Supplementary Materials to the article "Nonparametric imputation by data depth"

by Pavlo Mozharovskyi, Julie Josse and François Husson

1 Additional figures

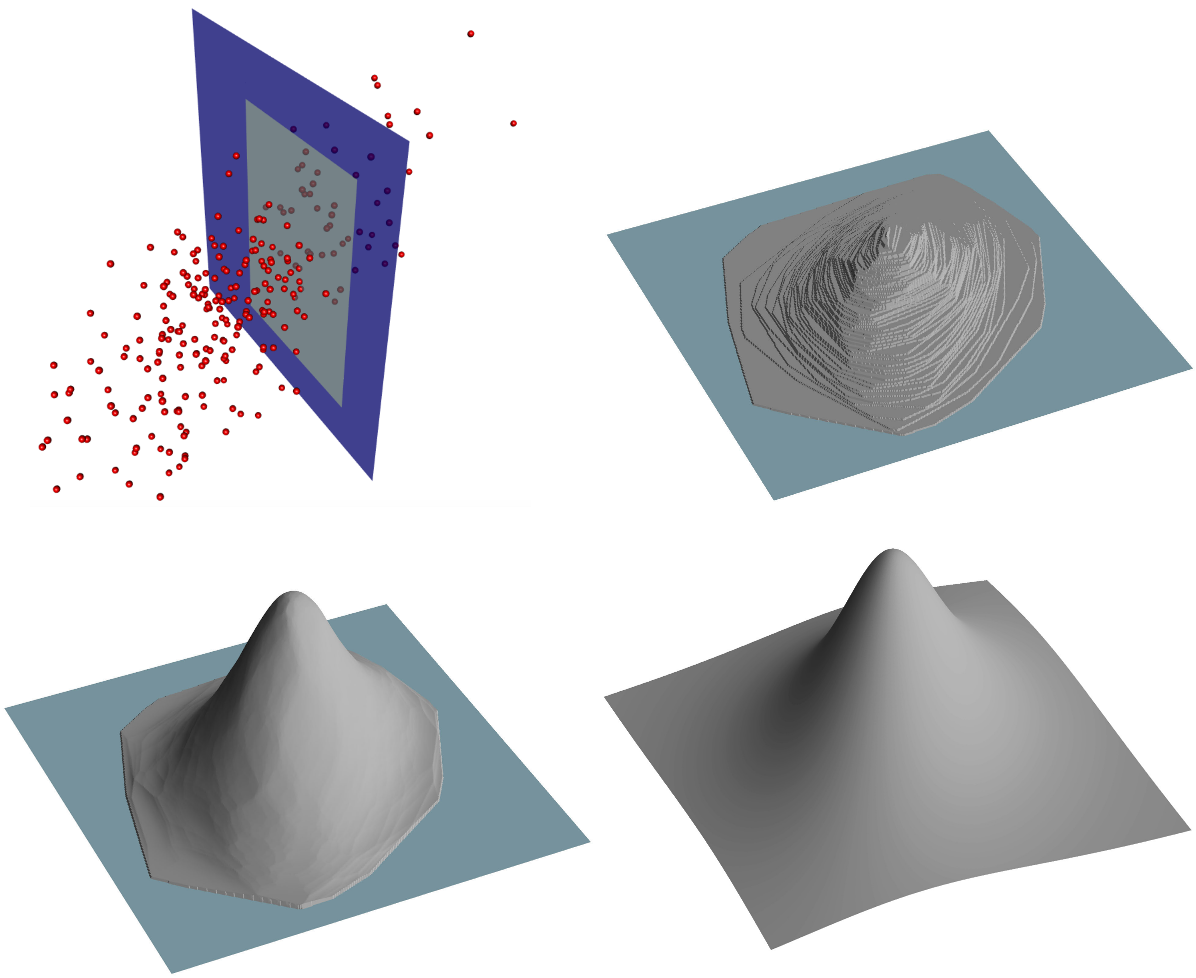

Figure 1: A Gaussian sample consisting of 250 points and a hyperplane of two missing coordinates (top, left), and the function $f(\boldsymbol{z}_{miss})$ to be optimized on each single iteration of Algorithm 1, for the smaller rectangle, for Tukey (top, right), zonoid (bottom, left), and Mahalanobis (bottom, right) depth. For the Tukey depth the maximum is not unique, and forms a polygon.

