Stat Papers (2011) 52:511–522
DOI 10.1007/s00362-009-0265-1

REGULAR ARTICLE

Generalized Mahalanobis depth in the reproducing kernel Hilbert space

Yonggang Hu · Yong Wang · Yi Wu · Qiang Li · Chenping Hou

Received: 21 November 2008 / Accepted: 26 June 2009 / Published online: 5 August 2009

Abstract In this paper, Mahalanobis depth (MHD) in the Reproducing Kernel Hilbert Space (RKHS) is proposed. First, we extend the notion of MHD to a generalized version, i.e., the generalized MHD (GHMD), to make it suitable for the small sample with singular covariance matrix. We prove that GMHD is consistent with MHD when the sample has a full-rank covariance matrix. Second, we further extend GMHD to RKHS, i.e, the kernel mapped GMHD (kmGMHD), and discuss its main properties. Numeric results show that kmGMHD can give a better depth interpretation for the sample with special shape, such as a non-convex sample set. Our proposed kmGMHD can be potentially used as a robust tool for outliers detection and data classification. In addition, we also discuss the influence of parameters on the shape of the central regions.

Keywords Mahalanobis depth · Kernel method · Data depth · Statistical depth

Mathematics Subject Classification (2000) 62H05 · 62H30

1 Introduction

Statistical depth functions are widely used in nonparametric derivations for multivariate data. A depth function suitable for a distribution F in R^p, denoted by $D(\mathbf{x}; F)$, brings out the non-central ranking of $\mathbf{x}$ in R^p with respect to F. Depth

This work is supported by the NSFC Grant # 60673090 and # 60802079.

Y. Hu (✉) · Y. Wang · Y. Wu · Q. Li · C. Hou
Department of Mathematics and Systems Science, National University of Defense Technology,
410073 Changsha, People's Republic of China
e-mail: xd7688@hotmail.com

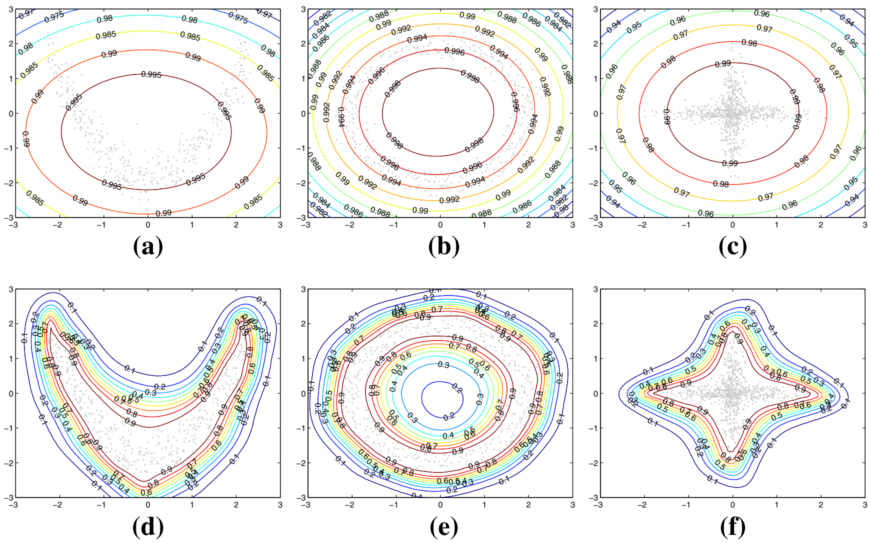

Fig. 1 The central regions of MHD and kmGMHD relative to the three types of data sets: **a** MHD w.r.t *clown* data, **b** MHD w.r.t *ring* data, **c** MHD w.r.t *cross* data, **d** kmGMHD with Gaussian kernel, $\sigma = 1.5$, **e** kmGMHD with Gaussian kernel, $\sigma = 1.5$, **f** kmGMHD with Gaussian kernel, $\sigma = 1.5$

functions have been successfully used in many fields, such as quality index (Liu 1992; Liu and Singh 1993), multivariable regression (Tian et al. 2002), limiting p values (Liu and Singh 1997), robust estimation (Pennacchi 2008; Chen and Tyler 2004), nonparametric test (Chenouri 2004) and discriminant analysis (Ghosh and Chaudhuri 2005b,a; Jörnsten 2004; Cui et al. 2008). A number of depth functions are available in the literature. A few are *halfspace depth* (Tukey 1975), *simplicial depth* (Liu 1990), *projection depth* (Stahel 1981; Donoho 1982), *spatial depth* (Vardi and Zhang 2000), *spatial rank depth* (Gao 2003), etc.

Among others, Mahalanobis depth (Mahalaobis 1936; Zuo and Serfling 2000) has attracted many scholars' attentions because of its convenience in computation. However, Mahalanobis depth is only suitable for the sample generated from elliptical distributions, or at least the convex data sets. For instance, as shown in Fig. 1a, b and c, whether the data sets are convex or not, the central regions of MHD are always convex. Obviously, such depth interpretation does not agree with our intuitions. Moreover, MHD requires the covariance matrix being positive definite, thus it can not deal with data sets with sparse samples of which the covariance matrix is singular. These deficiencies greatly limit the further applications of MHD.

