Genetics and Molecular Biology, 27, 4, 616-622 (2004)

www.sbg.org.br

Research Article

Performance and parameterization of the algorithm Simplified Generalized Simulated Annealing

Alcino Dall'Igna Júnior[1,2], Renato S. Silva[3], Kleber C. Mundim[4] and Laurent E. Dardenne[3]

[1]*Universidade Federal de Alagoas, Departamento de Tecnologia da Informação, Maceió, AL, Brazil.*
[2]*Laboratório Nacional de Computação Científica, Coordenação de Formação de Recursos Humanos, Petrópolis, RJ, Brazil.*
[3]*Laboratório Nacional de Computação Científica, Coordenação de Mecânica Computacional, Petrópolis, RJ, Brazil.*
[4]*Universidade de Brasília, Instituto de Química, Brasília, DF, Brazil.*

Abstract

The main goal of this study is to find the most effective set of parameters for the *Simplified Generalized Simulated Annealing* algorithm, **SGSA**, when applied to distinct cost function as well as to find a possible correlation between the values of these parameters sets and some topological characteristics of the hypersurface of the respective cost function. The **SGSA** algorithm is an extended and simplified derivative of the **GSA** algorithm, a Markovian stochastic process based on Tsallis statistics that has been used in many classes of problems, in particular, in biological molecular systems optimization. In all but one of the studied cost functions, the global minimum was found in 100% of the 50 runs. For these functions the best visiting parameter, q_v, belongs to the interval [1.2, 1.7]. Also, the temperature decaying parameter, q_T, should be increased when better precision is required. Moreover, the similarity in the *locus* of optimal parameter sets observed in some functions indicates that possibly one could extract topological information about the cost functions from these sets.

Key words: optimization, generalized simulated annealing.

Received: August 15, 2003; Accepted: August 20, 2004.

Introduction

A large number of problems in physics, chemistry and biology have as central point the minimization of an appropriate energy function for finding the global minimum of a particular target function. The protein folding and the ligand-receptor docking problems are two examples of challenges in the molecular biology field where the development of efficient and robust global optimization algorithms play a central role in order to find the conformational geometry associated to the global minimum of the molecular free energy hypersurface. Biological macromolecules and biomolecular complexes present a very complex free energy landscape with thousands of local minima. This fact dramatically increases the probability of an optimization process to be trapped in local minima and consequently turns the global minimum determination into a very difficult task. To cope with this problem one should choose a powerful optimization algorithm and understand it deeply to obtain a robust and efficient optimization protocol. This work investigates a simplified and extended version of the **GSA***, Generalized Simulated Annealing*, algorithm of optimization (Tsallis and Stariolo, 1995; 1996), called **SGSA** (*i.e., Simplified GSA*), with the main objective of understanding the role of the **SGSA** parameters in its performance in order to guide their choices in future biomolecular optimization studies.

The **GSA** algorithm or *Tsallis machine* is a Markovian stochastic process, based on Tsallis statistics (Tsallis, 1988; Curado and Tsallis, 1992), that has been used in many classes of problems, like physics and chemistry (Dorfman *et al.*, 2001; Ellis *et al.*, 2000; Mundim *et al.*, 2001; Gutterres *et al.*, 1999; Xiang *et al.*, 1997; Zhaoxian and Dang, 2003), and in particular in molecular systems optimization and protein folding problems (Andricioaei and Straub, 1996; Hansmann, 1997; Moret *et al.*, 1998; Moret *et al.*, 2001; Moret *et al.*, 2002; Mundim and Tsallis, 1996).

Send correspondence to Alcino Dall'Igna Júnior. Laboratório Nacional de Computação Científica, Coordenação de Formação de Recursos Humanos, Secretaria de Pós-graduação, Av. Getúlio Vargas 333, Quitandinha, 25651-075 Petrópolis, RJ, Brazil. E-mail: alcino@lncc.br.

The **GSA** is a generalization of the *Simulated Annealing* algorithm, **SA** (Kirkpatrick *et al.*, 1983), also known as *Boltzmann machine* because it is based on Boltzmann-Gibbs statistics, and of the *Fast Simulated Annealing* algorithm, **FSA** (Szu and Hartley, 1987), or *Cauchy machine*, based on Cauchy-Lorentz probabilistic distribution.

