Contributions to Discrete Mathematics

Volume 5, Number 2, Pages 185–194
ISSN 1715-0868

LEAST-SQUARES APPROXIMATION BY A TREE DISTANCE

RAVINDRA B. BAPAT

Abstract. Let T be a tree with vertex set $V(T) = \{1, \ldots, n\}$ and with a positive weight associated with each edge. The tree distance between i and j is the weight of the ij-path. Given a symmetric, positive real valued function on $V(T) \times V(T)$, we consider the problem of approximating it by a tree distance corresponding to T, by the least-squares method. The problem is solved explicitly when T is a path or a double-star. For an arbitrary tree, a result is proved about the nature of the least-squares approximation. Some properties of the incidence matrix of all the paths in the tree are proved and used. We also note similar results for the corresponding matrix of a directed graph and obtain a formula for the Moore-Penrose inverse of the all-paths matrix.

1. Introduction

Let T be a tree with $V(T) = \{1, \ldots, n\}$ and $E(T) = \{e_1, \ldots, e_{n-1}\}$. Let $\beta : E(T) \to [0, \infty)$. Thus β is an assignment of nonnegative weights to each edge of T. We extend β to a function on $V(T) \times V(T)$ as follows. We set $\beta(i, i) = 0$ for each i. If $i \neq j$, then $\beta(i, j)$ is defined to be the weight of the $ij-$path, where the weight of a path is the sum of the weights of the edges in the path. Note that $\beta(i, j) = \beta(j, i)$ for all i, j. The extended function $\beta : V(T) \times V(T) \to [0, \infty)$ will be called a tree distance, corresponding to T.

Suppose $w : V(T) \times V(T) \to [0, \infty)$ is a function satisfying $w(i, i) = 0$ and $w(i, j) = w(j, i)$ for all i, j. We will call w a dissimilarity. We consider the problem of approximating w by a tree distance β, corresponding to T, by the least-squares method. This problem is of interest and has been considered in the context of classification of species, see [2, Chapter 2], and [5]. A more recent reference is [8]. We now proceed to formulate this problem as a standard linear estimation problem.

Received by the editors June 11, 2009, and in revised form February 12, 2010.

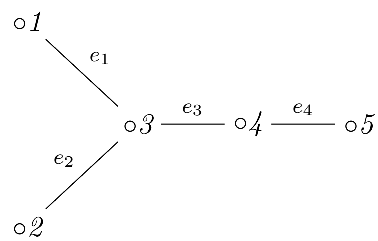
2000 *Mathematics Subject Classification.* 05C05, 05C12, 05C50, 62J05.

Key words and phrases. Tree, distance, weighted directed graph, Laplacian matrix, least-squares method, Moore-Penrose inverse.

The support of the JC Bose Fellowship, Department of Science and Technology, Government of India, is gratefully acknowledged.

It will be convenient to define the all-paths matrix S of T. The order of S is $\binom{n}{2} \times (n-1)$. The rows of S are indexed by $(i,j), 1 \leq i < j \leq n$, while the columns are indexed by $E(T)$. The entries of S are either 0 or 1. The row of S corresponding to (i,j) is the incidence vector of the ij-path in T. Thus the k-th entry in that row is 1 if e_k is on the ij-path, and 0 otherwise.

Example 1.1. *Consider the tree*



Then

$$S = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

If $w : V(T) \times V(T) \to [0, \infty)$, then let w also denote the vector of order $\binom{n}{2} \times 1$ with its components indexed by $(i,j), 1 \leq i < j \leq n$, where the component corresponding to $(i,j), i < j$, is set equal to $w(i,j)$. The problem of approximating a dissimilarity by a tree distance may be formulated as follows.

