Industrial Statistics and Manifold Data

E. del Castillo and Xueqi Zhao
Engineering Statistics and Machine Learning Laboratory
Dept. of Industrial & Manufacturing Engineering and Dept. of Statistics
The Pennsylvania State University, University Park, PA 16802

July 4, 2019

Abstract

Complex and not only big data exist everywhere in industry and how to control and optimize systems based on these data types is an important aspect of modern Quality Engineering. One fundamental type of complexity occurs when data lies on a lower dimensional, curved subspace or manifold. We review a new approach for statistical process monitoring of point cloud, mesh and voxel data based on intrinsic geometrical features of the 2-D manifold (surfaces) of scanned manufactured parts. Monitoring intrinsic properties avoids computationally expensive registration pre-processing of the data sets. We also present a review of recent approaches for analyzing and designing experiments where either the response or the covariates lie on manifolds.

1 Introduction: manifold data in industry

Quality Engineers working in industry must be acquainted with new ways to analyze not only larger datasets, but frequently, more complex datasets. In this paper we address the case the complexity of a dataset occurs because the data follow, apart of measurement error, a lower dimensional manifold, which can be understood informally as a curved space which, when looked over a small domain around any point in the space, it resembles euclidean space (as a typical example, consider the surface of the Earth or a spheroid in $\mathbb{R}^3$). We will focus on two of the main subfields of Industrial Statistics: Statistical Process Control (i.e., monitoring, hereafter referred to as SPC) and Experimental Design and Analysis. We are aware that the subject

of manifold data in industrial statistics is a wider theme than these subtopics and we do not aim at a comprehensive review; the subfields we concentrate are admittedly limited to our personal research interests. In the machine learning literature, manifold learning is a topic that has received considerable attention, and similarly, "shape analysis" and "*statistical* shape analysis" of 3D objects, whose surface is a 2-dimensional manifold, have been intensively studied in the computer vision and statistics fields, respectively, over the last two decades. We will discuss intersections between Industrial Statistics and these fields when appropriate. We concentrate in manufacturing industry, where discrete parts are produced and measured.

The wider availability of 3-dimensional (3D) non-contact (range) scanners in industry has produced large datasets of surface measurements, a canonical case of manifold data, where the manifold is the surface of a manufactured part and resides, or is embedded in, 3D euclidean space. The manifold (object of the surface) is *intrinsically* two dimensional, since, just as on the surface of the earth, it suffices to give two coordinates to locate a point on a surface. In its most unprocessed form, 3D scanner data have the form of *Point Clouds*, i.e., datasets where the (x_i, y_i, z_i) coordinates of hundreds of thousands of points $\mathbf{x_i}$ on the surface of a part are easily recorded (these could be unstructured or structured clouds, where the surface points follow a grid pattern). The scanner is not perfect, and what one measures is the true surface plus measurement error in all 3 coordinates. In addition to Point Cloud data, it is frequent that 3D scanners generate *mesh data*, and in particular, *triangulations*, where the measured points on the surface of a part correspond to vertices linked to nearby neighbors, in a network of triangles that defines the surface of an object. Most 3D scanners have built-in algorithms that directly generate a triangulation of the scanned data, useful for visualization purposes. Such triangulations are also common in Computer Aided Design (CAD) and in additive manufacturing (AM). Also prompted by AM needs, *voxel datasets* (three dimensional pixel data) are increasingly more frequent, as there is no better way to inspect internal features of a 3-D printed part than to use x-ray tomography Colosimo *et al.* (2018).

In this paper, we first discuss a recent approach for Statistical Process Monitoring of surface data based on "intrinsic" quantities of the surfaces of scanned parts, that is, quantities which depend only on the surface coordinates but not on distances across the 3D euclidean space where the part exists. Monitoring intrinsic properties avoids computationally expensive registration pre-processing of the point clouds. We propose to use properties of a 3D object that are *invariant* to rigid transformations,

i.e., rotations and translations so they are not affected by location or pose. Our review of recent work will make contact with Computer Graphics literature that uses a particular surface invariant, the Laplacian (or Laplace-Beltrami operator) of the surface. We next review experimental design and analysis for manifold data. We consider two cases, when the covariates ("X" data) lie on a low dimensional manifold embedded in a high dimensional euclidean space, and when the response data ("Y" data) lies on a manifold of high dimension.

2 Statistical Process Monitoring of Manifold (Surface) data

Consider a manufacturing process of discrete parts and the sequence of surfaces that are measured about which we wish to perform SPC (as mentioned below, the approach also applies to other complex data types, including voxel data, so SPC on internal features and not only of surface data is also possible). For surface data we assume a non-contact sensor acquires a large (hundreds of thousands) points per part, assuming points follow no particular pattern, with the number of points per part not necessarily equal, and not corresponding with each other. We could first think in applying techniques from the well established field of Statistical Shape Analysis (SSA), whose aim is statistical inference on objects in 2 and 3-dimensional euclidean space from data in the form of an $m \times d$ configuration matrix $\boldsymbol{X}$, where m is the number of points and $d = 2$ or 3 depending on whether the object is planar or 3-dimensional (see Kendall (1984). For an introduction to SSA, see Dryden & Mardia (2016) and the accompanying `shapes` R package, Dryden (2018). We further discuss SSA in section 3, with reference to analysis of experiments whose response is the shape of a manufactured object.)

A shape is classically defined as the information contained in $\boldsymbol{X}$ once we discount or "filter out" the effects of similarity transformations (translations, rotations, and changes in scale or dilatations, usually excluding reflections). This is a useful notion: whichever way we define the shape of an object, it should remain unchanged, or be *invariant* with respect to changes of location, orientation or changes of scale of the object (for process monitoring of manufactured parts, invariance to rigid transformations or isometries –rotations and translations, but not changes in scale– is perhaps more convenient. Two objects are then said to have the same *shape and size* if they

are congruent). SSA methods were initially applied in archeology and especially in paleontology and biology for classification of specimens in different species, testing whether two specimens have the same shape or not, or for the analysis of the main modes of variability in the shapes within a species. To perform either task, the set of points in $N \geq 2$ configuration matrices from different objects must first be superimposed or *registered*, i.e., the point clouds are put into the same location, orientation and scale via the *Generalized Procrustes Algorithm* (GPA, Dryden & Mardia (2016)).

The GPA registers or superimposes all the n objects by finding scaling factors $\beta_i \in \mathbb{R}$, rotation matrices $\boldsymbol{\Gamma}_i$ and d dimensional translation vectors $\boldsymbol{\gamma}_i, i = 1, ..., N$, such that they minimize the sum of squared "full procrustes" distances between all pairs of configuration matrices ($d_F(\boldsymbol{X}_i, \boldsymbol{X}_j)$).

