Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 227 (2007) 1500–1509
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Adaptive extremal optimization by detrended fluctuation analysis

K. Hamacher *

Bioinformatics and Theoretical Biology Group, Technical University Darmstadt, Schnittspahnstr. 10, 64287 Darmstadt, Germany

Received 20 March 2007; received in revised form 17 August 2007; accepted 19 September 2007
Available online 5 November 2007

Abstract

Global optimization is one of the key challenges in computational physics as several problems, e.g. protein structure prediction, the low-energy landscape of atomic clusters, detection of community structures in networks, or model-parameter fitting can be formulated as global optimization problems. Extremal optimization (EO) has become in recent years one particular, successful approach to the global optimization problem. As with almost all other global optimization approaches, EO is driven by an internal dynamics that depends crucially on one or more parameters. Recently, the existence of an optimal scheme for this internal parameter of EO was proven, so as to maximize the performance of the algorithm. However, this proof was not constructive, that is, one cannot use it to deduce the optimal parameter itself *a priori*. In this study we analyze the dynamics of EO for a test problem (spin glasses). Based on the results we propose an online measure of the performance of EO and a way to use this insight to reformulate the EO algorithm in order to construct optimal values of the internal parameter online without any input by the user. This approach will ultimately allow us to make EO parameter free and thus its application in general global optimization problems much more efficient.
© 2007 Elsevier Inc. All rights reserved.

PACS: 02.60.Pn; 02.70.−c; 02.70.Uu

Keywords: Global optimization; Potential energy surface; Stochastic processes; Detrended fluctuation analysis; Spin glasses; Monte-Carlo; Parameter free algorithms

1. Introduction

Determining global minima of an objective function in a high-dimensional search space is a challenge for many scientists, e.g. in research concerned with spin glasses [1], in protein structure prediction [2–7,9], in the effective parameterization of biomolecular simulations [10,11], or in ligand docking [12,13]. In a physical context these objective functions are referred to as potential energy surfaces. Finding such a global extremum for a given system is the global optimization problem [14,15].

* Tel.: +49 6151 16 5318.
E-mail address: hamacher@bio.tu-darmstadt.de
URL: http://www.kay-hamacher.de

0021-9991/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2007.09.013

There are numerous ways to determine global minima *probabilistically* [16] which perform very well – the most prominent being Simulated Annealing [17,18]. Other, more powerful approaches were suggested afterwards, such as genetic algorithms [8], basin hopping [9], or parallel tempering [19]. A current topic of debate is whether a deterministic, reliable and efficient approach to the global optimization problem is feasible or whether global optimization is NP-complete also for physical system [20–23]. Whatever the outcome, experience tells us that global optimization is a computationally very challenging enterprise and therefore there is still a pressing need to develop better/efficient methodologies. Here we describe a new, parameter free approach that constitutes a self-contained, adaptive methodology alleviating several problems of established algorithms.

2. Extremal optimization

Since its advent, extremal optimization (EO) has turned out to be a very powerful technique for solving global optimization problems. Starting with the first study by Boettcher et al. [24] the method was later applied to problems in artificial intelligence [25], continuous problems such as Lennard–Jones atomic clusters [26], spin glasses [1,27], and detection of community structures in networks [28].

The basic idea behind EO is derived from the Bak–Sneppen model of evolution [29]. The landmark insight of this theory is the importance of species interaction: the fitness of one species is not solely determined by the environment (which would be the static picture put forward by the genetic algorithm community) but also by the other species in the ecology, that is, its competitors. This competition is then responsible for chain reactions, e.g. after a particular species crosses a certain fitness threshold. In this model of *macroevolution* the resulting dynamics displays self-organized criticality, which is the outcome of co-evolution, and sudden bursts or avalanches of evolutionary activity. The volatility caused by self-organized criticality makes the whole 'configurational space' available to the evolutionary dynamics – a trait we also would like to have in a global optimization framework. EO was developed in analogy to this paradigm: the various degrees of freedom of a global optimization problem 'compete' against each other in evolving better solutions, while at the same time selection against *bad* solutions should also occur.

In more detail: EO is, in general, defined for an objective function $E(\vec{x})$, where $\vec{x}$ is an element of some set $\mathcal{X}^n$ and $\mathcal{X}^n$ is the set of all possible states (e.g. a protein configuration in dihedral space or a spin configuration). On $\mathcal{X}^n$ there exists a neighborhood relation $N(\vec{x}) \subseteq \mathcal{X}^n$ which denotes the set of configurations that can be reached in one 'step'. A state $\vec{x}$ is given by a particular choice of entries $\vec{x} = (x_1, x_2, \ldots, x_n)$. For each of the n degrees of freedom of a global optimization problem, EO assumes a fitness $\lambda_i(x_i|\vec{x})$ that resembles the contribution of the degree of freedom i to the overall function value $E(\vec{x})$ while keeping the remaining $n-1$ components of $\vec{x}$ fixed.

