13th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13
Seoul, South Korea, May 26-30, 2019

High-dimensional interpolation on the Grassmann manifold using Gaussian processes

Dimitris G. Giovanis
Postdoc, Dept. of Civil Engineering , Johns Hopkins University, Baltimore, MD, USA
Michael D. Shields
Ass. Professor, Dept. of Civil Engineering , Johns Hopkins University, Baltimore, MD, USA

ABSTRACT: This paper proposes a novel method for performing interpolation of high-dimensional systems. The proposed method projects the high-dimensional full-field solution into a lower-dimensional space where interpolation is computationally more tractable. The method combines the spectral clustering technique, which refers to a class of machine learning techniques that utilizes the eigen-structure of a similarity matrix to partition data into disjoint clusters based on the similarity of the points, in order to effectively identify areas of the parameter space where sharp changes of the solution field are resolved. In order to do this, we derive a similarity matrix based on the pairwise distances between the high-dimensional solutions of the stochastic system projected onto the Grassmann manifold. The distances are calculated using appropriately defined metrics and the similarity matrix is used in order to cluster the data based on their similarity. Points that belong to the same cluster are projected onto the tangent space (which is an inner-product flat space) defined at the Karcher mean of these points and a Gaussian process is used for interpolation on the tangent of the Grassmann manifold in order to predict the solution without requiring full model evaluations.

1. INTRODUCTION

Using high-fidelity computational models to accurately capture the underlying physics of complex physical systems is a constraining factor of computational mechanics. Despite the huge strides made over the last decades in terms of developing advanced solution algorithms and increasing the available computational resources, there are still major limitations when it comes to large-scale modeling. However, efforts to overcome this limitation are continuous. Along this line of research, reduced-order approaches have been developed, targeting to reduce the degrees of freedom of the system without losing critical information. A widely used class of methods that belongs in this category is the nonlinear projection methods. These methods are based on geometric relations between high-dimensional data by projection onto low-dimensional manifold spaces, i.e. trajectories between solutions computed using high-dimensional models are contained and may be interpolated in low-dimensional subspaces.

When the solution is temporally and/or spatially varying, linear algebra can be utilized in terms of matrices in order to study its behavior. Since an orthogonal matrix represents a linear Euclidean subspace, it can be viewed as a point on a very popular manifold of differential geometry, namely the Grassmann manifold (Absil et al. (2004)). Thus, understanding the intrinsic geometric structure of this manifold is necessary in order to have an insight to the physics of the high-dimensional system.

Clustering is one of the most widely used techniques when grouping of data based on "similar behavior" is seeked, with a wide range of applications. A wide variety of clustering al-

gorithms can be found in literature such as the K-Means, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Expectation–Maximization (EM) Clustering using Gaussian Mixture Models (GMM), Agglomerative Hierarchical Clustering and Spectral Clustering, among other. While traditional algorithms such as K-means in some cases fail to perform well, spectral clustering (Von Luxburg (2007)) is very simple to implement and can be solved efficiently by standard linear algebra methods.

In this work we combine a nonlinear projection method that acts on Grassmann manifolds, coupled with spectral clustering in order to enhance the computational performance of the interpolation of high-dimensional data in terms of computational time and quality using a Gaussian process model.

2. THE GRASSMANNIAN

A manifold of special interest in modern mathematics is the Grassmann manifold (or Grassmannian) $\mathscr{G}_{p,n}$ (Absil et al. (2004)). What makes the Grassmannian appealing is that it can be cast into matrix form, i.e each point on $\mathscr{G}_{p,n}$ is a matrix of size $n \times p$. However, it cannot be defined uniquely and thus, different representations of the Grassmann manifold exist (Edelman et al. (1998)). The Stiefel representation consider each point $\mathscr{X}$ on $\mathscr{G}_{p,n}$ to be an orthonormal matrix, $\mathbf{X} \in \mathbb{R}^{n\times p} \rightarrow \mathbf{X}^\intercal\mathbf{X} = \mathbf{I}_p$

