

# Optimization to Predict Oncogenes in Genome-Scale Metabolic Networks of Non-Small-Cell Lung Cancers

You-Tyun Wang, Min-Ru Lin, Wei-Chen Chen, Wu-Hsiung Wu and Feng-Sheng Wang\*

Department of Chemical Engineering, National Chung Cheng University, Chiayi, Taiwan

## Introduction to Nested Hybrid Differential Evolution (NHDE)

The oncogene inference optimization framework is used for in silico experiments to simulate gene screening procedures and predict oncogenes. The framework is a trilevel optimization problem (TLOP), a mixed-integer optimization formulation, and described as

$$\left\{ \begin{array}{l} \text{Outer optimization problem:} \\ \text{Similarity ratio of metabolite-flow rates and fluxes to the template:} \\ \max_{\delta, z_i} SR_M, \max_{\delta, z_i} SR_F \\ \text{Fuzzy equal grade of metabolite-flow rates and fluxes} \\ \text{compared to the template:} \\ \widetilde{EqualLFC}_M^{MUBL} \approx LFC_M^{CABL}, \widetilde{EqualLFC}_F^{MUBL} \approx LFC_F^{CABL} \\ \text{subject to the inner optimization problems:} \\ \left\{ \begin{array}{l} \text{FBA problem} \\ \max_{\mathbf{v}_{f/b}} obj = (w_{ATP} v_{ATP} + w_{biomass} v_{biomass}) \\ \text{subject to} \\ \mathbf{N}(\mathbf{v}_f - \mathbf{v}_b) = \mathbf{0} \\ v_{f/b,i}^{LB} \leq v_{f/b,i} \leq v_{f/b,i}^{UB}, z_i \notin \Omega^{MU} \\ v_{f/b,j}^{LB,MU} \leq v_{f/b,j} \leq v_{f/b,j}^{UB,MU}, z_j \in \Omega^{MU} \end{array} \right. \\ \left\{ \begin{array}{l} \text{UFD problem} \\ \min_{\mathbf{v}_{f/b}} \sum_{i \in \Omega^{int}} (v_{f,k})^2 + (v_{b,k})^2 \\ \text{subject to} \\ \mathbf{N}(\mathbf{v}_f - \mathbf{v}_b) = \mathbf{0} \\ v_{f/b,i}^{LB} \leq v_{f/b,i} \leq v_{f/b,i}^{UB}, z_i \notin \Omega^{MU} \\ v_{f/b,j}^{LB,MU} \leq v_{f/b,j} \leq v_{f/b,j}^{UB,MU}, z_j \in \Omega^{MU} \\ obj \geq obj^* \end{array} \right. \end{array} \right.$$

The optimization is rewritten as the following simplified formulation for easily explaining the NHDE algorithm.

$$\left\{ \begin{array}{l} \text{Outer optimization problem:} \\ \max_{\mathbf{x}, \mathbf{z}} f(\mathbf{x}, \mathbf{z}) \\ \text{subject to the inner optimization problems:} \\ \left\{ \begin{array}{l} \text{Linear programming (LP) problem} \\ \max_{\mathbf{x}} obj = \mathbf{c}^T \mathbf{x} \\ \text{subject to} \\ \mathbf{Ax} = \mathbf{0} \\ \mathbf{x}_{LB} \leq \mathbf{x} \leq \mathbf{x}_{UB}, \mathbf{z} \notin \Omega^{MU} \\ \mathbf{x}_{LB}^{MU} \leq \mathbf{x} \leq \mathbf{x}_{UB}^{MU}, \mathbf{z} \in \Omega^{MU} \end{array} \right. \\ \left\{ \begin{array}{l} \text{Quadratic programming (QP) problem} \\ \min_{\mathbf{x}} \mathbf{x}^T \mathbf{x} \\ \text{subject to} \\ \mathbf{Ax} = \mathbf{0} \\ \mathbf{x}_{LB} \leq \mathbf{x} \leq \mathbf{x}_{UB}, \mathbf{z} \notin \Omega^{MU} \\ \mathbf{x}_{LB}^{MU} \leq \mathbf{x} \leq \mathbf{x}_{UB}^{MU}, \mathbf{z} \in \Omega^{MU} \\ \mathbf{c}^T \mathbf{x} \geq \mathbf{c}^T \mathbf{x}^* \end{array} \right. \end{array} \right.$$

The inner optimization problem consists of LP and QP problems, which is a sequential relationship.