Let $w : V(T) \times V(T) \to [0, \infty)$ be a dissimilarity. Then the problem is to find $\beta : E(T) \to [0, \infty)$ such that $||S\beta - w||$ is minimized. Here $||x||$ denotes the usual Euclidean norm. It is well-known from the theory of least squares estimation that the minimizing vector β is a solution of the normal equations $S'S\beta = S'w$. We first make an elementary observation

Lemma 1.2. *$S'S$ is nonsingular.*

Proof. Consider the submatrix B of S, indexed by the rows (i,j), where $\{i,j\} \in E(T)$. If $\{i,j\} \in E(T)$, then the row of S corresponding to (i,j) has all zeros except a 1 in the position corresponding to the column indexed by the edge $\{i,j\}$. Thus B is a permutation matrix. It follows that the columns of S are linearly independent. Hence $\text{rank}(S'S) = \text{rank}(S) = n-1$ and therefore $S'S$ is nonsingular. $\square$

We refer to [3, 4] for background material on generalized inverses, and particularly, the Moore-Penrose inverse. It follows from Lemma 1.2 that the unique solution of the normal equations $S'S\beta = S'w$ is given by $\widehat{\beta} = (S'S)^{-1}S'w$. Note that $(S'S)^{-1}S'$ equals the Moore-Penrose inverse S^+. In the rest of the paper we obtain an explicit formula for $(S'S)^{-1}$ when T is a path or a double star. For an arbitrary tree we show that the (i,j)-element of $(S'S)^{-1}$ is 0 if the corresponding edges of T have no common vertex. This leads to some observations regarding the least squares solution $\widehat{\beta}$ in case of an arbitrary tree T. In the final section some results for a directed tree are described. A formula for the Moore-Penrose inverse of the all-paths matrix is obtained.

2. Path and double-star

In the following Theorem we provide a formula for $(S'S)^{-1}$ where S is the all-paths matrix of a path.

Theorem 2.1. *Let T be the path with $V(T) = \{1,\dots,n\}$ and $E(T) = \{e_1,\dots,e_{n-1}\}$, where e_i is the edge $\{i,i+1\}, i = 1,\dots,n-1$. Let S be the all-paths matrix of T. Then the (i,j)-entry of $(S'S)^{-1}$ is given by $\frac{2}{n}$, if $i = j$, $-\frac{1}{n}$ if $i \neq j$, and 0 otherwise.*

Proof. Let $B = S'S$. Then it is easy to see that

$$b_{ij} = \begin{cases} i(n-j) & \text{if } i \leq j, \\ j(n-i) & \text{if } i > j. \end{cases}$$

Thus the i-th row of B is given by

$$[n-i, 2(n-i), \dots, (i-1)(n-i), i(n-i), i(n-i-1), \dots, 2i, i].$$

Let C be the $(n-1)\times(n-1)$ matrix with

$$c_{ij} = \begin{cases} \frac{2}{n} & \text{if } i = j, \\ -\frac{1}{n} & \text{if } |i-j| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

We have, for $2 \leq j \leq n-2$,

$$(2.1)\quad \sum_{k=1}^{n-1} b_{ik}c_{kj} = b_{ij-1}c_{j-1i} + b_{ij}c_{jj} + b_{ij+1}c_{jj+1} = \frac{1}{n}(2b_{ij} - b_{ij-1} - b_{ij+1})$$

If $i = j$, the it follows from (2.1) that

$$\begin{aligned} \sum_{k=1}^{n-1} b_{ik}c_{ki} &= \frac{1}{n}(2b_{ii} - b_{ii-1} - b_{ii+1}) \\ &= \frac{1}{n}(2i(n-i) - (i-1)(n-i) - i(n-i-1) = 1. \end{aligned}$$

If $i > j$, then

$$\sum_{k=1}^{n-1} b_{ik}c_{kj} = \frac{1}{n}(2(n-i)j - (n-i)(j-1) - (n-i)(j+1)) = 0.$$

Finally, If $i < j$, then

$$\sum_{k=1}^{n-1} b_{ik}c_{kj} = \frac{1}{n}(2i(n-j+1) - i(n-j) - i(n-j)) = 0.$$

It can similarly be shown that if $j = 1$ or if $j = n-1$, then

$$\sum_{k=1}^{n-1} b_{ik}c_{kj} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Thus BC is the identity matrix and hence $C = (S'S)^{-1}$. □

We now derive a formula for the Moore-Penrose inverse S^+ of S.

