Global Optimization
With the Use of Optimal Covering

Adam Woźniak

Warsaw University of Technology, Institute of Control and Computation Eng., Warsaw, Poland,
e-mail: wozniak@ia.pw.edu.pl

Abstract. The paper describes global optimization algorithm based on Stratified Covering. Stratified means that feasible set is divided into M disjoint subsets of equal volume, and in each subset N sampling points are uniformly generated. Covering concerns method of uniform generation of points and means that sampling grid is the set of centers of N balls, which cover in the finest manner the subset. An abridged description of optimal stratified sampling and optimal covering algorithms containing only the essential of the methods is presented. For the purpose of illustrating both the actual working and the potentialities of the method, a set of computational results is presented.

1 Introduction

The simplest method of finding an extremum of function f numerically given on a feasible compact set $\boldsymbol{X}$ in n-dimensional space $\mathbb{R}^n$ is the Simple Monte Carlo Method (SMC Method) called also the Independent Sample Method (IS Method). The calculations involve two steps. The first one is the independent generation of N sampling points $\{x^1,..., x^N\}$ uniformly distributed on $\boldsymbol{X}$. Next, the best of them is taken as a solution $\bar{x}(N)$. For minimization problem it means that

$$\bar{x}(N) = \arg\min_{x^i \in X(N)} f(x^i)$$

and $\bar{x}(N)$ is the approximation of true minimizer x^*. It is obvious that accuracy of approximation depends on length N of the grid $\{x^1,..., x^N\} = X(N)$ and on the way in which points fill up the set $\boldsymbol{X}$. The IS Method is the simplest one, but many examples showed that its efficiency is rather moderate and there is a need for devising more efficient sampling scheme.

As mentioned above, two factors impact on grid efficiency: the length of sample N and its uniformity. Let us discuss shortly influence of the first factor. Basing on simple probabilistic considerations it is easy to show (e.g. [7]) that to obtain with given probability $\gamma \in (0,1)$ the solution $\bar{x}(N)$ with appropriately defined (volume) accuracy $\varepsilon \in (0,1)$ we have to choose number of samples N such that

$$N(\varepsilon, \gamma) = \mathrm{ceil}[\frac{\log(1-\gamma)}{\log(1-\varepsilon)}]\,. \tag{1}$$

Many different definitions of measures of filling quality (uniformity) exist in the literature. When we think about point of the grid, x^i, as center of a ball, $B(x^i,\rho)$ (in some metric ρ), it seems that situation when points are arranged in such a way that union of equal balls with minimal radius covers feasible set, $\boldsymbol{X}$, is a good filling. The situation is referred as covering by equal balls (with minimal radius). The next section gives formal definition and presents basic properties of point set defined by coverings of this type.

2 Grids in Global Optimization, Uniformity Characteristic

Let ρ be a metric on $\mathbb{R}^n$, feasible set $\boldsymbol{X} \subset \mathbb{R}^n$ be compact and $X(N) = \{x^1,..., x^N\}$ be N-element grid in $\boldsymbol{X}$. With each point of the grid we associate the ball of radius r with the center at x^i

$$B(x^i, r) = \{x \mid \rho(x^i, x) \le r\}.$$

The system of balls $\{B(x^i, r)\}_{x^i \in X(N)}$ with radius r, covers the set $\boldsymbol{X}$ when

$$\boldsymbol{X} \subseteq \bigcup\nolimits_{x^i \in X(N)} B(x^i, r).$$

Bearing in mind quoted above intuitive meaning of good filling, we define optimal grid, $X^*(N)$, as grid producing covering with minimal radius $\delta(\boldsymbol{X}, N)$

$$\delta(\boldsymbol{X}, N) = \min\nolimits_{X(N) \subseteq \boldsymbol{X}} \min\{r \mid \bigcup\nolimits_{x^i \in X(N)} B(x^i, r) \supseteq \boldsymbol{X}\}$$

$$X^*(N) = \arg\min\nolimits_{X(N) \subseteq \boldsymbol{X}} \min\{r \mid \bigcup\nolimits_{x^i \in X(N)} B(x^i, r) \supseteq \boldsymbol{X}\}.$$

One of the commonly used uniformity characteristics of point set $X(N)$ is its metric-dispersion, defined as

$$d(X(N), \boldsymbol{X}) = \max\nolimits_{x \in \boldsymbol{X}} \min\nolimits_{x^i \in X(N)} \rho(x, x^i).$$

It is easy to observe that

$$\min\nolimits_{X(N) \subseteq \boldsymbol{X}} d(X(N), \boldsymbol{X}) = \delta(\boldsymbol{X}, N).$$

Another reason for the significance of metric-dispersion as a measure of uniformity is the following property [3], [6], stated here in a little simplified version.

