Local intrinsic dimensionality estimators based on concentration of measure

Jonathan Bac

Institut Curie, PSL Research University,
Mines ParisTech, INSERM, U900
Centre de Recherches Interdisciplinaires,
Paris, France
jonathan.bac@cri-paris.org

Andrei Zinovyev

Institut Curie, PSL Research University,
Mines ParisTech, INSERM, U900
Paris, France
Lobachevsky University
Nizhni Novgorod, Russia
andrei.zinovyev@curie.fr

Abstract—**Intrinsic dimensionality (ID) is one of the most fundamental characteristics of multi-dimensional data point clouds. Knowing ID is crucial to choose the appropriate machine learning approach as well as to understand its behavior and validate it. ID can be computed globally for the whole data point distribution, or computed locally in different regions of the data space. In this paper, we introduce new local estimators of ID based on linear separability of multi-dimensional data point clouds, which is one of the manifestations of concentration of measure. We empirically study the properties of these estimators and compare them with other recently introduced ID estimators exploiting various effects of measure concentration. Observed differences between estimators can be used to anticipate their behaviour in practical applications.**

Index Terms—**high-dimensional data, intrinsic dimension, effective dimension, concentration of measure, linear separability**

I. INTRODUCTION

Datasets used in applications of machine learning frequently contain objects characterized by thousands or millions of features. In this respect, the well-known *curse of dimensionality* is frequently discussed which states that many problems become exponentially difficult in high dimensions [1]. However, in what concerns the application of machine learning methods, the curse of dimensionality is not automatically manifested when the number of features is large: this depends rather on the dataset's intrinsic dimensionality (ID). If the features of a dataset are correlated in linear or non-linear fashion then the data point cloud can be located close to a subspace of relatively low ID. This makes appropriate the application of dimension reduction algorithms to obtain a lower-dimensional representation of data. By contrast, if the value of ID is high then the data point cloud becomes sparse and its geometrical and topological properties can be highly non-intuitive, in

particular, due to various manifestations of concentration of measure [2].

Therefore, the estimation of ID is crucial to the choice of machine learning methodology and its applications, including validation, explainability and deployment. Indeed, ID determines to a large extent the strategy and feasibility of validating an algorithm, estimating the uncertainty of its predictions and explaining its decisions. Such explanations can even be legally required for its use in sensitive applications [3].

There has been recent progress on these topics, e.g. with several approaches to explain and validate classifiers by looking at various data properties related to ID [3]–[5]. ID can pose fundamental limits on the robustness of classifiers; this has been illustrated recently with adversarial examples, where a minimal perturbation to the input can lead to a misclassification. Recent theoretical results have shown that such adversarial examples are inevitable for a concentrated metric probability space [6]–[8] and algorithms have been proposed that estimate adversarial risk by quantifying concentration [9].

From a more general perspective, theory and algorithms for correcting artificial intelligence-based systems have been developed that exploit properties specific to spaces possessing large ID [10], [11]. In this respect, a complementarity principle has been formulated [12]–[14]: the data space can be split into a low volume (low dimensional) subset, where nonlinear methods are effective, and a high-dimensional subset, characterized by measure concentration and simplicity, allowing the effective application of linear methods.

One important observation is that in real life the ID of the complete dataset might not be equal to the ID of its parts. Therefore, ID can be considered a local characteristic of the data space, defined in each data neighborhood. In this case we refer to it as *local intrinsic dimensionality*. Even when there are no ID variations in a dataset, the relation between global and local intrinsic dimensionality can be non-trivial - for example, one can easily construct examples of datasets which are low-dimensional locally but globally possess large linear dimension [14].

Methods for estimating global ID can be applied locally

in a data neighbourhood, and, vice versa, local ID estimators can be used to derive a global ID estimate. However, local ID estimation presents specific challenges, such as dealing with restricted cardinality, which render many global methods ineffective in practice. These challenges explain the rise in recent works dedicated specifically to developing local ID estimators [15]–[21].

Well-known methods for estimating global or local ID are based on Principal Component Analysis (PCA) and quantifications of the covariance matrix's eigenspectrum using various heuristics (thresholding total explained variance, limiting the conditional number, using reference spectra such as broken stick distribution, etc.) [22]–[25]. Other famous examples include maximum likelihood estimation (MLE) and the correlation dimension, based on counting the number of objects in a growing neighbourhood [26], [27]. Since PCA is a linear method, it generally tends to overestimate ID, while MLE and correlation dimension both tend to underestimate larger ID values.

In order to overcome these limitations, several new ID estimators have been recently introduced that exploit concentration of measure [12], [17]–[20], [28], [29]. Increasing ID results in various manifestations of measure concentration that scale differently with respect to dimensionality. For example, linear separability of points increases rapidly, such that 30 dimensions can be already considered large, while the appearance of exponentially large quasi-orthogonal bases or hubness are usually manifested in higher dimensions [1], [30]. These differences lead to different properties of ID estimators, e.g. sensitivity to different dimensions and dependence of the estimated ID on sample size.

