# DD$\alpha$-classification of asymmetric and fat-tailed data

Tatjana Lange* Karl Mosler**
Pavlo Mozharovskyi**
*Hochschule Merseburg, 06217 Merseburg, Germany
**Universität zu Köln, 50923 Köln, Germany

January 8, 2013

**Abstract**

The DD$\alpha$-procedure is a fast nonparametric method for supervised classification of $d$-dimensional objects into $q \geq 2$ classes. It is based on $q$-dimensional depth plots and the $\alpha$-procedure, which is an efficient algorithm for discrimination in the depth space $[0, 1]^q$. Specifically, we use two depth functions that are well computable in high dimensions, the zonoid depth and the random Tukey depth, and compare their performance for different simulated data sets, in particular asymmetric elliptically and $t$-distributed data.

## 1 Introduction

Classical procedures for supervised learning like LDA or QDA are optimal under certain distributional settings. To cope with more general data, nonparametric methods have been developed. A point may be assigned to that class in which it has maximum depth (Ghosh and Chaudhuri (2005), Hoberg and Mosler (2006)). Moreover, data depth is suited to reduce the dimension of the data and aggregate their geometry in an efficient way. This is done by mapping the data to a depth-depth (DD) plot or, more generally, to a DD-space: a unit cube of dimension $q \geq 2$, where each axis indicates the depth w.r.t. a certain class (e.g. see Fig. 1, left and middle). A proper classification rule is then constructed in the DD-space, see Li et al. (2012). In Lange et al. (2012) the DD$\alpha$-classifier is introduced, which employs a modified version of the $\alpha$-procedure (Vasil'ev (1991), Vasil'ev (2003), Vasil'ev and Lange (1998) and Lange and Mozharovskyi (2012)) for classification in the DD-space. For other recent depth-based approaches, see Dutta and Ghosh (2012a), Dutta and Ghosh (2012b), Paindaveine and Van Bever (2012); all need intensive computations.

However, when implementing the DD$\alpha$-classifier a data depth has to be chosen and the so called 'outsiders' have to be treated in some way. (An *outsider* is a data point that has depth 0 in each of the classes.) The present paper also addresses the question which notion of data depth should be employed. To answer it we consider two depth notions that can be efficiently calculated in higher dimensions and explore their sensitivity to fat-tailedness and asymmetry of the underlying class-specific distribution. The two depths are the zonoid depth (Koshevoy and Mosler (1997), Mosler (2002)) and the location depth (Tukey (1975)). The zonoid depth is always exactly computed, while the location depth is either exactly or approximately calculated. In a large simulation study the average error rate of different versions of the DD$\alpha$-procedure is contrasted with that of standard classifiers, given data from asymmetric and fat-tailed distributions. Similarly the performance of different classifiers is explored depending on the distance between the classes, and their speed both at the training and classification stages is investigated. We restrict ourselves to the case $q = 2$, see Lange et al. (2012) for $q > 2$. Outsiders are randomly assigned with equal probabilities; for alternative treatments of outsiders, see Hoberg and Mosler (2006) and Lange et al. (2012).

The rest of the paper is organized as follows. Sect. 2 shortly surveys the depths notions used in the DD$\alpha$-procedure and their computation. In Sect. 3 the results of the simulation studies are presented and analyzed, regarding performance, performance dynamics and speed of the proposed classifiers. Sect. 4 concludes.

# 2 Data depths for the DD$\alpha$-classifier

Briefly, a data depth measures the centrality of a given point $\mathbf{x}$ in a data set $X$ in $\mathbb{R}^d$; see e.g. Zuo and Serfling (2000) and Mosler (2002) for properties.