$$\mathscr{G}_{p,n} = \{\mathbf{X} \in \mathbb{R}^{n\times p} : \mathbf{X}^\intercal\mathbf{X} = \mathbf{I}_p\}. \tag{1}$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix. Since the Grassmannian has a globally defined differential structure and locally resembles the Euclidean space, we can define on the manifold features such as the exponential map, the Riemmannian center of mass and the geodesic path and distance. The geodesic path $\gamma : t \in [0,1] \rightarrow \mathscr{G}(p,n)$ for any two points $\mathscr{X}_0, \mathscr{X}_1 \in \mathscr{G}_{p,n}$ spanned by matrices $\mathbf{X}_0$ and $\mathbf{X}_1$, respectively, is defined as the shortest between all smooth paths. Given initial conditions $\gamma(0) = \mathscr{X}_0 \in \mathscr{G}(p,n)$ and $\dot{\gamma}(0) = \dot{\mathscr{X}}_0 = \mathscr{H}_0 \in \mathscr{T}_{\mathscr{X}}\mathscr{G}_{p,n}$ we can define it uniquely as

$$\gamma(t) = \text{span}\left[\left(\mathbf{X}_0\mathbf{V}\cos(t\Sigma) + \mathbf{U}\sin(t\Sigma)\right)\mathbf{V}^\intercal\right] \tag{2}$$

where $\mathbf{U}, \mathbf{V}$ and Σ are the result of the factorizing (SVD) matrix $(\mathbf{X}_1 - \mathbf{X}_0\mathbf{X}_0^\intercal\mathbf{X}_1)(\mathbf{X}_0^\intercal\mathbf{X}_1)^{-1} = \mathbf{U}\Sigma\mathbf{V}^\intercal$. The explicit (logarithmic) mapping from $\mathbf{X}_1$ onto the tangent space $\mathscr{T}_{\mathscr{X}_0}$ defined at $\mathscr{X}_0$ is defined locally and given by the matrix $\Gamma = \log_{\mathscr{X}_0}(\mathbf{X}_1) = \mathbf{U}\tan^{-1}(\Sigma)\mathbf{V}^\intercal$. The inverse mapping, i.e from the tangent space back to the manifold, is called exponential mapping $\exp_{\mathscr{X}_0}(\Gamma) = \mathbf{X}_1 \Leftrightarrow \log_{\mathscr{X}_0}(\mathbf{X}_1) = \Gamma$.

Since the geodesic path does not encompass the information on how "far" away two points are $\mathscr{X}_0$ and $\mathscr{X}_1 \in \mathscr{G}_{p,n}$, there is a need for defining a metric for "distance". The geodesic distance is a nonlinear function of the principal angles θ_i that are calculated from the full SVD of the matrix product $\mathbf{X}_0^\intercal\mathbf{X}_1 = \mathbf{U}\Sigma\mathbf{V}^\intercal$, i.e $\theta_i = \cos^{-1}\sigma_i$ for $i = 1,\ldots,p$ where $\Sigma = \text{diag}(\sigma_1,\ldots,\sigma_p) \in \mathbb{R}^{p\times p}$ $(0 \leq \theta_1 \leq \ldots \theta_p \leq \pi/2)$. Table 1 depicts the most widely used distances.

Name	$d_{\mathscr{G}_{p,n}}(\mathbf{X}_1,\mathbf{X}_2)$
Grassmann	$\left(\sum_{i=1}^p \theta_i^2\right)^{\frac{1}{2}}$
Procrustes	$2\left(\sum_{i=1}^p \sin^2(\theta_i/2)\right)^{1/2}$
Projection	$\left(\sum_{i=1}^p \sin^2\theta_i\right)^{1/2}$

Table 1: Geodesic distances as functions of principal angles.