The simulated annealing algorithms family depends on a *visiting function* that determines how the domain of the function is searched, and on an *acceptance function* that says if a result of higher "energy" should be accepted or rejected.

In the original **SA**, the visiting function was simply a random variable choice, due to the binary nature of the variables. The acceptance function,

$$P(\Delta E) = e^{-\Delta E / k_B T(t)} \tag{1}$$

where k_B is the Boltzmann constant, gives the Boltzmann-Gibbs distribution nature of the movement of the reference point. In problems with more complex domains, the Boltzmann-Gibbs probability distribution function is also used as a visiting function. It was demonstrated (Geman and Geman, 1984) that in this case the maximum temperature decaying ratio should be $T(t) = T_0 / \log(1 + t)$, where the "time" t is the iteration step, to guarantee the theoretical convergence of the algorithm.

The **FSA** algorithm uses as its visiting function a Cauchy-Lorentz probability distribution function:

$$G_{T_c} = \frac{T_c(t)}{x^2 + T_c^2(t)} \tag{2}$$

where x is the variable of interest, and maintains the same acceptance function of the **SA** algorithm. Szu and Hartley proved that in this case T could decay with the inverse of the computing step, $T_c(t) = T_0 / (1 + t)$, because even in relatively low temperatures long range jumps are still possible, which made the annealing faster than in the **SA** algorithm.

The *Simplified Generalized Simulated Annealing* algorithm, **SGSA**, is an extended and simplified derivative of the **GSA** algorithm, with a reduced computational cost and the capacity to deal successfully with finite domain problems such as grid based receptor-ligand docking methodologies (Meng *et al.*, 1992; Luty *et al.*, 1995; Garrett *et al.*, 1998).

Material and Methods

Given a cost function, the simulated annealing family of algorithms works as follow:

1. From an initial set of values of the parameters of the given cost function, generally randomly chosen, an initial "energy", E_{ref}, of the system is evaluated and an initial "temperature" $T = T_0$ is selected;

2. a random perturbation is generated into the parameters of the cost function using the *visiting function*, and the new "energy" of the system, E_{new}, is then calculated;

3. if $\Delta E = E_{new} - E_{ref} \leq 0$, the new point is better or at least of the same quality as the previous one, the new set of values of the parameters of the function become the reference set;

4. if $\Delta E > 0$, the new point is worse than the reference point but still could be accepted depending upon the *acceptance probability function* and a random number, as defined in the Metropolis criteria (Metropolis *et al.*, 1953);

5. the "temperature" T is decreased, according to a *temperature decaying function*;

6. step 2 thru step 5 are repeated during a giving number of steps or until some other stopping criteria becomes satisfied.

The GSA algorithm

The **GSA** algorithm uses for the acceptance probability function in the cases where $E(x_{t+1}) > E(x_t)$, the expression

$$P_{q_A}(\Delta E(x)) = \frac{1}{1 + \left\{ 1 + \dfrac{(q_A - 1)[E(x_{t+1}) - E(x_t)]}{T_{q_A}(t)} \right\}} \tag{3}$$

where q_A ($1.0 < q_A < 3.0$) is the *acceptance parameter*.

The visiting function depends on the Tsallis probability density function:

$$g_{q_V}(x) = \left(\frac{q_V - 1}{\pi} \right)^{D/2} \times \frac{\Gamma\left(\dfrac{1}{q_V - 1} + \dfrac{D-1}{2} \right)}{\Gamma\left(\dfrac{1}{q_V - 1} - \dfrac{1}{2} \right)} \times \frac{T_{q_V}^{D/(q_V - 3)}}{\left[1 + \dfrac{(q_V - 1)x^2}{T_{q_V}^{2/(q_V - 3)}} \right]^{\frac{1}{q_V - 1} + \frac{D-1}{2}}} \tag{4}$$

where D is the dimension of the cost function, and q_V ($1.0 < q_V < 3.0$) is the *visiting parameter*. From the probability distribution function, $G(\Delta x_t)$,

