Scaling Up Ordinal Embedding: A Landmark Approach (Appendix)

Jesse Anderton [1] **Javed Aslam** [1]

A. Additional Proofs

Notation Guide

- $\mathcal{X}$ are the original, unknown point positions.
- $\mathcal{Y}$ are the embedded positions.
- $\|x_x - x_y\| := \|x - y\|$ is the Euclidean distance between points x and y.
- $\vec{xy}$ is the displacement vector from x to y.

A.1. Distance Interval Bounds

The proof of Theorem 1 relies on the following lemma.

Lemma 1. *For any randomly drawn sample of size $m \geq \frac{2}{\epsilon} \ln \frac{2}{\epsilon\delta}$, all contiguous intervals of weight ϵ contain at least one point with probability at least $1 - \delta$.*

Proof of Lemma 1. We partition the range into $2/\epsilon$ contiguous intervals of weight $\epsilon/2$ each. Drawing m samples at random from the underlying distribution, the probability that all samples miss a particular interval is $(1 - \epsilon/2)^m$. By the union bound, the overall error probability of *any* interval being missed by all samples is upper-bounded by the sum of error probabilities for each interval, $\frac{2}{\epsilon}(1 - \epsilon/2)^m$, which we show below is at most δ.

$$\begin{aligned} m &\geq \frac{2}{\epsilon} \ln \frac{2}{\epsilon\delta} \quad = \quad \frac{2}{\epsilon} \ln \frac{2}{\epsilon} - \frac{2}{\epsilon} \ln \delta \\ \Leftrightarrow \quad \ln \delta &\geq \ln \frac{2}{\epsilon} - \frac{\epsilon}{2} m \\ \Leftrightarrow \quad \delta &\geq \frac{2}{\epsilon} (e^{-\epsilon/2})^m \quad \geq \quad \frac{2}{\epsilon}(1 - \epsilon/2)^m \end{aligned}$$

where the last step makes use of the fact that $e^{-x} \geq 1 - x$ for $x = \epsilon/2$. Finally, we note that any contiguous interval of weight ϵ must entirely span at least one of the consecutive intervals of weight $\epsilon/2$, thus containing a point with probability at least $1 - \delta$. $\square$

[1] College of Computer and Information Science, Northeastern University, Boston, Massachusetts. Correspondence to: Jesse Anderton <jesse@ccs.neu.edu>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

Proof of Theorem 1. By contradiction, assume that with probability greater than δ, the largest gap has probability weight $\epsilon \geq \frac{2}{m} \ln \frac{m}{\delta}$. Assuming that $\ln \frac{m}{\delta} \geq 1$, which is true for all δ and any $m \geq 3$, we have $\epsilon \geq 2/m$ or equivalently $m \geq 2/\epsilon$. We then have

$$\begin{aligned} \epsilon &\geq \frac{2}{m} \ln \frac{m}{\delta} \\ \Leftrightarrow \quad m &\geq \frac{2}{\epsilon} \ln \frac{m}{\delta} \quad \geq \quad \frac{2}{\epsilon} \ln \frac{2}{\epsilon\delta} \end{aligned}$$

But for such a sample size, Lemma 1 states that any such gap will be hit by at least one point with high probability, thus yielding a contradiction. The remainder of the theorem simply converts an interval of measure ϵ into a distance interval of length at most $\epsilon d_m L_{\mathcal{D}}$. $\square$

A.2. Shell Intersection Embedding Accuracy

We prove Theorem 2 by means of the following lemmas. We will first prove our vector projection order lemma.

Proof of Lemma 1. This follows from the Pythagorean theorem and the fact that $\angle px'x$ and $\angle qx'x$ are right for any $x \in X$ not collinear with p and q, and that $\|x_x - x_{x'}\| = 0$ when x is collinear with p and q. We will use x' to denote the (signed) distance from p to x' along $\vec{v}$, and $\|x_q - x_{x'}\|$ to denote the (signed) distance from q to x'. We begin by using $\|x_q - x_y\| < \|x_q - x_x\|$. For notational convenience, we will denote $\|x_x - x_y\|$ as $\mathrm{d}_{x,y}$.

$$\begin{aligned} \mathrm{d}_{q,y}^2 &< \mathrm{d}_{q,x}^2 \\ \Rightarrow \quad \mathrm{d}_{q,y'}^2 + \mathrm{d}_{y,y'}^2 &< \mathrm{d}_{q,x'}^2 + \mathrm{d}_{x,x'}^2 \\ \Rightarrow \quad (\mathrm{d}_{p,q} - y')^2 + \mathrm{d}_{y,y'}^2 &< (\mathrm{d}_{p,q} - x')^2 + \mathrm{d}_{x,x'}^2 \\ \Rightarrow \quad y'^2 + \mathrm{d}_{y,y'}^2 - 2(\mathrm{d}_{p,q})(y') &< x'^2 + \mathrm{d}_{x,x'}^2 - 2(\mathrm{d}_{p,q})(x') \\ \Rightarrow \quad \mathrm{d}_{p,y}^2 &< \mathrm{d}_{p,x}^2 + 2\,\mathrm{d}_{p,q}(y' - x') \end{aligned}$$

