Global optimization – What is it all about ?

SK Mishra
Dept. of Economics
North-Eastern Hill University
Shillong, Meghalaya (India)

A Brief History of Optimization Research: The history of optimization of real-valued non-linear functions (including linear ones), unconstrained or constrained, goes back to Gottfried Leibniz, Isaac Newton, Leonhard Euler and Joseph Lagrange. However, those mathematicians often assumed differentiability of the optimand as well as constraint functions. Moreover, they often dealt with the equality constraints. Richard Valentine (1937) and William Karush (1939), however, were perhaps the first mathematicians to study optimization of nonlinear functions under inequality constraints. Leonid Kantorovich and George Dantzig are well known for developing and popularizing linear programming, which ushered a new era of 'operations research', a branch of mathematical science that specializes in optimization. The development of linear programming soon prompted the study of the optimization problem of nonlinear functions (often under linear or nonlinear constraints). The joint work of Harold Kuhn and Albert Tucker (1951) – that was backed up by the work of Karush – is a landmark in the history of optimization of nonlinear functions.

Initially, optimization of nonlinear functions was methodologically based on the Leibniz-Newton principles and therefore could not easily escape local optima. Hence, its development to deal with nonconvex (multimodal) functions stagnated until the mid 1950's. Stanislaw Ulam, John von Neumann and Nicolas Metropolis had in the late 1940's proposed the Monte Carlo method of simulation and it was gradually realized that the simulation approach could provide an alternative methodology to mathematical investigations in optimization. George Box (1957) was perhaps the first mathematician who exploited the idea and developed his evolutionary method of nonlinear optimization. Almost a decade later, MJ Box (1965) developed his complex method, which strews random numbers over the entire domain of the decision variables and therefore has a great potentiality to escape local optima and locate the global optimum of a nonlinear function. The simplex method of John Nelder and Roger Mead (1964) also incorporated the ability to learn from its earlier search experience and adapt itself to the topography of the surface of the optimand function.

Global Optimization: The 1970's evidenced a great fillip in simulation-based optimization research due to the invention of the 'genetic algorithm' by John Holland (1975). A genetic algorithm is a class of population-based adaptive stochastic optimization procedures, characterizing the presence of randomness in the optimization process. The randomness may be present as either noise in measurements or Monte Carlo randomness in the search procedure, or both. The basic idea behind the genetic algorithm is to mimic a simple picture of the Darwinian natural selection in order to find a good algorithm and involves the operations such as 'mutation', 'selection' and 'evaluation of fitness' repeatedly. The genetic algorithms may claim to have ushered the new era of global optimization.

A little later, in 1978, Aimo Törn introduced his "Clustering Algorithm" of global optimization. The method improves upon the earlier local search algorithms that needed 'multiple start' from several points distributed over the whole optimization region. Multi-start is certainly one of the earliest global procedures used. It has even been used in local optimization for increasing the confidence in the obtained solution. However, one drawback of Multi-start is that when many starting points are used, the same minimum will eventually be determined several times. In order to improve the efficiency of Multi-start this should be avoided. The clustering method of Törn avoids this repeated determination of local minima. This is realized in three steps, which may be iteratively used. The three steps are: (i) sample points in the region of interest, (ii) transform the sample to obtain points grouped around the local minima, and (iii) use a clustering technique to recognize these groups (i.e. neigbourhoods of the local minima). If the procedure employing these steps is successful, then, starting a single local optimization from each cluster would determine the local minima and, thus, also the global minimum. The advantage in using this approach is that the work spared by computing each minimum just once can be spent on computations in (i) and (ii), which will increase the probability that the global minimum will be found.

