Enhancing Gaussian Mixture Model Fitting via Equiprobable Binning and Adaptive Differential Evolution

Wojciech Achtelik[1][0009−0001−7363−4448] and Maciej Smołka[1][0000−0002−3386−0555]

AGH University of Krakow, Kraków, Poland
{wachtelik,smolka}@agh.edu.pl

Abstract. Fitting Gaussian Mixture Models (GMMs) to one-dimensional data is a fundamental task in machine learning, traditionally addressed using the Expectation-Maximization (EM) algorithm. However, EM lacks inherent mechanisms to enforce separation between mixture components, a critical requirement in domains like medical research where distinct subgroups must be identified. Recently, the Distribution Optimization (DO) framework addressed this limitation by reformulating GMM estimation as a chi-squared goodness-of-fit minimization problem with an overlap penalty to enhance separation. However, its reliance on equiwidth binning and genetic algorithms can limit accuracy and scalability. In this paper, we refine the DO framework in two key ways: (1) replacing equiwidth binning with Mann–Wald's equiprobable cells to improve estimation accuracy, and (2) adopting advanced Differential Evolution (DE) for more robust optimization of the high-dimensional parameter space. Through extensive experiments on synthetic and real-world datasets, we demonstrate that our refined approach significantly enhances accuracy, stability, and scalability compared to the original DO method.

Keywords: Gaussian Mixture Model · Minimum Chi-squared Estimation · Expectation Maximization · Differential Evolution · Distribution Optimization

1 Introduction

Gaussian Mixture Models are a cornerstone of modern statistical modeling, offering a powerful framework for capturing complex, multimodal distributions. By representing a probability density as a weighted sum of Gaussian components, GMMs have become essential across a wide array of disciplines, including machine learning, signal processing, bioinformatics, and medicine. In medical research, GMMs are well-suited for uncovering latent structures in single-dimensional data, with applications such as clustering patient scores, analyzing pain levels in rheumatoid arthritis [10], assessing diabetes risk [12], and evaluating olfactory function in clinical settings [11]. This suitability stems from their

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

theoretical foundation: many biological indicators arise from the cumulative effect of multiple underlying processes, aligning with the justification for normal distributions based on the central limit theorems.

Despite their widespread use, estimating GMM parameters accurately remains challenging, especially when overlap between components must be controlled or when reliable initial guesses are difficult to obtain. The classical approach is based on maximum likelihood estimation (MLE) and typically realized via the Expectation-Maximization algorithm. While EM offers computational efficiency and theoretical guarantees of convergence to a local optimum [24], it lacks built-in mechanisms for enforcing separation between mixture components, potentially complicating model interpretation in clinical applications where distinct, well-defined clusters are desired [7].

To overcome these limitations, alternative methods based on minimum chi-squared estimation (MCSE) have emerged. Notably, Lerch et al. [7] introduced the Distribution Optimization (DO) framework, which reimagines GMM parameter estimation as a constrained optimization problem. By minimizing a chi-squared goodness-of-fit statistic and incorporating an explicit penalty for component overlap, DO provides an intuitive mechanism to enforce mode separation, an advantage particularly relevant to medical applications where distinct subgroups must be identified. Yet, the original DO formulation has shortcomings: it employs equiwidth binning (Keating's formula), which can misalign with the data distribution and introduce bias, especially when components are well-separated or exhibit unequal variances, and it relies on genetic algorithms (GAs), which may lack the robustness required for higher-dimensional optimization landscapes.

This paper refines and extends the DO framework by addressing these critical challenges through two key innovations:

1. We replace equiwidth binning with Mann–Wald's equiprobable binning [16], which aligns bin boundaries with the underlying data distribution. This approach reduces bias and enhances estimation accuracy across diverse scenarios, from well-separated to overlapping components.
2. We substitute the original GA with advanced Differential Evolution (DE) variants, such as SHADE, L-SHADE, and iL-SHADE [20], [21], [2], known for their superior performance in continuous optimization. These adaptive algorithms improve convergence and scalability, particularly as the complexity of the mixture model grows.

Through rigorous experiments on synthetic and real-world datasets, we demonstrate that these refinements, equiprobable binning and DE-based optimization, significantly enhance the robustness, accuracy, and scalability of chi-squared-based GMM estimation.