Enlightened by the excellent performance of kernel method (Taylor and Cristianini 2004; Hofmann et al. 2008), where the data are processed in a Reproducing Kernel Hilbert Space (RKHS, denoted as $\mathcal{H}_k$), we firstly extend the definition of MHD to the sparse data sets with singular covariance matrix, i.e., the generalized MHD (GMHD), and then extend GMHD to the kernel mapped feature space, denoted as kmGMHD. The main properties of GMHD and kmGMHD are also discussed.

We will now describe our notation and give some background materials. All vectors will be column vectors unless transposed to a row vector by a ^{T}. $\mathbf{x}$ denotes a vector in R^p and $Y = \{Y_i, i = 1, \ldots, n\}$ is a data set in R^p, where n is the sample size and p is the dimensionality. The notation $k(\mathbf{x}, \mathbf{y})$ is a kernel function for $\mathbf{x}, \mathbf{y} \in R^p$, and $\phi(\cdot)$ is its associated mapping function. We will use $\mathscr{H}_k$ to denote the corresponding RKHS (also called *feature space*) generated by $\phi(\cdot)$, and $\mathscr{Y} = \{\mathscr{Y}_i, i = 1, \ldots, n\}$ is the image set of Y in $\mathscr{H}_k$. And through out this paper, we mainly consider the inner product kernel functions, especially the following two kernel functions:

1. Gaussian kernel function (also called the RBF kernel function)

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right), \quad \sigma > 0.$$

2. Polynomial kernel function: $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T\mathbf{y} + 1)^d, d \geq 1$.

The paper is arranged as follows. Section 2 briefly introduces the Mahalanobis Depth. Section 3 extends the definition of MHD to GMHD. Section 4 gives the definition of kmGMHD. Section 5 analyzes the main properties of kmGMHD and gives some numeric experiments. The influence of parameters in kernel function is discussed in Section 6, and Section 7 is a conclusion.

2 Mahalanobis depth

Zuo and Serfling (2000) proposed four desirable properties for the depth functions and discussed how to construct a depth function. As an example, they defined the Mahalanobis depth. Suppose X is a p-variate random variable, and $X \sim F$, then the Mahalanobis depth of $\mathbf{x} \in R^p$ with respect to F is defined as

$$\mathrm{MHD}(\mathbf{x}; F) \doteq \left(1 + d^2_{\Sigma_F}(\mathbf{x}, \mu_F)\right)^{-1}, \tag{1}$$

where $d^2_{\Sigma_F}(\mathbf{x}, \mu_F) = (\mathbf{x} - \mu_F)^T \Sigma_F^{-1}(\mathbf{x} - \mu_F)$, and μ_F and Σ_F are any corresponding location and covariance measures, respectively. The case that μ_F and Σ_F are the mean and covariance matrix of F was suggested by Liu (1992). For these choices, however, $\mathrm{MHD}(\mathbf{x}; F)$ is not robust, and it can fail to achieve maximum value at the center of A-symmetric distributions.

The sample version of $\mathrm{MHD}(\mathbf{x}; F)$ is $\mathrm{MHD}(\mathbf{x}; F_n)$, where F_n is the empirical distribution of $Y_1, \ldots, Y_n$, specifically,

$$\mathrm{MHD}(\mathbf{x}; F_n) = \left(1 + d^2_{\Sigma_{F_n}}(\mathbf{x}, \mu_{F_n})\right)^{-1}. \tag{2}$$

MHD satisfies all of the four desirable properties of Zuo and Serfling (2000).

3 Generalized Mahalanobis depth

In the definition of Mahalanobis depth, the covariance matrix Σ_{F_n} is required to be positive, which may not be satisfied in practice. For example, suppose $F_n = \{X_1, \ldots, X_n\}$,

$X_i \in R^p$ and $p > n$, then the covariance matrix Σ_{F_n} is singular, i.e., it has no inverse. Thus we can not directly use formula (2) to compute MHD($\mathbf{x}$, F_n).

However, suppose the rank of $A = (X_1 - \overline{X}, \ldots, X_n - \overline{X})^T$ is r, where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then there exists an orthogonal transformation $\mathbf{H}_{p\times p}$ such that $(X_i - \overline{X})^T\mathbf{H} = (\hat{x}_i^1, \ldots, \hat{x}_i^r, 0, \ldots, 0)_{1\times p}^T$, namely, we can completely describe F_n in a r-dimensional subspace. Let F_n^r denote the corresponding low dimensional expression in R^r, then the Mahalanobis depth of X_i with respect to F_n can be naturally represented by the depth of $X_i^r \doteq (\hat{x}_i^1, \ldots, \hat{x}_i^r)$ with respect to F_n^r since they have the same structure. Now we will give the definition of a generalized Mahalanobis depth as

Definition 1 Suppose $F_n = \{X_1, \ldots, X_n\}$, $X_i \in R^p$, r, $\mathbf{H}$, X_i^r and F_n^r are same as the above notations, then the generalized Mahalanobis depth (GMHD) is defined as

$$\mathrm{GMHD}(\mathbf{x}, F_n) = \mathrm{MHD}\left((\mathbf{x} - \overline{X})^T\mathbf{H}_{p\times r}, F_n^r\right), \tag{3}$$

where $\mathbf{H}_{p\times r}$ is the first r columns of $\mathbf{H}$.

