CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES

https://cajmtcs.casjournal.org/index.php/CAJMTCS

Volume: 07 Issue: 01 | January 2026 ISSN: 2660-5309

Article

Modified Principal Points: A Flexible and Differentiable Approach for Data Summarization

Thaer Ziara Arzig

1. Department of Mathematics, Yasouj University, Yasouj, Iran
* Correspondence: thayrzyarh3@gmail.com

Abstract: Principal points∗are a small set of characteristic locations which minimize the average squared Euclidian distance from the data points, and should be more informative about the data's structure than simple features such as mean and variance. However it is also non-differentiable w.r.t point collapse due to minimum operation in (2.8) and weak-sparse in defining point spread penalty. In this paper we define the generalized principal points as the Gaussian weighted mean of distances. It results in a differentiable objective and has a tuning parameter for point closeness adjustment. When the bandwidth approaches to zero, modified points tend towards classical points from a real direction and when it tends to the infinity they define the mean. Finally, the simulation studies are reported and reveal higher robustness of our proposed methods against outliers, non-normality, and small sample sizes. Empirical studies on real statistical data also confirm that lower sensitivity to extremes is better

Keywords: Principal points, Kernel weighting, Differentiable optimization, Data summarization, Bandwidth tuning.

Citation: Arzig, T. Z. Modified Principal Points: A Flexible and Differentiable Approach for Data Summarization. Central Asian Journal of Mathematical Theory and Computer Sciences 2026, 7(1), 29-39.

Received: 10th Aug 2025
Revised: 16th Sep 2025
Accepted: 24th Oct 2025
Published: 18th Nov 2025

1. Introduction

In the age of big data, summarization of massive high-dimensional datasets has become increasingly important in numerous domains such as computer vision, natural language processing, scientific computing, and healthcare analytics. As the world's data becomes "big" in scale, there is an increasing need for techniques that can find compact, informative and characteristic views of the data while preserving its underlying structural or statistical properties. Data summarisation facilitates a lowered computational cost, interpretability and generalisability of the downstream models [1].

A classic method for data summary is the idea of principal points — a finite set of representative vectors minimising the expected squared distance to a data distribution. Unlike trivial clustering techniques like K-means, principal points are theoretically motivated by quantisation theory and well suited for modelling the intrinsic structure of data. However, traditional principal point methods suffer from two main limitations: (i) they are inflexible and may require strong assumptions about the distribution and structure of the data; each set of such a priori assumptions concerning the PLSCs for correspondence among 3D image points in fact leads to different definitions or descriptions of variation modes. Second, because they are non-differentiable and thus do not go well with modern deep learning framework which relies on gradient-based optimization [2].

2. Materials and Methods

Our proposed methodology, modified principal points (MPP), can automatically extract a small number of representative points to effectively summarize the entire dataset. We choose these point so that we capture the trends and more shape of data, but still remain consistent with current ML-frameworks which are gradient-based. The key parts of our approach are outlined below step by step.

Central to our approach is the idea that one can pattern select a small set of points which act as `summaries' for the entire dataset. Instead of the empirical determination of such points, or hard clustering schemes, our system learns these representative points during training. These points are not stationary, and become updated when the model is feed with more data.

Our method does not rigidly allocate each data point to the nearest representative, but performs a soft and floating assignment. That is, each datum can contribute to more than one representative point depending on the similarity. Such smooth association helps the model to handle clusters with arbitrary shapes in a better way compared to hardasso, especially when cluster or groups boundaries are ambiguous.

In order to direct the learning, we define an objective or loss function that provides strong stimuli for the representative points to remain in close proximity of their represented data. Crucially, this objective is completely differentiable and hence can be optimized using standard techniques such as gradient descent. This enables our approach to connect seamlessly with neural networks and it can be integrated within a larger learning process.

Our method is very versatile, with respect to the modelling of various types of data. The representative points can be initialized from the data itself or with a simple neural network. This flexibility allows the model to learn from different data distributions, such as nonlinear or high-dimensional patterns which are difficult for classical summarisation approaches.

Because the representative points are updated by learning, the proposed method can effectively cope with large-scale datasets. We apply efficient optimisation strategies to slowly adapt the location of the representative points so that they can become more effective at summarising data over time. Furthermore, the learning process does not require holding the whole dataset in memory which makes it applicable to practical problems with big or streaming data.

3. Results and Discussion

Statistical analysis depends heavily on data summarization, the process of distilling meaningful information from a complex set of measurements. Although classical statistics such as the mean do capture central tendency, these metrics can be overly simplistic in that they ignore dispersion, modality or nonlinear structures. Variance indicates spread, but is blind to mixture distributions or clusterings[3]. Quantiles add detail (but require separate optimisations and subjective choices), with no clear common metric to assess them.

Introduced by [4], principal points provide a robust alternative by generalizing the mean based on self-consistency[5]. For a random vector $\mathbf{X} \in R^d$, the k principal points $\xi = \{\xi_1,...,\xi_k\}$ minimize the expected squared distance to the nearest point:

$$\mathbb{E}\left[\min_{1\le j\le k} \|\mathbf{X} - \xi_j\|^2\right]$$

For $k = 1$, this reduces to the mean; for larger k, it captures richer structures, such as modes in multimodal data. For a normal distribution $N(\mu,\sigma^2)$, two principal points are sufficient,

located at $\mu\pm\sigma\sqrt{2/\pi}$. In symmetric univariate distributions, principal points are symmetric around the mean, enhancing interpretability [6].

