

Poisson Disk Sampling on the Grassmannian: Applications in Subspace Optimization

Rushil Anirudh, Bhavya Kailkhura, Jayaraman J. Thiagarajan, Peer-Timo Bremer
Lawrence Livermore National Labs

{anirudh1,kailkhura1,jayaramanthil,bremer5}@llnl.gov *

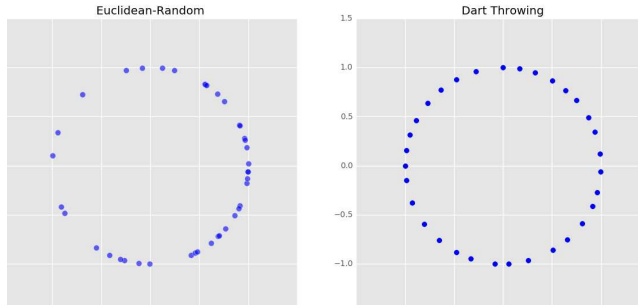
Abstract

To develop accurate inference algorithms on embedded manifolds such as the Grassmannian, we often employ several optimization tools and incorporate the characteristics of known manifolds as additional constraints. However, a direct analysis of the nature of functions on manifolds is rarely performed. In this paper, we propose an alternative approach to this inference by adopting a statistical pipeline that first generates an initial sampling of the manifold, and then performs subsequent analysis based on these samples. First, we introduce a better sampling technique based on dart throwing (called the Poisson disk sampling (PDS)) to effectively sample the Grassmannian. Next, using Grassmannian sparse coding, we demonstrate the improved coverage achieved by PDS. Finally, we develop a consensus approach, with Grassmann samples, to infer the optimal embeddings for linear dimensionality reduction, and show that the resulting solutions are nearly optimal.

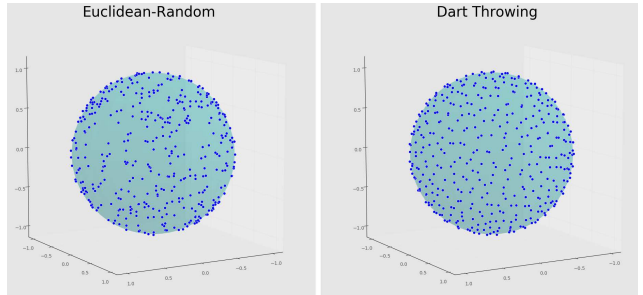
1. Introduction

A wide variety of applications in computer vision and machine learning rely on inference from data defined on embedded manifolds, such as, the Grassmannian manifold of subspaces. Examples range from dynamical system modeling [21] and activity recognition [16, 2] to subspace learning [10, 20] and domain adaptation [7]. In its most generic form the inference problem can be described as recovering or analyzing a multi-variate, smooth function $f : \mathcal{M} \mapsto \mathbb{R}^d$ defined on some manifold $\mathcal{M} \subset \mathbb{R}^n$. Conventional approaches in vision employ a variety of optimization tools to recover f , by incorporating the characteristics of known manifolds as additional constraints. While convexity of the quality metric on function recovery is often desired, the nature of that metric on \mathcal{M} is rarely analyzed [15]. Hence,

*This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. LLNL-CONF-730384-DRAFT



(a) Sampling on S^1 , with $N = 40$ samples.



(b) Sampling on S^2 , with $N = 400$ samples.

Figure 1. Motivating the need for better random sample design on embedded manifolds. Projecting random samples created in the Euclidean space results in poor coverage of the manifold. In comparison, our proposed approach produces highly uniform samples, thereby leading to improved analysis and optimization.

an alternative approach to this inference is to adopt a statistical pipeline that first generates an initial sampling of \mathcal{M} to create a baseline of knowledge, and then performs subsequent analysis based on these samples. Note that, this initial sampling is independent of the application-specific goals.

The goal of sampling is to produce the maximal amount of information with the minimal number of samples, since the function f can be quite complex and \mathcal{M} can be large. Naturally, the quality of initial sampling is crucial to the efficacy of the inference. Sampling low dimensional, Euclidean domains has been an active research area and a vari-

ety of techniques have demonstrated that carefully designed sampling patterns, such as, Blue noise [22, 12] or Poisson Disk [14, 5, 4, 6] can significantly outperform uniform random samples. However, extending these concepts to sampling embedded manifolds, for example, optimizing a loss function on the Grassmannian, has proven difficult.

A common strategy for sampling manifolds is to generate random samples in some subset of a euclidean domain and projecting them onto the manifold. However, since the projection, in general, is not volume preserving the resulting samples often cluster and thus need not sufficiently cover the domain. Consider examples of a circle \mathcal{S}^1 and a sphere \mathcal{S}^2 shown in Figure 1(a) and 1(b) respectively. On the left we sample a unit line or square and project samples to the manifolds by the method of normalization. This will map a point on the line or square to the corresponding point on the unit circle or sphere. Clearly, these samples do not uniformly cover the manifold and create undersampled regions. The challenge is that the projection will often introduce distortions that may result in regions with clustered points or in other cases undersampled regions. Since, in general, projections are somewhat opaque, it is often difficult to analyze let alone compensate for these distortions. Nevertheless, without a more formal analysis of manifold sampling, projection based techniques are used as is. A good sampling pattern is expected to have two main properties – (a) the sampling should be random, i.e. one wants to have an equal chance of finding features of interest, e.g., local minima in an optimization problem, anywhere in the domain \mathcal{M} ; and (b) the samples should uniformly cover all \mathcal{M} to minimize the distance to the optimal solutions.

