2nd International Conference on Autonomous Robots and Agents
December 13-15, 2004 Palmerston North, New Zealand

A Hybrid Method of Evolutionary Algorithms and Gradient Search

Hyun Wook Woo, Hyuk Hoon Kwon, and Min-Jea Tahk
Dept. of Aerospace Engineering,
Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
hwwoo@fdcl.kaist.ac.kr

Abstract

This paper proposes a hybrid evolutionary algorithm. It is based on a normal evolutionary algorithm and modified with gradient search technique. The gradient individual, which is the best individual of population, propagates to the next generation's one using gradient information. Other individuals except the best individual are distributed symmetrically near the best individual for gradient calculation. The central difference method is used for the gradient calculation and BFGS algorithm for Hessian. The gradient estimation accuracy and the overall performance of the proposed method are tested with numerical examples.

Keywords : Hybrid Algorithm, Evolutionary Algorithm, Convergence Acceleration, Gradient Estimation, Quasi-Newton Method

1 Introduction

Search space for general optimization problem can be divided into several convex regions and each region has a local minimum. Classical gradient-based optimization methods show better performance than evolutionary algorithm(EA) in finding the minimum of a convex region. However, they are not good at finding the global minimum of the problem with multiple local minima. On the other hand, EA has good ability in finding the global minimum[1], although it has slower convergence speed in convex region than the gradient-based method.

This paper proposes an accelerated hybrid evolutionary method exploiting gradient search. The new method is good at finding the global minimum since it is based on EA, moreover it shows good convergence speed near minimum points since it utilizes the gradient information of the best individual at each generation.

The estimation method of the gradient will be explained first, then the gradient search for hybrid algorithm, and we will summarize the algorithm. Numerical results for benchmark problems will be shown last.

2 Gradient Estimation

The new algorithm needs gradient information for accelerating EA, gradient at best individual can be the best choice among all individuals. The calculation of gradient vector needs proper distribution of individuals and adequate numerical method.

Well-known numerical methods for gradient calculation are forward difference, backward difference and central difference method. Since the central difference method has a higher order of error($O(h^2)$, h is the step size) than others($O(h)$)[2], the central difference method is used for gradient calculation and individuals are distributed suitably.

2.1 Offspring Generation

To keep the merits of evolutionary algorithm, the same selection and cost evaluation methods are used. However, the offspring generation process has some modification. The process is divided into two stages for gradient calculation.

For the first stage, m individuals are generated using selection, recombination, and mutation process and and constitute the first stage offspring population, which are same with typical EA. Single individual, which is propagated from previous generation's best individual, is added to the offspring population. It is propagated using the gradient method. Then a new best individual is selected from $m+1$ individuals and it will be propogated to the next generation. At the second stage, m more individuals are generated at symmetrical points of the first stage's m individuals(except the best individual) w.r.t the best individual. Even if a second stage individual is better than the first stage's best individual, it does not substitue the best individual. [3]

2.2 Least-square Estimation of Gradient

Let the cost of each individual x_i be $f(x_i)$. Then the cost difference of symmetrical pair Δf_i can be written as the following equation by the central difference method:

$$\Delta f_i = \Delta x_i g \tag{1}$$

where $\Delta f_i = f(x_{m+i}) - f(x_i), \Delta x_i = x_{m+i} - x_i$, and g is the gradient, which is calculated numerically.

Cost difference equations for m symmetrical pair is as follows:

$$\Delta f = \Delta X g \tag{2}$$

where

$$\Delta f = \begin{bmatrix} \Delta f_1 \\ \vdots \\ \Delta f_i \\ \vdots \\ \Delta f_m \end{bmatrix} \quad , \quad \Delta X = \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_i \\ \vdots \\ \Delta x_m \end{bmatrix} \tag{3}$$

Equation(2) is called the measurement equation. The least-square solution of gradient using equation(2) is

$$g = (\Delta X^T \Delta X)^{-1} \Delta X^T \Delta f \tag{4}$$

If the error information of the measurement equation is known, a better estimation result is given by a weighted least-square solution as[3]

$$g = (\Delta X^T R^{-1} \Delta X)^{-1} \Delta X^T R^{-1} \Delta f \tag{5}$$

Adding measurement noise v to equation(2) makes following equation

$$\Delta f = \Delta X g + v \tag{6}$$

The truncation error of the central difference method increases in proportion to the distance of the symmetrical pair. Therefore, with the assumption that each pair's error is uncorrelated, we can let the measurement error covariance matrix R be a diagonal matrix with the diagonal element $r_{ii} = a h_i^4$ where a is a constant and h_i is the distance of ith pair.

2.3 Estimation Results

The Rosenbrock function(n=2) was used to test two estimation methods. the function is given as

$$f(x_1, x_2) = 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2 \tag{7}$$

Evolutionary strategy(ES) is used for optimization and two estimation processes were performed. The estimation processes were separated with ES and gave no effect on the optimization process.