However, in most practical applications we have data where their representatives on the Grassmann are subspaces of different dimension. So, Ye and Lim (Ye and Lim (2016)) introduced variants of the distance metrics defined on the so-called double infinite Grassmannian $\mathscr{G}_{\infty,\infty}$, that are able to capture the geometry of $\mathscr{G}_{p,n}$ $\forall p < n$. For $\mathbf{X}_1 \in \mathbb{R}^{p\times n}$ and $\mathbf{X}_2 \in \mathbb{R}^{k\times n}$ with $p < k$, in order to measure the distance between $\mathbf{X}_1$ and $\mathbf{X}_2$, $\mathbf{X}_1$ is completed to a k dimensional subspace of $\mathbb{R}^n$, by adding $k-p$ vectors orthonormal to the subspace $\mathbf{X}_2$. Table 2 shows the corresponding to Table 1 distances.

Another important feature that can be defined on the Grassmannians is the so-called Riemannian center of mass or Karcher mean (Karcher (2008)). For of a set of independent sample points $\{\mathscr{X}_i\}_{i=1}^N \subset \mathscr{G}_{p,n}$, it is defined as the point which

Name	$d_{\mathscr{G}_{\infty,\infty}}(\mathbf{X}_1\mathbf{X}_2)$
Grassmann	$\left(\lvert k-p\rvert\pi^2/4+\sum_{i=1}^{\min(p,k)}\theta_i^2\right)^{1/2}$
Procrustes	$\left(\lvert k-p\rvert+\sum_{i=1}^{\min(p,k)}\sin^2\theta_i\right)^{1/2}$
Projection	$\left(\lvert k-p\rvert+\sum_{i=1}^{\min(p,k)}\sin^2(\theta_i/2)\right)^{1/2}$

Table 2: Geodesic distances between subspaces of different dimension.

minimizes locally the cost function $\lambda : \mathscr{G}_{p,n} \to \mathbb{R}_{\geq 0}$:

$$\lambda(\mathscr{Y}) = \frac{1}{N}\sum_{i=1}^{N} d^2_{\mathscr{G}_{p,n}}(\mathscr{X}_i, \mathscr{Y}) \tag{3}$$

where $d_{\mathscr{G}_{p,n}}(\cdot)$ are pairwise distances on the Grassmannian. However, since the Karcher mean is obtained through an optimization procedure, the existence and uniquenes of a solution is not ensured in Riemannian manifolds.

3. Spectral clustering

Clustering is a machine learning technique that involves the grouping of data points $\{\mathbf{x}_1, ...\mathbf{x}_n\}$ based on similar properties and/or features. It belongs to the family of unsupervised learning and can be utilized in order to gain some valuable insights from our data. Spectral clustering, which originates from the graph theory field (Chung (1997)), appears very attractive and is used in this work. Given a set of data points and a metric that expresses pairwise similarity (or dissimilarity) $s_{ij} \geq 0$ $(i, j \in \{1, \ldots, n\})$, the graph Laplacian matrices are used for performing the clustering. The unnormalized graph Laplacian matrix is defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \tag{4}$$

where $\mathbf{D}$ is a diagonal matrix defined as

$$\mathbf{D} = \begin{bmatrix} \sum_{j=1}^{n} w_{1j} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sum_{j=1}^{n} w_{nj} \end{bmatrix}$$

and $\mathbf{W}$ is the weighted adjacency matrix, i.e $\mathbf{W} = w_{ij},\ i, j = 1, \ldots, n$. There exist two unnormalized graph Laplacian matrices in literature (Shi and Malik (2000)) which are related to each other and defined as

$$\mathbf{L}_{\text{sym}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} \tag{5a}$$

$$\mathbf{L}_{\text{rw}} = \mathbf{D}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{W} \tag{5b}$$

where $\mathbf{L}_{\text{sym}}$ is a symmetric matrix, and $\mathbf{L}_{\text{rw}}$ is closely related to a random walk.

