The Anticipated Mean Shift and Cluster Registration in Mixture-based EDAs for Multi-Objective Optimization

Peter A.N. Bosman[a]

[a] *Centrum Wiskunde & Informatica, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands,* Peter.Bosman@cwi.nl

The full version of this paper has been accepted for publication at the *2010 Genetic and Evolutionary Computation Conference — GECCO*.

1 Introduction

Many optimization problems in practice are multi-objective (MO). In MO optimization, the optimum is a *set* of solutions, called the optimal Pareto front, because many solutions may be equally good, e.g. solution $\boldsymbol{a}$ may be better in the first objective than solution $\boldsymbol{b}$, but worse in the second objective. Population-based methods such as evolutionary algorithms (EAs) are among the state-of-the-art in solving MO optimization problems. Estimation-of-distribution algorithms (EDAs) are a particular type of EA that aim to exploit features of a problem's structure in a principled manner via probabilistic modelling. For single-objective (SO) optimization, EDAs are already highly efficient. Here, we consider three ways to increase the efficiency of MOEDAs: overlapping clusters for more smoother density estimates when using mixture distributions, cluster registration for smoother cluster trajectories over generations, thereby increasing the efficiency of using the anticipated mean shift and running equal-capacity SO optimizers in synchronous parallel.

2 Clustered Variation, Registration and the Anticipated Mean Shift

A mixture distribution, i.e. a weighted sum of k distributions, is of particular interest because it can spread the search intensity along the Pareto front, allowing more focused exploitation of problem structure in different regions of the objective space. In EDAs, mixture distributions are often estimated by clustering, followed by density estimation per cluster. For a straightforward extension of existing EDAs, it is convenient to have uniformly sized clusters. To this end, we propose balanced k-leader-means (BKLM) clustering in which first k leaders are chosen far apart, followed by k-means clustering. The final clusters are obtained by growing the clusters equally from the k-means cluster means until the sum of cluster sizes is twice the total number of points, resulting in overlapping clusters and consequently a smoother density estimate (see Figure 1).

An important part of state-of-the-art variation operators are adaptive mechanisms that span multiple generations such as the Anticipated Mean Shift (AMS). To work properly, a registration is required that determines the best correspondence between clusters in subsequent generations. Existing approaches use an *implicit* form of registration by assigning each newly generated solution to the cluster to which it is nearest (i.e. the highest density). Over multiple generations clusters can then however move across the Pareto front (see Figure 1). The goal of *explicit* cluster registration is to re-assign the cluster indices of the current generation t such that cluster i in generation t is the cluster that is closest to cluster i in the previous generation $t-1$. We have implemented an algorithm that achieves this. Figure 1 clearly shows the superiorly smooth front and stable registration over many generations for explicit registration, but so is the lack of front progress as a result of overfitting the selected solutions with more involved mixture estimates.

Front progress can be improved by use of AMS. The AMS is computed per cluster as the difference between the means of subsequent generations. A part, specifically $\alpha 100\%$, of the newly sampled solutions is then moved in the direction of the AMS. Inefficient front progress using one Gaussian is further detailed in Figure 1 on a two-dimensional and two-objective minimization problem defined by $f_0(\boldsymbol{x}) = \frac{1}{2}(x_0^2 + (x_1 - 1.0)^2)$ and $f_1(\boldsymbol{x}) = \frac{1}{2}((x_0 - 1.0)^2 + x_1^2)$. The optimal Pareto front is convex and defined by $x_0 = 1 - x_1$ and $f_1 = f_0 - 2\sqrt{f_0} + 1$. By initializing the population in $[0.9; 1.0]^2$ and not using AMS, the distribution quickly becomes misaligned with the direction of improvement. By adding AMS the Gaussian becomes re-aligned in parameter space with the direction of Pareto-improvement. In objective space the density already starts spreading along the optimal Pareto front within the first 7 generations.

