OPTIMIZED INTRINSIC DIMENSION ESTIMATOR USING NEAREST NEIGHBOR GRAPHS

Kumar Sricharan[1*], *Raviv Raich*[2], *Alfred O. Hero III*[1]

[1] Department of EECS, University of Michigan, Ann Arbor, MI 48109
[2] School of EECS, Oregon State University, Corvallis, OR 97331
{kksreddy,hero}@umich.edu, raich@eecs.oregonstate.edu

ABSTRACT

We develop an approach to intrinsic dimension estimation based on k-nearest neighbor (kNN) distances. The dimension estimator is derived using a general theory on functionals of kNN density estimates. This enables us to predict the performance of the dimension estimation algorithm. In addition, it allows for optimization of free parameters in the algorithm. We validate our theory through simulations and compare our estimator to previous kNN based dimensionality estimation approaches.

Index Terms— intrinsic dimension, manifold learning, k nearest neighbor , kNN density estimation, geodesics

1. INTRODUCTION

Intrinsic dimensionality is an important concept in high dimensional datasets whose principal modes of variation lie on a subspace of substantially lower dimension, the intrinsic dimension d. In such cases dimensionality reduction can be accomplished without loss of information. An accurate estimator of intrinsic dimension is a prerequisite for setting the embedding dimension of DR algorithms such as principal components analysis (PCA), ISOMAP, and Laplacian eigenmaps. Until recently the most common method for selecting an embedding dimension for these algorithms was to detect a knee in a residual error curve, e.g., scree plots of sorted eigenvalues. In this paper we introduce a new dimensionality estimator that is based on fluctuations of the sizes of nearest neighbor balls centered at a subset of the data points. In this respect it is similar to Costa's k-nearest neighbor (kNN) graph dimension estimator [1] and to Farahmand's dimension estimator based on nearest neighbor distances [2]. The estimator can also be related to the Leonenko's Rényi entropy estimator [3]. However, unlike these estimators, our new dimension estimator is derived directly from a mean squared error (M.S.E.) optimality condition for partitioned kNN estimators of multivariate density functionals. This guarantees that our estimator has the best possible M.S.E. convergence rate among estimators in its class. Empirical experiments are presented that show that this asymptotic optimality translates into improved performance in the finite sample regime.

The paper is organized as follows. In Sec. 2.1 we introduce the the general form of the new dimension estimator. In Sec. 2.2. we show that the estimator is related to a general class of kNN density estimators. In Sec. 3 we review results on the statistical properties of functionals of kNN density estimators and in Sec. 4 we use this theory to obtain expressions for the asymptotic bias and variance of the new dimension estimator, in addition to establishing that it satisfies a central limit theorem. The analytical expressions for bias and variance allow us to optimize over the tuning parameters of the dimension estimator. This is shown in Sec. 5. Next, motivated by the analysis in Sec. 3, in Sec. 6 we propose a modified dimension estimator with reduced variance. Finally, in Sec. 7 we report empirical comparisons that illustrate the improved performance of the new dimensionality estimator relative to previous approaches.

In this paper, bold face type will be used to indicate random variables and random vectors. We denote the expectation operator by $\mathbb{E}[.]$ and the variance operator by $\mathbb{V}[.]$.

2. PROBLEM FORMULATION

Let $\mathcal{Y} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_T\}$ be T independent and identically distributed sample realizations in $\mathbb{R}^D$ distributed according to density f. Assume the random vectors in $\mathcal{Y}$ are constrained to lie on a d-dimensional Riemannian submanifold M of $\mathbb{R}^D$ $(d < D)$. We are interested in estimating the intrinsic dimension d.

2.1. Log-length statistics

Let $\gamma > 0$ be any arbitrary number and $\alpha = \gamma/d$. Partition the T samples in $\mathcal{Y}$ into two disjoint sets $\mathcal{X}$ and $\mathcal{Z}$ of size $\lfloor T/2 \rfloor$ each. Denote the samples of $\mathcal{X}$ as $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_{\lfloor T/2 \rfloor}\}$ and $\mathcal{Z}$ as $\mathcal{Z} = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_{\lfloor T/2 \rfloor}\}$.