EO now proceeds as follows: for a configuration $\vec{x}$ we compute all the λ_i-values. These values give rise to a ranking k_i such that $\forall \text{pairs}(i,j) k_i \leqslant k_j$ iff $\lambda_j \leqslant \lambda_i$. We now provide a probability distribution p_{k_i} from which a rank k_i is drawn. The corresponding degree of freedom i is then changed in $\vec{x}$ to a new value (in Ising spin systems this corresponds to a spin flip, in systems with higher spins we just change x_i to one of the other spin states with equal probability, see [30]). The new configuration is now taken to be our current $\vec{x}$ and we iterate until some given maximal iteration number is reached. Using the existence of such fitness enables us here to select against *bad* contributions by a particular i (if we assume $p_k \leqslant p_l$ for all $\lambda_k \geqslant \lambda_l$). At the same time we have also implemented a potential co-evolution, as changing one i immediately influences the fitness values of the other degrees of freedom, as in the Kim–Sneppen model.

The original formulation of EO chooses a rank k according to a probability $p_k := \mathcal{N}^{-1} \cdot k^{-\tau}$, with some scaling exponent τ and $\mathcal{N}$ an appropriate normalization constant. We refer to this variant as the τ-EO algorithm in subsequent parts of this paper.

Middleton [31] introduced a memory kernel in the dynamics by τ-ranking the $\tilde{\lambda}_i := \lambda_i + \Gamma \cdot h_i$ instead of the original fitnesses λ_i, where h_i is the number of times the degree of freedom i was chosen. This leads to a weak self-avoidance in the space of degrees of freedom and always pushes the algorithm into unexplored regions. However, this happens at the cost of the introduction of another internal parameter Γ that crucially influences the performance of this algorithm.

A first study [32] on the performance of τ-EO under changing internal parameters revealed two distinct phases of greedy and random exploration. The optimal behavior, in the sense of efficiency, was found to lie on the boundary, that is, on a critical point of the sole τ-EO parameter τ.

To study the performance of EO Heilmann, Hoffmann, and Solamon [30] applied linear programming arguments to investigate potential increases in the *average* minimal energy obtained in several runs. They were able to give a rigorous proof on the existence of a *time-dependent* distribution of ranks to be chosen. According to their study one should employ in each step a rectangular distribution of ranks up to some maximal ranking number determined implicitly by a given fitness threshold λ_{TA}. This threshold is in general time-dependent. We note that there always exists an isomorphic, non-trivial, time-dependent mapping between fitness thresholds λ_{TA} and ranking thresholds k_{TA}. We will refer to this procedure as threshold-acceptance EO.

3. The test problem

We focus here on test cases, defined on a high-dimensional domain, as only those problems allow for a judgment of the performance of global optimization algorithms [33]. A particular, striking example of such a problem is a spin glass [34] with a complex potential energy surfaces that shows a broad distribution of barrier heights. The energy of a configuration $\vec{s}$ of Ising spins $s_i \in [-1;+1]$ in the absence of an external field is $E(\vec{s}) = \sum_{\langle i,j\rangle} J_{ij} s_i s_j$, where the summation $\langle i,j\rangle$ runs over nearest neighbors. We draw the couplings J_{ij} from a Gaussian distribution. The dynamics of e.g. a Monte-Carlo process on this potential energy surfaces displays critical slowing down [34]. This is a ramification of the fact that a spin glass is a very difficult optimization problem and therefore a good test case for this study.

We obtained exact ground states and their respective energies E_i^{o} from the Spin Glass-Server [35] of the Jünger group for comparison with our results. We constructed 50 independent two-dimensional spin glass replica of size 20×20 spins. On each replica we performed three independent optimization runs, thus dealing with statistically relevant numbers (150 independent tests). The proposed energies during a sampling give rise to a time series of energies E_i. Note that time here refers to the steps in the algorithm and not to the dynamics of the spin system in reality. The energies in this time series (see below) E_i are expressed as relative differences with the real ground state energy E_i^{o} of the respective replica i as $\epsilon_i := \left|\frac{E_i - E_i^{\mathrm{o}}}{E_i^{\mathrm{o}}}\right|$.

In a second step, we applied the overall framework to spin systems with sizes $N \times N$ spins with $N \in [25; 30; 35; 40; 45; 50; 60]$. This allowed us to examine any size/complexity dependence of our results.

For an Ising spin system the definition of the fitness λ_i for a particular degree of freedom (single spin state) is straightforward: $\lambda_i := \sum_j' J_{ij} s_i s_j$, where the restricted sum $\sum'$ only takes nearest neighbors into account.

