Optimisation with Simulated Annealing through Regularisation of the Target Function

Enrique C. Segura
Departamento de Computación,
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires
Ciudad Universitaria, Pab. I, (1428) Buenos Aires, Argentina
esegura dc.uba.ar

Abstract

A method is presented for function optimisation that generalises the Simulated Annealing algorithm by applying convolutions of the target function with smooth, infinitely differentiable kernels. Hence the search for a global optimum is performed over a sequence of functions that preserve the structure of the original one and converge to it pointwise. From an experimental point of view, the purpose of this paper was to compare the efficiency of this approach with that of the conventional Simulated Annealing. To do this, the proposed technique was tested both on complex combinatorial (discrete) problems (e.g. the Travelling Salesman Problem) and on the search of global minima for continuous functions. In some cases, performance was improved in terms of final results, while in other ones, even if no improvements were attained over the usual Simulated Annealing algorithm, the proposed method shows interesting abilities to provide fairly good approximations in relatively few iterations, i.e. at early stages of the search process.

eywords: simulated annealing, stochastic optimisation, function regularisation, smooth kernels.

1 I TRO CTIO

The difficulties presented by stochastic optimisation techniques and, in particular, Simulated Annealing (SA) [4][6][9] from the point of view of their applicability for general cases, is a recurrent field of study and experimentation in the area of stochastic optimisation. One of these problems is its high sensitivity regarding the behaviour of the function to be minimised (maximised): if this function is not smooth enough (i.e. for the continuous case, differentiable including its derivatives up to a certain order), the algorithm tends to get stuck on a valley (a deep local minimum), remaining there for a long time, which degrades the efficiency of the method. More specifically, the difficulty arises when local minima appear satisfying the following condition:

$$\min_{\in N(i)} df(i, \ ) \quad L$$

with

$$df(i, \ ) = \ f(i) - f(\ )$$

where f is the function to minimise, i is a local minimum, $N(i)$ is the set of states neighbour to i according to the connection graph and L is the Lipschitz constant defined as:

$$L = \max_{k \in S} \{ \max_{l \in N(k)} df(k,l) \}$$

being *S* the space of configurations of the problem. For the continuous case, the neighbourhood of a given point in the space of configurations is defined as the set of elements at a certain distance apart from it, i.e.

$$N(x) = \{ y \in S / \quad x - y \quad = \quad \} .$$

and the notion of connection graph is similar to that for the discrete case.

In other words, the function presents deep valleys, as deep as the biggest jump in the whole state space, which prevents control of the algorithm by means of its parameter (temperature), since this parameter is adjusted according to the information provided by *L*.

2 THEORETICAL GRO I G

From this motivation, the idea arises of applying some technique for smoothing the function to minimise, so as to preserve its structure, that is, if *f* is the original function and *F* is the transformed (smoothed) one, then, for every pair of states *i* and in *S* with *N(i)* it holds:

$$dF(i, \quad) \leq df(i, \quad)$$

More specifically, we look for a sequence of smooth kernels F_T such that

$$F_T(x) \quad \overset{T\to}{} \to f(x) \tag{1}$$

for any *x* in *S*, being *T* a discrete control parameter (i.e. a temperature taking only on positive integer values).

The method consists essentially of convolving the target function with smooth kernels of the form:

$$g_T(x) = \frac{1}{T} g(x/T)$$

where

$$\int_S g(x)dx = 1 \tag{2}$$

and

$$g \in C_0 \ (S) \tag{3}$$

It is a well known fact in functional analysis [10] that under these assumptions the functions

$$F_T(x) = f \quad g_T = \int_S f(y)g(x-y)dy$$

are infinitely differentiable and satisfy condition (1).

