Effective hybrid evolutionary computational algorithms for global optimization and applied to construct prion AGAAAAGA fibril models

Jiapu Zhang

Centre of Informatics and Applied Optimization &
Graduate School of Information Technology and Mathematical Sciences,
University of Ballarat, Mount Helen, VIC 3350, Australia
Email: j.zhang@ballarat.edu.au, Phone: 61 - 423 487 360

Abstract: Evolutionary algorithms are parallel computing algorithms and simulated annealing algorithm is a sequential computing algorithm. This paper inserts simulated annealing into evolutionary computations and successful developed a hybrid Self-Adaptive Evolutionary Strategy $\mu + \lambda$ method and a hybrid Self-Adaptive Classical Evolutionary Programming method. Numerical results on more than 40 benchmark test problems of global optimization show that the hybrid methods presented in this paper are very effective. Lennard-Jones potential energy minimization is another benchmark for testing new global optimization algorithms. It is studied through the amyloid fibril constructions by this paper.

To date, there is little molecular structural data available on the AGAAAAGA palindrome in the hydrophobic region (113-120) of prion proteins.This region belongs to the N-terminal unstructured region (1-123) of prion proteins, the structure of which has proved hard to determine using NMR spectroscopy or X-ray crystallography due to the insoluble and noncrystalline nature of the amyloid fibril. However, computer computational optimization can easily obtain a description of this peptide at a submicroscopic level. The hybrid methods presented in this paper will be applied to construct amyloid fibril molecular structures of the prion 113–120 region.

1 Introduction

Computer computational experiences have shown that simulated annealing (SA) and genetic algorithm (GA) are successful heuristic methods (Zhang 2004). SA is based upon the physical analogy of cooling crystal structures (including the case of quenching) that spontaneously arrives at a stable configuration, characterized by globally or locally minimal potential energy. It can accept bad solutions according to the Metropolis cri-

terion; X-ray crystallography finds the X-ray final structure of a protein, which usually need refinements using a SA protocol in order to produce a better structure. GA is inspired by the evolutionary process of species. Based on GA, evolutionary strategies (ESs) and evolutionary programming (EP) were developed in the 1990s (Zhang 2004). Numerical results in this paper show that the self-adaptive ES $\mu + \lambda$ (SAES($\mu + \lambda$)) method and the self-adaptive classical EP (SACEP) method can successfully get local minimums of many global optimization test problems. It is worth noticing that SA is a sequential computing algorithm and evolutionary algorithms are parallel computing algorithms. Applying SA (Bagirov and Zhang 2003) to all individuals in the population, in this paper we find that the hybrid with SA can furthermore improve the SAES($\mu + \lambda$) and SACEP methods and can get global minimal solutions of extensive global optimization testing problems, including the Lennard-Jones cluster (LJC) problem (Locatelli and Schoen 2008).

The LJC problem always serves as a benchmark for testing new optimization algorithms. LJC problem is to minimize the LJ potential energy of van der Waals (vdw) contacts of atoms. Constructions of the amyloid fibril molecular structures of prion 113–120 region is just a LJC problem. The constructions will be based on the most recently released experimental molecular structures of human M129 prion peptide 127–132 (PDB entry 3NHC released into Protein Data Bank (http://www.rcsb.org) on 04-AUG-2010). The atomic-resolution structure of this peptide is a steric zipper, with strong vdw interactions between β-sheets and hydrogen bonds to maintain the β-strands (Figure 1, where the purple dashed lines denote the hydrogen bonds). The two hybrid methods of this paper have succeeded in constructing the amyloid fibril molecular structures of prion (113–120) AGAAAAGA peptide.

This paper is organized as follows. In Section 2 the two hybrid SA-SAES($\mu + \lambda$) and SA-SACEP algorithms will be given, and numerous numerical results are shown and discussed. Three amyloid fibril models are successfully constructed applied by the two algorithms in Section 3. Section 4 concludes that the two SA-SAES($\mu + \lambda$) and SA-SACEP algorithms are very effective for solving global optimization problems, including the LJC global optimization problem of prion AGAAAAGA amyloid fibril model constructions.