2 Gaussian Mixture Model

The Gaussian Mixture Model is a parametric statistical model that represents a probability distribution as a weighted sum of multiple Gaussian components

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

[18]. In the one-dimensional case with M components, the probability density function is given by:

$$p(x \mid \lambda) = \sum_{i=1}^{M} \alpha_i \mathcal{N}(x \mid \mu_i, \sigma_i^2),$$

where $\mathcal{N}(x \mid \mu_i, \sigma_i^2)$ represents the i-th Gaussian component with mean μ_i and variance σ_i^2, and α_i is the corresponding mixing coefficient. The complete parameter set λ is defined as:

$$\lambda = \{\alpha_1, \ldots, \alpha_M,\ \sigma_1, \ldots, \sigma_M,\ \mu_1, \ldots, \mu_M\} \in \mathbb{R}^{3M},$$

subject to the constraints:

$$\alpha_i \geq 0, \quad \sum_{i=1}^{M} \alpha_i = 1, \quad \sigma_i > 0.$$

For notational convenience, we denote the parameters associated with the i-th component as:

$$\lambda_i = (\alpha_i,\ \sigma_i,\ \mu_i) \in \mathbb{R}^3.$$

2.1 Parameter estimation

Typically, the parameters of a GMM are estimated through maximum likelihood estimation, which is commonly carried out using the Expectation-Maximization algorithm. This work builds upon and extends the chi-squared minimization framework for GMM parameter estimation introduced by Lerch et al. [7]. We first summarize their core approach.

Lerch et al. [7] redefined the task of parameter estimation in GMMs as an optimization problem, leveraging minimum chi-squared. Their approach was motivated by the need to directly address the challenge of mode separation within the GMM, a limitation often overlooked in conventional methods. The core of their method is an objective function that integrates two key metrics: the discrepancy between theoretical and empirical distributions quantified by χ^2 statistic, and a measure of overlap error between components. Conceptually, this can be viewed as a constrained optimization problem that penalizes mixtures exhibiting excessive overlap. In doing so, Lerch et al. introduce two important innovations. First, they minimize χ^2 instead of maximizing likelihood. Second, they incorporate an explicit overlap penalty to improve mode separation. While this penalty could theoretically be applied to a likelihood-based objective, the choice of χ^2 stands out for its computational simplicity. Specifically, χ^2 can be calculated efficiently using binned data, eliminating the need to evaluate likelihood function for every data point individually. As a result, the χ^2-based method offers a practical advantage, especially when working with large datasets.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

To define the objective function we shall introduce some auxiliary notions. Let us denote the (one-dimensional) dataset on which the model is trained as X and

$$R_X = M_X - m_X, \quad M_X = \max(X), \quad m_X = \min(X), \quad n = \#X.$$

The χ^2 statistic assessing the goodness of fit is defined in the following way

$$\chi^2(\lambda) = \sum_{j=1}^{K} \frac{\left(\#(X \cap k_j) - P(x \in k_j|\lambda) \cdot n\right)^2}{P(x \in k_j|\lambda) \cdot n}, \tag{1}$$

where k_j denotes the j-th bin (further details are provided in the next section) and K is the total number of bins. The second component of the objective

$$\epsilon_{overlap}(\lambda) = \max_{m=1,\ldots,M} \sum_{i=1}^{n} \min\left(\frac{\max_{j=1,\ldots,M;\, j\neq m}\left(\alpha_j \cdot p(x_i|\lambda_j)\right)}{\alpha_m \cdot p(x_i|\lambda_m)}, 1\right) \tag{2}$$

approximates the area overlapped between neighboring mixture components. The final objective function is calculated as

$$f(\lambda) = \begin{cases} +\infty & \text{if } \epsilon_{overlap}(\lambda) > \tau_{overlap}, \\ \chi^2(\lambda) & \text{otherwise.} \end{cases} \tag{3}$$

For a particular dataset, the value of the overlap threshold $\tau_{overlap}$ can be chosen in such a way that components are properly separated. To make the optimization problem tractable, Lerch et al. [7] further constrain the search domain to the following subset of a hypercube in $\mathbb{R}^{3M}$:

$$\mathcal{D} = \Big\{\lambda \in \mathbb{R}^{3M} : \alpha_i \in [0,1],\ \sigma_i \in [0.001 \cdot R_X, 0.1 \cdot R_X],\ \mu_i \in [m_X, M_X],\ \sum_{i=1}^{M} \alpha_i = 1\Big\} \tag{4}$$

Thus, the final problem is to find $\lambda^* \in \mathcal{D}$ such that

$$f(\lambda^*) = \min_{\lambda \in \mathcal{D}} f(\lambda). \tag{5}$$

To solve this problem, Lerch et al. [7] propose the DO algorithm. It is a genetic algorithm (GA) that employs the uniform mutation, the simple arithmetic crossover and the tournament selection. The weights and the standard deviations of the initial population are sampled from the uniform distribution, the means are sampled from the following normal distribution

$$\mu_i \sim \mathcal{N}\left(m_X + i \cdot \frac{R_X}{M+1}, \frac{R_X}{5M}\right), \quad i = 1, \ldots, M. \tag{6}$$

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

3 Minimum Chi-Squared Estimation

Minimum chi-squared estimation is a classical technique for parameter estimation based on minimizing Pearson's chi-squared test statistic:

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}. \tag{7}$$

Although it is applied less frequently than MLE, many statisticians have highlighted its theoretical and practical merits. In particular, Berkson [1] argued that MCSE represents a fundamental principle of statistical estimation.

