EEDA : A New Robust Estimation of Distribution Algorithms

Michael Wagner, Anne Auger, Marc Schoenauer

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

EEDA: A New Robust Estimation of Distribution Algorithms

Michael Wagner — Anne Auger — Marc Schoenauer

N° 5190

May 2004

Thème COG

ISSN 0249-6399 ISRN INRIA/RR--5190--FR+ENG

EEDA: A New Robust Estimation of Distribution Algorithms *

Michael Wagner †, Anne Auger ‡, Marc Schoenauer §

Thème COG — Systèmes cognitifs
Projet TAO

Rapport de recherche n° 5190 — May 2004 — 16 pages

Abstract: In this report we address a subtle but important limitation found in the literature for Estimation of Distribution Algorithms (EDAs): symmetric initializations of the EDAs around the optimal solution. We focus our study on the performance of certain EDAs ($\text{EMNA}_{\text{global}}$ and PBIL_{C}) that are asymmetrically initialized far from the optimum. We show and explain the failure of these EDAs under these conditions. These observations lead us to develop a new EDA based on an eigenspace analysis, which we denote by EEDA (Eigenspace EDA). We conclude by analyzing this new EDA and by showing its strengths when compared with $\text{EMNA}_{\text{global}}$ and PBIL_{C} when the optimal solution is unknown.

Key-words: Optimization, Artificial Evolution, Distributions

* Research undertaken at and supported by the Institut National de Recherche en Informatique et en Automatique (INRIA), Rocquencourt, France.

† Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, mikew@mit.edu

‡ Now at Equipe TAO – INRIA Futurs, LRI - Bat. 490, Université de Paris Sud,1405 Orsay Cedex, France, anne.auger@inria.fr – formerly at Projet Fractales – INRIA Rocquencourt

§ Now at Equipe TAO – INRIA Futurs, LRI - Bat. 490, Université de Paris Sud,1405 Orsay Cedex, France, anne.auger@inria.fr – formerly at Projet Fractales – INRIA Rocquencourt

Unité de recherche INRIA Futurs
Parc Club Orsay Université, ZAC des Vignes,
4, rue Jacques Monod, 91893 ORSAY Cedex (France)
Téléphone : +33 1 72 92 59 00 —Télécopie : +33 1 72 92 59 ??

EEDA: un algorithme d'estimation de distribution robuste.

Résumé : Une faiblesse importante passée inaperçue de certains algorithmes d'estimation de distributions (EDAs) est leur fragilité par rapport à l'initialisation de la première distribution, souvent choisie symétrique et centrée sur la solution cherchée. Nous montrons que lorsque l'on initialise par exemple les algorithmes $\text{EMNA}_{\text{global}}$ et PBIL_{C} loin de la solution et de manière asymétrique, les performances de l'algorithme chutent. Après une analyse des raisons des ces échecs, nous proposons EEDA, une variante de $\text{EMNA}_{\text{global}}$ basée sur l'analyse des valeurs propres de la matrice de covariance. Enfin, nous montrons expérimentalement que cet algorithme est effectivement plus performant que $\text{EMNA}_{\text{global}}$ et PBIL_{C} dans le cas d'initialisation asymétrique non centré, tout en restant équivalent dans le cas symétrique.

Mots-clés : Optimisation, Évolution Artificielle, Distributions

1 Introduction

The class of algorithms referred to as Estimation of Distribution Algorithms (EDAs) was developed from standard evolutionary algorithms, the difference being that a distribution on the space of solutions is evolved as opposed to an actual population of solutions. The focus of this report is strictly limited to EDAs, specifically a subtle limitation that is prevalent in the literature concerning the initialization of these algorithms. We will specifically be analyzing EDAs applied to optimizing functions of the type $f : \mathbb{R}^n \to \mathbb{R}$; these functions will be referred to as fitness functions.

Though the study of EDAs is relatively new, there already exists a plethora of different types of EDAs. We shall limit ourselves to mention only the types of EDAs specifically utilized in this report. Binary and continuous versions of Population-Based Incremental Learning (PBIL) EDAs have been designed and analyzed by Baluja [1] and Sebag and Ducoulombier [3], respectively. However, the PBIL algorithms do not include any dependencies between variables since the underlying distribution structure is the product of univariate Gaussian distributions. The Estimation of Multivariate Normal Algorithm (EMNA) class of EDAs, presented by Larrañaga and Lozano [2], utilizes a multivariate Gaussian distribution structure which specifically allows the exploitation of inter-dependencies between variables. To present a concise study, we have chosen two representative EDAs from these two classes: continuous PBIL (PBIL_C) and a standard EMNA EDA ($\text{EMNA}_\text{global}$).