Theorem. Let ρ be a metric on $\mathbb{R}^n$, a feasible set $\boldsymbol{X} \subset \mathbb{R}^n$ be compact, a function f from $\mathbb{R}^n$, be Lipschitz-continuous on $\boldsymbol{X}$ with constant L ($f \in Lip(\boldsymbol{X}, L)$) and N be fixed.
Then

$$\min\nolimits_{X(N) \subseteq \boldsymbol{X}} \sup\nolimits_{f \in Lip(\boldsymbol{X},L)} | f^* - f(\bar{x}(N)) | = L \min\nolimits_{X(N) \subseteq \boldsymbol{X}} d(x(N), \boldsymbol{X}) = L\, \delta(\boldsymbol{X}, N), \qquad (2)$$

where

$$f^* = \min\nolimits_{\boldsymbol{X}} f(\cdot) \;\text{ and }\; f(\bar{x}(N)) = \min\nolimits_{x^i \in X(N)} f(x^i).$$

In words: "the covering with minimal radius gives worst-case optimal N-element grid".

It is interesting to know the dependence of lower bound of grid fitness defined by (2) on grid length N. When feasible set $\boldsymbol{X}$ is a unit cube $\boldsymbol{K}_n$, H. Niederreiter [3] proved that

$$\delta(\boldsymbol{K}_n, N) \ge \frac{1}{(V_B N)^{\frac{1}{n}}},$$

where V_B is the volume of unit ball in $(\mathbb{R}^n, \rho)$. It means that for plane with Euclidean metric

$$\delta(\boldsymbol{K}_2, N) \ge \frac{1}{\sqrt{\pi}\sqrt{N}} \cong 0.5642 \frac{1}{\sqrt{N}},$$

so: $\delta(\boldsymbol{K}_2,10) \ge 0.1784$ and $\delta(\boldsymbol{K}_2,100) \ge 0.0564$. As we will see later, the real minimal radius $\delta(\boldsymbol{K}_2,10)$ is only little greater than bound. For 3-dimansional space with Euclidean metric

$$\delta(\boldsymbol{K}_3, N) \geq \frac{1}{\sqrt[3]{\frac{4}{3}\pi}\sqrt[3]{N}} \cong 0.6204\frac{1}{\sqrt[3]{N}}\ .$$

As expected, this bound decreases with increase of N slower than previous one.

3 Optimal Covering

When we adopt J. Kifer [2] worst case approach to assessing the efficiency of optimization algorithm, it follows from presented theorem that minimax problem of devising optimal sampling sequence is equivalent to the problem of finding optimal covering. But, is this new problem solvable? The problem belongs to rather hard issues of discrete geometry, and the positive answer to our question is restricted to special classes of covering problem. First assumption is that feasible set $\boldsymbol{X}$ has simple structure. Particularly it is required that $\boldsymbol{X}$ is n-dimensional cube. It is not restrictive assumption for our global optimization problem, because we always can imbed original irregular feasible set in a suitable cube. With this assumption the problem of optimal covering of interval is the simplest, and its solution is trivial. The next is the problem of optimal covering of a square. In the sequel the essentials of algorithm solving this problem are presented. The algorithm derives from theoretical prototype presented by S.A. Brusov and S.A. Pijavski in 1970 [1]. Algorithm for n-dimensional cube was worked out also [5], but its efficiency is still unsatisfactory and improvements are needed.

Before we present the essential of the algorithm we have to state some definitions.