## 2.1 Zonoid *vs.* location depth

In the sequel we employ two data depths that can be efficiently computed for high-dimensional data ($d = 20$ and higher): the zonoid depth and the location depth. The computational aspect plays a significant role here as the depths have to be calculated for each point of each class at the training stage, and they still have to be computed for a

new point w.r.t. each class at the classification stage.
**Zonoid depth.** For a point $\mathbf{x}$ and a set $X = \{\mathbf{x}^1, \dots, \mathbf{x}^n\} \in \mathbb{R}^d$ the zonoid depth is defined as

$$ZD(\mathbf{x}, X) = \begin{cases} \sup\{\alpha : \mathbf{x} \in ZD_\alpha(X)\} & \text{if } \mathbf{x} \in ZD_\alpha(X) \text{ for some } 0 < \alpha \leq 1, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

with the zonoid region (2)

$$ZD_\alpha(X) = \left\{ \sum_{i=1}^{n} \lambda_i \mathbf{x}^i \,:\, 0 \leq \lambda_i \leq \frac{1}{n\alpha}, \sum_{i=1}^{n} \lambda_i = 1 \right\}; \tag{2}$$

see Koshevoy and Mosler (1997) and Mosler (2002) for properties.
**Location depth,** also known as halfspace or Tukey depth, is defined as

$$HD(\mathbf{x}, X) = \frac{1}{n} \min_{\mathbf{u} \in S^{d-1}} \sharp\{i : \langle \mathbf{x}^i, \mathbf{u} \rangle \geq \langle \mathbf{x}, \mathbf{u} \rangle\}, \tag{3}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. The location depth takes only discrete values and is robust (having a large breakdown point), while the zonoid depth takes all values in the set $\{[\frac{1}{n}, 1] \cup \{0\}\}$, is maximum at the mean $\frac{1}{n} \sum_i \mathbf{x}^i$, and therefore less robust. For computation of the zonoid depth we use the exact algorithm of Dyckerhoff et al. (1996).

## 2.2 Tukey depth *vs.* random Tukey depth

The location depth can be exactly computed or approximated. *Exact computation* is described in Rousseeuw and Ruts (1996) for $d = 2$ and in Rousseeuw and Struyf (1998) for $d = 3$. For bivariate data we employ the algorithm of Rousseeuw and Ruts (1996) as implemented in the R-package 'depth'. In higher dimensions exact computation of the location depth is possible (Liu and Zuo (2012)), but the algorithm involves heavy computations. Cuesta-Albertos and Nieto-Reyes (2008) instead propose to approximate the location depth, using (3), by minimizing the univariate location depth over randomly chosen directions $u \in S^{d-1}$. Here we explore two different settings where the set of randomly chosen $u$ is either *generated once and for all* or *generated instantly* when computing the depth of a given point. By construction, the random Tukey depth is always greater or equal to the exact location depth. Consequently, it yields fewer outsiders.

# 3 Simulation study

A number of experiments with simulated data is conducted. Firstly, the error rates of 17 different classifiers (see below) are evaluated on data from asymmetric $t$- and exponential distributions in $\mathbb{R}^2$. Then the performance dynamics of selected ones is visualized as the classification error in dependence of the Mahalanobis distance between the two classes. The third study explores the speed of solutions based on the zonoid and the random Tukey depth.

## 3.1 Performance comparison

While the usual multivariate $t$-distribution is elliptically symmetric, it can be made asymmetric by conditioning its scale on the angle from a fixed direction, see Frahm (2004). For each degree of freedom, $\infty$ (= Gaussian), 5 and 1 (= Cauchy), two alternatives are investigated: one considering differences in location only (with $\mu_1 = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$, $\mu_2 = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ and $\Sigma_1 = \Sigma_2 = \left[\begin{smallmatrix}1&1\\1&4\end{smallmatrix}\right]$) and one differing in both location and scale (with the same $\mu_1$ and $\mu_2$, $\Sigma_1 = \left[\begin{smallmatrix}1&1\\1&4\end{smallmatrix}\right]$, $\Sigma_2 = \left[\begin{smallmatrix}4&4\\4&16\end{smallmatrix}\right]$), skewing the distribution with reference vector $v_1^+ = (\cos(\pi/4), \sin(\pi/4))$, see Frahm (2004) for details. Further, the bivariate Marshall-Olkin exponential distribution (BOMED) is looked at: $(\min\{Z_1, Z_3\}, \min\{Z_2, Z_3\})$ for the first class and $(\min\{Z_1, Z_3\} + 0.5, \min\{Z_2, Z_3\} + 0.5)$ for the second one with $Z_1 \sim Exp(1)$, $Z_2 \sim Exp(0.5)$, and $Z_3 \sim Exp(0.75)$. Each time we generate a sample of 400 points (200 from each class) to train a classifier and a sample containing 1000 points (500 from each class) to evaluate its performance (= error rate).