We have chosen the variant of PBIL_C which performed the best, according to the results in [3]. As previously mentioned, the underlying structure of PBIL_C is a multivariate normal distribution which is the product of n (the dimension of the problem) independent univariate normal distributions. In other words, PBIL_C evolves the distribution $\prod_{i=1}^{n} f_i(x_i)$, where $f_i(x_i)$ are gaussian densities (of varying parameters) and $x = (x_1, \ldots, x_n)$ is the vector optimization variable. At each iteration k, the mean and standard deviation of each of the univariate distributions are updated in the following ways, where μ and σ are the vectors of means and standard deviations, respectively:

$$\mu^{k+1} = (1-\alpha)\mu^k + \alpha(\mu^k_{\text{best}} + \mu^k_{\text{2nd best}} - \mu^k_{\text{worst}}),$$

where μ^k_{best}, $\mu^k_{\text{2nd best}}$ and μ^k_{worst} are the values of the best, second best and worst individuals (with respect to the fitness function) at iteration k. The update rule for each of the standard deviations $\sigma_i,\ i = 1, \ldots, n$ is

$$\sigma_i^{k+1} = (1-\alpha)\sigma_i^k + \alpha\sqrt{\frac{\sum_{j=1}^{K}(\tilde{\mu}_i^j - \tilde{\mu}_i)^2}{K}},$$

where $\tilde{\mu}^j,\ j = 1, \ldots, K$ are the K best individuals (with respect to the fitness function) and $\tilde{\mu}$ is their mean. For the remainder of this report, K is chosen to be half of the population size. For both update rules, the α parameter is referred to as the learning rate.

As for the EMNA class of EDAs, we have chosen the simplest, $\text{EMNA}_\text{global}$, which also has good practical performance, as presented in [2]. The pseudo-code of this EDA is as follows:

1. At generation 0, we are given a multivariate Gaussian distribution of mean μ^0 and covariance matrix Σ^0. Generate a population of N individuals.
2. For $k \geq 0$,
   (a) At generation $k+1$, select the best $M < N$ individuals (with respect to fitness function) from generation k's population.
   (b) Form a new multivariate Gaussian distribution using the calculated mean μ^{k+1} and covariance matrix Σ^{k+1} of the M selected individuals.
   (c) Generate a new population of N individuals.

The rest of this report is organized as follows: in Section 2 we discuss important limitations found in the literature; in Section 3 we show that these limitations cause the previously mentioned EDAs to fail; in Section 4 we develop our new EDA, EEDA; in Section 5 we present a rigorous computational comparison of $\text{EMNA}_{\text{global}}$, PBIL_{C} and EEDA, showing the strengths of the latter; in Section 6 we present computational studies of EEDA's behavior as certain parameters are varied and we conclude in Section 7.

2 Subtle Limitations

We have found that the EDA literature contains computational studies that lack a complete description. In particular, the references cited in this report do not detail the initial distributions used in their computational studies. In our own studies, these EDAs succeeded only when the initial distributions were initialized symmetrically around the optimal solution. For example, if the optimal solution were at the zero vector, a possible symmetric initial distribution would be a uniform distribution over the unit ball (of appropriate dimension), centered at the zero vector. We have intentionally skewed our initial distribution to preclude this simplification: for the example of the zero vector being optimal, we would initialize on a unit ball centered at a point far from the origin. It will be seen that asymmetrical initializations cause well known EDAs to fail.

Another (less serious) limitation we have noticed is that the majority of the test functions in the EDA literature have the zero vector as the optimum. At times, it can be unclear whether the EDA in question actually solved the optimization problem or only just converged to the zero vector (leading to a false positive run). To eliminate this possibility, we have modified common test functions such that the solution is now at a nonzero value. The functions that we shall employ in our computational studies are:

$$
\begin{aligned}
f_1(x) &= \frac{100}{10^{-5} + \sum_{i=1}^{n} |y_i|}, \text{ where } y_i = (x_i - i + 1) + y_{i-1} \ (y_1 = x_1) \\
f_2(x) &= \sum_{i=1}^{n} (x_i - i + 1)^2 \\
f_3(x) &= 1 + \sum_{i=1}^{n} (x_i - i + 1)^2 - \prod_{i=1}^{n} \cos \frac{(x_i - i + 1)}{\sqrt{i+1}}.
\end{aligned}
$$

We maximize $f_1(x)$ and minimize $f_2(x)$ and $f_3(x)$. The optimum of all three of the above functions occurs at $x^* = (0, 1, \ldots, n-1)$. The optimal values of these functions are $f_1(x^*) = 10^7$, $f_2(x^*) = 0$ and $f_3(x^*) = 0$.

We address these issues due to the fact that in practice, these assumptions and simplifications would not hold in general. The main motivation of this report is that we wish to get a realistic view of the performance of EDAs and avoid false hope.

3 The Failure of EMNA$_{\text{global}}$ and PBIL$_{\text{C}}$

We have tested two EDAs, EMNA$_{\text{global}}$ and PBIL$_{\text{C}}$ (detailed in Section 1), under the generalizations we have mentioned. While the nonzero optimal solution didn't appear to affect the convergence, the asymmetrical initialization resulted in a premature convergence to a non-optimal value. Under a slight perturbation of the initialization, EMNA$_{\text{global}}$ failed. While more robust, PBIL$_{\text{C}}$ also failed under a more pronounced shift in the initial distribution. Please see the appendix for the validation of our implementations of EMNA$_{\text{global}}$ and PBIL$_{\text{C}}$.