Theorem 1 *Let σ_i^2 be the nonzero eigenvalue of A^TA ($i \le r$), and $V_{p\times r} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r)$, where $\mathbf{v}_i$ is the corresponding eigenvector of σ_i^2 ($i = 1, \ldots, r$) and r is the rank of Σ_{F_n}, then*

$$\mathrm{GMHD}(\boldsymbol{x}, F_n) = \left(1 + \sum_{i=1}^{r} \frac{\left((\boldsymbol{x} - \overline{X})^T\boldsymbol{v}_i\right)^2}{\sigma_i^2}\right)^{-1}. \tag{4}$$

Proof Consider the singular value decomposition of A^TA. There exists an orthogonal matrix $V_{p\times p}$ such that $A^TA = V diag(D, \mathbf{0})V^T$, where $D = diag(\sigma_1^2, \ldots, \sigma_r^2)$.

Obviously, $(X_i - \overline{X})^T\mathbf{v}_i = 0$ for any $r < i \le p$. Then for any point $\mathbf{x} \in R^p$, its corresponding low-dimensional expression in the space R^r, spread by $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is $(\mathbf{x} - \overline{X})^T\ (\mathbf{v}_1, \ldots, \mathbf{v}_r) \doteq (\mathbf{x} - \overline{X})^T V_{p\times r}$. Thus

$$\begin{aligned}
&(\mathbf{x} - \overline{X})^T V_{p\times r}\left[(AV_{p\times r})^T AV_{p\times r}\right]^{-1} V_{p\times r}^T(\mathbf{x} - \overline{X}) \\
&= (\mathbf{x} - \overline{X})^T V_{p\times r} D^{-1} V_{p\times r}^T(\mathbf{x} - \overline{X}) \\
&= \sum_{i=1}^{r} \frac{\left((\mathbf{x} - \overline{X})^T\mathbf{v}_i\right)^2}{\sigma_i^2}.
\end{aligned}$$

□

Theorem 2 *If A^TA has full rank, then $GMHD(\boldsymbol{x}, F_n) = MHD(\boldsymbol{x}, F_n)$.*

Remark 1 Theorem 2 shows that for nonsparse samples, GMHD is consistent with MHD. Thus, GMHD is an extended version of MHD, and is suitable for any data set.

Proposition 1 *Suppose r is the rank of Σ_{F_n}, and p is the dimensionality of the sample, then we have*

1. *If $r < p$, then GMHD satisfies three of the four desirable properties* (Zuo and Serfling 2000): vanishing at infinity*,* maximality at center *and* monotonicity relative to deepest point.
2. *If $r = p$, then GMHD satisfies all the four desirable properties like MHD.*

Proof (1). *Vanishing at infinity* and *maximality at center* are straightforward.

Monotonicity relative to the deepest point. Let $\mathbf{y} = \mathbf{x} - \overline{X} = t\mathbf{u}$, $t \in R^+$, $\mathbf{u}$ is an arbitrary vector, and $\|\mathbf{u}\| = 1$. Then,

$$\sum_{i=1}^{r} \frac{\left((\mathbf{x}-\overline{X})^T \mathbf{v}_i\right)^2}{\sigma_i^2} = \sum_{i=1}^{r} \frac{\left(t\mathbf{u}^T \mathbf{v}_i\right)^2}{\sigma_i^2} = t^2 \sum_{i=1}^{r} \frac{\left(\mathbf{u}^T \mathbf{v}_i\right)^2}{\sigma_i^2},$$

and the result follows.

(2). By Theorem 2, the conclusion is straightforward. □

Remark 2 If the rank of Σ_{F_n} is not full and $\mathbf{x} \notin F_n$, the structure between $\mathbf{x}$ and F_n is actually changed ($\mathbf{x}^T V_{p\times r}$ with F_n^r is only an approximation of $\mathbf{x}$ with F_n). In this case, GMHD does not satisfy *affine invariance* since $GMHD(\mathbf{x}, F_n)$ in nature is $MHD(\mathbf{x}^T V_{p\times r}, F_n^r)$. However, it still satisfies *translation invariance* and *orthogonal invariance*.

4 Kernel mapped GMHD

Now we consider the GMHD in $\mathcal{H}_k$. As we have pointed out, the specified expression of mapping function ϕ is not yet defined, so we usually do not know the exact coordinate of $\phi(Y_i)$. Thus, we can not directly use formula (4) to compute the GMHD depth in $\mathscr{H}_k$. However, since $\mathscr{H}_k$ is a *Hilbert space*, all the computations in $\mathscr{H}_k$ can be achieved by the kernel function $k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y})\rangle$.