4. A measure of performance – time series analysis

The scaling of fluctuations in a time series with respect to time can be described by a single exponent γ. It is straightforward to show a relation between such fluctuations and the correlation associated with such a time series. In general it holds for the squared fluctuations F^2 and the correlation C that $C(n_{\mathrm{iter}}) \sim n_{\mathrm{iter}}^{-\gamma}$ and $F^2(n_{\mathrm{iter}}) \sim n_{\mathrm{iter}}^{2-\gamma}$. Trends can, in particular, introduce a systematic bias in γ and one usually wants to eliminate such effects. The detrended fluctuation analysis (DFA) [36–38] procedure allows for the determination of exponents γ by omitting trends of polynomial order and recording the averaged squared fluctuation in the values encountered.

Here one decomposes the entire recorded history of e.g. energies E_n into N/l non-overlapping regions for a time scale l and performs polynomial fitting of $p^{(i)}(n)$ of degree N to the data points. Detrending the time series in each region i is done by setting $E_{\mathrm{DFA}}^{(i)}(n) := E_n - p^{(i)}(n)$. One then computes the variances of this reduced time series. The exponent γ is finally determined from the scaling of the averaged variance with respect to the segment size l by $F^2(l) \sim l^{2-\gamma}$.

Note that $\gamma \sim 1$ indicates a random-walk like behavior in the time series. If a computational ansatz for finding minima shows this behavior in the suggested function values, the performance is obviously inferior and unsatisfactory as we eventually observe 'random guessing'. The rationale to use DFA in the analysis of global optimization performance stems from the two potential phases of any stochastic global optimization run:

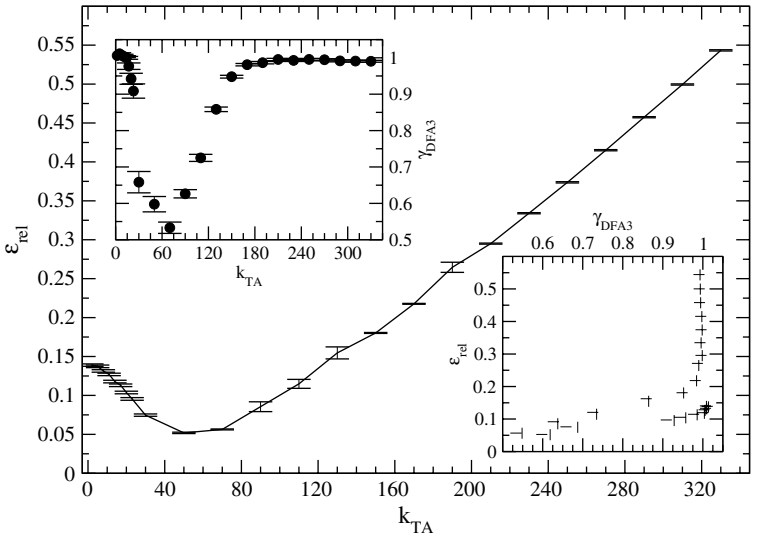

Fig. 1. The averaged relative error ϵ_{rel} for the threshold-acceptance version of EO with respect to the threshold k_{TA}. Error bars indicate the standard deviation in the results taken over the test set. The left inset shows the DFA exponent γ in third order DFA. Note the location of the minimum in both data sets. The right inset show the DFA exponent versus the error. Obviously there is a strong connection between small γ and small ϵ_{rel}.

(a) super-diffusive search (with trend, but 'guided') and (b) inefficient performance in a random-walk like fashion ('pure guessing'). To distinguish between these regimes one needs to obtain γ.

DFA has been successfully used in the study of other global optimization algorithms. For example, we showed that another stochastic global optimization approach, called Energy Landscape Paving, is optimal under such an analysis [39]. DFA was also used [40] to analyze the performance of the Stochastic Tunneling global optimization scheme [41,42] and helped to increase its performance. One further application is the investigation of the scaling of fluctuations with respect to system size in the original Bak–Sneppen model [43].

In Fig. 1, we show the performance of threshold-acceptance EO with respect to a constant k_{TA}. As noted already above there exists a one-to-one-mapping to threshold-acceptance in fitness values, so keeping k_{TA} fixed is one of the potential 'annealing' schedules for the distributions p_k^t. The analysis with respect to k_{TA} is a *generic* way to introduce a threshold while a fitness threshold is *not generic* – it is instead problem dependent.

We indeed observe in Fig. 1 a best choice of k_{TA} of around 50 which decreases the expected error in the suggested global minimum ϵ_{rel}. The existence of such a choice was proven in [30].