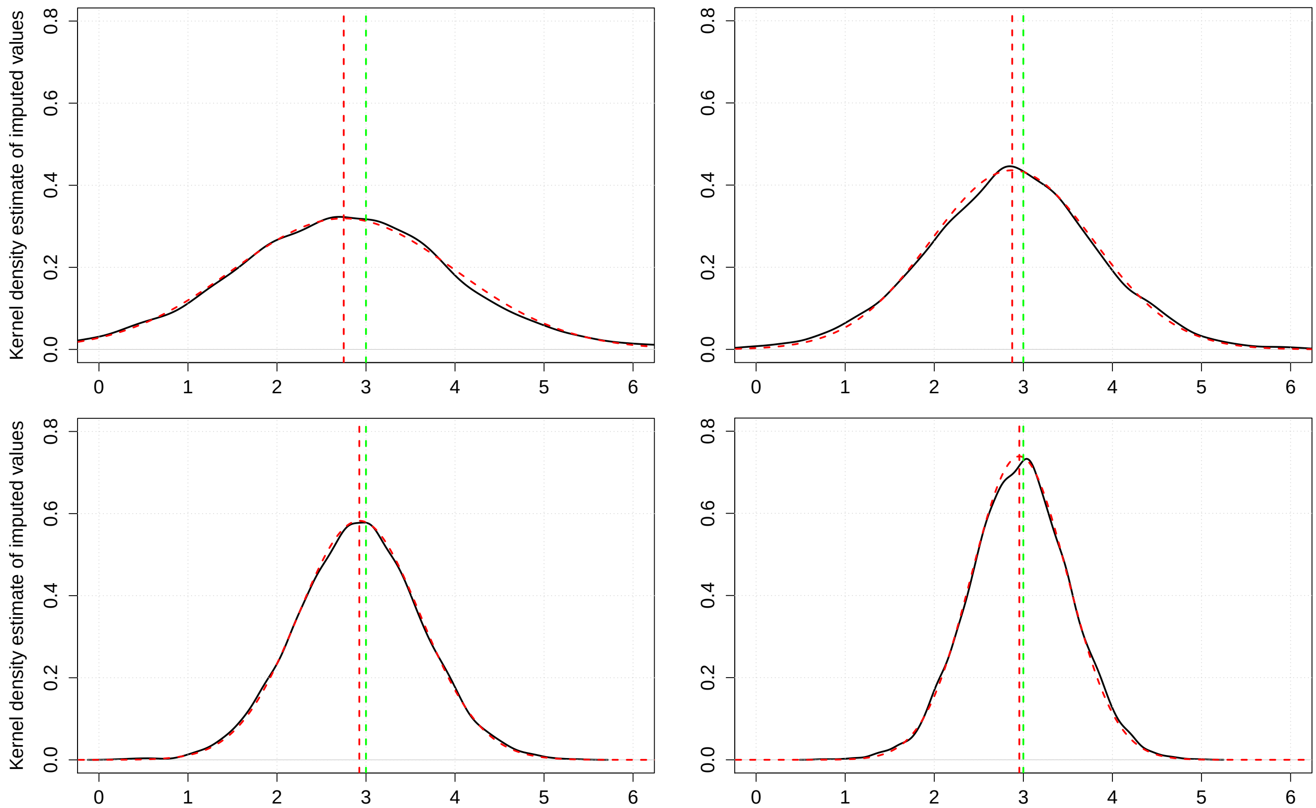

Figure 2: Samples of size 100 (top, left), 200 (top, right), 500 (bottom, left), and 1000 (bottom, right) are drawn from the bivariate Cauchy distribution with the location and scatter parameters $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ from the introduction. Single point with one missing coordinate is imputed with the Tukey depth. Its kernel density estimate (solid) and the best approximating Gaussian curve (dashed) over $10,000$ repetitions are plotted. The population's conditional center given the observed value equals 3.

