CREAT. MATH. INFORM. 28 (2019), No. 2, 97 - 104
Online version at http://creative-mathematics.ubm.ro/
Print Edition: ISSN 1584 - 286X Online Edition: ISSN 1843 - 441X

A comparative study of the PL homotopy and BFGS methods for some nonsmooth optimization problems

ANDREI BOZÂNTAN[1] and VASILE BERINDE[1,2]

ABSTRACT. We consider some non-smooth functions and investigate the numerical behavior of the Piecewise Linear Hompotopy (PLH) method ([Bozântan, A., *An implementation of the piecewise-linear homotopy algorithm for the computation of fixed points*, Creat. Math. Inform., **19** (2010), No. 2, 140–148] and [Bozântan, A. and Berinde, V., *Applications of the PL homotopy algorithm for the computation of fixed points to unconstrained optimization problems*, Creat. Math. Inform., **22** (2013), No. 1, 41–46]). We compare the PLH method with the BFGS with inexact line search, a quasi-Newton method, having some results reported in [Lewis, A. S. and Overton, M. L., *Nonsmooth optimization via BFGS*, submitted to SIAM J. Optimiz, (2009)]. For most of the considered cases, the characteristics of the PLH method are quite similar to the BFGS method, that is, the PLH method converges to local minimum values and the convergence rate seems to be linear with respect to the number of function evaluations, but we also identify some issues with the PLH method.

1. INTRODUCTION

In a previous paper ([8]), we tested numerically the efficiency and robustness of the PL homotopy algorithm (which has been designed and implemented for fixed point approximation, see [5], [6] and [7]), for some typical unconstrained optimization problems and compared it with some of the well known and widely used methods in optimization: Newton's method, Broyden-Fletcher-Goldfarb-Shanno (BFGS), conjugate gradient method, and nonlinear conjugate gradient method.

Efficiency is an important feature of any iterative procedure, since in concrete problems for more than three or four variables trial and error becomes impractical because, in some regions, the optimization algorithm may progress very slowly toward the optimum, requiring excessive computer time.

Robustness, i.e., the ability to achieve a solution, is equally or even more important because a general nonlinear function is unpredictable in its behavior: there may be local maxima or minima, saddle points, regions of convexity, concavity, and so on. Therefore, it is of great theoretical and practical importance to draw an extensive experience in testing optimization algorithms for unconstrained functions to evaluate their efficiency and robustness.

The numerical results from [8], clearly illustrated the fact that, for almost all test functions we have considered, but especially for the non differentiable ones, the piecewise-linear homotopy method is more robust (even though not always more efficient) than the methods enumerated above. The main generic advantages of the PL homotopy methods is that they don't require smoothness of the underlying map and this fact is also clearly illustrated by the numerical results from [8]. Also another important feature of these methods is that they can be applied when no a priori knowledge regarding the solutions of

Received: 21.03.2019. In revised form: 23.03.2019. Accepted: 30.03.2019
2010 *Mathematics Subject Classification.* 47H10, 47J25, 49M15, 49M30.
Key words and phrases. *nonsmooth optimization, piecewise-linear homotopy algorithm, quasi-Newton method, BFGS method.*
Corresponding author: Vasile Berinde; vberinde@cunbm.utcluj.ro

the system to be solved is available, and thus to choose a suitable starting point for the iterative method.

In [18] and [19] the authors investigate the BFGS (Broyden-Fletcher-Goldfarb-Shanno) variable metric (quasi-Newton) method with an inexact line search, when applied to nonsmooth functions. The numerical behavior of this method is tested on various classes of examples. The method routinely converges to local minimizers on all but the most difficult class of examples, and the convergence rate is observed to be linear with respect to the number of function evaluations, with a rate of convergence that varies consistently the problem parameters.