$$G(\Delta x_t) = \int_{-\infty}^{\Delta x_t} g_{q_V}(x) dx \tag{5}$$

to which a randomly chosen value is attributed, a perturbation Δx_t is determined in every iteration

$$x_{t+1} = x_t + \Delta x_t, \quad \Delta x_t = G^{-1}(\Delta x_t) \tag{6}$$

with the temperature decaying in Equation 4 controlled by

$$T_{q_V}(t) = \begin{cases} \dfrac{T_{q_V}(1)}{1+t} & \text{case } q_V = 2 \\ T_{q_V}(t)\dfrac{2^{q_V-1}-1}{(1+t)^{q_V-1}-1} & \text{otherwise} \end{cases} \quad (7)$$

where $q_V = 2$ is the **FSA** temperature decaying case.

Generally Equation 5 has no analytic solution and Equation 6 must be resolved numerically by means of the inversion of a power series (Moret *at al.*, 1996, 1998).

Usually $T_{q_A}(t) = T_{q_V}(t)$, but there is no specific reason that enforces that.

The idea of *generalized* in the algorithm comes from the fact that in the parameters limit (q_A; q_V) = (1; 1) reproduces the **SA** or *Boltzmann machine*; and (q_A; q_V) = (1; 2) reproduces the **FSA** or *Cauchy machine*.

As x was D-dimensional, Δx_{t_i} for every dimension was originally determined using products of sine and cosine functions, that introduces an artifact in the visiting function. Moret (1996) suggested the application of $g_{q_V}(x)$ independently in every dimension. In any case, two problems arise: the computational cost of the calculus of the inverse of the integral of Equation 5; and second, Δx_t computed in this way is not limited and when the domain is finite it must be normalized.

To cope with these problems, two main simplifications are used in the **SGSA**. The first is to make, for every dimension x_i,

$$\Delta x_i \propto g_{q_V}(r_i) \quad (8)$$

with r_i randomly chosen, greatly reducing the computational cost, because there is no power series to invert. The second is to make $D = 0$, that guarantees $0 \le g_{q_V} \le 1$, in order to maintain x_i always inside a given domain. In this case,

$$g_{q_V}(r_i) = \frac{1}{\left[1+\dfrac{(q_V-1)r_i^2}{T_{q_V}^{2/(q_V-3)}}\right]^{\frac{1}{q_V-1}-\frac{1}{2}}} \quad (9)$$

A lower value of q_V in the $g_{q_V}(r_i)$ gives a more global profile for the visiting function, where long jumps have a greater probability of occurring when compared to the probability given by greater q_V values. On the other hand, a greater value of q_V gives a more local visiting function profile, with high short jump probability and a very fast decreasing of the long jump probability.

Another difference from traditional **GSA** was the introduction of an additional temperature decaying parameter, q_T, in place of q_V in Equation 7, to maintain better and more independent control over the annealing process. A larger value of q_T causes a very fast T_{q_T} decaying with two possible effects: either the convergence to the global minimum is very fast or the algorithm is trapped in a local minimum.

Results and Discussion

With the objective of understanding the role of the **SGSA** parameters sets, (q_V, q_A, q_T), in the algorithm performance, the optimization procedure using six two-dimensional functions as case studies is investigated. The choice of two-dimensional functions permits the comparison between the *locus* of the best **SGSA** parameters for a particular function and its topology. The six functions studies are: Ackley (Solomatine, 1995), Figure 1(a); log-trigonometric (Kvasnicka and Pospíchal, 1997), Figure 1(b); Lavor (Lavor, 2001), Figure 1(c); Schwefel (Schwefel, 1981), Figure 1(d); Goldstein-Price (Solomatine, 1995), Figure 1(e); and De Jong F5 function (De Jong, 1975), Figure 1(f).

The Ackley and log-trigonometric functions have in common a unique and deep global minimum with several local minima around it. Lavor and Schwefel functions have both a smooth profile with an almost undistinguishable global minimum, because many of the local minima basins are very similar to the global minimum basin. The Goldstein-Price function seems at first an easy objective function, but presents a scale problem with a difference of many magnitude orders between the domain and the function hypersurface. Finally, the De Jong F5 function is, as could be easily seen, a nightmare for optimization algorithms, many deep and small minima basins with minimum values close to the global minimum.