A key issue in applying MCSE to continuous distributions is the necessity of discretizing data into bins, given that the chi-squared test is inherently designed for categorical data. This binning process can introduce bias and variability unless the bins are carefully selected. To address this, Mann and Wald [16] proposed using equiprobable bins and derived a formula for selecting the optimal number of such bins:

$$K = 4 \left(\frac{2n^2}{c(\alpha)^2} \right)^{\frac{1}{5}}, \tag{8}$$

where $c(\alpha)$ is the $(1-\alpha)$ quantile of the standard normal distribution (commonly $\alpha = 0.05$). Subsequent studies [3,17] later proved that equiprobable binning yields an unbiased chi-squared test, whereas bins with unequal probabilities do not. Intuitively, equal-probability boundaries avoid having bins with extremely low counts, thus reducing the distortion in the chi-squared statistic. This approach, often referred to as the Mann and Wald formula, remains widely used in both minimum chi-squared estimation [6] and goodness-of-fit testing [9].

By contrast, Lerch et al. [7] used equiwidth bins rather than equiprobable bins. They selected the total number of bins, K, using a variant of Keating's formula [5], which first computes

$$\hat{\sigma} = \min\Big(s,\ \frac{\mathrm{IQR}}{1.349}\Big), \tag{9}$$

where IQR is the interquartile range, s is the sample standard deviation. Then K is set by

$$K = \Big\lceil \frac{R_X\, n^{\frac{1}{3}}}{3.49\,\hat{\sigma}} \Big\rceil. \tag{10}$$

Although these binning strategies might appear superficially similar, they can produce substantial differences in MCSE outcomes.

Figure 1 illustrates how these differing binning choices lead to noticeably different fitted GMMs. Specifically, using equiwidth bins, as in Lerch et al. [7], tends to shift the estimated mean values relative to those obtained via Mann and Wald's equiprobable bins.

Lerch's [7] equiwidth strategy is particularly problematic when some components of the distribution are well-separated or overlapping. In well-separated scenarios, bins can end up near-empty in regions between modes. Such bins with

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

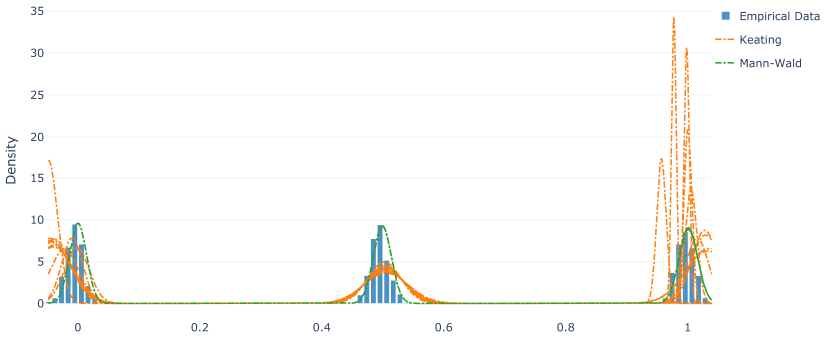

Fig. 1. Fitted GMMs under two different binning strategies, each run 10 times with different random seeds. The orange lines represent solutions using equal-width bins with Keating's formula, while green lines show solutions using equal-probability bins with the Mann-Wald method. The consistent pattern across multiple runs demonstrates that Keating's method systematically shifts the estimated mean values compared to the equal-probability approach.

expected counts near zero destabilize the chi-squared statistic, as χ^2 become more sensitive to small fluctuations when E_i is small. Conversely, if two overlapping components (especially those with small variance) end up in the same bin, MCSE struggles to resolve these components and yields parameter estimates with high variance. Furthermore, because equiwidth bin edges do not adapt to the observed data distribution, they frequently occur in low-probability regions. This placement systematically biases parameter estimates, pulling component means toward less accurate values, a phenomenon clearly illustrated in Figure 1. In contrast, the equiprobable binning strategy proposed by Mann and Wald naturally aligns bin boundaries with high-density regions of the data, leading to more robust and accurate parameter estimates.

4 Differential Evolution

Differential Evolution (DE) is a population-based stochastic optimization algorithm for continuous search spaces introduced by Storn and Price [19]. DE evolves a population of candidate solutions through successive generations by applying mutation, crossover, and selection operators. In the widely adopted *DE/rand/1* scheme, a mutant vector is created by:

$$\mathbf{v}_i = \mathbf{x}_{r_1} + F\left(\mathbf{x}_{r_2} - \mathbf{x}_{r_3}\right), \tag{11}$$

where r_1, r_2, r_3 are distinct random indices different from i, and $F \in (0, 2]$ is the scale factor controlling the mutation step size. A trial vector is then generated

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

through binomial crossover:

$$u_{i,j} = \begin{cases} v_{i,j}, & \text{if rand}(0,1) \leq CR \text{ or } j = j_{\text{rand}}, \\ x_{i,j}, & \text{otherwise,} \end{cases} \tag{12}$$

where $CR \in [0,1]$ is the crossover rate and j_{rand} ensures at least one component from the mutant vector is inherited. In the selection step, the trial vector replaces the target vector if it yields a better objective function value. Standard parameter settings such as a population size of 100, $F = 0.5$, and $CR = 0.5$ often serve as a baseline.

