КОНСТРУКТИВНАЯ ТЕОРИЯ ФУНКЦИЙ'84
София, 1984, с. 569—577

CONSTRUCTIVE THEORY OF FUNCTIONS'84
Sofia, 1984, p.569-577

ON PIECEWISE-LINEAR APPROXIMATION AND NUMERICAL MINIMIZATION OF CONVEX FUNCTIONS
/ SOME NEW OPTIMIZATION ALGORITHMS II /

Andrzej Maćkiewicz

1.Introduction. The present paper describes a new algorithm for minimization of convex function $f:R^n \to R$ on an arbitrary n-dimensional simplex, contained in R^n, based on simplicial methods not needing the reformulating of the problem to the fixed point problem. This approach allows avoiding the following basic difficulties characteristic for classical nonlinear optimization methods

- irregular behaviour of the methods in the vinicity of the optimal point,
- necessity of determining of the derivatives of the function f
- necessity of minimizing along the line.

The exactness of our algorithm depends only on the lenghts of machine words. This algorithm may also be used for improving results obtained by classical methods.

2. Basic concepts. If $w^1,\dots,w^{t+1}$ are t+1 affinely independent points of R^n then the convex hull of these points is called a t-simplex or t-dimensional simplex with vertices $w^1,\dots,w^{t+1}$. Such a simplex is denoted by $\sigma(w^1,\dots,w^{t+1})$ or simply σ /if this does not lead to ambiguity/. In the sequel we will consider algorithms minimizing a real function f defined on an arbitrary n-simplex $\sigma(w^1,\dots,w^{n+1})\subsetneq R^n$.

However for theoretical considerations presented here it is irrelevant, numerical nuances may in practice couse the necessity of imposing additional restrictions on affinely independent vectors w^i $(1\leqslant i\leqslant n+1)$ Such a restriction may require, for example that norms of vectors $w^{k+1}-w^k$ /k=1,...,n/ be more or less equal.

The optimal points will be approximated succesively by subsimplices of the initial simplex σ or by barycentric centers of this subsimplices.

In our considerations the way of triangulations of simplex σ may not be arbitrary but must be carefuly chosen. In numerical method of finding fixed points the most popular and useful method of triangulation of the standard unit n-simplex S^n called a regular subdivision is described in Kuhn [2,3,4]. In this method every vertex of a subsimplex in a regular subdivision of this unit simplex can be expressed as $(x_1/D,\dots,x_{n+1}/D)$ where D is the degree of subdivision, a positive integer and $x_1,\dots,x_{n+1}$ are nonnegative integers that sum to D. We will use this method for defining the most convenient for us triangulation of simplex σ.

2.1.Definition. The standard subdivision of the degree D of the simplex $\sigma(w^1,\dots,w^{n+1})$ is given by points $z=(1/D)(x_1w^1+\dots x_{n+1}w^{n+1})$ (called triangulation vertices) for all x_k nonnegative integers such that $\sum_{k=1}^{n+1} x_k=D$. A set of n+1 points $z^1,\dots,z^{n+1}$ (of the above form) spans a subsimplex of the triangulation if we can order the points so that

$$
\begin{aligned}
&z^2=z^1+(1/D)(w^{k_1}-w^{k_1-1}),\\
&z^3=z^2+(1/D)(w^{k_2}-w^{k_2-1}),\\
/2.2/\quad &\dots\dots\dots\dots\dots\dots\\
&z^{n+1}=z^n+(1/D)(w^{k_n}-w^{k_n-1}),\\
&z^1=z^{n+1}+(1/D)(w^{k_{n+1}}-w^{k_{n+1}-1}),
\end{aligned}
$$

where w^k is the k-th vertex of $\sigma(w^1,\dots,w^{n+1})$ and $(k_1,\dots,k_{n+1})=:\pi$ is a permutation of $(1,\dots,n+1)$. We will assume by definition that $k_i-1:=n+1$ if $k_i=1$ i.e. we will assume cyclic order of indices k_j.