The approach adopted in this study was an exhaustive search for the best parameters set, (q_V, q_A, q_T), for the **SGSA** algorithm. Using a stop criteria of 2,500,000 steps, the parameters were scanned using a 0.1 step in the intervals $1.1 \le q_V \le 2.9$, $1.1 \le q_A \le 2.9$, $1.1 \le q_T \le 3.5$, with initial "temperatures" $T_{q_A}(0) = T_{q_V}(0) = 1$. The entire process was repeated 50 times for every parameter set with different random initial conditions for each execution.

In Table 1 are shown some of the performance data in terms of mean number of cycles in 50 runs, with two different RMSD (*Root Mean Square Deviation*) limits, from the exact global minimum solution, for every function studied. The success index presented in the RMSD/Success column shows that for all functions except the De Jong F5 function, the algorithm was able to find the global minimum in all runs for a reasonable number of parameters sets (see cases after the success index). In the De Jong F5 function, in no more than 50% of the runs the algorithm was successful for a particular parameter set. These unsatisfactory results indicate that some work still must be done to improve the algorithm.

From the "Best case" columns it can be seen that the visiting parameter, q_V, belongs to the interval [1.2, 1.7]. An interval a bit larger, [1.1, 1.8], holds almost all good parameter set intervals (see the "Good cases intervals" columns in Table 1). In the same sense, the temperature decaying parameter, q_T, varies in a larger interval [1.2, 2.4], while the