The NHDE algorithm is a stochastic optimization based on hybrid differential evolution (HDE), which was extended from the original DE algorithm (Storn and Price, 1996; Storn and Price, 1997). The basic operations of original DE and modified NHDE are shown in Tab. S2-1. The detailed procedures have discussed by Wang (2017).

Table S2-1. Basic operations for the original DE and NHDE algorithms

Original DE	Modified NHDE
1. Representation and initialization	1. Representation and initialization
2. Mutation	2. Mutation with rounding operation
3. Crossover operation	3. Crossover operation
4. Selection and evaluation	4. Restriction operation
5. Repeat steps 2 to 4	5. Selection and evaluation
	6. Solve LP/QP problems for each candidate gene
	7. Compute fitness for each feasible design
	8. Migration operation performed naturally or enforced if necessary
	9. Repeat steps 2 to 6

The computational procedures of NHDE are listed in Tab. S2.2. NHDE is a parallel direct search algorithm (as shown in Fig.S2.1) that utilizes a population of  $N_p$  individuals (enzymes) to find an optimal solution. The initialization process randomly generates  $N_p$  individuals to cover the entire search space uniformly. Each individual in the population consists of a set of enzymes that are selected to be modulated.

The mutation operator of NHDE adopted from DE was an essential component compared with other evolutionary algorithms. Different from conventional evolutionary algorithms, the mutation operation of DE/NHDE uses the difference between two or four randomly chosen individuals as an evolutionary direction. The  $i^{th}$

mutant individual  $(\mathbf{z}^G)_i$  in generation G is obtained through the difference of two or four random individuals as expressed in the following form:

$$(\mathbf{z}^G)_i = \text{INT} \left\{ (\mathbf{z}^G)_p + \rho^G \left[ (\mathbf{z}^G)_j - (\mathbf{z}^G)_k + (\mathbf{z}^G)_l - (\mathbf{z}^G)_m \right] \right\}, i = 1, \dots, N_p$$

where random indices  $j, k, l, m \in \{1, \dots, N_p\}$  are mutually different. The operator INT in the equation is used to rounding the real vector into the integer vector. In DE, the differential mutation factor  $\rho^G \in [0, 1.2]$  is fixed and set by the user to obtain faster convergence. This factor is used to control the step length along the searching direction. A random mutation factor was used in NHDE to obtain more diversified individuals. NHDE also includes an additional mutation strategy that applying a linear crossover for the  $i^{\text{th}}$  individual and the best individual  $(\mathbf{z}^G)_b$  to generate the parent individual. The parent individual is therefore expressed as follows:

$$(\mathbf{z}^G)_p = \rho_p^G (\mathbf{z}^G)_b + (1 - \rho_p^G) (\mathbf{z}^{G-1})_i$$

where the factor  $\rho_p^G$  is a random number between zero and one generated by a uniform distribution generator, and  $(\mathbf{z}^{G-1})_i$  indicates the  $i^{\text{th}}$  mutant individual in the previous generation. The mutation operation may cause the mutant individual escape from the search domain. The mutation operation may cause the mutant individual to escape the search domain (i.e., bounds are violated). If this occurs, it is replaced by a random number within the lower and upper bounds of the particular decision variable, thus restricting to the search domain. The choice of mutation factor for DE/NHDE is heuristic and random. When population diversity is low, candidate individuals rapidly cluster together such that the individuals cannot be further improved, and premature convergence occurs. Similar to conventional evolutionary algorithms, the local population diversity could be increased by using a crossover operation such as a binomial crossover.

NHDE use the difference between two or four mutually independent individuals to determine the direction of search and obtain a mutant individual. This differential mutation converges quickly so that most individuals cluster around the best candidate individual in some generations. Consequently, the population diversity and exploration capability diminish and clustered individuals are unable to reproduce more diversified individuals through the mutation operation because the weighted difference is nearly zero. The recombination of mutant individuals and their clustered parents further prevents the reproduction of a diversified population. Therefore, all individuals quickly cluster together and superior individuals cannot be generated through mutation and crossover operations.