With the fixed degree D, if the point z^1 and the permutation π are known then the subsimplex described above will be denoted by $\sigma(z^1,\pi)$

Further the increasing sequence of natural numbers $\{D_j\}$ will be determined and regular subdivisions of simplex σ of degree D_j corresponding to subsequent elements of $\{D_j\}$ will be defined.Each of such triangulations will be characterized by the scalar

$$\text{mesh } M_j:=\sup_{\sigma\in M_j}\{\operatorname{diam}\sigma\}$$

Taking into account that for an arbitrary simplex $\sigma(z^1,\dots,z^{t+1})$ $\operatorname{diam}\sigma=\max_{1\leq i,j\leq t+1}(\|z^i-z^j\|)$ it is easily seen that mesh $M_j\to 0$ when $D_j\to\infty$.

Let us suppose now that $\sigma(z^1,\pi)$ is a subsimplex of the simplex obtained in triangulation M of degree D. Now we need a way of representing of the n-subsimplex adjacent to subsimplex $\sigma(z^1,\pi)$ if one of the vertices of the latter, say z^E, is replaced.

2.3. Pivot property. Given a subsimplex of the subdivision spanned by $z^1,\dots,z^{n+1}$ and a vertex z^E such that the opposite face does not lie on the boundary of $\sigma(w^1,\dots,w^{n+1})$, then there is a unique vertex $z^{E'}$ with $E'\neq E$ that spans a subsimplex with face opposite z^E and

$$z^{E'} = z^P + z^S - z^E,$$

where z^P and z^S are points immediately before and after z^E in the cyclic order.

Proof of that fact is analogus to the proof presented by Kuhn [3] for regular transformations of S^n. It is very important that in the described triangulations method the pivot operation is essentially trivial for this subdivision and does not need any information other than the $(n+1)$ points in the correct cyclic order.

The following theorem will be basic for us /see [1]/.

2.4. Theorem. Let $f:R^{n+1}\to R$ be a convex function, and S be a nonempty convex set in R^{n+1}. Consider the problem to minimize $f(z)$ subject to $z\in S$. The point $z^o\in S$ is an optimal solution to this problem if and only if f has a subgradient ξ at z^o such that $\xi^T(z-z^o)\geqslant 0$ for all $z\in S$.

It may be inferred from this theorem that in the case when f is differentiable this point is an optimal solution if and only if $\nabla f^T(z^o)(z-z^o)\geqslant 0$ for all $z\in S$.

Later we will describe a method of generating succesive approximations of the optimal point /which is the subject of the above theorem/ by subsimplices of triangulations M_j of the initial simplex of degree D_j /where $D_j\to\infty$/ or by barycentric centers of this subsimplices.

3. The principle of labeling and piecewise-linear approximation of a convex function.

Let now $\mathfrak{M}$ be a triangulation of R^n whose subset M is a triangulation of simplex σ /the existence of such triangulations follows from considerations of Todd [6]/.

3.1. Definition. Function $g:R^n\to R$ will be called piecewise-linear on the subdivision $\mathfrak{M}$ if

/3.1.1./ g is continuous,

/3.1.2./ given a subsimplex γ in $\mathfrak{M}$, there exists affine function $g_\gamma : R^n \rightarrow R$ such that g restricted to γ is g .

If the piecewise linear function g fullfils additional property

/3.1.3./ $f(z) = g(z) \; \forall z$ /z is a vertex of $\mathfrak{M}$ / we will call it a piecewise-linear approximation /p.l.a./ of function f on a subdivision $\mathfrak{M}$.

Of course from convexity of f follows convexity of its piecewise-linear approximation. Let us choose now an arbitrary point $z \in R^n$ which, according to the definition of triangulation of the space R^n /see [6] /, belongs to finite number of subsimplices $\sigma_1, \ldots, \sigma_r$ from $\mathfrak{M}$. Let g be a p.l.a. of function f and let $\nabla g_{\sigma_i}(z) = a_i$ /where ∇g_{σ_i} denotes the gradient of function g_{σ_i} . Then we define

$$/3.2./ \quad \partial g(z) = \text{hull}\{a_1, \ldots, a_r\}.$$

Let us note that $\partial g(z)$ is subdifferential of g at z /Rockafellar 7/

Now consider a subsimplex $\sigma(z^1, \pi)$ belonging to standard subdivision M_j, of degree D_j of simplex $\sigma(w^1, \ldots, w^{n+1})$. In order to find vector $\xi := (\xi_1, \ldots, \xi_n)^T = \nabla g_{\sigma(z^1, \pi)}$ the following system of linear equations may be considered

$$/3.3/ \quad \begin{aligned} (z^2 - z^1)^T \xi &= f(z^2) - f(z^1) \\ (z^3 - z^2)^T \xi &= f(z^3) - f(z^2) \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ (z^{n+1} - z^n)^T \xi &= f(z^{n+1}) - f(z^n) \\ (z^1 - z^{n+1})^T \xi &= f(z^1) - f(z^{n+1}) \end{aligned}$$

Each of the equations from the above system is linearly dependent on the remaining n equations. That is why an arbitrary equation may by omitted. The remaining equations form the system of n-linear independent equations with variables $\xi_1, \ldots, \xi_n$ and it may be solved by one of the classical methods.

