AIDIC *Conference Series*, Vol. 10, 2011
Editor Sauro Pierucci
DOI: 10.3303/ACOS1110002
ISBN 978-88-95608-58-7 ISSN 2036-5969

A HOMOTOPY METHOD FOR GLOBAL OPTIMIZATION OF CONTINUOUS MODELS

Karim Alloula[1,2*], Jean-Pierre Belaud[1,2], Jean-Marc Le Lann[1,2]

[1]Université de Toulouse, INPT-ENSIACET
4, allée Emile Monso, BP 44362, 31030 Toulouse Cedex 4, France
[2]CNRS, Laboratoire de Génie Chimique (UMR 5503)
4, allée Emile Monso, BP 44362, 31030 Toulouse Cedex 4, France
[*]Karim.Alloula@ensiacet.fr

An original approach to global optimization of continuous models is introduced. It belongs to the class of homotopy continuation methods, but "only" requires non linear equation systems to be solved. Unconstrained and non-linearly constrained optimization problems are specified nearly the same way. They are solved by coupling a robust Newton formulation for under determined systems and a heuristic estimating the global minimum value by means of the discrete Legendre-Fenchel biconjugate of the criterion. For the time being, the main drawback of the method is the excessive number of function evaluations near the global minimum. However, its success rate being very good on test problems, such as the global optimization of Lennard-Jones atomic clusters, it should be investigated further.

1. INTRODUCTION

Among the main classes of global optimization methods, reviewed for example in Floudas and Gounaris (2009), homotopy continuation methods remain quite confidential within the CAPE community. One reason is the difficulty to handle several optimization problems starting from a "simple" one, and moving to the problem of interest.

This paper introduces a homotopy method, suitable for continuous models only, which solves global optimization problems under non linear constraints. This original approach may be attractive because, instead of tackling intermediate optimization problems, it requires "only" to apply a Newton iterative scheme to non linear equation systems.

First, in part 2, we introduce the method principle by applying it to an unconstrained global optimization problem. Then, in part 3, a generalization to the non-linearly constrained case is detailed. Part 4 discusses the numerical challenges that had to be surpassed and the answers we propose. Part 5 illustrates the benefits and drawbacks of this homotopy formulation by means of a famous case study: the global optimization of Lennard-Jones atomic clusters.

2. UNCONSTRAINED GLOBAL OPTIMIZATION

Let us illustrate the method principle with the unconstrained global optimization of a continuous criterion $f(x)$, where $x \in \mathbb{R}^m$. We assume that we know some global minimum underestimate $\bar{\lambda}$ of the criterion. Instead of solving the initial optimization problem $min_{x\in\mathbb{R}^m} f(x)$, we are interested in a parameterized non linear problem (P_i): $g_{\lambda_i}(x) = f(x) - \lambda_i = 0$. For a given value of λ_i, we

Please cite this article as: Alloula K., Belaud J.P. and Le Lann J.M., (2011), An homotopy method for global optimization of continuous models, AIDIC Conference Series, 10, 11-18 DOI: 10.3303/ACOS1110002

compute a sequence of K Newton iterates $x_{i,1}$, $x_{i,2} = \mathrm{N}_{g_{\lambda_i}}(x_{i,1}), x_{i,3} = \mathrm{N}_{g_{\lambda_i}}(x_{i,2}), \ldots$ where $\mathrm{N}_{g_{\lambda_i}}$ is a Newton operator associated to g_{λ_i}. Among those Newton iterates, we consider only a subset $S_i = \{j \in \{1,2,\ldots,K\}; g_{\lambda_i}(x_{i,j}) \leq 0\}$. At any Newton iterate $x_{i,j}$, where j belongs to S_i, the criterion value is less than λ_i.