It is possible in principle to use GPA and the full procrustes distances $d_F(\boldsymbol{X}_i, \boldsymbol{X}_j)$ for SPC purposes. GPA can be used to register N in-control parts from which an in-control mean $\hat{\boldsymbol{\mu}}$ can be estimated from the centroid of the m registered configuration matrices representing the point clouds. GPA is implemented in function `procGPA` and the full procrustes distances $d_F(\boldsymbol{X}_i, \boldsymbol{X}_j)$ are computed in function `procdist` in the `shapes` R package Dryden (2018). For Matlab code, see Del Castillo & Colosimo (2011). Then, the full procrustes distance between the measured configuration of each new part $\boldsymbol{X}_i$ and $\hat{\boldsymbol{\mu}}$, $d_F(\boldsymbol{X}_i, \hat{\boldsymbol{\mu}})$ could be used for on-line monitoring purposes. Under the model $\boldsymbol{X}_i = \beta_i(\boldsymbol{\mu} + \boldsymbol{E}_i)\boldsymbol{\Gamma}_i + \mathbf{1}_m\boldsymbol{\gamma}_i'$, $i = 1, ..., N$, with $\text{vec}(\boldsymbol{E}_i) = N(\mathbf{0}, \sigma^2\boldsymbol{I})$ (isotropic normal errors) with σ not large, it is known that the test statistic for $H_0 : [\boldsymbol{\mu}_1] = [\boldsymbol{\mu}_2]$ vs. a two sided alternative is an F statistic (see Goodall (1991) and Dryden & Mardia (2016)). The assumptions of normality and isotropy can be relaxed and either permutation or bootstrapped tests can be performed (see function `testmeanshape` in the R package `shapes`).

Although GPA and the test for mean shape differences is not particularly expensive computationally, it requires:

1. *one-to-one corresponding* points between points in each object;

2. An equal number of points in each point point cloud

These are requisites frequently true in applications in Biology and Archeology, where they were developed and where the points on the objects under study are usually manually selected "landmarks" of scientific interest. However, neither of these requisites are true in manufacturing data obtained from non-contact sensors: first, the

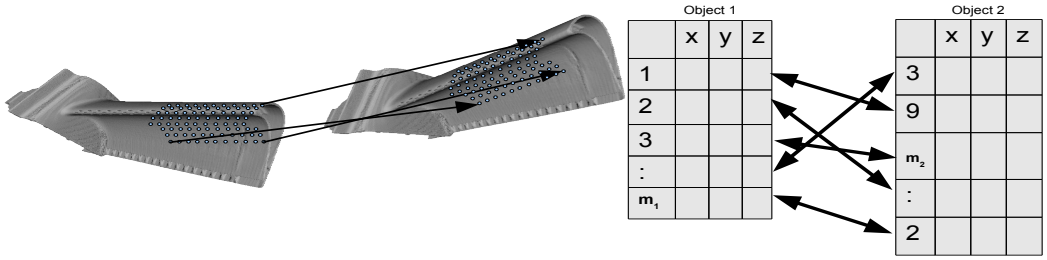

Figure 1: Diagrammatic representation of the traditional registration (matching) problem for two 3D point clouds of different size. Data points are either acquired by a non-contact sensor or could be acquired across the volume where the 3D object lies via computed tomography (CT), resulting in voxel data (see figure 2 for an example). Left: an instance of 2 turbine blades (from GATech (2019)) with different poses and with some of their point correspondences shown. Right: a sketch of the resulting configuration matrices with hypothetical correspondences (given by matrix $\boldsymbol{M}$) after applying the ICP algorithm.

number of points acquired by a scanner varies from part to part, even if parts are almost identical and almost identically posed with respect to the scanner, and second, the labels of each point on each configuration matrix do not correspond to the labels in other matrices. If surface measurements are obtained by a touch probe such as a coordinate measurement machine (CMM), then the assumptions of statistical shape analysis can be true. For this reason we return to SSA below in the designed experiments section (where careful measurements can be performed).

The problem of registering two 3-D objects with a large but different number of (non-corresponding) points has been known for a long time in the computer vision literature, where the Iterated Closest Point (ICP) algorithm (Besl & McKay, 1992; Zhang, 1992) is a standard. Consider the configurations of two distinct *unlabeled* objects $\boldsymbol{X}_q \in \mathbb{R}^{m_1 \times d}$ and $\boldsymbol{X}_p \in \mathbb{R}^{m_2 \times d}$, not necessarily having the same pose and assume $m_1 \leq m_2$. Let $\boldsymbol{M} \in \mathbb{R}^{m_1 \times m_2}$ with $M_{ij} = 1$ if $\boldsymbol{x}_{q,i} \in \boldsymbol{X}_q$ is matched with point $\boldsymbol{x}_{p,j} \in \boldsymbol{X}_p$, and zero otherwise. The objects may be located and oriented differently in space (see Figure 1), and hence the problem is not only to find a rigid body transformation but also the correspondences given by $\boldsymbol{M}$. This is a hard non-linear (not convex) discrete optimization problem.

Commercial Computer Aided Design (CAD) and inspection software use variants of the ICP method to align the cloud points of a scanned object and that of the CAD design, in order to determine regions in the manufactured part that are different

from nominal. Figure 2 shows an instance of a metal part CAD model and two color-coded comparisons between the CAD model and the 3D-printed part. This is actually voxel data, not point cloud point data on the surface of the object, but the registration problem is essentially identical. The deviations from nominal can be used for process monitoring purposes, in a similar way than traditional "DNOM" control charts (Farnum, 1994). The deviations from nominal are vectors, and either their norm or their individual components could be used for SPC (see Figure 2). The main issue is that aligning two or more point clouds (or meshes) using ICP with hundreds of thousands to millions of points in each mesh is not a task fast enough for *on-line* process monitoring. If in an effort to reduce the computational time in the ICP alignment problem one solves it only approximately returning a local optimal solution (given the hard global optimization problem involved) this can have the effect of inflating the variability in the data, an undesirable consequence if our goal is to detect true sources of extraneous variability.

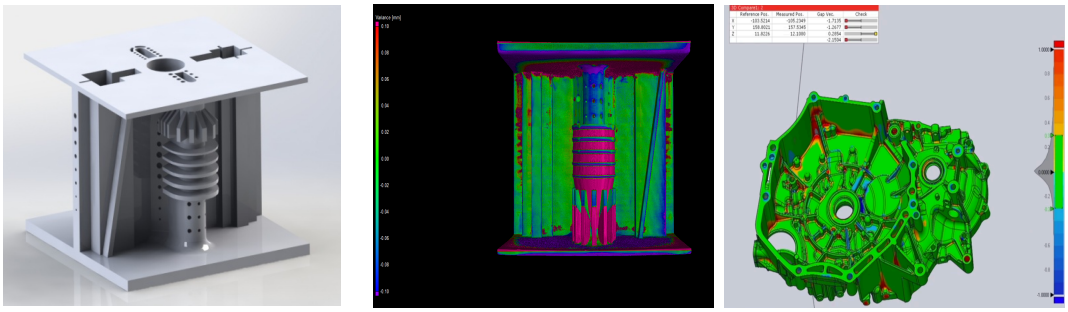

Figure 2: Left: CAD design 3D rendering of a metal part. Middle: CT image of the manufactured part with color contrast indicating differences in the x dimension between the CAD nominal dimension and the actual part x dimension. The CT software registers the CAD design and actual part using the ICP algorithm and color codes the deviations from nominal for visualization (VGStudio was used to create the first two figures). Right: deviations from nominal (or "gap", as called by the Geomagic inspection software) are in fact vectors in 3D.

