

Multiscale and Multiresolution methods for Sparse representation of Large datasets

Prashant Shekhar¹ and Abani Patra²

¹ State University of New York at Buffalo, New York, U.S.A
`pshekhar@buffalo.edu`

² State University of New York at Buffalo, New York, U.S.A
`abani@buffalo.edu`

Abstract

In this research we have presented a strategy which aims at benefitting from studying a dataset at different resolutions. The idea of resolutions here stems from the variation of the correlation strength among the different observation instances. The idea behind the approach is to make a large dataset as small as possible by removing all the redundant information so that as per the requirement, the original data could be retrieved with minimized losses on the information front. We have tested our approach on a carefully designed suite of analytical functions along with gravity and altimetry time series datasets from a section of the Greenland Icesheet. Besides providing with a good strategy for data compression, the proposed approach also finds application in efficient sampling procedures and error filtering in the datasets. The results presented in the article clearly establish the promising nature of the approach along with prospects of its application in different fields of data analytics in the scientific computing as well as other related domains

Keywords: Multiscale, Multiresolution, Kernel methods, Sparse representation

1 Introduction

The increase in the amount of data handling in the scientific domain since the past 2 decades has led the researchers to come up with innovative ways of data analysis. The disproportionate increase in the volume of the data availability and computational resources has motivated the research for new ways of data compression and learning procedures. The methodology proposed in this research directly follows from our previous research work [7] and exploits the correlation structure inherent in the dataset and uses this information to compress the dataset to a reduced number of observation instances while minimizing the amount of information loss. Data compression has been a crucial tool in several fields like image compression [9] and various techniques like signal processing has contributed well to this approach of analyzing data [10]. This approach by its very nature is relevant for a variety of data science domains. Like for

example [3] presented a bayesian approach where they have integrated various datasets with a Glacier dynamics model for inferring icesheet thickness in Western Antarctica. Besides being a data compression approach, the proposed methodology can also be thought of as an efficient sampling methods for the output analysis of various Numerical and data driven models. This could be understood as follows. Suppose we wish to generate the velocity profile of some fluid over a land area for which we already know the topography profile. Then based on the sparse representation of the DEM, we can approximately locate the critical locations over that land area which can explain the entire variability in the topography. Therefore if we can get velocity measurements over those critical coordinates, we can get a very good approximation of how the DEM would affect the velocity profile over the area without incorporating much noise through unnecessary measurements.

Similarly, the proposed strategy of efficient sampling from a dataset could be thought of as a way of minimizing the impact of observational errors in the inferences drawn from the observed data. This is due to the fact that when we remove redundant observations from our data, we also remove the observational uncertainty and error associated with them. Based on this efficient segregation of information from Noise ,the proposed methodology could also be used as a building block for development of approaches for Error filtering. [2] used Neural Networks for filtering and compression of their biomedical dataset.

In the Literature on Finite Elements and other modeling domains, researching have also been trying to develop efficient procedures for capturing of discontinuities either arising from the class of analytical functions used in the analysis part or due to the sheer nature of the physical process begin modeled. In this regard it is relevant to refer to the Extended Finite Element method [6] which was based on enriching the solution space for the differential equation being modeled by including several discontinuous functions. The proposed approach can also be thought of efficient way to generate basis functions which can efficiently model these discontinuities by optimal exploitation of correlation structures inherent in the data at different scales and resolutions. The exact meaning of scales and resolution with reference to the proposed procedure would be discussed in detail in the following sections.

Besides testing the approach on a set of analytical functions [12], we have also analyzed the procedure on two sets of time series data for the gravity measurements [5] and the altimetry measurements [11] at critical locations on the surface of the greenland icesheet. This dataset is a crucial parameter for the performance of our algoirhtm because of the high fluctuations inherent in this data. Also, the application on this dataset clearly brings out the capability of our approach to remove unnecessary noise from the information bulit into the data.

The Contributions of this paper can be summarized as follows. Firstly we are introducing this concept of sparse representation which analyzes the inherent structure of the datset and removes redundant information which is beneficial for memory savings and efficient computations. Secondly, unlike previously, here we are using a composition of scales and resolutions which represents the levels and sublevels at which a dataset can be studied. Due to this capability we are even able to model discontinuities through efficient basis construction.