Further partition $\mathcal{X}$ into N 'target' samples $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ and M 'reference' samples $\{\mathbf{X}_{N+1}, \ldots, \mathbf{X}_{\lfloor T/2 \rfloor}\}$ with $N + M = \lfloor T/2 \rfloor$. Partition $\mathcal{Z}$ in an identical manner. Now consider the following statistics based on the partitioning of sample space:

$$\mathbf{L_k}(\mathcal{X}) = \frac{\gamma}{N} \sum_{i=1}^{N} \log\left(\mathbf{R_k}(\mathbf{X}_i)\right),$$

where $\mathbf{R_k}(\mathbf{X}_i)$ is the k nearest neighbor (kNN) distance from target sample X_i to the M reference samples $\mathbf{X}_{N+1}, \ldots, \mathbf{X}_{N+M}$. This partitioning of samples is illustrated in Fig. 1.

2.2. Relation to kNN density estimates

Under the condition that k/M is small, the Euclidean kNN distance $\mathbf{R_k}(\mathbf{X}_i)$ approximates the kNN distance on the manifold. The kNN density estimate [4] of f at $\mathbf{X}_i$ based on the M samples $\mathbf{X}_{N+1}, \ldots, \mathbf{X}_{N+M}$ is then given by

$$\hat{\mathbf{f}}_{\mathbf{k}}(\mathbf{X_i}) = \frac{k-1}{M} \frac{1}{c_d \mathbf{R_k}(\mathbf{X}_i)^d} = \frac{k-1}{M} \frac{1}{\mathbf{V_k}(\mathbf{X}_i)},$$

*Acknowledgement: This work is partially funded by the Air Force Office of Scientific Research, grant number FA9550-09-1-0471.

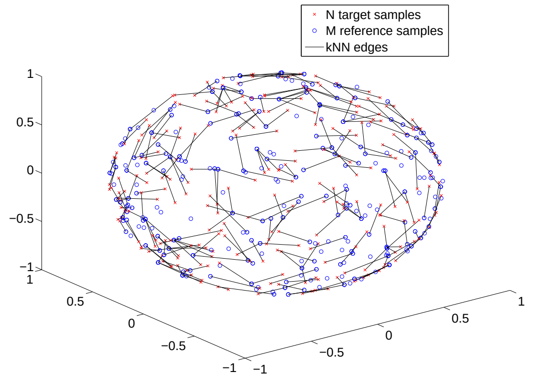


Fig. 1. kNN edges on sphere manifold with uniform distribution for $d = 2$, $D = 3$, and $k = 5$.

where $\mathbf{R_k}(\mathbf{X}_i)$ is the kNN distance from $\mathbf{X}_i$ to the M samples $\mathbf{X}_{N+1}, \ldots, \mathbf{X}_{N+M}$, c_d is the volume of the unit ball in d dimensions and therefore $\mathbf{V_k}(\mathbf{X_i})$ is the volume of the kNN ball. This implies that $\mathbf{L_k}(\mathcal{X})$ can be rewritten as follows:

$$\begin{aligned} \mathbf{L_k}(\mathcal{X}) &= \frac{\gamma}{N}\sum_{i=1}^{N} \log\left(\mathbf{R_k}(\mathbf{X}_i)\right) \\ &= \log\left(\frac{k-1}{Mc_d}\right)^{\alpha} + \frac{1}{N}\sum_{i=1}^{N}\log\left(\hat{\mathbf{f}}_\mathbf{k}(\mathbf{X}_i)\right)^{-\alpha} \\ &= \alpha\log(k-1) - \frac{\alpha}{N}\sum_{i=1}^{N}\log\hat{\mathbf{f}}_\mathbf{k}(\mathbf{X}_i) \\ &\quad -\alpha\log(c_d M). \end{aligned} \tag{1}$$

As eq. (1) indicates, the log-length statistics is linear with respect to $\log(k-1)$ with a slope of α. This prompts the idea of estimating α (and later d) from the slope of $\mathbf{L_k}(\mathcal{X})$ as a function of $\log(k-1)$.