Theorem 2.2. *Let T be the path with $V(T) = \{1, \ldots, n\}$ and $E(T) = \{e_1, \ldots, e_{n-1}\}$, where e_i is the edge $\{i, i+1\}, i = 1, \ldots, n-1$. Let S be the all-paths matrix of T. The rows of S^+ are indexed by the edges $\{i, i+1\}, i = 1, \ldots, n-1$ of T, while the columns are indexed by $(i, j), 1 \leq i < j \leq n$. The entry of S^+ corresponding to the edge $\{i, i+1\}$ and the pair (j, k) is given by $\frac{1}{n}$ if $i = j$ or $i = k-1$, $-\frac{1}{n}$ if $i = j-1$ or $i = j$, and 0 otherwise.*

Proof. We consider the case when $2 \leq i \leq n-2$. The cases $i = 1, n-1$ are similar. By Theorem 2.1 the row of $(S'S)^{-1}$ corresponding to the edge $\{i, i+1\}$ has all coordinates 0 except $\frac{2}{n}$ at coordinate i, and $-\frac{1}{n}$ at coordinates $i-1, i+1$. The column of S' corresponding to the pair (j, k) has 1 at coordinates $j, j+1, \ldots, k-1$ and zeros elsewhere. The inner product of these vectors gives the entry of $(S'S)^{-1}S' = S^+$ corresponding to the edge $\{i, i+1\}$ and the pair (j, k), and is seen to be as asserted in the statement of the theorem. □

Let T be the path with $V(T) = \{1, \ldots, n\}$ and $E(T) = \{e_1, \ldots, e_{n-1}\}$, where e_i is the edge $\{i, i+1\}, i = 1, \ldots, n-1$. Let $w : V(T) \times V(T) \to [0, \infty)$, be a dissimilarity. Consider the problem of finding $\beta : E(T) \to [0, \infty)$ such that $||S\beta - w||$ is minimized. Then we have the following result.

Theorem 2.3. *Let $\beta : E(T) \to [0, \infty)$ that minimizes $||S\beta - w||$ be $\widehat{\beta}$. The coefficients of $\widehat{\beta}$ are indexed by the edges $\{i, i+1\}, i = 1, \ldots, n-1$ of T. The coefficient of $\widehat{\beta}$ corresponding to $\{i, i+1\}, i = 1, \ldots, n-1$, is given by*

$$\sum_{j=i+1}^{n} w(i, j) + \sum_{k=1}^{i} w(k, i+1) - \sum_{j=i+2}^{n} w(i+1, j) - \sum_{k=1}^{i-1} w(k, i).$$

Proof. The β that minimizes $||S\beta - w||$ is given by S^+w. The result follows in view of the expression for S^+ given in Theorem 2.2. □

We consider another example in which the inverse of $S'S$ can be calculated explicitly. A double star is a tree in which all vertices have degree 1 except two vertices, which may have degree greater than 1. Consider the double star T with n vertices $\{1, \ldots, n\}, n = p+q+2$, in which vertices $1, \ldots, p+q$ are pendant, vertex $p+q+1$ is adjacent to $1, \ldots, p$, and vertex $p+q+2$ is adjacent to $p+1, \ldots, p+q$. Let S be the all-paths matrix of T. Let J_{rs} be the $r \times s$ matrix of all ones. Then it can be seen that

$$S'S = \begin{bmatrix} (n-2)I_p + J_{pp} & J_{pq} & (q+1)J_{p1} \\ J_{qp} & (n-2)I_q + J_{qq} & (p+1)J_{q1} \\ (q+1)J_{1p} & (p+1)J_{1q} & (p+1)(q+1) \end{bmatrix}. \tag{2.2}$$