4.1 SHADE variants

To improve DE's robustness, several adaptive strategies have been proposed that dynamically adjust control parameters based on search performance. SHADE [20] implements *success-history based parameter adaptation*. It maintains a historical memory (of a fixed size) which stores promising F and CR parameter settings. After each generation, SHADE identifies the F and CR values that successfully produced offspring superior to their parents. These *successful* parameters are aggregated, using a weighted mean that gives greater importance to those yielding larger fitness improvements. This aggregated result then updates one entry in the historical memory. When generating new candidate solutions, SHADE draws upon values stored in this memory to guide the selection of F and CR for the next generation. This data-driven approach, learning from recent successful parameters, provides more reliable and adaptive control compared to static settings. L-SHADE [21] extends SHADE by incorporating *linear population size reduction*, starting with a large population for exploration and gradually reducing it to focus the search. iL-SHADE [2] further refines L-SHADE through several modifications aimed at enhancing convergence speed and solution quality. It incorporates mechanisms to promote higher CR values and employs a refined parameter memory update rule. Additionally, iL-SHADE restricts high F and low CR values during the initial search phase. A key feature is the linear adaptation of the control parameter p for the `current-to-pbest/1` mutation, adjusting it based on the search progression. These combined enhancements contribute to iL-SHADE's standing as a state-of-the-art DE variant [23].

4.2 Motivation

In the IEEE CEC 2006 and 2010 competitions on constrained real-parameter optimization [8] [15], the majority of submitted algorithms were based on DE, and DE-variants ultimately claimed top positions in both events. The consistent success of DE in these competitions can be attributed to its robust mutation and recombination strategies, which balance exploration and exploitation effectively, even under complex constraints. Beyond constrained optimization, DE-based methods have exhibited state-of-the-art performance in single-objective

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

box-constrained problems, as demonstrated in large-scale benchmarking studies such as those conducted on the BBOB function suite [23]. Algorithms like SHADE, L-SHADE, and il-SHADE have repeatedly shown competitive performance across a variety of problem classes, underscoring the versatility and effectiveness of the DE framework. These advances make DE and its variants well-suited for tackling constrained optimization problems, including those under death-penalty constraints.

In addition to DE-based methods, we also performed preliminary studies using CMA-ES [4], a covariance-matrix-adaptation evolution strategy widely recognized for its effectiveness in continuous optimization. However, our initial experiments revealed that CMA-ES struggled to handle the constraint effectively, yielding inferior performance compared to DE variants. Given the poor performance in our pilot tests we excluded CMA-ES from further analyses.

5 Experimental Procedure

To comprehensively evaluate binning methods and optimization algorithms for fitting one-dimensional GMMs we chose to work with synthetically generated datasets. Synthetic data allow us to control the underlying distribution and assess fitted models against the ground truth. We measure performance with the Jensen-Shannon Divergence (JSD), because it is independent from the optimization objectives of the methods compared (χ^2 for MCSE, likelihood for EM). JSD's symmetry, finiteness, and information-theoretic interpretation make it well-suited for comparing continuous distributions.

Moreover, because established benchmarks for one-dimensional GMMs are scarce, we opted to produce test datasets with diverse parameter configurations. Our generation procedure is based on an iterative rejection sampling strategy to ensure that each chosen parameter set lies within the feasible domain of our optimization problem (Eq. (4)). Specifically, we generate candidate mixtures by sampling (i) the mixture weights (enforcing a minimum weight to avoid degenerate components), (ii) the component means within a specified range, and (iii) the corresponding standard deviations within the bounds stipulated by our domain constraints. For each candidate, we then sample data by drawing random samples from the resulting mixture distribution. A dataset is accepted only if it meets the criteria that all component weights, means, and standard deviations remain valid and that the mixture components exhibit an overlap in a specified interval (which can be adjusted). This generation process guarantees that the ground truth distributions are feasible solutions – a crucial aspect that ensures both MCSE-based methods and EM can theoretically achieve the optimal solution, addressing a limitation in previous comparative studies [13] where the feasibility of the underlying distribution was often overlooked, potentially biasing results in favor of one method over the other.

By generating synthetic datasets, we can further investigate how varying the number of mixture components influences each method's robustness. To this end, we construct 100 datasets for each component count between 2 and 10, result-

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

ing in a total of 900 datasets. For each dataset, we run each method 10 times with different random seeds to assess initialization effects. The performance of each run is then quantified by computing the Jensen–Shannon Divergence between the fitted model and the ground truth. We aggregate these results by comparing methods through rank-based analyses and by reporting the average Jensen–Shannon Divergence across all runs. Specifically, we benchmark the original Distribution Optimization approach (denoted as GA with Keating) against three alternatives: a Genetic Algorithm variant with Mann–Wald binning, an iL-SHADE algorithm with Mann–Wald binning, and the classical Expectation Maximization technique. This experimental design ensures a thorough assessment of how each method's stability and accuracy respond to changes in both the number of mixture components and the random seed.