2.3. Intrinsic dimension estimate based on varying bandwidth k

Let k_1 and k_2 be two different choices of bandwidth parameters. Let $\mathbf{L_{k_1}}(\mathcal{X})$ and $\mathbf{L_{k_2}}(\mathcal{Z})$ be the length statistics evaluated at bandwidths k_1 and k_2 using data $\mathcal{X}$ and $\mathcal{Z}$ respectively. A natural choice for the estimate of α would then be

$$\begin{aligned} \hat{\alpha} &= \frac{\mathbf{L_{k_2}}(\mathcal{Z}) - \mathbf{L_{k_1}}(\mathcal{X})}{\log(k_2-1) - \log(k_1-1)} \\ &= \alpha + \frac{\nu}{N}\sum_{i=1}^{N}\left(\log\hat{\mathbf{f}}_\mathbf{k_2}(\mathbf{Z}_i) - \log\hat{\mathbf{f}}_\mathbf{k_1}(\mathbf{X}_i)\right) \\ &= \alpha + \nu(\hat{\mathbf{E}}_\mathbf{k_2}(\mathcal{Z}) - \hat{\mathbf{E}}_\mathbf{k_1}(\mathcal{X})), \end{aligned}$$

where

$$\hat{\mathbf{E}}_\mathbf{k}(\mathcal{X}) = \frac{1}{N}\sum_{i=1}^{N}\log(\hat{\mathbf{f}}_\mathbf{k}(\mathbf{X_i})),$$

and $\nu = -\alpha/\log((k_2-1)/(k_1-1))$. The intrinsic dimension estimate is related to $\hat{\alpha}$ by the simple relation $\hat{\mathbf{d}} = \gamma/\hat{\alpha}$.

3. STATISTICAL PROPERTIES OF ESTIMATES OF FUNCTIONALS OF DENSITIES

We make the assumption that f is two times continuously differentiable and is bounded away from 0. Under this assumption, we have established the following the following theorems on the expectation, variance and asymptotic distribution of the density functional estimate $\hat{\mathbf{E}}_\mathbf{k}(\mathcal{X})$ in our work on statistical estimators of non-linear functionals of densities [5]. Let $\mathbf{Y}$ be distributed according to f. Denote $E = \mathbb{E}[\log(f(\mathbf{Y}))]$.

Theorem 3.1. *The bias of the log-length statistic $\hat{\mathbf{E}}_\mathbf{k}(\mathcal{X})$ is given by*

$$\mathbb{E}[\hat{\mathbf{E}}_\mathbf{k}(\mathcal{X})] - E = c_{b1}\left(\frac{k}{M}\right)^{2/d} + \frac{c_{b2}}{k} + o\left(\left(\frac{k}{M}\right)^{2/d} + \frac{1}{k}\right),$$

where $c_{b_1} = C_d\mathbb{E}[f^{(-1-2/d)}(\mathbf{Y})tr[\nabla^2(f(\mathbf{Y}))]]$, $c_{b2} = -0.5$ and the constant $C_d = (\Gamma^{(2/d)}((d+2)/2))/(\pi(d+2))$.

Theorem 3.2. *The variance of the log-length statistic $\hat{\mathbf{E}}_\mathbf{k}(\mathcal{X})$ is given by*

$$\mathbb{V}[\hat{\mathbf{E}}_\mathbf{k}(\mathcal{X})] = c_v\left(\frac{1}{N}\right) + o\left(\frac{1}{M} + \frac{1}{N}\right),$$

where $c_v = \mathbb{V}\left[\log(f(\mathbf{Y}))\right]$.