Let ρ be Euclidean metric, $\boldsymbol{K}$ denotes the unit square, $\boldsymbol{K} = [0,1]\times[0,1] \ni x = (x_1, x_2)$ and $X(N)$ be a grid in $\boldsymbol{K}$. The set $D(x^i)$ of all points in $\boldsymbol{K}$ closer or equally distant to a point x^i of $X(N)$ than to any other point of $X(N)$, that is closed convex polygon

$$D(x^i) = \{x \in \boldsymbol{K} \mid (\forall x^k \in X(N), x^k \neq x^i)\, \rho(x, x^i) \leq \rho(x, x^k)\},$$

is called the Dirichlet domain (its interior is recently known as Voronoi cell) for x^i.

We define the following sets:

$$\boldsymbol{K}_V = \{(0,0),(1,0),(1,1),(0,1)\},$$

$$L_1 = \{x \mid (\exists 0 \leq x_1 \leq 1)\ x = (x_1, 0)\}, \quad L_2 = \{x \mid (\exists 0 \leq x_1 \leq 1)\ x = (x_1, 1)\},$$

$$L_3 = \{x \mid (\exists 0 \leq x_2 \leq 1)\ x = (0, x_2)\}, \quad L_4 = \{x \mid (\exists 0 \leq x_2 \leq 1)\ x = (1, x_2)\}.$$

Let point x^i in $X(N)$ be selected. The sets below are related to this point

$$V(x^i) = \{x \mid (\forall x^k \in X(N), x^k \neq x^i)\, \rho(x, x^i) = \rho(x, x^k)\},$$

$$W(x^i) = \{x \mid (\forall (x^k, x^j) \in X(N) \times X(N), x^k \neq x^i, x^j \neq x^i x^k \neq x^j)\, \rho(x, x^i) = \rho(x, x^k) = \rho(x, x^j)\},$$

$$V_P(x^i) = \boldsymbol{K}_V \cup (V(x^i) \cap L_1) \cup (V(x^i) \cap L_2) \cup (V(x^i) \cap L_3) \cup (V(x^i) \cap L_4) \cup W(x^i).$$

The system of equations describing point set $V_P(x^i)$ is quadratic (metric is Euclidean) and direct formulae for its solution can be easily obtained.

Now it is easy to observe that vertices of $D(x^i)$ can be detected from set $V_P(x^i)$ as points v meeting conditions defining Dirichlet domain

$$VD(x^i) = V_P(x^i) \cap \{v \mid (\forall x^k \in X(N), x^k \neq x^i)\, \rho(v, x^i) \leq \rho(v, x^k)\}\ .$$ [1]

Having the set of vertices $VD(x^i)$ we can calculate the radius $r(x^i)$ of the circle, $B(x^i, r(x^i))$, circumscribing $D(x^i)$ with center at x^i as

---

[1] Family of sets $\{VD(x^i)\}_{x^i \in X(N)}$ gives, so called, Voronoi tessellation corresponding to the set $X(N)$.

$$r(x^i) = \max_{v \in VD(x^i)} \rho(v, x^i) \qquad v(x^i) = \arg\max_{v \in VD(x^i)} \rho(v, x^i).$$

The obtained system of circles $\{B(x^i, r(x^i))\}_{x^i \in X(N)}$ clearly covers $\boldsymbol{K}$, but circles have different radii. Therefore the next step of algorithm consists in improvement of a starting grid $X^{(k)}(N)$ to better one $X^{(k+1)}(N)$ with

$$\max_{x \in X^{(k+1)}(N)} r(x) < \max_{x \in X^{(k)}(N)} r(x), \tag{3}$$

using appropriately defined perturbation of suitable chosen $x^i \in X^{(k)}(N)$.

Due to lack of space we do not present details of choosing x^i for shift and method of grid improvement ensuring (3). For details we refer the reader to [1] or [5]. We point out only, that method of improvement is based on linearization of distance increments and somewhat complicated definition of feasible shifts set.

General-description of optimal covering algorithm

Initialization step. Choose integer N and initial grid $X^{(0)}(N)$ in $\boldsymbol{K}$; set k := 0.

Main step. Repeat until stopping criteria satisfied.

With $X(N) = X^{(k)}(N)$ for each x^i in $X(N)$ compute $VD(x^i)$ and $r(x^i)$.

Choose point p from $X(N)$ to shift, and move it to new place $p^{(k)}$ in such a way that

$$r(p^{(k)}) < r(p).$$

Set $X^{(k+1)}(N) = (X(N) \setminus \{p\}) \cup \{p^{(k)}\}$ and k := k +1.