4. Gaussian process

The Gaussian process (GP) model (Santner et al. (2003)) is a widely used stochastic interpolation technique used over some experimental design ($\mathbf{x} \in \mathscr{D} \subseteq \mathbb{R}^m$, $\mathbf{y}(\mathbf{x}) \in \mathbb{R}^n$). GP interprets the solution $\mathbf{y}(\mathbf{x}) \in \mathbb{R}^n$, for a $m-$dimensional input $\mathbf{x} \in \mathscr{D} \subseteq \mathbb{R}^m$, of a computational model $\Upsilon(\mathbf{x})$ as a realization of a regression function $\mathscr{F}$ and an underlying, unknown Gaussian process

$$\Upsilon(\mathbf{x}) \approx \tilde{\Upsilon}(\mathbf{x}) = \mathscr{F}(\beta^\intercal, \mathbf{x}) + \sigma^2 a(\mathbf{x}, \omega) \tag{6}$$

In Eq.(6), $a(\mathbf{x}, \omega)$ is the zero-mean Gaussian process with variance σ^2 (ω stands for a realization of the probability space), with covariance $\mathbb{E}[a(\mathbf{x}_i) \cdot a(\mathbf{x}_j)] = \sigma^2\mathbf{R}(|\mathbf{x}_i - \mathbf{x}_j|, \theta)$, where $\mathbf{R}(|\mathbf{x}_i - \mathbf{x}_j|, \theta)$ is the correlation function which is a function of $|\mathbf{x}_i - \mathbf{x}_j|$ with parameter θ. For the regression model $\mathscr{F}$ a linear combination of l chosen functions $\mathbf{f}_i : \mathbb{R}^n \mapsto \mathbb{R}$ can be used, i.e $\mathscr{F}(\beta^\intercal, \mathbf{x}) = \sum_{i=1}^{l} \beta_i f(\mathbf{x}_i) = \beta^\intercal f(\mathbf{x})$.

The regression parameters $\hat{\theta}$ can be estimated through the GP parameters β and σ_y^2

$$\hat{\theta} = \arg\min_{\theta} \left[\frac{1}{N}(\mathscr{Y} - \mathbf{F}\beta)^\intercal \mathbf{R}^{-1}(\mathscr{Y} - \mathbf{F}\beta)[\det(\mathbf{R})]^{1/N}\right] \tag{7}$$

where $\mathbf{R}$ is the correlation matrix of the experimental design. Solving a generalized least-squares problem we can estimate the GP parameters β, σ_y^2 as

$$\beta(\theta) = (\mathbf{F}^\intercal\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{F}\mathbf{R}^{-1}\mathscr{Y} \tag{8}$$

$$\sigma_y^2 = \frac{1}{N}(\mathscr{Y} - \mathbf{F}\beta)^\intercal\mathbf{R}^{-1}(\mathscr{Y} - \mathbf{F}\beta) \tag{9}$$

Since GP is a stochastic surrogate that follows a Gaussian probability distribution, its prediction at a new point $\mathbf{x}^*$ will also be normally distributed with mean value $\mu_{\tilde{Y}}(\mathbf{x}^*)$ and variance $\sigma_{\tilde{Y}}(\mathbf{x}^*)$.

5. REGRESSION ON THE MANIFOLD

For each point of the design experiment, the high-dimensional solution $\mathbf{y}(\mathbf{x}) \equiv \mathbf{v}_i \in \mathbb{R}^{n_{dof}}$, $i = 1,\ldots,N$ where n_{dof} is the degrees of freedom of the computational model, is projected onto the Grassmann manifold using SVD. That means that $\mathbf{v}_i$ is factorized by matrices $\mathbf{U}_i$, $\mathbf{V}_i$ - which by definition belong on some Grassmannians (orthonormal) - and scaled by its eigenvalues Σ. Spectral clustering is utilized in order to group the data based on their solution similarity and then, points that belong to the same group are projected onto the tangent space that is defined at the Karcher mean point for the specific cluster. A GP regression model is then trained for that cluster and used in order to predict the full solution $\tilde{\mathbf{v}}^*$ corresponding to a new point $\mathbf{x}^*$ by using the exponential mapping in order to move from the reduced-order space to the original space. The steps of the proposed method are:

1. Given N available data $(\xi^{(i)},\ \mathbf{v}_i)$, for $i = 1,\ldots,N$, cast each solution $\mathbf{v}_i$ into a matrix form:
$$\{\mathbf{v}_i \in \mathbb{R}^{n_{dof}}\} \rightarrow \{\mathbf{F}_i \in \mathbb{R}^{n_f \times m_f}\}\big|_{i=1}^{N} \tag{10}$$
where $n_{dof} = n_f \times m_f$.
2. Factorize $\{\mathbf{F}_i\}_{i=1}^{N}$ using SVD:
$$\{\mathbf{F}_i\} \rightarrow \{\mathbf{U}_i\Sigma_i\mathbf{V}_i^{\intercal}\}\big|_{i=1}^{N} \tag{11}$$
where $\mathbf{U}_i \in \mathbb{R}^{n_f \times r_i}$,$\Sigma_i \in \mathbb{R}^{r_i \times r_i}$ and $\mathbf{V}_i \in \mathbb{R}^{m_f \times r_i}$ and r_i is the rank of $\mathbf{F}_i$. In the general case, for $i \neq j$ it stands $r_i \neq r_j$ which means that $\mathbf{U}_i$ and $\mathbf{U}_j$ lie on manifolds of different dimension. Thus, we need to embed $\mathbf{U}_i$ and $\mathbf{V}_i$ into the doubly infinite manifolds $\mathcal{G}_{\infty,\infty}$ and calculate the distances there.
3. Calculate $d_{\mathcal{G}_{\infty,\infty}}(\mathbf{U}_k,\mathbf{U}_l)$ for $k,l \in \{1,\ldots,N\}$ and define the similarity matrix
$$(s_{ij}) = \exp\left(-\frac{d^2_{\mathcal{G}_{\infty,\infty}}(\mathbf{U}_k,\mathbf{U}_l)}{\alpha}\right) \tag{12}$$
where α is a scale parameter. In this work, we selected $\alpha = 1.0$.
4. Cluster points $\{\xi_i\}_{i=1}^{N}$ into n_{cl}. Each cluster C_i consists of N_i points ($\sum_{i=1}^{n_c} N_i = N$).
5. For each cluster C_i:
   a. Find its global rank:
   $r_c = \min(r_j : \{\mathbf{U}_j \in \mathcal{G}_{r_j,n_f}\}_{j=1}^{N_i})$
   b. For points $\{\mathbf{U}_i\}_{i=1}^{N_i}$ and $\{\mathbf{V}_i\}_{i=1}^{N_i} \in C_i$ calculate the corresponding centers of mass $\mu_{u,i}$ and $\mu_{v,i}$, respectively.
   $$\{\mathbf{U}_i\}_{i=1}^{N_i} \rightarrow \mu_{u,i} \tag{13a}$$
   $$\{\mathbf{V}_i\}_{i=1}^{N_i} \rightarrow \mu_{v,i} \tag{13b}$$
   c. Project all points of C_i onto the tangent space defined at the Karcher mean:
   $$\{\mathbf{U}_i \in \mathcal{G}_{r_c,n_f}\} \rightarrow \{\Gamma_{u,i} \in \mathcal{T}_{\mu_{u,i}}(\mathcal{G}_{r_c,n_f})\}\big|_{i=1}^{N_i} \tag{14a}$$
   $$\{\mathbf{V}_i \in \mathcal{G}_{r_c,m_f}\} \rightarrow \{\Gamma_{v,i} \in \mathcal{T}_{\mu_{v,i}}(\mathcal{G}_{r_c,m_f})\}\big|_{i=1}^{N_i} \tag{14b}$$
   where the indexes u,v of $\Gamma_{\cdot,i}$ and $\mu_{\ldots,i}$ represent actions regarding to the left eigenvector $\mathbf{U}_i$ and the right eigenvector $\mathbf{V}_i$ in the tangent space, respectively. So, the data now take the form of pairs where the solution lies in a low-dimensional flat space where interpolation is more feasible.
   $$\{\xi^{(i)},\gamma_{u,i}\}_{i=1}^{N_i} \rightarrow \{\mathbb{R}^{n_d},\mathbb{R}^{(r_c \times n_f)}\} \tag{15a}$$
   $$\{\xi^{(i)},\gamma_{v,i}\}_{i=1}^{N_i} \rightarrow \{\mathbb{R}^{n_d},\mathbb{R}^{(r_c \times m_f)}\} \tag{15b}$$
   $$\{\xi^{(i)},\lambda_i\}_{i=1}^{N_i} \rightarrow \{\mathbb{R}^{n_d},\mathbb{R}^{r}\} \tag{15c}$$
   where λ_i, $\gamma_{u,i}$ and $\gamma_{v,i}$ are vectors obtained from reshaping the matrices Σ_i, $\Gamma_{v,i}$ and $\Gamma_{v,i}$, respectively.

