

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

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

March 7, 2023

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur 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 disciplines. 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 $f : \mathbb{R}^d \rightarrow \mathbb{R}$. In many real world applications the dimension d of the domain is very large and many popular approaches 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)* [1], 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.

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 $P : \mathbb{R}^d \rightarrow G(d, n)$, 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 $\tilde{\mathbf{x}} \in \mathbb{R}^d$.

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

1.1 Related Work

1.2 Contributions

1.3 Structure of the Paper

1.4 Background

2 Grassmann Valued Gaussian Processes

3 Experiments

4 Conclusion

References

- [1] 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.

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)