In Zhao & del Castillo (2019), rather than solving registration problems via ICP or similar algorithms for on-line control, we propose instead to do SPC on properties of the surface (a 2-dimensional manifold) of an object that are not only invariant to rigid transformations (as the shape of an object is invariant) but also that are *intrinsic* properties. A geometric property of a manifold is intrinsic if it is computed

without any use of the coordinates or other information from the space it is embedded in, and only uses coordinates (or information, in general) defined on the manifold. For the surface of an object, surface coordinates and geodesic distances are intrinsic properties, so is the Gaussian curvature. They are all also invariant properties. In contrast, euclidean distances between points in an object are invariant but not intrinsic. By only using intrinsic properties computed from each part, it is possible to avoid the computationally expensive registration step (ICP registration could still be used for post alarm diagnostics, helping to locate the defect on the part, but not for on-line monitoring).

Using intrinsic properties is a central idea in computer vision and computer graphics (see e.g., Biasotti *et al.* (2016); Boscaini *et al.* (2015); Kimmel (2004)). These fields are usually interested in classifying different shapes of objects in broad classes. In most cases, the objects are not point clouds scanned from some real object (and subject to both manufacturing and measurement errors) but are meshes or point clouds obtained from high resolution graphical objects for the purpose of animation and computer graphics. Thus in these fields, the fundamental SPC issue of detecting differences in shape in a sequence of closely similar objects from noisy scanned data does not arise.

As we discuss in Zhao & del Castillo (2019), an intrinsic operator from Differential Geometry which contains considerable geometrical information about a manifold $\mathcal{M}$, and that is widely used for this reason in both machine learning and computer graphics/computer vision, is the so-called Laplace-Beltrami (LB) operator acting on a function f defined on $\mathcal{M}$:

$$\Delta_{\mathcal{M}} f = -\text{div}_{\mathcal{M}} \nabla_{\mathcal{M}} f$$

which reduces to the usual Laplace operator in case $\mathcal{M} = \mathbb{R}^n$. Here, f is a function of the manifold coordinates $x^1, x^2, ..., x^k$. In the case of a surface ($k = 2$) embedded in $\mathbb{R}^3$, the surface coordinates are usually denoted by $(u, v) \equiv (x^1, x^2)$. The intuition is that to measure how a function defined over a curvilinear space varies it is necessary to consider the curvature of the space, and the LB operator evaluated at a point $\boldsymbol{x} \in \mathcal{M}$ is a measure of local curvature of $\mathcal{M}$ (twice the so-called mean curvature) evaluated at $\boldsymbol{x}$. The LB operator appears in partial differential equations modeling difussion processes, and also in the wave equation. The heat equation, for instance, models how the heat at a location $\boldsymbol{x} \in \mathcal{M}$ varies both in space and in time. The

"spatial" part of the solution of this equation results in the eigenvalue problem:

$$\Delta_{\mathcal{M}} f = \lambda f \tag{1}$$

sometimes called the Helmholtz partial differential equation. The collection of eigenvalues $\{\lambda_i\}_{i=0}^{\infty}$ is called the *spectrum* of the LB operator. The spectrum of the LB operator is always discrete, non-negative, and contains considerable geometrical and topological information about a manifold that can be used for shape identification. The true spectrum of very few manifolds is known and in practice a discrete approximation based on either a network or voxel partition of the manifold needs to be used. When discretized, the infinite dimensional function f becomes a vector, and the LB operator $\Delta_{\mathcal{M}}$ reduces to a matrix $\boldsymbol{L}$. Then the Helmholtz equation becomes a standard linear algebra eigendecomposition problem.

One instance where the true spectrum of a 2-manifold is known is the case of a unit sphere. Using the sphere as an example, Figure 3 shows how the LB operator spectrum is *invariant* to rigid transformations, and therefore, it is potentially useful to detect changes in shape. The spectrum is also intrinsic, as its computation requires only information on the 2-manifold and not in the ambient space. In machine learning, LB operator-based methods have proved useful in both unsupervised and semi-supervised learning methods (e.g., see Belkin & Niyogi (2008)). For unsupervised learning, the spectrum of the LB operator is used. For semi-supervised learning, a discrete-valued function f defined on $\mathcal{M}$ is observed (labeled) at certain points in the point cloud dataset. The main goal is to fit this function, and an approximation to the LB operator is needed based on such dataset. Computer graphics and computer vision authors have also used the LB operator for shape classification (e.g., see Kimmel (2004); Levy (2006)).

2.1 Using the estimated LB spectrum as a tool for SPC

Several discrete approximations of the LB operator of a 2-manifold $\mathcal{M}$ have been developed over the past 15 years in the computer graphics literature. There are also discrete approximations for the case the manifold is a 3D solid and the data are in the form of voxels. In Zhao & del Castillo (2019) we use a *localized mesh Laplacian* (Li *et al.* , 2015) to approximate the LB operator on the triangulations that result from scanning each part. This approximation is useful for SPC purposes as it results

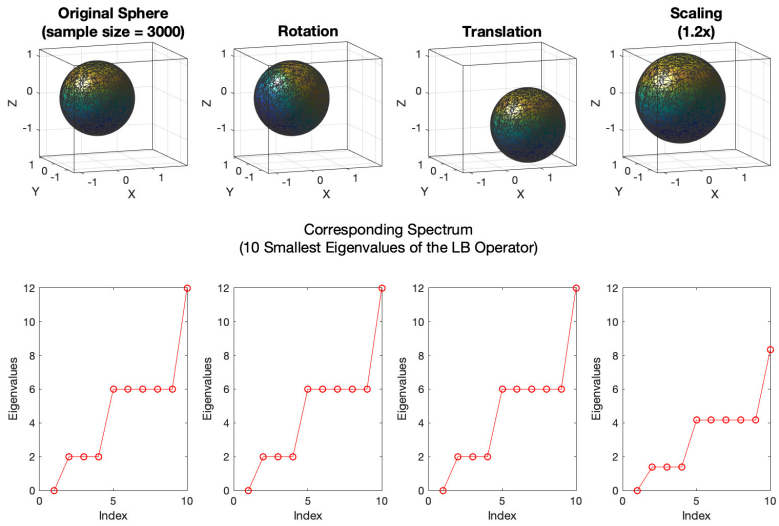

Figure 3: Lower part of the spectrum of the true LB operator of a unit sphere under similarity transformations. As can be seen, the true spectrum is invariant with respect to rigid transformations such as rotations and translations. Scaling the object by s will make the eigenvalues change by $1/s^2$ (here $s = 1.2$ made the eigenvalues decrease close to 30%).

in a sparse matrix $\boldsymbol{L}$, and important consideration given that sequences of $\boldsymbol{L}$ $(m \times m)$ matrices need to be processed for SPC thus not only computation (see below) but also storage issues arise as soon as one handles large, realistic meshes created from non-contact sensors. The lower part of the estimated spectrum is stable to noise (Zhao & del Castillo, 2019) and is known to converge to the spectrum of the true (continuous) LB operator (Li *et al.* , 2015).

Returning to the surface data SPC problem, suppose we have acquired meshes from a series of parts. A simple idea is to use the corresponding sequence of Mesh Laplacian spectra (sorted in natural order cropped up to a certain max eigenvalue) obtained from scanned parts, consider each spectrum a *profile*, and apply well-established SPC techniques for profile data. To illustrate the shape change detection capabilities of the LB spectrum, we count how many times the eigenvalues of a new part fall within an acceptable range. Consider Figure 4 where an acceptable and a defective part are shown: the defective part has a small chip in a corner. We simulated 100 realizations of the acceptable mesh adding isotropic $N(\mathbf{0}, 0.05^2\boldsymbol{I})$ noise

and computed the corresponding spectra from the approximate discrete LB operator. We also simulated one realization of the defective part with the same noise level. The variation of the spectra of the 100 acceptable parts is quite minor, but the defective part has a spectrum whose eigenvalues often fell outside the range of the acceptable eigenvalues. This indicates the potential for the spectrum to be used for SPC purposes.