As noted in [18], most of the local ID estimators use pairwise distances (relationship between two points) or angular information (relationship between three points) [19], [27], [28], whereas others take into account the volume of a d-simplex (i.e., the relationship between $d + 1$ points) [17], [18] or in our case, the separability probability of each point from the others [20]. The common approach is to assume local distributions of data points are close to a uniformly sampled unit n-ball B^n or their scaled vectors to a uniformly sampled unit sphere S^{n-1}. Then, various sample statistics are used whose dependence on n is theoretically established for uniform distributions on B^n or S^{n-1}. If the estimated statistics of a data sample are similar to the theoretical of B^n or S^{n-1} then the dimensionality of the sample is estimated to be n.

For example, DANCo [28] estimates the probability density function of normalized nearest neighbor distance from the center (exploiting concentration of norms), and the parameters of a Von-Mises distribution (exploiting concentration of angles). Expected Simplex Skewness (ESS) [17], in its default version, computes simplex skewness, defined as the ratio between the volume of a simplex with one vertex in the centroid and the others in data points, and the volume this simplex would have if edges incident to the centroid were orthogonal. For a d-simplex with $d = 1$, skewness is $sin(\theta)$, with θ the angle between the two edges incident to the centroid vertex. The mean sample skewness is compared to the ESS for uniformly distributed data on B^n. ANOVA [19] uses a U-statistic for the variance of the angle between pairs of vectors among uniformly chosen points in S^{n-1}.

In our previous work we introduced and benchmarked global estimators of ID based on Fisher separability, using theoretical results by Gorban, Tyukin et al. [10], [20]. The details of this approach are provided below. It appeared that for noisy samples from synthetic manifolds, the method was competitive with other ID estimators. In particular, its behavior was close to two recently introduced ID estimators based on concentration of measure, namely ESS and DANCo.

In this work we extend our previous study and introduce two ways to estimate ID locally, in each data point, based on Fisher separability properties. We compare the properties of these local dimensionality estimators with other estimators based on quantifying various manifestations of concentration of measure.

II. LOCAL ID ESTIMATION BASED ON FISHER SEPARABILITY

In the present work, we will follow the notations introduced in the works by A.Gorban, I.Tyukin and their colleagues [10]: we call a data vector $\mathbf{x} \in R^n$ *linearly separable* from a finite set of points $Y \subset R^n$ if there exists a linear functional l such that $l(\mathbf{x}) > l(\mathbf{y})$ for all $\mathbf{y} \in Y$. If for any point $\mathbf{x}$ there exists a linear functional separating it from all other data points, then such a data point cloud is called *linearly separable* or 1-convex. The separating functional l may be computed using the linear Support Vector Machine (SVM) algorithms, the Rosenblatt perceptron algorithm, or other comparable methods. However, these computations may be rather costly for large-scale estimates. Hence, in the works of Gorban, Tyukin and their colleagues it was suggested to use a non-iterative estimate of the linear functional using Fisher's discriminant which is computationally inexpensive, after a standard pre-processing: [10].

1) centering
2) performing linear dimensionality reduction by projecting the dataset into the space of k first principal components, where k may be relatively large. In practice, we select the largest k (in their natural ranking) such that the corresponding eigenvalue λ_k is not smaller that λ_1/C, where C is a predefined threshold. Under most circumstances, $C = 10$
3) whitening (i.e., applying a linear transformation after which the covariance matrix becomes the identity matrix)

After such normalization of X, it is said that a point $\mathbf{x} \in X$ is Fisher-linearly separable from the cloud of points Y with parameter α, if

$$(\mathbf{x}, \mathbf{y}) \leq \alpha(\mathbf{x}, \mathbf{x})$$

for all $\mathbf{y} \in Y$, where $\alpha \in [0, 1)$. If equation (II) is valid for each point $\mathbf{x} \in X$ such that Y is the set of points $\mathbf{y} \neq \mathbf{x}$

then we call the dataset X Fisher-separable with parameter α. In order to quantify deviation from perfect separability, let us introduce $p_\alpha(\mathbf{y})$, the probability that the point $\mathbf{y}$ is inseparable from all other points.

In order to associate a value of ID to a point $\mathbf{y}$ in the data space, we compare the empirical $p_\alpha(\mathbf{y})$ estimates to the p_α of some reference data distribution whose dimension is known and separability properties can be analytically derived. The simplest such distribution is the uniform distribution of vectors on the surface of a unit n-dimensional sphere. Since this distribution is uniform, p_α does not depend on a data point and equals in any point (see derivation in [1], [10]):

$$p_\alpha \lessgtr \frac{(1-\alpha^2)^{\frac{n-1}{2}}}{\alpha\sqrt{2\pi n}} \qquad (1)$$

By resolving this formula with respect to n, we derive the following value of dimensionality as a function of inseparability probability, for a uniform distribution on the surface of the unit n-dimensional sphere:

$$n_\alpha = \frac{W\left(\frac{-\ln(1-\alpha^2)}{2\pi p_\alpha^2 \alpha^2 (1-\alpha^2)}\right)}{-\ln(1-\alpha^2)} \qquad (2)$$

where $W(x)$ is the real-valued branch of the Lambert W function. As a reminder, the Lambert W function solves equation $v = we^w$ with respect to w, i.e. $w = W(v)$. The pseudo-code of the algorithm for computing n_α was provided by us earlier [20].