We can now use $\|x_p - x_x\| < \|x_p - x_y\|$ to complete our proof.

$$\begin{aligned} \mathrm{d}_{p,x}^2 &< \mathrm{d}_{p,y}^2 \\ \Rightarrow \quad \mathrm{d}_{p,x}^2 &< \mathrm{d}_{p,x}^2 + 2\,\mathrm{d}_{p,q}(y' - x') \\ \Rightarrow \quad 0 &< 2\,\mathrm{d}_{p,q}(y' - x') \\ \Rightarrow \quad x' &< y' \end{aligned}$$

□

The following lemma follows from Lemma 1 and from our anchor selection algorithm in Phase Two, providing a bound on the embedding error along individual vectors. We use this as the key lemma to prove our final embedding quality result.

Lemma 2 (Vector Projection Error Bound)**.** *For distinct $p, q, x, y \in \mathbb{R}^d$, let $\|x_p - x_x\|, \|x_p - x_y\| \in (l_p, u_p)$ and $\|x_q - x_x\|, \|x_q - x_y\| \in (l_q, u_q)$ hold for some $l_p, u_p, l_q, u_q \in \mathbb{R}$ such that $l_p + u_q \geq \|x_p - x_q\|$ and $l_q + u_p \geq \|x_p - x_q\|$. Then if x' and y' are the orthogonal projections of x and y (resp.) onto $\vec{pq}$ we have that*

$$|x' - y'| < \left(\frac{\|x_p - x_x\| + \|x_q - x_x\| + \gamma}{\|x_p - x_q\|}\right)\gamma$$

where $\gamma = \max\{u_p - l_p, u_q - l_q\}$.

Proof of Lemma 2. We will first show that the following inequality holds for both x' and y' by proving it holds without loss of generality for x'.

$$\frac{l_p^2 - u_q^2 + \|x_p - x_q\|^2}{2\|x_p - x_q\|} \leq x' (\text{or } y') \leq \frac{u_p^2 - l_q^2 + \|x_p - x_q\|^2}{2\|x_p - x_q\|}$$

To see that this inequality holds, consider any points $w \in \mathbb{R}^d$ such that $\|x_p - x_w\| = l_p$ and $\|x_q - x_w\| = u_q$ and $z \in \mathbb{R}^d$ such that $\|x_p - x_z\| = u_p$ and $\|x_q - x_z\| = l_q$. Such points must exist because the distances are restricted to satisfy the triangle inequality. By Lemma 1, we have $w' < x' < z'$ (and $w' < y' < z'$). The final interval follows from the Law of Cosines for $\triangle pqw$ and $\triangle pqz$.

We now observe that this interval has a gap of size $((u_p^2 - l_p^2) + (u_q^2 - l_q^2))/2\|x_p - x_q\|$, which goes to zero with ϵ. In particular, using $g_p = u_p - l_p$ and $g_q = u_q - l_q$, we have the following.

$$\begin{aligned}
&\frac{(u_p^2 - l_p^2) + (u_q^2 - l_q^2)}{2\|x_p - x_q\|} \\
=&\frac{(u_p - l_p)(u_p + l_p) + (u_q - l_q)(u_q + l_q)}{2\|x_p - x_q\|} \\
=&\frac{g_p(2l_p + g_p) + g_q(2l_q + g_q)}{2\|x_p - x_q\|} \\
\leq&\left(\frac{l_p + l_q}{\|x_p - x_q\|} + \frac{g_p + g_q}{2\|x_p - x_q\|}\right)\gamma \\
\leq&\left(\frac{\|x_p - x_x\| + \|x_q - x_x\|}{\|x_p - x_q\|} + \frac{\gamma}{\|x_p - x_q\|}\right)\gamma
\end{aligned}$$

□

We can now prove our embedding accuracy theorem.

Proof of Theorem 2 . Our set $\mathcal{A} \subset \mathcal{S}$ of anchors was chosen by farthest-first traversal, so it forms an ϵ-net. That is, for some $\epsilon \in \mathbb{R}$ we have $\forall a, b \in \mathcal{A}, \|x_a - x_b\| > \epsilon$ and $\forall x \in \mathcal{S}, \exists a \in \mathcal{A}$ s.t. $\|x_a - x_x\| < \epsilon$. Further, by the argument for Theorem **??** we have W.H.P. that for an arbitrary i.i.d. draw $x \in \mathcal{X}$ $\exists a \in \mathcal{A}$ s.t. $\|x_a - x_x\| < \epsilon$.