The computational bottleneck is of course obtaining the eigenvalues of $\boldsymbol{L}$, which depending on its sparseness is an $O(m)$ to $O(m^3)$ operation. In practice, only the lower part of the spectrum is needed, and this, together with the use of a sparse discrete LB operator (Li *et al.* , 2015) permits the on-line control of parts based on realistic (several tens of thousands) points per mesh-part.

For a full exposition of the concept of monitoring intrinsic geometrical properties for SPC purposes, including the LB spectrum for different data types, the discrete approximations used and a discussion of specific control charting schemes, see Zhao & del Castillo (2019).

3 Manifold data in experimental design and analysis

In addition to SPC problems for complex data types, Engineers need also to be acquainted with ways to handle manifold data that relate to designed experiments. Here we discuss recent approaches in two cases of particular importance in quality engineering, the case when the response data lies on a manifold (the shape space of configuration matrices) and the case when the regressors lie on a manifold.

3.1 Response data lying on a manifold

Here, the response is the shape of an object, assumed measured more accurately with a contact sensor in a R&D setting and represented by a configuration matrix $\boldsymbol{X}$. The goal of the experiment is to find controllable factors that have significant effects on the shape of a product or part, and estimate those effects. In this case the traditional techniques from statistical shape analysis, which are based on Procrustes registration and assume one to one correspondences, can be useful, although to our knowledge there is no work done beyond a 2-factor factorial for analyzing shapes using SSA

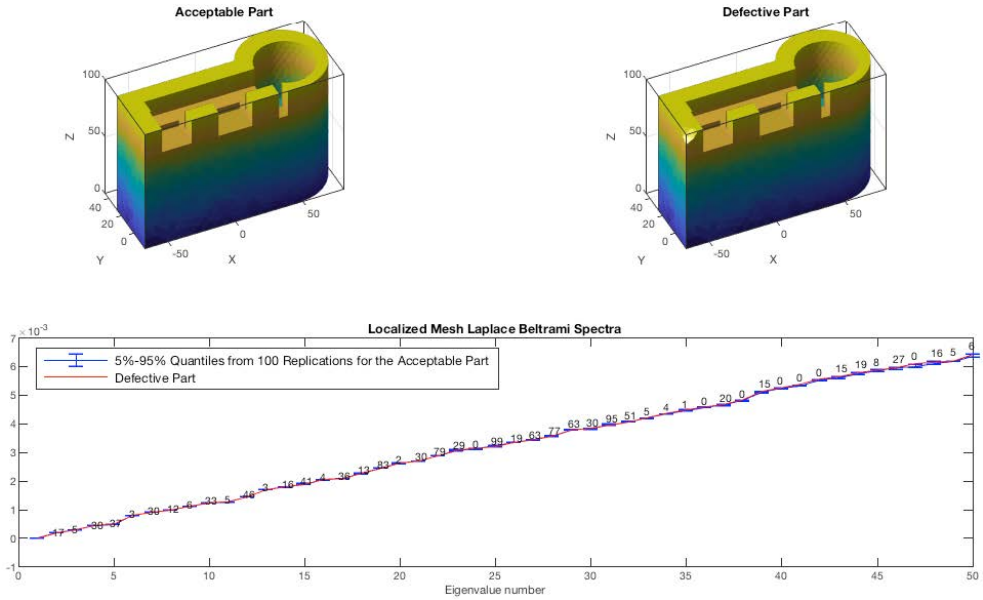

Figure 4: Spectra of 100 acceptable parts vs. spectrum of a defective part with a small chip in one corner. Graph shows bars at 5th and 95th percentiles of the eigenvalues of all acceptable parts vs. the spectra of a defective part. Numbers on graph are the number of times the ith eigenvalue realization of an acceptable part was larger than the ith eigenvalue from the defective part. Therefore, numbers closer to 0 or 100 indicate more power to detect the defect.

techniques. The "shape space", i.e., the space of configuration matrices $\boldsymbol{X}$ after discounting the effect of similarity or rigid transformations excluding reflections, is not an euclidean space but is a curved manifold (Kendall, 1984). When working with data items located on a curvilinear space, such as configuration matrices on shape space, if the registered matrices are considerably different and therefore far from each other, standard statistical tools such as ANOVA and PCA, based on euclidean concepts (e.g., Pythagoras theorem) do not apply and new tools are necessary to analyze these complex data.

An early reference in the analysis of experiments where the responses are shapes is Snee & Andrews (1971). Their work predated the modern developments in the statistical analysis of shapes by Kendall and others, but had the merit of pointing out the importance of experiments where it is necessary to characterize and optimize the shape of an object. These authors illustrated their ideas with application in

agriculture (characterization of the shapes of sweet potatoes as a function of planting date and variety; Snee (1972) also studied the shape of carrots). These authors used the type of analysis of shape data existing in the pre-SSA era: one analyzes ratios of distances between locations of interest in the object (what today are called the landmarks).

One way ANOVA tests for effects on the mean shape of an object were developed by Goodall based on the full procrustes distances $d(\boldsymbol{X}_i, \boldsymbol{X}_j)$, see Goodall (1991) and Dryden & Mardia (2016). The analysis is valid if the noise in the parts (as measured by σ below) is not too big and also if the difference between the shapes is not too large. Contrary to euclidean data, which lies in $\mathbb{R}^n$, we have that in a standard ANOVA using full procrustes distances as metric, the space of registered configuration matrices (shape space) is not euclidean, and therefore $SS_{total} \leq SS_{treatment} + SS_{error}$, with equality approaching as σ gets smaller and as the shapes of the different objects between treatments get closer. These are not unrealistic assumptions in manufactured products and an R&D environment, where the shape differences and noise will typically be moderate.

The one-way ANOVA for shapes was extended to the two-way case, including interaction, by Del Castillo & Colosimo (2011). Let the expected configuration $\boldsymbol{X}$ obtained under treatment 1 level i, treatment 2 level j for the lth replication be (following standard ANOVA notation):

$$E[\boldsymbol{X}_{ijl}] = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\beta}_j + (\boldsymbol{\tau}\boldsymbol{\beta})_{ij}, \quad i = 1, ..., a; j = 1, ..., b; l = 1, ..., n \tag{2}$$

