RAPID COMMUNICATIONS

PHYSICAL REVIEW E VOLUME 53, NUMBER 2 FEBRUARY 1996

Nonpolynomial fitting of multiparameter functions

Jürgen Schulte*
Electron Devices Department, Hitachi Central Research Laboratory, Hitachi, Ltd., Tokyo 184, Japan
(Received 15 November 1995)

A stochastic self-regulating simulated annealing optimization method is presented, and compared to other optimization methods such as the simplex, steepest descent, and the recently proposed fast fitting method by Penna [Phys. Rev. E **51**, R1 (1995)]. The presented method converges faster towards an acceptable set of optimization parameters than the other methods, and it is less susceptible to local minima of nonconvex functions. Examples are shown for fitting a simple two parameter Gaussian function and a complicated multiple parameter three-body interaction potential function.

PACS number(s): 02.70.−c, 02.50.Ey, 02.60.Pn, 05.70.Ln

Optimizing the parameters of a function to a given set of data is a frequent task to be performed in many areas in science and engineering. For simple functions with one, or only very few, parameters, and only few and well separated local minima, standard methods such as the simplex optimization and the steepest descent method are popular, and have proven to provide reasonable results, especially in cases where the shape of the function to be optimized, and the range of parameters close to the desired global minimum, is well known. With a large number of parameters (e.g., $k>5$) to be optimized, or with functionals comprising different families of function terms, optimizing a function to a given data set can become a tedious and computationally demanding undertaking, and is often accompanied by a superimposed initial parameter guess trial and error method.

In general, the simulated annealing method overcomes most of the problems that the simplex and steepest descent type methods present. Due to the inherent statistical nature of simulated annealing, local minima can be overcome much more easily, however, it requires careful handling of the quenching of the annealing temperature. Thus, standard simulated annealing may converge to a "better" global minimum, however, only if the parameter visit or jump range is closed in (quenched) fairly slowly, and thus resulting in slow, though better, convergence.

Recently, a fast simulating annealing optimization and fitting method based on Tsallis statistics has been presented by Penna [1,2]. This method has proven to be very fast in finding the (acceptable) global minimum when fitting a Gaussian function with two parameters to a given set of Gaussian function data. Though the method presented in [1,2] is a much faster optimization method than the simplex, steepest descent, or simple Metropolis annealing, it becomes less efficient when applied to multiple minima functions, and especially when applied to multiple minima functions with many parameters. In this Rapid Communication a more reliable, and even faster, Tsallis statistics simulated annealing optimization and fitting method is presented.

A well established method in optimization and function fitting is to minimize the sum over the squared deviation of function values and a given data sample,

$$\varepsilon^2=\sum_{i=1}^{N}[y(x_i,s_1,\ldots,s_k)-y_i]^2, \tag{1}$$

with $y(x,s_1,\ldots,s_k)$ being the function, and $s_1,\ldots,s_k$ the function's free parameter set to be optimized to the given data sample y_i.

It has been shown that in order to reach a global minimum with a Gaussian visiting distribution, the annealing temperature must decrease logarithmically with the cooling sequence [3]. This implies, however, slow convergence or less efficiency than other methods. Applying a Cauchy-Lorentz parameter visiting distribution improves the simulated annealing convergence significantly, i.e., the temperature decreases like the inverse of time (cooling sequence step) [4]. In order to further improve computational efficiency, the probability of a parameter set being accepted can be defined through a generalized Metropolis algorithm [1,2] based on the principles of a generalized statistical mechanics [5]. As has been shown by Penna [1,2], this approach results in very fast convergence, if the number of free parameters k is small. If the functional to be optimized, or fitted, is complicated and is comprised of parameter values with orders of magnitude differences, some drawbacks of the Penna approaches become apparent.