Choose $p \in \mathcal{A}$ s.t. $\|x_p - x_x\| < \epsilon$, and choose d arbitrary additional members $q_1, \ldots, q_d \in \mathcal{A}$. Since the points are affinely independent (as i.i.d. draws from a continuous measure), we have d linearly independent vectors $v_1, \ldots, v_d$ from p to each of the q_i. By Lemma 2, the distance between the projections of x and $\hat{x}$ onto each of these vectors will be at most

$$\left(\frac{\|x_p - x_x\| + \|x_q - x_x\| + \gamma}{\|x_p - x_q\|}\right)\gamma < 4\gamma,$$

where γ here is the size of the largest distance gap for any of the vectors. The inequality holds because $\|x_p - x_x\|/\|x_p - x_q\| < \epsilon/\epsilon = 1$, $\|x_q - x_x\|/\|x_p - x_q\| \leq (\|x_p - x_q\| + \|x_p - x_x\|)/\|x_p - x_q\| < 1 + \epsilon/\epsilon < 2$, and $\gamma/\|x_p - x_q\| < \gamma/\epsilon < 1$, where $\gamma < \epsilon$ is due to $m \gg |\mathcal{A}|$ and Theorem **??**.

We have thus bounded the displacement of $\hat{x}$ from x when projected along d linearly-independent vectors. It remains to bound the total distance between the points. This distance is the diagonal of the d-parallelotope defined by the vectors, whose length we can bound by the triangle inequality as the sum of the individual displacement vectors, leading to the bound

$$\|x - \hat{x}\| \leq 4d\gamma.$$

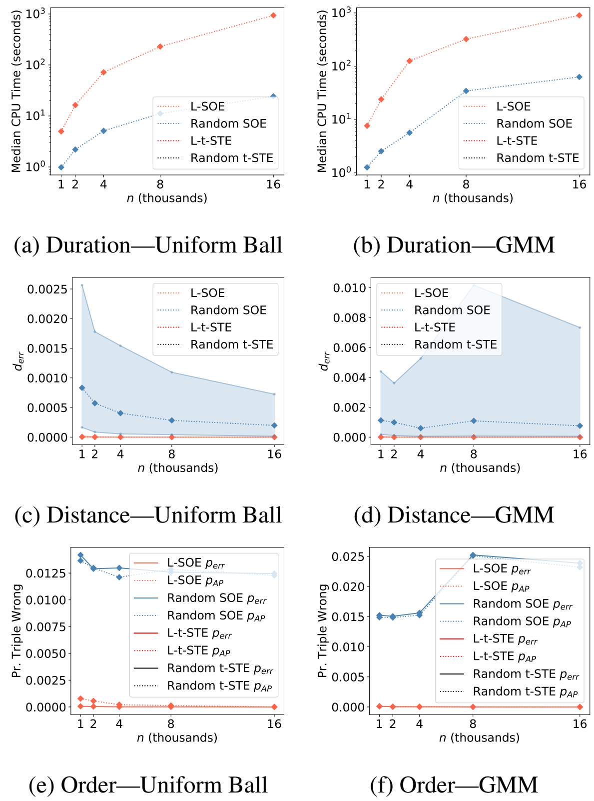
We arrive at the final result by applying Theorem **??** to upper bound γ with high probability. □

B. Additional Empirical Results

We here present the results of further experiments which supplement the results found in the main paper body.

B.1. L-SOE Performance

The plots in this section supplement those for L-SOE presented in the main paper. Figures 1, 2, 3, 4, 5, and 6 show embeddings of various dimensionalities. These results are very consistent with those in the main paper, and serve mainly to validate that the performance shown there holds for various values of n and d. We consistently observe that at these scales of n, random triplets can be embedded much faster and more consistently than can L-SOE triplets. However, when the L-SOE triplets are satisfied, and often even when they are not satisfied particularly well, the resulting embedding is of markedly higher quality. This holds until the embedding found violates the L-SOE triplets so badly that the distance error is sometimes worse than with random triplets, at which point sometimes p_{AP} (and thus also kNN performance) can be dramatically worse than for random triplets.