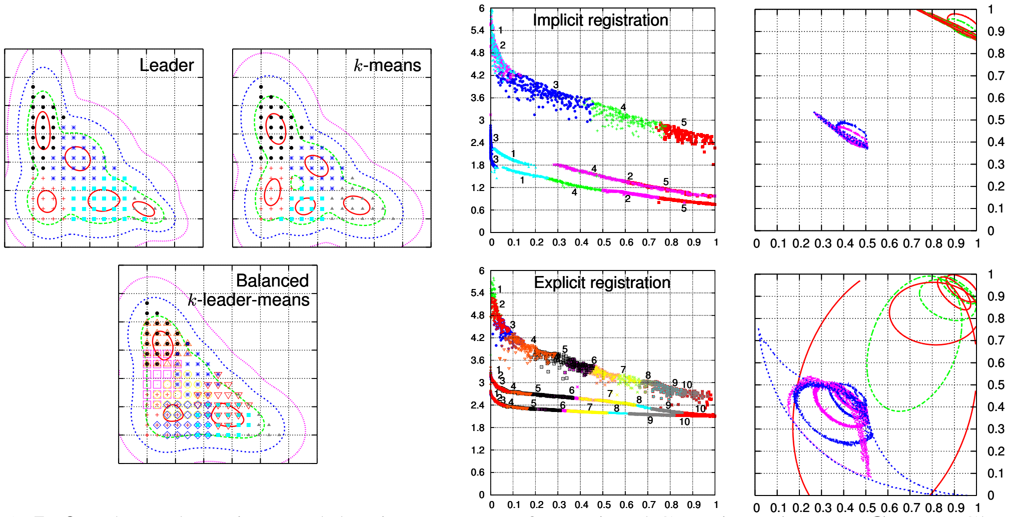

Figure 1: **Left**: Three clusterings and density contours of associated Gaussian mixtures. **Center**: Clustering using implicit and explicit registration of 5 and 10 clusters respectively in three different generations. **Right**: 95%-contours of the estimated distribution in the first 7 generations of typical MOEDA runs with a single Gaussian with (bottom) and without (top) AMS. Subsequent generations alternatingly use solid and dashed lines. Estimations are shown both in parameter space (red and green) and objective space (blue and pink).

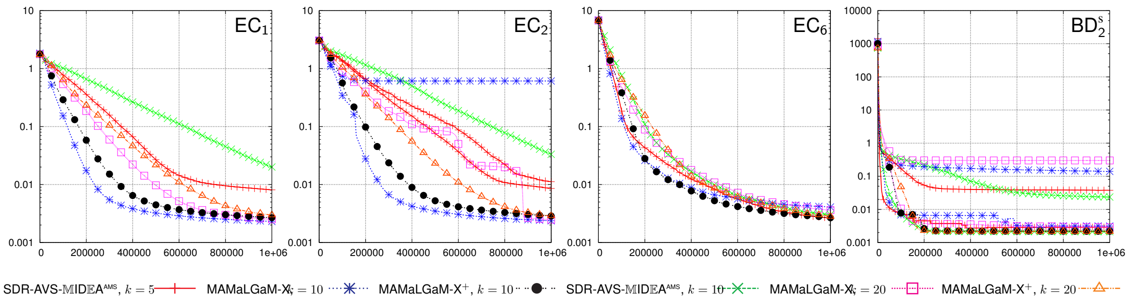

Figure 2: Average performance of various algorithms. Horizontal axis: number of evaluations (both objectives per evaluation). Vertical axis: distance to optimal Pareto front. Averages are shown both for successful runs and unsuccessful runs, giving double occurrences of lines if some runs were unsuccessful.

3 Synchronous Parallel Single Objective EDAs

Although clustered variation spreads the search bias, MO selection still focuses exploitation on all objectives at the same time. It may therefore be beneficial to add expert search bias in the form of separate SO optimization of the m objectives. We propose a scheme that differs from existing literature in that 1) m independent SO EDAs are run, one for each objective separately, 2) each SOEDA is similar to the MOEDA when considering a single cluster, resulting in similar convergence speed 3) the best solutions found by each SOEDA in each generation are integrated into the MOEDA, not vice versa.

The techniques described here are integrated into an existing MOEDA called SDR-AVS-$\mathbb{MIDEA}$, resulting in MAMaLGaM-X (Multi-objective AMaLGaM-miXture, where AMaLGaM stands for Adapted Maximum-Likelihood Gaussian Model). The added use of synchronous parallel SOEDAs (referred to as MAMaLGaM-X$^+$) was studied separately in an experimental setting. Convergence results averaged over 30 runs on well-known benchmark functions indicate better optimization performance (see Figure 2 for results on 4 functions; for the complete set of 9 functions, see the full version of this paper).

4 Conclusions

To find good approximations of the optimal Pareto front, continued pressure toward finding improvements is required. Once many solutions of a similar quality are selected, a MOEDA can easily converge prematurely due to overfitting a contour in the fitness landscape rather than exploring directions of improvement. Enlarging the capacity of the probabilistic model via mixture distributions and the modelling of dependencies only increases the probability of overfitting, contrary to what is the common expectation in EDA literature. The techniques described in this paper reduce this risk substantially and effectively. Moreover, using the proposed BKLM clustering technique any EDA can be extended to a mixture-based version straightforwardly.