Active Subspaces via Monte Carlo Sampling

Jordan R. Hall

Department of Mathematical and Statistical Sciences
University of Colorado Denver

Monday, December 3, 2018

Dimension Reduction and the Active Subspace

Active Subspaces via Monte Carlo

Quick Example!

References

- We consider functions $f : \Lambda \to \mathcal{D}$ where $\dim(\Lambda) = N$ is large and $\dim(\mathcal{D}) = M$ is such that $M < N$ or $M << N$.
  - Functions of interest may represent postprocessed quantities from the solution of complex physical models.
- It is not often that every parameter has equal impact on function values – usually some parameters matter more than others.
- The dimension reduction techniques considered seek to explain outputs $f(\Lambda)$ in an *active subspace* $\mathcal{A} \subset \Lambda$ for which $\dim(\mathcal{A}) < N$.
  - A lower-dimensional representation of f can save computational costs.

- $\nabla f(\lambda) \in \Lambda$ is a column vector with rows containing the N partial derivatives of f, which for this discussion we assume exist, and are square integrable in Λ equipped with some probability density that is positive everywhere in Λ and 0 otherwise.
  - We consider $\pi_\Lambda^{\text{prior}}(\lambda)$, the density describing our prior state of knowledge, which we abbreviate as π_Λ.
- One transforms inputs λ to the origin with some fixed variance, typically so that $\lambda \in [-1, 1]^N$. We define

Covariance in Gradient Space [1]

$$W = \int_\Lambda \nabla f(\lambda) \nabla f(\lambda)^\top \pi_\Lambda(\lambda) d\lambda, \tag{1}$$

which is an $N \times N$ symmetric positive semi-definite matrix.

[1] Constantine

- Interpreting W as a certain covariance structure over Λ leads one to the idea of computing the Singular Value Decomposition of W,

Singular Value Decomposition (SVD) of W

$$W = U\Sigma V^*, \tag{2}$$

where U is $N \times N$ unitary, Σ is $N \times N$ diagonal with the singular values of W along its diagonal, and V^* is $N \times N$ unitary.

- We plot the singular values, $\{\sigma_i\}_{i=1}^n$ and seek a drop-off in magnitude between some pair of singular values, σ_j and σ_{j+1}. The active subspace is the span of $u_1, \ldots, u_j$, which are the first j columns of U, the left singular vectors of W.

- For a point $\lambda \in \Lambda$, we define

Projection into $\mathcal{A}$, *active variables*

$$\mathcal{P}_{\mathcal{A}}(\lambda) = \sum_{i=1}^{j} \left(u_i^T \lambda\right) u_i \in \mathcal{A}, \tag{3}$$

which is the projection of λ in the active directions of f.

- We have arrived at the property that

Resolution of f in $\mathcal{A}$

$$f\left(\mathcal{P}_{\mathcal{A}}(\lambda)\right) \approx f(\lambda). \tag{4}$$

- Finding an active subspace requires forming an approximation to W via Monte Carlo. Here we consider techniques in [1] [2].
- We let $D_S = \{(\lambda_i, f(\lambda_i))\}_{i=1}^S$, which is a set of S pairs of samples $\lambda_i \in \Lambda$ and their function values. One may use D_S to approximate ∇f. We denote each estimation to $\nabla f(\lambda_i)$ with $\widehat{\nabla f}(\lambda_i)$.
- We form the $N \times S$ matrix $\tilde{W}$ (which we present as $\tilde{W}^\top$)

Monte Carlo Approximation to W [1]

$$\tilde{W}^\top := \left[\widehat{\nabla f}(\lambda_1) \cdots \widehat{\nabla f}(\lambda_S) \right]. \tag{5}$$

[1] Russi, T.M.
[2] Constantine, Eftekkari, Wakin

- Forming the SVD of $\tilde{W}$, $\tilde{W} = \tilde{U}\tilde{\Sigma}\tilde{V}^*$, we search for a drop off in the magnitude of the singular values $\{\tilde{\sigma}_i\}_{i=1}^S$. Assuming such a drop off occurs for an index $j : 1 < j < S$, we have the j corresponding left singular vectors,$\tilde{u}, \ldots, \tilde{u}_j$.
- Then we define