While the clustering algorithm was optimizing on the multi-start requirements of local search algorithms and the genetic algorithm simulated the Darwinian struggle for the survival of the fittest, the simulated annealing method (Kirkpatrick et al., 1983; Cerny, 1985) proposed to mimic the annealing process in metallurgy. In an annealing process a metal in the molten state (at a very high temperature) is slowly cooled so that the system at any time is approximately in thermodynamic equilibrium. As cooling proceeds, the system becomes more ordered – the liquid freezes or the metal re-crystallizes – attaining the ground state at T=0. This process is simulated through the Monte Carlo experiment (Metropolis et al. 1953). If the initial temperature of the melt is too low or cooling is done unduly fast the metal may become 'quenched' due to being trapped in a local minimum energy state (meta-stable state) forming defects or freezing out. The simulated annealing method of optimization makes very few assumptions regarding the function to be optimized, and therefore, it is quite robust with respect to irregular surfaces. In this method, the mathematical system describing the problem mimics the thermodynamic system. The current solution to the problem mimics the current state of the thermodynamic system, the objective function mimics the energy equation for the thermodynamic system, and the global minimum mimics the ground state. However, nothing in the numerical optimization problem directly mimics the temperature, T, in the thermodynamic system underlying the metallurgical process of annealing. Therefore, a complex abstraction mimics it. An arbitrary choice of initial value of a variable called 'temperature', how many iterations are performed at each 'temperature', the step length at which the decision variables are adjusted, and the rate of fall of 'temperature' at each step as 'cooling' proceeds, together make an 'annealing schedule'. This schedule mimics the cooling process. At a high 'temperature' the step lengths at which the decision variables are adjusted are larger than those at a lower 'temperature'. Whether the system is trapped into local minima (quenching takes place) or it attains the global minimum (faultless crystallization) is dependent on the said annealing schedule. A wrong choice of the initial 'temperature', or the rate of fall in the ‘temperature’ leads to quenching or entrapment of the solution in the local minima. The method does not provide any clear guideline as to the choice of the ‘annealing schedule’ and often requires judgment or trial and error. If the schedule is properly chosen, the process attains the global minimum. It is said that using this method is an art and requires a lot of experience and judgment.

A little later, Fred Glover (1986) introduced his ‘Tabu Search’ method. This method economizes on repeated visits to the already visited points and in some sense is close to the clustering algorithms. Glover attributes it's origin to about 1977. The basic concept of Tabu Search as described by Glover is "a meta-heuristic superimposed on another heuristic. The overall approach is to avoid entrainment in cycles by forbidding or penalizing moves which take the solution, in the next iteration, to points in the solution space previously visited (hence "tabu"). The Tabu method was partly motivated by the observation that human behavior appears to operate with a random element that leads to inconsistent behavior given similar circumstances. As Glover points out, the resulting tendency to deviate from a charted course, might be regretted as a source of error but can also prove to be source of gain. The Tabu method operates in this way with the exception that new courses are not chosen randomly. Instead the Tabu search proceeds according to the supposition that there is no point in accepting a new (poor) solution unless it is to avoid a path already investigated. This insures new regions of a problems solution space will be investigated in with the goal of avoiding local minima and ultimately finding the desired solution.

The pace of research in global optimization (GO) by stochastic process accelerated considerably in the 1990’s. Marco Dorigo in his Ph.D. thesis (1992) introduced his “Ant Colony” method of global optimization. It studies artificial systems that take inspiration from the behaviour of real ant colonies. Ants use pheromones that guide other fellow ants to identify the path that leads to a success. The chemical properties of pheromones and the ability of ants to gather information and use them are simulated in the Ant Colony method to reach at the global optimum. This method is well suited to combinatorial (discrete) optimization problems.

A couple of years later, James Kennedy and Russell Eberhart (1995) introduced their “Particle Swarm” method of global optimization. In the animal world we observe that a swarm of birds or insects or a school of fish searches for food, protection, etc. in a very typical manner. If one of the members of the swarm sees a desirable path to go, the rest of the swarm will follow quickly. The Particle Swarm method mimics this behaviour. Every individual of the swarm is considered as a particle in a multidimensional space that has a position and a velocity. These particles fly through hyperspace and remember the best position that they have seen. Members of a swarm communicate good positions to each other and adjust their own position and velocity based on these good positions. There are two main ways this communication is done: (i) “swarm best” that is known to all (ii) “local bests” are known in neighborhoods of particles. Updating of the position and velocity are done at each iteration such that the solution often converges to the global optimum of the function. Interestingly, this method has a very sound and welldocumented philosophical literature behind it (the British empiricist philosophy, the American pragmatism and others like those of Friedrich Hayek, Herbert Simon, etc.).

The method of Differential Evolution (DE) grew out of Kenneth Price's attempts to solve the Chebychev Polynomial fitting Problem that had been posed to him by Rainer Storn. A breakthrough happened (1996), when Price came up with the idea of using vector differences for perturbing the vector population. The crucial idea behind DE is a scheme for generating trial parameter vectors. Initially, a population of points (p in d-dimensional space) is generated and evaluated (i.e. f(p) is obtained) for their fitness. Then for each point (p_i) three different points (p_a, p_b and p_c) are randomly chosen from the population. A new point (p_z) is constructed from those three points by adding the weighted difference between two points ($w(p_b-p_c)$) to the third point (p_a). Then this new point (p_z) is subjected to a crossover with the current point (p_i) with a probability of crossover (c_r), yielding a candidate point, say p_u. This point, p_u, is evaluated and if found better than p_i then it replaces p_i else p_i remains. Thus we obtain a new vector in which all points are either better than or as good as the current points. This new vector is used for the next iteration. This process makes the differential evaluation scheme completely self-organizing.