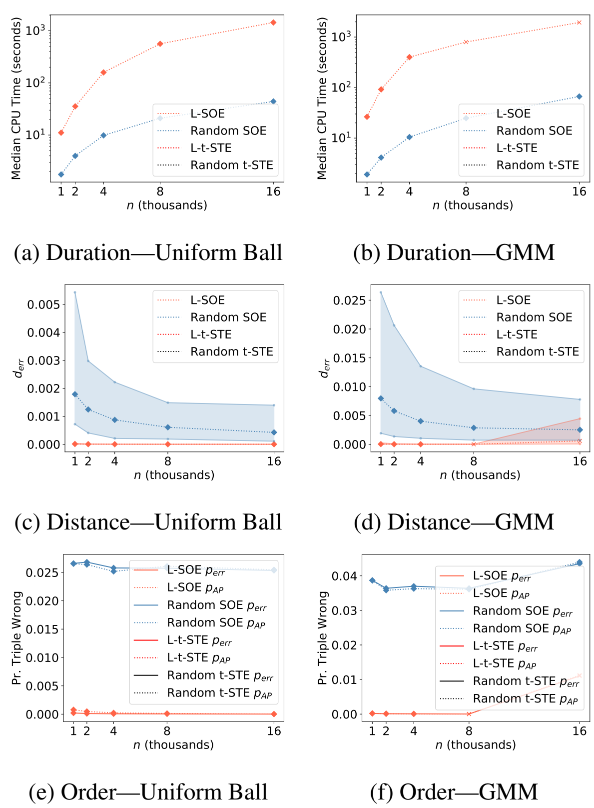


(a) Duration—Uniform Ball (b) Duration—GMM

(c) Distance—Uniform Ball (d) Distance—GMM

(e) Order—Uniform Ball (f) Order—GMM

Figure 1: L-SOE Performance in $\mathbb{R}^5$.

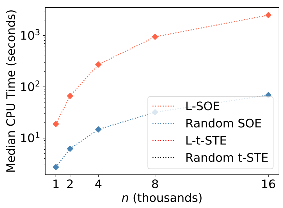


(a) Duration—Uniform Ball (b) Duration—GMM

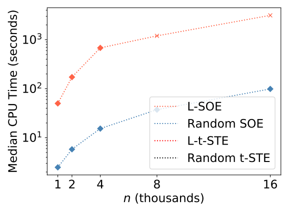
(c) Distance—Uniform Ball (d) Distance—GMM

(e) Order—Uniform Ball (f) Order—GMM

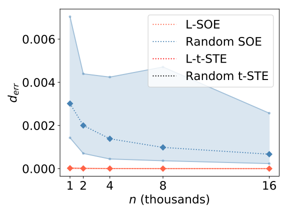
Figure 2: L-SOE Performance in $\mathbb{R}^{10}$.



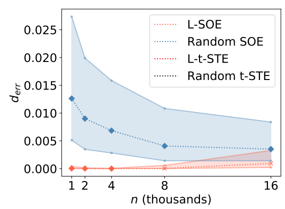
(a) Duration—Uniform Ball



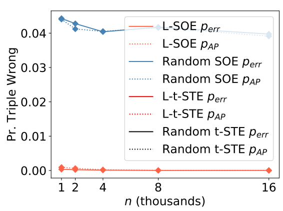
(b) Duration—GMM



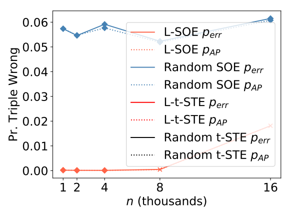
(c) Distance—Uniform Ball



(d) Distance—GMM

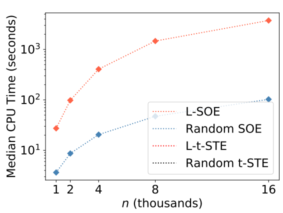


(e) Order—Uniform Ball

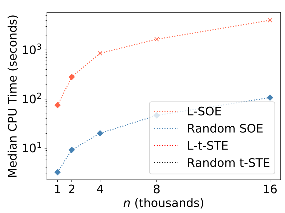


(f) Order—GMM

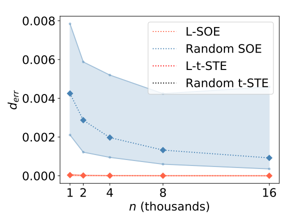
Figure 3: L-SOE Performance in $\mathbb{R}^{15}$.



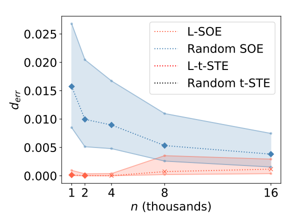
(a) Duration—Uniform Ball



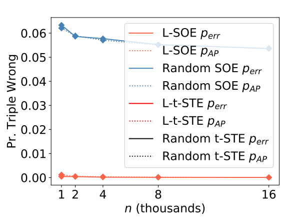
(b) Duration—GMM



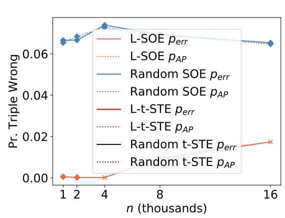
(c) Distance—Uniform Ball



(d) Distance—GMM

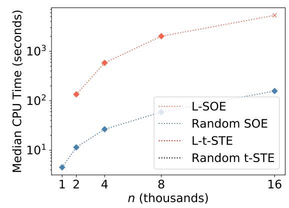


(e) Order—Uniform Ball

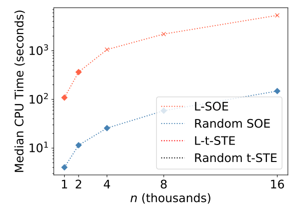


(f) Order—GMM

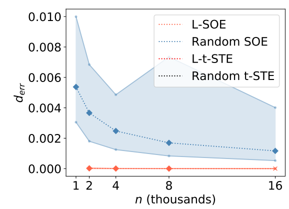
Figure 4: L-SOE Performance in $\mathbb{R}^{20}$.