Here we should make several important notes. Firstly, in order to apply the ID estimate (2) to a dataset X, one should apply an additional scaling step besides the preprocessing steps (1)-(3) described above. The scaling consists in normalizing each vector to the unit length, which corresponds to the projection onto a unit sphere. It means that in practice we do not distinguish an n-ball B^n from n-sphere S^{n-1}, both give ID=n in our case. This can lead to shifting by value 1 the estimations of ID in small dimensions, especially in artificial benchmark examples.

Secondly, as mentioned in [1], the formula (1) is an estimation from above, meaning that the actual empirical value of p_α is strictly less than the right hand side of (1). In particular, for some points, data point density, and values of α one can have empirical estimate $p_\alpha = 0$ which makes (2) inapplicable. The value of α should be adjusted in order to avoid the mean of p_α being too close to zero and in order to avoid too strong finite sampling effects (see [20] and Figure 1). This consideration also provides some theoretical limits on the maximally detectable dimensionalities, as shown hereafter in Figure 2.

Based on the above definitions, the separability properties of the data point cloud can be globally characterized by the histogram of empirical p_α distribution (probabilities of individual point inseparability) and the profile of intrinsic dimensions n_α (2) for a range of α values (e.g., $\alpha \in [0.6, ..., 1.0]$).

Let us denote $\bar{p}_\alpha(\mathbf{X})$ the mean value of the distribution of $p_\alpha(\mathbf{x})$ over all data points. We can introduce the global estimate of ID as in our previous work ([20]):

$$n_\alpha^{global} = \frac{W\left(\frac{-\ln(1-\alpha^2)}{2\pi \bar{p}_\alpha^2 \alpha^2 (1-\alpha^2)}\right)}{-\ln(1-\alpha^2)}. \qquad (3)$$

Now let us specify two local ID estimates based on Fisher separability. The first one will simply use the formula (2) in order to estimate n_α in a data point $\mathbf{y}$. We will call this estimate *global pointwise intrinsic dimension*, since in its definition the global separability properties of X are exploited, but the ID is computed in a data point:

$$n_\alpha(y) = \frac{W\left(\frac{-\ln(1-\alpha^2)}{2\pi p_\alpha(y)^2 \alpha^2 (1-\alpha^2)}\right)}{-\ln(1-\alpha^2)}. \qquad (4)$$

At the same time we can follow the standard approach of defining data neighborhoods and compute the n_α^{global} for these fragments of data. One of the simplest way to define a local neighbourhood is to determine the first k nearest neighbours (kNN) for a data point. This naive approach has well-known drawbacks, such as the edge effect : the ratio between points close to the border of the manifold and points inside it increases with dimensionality, meaning kNN data neighbourhoods can deviate from uniform distributions. [28], [31]. Forming kNN using distances of the original space can also create neighborhoods that do not reflect geodesic distance on the manifold. Recent work has suggested new ways to tackle these issues [15], [16]. Nonetheless this basic approach provides an easy way to start applying methods locally. Therefore, we define *local kNN ID* as:

$$n_\alpha^{kNN}(y) = \frac{W\left(\frac{-\ln(1-\alpha^2)}{2\pi p_\alpha^{kNN}(y)^2 \alpha^2 (1-\alpha^2)}\right)}{-\ln(1-\alpha^2)}, \qquad (5)$$

where $p_\alpha^{kNN}(y)$ is computed for a dataset comprising k nearest neighbours of y. In practical applications, k is chosen of few hundreds by the order of magnitude.

III. NUMERICAL RESULTS

A. Estimation of ID for n-dimensional balls

We illustrate our approach for a simple example of n-ball B^n (see Figure 1), sampled by 2000 points. The empirically estimated values of p_α are shown together with the theoretical dependence (1) in Figure 1A for $n = 2...10$. Figure 1B shows the empirically estimated distributions of p_α for $n = 2...10$. In Figure 1C we show the empirically estimated value of n as a function of α. One can see that smaller values of α create biased estimates of n. At the same time, maximum possible values of α (for which the mean of p_α is still non-zero) create unstable estimates suffering from finite sampling effects, starting from $n = 7$. Therefore, for the unique definition of n, our rule of thumb is to select p_α which would equal 0.8 multiplied by the maximally measured p_α (shown by crosses in Figure 1C).

Finite sample effects pose theoretical limits on maximum measurable values of ID from separability analysis as a function of the number of points N. The minimum measurable p_α equals $\frac{1}{N}$ (here we neglect small difference between N and $N-1$): therefore, maximally measurable $n_\alpha(y)$ (4) equals $\frac{W(\frac{-\ln(1-\alpha^2)N^2}{2\pi\alpha^2(1-\alpha^2)})}{-\ln(1-\alpha^2)}$ (shown in Figure 1D). At the same time, the mean of p_α is non-zero except if the dataset is completely Fisher-separable. This defines the limit for measuring (3) in $\frac{W(\frac{-\ln(1-\alpha^2)N^4}{2\pi\alpha^2(1-\alpha^2)})}{-\ln(1-\alpha^2)}$ (shown in Figure 1E) since the minimally measurable $\bar{p}_\alpha$ can be estimated as $\frac{1}{N^2}$. This conclusion allows us to estimate the maximally measurable $n_\alpha^{kNN}(y)$ (5) in $\frac{W(\frac{-\ln(1-\alpha^2)k^4}{2\pi\alpha^2(1-\alpha^2)})}{-\ln(1-\alpha^2)}$ which can be read from Figure 1E for cardinality equals k. Notice from Figure 1D,E that the maximum measurable ID quickly saturates with the number of points N but remains relatively high for smaller α.