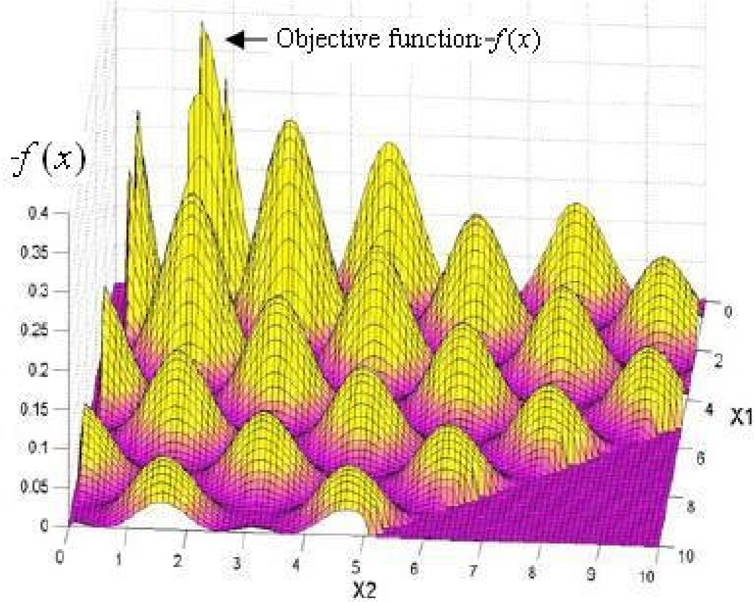
III. The Characteristic Features of Population-Based GO Methods: All population-based methods of global optimization partake of the probabilistic nature inherent to them. As a result, one cannot obtain certainty in their results, unless they are permitted to go in for indefinitely large search attempts. Larger is the number of attempts, greater is the probability that they would find out the global optimum, but even then it would not reach at the certainty. Secondly, all of them adapt themselves to the surface on which they find the global optimum. The scheme of adaptation is largely based on some guesswork since nobody knows as to the true nature of the problem (environment or surface) and the most suitable scheme of adaptation to fit the given environment. Surfaces may be varied and different for different functions. A particular type of surface may be suited to a particular method while a search in another type of surface may be a difficult proposition for it. Further, each of these methods operates with a number of parameters that may be changed at choice to make it more effective. This choice is often problem oriented and for obvious reasons. A particular choice may be extremely effective in a few cases, but it might be ineffective (or counterproductive) in certain other cases. Additionally, there is a relation of trade-off among those parameters. These features make all these methods a subject of trial and error exercises.

Keane’s bump function (Keane, 1994) may provide a fitting example of a multi-modal function that requires methods of global optimization to obtain a ‘solution’. It is given as:

$$Minimize\ \ f(x) = -\left| \left\{ \sum_{i=1}^{m} \cos^4(x_i) - 2\prod_{i=1}^{m} \cos^2(x_i) \right\} \Big/ \left(\sum_{i=1}^{m} i x_i^2 \right)^{0.5} \right|$$

$$\text{subject to: } g_1(x):\ 0.75 - \prod_{i=1}^{m} x_i < 0; \quad g_2(x):\ \sum_{i=1}^{m} x_i - 7.5m < 0; \quad 0 < x_i < 10.$$

Fig.-1: Keane's Bump Function in 2 Dimensions



Source: Hacker, Eddy and Lewis (2002)

A visual appreciation of Keane's two-dimensional (m=2) bump function may be obtained from the graphical presentation (Fig.-1; Hacker et al., 2002). As the dimension (m) grows larger, the optimum value of the function becomes more and more difficult to obtain. Keane (1994) observed that for m=20 the value of min[f(x)] could be about -0.76 and for m=50 it could be about -0.835 but did not know this to be the case.

Keane's bump function is considered as a standard benchmark for nonlinear constrained optimization. It is highly multi-modal and its optimum is located at the non-linear constrained boundary. Emmerich (2005, p. 116) noted that the true minimum of this function is perhaps unknown.