The principle is to build a strictly decreasing sequence of criterion values from S_i subsets. First of all, we try to exhibit some $x_{0,1}$ verifying $\|Df(x_{0,1})\|_\infty \geq \varepsilon$. We define $\lambda_0 = f(x_{0,1})$ and $S_0 = \{1\}$. The next subsets are built according to the following:

- If S_i is not empty, pick an iterate x_{i,j^*} from S_i according to some strategy. An obvious strategy is to pick an iterate minimizing f. Whatever this strategy is, $f(x_{i,j^*}) = {\lambda_i}^* \leq \lambda_i$.
  - If $\|Df(x_{i,j^*})\|_\infty < \varepsilon$ then consider that ${x_i}^* = x_{i,j^*}$ approximates a global minimum of f;
  - If $\|Df(x_{i,j^*})\|_\infty \geq \varepsilon$ then, from $x_{i+1,0} = x_{i,j^*}$, compute the K first Newton iterates associated to $g_{\lambda_{i+1}}(x) = g_{\bar{\lambda}}(x) = f(x) - \bar{\lambda} = 0$ to produce S_{i+1};
- If S_i is empty, consider the last non empty set S_a and define $\lambda_{i+1} = \frac{{\lambda_a}^* + \lambda_i}{2}$.
  - If $|\lambda_{i+1} - {\lambda_a}^*| < \varepsilon$ then consider that ${x_a}^*$ approximates a global minimum of f;
  - If $|\lambda_{i+1} - {\lambda_a}^*| \geq \varepsilon$ then, from $x_{i+1,0} = {x_a}^*$, compute the K first Newton iterates associated to $g_{\lambda_{i+1}}(x) = f(x) - \lambda_{i+1} = 0$ to produce S_{i+1}.

3. NON-LINEARLY CONSTRAINED GLOBAL OPTIMIZATION

The previous method can be easily extended to non-linearly constrained global minimization. Equality constraints are added to non linear problems (P_i). Inequality constraints are transformed into equality constraints by introducing slack variables. Then, those equality constraints are added to (P_i) problems.

Once again we assume that we know some global minimum underestimate $\bar{\lambda}$ of the criterion. Instead of solving the initial optimization problem $min_{x \in \mathbb{R}^m} f(x)$ under the constraints $c(x) = 0$, we are interested in a parameterized non linear problem (P_i): $g_{\lambda_i}(x) = \begin{pmatrix} f(x) - \lambda_i \\ c(x) \end{pmatrix} = 0$. For a given value of λ_i, we compute a sequence of K Newton iterates $x_{i,1}$, $x_{i,2} = \mathrm{N}_{g_{\lambda_i}}(x_{i,1}), x_{i,3} = \mathrm{N}_{g_{\lambda_i}}(x_{i,2}), \ldots$ where $\mathrm{N}_{g_{\lambda_i}}$ is a Newton operator associated to g_{λ_i}. Among those Newton iterates, we consider only a subset $S_i = \{j \in \{1,2,\ldots,K\}; f(x_{i,j}) - \lambda_i \leq 0 \ \wedge c(x_{i,j}) = 0\}$. At any Newton iterate $x_{i,j}$, where j belongs to S_i, the criterion value is less than λ_i.
The principle is to build a strictly decreasing sequence of criterion values from S_i subsets. First of all, we try to exhibit some $x_{0,1}$ verifying $\|Df(x_{0,1})\|_\infty \geq \varepsilon \wedge c(x_{0,1}) = 0$. We define $\lambda_0 = f(x_{0,1})$ and $S_0 = \{1\}$. The next subsets are built according to the following:

- If S_i is not empty, pick an iterate x_{i,j^*} from S_i according to some strategy. $f(x_{i,j^*}) = {\lambda_i}^* \leq \lambda_i$.
  - If $\|Df(x_{i,j^*})\|_\infty < \varepsilon$ then consider that ${x_i}^* = x_{i,j^*}$ approximates a global minimum of f;
  - If $\|Df(x_{i,j^*})\|_\infty \geq \varepsilon$ then, from $x_{i+1,0} = x_{i,j^*}$, compute the K first Newton iterates associated to $g_{\lambda_{i+1}}(x) = g_{\bar{\lambda}}(x) = \begin{pmatrix} f(x) - \bar{\lambda} \\ c(x) \end{pmatrix} = 0$ to produce S_{i+1};

- If S_i is empty, consider the last non empty set S_a and define $\lambda_{i+1} = \frac{\lambda_a{}^* + \lambda_i}{2}$.
  - If $|\lambda_{i+1} - \lambda_a{}^*| < \varepsilon$ then consider that $x_a{}^*$ approximates a global minimum of f;
  - If $|\lambda_{i+1} - \lambda_a{}^*| \geq \varepsilon$ then, from $x_{i+1,0} = x_a{}^*$, compute the K first Newton iterates associated to $g_{\lambda_{i+1}}(x) = \begin{pmatrix} f(x) - \lambda_{i+1} \\ c(x) \end{pmatrix} = 0$ to produce S_{i+1}.

