17th World Conference on Earthquake Engineering, 17WCEE
Sendai, Japan - September 13th to 18th 2020

NUMERICALLY EFFICIENT METHODOLOGY FOR DEVELOPING NON-ERGODIC GROUND-MOTION MODELS USING LARGE DATASETS

M. Lacour [1], N. A. Abrahamson [2]

[1] *Post-doctoral researcher, University of California, Berkeley, maxlacour@berkeley.edu*
[2] *Adjunct Professor, University of California, Berkeley, abrahamson@berkeley.edu*

Abstract

The field of Probabilistic Seismic Hazard Analysis (PSHA) is currently moving from the ergodic assumption, which assumes that the ground-motion model applies to all sites in a broad region, to the non-ergodic assumption, in which the coefficients in the ground-motion models are allowed to vary with spatial locations called Varying-Coefficient Models (VCM). The VCM approach uses the available dataset to include source, site, and path conditions that are specific to each spatial location. To have enough local data, the recorded ground-motion data sets are expanded to include data from small magnitude earthquakes. As the small magnitude data is included for a region, the dataset gets very large (e.g. up to 100,000 recordings in California), or if numerical simulations are used (e.g. 100,000 to 1,000,000 simulated ground motions) some efficient numerical methods are required to compute the parameters of the VCM model and forward predictions of the median ground motion and its epistemic uncertainty. For a dataset of 100,000 ground motions, the covariance matrix involved in the predictions is of size 10^5 x 10^5, which requires about 100 GB of memory storage with double precision, and large amounts of computational power to obtain its inverse and use it in forward predictions, which does not make the method practical.

To avoid such large memory and computational requirements, we use Structured Kernel Interpolation for Products (SKIP), to obtain a low-rank decomosition of the large covariance matrix, via interpolation of the covariance between some reference (or inducing) points. SKIP removes large memory requirements, and brings down the computational cost of ground-motion predictions of several orders of magnitude. This numerical method makes the development of non-ergodic ground-motion models practical. Large scale predictions can be obtained using SKIP in a run time of a few minutes using a regular laptop. We present the method and illustrate it with an application to ground-motion data from California.

Keywords: non-ergodic, seismic hazard, large-scale data

17th World Conference on Earthquake Engineering, 17WCEE
Sendai, Japan - September 13th to 18th 2020

1. Introduction

The field of Probabilistic Seismic Hazard Analysis (PSHA) is currently moving from the ergodic assumption, which assumes that the ground-motion model applies to all sites in a broad region, to the non-ergodic assumption, in which the coefficients in the ground-motion models are allowed to vary depending on the earthquake location and the site location, called Varying-Coefficient Models (VCM). The VCM approach uses the available dataset to include source, site, and path conditions that are specific to each spatial location. To have enough local data, the recorded ground-motion data sets are expanded to include data from small magnitude earthquakes. As the small magnitude data is included for a region, the data set gets very large, up to 100,000 recordings in California, or if numerical simulations are used, 100,000 to 1,000,000 simulated ground motions. VCM models are based on Gaussian Processes methods, where the equations involved for inference (learn from the data the parameters that govern the ground-motion model, also called hyperparameters) and predictions have a complexity of O(N^3) in computational time and O(N^2) in memory. Such computational complexities limit the use of these equations to a small amount of data only, in practice, to a few thousand earthquake scenarios. For a data set of 100,000 ground motions, the covariance matrix involved in the predictions is of size 10^5 x 10^5, which requires about 100 GB of memory storage with double precision, and large amounts of computational power to obtain its inverse and use it in forward predictions, which does not make the method practical. Therefore, some efficient numerical methods are required for inference and forward predictions of the median ground motion and its epistemic uncertainty.

An efficient numerical method that deals with large amounts of data was recently developed, called Scalable Kernel Interpolation (SKI) [1]. SKI performs sparse and accurate approximations of a covariance matrix, via kernel interpolation from some reference points, also called inducing points. Additionally, SKI allows the use of fast numerical methods for matrix-vector multiplications, that take advantage of the properties of the covariance functions such as stationary kernels, and grid-based input data points. Combining sparse approximations and fast matrix-vector multiplication algorithms makes SKI a numerically efficient tool for fast hyper-parameter learning and forward prediction with little memory requirements. However, SKI performs well only for covariance matrices of low dimensions, for example, using the covariance between the data along latitude or longitude only. When several dimensions are considered (latitude, longitude, magnitude, distance, ...), SKI does not perform well, because each dimension requires a set of inducing points, leading to a total the number of inducing points growing exponentially with the number of dimensions. We therefore refer to an extension of SKI, called Structured Kernel Interpolation for Products (SKIP) [2], that alleviates the exponential complexity of SKI with increasing numbers of dimensions. SKIP approximates covariance matrices along each dimension separately using SKI, and then combines them by exploiting mathematical properties of kernel products. SKIP results in linear rather than exponential scaling in complexity with dimension compared to SKI, and leads to low-rank decompositions of the covariance matrices involved. We introduce SKI and SKIP in the next sections, and further describe their application to non-ergodic ground motion prediction.