Algorithm described above in essentials was implemented in detail in cooperation with M. Pysiak in MATLAB. Figure 1 shows graphic depictions of results obtained for N = 10 and

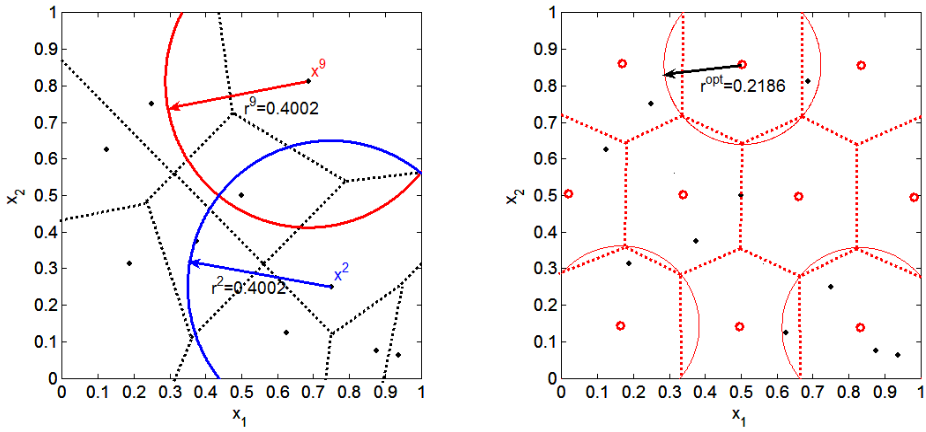

Figure 1. Voronoi tessellation and minimal covering radius for 10 sampling points of Sobol sequence (left) and for calculated optimal covering (right).

Sobol quasi-random sequence as initial grid. The optimal covering was obtained after 1600 main iterations. So large number of iteration was caused by exceptionally poor choice of initial grid. Gathered experience has showed that, when initial grid is more uniformly dispersed on the square, the number of iterations considerably decrease.

The discussed example is interesting from another point of view. Basing on simple considerations on symmetry of the ten-circle set I derived, probably the best covering.[2] The coordinates of this grid are showed in Table 1 together with those obtained by presented algorithm.

Table 1. Comparison of coverings and minimal radii.

Optimal (?) covering		Obtained covering
$a = 1/3 - 1/9\sqrt{3}$	Decimal expansion	
$(1/6, a)$	(0.16666666666667, 0.14088324360346)	(0.1644, 0.1435)
$(1/2, a)$	(0.5, 0.14088324360346)	(0.4963, 0.1406)
$(5/6, a)$	(0.83333333333333, 0.14088324360346)	(0.8319, 0.1390)
$(0, 1/2)$	(*0, 0.5*)	(*0.0174, 0.5053*)
$(1/3, 1/2)$	(0.33333333333333, 0.5)	(0.3382, 0.5016)
$(2/3, 1/2)$	(0.66666666666667, 0.5)	(0.6607, 0.4971)
$(1, 1/2)$	(*1, 0.5*)	(*0.9826, 0.4961*)
$(1/6, 1-a)$	(0.16666666666667, 0.85911675639654)	(0.1689, 0.8612)
$(1/2, 1-a)$	(0.5, 0.85911675639654)	(0.5036, 0.8586)
$(5/6, 1-a)$	(0.83333333333333, 0.85911675639654)	(0.8347, 0.8570)
$\delta(\boldsymbol{K}_2, 10) =$ $= -1/6 + 2/9\sqrt{3}$	0.218233512793084	$r^{\mathrm{opt}} =$ $= 0.2186$

Comparing figures in second and third column of Table 1 it is easy to conclude that output of presented algorithm well coincides with optimal points even for so hard instances as these with first coordinate on bounds of unit segment.

In the paper we use Optimal Covering Algorithm as a tool for generation of sampling points in global optimization. Of course it can be used for solving plethora of different problems. For example in the mobile communication, the problem of positioning the base-stations and the assignment of transmission ranges such the entire given area is covered and some fitness function dependent on distances between stations has to be minimized is in fact optimal covering problem.