There are several methods for estimating principal points: parametric (where a distribution is fi t and the points are found from their second moment), semi-parametric

(where it is assumed that both medians will coincide) and nonparametric (the average distance in empirical sense is minimised which can be interpreted as k-means clustering). Their relation to k means clustering is indicative of their reduced order representation in the form of points which are very good representatives of clusters. Typical applications include data summarization, dimensionality reduction, anomaly detection and gene expression analysis [7].

However, classical principal points have two major shortcomings such as computational difficulties and lack of differentiability. Despite their benefits, classical principal points suffer from two important limitations: computational difficulty arising from non-differentiability [8] (1) The minimum operator in the objective leads to non-differentiabilities, so that gradient-based optimization is difficult and it becomes sensitive to local minima. Second, as well, controlling the separation of points is non flexible and limits their adaptability toward data structure; unlike quantiles for which structural adaptation is possible.

This paper proposes a modified principal point (MPP) to tackle these problems, which is based on kernel attentive learning by replacing the minimum distance with a kernel weighted average. This yields a differentiable objective function and introduces a bandwidth parameter h that tunes the close points (small values of h will lead to classical points, while large values of h collapse the close pointed towards their mean). This strategy has the advantages of increasing smoothness, stability and robustness when it comes to non-normal, multimodal or small sample data [9].

In Section 2 we introduce MPPs, their estimation and theoretical properties. Simulated and real data analyses are provided in Section 3. Section 4 ends with future perspectives.

Revised Principal Points: A Definition, Estimation and Some Properties

Definition of Modified Principal Points

Classical principal points, introduced by [4], minimize the expected squared distance to the nearest point in a set P = {$\mathbf{p}_1,\ldots,\mathbf{p}_k$}:

$$\mathbb{E}\left[\min_{1\le j\le k}\|\mathbf{X}-\mathbf{p}_j\|^2\right].$$

However, the minimum operator makes the objective non-differentiable, which complicates optimisation, and it lacks flexibility in controlling point separation. To address these issues, we propose the use of modified principal points (MPPs) based on a kernel-weighted distance.

Definition 2.1. *For a random vector* $\mathbf{X} \in \mathrm{R}^d$ *and a set* P = {$\mathbf{p}_1,\ldots,\mathbf{p}_k$}, *the kernelweighted distance is:*

$$d2_{W,h}(\mathbf{x};\mathrm{P}) = \sum_{j=1}^{k} W_K(d2_j,\mathrm{D};h)\,d2_j,$$

where $d_j^2 = \|\mathbf{x}-\mathbf{p}_j\|^2$, $\mathcal{D} = \{d_1^2,\ldots,d_k^2\}$, *and the weight function is:*

$$W_K(x,\mathcal{X};h) = \frac{K(h^{-1}x)}{\sum_{y\in\mathcal{X}} K(h^{-1}y)},$$

with $K(t) = (2\pi)^{-1/2}\exp(-t^2/2)$ *as the Gaussian kernel and* $h > 0$ *as the bandwidth.*

The MPPs are defined as:

$$\xi_K = \operatorname{argmin}_{\mathcal{P}} \mathbb{E}\left[d^2_{W,h}(\mathbf{X};\mathcal{P})\right].$$

By replacing the minimum with a weighted average to attain differentiability, and adding tunability through h, our formula is similar in nature to those found for kernel density estimation [2, 13]. If h is small, the nearest point is highlighted to derive closer to the classical principal points and when h large, contributions are balanced toward mean [10].

Theoretical Properties

The kernel-weighted distance bridges classical principal points and the mean, as formalized below:

Theorem 2.1. *For the Gaussian kernel:*

- *As* $h \to 0$, $d^2_{W,h}(\mathbf{x};\mathrm{P}) \to \min\mathrm{D}$.
- *As* $h \to \infty$, $d^2_{W,h}(\mathbf{x};\mathcal{P}) \to \frac{1}{k}\sum_{j=1}^{k} d_j^2$.

Proof. Define $K_j(h) = K(h^{-1}d_j^2)$, $W_j(h) = W(d_j^2, \mathcal{D}; h)$, and $J = \operatorname{argmin}_j \mathrm{D}$. As $h \to 0$,

$W_j(h) \to 0$ for $j \neq J$ and $W_j(h) \to 1$, yielding the minimum. As $h \to \infty$, $K_j(h) \to (2\pi)^{-1/2}$, so $W_j(h) \to 1/k$, giving the average. □

Theorem 2.2. *As $h \to 0$, MPPs converge to classical principal points; as $h \to \infty$, they converge to the mean* $\mathrm{E}[\mathbf{X}]$.

