Systems biology

libSRES: a C library for stochastic ranking evolution strategy for parameter estimation

Xinglai Ji[1] and Ying Xu[2,*]

[1]Computational Biology Institute, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA and [2]Department of Biochemistry and Molecular Biology and Institute of Bioinformatics, University of Georgia, Athens, GA 30602-7229, USA

ABSTRACT

Summary: Estimation of kinetic parameters in a biochemical pathway or network represents a common problem in systems studies of biological processes. We have implemented a C library, named libSRES, to facilitate a fast implementation of computer software for study of non-linear biochemical pathways. This library implements a (μ, λ)-ES evolutionary optimization algorithm that uses stochastic ranking as the constraint handling technique. Considering the amount of computing time it might require to solve a parameter-estimation problem, an MPI version of libSRES is provided for parallel implementation, as well as a simple user interface. libSRES is freely available and could be used directly in any C program as a library function. We have extensively tested the performance of libSRES on various pathway parameter-estimation problems and found its performance to be satisfactory.

Availability: The source code (in C) is free for academic users at http://csbl.bmb.uga.edu/~jix/science/libSRES/

Contact: xyn@bmb.uga.edu

Supplementary information: Detailed documentation for libSRES is available at http://csbl.bmb.uga.edu/~jix/science/libSRES/

A basic challenge in systems biology is to understand the dynamical behaviors of biochemical pathways or networks. Many approaches have been developed to determine the network structure (Li *et al.*, 2005; Gardner *et al.*, 2003). However, the network topology alone is not sufficient to define its dynamics. The kinetic parameters are also needed to specify the network properties. In order to effectively determine the kinetic parameters of a biochemical network, one of the popular approaches is to find a most probable set of model parameters that would produce the behavior of the experimental system. Therefore, given a set of experimental data, the estimated parameters will reproduce the experimental results in the best possible way. The kinetic parameters can then be used to study the dynamic properties of the pathway, such as robustness and sensitivity. And more importantly, the parameters can be used to examine whether the inferred network topology or proposed pathway mechanism fits the experimental data and to set up subsequent working hypotheses.

This class of parameter-estimation problems can be stated as optimization problems with non-linear dynamic systems, many of which are, or can be transformed to, non-linear programming problems subject to differential algebraic constraints. Up to now, several methods and softwares have been developed to solve these problems, such as Monte Carlo method (Battogtokh *et al.*, 2002), Stochastic GO Methods (Moles *et al.*, 2003), GEPASI (Mendes and Kell, 1998), DYNAFIT (Kuzmič, 1996), etc. According to a few comparative studies (Banga *et al.*, 2003a, b; Moles *et al.*, 2003), with a non-linear biochemical dynamic model taken as a benchmark, the stochastic ranking evolution strategy (SRES) (Runarsson and Yao, 2000, 2005) was able to solve the benchmark problem successfully and showed better performance than other methods, which indicated that SRES might be the most competitive optimization method for estimation of parameters in a biochemical pathway. Here we contribute a C library 'libSRES' that implements SRES.

The core of libSRES is composed of a parameter estimator mainly implementing stochastic ranking evolution strategy (Runarsson and Yao, 2005). The SRES is a (μ, λ)-ES evolutionary optimization algorithm that uses stochastic ranking as the constraint handling technique. The stochastic ranking is based on the bubble-sort algorithm and is supported by the idea of dominance (Runarsson and Yao, 2000). It adjusts the balance between the objective and penalty functions automatically during the evolutionary search. (For parameter estimation in biological pathway, the objective function might be the difference between experimental and predicted data; the penalty functions represent the constraints on this pathway.) In the (μ, λ)-ES algorithm, the evaluated objective and the penalty functions for each individual are used to rank the individuals in a population, and the best (highest ranked) μ individuals out of λ are selected for the next generation. A total of 13 benchmark problems have been tested in previous works (Runarsson and Yao, 2000, 2005) and the results showed that it was capable of improving the performance significantly. The algorithm is also described in detail in our online documentation, as well as in the original papers.