Theorem 2.4. *Let $S'S$ be as in (2.2). Then*

$$(S'S)^{-1} = \begin{bmatrix} \frac{1}{n-2}I_p + v_1 J_{pp} & 0_{pq} & cJ_{p1} \\ 0_{qp} & \frac{1}{n-2}I_q + v_2 J_{qq} & dJ_{q1} \\ cJ_{1p} & dJ_{1q} & e \end{bmatrix},$$

where

$$c = -\frac{1}{p+q+2p^2}, \quad d = -\frac{1}{p+q+2q^2}, \quad v_1 = \frac{c(p-q)}{n-2},$$
$$v_2 = \frac{d(q-p)}{n-2}, \quad \textit{and} \quad e = \frac{1 - p(q+1)c - q(p+1)d}{(p+1)(q+1)}.$$

We omit the proof as it follows by simple verification. Once we have a formula for $(S'S)^{-1}$, an explicit expression for the least-squares approximation can also be obtained.

3. All-paths matrix of a tree

We now consider the all-paths matrix S of an arbitrary tree. We first prove some preliminary results.

Lemma 3.1. *Let T be a tree with $V(T) = \{1, \ldots, n\}$ and $E(T) = \{e_1, \ldots, e_{n-1}\}$. Let $e_i \in E(T)$ and let T_1 and T_2 be the components of $T \setminus \{e_i\}$. Let X be the submatrix of $S'S$ formed by the rows indexed by $E(T_1) \cup \{e_i\}$ and columns indexed by $E(T_2) \cup \{e_i\}$. Then* $\operatorname{rank} X = 1$.

Proof. For $e_j \in E(T), e_j \neq e_i$, let $f(e_j)$ denote the number of vertices in the component of $T \setminus \{e_j\}$ that does not contain e_i. Note that if $e_j \in E(T_1)$ and $e_k \in E(T_2)$, then the (e_j, e_k)-entry of X is $f(e_j)f(e_k)$. If $e_j \in E(T_1)$, then the (e_j, e_i)-entry of X is $f(e_j)|V(T_2)|$, while if $e_k \in E(T_2)$, then the (e_i, e_k)-entry of X is $f(e_k)|V(T_1)|$. It follows that $\operatorname{rank} X = 1$. □

Lemma 3.2. *Let A be an $m \times m$ matrix, $m \geq 2$. Let B be an $r \times s$ submatrix of A such that $r + s = m + 2$ and* $\operatorname{rank} B = 1$. *Then A is singular.*

Proof. We may assume, without loss of generality, that

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}.$$

Then

$$\begin{aligned} \operatorname{rank} A &\leq \operatorname{rank}[B, C] + \operatorname{rank}[D, E] \\ &\leq \operatorname{rank} B + \operatorname{rank} C + m - r \\ &\leq 1 + m - s + m - r \\ &\leq m - 1, \end{aligned}$$

and hence A is singular. □

Denote by $A(i|j)$ the submatrix obtained by deleting row i and column j of A. We now prove the main result of this section.

Theorem 3.3. *Let T be a tree with $V(T) = \{1, \dots, n\}$ and $E(T) = \{e_1, \dots, e_{n-1}\}$. Let S be the all-paths matrix of T. The rows and the columns of $S'S$ are indexed by $E(T)$. If $e_j, e_k \in E(T)$ have no vertex in common, then $S'S(j|k)$ is singular, and hence, the (j,k)-entry of $(S'S)^{-1}$ is zero.*