Theorem 3.3. *Let $\mathbf{Z}$ be a standard normal random variable. Then,*

$$\lim_{N,M\to\infty} Pr\left(\frac{\hat{\mathbf{E}}_\mathbf{k}(\mathcal{X}) - \mathbb{E}[\hat{\mathbf{E}}_\mathbf{k}(\mathcal{X})]}{\sqrt{c_v/N}} \le \alpha\right) = Pr(\mathbf{Z} \le \alpha).$$

4. STATISTICAL PROPERTIES OF INTRINSIC DIMENSION ESTIMATE

We can relate the error in estimation of α to the error in dimension estimation as follows:

$$\begin{aligned} \hat{\mathbf{d}} - d &= \gamma\left(\frac{1}{\hat{\alpha}} - \frac{1}{\alpha}\right) \\ &= \gamma\frac{\alpha - \hat{\alpha}}{\hat{\alpha}\alpha} \\ &= -\frac{\gamma}{\alpha^2}(\hat{\alpha} - \alpha) + o(\hat{\alpha} - \alpha). \end{aligned}$$

Define $\kappa = -\gamma\nu/\alpha^2$. Using the results on the length statistic $\hat{\mathbf{E}}_\mathbf{k}(\mathcal{X})$ from the previous section and the above relation between the errors $\hat{\mathbf{d}} - d$ and $\hat{\alpha} - \alpha$, we have the following statistical properties for the estimate $\hat{\mathbf{d}}$:

Estimator bias

$$\begin{aligned} \mathbb{E}[\hat{\mathbf{d}}] - d &= \kappa c_{b_1}\left(\left(\frac{k_2}{M}\right)^{2/d} - \left(\frac{k_1}{M}\right)^{2/d}\right) \\ &+ \kappa c_{b_2}\left(\left(\frac{1}{k_2}\right) - \left(\frac{1}{k_1}\right)\right) \\ &+ o\left(\frac{1}{k_1} + \frac{1}{k_2} + \left(\frac{k_1}{M}\right)^{2/d} + \left(\frac{k_2}{M}\right)^{2/d}\right). \end{aligned}$$

Estimator variance

$$\mathbb{V}(\hat{\mathbf{d}}) = 2\kappa^2 c_v\left(\frac{1}{N}\right) + o\left(\frac{1}{M} + \frac{1}{N}\right).$$

Central limit theorem

Let $\mathbf{Z}$ be a standard normal random variable. Then,

$$\lim_{N,M\to\infty} Pr\left(\frac{\hat{\mathbf{d}} - \mathbb{E}[\hat{\mathbf{d}}]}{\sqrt{2\kappa^2 c_v/N}} \leq \alpha\right) = Pr(\mathbf{Z} \leq \alpha).$$

5. OPTIMAL SELECTION OF PARAMETERS

We have theoretical expressions for the mean square error (M.S.E) of the dimension estimate $\hat{\mathbf{d}}$, which we can optimize over the free parameters k_1, k_2, N and M [6]. We restrict our attention to the case $k_2 = 2k$; $k_1 = k$. The M.S.E. of $\hat{\mathbf{d}}$ (ignoring higher order terms) is given by

$$\begin{aligned} \text{M.S.E.}(\hat{\mathbf{d}}) &= (\mathbb{E}[\hat{\mathbf{d}}] - d)^2 + \mathbb{V}[\hat{\mathbf{d}}] \\ &= \left(C_{b_1}\left(\frac{k}{M}\right)^{2/d} + C_{b_2}\left(\frac{1}{k}\right)\right)^2 \\ &+ C_v\left(\frac{1}{N}\right). \end{aligned} \quad (2)$$

where $C_{b_1} = \kappa 2^{(2/d-1)}$, $C_{b_2} = \kappa/4$ and $C_v = 2\kappa^2 c_v$.

5.1. Optimal choice of bandwidth

The optimal value of k w.r.t the M.S.E. is given by

$$k_{opt} = \lfloor k_0 M^{\frac{2}{2+d}} \rfloor. \quad (3)$$

where the constant $k_0 = (|C_{b_2}|d/2|C_{b_1}|)^{\frac{d}{d+2}}$.

5.2. Optimal partitioning of sample space

Under the constraint that $N + M = \lfloor T/2 \rfloor$ is fixed, the optimal choice of N as a function of M is then given by

$$N_{opt} = \lfloor N_0 M^{\frac{6+d}{2(2+d)}} \rfloor, \quad (4)$$

where the constant $N_0 = \frac{\sqrt{C_v(2+d)}}{2b_0}$.