Monte Carlo approximation to $\mathcal{A}$

$$\mathcal{A}\left(f; D_S\right) := \text{span}\{\tilde{u}, \ldots, \tilde{u}_j\},$$

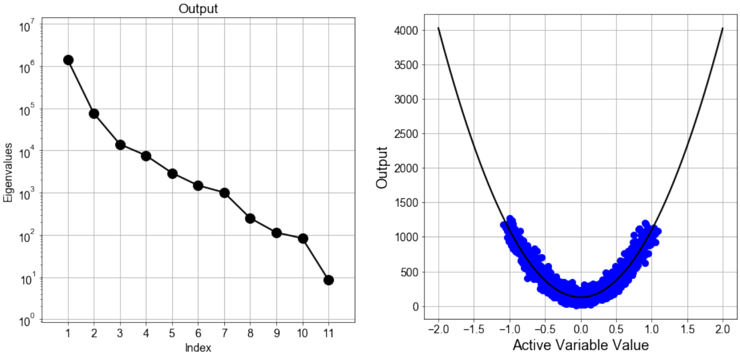
the active subspace of f with respect to the samples D_S.

- We check the extent to which the active subspace accounts for functions values $f(\lambda)$ by checking for resolution in a *sufficient summary plot*, where we plot active variables against function values.

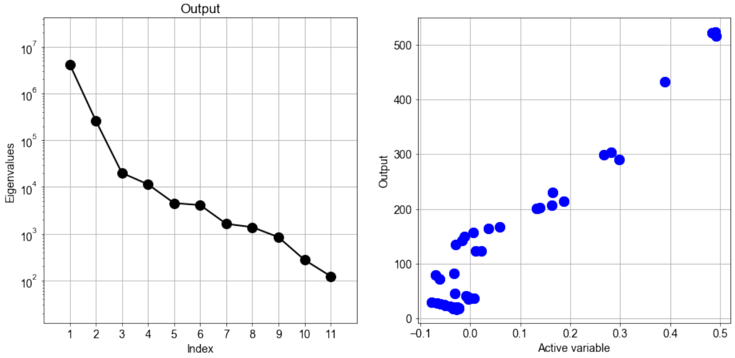
Example 1. Let $\Lambda = [-1, 1]^{11}$ and define

$$f(\lambda) = \sum_{i=0}^{10} 2^{(-1)^i i} \lambda_i^2 + \epsilon(\lambda),$$

where $\epsilon(\lambda)$ is a draw of additive noise corresponding to the input λ; here, we take draws of ϵ of order 10^{-4}. We see that $\mathcal{D} = [0, 2^{10}]$ and $N = 11$, $M = 1$. Note that the minimum of f is given by $0 \in \Lambda$. Here, as i increases, terms in f become either more important or less important, depending on whether i is even or odd.



Left: A plot of the eigenvalues of the matrix $\hat{W}$ formed from 1000 Monte Carlo samples in Λ. We see one dominant eigenvalue on the order of 10^6. Right: A sufficient summary plot where all 1000 samples are projected into $\mathcal{A}$ and plotted against their function values; a quadratic surrogate fits the projected data with $R^2 \approx 0.9$.



Left: A plot of the eigenvalues of the matrix $\hat{W}$ formed from using 100 DFO iterates as samples in Λ. We again see one dominant eigenvalue between orders 10^6 and 10^7. Right: A sufficient summary plot where all 100 samples are projected into $\mathcal{A}$ and plotted against their function values.

- Constantine, Paul G. “Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies." SIAM, 2015.
- Constantine, Eftekhari, Wakin. “Computing Active Subspaces Efficiently with Gradient Sketching." Conference paper, 2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP).
- Russi, Trent M. “Uncertainty Quantification with Experimental Data and Complex System Models." Dissertation, University of California Berkeley. 2010.