Proof. Since e_j and e_k have no vertex in common, there exists an edge e_i, distinct from e_j and e_k, on the path from e_j to e_k. Let T_1 and T_2 be the components of $T \setminus \{e_i\}$. As in Lemma 3.1, let X be the submatrix of $S'S$ formed by the rows indexed by $E(T_1) \cup \{e_i\}$ and the columns indexed by $E(T_2) \cup \{e_i\}$. By Lemma 3.1, $\operatorname{rank} X = 1$. Note that X is a matrix with $|E(T_1)| + 1$ rows and $|E(T_2)| + 1$ columns, and it is a submatrix of $S'S(j|k)$, which is of order $(n-2) \times (n-2)$. Since $|E(T_1)| + |E(T_2)| + 2 = n$, the result follows by Lemma 3.2. □

Theorem 3.3 has the following implication in terms of the problem of least-squares approximation by a tree distance. Let T be a tree with $V(T) = \{1, \dots, n\}$ and $E(T) = \{e_1, \dots, e_{n-1}\}$. Let $w : V(T) \times V(T) \to [0, \infty)$, be a dissimilarity. Consider the problem of finding $\beta : E(T) \to [0, \infty)$ such that $||S\beta - w||$ is minimized. Then we have the following result.

Theorem 3.4. *Let $\beta : E(T) \to [0, \infty)$ that minimizes $||S\beta - w||$ be $\widehat{\beta}$. Let $k \in \{1, \dots, n-1\}$. Let F be the set of edges of T which have a vertex in common with e_k. The least-squares estimate $\widehat{\beta}_k$ of β_k is a linear combination*

$$\sum_{i,j} \alpha(i,j) w(i,j),$$

such that

(i) *if the ij-path has no intersection with F, then $\alpha(i,j) = 0$.*
(ii) *if the intersection of the ij-path with F is the same as the intersection of the uv-path with F, then $\alpha(i,j) = \alpha(u,v)$.*

Proof. The β that minimizes $||S\beta - w||$ is given by $S^+w = (S'S)^{-1}S'w$. The coefficient $\widehat{\beta}_k$ is given by the inner product of the k-th row of $(S'S)^{-1}$ and $S'w$.

First suppose the ij-path has no intersection with F. By Theorem 3.3, the coordinates of the k-th row of $(S'S)^{-1}$ corresponding to edges not in F are all zero. Also the row of S corresponding to (i, j) has zeros at the places which correspond to edges in F. Thus the inner product of the k-th row of $(S'S)^{-1}$ and the row of S indexed by (i, j) is zero. This inner product equals $\alpha(i, j)$, which must then be zero. The second part follows similarly in view of the fact that the coordinates of the k-th row of $(S'S)^{-1}$ corresponding to edges not in F are all zero. □

Recall that a phylogenetic tree is a binary tree whose leaves are labelled by the species in a set X, and the internal vertices represent the unknown ancestors. An examination of the proof reveals that the results in this section apply equally well to a phylogenetic tree. It is known that the matrix S is nonsingular for any phylogenetic X-tree, see, for example, [7]. In fact, (i) of Theorem 3.4 has been observed in the context of a phylogenetic tree by Vach [9] and the property has been termed the *independence of irrelevant pairs* property in [7].

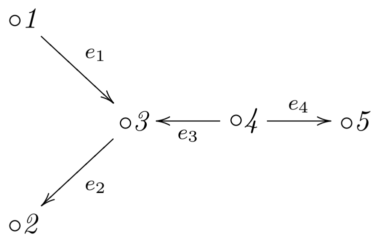
We also remark that the least-squares solution may not be nonnegative, a property required in practical applications in bioinformatics. The nonnegative least-squares problem must be approached by heuristic methods such as those in [2, 5]. Our results might be useful in that the least-squares solution, after rounding the negative entries to zero, can provide a good initial guess for such iterative methods. Our emphasis is on providing exact results for the least-squares solution.

We further remark that the least-squares method, without the nonnegativity constraint, involves inverting a matrix, or equivalently, solving a system of linear equations. The algorithmic complexity of matrix inversion by Gaussian elimination is known to be of the order $O(n^3)$. There exist faster methods which bring down the complexity to around $O(n^{2.8})$.

