A Probabilistic Approach To Determine Sparsity Patterns In Large Matrices

Shinjini Nandi

Hosting Site: Argonne National Laboratory, IL

Mentors: Paul Hovland, Sri Hari Krishna Narayanan

August 4, 2017

Abstract

The aim of this project is to investigate and improve on existing probabilistic methods to determine the sparsity patterns in matrices. The underlying concept involves sequentially updating the conditional probabilities associated with non-zero elements in the sparse matrix, subject to minimization of expected number of discrepancies between actual and estimated sparse matrix. Single probing methods based on this idea seek to sequentially determine a minimal number of vectors that identify the sparsity pattern. The batch probing approach arbitrates a set of vector probes to be used simultaneously and determine the sparsity pattern using a minimum number of such sets. We analyze the methods suggested by Griewank and Mitev (2002) and suggest some improvements in choices of the initial probability matrix and in the batch probing algorithm that determines the sparsity pattern using fewer number of probes.

1 My Internship Project

1.1 Introduction

Computation of Jacobian matrices is a common requirement in many optimization and numerical problems. Often these Jacobian matrices are sparse with few non-zero elements. It is advantageous to determine the sparsity pattern of these matrices for their efficient use in subsequent numerical calculations.Griewank and Mitev (2002) [2] provided a discussion on the sparsity detection problem using a probabilistic approach. The algorithm uses a probabilistic model for the sparsity pattern. For a given probing vector, information is obtained from the sparse matrix indirectly through multiplication of the matrix to the vector. The authors of this paper describe methods that use single probes and batches of probing vectors to estimate the sparse matrix applying as few matrix-vector products as possible. In this project, we discuss these probabilistic methods and extend on their work. We suggest some improvements that further reduce the number of iterations required to achieve our goal.

We implemented the methods in MATLAB $(2017a$, The MathWorks) [3]. The codes to reproduce all analyses in this paper are available at `https://github.com/shinjini01/Bayesian-Probing.git`. This work was presented in Summer Argonne Students' Symposium, 2017, Part II, conducted by LANS, Mathematics and Computer Science Division, Argonne National Laboratory.

1.1.1 Theoretical Background of the Probabilistic approach

The probing procedure described in [2] does not consider the actual sparse matrix. From the actual matrix, a Boolean representation S is obtained which is utilized for further calculations. Suppose

$$S = ((s_{ij})),\ i = 1, \ldots, m,\ j = 1, \ldots, n \tag{1}$$

be the given sparse representative matrix of order $m \times n$, where $s_{ij} = 1$ implies the (i, j)th element in the actual matrix is non-zero. Each $s_{ij} \sim \text{Bin}(1, p_{ij})$, independently $\forall i, j$. Obviously,

$$p_{ij} = \text{P}(s_{ij} = 1) = 1 \Rightarrow s_{ij} = 1, \text{for any } i, j. \tag{2}$$

Initially S is estimated conservatively by a completely dense matrix $\widehat{S}$. The estimate of the associated probability matrix $\widehat{P}$ is updated iteratively subject to minimization of an error measure, until $\hat{p}_{ij} = 0$ or $1,\ \forall\, i, j$. Equation 2 ensures $\widehat{P} = S$ and the sparsity pattern is completely determined using a probabilistic model.
At the kth iterative step, the estimated sparse matrix is denoted by $\widehat{S}_k$ and the corresponding probability matrix as P_k. The error measure is defined as the expected total number of zeros not discovered by the estimate $\widehat{S}_k$.

$$\begin{aligned} U &= \text{E}(\sum_{i=1}^{m} \sum_{j=1}^{n} \hat{s}_{ij}^{k} - s_{ij} \mid s_{ij} \in \Omega_k) \\ &= \sum_{p_{ij}^{k} > 0} (1 - p_{ij}^{k}) \end{aligned} \tag{3}$$

The matrix S is not directly used to determine the sparsity pattern. Information is obtained indirectly from it through the matrix vector multiplication result $r = St$.