2 The hybrid evolutionary computational algorithms

Algorithm 1 *SA-SAES($\mu + \lambda$).*

Step 0. Randomly generate μ parents, where each parent $z_k = (\vec{x}_k, \vec{\sigma}_k)$.

Step 1. Apply SA on each parent $\vec{x}_k$.

Step 2. Set $\tau = \left(\sqrt{\left(2\sqrt{(n)}\right)}\right)^{-1}$ and $\tau' = \left(\sqrt{(2n)}\right)^{-1}$.

Step 3. Until λ children are generated, do

Step 4. Select two parents $z_k = (\vec{x}_k, \vec{\sigma}_k)$ and $z_l = (\vec{x}_l, \vec{\sigma}_l)$ at random to generate child $\vec{y}_j = (\vec{x}_j, \vec{\sigma}_j)$.

Step 5. Discrete recombination: for each variable x_{ji} and step size σ_{ji} in $\vec{y}_j$, do ($x_{ji} = x_{ki}$ and $\sigma_{ji} = \sigma_{ki}$) or ($x_{ji} = x_{li}$ and $\sigma_{ji} = \sigma_{li}$)

Step 6. Mutation: For each x_{ji} and step size σ_{ji} in $\vec{y}_j$

$$x'_{ji} = x_{ji} + \sigma_{ji} N_j(0,1)$$

$$\sigma'_{ji} = \sigma_{ji} \exp(\tau' N(0,1) + \tau N_j(0,1))$$

Step 7. If the number of children is less than λ, go to Step 4.

Step 8. Select the best μ individuals among all the $\mu + \lambda$ parents and children.

Step 9. Apply SA on the best individual among the selected μ individuals.

Step 10. If the stopping criteria are satisfied, stop, else go to step 2.

Algorithm 2 *SA-SACEP.*

Step 0. Randomly generate μ parents and evaluate them, where each parent $z_k = (\vec{x}_k, \vec{\sigma}_k)$.

Step 1. Apply SA on each parent $\vec{x}_k$.

Step 2. Set $\tau = \left(\sqrt{\left(2\sqrt{(n)}\right)}\right)^{-1}$ and $\tau' = \left(\sqrt{(2n)}\right)^{-1}$.

Step 3. For each parent, generate a child as follows

$$x'_{ji} = x_{ji} + \sigma_{ji} N_j(0,1)$$

$$\sigma'_{ji} = \sigma_{ji} \exp(\tau' N(0,1) + \tau N_j(0,1))$$

Step 4. Evaluate all children

Step 5. Undertake a tournament y for each parent and child as follows: select ζ individuals with replacement from the joint set of parents and children. For each individual z of the ζ individuals, if y is better than z, add 1 to the fitness of y.

Step 6. Select the best μ individuals among all parents and children with the highest fitness.

Step 7. Apply SA on the best individual among the selected μ individuals.

Step 8. If the stopping criteria are satisfied, stop, else go to step 1.

Numerical results (in Table 1 and Table 2) show that the above two algorithms can find global minimums. However, SAES($\mu + \lambda$) and SACEP algorithms just can get local minimal solutions. Thus, SA-SAES($\mu + \lambda$) and SA-SACEP algorithms greatly improve the SAES($\mu + \lambda$) and SACEP algorithms. This shows the effectiveness of SA and the hybrid technique.

3 Prion AGAAAAGA amyloid fibril constructions

In Figure 1 we see that G (H) chains (i.e. β-sheet 2) of 3NHC.pdb can be obtained from A (B) chains (i.e. β-sheet 1) by

$$G(H) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} A(B) + \begin{pmatrix} 9.07500 \\ 4.77650 \\ 0.00000 \end{pmatrix}, \tag{1}$$

and other chains can be got by

$$I(J) = I_3 G(H) + \begin{pmatrix} 0 \\ 9.5530 \\ 0 \end{pmatrix}, K(L) = I_3 G(H) + \begin{pmatrix} 0 \\ -9.5530 \\ 0 \end{pmatrix}, \tag{2}$$

