A Maximum-Likelihood Approach to Localization from Mere Connectivity Information

Aditya Srinivas Timmaraju
with Prof. Andrea Montanari

EE 391 Report, Fall 2013
Department of Electrical Engineering
Stanford University

adityast@stanford.edu

December 16, 2013

Abstract

In the past decade, there has been a lot of work concerning the localization problem. It arises in many contexts ranging from wireless sensor networks to social networks. In this report, we present a Maximum-Likelihood approach towards solving the localization problem, given only the connectivity information between the comprising nodes. Though an explicit closed-form solution is not obtained, we present a gradient-descent based approach. We also describe the Fisher Information Matrix (FIM) for this estimation problem and characterize its singularity.

1 Introduction

In the past decade, there has been a lot of work around solving the localization problem [1-4]. This is in part due to the unprecedented surge in the application of wireless sensor networks (WSNs) in a wide range of fields. In a lot of WSN applications, particularly when used in monitoring, there is a need for knowing the location of a sensor node, without knowing which the data obtained from it will be irrelevant. There are other contexts in which the localization arises naturally, viz., locating people who are in a social network, given only the information about whether or not they are connected to each other on the social network.

Broadly speaking, there are two variants in localization problems. One of them is known as the "range-based" class and the other is known as the "range-free" class. In the "range-based" class, we are required to estimate the positions of nodes, given information about the absolute point-to-point distances between them. In the "range-free" class, there is no assumption about the availability of such data and mere connectivity information between the constituent nodes is provided. In the first of above illustrated examples (WSNs), it is possible to obtain distance measurements between constituent nodes and so, "range-based" algorithms can be applied. However, in the latter example (social networks), it is more natural to use the connectivity information between people for localization; so, "range-free" class of algorithms are more applicable. Drawing motivation from this, in this report, we concentrate on the "range-free" class.

It should be noted that the locations of each of the nodes can be estimated only upto a rigid transformation of the actual locations. This is because the relative information (distance) between the nodes remains invariant under a rigid transformation. Due to this reason, for complete location estimation, there is a need for the presence of nodes referred to as "anchors", whose locations are pre-determined.

The remainder of the report is organized as follows. Section 2 briefly mentions prior work on the localization problem. Section 3 formally introduces the problem description and elucidates our approach towards localization using Maximum-likelihood Estimation (MLE). In Section 4, we illustrate the results obtained from MLE localization. Section 5 concludes the report.

2 Prior Work

In [1], bounds are characterized on the performance of the MDS-MAP algorithm. MDS-MAP is a "range-free" algorithm that essentially uses the Djikstra's shortest paths algorithm followed by the MDS (Multi-dimensional Scaling) algorithm for location estimation. In [2], an SDP-based algorithm (for the "range-based" class) is introduced and performance bounds are characterized. In [3], the authors adopt an SDP-relaxation based method for localization.

In [4], the authors use a maximum-likelihood approach, but they limit their estimation to relative-location estimation, and not absolute location of each of the nodes. In contrast, the current work addresses the problem of finding the locations of each node (upto a rigid transformation). Apart from [4], there has not been much exploration into applying the maximum-likelihood approach to the localization problem.

3 Maximum Likelihood Approach

Given a set of "n" nodes (distributed uniformly at random) in $\mathbb{R}^d$, the range-free localization problem requires to estimate the locations of each of the points, using a set of pairwise measurements Y_{ij} with $(i, j) \in \{1, 2...n\} \times \{1, 2...n\}$. We use the following noisy model for the p.d.f. of Y_{ij} (Bernoulli random variable). Let $X_1, X_2...X_n$ be the 'n' nodes, whose locations are to be estimated. Then, we model

$$P(Y_{ij} = 1) = \frac{1}{(1 + exp(\beta(\|X_i - X_j\| - k_0))} = \phi(say)$$

where β and k_0 are real scalars.