To summarize, the non linearly constrained global optimization algorithm differs from the unconstrained global optimization procedure only in the following points:

1. An initial guess $x_{0,1}$ has to be found on the constraints variety;
2. Homotopy functions g_{λ_i} incorporate the residuals associated to the equality constraints. g_{λ_i} is a vectorial function, defined from $\mathbb{R}^m$ to $\mathbb{R}^n$, where $n < m$ (criterion value is optimized within a constraints variety not restricted to a single point). Consequently $Dg_{\lambda_i}(x)$, derivative of g_{λ_i} at x, is an n by m matrix. The number of variables m is equal to the number of state variables in the initial optimization problem formulation plus the number of slack variables, that is to say the number of inequality constraints;
3. S_i subsets definition is modified to take constraints validation into account. One should notice that any of the selected points in the S_i subsets is a feasible point.

4. NUMERICAL CHALLENGES AND ANSWERS

The homotopy method introduced here seems to be easier to implement than other homotopy continuation methods. First, it can be viewed as a Newton homotopy, which is simple and efficient (Yakoubsohn, 2003). Second, instead of tackling intermediate optimization problems, it requires "only" to apply a Newton iterative scheme to a family of non linear equation systems. However, some numerical and software challenges have to be tackled.

4.1 Under determined non linear systems

The first numerical challenge comes from the fact that the non linear systems are under determined, with more variables than equations. Consequently, a generalized Newton formulation, adopted from Dedieu (2006), is required. Assuming F is the residual function to nullify, the generalized Newton operator associated to F in the surjective case is defined by $\mathrm{N}_F(u) = u - [DF(u)]^{\dagger}.F(u)$, where $[DF(u)]^{\dagger}$ stands for the Moore-Penrose pseudo-inverse of the Jacobian matrix of F at u.

When F takes its value in $\mathbb{R}$ (unconstrained global optimization), the generalized Newton operator associated to F is $\mathrm{N}_F(u) = u - \frac{F(u)}{DF(u) \cdot DF(u)}.DF(u)$.

Dedieu (2006) demonstrated that the generalized Newton operator in the surjective case behaves like a projection on the sub-variety $\{u \in \mathbb{R}^m; F(u) = 0\}$. In practice, this theoretical result proves that, provided that the first Newton iterate satisfies the constraints, the next ones will do so. Consequently, the homotopy method we introduce here is a feasible path optimization method, well adapted to the context of chemical engineering where models are defined only within bounded domains.

In the case of non-linearly constrained global optimization, the generalized Newton operator evaluation involves the calculation of a Moore-Penrose pseudo-inverse. The IMSL® numerical library is partially in charge of the Moore-Penrose pseudo-inverse calculations. It provides the singular value decomposition of the Jacobian matrix $DF(u)$, more precisely the singular value decomposition of the normalized matrix $\frac{DF(u)}{\|DF(u)\|_{\infty}}$. Then, the pseudo-inverse is obtained from this

singular value decomposition after dropping the less significant singular values. For each matrix A, the accuracy of the numerical approximation $\widetilde{A^{\dagger}}$ of the Moore-Penrose pseudo-inverse obtained this way may be checked by evaluating the relative error $\frac{\left\|A-A\cdot\widetilde{A^{\dagger}}\cdot A\right\|_{\infty}}{\|A\|_{\infty}}$. The pseudo-inverse matrices are very accurate but, because this calculation is based today on a CPU-time costly singular value decomposition, its efficiency should be improved. The method recently introduced by Natsikis (2011) for calculating the Moore-Penrose pseudo-inverse may be a good alternative for conciliating both accuracy and performance.