$$C(D) = I_3 A(B) + \begin{pmatrix} 0 \\ 9.5530 \\ 0 \end{pmatrix}, E(F) = I_3 A(B) + \begin{pmatrix} 0 \\ -9.5530 \\ 0 \end{pmatrix}, \tag{3}$$

where I_3 is the 3-by-3 identity matrix. Basing on the template 3NHC.pdb from the Protein Data Bank, three prion AGAAAAGA palindrome amyloid fibril models - an AAAAGA model (Model 1), a GAAAAG model (Model 2), and an AAAAGA model (Model 3) - will be successfully constructed in this paper. AB chains of Models 1-3 were respectively got from AB chains of 3NHC.pdb using the mutate module of the free package Swiss-PdbViewer (SPDBV Version 4.01) (http://spdbv.vital-it.ch). It is pleasant to see that almost all the hydrogen bonds are still kept after the mutations; thus we just need to consider the vdw contacts only. Making mutations for GH chains of 3NHC.pdb, we can get the GH chains of Models 1-3. However, the vdw contacts between A chain and G chain, between B chain and H chain are too far at this moment (Figure 2).

Seeing Figure 2, we may know that for Models 1-3 at least two vdw interactions between A.ALA3.CB-G.ALA4.CB, B.ALA4.CB-H.ALA3.CB should be maintained. Fixing the coordinates of A.ALA3.CB and B.ALA4.CB, letting the coordinates of G.ALA4.CB and H.ALA3.CB be variables, we may get a simple LJC problem just with six variables. Setting the coordinates of G.ALA4.CB and H.ALA3.CB as initial solutions, running the above two hybrid algorithms, for Models 1-3 we get

$$G(H) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} A(B) + \begin{pmatrix} -0.703968 \\ 7.43502 \\ -0.33248 \end{pmatrix}. \quad (4)$$

By (4) we can get close vdw contacts between A chain and G chain, between B chain and H chain (Figure 3).

Furthermore, we may employ the Amber 11 package (Case et al. 2010) to optimize Models 1-3 and at last get Models 1-3 with stable total potential energies (Figure 4). The other CDIJ and EFKL chains can be got by parallelizing ABGH chains in the use of mathematical formulas (2)-(3).

4 Conclusion

If a parallel computing algorithm hybridizes with a sequential computing algorithm, then the new hybrid algorithm performs much better than they work alone separately. In this paper, the parallel computing algorithms used are the SAES($\mu + \lambda$) and SA-CEP algorithms and the sequential computing algorithm used is the SA algorithm. We successfully tested the new hybrid algorithms by extensive more than 40 benchmark global optimization problems. The hybrid evolutionary computational algorithms were successfully applied to construct three prion AGAAAAGA amyloid fibril models.

Acknowledgments: This research was supported by a Victorian Life Sciences Computation Initiative (VLSCI) grant number VR0063 on its Peak Computing Facility at the University of Melbourne, an initiative of the Victorian Government.

References

Bagirov, A.M., and Zhang, J.P. 2003. Comparative analysis of the cutting angle and

simulated annealing methods in global optimization. *Optimization* 52, 363–378.

Case, D.A., T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B.P. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P.A. Kollman. 2010. *AMBER 11*, University of California, San Francisco.

Locatelli M., and Schoen F. 2008. Structure prediction and global optimization. *Optima Mathematical Programming Society Newsletter USA* 76, 1-8.

Neumaier A. 2004. Neumaier's Global Optimization website:
solon.cma.univie.ac.at/˜neum/glopt.html

Neumaier A. 2010. Neumaier's Global Optimization test problems:
http://www.mat.univie.ac.at/˜neum/glopt/janka/funcs.html

Zhang, J.P. 2004. Derivative-free hybrid methods in global optimization and their applications. PhD thesis, The University of Ballarat, Australia:
http://sites.google.com/site/jiapuzhang/

Table 1: The Optimal objective function values of SAES($\mu + \lambda$) Algorithm and SA-SAES($\mu + \lambda$) Algorithm, and SACEP Algorithm and SA-SACEP Algorithm