We first examine EMNA$_{\text{global}}$. We borrow from [2] the parameters that resulted in success and modified the initial distribution. The parameters that were not changed are the dimension $n = 10$, population size $N = 2000$, sample size $M = 1000$, and untranslated versions of functions $f_1(x)$, $f_2(x)$, and $f_3(x)$ (where the optimum occurs at the zero vector), which we denote by $\tilde{f}_1(x)$, $\tilde{f}_2(x)$, and $\tilde{f}_3(x)$. We introduced an asymmetrical initialization by choosing $\mu^0 = (2, \ldots, 2)$ and $\Sigma^0 = I$, the identity matrix. The termination criteria was set to 76,000 function evaluations. We present the median and standard deviation of the final value of the fitness function in an ordered pair format (median, standard deviation), which are calculated from 20 runs. Please see Table 1.

$n = 10$	$\tilde{f}_1(x)$	$\tilde{f}_2(x)$	$\tilde{f}_3(x)$
EMNA$_{\text{global}}$	(1.412, 0.0236)	(15.966, 0.330)	(17.003, 0.662)

Table 1: Performance analysis for EMNA$_{\text{global}}$ under an asymmetrical initialization, demonstrating failure.

We notice that the results for $\text{EMNA}_{\text{global}}$ are not optimal. While some might argue that longer runs (with respect to a larger number of evaluations allowed) might result in a optimal solution, any prolongation of the studies resulted in numerical errors due to the covariance matrix approaching the zero matrix. Furthermore, the relative magnitudes of the termination value of the fitness function compared with the termination standard deviation of these values provide evidence that the optimal solution will not be reached. One reason for the premature convergence of $\text{EMNA}_{\text{global}}$ is that it was simply not close enough to converge: if $\text{EMNA}_{\text{global}}$ starts sufficiently far away from the solution, each step toward the solution results in a distribution with less variance (since the distributions at each iteration are formed from a finite sample); information is lost at each iteration. This EDA simply exhausts its information (i.e., the covariance matrix prematurely converges to the zero matrix) and further progress is impossible.

We next consider a similar perturbation to PBIL_{C} and show that it also fails. We now borrow from [3] the parameters that resulted in success and again modify the initial distribution. We retained the dimension $n = 100$, population size $N = 50$, sample size $K = 25$, and learning rate $\alpha = 0.01$. We introduced a modified initial distribution with mean $\mu^0 = (25, \ldots, 25)$ and standard deviations $\sigma_i^0 = 1$, $\forall i$. To be absolutely sure of the failure of PBIL_{C} and eliminate the possibility that we did not let the algorithm run for a sufficiently long time, we have extended the termination criteria to 1,000,000 function evaluations. Again, we present the median and standard deviation in an ordered pair format (median, standard deviation), which are calculated from 5 runs. We also present comparable results, retaining the majority of the above parameters and only changing the learning rate to $\alpha = 0.1$. For this case, we lower the termination criteria to 200,000 evaluations and present the statistics for 5 runs. Please see Table 2.

$n = 100$	$\tilde{f}_1(x)$	$\tilde{f}_2(x)$	$\tilde{f}_3(x)$
PBIL_{C} ($\alpha = 0.01$)	(0.0043, 0.0001)	(4464.17, 133.812)	(4456.09, 44.527)
PBIL_{C} ($\alpha = 0.1$)	(0.004, 0.0003)	(5920.41, 185.933)	(5812.23, 369.000)

Table 2: Performance analysis for PBIL_{C} under an asymmetrical initialization, demonstrating failure.

We see that PBIL_{C} also failed to converge to optimality under similar, but more extreme, initial conditions to those that we used in initializing $\text{EMNA}_{\text{global}}$. Given the very liberal termination condition, there is no doubt that PBIL_{C} has failed. The differences in magnitude between the final fitness function value and the corresponding standard deviations are further evidence that success is not possible under these initial conditions. We do note that PBIL_{C} is more robust than $\text{EMNA}_{\text{global}}$ – given the initial conditions that caused $\text{EMNA}_{\text{global}}$ to fail, PBIL_{C} easily converged to the optimal solution. A plausible reason for this greater robustness is that the update rules for PBIL_{C} contain memory; while the new sample might have little variety, the overall variance is maintained from the previous generation. The small value of the learning rate $\alpha = 0.01$ further adds to the robustness. However, the learning

rate only delayed the premature convergence to a non-optimal solution; while varying α, the critical parameters of the initial distribution for failure remained the same; c.f. Table 2.

A simple statement useful for understanding these failures is the following: If the initial distribution is centered more than a few standard deviations from the optimal solution (the number of standard deviations varying by EDA), failure will occur. There is no mechanism for boosting variance. We do note, that for both $\text{EMNA}_{\text{global}}$ and PBIL_{C}, if we initialize with a distribution that has a sufficiently large variance component, these EDAs will successfully converge. We shall provide some computational results concerning unequal initializations in our comparative study in Section 5. In practice the issue also arises concerning the question of what is a sufficient quantity of variance with which to initialize the EDA. In the sequel, we will see that our new EDA will not be encumbered by any such deliberations.