1.1.2 Criterion to select a probing vector

Suppose a random probe on S results in N mutually exclusive results $r^{(n)}$, $n = 1, \ldots, N$. The expected reduction in the error measure U is shown to be

$$\Delta U = \sum_{n=1}^{N} p^{*(n)} \mid \{(i, j) : p_{ij}^{(n)} = 0 \neq p_{ij}\} \mid \tag{4}$$

where $p^{*(n)} = $ P(the random probe on S has result $r^{(n)}$), $p_{ij} = \mathrm{P}(s_{ij} = 1)$ and the conditional probability

$$p_{ij}^{(n)} = \mathrm{P}(\{s_{ij} = 1\} \mid \text{the probe resulted in } r^{(n)}) \tag{5}$$

Since minimization of the error measure is equivalent to maximizing the reduction in it due to a random probe, the algorithm concentrates on the expected reduction in the error measure due to a random probe. A single probe vector $t \in \{0,1\}^n$ applied to the ith row of S is shown to obtain the following expected reduction according to equation 4

$$\Delta U_i = \prod_{j \in \tau_i \; p_{ij} > 0} (1 - p_{ij}) \mid \tau_i \mid \text{ where } \tau_i \text{ is the support of } t, \text{such that } p_{ij} > 0, \; i = 1, \ldots, m \tag{6}$$

Since s_{ij} are independently distributed for all $i, \; j$, it follows that the rows of S are independent and the total expected reduction in error for the matrix S is the sum of all row-wise expected reductions

$$\Delta U = \sum_{i=1}^{m} \Delta U_i \tag{7}$$

1.1.3 Updating the probability matrix

The set of indexes of the columns that maximize ΔU comprise the support of the vector probe t and is denoted by τ. Using t, we obtain $r = St$.
Consider the following events in the ith row of P for the probe vector $t^{n \times 1}$ and $r_i = s_i^T t$.
For a particular $j^* \in \{1, \ldots, n\}$, if the j^*th element of t, i.e., $t_{j^*} = 0, $ then $s_{ij^*} = 0$ or 1, independent of $r_i = \sum_{j=1}^{n} s_{ij} t_j = 0$ or 1. Hence $\mathrm{P}(\{s_{ij^*} = 1\} \mid t_{j^*} = 0, r_i) = \mathrm{P}(\{s_{ij^*} = 1\})$.

If $t_{j^*} = 1$ and $r_i = 0 \Rightarrow s_{ij^*} = 0$ with probability 1. If $t_{j^*} = 1$ and $r_i = 1$, then $s_{ij^*} t_{j^*} = 1 \Rightarrow s_{ij^*} = 1$. This information is used in the following specific form of the conditional probability in equation 5.

$$p_{ij}^{k+1} = \mathrm{P}(\{s_{ij} = 1\} \mid t_j, r_i) = \begin{cases} p_{ij}^k, & \text{if } t_j = 0 \\ 0, & \text{if } t_j = 1,\ r_i = 0 \\ \frac{\mathrm{P}(\{s_{ij}=1\})}{\mathrm{P}(t_j=1,\, r_i=1)} = \frac{p_{ij}^k}{1 - \prod\limits_{j \in \tau} (1 - p_{ij}^k)}, & \text{if } t_j = 1,\ r_i = 0 \end{cases}$$
$$\forall\ k \geq 2. \tag{8}$$

Here we use the fact that $\mathrm{P}(t_j = 1,\ r_i = 1) = \mathrm{P}(\text{At least one } s_{ij} = 1) = 1 - \prod\limits_{j \in \tau} (1 - p_{ij})$. Equation 8 gives us the $(k+1)$th update of the (i, j)th element in the probability matrix.

The probing method starts from the specified initial probability matrix and iteratively finds the best probe and updates the probability matrix till we get $U = 0$ or $\widehat{P} = S$.

1.2 The single probing method

The single probing method as proposed by Griewank and Mitev [2] is a sequential procedure to determine probe vectors and update the probability matrix accordingly. Starting from an user specified estimate of the probability matrix, the process iterates till the sparsity pattern is completely determined. Description of the algorithmm is given below.

1: Steps to implement Single Probing Method

1. The probability matrix $\widehat{P}$ is initialized subject to equation 2.

2. The set of columns $\tau =$ maximizing ΔU is chosen as the first probe. In order to determine the support τ of the probe, the following steps are used

   (a) Choose a column that maximizes expected reduction in error. If each column is considered as an $m \times 1$ matrix, then using equations 6 and 7 we have $\Delta U = \max_j \{\sum_{i=1}^{m}(1 - p_{ij})\}$. The first element of τ is $\arg\max_j \{\sum_{i=1}^{m}(1 - p_{ij})\}$.

