# Grassmann Valued Gaussian Processes (GrassGP) for Probabilistic Active Subspaces *

Yanni Papandreou[†] Jon Cockayne [‡] Andrew B. Duncan[§]

March 7, 2023

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

## 1 Introduction

Dimension reduction is a task of vital importance in the modern day fields of data analysis and scientific computing. Simulations and mathematical models are increasingly important tools in the study of complex physical processes in many scientific and engineering discplines. Unfortunately, complex mathematical models often involve a large number of parameters, leading to increased computational intensity as well as to

---

*

**Funding:** This work was supported by Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1 and EPSRC Grant EP/W006022/1, particularly the "Ecosystems of Digital Twins" theme within those grants & The Alan Turing Institute. YP was supported by a Roth Scholarship funded by the Department of Mathematics, Imperial College London.

[†]Imperial College London (john.papandreou18@imperial.ac.uk)

[‡]University of Southampton (jon.cockayne@soton.ac.uk)

[§]Imperial College London and The Alan Turing Institute (a.duncan@imperial.ac.uk)

a degradation in model performance, a phenomenon infamously referred to as *the curse of dimensionality*.

Many statistical/machine learning tasks boil down to approximating some unknown function $f : \mathbb{R}^d \to \mathbb{R}$. In many real world applications the dimension $d$ of the domain is very large and many popular aproaches for estimating $f$, such as Gaussian process (GP) regression, suffer from the aforementioned *curse of dimensionality*. However, suppose that our function $f$ is such that not all of the $d$ input variables are significant, and only a subset $M$ of $\mathbb{R}^d$ matters. Learning such a subspace $M$ allows us to perform *model order reduction* which can enable us to emulate the true function $f$ more efficiently. This idea originates in the study of *Active subspaces (AS)* [2], where the active subspace can be thought of as the collection of directions along which $f$ is most sensitive to perturbations (on average). This concept will be expanded on in more detail in Section 1.4.

add references to development of actives subspaces in related work section, see Pranay's paper.

In the original formulation of active subspaces one makes the assumption that there exists a **single global** dimension reducing subspace and any emulators based on AS approximations are constructed within this single subspace. In practice such an assumption does not always hold, especially when dealing with highly complex models which can exhibit diverse behaviour across different parameter regimes. This observation naturally motivates one to instead consider finding **different** low dimensional subspaces which are only valid locally in some subregion of the original input space. This task leads one to consider the formal problem of learning a subspace-valued mapping $P : \mathbb{R}^d \to G(d, n)$, where $G(d, n)$ is the Grassmann manifold of all $n$-dimensional (linear) subspaces of $\mathbb{R}^d$. This problem is a member of the more general field of research which deals with the statistical analysis of data lying in a Riemannian manifold. In particular, in this paper we will develop and investigate a probabilistic model based on GPs which can be used to interpolate a field of Grassmann-valued response variables. I.e. given a set of $N$ spatially indexed points in $G(d, n)$, $\{(\mathbf{x}_i, P(\mathbf{x}_i))\}_{i=1}^N$, we will develop a probabilistic model capable of making predictions for $P(\tilde{\mathbf{x}})$ at unobserved test points $\tilde{\mathbf{x}} \in \mathbb{R}^d$.

### 1.1 Related Work

Read these references more closely and expand on them as well as contrasting them with our method.

The main motivation of this paper stems from the literature surrounding active subspaces. The early development of active subspaces can be attributed to the work of Samarov [5], while more recent developments are due to the works of Constantine et al. [2, 1, 3].

There has been extensive research into the problem of learning a subspaced-valued mapping. These works include non-parametric methods [12, see also references therein] as well as semi-parametric methods [8]. The approach we investigate in this paper is motivated by that taken in [4] where we utilise a GP instead of a simple linear regression to allow us to more flexibly discover the spatial dependence of our manifold valued field. The linear model approach taken in [4] builds on the work of [9]. In [13] another GP based probabilistic model is considered for this problem.

Read this and make sure this is an accurate statement.

There are also several works which lie at the intersection of the study of ridge functions and active subspaces. Such works include [7, 10, 6].