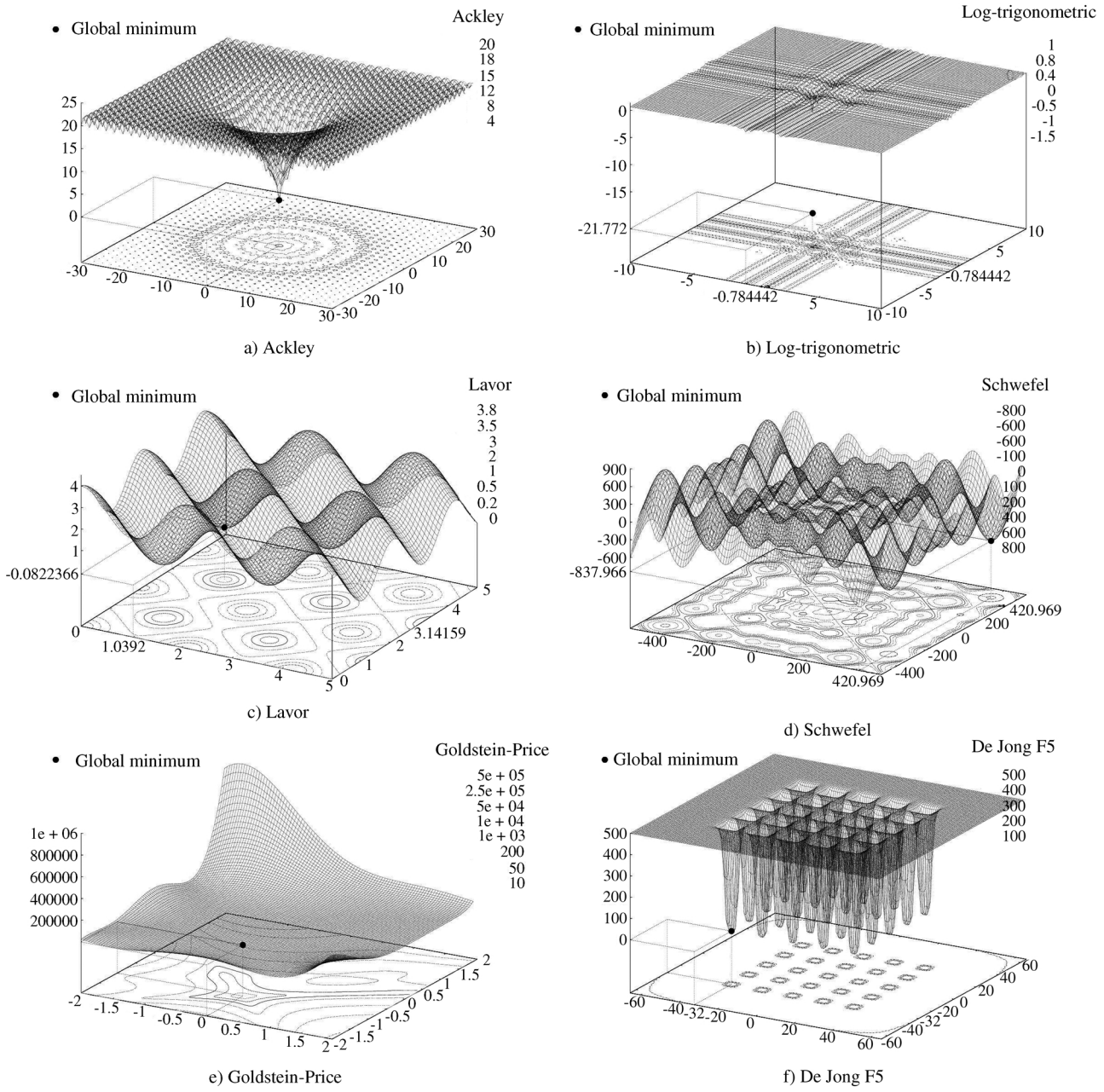

Figure 1 - Test functions used in this work.

good cases intervals are in the interval [1.1, 2.5]. It should be noted that when higher precision is required a greater q_T should be used in most of the cases, which increases the ability of the algorithm to act as a local search engine.

It was observed that the acceptance parameter, q_A, was almost ineffective, indicating that probably the initial acceptance temperature, T_{q_A}, should be increased independently of T_{q_V}. Another non-exclusive option to cope with this ineffectiveness is the use of negative values for q_A, as had been pointed out by Tsallis and Stariolo (1996).

In Figure 2 shows the profile of the **SGSA** parameters q_V and q_T for the best q_A value (see the "Best case" columns in Table 1), that achieve 100% success in 50 runs (for the De Jong F5 function a success index of more than 20% is used) for each studied function. Some degree of similarity in the profile of optimal parameter sets observed in the functions of Ackley and log-trigonometric, and Lavor and Schwefel (Figure 2), could also be observed in the hypersurface of the respective functions (Figure 1), indicating that possibly one could extract topological information about the cost function from these optimal parameter sets. If some information from a cost function hypersurface is known, an *a priori* **SGSA** parameters set range could be chosen. In a case where the function hypersurface is com-

Table 1 - Performance data of the **SGSA** algorithm for the selected two-dimension functions.