In particular, with many parameters to be optimized, and a good initial guess of parameters, or even the order of magnitude of the parameters, not known beforehand, the Cauchy-Lorentz distribution annealing procedure, and the annealing procedure of the generalized Metropolis parameter acceptance probability, may cause some parameters to converge very slowly while others get trapped in a local minimum. Penna suggested that the annealing sequences of the Cauchy-Lorentz distribution and the generalized Metropolis parameter acceptance probability may be decoupled by introducing different annealing temperatures for the respective probabilities. However, the fundamental problem of trial parameters being subject to the same respective probability distribution and annealing quenching sequence is not solved. Thus, slow convergence may occur when many parameters are to be optimized.

Analyzing various simulated annealing methods, we found the trial parameter selection and subsequent acceptance procedure following all the same algorithm time scales

*FAX: +81-423-27-7748. Electronic address: schulte@crl.hitachi.co.jp

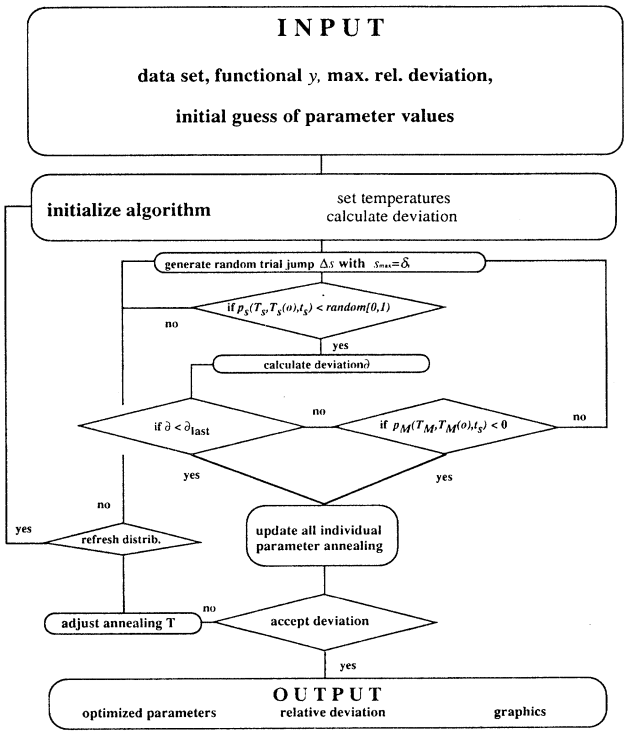

FIG. 1. Flowchart of the presented fitting and optimization method [8]. Note the differences compared to the Penna method (Ref. [2], Fig. 1).

(or annealing sequence) as one of the major reasons for slow convergence. Thus, we introduce a parameter individual Cauchy-Lorentz type visiting distribution with parameter individual annealing temperatures and parameter individual time scales, i.e., annealing sequence. In addition, every parameter observes an individual generalized Metropolis acceptance probability, with the annealing temperature decoupled from the Cauchy-Lorentz type annealing sequence.

In detail, the Cauchy-Lorentz visiting probability distribution of Stariolo-Tsallis form [6] written in terms of the probability of a parameter trial value s changing from s_t to s_{t+1} is

$$p_{q_\nu}(\Delta s_{t_s}) \propto \frac{T_s(t_s)^{-D/(3-q_\nu)}}{[1+(q_\nu-1)(\Delta s_t)^2/T_s(t_s)^{2/(3-q_\nu)}]^{[1+(q_\nu-1)]+[(D-1)/2]}} \quad (2)$$

where $T_s(t_s)$ indicates that every parameter and configuration vector element s observes its individual time scale t_s, and D being the dimension of the parameter space, and $\Delta s_{t_s}=s_{t_s+1}-s_{t_s}$ being the size of a respective parameter trial jump, and q_ν controlling the shape of the Cauchy-Lorentz type distribution. In the case $q_\nu=2$ Eq. (2) represents the proper Cauchy-Lorentz visiting distribution; with $q_\nu\to 1$ it changes to the Gaussian visiting distribution of the classical simulated annealing. The individual annealing time scales t_s are set according to the algorithm flowchart in Fig. 1. The individual generalized Metropolis parameter acceptance probability is

$$p_{s(t_s)}=\min\{1,[1-(1-q_a)(\Delta\varepsilon^2)/T_{M_s}(t_s)]^{1/(1-q_a)}\}, \quad (3)$$

where $\Delta\varepsilon^2$ is the difference of the sum over the squared deviations achieved with the parameter set s_{t_s} at time t_s and respective deviations achieved with the parameter set at the previous time step t_{s-1}. T_{M_s} is the generalized Metropolis individual parameter annealing temperature. The parameter range of q_a is $q_a\in[-10,1)$ which provides sufficient flexibility of fast annealing, and avoids numerical instability. In the presented examples $q_a=-5$ and $q_\nu=2$ have been found reasonable values, confirming the results in [1].