Function	# of variables	SAES($\mu + \lambda$)	SA-SAES($\mu + \lambda$)	SACEP	SA-SACEP
F1 (Neumaier 2004)	2	-186.731	-186.731	-186.731	-186.731
F2 (Neumaier 2004)	5	1.0	1.0	1.0	1.0
	10	1.0	1.0	1.0	1.0
	20	1.28551	1.0	24.5297	1.0
	30	1.02754	1.0	1.13336	1.0
	50	1.00388	1.00001	9.28671	1.00001
F3 (Ackleys)	2	0.0	0.0	0.0	0.0
	3	0.0	0.0	0.0	0.0
	5	0.0	2.41563e-05	0.0	2.41563e-05
	7	2.13384	4.86888e-05	1.72382	4.86888e-05
	10	3.90647	7.6222e-05	1.08046	8.82517e-05
	20	5.1886	0.000190629	2.24666	0.000224306
	30	5.47366	0.0003507	4.92406	0.000406911
F4 (Bohachevsky Nr.1)	2	0.11754	0.117535	0.117548	0.117535
F5 (Bohachevsky Nr.2)	2	0.0	0.0	0.0	0.0
F6 (Bohachevsky Nr.3)	2	0.0	0.0	0.0	0.0
F7 (Branin)	2	0.398891	0.397887	0.398055	0.397887
F8 (De Joung)	3	0.0	0.0	0.0	0.0
F9 (Easom)	2	-0.999725	-1.0	-0.98863	-1.0
F10 (Goldstein Price)	2	3.00006	3.0	3.00002	3.0
F11 (Hartman with $n = 3$)	3	-3.86271	-3.86278	-3.86277	-3.86278
F12 (Hartman with $n = 6$)	6	-1.84847	-3.32237	-3.32192	-3.32237
F13 (Hump)	2	8.86897e-05	4.65327e-08	0.000439177	0.0
F14 (Hyper-Ellipsoid)	30	1697.83	4.20078e-06	1.76103	0.0
F15 (Levy Nr.2)	5	0.0257144	1.02076e-10	0.0120519	0.0
	10	0.0129742	9.06744e-10	0.0317808	0.0
	20	2.34247e-06	5.48692e-09	0.0136671	0.0
	30	0.00193177	2.12137e-08	0.785024	0.0
	50	0.616365	6.12211e-08	2.07428	0.0
F16 (Levy Nr.3)	5	0.0218405	3.8796e-09	0.000743298	0.0
	10	0.00617594	1.35077e-08	0.000173664	0.0
	20	0.0	1.28154e-07	0.00358961	0.0
	30	0.000140932	4.54418e-07	0.000992482	0.0
	50	1.20497	1.68793e-06	1.32839e+06	1.60169e-06
F17 (Michalewicsz)	2	-1.95063	-1.8013	-1.95217	-1.8013
F18 (Neumaier Nr.2)	4	0.00245258	0.000487242	0.0766711	0.000174267
F19 (Neumaier Nr.3)	10	-21.0244	-209.998	-203.925	-209.999
F20 (Rastringins Nr.1)	2	0.0	2.36476e-10	0.0	0.0
	3	0.0	3.91857e-10	0.995047	0.0
	5	0.0	3.25394e-08	5.97189	0.0
	7	0.0	1.93565e-07	8.95636	0.0
	10	1.99124	1.98263e-06	32.8386	1.98263e-06

Table 2: The Optimal objective function values of SAES($\mu + \lambda$)Algorithm and SA-SAES($\mu + \lambda$) Algorithm, and SACEP Algorithm and SA-SACEP Algorithm (continuation)