For the n-ball example, we also studied how well different estimators can characterize the ID of high dimensional samples, depending on their cardinality (see Figure 1F). In the case of this simple benchmark, ESS and DANCo show impressive results, giving the exact value of ID even for very small samples (ESS is even able to estimate dimensionality higher than the number of points, which seems to be counter-intuitive at first glance but this is discussed in the paper [17]). Fisher separability analysis tends to overestimate the ID in case of small sample sizes but works as well as DANCo and ESS for larger number of points. The overestimation is connected to the point previously discussed of maximally measurable dimensionality for a given α value. In case of a small sample in very high dimension, the value of α, adjusted using the heuristics described above, becomes small in order to avoid full separability: in this case, the estimate (3) becomes less accurate.

We also note that many widely used ID estimators such as the correlation dimension or MLE heavily ID, showing saturation at few tens of dimensions. The ANOVA estimator showed surprising underestimation of larger ID even for large number of points (Figure 1F).

B. Synthetic and real-life examples

In Figure 1G we show visualizations of ID values for a simple synthetic '10 balls' dataset, which represents 10 unit n-balls of increasing $n = 2..11$ embedded without intersection in 11-dimensional space such that the n-ball shares $n-1$ dimensions with the $(n-1)$-ball. The balls are separated by increments of 1 along the first axis. From this example, one can see that both pointwise global and local kNN ID based on Fisher separability characterize well the dimensionality of the balls, however, pointwise global estimates are more heterogeneous within a ball while the local kNN estimate suffers when the data neighborhood includes points from different balls.

In Figure 2, we visualize local ID values from several ID estimators on real-life datasets. A first observation is that local ID can be highly variable from one data cluster to another. For example, in the MNIST datasets, local ID highlights some clusters which possess much lower local ID than the others (digit '1' - for the MNIST digit dataset, letters 'i' and 'l' for the MNIST letters dataset, and cluster 'Trousers' for the fashion MNIST dataset). This might reflect the intrinsic number of degrees of freedom in the distribution of different realizations (e.g., writing or design) of the same object type (e.g., a digit). At the same time, in the case of the ISOMAP Faces dataset, the distribution of local ID is more uniform and close to 3 as expected, although there is variation. This dataset is more challenging for local ID estimation since it only contains 700 data points. ESS and TLE in particular seem to overestimate ID.

The synthetic balls dataset (Figure 1G) and real data examples (Figure 2) demonstrate possible relations between global pointwise and local (kNN) ID estimates based on Fisher separability. Global pointwise ID estimates are computed based on the separability of a point from the rest of the data point cloud, while local kNN ID estimates are based on the mean inseparability of points in each local neighbourhood. These two estimates do not have to be similar: strong deviations of one from another can indicate complex data topology, when distant parts of the data point cloud are co-localized in close data subspaces. Local inseparability usually imposes global inseparability but not vice versa. Global pointwise ID can thus provide insights about the structure of the point cloud but will usually give less accurate ID estimates.

C. Dependence of ID estimators on the subsample size

A desirable property of an ID estimator is the ability to quantify a wide range of ID values from a relatively small sample size. In order to study these properties, we took several synthetic and real-life datasets and computed global ID and local ID with one hundred nearest neighbors for five estimators, on different subsample sizes. Running these estimations 10 times also gave us an idea about the uncertainty in the estimates for a given subsample size.

The results of this analysis are shown in Figure 3. Notice that for the uniformly sampled S^{10} sphere FisherS determines 11 intrinsic dimensions (as we noticed earlier, FisherS does not distinguish a ball and a sphere). For other uniformly sampled distributions (hypercube, multivariate Gaussian), 3 out of 5 estimators provide consistent results but TwoNN and TLE significantly underestimate the dimension. For the Swiss Roll, global estimators (DANCo and TwoNN) work well while ESS and FisherS show strong dependence of the mean local ID on the size of the subsample. This is expected since forming k-nearest neighborhoods with $k = 100$ will not respect geodesic distances for lower sample sizes. On the other hand the mean pointwise global estimate is usually biased towards smaller ID.