The existence and the uniqueness of the solution of such a system of equations is ensured by the assumption of affine independence of vectors $z^1, \ldots, z^{n+1}$. Let us notice however, that taking /2.2/ into account after simple transformations and deleting the equation with vector $(w^1 - w^{n+1})$ on the left hand side, and possibly after changing the order of the remaining equations we obtain a new system of equations of the form

$$/3.4/ \quad \begin{aligned} (w^2 - w^1)^T \xi &= D_j \left(f(z^{l_1}) - f(z^{l_1 - 1}) \right) \\ (w^3 - w^2)^T \xi &= D_j \left(f(z^{l_2}) - f(z^{l_2 - 1}) \right) \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

$(w^{n+1}-w^{n})^T = D_j(f(z^{l_n})-f(z^{l_{n-1}}))$

where $\pi' = (l_1, l_2, \ldots, l_{n+1})$ is the permutation of numbers $(1,2,\ldots,n+1)$ which may be easily generated.

Let us note that the left sides of the system do not depend on from which triangulation of simplex σ the subsimplex $\sigma(z^1,\pi)$ comes from. It is very important observation, because on its basis we conclude that for evaluating ∇g_γ for an arbitrary subsimplex γ obtained from an arbitrary regular subdivision of simplex σ it is enough to solve the system of linear equations $Ax=b(\gamma)$, where $b(\gamma)$ is the vector of right hand sides /depending on γ/ and matrix A /independent of γ/ may be generated only once. After determining of matrix A decomposition $PA=LR$ may be performed /also ones only/ with the cost of $\approx n^3/3$ additionals and multiplications /see [8]/. Having such decomposition we may determined ∇g_γ with the cost $O(n^2)$ additionals and multiplications each time it is necessary /it may be done by solving two systems of linear eguations with triangular matrices/.

Let now $z=x_1w^1+\ldots+x_{n+1}w^{n+1}$ /where $x_i \geq 0$ for $1 \leq i \leq n+1$ and $\sum x_i = 1$ / denote an arditrary vertex of subsimplex obtained from standard subdivision of the simplex σ. Having described above efficient way of determining $\xi = \nabla g_{\gamma} \in \partial g(z)$ we mayproceed to outlining the principle of labeling of triangulation vertices.

3.5. The principle of labeling. Let ξ be a fixed vector belonging to $\partial g(z)$ .Let us assign $L(z):=k$ where k is the smallest natural number for which $x_k > 0$ and $\xi^T(w^k-z) \geq 0$.

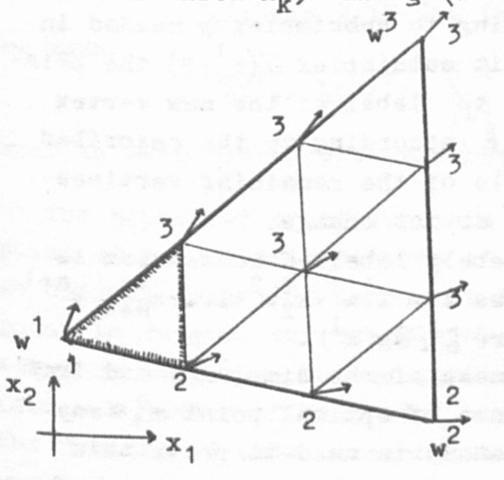

Fig.1

In figure 1 we have presented a standard subdivision of degree 3 of simplex $\sigma(w^1,w^2,w^3)$. In each vertex of this triangulation we have drawn a normalized /and shifted/ gradient of function $f(x)=\|x\|_2^2$ belonging of course to $\partial g(z)$ and determined for this point. Integers assigned to the vertices of triangulation illustrate the principle of labeling.