2 Multiscale/Multiresolution approach with Kernels

Multiscale and Multiresolution strategies are an effective tool for analysing the underlying structure of a dataset. It has been concluded by several researchers in the past that it is crucial to understand the underlying space in the dataset from where the the given data points may have been sampled. The approach discussed here exploits this idea of understanding the underlying covariance structure. In approximate terms, the idea is to test the extent to which

an observation at a point affects the observation at a different point at a test distance and then tuning this distance to explain the variation in the data. In our previous work [7], we proposed a distributed implementation of this approach so that we could handle the excessively large datasets and draw useful inferences from the given data. Before [7], [1] presented an application of the approach on some chosen test functions. However, their motive in that research was to use the procedure for sampling from a dataset at different scales and they did not explore the idea of having a sparse representation of a dataset. In our previous work we based our experiments on one particular type of kernel which had its properties derived from the exponential covariance structure. The idea was to have a structure where the effect of one point on another point in the considered space was inversely proportional to distance between them. Also the decay in the correlation strength was exponential. The expression for the covariance kernel is shown in equation 1. Here ϵ is the parameter which depends on the length scale of the covariance kernel. For more details on the approach the reader should refer to [7]. And for a detailed treatment on covariance kernel [8] should be referred.

$$g_{\epsilon}(r) = \exp\left(-\frac{r^2}{\epsilon}\right) \quad (1)$$

For the current research, we further analyzed this covariance criterion and tried to incorporate a more general covariance structure so that we can manipulate the degree of freedom for the variation of scales or the levels at which we analyze the data. In this research we have proposed the idea of using the rational quadratic kernel in order to map the covariance between the points. The rational quadratic covariance kernel makes sense for this approach as it allows a criterion for multiple scales at once in the analysis. This property of increased degrees of freedom for the covariance kernels has led us to incorporate this concept of multiple resolutions at each scale of the analysis.

$$k_{RQ}(r) = \left(1 + \frac{r^2}{2\alpha \epsilon^2}\right)^{-\alpha} \quad (2)$$

Through rational quadratic kernel our approach is able to study and model the data by studying the behavior of the data at different scales (variation of ϵ) and then adaptively varying the resolution at each scale in order to have a higher accuracy at critical points (for example at a discontinuity) in the domain. The expression for the rational quadratic kernel could be seen in equation 2. Here r represents the distance between two observation points through an appropriate metric and ϵ represents the length scale similar to the square exponential kernel shown in eq. 1. The term α here represents the second variable which also allows to vary the behavior of the covariance kernel. The presence of two scale parameters present in the kernel expression (ϵ and α) enables it to have a mixture of scales and resolutions for different points in the domain.

3 Sparse Representation

The idea of sparse representation introduced in this work is suppose to target the need of Data compression and optimal storage for a given large dataset. Data compression has been a well explored topic in the recent past. [4] discussed it with respect to image classification. [13] have implemented this concept in reference to classification problem. One thing to note here is that our approach is quite different from approaches like PCA where a higher dimensional dataset is compressed to a lower dimension with the new coordinates expected to preserve most of the

original information. In the current approach we are not reducing the size of the given data by compressing the dimensions, however our data compression is based on reducing the number of observations itself. Conceptually, the process aims to remove the observations from the dataset which have negligible information contribution given other observation points. Our Multiscale and Multiresolution based approach achieves this objective of data compression by creating a new data structure which is a compressed representation of the original dataset. This Data structure has 4 components

- Representative data points in the original dataset
- Scale ID
- Resolution ID
- Coordinates for the representative Basis functions

Here Representative data points refer to the selected observations from the dataset. These representative points have also been referred to as the sparse representation in the current work. Scale ID here represents the scale value inbuilt into ϵ as shown in [7]. Resolution ID here is α (when α is assumed to be constant for the whole domain at each scale). The case for variable resolution (α) has been discussed in the results section. The fourth component of this data structure refers to the weightage which needs to be given to the covariance values of the prediction points with respect to the representative points. Therefore for prediction at an unknown point, one just needs to find the covariance of this new point with respect to the representative points by using the scale id and the resolution id in the rational quadratic kernel and then doing a linear combination of these covariance values by using the coordinates as the weights.

