

On the Min4PC Matrix of a Tree

Ali Azimi, Rakesh Jana, Mukesh K Nagar, and Sivaramakrishnan
Sivasubramanian*

(Communicated by Sudipta Mallik)

Abstract

The Four point condition (abbreviated as 4PC) is a condition used to test if a given distance matrix arises from shortest path distances on trees. From a tree T , Bapat and Sivasubramanian defined a matrix Min4PC_T based on this condition. They also gave a basis B for the row space of Min4PC_T and determined its Smith Normal Form. In this paper, we consider the matrix $\text{Min4PC}_T[B, B]$ restricted to a basis B and give an explicit inverse for it. It is known that the distance matrix D_T of a tree T , is invertible and that its inverse is a rank-one update of its scaled Laplacian. Our inverse has a similar form and is a rank-one update of a Laplacian like matrix.

1 Introduction

In the field of *Phylogenetics*, one explores the evolutionary relationships among species and studies a central concept of an *evolutionary tree*. This is a rooted tree structure where each node corresponds to a species and an edge signifies an evolutionary connection between two species, with the child node being considered as descending from its parent node. Typically, a new vertex is represented as a child when a genetic mutation occurs within a species. Thus, we denote species by nodes and connect by an edge all species obtained by mutation from that node. By measuring quantities such as the frequency of alleles in a population (that is, gene variations), it is possible to define a distance on this tree that quantifies the genetic difference between two species.

Typically, one can measure and hence define a distance between any pair of species and wants to determine the evolution tree which gives rise to these distances among its nodes. Due to possible inaccuracies in distance measurement, we get approximate distances and it is not obvious that an underlying tree exists for a given measured distance among pairs of

* Corresponding author

MSC2020: 05A15, 05C05; Keywords: Distance, Tree, Four point condition, Inverse

Received Nov 4, 2023; Revised Feb 8, 2024; Accepted Feb 9, 2024; Published Feb 14, 2024

© The author(s). Released under the CC BY 4.0 International License

vertices. Thus, our first aim after getting pairwise distances is to check if the distances arise from some tree. Here, when we write distance between two vertices, we mean the distance induced by the shortest path in the tree. For $u, v \in V(T)$, we denote the shortest path distance between them as $\text{dist}_T(u, v)$. Mathematically, our problem translates as follows: given a finite set X with a metric d_X on it, we want to determine whether there exists a tree T and an isometric embedding $\iota : X \rightarrow V(T)$ where $V(T)$ is the vertex set of T with $d_X(x, x') = \text{dist}_T(\iota(x), \iota(x'))$ for each $x, x' \in X$.

A classical theorem in this subject, due to Buneman [5], characterizes shortest path distances (that is metrics) derived from trees and is referred to as the *four-point condition* (abbreviated henceforth as 4PC). The 4PC states that for any four elements w, x, y and z from the given metric space (X, d_X) , among the three terms $d_X(x, y) + d_X(z, w)$, $d_X(x, z) + d_X(y, w)$, and $d_X(x, w) + d_X(y, z)$, the maximum value equals the second maximum value. It is noteworthy that the 4PC is more stringent than the triangle inequality as the triangle inequality can be derived by setting $z = w$. A metric d_X on X which satisfies the 4PC also known as *additive*, see Deza and Deza [6, page 16] and satisfies interesting properties. The following result is known.

Theorem 1.1 (Zaretskii, 1965). *Let d_X be a metric on a finite, nonempty set X that satisfies the 4PC. Then, there exists a unique weighted tree T whose leaves are precisely X , such that the weighted tree distance on the leaves X equals d_X .*

Theorem 1.1 was initially proven by Zaretskii in 1965 [13] when d_X had only integer distances. Subsequently, in 1969, Pereira in [11] extended it to cover non-integer distances. Independently, Buneman also gave a proof in [5] with no restriction on the distances. A further generalization of Theorem 1.1 has been studied to the case when the set X is not necessarily finite. In such cases, X might not be embeddable in a conventional tree, but is instead embedded in a more general structure called an \mathbb{R} -tree. For more details, see the paper by Gómez and Mémoli [7]. Another generalization of the 4PC can be found in the paper by Petrov and Salimov [12].

