Title	Heterogeneous Kinase Computing : A Novel Class of Optimization Algorithms Inspired by Molecular Biology (Mathematical Optimization Theory and its Algorithm)
Author(s)	Liu, Jian-Qin; Shimohara, Katsunori
Citation	数理解析研究所講究録 (2001), 1241: 1-9
Issue Date	2001-12
URL	http://hdl.handle.net/2433/41622
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

数理解析研究所講究録 1241 巻 2001 年 1-9

Heterogeneous Kinase Computing:

A Novel Class of Optimization Algorithms Inspired by Molecular Biology

Jian-Qin Liu and Katsunori Shimohara
(劉　健勤　下原勝憲)

Information Sciences Division, ATR International
2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan
((株) 国際電気通信基礎技術研究所　先端情報科学研究部)

Abstract **--** In order to create formal tools for designing efficient optimization algorithms, a new formal model of molecular computing with a parallel mechanism for optimization is proposed and its corresponding computability, which is equivalent to a Turing machine and schemes toward global optimum are discussed.

Keyword: Optimization, Molecular Computing, LS-systems.

1. Introduction

From the viewpoint of solving optimization problems using parallel methods, biologically inspired, unconventional computing methods have received great attention, owing to the advantages of living systems such as their adaptation, self-organization, evolvability (which here refers to a kind of mechanism that enables the living system to self-adjust its structure). For example, in DNA computing, massively parallel NP problem solving processes are implemented by the self-assembly of DNA molecules. The following three characteristics of biological mechanisms motivate us to further explore molecular computing approaches in order to solve optimization problems using parallel methods:
(1) The mechanism of signal molecules for communication inside cells implies a kind of parallel signal processing principle.
(2) The biological pathways for signal transduction in cells can be formalized as a kind of hypergraph in terms of theoretical computer science.
(3) The living system has a perfect self-regulation structure for energy minimization. The complexity increases greatly when the interactions of pathways are carried out. This shows us that we can embed the mechanism of complexity increases into optimum algorithms.

The core of our idea is to operate on large-scale molecule objects in parallel and efficiently towards the output that we design in optimization, only within the robust shell of the cells, and to carry out computation by controlled kinase-guided biochemical reactions in cells (here we select Rho-GTPases and related signaling pathways in cells). This proposal is inspired by Rho family GTPases[1] and the related functional information processing mechanism in signaling pathways and cell communications. Here, the computing reflected in the forms of the pathways' representation, is carried out by kinases,

enzymes and functional-proteins in cells and controlled by the internal mechanism of cells *in situ*. Inputs and outputs of this computing are defined as a subset of the pathways, i.e., the subset of hypergraphs or part of pathways. The reasons for selecting pathways in cells are briefly summarized as:
(1) In essence, cells are excellent natural and all-round "test-tubes" made by nature, which provide a functional and powerful frame for a cell communication environment in *an "in-situ* "way;
(2) To employ pathways as a candidate pool (population) for combinatorial solutions can adaptively increase the complexity of NP-problem-solving processes themselves;
(3) One of the latest advances in cell signaling systems in cell molecular biology is the discovery of the signal transduction mechanism of pathways of Rho family GTPases[1], that plays a key role in cell communication. As direct controllers that regulate the signaling pathway, the Rho-GTPase and related signaling mechanism in cells are good examples in the real world for us to study and then to use in designing an optimization algorithm.

We have been studying a kinase computing paradigm[2,3], which was initiated by us, in aspects such as:
(1) Computability that equals to a Turing machine;
(2) A non-Turing model of computing that is related to the concept of the super-Turing machine;
(3) A robust structure for controlling pathways in kinase computing;
(4) Schemes for large-scale NP problem solving;
(5) Algorithmic and logic units of kinase computing;
(6) The relationship between kinase computing and artificial life (algorithmic chemistry).

2. Formalization of Kinase Computing

In order to conduct theoretical analysis for optimization schemes by molecular computing with a biological medium of Rho-GTPases, we have worked out a formal description for the underlying mechanisms. Through formalization we have proven that it has Turing computability, which is necessary to ensure that a sound theoretical foundation is available for any concrete algorithms. This model can be formalized as the following construct:

$$\text{Wcells} = < V, T, D, V', E, Y, Z, PTs, Q >$$

where

V -- the alphabet set;

T -- the terminal set and $T \subset V$;

D -- the set of {0, 1};

V' -- the set of vertexes;

E -- the set of edges;

Y -- the set of hypergraphs <HE, V'>
in which HE is the set of hyperedges in Y
and the corresponding HR (hyperedge replacement)
and VR (vertex replacement) are defined based on HE;

Z -- the set of local concentration and $Z \subset V'$;