4 The Eigenspace EDA: EEDA

As was previously mentioned, $\text{EMNA}_{\text{global}}$ will converge prematurely if initialized with an asymmetrical distribution with respect to the optimum. However, at a given iteration there is information that can be exploited. The nature of this information is best explained with an example and its corresponding illustration, Figure 1.

Suppose we want to minimize the standard sphere function $\tilde{f}_2(x)$ in two dimensions. Consider a multivariate normal distribution centered far from the optimal solution; in Figure 1, the optimum is at (0,0) and we consider a normal distribution centered at (2,2) with the identity matrix as its covariance matrix. We create a population from this distribution (the white and black circles in the figure) and select the sample of the best individuals with respect to the sphere fitness function (the black circles). We notice that the shape of this sample is skewed; it is this information that we will exploit. Note the level set curve as well as its linear approximation in the figure. Further note that the sample is elongated along the linear approximation of the level set curve. As we do not want to trace the level set and we would rather descend against it, the direction in the sample where there is the least variance corresponds to the direction perpendicular to the level set – a descent direction of cost decrease, which in this case is the negative of the gradient of the fitness function.

While the preceding example was a simplified two dimensional abstraction, the idea is useful in arbitrary dimension. In dimension n, we again attain a sample which we expect to be skewed in the sense that there is little variance in a direction of cost decrease. A simple way to find a finite set of candidate directions with which to choose a descent direction is to calculate an eigenvalue decomposition of the covariance matrix of the sample of the best M individuals. Then, we simply choose the eigenvector with the smallest eigenvalue; *among the eigenvector directions*, the eigenvector with the smallest eigenvalue will correspond to the direction in the sample having the least variance. We choose this eigenvector as our descent direction.

In the general setting, this direction might not correspond to the negative gradient as our simple example did; the direction that we will use will in general be different than the negative of the gradient. Simple computational studies have confirmed this statement. We

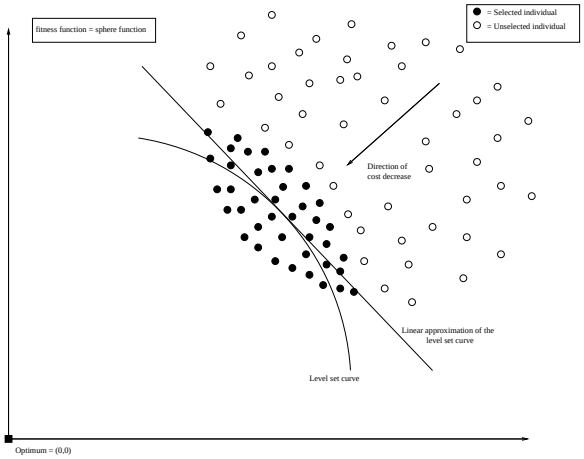

Figure 1: A 2-dimensional illustration of the central idea behind EEDA.

have examined EEDA in dimensions $n = 2$ and $n = 10$, starting from $\mu^0 = (100, \ldots, 100)$ and $\Sigma^0 = I$. We present the median, mean and the standard deviation of the angle between these two directions; these statistics are taken from 20 runs, each of which consisted of 10,000 function evaluations. We calculate the angle between these vectors using the definition of the Euclidean inner product: $x \cdot y = \|x\|\|y\| \cos\theta$, where θ is the angle between vectors x and y. Please see Table 3.

	$n = 2$	$n = 10$
median	8.375	81.308
mean	11.858	82.126
standard deviation	12.048	26.345

Table 3: Statistics for the angle between the eigenspace extension direction and the true gradient.

We note that the underlying principle of our development is that we only need to find a direction of cost decrease; finding the exact gradient would usually be too costly or impossible.

4.1 Modifying $\text{EMNA}_{\text{global}}$ to create EEDA

Once we have found the eigenvector with the minimum eigenvalue, we exploit this information by modifying $\text{EMNA}_{\text{global}}$ in order to create the Eigenspace EDA (EEDA). We retain the majority of the structure of $\text{EMNA}_{\text{global}}$ but modify it in the following way: **At each**

iteration, we redefine the sample covariance matrix by *extending* the original matrix in the direction of the eigenvector corresponding to the smallest eigenvalue. Specifically, in the eigenvalue decomposition, we reset the minimum eigenvalue to the value of the maximum eigenvalue. The final description of EEDA is presented in the following pseudo-code:

1. At generation 0, we are given a multivariate Gaussian distribution of mean μ^0 and covariance vector Σ^0. Generate a population of N individuals.
2. For $k \geq 0$,
   (a) At generation $k+1$, select the best $M < N$ individuals (with respect to fitness function) from generation k's population.
   (b) Calculate the mean μ^{k+1} and covariance matrix Σ^{k+1} of the M selected individuals.
   (c) **Calculate an eigenvalue decomposition of the covariance matrix:** $\Sigma^{k+1} = V\Lambda V^T$.
   (d) **Perform an extension as defined above:** $\Sigma^{k+1} = V\Lambda V^T + (\lambda_{max} - \lambda_{min}) v_{min} v_{min}^T$.
   (e) Generate a new population of N individuals.