The idea of considering several different low dimensional subspaces instead of one global active subspace is not new and has been investigated in [11].

### 1.2 Contributions

### 1.3 Structure of the Paper

## 1.4 Background

## 2 Grassmann Valued Gaussian Processes

## 3 Experiments

# 4 Conclusion

## References


[1] P. Constantine and D. Gleich. Computing active subspaces with Monte Carlo, July 2015.

[2] P. G. Constantine. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. Number 2 in SIAM Spotlights. Society for Industrial and Applied Mathematics, Philadelphia, 2015. ISBN 978-1-61197-385-3.

[3] P. G. Constantine, E. Dow, and Q. Wang. Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces. *SIAM J. Sci. Comput.*, 36(4): A1500–A1524, Jan. 2014. ISSN 1064-8275. doi: 10.1137/130916138.

[4] D. Pigoli, A. Menafoglio, and P. Secchi. Kriging prediction for manifold-valued random fields. *Journal of Multivariate Analysis*, 145:117–131, Mar. 2016. ISSN 0047259X. doi: 10.1016/j.jmva.2015.12.006.

[5] A. M. Samarov. Exploring Regression Structure Using Nonparametric Functional Estimation. *Journal of the American Statistical Association*, 88(423):836–847, Sept. 1993. ISSN 0162-1459. doi: 10.1080/01621459.1993.10476348.

[6] A. Scillitoe, P. Seshadri, C. Y. Wong, and A. Duncan. Polynomial ridge flowfield estimation. *Physics of Fluids*, 33(12):127110, Dec. 2021. ISSN 1070-6631, 1089-7666. doi: 10.1063/5.0064000.

[7] P. Seshadri, S. Yuchi, and G. T. Parks. Dimension Reduction via Gaussian Ridge Functions. *SIAM/ASA J. Uncertainty Quantification*, 7(4):1301–1322, Jan. 2019. ISSN 2166-2525. doi: 10.1137/18M1168571.

[8] X. Shi, H. Zhu, J. G. Ibrahim, F. Liang, J. Lieberman, and M. Styner. Intrinsic Regression Models for Medial Representation of Subcortical Structures. *Journal of the American Statistical Association*, 107(497):12–23, Mar. 2012. ISSN 0162-1459. doi: 10.1080/01621459.2011.643710.

[9] P. Thomas Fletcher. Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds. *Int J Comput Vis*, 105(2):171–185, Nov. 2013. ISSN 1573-1405. doi: 10.1007/s11263-012-0591-y.

[10] C. Y. Wong, P. Seshadri, G. Parks, and M. Girolami. Embedded Ridge Approximations. *arXiv:1907.07037 [cs, math]*, Aug. 2020.

[11] J. Xiong, X. Cai, and J. Li. Clustered active-subspace based local Gaussian Process emulator for high-dimensional and complex computer models, Nov. 2021.

[12] Y. Yuan, H. Zhu, W. Lin, and J. S. Marron. Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(4):697–719, 2012. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2011.01022.x.

[13] R. Zhang, S. Mak, and D. Dunson. Gaussian Process Subspace Regression for Model Reduction, July 2021.

# Supplementary Material

## For: Grassmann Valued Gaussian Processes (GrassGP) for Probabilistic Active Subspaces*


Yanni Papandreou[†] Jon Cockayne [‡] Andrew B. Duncan[§]


March 7, 2023

## S1 Implementation Details for Experiments

In this section we present remarks which give further details on the implementation of the experiments in Section 3.

## S2 Proofs


*

**Funding:** This work was supported by Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1 and EPSRC Grant EP/W006022/1, particularly the "Ecosystems of Digital Twins" theme within those grants & The Alan Turing Institute. YP was supported by a Roth Scholarship funded by the Department of Mathematics, Imperial College London.



[†]Imperial College London (john.papandreou18@imperial.ac.uk)

[‡]University of Southampton (jon.cockayne@soton.ac.uk)

[§]Imperial College London and The Alan Turing Institute (a.duncan@imperial.ac.uk)