Published in Workshop on Evolutionary Computation and Parallel Processing (part of GECCO-2000)

On the Effects of Migration on the Fitness Distribution of Parallel Evolutionary Algorithms

Erick Cantú-Paz
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
Livermore, CA 94551
cantupaz@llnl.gov

Abstract

Migration of individuals between populations may increase the selection pressure. This has the desirable consequence of speeding up convergence, but it may result in an excessively rapid loss of variation that may cause the search to fail. This paper describes the effects of migration on the distribution of fitness. The calculations consider finite populations, arbitrary migration rates, and topologies with different numbers of neighbors. The paper shows that even if different algorithms are configured to produce the same selection intensity, they change the composition of the population in different ways. The results suggest that migration preserves more diversity as the number of neighbors increases.

1 INTRODUCTION

A popular method to parallelize evolutionary algorithms (EAs) is to use multiple populations (also called demes) and allocate each to a different processor. In this method, the populations periodically exchange a few individuals in a process analogous to migration of natural organisms. In EAs, migration is controlled by several parameters: the migration rate that determines how many individuals migrate from a population; the migration frequency that determines how often migrations occur; the migration topology that determines the destination of the migrants; and the migration policy that determines which individuals migrate and which are replaced at the receiving deme. The objective of this paper is to examine how the migration rate and the number of neighbors of a deme affect the distribution of fitness.

In a previous study, Cantú-Paz (In press) determined that the selection pressure depends on the migration policy. The pressure increases when the migrants or the individuals that are replaced at the receiving deme are chosen according to their fitness. The previous study quantified the increase in the selection intensity, and suggested that comparing algorithms with different selection intensities could explain many of the frequent claims of superlinear parallel speedups. A more appropriate comparison would be between serial and parallel algorithms that have the same selection intensity. However, that study also recognized that different algorithms should not be considered equivalent only because they have the same selection intensity. We should also take into account higher order cumulants, such as the variance, in comparisons. After all, one of the anecdotal explanations of superlinear speedups is that the search is better because migration introduces diversity to the demes. This paper extends the previous study and shows that even if different algorithms have the same selection intensity, they change the population in different ways.

This paper is organized into five sections. Section 2 presents the approach used to describe the distribution of fitness. Section 3 reviews the previous results that quantify the selection intensity caused by migration. Section 4 examines the effect of migration on the higher cumulants of the distribution of fitness. Finally, section 5 summarizes the paper.

2 THE FITNESS DISTRIBUTION

The approach of this paper is to describe the distribution of fitness using its cumulants. The cumulants of a distribution are related to its central moments. The r-th central moment of the distribution of fitness of a population of size n is

$$\mu_r = \frac{1}{n}\sum_{i=1}^{n}(f_i - \bar{f})^r, \tag{1}$$

where f_i is the fitness of the i-th individual, and $\bar{f} = \frac{1}{n}\sum_{i=1}^{n} f_i$ is the mean fitness. The first three cumulants are equal to the first central moments. The fourth cumulant is $\kappa_4 = \mu_4 - 3\mu_2^2$.

The first cumulant is the mean ($\kappa_1 = \mu_1 = \bar{f}$), and the second is the variance ($\kappa_2 = \mu_2 = \sigma_f^2$). The third and fourth cumulants give additional information about the shape of the distribution, and sometimes they are divided by $\kappa_2^{r/2}$ to obtain the skewness and kurtosis coefficients. The skewness measures the asymmetry of the distribution; it is negative if a distribution skewed to the left, and it is positive if the distribution is skewed to the right. The kurtosis measures the "peakedness" of the distribution; a negative kurtosis indicates that the distribution is flatter than a normal, and a positive kurtosis indicates that the distribution is more peaked than a normal.

The method of this paper is based on order statistics. The critical observation is that we may interpret the fitness values f_i as samples of random variables F_i with a common distribution. We obtain the order statistics of the F_i variables by arranging them in increasing order:

$$F_{1:n} \leq F_{2:n} \leq \ldots \leq F_{n:n}.$$

Without loss of generality, in the remainder we assume a maximization problem, and we normalize the random variables as follows:

$$Z_{i:n} = \frac{F_{i:n} - \bar{F}}{\sigma_F}. \tag{2}$$

The expected value of the i-th order statistic is

$$\begin{aligned} \mu_{i:n} &= \mathrm{E}(Z_{i:n}) \\ &= n\binom{n-1}{i-1}\int_{-\infty}^{\infty} z\phi(z)[\Phi(z)]^{i-1}[1-\Phi(z)]^{n-i}dz, \end{aligned} \tag{3}$$

where $\phi(z)$ and $\Phi(z)$ are the PDF and CDF of the distribution of fitness, respectively.