4. All-paths matrix of a directed tree

We consider directed graphs in this section. Let T be a directed tree with $V(T) = \{1, \ldots, n\}$ and $E(T) = \{e_1, \ldots, e_{n-1}\}$. We define the all-paths matrix P of T, which is a natural analogue of the undirected case. The order of P is $\binom{n}{2} \times (n-1)$. The rows of P are indexed by $(i, j), 1 \leq i < j \leq n$, while the columns are indexed by $E(T)$. The entries of P are either 0 or ± 1. The row of P corresponding to (i, j) is the incidence vector of the ij-path in T, where the directions of the edges are taken into account. Thus the k-th entry in that row is 1 if e_k is on the ij-path, and e_k is directed in the same way as we go from i to j along the path, it is -1 if e_k is on the ij-path, and e_k is directed in the opposite way as we go from i to j along the path, and it is 0 otherwise.

Example 4.1. *Consider the directed tree*



Then

$$P = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & -1 & -1 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Recall the definition of the (vertex-edge) incidence matrix of T, denoted by Q. It is a matrix of order $n \times (n-1)$, with its rows and columns indexed by $V(T)$ and $E(T)$ respectively. The (i, j)-entry of Q is 0 if vertex i and edge e_j are not incident, and otherwise it is 1 or -1 according as e_j originates or terminates at i, respectively. The incidence matrix of the tree T in Example 4.1 can be seen to be

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

The matrix $K = Q'Q$ has been termed the edge-Laplacian matrix of T by Merris [6] where a remarkable formula for K^{-1} is obtained. It is evident that the formula obtained by Merris can be expressed in the following equivalent form.

Theorem 4.2. $K^{-1} = (Q'Q)^{-1} = \frac{1}{n}P'P$.

The rows and the columns of $P'P$ are indexed by $E(T)$. It follows from Theorem 4.2 that $(P'P)^{-1} = \frac{1}{n}Q'Q$. Thus if edges e_i and e_j have no vertex in common, then the (i, j)-element of $(P'P)^{-1}$ is zero. This property holds in the undirected case as well, as observed in Theorem 3.3. In the directed case, an explicit formula is available for $(P'P)^{-1}$, namely $(P'P)^{-1} = \frac{1}{n}Q'Q$. However such a formula seems difficult to obtain in the case of an undirected tree.

We mention in passing that the matrices S' and P' may also be viewed as the fundamental cut-set matrices, over integers modulo 2 and over reals respectively, of the complete graph K_n, with respect to the spanning tree T.

It follows from Theorem 4.2 that $Q'QP'P = nI$, and hence $P^+ = \frac{1}{n}Q'QP'$. It is possible to give a graph-theoretic description of the entries of P^+ as we proceed to show.

The rows of P^+ are indexed by $E(T)$, while the columns of P^+ are indexed by $\{(i,j) : 1 \le i < j \le n\}$. Let $e_k \in E(T)$ have end-vertices u and v, and suppose e_k is directed from u to v. Fix $(i,j), i < j$. Let the entry of P^+ in the row indexed by e_k, and the column indexed by (i,j) be θ. We consider the following cases:

(i) If $i = u, j \neq v$ and the ij-path in T contains e_k, then $n\theta = 1$.
(ii) If $i = u$ and the ij-path in T does not contain e_k, then $n\theta = 1$.
(iii) If $i = v, j \neq u$ and the ij-path in T contains e_k, then $n\theta = -1$.
(iv) If $i = v$ and the ij-path in T does not contain e_k, then $n\theta = -1$.
(v) If $j = u, i \neq v$ and the ij-path in T contains e_k, then $n\theta = -1$.
(vi) If $j = u$ and the ij-path in T does not contain e_k, then $n\theta = -1$.
(vii) If $i = u, j = v$, then $n\theta = 2$.
(viii) If $i = v, j = u$, then $n\theta = -2$.

If none of the Cases (i)–(viii) hold, then e_k does not have even one vertex from i and j as an end-vertex and in that event, $\theta = 0$.