Furthermore, our experimental design enables a detailed analysis of the individual contributions of binning strategies and optimization algorithms to the overall performance. By comparing the original GA with Keating and the GA with Mann–Wald binning, we can isolate and quantify the impact of the binning strategy, as both methods employ the same optimization algorithm. Any performance differences between these two methods can thus be attributed to the choice of binning. Similarly, the comparison between GA with Mann–Wald and iL-SHADE with Mann–Wald allows us to evaluate the effect of employing a more sophisticated optimization algorithm, iL-SHADE, while maintaining the same binning approach. In addition, by incorporating the classical EM algorithm into our benchmarking suite, we can assess the performance of the Distribution Optimization approach relative to a well-established standard in the field. Previous studies comparing the DO approach with EM have often relied on a much smaller number of datasets, which may not fully capture their behavior across diverse scenarios.

Finally, we compare optimization algorithms[1] (GA, DE, SHADE, L-SHADE, iL-SHADE) under fixed Mann-Wald binning. Each algorithm runs 10 times per dataset with different seeds, recording the best objective value f (Eq. (3)) after 10,000 evaluations – a standard budget sufficient for convergence across algorithms in our setting. We evaluate consistency, effectiveness, and robustness to the number of components, which scales the solution space dimensionality.

5.1 Real-World Datasets

To complement our synthetic experiments and evaluate the generalizability of the algorithms, we analyze three real-world datasets. These datasets were previously used to benchmark Gaussian Mixture Model implementations in R packages [14]. In the absence of ground truth distributions, we perform visual and qualitative assessments of the fitted GMMs.

[1] All implementation details are available in the accompanying GitHub repository at https://github.com/agh-a2s/distribution-optimization.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

1. **Truck Driving**: Sourced from the `AdaptGauss` R package [22], this dataset captures the time taken by trucks to reach seaports. We model it with a three-component GMM.
2. **Iris ICA**: This dataset comprises the first independent component obtained via Independent Component Analysis (ICA) applied to Fisher's Iris dataset. The one-dimensional projection corresponds to the three Iris species, justifying a three-component GMM to model the species-specific distributions.
3. **Chromatogram Time**: Available in the `opGMMassessment` R package [14], this dataset contains chromatograms of five distinct lysophosphatidic acids. A five-component GMM is fitted to represent the separate peaks associated with each acid.

In our real-world study, we narrow our focus to two algorithms: the original DO approach, which uses a GA with Keating binning, and our enhanced version, which employs Mann-Wald binning and the iL-SHADE. This deliberate restriction allows us to directly evaluate the practical impact of our proposed modifications. By comparing these two methods, we can assess their effects on convergence stability and the variability of fitted mixture models, determining whether the improvements seen in synthetic experiments translate to real-world scenarios.

6 Results

Figure 2 evaluates the performance of four methods for GMM estimation: GA with Keating (original Distribution Optimization algorithm), GA with Mann-Wald, iL-SHADE with Mann-Wald, and EM across synthetic datasets with mixture components ranging from 2 to 10. The left panel displays the average rank based on JSD, where a lower rank indicates better performance, while the right panel shows the average JSD value, with lower values reflecting closer fits to the ground truth. For datasets with fewer components (2–4), methods using Mann-Wald binning (GA with Mann-Wald and iL-SHADE with Mann-Wald) achieve JSD values and ranks comparable to EM. However, as the number of components increases beyond 4, EM consistently outperforms the evolutionary methods, securing the lowest rank (near 1) and JSD values (0.01). This trend highlights EM's superior efficacy in optimizing complex mixtures. Among the evolutionary methods, iL-SHADE with Mann-Wald maintains the best performance, with JSD values rising modestly to approximately 0.02 by 10 components. In contrast, GA with Mann-Wald and GA with Keating show declining performance, with JSD values climbing to 0.06 and 0.08 respectively, as complexity grows. The transition from Keating to Mann-Wald binning with the GA delivers a consistent improvement, reducing JSD values by a nearly constant margin across all component counts. More striking, however, is the optimization algorithm's influence on performance scalability. While GA with Mann-Wald exhibits a rapid increase in JSD values with growing mixture complexity, iL-SHADE with Mann-Wald demonstrates a much slower growth rate, indicating

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

that iL-SHADE's adaptive differential evolution enhances search efficiency in higher-dimensional spaces compared to genetic algorithms, enabling better handling of increasingly complex mixture models.