Based on the observations Y_{ij}, we wish to estimate the location parameter vector given as $(X_1, ..., X_n)$. It is also a standard procedure to describe them as a matrix $X \in \mathbb{R}^{n \times d}$, whose rows comprise of the individual node locations. The likelihood of the observations (assuming independence between each observation) conditioned on the parameters can be expressed as,

$$P(Y/X) = \prod_{i,j} P(Y_{ij}/X) = \prod_{i,j} \phi^{Y_{ij}} * (1 - \phi)^{1-Y_{ij}}$$

We wish to maximize the likelihood function, which is equivalent to minimizing the negative of the log-likelihood function (say J). We can express J as (after eliminating terms independent of X_i's)

$$J = \beta * \sum_{i,j} Y_{ij} * \|X_i - X_j\| + \sum_{i,j} log(1 + exp(-\beta(\|X_i - X_j\| - k_0))$$

Since we need to minimize J, we first take its derivative with respect to each of the parameters.

$$\frac{\partial J}{\partial X_w} = \beta * \sum_{i=1, i\neq w}^{n} (Y_{iw}+Y_{wi}) * \frac{X_w - X_i}{\|X_w - X_i\|} - 2*\beta \sum_{i=1, i\neq w}^{n} \frac{1}{1+exp(\beta(\|X_w - X_i\| - k_0)} * \frac{X_w - X_i}{\|X_w - X_i\|} \quad (1)$$

for $w = 1, 2, ..n$.

Notice that our analysis allows for Y_{ij} and Y_{ji} to be potentially different. Intuitively speaking, if one of these observations is '0' and the other is '1', then there is a very good chance that the value $\|X_i - X_j\|$ is around k_0. This is due to the nature of the sigmoid function, which arises from the chosen Bernoulli distribution.

Since there is no known closed-form solution to the parameters X_i in (1), we employ the Gradient-Descent method to estimate the parameter vector $(X_1, ..., X_n)$.

3.1 Initialization of Gradient Descent

The key issue in using gradient descent is the question concerning the initialization of the parameters $X_1, ..., X_n$. We now propose the following simple algorithm to initialize the matrix X_{init}.

input : Observations Y_{ij} for $(i, j) \in \{1, 2...n\}$x$\{1, 2...n\}$, k_0
output: Initializations X_i for $i = 1, ..., n$, represented as a matrix X_{init}

```
for i ← 1 to n do
    for j ← 1 to n do
        if XOR(Y_ij, Y_ji) = 1 then
            R(i, j) ← k_0;
            R(j, i) ← k_0;
        end
        if AND(Y_ij, Y_ji) = 1 then
            R(i, j) ← k_0/2;
            R(j, i) ← k_0/2;
        end
        if AND(Y_ij, Y_ji) = 0 then
            R(i, j) ← 1;
            R(j, i) ← 1;
        end
    end
end
X_init = MDS(R);
```

Algorithm 1: Initialization method for Gradient Descent

The above algorithm can be derived by regarding $d_{ij} = \|x_i - X_j\|$ as the parameter to be estimated and taking the derivative of J w.r.t. d_{ij}. By following this approach, if both Y_{ij} & Y_{ji} are '1', we get d_{ij} to be '0' and if both Y_{ij} & Y_{ji} are '0', we get d_{ij} to be '0'. If exactly one of Y_{ij} or Y_{ji} is '1', we get $d_{ij} = k_0$.

3.2 Fisher Information Matrix

Denote our parameter vector as $\theta = (X_1, ..., X_n)$. Then the Fisher Information Matrix (FIM) is a matrix $F \in \mathbb{R}^{nd \times nd}$. We can view F as an $n \times n$ block matrix, with each block of size $d \times d$. Then, the $(w, k)^{th}$ block in the matrix F can be expressed as,

$$F_{wk} = E_Y(\frac{\partial}{\partial X_w}(\frac{\partial J}{\partial X_k}))$$

(since J is the negative of log-likelihood function)