5. Internal construction of optimal parameter

We can now use our finding on the fluctuations of energies to tune the internal parameter k_{TA} and make it time-dependent. As the algorithm improves the choice of k_{TA}^t on its own, this EO formulation becomes *parameter free* and is thus much more preferable over any other known EO implementation, where one needs to change the parameter in an *a priori* unknown manner to find the optimal regime.

In order to achieve this *parameter free* behavior, we define the adaptive extremal optimization (Adaptive EO from now on) algorithm: (a) perform for given k_{TA} a threshold-acceptance EO for n steps, (b) determine the DFA exponent γ for the time series of energies encountered, (c) fit a polynomial of degree two to the γ of the last three threshold-acceptance EO runs, (d) determine the minimum of this polynomial and extract a new k_{TA} that minimizes the approximated $\gamma(k_{\text{TA}})$-curve and (e) iterate until satisfied. The first three γ-values are obtained in an initialization phase with equally spaced k_{TA}. We note in passing that we also kept track of the generated k_{TA}. To avoid cycles in k_{TA}-space we set k_{TA} to a uniformly chosen value whenever a suggested k_{TA} had been visited before. A more detailed version of the procedure is shown in Appendix A.

The rationale behind a second order polynomial in Adaptive EO is the following: Fig. 1 shows a monotonic tendency in the curve $\gamma(k_{\mathrm{TA}})$, thus we would also be doing fine with a steepest decent approach in k_{TA}. Nevertheless it is most advantageous to increase the speed of convergence. This can obviously be achieved by taking the (local) curvature of $\gamma(k_{\mathrm{TA}})$ into account. Higher order terms would increase the rate of convergence even more, but are harder to come by because of a longer initialization phase – and a longer initialization phase might waste CPU time.

6. Results

Here we show the increased performance of such an adaptive choice over a rectangular rank distribution with a fixed threshold. To this end Fig. 2 shows a comparison between the estimated errors ϵ_{rel} for the fixed threshold $k_{\mathrm{TA}} = 50$, as derived from Fig. 1 and the error made by the adaptive version Adaptive EO. What is astonishing is that at first glance Adaptive EO performs on average better than threshold-acceptance EO with the optimal choice of $k_{\mathrm{TA}} = 50$. One would naively expect that Adaptive EO can at best match the performance of threshold-acceptance EO. However, this is incorrect due to the various replica presented to the algorithm: for each realization of a spin glass the optimal choice of k_{TA} is in general different. Therefore the choice $k_{\mathrm{TA}} = 50$ increases the performance of threshold-acceptance EO *on the average*. Adaptive EO is, on the other hand, able to exploit the differences between the various replicas and to thus adapt to the actual energy landscape by choosing a different trajectory in k_{TA}-space for each replica. A detailed analysis of the dynamics in k-space must be postponed to a forthcoming publication as we want to focus on method development aspects in this publication.

Keep in mind that for the Adaptive EO results we just performed 150 computations as the algorithm adapts itself. To achieve such good performance with threshold-acceptance EO (open symbols in Fig. 2) we tested the whole range of k_{TA} as show in Fig. 1, thus eventually performing $24 \times 150 = 3600$ simulation runs. We, therefore, not only gained more resolution in proposed energies but also reduced the CPU demand tremendously.

As mentioned in Section 3, we repeated the whole study with larger systems to investigate size dependency. Fig. 3 shows that the superior performance of Adaptive EO which appears to be independent of the system

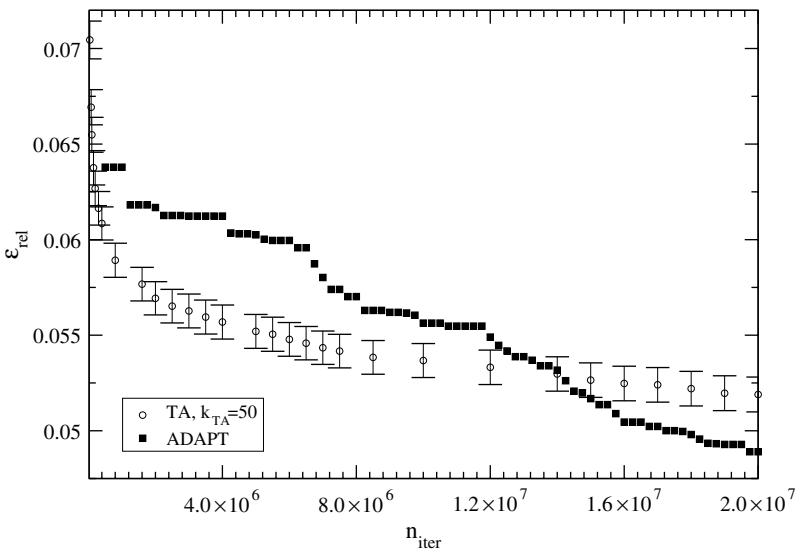

Fig. 2. The averaged relative error ϵ_{rel} for threshold-acceptance EO with a fixed threshold $k_{\mathrm{TA}} = 50$ (open circles) and the Adaptive EO variant (filled boxes). Error bars indicate again the standard deviation in the results taken over the test set. The error bars for Adaptive EO are of the same size, thus separating the mean values at $n_{\mathrm{iter}} = 2 \times 10^7$ by roughly two standard deviations, which corresponds to a significance of 75%.