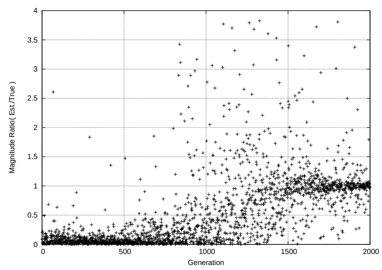
Since the Rosenbrock function is analytically differentiable, the estimated gradient is easily compared with the true one

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} -400 x_1 \left(x_2 - x_1^2\right) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} \tag{8}$$

Estimation results(gradient vectors) are compared with the true one in magnitude ratio and the angle between the two vectors(the true and the estimated).

Figures 1, 2, 3, and 4 shows the magnituge ratio and the angle between vectors for both estimation methods.

Since the individuals are widely spread over non-convex search space in early generations, the estimation result is not accurate at early stages in figures. However it converges to the true values gradually. The inaccurate gradient vector in the early phase doesn't degrade the performance of the hybrid method. It affects only one individual and the method works like a normal EA if the gradient information degrades the individual propagating through gradient search process.



[3]

Figure 1: Magnitude Ratio(Least-square)

3 Gradient Individual

To improve EA with a gradient-based algorithm, the proposed algorithm introduces a gradient individual. The best individual propagates to the best individual of the next generation using the gradient search method.

Each generation's gradient individual is calculated from that of previous generation with gradient g

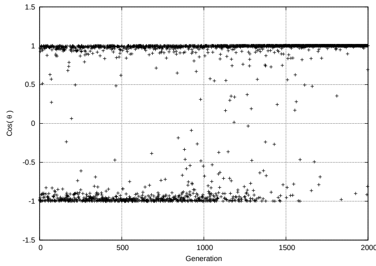


Figure 2: Angle(Least-square)

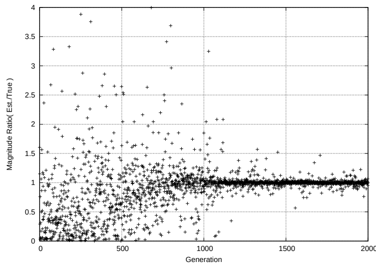


Figure 3: Magnitude Ratio(Weighted least-square)

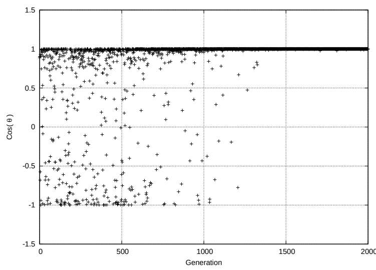


Figure 4: Angle(Weighted least-square)

and inverse Hessian H. The calculated gradient individual is compared with m individuals generated by selection, recombination, and mutation of typical EA. If any of m individuals has better cost than the gradient individual, it will substitute the gradient individual and become a new gradient individual.

3.1 Search with Quasi-Newton method

Let the gradient individual of kth generation x_k and the gradient and inverse Hessian is known, then $k+1$th generation's gradient individual x_{k+1} can be obtained by following equation

$$x_{k+1} = x_k + g_k H_k \tag{9}$$

Calculating equation(9) requires inverse Hessian and BFGS algorithm is the most popular for calculating inverse Hessian numerically.

BFGS algorithm for inverse Hessian[4]

$$\begin{aligned} H_{k+1} &= H_k + \frac{1}{\delta g_k^T \delta x_k} \delta x_k \delta x_k^T \\ &\quad - \frac{1}{\delta g_k^T H_k \delta g_k} H_k \delta g_k (H_k \delta g_k)^T \\ &\quad + (\delta g_k^T H_k \delta g_k) v_k v_k^T \end{aligned} \tag{10}$$

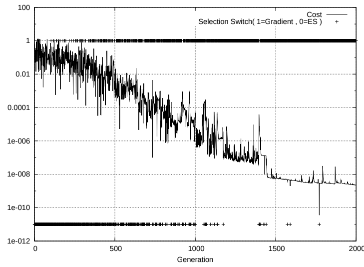


Figure 5: Cost and Flag History

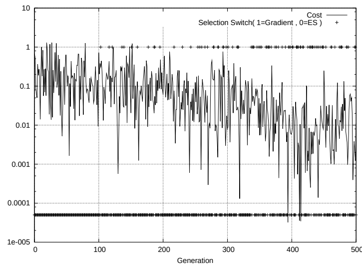


Figure 6: Cost and Flag History for First Quarter

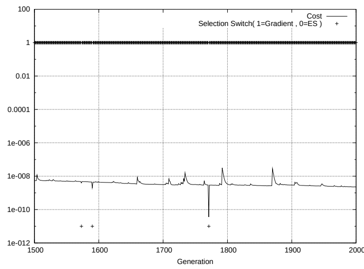


Figure 7: Cost and Flag History for Last Quarter

where

$$v_k = \frac{1}{\delta g_k^T \delta x_k} \delta x_k - \frac{1}{\delta g_k^T H_k \delta g_k} H_k \delta g_k \tag{11}$$

δx_k and δg_k are the difference of individuals and gradients of $k+1$th and kth generations.