The migration operation of the NHDE algorithm is used to help individuals escape from the local cluster, but this operation is performed only if the population diversity falls below a desired level. The degree of population diversity  $\zeta$  is introduced to check whether the migration operation should be performed. Each element of the  $i^{\text{th}}$

individual  $(z^G)_i$  in generation  $G$  is referred to as a gene of the individual, and the gene diversity index  $dz_{ji}$  is given by

$$dz_{ji} = \begin{cases} 0, & \text{if } z_{ji}^G = z_{jb}^G, j = 1, \dots, n; i = 1, \dots, N_p; i \neq b \\ 1, & \text{otherwise,} \end{cases}$$

where  $z_{ji}^G$  and  $z_{jb}^G$  are the  $j^{\text{th}}$  gene of the  $i^{\text{th}}$  and best individual at the  $G^{\text{th}}$  generation, respectively.  $dz_{ji}$  is set to zero if the  $j^{\text{th}}$  gene of the  $i^{\text{th}}$  individual is identical to the best gene; otherwise it is set to one (Chiou and Wang, 1999; Liao, et al., 2001).  $\zeta$  is defined as the ratio of total gene diversities to the total number of genes other than those of the best individual:

$$\zeta = \frac{\sum_{i=1, i \neq b}^{N_p} \sum_{j=1}^n dz_{ji}}{n(N_p - 1)}$$

The value of population diversity degree ranges between zero and one. A value of zero implies that all of the genes are clustered around the best individual. On the other hand, a value of one indicates that current candidate individuals are a completely diversified population. The desired tolerance for population diversity is assigned by the user. A tolerance value of zero implies that the migration operation in NHDE is switched off, and one implies that the migration operation is performed at every generation. Consequently, the user can set a tolerance value for population diversity degree,  $\varepsilon \in (0, 1)$ . If  $\zeta$  is smaller than  $\varepsilon$ , then NHDE performs migration operations to regenerate a new population in order to escape from a local point; otherwise, NHDE suspends the migration operation and maintains a constant search direction toward finding a new solution.

Table S2-2. The NHDE algorithm for iteratively selecting a set of candidate enzymes and to infer optimal oncogenes

<b>NHDE</b>	
1.	<p>Representation and initialization</p> $(\mathbf{z}^0)_i = \text{uniformInt}(\mathbf{z}^{\min}, \mathbf{z}^{\max}), i = 1, \dots, N_p$ <p>Each individual is generated by an integer random number between <math>\mathbf{z}_{\min}</math> and <math>\mathbf{z}_{\max}</math> with uniform distribution</p>
2.	<p>Mutation with rounding operation</p> $(\hat{\mathbf{z}}^G)_i = \text{INT} \left\{ (\mathbf{z}^G)_p + \rho^G \left[ (\mathbf{z}^G)_j - (\mathbf{z}^G)_k + (\mathbf{z}^G)_l - (\mathbf{z}^G)_m \right] \right\}$
3.	<p>Crossover operation</p> $z_{ji}^G = \begin{cases} z_{ji}^{G-1}, & \text{if a random number} > C_R \\ \hat{z}_{ji}^G, & \text{otherwise, } j = 1, \dots, n; i = 1, \dots, N_p \end{cases}$
4.	<p>Restriction operation</p> $z_{ji}^G = \begin{cases} z_{ji}^G, & z_{ji}^G \in [z_j^{\min}, z_j^{\max}] \\ \text{uniformInt}(z_j^{\min}, z_j^{\max}), & z_{ji}^G \notin [z_j^{\min}, z_j^{\max}] \end{cases}$
5.	<p>Selection and evaluation</p> <p>(a) For each enzyme, solve the inner LP/QP problem by LP/QP solver, respectively</p> <p>(b) Compute fitness for each feasible solution</p> $\text{fitness} = f(\mathbf{x}, \mathbf{z}) + \text{penalty}$
6.	<p>Migration operation performed naturally or enforced if necessary</p> $(\mathbf{z}^G)_i = \text{uniformInt}(\mathbf{z}^{\min}, \mathbf{z}^{\max}), \text{ if } \zeta \leq \varepsilon = [0,1]$
7.	<p>Repeat steps 2 to 6</p>

