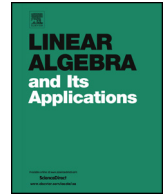Contents lists available at ScienceDirect

Linear Algebra and its Applications

Completion of tree metrics and rank 2 matrices

Daniel Irving Bernstein

Department of Mathematics, North Carolina State University, Raleigh, NC 27695, United States

A R T I C L E   I N F O

A B S T R A C T

Motivated by applications to low-rank matrix completion, we give a combinatorial characterization of the independent sets in the algebraic matroid associated to the collection of $m \times n$ rank-2 matrices and $n \times n$ skew-symmetric rank-2 matrices. Our approach is to use tropical geometry to reduce this to a problem about phylogenetic trees which we then solve. In particular, we give a combinatorial description of the collections of pairwise distances between several taxa that may be arbitrarily prescribed while still allowing the resulting dissimilarity map to be completed to a tree metric.

Published by Elsevier Inc.

1. Introduction

Given a matrix where only some of the entries are known, the low-rank matrix completion problem is to determine the missing entries under the assumption that the matrix has some low rank r. One can also assume additional structure such as (skew) symmetry or positive definiteness. Practical applications of the low-rank matrix completion problem abound. A well-known example is the so-called "Netflix Problem" of predicting an

individual's movie preferences from ratings given by several other users. A brief survey of other applications appears in [2].

Singer and Cucuringu show how ideas from rigidity theory can be applied to this problem in [16]. Jackson, Jordán, and Tibor further develop these ideas in [8,9]. Király, Theran, and Tomioka incorporate ideas from algebraic geometry into this rigidity-theoretic framework in [11] and Király, Theran and Rosen further develop these ideas in [12]. We add tools from tropical geometry to this picture.

Let V be a determinantal variety over some algebraically closed field $\mathbb{K}$. The results in this paper concern the cases were $V = \mathcal{S}_r^n(\mathbb{K})$, the collection of $n \times n$ skew-symmetric $\mathbb{K}$-matrices of rank at most r, or $V = \mathcal{M}_r^{m \times n}(\mathbb{K})$, the collection of $m \times n$ $\mathbb{K}$-matrices of rank at most r. A *masking operator* corresponding to some $S \subseteq \binom{[n]}{2}$ in the skew symmetric case, or $S \subseteq [m] \times [n]$ in the rectangular case, is a map $\Omega_S : V \to \mathbb{K}^S$ that projects a matrix M onto the entries specified by S. In the case of skew-symmetric $n \times n$ matrices, we view S as the edge set of a graph on vertex set $[n]$, which we denote $G(S)$. In the case of rectangular matrices, we view S as the edge set of a bipartite graph on partite sets of size m and n which we also denote $G(S)$. Context will make the proper interpretation of $G(S)$ clear.

Low-rank matrix completion problems can now be phrased as: given $\Omega_S(M)$ can we recover M if we know $M \in V$? For generic M the answer to this question only depends on the observed entries S and not the particular values observed. Namely, given $\Omega(M)$ for generic $M \in V$, M may be recovered up to finitely many choices if and only if S is a spanning set of the algebraic matroid associated to V. Hence it is useful to find combinatorial descriptions of the algebraic matroids associated to various determinantal varieties. We obtain such combinatorial descriptions for the cases where $V = \mathcal{S}_2^n(\mathbb{C})$ and $V = \mathcal{M}_2^{m \times n}(\mathbb{C})$. The most natural way to phrase our characterization is in terms of the independent sets of V. Note that a subset S of entries is independent in the algebraic matroid underlying V if and only if $\Omega_S : V \to \mathbb{C}^S$ is a dominant morphism. The salient feature for independent sets is that $\mathbb{C} \setminus \Omega_S(V)$ has Lebesgue measure zero. We also note that our result here answers a question of Kalai, Nevo, and Novik in [10] to find a combinatorial classification of what they call "minimally $(2, 2)$-rigid graphs" (posed in the paragraph after the proof of their Example 5.5). Using our language, these are the maximal independent sets in the algebraic matroid underlying $\mathcal{M}_2^{m \times n}(\mathbb{C})$.

An alternating closed trail in a directed graph is a walk $v_0, v_1, \ldots, v_k$ such that each edge appears at most once, $v_k = v_0$, and adjacent edges $v_{i-1}v_i$ and v_iv_{i+1} have opposite orientations (indices taken modulo $k + 1$). We now state our main result.