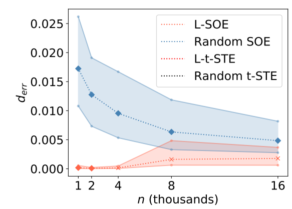
(a) Duration—Uniform Ball



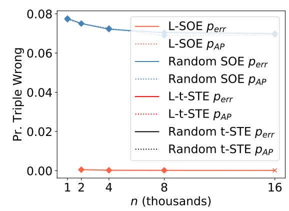
(b) Duration—GMM



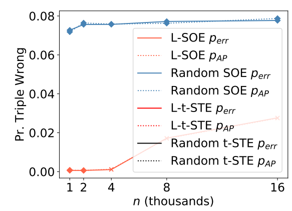
(c) Distance—Uniform Ball



(d) Distance—GMM

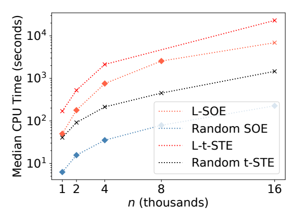


(e) Order—Uniform Ball

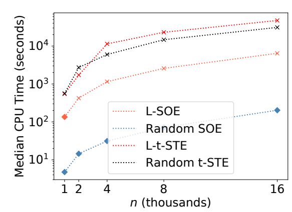


(f) Order—GMM

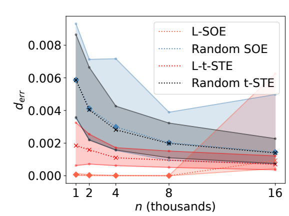
Figure 5: L-SOE Performance in $\mathbb{R}^{25}$.



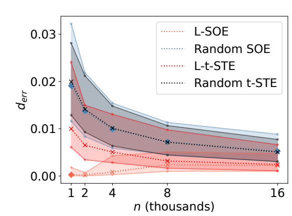
(a) Duration—Uniform Ball



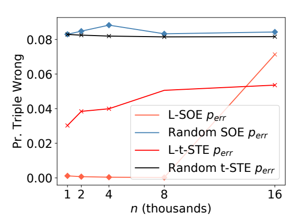
(b) Duration—GMM



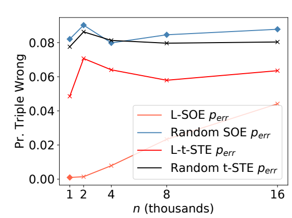
(c) Distance—Uniform Ball



(d) Distance—GMM



(e) Order—Uniform Ball

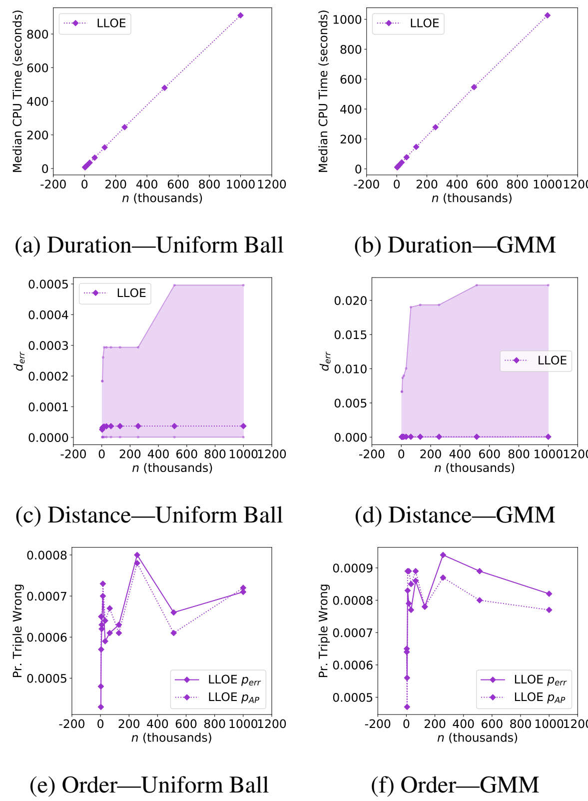


(f) Order—GMM

Figure 6: L-SOE Performance in $\mathbb{R}^{30}$.

B.2. LLOE Performance

Here we present supplemental empirical results for LLOE. Figures 7, 8, 9, 10, 11, and 12 show embeddings of our synthetic datasets at various dimensionalities for $m = 1,000$. We show in Figures 13 and 14 results in $d = 30$ for larger values of m. Finally, Figures **??**, **??**, **??**, **??**, 15, and 16 show results for our real datasets at various dimensionalities. These results are largely consistent with those presented in the main paper.

