Pre-Processing of Cross-Sections Using Dimensionality Reduction Techniques

D. Mandelli[*], *C. Rabiti*[†], *A. Alfonsi*[‡]

Idaho National Laboratory

[*]*diego.mandelli@inl.gov*, [†]*cristian.rabiti@inl.gov*, [‡]*andrea.alfonsi@inl.gov*

INTRODUCTION

Reactor physics codes rely on cross-section data to determine, for example, flux profiles or burn-up calculations. This data contains cross section values for different neutron-atom interactions (e.g., absorption, fission and scattering) for a large number of isotopes as function of energy and it is tabulated for a specific set of parameters (e.g., moderator temperature, fuel temperature, moderator density). In order to perform such neutronic calculations, the amount of this data that needs to be stored is very large due to the high number of isotopes and many tabulation points are needed. This data needs to be retrieved at each time iteration of the code from the machine memory which negatively affects the computational time of the code itself.

This paper aims to reduce the amount of data that needs to be stored in order to decrease computational time of neutronic codes. More specifically a reduction on the number of tabulation points is performed. This reduction is performed by identifying and eliminating redundant information in the tabulated cross section data. This reduction is performed by employing dimensionality reduction techniques and, more specifically, Principal Component Analysis [1] (PCA). We will present the structure of the algorithm and we will evaluate error generated by the data reduction process.

DIMENSIONALITY REDUCTION

Dimensionality reduction is the process of finding a bijective mapping function $\mathfrak{F}$:

$$\mathfrak{F} : \mathbb{R}^N \mapsto \mathbb{R}^M \text{ (where } M < N) \tag{1}$$

which maps the data points from the N-dimensional space into a reduced M-dimensional space (i.e., embedding on a manifold) in such a way that the distances between each point and its neighbors are preserved.

Linear algorithms for dimensionality reduction, such as PCA [1] or multidimensional scaling (MDS) [2], have the advantage that they are easier to implement but they can only identify linear correlations among variables. More advanced algorithms such as ISOMAP [3], Kernel PCA [4], Laplacian Eigenmaps [5] and Local Linear Embedding [6] are however able to identify non-linear correlations among variables. However, a major disadvantage of non-linear algorithms is the difficulty to construct the inverse function $\mathfrak{F}^{-1} : \mathbb{R}^M \mapsto \mathbb{R}^N$.

In our application, the ability to build both $\mathfrak{F}$ and $\mathfrak{F}^{-1}$ is essential and that is the main reason we chose linear algorithms and, in particular, PCA due to its high computational speed and flexibility.

Dimensionality reduction through PCA is accomplished by determining the eigenvectors and their corresponding eigenvalues of the data covariance matrix[1] S. The eigenvectors that correspond to the largest eigenvalues (i.e., the principal components) can be used as a set of basis functions. Thus, the original space is reduced to the space spanned by few eigenvectors and the original data points are projected into this new reduced space.

Figure 1 shows an example of dimensionality reduction using PCA for a data set distributed in a 2-dimensional space. After performing the eigenvalue-eigenvector decomposition of the covariance matrix, the algorithm chooses the eigenvector having the largest eigenvalue (i.e., λ_1) as subspace to project the original data.

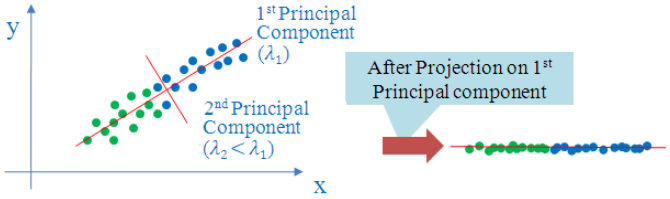

Fig. 1. Example of dimensionality reduction using PCA (reduction from $D = 2$ to $d = 1$).

DATA SET

The methodology that will be described in detail in the next section has been tested on a set of cross-sections data for neutron reactions. These data have been retrieved by an internal tool of the Reactor Physics toolkit PHISICS [7] which is capable of translating the binary AMPX cross section files into a readable format. The binary files have been generated through a SCALE 6.1/TRITON [8] depletion calculation of a standard 17×17 UOx (4.2% enrichment) fuel assembly. Cross sections considered are: n-tot, n-fis,n-abs,n-2n,n-α,n-p. Table I summarizes the characteristics of the cross sections used in this paper.