Included with the libSRES are the 13 benchmark examples that can serve as templates for user programs. In addition, libSRES has been applied to estimate parameters in two biochemical pathway models (not included in the distribution): the three-step pathway model with 36 parameters described in the work by Moles *et al.*, 2003, and the model of irreversible inhibition of HIV proteinase, which has 20 parameters and has been described in previous works (Kuzmič 1996; Mendes and Kell, 1998). In these two applications, an ordinary differential equation (ODE) solver is required to solve the ODEs converted from biochemical reactions in the model. CVODE (Cohen and HindMarsh, 1994) has been used here,

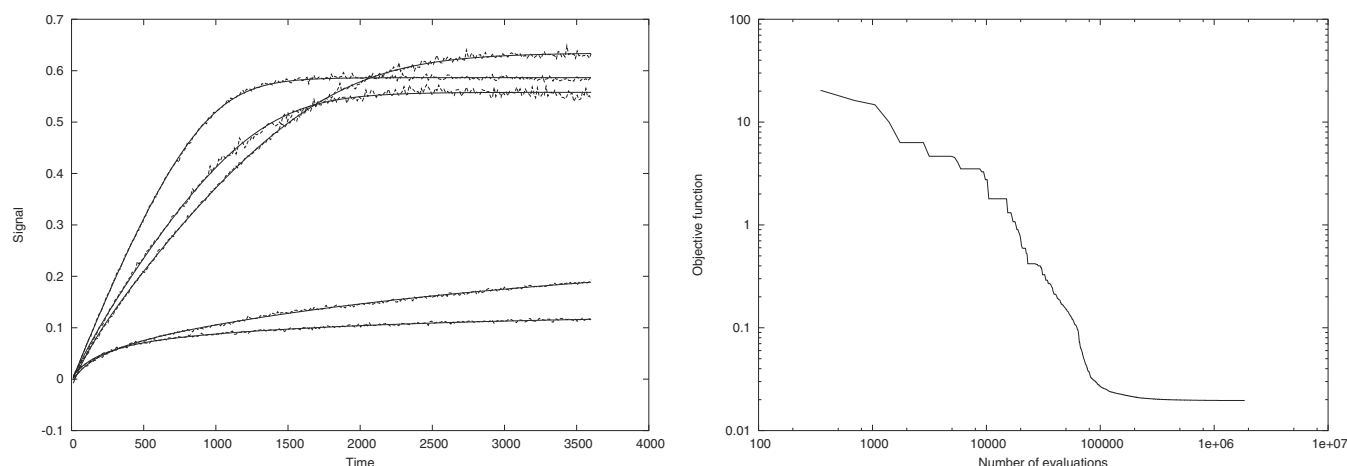*To whom correspondence should be addressed.

Fig. 1. Parameter estimation of HIV model. Figure on the left represents the simulated (smooth curves) and experimental data (noisy curves); figure on the right is the convergence curve (objective function versus number of evaluations).

which is a solver suitable for both non-stiff and stiff problems that often appear in equations of biological systems and is written in C. The libSRES has been successfully applied to these two models. For the three-step model, we repeated the experiment by Moles *et al.*, 2003 and found the nominal values successfully. For the HIV model, we found a few sets of parameters that would fit the experimental data very well. The simulation curves produced by one of these sets are represented in Figure 1. Not only these sets of parameters are comparable with (if not better than) previous works (Kuzmič, 1996; Mends and Kell, 1998), but also much less evaluations are required, which means less time consumed (see convergence curve in Fig. 1). These results indicate the good performance of libSRES for biochemical problems. The details are described in the documentation.

The library provides a simple interface for user programs. A manual for using libSRES is described in detail in the documentation at the URL provided. Here we give a simple description. As described in Figure 2, the optimization problem is coded in function 'fitness', in which the objective and penalty functions should be defined. For biochemical dynamic system, an ODE solver may be introduced into this user-defined function. And then, the user's program should call the libSRES to do the optimization work. The libSRES provides three routines to implement the SRES algorithm: ESInitial, which initializes the optimization process (i.e. the evolution process) based on defined variables. These algorithm variables may vary with the problems; ESStep, which evolves (or optimizes) generation by generation until the termination criteria are satisfied. In each generation, the highest ranked individuals will be selected to produce the next generation; ESDeInitial, which frees memory allocated by libSRES when the evolution process is stopped. When the program is running, it outputs statistical information at each generation, including the estimated parameters and fitness value. In a word, what the user needs to do is to define the special problems in 'fitness' function and to call the three routines in sequence.