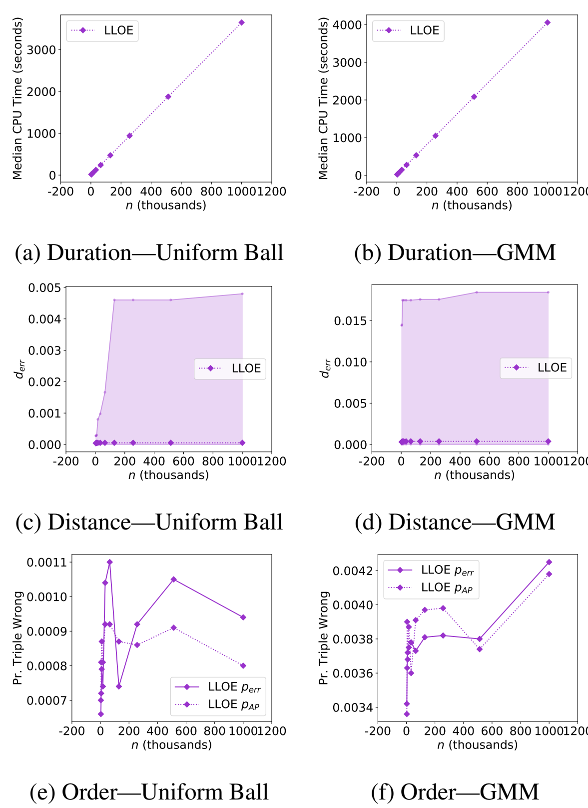


(a) Duration—Uniform Ball (b) Duration—GMM

(c) Distance—Uniform Ball (d) Distance—GMM

(e) Order—Uniform Ball (f) Order—GMM

Figure 7: LLOE in $\mathbb{R}^5$ with $m = 1,000$.

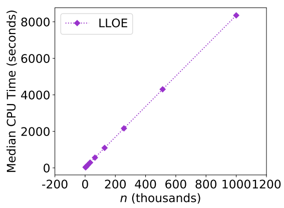


(a) Duration—Uniform Ball (b) Duration—GMM

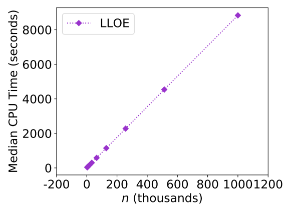
(c) Distance—Uniform Ball (d) Distance—GMM

(e) Order—Uniform Ball (f) Order—GMM

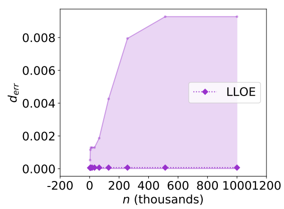
Figure 8: LLOE in $\mathbb{R}^{10}$ with $m = 1,000$.



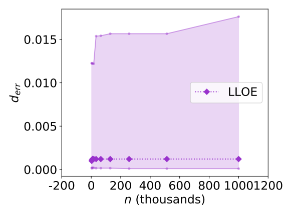
(a) Duration—Uniform Ball



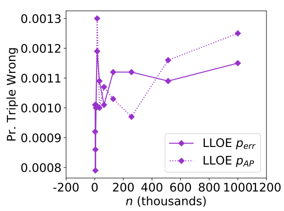
(b) Duration—GMM



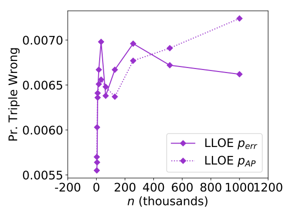
(c) Distance—Uniform Ball



(d) Distance—GMM

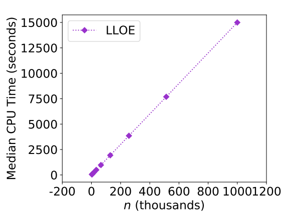


(e) Order—Uniform Ball

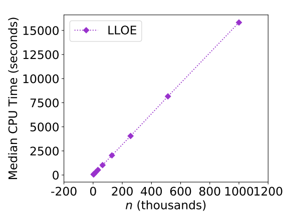


(f) Order—GMM

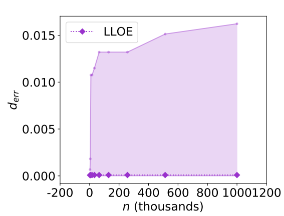
Figure 9: LLOE in $\mathbb{R}^{15}$ with $m = 1,000$.