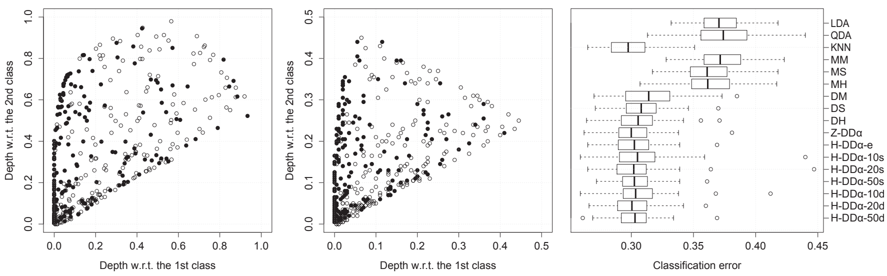


Figure 1: DD-plots using zonoid (left) and location (middle) depth with black and open circles denoting observations from the two different classes and combined box-plots (right) for Gaussian location alternative

DD-plots of a training sample for *the Gaussian location alternative* using zonoid (left) and location (middle) depth are shown in Fig. 1. For each classifier, training and testing is performed on 100 simulated data sets, and a box-plot of error rates is drawn; see Fig. 1 (right). The first group of non-depth classifiers includes linear (LDA) and quadratic (QDA) discriminant analysis and $k$-nearest-neighbors classifier (KNN). Then the maximal depth classifiers (MM, MS and MH; cf. Ghosh and Chaudhuri (2005)) and the DD-classifiers (DM, DS and DH; cf. Li et al. (2012)) are regarded. Each triplet uses the Mahalanobis (Mahalanobis (1936), Zuo and Serfling (2000)), simplicial (Liu (1990)) and location depths, respectively. The remaining eight classifiers are DD$\alpha$-classifiers based on zoonoid depth (Z-DD$\alpha$), exactly computed location depth (H-DD$\alpha$-e), random Tukey depth for once-only (H-DD$\alpha$-$\sharp$s) and instantly (H-DD$\alpha$-$\sharp$d) generated directions, each time using $\sharp = 10, 20, 50$ random directions, respectively. The combined box-plots together with corresponding DD-plots using zonoid and location depth are presented for *the Cauchy location-scale alternative* (Fig. 2) and *the BOMED location alternative* (Fig. 3).

Based on these results (and many more not presented here) we conclude: In many cases DD$\alpha$-classifiers, both based on the zonoid depth and the random Tukey depth, are better than their competitors. The versions of the DD$\alpha$-classifier that are based on the random Tukey depth are not outperformed by the exact computation algorithm. There is no noticeable difference between the versions of the DD$\alpha$-classifier based on the random Tukey depth using same directions and an instantly generated direction set. The statement 'the more random directions we use, the better classification we achieve' is not necessarily true with the DD$\alpha$-classifier based on the random Tukey depth, as the portion of outsiders and their treatment are rather relevant.