Theorems 4.2 and 4.4. *Let $V = \mathcal{S}_2^n(\mathbb{C})$ be the variety of skew-symmetric $n \times n$ matrices of rank at most 2, or $V = \mathcal{M}_2^{m \times n}(\mathbb{C})$ be the variety of rectangular $m \times n$ matrices of rank at most 2. A subset of observed entries $S \subseteq \binom{[n]}{2}$ (skew symmetric case) or $S \subseteq [m] \times [n]$ (rectangular case) is independent in the algebraic matroid underlying V if and only if there exists some acyclic orientation of $G(S)$ that has no alternating closed trail.*

Using techniques of [11], one can see that deciding whether a given $S \subseteq \binom{[n]}{2}$ or $S \subseteq [m] \times [n]$ is independent in the algebraic matroid underlying $\mathcal{S}_2^n(\mathbb{C})$ or $\mathcal{M}_2^{m \times n}(\mathbb{C})$ is in the complexity class RP. Hence both decision problems are also in NP. We provide an explicit combinatorial certificate of this fact. Existence of a polynomial time algorithm for solving either decision problem remains open.

The key to our approach is to use tropical geometry to allow us to reduce to the following easier question: Which entries of a dissimilarity map may be arbitrarily specified such that the resulting partial dissimilarity map may be completed to a tree metric? Our result here is as follows.

Theorem 3.6. *Let $S \subseteq \binom{[n]}{2}$. Any partial dissimilarity map whose known distances are given by S can be completed to a tree metric, regardless of what those specified values are, if and only if there exists some acyclic orientation of $G(S)$ that has no alternating closed trail.*

As in the case of partial matrices, we give an explicit combinatorial certificate showing that the corresponding decision problem is in NP but we do not know whether a polynomial time algorithm exists.

The problem of deciding whether a *particular* partial dissimilarity map is completable to a tree metric was shown to be NP-complete in [5]. Note that this is distinct from our decision problem here, because we are not setting values of the observed entries. Special cases that allow a polynomial time algorithm were investigated in [7,6]. Questions about whether a partial dissimilarity map can be completed to a tree metric with a particular topology have been addressed in [3,4].

The outline of the paper is as follows. Section 2 lays out some general theory for using tropical geometry to characterize algebraic matroids. Section 3 contains our results relating to completion of tree metrics. Section 4 shows how our results on partial matrices are easily obtained from our results on trees.

2. Completion and tropical varieties

We begin with the necessary preliminaries from tropical geometry. The most important parts of this section are Lemmas 2.4 and 2.5 which enable us to use the polyhedral structure of a tropical variety to gain insight into the corresponding algebraic matroid.

Let $\mathbb{K}$ be a field and let $V \subset \mathbb{K}^n$. We let $M(V)$ denote the *independence complex* of S which we define to be the collection of subsets S of $\{1, \ldots, n\}$ such that the projection of V onto the coordinates indicated by S is full-dimensional in $\mathbb{K}^S$. In the case that the sets in $M(V)$ form a matroid (e.g. V is an irreducible affine variety) we refer to $M(V)$ as the *matroid underlying V*.

Denote by $\mathbb{C}\{\{t\}\}$ the field of complex formal Puiseux series. That is, $\mathbb{C}\{\{t\}\}$ is the set of all formal sums $\sum_{\alpha \in J} c_\alpha t^\alpha$ where $J \subset \mathbb{Q}$ such that J has a smallest element and the elements of J can be expressed over a common denominator. The valuation map

$\mathrm{val}(\cdot) : \mathbb{C}\{\{t\}\} \to \mathbb{Q}$ sends $\sum_{\alpha \in J} c_\alpha t^\alpha$ to $\min\{\alpha \in J : c_\alpha \neq 0\}$. For any affine variety V over $\mathbb{C}\{\{t\}\}$, the corresponding *tropical variety* is

$$\mathrm{trop}(V) = \overline{\{(-\mathrm{val}(x_1), \ldots, -\mathrm{val}(x_n)) : (x_1, \ldots, x_n) \in V\}} \subseteq \mathbb{R}^n$$

where the overline indicates closure in the Euclidean topology on $\mathbb{R}^n$. We sometimes refer to $\mathrm{trop}(V)$ as *the tropicalization of V*.

We can also tropicalize varieties over $\mathbb{C}$ by lifting to $\mathbb{C}\{\{t\}\}$ and tropicalizing there. More specifically, let $V \subseteq \mathbb{C}^n$ be an affine variety over $\mathbb{C}$ with ideal $I \subseteq \mathbb{C}[x_1, \ldots, x_n]$. By lifting this ideal into $\mathbb{C}\{\{t\}\}[x_1, \ldots, x_n]$ we obtain a variety $V' \subseteq (\mathbb{C}\{\{t\}\})^n$. The tropical variety $\mathrm{trop}(V)$ corresponding to V is simply $\mathrm{trop}(V')$.