Since we previously obtained some encouraging results, the objective of this paper is to continue the study and investigate the numerical behavior the PLH for some **nonsmooth** optimization problems. We will consider several of the problems described in [18] and solve them using the PLH method. The numerical results of the PLH method are then compared with the results obtained by means of the BFGS method.

2. The piecewise-linear homotopy method

This new approach to unconstrained optimization problems is based on the fact that Newton methods can be regarded as a particular case of a classic fixed point iterative method, that is, of the Picard iteration or successive approximations method associated to a certain nonlinear fixed point equation

$$x = Tx, \tag{2.1}$$

where T is a given self operator of a space X. Suppose X and T are such that the equation (2.1) has at least one solution (usually called a *fixed point* of T). A typical situation of this kind is illustrated by the well known Brouwer's fixed point theorem, see [12].

Theorem 2.1. *Every continuous mapping f from a convex compact subset K of a Euclidean space to K itself has a fixed point.*

Under the assumptions of Theorem 2.1, the Picard iteration associated to (2.1), defined by $x_0 \in X$ and

$$x_{n+1} = Tx_n,\ n = 0, 1, 2, \dots, \tag{2.2}$$

does not converge, in general, even though in many cases (e.g., for contractive type mappings, see [4]) it is a useful method to solve nonlinear fixed point equations.

In this context, remind that, if X is a nonempty set and $T : X \to X$ is a self mapping and we denote the set of fixed points of T by $Fix\,(T)$, i.e., $Fix\,(T) = \{a \in X : T(a) = a\}$, then T is said to be a *Picard operator*, see for example Rus [22], if

(i) $Fix\,(T) = \{p\}$;

(ii) $T^n(x_0) \to p$ as $n \to \infty$, for any x_0 in X.

The most useful class of Picard operators, that play a crucial role in nonlinear analysis and was mentioned above, is the class of mappings known in literature as *Picard-Banach contractions*, first introduced by Banach in [3], in the case of what we call now a Banach space, and then extended to complete metric spaces by Caccioppoli [9].

Banach contraction mapping principle, see [4] for a complete form, essentially states that, in a complete metric space (X, d), any Picard-Banach contraction $T : X \to X$, that is, any mapping for which there exists $c \in [0, 1)$ such that

$$d(Tx, Ty) \le c \cdot d(x, y), \forall x, y \in X, \tag{2.3}$$

is a Picard operator.

It is obvious that any contraction mapping is continuous but the reverse is not true. This is the reason why several authors tried to find specific algorithms that could be successfully used to compute fixed points of continuous but not necessarily contractive mappings.

In this context, in 1967 Herbert Scarf proposed a method for approximating fixed points of continuous mappings [23]. The algorithm proposed by Scarf, which is also a numerically implementable constructive proof of the Brouwer fixed point theorem, has its origins in the Lemke-Howson complementary pivoting algorithm for solving linear complementarity problems [17]. Beside the generalization and applications in fixed point theory, the Lemke-Howson algorithm is also famous for its applications in finding Nash equilibrium points for bimatrix games.

Several improvements to the algorithm developed by Scarf were made by Terje Hansen in 1967, see [24] and by Harold W. Kuhn in 1968 [15]. But the decisive advancements came in 1972, when Eaves [10] and then Eaves and Saigal [11] described a piecewise-linear (PL) homotopy deformation algorithm as an improvement for the algorithm proposed by Scarf.

Another PL algorithm, related to the one proposed by Eaves and Saigal, was presented by Orin H. Merrill in 1972 [21]. The main practical advantage of the PL homotopy methods is that they don't require smoothness of the underlying map, and in fact they can be used to calculate fixed points of set-valued maps. Although PL methods can be viewed in the more general context of complementary pivoting algorithms, they are usually considered in the special class of homotopy or continuation methods [2].

The homotopy methods are useful alternatives and aides for the Newton methods in solving systems of n nonlinear equations in n variables:

$$F(x) = 0, \quad F : \mathbb{R}^n \to \mathbb{R}^n. \tag{2.4}$$

mainly when very little a priori knowledge regarding the zero points of F is available and so, a poor starting value could cause a divergent Newton iteration sequence.