For some distributions, we can use tables to find the expected values of the normalized order statistics (Harter, 1970). If we make the assumption that the fitness has a unit Gaussian distribution (with $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$ and $\Phi(z) = \int_{-\infty}^{z}\phi(x)dx$), we can use the following approximation (Harter, 1970):

$$\mu_{i:n} \approx \Phi^{-1}\left(\frac{1-\alpha_i}{n-2\alpha_i+1}\right), \tag{4}$$

where $\Phi^{-1}(x)$ is the inverse of the CDF (i.e., it returns the value of z such that $\Phi(z) = x$), and α_i is defined as

$$\alpha_i = \begin{cases} 0.315065 + 0.057974\log n - 0.009776(\log n)^2 \\ \hfill \text{if } i = 1, \\ 0.327511 + 0.058212\log n - 0.007909(\log n)^2 \\ \hfill \text{otherwise.} \end{cases}$$

Harter (1970) discourages the use of the approximation above for $n > 400$. However, even for much larger populations, the approximation is sufficiently accurate for our purposes (Cantú-Paz, 2000).

3 MIGRATION AND SELECTION INTENSITY

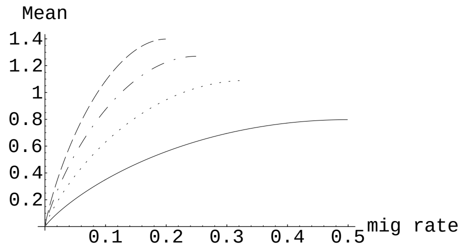
There are two popular choices to select the individuals that migrate: choose them randomly or pick the best individuals in the population. Likewise, there are two popular choices at the receiving deme to replace existing individuals with the incoming migrants: choose them randomly or replace the worst. Migrants or replacements can also be chosen by any selection method (e.g., tournaments, ranking). The point is that when the migrants or replacements are chosen according to their fitness, the selection pressure increases.

We restrict our attention to the case when the best individuals in a deme are selected (deterministically) to migrate, and replace the worst individuals in the receiving deme. This is the most frequently used migration policy in parallel EAs, but other policies can be studied easily adapting the framework of this paper.

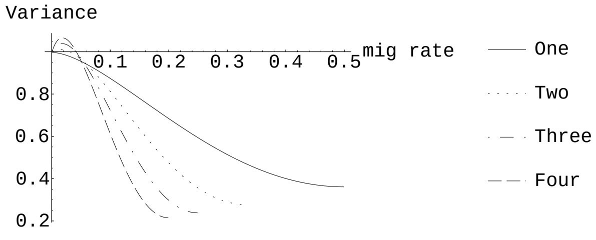
Using δ to denote the number of neighbors of a deme (i.e., the degree of the connectivity graph) and ρ to denote the migration rate (i.e., the fraction of the population that migrates every generation), we can calculate the selection intensity caused by migration as (Cantú-Paz, In press)

$$I_m = \frac{\delta}{n}\sum_{i=n-n\rho+1}^{n}\mu_{i:n} + \frac{1}{n}\sum_{i=\delta\rho n+1}^{n}\mu_{i:n}. \tag{5}$$

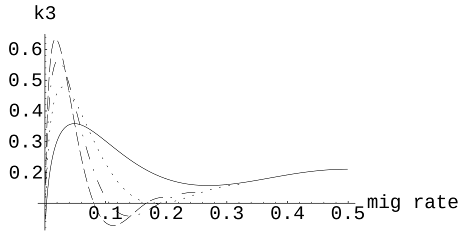
The first term is the selection intensity caused by selecting the emigrants, and the second term is the intensity caused by selecting replacements in the receiving deme. The selection intensity is equal to the mean fitness of the normalized population after migration, which we denote as $\bar{Z}^{\mathrm{mig}} = I_m$. In the next section we use it to calculate the higher cumulants of the fitness distribution.



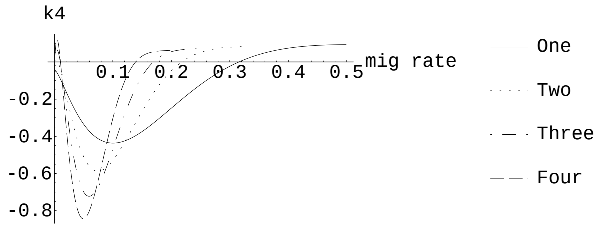
(a) Mean



(b) Variance



(c) κ_3



(d) κ_4

Figure 1: The first four cumulants of the fitness distribution (Eq. 6) varying the migration rate and the number of neighbors.