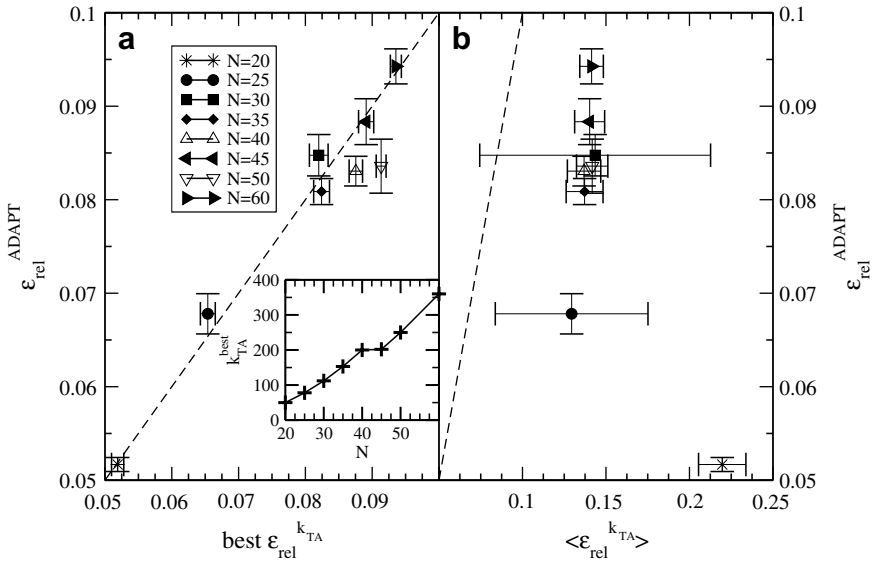

Fig. 3. Comparison of the relative error in optima at the end (iteration number $n_{\text{iter}} = 2 \times 10^7$) proposed by the threshold-acceptance EO and the Adaptive EO algorithm for various system sizes $N \times N$: (a) between Adaptive EO and the threshold-acceptance EO with the best k_{TA} found for a particular system size and (b) between Adaptive EO and the average of all threshold-acceptance EO runs for all the selected k_{TA}. Note that for threshold-acceptance EO in part (a) we had to test various parameters k_{TA}, thus spending orders of magnitudes more CPU time than for Adaptive EO, which adopts its k_{TA} in a time-dependent fashion. Clearly for almost all systems Adaptive EO performed as well or even better than threshold-acceptance EO, while the average threshold-acceptance EO (right) did much worse in comparison to Adaptive EO. Note the different scales on the x-axis. The dashed lines indicate the identity, thus dividing areas of superior performance of one of the algorithms. Error bars indicate the variance over the samples. The inset on the left side shows the scaling of the best k_{TA} with the system size.

size. Adaptive EO is capable of the same performance as the *best* threshold-acceptance EO runs. However, in real applications it is prohibitively expensive to scan for $k_{\text{TA}}^{\text{best}}$. Therefore, one would expect to end up with an average performance for some sub-optimally performing k_{TA} (right hand side of Fig. 3).

7. Comparison to simulated annealing

To analyze the dynamics of Adaptive EO further and show the qualitative difference to other well-known global optimization algorithms simulated annealing [17] optimization runs on the spin glass replica of size 20×20 were performed. We searched for optimal parameters and an efficient cooling schedule (exponential versus linear etc.) and identified a good cooling regime.

In comparison to Adaptive EO (Fig. 4) the performance of simulated annealing, measured in terms of the relative error ϵ_{rel}, was inferior. We found $\epsilon_{\text{rel}} = 11.4\%$ for simulated annealing and $\epsilon_{\text{rel}} = 5.1\%$ for Adaptive EO at maximal number of iterations.

It is interesting to note the difference in the dynamics between EO and Adaptive EO on the one hand and simulated annealing on the other hand: the times when simulated annealing finds the best minimum in a particular run exhibits a broad distribution. What is more striking is the fact that at larger iteration numbers the 'hits' vanish altogether – indicating that one cannot expect simulated annealing to perform better by allowing for more computations since the procedure has saturated and better minima are only created by chance. The situation is completely different for EO and Adaptive EO. The best minima ever encountered in a single run occur mostly in the final phase of the respective run (close to the maximal number of iterations 2×10^7). Thus increasing the maximal number of iterations is likely to create even better minima.