Bibliography

- Bauer, J.M.: "Harnessing the Swarm: Communication Policy in an Era of Ubiquitous Networks and Disruptive Technologies", *Communications and Strategies*, 45, 2002.
- Box, M.J.: "A new method of constrained optimization and a comparison with other methods". *Comp. J.* 8, pp. 42-52, 1965.
- Bukin, A. D.: *New Minimization Strategy For Non-Smooth Functions*, Budker Institute of Nuclear Physics preprint BUDKER-INP-1997-79, Novosibirsk 1997.
- Cerny, V.: "Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm", *J. Opt. Theory Appl.*, 45, 1, 41-51, 1985.
- Eberhart R.C. and Kennedy J.: "A New Optimizer using Particle Swarm Theory", *Proceedings Sixth Symposium on Micro Machine and Human Science*, pp. 39–43. IEEE Service Center, Piscataway, NJ, 1995.
- Emmerich, MTM: *Single- and Multi-objective Evolutionary Design Optimization Assisted by Gaussian Random Field Metamodels*, Dissertation for Doctoral Degree in Natural Sciences, University of Dortmund, Dortmund. 2005.
- Fleischer, M.: "Foundations of Swarm Intelligence: From Principles to Practice", Swarming Network Enabled C4ISR, arXiv:nlin.AO/0502003 v1 2 Feb 2005.
- G.E.P. Box, "Evolutionary operation: A method for increasing industrial productivity", *Applied Statistics*, 6 , pp. 81-101, 1957.
- Glover F.," Future paths for Integer Programming and Links to Artificial Intelligence", *Computers and Operations Research*, 5:533-549, 1986.

- Hacker, KA, Eddy, J and Lewis, KE: "Efficient Global Optimization using Hybrid Genetic Algorithms," in *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Atlanta, Georgia, 4-6 September 2002.
- Hayek, F.A.: *The Road to Serfdom*, Univ. of Chicago Press, Chicago, 1944.
- Holland, J.: *Adaptation in Natural and Artificial Systems,* Univ. of Michigan Press, Ann Arbor, 1975.
- Karush, W. *Minima of Functions of Several Variables with Inequalities as Side Constraints*. M.Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois, 1939.
- Keane, AJ: "Bump: A Hard (?) Problem" http://www.soton.ac.uk/~ajk/bump.html, 1994.
- Keane, AJ: "Experiences with optimizers in structural design," in Parmee, IC (ed) *Proceedings of the 1st Conf. on Adaptive Computing in Engineering Design and Control*, University of Plymouth, UK, pp. 14-27, 1994.
- Kirkpatrick, S., Gelatt, C.D. Jr., and Vecchi, M.P.: "Optimization by Simulated Annealing", *Science*, 220, 4598, 671-680, 1983.
- Kuhn, H.W. and Tucker, A.W.: "Nonlinear Programming", in Neymann, J. (ed) *Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probability,* Univ. of California Press, Berkrley, Calif. pp. 481-492, 1951.
- Liu, P and Lewis, MJ: "Communication Aspects of an Asynchronous Parallel Evolutionary Algorithm", *Proceedings of the Third International Conference on Communications in Computing* http://grid.cs.binghamton.edu/papers/LiuAPEAComm_CIC.pdf *(CIC 2002),* pp. 190-195, Las Vegas, NV, June 24-27, 2002.
- Metropolis, N. The Beginning of the Monte Carlo Method. *Los Alamos Science*, No. 15, Special Issue, pp. 125-130, 1987.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E.: "Equation of State Calculations by Fast Computing Machines", *J. Chem. Phys.*,21, 6, 1087-1092, 1953.
- Mishra, SK (a): "Global Optimization by Particle Swarm Method: A Fortran Program" (August 1, 2006). Available at SSRN: http://ssrn.com/abstract=921504
- Mishra, SK (b): "Global Optimization by Differential Evolution and Particle Swarm Methods: Evaluation on Some Benchmark Functions" (September 30, 2006). Available at SSRN: http://ssrn.com/abstract=933827
- Mishra, SK (c): "Performance of Differential Evolution and Particle Swarm Methods on Some Relatively Harder Multi-Modal Benchmark Functions" (October 13, 2006). Available at SSRN: http://ssrn.com/abstract=937147
- Mishra, S.K. (d): "Least Squares Fitting of Chacón-Gielis Curves by the Particle Swarm Method of Optimization", *Social Science Research Network* (SSRN), Working Papers Series, http://ssrn.com/abstract=917762 , 2006.
- Mishra, S.K. (e): "Performance of Repulsive Particle Swarm Method in Global Optimization of Some Important Test Functions: A Fortran Program" , *Social Science Research Network* (SSRN), Working Papers Series, http://ssrn.com/abstract=924339 , 2006
- Mishra, S.K. (f): "Repulsive Particle Swarm Method on Some Difficult Test Problems of Global Optimization" ,SSRN: http://ssrn.com/abstract=928538 , 2006
- Mishra, S.K. (g): "Some Experiments on Fitting of Gielis Curves by Simulated Annealing and Particle Swarm Methods of Global Optimization", *Social Science Research Network* (SSRN): http://ssrn.com/abstract=913667, Working Papers Series, 2006
- Mishra, S.K. (h): "Some New Test Functions for Global Optimization and Performance of Repulsive Particle Swarm Method", *Social Science Research Network* (SSRN) Working Papers Series, http://ssrn.com/abstract=927134, 2006
- Mishra, SK (i): "Minimization of Keane's Bump Function By the Repulsive Particle Swarm and the differential Evolution Methods" SSRN: http://ssrn.com/abstract=983836, May 2007
- Nagendra, S.: *Catalogue of Test Problems for Optimization Algorithm Verification*, Technical Report 97-CRD-110, General Electric Company, 1997.
- Nelder, J.A. and Mead, R.: "A Simplex method for function minimization" *Computer Journal*, 7: pp. 308-313, 1964.