Function	# of variables	SAES($\mu + \lambda$)	SA-SAES($\mu + \lambda$)	SACEP	SA-SACEP
F21 (Rosenbrock)	2	0.0079492	5.68257e-07	0.00856004	2.096e-06
	5	0.915901	0.000190216	0.00588099	3.13482e-05
	10	4.104	3.83856e-05	2.15272	0.000239605
F22 (Schaffer Nr.1)	2	0.0	0.0	0.0	0.0
F23 (Schaffer Nr.2)	2	0.0	0.195296	0.0	0.195296
F24 (Shekel-5)	4	-5.04985	-5.27766e+13	-5.05082	-5.27766e+13
F25 (Shekel-7)	4	-5.0606	-5.27766e+13	-5.05484	-5.27766e+13
F26 (Shekel-10)	4	-5.1273	-5.27766e+13	-5.11435	-5.27766e+13
F27 (Shubert Nr.1)	2	-186.731	-186.731	-186.731	-186.731
F28 (Shubert Nr.2)	2	-186.341	-186.731	-186.731	-186.731
F29 (Step)	5	-144.0	0.0	-2848	0.0
	10	-366	0.0	-1.18937e+07	0.0
	50	-13864	0.0	-7.19852e+34	0.0
F31 (Zimmermanns)	2	-103.806	-494.741	-494.748	-494.735
F32 (Neumaier, 2010)	2	0.0	4.19095e-06	0.0	4.19095e-06
	5	0.0	6.11739e-05	0.0	6.11739e-05
	10	0.0	0.000461433	0.00287121	0.000553783
	50	0.681216	6.26669e-13	15.1833	0.017286
F33 (Neumaier, 2010)	2	0.0	6.26669e-13	0.0	0.0
	5	0.0	4.16862e-06	0.0	4.16862e-06
	10	8.06556	0.0113471	0.0	0.0322543
	50	11206.3	1030.77	6733.64	894.608
F34 (Neumaier, 2010)	2	0.0	1.78348e-05	0.0	1.78348e-05
	5	0.0	0.000742218	0.0	0.000742218
	10	0.0	0.00371919	0.0	0.00371919
	50	79.0741	0.189135	14.1295	0.189135
F35 (Neumaier, 2010)	2	0.0	0.0	0.0	0.0
	5	0.0	0.0	0.0	0.0
	10	0.0	0.0	0.0	0.0
	50	52.0	0.0	33.0	0.0
F36 (Neumaier, 2010)	2	5.03179e-06	8.18303e-07	4.59426e-06	8.18303e-07
	5	2.03186e-05	1.28561e-05	0.000226618	6.52515e-06
	10	0.000277681	6.09379e-05	0.001168	6.73284e-05
	50	415.836	0.00306063	380.029	0.00440367
F37 (Neumaier, 2010)	2	-837.931	-837.966	-3947.21	-837.966
	5	-1796.66	-2094.91	-1513.87	-2094.91
	10	-3809.75	-4189.83	-3245.56	-4189.83
	50	-18813.3	-20949.1	-12809.6	-20949.1
F38 (Neumaier, 2010)	2	0.0179898	6.08096e-11	0.00202397	0.0
	5	0.0744221	9.45082e-09	0.0409532	0.0
	10	0.0019571	1.13757e-07	0.0114677	0.0
	50	7.73384	4.46473e-05	10.5956	4.02717e-05
F41 (Neumaier, 2010)	2	-4.12397	-4.12398	-4.12373	-4.12398
F42 (Neumaier, 2010)	2	0.398891	0.397887	0.398055	0.397887
F43 (Neumaier, 2010)	2	3.00006	3.0	3.00002	3.0

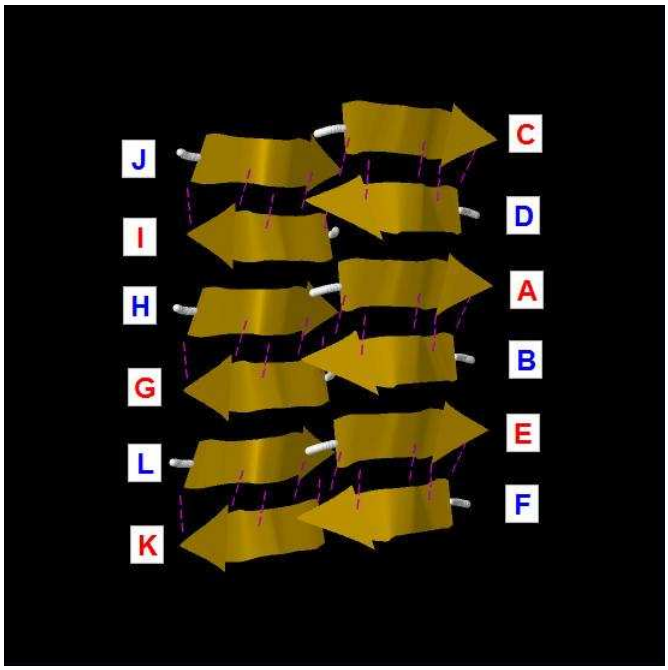

Figure 1: Protein fibril structure of human M129 prion GYMLGS (127–132).