4 Stratified Covering Algorithm

At the end of Introduction we quote formula describing dependence of sampling sequence length, N, on desired accuracy. The use of this formula needs explanation. It uses notions from

[2] Obtained radius of covering is the same as computed in [4], where time consuming algorithm for optimal covering is presented.

probability theory, hence accuracy used is, so-called, volume accuracy (measure). It means that when we demand ε-accuracy with probability γ, in fact we demand that ratio of the "volume" in which with probability γ optimal solution occurs to the "volume" of the whole feasible set ***X*** equals ε. Usually we reason using the measure of "length" instead of "volume"; therefore we ought to translate accuracy in "volume measure" to accuracy in "length measure". As we know for n-dimensional cube with length of edge α, volume equals α^n. It means that for feasible set which is hypercube, $\boldsymbol{X} = [a, b]^n$, $a < b$, when we intend to localize optimal solution with accuracy $1/\alpha$ of range $b - a$ of each variable, we should put $\varepsilon = (1/\alpha)^n$.

In our 2-dimensional case it means, that for "length accuracy" $1/\alpha = 10^{-4}$, $\varepsilon = 10^{-8}$ and for probability $\gamma = 0.01$ formula (1) gives $N = 1\ 005\ 034$. To generate randomly more than million points is possible, but it seems that at present and in near future to find optimal covering by million circles is impossible.

Classical way out of this difficult situation is to use combination of stratified sampling and clustering. We follow this way.

In the sequel, for the sake of simplicity, we assume that in considered optimization problem feasible set is a square, $\boldsymbol{X} = [a, b]^n$, $a < b$, and minimized function f is continuous. The generalization to feasible set with any shape is straightforward.

Below we shall describe in simplified form proposed global optimization algorithm based on Stratified Covering.

Simplified-description of stratified covering algorithm

Initialization step. Choose length of basic sample, N. Compute (import) optimal covering $X^*(N)$ of unit square. Choose number of each level strata, R. Divide square ***X*** into $M = R^2$ equal sub-squares $C_m^{(1)}, m = 1,...,M$.
Set $S^{(1)} = M$, and k := 1.

Main step. Repeat until stopping criteria satisfied. When stop, adopt current $\bar{x}$ as approximation of true minimizer.

With $A_m = C_m^{(k)}$, $m = 1,...,S^{(k)}$ rescale $X^*(N)$ to bounds of A_m , and $XA_m(N)$, $m = 1,...,S^{(k)}$, be the grid obtained.

Evaluate function f in all points of k-level grid $G^{(k)} = \bigcup_{1 \le m \le S^{(k)}} XA_m(N)$.

Using adopted measures based on evaluations of f , identify Q promising sub-squares, i.e., squares where perhaps minimum lies, and $\{S_1,...,S_Q\}$ be the set of sub-squares obtained.

Divide each sub-square S_q into M sub-subsquares, and $C_1^{(k+1)},...,C_{Q \cdot M}^{(k+1)}$ be the set of sub-subsquares obtained.

Set $\bar{x} = \arg\min_{x \in G^{(k)}} f(x)$, set $S^{(k+1)} = Q \cdot M$ and k := k +1.

Different modifications of this prototype algorithm are possible, and we leave them to readers' invention. The key issue is a way in which sub-squares will be counted to promising ones.

In order to illustrate the functioning of the method, we present computational result on one small test problem, which optimal solutions are known beforehand. The so-called Himmelblau test was chosen. It consists in finding all minimizers of the function

$$(x_1,x_2) \mapsto f(x_1,x_2) = ((x_1)^2 + x_2 - 11)^2 + (x_1 + (x_2)^2 - 7)^2.$$

In fact, Himmelblau test is the problem of fining solution of the system of quadratic equations:

$$\begin{cases} (x_1)^2 + x_2 - 11 = 0 \\ x_1 + (x_2)^2 - 7 = 0 \end{cases}.$$

The system has perhaps four real solutions (one obvious is (3,2)) and as a consequence function f has potentially four global minima with value 0.

We assume, that initial square is rather large: $\boldsymbol{X} = [-100,100]\times[-100,100]$.

Before we present obtained results we calculate number of samples guaranteeing according to formula (1), length accuracy 0.01 (what means in the case that $\varepsilon = 2.5\cdot 10^{-9}$) with probability 0.9. It equals near billion, exactly 921 034 042 points. And how it will be looking for stratified covering algorithm?