The idea of the homotopy method is to consider a new function $G : \mathbb{R}^n \to \mathbb{R}^n$, related to F, with a known solution, and then to gradually deform this new function into the original function F. Typically one can define the convex homotopy:

$$H(x,t) = t \cdot G(x) + (1-t) \cdot F(x) \tag{2.5}$$

and can try to trace the implicitly defined curve

$$H^{-1}(0) = \{x \in \mathbb{R}^n \mid \exists\, t \in [0,1] \text{ such that } H(x,t) = 0\} \tag{2.6}$$

from a starting point $(x_0, 1)$ to a solution point $(x^*, 0)$. The implicit function theorem ensures that the set $H^{-1}(0)$ is at least locally a curve under the assumption that $(x_0, 1)$ is a regular value of H, i.e. the Jacobian $H'(x_0, 1)$ has full rank n. However, because there is no smoothness condition on F, a more complex approach involving piecewise-linear approximations is needed.

A detailed description of the implementation and use of the PLH method for solving some unconstrained optimization problems is given in [5], [6] and [7] and here we will present just some important details of the PLH method.

We define the "refining" triangulation J_3 of $\mathbb{R}^n \times (0,1]$ such that the vertices of this triangulation are given by the set of points:

$$J_3^0 = \{(v_1, \ldots, v_{n+1}) \mid v_{n+1} = 2^{-k}, k \in \mathbf{N} \text{ and } \frac{v_i}{v_{n+1}} \in \mathbb{Z}\}.$$

So, every $(n+1)$-simplex of this triangulation is contained in some slab $\mathbb{R}^n \times [2^{-k}, 2^{-k-1}]$, $k \in \mathbf{N}$.

Let $\sigma = [v_1, v_2, \ldots, v_{n+1}, v_{n+2}] \in J_3$ be an $(n+1)$-simplex and let $\pi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be the following canonical projection: $\pi(x,t) = t$. We define the level of σ as $\max_{i=\overline{1,n+2}} \pi(v_i)$, which is the maximum of the last co-ordinates of all vertices of σ. We call J_3 a refining triangulation of $\mathbb{R}^n \times \mathbb{R}$ because the diameter of σ tends to zero as the level of σ tends to zero.

We define the piecewise linear homotopy map H_{J_3} which interpolates H on the vertices of the given refining triangulation J_3:

- $H_{J_3}(x,1) = G(x)$, $H_{J_3}(x,0) = F(x)$;
- $H_{J_3}(x,t) = \sum_{i=1}^{n+2} \lambda_i H(v_i,t)$, where: (v_i,t) are vertices of $\sigma \in J_3$ and

$$(x,t) = \sum_{i=1}^{n+2} \lambda_i (v_i,t), \quad \sum_{i=1}^{n+2} \lambda_i = 1, \lambda_i \geq 0.$$

The algorithm will trace the unique component of the polygonal path $H_{J_3}^{-1}(0)$ which contains $(x_0,1)$, with nodes on the n-faces of the triangulation J_3.

The algorithm starts with the unique simplex σ_0 which contains the initial point $(x_0,1)$. Then, for each $i = 0,1,\ldots$ it will perform the following steps in a loop:

- It will trace the restriction of $H_{J_3}^{-1}(0)$ to the current simplex σ_i, from the point (x_i,t_i) and finds the intersection point (x_{i+1},t_{i+1}) with some other facet of σ_i. This step is called "door-in-door-out step", see [2]. Sometimes this step is also called linear programming step because it involves the solving of linear equations in a manner typical for linear programming methods.
- It performs a pivoting step, which means to find the new simplex σ_{i+1} which is adjacent to the current simplex and which contains the point (x_{i+1},t_{i+1}). This step is usually performed using only a few operations which define the pivoting rules of the triangulation.