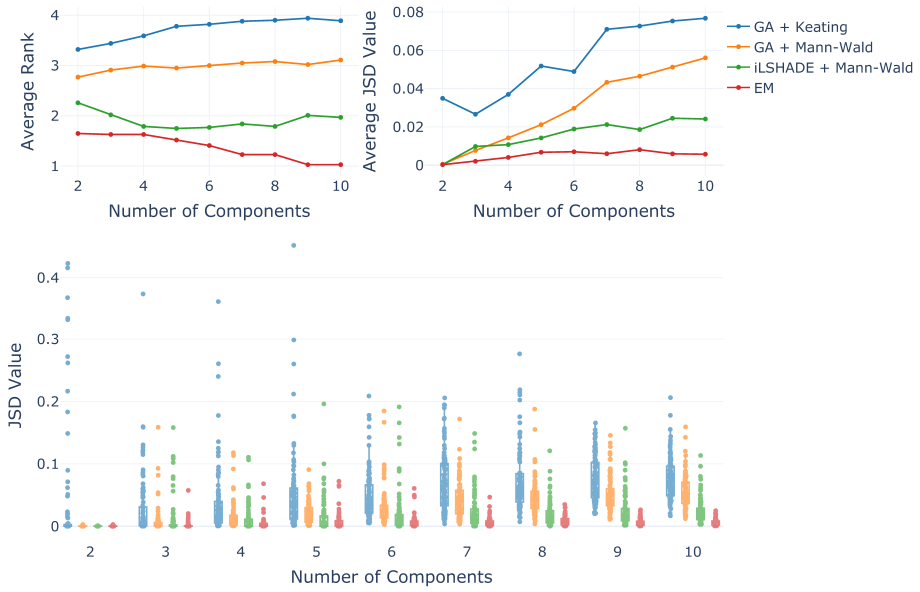

Fig. 2. The top left panel reports the average rank (based on JSD), while the top right panel presents the corresponding average JSD value. Both metrics are aggregated over 100 synthetic datasets per component count, with 10 runs per dataset. The plot in the second row presents distributions of JSD values.

Figure 3 compares five optimization algorithms – DE, GA, SHADE, LSHADE, and iL-SHADE – using the final objective function value (Eq. 3) under fixed Mann-Wald binning. The left panel shows the average rank (lower is better), and the right panel presents the average function value (lower indicates better solutions) across component counts from 2 to 10. iL-SHADE consistently achieves the lowest rank and function values, demonstrating robust and effective optimization. LSHADE follows closely, maintaining strong performance, while SHADE and GA rank progressively higher (worse) and yield poorer function values, especially as the number of components increases. These findings underscore iL-SHADE's superior ability to handle multidimensional optimization problems.

6.1 Real-World Datasets

Figure 4 presents three real-world datasets, each overlaid with the Gaussian Mixture Model fits from the GA using Keating and from iL-SHADE employ-

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

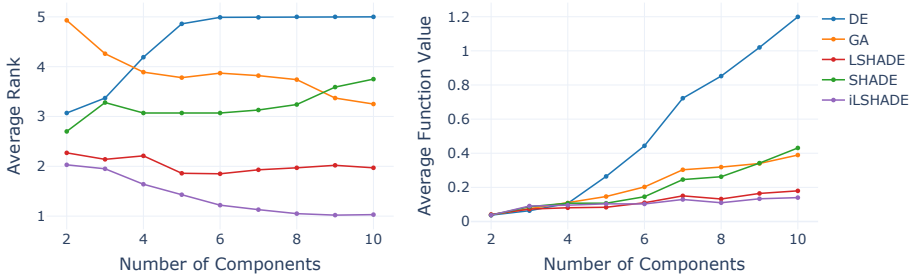

Fig. 3. The left panel reports the average rank (based on objective function value), while the right panel presents the corresponding average objective function value. Both metrics are aggregated over the same datasets as in Fig.2.

ing Mann-Wald, repeated ten times with distinct random seeds. Several clear patterns emerge:

- **Truck Driving (top).** The iL-SHADE solutions demonstrate consistently lower variability across runs and exhibit a closer alignment with the peaks. However, both models effectively capture the dataset's key structural features While subtle differences are present, it remains unclear which method offers a superior fit.
- **Iris ICA (middle).** Both methods capture the main mode near 1.0, but GA with Keating often broadens this peak excessively or shifts it (to the right). Meanwhile, iL-SHADE with Mann-Wald consistently converges on a sharper peak that more accurately matches the empirical histogram. Across repeated runs, GA-based fits fluctuate considerably and produce both underestimated and over-estimated standard deviations, whereas iL-SHADE remains stable and yields nearly identical fits each time.
- **Chromatogram Time (bottom).** Here, the data contain five well-separated peaks. iL-SHADE with Mann-Wald again provides cohesive mixtures with tightly estimated component means and variances, resulting in visually robust fits for each distinct mode. In contrast, GA with Keating occasionally merges or splits peaks unnecessarily and shows a tendency to decrease the component widths.