   (b) The second column index or subsequent column indexes are added to the subset τ if they succeed in increasing the ΔU value.
   Observe that τ and $\{1, \ldots n\} \setminus \tau$ are mutually exclusive. Hence for $j^* \in \{1, \ldots n\} \setminus \tau$, deriving from equations 6 and 7 we have

   $$\Delta U(P, \tau \cup j^*) = \sum_i \{ \prod_{j \in \tau_i \; p_{ij} > 0} (1 - p_{ij})(1 - p_{ij^*}) \} \{ \mid \tau_i \mid + I(p_{ij^*} > 0) \}$$

   where $I(p_{ij^*} > 0) = \begin{cases} 1 & \text{if } p_{ij^*} > 0 \\ 0, & \text{otherwise} \end{cases}$
   If $\Delta U(P, \tau \cup j^*) > \Delta U(P, \tau)$, j^* is added to τ.

   (c) Step (b) is executed for all elements of the set of column indexes $\{1, \ldots n\} \setminus \tau$ till all column indexes are tested for and possibly included in τ.

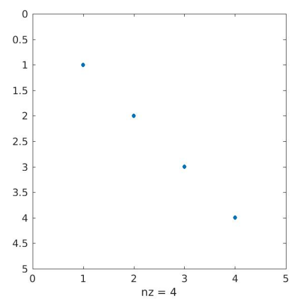
   (d) Once τ is determined, it is used to define the probe vector t.

3. Using the result vector $r = St$, update the probability matrix $\widehat{P}$ according to equation 8.

4. The process is repeated until we obtain $U = 0$, i.e., $\widehat{P} = S$.

1.2.1 Illustration of the Single Probing Method

We consider a 4×4 diagonal matrix as the sparse matrix for illustration.



Correspondingly, we have

$$S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \ \widehat{P} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix},$$

Here we have $m, n = 4$ and we use the naive initialization of $p_{ij} = 1/n, \forall\ i, j$ as the first estimate of the P matrix. The transition of the probability matrix from this initial estimate to convergence to the matrix S using different probes and result vectors is shown in figure 4.

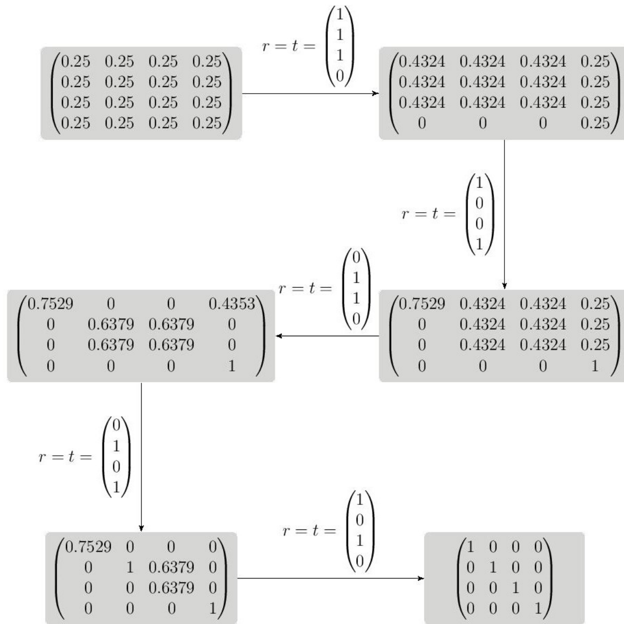


Figure 1: Stepwise illustration of the single probing method to determine the sparsity pattern in an 4×4 diagonal matrix.

1.3 The Batch Probing method

1.3.1 The batch probing method suggested by Griewank and Mitev

The batch probing method suggested in [2] is an extension of the single probing method. In this procedure, a batch of K probe vectors is selected that when applied to the estimated probability matrix of size $m \times n$ sequentially, maximizes the expected reduction in error. The probes explore the matrix in parallel to ascertain the sparsity pattern in a fewer number of moves than the single probing method. However, to curb complexities in calculations of the error measure, it is necessary to keep the probes mutually disjoint. This is ensured by eliminating the column indexes included in any probe from further consideration for inclusion in other probes. Each probe is restricted to include a maximum of n/K column indexes so that all K probes investigate at least one column.

2: Steps to implement Batch Probing Method

---

1. **Selection of K probes in a batch.** Selection of the first probe is based on columns that maximize ΔU, restricted to at most n/K columns.The concept of probe selection from the single probing method described in the previous section is applied here.The next and subsequent probes are determined similarly subject to maximizing ΔU from the remaining columns.