4 Experimental setup

In this section we have discussed our entire experimentation setup for the demonstration of the capabilities of our current approach. For the first part of the analysis we have generated the sparse representations for some of the carefully chosen analytical functions which can bring out the crucial properties of our approach. Each of these test functions have several characteristics which make them representative of a big class of practical cases. Each of these test functions would be specifically dealt with in the following part of the current section.

As mentioned earlier, in order to demonstrate the application of our approach, we have also presented the results of the multi-scale and multi-resolution approach on two datasets obtained from physical domains. The first dataset consists of NASA GSFC solution for monthly estimates of time-variable gravity in units of cm equivalent water height (cm w.e.) for a global set of 41,168 1x1 arc-degree equal-area mass concentration blocks. The second dataset consists of Altimetry measurements over the surface of the icesheet at different locations over time.

4.1 Analytical Functions

In this section, we will briefly go over all the test functions [12] which we have used to demonstrate the performance of our approach. Firstly we will go over the 1 D functions followed by the 2 D test functions. Figure 1 shows the test functions used in this study to demonstrate the sparse representation generation capacity of our approach. As mentioned earlier, each of these test functions represent a particular type of complexity which may occur while generating

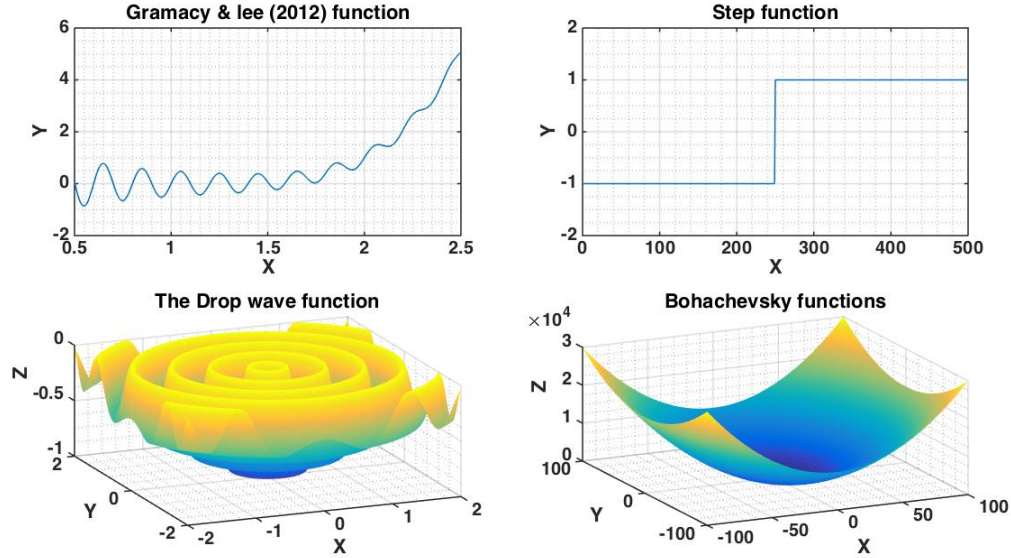


Figure 1: Test Functions

a compressed data representation of a function. Here compressed data representation refers to the procedure where we have some sampled points form an underlying function and then we carefully choose a minimal number of points which can explain the behavior pattern of the functions, thus minimizing the storage and computational resource requirement in the further analysis. The specific functions which we will tackle for this research are as follows:

- 1 dimensional function with global and local trends (Gramacy and Lee (2012) test function)

$$f(x) = \frac{\sin(10\pi x)}{2x} + (x-1)^4 \quad (3)$$

- 1 dimensional function with a discontinuity (Step function)

$$f(x) = \begin{cases} -1 & x < 250 \\ 1 & x \geq 250 \end{cases} \quad (4)$$

- 2 dimensional Multimodal function (The Drop Wave function)

$$f(x, y) = -\frac{1 + \cos(12\sqrt{x^2 + y^2})}{0.5(x^2 + y^2) + 2} \quad (5)$$

- 2 dimension unimodel function (Bohachevsky function)

$$f(x, y) = x^2 + 2y^2 - 0.3\cos(3\pi x) - 0.4\cos(4\pi y) + 0.7 \quad (6)$$

With these test functions we will be able to test if our approach is able to capture the different trends which may be inherent in the function at different resolutions. Gramacy and Lee function