Overall, the GA with Keating approach often struggles to estimate the correct mean locations and realistic standard deviations for the mixture components, often generating overly narrow or broadened peaks that fail to capture the true distributions. By comparison, iL-SHADE with Mann-Wald not only yields more stable solutions over multiple runs but also offers more precise alignment with the empirical data, indicating stronger convergence properties and more reliable mixture fits.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

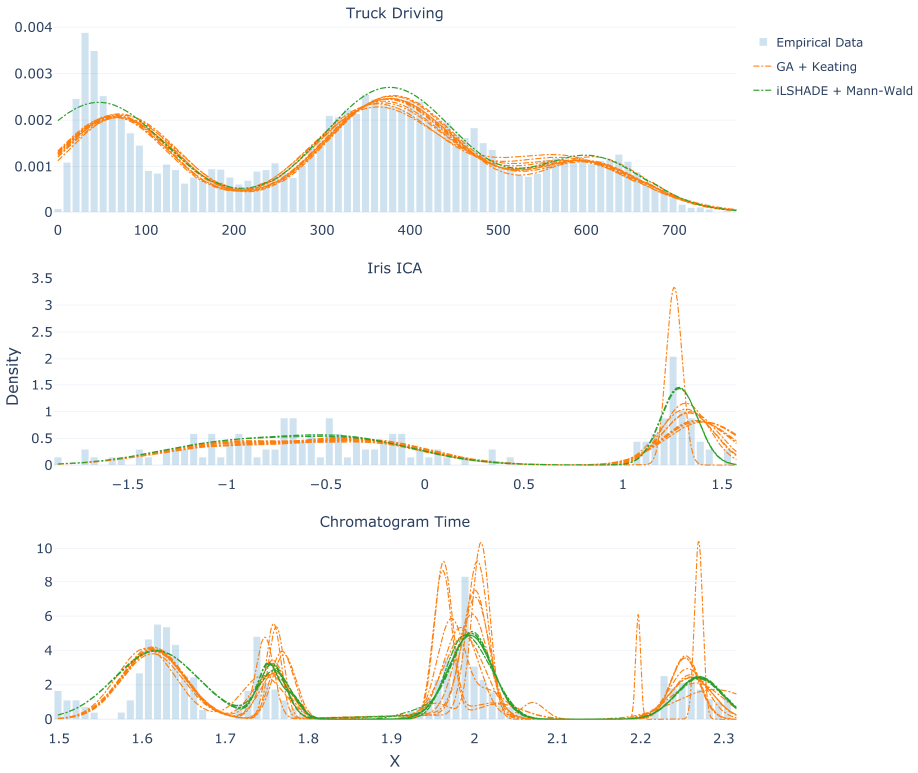

Fig. 4. Three real-world datasets (histograms) alongside the ten GMM fits of two methods: GA with Keating and iL-SHADE with Mann-Wald.

7 Conclusions

This paper examined strategies for estimating one-dimensional GMMs with a death-penalty constraint by focusing on two main aspects: the choice of binning scheme and the selection of optimization algorithm. Our study can be viewed as both an in-depth investigation and a refinement of the Distribution Optimization approach [7]. A key contribution of our work is the adoption of a binning method better suited to chi-squared-based estimation. Whereas Keating's equiwidth bins frequently misalign with the data, Mann–Wald's equiprobable bins adapt naturally to the underlying distribution. This refinement addresses previous difficulties in fitting GMMs by reducing bias. Among the evolutionary algorithms evaluated, iL-SHADE consistently outperformed the Genetic Algorithm and other DE variants, achieving lower objective function and JSD values. Furthermore, both iL-SHADE and LSHADE exhibit superior scalability, maintaining effective performance as the number of mixture components increases. While evolutionary methods were competitive for low-dimensional mixtures (fewer than four com-

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

ponents), the EM algorithm achieved notably better fits for complex mixtures, consistently obtaining lower JSD values. This highlights EM's strengths particularly when the number of mixture components grows and the search space becomes more challenging. Nevertheless, evolutionary algorithms remain appealing when constraints or custom objective functions are required, scenarios in which EM might not be directly applicable or would require major modifications.

Building upon these findings, several avenues for future research emerge. A primary direction involves extending the proposed methodology to estimate multidimensional GMMs. Furthermore, the current work relied exclusively on the death-penalty approach for handling constraints. Future investigations could benefit from analyzing other constraint handling techniques.