6. IMPROVED ESTIMATOR BASED ON CORRELATED ERROR

Consider the following alternative estimator for α:

$$\begin{aligned} \tilde{\alpha} &= \frac{\mathbf{L_{k_2}}(\mathcal{X}) - \mathbf{L_{k_1}}(\mathcal{X})}{\log(k_2 - 1) - \log(k_1 - 1)} \\ &= \alpha + \kappa(\hat{\mathbf{E}}_{\mathbf{k_2}}(\mathcal{X}) - \hat{\mathbf{E}}_{\mathbf{k_1}}(\mathcal{X})), \end{aligned}$$

and the corresponding density estimate $\tilde{\mathbf{d}}$ which satisfies

$$\tilde{\mathbf{d}} - d = -\frac{\gamma}{\alpha^2}(\tilde{\alpha} - \alpha) + o(\hat{\alpha} - \alpha),$$

where both the length statistics at bandwidths k_1 and k_2 are evaluated using the same sample X. The density functional estimates $\hat{\mathbf{E}}_{\mathbf{k_1}}(\mathcal{X})$ and $\hat{\mathbf{E}}_{\mathbf{k_2}}(\mathcal{X})$ will be highly correlated (as compared to the independent quantities $\hat{\mathbf{E}}_{\mathbf{k_1}}(\mathcal{X})$ and $\hat{\mathbf{E}}_{\mathbf{k_2}}(\mathcal{Z})$). This implies that the variance of the difference $\hat{\mathbf{E}}_{\mathbf{k_2}}(\mathcal{X}) - \hat{\mathbf{E}}_{\mathbf{k_1}}(\mathcal{X})$ will be smaller when compared to $\hat{\mathbf{E}}_{\mathbf{k_2}}(\mathcal{Z}) - \hat{\mathbf{E}}_{\mathbf{k_1}}(\mathcal{X})$, (while the expectation remains the same).

Since the estimator bias is unaffected by this modification, the variance reduction suggests that $\tilde{d}$ will be an improved estimator as compared to $\hat{\mathbf{d}}$ in terms of M.S.E.. In order to obtain statistical properties for the improved estimator $\tilde{\mathbf{d}}$ (equivalent to the properties developed in Section 4 for the original estimator $\hat{\mathbf{d}}$), we need to analyze the joint distribution between $\hat{\mathbf{f}}_{\mathbf{k_1}}(X_i)$ and $\hat{\mathbf{f}}_{\mathbf{k_2}}(X_j)$ for two distinct values k_1 and k_2. Our theory, at present, cannot address the case of distinct bandwidths k_1 and k_2.

Since the estimate $\tilde{\mathbf{d}}$ has smaller M.S.E. compared to $\hat{\mathbf{d}}$, M.S.E. predictions for the estimate $\hat{\mathbf{d}}$ can serve as upper bounds on the M.S.E. performance of the improved estimate $\tilde{\mathbf{d}}$.

7. SIMULATIONS

We generate $T = 10^5$ samples $\mathcal{B}$ drawn from a $d = 2$ mixture density $f_m = .8f_\beta + .2f_u$, where f_β is the product of two 1 dimensional marginal beta distributions with parameters $\alpha = 2$, $\beta = 2$ and f_u is a uniform density in 2 dimensions. These samples are then projected to a 3-dimensional hyperplane in $\mathbb{R}^3$ by applying the transformation $\mathcal{Y} = U\mathcal{B}$ where U is a 3×2 random matrix whose columns are orthonormal. We apply our intrinsic dimension estimates on the samples $\mathcal{Y}$.

7.1. Optimal selection of free parameters

In our first experiment, we theoretically compute the optimal choice of k for a fixed partition with $M = 3.5\times10^4$ and $N = 1.5\times10^4$. We then show the variation of the theoretical and experimental M.S.E. of the estimate $\hat{\mathbf{d}}$ and the experimental M.S.E. of the improved estimate $\tilde{\mathbf{d}}$ with changing bandwidth k in Fig. 2(a). In our second experiment, we compute the optimal partition according to eq. (4) and show the variation of M.S.E. with varying choices of partition in Fig. 2(b).