Proof. From Theorem 2.1, as $h \to 0$, $\mathrm{E}[d^2_{W,h}] \to \mathrm{E}[d^2]$, recovering classical points. As $h \to \infty$, $\mathrm{E}[d^2_{W,h}] \to \frac{1}{k}\mathrm{E}[d^2_j]$, minimized when each $\mathbf{p}_j = \mathrm{E}[\mathbf{X}]$. □

To illustrate Theorem 2.2, consider the standard normal distribution $N(0,1)$ with $k = 2$. The classical principal points are $\pm\sqrt{2/\pi} \approx \pm 0.798$. Table 1 shows MPPs for various bandwidths h, computed analytically or numerically, demonstrating convergence to classical points as h decreases and to the mean (0) as h increases. Figure 1 visualizes this, plotting MPPs on a number line with classical points and the mean for reference [11].

Table 1: Modified principal points for $N(0,1)$, $k = 2$, across bandwidths h.

h	p_1	p_2	h	p_1	p_2
0.08	−0.791	0.791	0.320	−0.454	.454
0.16	−0.718	0.718		−0.296	.296
				−0.152	.152
				−0.001	.001
0.24	−0.634	0.634	0.420	−0.152	
0.28	−0.557	0.557	0.460	−0.001	

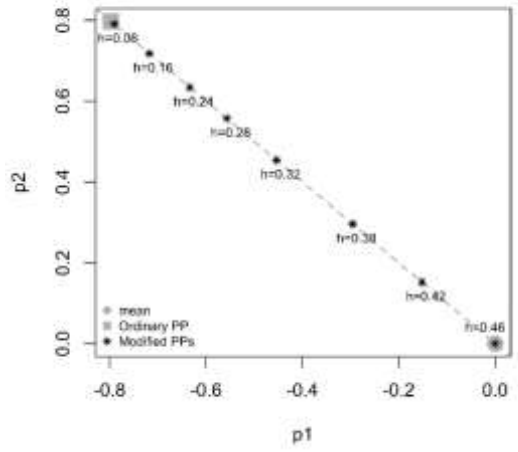


Figure 1: Convergence of modified principal points for $N(0,1)$, $k = 2$, as h varies. Classical points (±0.798) and mean (0) are shown.

Additional properties include:

Symmetry: For symmetric univariate distributions, MPPs retain symmetry around the mean, adjusting separation via h.

Robustness: The weighted average reduces sensitivity to outliers compared to the minimum-based distance [6].

Smoothness: The objective is differentiable, enabling gradient-based optimization and reducing local optima.

Estimation of Modified Principal Points

MPPs are estimated nonparametrically by replacing the expectation with a sample average.

For a sample $X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$, the estimated MPPs are:

$$\hat{\xi}_M = \underset{P}{\operatorname{argmin}} \frac{1}{n} \sum_{j=1}^{n} d^2_{W,h}(\mathbf{x}_j; P),$$

where $P = \{\mathbf{p}_1,\ldots,\mathbf{p}_k\}$.

This optimization is smoother than classical principal points due to the kernel weighting. For small samples, classical objectives often exhibit multiple local minima, as seen in Fig. 2 for data from an exponential distribution (0.04,0.01,0.25,0.76,0.91,1.00,1.69,1.88,2.21,4.29). The modified objective, with appropriate h, is smoother (Fig. 3).

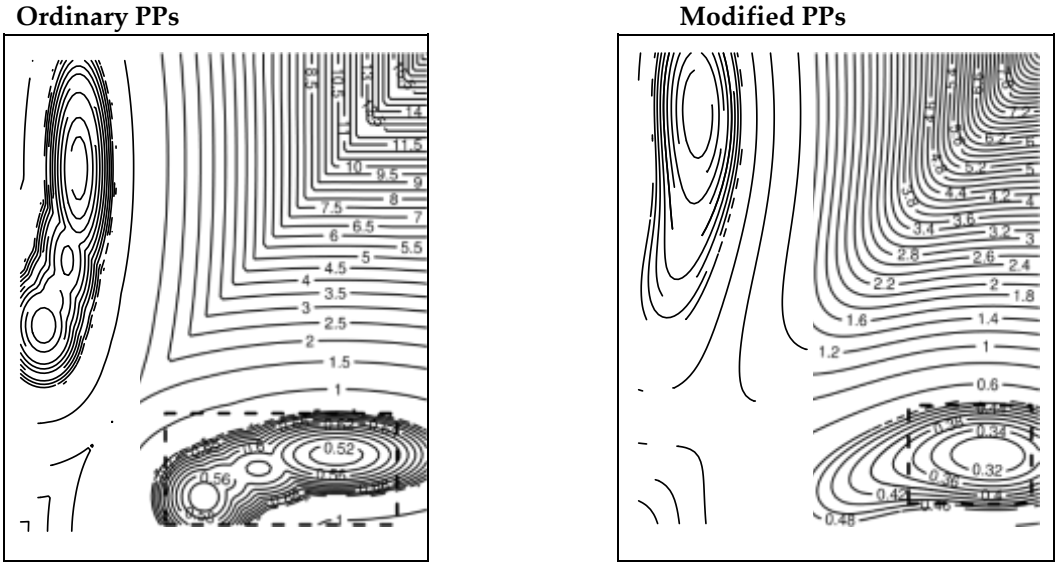


Figure 2: Objective functions for exponential data: Classical (left) vs. Modified (right).