Finally we can configure a GP model for every one of the three data pairs, where we assume that $(\gamma_u,\gamma_v,\lambda)$, respectively, are second-order stationary processes with a stationary and square-integrable

covariance function a_u, a_v and a, respectively. For any new realization $\mathbf{x}^\star$ of the parameter vector we can predict the corresponding $\tilde{\mathbf{v}}^*$ using the following steps:

1. Assign $\mathbf{x}^\star$ to a cluster C_k, $k = 1,\ldots,n_{cl}$ based on the minimum euclidean distance between $\mathbf{x}^\star$ and $\{\xi^{(i)}\}_{i=1}^N$:

$$\xi_{\text{closest}} : \min\{d(\xi^\star, \xi^{(i)})\}, \quad i = 1,\ldots,N \tag{16a}$$

$$d(\xi^\star, \xi^{(i)}) = \sqrt{(\xi_1^\star - \xi_1^{(i)}) + \ldots + (\xi_{n_d}^\star - \xi_{n_d}^{(i)})} \tag{16b}$$

2. Using the corresponding trained GP model evaluate the quantities $\tilde{\gamma}_u(\xi^\star), \tilde{\gamma}_v(\xi^\star), \tilde{\lambda}(\xi^\star)$.
3. Reshape vectors $\tilde{\gamma}_u, \tilde{\gamma}_v$ back into the original matrix form i.e $\tilde{\Gamma}_u$ and $\tilde{\Gamma}_v$, respectively, and project them back to their Grassmann representatives $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$. The approximate $\tilde{\Sigma}$ is obtained through transformation of the vector $\tilde{\lambda}$ into a diagonal matrix.
4. The approximate full-field solution at $\xi^\star$ is given by

$$\tilde{\mathbf{v}}^* \leftarrow \mathbf{F}(\xi^\star) = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^\intercal \tag{17}$$

6. Numerical Example

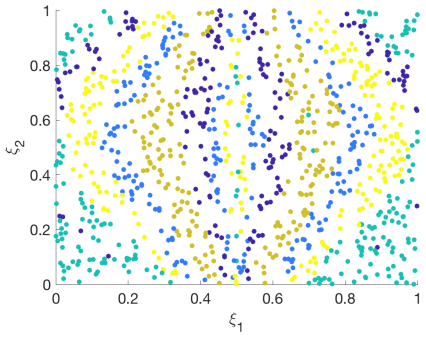
The Kraichnan-Orszag (K-O) three mode problem (Orszag and Bissonnette (1967)), a nonlinear three-dimensional stochastic ordinary differential equations (SODE) system is studied. More specifically, we are utilizing a transformed variant of the original SODE (Wan and Karniadakis (2005)):

$$\begin{aligned} \frac{dv_1}{dt} &= v_1 v_3 \\ \frac{dv_2}{dt} &= -v_2 v_3 \\ \frac{dv_3}{dt} &= -v_1^2 + v_2^2 \end{aligned} \tag{18}$$