Lemma 2.1 *([19], Lemma 2). Let V be an irreducible affine variety over either $\mathbb{C}$ or $\mathbb{C}\{\{t\}\}$. Then the independence complex of V and $\mathrm{trop}(V)$ are the same.*

Many matrix completion problems ask for a combinatorial description of the algebraic matroid associated to a particular irreducible affine variety. Lemma 2.1 says that we can tackle such problems by looking at the corresponding tropical variety instead. The advantage in this is that tropical varieties have a useful polyhedral structure which we now describe.

Definition 2.2. Let Σ be a rational fan in $\mathbb{R}^n$ of pure dimension d. We say that Σ is *balanced* if we can associate a positive integer $m(\sigma)$, called the *multiplicity* to each full-dimensional cone σ in a way such that for each cone $\tau \in \Sigma$ of dimension $d-1$,

$$\sum_{\sigma \supsetneq \tau} m(\sigma) v_\sigma \in \mathrm{span}(\tau)$$

where v_σ is the first lattice point on the ray $\sigma/\mathrm{span}(\tau)$. We say that Σ is *connected through codimension one* if for any d-dimensional cones $\sigma, \rho \in \Sigma$, there exists a sequence of d-dimensional cones $\sigma = \sigma_0, \sigma_1, \ldots, \sigma_k = \rho \in \Sigma$ such that $\sigma_i \cap \sigma_{i+1}$ has dimension $d-1$.

The well-known structure theorem for tropical varieties applied to the special case where the defining equations have constant coefficients gives us the following.

Theorem 2.3 *([14], Theorem 3.3.5). Let V be an irreducible d-dimensional affine variety over $\mathbb{C}$. Then $\mathrm{trop}(V)$ is the support of a balanced fan of pure dimension d that is connected through codimension one.*

The following two lemmas give us ways that one can use the polyhedral structure of $\mathrm{trop}(V)$ to gain insight into the structure of $M(V)$. When we refer to a cone in $\mathrm{trop}(V)$, we mean in a polyhedral subdivision of $\mathrm{trop}(V)$ that is balanced and connected through codimension one.

Lemma 2.4. *Let $V \subseteq \mathbb{C}^n$ be an irreducible affine variety of dimension d. Then $S \subseteq [n]$ is independent in $M(V)$ if and only if S is independent in $M(\mathrm{span}(\sigma))$ for some d-dimensional cone σ in $\mathrm{trop}(V)$.*

Proof. Let S be independent in $M(V)$. By Lemma 2.1, the projection of $\mathrm{trop}(V)$ onto $\mathbb{R}^S$ is full dimensional. In particular, the projection of some maximal dimensional cone $\sigma \in \mathrm{trop}(V)$ onto $\mathbb{R}^S$ is full dimensional and therefore S is independent in the matroid $M(\mathrm{span}(\sigma))$.

Now let S be independent in $M(\mathrm{span}(\sigma))$ for some maximal cone $\sigma \in \mathrm{trop}(V)$. Then the projection of $\mathrm{span}(\sigma)$ onto $\mathbb{R}^S$ is full dimensional. Therefore the same holds for σ and therefore $\mathrm{trop}(V)$. So S is in the independence complex of $\mathrm{trop}(V)$ and therefore independent in $M(V)$ by Lemma 2.1. $\square$

Lemma 2.5. *Let $V \subseteq \mathbb{C}^n$ be an irreducible d-dimensional affine variety. Let τ be a $d-1$ dimensional cone of $\mathrm{trop}(V)$ and let $\sigma_1, \ldots, \sigma_k$ be the d-dimensional cones in $\mathrm{trop}(V)$ containing τ. If $B \subseteq [n]$ is a basis of $M(\mathrm{span}(\sigma_1))$ then B is also a basis of $M(\mathrm{span}(\sigma_i))$ for some $i \neq 1$.*