After simplification, we arrive at the following,

$$F_{wk} = -\frac{\beta^2}{1 + cosh(\beta(\|X_w - X_k\| - k_0))} \times \frac{(X_w - X_k) * (X_w - X_k)^T}{\|X_w - X_k\|^2}; w \neq k$$

$$F_{ww} = \sum_{i=1, i\neq w}^{n} \frac{\beta^2}{1 + cosh(\beta(\|X_w - X_i\| - k_0))} \times \frac{(X_w - X_i) * (X_w - X_i)^T}{\|X_w - X_i\|^2}$$

3.2.1 1-Dimensional Case

For simplicity, let us consider the case when $n = 2$ (we later generalize). The FIM becomes,

$$F = \begin{pmatrix} -b_{12} & b_{12} \\ b_{12} & -b_{12} \end{pmatrix}$$

where $b_{wk} = -1 * \frac{\beta^2}{1+cosh(\beta(\|X_w - X_k\| - k_0))}$ (this definition of b_{wk} is used throughout the report)

Clearly, the eigenvalues of F are '0' and 'k_0', making it a singular matrix.

For general 'n', the matrix F can be given as,

$$F = \begin{pmatrix} -\sum\limits_{i=1,i\neq 1}^{n} b_{1i} & b_{12} & \cdots & b_{1n} \\ b_{21} & -\sum\limits_{i=1,i\neq 2}^{n} b_{2i} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & -\sum\limits_{i=1,i\neq n}^{n} b_{ni} \end{pmatrix}$$

We can see that all rows and all columns in F sum to '0'. To evaluate the eigenvalues, we find λ for which $|det(F - \lambda I)| = 0$

$$|det(F - \lambda I)| = 0$$

$$\implies \begin{vmatrix} -\sum\limits_{k=1,k\neq 1}^{n} b_{1k} - \lambda & b_{12} & \cdots & b_{1n} \\ b_{21} & -\sum\limits_{k=1,k\neq 2}^{n} b_{2k} - \lambda & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & -\sum\limits_{k=1,k\neq n}^{n} b_{nk} - \lambda \end{vmatrix} = 0$$

We do the following Gaussian-elimination step: $R_1 \rightarrow R_1 + R_2 + ... + R_n$, where R_i denotes the i^{th} row. We then obtain,

$$\implies \begin{vmatrix} -\lambda & -\lambda & \cdots & -\lambda \\ b_{21} & -\sum\limits_{k=1,k\neq 2}^{n} b_{2k} - \lambda & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & -\sum\limits_{k=1,k\neq n}^{n} b_{nk} - \lambda \end{vmatrix} = 0$$

$$\implies \lambda * \begin{vmatrix} 1 & 1 & \cdots & 1 \\ b_{21} & -\sum\limits_{k=1,k\neq 2}^{n} b_{2k} - \lambda & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & -\sum\limits_{k=1,k\neq n}^{n} b_{nk} - \lambda \end{vmatrix} = 0$$

Hence, we see that again $\lambda = 0$ is an eigenvalue, which means F is a singular matrix.

3.2.2 2-Dimensional Case

Like before, first consider the case when $n = 2$. The matrix F will be now comprised of four 2×2 block matrices.

$$F = \begin{pmatrix} -F_{12} & F_{12} \\ F_{12} & -F_{12} \end{pmatrix}$$

where, $F_{wk} = b_{wk} * \frac{1}{1+m_{wk}^2} \begin{pmatrix} -1 & m_{wk} \\ m_{wk} & -m_{wk}^2 \end{pmatrix}$

Here, m_{wk} is the slope of line joining X_w and X_k. From this, it is clear that F_{12} is a singular matrix. Thereby, F is a singular matrix too.