Since the parts may not be initially registered, the first step of the ANOVA on shapes is to register all $abn = N$ shapes with the GPA algorithm. The procrustes fits will then be given by $\boldsymbol{X}^p_{ijl} = \widehat{\beta}_{ijl}\boldsymbol{X}_{ijl}\widehat{\boldsymbol{\Gamma}}_{ijl} + 1_k\widehat{\gamma}'_{ijl}$. The estimated overall mean shape is $\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{X}}_{\bullet\bullet\bullet} = 1/N \sum_i \sum_j \sum_l \boldsymbol{X}^p_{ijl}$ where we use the standard "dot" notation in two-way ANOVA. Provided the shapes of the objects are not too far from the mean shape, we have the approximate ANOVA partition:

$$SS_{total} \approx SS_A + SS_B + SS_{AB} + SS_{error}$$

where $SS_{total} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{l=1}^{n} d_F^2(\boldsymbol{X}^p_{ijl}, \overline{\boldsymbol{X}}_{\bullet\bullet\bullet})$,

$$SS_A = bn \sum_{i=1}^{a} d_F^2(\overline{\boldsymbol{X}}_{i\bullet\bullet}, \overline{\boldsymbol{X}}_{\bullet\bullet\bullet}), \quad SS_B = an \sum_{j=1}^{b} d_F^2(\overline{\boldsymbol{X}}_{\bullet j\bullet}, \overline{\boldsymbol{X}}_{\bullet\bullet\bullet}),$$

$$SS_{AB} = n\sum_{i=1}^{a}\sum_{j=1}^{b} d_F^2(\overline{\boldsymbol{X}}_{ij\bullet} - (\overline{\boldsymbol{X}}_{i\bullet\bullet} - \overline{\boldsymbol{X}}_{\bullet\bullet\bullet}) - (\overline{\boldsymbol{X}}_{\bullet j\bullet} - \overline{\boldsymbol{X}}_{\bullet\bullet\bullet}), \overline{\boldsymbol{X}}_{\bullet\bullet\bullet})$$

and $SS_{error} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{n} d_F^2(\boldsymbol{X}_{ijl}^p, \overline{\boldsymbol{X}}_{ij\bullet})$. An extension of the one-way ANOVA in Goodall (1991) for testing $Ho^{(1)} : \boldsymbol{\tau}_i = \mathbf{0}$, $Ho^{(2)} : \boldsymbol{\beta}_j = \mathbf{0}$ and $Ho^{(3)} : (\boldsymbol{\tau\beta})_{ij} = \mathbf{0}$ is based on the statistics: $F_0^{(1)} = MS_A/MS_{error}$, $F_0^{(2)} = MS_B/MS_{error}$, $F_0^{(3)} = MS_{AB}/MS_{error}$. Similarly as for the one-way ANOVA case, the distributions of these three statistics (under their null hypothesis for small σ) are a $F_{(a-1)M,ab(n-1)M}$, a $F_{(b-1)M,ab(n-1)M}$ and a $F_{(a-1)(b-1)M,ab(n-1)M}$, respectively, where $M = (k-1)m - 1 - m(m-1)/2$ is the dimension of the shape space for k landmarks in $\mathbb{R}^m$. But just as for the one-way ANOVA-shape case, given the restrictive assumptions in the isotropic, normal model, permutation tests were recommended (Del Castillo & Colosimo, 2011). As an aid for visualizing main and interaction effects on the shape of the object the authors presented graphs based on "quiver" plots (see Figure 5). For non-planar surfaces, contrast plots similar to those produced by CAD inspection software (Figure 2) could be used for visualization (the ANOVA for shapes is valid for surface and solid manifolds $\mathcal{M} \subset \mathbb{R}^3$). MATLAB code that implements these methods is available at PSU's Engineering Statistics and Machine Learning Lab web site.

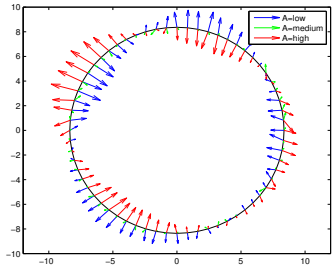

Figure 5: Shape main effects plot for the depth of cut in a cross-section of a titanium part manufactured in a lathe (from Del Castillo & Colosimo (2011)). The circle denotes the overall mean shape $\overline{\boldsymbol{X}}_{\bullet\bullet\bullet} = \widehat{\boldsymbol{\mu}}$. The effects have been increased 1000 times to allow visualization.

Power analysis of the SSA methods for the analysis of experiments presented in Del Castillo & Colosimo (2011) were studied by Alshraideh & Del Castillo (2013).

Comparisons with euclidean distance based methods (Lele & Richtsmeier, 1991) and with the common use of manufacturing *form errors* (cylindricity, circularity) indicate that the ANOVA tests shown above (both the normal-error ANOVA and for general noise distributions, the permutation ANOVA test) provide highest power to detect changes in shape among these methods for a variety of simple 2D and 3D objects.

3.2 Covariate data on a Manifold

Optimal experimental design methods for manifold covariate data have received increased attention over the last decade in the area of *Active Learning* (AL) (Cai & He (2012); He (2010)). The AL problem can be described as follows: given a set of points $\{x_1, x_2, ..., x_n\} \in \mathcal{M} \subset \mathbb{R}^r$, find a subset $\{v_1, v_2, ..., v_m\} = \mathcal{V} \subset \mathcal{M}$ which are the 'most informative' to be *labeled* next. In standard machine learning terminology, a point x_j is labeled if its response value y_j is observed, and *learning* refers to fitting a function from data. AL aims at accelerating learning by choosing the instances to label, essentially an experimental design problem. Figure 6 shows a diagram of a typical AL cycle in which a human annotator acts as an "oracle" providing the labels for instances, in a process that is repeated starting from a state where very little to no labeled data are available. AL is useful when labeling is expensive (precisely because it requires human intervention) and when unlabelled data are in contrast abundant and inexpensive. There are many analogies between AL and the theory of optimal design of experiments (ODOE), as it can be seen in Table 1. ODOE can be used to select the next instance (point) to label, i.e., to select the most informative"query" to the "oracle", in an iterative process. Similar to the case of ODOE, different criteria are used to define 'most informative' and result in different AL strategies. The main issue we wish to discuss occurs when all points, labelled and unlabelled, are in a curved manifold of much lower dimension that their ambient space.

The optimal AL strategy is frequently based on regularized least squares estimation. As far as we know, the first discussion of regularized least squares estimation in the ODOE literature is a paper by Vuchkov (1977). Classical ODOE considers fitting a linear regression model of the form

$$y = f(x, \theta) + \varepsilon \;=\; \theta' z(x) + \varepsilon, \qquad z(x) \in \mathbb{R}^p, \quad \varepsilon \sim N(0, \sigma^2)$$

where $z(x)$ is a vector of p known basis functions evaluated at the vector of r experimental design factors $x \in \mathbb{R}^r$, and $\theta \in \mathbb{R}^p$ is a vector of unknown parameters. In the

Active Learning	Optimal Experimental Design
Active learning strategy	Sequential optimal experimental design
Learning	Model fitting
Labels	Response values y_i
Labeled data	Points x_i *and* their observed response value y_i
Unlabeled data	Candidate list

Table 1: Similitudes between Active learning and optimal experimental design

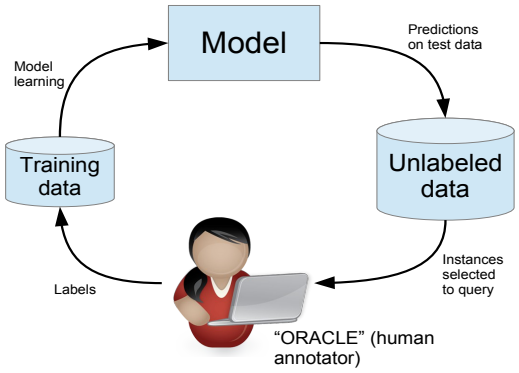

Figure 6: An Active Learning Cycle.

machine learning literature, $z(x)$ is a nonlinear *feature* mapping from the ambient space ($\mathbb{R}^r$) to a higher dimensional feature space.