Proof. Let $L = \mathrm{span}(\tau)$ and $L_i = \mathrm{span}(\sigma_i)$. By Theorem 2.3 there exist $v_i \in \sigma_i \setminus L$ such that $\sum_{i=1}^k v_i \in L$. Let B be a basis of L_1. Then the projection of L_1 onto $\mathbb{R}^B$ is all of $\mathbb{R}^B$ and the projection of L onto $\mathbb{R}^B$ is some hyperplane $\{x \in \mathbb{R}^B : c^T x = 0\}$. By padding with extra zeros, we can extend c to an element of $\mathbb{R}^n$. Then $L = L_1 \cap \{x \in \mathbb{R}^n : c^T x = 0\}$ and $c^T v_i \neq 0$. Since $\sum_{i=1}^k v_i \in L$, we must have $\sum_{i=1}^k c^T v_i = 0$. Therefore there must exist some $i \neq 1$ such that $c^T v_i \neq 0$. Since $L_i = L + \mathrm{span}(v_i)$, the projection of L_i onto $\mathbb{R}^B$ is all of $\mathbb{R}^B$. So B is a basis of L_i. $\square$

We end this section by noting a nice feature about projections of tropical varieties which was noted by Yu in [19].

Proposition 2.6. *Let $\mathbb{K}$ be either $\mathbb{C}$ or $\mathbb{C}\{\{t\}\}$ and let $V \subseteq \mathbb{K}^n$ be an irreducible affine variety. If S is independent in $M(V)$ then the projection of $\mathrm{trop}(V)$ onto $\mathbb{R}^S$ is all of $\mathbb{R}^S$.*

3. Tree metrics and tree matroids

In this section we determine which entries of a dissimilarity map can be arbitrarily specified while still allowing completion to a tree metric. We begin with the necessary preliminaries on tree metrics. Proposition 3.2 describes the tropical and polyhedral structure on the set of tree metrics. Lemma 3.3 reduces our question about completion to an arbitrary tree metric to the question about completion to a tree metric whose topology is a caterpillar tree. We answer this simpler question in Proposition 3.4 and give the section's main result in Theorem 3.6.

We now give the necessary background on tree metrics. For a more detailed account, see [15]. Let X be a set. A *dissimilarity map on X* is a function $\delta : X \times X \to \mathbb{R}$ such that
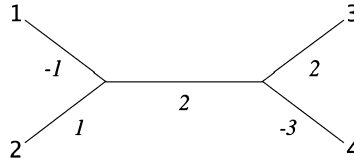
Fig. 1. An edge-weighted tree with all internal edge weights positive whose leaves are labeled by $\{1, 2, 3, 4\}$ showing that $d = (d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34}) = (0, 3, -2, 5, 0, -1)$ is a tree metric.

$\delta(x, x) = 0$ and $\delta(x, y) = \delta(y, x)$ for all $x, y \in X$. If $X = \{x_1, \ldots, x_n\}$ and $[n] = \{1, \ldots, n\}$ then there is a natural bijection between dissimilarity maps on X and points in $\mathbb{R}^{\binom{[n]}{2}}$. Namely, denoting the entry of a point $d \in \mathbb{R}^{\binom{[n]}{2}}$ corresponding to $\{i, j\}$ by d_{ij} or d_{ji}, we associate a dissimilarity map δ on X with the point $d \in \mathbb{R}^{\binom{[n]}{2}}$ such that $d_{ij} = \delta(x_i, x_j)$. Hence we will often speak of dissimilarity maps as if they were merely points in $\mathbb{R}^{\binom{[n]}{2}}$.

A tree with no vertices of degree 2 with leaf set X is called an X-tree. Let T be an X-tree and let w be an edge-weighting on T such that $w(e) > 0$ for all internal edges e of T. Some authors require that $w(e) > 0$ for *any* edge e of T, but we do not. The triple (T, X, w) gives rise to a dissimilarity map d where d_{ij} is the sum of the edge weights given by w along the unique path from x_i to x_j in T. Any dissimilarity map d that arises from a triple in this way is called a *tree metric*. For example, if $X = \{1, 2, 3, 4\}$, then the dissimilarity map $d = (d_{12}, d_{13}, d_{14}, d_{23}, d_{24}, d_{34}) = (0, 3, -2, 5, 0, -1)$ is a tree metric because it can be displayed on a tree as in Fig. 1. Given a tree metric d on X, there is a unique X-tree T that realizes d as such [15, Theorem 7.1.8]. This tree T is called the *topology* of d. A tree T is *binary* if all internal vertices have degree three.

Proposition 3.1 *([15], Theorem 7.2.6). A dissimilarity map $d \in \mathbb{R}^{\binom{[n]}{2}}$ is a tree metric if and only if it satisfies*

$$d_{ij} + d_{kl} \leq \max\{d_{ik} + d_{jl}, d_{il} + d_{jk}\}$$

for all distinct $i, j, k, l \in [n]$.