Now, let us consider the case for general 'n'. The FIM for this case can be expressed as,

$$F = \begin{pmatrix} -\sum\limits_{k=1,k\neq 1}^{n} F_{1k} & F_{12} & \cdots & F_{1n} \\ F_{21} & -\sum\limits_{k=1,k\neq 2}^{n} F_{2k} & \cdots & F_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ F_{n1} & F_{n2} & \cdots & -\sum\limits_{k=1,k\neq n}^{n} F_{nk} \end{pmatrix}$$

3.2.3 n-Dimensional Case

In this case, as observed before, we will have n^2 blocks, each of size $d \times d$. Let $F_{ij,wk}$ denote the $(i,j)^{th}$ element in the $(w,k)^{th}$ block matrix. Then,

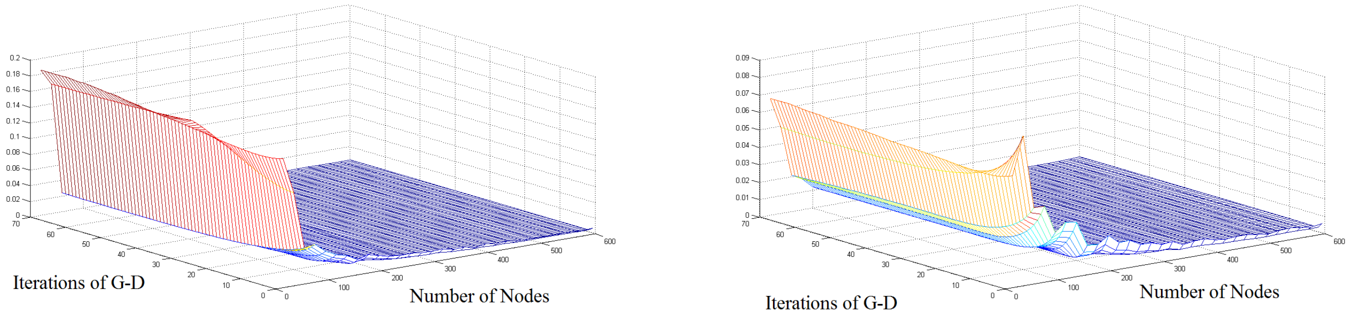
$$F_{ij,wk} = \frac{1}{b_{wk}} * \frac{1}{m_{ij,wk} + m_{ji,wk} + \sum\limits_{l=1;l\neq i,j}^{d} m_{li,wk} * m_{lj,wk}}; w \neq k$$

$$F_{ij,ww} = -\sum_{k=1,k\neq w}^{n} \frac{1}{b_{wk}} * \frac{1}{m_{ij,wk} + m_{ji,wk} + \sum\limits_{l=1;l\neq i,j}^{d} m_{li,wk} * m_{lj,wk}}$$

where,

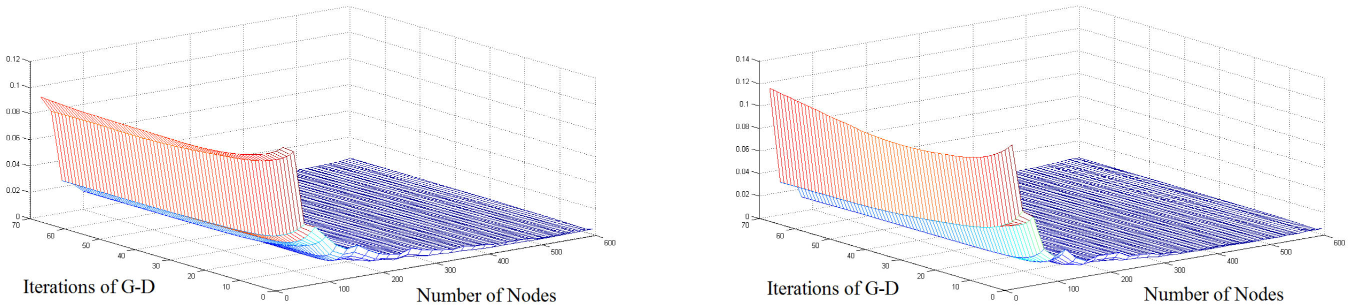
$$m_{ij,wk} = \frac{X_k^{(j)} - X_w^{(j)}}{X_k^{(i)} - X_w^{(i)}}$$

$s^{(q)}$ denotes the q^{th} component in a vector s.