4.2 Global minimum underestimate

The second numerical challenge is related to finding a good underestimate of the global minimum. Obviously, most of the time, a good approximation of the global minimum is not available when starting the method. So, instead of using a single global minimum underestimate $\bar{\lambda}$, as stated in parts 2 and 3, a first refinement is to use a sequence of global minimum underestimates $\left(\bar{\lambda}_{\mathrm{i}}\right)_{\mathrm{i}}$, each one being predicted from the criterion values already computed $\left(\lambda_j{}^*\right)_{\mathrm{j}}$. When $\bar{\lambda}_{\mathrm{i}}$ prediction is based only on the last two criterion values $\lambda_i{}^*$ and $\lambda_{i-}{}^*$, a strategy may be to try to decrease the criterion value twice more than previously done. In such a case, $\bar{\lambda}_{\mathrm{i}} = \lambda_i{}^* - 2(\lambda_{i-}{}^* - \lambda_i{}^*)$. To initiate the process, $\bar{\lambda}_0$ is set to a numerical value representing $-\infty$. Assuming S_1 is not empty, $\bar{\lambda}_1$ is set to $\bar{\lambda}_1 = \lambda_1{}^* - 2(\lambda_0 - \lambda_1{}^*)$.

This first refinement is not always efficient for leaving the basins of attraction of local minima. We have to incorporate some global knowledge to the previous prediction which is based only on the local behavior of the criterion. So, the second refinement consists of estimating a convex hull of the criterion by applying twice the Legendre-Fenchel transform to it. Let's remind that the Legendre-Fenchel transform of a function f is the function f^* defined by $f^*(k) = \sup_x \{k \cdot x - f(x)\}$. A very important property of the double Legendre-Fenchel transform f^{**} of f is the following one: f^{**} is the convex envelope of f. Unfortunately, calculating the Legendre-Fenchel bi-conjugate f^{**} from the analytical expression of the criterion f may be harder than the initial global optimization problem: instead of one optimization problem, one has to solve two parameterized optimization problems! In practice, instead of working with the continuous Legendre-Fenchel transform, a discrete Legendre-Fenchel transform may be defined, the bi-conjugate of the criterion function being estimated only at points where the criterion has already been calculated. Instead of calculating the conjugate f^* of a real function f from $f^*(x) = \sup_y \{x \cdot y - f(y)\}$, a discrete Legendre-Fenchel transform $f^{\tilde{*}}$ is evaluated over a set of points $\{x_1, x_2, \dots, x_K\}$ for which previous criterion evaluations $\{f(x_1), f(x_2), \dots, f(x_K)\}$ are available. More precisely, for any x_k in $\{x_1, x_2, \dots, x_K\}$, $f^{\tilde{*}}(x_k) = \max_l \{x_k \cdot x_l - f(x_l)\}$. One can easily show that the discrete Legendre-Fenchel bi-conjugate still has an interesting property from our point of view: $f^{\widetilde{**}}$ is a convex function over the convex envelope of $\{x_1, x_2, \dots, x_K\}$, which underestimates f at any point of $\{x_1, x_2, \dots, x_K\}$. Corrias (1996) presents results about convergence of the discrete Legendre-Fenchel transform to the continuous Legendre-Fenchel transform. According to those results, we can hope that $\min_{k \in \{1, \dots, K\}} f^{**}(x_k)$ may give a good estimate of the criterion minimum value over the convex span of $\{x_1, x_2, \dots, x_K\}$, i.e. $\min_{k \in \{1, \dots, K\}} f^{**}(x_k) \approx \min_{x \in span(x_1, x_2, \dots, x_K)} f(x)$. So, the first prediction formula $\bar{\lambda}_{\mathrm{i}} = 2(\lambda_{i-}{}^* - \lambda_i{}^*)$ is replaced by the following one:

$$\bar{\lambda}_{\mathrm{i}} = min\left[2(\lambda_{i-}{}^* - \lambda_i{}^*), \min_{k \in \{1, \dots, K\}} f^{**}(x_k)\right].$$

In practice, such a prediction leads to an underestimation of any minimum, either local or global. For local minima, this strategy is adequate; but, when reaching the basin of attraction of the global minimum, it leads to unnecessary criterion evaluations because we are looking for points $x_{i,j}$ satisfying $f(x_{i,j}) - \lambda_i \leq 0$ without success, until λ_i becomes slightly greater than the global minimum.