In this paper, we introduce Poisson Disk Sampling (PDS) for the Grassmann manifold and show that just as for Euclidean domains a PDS sampling of the Grassmannian can significantly outperform uniform random samples. Motivated by its success in the Euclidean-domain, we propose to use the dart throwing technique [3] to generate approximate PDS samples on the Grassmannian. This is expected to improve the quality of samples, thus resulting in a higher chance of finding solutions to optimization problems. We illustrate this on the right of Figure 1(a) and 1(b), where we show that the corresponding samples obtained using the dart throwing algorithm result in a much better coverage. Note that, one might argue that first generating PDS samples in Euclidean-domain and then projecting it to Grassmann can be a good sampling strategy. However, this approach cannot perform better than the naive random sampling of Grassmann as it is well known that PDS properties are lost under projection and in practice it ends up performing poorly.

In order to study the impact of sample quality on optimization problems, we focus on the Grassmann manifold, which is key to a number of optimization problems in machine learning. In particular, we consider the problem

of linear subspace learning based on different embedding objectives: Principal Component Analysis (PCA), Locality Preserving Projections (LPP) and Linear Discriminant Analysis (LDA). In general, these embedding techniques can be posed as searching for a point on the Grassmannian that minimizes the corresponding embedding cost. We develop a novel consensus technique based on the generated samples to compute the low-dimensional embeddings, and demonstrate that the proposed PDS samples produce solutions close to the true optimum (in terms of the embedding cost), while outperforming uniform random samples.

Our main contributions are summarized as follows:

- We demonstrate the effectiveness of Poisson disk sampling (PDS) on the Grassmann manifold.
- We propose a dart throwing based algorithm to generate approximate PDS for the Grassmann manifolds in different dimensions – \mathcal{G}_n^k .
- Using Grassmannian sparse coding, we demonstrate the improved coverage achieved by PDS.
- We develop a consensus approach, with Grassmann samples, to infer the optimal embeddings for PCA, LPP and LDA objectives, and show that the resulting solutions are nearly optimal.

2. Sampling Euclidean Spaces

We briefly discuss the sampling problem in Euclidean and review the idea of Poisson Disk Sampling, which we generalize to the Grassmannian in Section 3.

Uniform sampling in low-dimensional Euclidean spaces has been of significant interest to both the computer graphics and uncertainty quantification researchers. In many scenarios, the Poisson-disk sampling (PDS) process is often considered to be an optimal choice. Formally, PDS distributes uniform random point samples on a domain of d -dimensional space based on a minimum distance criterion between point samples [6].

Definition A set of N point samples $\{\mathbf{S}_i\}$ in the Euclidean domain \mathcal{D} are Poisson disk samples, if $\mathbf{S} = \{\mathbf{S}_i \in \mathcal{D}; i = 1, \dots, N\}$ satisfy the following two conditions:

- $\forall \mathbf{S}_i \in \mathbf{S}, \forall A \subseteq \mathcal{D} : P(\mathbf{S}_i \in A) = \int_A d\mathbf{s}$
- $\forall \mathbf{S}_i, \mathbf{S}_j \in \mathbf{S} : \|\mathbf{S}_i - \mathbf{S}_j\| \geq r_{\min}$

where r_{\min} is the Poisson disk radius.

The first condition states that the probability of a uniformly distributed random sample $\mathbf{S}_i \in \mathbf{S}$ falling inside a subset A of \mathcal{D} is equal to the area of A . On the other hand, the second condition enforces the minimum distance constraint

between point sample pairs. Note, a Poisson sampling process enforces the first condition alone, in which case the number of samples that fall inside any subset $A \subseteq \mathcal{D}$ obeys a discrete Poisson distribution. Though easier to implement, Poisson sampling often produces distributions where the samples are grouped into clusters. Consequently, a sampling process that distributes random samples in an even manner across \mathcal{D} is preferred, so that no clustering is observed. The disk condition helps to eliminate the clustering behavior by preventing samples from being closer than r_{\min} .

The primary ingredient to design and analyze sample distributions is a suitable spatial statistic. One such measure in the Euclidean domain is the pair correlation function [19], which describes the joint probability of having points at two locations \mathbf{S}_i and \mathbf{S}_j respectively. For example, the pair correlation function for PDS can be defined as [13]:

$$G(r) = \begin{cases} 0 & \text{if } r < r_{\min} \\ 1 & \text{if } r \geq r_{\min}. \end{cases} \quad (1)$$

Generalizing the idea of Poisson disk sampling to embedded manifolds requires the definition of an equivalent spatial statistic, which does not currently exist. Hence, we propose to generate an approximate PDS distribution using dart throwing on the Grassmannian based on a valid geodesic distance metric. While dart throwing techniques in Euclidean domains have been well-studied [3], their behavior on the Grassmannian is unknown.

3. Generating Samples on the Grassmannian

3.1. The Grassmann manifold

The Grassmann manifold denoted by \mathcal{G}_n^k is defined as the space of all k dimensional subspaces in \mathbb{R}^n . The application of interest in this paper is linear dimensionality reduction schemes, to which the Grassmann manifold naturally lends itself as the subspaces used for low dimensional projection dimension can be represented as a point on the Grassmannian. In other words, for a set of points in a higher dimensional space $\mathbf{Z}_H \in \mathbb{R}^n$, $\mathbf{Z}_l = \mathbf{S}^T \mathbf{Z}_H$, where $\mathbf{S} \in \mathcal{G}_n^k$, $\mathbf{Z}_l \in \mathbb{R}^k$. For each subspace $\mathbf{S} \in \mathcal{G}_n^k$, that is represented by an orthogonal basis, there is a corresponding projection matrix $\mathbf{P} = \mathbf{S}\mathbf{S}^T \in \mathbb{R}^{n \times n}$. The set, \mathcal{P}_n of all such projection matrices is diffeomorphic to \mathcal{G}_n^k . The advantage of using this representation is that the Riemannian metric in \mathcal{P}_n , is just the inner product $\langle \mathbf{P}_1, \mathbf{P}_2 \rangle = \text{trace}(\mathbf{P}_1 \mathbf{P}_2^T)$, which is computed using the Frobenius norm between two $n \times n$ matrices of rank k (also referred to as the *chordal* or *extrinsic* distance in literature). The inverse mapping, $\Pi : \mathcal{P}_n \mapsto \mathcal{G}_n^k$, is given by $\Pi(\mathbf{P}) = \mathbf{U}\mathbf{U}^T$, where $\mathbf{P} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^T$ is the k -rank SVD of \mathbf{P} .