In fact, we have the following result:

Theorem 3 *For a data set Y in R^p, if Y is mapped into $\mathscr{H}_k$ by a mapping function $\phi(\cdot)$, and $\phi(\cdot)$ is associated with the kernel function $k(\cdot,\cdot)$, then for any point $\boldsymbol{x} \in R^p$, the GMHD of $\phi(\boldsymbol{x})$ with respect to $\mathscr{Y}$ can be expressed as follows*

$$\mathrm{GMHD}(\phi(\boldsymbol{x}), \mathscr{Y}) = \left(1 + \sum_{i=1}^{r} \frac{\left(\left(S(\boldsymbol{x}, \boldsymbol{x}_j)\right)_{1\times n}^{j=1,\ldots,n} \boldsymbol{u}_i\right)^2}{\sigma_i^4}\right)^{-1}, \tag{5}$$

where $S(\boldsymbol{x}, \boldsymbol{x}_j) = k(\boldsymbol{x}, \boldsymbol{x}_j) - \frac{1}{n}\sum_{\ell=1}^{n} k(\boldsymbol{x}, \boldsymbol{x}_\ell) - \frac{1}{n}\sum_{\ell=1}^{n} k(\boldsymbol{x}_j, \boldsymbol{x}_\ell) + \frac{1}{n^2}\sum_{k=1}^{n}\sum_{\ell=1}^{n} k(\boldsymbol{x}_k, \boldsymbol{x}_\ell)$, σ_i^2 is the nonzero eigenvalue of $(S(\boldsymbol{x}_k, \boldsymbol{x}_\ell))_{n\times n}^{k,\ell=1,\ldots,n}$ with eigenvector $\boldsymbol{u}_i$, and r is the rank of $(S(\boldsymbol{x}_k, \boldsymbol{x}_\ell))_{n\times n}$.

Lemma 1 *If $A_{n\times m} \in R^{n\times m}$, r is the rank of A, σ_i^2 ($\sigma_i > 0, i \le r$) is the nonzero eigenvalue of AA^T and $\boldsymbol{v}_i \in R^n$ is its corresponding eigenvector, then σ_i^2 is also the eigenvalue of $A^T A$, and its corresponding eigenvector is $\frac{A^T \boldsymbol{v}_i}{\sigma_i}$.*

Proof (Theorem 3) Let $A = (\phi(\mathbf{x}_1) - \overline{\phi(\mathbf{x})}, \ldots, \phi(\mathbf{x}_n) - \overline{\phi(\mathbf{x})})^T_{n\times m}$, where $\overline{\phi(\mathbf{x})} = \frac{1}{n}\sum_{i=1}^n \phi(\mathbf{x}_i)$ and suppose $m \geq n$, then there exists an orthogonal matrix $U_{n\times n}$, such that

$$AA^T = (\mathbf{u}_1, \ldots, \mathbf{u}_n)\begin{pmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}_{n\times n}(\mathbf{u}_1, \ldots, \mathbf{u}_n)^T$$

$$\doteq U diag(D_{r\times r}, \mathbf{0})U^T,$$

where σ_i^2 is the eigenvalue of AA^T, and $\sigma_i^2 \neq 0$.

Then, by Lemma 1, $\frac{A^T\mathbf{u}_i}{\sigma_i}$ is the eigenvector of A^TA with respect to the eigenvalue σ_i^2. Thus, by Theorem 1, we have

$$\frac{\left((\phi(\mathbf{x}) - \overline{\phi(X)})^T\mathbf{v}_i\right)^2}{\sigma_i^2} = \frac{\left((\phi(\mathbf{x}) - \overline{\phi(X)})^T A^T\mathbf{u}_i\right)^2}{\sigma_i^4} = \frac{\left((S(\mathbf{x}, \mathbf{x}_j))_{1\times n}^{j=1,\ldots,n}\,\mathbf{u}_i\right)^2}{\sigma_i^4},$$

where $S(\mathbf{x}, \mathbf{x}_j) = k(\mathbf{x}, \mathbf{x}_j) - \frac{1}{n}\sum_{\ell=1}^n k(\mathbf{x}, \mathbf{x}_\ell) - \frac{1}{n}\sum_{\ell=1}^n k(\mathbf{x}_j, \mathbf{x}_\ell) + \frac{1}{n^2}\sum_{l=1}^n\sum_{\ell=1}^n k(\mathbf{x}_l, \mathbf{x}_\ell)$. Substituting it into (4) gives the result. □

Formula (5) actually defines a new depth function for $\mathbf{x} \in R^p$ relative to Y. Thus, we redefine it as:

Definition 2 The depth function defined in (5) is called *kernel mapped GMHD* (kmGMHD) and is denoted as $GMH_kD(\mathbf{x}, F_n)$.

Its distribution version can be defined as

Definition 3 $GMH_kD(\mathbf{x}, F) \doteq (1 + \sum_{i=1}^r \frac{((\phi(\mathbf{x})-E\phi(X))^T\mathbf{v}_i)^2}{\sigma_i^2})^{-1}$, where σ_i^2 and $\mathbf{v}_i$ are the nonzero eigenvalue and its corresponding eigenvector of $E(\phi(X) - E\phi(X))(\phi(X) - E\phi(X))^T$, respectively.