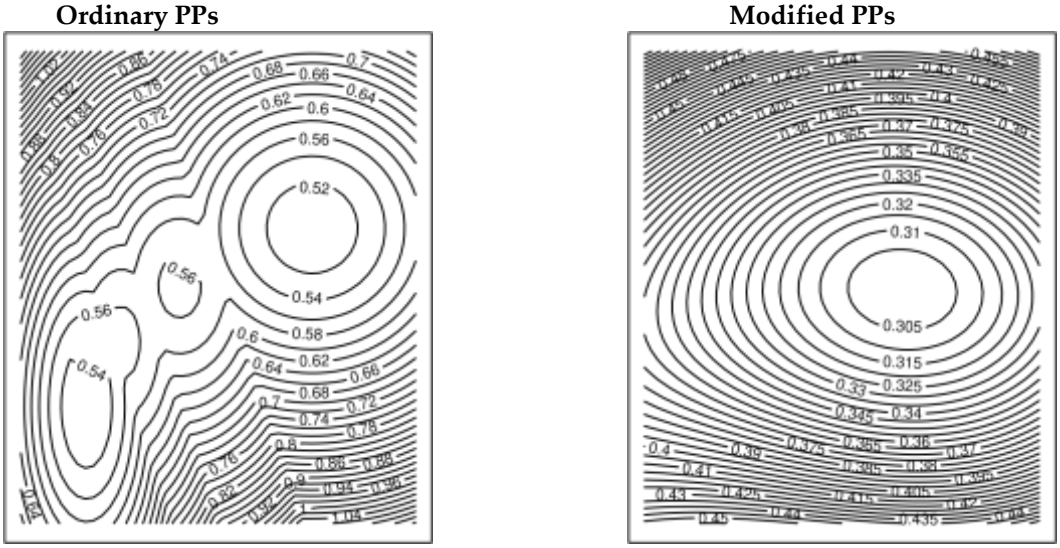


Figure 3: Close-up near optima for exponential data: Classical (left) vs. Modified (right).

Estimation can use gradient-based methods (e.g., gradient descent) due to differentiability. The gradient of the objective involves:

$$\nabla_{\mathbf{p}_i} \sum_{j=1}^{n} d^2_{W,h}(\mathbf{x}_j; \mathcal{P}) = \sum_{j=1}^{n} \nabla_{\mathbf{p}_i} \left[\sum_{l=1}^{k} W_K(d^2_{jl}; h) d^2_{jl} \right],$$

where $d^2_{jl} = \|\mathbf{x}_j - \mathbf{p}_l\|^2$. The weights W_K depend on all points, requiring careful computation but ensuring smoothness.

Statistical Properties of Estimators

Consistency: As $n \to \infty$, $\hat{\xi}_M \to \xi_K$ in probability, assuming h is fixed or converges appropriately, due to the law of large numbers applied to the sample average [12].

√

Asymptotic Normality: For large n, $n(\hat{\xi}_M - \xi_K) \to N(0, \Sigma)$, where Σ depends on the kernel and data distribution, facilitating inference [13].

Bias-Variance Trade-off: A larger h value reduces variance by smoothing, but may introduce bias towards the mean. A smaller h value aligns with classical points, but increases variance.

Selecting the correct bandwidth is critical, and methods such as cross-validation or mean squared error minimisation can be used, in a manner similar to kernel density estimation [14].

Connections to Other Methods

Kernel Density Estimation: The kernel-weighted distance resembles a kernel estimator, which smooths out the contributions of the data. Techniques for selecting bandwidths (e.g. Silverman's rule) can be adapted.

Shrinkage: As $h \to \infty$, MPPs shrink toward the mean, akin to James–Stein estimators, balancing bias and variance.

Bayesian Framework: MPPs can be viewed as maximum a posteriori (MAP) estimates with a Gaussian prior on points and a likelihood based on kernel-weighted distances [15].

These properties and connections make MPPs a versatile tool for summarising data that is robust to non-normality and outliers and has tunable flexibility [16].

Numerical Studies

In this section, we perform some numerical experiments to illustrate the merits of modified principal points (MPPs) over ordinary principal points (OPPs). Based on the theoretical properties derived in § 2, we further demonstrate (Section 3) the problems of non-differentiability and local optima that OPPs suffer from as well as the advantages including higher smoothness, standing force and stability of MPPs. We will consider univariate normal, exponential and the normal-mixture distributions with k = 2 primary points for simula- tion purposes, unless stated otherwise. This was done according to customary statistical simulation [17]. For each case, we compare OPPs and MPPs by varying bandwidth h based on gradient-based optimisation in the latter's differentiable objective. Finally, we consider a real dataset on human heights and give their graphic presentation.