Data set is then composed of $(N_p^{N_{param}}) = 32$ tabulation

[1]Given a data set in the form of a matrix Z (size $D \times \Lambda$), rows correspond to data dimensions (D) and columns correspond to the number data observations (Λ), the covariance matrix S is determined as: $S = \frac{1}{\Lambda - 1} Z'Z$.

TABLE I. Cross Section Data: Summary

Energy groups N_{EG}	6
Cross sections N_{XS}	6
Tabulation parameters N_{param}	5
Tabulation points N_p	2
Tabulation matrices	$(N_p^{N_{param}}) = 32$
Isotopes N_{iso}	266

TABLE II. Tabulation Values

Parameter p_i	Value 1	Value 2
p_1: Moderator Density (kg/cc)	700	950
p_2: Moderator Temperature (K)	558	589
p_3: Fuel Temperature (K)	900	1200
p_4: Burn-up (GWd/MtHM)	0.0	25.0
p_5: Control Rod Position[a]	0.0	1.0

[a]Ranging from 0 (fully withdrawn) to 1.0 (fully inserted)

matrices (having size $N_{iso} \times (N_{XS} \cdot N_{EG})$):

$$\mathbf{A}(\mathbf{p}), \mathbf{p} = [p_1, \ldots, p_{N_{param}}] \tag{2}$$

each of them composed of 11,172 ($N_{iso} \cdot N_{XS} \cdot N_{EG}$) elements. Figure 2 shows how each element of the tabulated matrices is distributed: elements of the matrix with high and low variability are pictured in red and blue respectively. Our dimensionality eduction algorithm will act on the regions of the matrix with low variability.

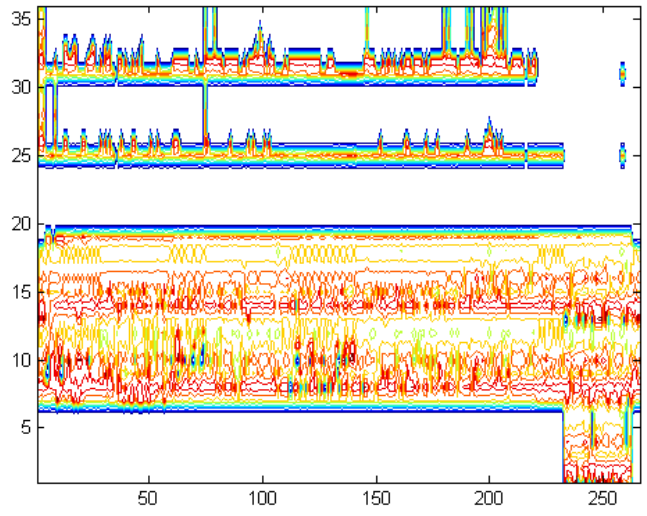

Fig. 2. Distribution of each element of the tabulated matrices: elements of the matrix with high and low variability are pictured in red and blue respectively

Interpolation is performed linearly:

$$\begin{aligned}\mathbf{A}(\widetilde{\mathbf{p}}) &= \sum_{i=1}^{5} \alpha_i \mathbf{A}(p(i)) \\ &= \sum_{i=1}^{5} \frac{\mathbf{A}(p_i(2)) - \mathbf{A}(p_i(1))}{p_i(2) - p_i(1)} (\tilde{p}_i - p_i(1))\end{aligned} \tag{3}$$

The reduction process will focus on the reduction of the number of tabulation matrices. Before performing the data reduction we investigated the distribution of each element of the tabulated matrices. Figure 2 plots this analysis: elements of the matrix with high and low variability are pictured in red and blue respectively.

ALGORITHM IMPLEMENTATION

As mentioned earlier, we chose a linear algorithm for dimensionality reduction based on PCA [1]. The algorithm is structured as follows:

Algorithm 1 PCA based Algorithm

1: Given N $R \times C$ matrices, view each matrix X_n $(n = 1, \ldots, N)$ as an RC-dimensional vector
2: Compute the mean $\bar{X}$ of the X_n matrices
3: Compute deviation matrix $U = [U_1, \ldots, U_n, \ldots, U_N]$ where $U_n = \bar{X} - X_n$,
4: Compute the covariance matrix $Cov = U^T U$
5: Perform eigenvalue/eignevector decomposition of Cov; N eigenvectors w_n and N corresponding eigenvalues λ_n are generated
6: Sort eigenvalues in decreasing order
7: Choose the first M eigenvalues and their corresponding eigenvectors
8: Compute $v_m = U w_m$ $(m = 1, \ldots, M)$
9: Each matrix X_n can be approximated as:

$$X_n \approx \sum_{m=1}^{M} y_{n,m} v_m \tag{4}$$

where $y_{n,m} = v_m^T U_n$

In our application Step 7 is performed by looking at the reconstruction relative error: M is such that such error is below 1%:

$$\left\| \frac{X_n - \sum_{m=1}^{M} y_{n,m} v_m}{\sum_{m=1}^{M} y_{n,m} v_m} \right\|_\infty < 1\% \tag{5}$$