Lemma 2 *If $k(\cdot,\cdot)$ is the Gaussian kernel, or the polynomial kernel with $E\|X\|^d < \infty$, we have*

$$\Sigma_{F\circ\phi_n} \xrightarrow{a.s.} \Sigma_{F\circ\phi}, \tag{6}$$

Proof Let $\phi(X) = (\mathscr{X}_1, \ldots, \mathscr{X}_q)$, where q is the dimensionality of the feature space. Since

$$\int \|\phi(X)\|\, dF_X = \int \sqrt{k(X, X)} dF_X \doteq I,$$

Case I: $k(\cdot,\cdot)$ is the Gaussian kernel, then $k(X,X)=1$. Thus

$$I=\int dF_X=1<\infty.$$

Case II: $k(\cdot,\cdot)$ is the polynomial kernel, $k(X,X)=(X^TX+1)^d=(\|X\|^2+1)^d$. Since $E\|X\|^d<\infty$, then

$$I=\int\sqrt{\left(\|X\|^2+1\right)^d}dF_X<\infty.$$

Hence, by the Bochner integral, there exists a $\mathscr{X}$, such that $\mathscr{X}=\int\phi(X)dF_X\in\mathscr{H}_k$. Thus, for any $i,j=1,\ldots,q$, $E\mathscr{X}_i$ and $E(\mathscr{X}_i-E\mathscr{X}_i)(\mathscr{X}_j-E\mathscr{X}_j)$ exists. By the law of large numbers, the result follows. □

Theorem 4 *If $k(\cdot,\cdot)$ is the Gaussian kernel, or the polynomial kernel with $E\|X\|^d<\infty$, then*

$$|GMH_kD(\boldsymbol{x},F_n)-GMH_kD(\boldsymbol{x},F)|\xrightarrow{a.s.}0 \text{ as } n\to\infty. \tag{7}$$

Proof Suppose $\Sigma_{F\circ\phi_n}=V_nD_nV_n^T$, and $\Sigma_{F\circ\phi}=VDV^T$. According to Lemma 2 and Li and Zhang (1998), $V_n\xrightarrow{a.s.}V$ and $D_n\xrightarrow{a.s.}D$, that is, $\sigma_{ni}\xrightarrow{a.s.}\sigma_i$ and $\mathbf{v}_{ni}\xrightarrow{a.s.}\mathbf{v}_i$. Thus, by the Definition 3, the result follows. □

5 Main properties of kmGMHD

5.1 Properties for distribution version

Proposition 2 *For the depth function $GMH_kD(\boldsymbol{x};F)$, if $E(\phi(X)-E\phi(X))(\phi(X)-E\phi(X))^T$ exists, the following properties are satisfied*

1. *If $\|\phi(\boldsymbol{x})\|\to\infty$ as $\|\boldsymbol{x}\|\to\infty$, then $\lim_{\|\boldsymbol{x}\|\to\infty}GMH_kD(\boldsymbol{x};F)=0$; If for any $\boldsymbol{x}\in R^p$, $\|\phi(\boldsymbol{x})\|<C$, where C is a positive constant in R, $GMH_kD(\boldsymbol{x};F)$ has a nonzero depth at infinity;*
2. *If F is absolutely continuous, $GMH_kD(\boldsymbol{x};F)$ is continuous;*
3. *Let $\boldsymbol{H}$ be a $p\times p$ matrix satisfying $\boldsymbol{H}'\boldsymbol{H}=I_p$, and $b\in R^p$. If $k(\cdot,\cdot)$ is the Gaussian kernel, $GMH_kD(\boldsymbol{x};X)=GMH_kD(\boldsymbol{Hx}+b;\boldsymbol{H}X+b)$;*
4. *For any inner product kernel functions, if $\boldsymbol{H}'\boldsymbol{H}=I_p$, we always have $GMH_kD(\boldsymbol{Hx},\boldsymbol{H}X)=GMH_kD(\boldsymbol{x},X)$.*

Proof (1), (2) Straightforward.

(3) For the Gaussian kernel, $k(\mathbf{x},\mathbf{y})=k(\|\mathbf{x}-\mathbf{y}\|)$, thus

$$\begin{aligned}
k(\mathbf{Hx}+b,\mathbf{Hy}+b)&=k(\|\mathbf{Hx}-\mathbf{Hy}\|)\\
&=k(\sqrt{(\mathbf{Hx})'(\mathbf{Hx})+(\mathbf{Hy})'(\mathbf{Hy})-2(\mathbf{Hx})'(\mathbf{Hy})})\\
&=k(\sqrt{\mathbf{x}'\mathbf{x}+\mathbf{y}'\mathbf{y}-2\mathbf{x}'\mathbf{y}})=k(\|\mathbf{x}-\mathbf{y}\|)=k(\mathbf{x},\mathbf{y}).
\end{aligned}$$

Then, in the formula 5, $(S(\mathbf{Hx}+b, \mathbf{Hx}_j+b))_{1\times n}^{j=1,\ldots,n} = (S(\mathbf{x}, \mathbf{x}_j))_{1\times n}^{j=1,\ldots,n}$ and

$$(S(\mathbf{Hx}_k+b, \mathbf{Hx}_\ell+b))_{n\times n}^{k,\ell=1,\ldots,n} = (S(\mathbf{x}_k, \mathbf{x}_\ell))_{n\times n}^{k,\ell=1,\ldots,n}.$$

Thus, for any F_n, $GMH_kD(\mathbf{Hx}+b, \mathbf{H}F_n+b) = GMH_kD(\mathbf{x}, F_n)$. Finally, by Theorem 4, the result follows.

