1   The Problem

One of the most challenging difficulties facing high-fidelity modeling is the treatment of high-dimensional parameter spaces: the Curse of Dimensionality. Consider a parameter study on some quantity of interest (QoI) f in a space of dimension m; a simple heuristic is to use 10 points per dimension, in order to well represent the parameter space. Then the total number of sample points is 10^m. If a computer code implementing our model executes in a fixed time of 1 second, then our parameter study execution time scales exponentially. Figure 1 depicts the aforementioned scenario.
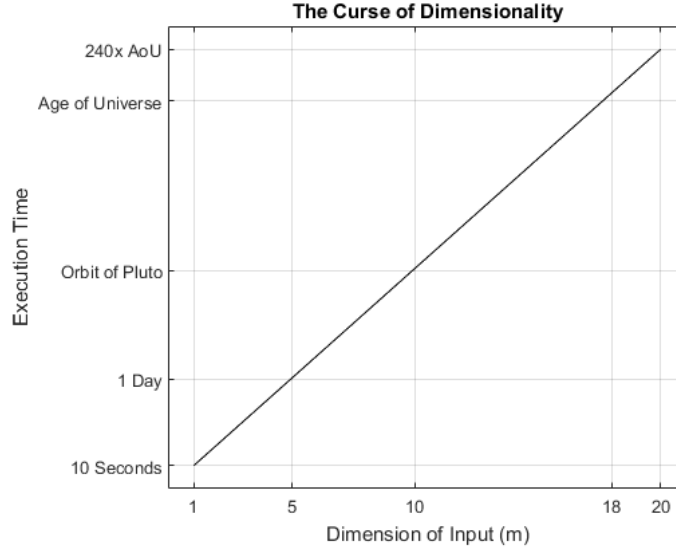

Figure 1: The Curse of Dimensionality; as dimensionality increases, the cost of many dimension-dependent studies (parameter studies, integration, etc.) increases exponentially. In this example, for modestly high-dimensional systems (say 10) the study is already completely intractable.

The only reasonable strategy to mitigate this challenge is to perform *dimension reduction*, that is, to reduce m. One scheme for dimension reduction of this sort is to seek *Active Subspaces* – linear subspaces in parameter space along which the majority of variation in our QoI is captured. [1] Active Subspaces gives a 'perfect' dimension reduction in the case that our QoI is a ridge function; that is, for $\boldsymbol{x} \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times k}$ with $k < m$, we have

$$f(\boldsymbol{x}) = g(A^T \boldsymbol{x}), \tag{1}$$

where $g : \mathbb{R}^k \to \mathbb{R}$. Such a function varies only along the range of A, which can be seen clearly from the gradient $\nabla f = A \nabla g$. Since ∇f lies in the range of A, the gradient is zero along directions orthogonal to the range of A. Active Subspaces have already proven to be a useful strategy in numerous engineering and scientific computing applications, though the technique has some limitations. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(\boldsymbol{x}) = x_1^2 - 2x_2^2. \tag{2}$$

Note that such a function does not fit the definition of a ridge function above. While approximate Active Subspaces are possible and indeed useful in practice, for Equation 2 over any modestly big region of $\mathbb{R}^2$, no accurate Active Subspace is possible. Note that for our example function, we can always move orthogonal to the gradient and see no change in our QoI – in this sense the function can be parameterized by only one variable.

2 Approach

There are many challenges associated with this idea, so the scope of the project will be limited to keep things tractable. The focus will be on identifying manifolds in parameter space $\mathbb{R}^m$ along which the quantity of interest does not change – here called *Inactive Manifolds*. A separate problem is using these manifolds to efficiently parameterize a function – this will not be addressed in this project. Stated mathematically, we seek $W(\boldsymbol{x}) \in \mathbb{R}^{m \times k}$ with $k < m$ such that