The generated sequence $(x_0,1),(x_1,t_1),\ldots$ will converge to a solution $(x^*,0)$ of the homotopy map H, that is, to a solution x^* of the equation $F(x) = 0$.

3. Experimental comparative study

In this section we investigate the numerical behavior the PLH method for some **nonsmooth** optimization problems described in [18] and compare the numerical results thus obtained with the corresponding results obtained by means of the BFGS method.

3.1. **The tilted norm function.**

We consider the Euclidean norm function on $\mathbb{R}^n$ tilted by a linear term

$$f(x) = w||x|| + (w-1)e_1^T x$$

where e_1 is the first coordinate vector and $w \geq 1$. The only minimizer is the origin.

In Figure 1 we compare the results of the BFGS and PLH methods. For the tilted norm function, the characteristics of the PLH method are quite similar to the BFGS method, for all the cases considered. We observe that the number of iterations required for the PLH method grows quicker with the dimension of the problem n, when compared to the iterations required by the BFGS method. This is actually a known weakness of the method, reported from the earliest descriptions. Another issue identified in the results, is that the PLH method breaks when the precision is smaller than around 10^{-11}, an issue which is visible in the visual data from figure 1.

3.2. **A nonsmooth Rosenbrock function.**

We consider the following nonsmooth variant of the well known Rosenbrock function in two variables:

$$f(x_1,x_2) = w|x_2 - x_1^2|^2 + (1-x_2)^2, \; (x_1,x_2) \in \mathbb{R}^2.$$

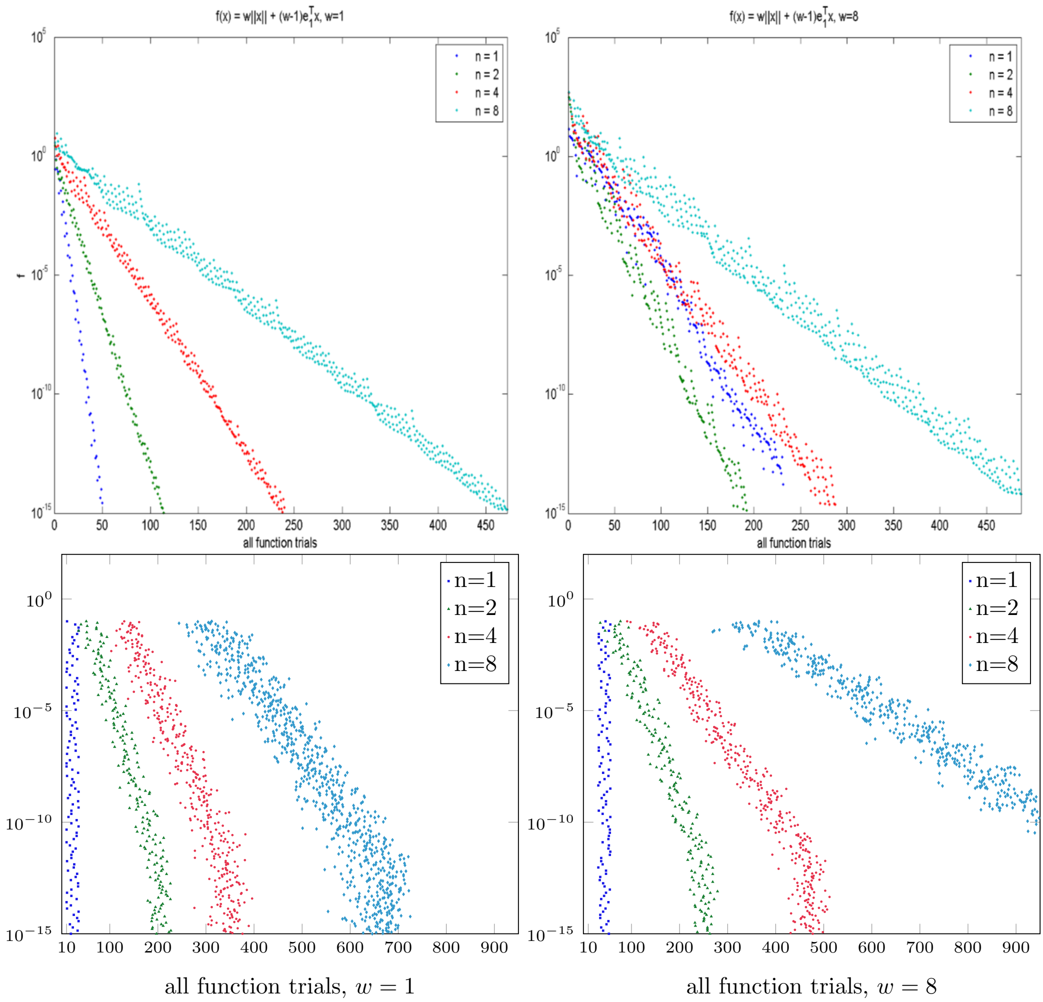

FIGURE 1. Tilted Norm Function - BFGS (top) and PLH (bottom)

The function is unimodal, and the global minimum value, $f(1,1) = 0$, lies in a narrow, parabolic valley.

In Figure 2, we see linear convergence with respect to the number of function trials for typical runs with $w = 1, 2, 4, 8$, for both PLH and BFGS methods. The value of w and the required result precision do not influence too much the number of iterations for the PLH algorithm. The initial value seems to play a more important role in the number of iterations of the PLH method for this problem. The PLH algorithm requires a bigger number of minimal steps for lower precision of the result, but the number of steps for higher precisions is smaller than the steps of the BFGS method, especially for bigger values of the w parameter.

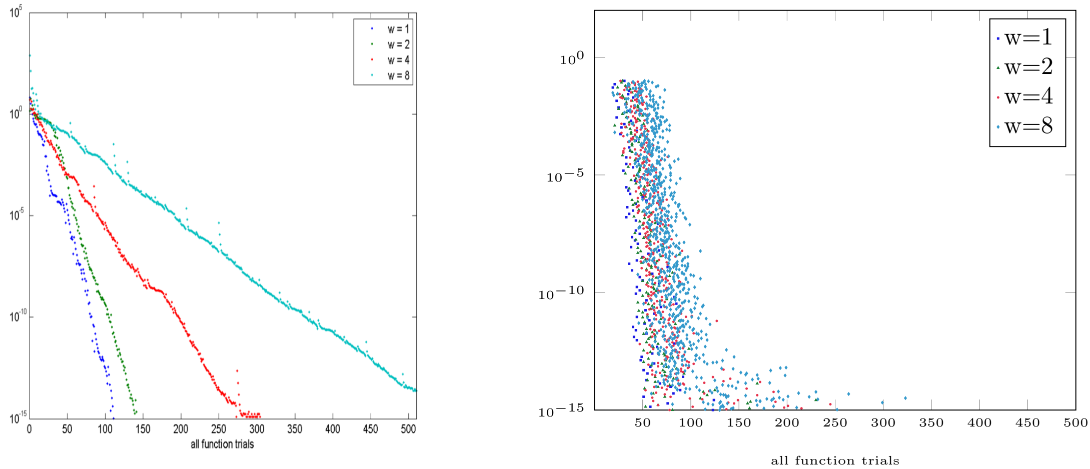

FIGURE 2. Nonsmooth Rosenbrock Function - BFGS (left) and PLH (right)

3.3. **Nesterovs Chebyshev-Rosenbrock Functions.**
We consider the following smooth function introduced by Nesterov:

$$\tilde{f}(x) = \frac{1}{4}(x_1 - 1)^2 + \sum_{i=1}^{n-1}(x_{i+1} - 2x_i^2 + 1)$$

and a non smooth variant:

$$f(x) = \frac{1}{4}(x_1 - 1)^2 + \sum_{i=1}^{n-1}|x_{i+1} - 2x_i^2 + 1|$$

In both cases the minimizer is $\bar{x} = [1, 1, ..., 1]^T$.