Algorithm 1 Dart Throwing on the Grassmann manifold

Require: Dimensions (n, k) , number of samples N and r_{\min} ,
 $\mathbf{S} = \emptyset$
1: **while** $|\mathbf{S}| \leq N$ **do**
2: Throw a Dart i :
 • Generate random matrix $\mathbf{Z}_i \in \mathbb{R}^{n \times k}$
 • Obtain the corresponding point $Q \in \mathcal{G}_{n,k}$ as the QR decomposition of \mathbf{Z}_i .
 • Assign $\mathbf{S}_i \leftarrow Q$
3: **if** $d_G(\mathbf{S}_i, \mathbf{S}_j) \leq r_{\min}, \forall \mathbf{S}_j \in \mathbf{S}$ **then**
4: Add sample \mathbf{S}_i to the point set \mathbf{S}
5: **end if**
6: **end while**

3.2. Dart throwing for Poisson Disk Sampling

One of the most straightforward ways of generating PDS in Euclidean domains is by using the *dart throwing* algorithm [3], which generates random candidate points over the domain, however, accepts each candidate *iff* it lies at a minimum distance r_{\min} from every previously accepted sample. A practical variant of the dart throwing algorithm is where the dart radius is gradually decreased as more samples are placed. The dart throwing algorithm is therefore inherently sequential and slow due to its quadratic complexity, but the quality of the results is nevertheless high.

Furthermore, an important property of any distribution generated with dart throwing is its progressive nature: Similar to low-discrepancy sequences, if we put all the generated points into a list of N elements, and draw only a portion of that list, the resulting point distribution also covers the entire plane (as opposed to sweeping it), but simply does it with a lower density. Motivated by these observations, we employ the dart throwing approach to generate poisson disk sample directly on Grassmann manifold, referred to hereon as PDS samples, while the baseline samples are referred to as random samples.

The algorithm described in 1 describes the procedure to perform dart throwing on the Grassmann manifold for PDS. Note that, for all experiments and analysis presented in this paper, we use the *chordal distance* to compare the points on the Grassmannian, defined as

$$d_G(\mathbf{S}_1, \mathbf{S}_2) = \frac{1}{\sqrt{2}} \|\mathbf{S}_1 \mathbf{S}_1^T - \mathbf{S}_2 \mathbf{S}_2^T\|_F. \quad (2)$$

The baseline for comparison is generating random samples, which is equivalent to throwing N darts with an $r_{\min} = 0$, i.e. accept all darts as good samples. The computational complexity of a naive implementation is $\mathcal{O}(N^2)$ for N darts. However, more efficient algorithms are available that can achieve the same in $\mathcal{O}(N \log N)$ complexity, with spatial data structures such as hashing grids or trees to reduce the cost of nearest-neighbor lookups [4].

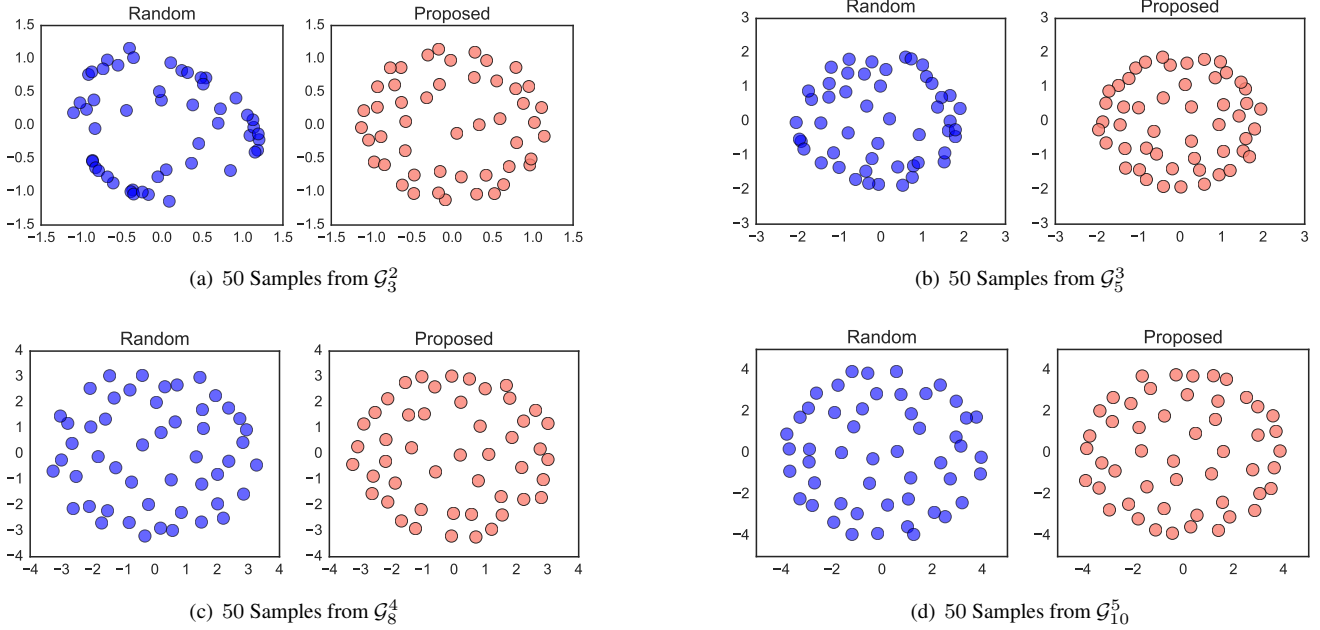


Figure 2. Visualizing the sample distributions, obtained with uniform random and dart throwing methods, using Multidimensional Scaling.