3.6. Theorem. The principle 3.5. leads to the proper labeling of the vertices of triangulation /in Sperner sense/.

Proof: Obviously, if $x_k=0$ vertex z never gets the label k. Thus we shall show that the principle /3.5./ leads to the proper labeling of the vertices of triangulation /in Sperner sense/ if we prove that each such vertex assigns an index from the set $\{1,\ldots,n+1\}$. If $z=w^k$ /$k=0,\ldots,n+1$/ then obviously $x_j=0$ for $j<k$, $x_k=1>0$ and $\xi^T(w^k-z)=\xi^T(w^k-w^k)=0\geqslant 0$ i.e. $L(w^k)=k$. When however z is a point belonging to a relative interior of j-dimensional /j>0/ face of simplex σ then it can be represented in the form

$$z=x_{i_1}w^{i_1}+x_{i_2}w^{i_2}+\ldots+x_{i_j}w^{i_j}$$

where $x_{i_l}>0$ for $l=1,\ldots,j$ and $\sum_{l=1}^{j} x_{i_l}=1$.

Of course in this case z may not be assigned any index from the set $\{[1,\ldots,n+1]-[i_1,\ldots,i_j]\}$. If now $\xi^T(w^{i_l}-z)<0$ for $l=1,\ldots,j$ then simultaneously we obtain

$$0=\xi^T(z-z)=\xi^T\Big(\sum_{l=1}^{j} x_{i_l}(w^{i_l}-z)\Big)=\sum_{l=1}^{j} x_{i_l}\,\xi^T(w^{i_l}-z)<0$$

This calculation proves that among indices $i_1,\ldots,i_l$ there is a smallest /e.g. i_m/ for which $x_{i_m}>0$ and $\xi^T(w^{i_m}-z)\geqslant 0$. Now, by virtue of Sperners lemma /see [6] / /in a nonconstructive way/ or e.g. Kuhn's "sandwich" algorithm [2,3,4] /in a constructive way/ we infer about the existence of such subsimplex of a given triangulation, whose vertices are labelled respectively by 1,2,...,n+1. Such subsimplex will be called completely labelled. ∎

It should be emphasized that to moving to subsimplex γ needed in the sandwich method from adjacent to it subsimplex $\sigma(z^1,\pi)$ the principle /2.3/ is applied and determining the label of the new vertex z^E requires at most finding of vector ξ /according to the described above way/. Previously determined labels of the remaining vertices of γ and which were also vertices of σ do not change.

Obviously the uniqueness of a completely labelled subsimplex is not guaranted. This simplex has vertices $z^i=x_1^i w^1+x_2^i w^2+\ldots+x_{n+1}^i w^{n+1}$ such that $x_i^i>0$ and $(\xi^i)^T(w^i-z^i)\geqslant 0$, where $\xi^i\in\partial g(z^i)$.

Let us notice that from the compactness of the simplex σ and from the continuity of mapping f the existence of optimal point $x^o\in\sigma$ may be inferred. However, the theorems of analysis used to prove this fact are purely existential nature. Below we shall present a proof which can be used to construct effectively working computer programs.

3.7. Theorem. Let f be finite, differentiable convex function on R^n. Then for arbitrary n-simplex $\sigma(w^1,\dots,w^{n+1})\subseteq R^n$ there is a point $z^o \in \sigma$ such that $f(z^o)\leq f(z)\ \forall z\in\sigma$.

Proof: Let $\{D_j\}_{j=1}^{\infty}$ be an incteasing sequence of natural numbers tending to infinity and denoting degrees of succesive regular transformations of simplex $\sigma(w^1,\dots,w^{n+1})$. We can apply the rule of labeling /3.5./ to the vertices of j-th /j=1,2,.../ triangulation and thus constructively achieve a completely labeled subsimplex $\mathfrak{S}_j$ /using e.g. simple modification of the sandwich method/. In this way we get a family $\mathfrak{S}:=\{\mathfrak{S}_j\}_{j=1}^{\infty}$ of completely labeled subsimplices of simplex σ with diameter /mesh/ tending to zero. Since σ is a compact set, there is a point x^o in whose arbitrary neighbourhood a simplex from the family $\mathfrak{S}$ is contained. Thus we infer about existence of sequences $\{z^{ij}\}_{j=1}^{\infty}$ /$1\leq i\leq n+1$/ convergent to a point $z^o\in\sigma$, and which elements are vertices of completely labeled subsimplices obtained in the triangulation process. According to the principle of labeling we have $(\xi^{ij})^T(w^i-z^{ij})\geq 0$. From the assumption of differentiability of function f we know that $\nabla f(z^o)$ exists. Tending with both sides to infinity we may infer /see [7], theorem 24.5./ that

$$\lim_{j\to\infty}(\xi^{ij})^T(w^i-z^{ij}) = \nabla f^T(z^o)(w^i-z^o)\geq 0$$

for i=1,...,n+1 /due to arbitrariness of the selection of index i in reasoning presented above/.