(4) In fact, for any inner product kernel function, it can always be expressed as

$$k(\mathbf{x}, \mathbf{y}) = k(\mathbf{x}'\mathbf{x}, \mathbf{y}'\mathbf{y}, \mathbf{x}'\mathbf{y}).$$

Thus, from (3), the result follows. □

Corollary 1 *KmGMHD which is based on the polynomial kernel vanishes at infinity; however, for Gaussian kernel, kmGMHD has a nonzero depth at infinity.*

Proposition 3 *For the depth function $GMH_kD(\boldsymbol{x}; F)$ defined in Definition 3, then if F is symmetric about the origin, $GMH_kD(\boldsymbol{x}; F) = GMH_kD(-\boldsymbol{x}; F)$.*

Proof By the Definition 3, we just prove that $(\phi(-\mathbf{x}) - E\phi(X))^T\mathbf{v}_i = (\phi(\mathbf{x}) - E\phi(X))^T\mathbf{v}_i$. Firstly, consider the following equations.

$$\begin{aligned}
&(\phi(\mathbf{x}) - E\phi(X))^T E(\phi(X) - E\phi(X))(\phi(X) - E\phi(X))^T(\phi(\mathbf{x}) - E\phi(X)) \\
&= E\left|(\phi(\mathbf{x}) - E\phi(X))^T(\phi(X) - E\phi(X))\right|^2 \\
&= \int \left|k(\mathbf{x}, X) - Ek(\mathbf{x}, X) - (E\phi(X))^T\phi(X) + \|E\phi(X)\|^2\right|^2 dF_X \\
&= \int \left|k(\mathbf{x}, Y) - Ek(\mathbf{x}, Y) - (E\phi(X))^T\phi(Y) + \|E\phi(X)\|^2\right|^2 dF_Y \\
&= \int \left|k(\mathbf{x}, Y) - Ek(\mathbf{x}, Y) - (Ek(X, Y))^T + \|E\phi(X)\|^2\right|^2 dF_Y \\
&\doteq f(\mathbf{x}).
\end{aligned}$$

Since

$$\begin{aligned}
Ek(-\mathbf{x}, X) &= \int k(-\mathbf{x}, X)dF_X = \int k\left(\|-\mathbf{x}\|^2, -\mathbf{x}^T X, \|X\|^2\right)dF_X \\
&= \int k\left(\|\mathbf{x}\|^2, \mathbf{x}^T(-X), \|-X\|^2\right)dF_X = \int k\left(\|\mathbf{x}\|^2, \mathbf{x}^T X, \|X\|^2\right)dF_X \\
&= Ek(\mathbf{x}, X)
\end{aligned}$$

and

$$E\phi(-X) = \int \phi(-X)dF_X = \int \phi(X)dF_X$$

then

$$
\begin{aligned}
f(-\mathbf{x}) &= \int \left|k(-\mathbf{x}, Y) - Ek(-\mathbf{x}, Y) - (Ek(X, Y))^T + \|E\phi(X)\|^2\right|^2 dF_Y \\
&= \int \left|k(\mathbf{x}, -Y) - Ek(\mathbf{x}, -Y) - (Ek(X, Y))^T + \|E\phi(X)\|^2\right|^2 dF_Y \\
&= \int \left|k(\mathbf{x}, Y) - Ek(\mathbf{x}, Y) - (Ek(X, -Y))^T + \|E\phi(X)\|^2\right|^2 dF_Y \\
&= \int \left|k(\mathbf{x}, Y) - Ek(\mathbf{x}, Y) - (Ek(X, Y))^T + \|E\phi(X)\|^2\right|^2 dF_Y = f(\mathbf{x}).
\end{aligned}
$$

At the same time, suppose the spectral decomposition of $\Sigma_{F\circ\phi} = E(\phi(X) - E\phi(X))(\phi(X) - E\phi(X))^T$ is denoted as

$$\Sigma_{F\circ\phi} = VDV^T,$$

where $D = diag(\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0)$ and σ_i^2 is the nonzero eigenvalue of $\Sigma_{F\circ\phi}$. then, $f(\mathbf{x})$ can be rewritten as

$$f(\mathbf{x}) = \sum_{i=1}^{r} \left((\phi(\mathbf{x}) - E\phi(X))^T \mathbf{v}_i\right)^2 \sigma_i^2.$$

Since for any $\mathbf{x} \in R^p$, $f(\mathbf{x}) = f(-\mathbf{x})$. Hence, we have

$$\left((\phi(\mathbf{x}) - E\phi(X))^T \mathbf{v}_i\right)^2 = \left((\phi(-\mathbf{x}) - E\phi(X))^T \mathbf{v}_i\right)^2 \quad i = 1, \ldots, r.$$

Thus, by formula (3), $GMH_kD(\mathbf{x}, F) = GMH_kD(-\mathbf{x}, F)$. □

5.2 The shape of the central regions

For the Gaussian kernel, $\|\phi(\mathbf{x}) - \phi(\mathbf{y})\| \to 0$ as $\|\mathbf{x} - \mathbf{y}\| \to 0$. Since $k(\mathbf{x}, \mathbf{x}) = \phi'(\mathbf{x})\phi(\mathbf{x}) = 1$, $\phi(\mathbf{x})$ is on the surface of the unit hypersphere. At the same time, Abe (2005) pointed out that $\phi(\cdot)$ is one-to-one, then G is homeomorphous with R^p, where G denotes the image of R^p. Thus if G enclosed by the $D^\alpha(\mathscr{Y})$ is discontinuous, its preimage in R^p is also discontinuous. Therefore, unlike the sample depth of $GMRD(\mathbf{x}, F_n)$, the central region of $GMH_kD(\mathbf{x}, F_n)$ may not be simply connected. However, we have