2. Scalable Kernel Interpolation (SKI)

SKI allows to obtain a sparse and accurate approximation of a large covariance matrix by interpolation from the values of a smaller covariance matrix of some fixed, reference points, also called inducing points. The idea behind SKI is that the values from a large covariance matrix K_{XX} between some large data set X should be close enough from the values of the small covariance matrix of the inducing points K_{UU}, so that we should be able accurately approximate the values of K_{XX} by interpolation from the values of K_{UU}. We therefore only need to compute the small covariance matrix K_{UU} and the interpolation weights to obtain a sparse

approximation of K_{XX} using SKI, which requires few computations and memory. Mathematically, this sparse approximation is given by:

$$K_{XX} \approx K_{SKI} = W * K_{UU} * W^T \quad (1)$$

where W is a sparse matrix of interpolation weights, with n×4 non-zero entries only using cubic interpolation. We illustrate in Fig. 1 the placement of inducing points for SKI over the Bay Area, and in Fig. 2 the error of the approximation between SKI and the direct approach for a squared exponential covariance function $k(x, x') = \theta e^{-\frac{\|x-x'\|^2}{2\rho}}$ with parameters θ = 0.5 and ρ = 1.

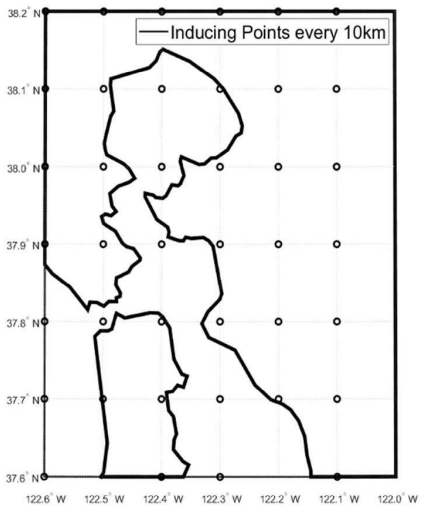

Fig. 1 – Inducing points placed every 10km over the Bay Area

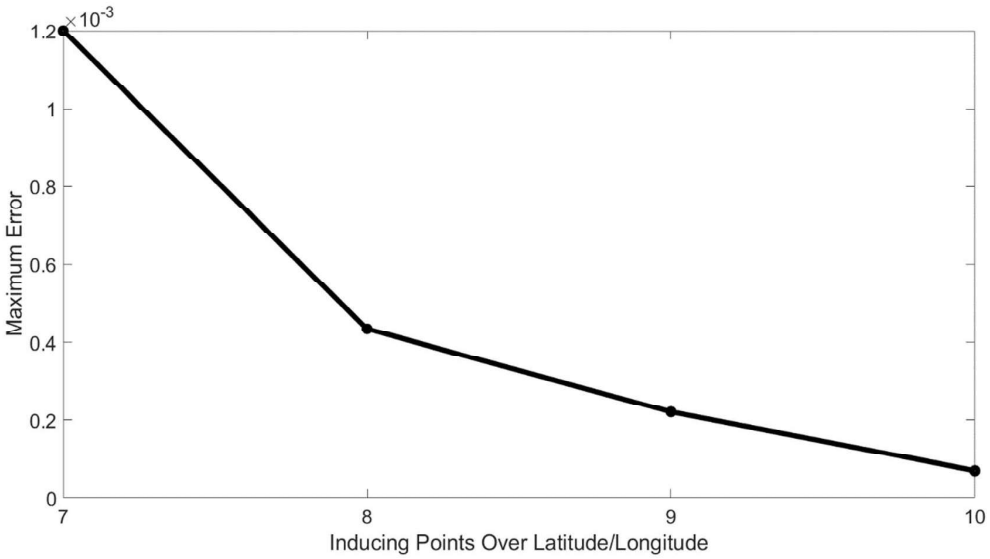

Fig. 2 – Error Between Covariance Matrix from Direct Approach and SKI with Number of Inducing Points

SKI gives an accurate and sparse approximation of large covariance matrices over small domains such as the Bay Area; however, it does not remain efficient for large domains, such as the State of California, due to an exponentially growing number of inducing points per dimension. We next introduce SKIP, an improvement of SKI, which reduces this computational complexity.