PTs -- the set of pathways -- the directed graphs that have inputs and outputs in V and contain hyperedges in HE with interactions that fall into HE;

It is defined as:
[pathway] ::= [pathway] U [single-hyperedge with two vertexes]
under the operation of Q,

Q -- the set of operators for operations on hypergraphs in HE from V and E,
i.e., the objects in PTs;

$$Q = \{Q1, Q2, Q3, Q4\}.$$

The operation processes carried out by the operator set of Q is formalized as four rules in terms of graph rewriting on hypergraphs:

(1) For Q1 (the rule of "interaction"):

$\alpha \rightarrow \chi,\ \beta \rightarrow \delta,\ \varepsilon \rightarrow \phi, \ldots,\ \gamma \rightarrow \eta$
where $\alpha, \chi, \beta, \delta, \varepsilon, \phi, \ldots, \gamma, \eta$ refer to pathways as we defined in PTs.

(2) For Q2 (the rule of "feedback-making"):

$\varphi \rightarrow \kappa,\ \lambda \rightarrow \mu$
where $\varphi, \kappa, \lambda, \mu$ refer to pathways and $\varphi = \mu, \kappa \neq \lambda$.

(3) For Q3 ("addition of new pathways")

$\nu \rightarrow \pi\theta$

where ν, π and θ refer to pathways.

(4) For Q4 (deletion):

$\rho \rightarrow \sigma$

where ρ, σ refer to pathways.

So from the above model we can derive the following writing on hypergraphs:

(1) PTs (Gh) -> PTs (Gh') s.t. rewriting on hypergraphs by Q,

(2) Gh" -> Gh "' s.t. rewriting on hypergraphs by Q,

where Gh, Gh', Gh" and Gh"' refer to hypergraphs in Y and rewriting is carried out through executing Q operators.

In the context of biological computing, we define:

V -- used for symbolic representation for all of the chemicals in cells,
e.g., the name of the reactants in biochemical reactions,
which lead to signaling pathways where Rho-GTPases are involved;

D -- introduced for digital denotation in tapes of equivalent Turing-machines;

Z -- the chemicals in cells;

and

D -- the measurement decided by the threshold quantity represented by symbols,
which belong to the set of V.

3. Computability

By imitating biological processes of the information processing mechanism for biochemical signaling activities activated by Rho GTPases in cells, the result of a corresponding Turing-machine can be obtained for the string, e.g. "1 0 1 0 1 1 1 1 1 1 1". The bio-chemical processes from the Rho-GTPases are equivalent to this computation in the abstract meaning of computers.
In essence, the core of our work is to convert the pathways defined here -- a kind of special directed graphs with constraints -- into strings with special representation. Relevant logical constraints for graphs should also be taken into consideration. All of the constraints are reflected in the rules of our formal system where the string representation from "pathways" is defined.
For the string representation of the pathway, we have

$$Gs = f_{ls}(Path_i) \ , \ Path_i \in PTs \text{ and } i = 0,1,\ldots$$

and

$$Gs = f_{ls}(f^{-1}{}_{ls}(Gs)),$$

where Gs is a Gödel number and $Path_i$ is the string representation for the pathways and the explanation is given below. Uniqueness is thus obtained (injective mapping) and the inputs

and outputs of pathways are never repeated as the same in our domains or in the set of all the objects. The resulting string takes the form of

"input-Vs + Es + output-Vs",

e.g.

| E00 E01 ... E0m | | Eh0,Eh1,Eh2,Eh3,Eh4... | | En0 En1 ... Enl |,

i.e.

Vs 0 | Es 01 | Vs 1 with length at least 3,

where

Gs is with the length of all the strings corresponding to sub-hypergraphs.

Structural features show that the above representation is consistent with the notation used in string rewriting. Rewriting the above strings can be represented as a corresponding mapping in the domain of the Gödel number as

$$N \rightarrow N.$$

It is obvious that Gödel number $\neq 0$ means that pathways exist and that Gödel number = 0 means that pathways have the meaning of non-existence, which shows that all of the 0s in the Gödel number's representation and the symbol of the empty λ in the string have the same meaning.
Furthermore, uniqueness is guaranteed and the bijection function (mapping) for the following also exists:

$$Gs \leftrightarrow Gs'$$

Therefore, the whole set of all of the pathways can be covered in the operation carried out by operators in the Q set. All of the elements that we got from the Gödel numbering fall into the set described by the predicates of Q. The predicate set includes the predicates of Q1~4 which correspond to the operations for handling the replacement of components in the pathways. Here the term "replacement" for input, output, middle, and related representations is assigned meanings in graph rewriting according to the "replacement" for the updated two measures (i.e., Gödel numbers before and after rewriting in the left and right side of the rules). The mapping from N into N can be obtained by Gödelization after the pathways' operations executed to all of the objects in the formal system we are discussing.