4 MIGRATION AND THE HIGHER CUMULANTS

The r-th cumulant of the fitness distribution after migration has two components that correspond to the migrants and to the native individuals that are not replaced at the receiving deme:

$$\begin{aligned}\mu_r^{\mathrm{mig}} =& \frac{\delta}{n} \sum_{i=n-n\rho+1}^{n} (\mu_{i:n} - \bar{Z}^{\mathrm{mig}})^r + \\ & \frac{1}{n} \sum_{i=\delta\rho n+1}^{n} (\mu_{i:n} - \bar{Z}^{\mathrm{mig}})^r .\end{aligned} \tag{6}$$

Figure 1 shows plots of the first four cumulants of the distribution of fitness varying the migration rate and the number of neighbors. The maximum migration rate is $1/(\delta+1)$, and the mean increases monotonically with the migration rate. Note that in configurations with more than one neighbor and low migration rates, the variance is higher than the original value of one. This occurs at approximately $\rho = 0.02$ or 0.03, regardless of the number of neighbors.

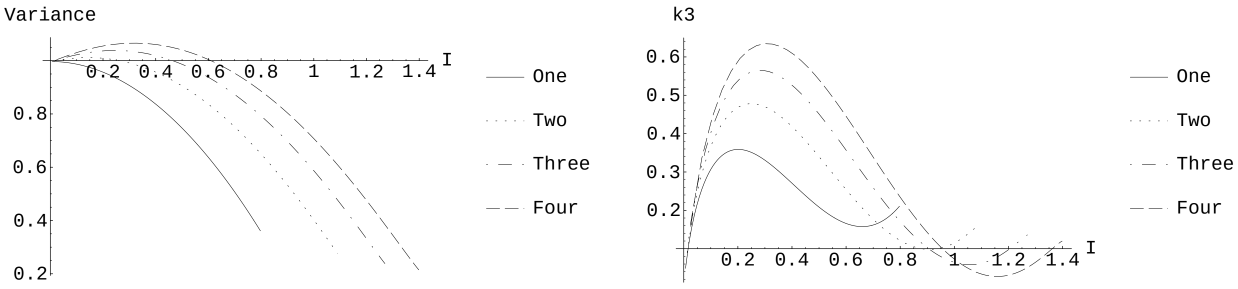
To facilitate comparisons between different algorithms, Figure 2 plots the variance and the third cumulant of the fitness distribution against the selection intensity. This figure clearly shows that different configurations, even if they have the same selection intensity, affect the distribution in different ways.

Preserving (or increasing) the diversity in a deme is desirable for at least two reasons. The first is that the increased diversity will delay the convergence of the algorithm. This may give enough time to the crossover operator to mix BBs together into solutions of high quality[1]. The other effect of preserving diversity in a deme is that it may be possible to evolve partial solutions independently in different demes and integrate them after migration.

5 CONCLUSIONS

The calculations in this paper describe how the migration rate and the number of neighbors of a deme affect the distribution of fitness. We found that increasing the number of neighbors and the migration rate results in a higher selection intensity. However, even when

[1]See the papers by Goldberg et al. (1993) and Thierens and Goldberg (1993) for a discussion on the time required to mix BBs into good solutions and the relation of this "innovation time" with the success of the search.



(a) Variance

(b) κ_3

Figure 2: The second and third cumulants of the fitness distribution in configurations with the same selection intensity.

parallel EAs are configured to have the same selection intensity, they modify the population in different ways.

In cases with more than one neighbor and low migration rates, the variance after migration is higher than its original value. This is desirable, but we must be cautious, because the increase in variance comes together with a relatively skewed distribution and higher cost of communications. Also, we must consider that this paper deals with the distribution of the fitness values, not the distribution of alleles, and it is not clear if the increase in fitness variance translates to improvements in search quality.

In any case, this paper represents a step toward a better understanding of the effects of migration on the population. Migration acts as a form of selection: It reduces the variance and biases the population towards specific types of individuals. The results of this paper can be used to choose appropriate genetic operators that ensure that the algorithm does not converge prematurely because of lack of diversity.

Acknowledgments

UCRL-JC-138729. This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48.

References

Cantú-Paz, E. (2000). *Comparing selection methods of evolutionary algorithms using the distribution of fitness.* Unpublished manuscript.

Cantú-Paz, E. (In press). Migration policies, selection pressure, and parallel evolutionary algorithms. *Journal of Heuristics*.

Goldberg, D. E., Deb, K., & Thierens, D. (1993). Toward a better understanding of mixing in genetic algorithms. *Journal of the Society of Instrument and Control Engineers*, *32*(1), 10–16.

Harter, H. L. (1970). *Order statistics and their use in testing and estimation.* Washington, D.C.: U.S. Government Printing Office.

Thierens, D., & Goldberg, D. E. (1993). Mixing in genetic algorithms. In Forrest, S. (Ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms* (pp. 38–45). San Mateo, CA: Morgan Kaufmann.