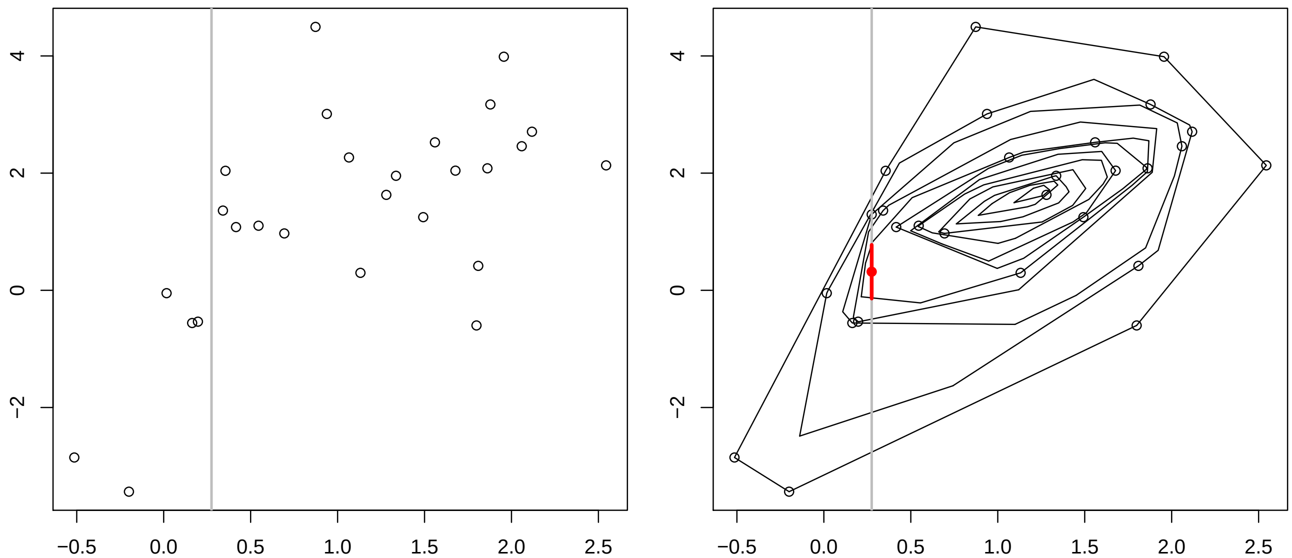

Figure 3: Illustration of imputation with the Tukey depth. When imputing the point with a missing second coordinate (left), the maximum of the constrained Tukey depth is non-unique (the red line segment), and an average over the optimal arguments (the red point) is used in equation (3) (right).

2 Simulation results with other percentages of missing values

When varying the percentage of missing values, the general trend remains unchanged. The small differences seen can be summarized as follows: with a decreasing percentage of missing values, the difference between EM and Mahalanobis depth imputation (and thus also the rank two PCA one) shrinks, and indeed the latter performs comparably to EM for 5% missingness. For the same percentage and the Cauchy distribution, nonparametric methods (kNN and random forest) perform comparably to the Tukey depth due to a sufficient quantity of available observations and an absence of correlation structure (outliers are generated from Cauchy distribution as well).

Distr.	D^{Tuk}	D^{zon}	D^{Mah}	$D^{Mah}_{MCD.75}$	EM	regPCA1	regPCA2	kNN	RF	mean	oracle
$t\,\infty$	1.577	1.547	*1.532*	1.537	**1.518**	1.596	*1.532*	1.684	1.681	2.058	1.487
	(0.2345)	(0.2128)	*(0.216)*	(0.2199)	**(0.2129)**	(0.2327)	*(0.2159)*	(0.2422)	(0.2445)	(0.2774)	(0.2004)
$t\,10$	1.748	1.718	*1.693*	1.709	**1.69**	1.769	*1.692*	1.853	1.871	2.275	1.642
	(0.287)	(0.2838)	*(0.2737)*	(0.2827)	**(0.2826)**	(0.3039)	*(0.2741)*	(0.3168)	(0.3085)	(0.3757)	(0.2771)
$t\,5$	1.993	1.971	*1.956*	*1.956*	**1.933**	2.017	*1.956*	2.125	2.134	2.565	1.874
	(0.378)	(0.3602)	*(0.3799)*	*(0.361)*	**(0.361)**	(0.3759)	*(0.3796)*	(0.4126)	(0.3976)	(0.4732)	(0.3492)
$t\,3$	2.417	2.434	2.39	**2.333**	*2.362*	2.431	2.39	2.55	2.592	3.045	2.235
	(0.5996)	(0.6032)	(0.5792)	**(0.5571)**	*(0.5808)*	(0.5734)	(0.5793)	(0.6154)	(0.612)	(0.6943)	(0.5319)
$t\,2$	*3.31*	3.373	3.431	**3.192**	3.366	3.437	3.422	3.538	3.555	4.155	2.986
	(1.191)	(1.273)	(1.314)	**(1.148)**	(1.249)	(1.343)	(1.289)	(1.33)	(1.321)	(1.466)	(1.063)
$t\,1$	**13.19**	15.13	15.17	*13.39*	14.86	14.82	15.22	14.09	13.91	16.77	11.17
	(10.83)	(12.06)	(11.74)	*(10.32)*	(11.57)	(11.64)	(11.94)	(11.28)	(11.06)	(13.22)	(8.901)

Table 1: Median and MAD of the RMSEs of the imputation for a sample of 100 points drawn from elliptically symmetric Student-t distributions with $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_2$ having 15% of MCAR values, over 1000 repetitions.

Distr.	D^{Tuk}	D^{zon}	D^{Mah}	$D^{Mah}_{MCD.75}$	EM	regPCA1	regPCA2	kNN	RF	mean	oracle
$t\,\infty$	**1.656**	1.754	1.853	*1.671*	1.817	1.855	1.853	1.793	1.762	2.182	1.514
	(0.2523)	(0.3142)	(0.4062)	*(0.2906)*	(0.3937)	(0.4158)	(0.4071)	(0.2974)	(0.282)	(0.3821)	(0.2121)
$t\,10$	**1.859**	1.973	2.048	*1.865*	2.027	2.05	2.044	1.995	1.968	2.45	1.677
	(0.3062)	(0.4031)	(0.511)	*(0.3917)*	(0.4933)	(0.5069)	(0.5119)	(0.3861)	(0.3402)	(0.4778)	(0.279)
$t\,5$	**2.09**	2.23	2.31	*2.109*	2.267	2.348	2.31	2.255	2.233	2.749	1.91
	(0.4275)	(0.5122)	(0.6217)	*(0.4841)*	(0.6006)	(0.6504)	(0.6219)	(0.4543)	(0.4476)	(0.6089)	(0.3742)
$t\,3$	**2.507**	2.697	2.772	*2.541*	2.737	2.791	2.779	2.707	2.699	3.32	2.239
	(0.6389)	(0.7977)	(0.8516)	*(0.7133)*	(0.8243)	(0.9306)	(0.8495)	(0.6946)	(0.7254)	(0.964)	(0.5497)
$t\,2$	**3.462**	3.68	3.733	*3.517*	3.669	3.807	3.736	3.709	3.762	4.476	3.061
	(1.35)	(1.577)	(1.6)	*(1.39)*	(1.589)	(1.648)	(1.601)	(1.413)	(1.444)	(1.794)	(1.136)
$t\,1$	**11.81**	14.12	14.22	*12.34*	13.78	13.73	14.31	12.58	13.64	15.73	10.37
	(9.738)	(12.09)	(12.05)	*(9.631)*	(11.48)	(11.01)	(12.12)	(10.36)	(11.44)	(12.65)	(8.249)