3. Structured Kernel Interpolation for Products (SKIP)

The main idea behind SKIP is that some kernel functions, like the ones used in the non-ergodic model, are separable over their dimensions. Therefore, the covariance matrix involved can be expressed as the element-wise product of one-dimensional covariance matrices. One-dimensional covariance matrices can efficiently be approximated using SKI, and their element-wise products require much less operations than their full products. SKIP takes advantage of these properties to efficiently obtain low-rank decomposition of large

covariance matrices from multidimensional kernels over large domains. For example, for a squared-exponential covariance function over latitude and longitude with parameters θ and ρ, we have:

$$k(lat, lon; lat', lon') = \theta e^{-\left[(lat-lat')^2+(lon-lon')^2\right]/2\rho^2} \quad (2)$$

which can be separated over latitude and longitude, such as:

$$k(lat, lon; lat', lon') = \theta e^{-(lat-lat')^2/2\rho^2} \times e^{-(lon-lon')^2/2\rho^2} \quad (3)$$

The resulting covariance matrix $\boldsymbol{K_{lat,lon}}$ can be expressed as the element-wise product:

$$\boldsymbol{K_{lat,lon}} = \boldsymbol{K_{lat}} \circ \boldsymbol{K_{lon}} \quad (4)$$

where $\boldsymbol{K_{lat}}$ and $\boldsymbol{K_{lon}}$ can efficiently be approximated using SKI. By taking advantage of fast element-wise operations, and using some efficient algorithm ([3]), we can further efficiently obtain a low-rank decomposition of $\boldsymbol{K_{lat,lon}}$, such as:

$$\boldsymbol{K_{lat,lon}} = \boldsymbol{UDV^T} \quad (5)$$

We now describe how to use SKIP for efficient non-ergodic ground motion prediction.

4. Non-Ergodic Ground Motion Model

The non-ergodic median ground-motion model is given by [4]:

$$y = \beta_{-1}(t_e) + \beta_0(t_s) + \beta_1 M + \beta_2 M^2 + (\beta_3(t_e) + \beta_4 M) ln\sqrt{(R_{JB}^2 + h^2)} + \beta_5(t_e) R_{JB} + \beta_6(t_s) lnVS_{30} + \beta_7 F_R + \beta_8 F_{NM} + \sigma_n \quad (5)$$

where the terms βi are assumed to be independent Gaussian processes with mean zero and spatial covariance functions given by:

$$k_{\beta_i}(t,t') = \theta_{i,}, i = 1,2,4,7,8$$

$$k_{\beta_i}(t_e,t_e') = \theta_i e^{-\left\|t_e - t_e'\right\|^2/2\rho^2} + i, i = -1,3,5 \quad (6)$$

$$k_{\beta_i}(t_s,t_s') = \theta_i e^{-\left\|t_s - t_s'\right\|^2/2\rho^2} + \pi_i, i = 0,6$$

Since the terms βi in Eq. (5) are assumed to be independent, the total covariance matrix $K_{X,X}$ between the data X from the non-ergodic ground-motion model is the sum of covariance matrices $K_{(X,X),i}$ coming from each term in Eq. (5), such as:

$$K_{X,X} = \sum_{i=-1}^{8} K_{(X,X),i} + \sigma_n^2 I_d \quad (7)$$

To compute the median ground motion at a new location t ∗ , with features X∗ , given the observed data set $\boldsymbol{y} = [\boldsymbol{y_1}, \dots, \boldsymbol{y_N}]$ and features $\boldsymbol{X} = [\boldsymbol{x_1}, \dots, \boldsymbol{x_N}]$at locations $\boldsymbol{t} = [\boldsymbol{t_1}, \dots, \boldsymbol{t_N}]$, we use the equations given by [4]:

$$\mu = K_{X*,X} * \left(K_{X,X} + \sigma_0^2 * Id\right)^{-1} * y_{obs} \quad (8)$$

The epistemic uncertainty of the median ground motion is given by:

$$\psi = diag\left(K_{X*,X*} - K_{X*,X}\left(K_{X,X} + \sigma_0^2 * Id\right)^{-1} K_{X,X*}\right) \quad (9)$$

We approximate the large covariance matrices $K_{X,X}$ and $K_{X*,X*}$ by their the low-rank decomposition using SKIP, such as:

$$K_{X,X} = \mathrm{UDV}^{\mathrm{T}} \quad (10)$$

and plug them into Eq. (8) and Eq. (9) to approximate the median ground-motion and its epistemic uncertainty for a given earthquake scenario and site location. Fig. 3 illustrates a set of 500 × 500 = 25, 000 inducing points placed over the entire State of California, for the sparse approximation from SKIP. This approximation is accurate and efficient because of the large amount of inducing points used, and because the covariance over latitude and longitude are treated separately using SKIP. Fig. 4 shows the eigenvalues of the covariance matrix from the non-ergodic model using a dataset of size 2000, and compares it with the first 400 eigenvalues obtained using SKIP. The accuracy reached is of the order 10−4 in the spectral decomposition. The desired accuracy in the spectral decomposition can be prescribed by the user to guarantee that the smallest eigenvalue obtained by SKIP is under a certain threshold. For this example, with 2000 data points, the computational time is 9.8s for the direct approach, and 5.6s for SKIP. If the data set is increased to 100,000 points, then the computational time using SKIP will be about 5 minutes and require about a few hundred MB of memory, whereas the computational time using the direct method will be about 14 days and require about 100GB of memory.