Function	RMSD[1]/ success	Best case				Good cases intervals[6]				
		Mean cycle[3]	q_V	q_A	q_T	Limit[4]	Mean cycle[5]	q_V	q_A	q_T
(actual minimum)	(success / cases)[2]	(min. / max.)	(mean min. potential)			(cases)	(min. / max.)			
Ackley	10^{-6}	1630.0	1.5	1.4	2.4	< 5000	3626.9	1.3-1.7	1.1-2.9	2.1-2.5
(0.0)	(100%/348)	(654/2742)	(0.00000281)			(132)	(309/136799)			
	10^{-3}	446.7	1.4	1.2	2.2	< 800	660.8	1.3-1.6	1.1-2.6	1.8-2.4
	(100%/2687)	(145/990)	(0.00028682)			(117)	(118/12312)			
Log-trigonometric	10^{-6}	1008.7	1.2	1.3	1.7	< 5000	2736.8	1.1-1.5	1.1-2.8	1.3-2.1
	(100%/1281)	(390/3450)	(-21.60551500)			(324)	(242/49812)			
(-21.772042)	10^{-3}	624.3	1.2	1.2	1.4	< 1000	851.5	1.1-1.5	1.1-1.9	1.2-1.8
	(100%/2120)	(320/1250)	(-9.67091560)			(120)	(90/5356)			
Lavor	10^{-5}	17202.9	1.7	1.1	2.1	< 40000	32357.0	1.5-1.8	1.1-2.5	1.9-2.1
(-0.8223661)	(100%/1370)	(1379/176394)	(-0.08223661)			(48)	(601/1147572)			
	10^{-2}	264.9	1.4	1.3	1.6	< 500	421.6	1.1-1.8	1.1-2.1	1.3-1.9
	(100%/2856)	(66/933)	(-0.08170597)			(177)	(11/5811)			
Schwefel	10^{-3}	3001.3	1.4	1.1	2.0	< 6000	680.1	1.2-1.6	1.1-2.5	1.8-2.2
(-837.96577)	(100%/1832)	(92/10626)	(-837.96577)			(108)	(92/96094)			
	10^{-2}	732.3	1.3	1.3	2.0	< 1200	1043.8	1.2-1.5	1.1-2.4	1.6-2.0
	(100%/2096)	(72/8441)	(-837.965760)			(99)	(72/8441)			
Goldstein-Price	10^{-4}	3228.9	1.3	1.1	1.5	< 10000	7154.4	1.1-1.6	1.1-2.9	1.3-1.6
	(100%/675)	(422/9302)	(3.00000360)			(184)	(120/210872)			
(3.0)	10^{-2}	252.8	1.2	2.1	1.2	< 400	67.0	1.1-2.7	1.1-2.9	1.1-2.3
	(100%/1849)	(48/782)	(3.03205290)			(138)	(13/6066)			
De Jong F5	10^{-3}	521480.6 [7]	1.4	1.4	1.4	30%[8]	512468.2	1.1-1.5	1.2-2.7	1.1-1.8
(0.0)	(20%/1140)	(10586/2084571)	(1.0918×10^{-18})			(12)	(1683/2481019)			
		(36%/18)								
	10^{-2}	153853.4 [7]	1.4	1.6	1.4	40%[8]	266739.8	1.1-1.7	1.1-2.9	1.1-1.7
	(20%/1143)	(1287/1486847)	(7.4837×10^{-14})			(58)	(184/2481019)			
		(50%/25)								

1. Maximum Root Mean Square Deviation of the function parameters from the exact global minimum solution values. This value is used as success criterion in the optimization process.
2. Minimum SGSA percentage of success in the 50 runs for a particular (q_V, q_A, q_T) set to be counted and the number of sets that met this requirement.
3. Mean number of cycles where the global minimum was reached among the 50 runs for the best (q_V, q_A, q_T) set case and the minimum/maximum number of cycles among them.
4. The cases represent the number of parameter sets for which the global minimum was reached in a mean number of cycles lower than the chosen limit for every function.
5. Mean number of cycles where the global minimum was reached among the 50 runs for the cases of the previous column and the minimum/maximum number of cycles among them.
6. One must note that the parameters intervals are independent of one another.
7. In the De Jong F5 function case, the parameter set considered to have the better success index is represented on the additional line.
8. In the De Jong F5 function case, the limit is determined in function of the success index.

pletely unknown, some insight about it could be obtained by means of an exhaustive search in the parameters space.

The results obtained are useful in indicating a direction for the use of this algorithm in problems like protein folding or ligand-protein docking, reducing significantly the number of algorithm parameter choices as well as giving hints about the effect of parameters on the behavior of the algorithm. With these results some improvements were already achieved in studies in progress using the **SGSA** algorithm in the ligand-protein docking studies in progress, which will be published elsewhere.

Acknowledgments

This work was supported by Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Universidade Federal de Alagoas, UFAL, Conselho Nacional de Pesquisa, CNPq (grant number 40.2003/2003.9), Fundação de Amparo a Pesquisa do Estado do Rio de Janeiro, FAPERJ (grant number E26/171.401/01), and Laboratório Nacional de Computação Científica, LNCC. Special thanks to the Carcara Project at LNCC for the computational cluster resources provided.