The data in Fig. 4 allows for the following interpretation: simulated annealing is 'frozen'. It cannot escape local minima anymore at some intermediate iteration number, therefore increasing the overall computational

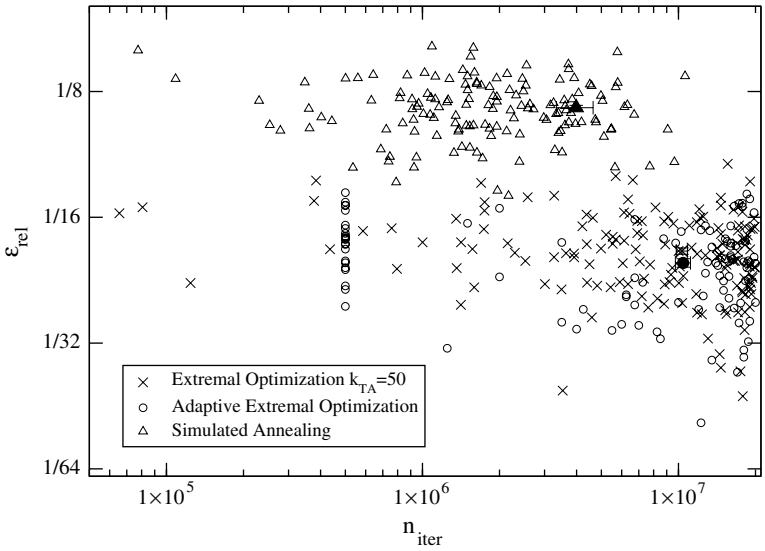

Fig. 4. Comparison of the relative error ϵ_{rel} in the best, visited (local) optima and the respective first-passage-time measured in number of iterations n_{iter}. The symbols with error bars show the mean and the variance of the respective data set. Note the logarithmic scales.

effort will only gain so much. EO on the other hand is not doomed to freeze instead it is very efficiently creating better and better minima throughout the iterations.

The 'clustering' of Adaptive EO results at $n_{iter} = 5 \times 10^5$ in Fig. 4 is a direct result from our choice of n in the algorithm shown in Appendix A and the fact that our implementation reports results only every n iterations (to avoid too much I/O-load).

8. Conclusions

In this study, we have implemented a threshold-acceptance procedure for the extremal optimization algorithm in accordance with analytical results of Heilmann et al. [30]. We have then motivated the usage of standard protocols of time series analysis in order to understand the dynamics of the optimization run. The results show that for a certain choice of the threshold parameter optimal performance is achieved. Arguments from random-walk theory enabled us to assign a 'guided' dynamics to this parameter regime, while insufficient performance was observed for random-walk like behavior. This can be attributed to random guessing.

Using this insight we formulated a *parameter free* extremal optimization procedure that adjusts its internal parameter to avoid random-walk like behavior. We were able to show that in 1/24 of the CPU time we found with 75% significance *better* results. Thereby one can eventually boost the speed and increase the accuracy of the method at the same time. It was also shown that this behavior has no size-dependency, thus rendering the adaptive extremal optimization procedure proposed in this study a versatile, generic and parameter free approach to global optimization.

The results were obtained for a particular difficult optimization problem (spin glass) suggesting that the approach is transferable to 'easier' problems without further thought. However, one should keep in mind that ''There's no free lunch'' as the saying goes. Wolpert [44] and Sharpe [45] have discussed in length how this applies also to the global optimization problem. In particular Wolpert has shown rigiously that what an algorithm gains in performance on one problem it will loose for another. Therefore, it is *always* advisable to check the performance of several algorithms for every new optimization class at hand. There would not be the 'one-and-only' optimization algorithm. Nevertheless, this paper alleviates the burden of such efforts by two main contributions: first, it demonstrates the applicability of the proposed analysis technique (time series analysis of visited minima and DFA) to all algorithms and all optimization problem classes. Second, it makes EO effectively *parameter free* which reduces the CPU time significantly.

Acknowledgments

K.H. is grateful to Dáibhid Ó Maoiléidigh, Thilo Gross, and Abigail Klopper for helpful comments on the manuscript and to the Fonds der chemischen Industrie which supported this study in part through a Liebig-Fellowship awarded to K.H. K.H. also gratefully acknowledges the hospitality of J.A. McCammon's group at UCSD where initial parts of this work were done. K.H. thanks F. Liers (Cologne) for discussions on the technical details of the spin glass server. The author thanks unknown referees for their helpful comments and suggestions which led to an even more clear and detailed picture of the usefulness and superior performance of the proposed approach.