2. **Updating the probability matrix.** Once the batch is determined, the probability matrix is updated cumulatively for each of the K probes and K result vectors $r_k = St_k,\ k = 1, \ldots, K$, using equation 8.

3. The updated matrix $\widehat{P}$ is again used to select the second batch of probes and update the probability matrix and reiterate the steps till the total error is zero, i.e., $\widehat{P} = S$.

1.3.2 Illustration of the batch probing method

We consider an 8×8 diagonal matrix for illustration of the method using batches of 2 probes. Corresponding to the sparse matrix, we have

$$S^{8\times 8} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

and the initial estimate of the probability matrix

$$\widehat{P}^{8\times 8} = \begin{pmatrix} 0.125 & \cdots & 0.125 \\ \vdots & \ddots & \vdots \\ 0.125 & \cdots & 0.125 \end{pmatrix}$$

The process takes 4 batches of 2 probes to converge. This is illustrated in figure 2. From an estimate of the probability matrix P, a batch of two probes T is determined, which, along with the result matrix $R = S.T$ to update the probability matrix till the entire pattern is determined.

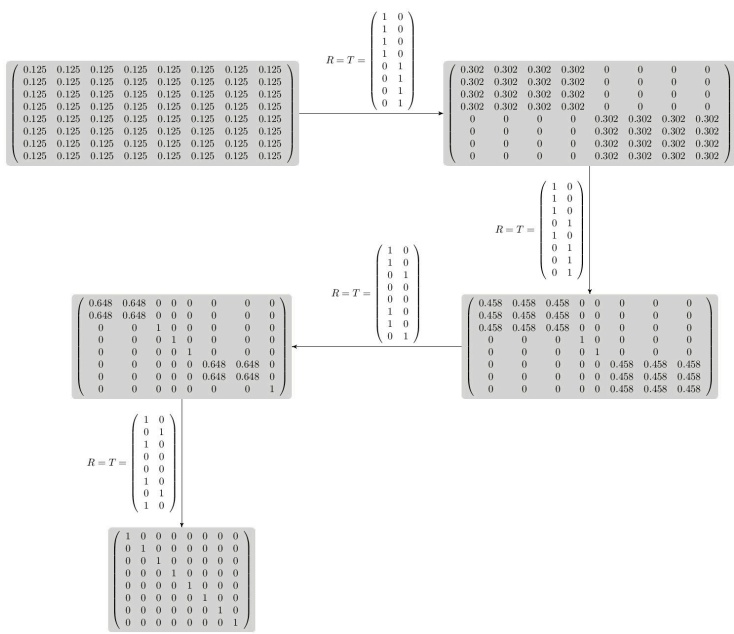


Figure 2: Stepwise illustration of the batch probing method proposed by Griewank and Mitev to determine the sparsity pattern in an 8×8 diagonal matrix.

1.3.3 Proposed batch probing method

We consider a variation in the selection procedure of a batch described in the method suggested by Griewank and Mitev. Instead of determining the probes sequentially, we propose to identify them simultaneously. This has given better performance than the sequential method at least for some common sparse patterns.

3: Steps to implement Proposed Batch Probing Method

---

1. The set of column indexes of matrix P are sorted in descending order of ΔU. The first K of these sorted indexes are distributed over the K probes as the first elements in each of $\tau_k, k = 1, \ldots, K$, τ_k being the support of the kth probe.
   The remaining columns are searched to find the index that maximizes ΔU for the first probe and so on. The process continues till all columns in P are exhausted and a batch of K probes is obtained.

2. The probability matrix is updated cumulatively.

3. The iterative process continues to find other batches of probes and update the probability matrix till the entire sparsity pattern is determined.

1.3.4 Illustration of the proposed batch probing method

To illustrate the working of our proposed method, we use the same matrix as in section 1.3.2. Beginning from the same estimated probability matrix,

$$\widehat{P} = \begin{pmatrix} 0.125 & \cdots & 0.125 \\ \vdots & \ddots & \vdots \\ 0.125 & \cdots & 0.125 \end{pmatrix}$$

the process is shown in figure 3. Unlike the former method discussed, this method takes fewer steps to converge.