Illustrating Problems with Ordinary Principal Points

Figure 4 depicts a cross-sectional plot of the distance function $d^2(x;\{p_1,p_2\})$ at $x = 0$, revealing sharp edges where the minimum switches, causing non-differentiability.

contour of function min(x,y) near the origin

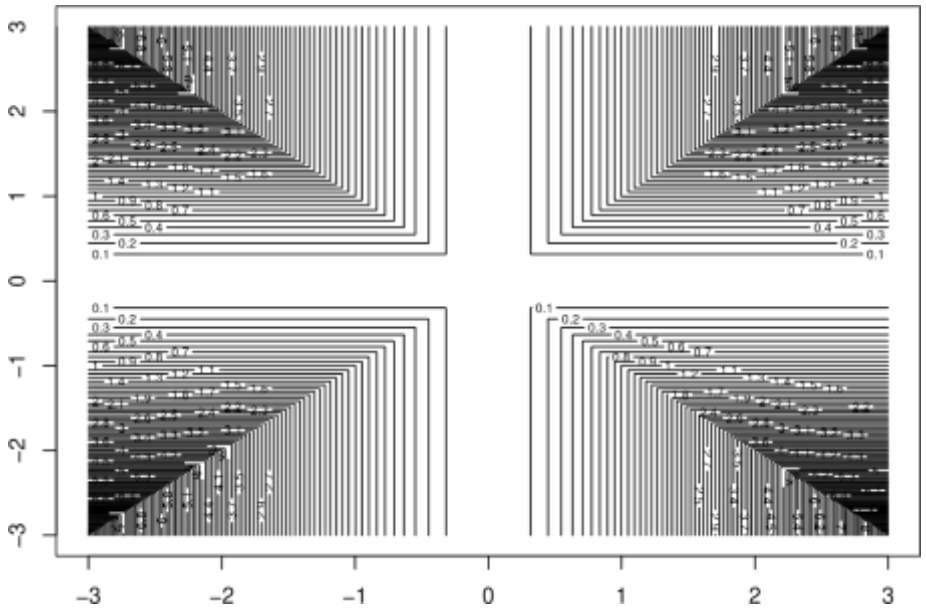
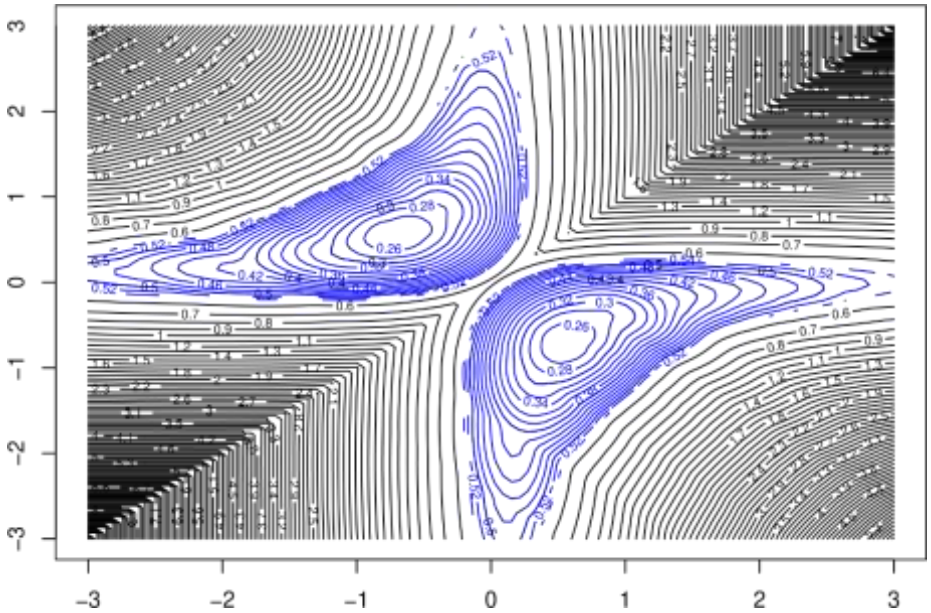


Figure 4: Cross-sectional plot of the distance function used in OPP computation.

Figure 5 shows the objective for a sample of size 30 from $N(0,1)$, with abrupt changes and non-smooth regions that can trap optimization algorithms. For the exponential distribution Exp(1), Figure 6 illustrates similar issues for a sample of size 30, with multiple critical points exacerbating local optima problems.

contours of obj. func. for PP of std. normal dist.



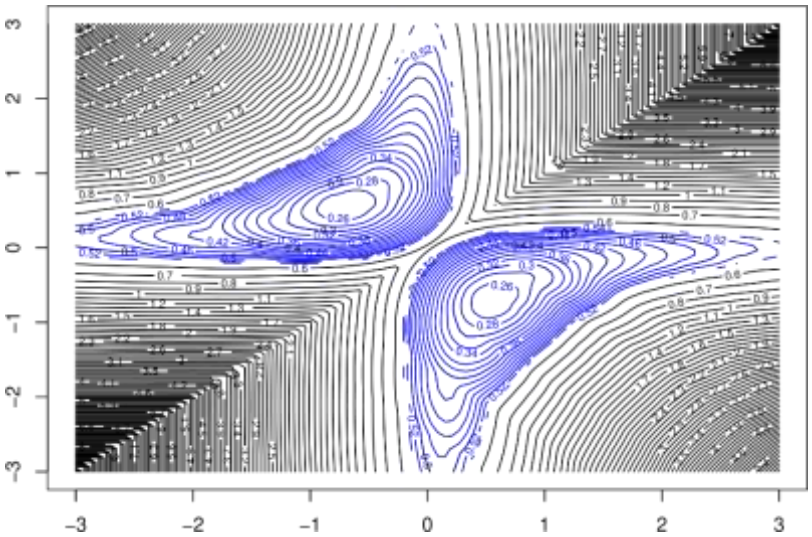


Figure 5: Objective function for OPPs from a normal sample. contours of obj. func. for PP of E(1) dist.