3.2 Gradient Effect on Algorithm

The gradient information's effect on the proposed hybrid algorithm is checked by the flag; the best individual propagated using gradient search is conserved or replaced by other individual evolved by EA. Figure 5 is the plot of the cost and flag history. The flag is 0 for EA or 1 for Gradient. Figure 6 and 7 are the same plots with figure 5 for the first quarter(Generation 1-500) and last quarter(Generation 1501-2000). Table 1 shows the best individual counts by gradient search and EA quantitatively.

Table 1: Best Individual Counts and Percentage

Generation	1st	2nd	3rd	4th
By Gradient Search	89	331	474	497
By EA	411	169	26	3
Percentage(%)	17.8	66.2	94.8	99.4

It is clear that EA is dominant in early generations, however the gradient search becomes dominant as generation increases. The accuracy of the gradient vector depends on the generation. The search space becomes convex after sufficient generations, which explains the dominance of the gradient individual.

4 Algorithm Summary

Since the new algorithm is based on EA, it has same selection, recombination and mutaion processes of EA. The modification using gradient-based algorithm has made in evaluation and offspring generation processes. Evaluation and offspring generation processes were separated in typical EA. However they are merged in the hybrid method since offspring generation process is divided into two stages and evaluation process is performed between two stages.

1. Generate initial populations.
   (a) Generate $m+1$ individuals and calculate the costs.
   (b) Choose the best individual x_{B0}.
   (c) Generate m more individuals using $x_{m+i} = 2 \cdot x_{B0} - x_i$.
   (d) Calculate the initial gradient $g_0 = (\Delta X^T \Delta X)^{-1} \Delta X^T \Delta f$ and set $H_0 = I$.
2. Select μ parents from total offsprings($2m+1$).
3. Generate the offsprings and evaluate the costs.
   (a) Make m individuals through recombination of μ parents.
   (b) Mutate the individuals and make m offsprings.
   (c) Generate one more offspring using $x_{m+1} = x_{Bk} - H_k \cdot g_k$.
   (d) Evaluate $m+1$ offsprings' cost
   (e) Choose the best one and set it as x_{Bk+1}.
   (f) Generate m more offsprings using $x_{m+i} = 2 \cdot .x_{Bk+1} - x_i$ and evaluate the costs.
4. Calculate g using the least-square method or the weighted least-square method.

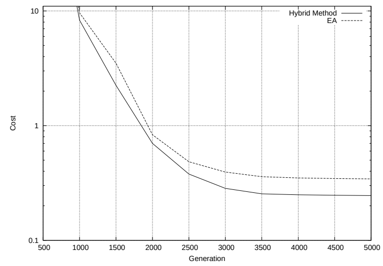


Figure 8: Cost History of Best

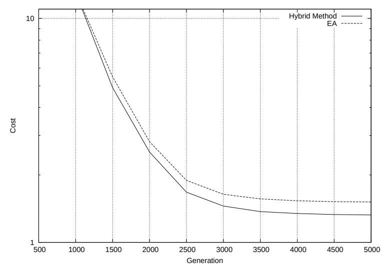


Figure 9: Cost History of Median

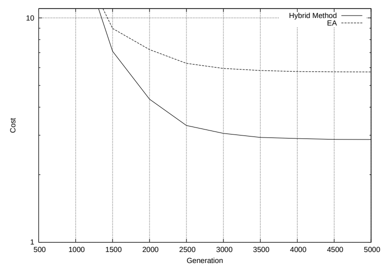


Figure 10: Cost History of Worst

5. Calculate H using the BFGS algorithm.
6. Go to Step 2.

5 Numerical Results

The proposed hybrid method's performance is compared with normal ES using a higher order Rosenbrock function($n = 10$). Both methods were run 20 times each. Figure 8, 9, and 10 are the plots of best, worst and median costs history.

The hybrid method shows better performance in every plots, especially in the plot of worst cost.

6 Conclusion

In this paper, the hybrid evolutionary algorithm is proposed. the hybrid algorithm is based on typical evolutionary algorithm, however it uses gradient search method for accelerating convergence of best individual.

The gradient estimation result that the proposed algorithm has acceptable gradient estimation ability. Moreover the gradient search works dominantly as generation increases.

The hybrid method works better than normal ES in numerical testing with a benchmark problem.

7 Acknowledgements

This study has been supported by 'Smart UAV Develpment - 21C Frontier R&D Project' of Korea Aerospace Research Institute.

References

[1] T. Back, *Evolutionary Algorithms in Theory and Practice,* Oxford University Press, 1996.

[2] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization,* Academic Press, 1981.

[3] A. Gelb, *Applied Optimal Estimation,* M.I.T Press, 1974.

[4] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming : Theory and Algorithms,* John Wiley & Sons, 1993.