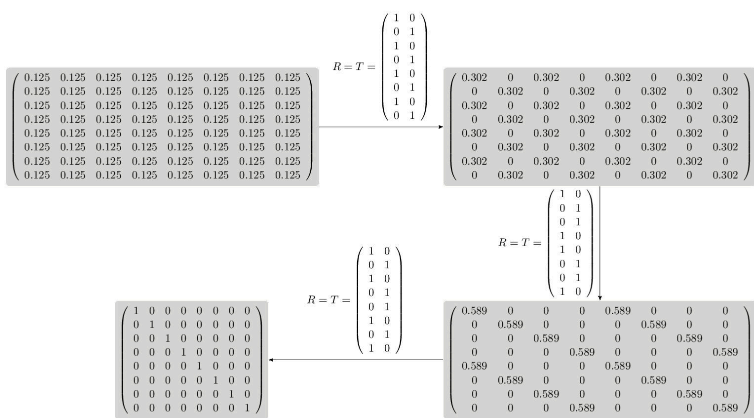


Figure 3: Stepwise illustration of the proposed batch probing method to determine the sparsity pattern in an 8×8 diagonal matrix.

1.4 Discussion on the naive initialization of the probability matrix

Since no specific information is available about the sparsity pattern, the authors in [2] suggested an uninformative uniform initialization of the estimate of the probability matrix $\widehat{P}$ and that is $\hat{p}_{ij} = 1/\min(m, n)$, $\widehat{P}$ being of order $m \times n$. We examined the performance of the single probing method with some naive initializations and initializations with some information about the row-wise sparsity patterns of the $\widehat{P}$ matrix. We applied these alternatives on several sparse matrices of different sizes and different sparsity patterns. We considered a diagonal matrix of size 100×100, an arrow head matrix of size 100×100, a 96×96 banded matrix obtained from Symmetric Pattern from Cannes, Lucien Marro, June 1981, which also belongs to the Harwell Boeing [1] group and a 32×32 Harwell Boeing [1] matrix. In general, we observed that even usage of minimal information about the sparsity pattern made the probing procedure more efficient. For the choice of initial probabilities using the average number of non-zero elements per row, we can determine the sparsity pattern using a minimum number of probes.

The following table summarizes the number of probes required to determine the sparsity pattern fro different choices of the initial probability matrix. For a choice of the probability matrix based on random samples, we considered samples of sizes 1000 or 500. The number of probes required is found out in 100 cases and the average number of probes is tabulated.

Table 1: Table showing number of iterations required according to the choice of the initial probability matrix.

Choice of P \ Sparse Matrix	Diagonal 100×100, 1%	Arrow head 100×100, 2.98%	Banded 96×96, 8.33%	Random HB 32×32, 12.3%
$p_{ij} = 1/n$	18	235	91	58
$p_{ij} = \frac{\text{number of non-zeros in row } i}{n}$	18	31	77	44
$p_{ij} =$ r.s. from $U(\frac{1}{n}, 1 - \frac{1}{n})$	65	178	93	40
$p_{ij} = \frac{1}{n}\min(nz_i, nz_j)$, nz_i, nz_j are r.s. from $(1, 2\log n)$	18	231	93	49
$p_{ij} = \bar{x}$, based on 500 samples from $X \sim \text{beta}(1 + 1/n, 1 - 1/n)$	100	100	96	32
$p_{ij} = \frac{\bar{x}_1 + \bar{x}_2}{2}$, based on 1000 r.s., $X_1 \sim \text{beta}(1 + 1/n, 1 - 1/n)$, $X_2 \sim \text{beta}(1/n, 2 - 1/n)$	37	190	79	41

1.5 Comparison of the Batch Probing methods

We applied the Griewank-Mitev's batch probing method and our proposed batch probing method to several sparse matrices of different sizes. Comparatively, our proposed method performs better than the former method for all regular sparse matrices. This is shown in figure 4.