For all estimators the dependence of the local ID on the sample size can be quite strong. Overall, it seems that DANCo gives ID estimates closer to TwoNN while ESS matches better FisherS.
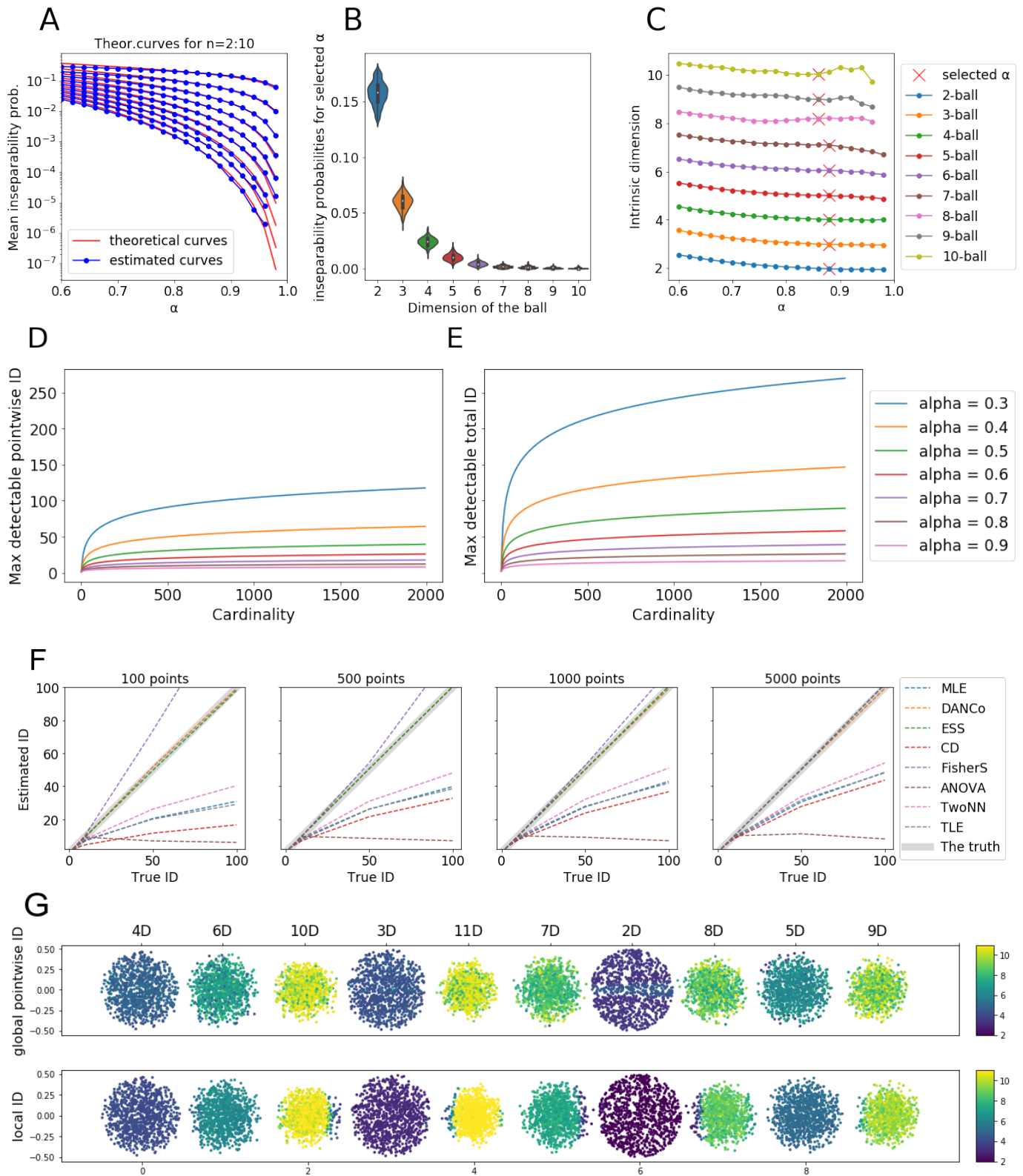
Fig. 1. **A-** theoretical vs estimated mean inseparability probability for balls of increasing dimension; **B-** complete inseparability probability histograms for selected $\alpha = .88$; **C-** intrinsic dimension estimated from the mean inseparability probability. Due to measure concentration, inseparability probability sharply decreases as a function of dimension. **D-** maximum detectable pointwise intrinsic dimension; **E-** maximum detectable "total" ID (estimate from the mean inseparability of all data points); **F-** ability of various estimators to quantify high intrinsic dimensionality from different sample size, tested on the simplest n-ball example. Here CD stands for the Correlation Dimension, FisherS is the global ID estimation based on Fisher separability, MLE stands for Maximum Likelihood Estimation of Intrinsic Dimension [27] and TLE stands for Tight Locality Estimator [15]. Other estimator names are referenced in the Introduction; **G-** local and global