3 THE ALGORITHM

Based on the preceding considerations, an algorithm was conceived to solve optimisation problems for functions with many local minima. The method consists of a SA where each element in the space of configurations is a "macrostate" formed with a fixed number of neighbour states. A cost associated to such a configuration results from the local convolution of the function with a kernel of the form:

$$g_T(x) = \frac{1}{T}\exp\{-(x/T)^2/2\}$$

where T $T(t)$ is the temperature at time t. It is easy to check that such function, properly normalised, satisfies conditions (2) and (3).

The convolution is discretised as:

$$F_T(\bar{x}) = \frac{1}{T}\sum_{i=1}^{N} f(x_N)\exp\{-(([N/2]-i)/T)^2/2\}$$

where $\bar{x} = (x_1,...,x_N)$ stands for the N-tuple of states that form the configuration. This means that the largest weight is assigned to the state equidistant to both ends of the "macrostate", defining the distance between two states as the number of edges from each other in the graph, according to the path followed when generating them.

In general terms, the method proceeds in a similar way as conventional SA. An initial configuration is generated as a sequence of N neighbour states (i.e. a path of length N in the graph). At each iteration one of both ends of the sequence is chosen (with equal probability or taking that of lowest individual cost) and a new neighbour is concatenated to it, eliminating at the same time the state corresponding to the end that was not chosen. The new value of F_T is computed convoluting the new sequence with the new g_T. If the cost of the new configuration $\bar{x}(t+1)$ is less than the previous one, it is accepted; otherwise, it is accepted only with probability $\exp\{-df/T\}$, being

$$df = F_{T(t+1)}(\bar{x}(t+1)) - F_{T(t)}(\bar{x}(t)).$$

4 E PERIME TS

The proposed algorithm was tried on both a) discrete and b) continuous problems, as follows.

a) Two instances of the classical Traveling Salesman Problem [1][3][5] (in ^{2}, Euclidean and with triangle inequality), both considerably complex and with known solution:

i) Map of 100 cities in a squared 10 x 10 array, with distance between neighbours equal to unity.

ii) Map of 120 cities with known minimum cost (example taken from [5]).

The reason to choose the TSP for the experiments in the discrete case is that it is, due to its complexity (a paradigmatic case of NP-complete problem), widely accepted as one of the more representative benchmarks of its type.

Table 1: average performance (over 10 runs for each value of N) for the TSP on problem a)i) (quadrangular lattice of 100 cities). Optimal solution f =100.

N	t=10000	T=50000	t=100000
1	145.1	111.2	105.3
10	183.1	116.4	104.8
20	189.6	112.7	104.2
30	196.5	109.8	103.9

Table 2: average performance (over 10 runs for each value of N) for the TSP on problem a)ii) (from [5]). Optimal solution f = 83.21.

N	t=10000	t=50000	t=100000
1	114.8	92.0	88.1
10	119.3	93.2	87.9
20	124.8	92.6	86.7
30	126.4	91.5	86.4

Tables 1 and 2 show results obtained in both cases for different values of N (case N=1 corresponds to the behavior of usual SA) and for different number of iterations.

First of all, it must be said that the initial hypothesis in the sense that the complexity of the problem could be substantially reduced as a function of the dimension (parameter N) was not confirmed by the experiments. The performance is slightly better than that of the conventional SA, but better solutions are obtained only after many iterations. Anyway, the efficiency grows monotonically, although slowly, with the order of the method (parameter N). The algorithm is still under evaluation; more experiments are currently in progress.

b) Minimisation of the function

$$f(x,y) = a\,x^2 + b\,y^2 - c\,\cos(\quad x) - d\,\cos(\delta\, y) + c + d$$

for $0 \le x \le 5$, $0 \le y \le 5$ with $a = 1$, $b = 2$, $c = 0.3$, $d = 0.4$, $\quad = 3$, $\quad = 4$.

This function has a unique global minimum at $f(0,0) = 0$.

For this continuous case, generation of a new neighbour from a given point (x,y) at each step was performed by randomly selecting a point (x',y') on the circumference defined by $\sqrt{(x-x)^2 + (y-y)^2} = \quad$. A value $\quad = 0.005$ was used for this experiment.