Trying to solve the smooth problem for $n = 8$ using $x_0 = [-1, 1, ..., 1]^T$ as starting point, the PLH method requires:

- $>$ 17000 steps for a precision of 10^{-3}
- $>$ 27000 steps for a precision of 10^{-4}
- $>$ 67000 steps for a precision of 10^{-5}
- $>$ 178000 steps for a precision of 10^{-6}

This suggests that the convergence rate of the PLH method is not linear. The BFGS method, for $n = 8$ and using the same initial value x_0 achieves a precision of 10^{-15} in just 6700 iterations, but it requires over 50000 iterations for $n = 10$.

We are able to solve the nonsmooth problem for $n = 2$ using the PLH method, but our tests for $n = 3$ and $n = 4$ are breaking due to numerical instability. We are still investigating if this is a problem in the particular implementation of the PLH algorithm. The above nonsmooth problem is challenging for the BFGS method also, which breaks down for $n >= 4$.

4. CONCLUSIONS

We applied the PLH method to several nonsmooth problems and compared the results with the BFGS method. For two of the examples, the tilted norm function and the Rosenbrock function, the results obtained with the PLH method are quite similar to the BFGS method, that is, the PLH method converges to local minimum values and the convergence rate seems to be linear with respect to the number of function evaluations.

When applied to more challenging problems, i.e., the Nesterov's Chebyshev-Rosenbrock functions presented in section 3.3, some issues of the PLH method become apparent. Firstly, it seems that the convergence rate when considering problems with higher dimensions (e.g. $n = 8$), is not linear.

Another issue identified in example 3.3, is that the PLH method breaks for the nonsmooth function when $n > 2$, due to numerical instability, an issue which has to be analyzed more carefully.

A third problem is that the method breaks when the precision is smaller than 10^{-11}, which is visible also in the visual data form Figure 1.

For other recent related research on that area we refer to [1], [13], [16], [20], [26], [25], [27].

REFERENCES

[1] Alavi Hejazi, M. and Nobakhtian, S., *Sensitivity analysis of the value function for nonsmooth optimization problems*, Oper. Res. Lett., **45** (2017), No. 4, 348–352

[2] Allgower, E. L. and Georg, K., *Introduction to numerical continuation methods*, Springer-Verlag, Berlin, 1990

[3] Banach, S., *Sur les opérations dans les ensembles abstraits et leurs applications aux équations intégrales*, Fund Math., **3** (1922), 133–181

[4] Berinde, V., *Iterative Approximation of Fixed Points*, Springer, Berlin Heidelberg New York, 2007

[5] Bozântan, A., *An implementation of the piecewise-linear homotopy algorithm for the computation of fixed points*, Creat. Math. Inform., **19** (2010), No. 2, 140–148

[6] Bozântan, A. and Berinde, V., *Applications of the PL homotopy algorithm for the computation of fixed points to unconstrained optimization problems*, Creat. Math. Inform., **22** (2013), No. 1, 41–46

[7] Bozântan, A., *New implementations of fixed point iterative algorithms and applications to nonsmooth optimization*, Ph.D. Thesis, North University Center at Baia Mare, Technical University of Cluj-Napoca, 2014

[8] Bozântan, A. and Berinde, V., *A numerical study on the robustness and efficiency of the PL homotopy algorithm for solving unconstrained optimization problems*, Creat. Math. Inform., **24** (2015), No. 2, 113–123

[9] Caccioppoli, R., *Un teorema generale sull'esistenza di elementi uniti in una transformazione funzionale*, Rend Accad dei Lincei., **11** (1930), 794–799