Authors who require that all edge weights be positive in the definition of a tree metric have a similar four point condition. The only difference is that i, j, k, l need not be distinct. We denote the set of all tree metrics on a set of size n by $\mathcal{T}_n$.

The following proposition summarizes many known facts about the polyhedral and tropical structure of $\mathcal{T}_n$ (references are given in the proof). In particular, it tells us how to realize $\mathcal{T}_n$ as a tropical variety along with the polyhedral subdivision guaranteed to exist by Theorem 2.3. We note that it would not be true had we taken the more restrictive definition of tree metric that required all (as opposed to just internal) edge weights to be positive. Recall that the Grassmannian $\mathrm{Gr}_{k,n} \subset \mathbb{K}^{\binom{[n]}{k}}$ is the irreducible affine variety parameterized by the $k \times k$ minors of a $k \times n$ matrix over $\mathbb{K}$.

Proposition 3.2. *The space of phylogenetic trees $\mathcal{T}_n$ is $\mathrm{trop}(\mathrm{Gr}_{2,n})$. We can give $\mathcal{T}_n$ a balanced fan structure connected through codimension one as follows. Each open cone is*
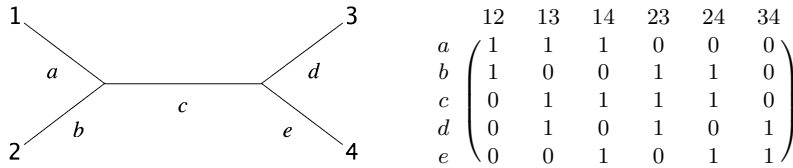
Fig. 2. Let T be the tree on the left with leaves $\{1, 2, 3, 4\}$ and edges labels $\{a, b, c, d, e\}$. The matrix on the right is A_T. Its columns are the $\lambda_T(ij)$s.

the collection of tree metrics whose topology is T for a particular tree T. Such a cone is maximal if and only if T is binary. The cone has codimension one if and only if T can be obtained from a binary tree by contracting exactly one edge. Given a cone τ of codimension one corresponding to tree T, the cones containing τ correspond to the three binary trees that can be contracted to T.

Proof. Let L_n be the set of tree metrics whose topology is a star tree – that is, a tree with no internal edges. Since an edge-weighting of a leaf-labeled tree gives rise to a tree metric if and only if all *internal* edge weights are nonnegative, L_n is the lineality space of $\mathcal{T}_n$. Theorem 3.4 in [17] implies that $\mathcal{T}_n/L_n = \text{trop}(\text{Gr}_{2,n})/L_n$. But as L_n is also the lineality space of $\text{Gr}_{2,n}$ (cf. remarks following Corollary 3.1 in [17]), this implies that $\mathcal{T}_n = \text{trop}(\text{Gr}_{2,n})$.

The polyhedral structure is given in [1]. Connectedness through codimension one follows from the fact that any two binary trees on the same leaf set can be reached from one another via a finite sequence of nearest-neighbor-interchanges (see [18, Theorem 2]). We can see that this polyhedral fan is balanced by assigning multiplicity 1 to each maximal cone (see Remark 4.3.11 and Theorem 3.4.14 in [14]). $\square$

Let T be a tree without vertices of degree two whose leaves are labeled by $[n]$. Denote its edge set by E and let $\mathbb{R}^E$ denote the vector space of edge-weightings of T. For each pair $ij \in \binom{[n]}{2}$, define $\lambda_T(ij) \in \mathbb{R}^E$ by $\lambda_T(ij)_e = 1$ whenever e is on the unique path from the leaf labeled i to the leaf labeled j and 0 otherwise. We denote the linear matroid underlying these $\lambda_T(ij)$ by $M(T)$ and call it the *matroid associated to T*. We will abuse language and say that a set $S \subseteq \binom{[n]}{2}$ is independent in $M(T)$ to mean that $\{\lambda_T(ij)\}_{ij \in S}$ is independent in $M(T)$. Let A_T be the matrix whose columns are the $\lambda_T(ij)$s. An example is given in Fig. 2. Note that rowspan(A_T) is the linear span of the cone in $\mathcal{T}_n$ containing all tree metrics with topology T. Moreover, $M(T)$ is the matroid of this linear space. These matroids were introduced and studied in [3].

Let T be a tree with leaf labels $[n]$. A *cherry* on a binary $[n]$-tree T is a pair of leaves ij such that i and j are adjacent to the same vertex of degree 3. A caterpillar tree on $n \geq 4$ vertices is a binary tree with exactly two cherries. See Fig. 3 for an illustration.