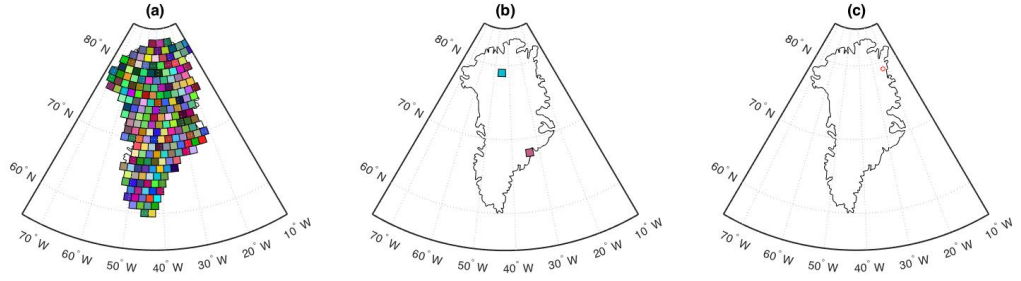


Figure 2: Gravity and Altimetry data a) Locations of all regions where Gravity time series is available b) The locations chosen for study c) Location of the coordinate (red circle in North East Greenland) for the time series of Altimetry measurements

tests the ability of the approach to capture global and local trends. The Step function will be able determine if our multiscale/multiresolution approach is capable of capturing sharp trends and variations in a function. The way in which our approach deals with the gibbs phenomena of ringing at the discontinuity is also interesting to see. The drop wave function represents the case of high frequency variation over the entire 2 dimensional domain. Therefore, the results in the next section would be important to determine if our approach is suitable for functions and datasets in higher dimensions. The Last function (Bohachevsky function) represents one of the simple cases for a 2 dimensional function. This function is supposed to show the level to which our approach could compress the data and how well our approach could generate a sparse representation for an underlying function.

4.2 Practical Application

The sparse representation generation for the mentioned analytical functions are meant to show if the algorithm is able to perform in an appropriate manner if the dataset is derived from a underlying function having a definite structure. Therefore, although these chosen functions are sufficiently complex based on their attributes, they lack the complexity level of a random dataset sampled from a physical domain because a practical dataset may even lack structure and may be even more difficult to model.

4.2.1 Gravity Dataset

Our first data is a gravity measurements dataset obtained from [5]. This dataset consists of 199 mascons (1x1 arc-degree equal-area mass concentration blocks or mascons) shown in the first figure in Figure 2. Each of these mascons have an associated time series of gravity measurements with them. This gravity data is in the form of monthly estimates of time-variable gravity in units of cm equivalent water height, which has been used as a proxy for the mass change measurements over the greenland icesheet. This data has been obtained after processing of the gravity measurements from the GRACE satellite mission. Since, this dataset consists of gravity measurements time series for each of the 199 mascons and each of these time series have a monthly resolution; Therefore, for the sparse representation in the time domain we have

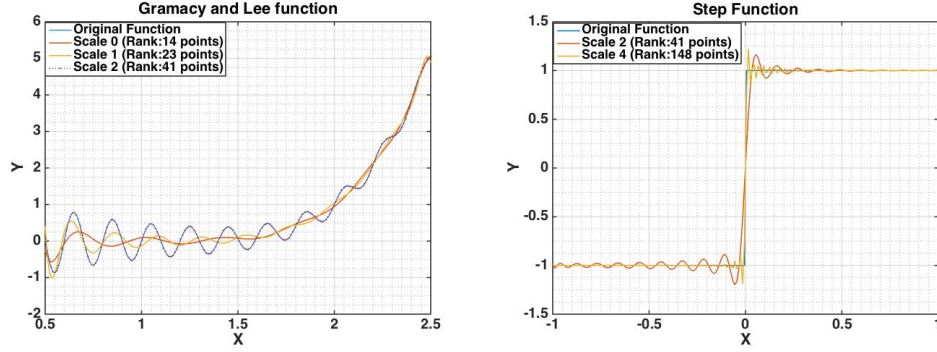


Figure 3: Prediction from sparse data (a):Gramacy and Lee function (b) Step function

chosen two particular mascons shown in the second figure in Figure 2 (Here the green Mascon would be referred to as M2 and the red Mascon as M1 henceforth). These specific mascons have been chosen because one of them is around the center of the icesheet and hence experiences less melting and therefore the gravity change dataset will not involve much fluctuation. However, the second mascon is expected to be associated with high fluctuations because it is closer to the edge of the icesheet and hence experiences a higher melting rate. In scientific terms, the mascon at the edge falls into an area known as the ablation zone and hence is a major contributor in the dynamic thinning of the icesheet.