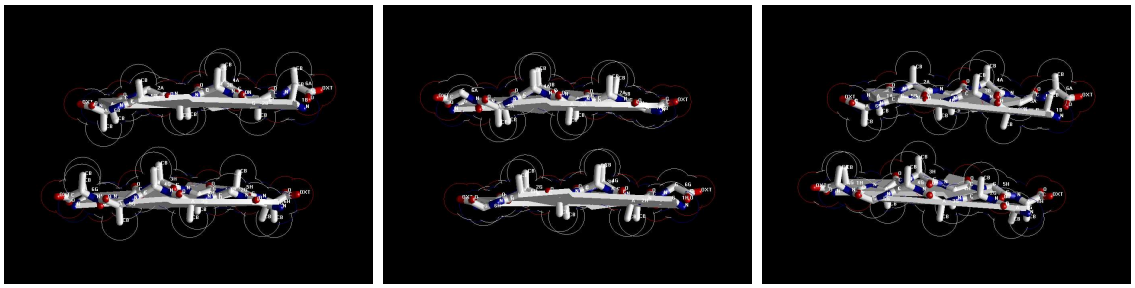

Figure 2: Far vdw contacts of AG chains and BH chains of Models 1-3.

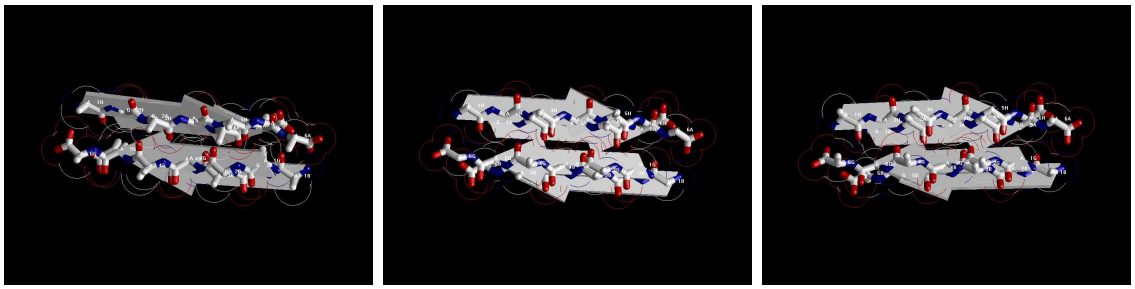

Figure 3: Close vdw contacts of AG chains and BH chains of Models 1-3.

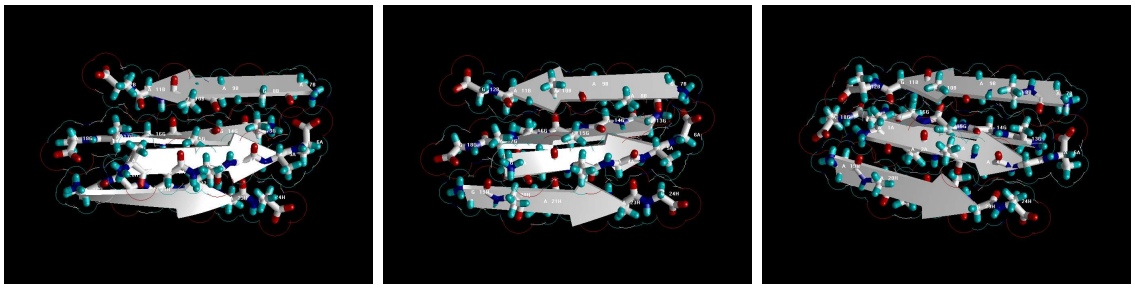

Figure 4: Optimal structures of prion AAAAGA amyloid fibril Models 1-3.