RESULTS

We applied the PCA-based dimensionality reduction Algorithm 1 based on the full data set with the scope of reducing the

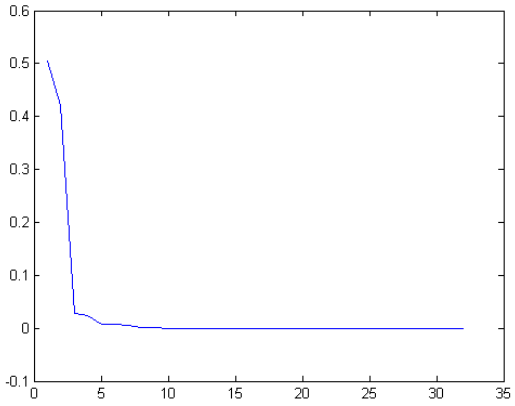

Fig. 3. Plot of the eigenvalues values for each of the 32 eigenvectors (sorted in descending order)

number of cross section matrices. In this respect, Fig. 3 shows the value of the eigenvalue for each of the 32 eigenvectors sorted in descending order. As also described in Algorithm 1, the reduction is performed by choosing the first M eigenvectors such that the relative error in the reconstruction in below 1% (see Eq. 5). For the data set described above we found that 14 was the minimum number of eigenvectors that satisfies Eq. 5. A lower number of eigenvectors would negatively affects the reconstruction error (see Eq. 5). That allowed us to reduce the amount of data needed by 56%.

It is also relevant to mention that using only two points by tabulation parameters created large variation in the data set. This impairs the algorithm from being more effective. When this algorithm will be applied to a finer sampling of the parameter space (i.e., a larger number of cross section tabulation matrices) it is expected to increase its effectiveness (i.e., a wider reduction of tabulation matrices) and also to provide an assessment of the minimal sampling to ensure a given accuracy.

CONCLUSIONS

This paper shows a first approach to reduce the amount of cross section data needed for neutronic codes. We implemented a PCA-based algorithm and despite the limitation of the algorithm to model only linear correlations among variables, we were able to considerably reduce the original set of cross section matrices from 32 to 14 with only limited error in the reconstruction process (below 1%). Further work will include the following:

- Testing the algorithm presented in this paper for much larger cross section data (in terms of both number and size of the matrices)
- Implementation and testing dimensionality reduction algorithms that can also model non-linear correlations among variables
- Reduce the actual size of the matrices by not considering regions of the matrices that have similar patterns (see Fig. 2)
- Perform dimensionality reduction for MOX fuels and for assembly located near control rods

REFERENCES

1. I. T. JOLLIFFE, *Principal Component Analysis*, Springer, second ed. (October 2002).
2. I. BORG and P. GROENEN, *Modern Multidimensional Scaling: Theory and Applications*, Springer-Verlag New York (2005).
3. J. B. TENENBAUM, V. DE SILVA, and J. C. LANGFORD, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, **290**, 2319–2323 (2000).
4. B. SCHLKOPF, A. SMOLA, and K. MÜLLER, "Kernel principal component analysis," *Advances in kernel methods: support vector learning*, pp. 327–352 (1999).
5. M. BELKIN and P. NIYOGI, "Laplacian Eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, **15**, 6, 1373–1396 (Jun. 2003).
6. S. T. ROWEIS and L. K. SAUL, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, **290**, 2323–2328 (2000).
7. C. RABITI, Y. WANG, G. PALMIOTTI, H. HIRUTA, J. COGLIATI, and A. ALFONSI, "PHISICS: a New Reactor Physics Analysis Toolkit," in "Proceeding ofAmerican Nuclear Society (ANS)," (2011), vol. 104, pp. 831–833.
8. A. ALFONSI, C. RABITI, and A. S. EPINEY, "PHISICS TOOLKIT: MULTI-REACTOR TRANSMUTATION ANALYSIS UTILITY - MRTAU," in "Proceeding of Physor Conference (ANS)," (2012).