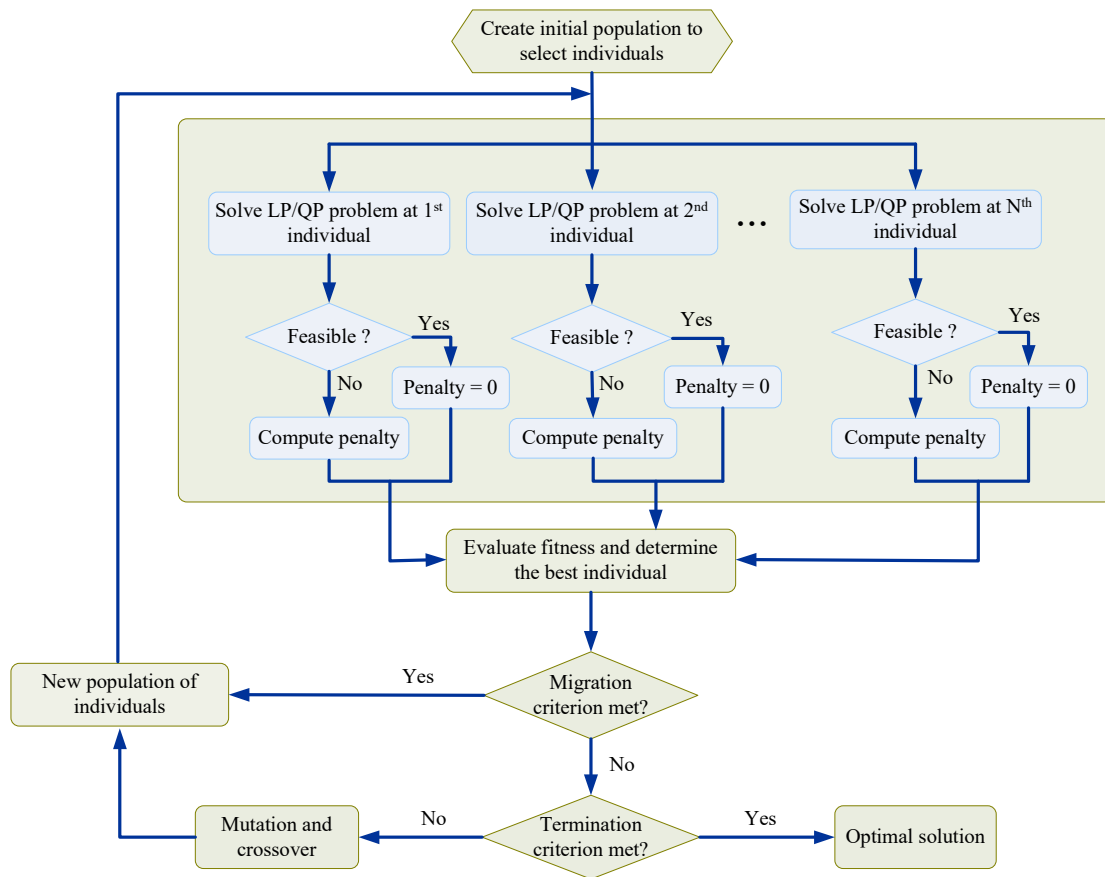


Figure S2-1. Flowchart of the parallel search algorithm in NHDE

## References

Storn, R. and Price, K. (1996) Minimizing the real functions of the ICEC'96 contest by differential evolution. *Evolutionary Computation*, 1996., Proceedings of IEEE International Conference on. IEEE, Nagoya, pp. 842 - 844.

Storn, R. and Price, K. (1997) Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization*, 11, 341-359.

Wang, FS (2017) Nested differential evolution for mixed-integer bi-level optimization for genome-scale metabolic networks, Ch.12, in *Differential evolution in chemical engineering* edited by Rangaiah and Sharma, World Scientific.

Chiou, J.P. and Wang, F.S. (1999) Hybrid method of evolutionary algorithms for static and dynamic optimization problems with application to a fed-batch fermentation process, *Computers & Chemical Engineering*, 23, 1277-1291.

Liao, C.T., Tzeng, W.J. and Wang, F.S. (2001) Mixed-integer hybrid differential evolution for synthesis of chemical processes, *Journal of the Chinese Institute of Chemical Engineers*, 32, 491-502.