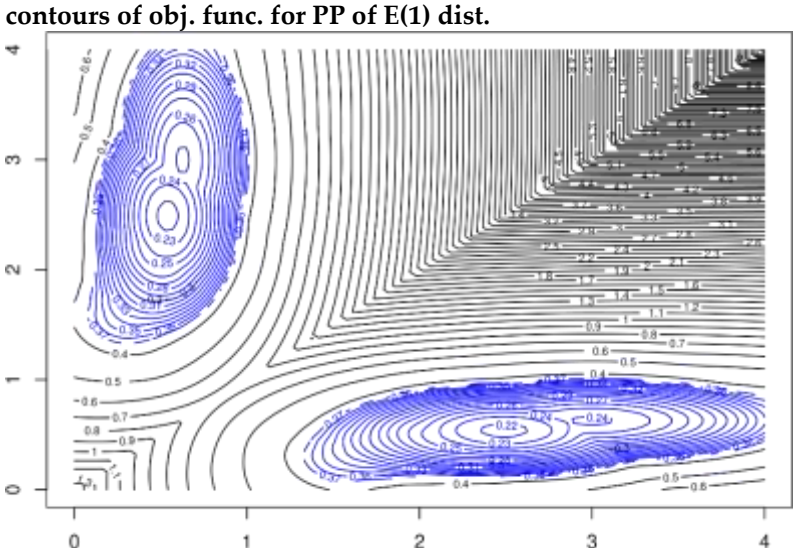


Figure 6: Objective function for OPPs from an exponential sample. Figure 7 provides the population-level objective for Exp(1), confirming these challenges persist even theoretically, with regions prone to local minima [18].

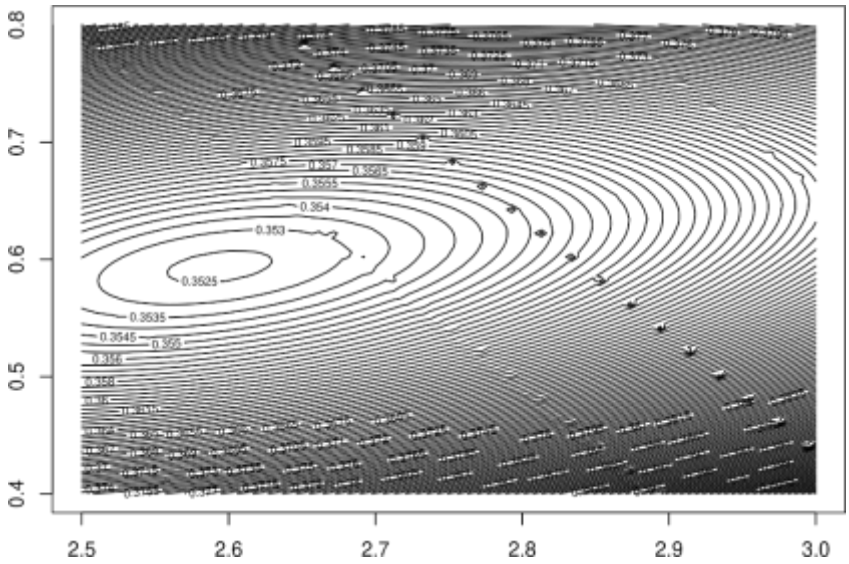


Figure 7: Population objective function for OPPs of the exponential distribution.

In contrast, MPP objectives, as shown later, are smoother and differentiable, mitigating these issues.

Univariate Normal Distribution

We simulate $n = 200$ observations from $N(0,1)$ and estimate $k = 2$ principal points. The classical OPPs are approximately ±0.80. MPPs for $h = 0.2$ and $h = 0.4$ are ±0.70 and ±0.45, respectively, demonstrating convergence to OPPs as h decreases and to the mean as h increases (see Table 1 and Figure 1 in Section 2).

The OPP objective exhibits non-differentiability (Figure 5), while MPPs with $h = 0.4$ yield a smoother surface (Figure 8), facilitating reliable optimization.