Proposition 4 *There exists $\alpha_0 \in (0, 1)$, such that for any $\alpha \le \alpha_0$, $D^\alpha(Y)$ is simply connected.*

Proof Suppose $Y = \{Y_1, \ldots, Y_n\}$ is a data set in R^p, and $\mathscr{Y} = \{\mathscr{Y}_1, \ldots, \mathscr{Y}_n\}$ is its images in the feature space $\mathscr{H}_k$. Consider GMHD in $\mathscr{H}_k$. Because the central region $D^\alpha(\mathscr{Y})$ in $\mathscr{H}_k$ is convex, there exists $\alpha_0 \in (0, 1)$ such that the unit hypersphere has only one continuous surface enclosed by $D^{\alpha_0}(\mathscr{Y})$. Thus there's also only one piece of G enclosed by $D^{\alpha_0}(\mathscr{Y})$. Denote such continuous area as $G_1 \subset G$, then $\phi^{-1}(G_1)$ is also continuous because $\phi(\cdot)$ is one-to-one and continuous. $\square$

Remark 3 Although the cases for other types of kernel functions are not discussed here, they have similar results with that of the Gaussian kernel.

5.3 Experiments

Figure 1 shows the central region of MHD and kmGMHD with respect to the following three types of data sets:

1. *clown*. Created by Canu et al. (2005), and we only take the *mouth* part;
2. *ring*. (ρ, θ), where $\theta \sim U(0, 2\pi)$, and $\rho \sim N(2, 0.2^2)$.
3. *cross*. Each half is created by $N(\mu, \Sigma_i)$, $i = 1, 2$, where

$$\Sigma_1 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.02 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 0.02 & 0 \\ 0 & 0.5 \end{pmatrix}, \quad \mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

For MHD, as we have discussed in Section 1, whatever the shape of the data's distribution is, the central regions are always convex. For the data cloud with convex shape, such depth distribution is reasonable. However, for the data with non-convex shape like those data sets in Fig. 1, the depth distribution of MHD does not agree with our intuition, such as Fig. 1a, b and c. For instance, in Fig. 1a, the MHD of point (0, 0) is relatively large, while from the figure we can see clearly that (0, 0) is really outside of the cloud, though it lies in the center of convex envelope. The same problem also exists in the other two figures. These depth interpretations for the special-shape data cloud are really worse.

Nonetheless, kmGMHD depth eliminates such embarrassments for us.The results are shown in Fig. 1d, e, and f. From these figures we can see that, the central regions of kmGMHD fit with the shape of the data cloud, and the deepest points of these three data sets no longer lie at the center of the convex hull, but at the center of data itself. However, it is important to note that, as discussed above, the central region of kmGMHD may not be simply connected for large values of the centrality parameter, and their medians are usually not unique.

Remark 4 It is important to note that the medians of kmGMHD have a little difference to those of MHD. Zuo and Serfling (2000) require that the deepest point should be the center of the distribution, while, from the above discussion we know that such definition sometimes is not proper, especially for the purpose of outlier detection or data classification.

6 The influence of the parameters of kernel function

The parameter in the kernel function, such as σ in the Gaussian kernel or d in the polynomial kernel, is mainly used to control the dimensionality of the feature space $\mathscr{H}_k$. For Gaussian kernel, the dimensionality of its feature space is infinity. However,

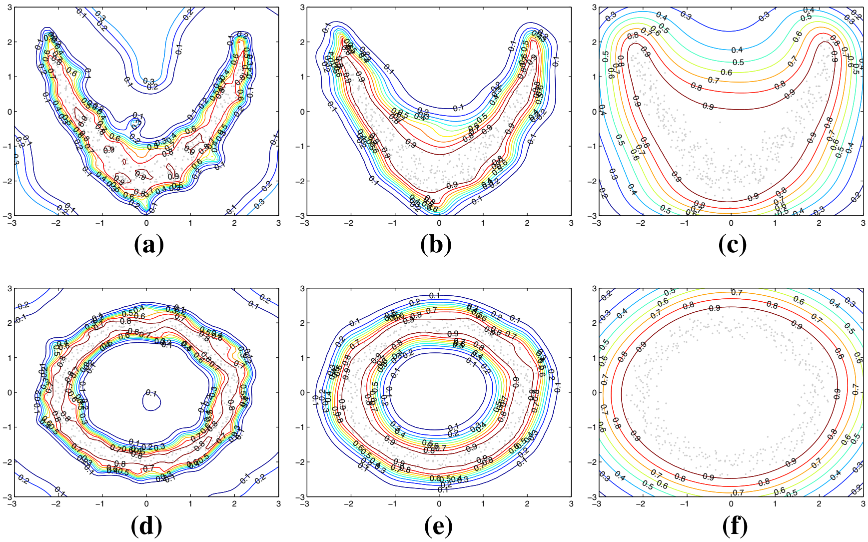

Fig. 2 The central regions of kmGMHD with respect to both clown data (*top row*) and ring data (*bottom row*) with different parameters of Gaussian kernel: **a** $\sigma = 0.5$, **b** $\sigma = 1$, **c** $\sigma = 4$, **d** $\sigma = 0.5$, **e** $\sigma = 1$, **f** $\sigma = 4$

in spite of that, most information of the mapped data are mainly limited in a subspace of $\mathscr{H}_k$, and the dimensionality of such subspace is called *effective dimensionality*. Parameter σ indeed controls the size of such effective dimensionality. For a large σ, the weight decays very fast on the higher-order features, and it corresponds to a smaller effective dimensionality, and vice versa. The central region of kmGMHD with respect to the *clown* data and the *ring* data with different σ for the Gaussian kernel are shown in Fig. 2.