When many parameters are to be solved at the same time, the parameter-estimation process requires high-performance computing systems because of huge search space, which cannot be provided
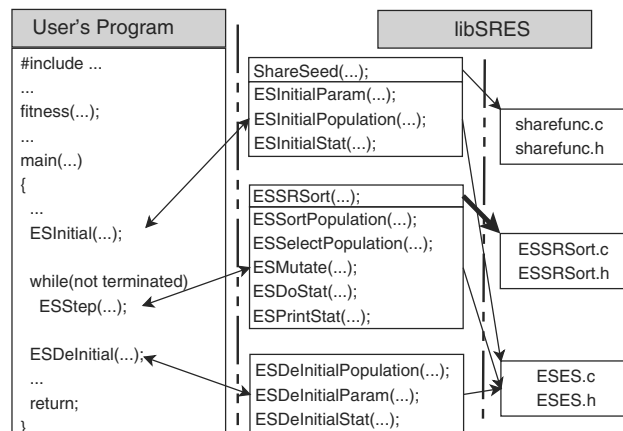

Fig. 2. Framework of user's program and libSRES.

by a single personal computer. Therefore, more computational resources are becoming necessary for large-scale biochemical pathway estimation. For this purpose, an MPI (Message Passing Interface) version has been developed for high performance on both massively parallel machines and on workstation clusters. The MPI version of libSRES has almost the same user interface as the normal version described above, which makes it very easy to be used and allows the programs developed with the normal version to be reused smoothly.

The distribution includes all of the C source code for the libSRES and the 13 example applications. The library has been successfully tested under Linux/Unix operating systems with GCC compiler. The MPI version has also been tested on Linux clusters with MPI implemented. For academic usage, the libSRES is free and falls under the GNU General Public License. The library is available at http://csbl. bmb.uga.edu/~jix/science/libSRES/. The web site also contains the test data file and complete documentation in the forms of both HTML pages and PDF file. The source code is explained in detail in the documentation, as well as the example programs describing how the library can be used.

ACKNOWLEDGEMENTS

Conflict of Interest: none declared.

REFERENCES

Banga,J.R. *et al.* (2003a) Dynamic optimization of bioreactors: a review. *Proc. Ind. Natl Sci. Acad.*, **69A**, 257–265.

Banga,J.R., Moles,C.G and Alonso,A.A. (2003b) Global optimization of bioprocesses using stochastic and hybrid methods. In Floudas,C.A. and Pardalos,P.M. (eds), *Frontiers in Global Optimization, vol. 74 of Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, Hingham, MA, USA, **74**, pp. 45–70.

Battogtokh,D. *et al.* (2002) An ensemble method for identifying regulatory circuits with special reference to the *qa* gene cluster of *Neurospora crassa*. *Proc. Natl Acad. Sci. USA*, **99**, 16904–16090.

Cohen,S.D and HindMarsh,A.C. (1994) *CVODE User Guide*. Lawrence Livermore National Laboratory, Numerical Mathematics Group, CA.

Gardner,T.S *et al.* (2003) Inferring genetic networks and indentifying compoud mode of action via expression profiling. *Science*, **301**, 102–105.

Kuzmič,P. (1996) Program (DYNAFIT) for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.*, **237**, 260–273.

Li,H. *et al.* (2005) Detection of parallel functional modules by comparative analysis of genome sequences. *Nat. Biotechnol.*, **23**, 253–260.

Mendes,P. and Kell,D.B. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, **14**, 869–883.

Moles,C.G. *et al.* (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.*, **13**, 2467–2474.

Runarsson,T.P. and Yao,X. (2000) Stochastic ranking for constrained evolutionary optimaziation. *IEEE Trans. Evol. Comput.*, **4**, 284–294.

Runarsson,T.P. and Yao,X. (2005) Search biases in constrained evolutionary optimization. *IEEE Trans. Sys. Man Cybern.*, **35**, 233–243.