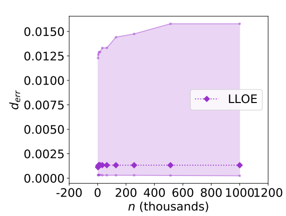
(a) Duration—Uniform Ball



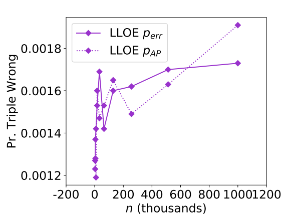
(b) Duration—GMM



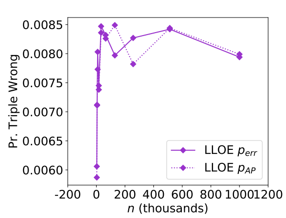
(c) Distance—Uniform Ball



(d) Distance—GMM

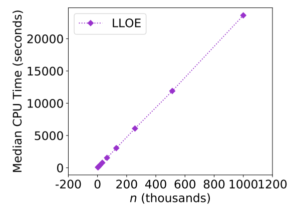


(e) Order—Uniform Ball

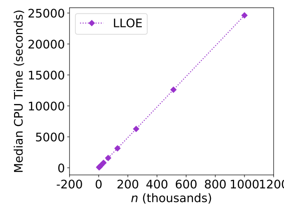


(f) Order—GMM

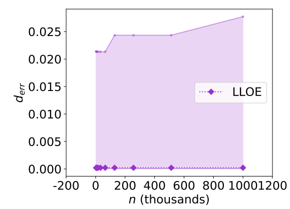
Figure 10: LLOE in $\mathbb{R}^{20}$ with $m = 1,000$.



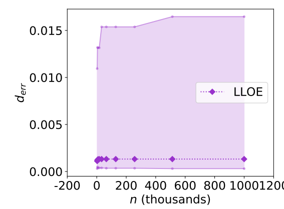
(a) Duration—Uniform Ball



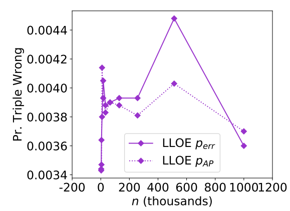
(b) Duration—GMM



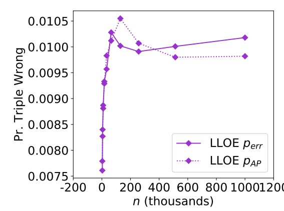
(c) Distance—Uniform Ball



(d) Distance—GMM

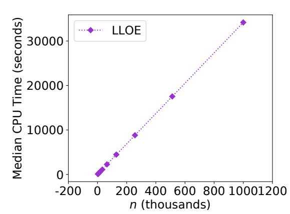


(e) Order—Uniform Ball

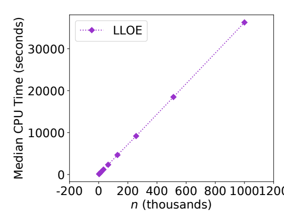


(f) Order—GMM

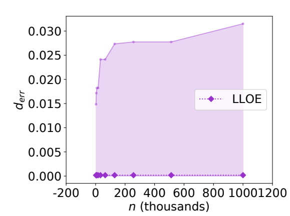
Figure 11: LLOE in $\mathbb{R}^{25}$ with $m = 1,000$.



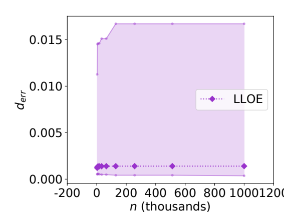
(a) Duration—Uniform Ball



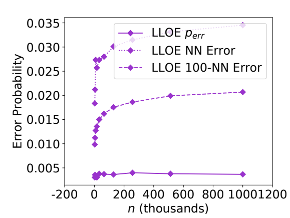
(b) Duration—GMM



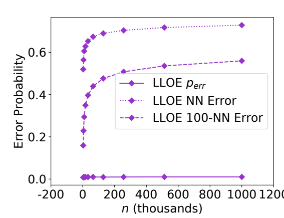
(c) Distance—Uniform Ball



(d) Distance—GMM

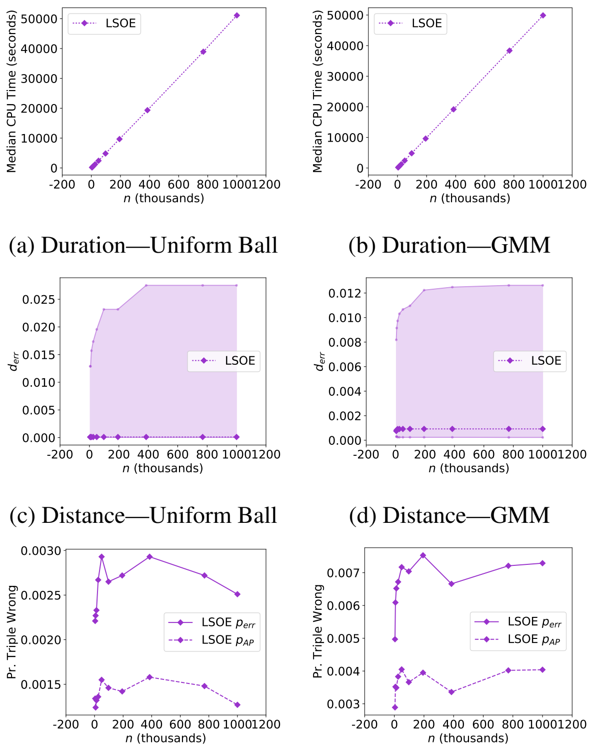


(e) Order—Uniform Ball



(f) Order—GMM

Figure 12: LLOE in $\mathbb{R}^{30}$ with $m = 1,000$.