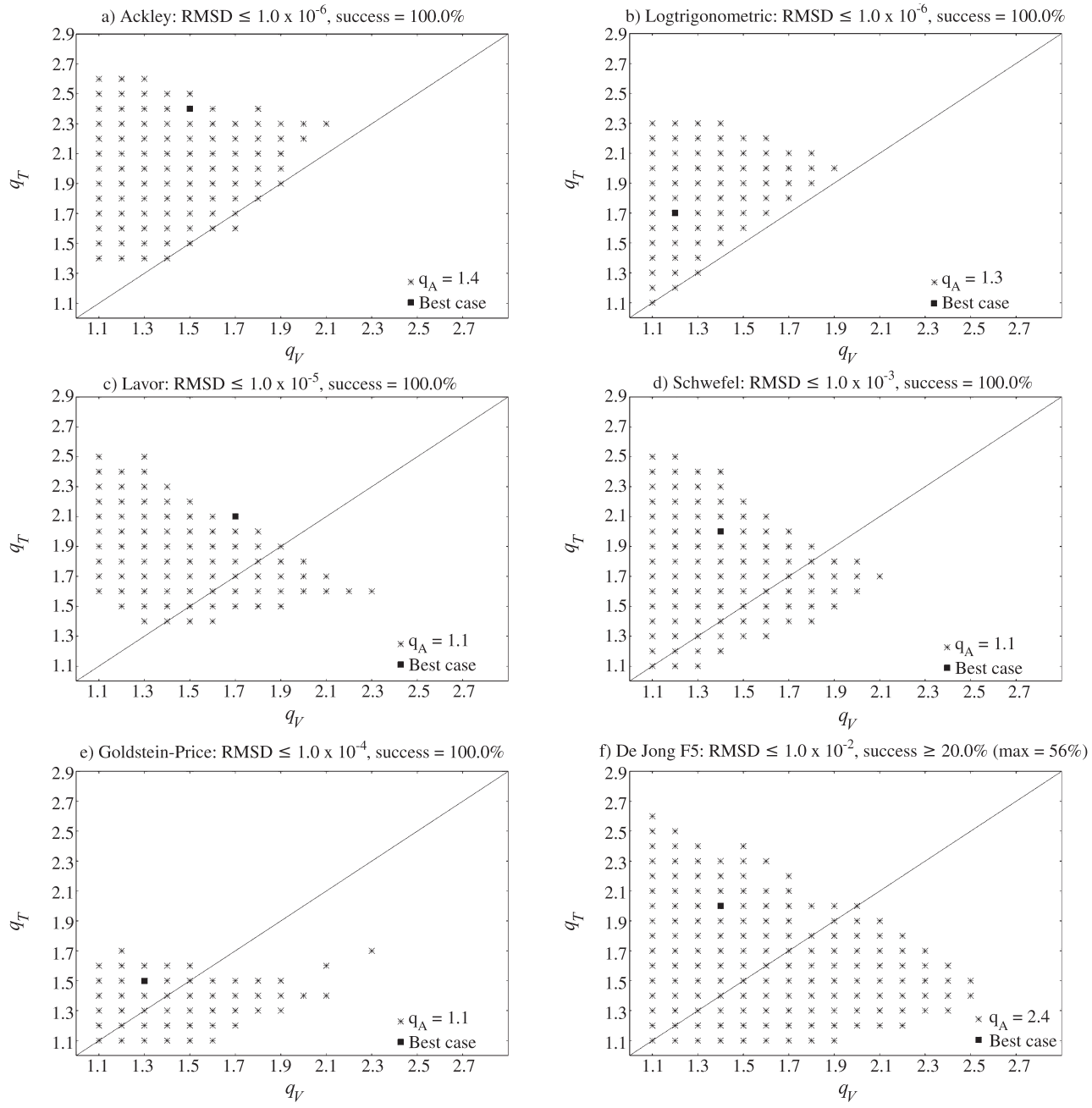

Figure 2 - SGSA parameters q_V and q_T for the best q_A value with 100% success in 50 runs. For the De Jong F5 function a success index of more than 20% was used.

References

Andricioaei I and Straub JE (1996) Generalized simulated annealing algorithms using Tsallis statistics: application to conformational optimization of a tetrapeptide of special interest. Phys Rev E 53:R3055-R3058.

