HYBRID SIMULATED ANNEALING

R. SALAZAR[1], R. TORAL[1,2]
(1) Departament de Física, Universitat de les Illes Balears
(2) Instituto Mediterráneo de Estudios Avanzados (IMEDEA, UIB-CSIC)
07071 Palma de Mallorca, Spain.

We propose a variant of the Simulated Annealing method for optimization in the multivariate analysis of differentiable functions. The method uses the Hybrid Monte Carlo algorithm for the proposal of new configurations. We show how this choice can improve the performance of simulated annealing methods by allowing much faster annealing schedules.

1 Introduction

We address here the problem of finding the value of the N–dimensional vector $\mathbf{x} = (x_1, x_2, ..., x_N)$, absolute minimum of the real function $E(\mathbf{x})$. For large N, a direct search method is not effective due to the large configuration–space available. Moreover, more sophisticated methods, such as those using the gradient of $E(\mathbf{x})$, might get stuck in local minima and never be able to reach the absolute minimum. One of the most effective methods devised to overcome these difficulties is that of "Simulated Annealing" (SA), which allows escaping from local minima through tunnelling and by accepting higher values of $E(\mathbf{x})$ with a carefully chosen probability[1]. This method is based on an analogy with Statistical Physics: a system with N degrees of freedom $(x_1, \ldots, x_N)$ at temperature T has a probability of being on the state with energy $E(\mathbf{x})$ given by the Gibbs factor:

$$P(\mathbf{x}) \sim \exp(-E(\mathbf{x})/T) \tag{1}$$

From this relation we can see that high energy states can appear with a finite probability at high T. If we now diminish the temperature slowly, the high energy states become less probable and, as $T \to 0$, only those states near the minimum of $E(\mathbf{x})$ have a finite probability to appear. In this way, by decreasing the temperature we can arrive, when $T \to 0$, to the (absolute) minimum energy state.

SA proceeds by interpreting the function $E(\mathbf{x})$ we want to minimize as the "energy" of a fictitious system. It then generates representative configurations of this system at "temperature" T and slowly lower in "time" the system temperature $T = T(t)$ until it reaches $T = 0$ (annealing schedule). The Monte-Carlo method is used to generate the representative configurations. It does so by proposing configuration changes $\mathbf{x} \to \mathbf{x}'$ with probability $g(\mathbf{x}'|\mathbf{x})$ and accepting those changes with probability $h(\mathbf{x}'|\mathbf{x})$. Some freedom is available in the proposal and the acceptance probabilities, as far as the detailed balance condition is satisfied:

$$g(\mathbf{x}'|\mathbf{x})h(\mathbf{x}'|\mathbf{x})P(\mathbf{x}) = g(\mathbf{x}|\mathbf{x}')h(\mathbf{x}|\mathbf{x}')P(\mathbf{x}') \tag{2}$$

In the standard SA methods (see below), the proposal probability is a symmetric function of the difference $\Delta\mathbf{x} \equiv \mathbf{x}' - \mathbf{x}$, $g(\Delta\mathbf{x}) = g(-\Delta\mathbf{x})$. A commonly used solution to the detailed balance equation is the Metropolis choice:

$$h(\mathbf{x}'|\mathbf{x}) = \min\left(1, \exp\left[-\left(E(\mathbf{x}') - E(\mathbf{x})\right)/T\right]\right) \tag{3}$$

The various SA methods differ in the choice of the proposal probability $g(\Delta\mathbf{x})$ and in the annealing schedule $T(t)$. The latter must be coherent with the choice for $g(\Delta\mathbf{x})$ in such a way that the configuration space is efficiently sampled. Amongst the many choices proposed in the literature, we mention the following two that will be used later:

-*Boltzmann Simulated Annealing* (BSA)[2]: Based on a functional form derived for many physical system belonging to the class of Gaussian-Markovian systems, the algorithm chooses a proposal probability given by a Gaussian distribution:

$$g(\Delta\mathbf{x}) \sim \exp\left[-\frac{|\Delta\mathbf{x}|^2}{2T(t)}\right] \tag{4}$$

This requires a particularly slow annealing schedule: $T(t) = T_0/\ln(t)$.

-*Fast Simulated Annealing* (FSA)[3]: The proposal probability is given by a N–dimensional Cauchy distribution:

$$g(\Delta\mathbf{x}) \sim T(t)(|\Delta\mathbf{x}|^2 + T(t)^2)^{-\frac{N+1}{2}} \tag{5}$$

which allows a much faster annealing schedule: $T(t) = T_0/t$.

2 Hybrid Simulated Annealing

The alternative method we propose –Hybrid Simulated Annealing (HSA)– uses the Hybrid Monte–Carlo (HMC)[4] to generate the representative configurations. The main advantage is that, by using a Hamiltonian dynamics for the proposal of the new configuration, HMC allows a much faster annealing schedule, so increasing the efficiency of the simulated annealing.