We call the new steps an eigenspace extension and the algorithm the Eigenspace EDA (EEDA). Also, by resetting the minimum eigenvalue to the value of the maximum eigenvalue, we still have a framework for convergence.

5 A Comparative Study

We now present an computational comparative study of EEDA, PBIL_C, and $\text{EMNA}_\text{global}$. We define our termination condition as $10,000$ function evaluations. For each EDA and set of parameters, we have the data for 20 independent runs. We present the *median* value of the fitness function over the 20 runs at the termination of the computation as well as the corresponding standard deviation in the form of an ordered pair: (median, standard deviation). We choose the population size $N = 40$[1] and sample sizes $K = M = 20$. We set the dimension $n = 10$. The initial covariance matrix Σ^0 for EEDA and $\text{EMNA}_\text{global}$ is the identity matrix I and PBIL_C's initial distributions have unit variance. PBIL_C's learning rate α is fixed at 0.1. We vary the fitness function as well as the initial mean μ^0. Please see Tables 4 and 5; the best results are in bold.

An F superscript denotes results contaminated by numerical problems; these problems occurred exclusively for $\text{EMNA}_\text{global}$. The reason for these numerical errors is that

[1]Some might argue that $\text{EMNA}_\text{global}$ requires a larger population to function properly. However, this objection is moot, as we have already shown the failure of $\text{EMNA}_\text{global}$ with a large population under less extreme initial conditions; c.f. the section entitled "The Failure of $\text{EMNA}_\text{global}$ and PBIL_C."

$\mu^0 = (10, \ldots, 10)$	$f_1(x)$	$f_2(x)$	$f_3(x)$
EEDA	(86.396, 110.038)	**(1.830e-26, 3.364e-23)**	**(0, 0)**
PBIL_C	**(244.984, 56.846)**	(0.0001, 7.546e-05)	(0.0002, 9.009e-05)
$\text{EMNA}_\text{global}$	$(0.295, 0.006)^F$	(314.036, 128.201)	(285.014, 159.722)

Table 4: Performance analysis for EEDA, PBIL_C and $\text{EMNA}_\text{global}$ in dimension 10 under a slight asymmetrical initialization.

$\mu^0 = (100, \ldots, 100)$	$f_1(x)$	$f_2(x)$	$f_3(x)$
EEDA	**(6.653, 4.437)**	**(1.226e-19, 5.188e-17)**	**(0, 2.419e-17)**
PBIL_C	(0.024, 0.0003)	(53357.7, 1292.2)	(52358.6, 1096.5)
$\text{EMNA}_\text{global}$	$(0.018, 3.275\text{e-}05)^F$	(89938.4, 44902.8)	$(0, 45041.778)^F$

Table 5: Performance analysis for EEDA, PBIL_C and $\text{EMNA}_\text{global}$ in dimension 10 under a pronounced asymmetrical initialization.

$\text{EMNA}_\text{global}$'s covariance matrix prematurely converged to zero, and the subsequent iterations encountered errors. These errors are insignificant since they only resulted after the EDA converged to a non-optimal solution.

We now explore the dependence on dimension, by repeating the above experiment with a few minor changes: we enlarge the termination criteria to $50,000$ function evaluations and enlarge the dimension to $n = 30$. We also enlarge the population size to $N = 80$ and sample sizes to $K = M = 40$. All other parameters remain the same. Please see Tables 6 and 7; again, the best results are in bold.

$\mu^0 = (10, \ldots, 10)$	$f_1(x)$	$f_2(x)$	$f_3(x)$
EEDA	(4.005, 1.186)	(0.00055, 0.0016)	(0.0012, 0.0020)
PBIL_C	**(31.503, 8.014)**	**(3.014e-05, 6.114e-06)**	**(3.521e-05, 6.745e-06)**
$\text{EMNA}_\text{global}$	$(0.075, 0.0003)^F$	$(2669.13, 26.016)^F$	$(2628.7, 19.05)^F$

Table 6: Performance analysis for EEDA, PBIL_C and $\text{EMNA}_\text{global}$ in dimension 30 under a slight asymmetrical initialization.

5.1 Analysis of the computational results

We begin by analyzing the results for dimension 10. Even when the initial distribution is centered at nearly the optimal solution ($\mu_0 = (10, \ldots, 10)$), EEDA performed well, being outperformed only by PBIL_C on $f_1(x)$. On the other two functions, EEDA greatly outperformed the other two EDAs. While the optimal solution was attained (within a numerical

$\mu^0 = (100, \ldots, 100)$	$f_1(x)$	$f_2(x)$	$f_3(x)$
EEDA	**(0.331, 0.152)**	**(0.212, 0.804)**	**(0.203, 0.370)**
PBIL_C	(0.004, 6.948e-05)	(77039.7, 2919.8)	(77266.6, 2317.107)
$\text{EMNA}_\text{global}$	$(0.0023, 1.556\text{e-}06)^F$	$(219900, 263.043)^F$	$(219828, 269.352)^F$

Table 7: Performance analysis for EEDA, PBIL_C and $\text{EMNA}_\text{global}$ in dimension 30 under a pronounced asymmetrical initialization.

error) by both EEDA and PBIL_C on $f_2(x)$ and $f_3(x)$, these algorithms failed to attain the optimal solution for $f_1(x)$. The reason for this is simply the cutoff of $10,000$ evaluations; given a larger cutoff criteria, under these current conditions, both algorithms eventually converged to the optimal solution (PBIL_C more quickly than EEDA). We note that $\text{EMNA}_\text{global}$ failed under all three functions; this is also the case if we enlarge the cutoff criteria. With even a slight disturbance from a symmetrical initialization, $\text{EMNA}_\text{global}$ fails.