For computational test the following figures for main parameters of the algorithm were chosen:

- length of basic sample $N = 9$;
- covering generated by optimal covering algorithm
  $X^*(9) = \{(0.167, 0.833)\ (0.5, 0.833)\ (0.833, 0.833)\ (0.167, 0.5)$
  $(0.5, 0.5)\ (0.833, 0.5)\ (0.167, 0.167)\ (0.5, 0.167)\ (0.833, 0.167)\}$;
- number of each level strata $R = 8$;
- maximal number of steps $T = 8$.

Left part of figure 2 depicts contours of minimized function together with underlying parabolas.

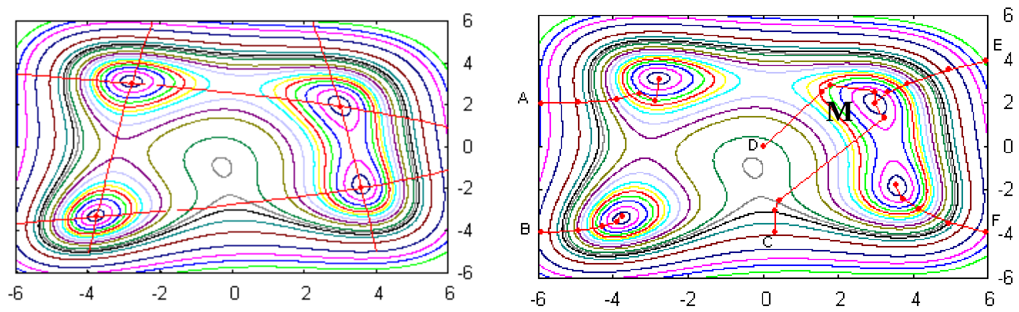

Figure 2. Contour plot of Himmelblau function near extrema (left) and iterations of quasi-Newton minimization (right).

It really has four minima and its shape with four basins of attraction makes it troublesome for gradient algorithms, what depicts right part of figure 2 where iterations of MATLAB `fminunc` function that uses BFGS quasi-Newton method of optimization are presented. To find all minima the starting points for multistart search have to be carefully selected because for gradient methods basin of attraction of minimum **M**, (3,2), is greater than others are.

Table 2 shows solutions obtained using implemented trial version of stratified covering algorithm. Their accuracy is very high, although during 26 iterations function was evaluated only in 26·(8·8·9) = 14 976 points.

Table 2. Test problem solution.

Point	Function value × 10^{-10}
(3.58442755937577, –1.84812693595886)	0.3662467076998
(2.99999947547913, 1.99999963045121)	0.1637786514753
(–3.77931174039840, –3.28318794965744)	2.1547638875875
(–2.80511790513992, 3.13131369352341)	0.5690748832914

In view of the result obtained, where the proposed algorithm performed efficiently, we believe that further research will lead to construction of the algorithm which be fast and robust and can be used to solve complicated, practical problems.

Moreover, when the problems with 3-dimensional coverings will be overcome the area of application will be widened to optimization problems with three variables.

Bibliography

[1] Brusov, S.A., Piyavskiy S.A. Vychislitelnye algoritmy optimal’nogo pokrytiya oblastey ploskosti. *Zh. vychisl. mat. i mat. fiziki* 11(2): 304–312, 1971.

[2] Kiefer J. Sequential minimax search for a maximum. *Proc. Amer. Math. Soc.* 4(3): 502–508, 1953.

[3] Niderreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia, PA: SIAM, 1992.

[4] Nurmela K.J., Östergård, P.R.J., *Covering a Square With up to 30 Equal Circles*. Research Report HUT-TCS-A62, Helsinki University of Technology, Laboratory for Theoretical Computer Science, 2000.

[5] Pysiak, M., *Metody optymalizacji globalnej wykorzystującej pokrycia*. Eng. Dissertation, Warsaw University of Technology, Institute of Control and Computation Eng., 2007.

[6] Zhigljavsky A.A., *Theory of Global Random Search*. Dortrecht: Kluwer Academic, 1991.

[7] Zieliński R., Neumann P. *Sochastische Verfahren zur Suche nach dem Minimum einer Funktion*. Berlin: Akademie Verlag, 1983 (polish translation: *Stochastyczne metody poszukiwania minimum funkcji*. Warszawa: WNT, 1986).