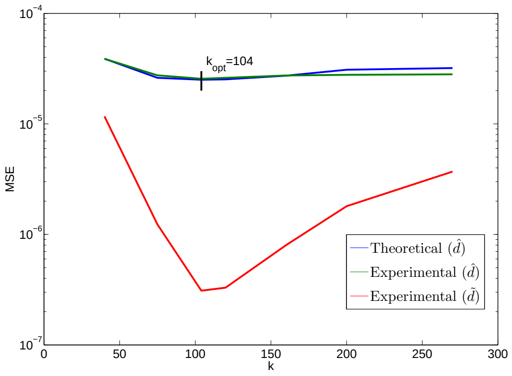
From our experiments, we see that there is good agreement between our theory and simulations. As a consequence, we find the theoretically predicted optimal choices of $k, N$and M to minimize the observed M.S.E.. In addition, as predicted by our theory, the modified estimator $\tilde{\mathbf{d}}$ significantly outperforms $\hat{\mathbf{d}}$. The theoretically predicted M.S.E. for $\hat{\mathbf{d}}$ therefore serves as a strict upper bound for the M.S.E. of the improved estimator $\tilde{\mathbf{d}}$.

7.2. Comparison of dimension estimation methods

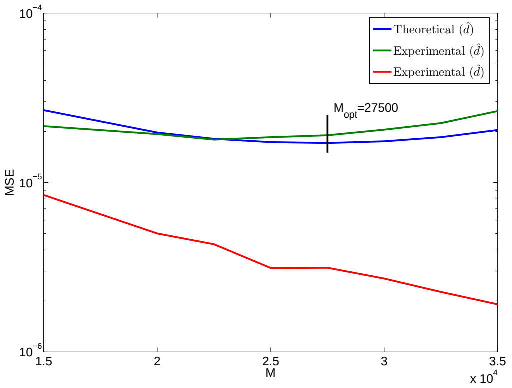
We compare the performance of our proposed dimension estimators to the estimated proposed by Frahmand et. al. [2] (denote as $\hat{\mathbf{d}}_f$) and Costa et. al. [1] (denote as $\hat{\mathbf{d}}_j$).

Expressions for the optimal bandwidth k (eq. (3)) and partition N, M (eq. (4)) depend on the unknown intrinsic dimension d and constants c_{b_1}, c_{b_2} and c_v which depend on unknown density f. The constants c_{b_1}, c_{b_2} and c_v can be estimated from the data using plug-in methods similar to the ones used by Raykar et. al. [7] for optimal bandwidth selection for kernel density estimation . To establish the potential advantages of our dimension estimators we compare an omniscient optimal form of our estimator, for which the true values of these constants are known, to a suboptimal form of our estimator that does not know the constants.

For the optimal estimator, we theoretically compute the optimal choice for k, N and M for different choices of total sample size T (sub-sampled from the initial 10^5 samples), and use these optimal parameters for the estimators $\hat{\mathbf{d}}$ and $\tilde{\mathbf{d}}$. We use this optimal choice of bandwidth k for the estimators $\hat{\mathbf{d}}_f$ and $\hat{\mathbf{d}}_j$ as well (partitioning not applicable). For the suboptimal estimator, we arbitrarily choose the parameters as follows: fixed k = 20, $N = T/50$, $M = \lfloor T/2 \rfloor - N$.



(a) M.S.E. as function of k ($M = 3.5 \times 10^4$, $N = 1.5 \times 10^4$).



(b) M.S.E. as function of M ($\lfloor T/2 \rfloor = N + M = 5 \times 10^4$).

Fig. 2. Comparison of theoretically predicted and experimental M.S.E. for varying choices of k, N and M. The experimental performance of the estimator $\hat{\mathbf{d}}$ is in excellent agreement with the theoretical expression and, as predicted by our theory, the modified estimator $\tilde{\mathbf{d}}$ significantly outperforms $\hat{\mathbf{d}}$.

The performance of these estimators as a function of sample size T is shown in Fig. 3. Estimators with optimal choice of parameters are indicated in solid line, and the suboptimal estimators are indicated in dashed lines.