pointwise ID estimates computed on a synthetic dataset containing 10 unit balls of different dimensions (2D to 11D, each sampled with 500 points) arranged in a line. Maximum ID is capped to the embedding dimension (11)
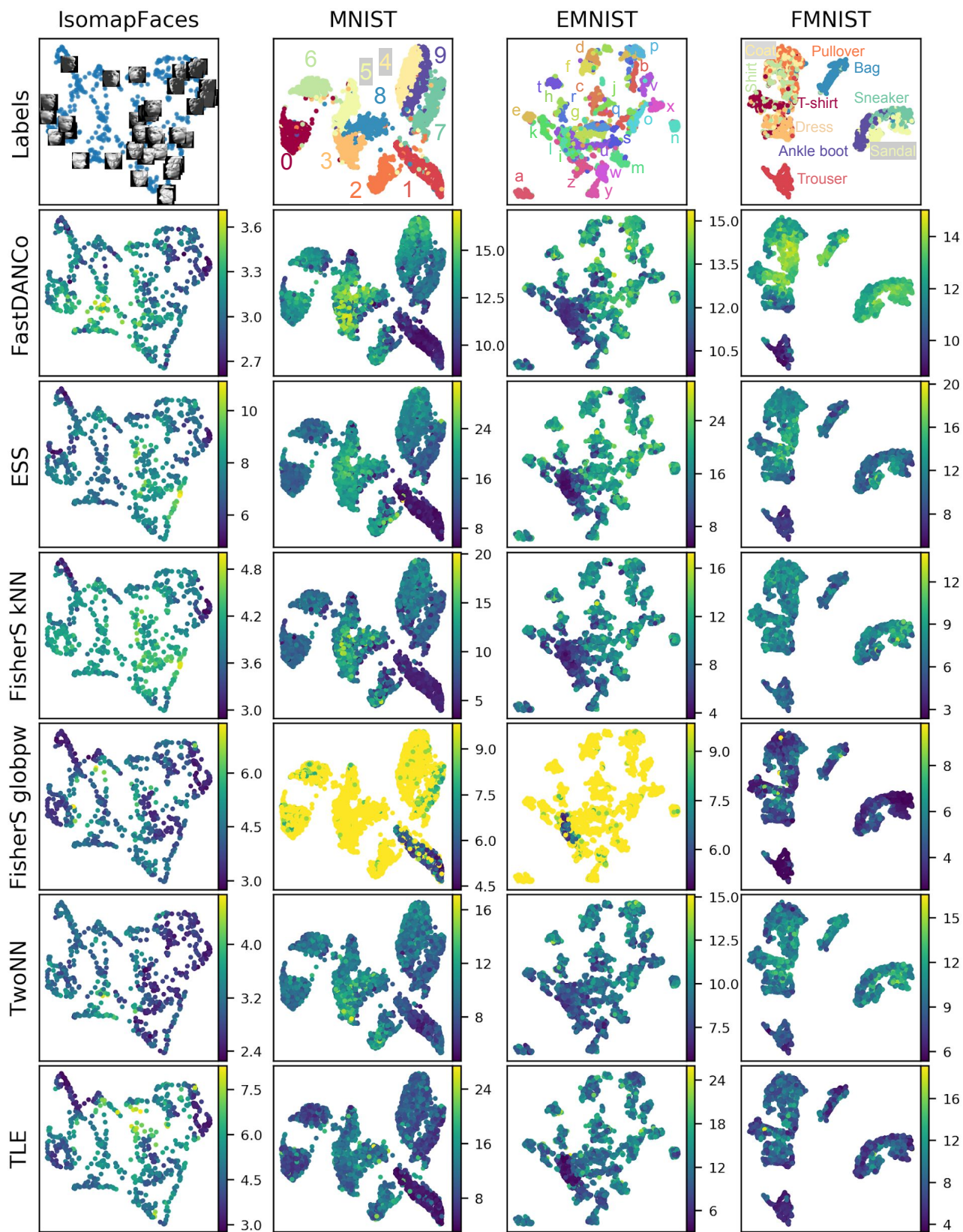
Fig. 2. Visualization of local ID measures (obtained with the 100 nearest neighbors of each point) on 2D layouts of real datasets projected onto their first 100 principal components and subsampled to 5000 points. For the MNIST datasets, the class labels are annotated directly on the UMAP scatter plot with colored labels located close to the corresponding clusters.
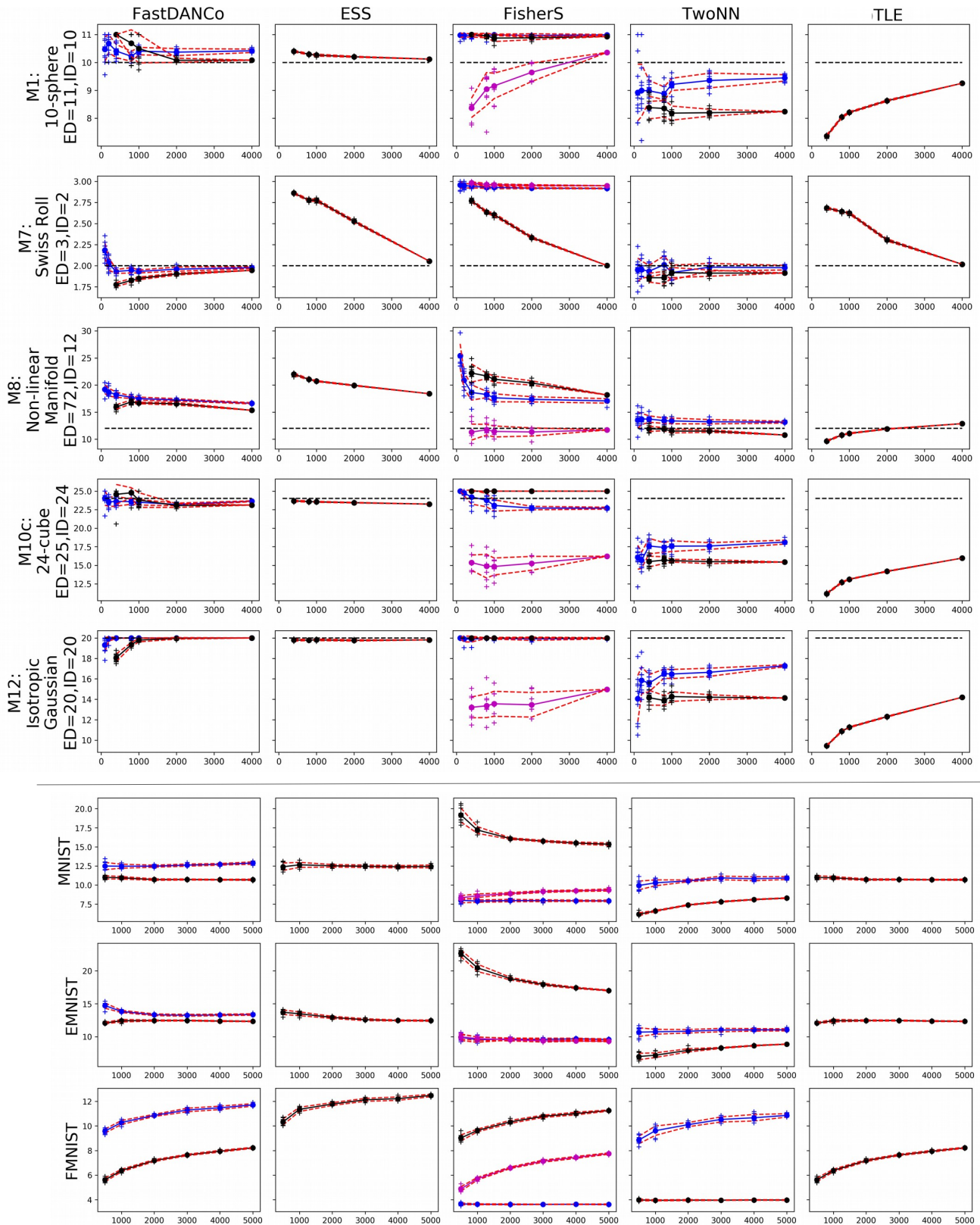
Fig. 3. Dependence of four ID estimators on subsample size. A selection of synthetic and real-life datasets used in this study is shown. The blue line shows the global ID estimate, while the black one shows the mean local ID. Red dashed lines show confidence intervals. In case of FisherS, the magenta line shows the mean global pointwise ID estimation. For the synthetic datasets, the dashed black line represents true ID. ESS and TLE are local estimators providing no specific procedure to produce global estimates, thus we only show mean local ID.