Acknowledgments. The research presented in this paper was partially supported by the Polish National Science Center under grant No. 2023/49/B/ST6/01404, by the funds of the Polish Ministry of Science and Education assigned to the AGH University of Krakow, and by the program "Excellence initiative - research university" for the AGH University of Krakow.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Berkson, J.: Minimum chi-square, not maximum likelihood! The Annals of Statistics **8**(3), 457–487 (1980)
2. Brest, J., Maučec, M.S., Bošković, B.: iL-SHADE: Improved L-SHADE algorithm for single objective real-parameter optimization. In: 2016 IEEE Congress on Evolutionary Computation (CEC). pp. 1188–1195 (2016). https://doi.org/10.1109/CEC.2016.7743922
3. Cohen, A., Sackrowitz, H.: Unbiasedness of the chi-square, likelihood ratio, and other goodness of fit tests for the equal cell case. The Annals of Statistics pp. 959–964 (1975)
4. Hansen, N.: The CMA evolution strategy: A tutorial. arXiv (2023). https://doi.org/10.48550/arXiv.1604.00772
5. Keating, J., Scott, D.: A primer on density estimation for the great home run race of 98. Stats **25**, 16–22 (1999)
6. Kominek, Z.: Minimum chi-squared estimation of stable distributions parameters: an application to the Warsaw Stock Exchange. Journal of Applied Statistics **29**(5), 729–744 (2002)
7. Lerch, F., Ultsch, A., Lötsch, J.: Distribution Optimization: An evolutionary algorithm to separate Gaussian mixtures. Scientific Reports **10**(1), 648 (2020). https://doi.org/10.1038/s41598-020-57432-w
8. Liang, J.J., Runarsson, T.P., Mezura-Montes, E., Clerc, M., Suganthan, P.N., Coello Coello, C.A., Deb, K.: Problem definitions and evaluation criteria for the CEC 2006 special session on constrained real-parameter optimization. Journal of Applied Mechanics **41**(8), 8–31 (2006)
9. Liu, T.C., Kalugin, P.N., Wilding, J.L., Bodmer, W.F.: GMMchi: gene expression clustering using [gaussian] mixture modeling. BMC Bioinformatics **23**(1), 457 (2022)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22

10. Loetsch, J., Alfredsson, L., Lampa, J.: Machine-learning–based knowledge discovery in rheumatoid arthritis–related registry data to identify predictors of persistent pain. Pain **161**(1), 114–126 (2020)
11. Lötsch, J., Hummel, T.: Data science-based analysis of patient subgroup structures suggest effects of rhinitis on all chemosensory perceptions in the upper airways. Chemical Senses **46** (01 2021). https://doi.org/10.1093/chemse/bjab001
12. Lötsch, J., Hähner, A., Schwarz, P.E.H., Tselmin, S., Hummel, T.: Machine learning refutes loss of smell as a risk indicator of diabetes mellitus. Journal of Clinical Medicine **10**(21) (2021). https://doi.org/10.3390/jcm10214971
13. Lötsch, J., Malkusch, S., Ultsch, A.: Comparative assessment of automated algorithms for the separation of one-dimensional Gaussian mixtures. Informatics in Medicine Unlocked **34**, 101113 (2022). https://doi.org/https://doi.org/10.1016/j.imu.2022.101113, https://www.sciencedirect.com/science/article/pii/S2352914822002507
14. Lötsch, J., Malkusch, S., Ultsch, A.: Comparative assessment of automated algorithms for the separation of one-dimensional Gaussian mixtures. Informatics in Medicine Unlocked **34**, 101113 (2022). https://doi.org/10.1016/j.imu.2022.101113
15. Mallipeddi, R., Suganthan, P.N.: Problem definitions and evaluation criteria for the CEC 2010 competition on constrained real-parameter optimization. Nanyang Technological University, Singapore **24**, 910 (2010)
16. Mann, H., Wald, A.: On the choice of the number of class intervals in the application of the chi square test. The Annals of Mathematical Statistics **13**(3), 306–317 (1942)
17. Moore, D.S.: Tests of chi-squared type. In: Goodness-of-fit-techniques, pp. 63–96. Routledge (2017)
18. Reynolds, D.: Gaussian mixture models. In: Li, S.Z., Jain, A. (eds.) Encyclopedia of Biometrics, pp. 659–663. Springer, Boston, MA (2009). https://doi.org/10.1007/978-0-387-73003-5_196
19. Storn, R., Price, K.: Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization **11**, 341–359 (1997)
20. Tanabe, R., Fukunaga, A.: Evaluating the performance of SHADE on CEC 2013 benchmark problems. In: 2013 IEEE Congress on Evolutionary Computation. pp. 1952–1959 (2013). https://doi.org/10.1109/CEC.2013.6557798
21. Tanabe, R., Fukunaga, A.S.: Improving the search performance of SHADE using linear population size reduction. In: 2014 IEEE Congress on Evolutionary Computation (CEC). pp. 1658–1665 (2014). https://doi.org/10.1109/CEC.2014.6900380
22. Ultsch, A., Thrun, M.C., Hansen-Goos, O., Lötsch, J.: Identification of molecular fingerprints in human heat pain thresholds by use of an interactive mixture model R toolbox (AdaptGauss). International Journal of Molecular Sciences **16**(10), 25897–25911 (2015)
23. Vermetten, D., Doerr, C., Wang, H., Kononova, A.V., Bäck, T.: Large-scale benchmarking of metaphor-based optimization heuristics. In: Proceedings of the Genetic and Evolutionary Computation Conference. p. 41–49. GECCO '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3638529.3654122, https://doi.org/10.1145/3638529.3654122
24. Wu, C.F.J.: On the convergence properties of the em algorithm. The Annals of Statistics **11**(1), 95–103 (1983), http://www.jstor.org/stable/2240463

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_22