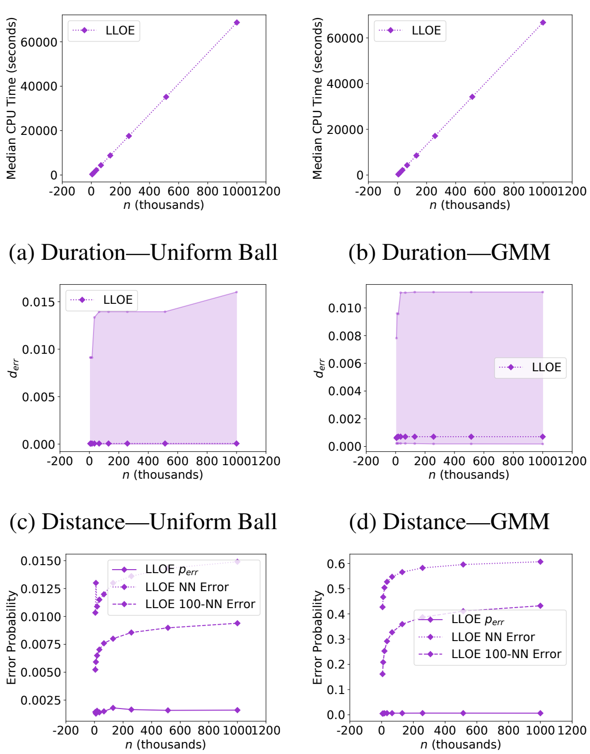


(a) Duration—Uniform Ball (b) Duration—GMM

(c) Distance—Uniform Ball (d) Distance—GMM

(e) Order—Uniform Ball (f) Order—GMM

Figure 13: LLOE in $\mathbb{R}^{30}$ with $m = 1,500$.

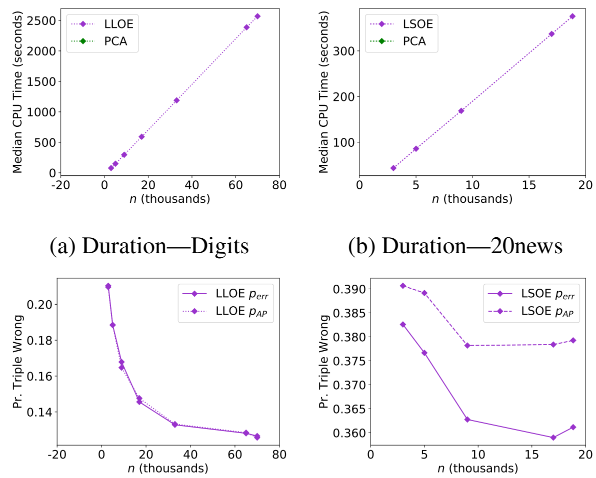


(a) Duration—Uniform Ball (b) Duration—GMM

(c) Distance—Uniform Ball (d) Distance—GMM

(e) Order—Uniform Ball (f) Order—GMM

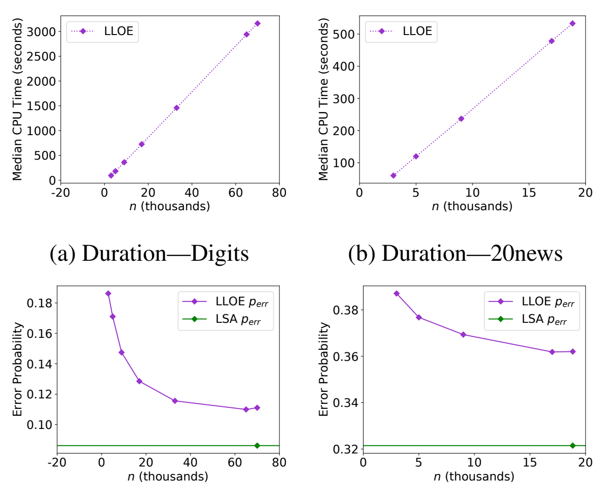
Figure 14: LLOE in $\mathbb{R}^{30}$ with $m = 2,000$.



(a) Duration—Digits (b) Duration—20news

(c) Order—Digits (d) Order—20news

Figure 15: LLOE on real data in $\mathbb{R}^{25}$ with $m = 1,000$.



(a) Duration—Digits (b) Duration—20news

(c) Order—Digits (d) Order—20news

Figure 16: LLOE on real data in $\mathbb{R}^{30}$ with $m = 1,000$.