Curado EMF and Tsallis C (1991) Generalized statistical mechanics: Connection with thermodynamics. J Phys A 24:L69-L72. Corrigenda: (1991) J Phys A 24:3187, and (1992) 25:1019.

De Jong K (1975) An analysis of the behavior of a class of genetic adaptive systems. PhD thesis, Dept. Comput. Sci., U. of Michigan, Ann Arbor.

Dorfman S, Mundim KC, Fuks D, Berner A, Ellis DE and Humbeeck JV (2001) Atomistic study of interaction zone at copper-carbon interfaces. Materials Science and Engineering C 15:191-193

Ellis DE, Mundim KC, Fuks D, Dorfman S and Berner A (2000) Modeling of copper-carbon solid solutions. Materials Science in Semiconductor Processing 3:123-127.

Geman S and Geman D (1984) Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell PAMI-6 6:721.

Gutterres RF, Argollo de Menezes M, Fellows CE and Dulieu O (1999) Generalized simulated annealing method in the analysis of atom-atom interaction. Chem Phys Letters 300:131-139.

Hansmann UHE (1997) Simulated annealing with Tsallis weights - A numerical comparison. Physica A 242:250-257.

Kirkpatrick S, Gelatt J and Vecchi MP (1983) Optimization by simulated annealing. Science 220:671-680.

Kvasnicka V and Pospíchal J (1997) A hybrid of simplex method and simulated annealing. Chemometrics Intell Lab Sys 39:161-173.

Lavor CC (2001) Uma abordagem determinística para minimização global da energia potencial de moléculas. PhD thesis, COPPE/UFRJ, Rio de Janeiro.

Luty BA, Wasserman ZR, Stouten PFW, Hodge CN, Zacharias M and McCammon JA (1995) A molecular mechanics/grid method for evaluation of ligand-receptor interactions. J Comput Chem 16:454-464.

Meng EC, Shoichet BK and Kuntz LD (1992) Automated docking with grid-based energy evaluation. J Comput Chem 13:505-524.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH and Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087-1092.

Moret MA (1996) Modelagem molecular clássica usando o método estocástico: Simulação por termalização - GSA. Master thesis, UFBA, Salvador.

Moret MA, Bisch PM, Mundim KC and Pascutti PG (1998) New stochastic strategy to analyze helix folding. Biophys J 82:1123-1132.

Moret MA, Pascutti PG, Bisch PM and Mundim KC (1998) Stochastic molecular optimization using generalized simulated annealing. J Comput Chem 19:647-657.

Moret MA, Pascutti PG, Mundim KC, Bisch PM and Nogueira Jr E (2001) Multifractality, Levinthal paradox, and energy hypersurface. Phys Rev E 63:R1-4.

Mundim KC, Liubich V, Dorfman S, Felsteiner J, Fuks D and Borstel G (2001) Nonempirical simulations of boron interstitials in tungsten. Physica B 301:239-254.

Mundim KC and Tsallis C (1996) Geometry optimization and conformational analysis through generalized simulated annealing. Int J Quantum Chem 58:373-381.

Schwefel H-P (1981) Numerical optimization of computer models. Chichester, Wiley & Sons, 398 p.

Solomatine PD (1995) Application of global optimization to models calibration. http://www.xs4all.nl/~dpsol/data-machine/papers/p_glob1/onglob06.htm (06/05/2003).

Szu H and Hartley R (1987) Fast simulated annealing. Phys Lett A 122:157-162.

Tsallis C (1988) Possible generalization of Boltzmann-Gibbs statistics. J Stat Phys 52:4798.

Tsallis C and Stariolo DA (1995) Optimization by simulated annealing: Recent progress. In: Stauffer D (ed) Annual Reviews of Computational Physics v. II, World Scientific, Singapore, pp 353-356.

Tsallis C and Stariolo DA (1996) Generalized simulated annealing. Physica A 233:395-406.

Xiang Y, Sun DY, Fan W and Gong XG (1997) Generalized simulated annealing algorithm and its application to the Thomson model. Phys Lett A 233:216-220.

Zhaoxian Y and Dang M (2003) Generalized simulated annealing algorithm applied in the ellipsometric inversion problem, Thin Solid Films 425:108-112.