If "labels" (response values) $\{y_i\}_{i=1}^n$ are available for a sample of n design points $\{x_i\}_{i=1}^n$, as it is well-known, under normal errors, the maximum likelihood estimator of the parameters θ is the least squares estimator:

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}^p} \left\{ \sum_{i=1}^{n} (y_i - \theta' z(x_i))^2 \right\} \; = \; (Z'Z)^{-1} Z'Y$$

where Z is $n \times p$ and Y is $n \times 1$. With this, we have that the fitted function is $\hat{f}(x) = \hat{\theta}' z(x)$. The sometimes called "alphabetic" design optimality criteria focuses on the variance-covariance matrix of the parameter estimates:

$$\mathrm{Var}(\hat{\theta}) = \sigma^2 (Z'Z)^{-1}$$

Optimal DOE seeks to optimize some scalar function of the matrix on the right, as a function of the experimental design. While the theory of convex optimal design

focuses in the case an experimental design ξ is a probability measure over the experimental region (assumed a subset of $\mathbb{R}^n$), in practice an "exact" or "finite" design ξ_n is desired, and this is defined as:

$$\xi_n = \left\{ \begin{array}{cccc} x_1 & x_2 & ... & x_n \\ p_1 & p_2 & ... & p_n \end{array} \right\}$$

where $p_i = r_i/n$ and r_i is the number of observations at point x_i (the notion of replications is practically non-existing in active learning, where most of the time $r_i = 1$), so $\sum_i r_i = n$. For instance, exact D-optimal designs attempt to maximize the determinant of the information matrix $M(\xi_n)$, defined as $|M(\xi_n)| = |\frac{Z'Z}{n}| = |\sum_i p_i z(x_i) z(x_i)'|$. Because algorithms for finding D-optimal designs are frequently iterative and based on the update formula for the determinant, they require to be started at iteration $p + 1$ from a p-point non-singular design. Avoiding an initial p-point design was the main reason Vuchkov (1977) proposed to modify the least squares estimator with $(Z'Z + \lambda I)^{-1} Z'Y$, which corresponds to the *ridge regularized* problem

$$\hat{\theta}_{ridge} = \arg\min_{\theta} \left\{ \sum_{i=1}^{n} (\theta' z(x_i) - y_i)^2 + \lambda \sum_{j=1}^{p} \theta_j^2 \right\} \; = \; (Z'Z + \lambda I)^{-1} Z'Y \qquad (3)$$

Regularization permits fitting complex models based on moderate or small data sets Bishop (2006). The ridge solution (3) is a particular case of a more general learning problem where model fitting is seen as a functional problem in a Hilbert space $\mathcal{H}_K$ endowed with a positive semi-definite kernel function $K(x_i, x_j) : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}$, i.e., a reproducible kernel Hilbert space (RKHS, for a thorough exposition of RKHS in smoothing splines, see Wahba (1990)).

Belkin *et al.* (2006) study how to use the geometry of the manifold where the covariate data resides for a semi-supervised machine learning problem. These authors make the following **manifold assumption**: the marginal distribution $P(y|x)$ is supported in a manifold $\mathcal{M}$ of intrinsic lower dimension than the "ambient space" of x (in our case, lower than r). Furthermore, the marginal distribution is assumed to be smooth as x varies within the data manifold.

If the manifold assumption holds for a dataset, Belkin *et al.* (2006) show how the geometry of the *unlabeled* data carries information that can be used to fit (learn about) a function supported in a manifold using both labeled and unlabelled data.

Their semi-supervised framework consists in solving the double regularized problem:

$$\hat{f} = \arg\min_{f \in \mathcal{H}_K} \left\{ \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda_A ||f||^2_{\mathcal{H}_K} + \lambda_I f' L f \right\} \tag{4}$$

where L is the $m \times m$ combinatorial Laplacian matrix associated with the graph G of all labeled and unlabeled data points x_i, defined as $L = D - W$ but where W is a weight matrix with weights given by the heat kernel $k_t(x_i, x_j)$ $(t = 1)$, and D is a diagonal matrix with ith entry equal to $\sum_{j=1}^{m} W_{ij}$. The first regularization is on the smoothness of f on the ambient space and uses a kernel function $K(x_i, x_j) : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}$, similar to the regularization used when fitting splines. The second regularization term is a penalization on the smoothness of f as it varies between adjacent points in G. Similar models than (4) have been discussed in the Statistics literature, used for fitting regression models on 2-manifolds which only consider the second type of regularization, see e.g., Sangalli *et al.* (2013). It can be shown (see Belkin (2003)) that this second regularization term equals:

$$f' L f = \sum_{i=1}^{m} \sum_{j=1}^{m} (f(x_i) - f(x_j))^2 W_{ij} \tag{5}$$

where m equals the sum of the n points to be labeled and the $m - n$ unlabelled points. Belkin *et al.* (2006) show that the solution of problem (4) in terms of all points, labeled and unlabeled (note the sum of squares in (4) is over the n labeled data points only) is:

$$\hat{f} = \sum_{i=1}^{m} \hat{\alpha}_i K(x_i, x)$$

Substituting this expression in the risk function (4) and using calculus, the solution to the *Laplacian regularized Least Squares* problem is (see Belkin *et al.* (2006)) :

$$\hat{\alpha} = (JK + \lambda_A I + \lambda_I L K)^{-1} Y \tag{6}$$

where K is the $m \times m$ Gram matrix over all labeled and unlabeled points, Y is an $m \times 1$ vector of response values (labels) with the last $m - n$ entries equal to zero, and J is a diagonal matrix with n ones followed by $m - n$ zeroes.

There are two threads in the extant literature on manifold-based AL methods. One thread is based on the "Laplacian Regularized Least Squares" (LapRLS) formulation above (4) and the "representer" (6). Methods in this first thread apply either

D or G (determinant or minimax) DOE optimality criteria to the covariance of the parameter estimates, and are described in section 3.2.1 below.

A second thread is based on the concept of finding a subset of points that is explicitly *representative* of the manifold of points, and are based on an A-optimality (average prediction variance) criterion. We comment in these so-called *transductive experimental design* (TED) methods for AL in section 3.2.2 below.

3.2.1 Laplacian regularized AL methods

Here we study the first thread of manifold AL methods. He (2010) considers the RKHS framework in model (4) and proposes to use the covariance matrix of the parameter estimates for AL purposes. Assume initially a linear regression model $y_i = f(x_i) + \varepsilon = \theta' x_i + \varepsilon$ with $\text{var}(\varepsilon) = \sigma^2$. Then the objective (4) is equal to

$$\begin{aligned} J(\theta) &= (Y - X\theta)'(Y - X\theta) + \lambda_I f(X)' L f(X) + \lambda_A ||\theta||^2 \\ &= (Y - X\theta)'(Y - X\theta) + \lambda_I \theta' X' L X \theta + \lambda_A \theta'\theta \end{aligned}$$

where Y is an $n \times 1$ vector of labels (responses) and we used $f(X) = X\theta$. The goal is to select the $n \times p$ design matrix Z from the "candidate list" of points (assumed to lie on a manifold) X (an $m \times p$) matrix. This is a convex function minimized at

$$\hat{\theta} = (\underbrace{Z'Z + \lambda_I X' L X + \lambda_A I}_{H})^{-1} Z' Y$$

which has a covariance matrix equal to $\text{cov}(\hat{\theta}) = \sigma^2 H^{-1} Z' Z H^{-1}$. Since $\arg\min |H^{-1}| = \arg\max |H|$ we can simply $\max |H|$ with respect to $\{z_1, z_2, ..., z_n\}$. Chen *et al.* (2010) proposed instead a minimax or G-optimal approach, that is:

$$\min_{Z} \max_{x_i \in X} \; x' H_k^{-1} Z_k' Z_k H_k^{-1} x$$

A Laplacian-regularized active learning scheme then corresponds to solving these optimization problems sequentially. Extending this to any kernel other than linear involves using a "kernel trick" (see Bishop (2006)).