Sub-sampling is a closely related strategy that can potentially be applied here. The dart throwing algorithm 1, samples the space directly instead of over-sampling first, and then picking the best set of samples. This can be achieved by clustering-like approaches or modified exemplar selection such as in Manifold Precis [18] in the context of *diverse* sampling from a given dataset. Conceptually, this is similar to the *farthest point optimization* [17] techniques in the sampling literature. Though this approach is known to satisfy the r_{min} condition, in the Euclidean space, it tends to produce clustering patterns unlike the PDS.

4. Sample Quality Evaluation

Analyzing statistical performance of high-dimensional point sets on embedded manifolds is a challenging task. For example, there do not exist techniques to estimate spatial statistics, such as pair correlation function [19], on the Grassmann manifold. Therefore in this section, we develop proxy strategies to evaluate the quality of samples produced using the dart throwing algorithm. We first visualize the sample distribution using Multi-Dimensional Scaling, and then quantitatively measure the representative power of the samples, as a proxy for coverage on the manifold.

Visualizing sample distribution in 2D: We employ Multi-dimensional scaling (MDS) to project Grassmann samples into 2D for visual comparisons. MDS is designed to represent high dimensional data in a low dimensional space while preserving the similarities between data points.

We apply MDS on a pairwise distance matrix for the dart throwing and random sampling, in order to observe their difference in quality (see figure 2). The samples obtained from dart throwing are observed to cover the embedded space uniformly, as opposed to the random samples which leave holes in the space. Further, random sampling contains the undesired grouping behavior of samples, illustrating the benefit of imposing the minimum distance criterion.

Quantifying coverage on the manifold: We evaluate the representative power of the generated samples, which reflects the coverage achieved by these samples. Specifically, we propose to oversample the Grassmannian, sparsely encode each of them by treating the set of generated samples (Section 3.2) as the dictionary, and analyze the statistics of the approximation error. If a sampling strategy produces a good set of samples in terms of covering the entire manifold, then, those samples should ideally result in high-fidelity approximation of the oversampled set.

The sparse encoding problem is as follows [8]: Given a subspace \mathbf{X} and a dictionary $\mathcal{S} = \{\mathbf{S}_i\}_{i=1}^N$ with N pre-designed samples, approximate \mathbf{X} using a linear combination of a small number of elements in \mathcal{S} . In order to perform sparse coding on the Grassmann manifold, we adopt the formulation proposed in [9]. This can be formally expressed as

$$\min_{\alpha} \|\mathbf{X}\mathbf{X}^T - \sum_{i=1}^N [\alpha]_i \mathbf{S}_i \mathbf{S}_i^T\|_F + \lambda \|\alpha\|_1, \quad (3)$$