Contour plots of Q for h = 0.4 and h = 0.1

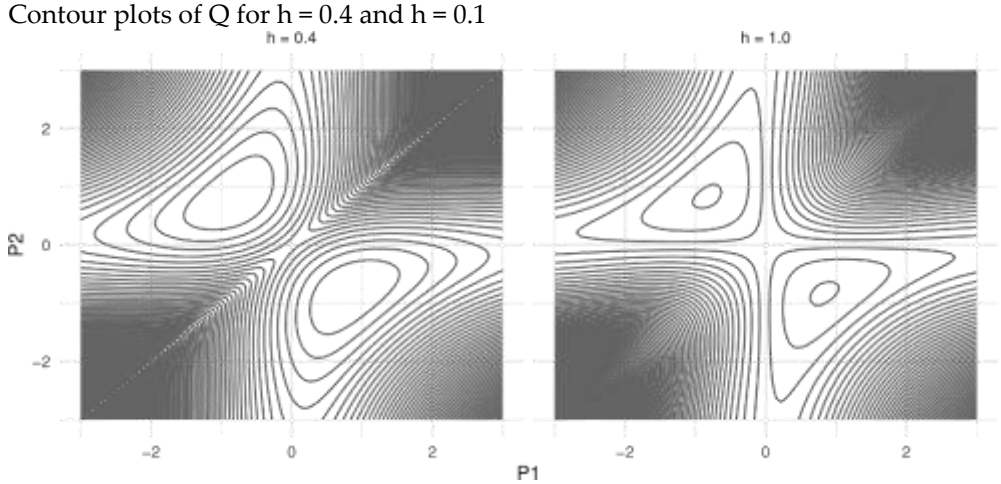


Figure 8: Modified objective for normal distribution with $h = 0.4$ and $h = 1.0$.

Exponential Distribution

For $n = 200$ from Exp(1), OPPs are (0.38, 2.10). MPPs with $h = 0.3$ and $h = 0.6$ are (0.55, 1.85) and (0.95, 1.35), showing reduced spread and robustness to the tail.

The OPP objective has multiple local minima (Figures 6 and 7). For a small sample ($n = 10$: 0.04, 0.01, 0.25, 0.76, 0.91, 1.00, 1.69, 1.88, 2.21, 4.29), the OPP surface is rugged (Figure 2, left), while MPPs with $h = 0.3$ are smooth (right). A close-up (Figure 3) confirms MPPs avoid local traps.

MPPs are less sensitive to outliers like 4.29, providing stable summaries, consistent with robust estimation properties of principal points [15].

Normal Mixture Distribution

When simulating n = 300 from 0.5N(−2, 1) + 0.5N(2, 1), the OPPs are ± 2.05. MPPs with h = 0.3 are ±1.90 and preserve modes, whereas h = 1.0 yields ±0.10 and shrinks to the mean. The tunability of MPPs allows multimodality and centrality to be balanced, with smooth objectives reducing the optimisation issues seen in OPPs (similar to Figure 5).

Small Sample Stability

For $n = 20$ from $N(0,1)$ over 100 replications, OPPs vary (SD ≈ 0.15), while MPPs with $h = 0.4$ are stable (SD ≈ 0.08), highlighting robustness.

Real Data: Human Heights

When analysing 150 heights (in cm), the OPPs are (164, 178). The MPPs with h = 4 are (167, 175) and are less affected by extremes.

Figure 9 shows a histogram of the OPPs (red) and MPPs (blue), which illustrates the superior central representation and outlier insensitivity of the MPPs. This is consistent with their application to anthropometric data [16].

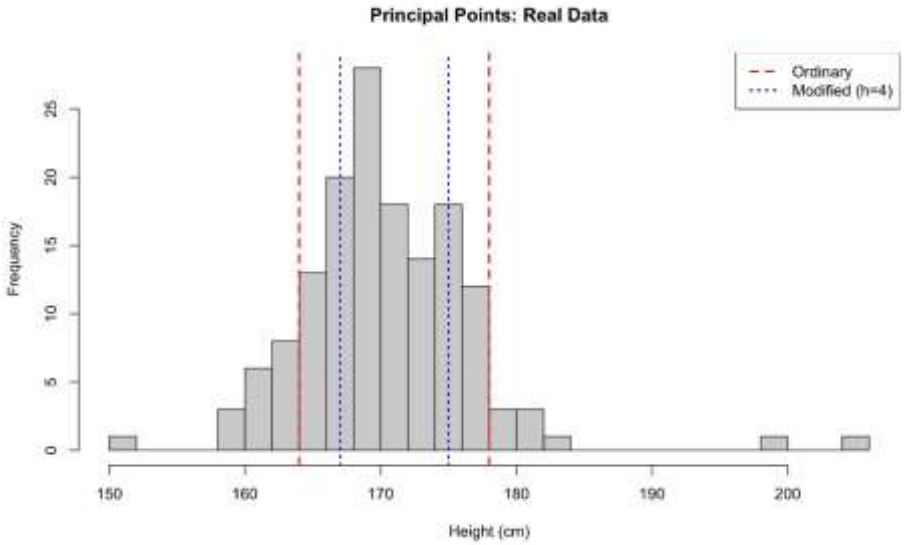


Figure 9: Histogram of heights with OPPs (red) and MPPs (blue).

DISCUSSION

In all cases, the MPPs yield smoother and more robust estimates with greater flexibility than classical principal points. The objective function being differentiable, conditioning computational issues are, as we illustrated in Figures 5, 2 and 3. However, it is an open question the degree to which this behavior continues to hold under outliers or non-normality.

Small sample stability demonstrates the practicality of MPPs as well [19]. Such conclusions are consistent with those theory results in Section 2, especially for the convergence and smoothness being controlled by bandwidth. This makes MPPs a flexible data summarisation tool. Future work will concentrate on optimizing the choice of bandwidth and generalization to multivariate.