IV. Conclusion

We introduced two new local intrinsic dimensionality (ID) estimators, based on quantifying the separability probability of data points and comparing this statistic to a reference uniform distribution on a unit n-dimensional sphere. We performed comparative benchmarking of a number of local ID estimators based on concentration of measure using synthetic and real-life datasets. These benchmarks assessed the ability of the estimators to compute the correct ID from a small sample, their stability to random subsampling and the consistency of their local estimates.

We can conclude that the recently introduced family of local ID estimators exploiting various aspects of concentration of measure, including ours based on linear Fisher separability, performs well on synthetic and real datasets. However, we can distinguish three estimators, FisherS, ESS and DANCo, which seem to be the most adapted to estimate a wide range of ID (from three to several tens) for a neighborhood size of a hundred points or more. ESS showed high agreement with FisherS and can give better results in synthetic cases such as uniform distribution of points in an n-dimensional ball with large n, while FisherS can give closer to expected estimates for real-life datasets such as ISOMAP Faces. In addition, the global pointwise variant of FisherS estimate combines information about both local and global data point cloud properties, since it quantifies the separability of a data point from all other points, and not only from the closest neighbours. This property might appear useful in some specific applications.

References

[1] A. N. Gorban, V. A. Makarov, and I. Y. Tyukin, "High-Dimensional Brain in a High-Dimensional World: Blessing of Dimensionality," *Entropy 2020, Vol. 22, Page 82*, vol. 22, no. 1, p. 82, jan 2020.

[2] M. Gromov, "Isoperimetry of waists and concentration of maps," *Geom. Funct. Anal.*, vol. 13, pp. 178–215, 2003.

[3] Y. Jia, J. Bailey, K. Ramamohanarao, C. Leckie, and M. E. Houle, "Improving the quality of explanations with local embedding perturbations," 2019.

[4] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2016, pp. 1135–1144.

[5] H. Jiang, B. Kim, M. Guan, and M. Gupta, "To trust or not to trust a classifier," in *Advances in neural information processing systems*, 2018, pp. 5541–5552.

[6] J. Gilmer, L. Metz, F. Faghri, S. S. Schoenholz, M. Raghu, M. Wattenberg, and I. Goodfellow, "Adversarial spheres," *arXiv preprint arXiv:1801.02774*, 2018.

[7] A. Fawzi, H. Fawzi, and O. Fawzi, "Adversarial vulnerability for any classifier," in *Advances in Neural Information Processing Systems*, 2018, pp. 1178–1187.

[8] A. N. Bhagoji, D. Cullina, and P. Mittal, "Lower bounds on adversarial robustness from optimal transport," in *Advances in Neural Information Processing Systems*, 2019, pp. 7496–7508.

[9] S. Mahloujifar, X. Zhang, M. Mahmoody, and D. Evans, "Empirically measuring concentration: Fundamental limits on intrinsic robustness," *arXiv preprint arXiv:1905.12202*, 2019.