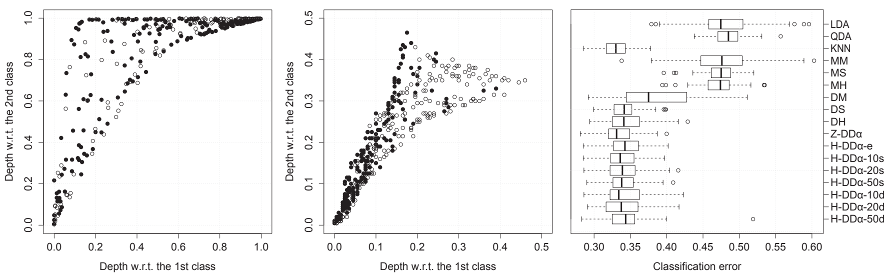


Figure 2: DD-plots using zonoid (left) and location (middle) depth and combined box-plots (right) for Cauchy location-scale alternative

## 3.2 Performance dynamics

To study the performance dynamics of the various DD$\alpha$-classifiers in contrast with existing classifiers we regard $t$-distributions with $\infty$, 5 and 1 degrees of freedom, each in a symmetric and an asymmetric version (see Sect. 3.1). The Mahalanobis distance between the two classes is systematically varied. At each distance the average error rate is calculated over 100 data sets and five shift directions in the range $[0, \pi/2]$.

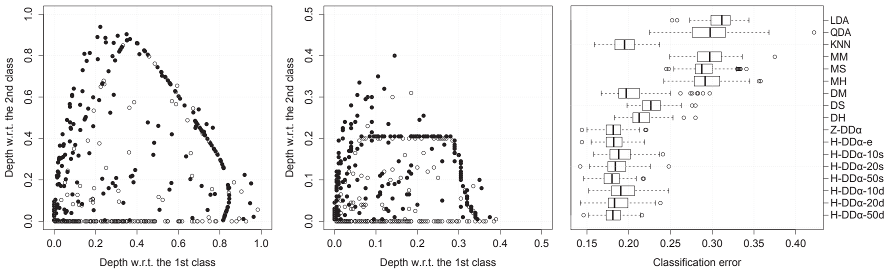


Figure 3: DD-plots using zoonid (left) and location (middle) depth and combined box-plots (right) for BOMED location alternative

(As we consider two classes and have one reference vector two symmetry axes arise.) By this we obtain curves for the classification error of some of the classifiers considered in Sect. 3.1, namely LDA, QDA, KNN, all DD$\alpha$-classifiers, and additionally those using five constant and instantly generated random directions. The results for two extreme cases, Gaussian distribution (left) and asymmetric conditional scale Cauchy distribution (right) are shown in Fig. 4.

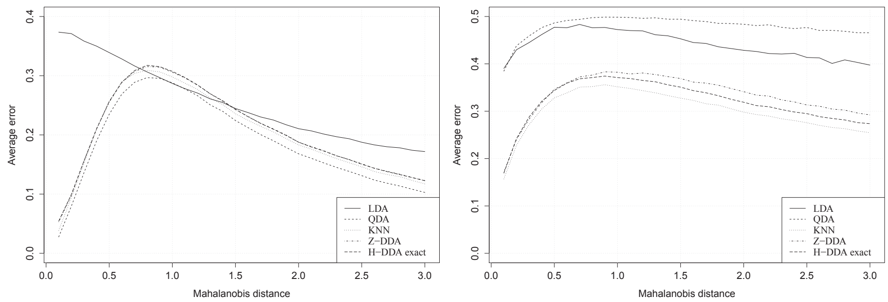


Figure 4: Performance dynamic graphs for Gaussian (left) and asymmetric conditional scale Cauchy (right) distributions

Under bivariate elliptical settings (Fig. 4, left) QDA, as expected from theory, outperforms other classifiers and coincides with LDA when the Mahalanobis distance equals 1. DD$\alpha$-classifiers suffering from outsiders perform worse but similarly, independent of the number of directions and the depth notion used; they are only slightly outperformed by KNN for the 'upper' range of Mahalanobis distances. (Note that KNN does not have the 'outsiders problem'.) But when considering an asymmetric fat-tailed distribution (Fig. 4, right), neither LDA nor QDA perform satisfactorily. The DD$\alpha$-classifiers are still outperformed by KNN (presumably because of the outsiders). They perform almost the same for different numbers of directions. The DD$\alpha$-classifier based on zonoid depth is slightly outperformed by that using location depth, which is more robust.