[10] Eaves, C. B., *Homotopies for computation of fixed points*, Mathematical Programming, **3** (1972), No. 1, 1–22

[11] Eaves, C. B. and Saigal, R., *Homotopies for computation of fixed points on unbounded regions*, Mathematical Programming, **3** (1972), No. 1, 225–237

[12] Istrăţescu, V. I., *Fixed Point Theory. An Introduction*, Kluwer Academic Publishers, 2001

[13] Jiang, B., Lin, T. Y, Ma, S. Q. and Zhang, S. Z., *Structured nonconvex and nonsmooth optimization: algorithms and iteration complexity analysis*, Comput. Optim. Appl., **72** (2019), No. 1, 115–157

[14] Karmitsa, N., Gaudioso, M. and Joki, K., *Diagonal bundle method with convex and concave updates for large-scale nonconvex and nonsmooth optimization*, Optim. Methods Softw., **34** (2019), No. 2, 363–382

[15] Kuhn, H. W., *Simplicial approximation of fixed points*, Proc. Nat. Acad. Sci. U.S.A., **61** (1968), 1238–1242

[16] Lanza, A., Morigi, S., Selesnick, I. and Sgallari, F., *Nonconvex nonsmooth optimization via convex-nonconvex majorization-minimization*, Numer. Math., **136** (2017), No. 2, 343–381

[17] Lemke, C. E., *Bimatrix equilibrium points and mathematical programming*, Management Science, **11** (1965), No. 7, 681–689

[18] Lewis, A. S. and Overton, M. L., *Nonsmooth optimization via BFGS*, Submitted to SIAM J. Optimiz., (2009) 1–35

[19] Lewis, A. S. and Overton, M. L., *Nonsmooth optimization via quasi-Newton methods*, Math. Program., **141** (2013), No. 1-2, Ser. A, 135–163

[20] Liuzzi, G., Lucidi, S. and Rinaldi, F., *A derivative-free approach to constrained multiobjective nonsmooth optimization*, SIAM J. Optim., **26** (2016), No. 4, 2744–2774

[21] Merrill, O. H., *Applications and extensions of an algorithm that computes fixed points of certain upper semi-continuous point to set mappings*, Ph.D. Thesis, University of Michigan., Ann Arbor, Michigan, 48106, U.S.A., (1972)

[22] Rus, I. A., *Weakly Picard operators and applications*, Semin. Fixed Point Theory Cluj-Napoca, **2** (2001), 41–57

[23] Scarf, H., *The approximation of fixed points of a continuous mapping*, SIAM J. Appl. Math., **15** 1967 1328–1343

[24] Scarf, H. and Hansen, T., *The computation of economic equilibria*, Yale University Press New Haven, Conneticut, (1973)

[25] Stella, L., Themelis, A. and Patrinos, P., *Forward-backward quasi-Newton methods for nonsmooth optimization problems*, Comput. Optim. Appl., **67** (2017), No. 3, 443–487

[26] Tang, C., Lv, J. and Jian, J., *An alternating linearization bundle method for a class of nonconvex nonsmooth optimization problems*, J. Inequal. Appl., **2018**, No. 101, 23 pp.

[27] Wang, Y., Yin, W. and Zeng, J., *Global convergence of ADMM in nonconvex nonsmooth optimization*, J. Sci. Comput., **78** (2019), No. 1, 29–63

[1] DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
TECHNICAL UNIVERSITY OF CLUJ-NAPOCA
NORTH UNIVERSITY CENTRE AT BAIA MARE
VICTORIEI 76, 430122 BAIA MARE, ROMANIA
E-mail address: andrei.bozantan@gmail.com
E-mail address: vberinde@cunbm.utcluj.ro

[2] ACADEMY OF ROMANIAN SCIENTISTS
E-mail address: vberinde@cunbm.utcluj.ro