3.2.2 "Transductive" manifold AL methods

Yu *et al.* (2006) consider a ridge-like regularized problem. The scaled covariance matrix of the parameter estimates is

$$\frac{\text{Var}(\hat{\theta}_{ridge})}{\sigma^2} = (Z'Z + \lambda I)^{-1}.$$

The prediction covariance matrix across all points in the candidate list X, C_p, is:

$$\begin{aligned} C_p &= X(Z'Z + \lambda I)^{-1}X' \\ &= \frac{1}{\lambda}\left[X'X - XZ'(\lambda I + Z'Z)^{-1}ZX'\right] \end{aligned}$$

These authors then consider minimizing the average prediction variance over the candidate list X, and idea they refer to as "transductive experimental design" (TED). The goal is to find an n-point experimental design Z such that:

$$Z = \arg\min_{Z \subset X} \ \mathrm{tr}(C_p) \ = \arg\max_{Z \subset X} \ \mathrm{tr}(X'Z(\lambda I + ZZ')^{-1}ZX')$$

Therefore, this is an "A"-optimal criterion. This problem is equivalent to

$$\min_{Z \subset X, A} \ \sum_{i=1}^{m} (||x_i - Z'a_i||^2 + \lambda ||a_i||^2) \tag{7}$$

where $A = [a_1, a_2, ..., a_m] \in \mathbb{R}^{m \times n}$. This implies that the average prediction criterion over the candidate points is equivalent to finding a subset of points Z from X that best approximates (in the least squares sense) the whole candidate list X using $\hat{x}_i = Z'a_i$.

Yu *et al.* (2008) then showed how given that the TED problem is NP hard (the TED problem is essentially a subset selection problem), they solve instead a convex relaxation of (7) for which a global optimum is guaranteed, and called their procedure "convex TED".

Finally, Cai & He (2012) and Zhang *et al.* (2014) propose to solve:

$$\min a_1, a_2, ..., a_n \ \sum_{i=1}^{m} ||\tilde{\phi}(x_i) - \tilde{\phi}(Z)'a_i||^2 + \lambda ||a_i||^2 \tag{8}$$

where $\tilde{\phi}(\cdot)$ is the feature vector associated with the *warped* kernel matrix $\tilde{K}^{-1} = K^{-1} + \lambda_I/\lambda_A L$. These authors then used the convex relaxation approach in Yu *et al.* (2008) to optimize (8) leading to what they referred to "Manifold Adaptive Experimental Design" (MAED).

It should be pointed out that all the AL methods reviewed in section 3.2.1 are based on either D or G optimality, whereas all the methods in section 3.2.2 are based on A-optimality.

3.2.3 Application of Manifold AL methods to a real data set

We applied the MAED and "Convex TED" AL methods (see section 3.2.2 and Cai & He (2012) and Yu *et al.* (2008), respectively) to the Wisconsin breast cancer image data set from the UCI Machine learning repository (Lichman, 2013). The dataset, widely studied in the machine learning literature, contains 2 classes of labels for the diagnosis (malignant or benign) associated with 10 attributes obtained from digitalized images of a fine needle aspirate procedure of a breast mass. There are 569 points in total.

The regularization parameters of each method were tuned using cross-validation. A Gaussian kernel was used for MAED and a linear one for Convex TED as these prove best classification in cross-validation. We follow the experimental procedure in Cai & He (2012) who used a different (text) dataset. We ran MAED and Convex TED to select different numbers (k) of points to label. Then the classifier is trained with these k points and their labels, and used to classify the rest of the dat

table 2

randomize the experiments, in each run, we restricted the training examples to be selected only among a subset of 50% of the total data and used the remaining 50% as test data over which the classification rate were computed. Table **??** shows the average and standard errors of the miss-classification errors for MAED, Convex-TED, and a random sampling strategy where the k points to label are simply randomly selected from the training data (this provides a baseline performance benchmark). Figure 7 displays the average classification error rates as a function of k.

As it can be seen, both Convex TED and MAED provide much better classification performance when the number of points to label k is small, but as k grows, the performance of both methods converges to that of randomly sampling new points. This is a desirable behavior in practice, since usually labeling is expensive or time-consuming, and only a few points are selected for labeling. Clearly, MAED performs better than Convex TED, especially for the first hundred iterations. Evidently, this example shares many features with image-based sequential inspection systems in industry.

The applications of ODOE tools for manifold covariate data did not have, until recently, theoretical support like their euclidean counterparts, where the Kiefer-Wolfowitz equivalence theorem between D and G optimality criteria for approximate designs provides some rationale about the discrete or exact application of these ideas. For analogous theoretical work on equivalence theorems for approximate designs on

No. of points to classify	Random Sampling	Convex TED	MAED
15	0.241 ± 0.088	0.189 ± 0.073	0.080 ± 0.022
35	0.140 ± 0.032	0.092 ± 0.024	0.064 ± 0.016
55	0.086 ± 0.028	0.079 ± 0.019	0.054 ± 0.012
75	0.086 ± 0.023	0.068 ± 0.009	0.055 ± 0.015
95	0.064 ± 0.008	0.052 ± 0.011	0.045 ± 0.011
115	0.066 ± 0.010	0.052 ± 0.011	0.044 ± 0.013
135	0.061 ± 0.010	0.052 ± 0.008	0.042 ± 0.009
155	0.060 ± 0.011	0.046 ± 0.005	0.042 ± 0.010
175	0.055 ± 0.012	0.043 ± 0.011	0.041 ± 0.010
195	0.056 ± 0.012	0.041 ± 0.008	0.040 ± 0.005

Table 2: Misclassification error percentages of different AL methods for the Breast Cancer dataset. The regularization parameters in the MAED and Convex TED methods were tuned using cross-validation. Table shows the average and standard deviation of the error percentages.

Riemmannian manifolds, and their use in discrete (exact) designs in an AL context for regression on manifolds, see Li & Del Castillo (2019).

4 Conclusions

We have reviewed some recent concepts related to manifold data and Quality Engineering, with focus on Statistical Process Control (monitoring) and Analysis and Design of Experiments. In the case of SPC, we review a fundamentally new approach for monitoring surfaces of manufactured products via intrinsic properties of the meshes or point clouds generated with non-contact scanners that we have proposed recently (Zhao & del Castillo, 2019). We discuss how the spectrum of the Laplace-Beltrami operator carries considerable geometry information about a closed surface (manifold) in $\mathbb{R}^3$. These methods, since they are intrinsic, do not necessitate expensive pre-processing for alignment of the part with either a CAD model or with other parts. Furthermore, the LB operator and its spectrum can be estimated not only from point cloud and mesh data, but also from voxel data, necessary for the inspection of parts in AM Reuter *et al.* (2007).