With respect to the computability of the model mentioned above, we have the following.

Proposition: Our proposed model of kinase computing is Turing-computable.

4. Outline of the Basic Algorithm

Here we present the scheme for optimization in terms of kinase computing. The two major points are biologically inspired:
(1) The heterogeneous architecture of biological molecular computers:
(2) The internal structure defined as pathways related to hypergraphs;
We find that the merits of kinase computing for designing optimization algorithms come from its parallelism, adaptivity, evolvability and emergeability. The parallelism that we employ is our work on "self-catalysis", which can be regarded as our effort towards exploring new theories related to auto-catalysis. The adaptive mechanism in kinase computing mainly contributes to the self-adjusting of the parameters in the pathways. The increasing complexity dynamically related to the internal structure is our consideration in the area of behavior as related to evolvability. To evollve the internal mechanism to reach the order of complexity that matches the problem to be solved, is guaranteed by the population arrangement and evolutionary processes basically in non-parametric programming, owing to the fact that the population size is not limited in advance.

The basic algorithm of kinase computing for optimization is given as follows:

Step 1: Initialize and set the population with hypergraphs and pathways.
Here the candidates for searching for potential solutions are defined as the vertexes in the hypergraphs and also the input nodes of the related pathways. The pathway is made to correspond to provide the output nodes with the value of the object function and condition constraints.
Step 2: Calculate the outputs
Step 3: If the optimal objects exist, then go to step 7.
Else go to the next step.
Step 4: Execute the Q operators on the objects in the population.
Step 5: Adjust the pathways and related hypergraphs.
Step 6: Go to step 2.
Step 7: Select the optimal objects as output.

5. Construction of the Schemes toward Global Optimum

According to[4], the global optimum is formulated in a manifold space with probabilistic measurement defined. Related notations are listed as follows:

M -- the number of the population,
Q -- a standard θ - generalized normal random matrix ($L \times L$)
with the distribution of $N_{L,L}(A,Y,C,\theta)$,
X -- a standard θ - generalized normal random matrix.
with the distribution of $N_{L,L}(S,V,U,\theta)$,
A,Y,C,S,V and U -- matrices,
θ -- the parameter of the corresponding distribution,
C -- the matrix with diagonal non-zero (taken as 1 in values) and the rest zeros.

The global optimum is inferred as

$$\sum_{i=0}^{p+1} \sum_{j=0}^{p+1} \psi_{ij} F_i F_j \alpha^2_{ij} = 0,$$

where

$$\psi_{0j} = \psi_{i0} = 2/M, \quad F_0 = \alpha_{0j} = \alpha_{i0} = 1,$$

and

$$\psi_{ij} = \frac{\Gamma(1/\theta)}{\Gamma(3/\theta) \| (VV^t) \otimes (UU^t) \|},$$

where $i \neq 0$ and $j \neq 0$.

The condition is

$$\sum_{l=1}^{p+1} y_{il} = d_i \sum_{l=1}^{p+1} g_l \partial / \partial x_i H_l(x),$$

$$y_{il} = 0 \quad \text{if} \quad p+1 < l \leq L,$$

$$E_\theta [q_{ij}^2] = \| E_w \partial i \partial j Z(B_l, W)] \|,$$

$F = \text{vec}(X^t)$, i.e., the tangency vector in the statistical manifold $\Im$.

Here the notation of $H_i(.)$ is:

H_0 = the object function,

H_j ($j > 0$) is the condition constraint in the underlying problem,

B_l is the measurement for the goodness or fitness of the individual in the population.

Based on the above forms for global optimum, we must design corresponding algorithms to "evolve" the population. In the quantity of the above proposition, we present the following scheme to design the searching procedure.

The mapping is made from hypergraph space to the statistical manifold space $\Im$ mentioned above as

$$Z(B_l, W) = \ln P(B_l, W)$$

where W = the set of the probability of all of the operations in the set Q.