Lemma 3.3. *A set $B \subseteq \binom{[n]}{2}$ is a basis of $M(\mathcal{T}_n)$ if and only if B is a basis in the matroid associated to some caterpillar tree.*
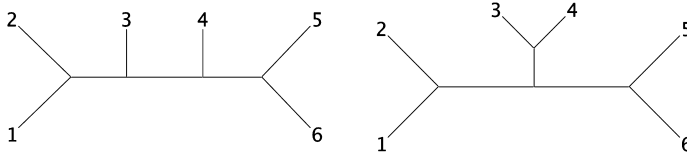
Fig. 3. The tree on the left has cherries 12 and 56 hence it is a caterpillar. The tree on the right has cherries 12, 34 and 56 and is therefore *not* a caterpillar.
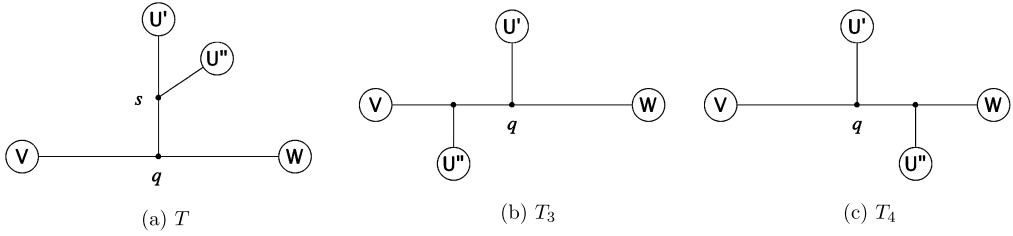

Fig. 4. Breaking T into subtrees.

Proof. Lemma 2.4 implies that S is independent in $M(\mathcal{T}_n)$ if and only if S is independent in $M(T)$ for some binary tree T. We proceed by showing that if B is a basis of $M(T)$ for a non-caterpillar binary tree T, then B is also the basis of $M(T')$ for some binary tree T' with one fewer cherry. It will follow by induction that B is a basis of $M(C)$ for some caterpillar C.

So let T be a binary tree with three distinct cherries ij, $i'j'$, $i''j''$ with corresponding degree-3 vertices p, p', p''. There is a single vertex q in the common intersection of the three paths p to p', p' to p'', and p to p''. If we delete q from T we get three trees U, V, W which contain the cherries ij, $i'j'$, and $i''j''$ respectively.

We now induct on the number of leaves in U. The base case is where U consists of the two leaves i and j. We can visualize this as in Fig. 4a letting U' be i and U'' be j. Lemma 2.5 and Proposition 3.2 imply that B must also be a basis of one of the trees T_1 or T_2, depicted in Figs. 4b and 4c respectively. Note that T_1 and T_2 each have fewer cherries than T.

Now assume U has more than two leaves. Let s be the first node on the path from q to p. Deleting s from U splits it into two subtrees which we denote U' and U''. This is depicted in Fig. 4a. Assume ij belongs to the subtree U'. Using Lemma 2.5 and Proposition 3.2 as before, we see that B must also be a basis of one of the trees T_3 or T_4, depicted in Figs. 4b and 4c respectively.

We can now repeat the process with U' taking the place of U. In the case that B is a basis of T_3 we expand V to include U'', and in the case of T_4 we instead expand W to contain U''. The result has strictly fewer leaves in U and so we induct. $\square$

Lemma 3.3 reduces the problem of describing the independent sets of $M(\mathcal{T}_n)$ to the simpler problem of describing the independent sets of $M(T)$ where T is a caterpillar tree. Luckily, such independent sets have a simple combinatorial description. Given a subset
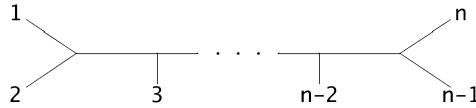
Fig. 5. The caterpillar Cat(n).

$S \subseteq \binom{[n]}{2}$, we let $G(S)$ denote the graph with vertex set $[n]$ and edge set S. For each n, we let Cat(n) denote the caterpillar with cherries 1, 2 and $n-1$, n such that the other leaf labels increase from 3 to $n-2$ along the path from the 1, 2 cherry to the $n-1$, n cherry. This is depicted in Fig. 5. Recall that a *closed trail* in a graph is a sequence of vertices $v_0, \ldots, v_k$ such that each (v_i, v_{i+1}) is an edge (indices taken mod $k+1$), no edge is repeated, and $v_0 = v_k$.