Under the more extreme initialization of $\mu_0 = (100, \ldots, 100)$, we see that EEDA clearly outperforms each of the other two EDAs for each function. Both $\text{EMNA}_\text{global}$ and PBIL_C failed under these conditions, even if more evaluations were allowed. We clearly see the benefit of EEDA in these cases.

We now examine the results in dimension 30. When the initial distribution is relatively close to the optimal solution ($\mu_0 = (10, \ldots, 10)$), EEDA is only slightly outperformed by PBIL_C for all three functions; this shows that EEDA does not need to be initialized far from the optimum to be effective. Again, under this slight disturbance, $\text{EMNA}_\text{global}$ fails.

Once we consider a more skewed initialization, we again see that EEDA greatly outperforms the other two EDAs on all test functions. Once again, the other two EDAs converged to a non optimal solution. We note that even EEDA is not undeniably at the optimal solutions; this is only due to the number of evaluations cutoff, which if increased, allows EEDA to get within more significant digits of the optimal solution.

We conclude this section with two plots of the rates of convergence of the medians of the three EDAs, corresponding to function $f_2(x)$, in dimension $n = 10$, with $\mu^0 = (10, \ldots, 10)$. Figure 2 details the study we have just completed, where the initializations are identical for each EDA. In Figure 3, we attempt to present a more equitable comparison, by differing the initializations for $\text{EMNA}_\text{global}$ and PBIL_C in their favor. For $\text{EMNA}_\text{global}$, we enlarge the population size to $N = 2000$, sample size to $M = 1000$ and initial covariance matrix to $\Sigma^0 = 100I$. Under these conditions, $\text{EMNA}_\text{global}$ is in fact successful, converging to the optimal solution. For PBIL_C, we only enlarge the initial standard deviations to $\sigma_i^0 = 10$, $\forall i$, to correspond with the initialization of $\text{EMNA}_\text{global}$. We note that while PBIL_C was successful in both cases, the change in the initial standard deviations didn't affect the convergence rate very much; in fact, the rate slightly decreased, supporting the claim that a larger initial variance will delay convergence. The reasons we have modified the initializations as we have are as follows: (1) $\text{EMNA}_\text{global}$ is tested in the literature for large population sizes and was likely designed with these sizes in mind; for this reason, we have enlarged its population

size; (2) initializations of both $\text{EMNA}_{\text{global}}$ and PBIL_{C} in the literature are vague and sometimes only mention "uniform initializations;" thus a large initial variance component will produce a population covering a larger portion of the search space (approximating the "uniform initialization") and, under the current conditions, would produce some individuals near the optimal solution with high probability. We stress that EEDA remained initialized with $\Sigma^0 = I$. The point to note is that even by giving $\text{EMNA}_{\text{global}}$ and PBIL_{C} much more initial variance (and larger populations in the case of $\text{EMNA}_{\text{global}}$), EEDA still outperformed them.

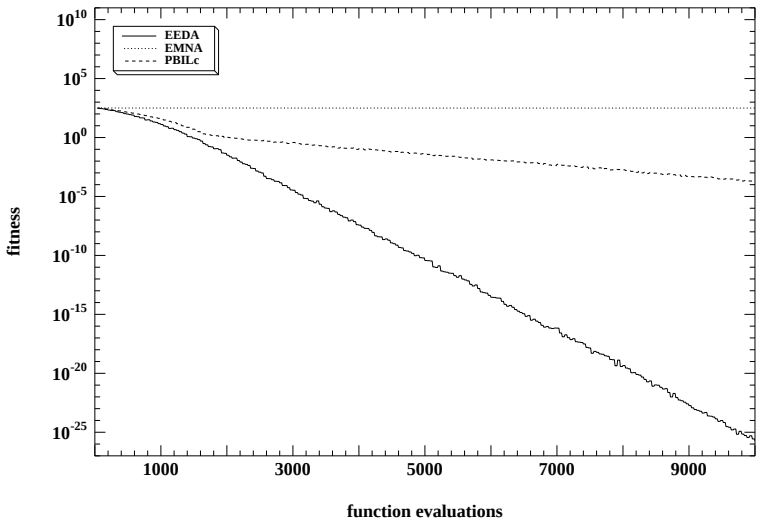

Figure 2: Evolution of convergence rates of EEDA, PBIL_{C} and $\text{EMNA}_{\text{global}}$ under identical initializations.