Note that the entries of nP^+ are all from $\{0, \pm 1, \pm 2\}$. Each row has exactly $2n - 3$ nonzero entries out of which one entry is ± 2.

We indicate an argument in justification of Case (i). Let e_k and e_ℓ be the first two edges on the (ij)-path. Let v and w be the end-vertices of e_ℓ.

Suppose $i = u, j \neq u$ and that the (ij)-path contains e_k. Let x be the row of $Q'Q$ indexed by e_k, and let y be the row of S indexed by (i,j). Since $nS^+ = Q'QS'$, $n\theta$ is given by the inner product x and y.

The elements of both x and y are indexed by $E(T)$. For $s \in \{1, \dots, n-1\}$, x_s is nonzero if and only if e_s has a vertex in common with e_k, while y_s is nonzero if and only if e_s is on the (ij)-path. Thus $x_s y_s \neq 0$ if and only if s equals either k or ℓ. Also $x_k = 2$ and $y_k = 1$.

If e_ℓ is directed from v to w, then $x_\ell = -1$ and $y_\ell = 1$. Thus

$$n\theta = \sum_{s=1}^{n-1} x_s y_s = x_k y_k + x_\ell y_\ell = 1.$$

Now suppose e_ℓ is directed from w to v. Then $x_\ell = 1$ and $y_\ell = -1$. Thus

$$n\theta = \sum_{s=1}^{n-1} x_s y_s = x_k y_k + x_\ell y_\ell = 1.$$

This completes the proof of the statement pertaining to Case (i). The proof is similar in the remaining cases.

The Moore-Penrose inverse of the all-paths matrix P of the tree in Example 4.1 is given by

$$P^+ = \frac{1}{5}\begin{bmatrix} 1 & 2 & 1 & 1 & 1 & 0 & 0 & -1 & -1 & 0 \\ 1 & -1 & 0 & 0 & -2 & -1 & -1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 & 1 & -1 & 0 & -2 & -1 & 1 \\ 0 & 0 & -1 & 1 & 0 & -1 & 1 & -1 & 1 & 2 \end{bmatrix}.$$

Consider the entry in row 3 and column 9. This corresponds to the edge e_3 and the pair $(3, 5)$. Setting $u = 4, v = i = 3$ and $j = 4$, we see from Case (iii) that the entry in $5P^+$ should be -1.

References

[1] R. B. Bapat, *Moore-Penrose inverse of the incidence matrix of a tree*, Linear and Multilinear Algebra **42** (1997), no. 2, 159–167.

[2] J.-P. Barthélemy and A. Guénoche, *Trees and Proximity Representations*, Wiley, 1991.

[3] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, Wiley-Interscience, 1974.

[4] S. L. Campbell and Jr. C. D. Meyer, *Generalized Inverses of Linear Transformations*, Pitman, 1979.

[5] V. Makarenkov and B. Leclerc, *An algorithm for the fitting of a phylogenetic tree according to a weighted least-squares criterion*, Journal of Classification **16** (1999), no. 1, 3–26.

[6] R. Merris, *An edge version of the matrix-tree theorem and the Wiener index*, Linear and Multilinear Algebra, vol. 25, 1989, pp. 291–296.

[7] R. Mihaescu and L. Pachter, *Combinatorics of least-squares trees*, `http://www.pnas.org/cgi/doi/10.1073/pnas.0802089105`.

[8] C. Semple and M. Steel, *Phylogenetics*, Oxford Lecture Series in Mathematics and its Applications, vol. 24, Oxford University Press, Oxford, 2003.

[9] W. Vach, *Least squares approximation of additive trees*, Conceptual and Numerical Analysis of Data, Springer, Heidelberg, 1990, pp. 230–238.

Indian Statistical Institute,
7, S. J. S. Sansanwal Marg, New Delhi 110016, India
E-mail address: rbb@isid.ac.in