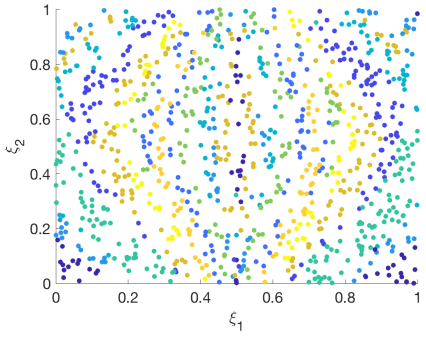
In this formulation, the discontinuity line exists along the plane $v_1 = 0$ and the initial conditions are considered to be stochastic, $v_1(0) = 1.0$, $v_2(0) = 0.1\xi_1(\omega)$ and $v_3(0) = \xi_2$, where ξ_1, ξ_2 are random variables, unifomly distributed in (-1, 1). The available data consists of $N = 1{,}000$ realizations of the parameter vector and its corresponding solution. Each SODE is solved for a period of T=30 with time steps $\Delta = 0.003$, resulting in a solution vector $\in \mathbb{R}^{10^4}$ that we broadcast into a matrix $\mathbb{R}^{100\times100}$. Moreover, $N_{\text{val}} = 200$ additional data were generated in order to quantify the quality of the interpolated solution $\tilde{\mathbf{v}}$ in terms of the root mean square error (RMS)

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^{N_{\text{val}}} (\mathbf{v}_i - \mathbf{v}_i)^2}{N_{\text{val}}}} \tag{19}$$

Figure 1 depicts groups of the parameters based on the similarity of their corresponding solution vectors for different number of number of clusters ($n_{cl} = 5, 10$). In this figure for $n_{cl} = 5$ the existence of symmetrical bands of samples with similar solution behavior is pronounced. However, if we increase the number of clusters to $n_{cl} = 10$ similarity patters are starting to appear which were not obvious before.



(a)

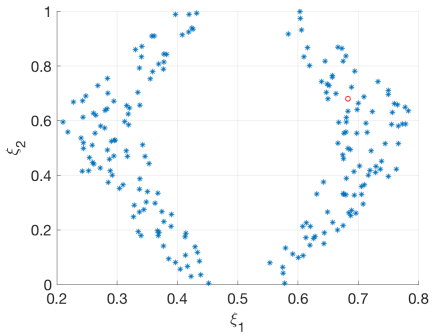


(b)

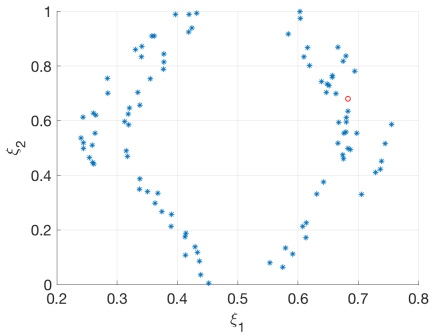
Figure 1: Clustering patterns of the $N = 1{,}000$ input vectors ξ using (a) $n_{cl} = 5$ and (d) $n_{cl} = 10$ clusters.

In order to visualize the quality of an approximate

solution, we select a point $\xi^\star = (0.67, 0.70)$ in the parameter space and we compare the exact solution with the one obtained using the surrogate model, for the two cases of number of clusters ($n_{cl} = 5, 10$). Figure 2(a-b) shows the location of $\xi^\star$ in the parameter space (red marker circle) along with the samples (blue marker) that belong to the respective cluster. As expected, for increasing number of clusters, the number of points that belong to the corresponding cluster decreases.



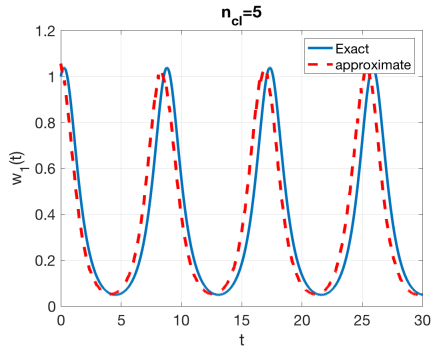
(a)



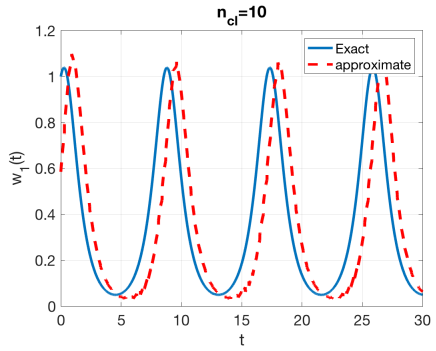
(b)

Figure 2: Samples (blue marker) used for the interpolation of the sample with the red circle marker for (a) $n_c = 5$ and (b) $n_c = 10$ clusters.