In order to overcome the problem of optimization parameters being trapped in local minima, the parameter distribution probabilities are quenched according to individual parameter annealing temperatures T_s with

$$T_s(t_s)=T_s^0\frac{2^{q_\nu-1}}{(1+t_s)^{q_\nu-1}-1}, \quad (4)$$

where $T_s(t_s)$ indicates that every parameter s observes its individual time scale t_s, and T_s^0 being the parameter individual starting temperatures that are functions of the respective initial trial parameters. The parameter range for q_ν is $q_\nu\in[2,3)$, with $q_\nu=2$ giving a Cauchy-Lorentz distribution, and for $q_\nu>2$ giving a somewhat distorted Cauchy-Lorentz distribution. With the individual time scales for every parameter, the visiting distributions and the acceptance probabilities can now close in at the respective optimal parameter value according to an annealing speed appropriate to the actual offset of the parameter from its optimal value. Thus, the algorithm becomes less susceptible to local minima, and therefore gains computational efficiency. Additional speedup in convergence can be achieved by introducing individual Metropolis annealing temperatures. In our examples, we have chosen the generalized Metropolis annealing sequence, and temperature, to observe the equivalent functional of Eq. (4). In order to easily confirm the reliability of optimized parameters we allow individual parameter trial distributions to refresh after a preset t_s to its original distribution (Fig. 1) with the latest optimized parameter as the new initial guess. A welcome additional effect of this self-regulating refreshing loop is that more local minima may be overcome, and that it allows convergence and annealing to be much faster than just being proportional to the inverse of T_s. Setting the confirmation, or refreshing, rate has been found not to be critical, and can be done in a quite general way. The presented optimization method can best be understood by comparing the flowchart of Fig. 1 here, and the C code listing in Fig. 1 of Ref. [2].

The central idea of the presented approach of allowing an entire set of temperatures to evolve along individual paths may be considered reminiscent of those experiments in which different degrees of freedom, e.g., electron spin and nuclear spin, tend to thermalize at different temperatures.

The performance of the presented optimization and fitting method, compared to other methods, is shown in Fig. 2 where a two parameter Gaussian function has been fitted to a set of exact Gaussian data, and in Fig. 3 where a multipa-

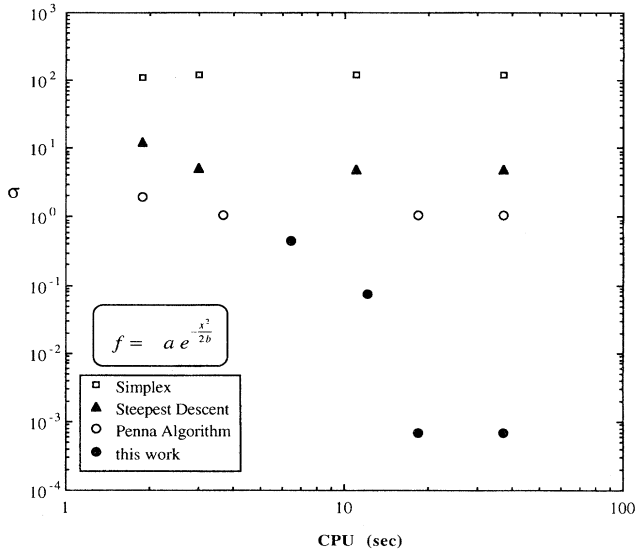

FIG. 2. The relative deviations σ of various optimization methods as a function of the CPU time required for the fit of a two parameter Gaussian function.