Appendix A. Pseudo-code for the adaptive extremal optimization method

In the following, we show in pseudo-code the adaptive extremal optimization procedure for N spins. Each spin i contributes an energy $\lambda_i := \sum'_j J_{ij} s_i s_j$ to the overall energy E.

In the pseudo-code *rand(x)* returns an uniformly distributed, integer number from 1 to x. n is some rule-of-thumb value that only influences how often the DFA-exponent γ is computed. It should be large enough to allow for sufficient sampling. We found no difference in the results for $n \approx 10^4$–5×10^5. Larger n will even improve the sampling more.

Algorithm 1 Adaptive extremal optimization

```
k_TA := rand(N)
computed := 0
for iter := 1 to maxiter
  // Threshold Extremal Optimization with threshold k_TA
  sort λ_i in descending order
  pick a random spin within the threshold j := rand(k_TA)
  flip that spin j in all λ_i's
  update energy E
  insert E into EnergyHistory
  if(E < E_best) then
    print E and iter
    E_best := E
  endif
  // DFA computation and adjustment of k_TA to find minimal γ
  if(iter mod n == 0) then
      computed := computed + 1
      γ := compute_DFA_exponent(EnergyHistory)
      insert γ into (k_TA, γ)-set
      clean EnergyHistory
      if(computed ⩾ 3) then // enough data
        fit f(k) = a·k^2 + b·k + d to (k_TA, γ)-set
        solve f(k_TA) = 0 for k_TA
        delete oldest entry in (k_TA, γ)-set
        if (k_TA) ∈ {k_visited} then
           k_TA := rand(N) // do not return to a previously visited value
        endif
        else // not enough data yet
         k_TA := rand(N)
       endif
      insert k_TA in {k_visited}
  endif
endfor
```