Figure 3(a-b) shows a comparison between the approximate time-history corresponding to $\xi^\star$ estimated using the reduced-order model for fitting the GP and the exact solution obtained through a full-model evaluation.



(a)



(b)

Figure 3: Exact vs approximated solutions for a) $n_c = 5$ and (b) $n_c = 10$ clusters.

Last but not least, the sensitivity of the proposed method to the size of solution vector, as a function of the number of clusters is depicted in Figure 4. In this figure we can see the convergence of the mean value of the RMS μ_{RMS}, for increasing number of clusters n_{cl}, for three distinct discretization schemes i.e $n_{dof} = 100, 2,500$ and $n_{dof} = 10^4$. It is obvious that the higher the size of the solution vector the higher the convergence of the RMS. This can be explained by the fact that during the factorization of the solution matrix, more information is projected onto the Grassmann manifold and becomes available to the surrogate.

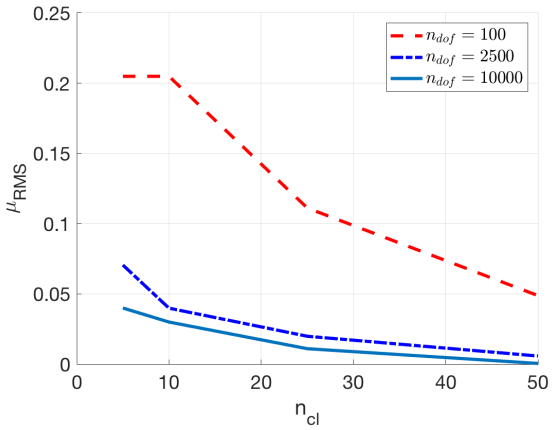


Figure 4: Convergence of the mean value of RMS μ_{RMS} for each discretization case for increasing number of clusters n_{cl}.

7. Conclusions

In this paper, a method for enhancing the computational performance of regression of high-dimensional data by utilizing concepts of differential geometry and machine learning is introduced. To this end, spectral clustering- a modern clustering algorithm- and theory of the Grassmann manifold were combined together with a Gaussian process model in order to perform regression of high-dimensional data after they are projected onto a lower-dimension manifold.

Acknowledgements

Methodological developments presented herein have been supported by the Office of Naval Research.

8. References

Absil, P.-A., Mahony, R., and Sepulchre, R. (2004). "Riemannian geometry of Grassmann manifolds with a view on algorithmic computation." *Acta Applicandae Mathematica*, 80(2), 199–220.

Chung, F. (1997). "Spectral graph theory." *CBMS Regional Conference Series in Mathematics*, Vol. 92.

Edelman, A., Arias, T. A., and Smith, S. T. (1998). "The geometry of algorithms with orthogonality constraints." *SIAM. J. Matrix Anal. & Appl.*, 20(2), 303–353.

Karcher, H. (2008). "Riemannian center of mass and mollifier smoothing." *Communications on Pure and Applied Mathematics*, 30, 509–541.

Orszag, S. A. and Bissonnette, L. (1967). "Dynamical properties of truncated wiener-hermite expansions." *Physics of Fluids*, 10(12), 2603–2613.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer.

Shi, J. and Malik, J. (2000). "Normalized cuts and image segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888–905.

Von Luxburg, U. (2007). "A tutorial on spectral clustering." *Statistics and computing*, 17(4), 395–416.

Wan, X. and Karniadakis, G. E. (2005). "An adaptive multi-element generalized polynomial chaos method for stochastic differential equations." *Journal of Computational Physics*, 209(2), 617–642.

Ye, K. and Lim, L. H. (2016). "Schubert varieties and distances between subspaces of different dimensions." *SIAM Journal on Matrix Analysis and Applications*, 37(3), 1176–1197.