(a) (b)

Figure 1: Error as function of ‘n’ and G-D iterations (a)$k_0 = 0.01$ (b)$k_0 = 0.03$



(a) (b)

Figure 2: Error as function of ‘n’ and G-D iterations (a)$k_0 = 0.05$ (b)$k_0 = 0.07$

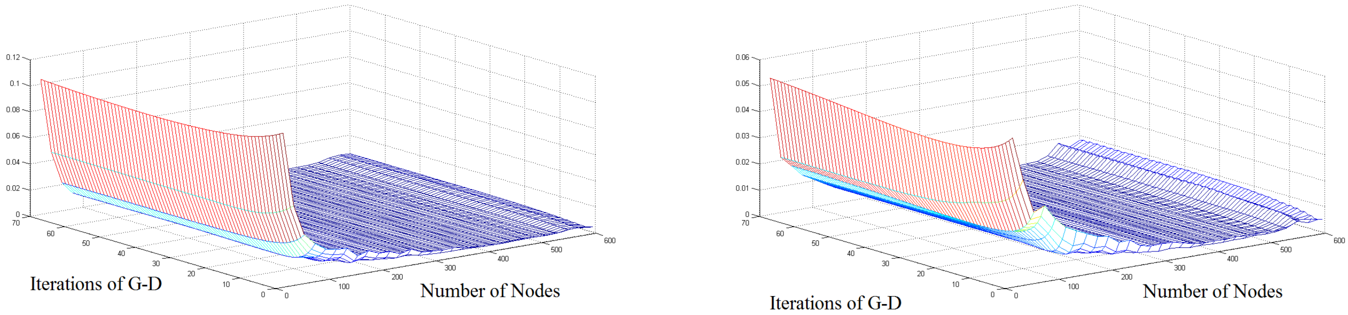
4 Experimental Results

In this section, we describe our experimental results considering the points $X_1, X_2, ..., X_n$ (distributed uniformly at random) from a unit square. The above elucidated gradient descent-based MLE has been implemented for a range of values of ‘k_0’ and ‘n’, the resulting plots are illustrated in Figures 1-10 (β, which governs the rate at which the sigmoid dies to ‘0’/rises to ‘1’, is chosen to be ‘5’). We compute the Minimum Mean-Squared error metric between the actual locations given by the matrix ‘X’ and the estimated matrix ‘$\hat{X}$’ as,

$$MMSE = \min_{t \in \mathbb{R}^2, R \in O(2)} \frac{1}{n} \sum_{i=1}^{n} \|X_i - t - R\hat{X}_i\|^2$$

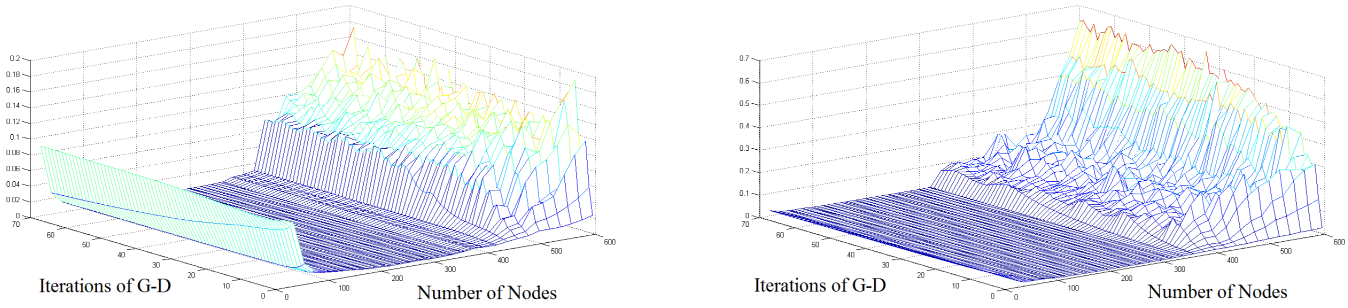
where $t \in \mathbb{R}^2$ and $R \in O(2)$, the set of orthogonal matrices $\in \mathbb{R}^{2\times 2}$, are chosen as the optimal rigid transformation between X and $\hat{X}$.