where $\{\mathbf{S}_i\}_{i=1}^N$ are points on Grassmann obtained using the

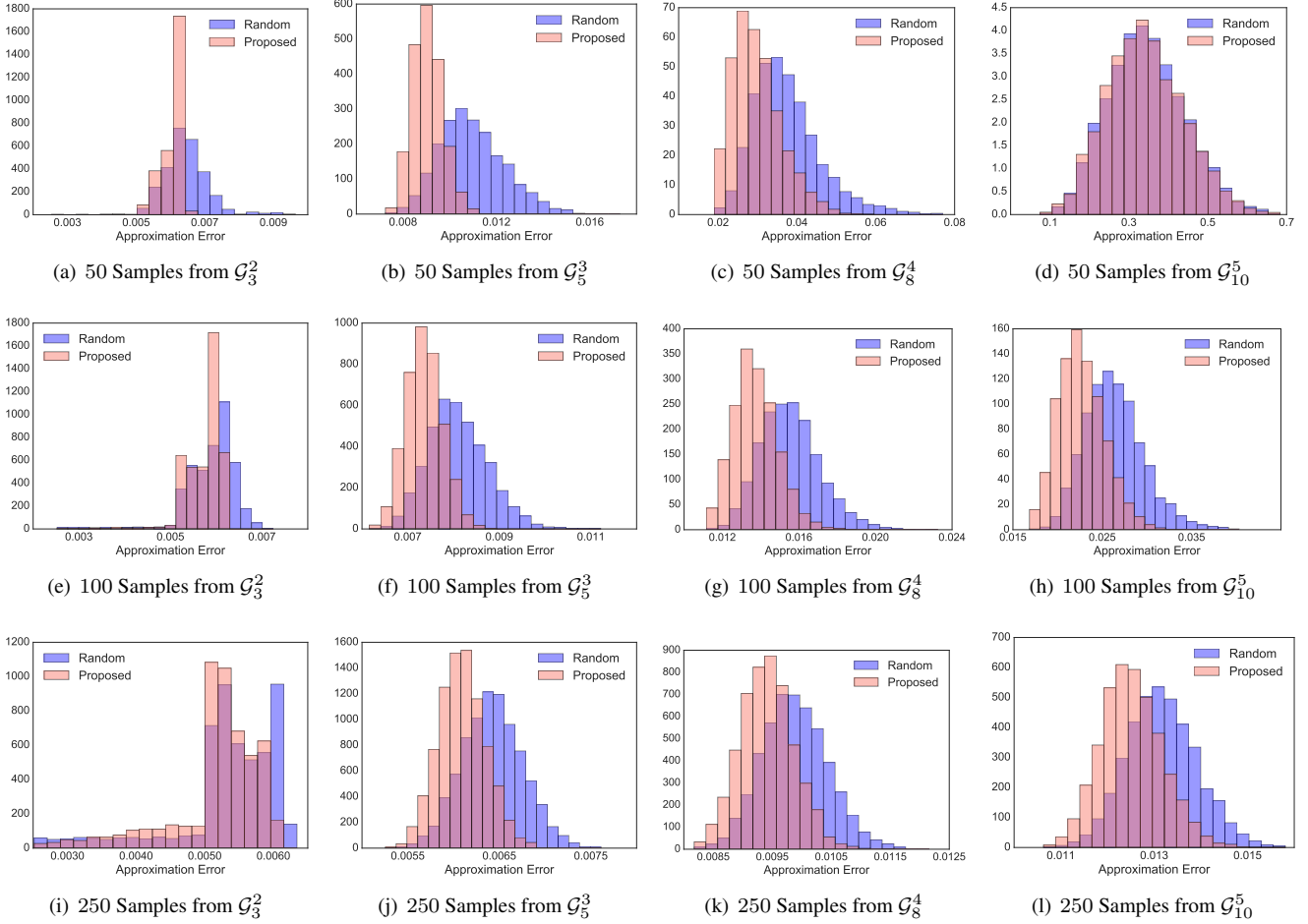


Figure 3. Coverage analysis of Grassmann samples - We adopt a sparse coding based empirical study to quantify the coverage achieved by different sampling techniques. In each case, we show the histogram of approximation errors, measured as the chordal distance, for an oversampled set of samples from the Grassmannian.

Table 1. Formulating common linear dimensionality reduction techniques using the unified graph embedding framework in [23]

Method	Intrinsic Graph	Penalty Graph
Principal Component Analysis	$W_{ij} = 1/T, i \neq j$	$\mathbf{B} = \mathbf{I}$
Linear Discriminant Analysis	$W_{ij} = \delta_{c_i, c_j} / n_{c_i}, \delta_{c_i, c_j} = 1 \text{ if } c_i = c_j$	$\mathbf{B} = \mathbf{I} - 1/Nee^T$
Locality Preserving Projection	$W_{ij} = \exp(-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2), \text{ if } i \in \mathcal{N}(j) \text{ or } j \in \mathcal{N}(i)$	$\mathbf{B} = \mathbf{D}$

sampling technique. Note that, finding the sparse codes with fixed S can be solved efficiently as the optimization problem is convex in α .

To evaluate the coverage achieved by the sampling techniques, we generate 50, 100 and 250 samples on the following Grassmannian manifolds: $\{\mathcal{G}_3^2, \mathcal{G}_5^3, \mathcal{G}_8^4, \mathcal{G}_{10}^5\}$ using both random and proposed dart throwing algorithms. In each case, we obtained 50 independent realizations and using each realization as the dictionary, we evaluated the sparse approximation errors for 200 randomly drawn samples. Note, we used the chordal distance to measure the difference between the true subspace and approximated returned by sparse coding. The histograms of the approxi-

mation errors from both the approaches, for all cases, are shown in Figure 3. As expected, with the increase in sample size, error decreases for both random and dart throwing sampling. Further, for the same sample size, the proposed dart throwing technique outperforms uniform random samples, thus evidencing improved coverage on the manifold. Interestingly, in the case of \mathcal{G}_{10}^5 , both approaches perform poorly, as number of samples need for a reasonable approximation grows exponentially and the curse of dimensionality starts affecting the performance. Furthermore, if sample sizes are too small both the techniques find it challenging to approximate the oversampled set from the Grassmannian and the performance difference between these sampling pat-

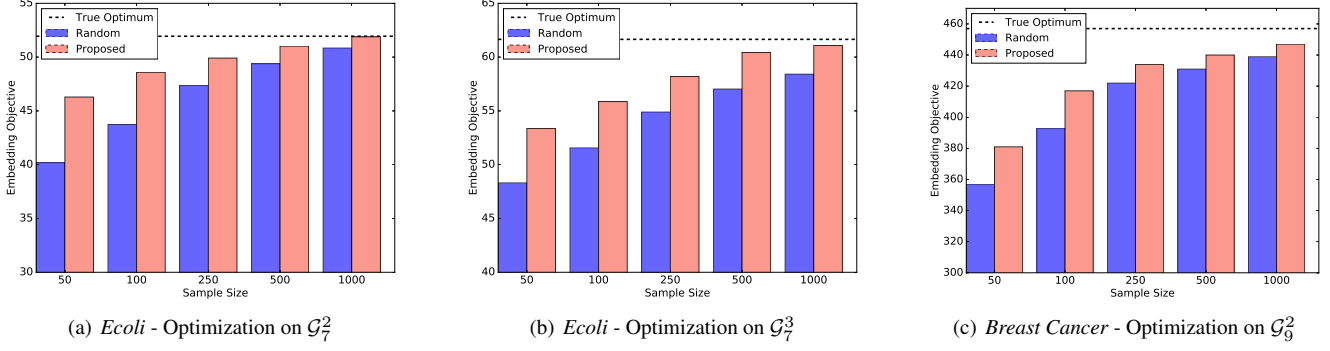


Figure 4. Principal Component Analysis cost optimization using consensus on the Grassmannian (higher is better).

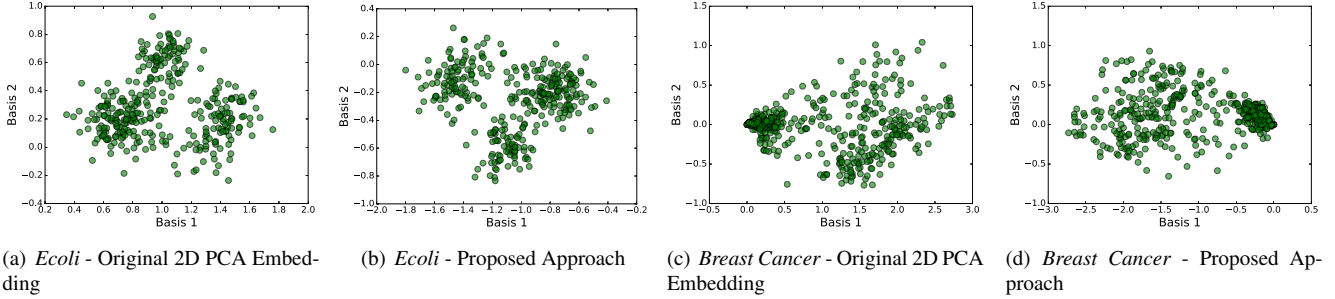


Figure 5. Comparison of the consensus embedding obtained using the PCA cost on the Grassmannian, to the true solution obtained using generalized eigenvalue decomposition.

terns becomes negligible. However, given reasonable sample sizes, the PDS samples perform consistently better compared to random samples in all dimensions.