Table 2: Median and MAD of the RMSEs of the imputation for 100 points drawn from elliptically symmetric Student-t distributions with $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_2$ contaminated with 15% outliers, and 15% of MCAR values on non-contaminated data, over 1000 repetitions.

Distr.	D^{Tuk}	D^{zon}	D^{Mah}	$D^{Mah}_{MCD.75}$	EM	regPCA1	regPCA2	kNN	RF	mean	oracle
$t\,\infty$	1.464	1.454	**1.447**	1.453	*1.449*	1.529	**1.447**	1.571	1.581	2.009	1.404
	(0.3694)	(0.3713)	**(0.3595)**	(0.3663)	*(0.3593)*	(0.401)	**(0.3594)**	(0.3892)	(0.3946)	(0.486)	(0.3399)
$t\,10$	1.649	1.597	**1.57**	*1.572*	**1.57**	1.665	**1.57**	1.755	1.754	2.2	1.529
	(0.4316)	(0.4285)	**(0.4163)**	*(0.4203)*	**(0.4206)**	(0.4502)	**(0.4163)**	(0.4565)	(0.46)	(0.5737)	(0.4278)
$t\,5$	1.816	1.799	**1.757**	*1.758*	**1.757**	1.876	**1.757**	1.955	1.972	2.402	1.712
	(0.5134)	(0.5129)	**(0.49)**	*(0.4991)*	**(0.4899)**	(0.5499)	**(0.4901)**	(0.555)	(0.5345)	(0.7318)	(0.4869)
$t\,3$	2.213	2.184	2.147	**2.101**	*2.139*	2.242	2.147	2.37	2.343	2.844	2.054
	(0.7882)	(0.8159)	(0.8016)	**(0.7618)**	*(0.8)*	(0.7782)	(0.801)	(0.8563)	(0.8357)	(1.011)	(0.7649)
$t\,2$	2.837	2.919	2.813	**2.68**	*2.8*	2.911	2.813	3.03	2.99	3.578	2.529
	(1.249)	(1.342)	(1.309)	**(1.196)**	*(1.287)*	(1.311)	(1.31)	(1.325)	(1.331)	(1.554)	(1.133)
$t\,1$	**7.806**	8.718	8.911	8.286	8.9	9.118	8.935	*8.135*	8.138	10.99	6.367
	(6.351)	(7.135)	(7.127)	(6.602)	(7.124)	(7.334)	(7.137)	*(6.605)*	(6.563)	(8.952)	(5.12)

Table 3: Median and MAD of the RMSEs of the imputation for a sample of 100 points drawn from elliptically symmetric Student-t distributions, with $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_2$ having 5% of MCAR values, over 1000 repetitions.

3 Proofs

Proof of Theorem 1:

Due to the fact that $D_{n,\alpha}(\boldsymbol{X}) \xrightarrow[n\to\infty]{a.s.} D_\alpha(X)$, in what follows we focus on the population version

Distr.	D^{Tuk}	D^{zon}	D^{Mah}	$D^{Mah}_{MCD.75}$	EM	regPCA1	regPCA2	kNN	RF	mean	oracle
$t\,\infty$	**1.552**	1.613	1.709	*1.553*	1.701	1.769	1.709	1.695	1.603	2.167	1.406
	(0.3693)	(0.4107)	(0.4867)	*(0.4379)*	(0.4788)	(0.5248)	(0.4877)	(0.407)	(0.3924)	(0.5981)	(0.3171)
$t\,10$	**1.706**	1.778	1.874	*1.73*	1.861	1.906	1.875	1.884	1.823	2.398	1.564
	(0.4415)	(0.5106)	(0.6104)	*(0.4912)*	(0.6032)	(0.6053)	(0.6111)	(0.5182)	(0.4797)	(0.7059)	(0.3938)
$t\,5$	**1.868**	1.951	2.038	*1.877*	2.027	2.172	2.039	2.077	1.995	2.57	1.698
	(0.5565)	(0.5843)	(0.6859)	*(0.5679)*	(0.6806)	(0.7747)	(0.6819)	(0.6256)	(0.6102)	(0.8625)	(0.491)
$t\,3$	*2.243*	2.348	2.421	**2.226**	2.42	2.525	2.421	2.429	2.392	3.05	2.016
	(0.8064)	(0.8694)	(0.9166)	**(0.8258)**	(0.9345)	(1.019)	(0.9237)	(0.8484)	(0.8521)	(1.171)	(0.7047)
$t\,2$	**2.902**	3.032	3.183	*2.933*	3.163	3.196	3.188	3.142	3.071	4.073	2.55
	(1.375)	(1.498)	(1.566)	*(1.421)*	(1.558)	(1.565)	(1.582)	(1.472)	(1.43)	(2.007)	(1.129)
$t\,1$	**7.464**	8.487	8.531	8.334	8.5	8.675	8.541	*7.958*	8.1	10.82	6.245
	(5.916)	(6.869)	(7.081)	(6.867)	(6.988)	(7.261)	(7.117)	*(6.509)*	(6.922)	(8.802)	(4.874)