4.2.2 Altimetry Dataset

Our second dataset consist of timeseries of altimetry measurements obtained from SERAC model [11]. These measurements refer to the height of the icesheet observed at different points in time at a particular location, thus making it a time series dataset. This dataset may be regarded a little more complex than the gravity dataset because here the sampling is not uniform in the time domain. For this study we have used the time series from the North Eastern side of Greenland shown as a red point in the third plot in Figure 2

5 Results

Following the testing procedure described in the previous section, here we will present and describe the results obtained after the implementation of our multiscale/multiresolution approach on the test function and the gravity data.

5.1 Constant Resolution (α)

For this first part, the resolution (α) is fixed at 0.5 and scale built into (ϵ) is varied during the analysis. Figure 3 shows the results for the first two test functions. For Gramacy and Lee function, it can be seen that as the scale is increased, ϵ is reduced and therefore the rank of the kernel is increased and thus more and more are chosen for the sparse representation. It is found that around 41 points are able to explain the variation in the function. The quantification of the accuracy would be presented in the following subsections.

For the step function, we employed a similiar procedure to find a sparse representation. However, it could be observed that even with 148 of the original 200 points, we can still see

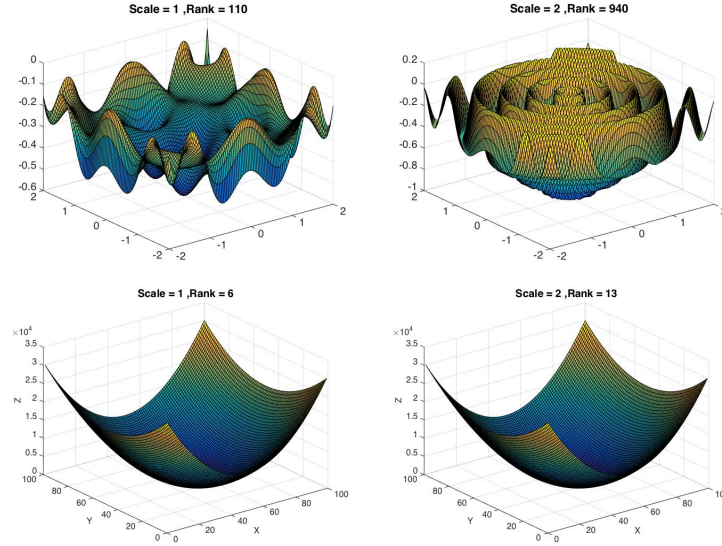


Figure 4: Prediction from sparse data: The drop Wave function (Top row) Bohachevsky function (Bottom row)

the ringing effect at the discontinuity. These fluctuation do not go away even if we use all the original points and construct the basis functions. Therefore for dealing with this we have used the concept of adaptive resolution discussed in the next subsection.

Figure 4 show the effect of variation of scales and predicting the underlying function from the sparse representation obtained at different scales for the Wave and Bohachevsky function. The top two figures show that when the number of critical points chosen are increased from 110 to 940 (as a result of increament in scale) for the drop wave function, the predicted function from this sparse representation becomes highly accurate as shown in Table 1. For the Bohachevsky function, the compression possible is very extreme and it able to reproduce a close approximate of the original function with just 13 points. Based on the behavior of these test function, the observation points chosen at strategic locations at differen scales have been shown in Figure 5. Here the blue points show the original points available and the chosen points for sparse representation are circled in red. For the 2 dimension functions, the points are shown in the XY plane instead of the 3 dimenion plane for better clarity