[10] A. Gorban, A. Golubkov, B. Grechuk, E. Mirkes, and I. Tyukin, "Correction of ai systems by linear discriminants: Probabilistic foundations," *Information Sciences*, vol. 466, pp. 303 – 322, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025518305607

[11] I. Y. Tyukin, A. N. Gorban, A. McEwan, and S. Meshkinfamfard, "Bringing the blessing of dimensionality to the edge," in *2019 1st International Conference on Industrial Artificial Intelligence (IAI)*, July 2019, pp. 1–5.

[12] A. N. Gorban and I. Y. Tyukin, "Blessing of dimensionality: mathematical foundations of the statistical physics of data," *Philos Trans A Math Phys Eng Sci*, vol. 376, no. 2118, Apr 2018.

[13] A. Gorban, V. Makarov, and I. Tyukin, "Symphony of high-dimensional brain. Reply to comments on "The unreasonable effectiveness of small neural ensembles in high-dimensional brain"," *Phys. Life Rev.*, vol. 29, pp. 115–119, 2019.

[14] J. Bac and A. Zinovyev, "Lizard Brain: Tackling Locally Low-Dimensional Yet Globally Complex Organization of Multi-Dimensional Datasets," *Frontiers in Neurorobotics*, vol. 13, jan 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnbot.2019.00110/full

[15] L. Amsaleg, O. Chelly, M. E. Houle, K.-i. Kawarabayashi, M. Radovanović, and W. Treeratanajaru, "Intrinsic dimensionality estimation within tight localities," in *Proceedings of the 2019 SIAM International Conference on Data Mining*. SIAM, 2019, pp. 181–189.

[16] O. Chelly, M. E. Houle, and K.-i. Kawarabayashi, "Enhanced estimation of local intrinsic dimensionality using auxiliary distances," Technical report, National Institute of Informatics, Tech. Rep., 2016.

[17] K. Johnsson, "Structures in high-dimensional data: Intrinsic dimension and cluster analysis," Ph.D. dissertation, Faculty of Engineering, LTH, 8 2016.

[18] D. R. Wissel, "Intrinsic dimension estimation using simplex volumes," Ph.D. dissertation, 2018. [Online]. Available: http://hss.ulb.uni-bonn.de/2018/4951/4951.htm

[19] M. Díaz, A. J. Quiroz, and M. Velasco, "Local angles and dimension estimation from data on manifolds," 2018.

[20] L. Albergante, J. Bac, and A. Zinovyev, "Estimating the effective dimension of large biological datasets using fisher separability analysis," in *Proc. IEEE IJCNN 2019 - International Joint Conference on Neural Networks*. Budapest Hungary: IEEE, 2019, pp. N–19 814.

[21] V. Erba, M. Gherardi, and P. Rotondo, "Intrinsic dimension estimation for locally undersampled data," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

[22] K. Fukunaga and D. Olsen, "An Algorithm for Finding Intrinsic Dimensionality of Data," *IEEE Transactions on Computers*, vol. C-20, no. 2, pp. 176–183, feb 1971. [Online]. Available: http://ieeexplore.ieee.org/document/1671801/

[23] J. Bruske and G. Sommer, "Intrinsic dimensionality estimation with optimally topology preserving maps," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 5, pp. 572–575, 1998.

[24] M. Fan, N. Gu, H. Qiao, and B. Zhang, "Intrinsic dimension estimation of data by principal component analysis," Tech. Rep., 2010. [Online]. Available: https://pdfs.semanticscholar.org/7d4d/936df2f628550123cef40996277628468e61.pdf

[25] A. V. Little, J. Lee, Y.-M. Jung, and M. Maggioni, "Estimation of intrinsic dimensionality of samples from noisy low-dimensional manifolds in high dimensions with multiscale SVD," in *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*. IEEE, 2009, pp. 85–88.

[26] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D: Nonlinear Phenomena*, vol. 9, no. 1-2, pp. 189–208, oct 1983. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0167278983902981

[27] E. Levina and P. J. Bickel, "Maximum Likelihood estimation of intrinsic dimension," pp. 777–784, 2004. [Online]. Available: https://dl.acm.org/citation.cfm?id=2976138

[28] C. Ceruti, S. Bassis, A. Rozza, G. Lombardi, E. Casiraghi, and P. Campadelli, "DANCo: An intrinsic dimensionality estimator exploiting angle and norm concentration," *Pattern Recognition*, vol. 47, no. 8, pp. 2569–2581, aug 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003132031400065X

[29] E. Facco, M. D'Errico, A. Rodriguez, and A. Laio, "Estimating the intrinsic dimension of datasets by a minimal neighborhood information," *Scientific Reports*, vol. 7, no. 1, p. 12140, dec 2017. [Online]. Available: http://www.nature.com/articles/s41598-017-11873-y

[30] A. N. Gorban, I. Y. Tyukin, D. V. Prokhorov, and K. I. Sofeikov, "Approximation with random bases: Pro et Contra," *Information Sciences*, vol. 364-365, pp. 129–145, oct 2016.

[31] P. J. Verveer and R. P. W. Duin, "An evaluation of intrinsic dimensionality estimators," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 17, no. 1, pp. 81–86, 1995.