Table 4: Median and MAD of the RMSEs of the imputation for 100 points drawn from elliptically symmetric Student-t distributions with $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_2$ contaminated with 15% of outliers, with 5% MCAR values on non-contaminated data, over 1000 repetitions.

only. For $X \sim \mathcal{E}_d(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X, F_R)$ allow the transform $X \mapsto Z = \boldsymbol{R}\boldsymbol{\Sigma}^{-1/2}(X - \boldsymbol{\mu})$, with $\boldsymbol{R}$ being a rotation operator such that w.l.o.g. $\boldsymbol{x} \mapsto \boldsymbol{z}$, such that missing values still constitute a $|miss(\boldsymbol{x})|$-dimensional affine space parallel to missing coordinates' axes. Since contours $D_\alpha(Z)$ are concentric spheres centered at the origin, $D^*_\alpha(Z)$ in (3) is of the form $\{\boldsymbol{v}\,|\,\boldsymbol{v} = \boldsymbol{z}' + \beta\boldsymbol{r}\,,\, \beta \geq 0\}$ with $\boldsymbol{z}'_{obs(\boldsymbol{z})} = \boldsymbol{z}_{obs}$ and $\boldsymbol{z}'_{miss(\boldsymbol{z})} = \boldsymbol{0}_{|miss(\boldsymbol{x})|}$, and $\boldsymbol{r} \in \mathcal{S}^{|miss(\boldsymbol{x})|-1}$, a unit sphere in the linear span of $miss(\boldsymbol{z})$. Because of the fact that $P\bigl(\{\boldsymbol{x} \in \mathbb{R}^d\,|\,D(\boldsymbol{x}|X) = \alpha\}\bigr) = 0$, $\beta = 0$ almost surely and thus $\boldsymbol{z}$ is imputed with $\boldsymbol{z}' = \boldsymbol{R}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})$. $\square$

Proof of Theorem 2:
(The challenge here is that the resulting distribution is not elliptical.)

For $X \sim \mathcal{E}_d(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X, F_R)$ allow the transform $X \mapsto Z = \boldsymbol{R}\boldsymbol{\Sigma}^{-1/2}(X - \boldsymbol{\mu})$, with $\boldsymbol{R}$ being a rotation operator such that w.l.o.g. $\boldsymbol{x} \mapsto \boldsymbol{z}$, such that $miss(\boldsymbol{z}) = 1$. (Z has spherical density contours and missing values are in the first coordinate only.)

Let $Z' = \bigl(0, (Z'')^\top\bigr)^\top$ with $Z'' \sim \mathcal{E}_{d-1}(\boldsymbol{0}, \boldsymbol{I}, F_R)$, where $\boldsymbol{I}$ is the diagonal matrix. Consider a random vector $U \sim (1-p)Z + pZ'$ which is a mixture of d- and $(d-1)$-dimensional spherical distributions. Z' corresponds to the imputed missing values—let us now show that this is true. Due to the fact that $D_{n,\alpha}(\boldsymbol{U}) \xrightarrow[n\to\infty]{a.s.} D_\alpha(U)$, in what follows we focus on the population version only. Missing values constitute one-dimensional affine subspaces parallel to the first coordinate. Thus, due to the affine invariance property (P1 in Definition 2), $D_\alpha(U) \cap \{\boldsymbol{u} \in \mathbb{R}^d\,|\,\boldsymbol{u}_1 \geq 0\} = D_\alpha(U) \cap \{\boldsymbol{u} \in \mathbb{R}^d\,|\,\boldsymbol{u}_1 \leq 0\} \times (-1, 0, \ldots, 0)^\top$. To see this, it suffices to note that the symmetric reflection of U w.r.t. the linear space normal to $(1, 0, \ldots, 0)^\top$ equals U. Now, for $\lambda \in \mathbb{R}$ let $\boldsymbol{u} = (\lambda, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_d)^\top$ be this one-dimensional affine subspace of missingness for a point. In (3), $\mathrm{ave}\bigl(D_\alpha(U) \cap \boldsymbol{u}\bigr) = \mathrm{ave}\bigl(D_\alpha(U) \cap \{\boldsymbol{v} \in \mathbb{R}^d\,|\,\boldsymbol{v}_1 > 0\} \cap \boldsymbol{u} \cup D_\alpha(U) \cap \{\boldsymbol{v} \in \mathbb{R}^d\,|\,\boldsymbol{v}_1 < 0\} \cap \boldsymbol{u} \cup D_\alpha(U) \cap \{\boldsymbol{v} \in \mathbb{R}^d\,|\,\boldsymbol{v}_1 = 0\} \cap \boldsymbol{u}\bigr) = D_\alpha(U) \cap \{\boldsymbol{v} \in \mathbb{R}^d\,|\,\boldsymbol{v}_1 = 0\} \cap \boldsymbol{u} = (0, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_d) = \boldsymbol{R}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})$ (with the obvious correspondence between $\boldsymbol{u}$ and $\boldsymbol{y}$). $\square$