Table 3 displays results obtained for different values of N and for different number of iterations. Note that in this case the necessary number of steps for the algorithm either to find a good solution or being stopped is more than one order of magnitude lower that in the discrete cases.

It is easy to observe that the best results with the proposed technique are obtained only for a parameter N = 20, and these are just slightly better than those reached with the SA in its conventional form. Moreover, for higher values of N the performance gets worse. This might suggest that the proposed technique is not quite convenient in this case, given its higher computational complexity with respect to the classical SA (N = 1).

Table 3: average performance (over 10 runs for each value of N) for minimisation of the continuous function given in b). Optimal solution $f(0,0) = 0$.

N	t=500	t=1000	t=2000
1	0.0038	0.0024	0.00016
10	0.0017	0.0009	0.00020
20	0.0013	0.00073	0.00014
30	0.0015	0.0022	0.00016

However, the values in the first and second columns of the table (t = 500 and t = 1000) could be indicating, in its turn, that for short runs of the algorithm (e.g. when not much computational time can be taken and/or a rough approximation to the optimal solution is considered good enough for certain particular purpose) the technique introduced here might be useful.

5 ISC SSIO

An algorithm for stochastic optimisation was presented, featuring a generalised formulation of the classical Simulated Annealing by applying convolutions of the target function with smooth, infinitely differentiable kernels. The search for a global optimum is performed over a sequence of functions that preserve the structure of the original one and converge to it pointwise.

Convergence to a global minimum (maximum) is, the same as in the conventional SA, guaranteed from a theoretical point of view [2][7][8].

As for practical concerns, the aim of this work was to compare the capabilities of the proposed method with those of the usual Simulated Annealing. In this sense, the results confirm that it avoids long permanence in local minima with deep valleys, thus providing optimal or sub-optimal solutions in lower times than the conventional SA algorithm.

On the other hand, from the point of view of absolute efficiency (i.e. the best solutions obtained regardless of the time elapsed), some difference must be remarked between the discrete and the continuous cases . While in the former this efficiency grows monotonically, although slowly, with the order of the method (parameter *N)*, in the latter (continuous) the results get worse beyond a certain value of *N*. In this case, however, faster convergence to coarse, approximate solutions is observed.

All these considerations are still subject to more extensive experimentation.

REFERE CES

[1] Cerny, V.; Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm; . *Opt. Th. and Appl.*, 45(1):41-51, 1985.

[2] Geman, S. and Geman, D.; Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images; *IEEE Trans. PAMI* 6(6):721-741, 1984.

[3] Gutin, G. and Punnen, A.P. (eds.); *The Traveling Salesman Problem and its Variations*, Springer, 2002.

[4] Kirkpatrick S., Gelatt C. D. and Vecchi M.P.; Optimization by Simulated Annealing; *Science*, 220: 671-680, 1983.

[5] Lawler E.L., Lenstra J.K., Rinooy Kan A. H. G. and Shmoys D.B. (eds.); *The traveling salesman problem. A guided tour of combinatorial optimization*; J. Wiley and Sons, 1985.

[6] Metropolis N., Rosenbluth A., Rosenbluth M., Teller A.H. and Teller E.; Equation of state calculations by fast computing machines; . *Chem. Phys.*, 21(6):1087-1091, 1953.

[7] Mitra D., Romeo F. and Sangiovanni-Vincentelli A.; Convergence and finite-time behavior of simulated annealing; *Adv. Appl. Prob.* 18: 747-771, 1986.

[8] Otten R.H.J.M. and Ginneken L.P.P.P. van; *The Annealing Algorithm.* The International Series in Engineering and Computer Science, Springer, 1989.

[9] Salamon P., Paolo Sibani P. and Frost R.; *Facts, Con ectures, and Improvements for Simulated Annealing*. SIAM Monographs on Mathematical Modeling and Computation, 2002.

[10] Yosida K.; *Functional Analysis*; Berlin, Springer-Verlag, 2003.