Proposition 3.4. *Let* $S \subseteq \binom{[n]}{2}$. *Then S is independent in $M(\mathrm{Cat}(n))$ if and only if $G(S)$ contains no closed trail with alternating vertices. That is, no closed trail of the form $v_0, v_1, \ldots, v_{2k-1}$ where $v_{2i-1}, v_{2i+1} < v_{2i}$ for each $i = 0, \ldots, k-1$ where indices are taken mod $2k$.*

Proof. Let $A_{\mathrm{Cat}(n)}$ be the matrix whose columns are given by the $\lambda_{\mathrm{Cat}(n)}(ij)$s. We claim that if $x \in \mathbb{R}^{\binom{[n]}{2}}$ then $x \in \mathrm{rowspan}(A_{\mathrm{Cat}(n)})$ if and only if $x_{ik} + x_{jl} - x_{il} - x_{jk} = 0$ for all $i < j < k < l$. To see this, note that $\mathrm{rowspan}(A_{\mathrm{Cat}(n)})$ is equal to the linear hull of the collection of tree metrics with topology Cat(n). If x is a tree metric with topology Cat(n), then each quartet $i < j < k < l$ must satisfy $x_{ij} + x_{kl} < x_{ik} + x_{jl} = x_{il} + x_{jk}$. Since the topology of a tree metric is determined by its quartets [15] the relations $x_{ik} + x_{jl} - x_{il} - x_{jk} = 0$ define the linear hull of the set of tree metrics with topology Cat(n).

Now let B be the vertex-edge incidence matrix of a complete bipartite graph on n vertices. That is, the columns of B are indexed by the set $[n] \times [n]$, the rows are indexed by two disjoint copies of $[n]$, and the column corresponding to (i, j) has 1s at the entries corresponding to i in the first copy of $[n]$ and j in the second copy, and 0s elsewhere. Then $x \in \mathrm{rowspan}(B)$ if and only if $x_{(i,k)} + x_{(j,l)} - x_{(i,l)} - x_{(j,k)} = 0$ for all i, j, k, l.

Notice that any functional vanishing on $\mathrm{rowspan}(A_{\mathrm{Cat}(n)})$ can be associated to a functional vanishing on $\mathrm{rowspan}(B)$ in a coordinate-wise way. In particular, for $i < j$ we associate each coordinate ij in $\mathrm{rowspan}(A_{\mathrm{Cat}(n)})$ with the coordinate (i, j) in $\mathrm{rowspan}(B)$. This means that $M(\mathrm{Cat}(n))$, which is the column matroid of $A_{\mathrm{Cat}(n)}$, is the restriction of the column matroid of B to ground set $\{x_{(i,j)}\}$ where $i < j$.

The column matroid of B is the polygon matroid on the complete bipartite graph on two disjoint copies of $[n]$. To realize $M(\mathrm{Cat}(n))$ as a submatroid, we restrict to the collection of edges (i, j) from $\{1, \ldots, n-1\}$ to $\{2, \ldots, n\}$ such that $i < j$. Closed trails supported on such edges in this bipartite graph correspond exactly to alternating closed trails in the complete graph on vertex set $[n]$. □

Remark 3.5. We now have a purely combinatorial proof that the problem of deciding whether a given $S \subseteq \binom{[n]}{2}$ is independent in $M(\mathcal{T}_n)$ is in the complexity class NP. Namely,

let H_n be the bipartite graph on partite sets $A = \{1, \ldots, n - 1\}$ and $B = \{2, \ldots, n\}$ where (i, j) is an edge from A to B if and only if $i < j$. By Proposition 3.4, $S \subseteq \binom{[n]}{2}$ is independent in $M(\mathrm{Cat}_n)$ if and only if the corresponding edges in H_n form no cycles. Therefore, by Lemma 3.3, a polynomial-time verifiable certificate that a given $S \subseteq \binom{[n]}{2}$ is independent in $M(\mathcal{T}_n)$ is a permutation σ of $[n]$ such that the edges $\{(\sigma(i), \sigma(j),$ $(i, j) \in S\}$ of H_n induce no cycles.

Recall that an acyclic orientation of a graph is an assignment of directions to each edge in a way such that produces no directed cycles. An alternating closed trail in a directed graph is a closed trail $v_0, \ldots, v_k$ such that each pair of adjacent edges $v_{i-1}v_i$ and $v_i v_{i+1}$ have opposite orientations. We remind the reader of Proposition 2.6 which implies that if $S \subseteq \binom{[n]}{2}$ is independent in $M(\mathcal{T}_n)$, then one may arbitrarily specify the distances among pairs in S and still be able to extend the result to a tree metric. We are now ready to prove Theorem 3.6 which we restate below using the language of this section.