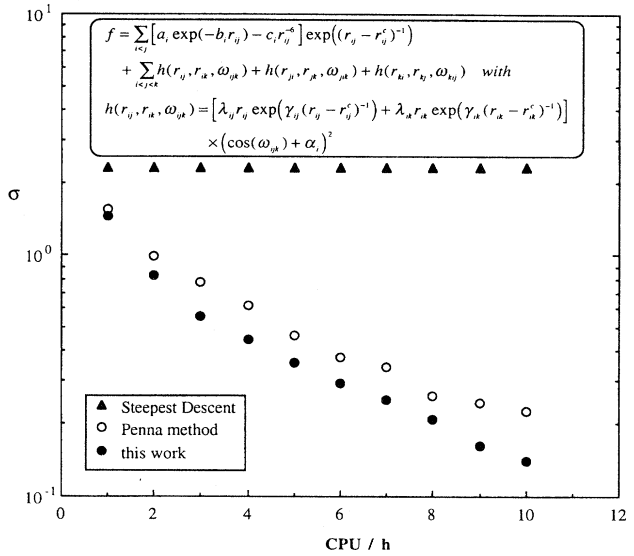

FIG. 3. The relative deviations σ as function of the CPU time required for the complicated multiparameter fit to an *ab initio* energy surface [7] comprising 600 data points.

rameter nontrivial three-body atomic interaction potential function ($D=9$, $k=10$) has been fitted to a large set of *ab initio* calculated energy data (details and results of the *ab initio* fit will be shown elsewhere [7]). In Figs. 2 and 3 the relative deviations of typical runs, as measured for accuracy and convergence, are plotted over the CPU time required on an ordinary PC 486. In the case of the Gaussian function, all parameter values were initially set to 1. The fitting methods being compared in Figs. 2 and 3 were initialized with identical initial guess parameters. Note the logarithmic scale of the relative deviation σ, and the linear time scale.

Figure 2 shows that the simplex and steepest descent fitting of a Gaussian function are very susceptible to local minima, resulting in large relative deviation and poor convergence. The fast annealing method proposed by Penna overcomes many local minima very fast, however, eventually gets trapped in difficult local minima as well. The reason for this behavior is that all parameter trial distributions have followed a uniform annealing sequence, in this case (as can be seen from Fig. 2) proportional to the inverse of the temperature. With the method presented here three orders of magnitude higher accuracy and orders of magnitude faster convergence compared to the Penna method could be achieved. In fact, the overall annealing sequence turned out to be proportional to somewhat faster than the inverse square of the temperature, a speed that a uniform application of Eq. (4) would not allow. The performance of the complicated multiparameter fit is shown in Fig. 3. It clearly shows the fast convergence of the Penna method tailing after 6 h and 8 h, while the method of this work consistently found better minima with a speed of convergence 30% faster than the original Penna method. The presented fitting method has been applied to Buckingham-type potentials ($D=5$, $k=5$) with long-range tails, and to the discontinuous three-body Tersoff potential ($D=9$, $k=18$) [9]. The discontinuity could be fitted well; the long-range tail, however, proved a difficult task, though the presented method outperformed the other simulated annealing methods by far.

[1] T. J. P. Penna, Phys. Rev. E **51**, R1 (1995).
[2] T. J. P. Penna, Comput. Phys. **9** (3), 341 (1995).
[3] S. Geman and G. Geman, IEEE Trans. Patt. Anal. Mach. Int. **PAMI-6**, 721 (1984).
[4] H. Szu and R. Hartley, Phys. Lett. A **122**, 157 (1987).
[5] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[6] D. A. Stariolo and C. Tsallis, in *Annual Review of Computational Physics II*, edited by D. Stauffer (World Scientific, Singapore, 1994).
[7] J. Schulte, J. Ushio, Y. Takemura, and T. Maruizumi (unpublished).
[8] Patent Receipt No. 319501416, Intellectual Property Office, Hitachi, Ltd., Kokubunji, Japan.
[9] J. Schulte, J. Ushio, Y. Takemura, and T. Maruizumi (unpublished).