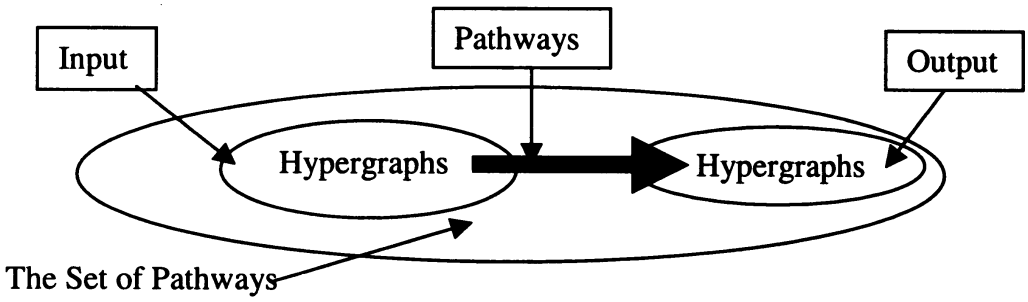

Fig. 1 The relationship among the objects

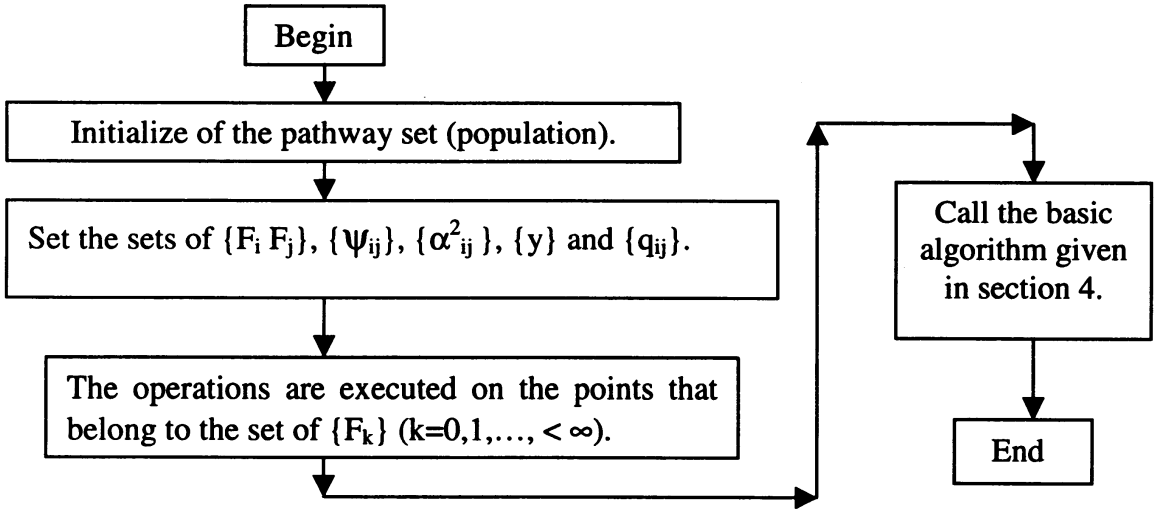

Fig. 2 The flow chart

And we have that

(1) The set of $\{F_i(F_j)\}$ is assigned to the hypergraph that denotes the input nodes for the pathway;

(2) The sets of $\{\psi_{ij}\}$ and $\{\alpha^2_{ij}\}$ are defined as parameters to be estimated by empirical ways;

(3) The set of $\{y\}$ is assigned to the pathways for those hypergraphs as their input and is generated by repeated operations guided by the Q rule in LS-systems;

(4) It is assumed that the forms of the $E_\theta [q_{ij}^2]$ are linear;

(5) The set of $\{q_{ij}\}$ is also assigned by the values in the hyperedges.

The algorithm for searching for the target is carried out in an iterative manner and will stop until the optimum objects emerge in the population. The objects concerned in the above processes are illustrated in figure 1 and a brief description for this algorithm is given in figure 2.

6. Conclusion

The formal model of kinase computing has provided a parallel structure that can increase complexity within the system itself. The processes that allow the individuals to evolve towards the global optimum are operable in theory. So we can employ interactions among the pathway-like structure that are updated in a complexity increasing manner. This offers potential for algorithm designing that is oriented to specific optimization problems.

Acknowledgements

J.-Q. Liu sincerely thanks Prof. Masami Ito, Prof. Shigeru Masuyama, and Prof. Reiji Nakajima for their helpful advice and fruitful discussions on formal languages and the theory of computation, optimization theory and parallel algorithms, and formalization and computability, respectively.

References

[1] Kozo Kaibuchi, Shinya Kuroda, and Mutsuki Amano, Regulation of the cytoskeleton and cell adhesion by the Rho family GTPases in mammalian cells, Annu. Rev. Biochem. 1999.68: 459-486.

[2] Jian-Qin Liu and Katsunori Shimohara, Extending proteomic computing to a parallel prototype, IPSJ SIG Notes, 2001-MPS-33, Vol. 2001, No.27, pp.73-76.

[3] Jian-Qin Liu and Katsunori Shimohara, Pathway graph models for molecular computing *in situ*, RIMS Kokyuroku, Kyoto University, 2001.

[4] Jian-Qin Liu and Min-Jie Wei, A global optimality criterion for evolutionary computation, J. Cent. South Univ. Technol., Vol.5, No.1, May 1998; 64-67.