5. Subspace Optimization using Grassmann Samples

In this section, we consider the problem of linear subspace learning to perform dimensionality reduction. In practice, there exists a wide variety of quality metrics to optimize for effective linear embeddings from high-dimensional datasets, e.g. class separation in linear discriminant analysis. More generally, these problems can be viewed under a unified graph embedding formulation and solved using common optimization techniques such as the generalized eigenvalue decomposition [23]. In contrast, we propose to infer optimal subspaces, for a given quality metric, using only the pre-designed set of samples in lieu of the actual optimization. Using empirical studies on real datasets, we investigate the impact of sample quality on the efficacy of the inferred solutions.

5.1. Subspace Learning as Graph Embedding

Dimensionality reduction techniques have become an integral part of several supervised and unsupervised learning systems in computer vision and pattern recognition. Among them, linear embedding techniques such as the Principal

Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are popular because of their relative simplicity and effectiveness. We adopt the general graph embedding framework proposed in [23] since it provides a unified view for formulating linear dimensionality reduction techniques.

In this approach, we represent each vertex of a graph as a low-dimensional vector that preserves relationships between the vertex pairs, where the relationship is measured by a similarity metric that characterizes certain statistical or geometric properties of the data set. Let $G = \{\mathbf{X}, \mathbf{W}\}$ denote a undirected graph (referred as the intrinsic graph), where the matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_T]$, $\mathbf{x}_i \in \mathbb{R}^n$ represents the data samples in n dimensions, while the matrix $\mathbf{W} \in \mathbb{R}^{T \times T}$ is the similarity matrix between all pairs of samples. The diagonal degree matrix \mathbf{D} and the Laplacian of a graph G can be defined as:

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \text{ where } D_{ii} = \sum_{j \neq i} W_{ij} \quad (4)$$

Denoting the corresponding low-dimensional representations for the set of data samples by $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_T]$, $\mathbf{y}_i \in \mathbb{R}^k$, where $k \ll n$, the problem of graph embedding can be posed as follows:

$$\mathbf{Y}^* = \arg \min_{tr(\mathbf{Y}\mathbf{B}\mathbf{Y}^T)=d} tr(\mathbf{Y}\mathbf{L}\mathbf{Y}^T). \quad (5)$$

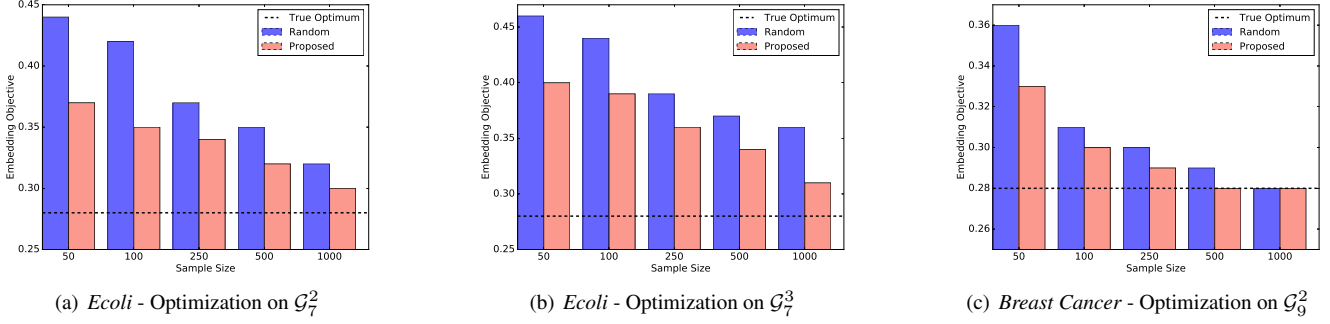


Figure 6. Linear Discriminant Analysis cost optimization using consensus on the Grassmannian (lower is better).

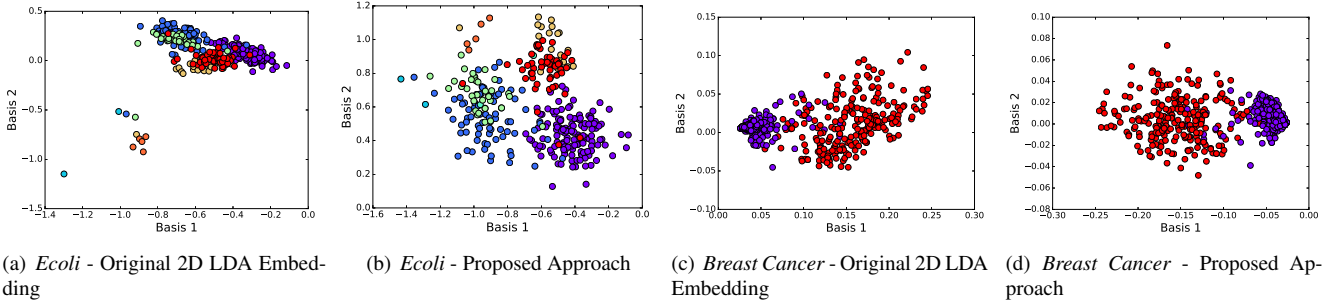


Figure 7. Comparison of the consensus embedding obtained using the LDA cost on the Grassmannian, to the true solution obtained using generalized eigenvalue decomposition.