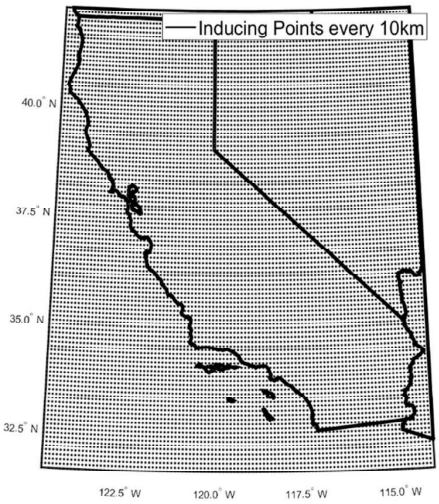

Fig. 3 – Inducing points placed over a grid every 10km in California

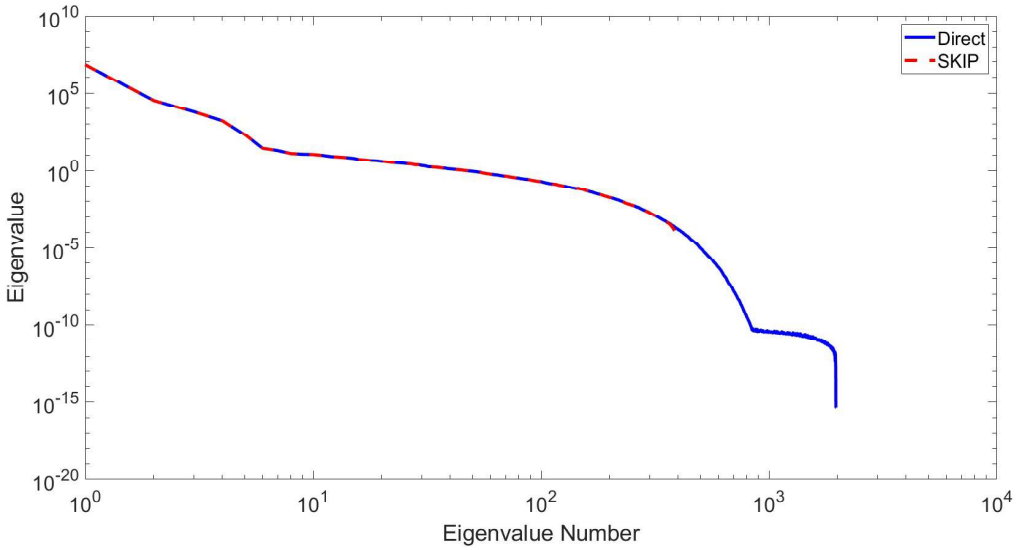

Fig. 4 – Eigenvalues of covariance matrix from full approach (N = 2000), first 400 eigenvalues from SKIP

5. Conclusion

The move to developing non-ergodic GMMs from large data sets of moderate and small earthquakes will require either faster and larger computers or more efficient algorithms to compute the non-ergodic terms using the GP regression. Our approach has been to use more efficient algorithms such as SKIP. With these sparse matrix approximations combined with matrix-vector multiplications provide tremendous improvements in computational speed and reduced memory requirements while still providing good accuracy. The method described in this paper allows non-ergodic GMM to be developed for large data sets using standard laptop computers which will make the non-ergodic approach accessible to a much larger group of ground-motion developers.

6. References

[1] Wilson AG, Nickisch H (2015): Kernel interpolation for scalable structured Gaussian processes (KISS-GP), *Proceedings of the 32nd International Conference on International Conference on Machine Learning,* Lille, France.

[2] Gardner J, Pleiss G, Ruihan W, Wilson AG (2018): Product Kernel Interpolation for Scalable Gaussian Processes. *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, Lanzarote, Spain.

[3] Lanczos, C. (1950). An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *United States Governm. Press Office,* Los Angeles, CA.

[4] Landwehr N, Kuehn NM, Scheffer T, Abrahamson NA (2016): A Nonergodic Ground-Motion Model for California with Spatially Varying Coefficients. *Bulletin of the Seismological Society of America*, Pacific Earthquake Engineering Research, Berkeley, USA.