5.2 Adaptive Resolution

Due to Gibbs phenomenon, the prominent ringing effect in the modeling of the step function has motivated us to use the second degree of freedom in the covariance kernel. This is done by adaptively varying the resolution parameter over the entire domain at each scale. In order to do so, at each scale, the resolution at a point is defined as a function of its distance from the point of discontinuity. Therefore, one piece of information which goes into modeling of discontinuity through this method is that the point of disconituity should be approximately known. The precision with which the modeler knows the location of disconituity, the more accurate would be results. For the results shown in Figure 6, we used the values of resolution at a point as 0.5 times the distance from the discontinuity at each scale. Therefore, at very

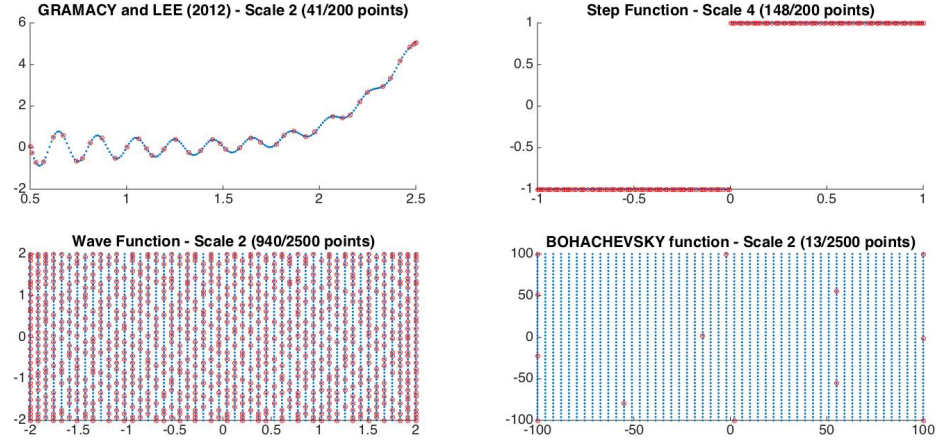
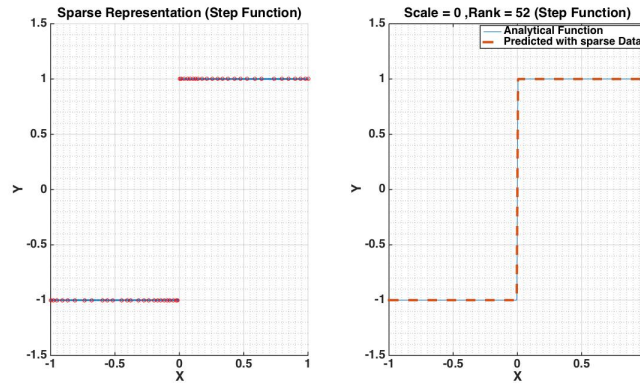


Figure 5: Sparse Representation for the test functions at Relevant scales

Figure 6: Sparse representation (Left) and Prediction (Right) for step function with variable α

close proximity to the discontinuity, α would be very small and hence, it will affect a very small region. Thus, this sort of formulation for the resolution ensures that in the limit towards the point of discontinuity, the region of impact of an observation becomes zero and therefore it doesn't cause any ringing or unnecessary fluctuations. This pattern of defining the sparse representation by using an adaptive combination of variation in scales and resolution also leads to a variation in the pattern in which points are chosen for sparse representation. It could be seen that unlike the top right plot in Figure 5, where points were uniformly chosen, here (left plot in Figure 6), more points are automatically chosen near the discontinuity, so that it can be accurately modeled. The result of prediction from the sparse representation has been shown in the right plot in Figure 6. The fact that ringing effect has almost completely disappeared with just 52 points justifies our use of an adaptive resolution approach.