With respect to the software implementation of the underestimator, the Legendre-Fenchel transform can be factorized in unidimensional transforms so that its discrete formulation can be calculated in a very efficient way, as detailed in Lucet (1996 and 1997).

4.3 Symbolic-numeric calculations

From the software point of view the challenge is to produce a piece of code able to obtain automatically from the problem specification $min_{x\in\mathbb{R}^m} f(x)$, and for any value of λ_i:

- the g_{λ_i} functions and their analytical derivatives Dg_{λ_i};
- the numerical values of $\left[Dg_{\lambda_i}(u)\right]^{\dagger}$ matrices;
- a numerical approximation of the Legendre-Fenchel bi-conjugate ${g_{\lambda_i}}^{**}$ from some values of g_{λ_i}.

Those symbolic-numeric calculations could have been achieved by some existing computer algebra system. The features provided by our in-house environment *eXMSL*, a symbolic and numerical calculation system (Alloula et al., 2009), proved to be very adequate to mix the symbolic processing steps and the numerical evaluations associated to this optimization method.

The residual functions g_{λ_i} are obtained automatically from the initial formulation, using slack variables when constraints are inequalities. From any residual function g_{λ_i}, *eXMSL* derives automatically the corresponding Jacobian matrix. The criterion expression can be made of symbols, numbers and function evaluations, either explicit or implicit. When no implicit function is involved in the criterion expression, *eXMSL* computes an explicit and analytical expression of Dg_{λ_i}. When implicit functions appear in the criterion expression, *eXMSL* computes an implicit and analytical expression of Dg_{λ_i}, which can be evaluated at any point by means of the implicit function theorem.

The numerical evaluations are handled by external software or coded inside *eXMSL*. The IMSL® numerical library is in charge of the singular value decompositions used in the Moore-Penrose pseudo-inverse calculations. The discrete Legendre-Fenchel transform has been coded by ourselves.

5. APPLICATION TO GLOBAL OPTIMIZATION OF LENNARD-JONES ATOMIC CLUSTERS

The global optimization of Lennard-Jones atomic clusters is a classical test problem for global optimization methods. The problem description can be found in Daven (1996). Briefly speaking, the problem is defined as finding the coordinates in $\mathbb{R}^3$ of a system of N atoms leading to a potential energy minimum. The main difficulty in solving this problem arises from the fact that the objective function is a non-convex and highly nonlinear function with an exponentially increasing number of local minima with N.

The strategy described in this paper, and coded within *eXMSL*, was applied to clusters up to 100 atoms. The initial guess was always randomly selected without any a priori knowledge, making the problems harder to solve. The number of degrees of freedom, which is $3N$ initially, is decreased to

$3N-6$ without loss of generality, by assuming that one of the N atoms stays at the origin point O, a second one stays on the x axis, and a third one stays on the Oxy plane ($x_1=0, y_1=0, z_1=0, y_2=0, z_2=0, z_3=0$). Our results were compared with the best minima reported in http://www-wales.ch.cam.ac.uk/~jon/structures/LJ/tables.150.html (Web site consulted on October, 3rd, 2011).

With an unconstrained formulation of the problem, we retrieved the best criterion values given within the literature only when the number of atoms was less than 20. The success rate decreases with N, being 80% for $N=6$, and vanishing to 0% around $N=15$.

In order to check the method validity on non-linearly constrained problems, we stated that all the inter-particle distances were greater than 0.6187 (Vinko, 2005). For a cluster of N atoms, this results in $N(N-1)/2$ constraints. With this constrained formulation of the problem, we retrieved all the best criterion values given within the literature for $3 \leq N \leq 100$. The success rate increases with N, being 70% for $N=6$, and reaching 100% when $N \geq 30$. Combined with the equation $f(x)-\lambda=0$, constraints on the inter-particle distances provide a good path for the homotopy method to reach a global minimum. When N is quite small, the feasible domain defined by "the inter-particle distances are to be greater than 0.6187" is important when compared to the whole variable space (for $N=6$, the inter-particle distances at the global minimizer are close to 0.995531). When $N \geq 30$, the feasible domain defined by "the inter-particle distances are to be greater than 0.6187" is small when compared to the whole variable space. Then, the main difficulty becomes finding an initial point belonging to the constraint variety. This problem consists of solving an under determined non linear system of $1+N(N-1)/2$ equations and $(3N-6)+N(N-1)/2$ variables. The generalized Newton formulation, cited previously, was applied successfully, even for initial guesses without any physical meaning.