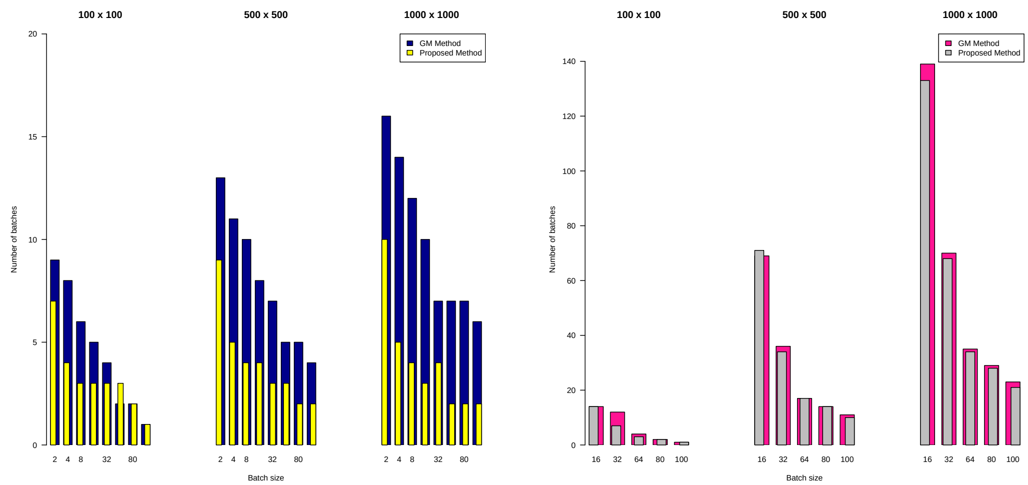


Figure 4: Comparison of the batch probing method proposed by Griewank amd Mitev (GM method) with our proposed method, Left panel shows the comparisons for diagonal matrices of orders 100, 500, 1000. Batches of sizes $2, 4, 8, 16, 32, 64, 80$ and 100 are used for diagonal matrices.Right panel is for arrow head matrices of same orders. Batches of size $16, 32, 64, 80$ and 100 are used for the arrow headed matrix

1.6 Conclusion and Future Work

In this project we elaborated on the characteristics of the sequential probing algorithm described by Griewank and Mitev [2] and implemented it to determine sparsity of several sparse matrices. There remains several aspects of the algorithm that can be further investigated and improved upon. Besides the simple single probing method discussed, the authors in [2] also discuss a combined process of probing based on the given matrix and those based on the transposed matrix. This combined process is faster than the simple probing method, in case of matrices with at least one dense row or column. In future, we would like to look into this method and consider extending its application to the batch probing method.

Our proposed batch probing method is more efficient than the one proposed by Griewank and Mitev for detecting some regular patterns of sparsity. However, we cannot claim its better performance for any random sparse matrix. We need delve deeper to possibly improve its performance and estimate its convergence rate for any kind of sparse matrix.
The choice of the initial probability matrix goes a long way to influence the convergence of any of these probabilistic probing methods. The naive initialization prescribed, or the assumption of a completely dense estimate of the matrix under inspection are very conservative. The algorithm can be improved with more radical assumptions about initial estimates as we have shown. Further research can be done to find the best possible initial estimates using minimal information from the sparse matrix. We would also like to consider batch probing methods that implement multiple initializations of the probability matrix aiming at faster convergence.

2 Impact of Internship on My Career

It was a rare opportunity and a very special experience to work at Argonne National Laboratory. Argonne has an excellent research environment conducive to learning and further development of science and technology. There were many seminars and discussions conducted by various divisions of the Laboratory that welcomed students and young scientists from all disciplines to participate and know more about the cutting edge research they pursue. I greatly benefited from the seminars conducted by the Mathematics and Computer Science division. I learned of various kinds of research conducted in Computer Science, Machine Learning and several branches of applied Mathematics that include large scale computation, automatic differentiation, various optimization techniques, to name a few. They have very significant applications that range from improving present computational techniques to developing better climate models to better predict climatic changes. This internship opened the doors to a world of highly interesting research areas which was previously unfamiliar to me. I hope to continue to communicate with the scientists and experts I came across and collaborate and share ideas on future research projects.

3 Acknowledgements

I am thankful to the NSF Mathematical Sciences Summer Internship Program sponsored by the National Science Foundation, Division of Mathematical Sciences (DMS), for taking the initiative to conduct this internship program. I am also thankful to Argonne National Laboratory, Mathematics and Computer Science Division for cordially hosting me as a summer intern. I am indebted to my supervisors Dr. Paul Hovland and Dr. Sri Hari Krishna Narayanan for their valued advice and support with my project. Lastly, I would like to acknowledge my new friends and everyone I came across who helped make this a very enjoyable and rewarding experience.

References

[1] I. S. Duff, Roger G. Grimes, and John G. Lewis. Sparse matrix test problems. *ACM Trans. Math. Softw.*, 15(1):1–14, March 1989.

[2] Andreas Griewank and Christo Mitev. Detecting jacobian sparsity patterns by bayesian probing. *Mathematical Programming*, 93(1):1–25, Jun 2002.

[3] MATLAB. *version 9.2.0.556344 (R2017a)*. The MathWorks Inc., Natick, Massachusetts, 2017.