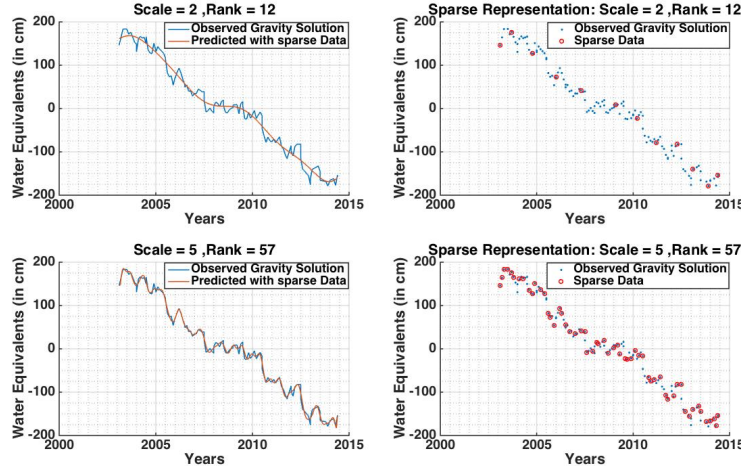


Figure 7: Gravity Dataset (Mascon M1)

Test Function	Number of originally sampled point	Scale	Number of Points in Sparse Rep.	RMSE of 20 random pred. points
Gram. & Lee	200	2	41	2.85e-05
Step (adaptive α)	200	0	52	6.10e-07
The Drop Wave	2500	2	940	5.83e-05
Bohachevsky	2500	2	13	7.6009

Table 1: Accuracy in terms of RMSE at 20 prediction points through the sparse representation

5.3 Accuracy Results

Table 1 shows the root mean square error of prediction at 20 random points chosen in the respective domains. It could be seen that at the scales discussed in Figures 3 and 4, the prediction accuracy is very good for all the test functions (evident from the small RMSE). Here the RMSE of the Bohachevsky function is found to be very high relatively. This is due to the nature of the function which produces values of the order 4 as shown in Figure 1.

5.4 Practical Application

For the gravity dataset, the sparse representation of the time series at Mascon M1 has been shown in Figure 7. Here, the right hand side plots show the points which were chosen for sparse representation at different scale. The corresponding left hand side plots show the prediction of the underlying function through these sparse datasets. It can be seen that as scale is increased the prediction is becoming accurate. Here, the original number of points available is 127 from which 57 points have been chosen at scale 5. This kind of sparse representation also helps as it may also help in noise removal as seen in the bottom left plot in Figure 7. Similar results have been obtained for Mascon M2 (shown in Figure 8).

For the altimetry dataset we have just shown it for one coordinate at the edge of the icesheet shown in the rightmost plot in Figure 2. Here also we have obtained similar results where we are able to generate the sparse representation from the original dataset (this time series contains

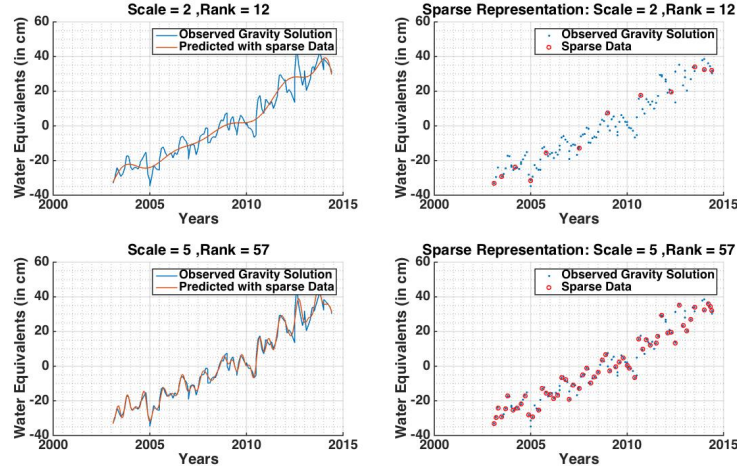


Figure 8: Gravity Dataset (Mascon M2)

