Control, Budapest.
- Frontini, F. (1975). Un metodo automatico per l' ottimizzazione globale in problemi di progettazione; C.I.S.E. Rept. N-172.
- Hartman, K. (1973). Some experiments in global optimization; Naval Postgraduate School, Tech. Rept. NPS55HH72051/A.
- Hartman, K. (1973). A new method for global optimization; Naval Postgraduate School, Tech. Rept. NPS55HH73041/A.
- Mikhail'skii, A.J. (1974). The method of averaged splines in the problem of approximating dependencies on the basis of empirical data ; Automation and Remote Control, Vol. 35, № 3.
- Opacic, J. (1973). A heuristic method for finding most extrema of a nonlinear function; IEEE Transactions on Systems, Man and Cybernetics; 102, 107.
- Rudin, W. (1970). Real and complex analysis, Mc. Graw Hill.
- Schumaker, L.L. (1969). Theory and applications of spline functions; Greville ed. (Academic Press, New York).

A MIXED STOCHASTIC-DETERMINISTIC TECHNIQUE FOR GLOBAL OPTIMISATION

Enrico Fagioli ^(°) - Paolo Pianca ^(°) - Marco Zecchin ^(°°)

ABSTRACT. A space-covering method for global optimisation is presented, based on the construction of a stochastic automaton. Three possible implementations of the method are suggested. Some numerical results are given for two of these algorithms.

ACKNOWLEDGEMENTS. This study has been partially done during the visit of E. Fagioli at the Numerical Optimisation Center of Hatfield Polytechnic, in the framework of the Italo-British research project on global optimisation techniques sponsored by S.R.C. and C.N.R.

(°) University of Venice.

(°°) University of Trieste.