Here, $tr(\cdot)$ computes the trace of a matrix, and the matrix \mathbf{B} corresponds to the Laplacian of an optional penalty graph, typically used to regularize the learning. In case of linear embeddings, we can rewrite (5) as

$$\mathbf{Y}^* = \arg \min_{tr(\mathbf{V}^T \mathbf{X} \mathbf{B} \mathbf{X}^T \mathbf{V}) = d} tr(\mathbf{V}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{V}), \quad (6)$$

where $\mathbf{V} \in \mathbf{R}^{n \times k}$ is the desired projection. The solution to (6) can be obtained using generalized eigenvalue decomposition.

We consider the optimization of three linear embedding techniques: PCA, LDA and locality preserving projections (LPP), that preserves local relationships within the dataset and uncovers its essential manifold structure. Table 1 lists the appropriate construction of the intrinsic graph and penalty graph Laplacians \mathbf{L} and \mathbf{B} for the three methods.

5.2. Proposed Consensus Approach

To optimize the linear projection for a desired quality metric, we evaluate the metric for each of the samples in \mathcal{S} . For example, in the case of PCA, we measure the total variance captured by a linear projection \mathbf{S}_i as $\mathbf{S}_i^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{S}_i$. Though the chance of finding the optimal solution in \mathcal{S} is low, using the observations in Section 4, we hypothesize that it can be approximated as a linear combination of the samples. In other words, we construct the optimal embedding as a weighted consensus of the samples in \mathcal{S} .

We employ a simple strategy which assigns larger weights to the samples with higher embedding quality and exponentially smaller weights to the rest of the samples. We optimize for the weighted consensus as:

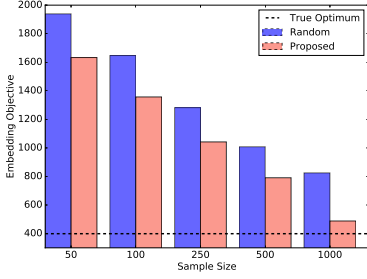
$$\mathbf{V}^* = \arg \min_{\mathbf{V}^T \mathbf{V} = \mathbf{I}} tr \left(\mathbf{V}^T \sum_{i=1}^N (\alpha_i \mathbf{I} - \alpha_i \mathbf{S}_i \mathbf{S}_i^T) \mathbf{V} \right) \quad (7)$$

where α_i are the positive weights. \mathbf{V} can be evaluated as the d eigen vectors corresponding to the smallest eigen values of $\sum_{i=1}^N (\alpha_i \mathbf{I} - \alpha_i \mathbf{S}_i \mathbf{S}_i^T)$.

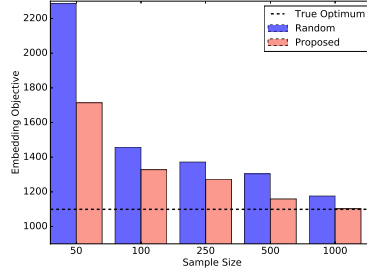
5.3. Results

In this section, we evaluate the effectiveness of Poisson disk samples in approximating the optimal embeddings for PCA, LDA and LPP quality metrics. For the results reported in this section, we used two datasets from the UCI repository: (a) *Ecoli* dataset (7 dimensions) and (b) *Breast Cancer* dataset (9 dimensions). For the *ecoli* dataset, we inferred subspaces of 2 and 3 dimensions, while for the *breast cancer* dataset we considered $k = 2$.

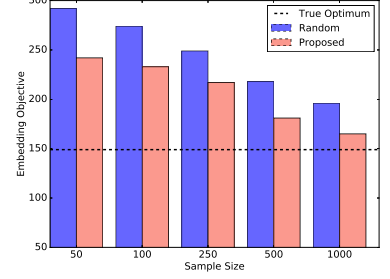
Figure 4 plots the embedding cost of the consensus subspace, obtained using both random and Poisson disk samples, for the case of PCA. Note that, unlike LDA and LPP, PCA attempts to maximize the embedding cost (variance) and hence larger the cost better is the inferred subspace. In addition, we show the embedding cost of the true PCA solution computed using generalized eigenvalue decomposition.



(a) *Ecoli* - Optimization on G_7^2



(b) *Ecoli* - Optimization on G_7^3



(c) *Breast Cancer* - Optimization on G_9^2

Figure 8. *Locality Preserving Projections* cost optimization using consensus on the Grassmannian (lower is better).

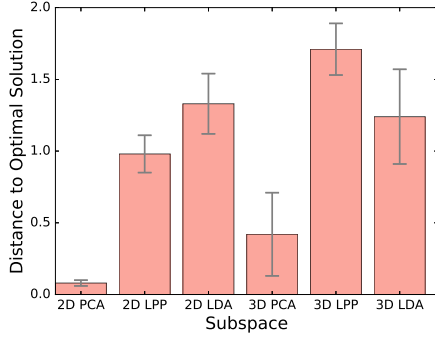


Figure 9. Grassmann distance between the optimal solution and the linear subspace inferred using the consensus from PDS samples for the *Ecoli* dataset.

The first striking observation is that with only around 500 Grassmann samples, we obtain solutions that are very close to the true optimum, in all cases. Second, samples generated using the dart throwing schemes significantly outperform random samples. Note that, the results reported in the figure were obtained as the average of 100 independent trials with different sampling realizations. Similar results are observed for the cases of LDA and LPP quality metrics as well, as observed in Figures 6 and 8 respectively.

Figure 5 shows the PCA embeddings when $k = 2$, for both datasets, obtained using PCA and the proposed consensus strategy. Similarly, Figure 7 shows the 2D embeddings that optimize the LDA quality metric, i.e. class separation. Observe that the solutions obtained using the Grassmann samples are different than the true solution and yet they are very effective in preserving the desired relationships. For example, the LDA embedding in Figure 7(b), the class separation is similar to the true embedding, while the crowding effect, commonly observed in linear embedding techniques, is clearly avoided. Further investigation of the closeness of the inferred solutions to their true counterparts in 9, reveals that in many cases there is evidence for other nearly optimal solutions that are far from the true optimum, and in some cases they could be more insightful. This is in agreement with the observations reported in [15], where the

authors identify multiple modes in the optimization surface of quality metrics on the Grassmannian.