We also reviewed recent work on manifold-based data that occurs in designed experiments, either when the response is a surface (manifold), of great importance for quality control and improvement, or when the regressors are concentrated in a

parenthesis

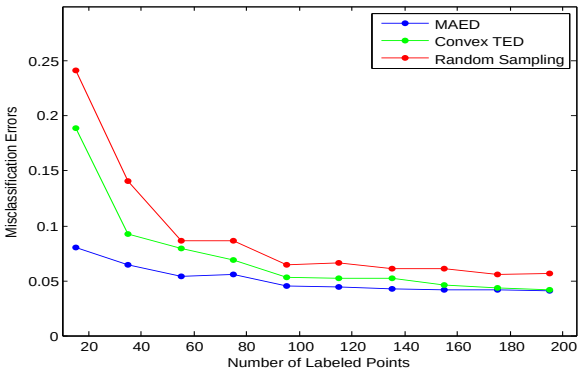

Figure 7: Average misclassification error percentages of two manifold-based AL methods as a function of k, the number of points selected for labeling, compared to a random selection strategy, breast cancer dataset. The regularization parameters in the MAED and Convex TED methods were tuned using cross-validation. Clearly, bigger gains are provided by the two AL methods when these are used to select few points for labeling, which is usually the case when labeling is expensive or time-consuming.

low-dimensional manifold within a very high dimensional ambient space. In the first case, we review an ANOVA approach for shapes, including visualization of effects for planar shapes. In the latter case, we reviewed work on Active Learning (AL) methods, popular in machine learning, applied for manifold data that can be used for quicker defective detection in a quality control or inspection environment.

Acknowledgement.- This research was partially supported by NSF grant CMMI 1537987 and CMMI 1634878. We thank Mr. Hang Li for providing the AL results of section 3.2.3.

References

Alshraideh, H., & Del Castillo, E. 2013. Statistical performance of tests for factor effects on the shape of objects with application in manufacturing. *IIE Transactions*, **45**(2), 121–131.

Belkin, M. 2003. *Problems of Learning in Manifolds*. Ph.D. thesis, The University of Chicago.

Belkin, M., & Niyogi, P. 2008. Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, **74**(8), 1289–1308.

Belkin, M., Niyogi, P., & Sindhwani, V. 2006. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of machine learning research*, **7**(Nov), 2399–2434.

Besl, P.J., & McKay, N. D. 1992. A Method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14(2)**, 239–256.

Biasotti, S., Cerri, A., Bronstein, A., & Bronstein, M. 2016. Recent trends, applications, and perspectives in 3D shape similarity asessment. *Computer Graphics forum*, 87–119.

Bishop, C.M. 2006. *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Boscaini, D., Eynard, D., Kourounis, D., & Bronstein, M. M. 2015. Shape-from-operator: recovering shapes from intrinsic operators. *Eurographcis 2015*, **34**(2), 265–274.

Cai, D., & He, X. 2012. Manifold Adaptive Experimental Design for Text Categorization. *IEEE Trans. on Knowledge and data engineering*, **24**(4), 707–719.

Chen, C, Chen, Z, Bu, Z, Wang, C, Zhang, L, & Zhang, C. 2010. G-optimal design with Laplacian Regularization. *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI-10)*, 413–418.

Colosimo, B., Huang, Q., Dasgupta, T., & Tsung, F. 2018. Opportunities and challenges of quality engineering for additive manufacturing. *Journal of Quality Technology*, **50**(3), 238–252.

Del Castillo, E., & Colosimo, B. 2011. Statistical Shape Analysis of Experiments for Manufacturing Processes. *Technometrics*, **53**(1), 1–15.

Dryden, I. L. 2018. *Shapes package.* R Foundation for Statistical Computing, Vienna, Austria. Contributed package, URL http://www.R-project.org.

Dryden, I. L., & Mardia, K. 2016. *Statistical Shape Analysis with Applications in R, 2nd edition*. Wiley.

Farnum, N. R. 1994. *Modern Statistical Quality Control and Improvement*. Belmont, CA: Duxbury press.

GATech. 2019. *Large Geometrics Model Archive, Georgia Institute of Technology*. URL https://www.cc.gatech.edu/projects/large-models/.

Goodall, C.R. 1991. Procrustes methods in the statistical analysis of shape (with discussion). *J. of the Royal Statistical Society, Series B*, **53**, 285–339.

He, Xiaofei. 2010. Laplacian Regularized D-Optimal Design for Active Learning and Its Application to Image Retrieval. *IEEE Transactions on Image Processing*, **19**(1), 254–263.

Kendall, D. G. 1984. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bulletin of the London Mathematical Society*, **16**, 81–121.

Kimmel, R. 2004. *Numerical geometry of images*. New York, NY: Springer-Verlag.

Lele, S.R., & Richtsmeier, J.T. 1991. Euclidean distance matrix analysis: a coordinate-free approach for comparing biological shapes using Euclidean distance matrix analysis: a coordinate-free approach for comparing biological shapes using landmark data. *Journal of Physical Anthropology*, **86**, 415–427.

Levy, B. 2006. Laplace-Beltrami Eigenfunctions. Towards an Algorithm that "understands" geometry. *Proceedings of the IEEE International Conference on Shape Modeling and Applications (SMI '06)*.

Li, H., & Del Castillo, E. 2019. An equivalence theorem on Manifolds and its use for Active Learning. *Working paper, ESMLab, Dept. of Industrial and Manufacturing Engineering, Penn State University*.

Li, X, Xu, G., & ZHang, Y.J. 2015. Localized discrete Laplace-Beltrami operator over triangluar mesh. *Computer Aided Geometric Design*, **39**, 67–82.

Lichman, M. 2013. *UCI Machine Learning Repository*.

Reuter, M., Niethammer, M, Wolger, F.E, Bouix, S., & Shenton, M. 2007. Global Medical Shape Analysis Using the Volumetric Laplace Spectrum. *2007 International Conference on Cyberworlds (CW'07)*, 417–426.

Sangalli, L., Ramsey, J.O., & Ramsay, T.O. 2013. Spatial Spline Regression Models. *JRSS-B*, **75**(4), 681–703.

Snee, R.D. 1972. A useful method for conducting carrot shape studies. *Journal of Horticultural Science*, **47**, 267–277.

Snee, R.D., & Andrews, H.P. 1971. Statistical design and analysis of shape studies. *Applied Statistics*, **20**(3).

Vuchkov, I.N. 1977. A ridge-type procedure for design of experiments. *Biometrika*, **64**(2), 147–150.

Wahba, G. 1990. *Spline models for observational data*. New York, NY.: SIAM.

Yu, K., Zhu, S., & Gong, Y. 2008. Non-greedy Active Learning for Text Categorization using Convex Transductive Experimental Design. *In: SIGIR '08*. ACM, Singapore.

Yu, Kai, Bi, Jinbo, & Tresp, Volker. 2006. Active Learning via Transductive Experimental Design. *Pages 1081–1088 of: Proceedings of the 23 International Conference in Machine Learning*. ACM.

Zhang, L., Wang, L., Lin, W., & Yan, S. 2014. Geometric optimum experimental design for collaborative image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, **24**(2), 346–359.

Zhang, Z. 1992. *Iterative point matching for registration of free-form curves*. Tech. rept. Reports de Recherche no. 1658, INRIA, Sophia Antipolis, France,.

Zhao, Xueqi, & del Castillo, Enrique. 2019. An Intrinsic Geometrical Approach for Statistical Process Control of Surface and Manifold Data. *arXiv preprint arXiv:1907.00111*.