For a particular ‘k_0’, we see that there exists a range of ‘n’ values for which G-D converges. We can further see that as ‘k_0’ increases upto 0.5, there is a decrease in the upper limiting ‘n’ value for convergence. When k_0 is increased beyond 0.5, the upper limiting ‘n’ value for convergence shows an increasing behavior. To confirm that this is indeed the case, we carry out a re-run of the experiment and plot the resulting curves, which again depict this ‘decreasing upto $k_0 = 0.5$ and then increasing’ behavior of the upper limiting value of ‘n’ for convergence.



(a) (b)

Figure 3: Error as function of ‘n’ and G-D iterations (a)$k_0 = 0.09$ (b)$k_0 = 0.1$



(a) (b)

Figure 4: Error as function of ‘n’ and G-D iterations (a)$k_0 = 0.3$ (b)$k_0 = 0.5$

5 Conclusion

We have summarized the experimental results and characterized the Fisher Information Matrix for MLE of the localization problem. One of the advantages of our approach is the avoidance of using Djiksta’s shortest paths algorithm, which takes $O(|E| + V log|V|)$ complexity (where $|V|$ is the number of vertices in the graph and $|E|$ is the number of edges). Also, at least empirically, we observe that the proposed algorithm does better compared to MDS-MAP in terms of MMSE.

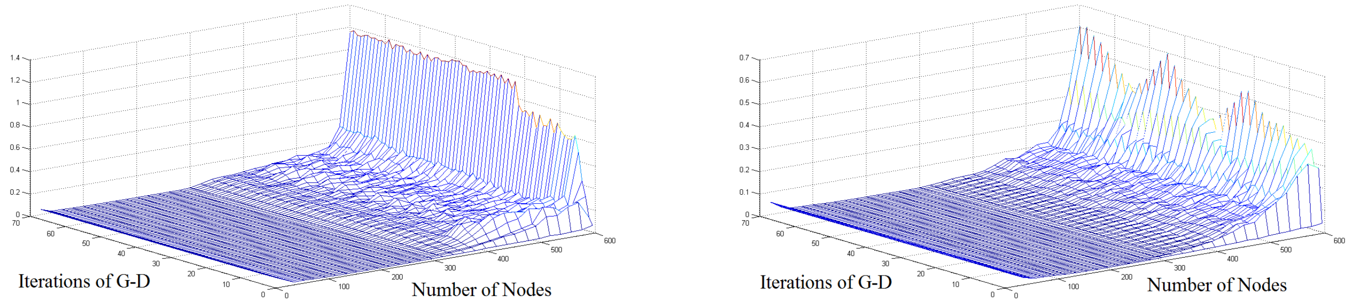
References

[1] Sewoong Oh, Andrea Montanari, and Amin Karbasi, “Sensor network localization from local connectivity: Performance analysis for the MDS-MAP algorithm” Information Theory Workshop (ITW), 2010 IEEE.

[2] Adel Javanmard, and Andrea Montanari, “Localization from incomplete noisy distance measurements” Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on, IEEE, 2011.