6. Discussion and Future Work

We introduced parameter inference approach on embedded manifolds that first performed an initial sampling of the manifold, and then subsequent analysis based on these samples. To generate samples with better coverage properties in embedded domains, we proposed the use of Poisson disk sampling. In addition to demonstrating improved coverage, we showed that the samples are highly effective in approximating the optimal solution for subspace learning problems. These results indicate that high-quality samples can serve as anchor points to search through the Grassmannian using conventional optimization strategies including gradient based methods [1] and random optimization [11].

A few observations warrant further study (a) The proposed dart throwing algorithm is expected to result in approximate PDS samples, the superior quality of which we have demonstrated with empirical studies. However, a more in-depth analysis is required to prove claim (1) from the definition of PDS, with regard to computing the hyper-volume of a subset on the manifold. A proxy could be to count the number of points that lie inside a unit ball of radius r ; (b) The curse of dimensionality ensures that in higher dimensions, even PDS samples tend to resemble random, in that they both cover the space very poorly. As a result an exponentially large number of samples are required to obtain a *good* sampling; and (c) As Figure 9 indicates, there are solutions to the subspace learning problem, which are far from the true solution in terms of the Grassmann distance. However, we realized that though those solutions are slightly sub-optimal in terms of the embedding cost, they might nevertheless be useful embeddings. This key observation demands further research to understand the topological characteristics of the loss function for designing better optimization techniques on the Grassmannian, and potentially other matrix manifolds.

References

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2007.
- [2] R. Anirudh, P. Turaga, J. Su, and A. Srivastava. Elastic functional coding of Riemannian trajectories. *IEEE transactions on pattern analysis and machine intelligence*, 2016.
- [3] R. L. Cook. Stochastic Sampling in Computer Graphics. *ACM Trans. Graph.*, 5(1):51–72, Jan. 1986.
- [4] D. Dunbar and G. Humphreys. A Spatial Data Structure for Fast Poisson-disk Sample Generation. *ACM Trans. Graph.*, 25(3):503–508, July 2006.
- [5] M. S. Ebeida, S. A. Mitchell, A. Patney, A. A. Davidson, and J. D. Owens. A Simple Algorithm for Maximal Poisson-Disk Sampling in High Dimensions. *Computer Graphics Forum*, 31(2pt4):785–794, 2012.
- [6] M. N. Gamito and S. C. Maddock. Accurate multidimensional poisson-disk sampling. *ACM Trans. Graph.*, 29(1):8:1–8:19, Dec. 2009.
- [7] R. Gopalan, R. Li, and R. Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 999–1006. IEEE, 2011.
- [8] M. Harandi, C. Sanderson, C. Shen, and B. Lovell. Dictionary learning and sparse coding on grassmann manifolds: An extrinsic solution. In *2013 IEEE International Conference on Computer Vision*, pages 3120–3127, Dec 2013.
- [9] M. T. Harandi, C. Sanderson, C. Shen, and B. C. Lovell. Dictionary learning and sparse coding on grassmann manifolds: An extrinsic solution. In *ICCV*, pages 3120–3127, 2013.
- [10] S. Hauberg, A. Feragen, and M. J. Black. Grassmann averages for scalable robust pca. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3810–3817, 2014.
- [11] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *J. of Global Optimization*, 13(4):455–492, Dec. 1998.
- [12] B. Kailkhura, J. J. Thiagarajan, P.-T. Bremer, and P. K. Varshney. Stair blue noise sampling. *ACM Trans. Graph.*, 35(6):248:1–248:10, Nov. 2016.
- [13] B. Kailkhura, J. J. Thiagarajan, P. T. Bremer, and P. K. Varshney. Theoretical guarantees for poisson disk sampling using pair correlation function. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2589–2593, March 2016.
- [14] A. Lagae and P. Dutre. A Comparison of Methods for Generating Poisson Disk Distributions. *Computer Graphics Forum*, 27(1):114–129, 2008.
- [15] S. Liu, P.-T. Bremer, J. Jayaraman, B. Wang, B. Summa, and V. Pascucci. The grassmannian atlas: A general framework for exploring linear projections of high-dimensional data. In *Computer Graphics Forum*, volume 35, pages 1–10. Wiley Online Library, 2016.
- [16] Y. M. Lui. Human gesture recognition on product manifolds. *Journal of Machine Learning Research*, 13(Nov):3297–3321, 2012.
- [17] T. Schlömer, D. Heck, and O. Deussen. Farthest-point Optimized Point Sets with Maximized Minimum Distance. In *Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics*, HPG ’11, pages 135–142, New York, NY, USA, 2011. ACM.
- [18] N. Shroff, P. Turaga, and R. Chellappa. Manifold precis: An annealing technique for diverse sampling of manifolds. In *Advances in Neural Information Processing Systems*, pages 154–162, 2011.
- [19] D. Stoyan, U. Bertram, and H. Wendrock. Estimation variances for estimators of product densities and pair correlation functions of planar point processes. *Annals of the Institute of Statistical Mathematics*, 45(2):211–221, 1993.
- [20] J. J. Thiagarajan and K. N. Ramamurthy. Subspace learning using consensus on the grassmannian manifold. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2031–2035, April 2015.
- [21] P. K. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa. Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition. 33(11):2273–2286, 2011.
- [22] D.-M. Yan, J.-W. Guo, B. Wang, X.-P. Zhang, and P. Wonka. A survey of blue-noise sampling and its applications. *Journal of Computer Science and Technology*, 30(3):439–452, 2015.
- [23] S. Yan, D. Xu, B. Zhang, H. j. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, Jan 2007.