Buneman showed that distances arising from a tree T satisfy the 4PC and this result motivates the definition of three matrices Min4PC_T (see Bapat and Sivasubramanian [4]), Avg4PC_T (also called the 2-Steiner distance matrix, see Azimi and Sivasubramanian [2]) and Max4PC_T (see Azimi, Jana, Nagar and Sivasubramanian [1]). We describe these matrices below.

Let T be a tree with vertex set $V(T) = [n]$, where $[n] = \{1, \dots, n\}$. Denote its distance matrix as $D = (\text{dist}_T(i, j))_{1 \leq i, j \leq n}$. Let \mathcal{V}_2 be the set of 2-element subsets of $V(T)$. Note that $|\mathcal{V}_2| = \binom{n}{2}$. We describe three $\binom{n}{2} \times \binom{n}{2}$ matrices each of whose rows and columns are indexed by elements of \mathcal{V}_2 .

The minimum-4PC matrix, average-4PC and maximum-4PC matrices are denoted as Min4PC_T , Avg4PC_T and Max4PC_T respectively. For $X, Y \in \mathcal{V}_2$ with $X = \{i, j\}$ and $Y = \{k, l\}$, the (X, Y) th entry of Min4PC_T and Max4PC_T are denoted as $\text{Min4PC}_T(X, Y)$ and $\text{Max4PC}_T(X, Y)$ respectively. The entry $\text{Min4PC}_T(X, Y)$ of the Min4PC_T matrix is defined to be the minimum value of the three terms $\text{dist}_T(i, j) + \text{dist}_T(k, l)$, $\text{dist}_T(i, k) + \text{dist}_T(j, l)$, and $\text{dist}_T(i, l) + \text{dist}_T(j, k)$. An identical definition of Max4PC_T can be given by changing minimum to maximum in the previous sentence. The (X, Y) th entry of Avg4PC_T

is denoted as $\text{Avg4PC}_T(X, Y)$ and is defined as $\text{Avg4PC}_T(X, Y) = \frac{1}{2}(\text{Min4PC}_T(X, Y) + \text{Max4PC}_T(X, Y))$.

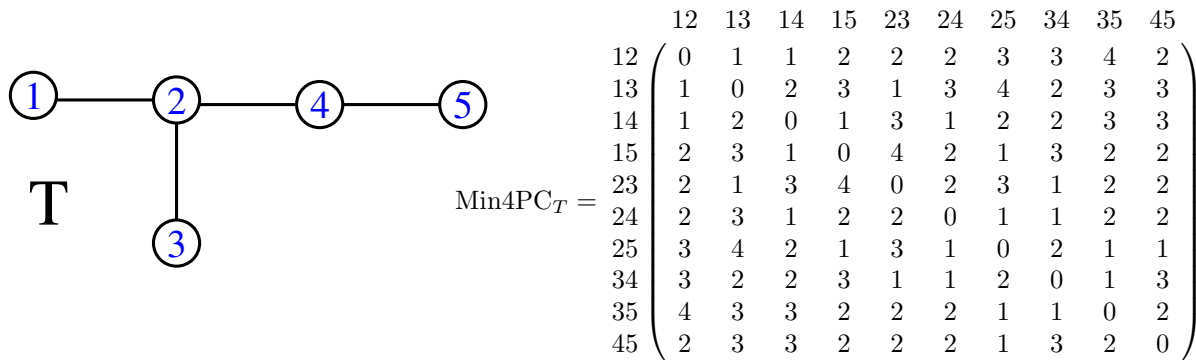


Figure 1: An example of the matrix Min4PC_T for the tree T given on the left

Interestingly, in [2, Lemma 4] it is shown that $\text{Avg4PC}_T(X, Y)$ is the *Steiner distance* between X and Y . The Steiner distance $d_{\text{ST}}(X, Y)$ between the subsets X and Y of $V(T)$ is defined as the number of edges in the smallest connected subtree of T that contains all the vertices of $X \cup Y$. It is noteworthy that when $X = \{x\}$ and $Y = \{y\}$, the Steiner distance between X and Y is the usual tree distance between x and y , that is, $d_{\text{ST}}(X, Y) = \text{dist}_T(x, y)$. For further information about Steiner distances in graphs, we refer the readers to Mao’s survey [10].

Bapat and Sivasubramanian in [4] studied the Min4PC_T matrix for a tree T and determined its rank, implicitly gave a basis for its row space and also determined its Smith Normal Form. For a square matrix M and a subset B of its rows (and columns), denote by $M[B, B]$ the restriction of M to the entries in the rows and columns indexed by elements of B . The following result about Min4PC_T is implicit in the work of Bapat and Sivasubramanian.