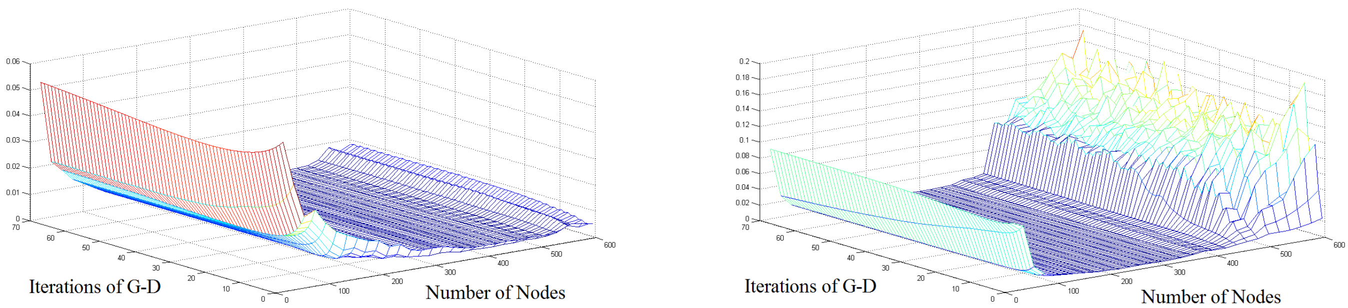
[3] Pratik Biswas, and Yinyu Ye, “Semidefinite programming for ad hoc wireless sensor network localization” Proceedings of the 3rd international symposium on Information processing in sensor networks, ACM, 2004.

[4] Neal Patwari, Alfred O Hero III, M Perkins, NS Correal and RJ O’dea, “Relative location estimation in wireless sensor networks” IEEE Transactions on Signal Processing 51(8), (2003).



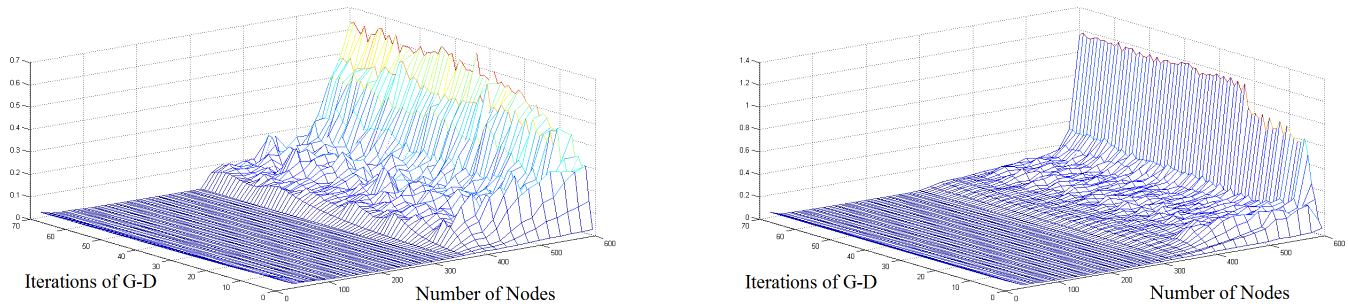
(a) (b)

Figure 5: Error as function of ‘n’ and G-D iterations (a)$k_0 = 0.7$ (b)$k_0 = 0.9$



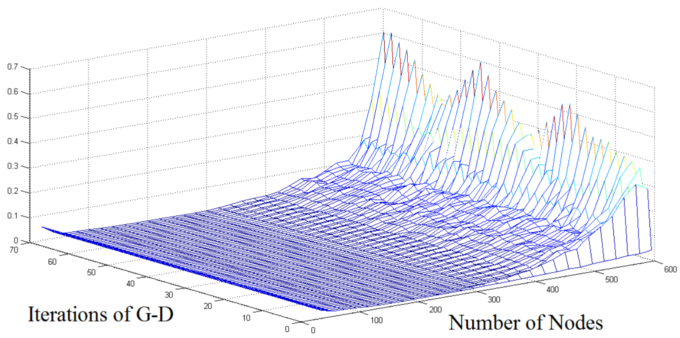
(a) (b)

Figure 6: Re-run experiment: Error as function of ‘n’ and G-D iterations (a)$k_0 = 0.1$ (b)$k_0 = 0.3$



(a) (b)

Figure 7: Re-run experiment: Error as function of ‘n’ and G-D iterations (a)$k_0 = 0.5$ (b)$k_0 = 0.7$



(a)

Figure 8: Re-run experiment: Error as function of ‘n’ and G-D iterations (a)$k_0 = 0.9$