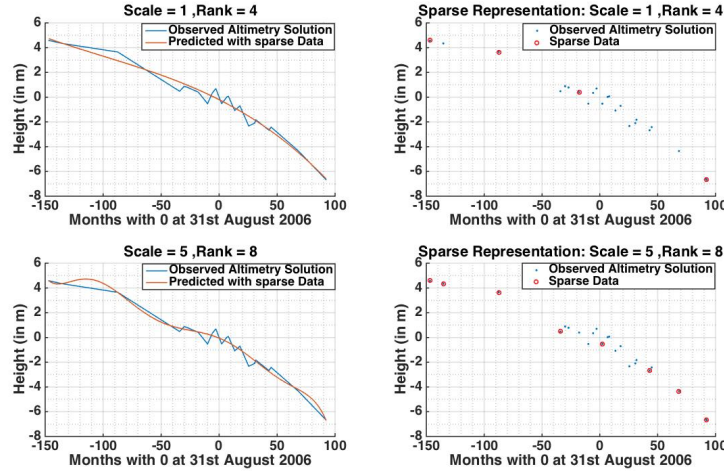


Figure 9: Altimetry Dataset

23 time instances). Here it could be seen that with around 8 points at scale 5 the algorithm is able to capture most of the variability in the dataset. However, at the starting time instances, due to very sparse sampling, the fit is not very accurate as there is clear lack of information during the starting months.

6 Concluding Remark

In this research we proposed a strategy in which we can compress a dataset by removing all the unnecessary observations and retaining only the valuable observation instances. The approach as discussed in our previous work [7] is based on a QR matrix decomposition procedure

which identifies the observation instances with maximum amount of information (through the decomposition of the covariance kernel) and uses them along with the kernel specified as a sparse representation for the original dataset. Then this sparse representation could be used to predict the original function at any time as required. The application of the approach has been shown on 4 analytical functions along with two real datasets from the geophysical domain.

The approach is shown to be highly useful in a variety of settings. However, there are still several ways in which it could be improved for making the implementation more practical. Firstly, although we introduced a distributed implementation of the approach in [7], still improvements have to be made there for better scaling. Due to the vast availability of literature in Kernel methods, there may be some way to capture the covariance between the data points in an even better way. Some Future work can also be done in extending the approach to the spatio-temporal modeling domain.

References

- [1] Amit Bermanis, Amir Averbuch, and Ronald R Coifman. Multiscale data sampling and function extension. *Applied and Computational Harmonic Analysis*, 34(1):15–29, 2013.
- [2] Kalyan Chatterjee, Nilotpal Mrinal, and S Dasgupta. Adaptive filtering and compression of bio-medical signals using neural networks. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2(3):323–327, 2013.
- [3] Yawen Guan, Murali Haran, and David Pollard. Inferring ice thickness from a glacier dynamics model and multiple surface datasets. *arXiv preprint arXiv:1612.01454*, 2016.
- [4] Bao-Di Liu, Yu-Xiong Wang, Bin Shen, Yu-Jin Zhang, and Martial Hebert. Self-explanatory sparse representation for image classification. In *European Conference on Computer Vision*, pages 600–616. Springer, 2014.
- [5] Scott B Luthcke, TJ Sabaka, BD Loomis, AA Arendt, JJ McCarthy, and J Camp. Antarctica, greenland and gulf of alaska land-ice evolution from an iterated grace global mascon solution. *Journal of Glaciology*, 59(216):613–631, 2013.
- [6] Nicolas Mos, John Dolbow, and Ted Belytschko. A finite element method for crack growth without remeshing. *International Journal for Numerical Methods in Engineering*, 46(1):131–150, 1999.
- [7] Abani Patra, Prashant Shekhar, and E Ramona Stefanescu. Multilevel methods for sparse representation of topographical data. *Procedia Computer Science*, 80:887–896, 2016.
- [8] Carl Edward Rasmussen. Gaussian processes for machine learning. 2006.
- [9] Awwal Mohammed Rufai, Gholamreza Anbarjafari, and Hasan Demirel. Lossy image compression using singular value decomposition and wavelet difference reduction. *Digital Signal Processing*, 24:117–123, 2014.
- [10] Aliaksei Sandryhaila and Jose MF Moura. Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure. *IEEE Signal Processing Magazine*, 31(5):80–90, 2014.
- [11] Tony Schenk and Beata Csatho. A new methodology for detecting ice sheet surface elevation changes from laser altimetry data. *IEEE Transactions on Geoscience and Remote Sensing*, 50(9):3302–3316, 2012.
- [12] S. Surjanovic and D. Bingham. Virtual library of simulation experiments: Test functions and datasets. Retrieved January 8, 2017, from <http://www.sfu.ca/~ssurjano>.
- [13] Jianchao Yang, Jiangping Wang, and Thomas Huang. Learning the sparse representation for classification. In *2011 IEEE International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2011.