### 3.3 The speed of training and classification

The third task tackled in this paper is comparing the speed of the DD$\alpha$-classifiers using zonoid and random Tukey depth, respectively. (For the latter we take 1000 random directions and do not consider outsiders.) Two distributional settings are investigated: $\mathrm{N}(\mathbf{0}_d, \mathbf{I}_d)$ *vs.* $\mathrm{N}(0.25 \cdot \mathbf{1}_d, \mathbf{I}_d)$ and $\mathrm{N}(\mathbf{0}_d, \mathbf{I}_d)$ *vs.* $\mathrm{N}((0.25\ \mathbf{0}'_{d-1})', 5 \cdot \mathbf{I}_d)$, $d = 5, 10, 15, 20$. For each *pair of classes* and *number of training points* and *dimension* we train the classifier 100 times and test each of them using 2500 points. Average times in seconds are reported in Table 1.

Table 1: The average speed of training and classification (in parentheses) using the random Tukey depth, in seconds

| | $\mathrm{N}(\mathbf{0}_d, \mathbf{I}_d)$ vs $\mathrm{N}(0.25 \cdot \mathbf{1}_d, \mathbf{I}_d)$ | | | | $\mathrm{N}(\mathbf{0}_d, \mathbf{I}_d)$ vs $\mathrm{N}((0.25\ \mathbf{0}'_{d-1})', 5 \cdot \mathbf{I}_d)$ | | | |
|---|---|---|---|---|---|---|---|---|
| | d=5 | d=10 | d=15 | d=20 | d=5 | d=10 | d=15 | d=20 |
| $n = 200$ | 0.1003<br>(0.00097) | 0.097<br>(0.00073) | 0.0908<br>(0.00033) | 0.0809<br>(0.00038) | 0.0953<br>(0.00098) | 0.0691<br>(0.0005) | 0.0702<br>(0.00034) | 0.0699<br>(0.00025) |
| $n = 500$ | 0.2684<br>(0.00188) | 0.2551<br>(0.00095) | 0.2532<br>(0.00065) | 0.252<br>(0.00059) | 0.2487<br>(0.0019) | 0.2049<br>(0.00096) | 0.1798<br>(0.00065) | 0.1845<br>(0.00049) |
| $n = 1000$ | 0.6255<br>(0.00583) | 0.6014<br>(0.00289) | 0.5929<br>(0.00197) | 0.5846<br>(0.00148) | 0.5644<br>(0.0058) | 0.5476<br>(0.00289) | 0.4414<br>(0.00197) | 0.4275<br>(0.00148) |

Table 1 and the distributional settings correspond to those in Lange et al. (2012), where a similar study has been conducted with the zonoid depth. We also use the same PC and testing environment. Note firstly that the DD$\alpha$-classifier with the random Tukey depth requires substantially less time to be trained than with the zonoid depth. The time required for training increases almost linearly with the cardinality of the training set, which can be traced back to the structure of the algorithms used for the random Tukey depth and for the $\alpha$-procedure. The time decreases with dimension, which can be explained as follows: The $\alpha$-procedure takes most of the time here; increasing $d$ but leaving $n$ constant increases the number of points outside the convex hull of one of the training classes, that is, having depth $= 0$ in this class; these points are assigned to the other class without calculations by the $\alpha$-procedure.

## 4 Conclusions

The experimental comparison of the DD$\alpha$-classifiers, using the zonoid depth and the random Tukey depth, on asymmetric and fat-tailed distributions shows that in general both depths classify rather well, the random Tukey depth performs not worse than the zonoid depth and sometimes even outperforms it (cf. Cauchy distribution), at least in two dimensions. Though both depths can be efficiently computed, also for higher dimensional data, the random Tukey depth is computed much faster. Still when employing the random Tukey depth the number of random directions has to be selected; this as well as a proper treatment of outsiders needs further investigation.