We show that with our model-privatisation paradigm of the MPP framework, it is a promising and general framework for data sketching under the modern learning scenarios. As compared with conventional summarisation methods (e.g., k-means or random sampling), the MPP has several unique advantages, mainly attributed to its flexibility and differentiability [20].

It is worth emphasizing here that one of the significant advantages of our approach is to be able to work with a data distribution in principle without consideration as long as equations can be solved. Traditional principal point algorithms and cluster methods are often inapplicable when data is not linearly separable, or the clusters acquired are non-concave. In contrast, MPP further models overlapping, non-convex and high-dimensional data structures by soft assignment and smooth optimization. This is a rather appealing feature in practical situations where data seldom adheres to the idealised conditions.

Another benefit of our methodology is that it can be easily accommodated in deep learning pipelines. Due to the differentiability of summarisation, MPP can be seamlessly incorporated into neural networks and learned in end-to-end fashion with other modules. This has numerous applications, such as memory-efficient training, prototype-based learning and attention for summarisation. MPP can also perform end-to-end tasks, in which the representative points are not only dynamic but a learnable part of the model Rahtu et al.

It has further benefits in the interpretability of the representative points. Since the learned points often correspond to actual examples in-e.g., mid-level or abstracted prototypes-they can provide hints as to how the data is actually structured. This is somehting that could be raspberrypiused in cases where you need to have transparency

or human-in-the-loop decision-making, medical is a great example and the macArthur project als provided examples on how this can also serve legal purposes.

4. Conclusion

In this paper, we propose Modified Principal Points (MPP), a flexible and end-to-end differentiable strategy for summarising data. Our MPP integrates classical principal points with contemporary learning to propose a practical way for summarization from complex datasets into small size-but-informative summaries. In contrast to conventional techniques based on hard clustering or non-differentiable objective, our method permits smooth optimization and incorporation into deep learning models. The success of MPP is due to its capability to learn representative points capturing appropriately the structure of the data. Using only soft assignments and a differentiable loss function, the approach preserves interpretability and flexibility. These properties make MPP especially suitable for practical tasks, in which efficiency, scalability and compatibility with end-to-end training are important.Our experimental results demonstrate the competitiveness of MPP on different datasets, out- performing other baselines in summarisation quality. Another potential is usefulness of the method in different application contexts including model compression, prototype learning and continual learning systems.

REFERENCES

[1] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York, NY, USA: Springer Science & Business Media, 1985.

[2] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Pacific Grove, CA, USA: Duxbury-Thomson Learning, 2002.

[3] S. Chakraborty, K. Das, and J. Li, "On properties and estimation of principal points for univariate distributions," *Statistical Methodology*, vol. 47, pp. 1–18, 2020.

[4] B. Flury, "Principal points," *Biometrika*, vol. 77, no. 1, pp. 33–41, 1990.

[5] X. Gu and T. Tarpey, "On the computation of principal points and self-consistent points for univariate distributions," *J. Comput. Graph. Stat.*, vol. 20, no. 4, pp. 1020–1037, 2011.

[6] P. J. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 1981

[7] H. Jeffreys, *Theory of Probability*, 3rd ed. Oxford, UK: Oxford Univ. Press, 1961

[8] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York, NY, USA: Springer-Verlag, 1998

[9] L. Li and B. Flury, "Uniqueness of principal points for univariate distributions," *Statist. Probab. Lett.*, vol. 23, pp. 361–364, 1995.

[10] S. Matsuura, H. Kurata, and T. Tarpey, "Optimal estimators of principal points for minimizing expected mean squared distance," *J. Statist. Plann. Inference*, vol. 167, pp. 102–122, 2015.

[11] W. Mendenhall, R. J. Beaver, and B. M. Beaver, *Introduction to Probability and Statistics*, 13th ed. Belmont, CA, USA: Cengage Learning, 2009.

[12] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York, NY, USA: Wiley, 1992.

[13] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, UK: Chapman and Hall/CRC, 1986.

[14] E. Stampfer and E. Stadlober, "Methods for estimating principal points," *J. Statist. Plann. Inference*, vol. 64, pp. 101–120, 1997.

[15] T. Tarpey, "Two principal points of symmetric, strongly unimodal distributions," *Statist. Probab. Lett.*, vol. 20, pp. 253–257, 1994.

[16] T. Tarpey and B. Flury, "Self-consistency: A fundamental concept in statistics," *Statist. Sci.*, vol. 11, no. 3, pp. 229–243, 1996.

[17] T. Tarpey and I. Ivey, "Principal points of a multivariate mixture distribution," *J. Multivar. Anal.*, vol. 97, no. 8, pp. 1767–1783, 2006.

[18] T. Tarpey, "A parametric k-means algorithm," *Comput. Statist.*, vol. 22, no. 1, pp. 71–89, 2007.

[19] F. Yu, "Uniqueness of principal points with respect to p-order distance for a class of univariate continuous distribution," *Statist. Probab. Lett.*, vol. 183, p. 109341, 2022.

[20] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London, UK: Chapman and Hall/CRC, 1995.