Since any $z\in\sigma$ may be represented in the form

$$z= z^o+\lambda_1(w^1-z^o)+\lambda_2(w^2-z^o)+\dots+\lambda_{n+1}(w^{n+1}-z^o)$$

/where $\lambda_i\geq 0$ for i=1,...,n+1/

we have

$$\nabla f^T(z^o)(z-z^o) = \sum_{i=1}^{n+1}\lambda_i\nabla f^T(z^o)(w^i-z^o)\geq 0 \quad \forall z\in\sigma,$$

what in combination with theorem 2.4. ends the proof. ∎

The presented proof supports assertion that every accomodation point of barycentric centers of completely labeled subsimplices is an optimal point. Of course at least one such point exists. When this point is unique, for example when f is strictly convex the discussed sequence converges to the optimal point. The present autor has performed a number of succesful numerical experiments. Comparison of the results of these experiments with the results obtained from the work of other algorithms used in similar situations will be published separetely.

4. Final remarks. Let us notice that the considerations described above and concerning piecewise-linear approximation and labeling principle did not require the assumption of differentiability of function f. In the proof of theorem /3.7./ only differentiability of function f in the point $z^0 \in \sigma$ is used. Further it is known that proper convexity of function f ensures its differentiability on a set D dense of R^n, with complement of measure zero. So if differentiability of function f is not assumed, and sequences $\{z^{ij}\}_{j=1}^{\infty}$ $/1 \leq i \leq n+1/$ from the proof of theorem /3.7./ converge to $z^0 \in D \cap \sigma$, then sequences $\{\xi^{ij}\}_{j=1}^{\infty}$ $/1 \leq i \leq n+1/$ converge to $\nabla f(z^0)$. It means that the further part of the proof is correct in this case as well and z^0 is an optimal point. However when z^0 belongs to set $\sigma \setminus D$, we can only conclude that subdifferential $\partial f(z^0)$ contains vectors $a^1, a^2, \ldots, a^{n+1}$ such that $(a^i)^T(w^i - z^0) \geq 0$, $/1 \leq i \leq n+1/$ which may not be sufficient to guarantee optimality of point z^0. Final decision whether z^0 is an optimal point or not must be reached by means of additional criteria known in optimization theory.

Acknowledgements. The author is grateful to Professor Dr. Jerzy Albrycht from Technical University of Poznań for his help in prepering this paper.

References.

1. M.S.Bazaraa and C.M.Shetty. Nonlinear Programming. J.Wiley, New York 1979.

2. H.W.Kuhn. Simplicial approximation of fixed points. Proc.Nat.Acad.Sci.USA. 1968, pp. 1238-1242

3. H.W.Kuhn. Pivotal exchange algorithms. L.C.W.Dixon, E.Spedicato, G.P.Szego /Eds./ Nonlinear optimization theory and algorithms. Birkhauser, Boston 1980.

4. H.W.Kuhn and J.G.MacKinnon. Sandwich Method for Finding Fixed Points. Journal of Optimization Theory and Applications, Vol. 17, Nos.3/4. 1975. pp.189-204.

5. A.Maćkiewicz. On sandwich method for minimization of convex functions and for finding fuzzy fixed points /to appear/.

6. M.Todd. The computation of fixed points. Lecture Notes in Economics and Mathematical System. Springer, Berlin-Heidelberg-New York, 1976.

7. R.T.Rockafellar. Convex Analysis. Princeton University Press. Princeton 1970.

8. J.Stoer and R.Bulirsch. Introduction to Numerical Analysis. Springer, Berlin-Heidelberg-New York, 1980.

Andrzej Maćkiewicz
A.Mickiewicz University
Matejki 48/49
Poznań, Poland.