Proof of Corollary 1:
(P1)–(P5) are obviously satisfied for the Tukey, zonoid and Mahalanobis depths. In Theorem 1,

$D_{n,\alpha}(\boldsymbol{X}) \xrightarrow[n\to\infty]{a.s.} D_\alpha(X)$ is clearly satisfied for the Mahalanobis depth, following Corollary 3.11 by Mosler (2002) for the zonoid depth, and by Theorem 4.2 in Zuo & Serfling (2000) for the Tukey depth. In Theorem 2, the same logic holds for the Mahalanobis and zonoid depths, but not for the Tukey depth as Z is not elliptical. Using techniques similar those in the proof of Theorem 3.4 in Zuo & Serfling (2000), one can show that $P\big(\{\boldsymbol{x} \in \mathbb{R}^d \,|\, D(\boldsymbol{x}|Z) = \alpha\}\big) = 0$, from which, together with the vanishing at infinity property (P4) and $\sup_{\boldsymbol{x}\in\mathbb{R}^d} |D_n(\boldsymbol{x}|\boldsymbol{Z}) - D(\boldsymbol{x}|Z)| \xrightarrow[n\to\infty]{a.s.} 0$ (see Donoho & Gasko 1992), it follows that $D_{n,\alpha}(\boldsymbol{Z}) \xrightarrow[n\to\infty]{a.s.} D_\alpha(Z)$. □

Proof of Proposition 1: w.l.o.g. we restrict ourselves to the case $i = 1$. Let $\boldsymbol{Z}$ be $\boldsymbol{X}$ transformed in such a way that it is an $n\times d$ matrix with $\boldsymbol{\mu_Z} = \boldsymbol{0}$ and $\boldsymbol{z}_{1,miss(1)} = \boldsymbol{\Sigma}_{\boldsymbol{Z}\,miss(1),obs(1)}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{Z}\,obs(1),obs(1)}\boldsymbol{z}_{1,obs(1)}$. Denote the argument $\boldsymbol{a} = (0,\ldots,0,\boldsymbol{y}^\top)^\top \in \mathbb{R}^d$. Replacing $\boldsymbol{z}_1$ with $\boldsymbol{z}_1 + \boldsymbol{a}$ and subtracting the column-wise average $\frac{\boldsymbol{a}}{n}$ from each row gives the covariance matrix estimate:

$$
\begin{aligned}
n\boldsymbol{\Sigma_Z}(\boldsymbol{y}) &= \boldsymbol{Z}^\top\boldsymbol{Z} - \boldsymbol{z}_1\boldsymbol{z}_1^\top + (\boldsymbol{z}_1+\boldsymbol{a})(\boldsymbol{z}_1+\boldsymbol{a})^\top - \frac{1}{n}\boldsymbol{a}\boldsymbol{a}^\top \\
&= \boldsymbol{Z}^\top\boldsymbol{Z} + 2\boldsymbol{z}_1\boldsymbol{a}^\top + \frac{n-1}{n}\boldsymbol{a}\boldsymbol{a}^\top\,.
\end{aligned}
$$

Since $\boldsymbol{z}_1^\top(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{a} = 0$ due to Mahalanobis orthogonality, by simple algebra for the determinant, one obtains:

$$
\begin{aligned}
n^d|\boldsymbol{\Sigma_Z}(\boldsymbol{y})| &= \Big|\boldsymbol{Z}^\top\boldsymbol{Z} + \sqrt{2}\boldsymbol{z}_1\boldsymbol{a}^\top\sqrt{2} + \sqrt{\frac{n-1}{n}}\boldsymbol{a}\boldsymbol{a}^\top\sqrt{\frac{n-1}{n}}\Big| \\
&= \big|\boldsymbol{Z}^\top\boldsymbol{Z} + \sqrt{2}\boldsymbol{z}_1\boldsymbol{a}^\top\sqrt{2}\big|\Big(1 + \sqrt{\frac{n-1}{n}}\boldsymbol{a}^\top\big(\boldsymbol{Z}^\top\boldsymbol{Z} + \sqrt{2}\boldsymbol{z}_1\boldsymbol{a}^\top\sqrt{2}\big)^{-1}\boldsymbol{a}\sqrt{\frac{n-1}{n}}\Big) \\
&= |\boldsymbol{Z}^\top\boldsymbol{Z}|\big(1 + \sqrt{2}\boldsymbol{a}^\top(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{z}_1\sqrt{2}\big)\Big(1 + \sqrt{\frac{n-1}{n}}\boldsymbol{a}^\top\big(\boldsymbol{Z}^\top\boldsymbol{Z}\big)^{-1}\boldsymbol{a}\sqrt{\frac{n-1}{n}} - \\
&\quad - \frac{\sqrt{\frac{n-1}{n}}\boldsymbol{a}^\top(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{z}_1\sqrt{2}\cdot\sqrt{2}\boldsymbol{a}^\top(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{a}\sqrt{\frac{n-1}{n}}}{1 + \sqrt{2}\boldsymbol{a}^\top(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{z}_1\sqrt{2}}\Big) \\
&= |\boldsymbol{Z}^\top\boldsymbol{Z}|\big(1 + \frac{n-1}{n}\boldsymbol{a}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{a}\big)\,.
\end{aligned}
$$

Thus $|\boldsymbol{\Sigma_Z}(\boldsymbol{y})|$ is a quadratic function of $\boldsymbol{y}$, which is clearly minimized in $\boldsymbol{y} = (0,\ldots,0)^\top$. □

Proof of Theorem 3: The first point can be checked by elementary algebra. The second point follows from the coordinate-wise application of Proposition 1. For the third point, it suffices to prove the single-output regression case. The regularized PCA algorithm will converge if

$$
\boldsymbol{y}_{id} = \sum_{s=1}^{d} u_{is}\sqrt{\lambda_s}v_{ds} = \sum_{s=1}^{d} u_{is}\big(\sqrt{\lambda_s} - \frac{\sigma^2}{\sqrt{\lambda_s}}\big)v_{ds}
$$

for any $\sigma^2 \leq \lambda_d$. W.l.o.g. we prove that

$$\boldsymbol{y}_d = \boldsymbol{\Sigma}_{d\,(1,\ldots,d-1)} \boldsymbol{\Sigma}^{-1}_{(1,\ldots,d-1)\,(1,\ldots,d-1)} \boldsymbol{y}_{(1,\ldots,d-1)} \iff \sum_{i=1}^{d} \frac{u_i v_{di}}{\sqrt{\lambda_i}} = 0,$$

denoting $\boldsymbol{\Sigma}(\boldsymbol{Y})$ simply $\boldsymbol{\Sigma}$ for the centered $\boldsymbol{Y}$, and an arbitrary point $\boldsymbol{y}$. Using matrix algebra,

$$\boldsymbol{y}_d = \boldsymbol{\Sigma}_{d\,(1,\ldots,d-1)} \boldsymbol{\Sigma}^{-1}_{(1,\ldots,d-1)\,(1,\ldots,d-1)} \boldsymbol{y}_{(1,\ldots,d-1)} = -\big((\boldsymbol{\Sigma}^{-1})_{dd}\big)^{-1} (\boldsymbol{\Sigma}^{-1})_{d\,(1,\ldots,d-1)} \boldsymbol{y}_{(1,\ldots,d-1)},$$

$$\begin{aligned}\sum_{i=1}^{d} u_i \sqrt{\lambda_i} v_{di} = -\Big(\sum_{i=1}^{d} \frac{v_{di}^2}{\lambda_i}\Big)^{-1} \Big(\sum_{i=1}^{d} \frac{v_{di} v_{1i}}{\lambda_i}, \sum_{i=1}^{d} \frac{v_{di} v_{2i}}{\lambda_i}, \ldots, \sum_{i=1}^{d} \frac{v_{di} v_{(d-1)\,i}}{\lambda_i}\Big) \times \\ \times \Big(\sum_{i=1}^{d} u_i \sqrt{\lambda_i} v_{1i}, \sum_{i=1}^{d} u_i \sqrt{\lambda_i} v_{2i}, \ldots, \sum_{i=1}^{d} u_i \sqrt{\lambda_i} v_{(d-1)\,i}\Big)^{\top}.\end{aligned}$$

After reordering the terms, one obtains

$$\sum_{i=1}^{d} u_i \sqrt{\lambda_i} \sum_{j=1}^{d} \frac{v_{dj}}{\lambda_j} \sum_{k=1}^{d} v_{ki} v_{kj} = 0.$$

Due to the orthogonality of $\boldsymbol{V}$, $d^2 - d$ terms from the two outer sum signs are zero. Gathering non-zero terms, i.e., those with $i = j$ only, we have that

$$\sum_{i=1}^{d} u_i \sqrt{\lambda_i} \frac{v_{di}}{\lambda_i} = \sum_{i=1}^{d} \frac{u_i v_{di}}{\sqrt{\lambda_i}} = 0.$$