6. CONCLUSION AND FURTHER WORK

This paper introduces a deterministic method for global optimization. Applied to non-linearly constrained continuous models, this homotopy method produces a feasible path along which the criterion value decreases. Points of the homotopy curve come from a Newton process where a generalized Newton operator tries to keep iterates on the constraint variety. Two benefits are associated with such a feasible path optimization:

- the variables remain in the validity domain of the (thermodynamic) models;
- the search space being limited to the constraint variety, the probability for the algorithm to reach and stay at a non global minimum is reduced.

This method was applied successfully to a collection of test problems. Unfortunately, for the time being, the number of function evaluations remains large, mainly because of a costly bisection process: the homotopy parameter λ_i is successively increased or decreased, depending on whether some $x_{i,j}$ verifying $f(x_{i,j}) \leq \lambda_i$ is found or not. For the method to be not only robust, but also efficient, the bisection strategy has to be improved. Two ways are investigated:

- find the "best" bisection parameter value to be applied when some $x_{i,j}$ verifying $f(x_{i,j}) \leq \lambda_i$ is found. From this study, the discrete Legendre-Fenchel transform appears to be a good tool for approximating a local convex underestimator from some previous criterion evaluations. The closer this convex hull is to the criterion representative surface, the better is the prediction of the next homotopy parameter value;
- find the "best" bisection parameter value to be applied when some $x_{i,j}$ verifying $f(x_{i,j}) \leq \lambda_i$ is not found. Until now, this bisection parameter is set to 0.5. Its value may be related to

$\frac{\lambda_a{}^* - f(x_{i,j})}{\lambda_a{}^* - \lambda_i}$, ratio between the criterion decrease we obtain and the criterion decrease we expected.

A generalization of this optimization method to mixed integer non linear programming is being considered. Such an extension seems justified by two facts:

- discrete variables could be considered as continuous ones by adding constraints in order to restrict their allowed values;
- any iterate is a sub-optimal point for which discrete variable values make sense, because they satisfy the associated constraints (feasible-path method).

The trim-loss problem, described in Harjunkoski (1998), and to which a wide range of optimization methods have already been applied, has been chosen as a benchmark for the generalization of the homotopy method presented in this paper.

References

Alloula K., Belaud J.-P. and Le Lann J.-M., 2009, A co-operative model combining computer algebra and numerical calculation for simulation, Computer-Aided Chemical Engineering, 26, 889-894.

Corrias L., 1996, Fast Legendre-Fenchel transform and applications to Hamilton-Jacobi equations and conservation laws, SIAM Journal on Numerical Analysis, 33, n°4, 1534-1558.

Daven D.M., Tit N., Morris J.R. and Ho K.M., 1996, Structural optimization of Lennard-Jones clusters by a genetic algorithm, Chemical Physics Letters, 256, n°1-2, 195-200.

Dedieu J.-P., 2006, Points fixes, zéros et la méthode de Newton, Springer Berlin Heidelberg.

Floudas C. A., Gounaris C. E., 2009, A review of recent advances in global optimization, Journal of Global Optimization, 45, n°1, 3-38.

Harjunkoski I., Westerlund T., Pörn R. and Skrifvars H., 1998, Solving trim-loss problems with MINLP methods, AIDIC Conference Series, 2, 303-310.

Lucet Y., 1996, A fast computational algorithm for the Legendre–Fenchel transform, Computational Optimization and Applications, 6, 27–57.

Lucet Y., 1997, Faster than the fast Legendre transform, the linear-time Legendre transform, Numerical Algorithms, 16, 171–185.

Natsikis V. N., Pappas D., Petralias A., 2011, An improved method for the computation of the Moore–Penrose inverse matrix, Applied Mathematics and Computation, 217, 9828-9834.

Vinko T., 2005, Minimal inter-particle distance in atom clusters, Acta Cybernetica, 0, 1-15.

Yakoubsohn J.-C., 2003, Numerical Elimination, Newton Method and Multiple Roots, Algorithms Seminar 2001–2002, F. Chyzak (ed.), INRIA, 49–54.