HMC introduces a set of auxiliary momenta variables $\mathbf{p} \equiv (p_1, \ldots, p_N)$ and a Hamiltonian function $\mathcal{H}(\mathbf{x}, \mathbf{p})$:

$$\mathcal{H}(\mathbf{x}, \mathbf{p}) = E(\mathbf{x}) + \mathbf{p}^2/2 \tag{6}$$

From the Gibbs factor:

$$P(\mathbf{x},\mathbf{p}) \sim \exp[-\mathcal{H}(\mathbf{x},\mathbf{p})/T] = \exp[-E(\mathbf{x})/T]\exp[-\mathbf{p}^2/2T] \tag{7}$$

we see that, from the statistical point of view, the momenta $\mathbf{p}$ are nothing but a set of independent Gaussian distributed random variables of zero mean and variance equal to the system temperature.

For the proposal probability $g(\mathbf{x}'|\mathbf{x})$ HMC generates $\mathbf{p}$ using the Gaussian distribution $\exp[-\mathbf{p}^2/2T]$ and then integrates numerically Hamilton's equations of motion by using the *leap–frog* algorithm:

$$\mathbf{x}' = \mathbf{x} + \frac{\delta t^2}{2}\mathbf{F}(\mathbf{x}) + \delta t\mathbf{p} \tag{8}$$

$$\mathbf{p}' = \mathbf{p} + \frac{\delta t}{2}(\mathbf{F}(\mathbf{x}) + \mathbf{F}(\mathbf{x}')) \tag{9}$$

where $F_i(\mathbf{x}) = -\partial E(\mathbf{x})/\partial x_i$ is the "force" acting on the variable x_i. This proposal must now be accepted with a probability given by:

$$h(\mathbf{x}'|\mathbf{x}) = \min\left(1, \exp\left[-\left(\mathcal{H}(\mathbf{x}') - \mathcal{H}(\mathbf{x})\right)/T\right]\right) \tag{10}$$

Notice that in the case $T \to 0$ the random component of the evolution (the momenta variables) in Eq.(8) is zero and then the proposal coincides with that of gradient methods.

The HMC has been extensively used in problems of Statistical Physics[5]. For our purpose here, we have found that the use of the Hamiltonian dynamics of the statistical system associated with the energy $E(\mathbf{x})$, allows a much more effective annealing schedule than, for instance, the Boltzmann and Fast annealing methods mentioned above. In particular we have been able to use an exponential annealing schedule: $T(t) = T_0 e^{-ct}$. This fast annealing schedule can not be efficiently used with the FSA and BSA methods.

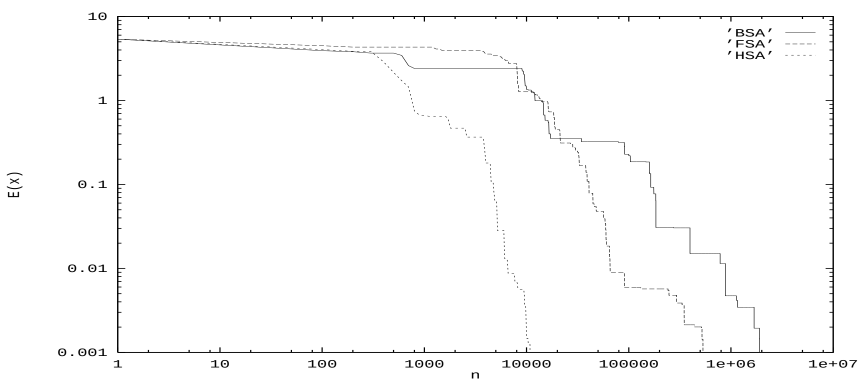
To make this point clear, we present results of the optimization of a typical test function:

$$E(\mathbf{x}) = b\sum_{i=1}^{N} v[a(x_i - i + N/2), \beta] \tag{11}$$

$$v(z,\beta) = z^2[1 - \beta cos(z)] \tag{12}$$

performed with the three methods described: HSA, BSA and FSA. The chosen function presents many local minima but only one absolute minimum at $\mathbf{x}_i = i + N/2$ of value $E_{minimum}(\mathbf{x}) = 0$ (we have used $N = 10$, $\beta = 0.5$, $a = 10$ and $b = 0.0001$). In the figure we plot the remaining energy versus the number n

of function evaluations for each of the three methods: Hybrid, Boltzmann and Fast Simulated annealing. It is clear from this figure that HSA requires orders of magnitude less evaluations of the function than the other two methods and can, therefore, give a precise answer in less computer time. This conclusion remains despite the fact that HSA requires some extra work when computing the evolution equations since it needs to compute also the forces F_i acting on the different variables.



In conclusion, we have shown how the use of the Hybrid Monte Carlo algorithm can indeed improve the performance of simulated annealing methods by allowing much faster annealing schedules. It is conceivable also that one could then use efficiently some of the acceleration schemes (Fourier, wavelet, etc.) available for Monte Carlo methods in order to improve upon the convergence of the simulated annealing techniques.

We acknowledge financial support from DGICyT, grants PB94-1167 and PB94-1172. R. Salazar is supported by the Agencia Española de Cooperación Internacional in the Mutis program.

References

1. S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, *Science* **220**, 671 (1983).
2. S. German and D. German, *IEEE Trans. Patt. Anan. Mach. Int.* **PAMI-6**, 721 (1984).
3. H. Szu and R. Hartley, *Phys. Lett.* A **122**, 157 (1987).
4. S. Duane *et al.*, *Phys. Lett.* B **195**, 216 (1987).
5. R. Toral, in *3th Granada Lectures in Computational Physics*, P.L. Garrido, J. Marro, eds. Springer (1995).