Theorem 1.2 (Bapat and Sivasubramanian, 2020). *Let T be a tree on n vertices with the edge set $E(T)$. Then, $\text{rank}(\text{Min4PC}_T) = n$. Further, if $f = \{k, l\} \notin E(T)$, then $B_f = E(T) \cup \{f\}$ is a basis of Min4PC_T and*

$$\det \text{Min4PC}_T[B_f, B_f] = (-1)^{n-1} 2^{n-2} (n-1) (\text{dist}(k, l) - 1)^2.$$

If D_T denotes the distance matrix of tree T on n vertices, the remarkable result of Graham and Pollak [9] shows that $\det(D_T) = (-1)^{n-1} (n-1) 2^{n-2}$ and hence when $n > 1$, we have $\text{rank}(D_T) = n$. Apart from the paper by Graham and Pollak, we also refer the reader to the book by Bapat [3, Chapter 8] for a proof of this result. Thus, we have $\text{rank}(D_T) = \text{rank}(\text{Min4PC}_T) = n$. Further, if we take $f = \{k, l\}$ with $\text{dist}(k, l) = 2$, then by Theorem 1.2, we get $\det \text{Min4PC}_T[B_f, B_f] = \det D_T = (-1)^{n-1} 2^{n-2} (n-1)$. Such unexpected coincidences hint that there is more in the matrix Min4PC_T than meets the eye.

In this short paper, we give the inverse of the matrix $M_f = \text{Min4PC}_T[B_f, B_f]$ and observe another similarity between the inverses of M_f and D_T . Graham and Lovász in [8] showed for a tree T that the inverse of D_T is a rank-one update of its scaled Laplacian matrix. Their result is the following.

Theorem 1.3 (Graham and Lovász, 1978). *Let T be a tree on $n \geq 2$ vertices and let D and L be its distance matrix and Laplacian respectively. Define the $n \times 1$ column vector τ by $\tau(v) = 2 - \deg(v)$ where $\deg(v)$ is the degree of vertex v in T . Then,*

$$D^{-1} = -\frac{1}{2}L + \frac{1}{2(n-1)}\tau\tau^\top \tag{1.1}$$

Our main result of this paper (proved in Section 2) is the following similar inverse of $M_f = \text{Min4PC}_T[B_f, B_f]$ as a rank-one update of a Laplacian type matrix.

Theorem 1.4. *Let T be a tree on n vertices. Suppose $f = \{k, \ell\}$ with $\text{dist}(k, \ell) = d > 1$. Then, there exists an $n \times n$ matrix L_f with zero row and column sums and an $n \times 1$ column vector τ_f such that*

$$M_f^{-1} = -\frac{1}{2(d-1)}L_f + \frac{1}{2(n-1)(d-1)^2}\tau_f\tau_f^\top. \tag{1.2}$$

2 Proof of Theorem 1.4

We start by looking at a principal submatrix of M_f and then define our Laplacian type matrix L_f . We need the following Lemma from Bapat and Sivasubramanian [4, Lemma 3, 4]. Let J be the all-ones matrix of appropriate dimension, all of whose entries are 1 and let I denote the identity matrix of appropriate dimension. Let $\mathbf{1}$ be a column vector of an appropriate dimension all of whose components are 1. As the dimensions of the matrices J, I and the vector $\mathbf{1}$ will be clear from the context, we take the liberty of mildly abusing this notation.

Lemma 2.1 (Bapat and Sivasubramanian). *Let T be a tree on n vertices. Let $K = M_f[E(T), E(T)]$ be the submatrix of M_f restricted to the entries indexed by $E(T)$. Then, $K = 2(J - I)$ and $2K^{-1} = -I + \frac{1}{n-2}J$.*

Let T be a tree on n vertices. Let $f = \{k, \ell\} \notin E(T)$ not be an edge of T and let $d = \text{dist}(k, \ell)$ be the distance between the vertices k and ℓ in T . Let $B_f = E(T) \cup \{f\}$. Let E_f denote the set of edges of T that are on the (unique) k, ℓ -path. Let E_f^c denote the set of edges of T that are **not** on the k, ℓ -path. Define the $n \times 1$ column vector τ_f in \mathbb{R}^n with entries indexed by B_f as follows

$$\tau_f(e) = \begin{cases} d-1 & \text{if } e \in E_f^c, \\ -(n-d-2) & \text{if } e \in E_f, \\ n-3 & \text{if } e = f. \end{cases}$$