From our experiments we see that the performance of the original estimator $\hat{\mathbf{d}}$ with suboptimal choice of parameters is marginally inferior when compared to the estimator with optimal choice of parameters. This does not hold for the other estimators as can be expected since the parameters are optimized w.r.t. the performance of $\hat{\mathbf{d}}$.

We note that the improved estimator $\tilde{\mathbf{d}}$ outperforms all other estimators while the performance of our original estimator $\hat{\mathbf{d}}$ is sandwiched between $\hat{\mathbf{d}}_f$ and $\hat{\mathbf{d}}_j$. We conjecture that the performance of $\hat{\mathbf{d}}_j$ is superior to $\hat{\mathbf{d}}$ for the same reason that $\tilde{\mathbf{d}}$ outperforms $\hat{\mathbf{d}}$: correlated error between different length statistics.

8. CONCLUSIONS

We proposed a new estimator $\hat{\mathbf{d}}$ for intrinsic dimension estimation based on our theory on multivariate functionals of kNN density estimates. We present results on the bias, variance and asymptotic distribution of our proposed estimate. Using these results, we obtain theoretical expressions for optimal selection of free parameters - bandwidth k and partition scheme N, M - for minimum M.S.E.. We further improve upon the parameter optimized dimension estimator by applying a variance reducing correction that was motivated directly by our theory. Simulations validate the theoretical results presented in this paper. Furthermore, the improved estimator $\tilde{\mathbf{d}}$ is shown to have the best performance among other kNN based intrinsic dimension estimates.

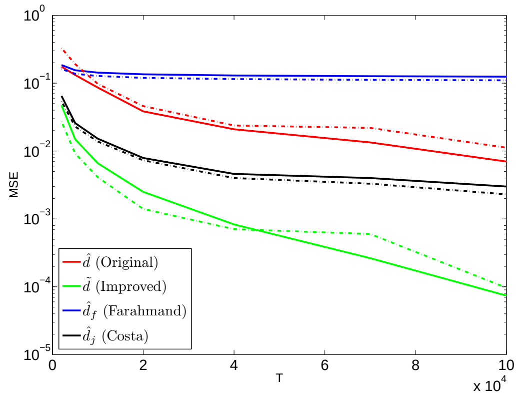


Fig. 3. Comparison of performance of dimension estimates (Solid line: Optimal (optimal choice of k,N and M as per eq. (3) and eq. (4)); Dashed line: Suboptimal (fixed $k = 20$, $N = T/50$, $M = \lfloor T/2 \rfloor - N$)): The proposed improved kNN distance estimator outperforms all other estimators considered.

9. REFERENCES

[1] J.A. Costa, A. Girotra, and A.O. Hero, "Estimating local intrinsic dimension with k-nearest neighbor graphs," in *2005 IEEE/SP 13th Workshop on Statistical Signal Processing*, 2005, pp. 417–422.

[2] A.M. Farahmand, C. Sepesvari, and J-Y Audibert, "Manifold-adaptive dimension estimation," *Proc of 24th Intl Conf on Machine Learning*, pp. 265–272, 2007.

[3] N. Leonenko, L. Prozanto, and V. Savani, "A class of rényi information estimators for multidimensional densities," *Annals of Statistics*, vol. 36, pp. 2153–2182, 2008.

[4] K. Fukunaga and L. Hostetler, "Optimization of k nearest neighbor density estimates," *IEEE Transactions on Information Theory*, vol. 19, no. 3, pp. 320–326, 1973.

[5] K. Sricharan, R. Raich, and A. O. Hero, "Plug-in estimators for non-linear functionals of densities," *Technical Report, Communications and Signal Processing Laboratory, The University of Michigan*, September 2009, (To appear).

[6] K. Sricharan, R. Raich, and A. O. Hero, "Global performance prediction for divergence-based image registration criteria," in *Proc. IEEE Workshop on Statistical Signal Processing*, 2009, (To appear).

[7] V. C. Raykar and R. Duraiswami, "Fast optimal bandwidth selection for kernel density estimation," in *Proceedings of the sixth SIAM International Conference on Data Mining*, J. Ghosh, D. Lambert, D. Skillicorn, and J. Srivastava, Eds., 2006, pp. 524–528.