6 A Study of EEDA

We now solely study the behavior of EEDA with respect to the size of the underlying populations as well as the degree of skewness of the initial distribution.

6.1 Population size

We begin this section with a study concentrating on the behavior of EEDA with respect to the population size. We consider the following situation: dimension $n = 10$, termination cutoff of 100,000 function evaluations on each of 20 runs, an initial mean $\mu^0 = (100, \ldots, 100)$, and an initial covariance matrix $\Sigma^0 = I$. We vary the population size, considering $N \in$

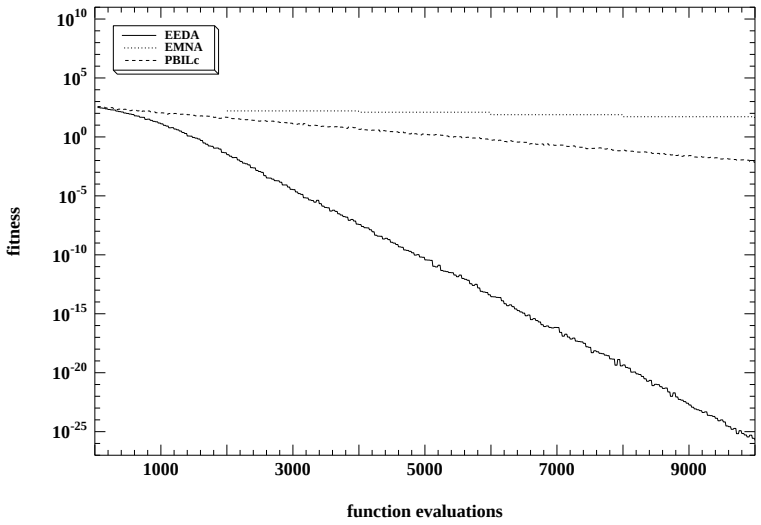

Figure 3: Evolution of convergence rates of EEDA, PBIL_C and $\text{EMNA}_\text{global}$ under different initializations, giving both PBIL_C and $\text{EMNA}_\text{global}$ larger initial variances and $\text{EMNA}_\text{global}$ a larger population.

$\{40, 400, 4000\}$. The sample size M is fixed at one half of the population size, in accordance with preceding tests. Please see Table 8 for medians and standard deviations.

EEDA	$f_1(x)$	$f_2(x)$	$f_3(x)$
$(N = 40, M = 20)$	(36.230, 5949.895)	(0, 5.242e-34)	(0, 0)
$(N = 400, M = 200)$	(1.772, 1.053)	(1.856e-23, 1.808e-22)	(0, 0)
$(N = 4000, M = 2000)$	(0.0207, 5.326e-05)	(76143.3, 343.002)	(76080.8, 295.580)

Table 8: Performance analysis for EEDA in dimension 10 as a function of underlying population size.

We now examine the results of the computational study on the behavior of EEDA as a function of the underlying population size. We first note that EEDA requires a relatively small population to function properly. Also, the smaller populations resulted in better performance. In theory, we only need the sample of this population to provide us with a reasonable estimate of the level set of the fitness function.

6.2 Initial distribution

We now present a study of the performance of EEDA as the initial distribution is varied. We concentrate on the skewness of the initial distribution; i.e., we vary the degree of the asymmetry of the initial mean. We retain the initial identity covariance matrix $\Sigma^0 = I$ and vary the initial mean $\mu^0 = (10^m, \ldots, 10^m)$, where $m \in \{1, 2, 3, 4, 5, 6\}$. We conduct this study in dimension $n = 10$ with a termination condition of 20,000 function evaluations in each of 20 runs. Please see Table 9 for medians and standard deviations.

EEDA	$f_1(x)$	$f_2(x)$	$f_3(x)$
$\mu^0 = (10^1, \ldots, 10^1)$	(484.303, 2426.322)	(2.031e-198, 3.722e-33)	(0, 0)
$\mu^0 = (10^2, \ldots, 10^2)$	(38.877, 80.535)	(3.146e-143, 5.079e-33)	(0, 0)
$\mu^0 = (10^3, \ldots, 10^3)$	(3.407, 6.774)	(8.832e-115, 1.750e-33)	(0, 0)
$\mu^0 = (10^4, \ldots, 10^4)$	(0.313, 0.854)	(3.80919e-72, 1.106e-32)	(0, 0)
$\mu^0 = (10^5, \ldots, 10^5)$	(0.036, 0.085)	(8.8553e-31, 8.012e-29)	(0, 0)
$\mu^0 = (10^6, \ldots, 10^6)$	(0.002, 0.002)	(3.63196e-25, 3.138e-19)	(0, 0)

Table 9: Performance analysis for EEDA in dimension 10 as a function of the degree of asymmetry of the initial distribution.

We conclude by analyzing the behavior of EEDA as a function of the initial distribution. On functions $f_2(x)$ and $f_3(x)$, EEDA performs excellently, even for an extremely skewed initial distribution. We do note that EEDA does not maintain the same quality of solution on $f_1(x)$. A plausible reason for this is the following: Populations far from the optimal solution of $f_1(x)$ do not contain much useful information as $f_1(x)$'s level sets are very flat far from the optimal solution; EEDA has a difficult time approximating the level sets.