$\square$

Derivation of (4): The integrated quantity is the conditional depth density that can be obtained from the joint one by the volume transformation (denoting $d_M\big(z, \boldsymbol{\mu}\big)$ the Mahalanobis distance between a point of depth z and $\boldsymbol{\mu}$):

$$\begin{aligned} f_{D((X|X_{obs}=\boldsymbol{x}_{obs})|X)}(z) = f_{D(X|X)}(z) \cdot C \cdot T_{down}\big(d_M(z, \boldsymbol{\mu})\big) \cdot T_{up}\big(d_M(z, \boldsymbol{\mu}^*)\big) \times \\ \times T_{angle}\big(d_M(z, \boldsymbol{\mu}), d_M(z, \boldsymbol{\mu}^*)\big)\,. \end{aligned}$$

Any constant C is ignored as it is unimportant when drawing. The three terms below correspond to descaling the density to dimension one (downscaling), re-scaling it to the dimension of the missing values (upscaling), and the linear transformation from dimension d to dimension $|miss|$ =number

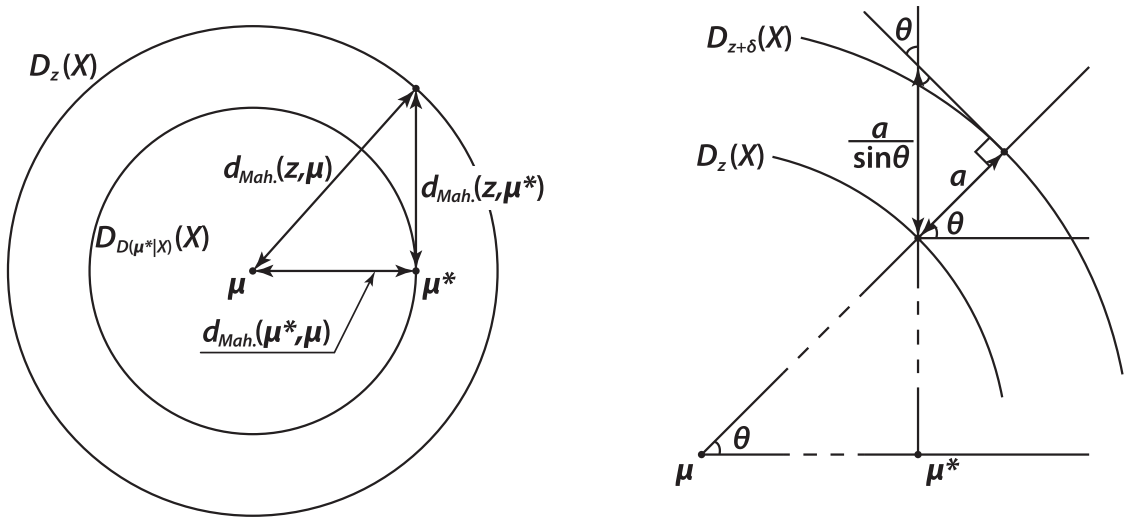

Figure 4: Illustration of the derivation of (4).

of missing coordinates of a point (angle transformation):

$$
\begin{aligned}
T_{down}\big(d_M(z,\boldsymbol{\mu})\big) &= d_M^{1-d}(z,\boldsymbol{\mu}) = \frac{1}{d_M^{d-1}(z,\boldsymbol{\mu})}\,. \\
T_{up}\big(d_M(z,\boldsymbol{\mu}^*)\big) &= d_M^{|miss(\boldsymbol{x})|-1}(z,\boldsymbol{\mu}^*) \\
&= \Big(\sqrt{d_M^2(z,\boldsymbol{\mu}) - d_M^2\big(D(\boldsymbol{\mu}^*|X),\boldsymbol{\mu}\big)}\Big)^{|miss(\boldsymbol{x})|-1}\,. \\
T_{angle}\big(d_M(z,\boldsymbol{\mu}), d_M(z,\boldsymbol{\mu}^*)\big) &= \frac{1}{\sin\theta} = \frac{1}{\frac{d_M(z,\boldsymbol{\mu}^*)}{d_M(z,\boldsymbol{\mu})}} \\
&= \frac{d_M(z,\boldsymbol{\mu})}{\sqrt{d_M^2(z,\boldsymbol{\mu}) - d_M^2\big(D(\boldsymbol{\mu}^*|X),\boldsymbol{\mu}\big)}}\,.
\end{aligned}
$$

T_{down} and T_{up} are illustrated in Figure 4 (left); for T_{angle} see Figure 4 (right). Setting $d_M(z,\boldsymbol{\mu}) = d_M(z)$ to shorten notation gives (4).

References

Donoho, D. L. & Gasko, M. (1992), 'Breakdown properties of location estimates based on halfspace depth and projected outlyingness', *The Annals of Statistics* **20**(4), 1803–1827.

Mosler, K. (2002), *Multivariate Dispersion, Central Regions, and Depth: The Lift Zonoid Approach*, Lecture Notes in Statistics, Springer New York.

Zuo, Y. & Serfling, R. (2000), 'Structural properties and convergence results for contours of sample statistical depth functions', *The Annals of Statistics* **28**(2), 483–499.