Lemma 2.2. *Let T be a tree on n vertices. Suppose $f = \{k, \ell\}$ with $\text{dist}(k, \ell) = d > 1$. Then, we assert that $\mathbf{1}^\top \tau_f = 2(d-1)$ and $M_f \tau_f = (n-1)(d-1)\mathbf{1}$.*

Proof. Partition the set B_f as $B_f = E_f^c \sqcup E_f \sqcup \{f\}$ where \sqcup denotes a disjoint union. We consider this partition as edges e on the k, ℓ path have the same $\tau_f(e)$ value and edges e not on the k, ℓ path also have the same $\tau_f(e)$ value. With respect to this partition, τ_f^\top can be written as

$$\tau_f^\top = [(d-1)\mathbf{1}, \quad -(n-d-2)\mathbf{1}, \quad n-3]^\top. \quad (2.1)$$

Since $|E_f| = d$ and $|E_f^c| = n-d-1$, it follows that

$$\mathbf{1}^\top \tau_f = (n-d-1)(d-1) - d(n-d-2) + n-3 = 2(d-1).$$

Further, note that with respect to the partition $B_f = E_f^c \sqcup E_f \sqcup \{f\}$, the matrix M_f will be

$$M_f = \left[\begin{array}{c|c|c} 2(J-I) & 2J & (d+1)\mathbf{1} \\ \hline 2J & 2(J-I) & (d-1)\mathbf{1} \\ \hline (d+1)\mathbf{1}^\top & (d-1)\mathbf{1}^\top & 0 \end{array} \right].$$

To see the last row and column of M_f , if $e = \{i, j\}$, it is simple to note that

$$M_f(e, f) = \begin{cases} d-1 & \text{if } e \in E_f, \\ d+1 & \text{if } e \notin E_f. \end{cases}$$

We thus have

$$\begin{aligned} M_f \tau_f &= \left[\begin{array}{c|c|c} 2(J-I) & 2J & (d+1)\mathbf{1} \\ \hline 2J & 2(J-I) & (d-1)\mathbf{1} \\ \hline (d+1)\mathbf{1}^\top & (d-1)\mathbf{1}^\top & 0 \end{array} \right] \begin{bmatrix} (d-1)\mathbf{1} \\ -(n-d-2)\mathbf{1} \\ n-3 \end{bmatrix} \\ &= \begin{bmatrix} 2(d-1)(n-d-2)\mathbf{1} - 2d(n-d-2)\mathbf{1} + (d+1)(n-3)\mathbf{1} \\ 2(d-1)(n-d-1)\mathbf{1} - 2(n-d-2)(d-1)\mathbf{1} + (d-1)(n-3)\mathbf{1} \\ (d+1)(d-1)(n-d-1) - d(d-1)(n-d-2) \end{bmatrix} \\ &= (n-1)(d-1)\mathbf{1}. \end{aligned}$$

Our proof is complete. □

2.1 A Laplacian type matrix L_f

For a tree T with n vertices with $f = \{k, \ell\} \notin E(T)$, we define a symmetric matrix L_f with rows and columns indexed by elements of B_f . Let $E(T) = \{e_1, \dots, e_{n-1}\}$ and $\text{dist}(k, \ell) = d$. When $i \neq j$, we define

$$L_f(e_i, e_j) = \begin{cases} 0 & \text{if } e_i, e_j \in E_f^c, \\ -1 + \frac{n-d}{d-1} & \text{if } e_i, e_j \in E_f, \\ -1 & \text{otherwise,} \end{cases} \quad \text{and} \quad L_f(f, e_i) = \begin{cases} 1 & \text{if } e_i \in E_f^c, \\ -\frac{n-d}{d-1} & \text{if } e_i \in E_f. \end{cases}$$

We define the diagonal entries of L_f so that L_f has zero row and column sums. Therefore, for $g \in B_f$, we have

$$L_f(g, g) = \begin{cases} d - 1 & \text{if } g \in E_f^c, \\ \frac{n - d}{d - 1} + d - 2 & \text{if } g \in E_f, \\ \frac{n - 1}{d - 1} & \text{if } g = f. \end{cases}$$

Remark 2.3. Recall the partition $B_f = E_f^c \sqcup E_f \sqcup \{f\}$. With respect to this partition, the matrix L_f can be written in block form as

$$L_f = \left[\begin{array}{c|c|c} (d - 1)I & -J & \mathbf{1} \\ \hline -J & (d - 1)I + (q - 1)J & -q\mathbf{1} \\ \hline \mathbf{1}^\top & -q\mathbf{1}^\top & 1 + q \end{array} \right],$$

where $(d - 1)q = (n - d)$.