$$W^T(\boldsymbol{x})\nabla f(\boldsymbol{x}) = 0, \tag{3}$$

where $W(\boldsymbol{x})$ defines Inactive Manifolds, and k is as large as possible. The space of all possible W is infinite-dimensional, and thus not computationally tractable. In order to attack this problem computationally, we will parameterize W with a number of test manifolds, and find a subset which satisfy the requirements above. Define

$$W_\alpha(\boldsymbol{x}) \equiv \left[\nabla \phi_1(\boldsymbol{x})\alpha_1, \ldots, \nabla \phi_l(\boldsymbol{x})\alpha_l \right] \tag{4}$$

where W_α defines l manifolds; more than the k that exist. We then enforce Equation 3 at a finite number of sampled points $\boldsymbol{x} \in \mathbb{R}^m$. Since W_α is linear in α, we effectively have a nullspace equation for α

$$M\boldsymbol{\alpha} = 0. \tag{5}$$

Equation 5 only holds when there exists an exact family of Inactive Manifolds; in practice we must solve a minimization problem to arrive at an approximate answer. However, since Equation 5 is a nullspace computation, there always exists a trivial answer $\boldsymbol{\alpha} = 0$. Also note that we seek to find a dense solution for α; that is, one which contains many nonzero entries – the more Inactive Manifolds we can find, the better. It is well known that $l1$ minimization encourages sparsity; for instance, see Reference [2]. Conversely, it seems that $l1$ *maximization* should encourage density – this is a conjecture that will be tested in the course of the project. Assuming that $l1$ maximization works as suggested, the proposed approximation to Equation 5 is given by

$$\begin{aligned} \min \ & \|M\boldsymbol{\alpha}\|_2 - \beta\|\alpha\|_1, \\ \text{s.t.} \ & \|\alpha\|_2 \leqslant 1. \end{aligned} \tag{6}$$

The $l2$ inequality constraint is added to prevent the solution from running to infinity (at which point the $l1$ term would dominate), and β is a tunable parameter added to achieve the desired scaling between scalarized objectives. This is a non-linear optimization problem with convex constraints, but a non-convex objective function. The function is continuous over a compact set Ω, thus it attains a minimum in Ω (Extreme Value Theorem). However, since the objective is non-convex, a unique global minimum is not guaranteed to exist.

To solve Formulation 6, I propose to use an interior point method using BFGS and log-barrier functions. I have already implemented such a method for the previous project, so applying the optimization routine should be straightforward. It is likely that I will replace my own BFGS implementation for that of SciPy, as theirs is better tuned. [3] The primary difficulties I anticipate are in the problem formulation – I suspect Formulation 6 is not well-posed, with different local minima. To help detect local minima, I may explore solving the dual problem to study the duality gap, or employ a stochastic optimization routine. More importantly, I anticipate needing to make modifications to the problem statement; one simple addition would be to restrict Ω to a half-space, as the negative of any vector is simply repeat information, with respect to Equation 5.

3   Measuring Success

Measuring success in this project will proceed in a few stages:

First, I will compare against Active Subspaces. One can show that with a linear choice of basis functions $\{\phi_i = x_i\}_{i=1}^m$, we recover Active Subspaces by solving Equation 5 with an SVD. If Formulation 6 gives the same results as Active Subspaces with the choice of ϕ_i given above, that is a good indication that the optimization problem well-approximates Equation 5. This will establish minimal functionality for the project.

Second, I will use different basis functions ϕ_i to search for Inactive Manifolds on different QoI test cases. I have already built some cases which can be solved exactly under Equation 5 – I will check that the procedure can recover these solutions. Then, I will design some test cases which demand approximate Inactive Manifolds, and compare the results to Active Subspaces. The conjecture is that, in the case of Inactive Manifolds, the $l2$ error in $M\boldsymbol{\alpha}$ should be smaller and dimensionality larger than that of the Inactive Subspace. This would show that Active Manifolds (which lie orthogonal everywhere to Inactive Manifolds) could potentially reduce the dimensionality of more general functions than can Active Subspaces – this would be a pretty sweet result.

References

[1] P. Constantine. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. SIAM Philadelphia, 2015.

[2] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[3] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.