References

[1] S. Boettcher, Extremal optimization for sherrington–kirkpatrick spin glasses, Eur. Phys. J. B 46 (2005) 501–505.
[2] P. Derreumaux, A diffusion process-controlled Monte-Carlo method for finding the global energy minimum of a polypeptide chain. I. Formulation test on a hexadecapeptide, J. Chem. Phys. 106 (12) (1997) 5260–5270.
[3] H. Arkin, Searching low-energy conformations of two elastin sequences, Eur. Phys. J. B 37 (2004) 223–228.
[4] Silvia Crivelli, Richard Byrd, Elizabeth Eskow, Robert Schnabe, Richard Yu, Thomas M. Philip, Teresa Head-Gordon, A global optimization strategy for predicting α-helical protein tertiary structure, Comput. Chem. 24 (3–4) (2000) 489–497.
[5] U. Hansmann, L.T. Wille, Global optimization by energy landscape paving, Phys. Rev. Lett. 88 (23) (2002) 068105.
[6] A. Mitsutake, Y. Okamoto, Replica-exchange extensions of simulated tempering method, J. Chem. Phys. 121 (6) (2004) 2491–2504.
[7] C. Chou, R. Hand, S. Li, T. Lee, Guided simulated annealing method for optimization problems, Phys. Rev. E 67 (2003) 066704.
[8] D.E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, 1989.
[9] D.J. Wales, H.A. Scheraga, Global optimization of clusters, crystals and biomolecules, Science 285 (5432) (1999) 1368–1372.
[10] K. Hamacher, J.A. McCammon, Computing the amino acid specificity of fluctuations in biomolecular systems, J. Chem. Theory Comput. 2 (3) (2006) 873–878.
[11] K. Hamacher, A. Hübsch, J.A. McCammon, A minimal model for stabilization of biomolecules by hydrocarbon cross-linking, J. Chem. Phys. 124 (16) (2006) 164907.
[12] H. Merlitz, W. Wenzel, Comparison of stochastic optimization methods for receptor-ligand docking, Chem. Phys. Lett. 362 (2002) 271–277.
[13] A. Perryman, J. McCammon, Autodocking dinucleotides to the HIV-1 integrase core domain: exploring possible binding sites for viral and genomic DNA, J. Med. Chem. 45 (2002) 5624–5627.
[14] R. Horst, P.M. Pardalos, N.V. Thoai, Introduction to global optimization, Nonconvex Optimization and Its Applications, second ed., vol. 48, Kluwer Academic Publishers, 2000.
[15] R. Horst, P.M. Pardalos, Handbook of Global Optimization, Kluwer Academic Publishers, 1995.
[16] P.M. Pardalos, E. Romeijn, Handbook of Global Optimization – Heuristic Approaches, vol. 2, Kluwer Academic Publishers, 2002.
[17] S. Kirkpatrick, C. Gelatt, M. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.
[18] A. Franz, K.H. Hoffmann, Optimal annealing schedules for a modified tsallis statistics, J. Comp. Phys. 176 (2002) 196–204.
[19] U. Hansmann, Parallel tempering algorithm for conformational studies of biological molecules, Chem. Phys. Lett. 281 (1997) 140–150.
[20] A.B. Adib, Np-hardness of the cluster minimization problem revisited, J. Phys. A: Math. Gen. 38 (40) (2005) 8487–8492.
[21] J. Ngo, J. Marks, Computational complexity of a problem in molecular structure prediction, Protein Eng. 5 (4) (1992) 313–321.
[22] L. Wille, J. Vennik, Computational complexity of the ground-state determination of atomic clusters, J. Phys. A: Math. Gen. 18 (1985) L419–L422.
[23] N.D. Jones, Computability Theory – An Introduction, ACM Monograph Series, Academic Press, New York, 1973.
[24] S. Boettcher, A.G. Percus, Optimization with extremal dynamics, Phys. Rev. Lett. 86 (23) (2001) 5211–5214.
[25] S. Boettcher, A. Percus, Nature's way of optimizing, Artif. Intell. 119 (2000) 275–286.
[26] T. Zhou, W.-J. Bai, L.-J. Cheng, B.-H. Wang, Continuous extremal optimization for Lennard–Jones clusters, Phys. Rev. E 72 (1) (2005) 016702.
[27] S. Boettcher, P. Sibani, Comparing extremal and thermal explorations of energy landscapes, Eur. Phys. J. B 44 (2005) 317–326.
[28] J. Duch, A. Arenas, Community detection in complex networks using extremal optimization, Phys. Rev. E 72 (2) (2005) 027104.
[29] K. Sneppen, P. Bak, H. Flyvbjerg, M. Jensen, Evolution as a self-organized critical phenomenon, PNAS 92 (11) (1995) 5209–5213.
[30] F. Heilmann, K. Hoffmann, P. Salamon, Best possible probability distribution over extremal optimization ranks, Europhys. Lett. 66 (3) (2004) 305–310.
[31] A.A. Middleton, Improved extremal optimization for the Ising spin glass, Phys. Rev. E (Stat. Nonlinear Soft Matter Phys.) 69 (5) (2004) 055701.
[32] S. Boettcher, M. Frank, Optimizing at the ergodic edge, Physica A 367 (2006) 220–230.
[33] K. Hamacher, On stochastic global optimization of one-dimensional functions, Physica A 354 (2005) 547–557.
[34] K. Binder, A. Young, Spin glasses: experimental facts, theoretical concepts, and open questions, Rev. Mod. Phys. 58 (4) (1986) 801–976.
[35] C. Simone, M. Diehl, M. Jünger, P. Mutzel, G. Reinelt, Exact ground states of Ising spin glasses: bew experimental results with a branch-and-cut algorithm, J. Stat. Phys. 80 (1995) 487.
[36] A. Bunde, J. Kantelhardt, Langzeitkorrelationen in der natur: von klima, erbgut und herzrhythmus, Phys. Bl. 57 (5) (2001) 49–54.
[37] C.-K. Peng, S. Buldyrev, S. Havlin, M. Simons, H. Stanley, A. Goldberger, Mosaic organization of dna nucleotides, Phys. Rev. E 49 (1994) 1685.
[38] K. Hu, P.C. Ivanov, Z. Chen, P. Carpena, H. Eugene Stanley, Effect of trends on detrended fluctuation analysis, Phys. Rev. E 64 (1) (2001) 011114.
[39] K. Hamacher, Energy landscape paving as a perfect optimization approach under detrended fluctuation analysis, Physica A 378 (2) (2007) 307–314.
[40] K. Hamacher, Adaptation in stochastic tunneling global optimization of complex potential energy landscapes, Europhys. Lett. 74 (6) (2006) 944–950.
[41] K. Hamacher, W. Wenzel, The scaling behaviour of stochastic minimization algorithms in a perfect funnel landscape, Phys. Rev. E 59 (1) (1999) 938–941.

[42] W. Wenzel, K. Hamacher, A stochastic tunneling approach for global minimization, Phys. Rev. Lett. 82 (15) (1999) 3003–3007.
[43] K. Ma, C. Yang, X. Cai, Correlations in the Bak–Sneppen model from detrended fluctuation analysis, Physica A 357 (3–4) (2005) 455–465.
[44] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Trans. Evol. Comput. 1 (1) (1997) 67–82.
[45] O. Sharpe, Beyond NFL: a few tentative steps, in: J.R. Koza, W. Banzhaf, K. Chellapilla, K. Deb, M. Dorigo, D.B. Fogel, M.H. Garzon, D.E. Goldberg, H. Iba, R. Riolo (Eds.), Genetic Programming 1998: Proceedings of the Third Annual Conference, Morgan Kaufmann, University of Wisconsin-Madison, Wisconsin, USA, 1998, pp. 593–600.