Lemma 2.4. Let T be a tree on n vertices and let $f = \{k, \ell\}$ with $\text{dist}(k, \ell) = d > 1$. Then, we have $L_f \mathbf{1} = \mathbf{0}$.

Proof. The proof is immediate from the definition of L_f . □

Lemma 2.5. Let T be a tree on n vertices and let $f = \{k, \ell\}$ with $\text{dist}(k, \ell) = d > 1$. Then

$$M_f L_f + 2(d - 1)I = \mathbf{1} \tau_f^\top.$$

Proof. Once again, recall that $B_f = E_f^c \sqcup E_f \sqcup \{f\}$ and also recall (2.1). Using Remark 2.3, we have

$$M_f L_f = \left[\begin{array}{c|c|c} 2(J - I) & 2J & (d + 1)\mathbf{1} \\ \hline 2J & 2(J - I) & (d - 1)\mathbf{1} \\ \hline (d + 1)\mathbf{1}^\top & (d - 1)\mathbf{1}^\top & 0 \end{array} \right] \left[\begin{array}{c|c|c} (d - 1)I & -J & \mathbf{1} \\ \hline -J & (d - 1)I + (q - 1)J & -q\mathbf{1} \\ \hline \mathbf{1}^\top & -q\mathbf{1}^\top & 1 + q \end{array} \right], \tag{2.2}$$

where $(d - 1)q = (n - d)$.

Let $N = M_f L_f + 2(d - 1)I$. Since N is a 3×3 block matrix, we denote its blocks as the (i, j) -th block where $1 \leq i, j \leq 3$. For positive integers r, s and t , we will use the two easy facts that $J_{r \times s} J_{s \times t} = s J_{r \times t}$ and $\mathbf{1}_r \mathbf{1}_t^\top = J_{r \times s}$. Clearly, the

$$\begin{aligned} (1, 1)\text{-th block of } N &= 2(d - 1)(J - I) - 2dJ + (d + 1)J + 2(d - 1)I = (d - 1)J, \text{ the} \\ (1, 2)\text{-th block of } N &= -2(n - d - 1)J + 2J + 2(d - 1)J + 2(q - 1)dJ - q(d + 1)J \\ &= -2(n - d - 1)J + q(d - 1)J = -(n - d - 2)J, \text{ and the} \\ (1, 3)\text{-th block of } N &= 2(n - d - 1)\mathbf{1} - 2\mathbf{1} - 2qd\mathbf{1} + (d + 1)(1 + q)\mathbf{1} \\ &= (2n - 3 - d - qd + q)\mathbf{1} = (n - 3)\mathbf{1}. \end{aligned}$$

Similarly, by (2.2), we get the

$$\begin{aligned}
 (2, 1)\text{-th block of } N &= 2(d-1)J - 2dJ + 2J + (d-1)J = (d-1)J, \text{ the} \\
 (2, 2)\text{-th block of } N &= -2(n-d-1)J + 2(d-1)(J-I) + 2(q-1)dJ - 2(q-1)J \\
 &\quad - q(d-1)J + 2(d-1)I \\
 &= -2(n-2d)J + (q-2)(d-1)J = -(n-d-2)J, \text{ and the} \\
 (2, 3)\text{-th block of } N &= 2(n-d-1)\mathbf{1} - 2qd\mathbf{1} + 2q\mathbf{1} + (d-1)(1+q)\mathbf{1} \\
 &= 2(n-d-1)\mathbf{1} + (d-1)(1-q)\mathbf{1} = (n-3)\mathbf{1}.
 \end{aligned}$$