7 Conclusion

In this report we have addressed a serious limitation that we have found in the literature concerning the initialization of EDAs. While some might consider this issue trivial, as it is not even detailed in the papers we have referenced, it is in fact quite important. By varying the initial distribution of $\text{EMNA}_{\text{global}}$ and PBIL_{C}, we demonstrate their failure as they converged to non-optimal solutions.

Using observations from the failures of the previous EDAs, we have designed a new EDA which we have denoted the Eigenspace EDA (EEDA). We have compared EEDA with the previously mentioned EDAs and have shown that its performance, with respect to the other EDAs, improves as the skewness of the initialization increases. We concluded our computational studies by solely analyzing EEDA as a function of various parameters.

The main strength of this new method is twofold: (1) We exploit information at each iteration in order to find a direction in which to proceed and (2) we have an inherent mechanism for boosting the variance of the underlying distribution if there is still a distance

to proceed to the optimal solution. These two properties make EEDA quite robust in its performance. In particular, it succeeds in finding the optimal solution under conditions where the other EDAs we examined failed. A possibility for further research would be to investigate more sophisticated algorithms that aim to improve EEDA's basic strengths.

8 Acknowledgements

The first author was visiting Projet Fractales at INRIA Rocquencourt during Summer 2003 when this work has been done, and wishes to thank the MIT-INRIA collaboration program that partly funded this internship.

References

[1] S. Baluja. An empirical comparison of seven iterative and evolutionary function optimization heuristics. Technical Report CMU-CS-95-193, Carnegie Mellon University, 1995.

[2] Pedro Larrañaga and Jose A. Lozano. *Estimation of Distribution Algorithms.* Kluwer Academic Publishers, 2002.

[3] Michéle Sebag and Antoine Ducoulombier. Extending population-based incremental learning to continuous search spaces. In Th. Baeck, G. Eiben, M. Schoenauer, and H.-P. Schwefel, editors, *Proceedings of the 5th Conference on Parallel Problems Solving from Nature*, pages 418–427, 1998.

A Appendix: Validation of the EDA implementation

In this appendix, we validate our implementations of $\text{EMNA}_{\text{global}}$ and PBIL_{C}.

The published computational results for $\text{EMNA}_{\text{global}}$ do not detail completely the initialization of the algorithm. Particularly, the initial distribution is not explicitly characterized. While keeping everything else constant, we have initialized $\text{EMNA}_{\text{global}}$ with $\mu^0 = 0$ and $\Sigma^0 = 100I$, which possibly matches the "random generation procedure based on a uniform distribution" referred to in [2]. Under this initial distribution, we achieve results closely matching published results. Lastly, we note that we have omitted the function $\tilde{f}_3(x)$ since the published results utilized a variant that we did not use. The parameters that were presented in [2] were kept the same to validate our implementation of $\text{EMNA}_{\text{global}}$: dimension $n = 10$, population size $N = 2000$, sample size $M = 1000$, approximately $192,000$ function evaluations for $\tilde{f}_1(x)$ and approximately 95,000 for $\tilde{f}_2(x)$ in each of 10 runs. Here, we compare means to accommodate the published results.

Next, we validate PBIL_{C}. Since the authors of [3] do not investigate the performance of $\tilde{f}_2(x)$, we can only validate this EDA on $\tilde{f}_1(x)$ and $\tilde{f}_3(x)$. Once again, it is unclear

which initial parameters are used. Using initial standard deviations of $\frac{1}{2}$ (and zero mean), we attained results which again closely match published results. We matched the details presented in [3]: dimension $n = 100$, population size $N = 50$, sample size $K = 25$ and $200,000$ function evaluations for both $\tilde{f}_1(x)$ and $\tilde{f}_3(x)$ in each of 20 runs. Again, we compare means to accommodate the published results.

We present our results as well as the published results for comparison, in the form "mean $\pm$ standard deviation". Please see Table 10.

	Published	Current
$\text{EMNA}_{\text{global}}$ on $\tilde{f}_1(x)$	9999910 $\pm$ 10.3	9927294 $\pm$ 13115.4
$\text{EMNA}_{\text{global}}$ on $\tilde{f}_2(x)$	7.335e-06$\pm$ 2.24e-06	2.111e-07 $\pm$ 3.651e-08
PBIL_{C} on $\tilde{f}_1(x)$	4.65 $\pm$ 0.49	4.632 $\pm$ 0.483
PBIL_{C} on $\tilde{f}_3(x)$	11e-06$\pm$ 1e-06	0.021 $\pm$ 0.0013

Table 10: Validation of the current implementations of PBIL_{C} and $\text{EMNA}_{\text{global}}$.

The discrepancy in the validation of PBIL_{C} on $\tilde{f}_3(x)$ is likely due to a difference in initialization; $\tilde{f}_3(x)$ seems to be more sensitive to the initial distribution than $\tilde{f}_1(x)$, which matches the published results closely.

Unité de recherche INRIA Futurs
Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399