# References


[1] CUESTA-ALBERTOS, J.A. and NIETO-REYES, A. (2008): The random Tukey depth. *Computational Statistics and Data Analysis, 52, 4979–4988.*

[2] DUTTA, S. and GHOSH A.K. (2012a): On robust classification using projection depth. *Annals of the Institute of Statistical Mathematics, 64, 657–676.*

[3] DUTTA, S. and GHOSH A.K. (2012b): On classification based on $L_p$ depth with an adaptive choice of $p$. *Technical Report Number R5/2011, Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata, India.*

[4] DYCKERHOFF, R, KOSHEVOY, G and MOSLER, K. (1996): Zonoid data depth: Theory and computation. In: A. Prat (Ed.): *COMPSTAT 1996 - Proceedings in Computational Statistics.* Physica, Heidelberg, 235–240.

[5] FRAHM, G. (2004): Generalized elliptical distributions: theory and applications. *Doctoral thesis.* University of Cologne.

[6] GHOSH, A.K. and CHAUDHURI, P. (2005): On maximum depth and related classifiers. *Scandinavian Journal of Statistics, 32, 327–350.*

[7] HOBERG, R. and MOSLER, K. (2006): Data analysis and classification with the zonoid depth. In: R. Liu, R. Serfling and D. Souvaine (Eds.): *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications.* American Mathematical Society, 49–59.

[8] KOSHEVOY, G. and MOSLER, K. (1997): Zonoid trimming for multivariate distributions. *Annals of Statistics, 25, 1998–2017.*

[9] LANGE, T., MOSLER, K. and MOZHAROVSKYI, P. (2012): Fast nonparametric classification based on data depth. *Statistical Papers* (in print).

[10] LANGE, T., MOZHAROVSKYI, P. (2012): The Alpha-Procedure - a nonparametric invariant method for automatic classification of $d$-dimensional objects. *36th Annual Conference of the German Classification Society, Hildesheim.*

[11] LI, J., CUESTA-ALBERTOS J.A. and LIU R.Y. (2012): *DD*-classifier: Nonparametric classification procedure based on *DD*-plot. *Journal of the Americal Statistical Association, 107, 737–753.*

[12] LIU, R. (1990): On a notion of data depth based on random simplices. *Annals of Statistics, 18, 405–414.*

[13] LIU, X. and ZUO, Y. (2012): Computing halfspace depth and regression depth. Mimeo.

[14] MAHALANOBIS, P. (1936): On the generalized distance in statistics. *Proceedings of the National Institute of Science of India, 2, 49–55.*

[15] MOSLER, K. (2002): *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach.* Springer Verlag, New York.

[16] PAINDAVEINE, D. and VAN BEVER, G. (2012): Nonparametrically consistent depth-based classifiers. `arXiv:1204.2996v1 [math.ST]`.

[17] ROUSSEEUW, P.J. and RUTS, I. (1996): Bivariate location depth. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 45, 516–526.*

[18] ROUSSEEUW, P.J. and STRUYF, A: (1998): Computing location depth and regression depth in higher dimensions. *Statistics and Computing, 8, 193–203.*

[19] TUKEY, J.W. (1975): Mathematics and the picturing of data. *Proc. Inter. Cong. Math., Vancouver, 523–531.*

[20] VASIL'EV, V.I. (1991): The reduction principle in pattern recognition learning (PRL) problem. *Pattern Recognition and Image Analysis 1, 23–32.*

[21] VASIL'EV, V.I. (2003): The reduction principle in problems of revealing regularities I. *Cybernetics and Systems Analysis 39, 686–694.*

[22] VASIL'EV, V.I. and LANGE, T. (1998): The duality principle in learning for pattern recognition (in Russian). *Kibernetika i Vytschislit'elnaya Technika 121, 7–16.*

[23] ZUO, Y. and SERFLING, R. (2000): General notions of statistical depth function. *Annals of Statistics 28, 462–482.*