Finally, using (2.2) again, we have the

$$\begin{aligned}
 (3, 1)\text{-th block of } N &= (d+1)(d-1)\mathbf{1}^\top - d(d-1)\mathbf{1}^\top = (d-1)\mathbf{1}^\top, \text{ the} \\
 (3, 2)\text{-th block of } N &= -(d+1)(n-d-1)\mathbf{1}^\top + (d-1)^2\mathbf{1}^\top + d(d-1)(q-1)\mathbf{1}^\top \\
 &= -(n-d-2)\mathbf{1}^\top, \text{ and the} \\
 (3, 3)\text{-th block of } N &= (d+1)(n-d-1) - qd(d-1) + 2(d-1) = n-3.
 \end{aligned}$$

Hence, by (2.1), it follows that

$$M_f L_f + 2(d-1)I = \left[\begin{array}{c|c|c} (d-1)J & -(n-d-2)J & (n-3)\mathbf{1} \\ \hline (d-1)J & -(n-d-2)J & (n-3)\mathbf{1} \\ \hline (d-1)\mathbf{1}^\top & -(n-d-2)\mathbf{1}^\top & (n-3) \end{array} \right] = \mathbf{1}\tau_f^\top.$$

Our proof is complete. □

We are now ready to prove Theorem 1.4.

Proof. (Of Theorem 1.4) By Lemmas 2.2 and 2.5, we have

$$M_f \left(-L_f + \frac{1}{(n-1)(d-1)} \tau_f \tau_f^\top \right) = -M_f L_f + \mathbf{1}\tau_f^\top = 2(d-1)I.$$

This completes the proof. □

Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments which made our article better.

References

- [1] Ali Azimi, Rakesh Jana, Mukesh Kumar Nagar, and Sivaramakrishnan Sivasubramanian, *The maximum four point condition matrix of a tree*, Preprint available at <https://arxiv.org/abs/2308.08237>.

- [2] Ali Azimi and Sivaramakrishnan Sivasubramanian, *The 2-steiner distance matrix of a tree*, Linear Algebra Appl. 655 (2022) 65–86.
- [3] R. B. Bapat, *Graphs and Matrices*, Hindustan Book Agency, Second Edition (2014).
- [4] R.B. Bapat and Sivaramakrishnan Sivasubramanian, *Smith normal form of a distance matrix inspired by the four-point condition*, Linear Algebra Appl. 603 (2020) 301–312.
- [5] Peter Buneman, *A note on the metric properties of trees*, Journal of Combinatorial Theory, Series B 17(1) (1974) 48–50.
- [6] Michel Marie Deza and Elena Deza, *Encyclopedia of distances*, Springer Verlag, Fourth Edition (2016).
- [7] Mario Gómez and Facundo Mémoli, *The four point condition: An elementary tropicalization of Ptolemy’s inequality*, Preprint available at <https://arxiv.org/abs/2111.10328>
- [8] Ronald L Graham and Lovasz Laszlo, *Distance Matrix Polynomials of Trees*, Advances in Mathematics, 29(1) (1978) 60–88.
- [9] Ronald L Graham and H. O. Pollak, *On the addressing problem for loop switching*, Bell System Tech. J 50 (1971) 2495–2519.
- [10] Yaping Mao, *Steiner distance in graphs—a survey*, Preprint available at <https://arxiv.org/abs/1708.05779>, 2017.
- [11] J.M.S. Simões Pereira, *A note on the tree realizability of a distance matrix*, Journal of Combinatorial Theory 6(3): (1969), 303–310.
- [12] Evgeniy A. Petrov and Ruslan R. Salimov, *A note on generalized four-point inequality*, Journal of Mathematical Sciences 273(3): 2023, 414–426.
- [13] KA Zaretskii, *Postroenie dereva po naboru rasstoianii mezhdu visiacimi*, Uspekhi Matematicheskikh Nauk 20 (1965) 94–96.

Contact Information

Ali Azimi	Department of Mathematics and Applied Mathematics
ali.azimi@xmu.edu.my	Xiamen University Malaysia
ali.azimi61@gmail.com	43900 Sepang, Selangor Darul Ehsan, Malaysia


Rakesh Jana	Department of Mathematics
rjana@math.iitb.ac.in	Indian Institute of Technology, Bombay
	Mumbai 400076, India

Mukesh Kumar Nagar
mukesh.kr.nagar@gmail.com

Department of Mathematics
Jaypee Institute of Information Technology
Noida 201309, India

Sivaramakrishnan Sivasubramanian
krishnan@math.iitb.ac.in

Department of Mathematics
Indian Institute of Technology, Bombay
Mumbai 400076, India

 <https://orcid.org/0000-0003-0463-4409>