As seen from Fig. 2a and d, for a smaller σ, the central regions shrink to each data point. While for a larger σ, kmGMHD almost degenerates to MHD, which are shown in Fig. 2c and f. In practice, the value of the kernel function's parameter is mainly depended on the specified data sets. Generally speaking, a simpler structure of data corresponds to a larger σ in the Gaussian kernel function or a smaller d in the polynomial kernel.

7 Conclusion

In this paper, we proposed the kernel mapped GMHD to extend the applications of MHD. KmGMHD in nature is GMHD of the mapped data in its feature space. Compared with MHD which only has a good depth interpretation for elliptical distribution, kmGMHD performs well in dealing with many types of data, especially when the data cloud has complex shape, such as the data set with non-convex shape. The contour of kmGMHD can shrink to the shape of data cloud or its distribution and give a more rational interpretation with depth.

Although, for the sample version of kmGMHD, the central regions of GMH_kD $(\mathbf{x}, F_n)$ may not be simply connected for large values of the centrality parameter (containing more than one peaks), we can improve it by adjusting the parameter in the kernel function. Further more, kmGMHD has a good distribution in the edge area of data cloud, which makes it a potential tool for robust outliers detection and data classification.

Acknowledgments The authors thank the anonymous referees for their helpful comments and suggestions, which led to a significant improvement in this paper.

References

Abe S (2005) Support vector machines for pattern classification, 1st edn. Springer-Verlag London Limited, Reading, subsection 2.3, pp 32–33

Canu S, Grandvalet Y, Guigue V, Rakotomamonjy A (2005) Svm and kernel methods matlab toolbox. Perception Systèmes et Information, INSA de Rouen, Rouen, France

Chen Z, Tyler DE (2004) On the finite sample breakdown points of redescending m-estimates of location. Stat Probab Lett 69:233–242

Chenouri S (2004) Multivariate robust nonparametric inference based on data depth. Univ Waterloo 90: 67–89

Cui X, Lin L, Yang GR (2008) An extended projection data depth and its applications to discrimination. Commun Stat Theory Methods 37(14):2276–2290

Donoho DL (1982) Breakdown properties of multivariate location estimators. PhD thesis, Deptment of Statistics, Harvard University

Gao Y (2003) Data depth based on spatial rank. Stat Probab Lett 65:217–225

Ghosh AK, Chaudhuri P (2005a) On data depth and distribution-free discriminant analysis using separating surfaces. Bernoulli 11:1–27

Ghosh AK, Chaudhuri P (2005b) On maximum depth and related classifiers. Board Found Scand Stat 32:327–350

Hofmann T, SchöLkopf B, Smola AJ (2008) Kernel methods in machine learning. Ann Stat 36(3): 1171–1220

Jörnsten R (2004) Clustering and classification based on the l_1 data depth. J Multivar Anal 90:67–89

Li G, Zhang J (1998) Sphering and its properties. Sankhy$\bar{a}$ Indian J Stat 60(Series A, Pt. 1):119–133

Liu R, Singh K (1993) A quality index based on data depthand multivariate rank tests. J Am Stat Assoc 88:252–260

Liu R, Singh K (1997) Notionso of limiting p values based on data depth and bootstrap. J Am Stat Assoc 92:266–277

Liu RY (1990) On a notion of data depth based on random simplices. Ann Stat 18:405–414

Liu RY (1992) Data depth and multivariate rank tests. In: Dodge Y (ed) L_1-Statistics and related methods. North-Holland, Amsterdam, pp 279–294

Mahalaobis PC (1936) On the generalized distance in statistics. Proc Natl Acad Sci India 12:49–55

Pennacchi P (2008) Robust estimate of excitations in mechanical systems using m-estimators. J Sound Vib 310:923–946

Stahel WA (1981) Robuste schatzungen: Infintesimale optimalitat and schatzungen von kovarianzmatrizen. PhD thesis, Zurich

Taylor JS, Cristianini N (2004) Kernel methods for pattern analysis, 1st edn. Cambridge University Press, Reading

Tian X, Vardi Y, Zhang C (2002) l_1-depth, depth realtive to a model, and robust regression. In: Dodge Y (ed) In Statistical data analysis based on the L_1-norm and Related Methods. Birkhauser, Basel, pp 285–299

Tukey JW (1975) Mathematics and picturing of data. In: James RD (ed) Proceedings of the international congress on mathematics, vol 2. Canadian Mathematics Congress, Vancouver, pp 523–531

Vardi Y, Zhang C (2000) The multivariate l_1-median and associated data depth. In: Proceedings of National Academy of Science, vol 97, pp 1423–1426

Zuo Y, Serfling R (2000) General notions of statistical depth function. Ann Stat 28(2):461–482