- Ong, YS and Keane, AJ: "Meta-Lamarckian Learning in Memetic Algorithms" http://ntu-cg.ntu.edu.sg/ysong/journal/IEEE_EC_Ysong2003.pdf, 2003.
- Ong, YS, Lim, MH, Zhu, N and Wong, KW: "Classification of Adaptive Memetic Algorithms: A Comparative Study", http://ntu-cg.ntu.edu.sg/ysong/journal/AdaptiveMA.pdf Un-dated working paper, possibly written in 2005.
- Parsopoulos, K.E. and Vrahatis, M.N., "Recent Approaches to Global Optimization Problems Through Particle Swarm Optimization", *Natural Computing*, 1 (2-3), pp. 235- 306, 2002.
- Prigogine, I. and Strengers, I.: *Order Out of Chaos: Man's New Dialogue with Nature*, Bantam Books, Inc. NY, 1984.
- Silagadge, Z.K.: "Finding Two-Dimensional Peaks", Working Paper, Budkar Insttute of Nuclear Physics, Novosibirsk, Russia, arXive:physics/0402085 V3 11 Mar 2004.
- Simon, H.A.: *Models of Bounded Rationality*, Cambridge Univ. Press, Cambridge, MA, 1982.
- Smith, A.: *The Theory of the Moral Sentiments*, The Adam Smith Institute (2001 e-version), 1759.
- Sumper, D.J.T.: "The Principles of Collective Animal Behaviour", *Phil. Trans. R. Soc. B*. 361, pp. 5-22, 2006.
- Törn, A.A and Viitanen, S.: "Topographical Global Optimization using Presampled Points", *J. of Global Optimization*, 5, pp. 267-276, 1994.
- Törn, A.A.: "A search Clustering Approach to Global Optimization" , in Dixon, LCW and Szegö, G.P. (Eds) *Towards Global Optimization – 2*, North Holland, Amsterdam, 1978.
- Tsallis, C. and Stariolo, D.A.: "Generalized Simulated Annealing", *ArXive condmat/9501047 v1 12 Jan,* 1995.
- Valentine, R.H.: *Travel Time Curves in Oblique Structures*, Ph.D. Dissertation, MIT, Mass, 1937.
- Veblen, T.B.: "Why is Economics Not an Evolutionary Science" *The Quarterly Journal of Economics*, 12, 1898.
- Veblen, T.B.: *The Theory of the Leisure Class*, The New American library, NY. (Reprint, 1953), 1899.
- Vesterstrøm, J. and Thomsen, R.: "A comparative Study of Differential Evolution, Particle Swarm Optimization, and Evolutionary Algorithms on Numerical Benchmark Problems", *Congress on Evolutionary Computation, 2004. CEC2004*, 2, pp. 1980-1987, 2004.
- Whitley, D., Mathias, K., Rana, S. and Dzubera, J.: "Evaluating Evolutionary Algorithms", *Artificial Intelligence*, 85, pp. 245-276, 1996.
- Yao, X. and Liu, Y.: "Fast Evolutionary Programming", in Fogel, LJ, Angeline, PJ and Bäck, T (eds) *Proc. 5th Annual Conf. on Evolutionary programming*, pp. 451-460, MIT Press, Mass, 1996.