Theorem 3.6. *Let $S \subseteq \binom{[n]}{2}$. Then S is independent in $M(\mathcal{T}_n)$ if and only if some acyclic orientation of $G(S)$ has no alternating closed trails.*

Proof. Assume that some acyclic orientation of $G(S)$ has no alternating closed trails. Fix such an acyclic orientation. Choose a permutation $\sigma : [n] \to [n]$ such that edge ij is oriented from i to j if and only if $\sigma(i) < \sigma(j)$. Then by Proposition 3.4, applying σ to the vertices of $G(S)$ gives an independent set in $M(\mathrm{Cat}(n))$. Since independence in $M(\mathcal{T}_n)$ is invariant under permutation of the leaves, this implies that S is independent in $M(\mathcal{T}_n)$.

Now assume that every acyclic orientation of $G(S)$ has an alternating closed trail. Then there does not exist any permutation of $[n]$ that produces an independent set of $M(\mathrm{Cat}(n))$ when applied to S. Therefore S is not independent in $M(T)$ for any caterpillar T. Lemma 3.3 then implies that S is not independent in $M(\mathcal{T}_n)$. $\quad\square$

4. Rank two matrices

We show how Theorem 3.6 immediately characterizes the algebraic matroid underlying the set of $n \times n$ skew-symmetric matrices of rank at most 2. We then use this to characterize the algebraic matroid underlying the set of general $m \times n$ matrices of rank at most 2.

Denote by $\mathcal{S}_r^n(\mathbb{K})$ the collection of $n \times n$ skew-symmetric $\mathbb{K}$-matrices of rank at most r and denote by $\mathcal{M}_r^{m \times n}(\mathbb{K})$ the collection of $m \times n$ $\mathbb{K}$-matrices of rank at most r. Both are irreducible algebraic varieties. We will only be concerned with these sets when $\mathbb{K}$ is $\mathbb{C}$ or $\mathbb{C}\{\{t\}\}$. Lemma 2.1 implies that limited to these choices, the algebraic matroid does not depend on the ground field and so we will suppress $\mathbb{K}$ from notation.

Proposition 4.1. *The tropicalization of the set of $n \times n$ skew symmetric $\mathbb{C}\{\{t\}\}$-matrices of rank at most 2 is the set of all tree metrics on n leaves. That is, $\mathrm{trop}(\mathcal{S}_2^n(\mathbb{C})) = \mathcal{T}_n$.*

Proof. It is shown in [17] that $\mathrm{trop}(\mathrm{Gr}_{2,n}) = \mathcal{T}_n$. Therefore the proposition follows from the claim that the projection of $\mathcal{S}_2^n(\mathbb{C})$ onto the upper-triangular coordinates is $\mathrm{Gr}_{2,n}$. This claim seems to be a well-known fact, but it is difficult to find a precise reference so we give a proof here.

Let $\{x_{ij} : 1 \leq i < j \leq n\}$ be indeterminates and let M be the $n \times n$ skew-symmetric matrix whose ij entry is x_{ij} whenever $i < j$. Let $I_{2,n} \subseteq \mathbb{C}\{\{t\}\}[x_{ij} : 1 \leq i < j \leq n]$ be the radical of the ideal generated by the 3×3 minors of M. The variety of $I_{2,n}$ is the projection onto the upper-triangular coordinates of the set of all $n \times n$ skew-symmetric matrices of rank at most 2. Moreover, $I_{2,n}$ contains all principal 4×4 minors of M, and these polynomials are the squares of the canonical generating set of $\mathrm{Gr}_{2,n}$. Therefore $\mathrm{Gr}_{2,n}$ contains the variety of $I_{2,n}$. Since $\mathrm{Gr}_{2,n}$ is irreducible, as it is parameterized by the 2×2 determinants of a generic $2 \times n$ matrix, it can be seen to be equal to the variety of $I_{2,n}$ by showing that the variety of $n \times n$ skew-symmetric matrices of rank at most 2 has dimension at least that of $\mathrm{Gr}_{2,n}$. Since the dimension of $\mathcal{T}_n$ is $2n - 3$ (see e.g. [15]), Theorem 2.3 implies that $\mathrm{Gr}_{2,n}$ has dimension $2n - 3$ as well. We can see that the set of skew-symmetric matrices of rank at most 2 has at least this dimension because in constructing such a matrix we may arbitrarily specify all but the first entry of the first row and all but the first and second entries of the second row. $\square$

Theorem 4.2. *Let $S \subseteq \binom{[n]}{2}$. Then S is independent in $M(\mathcal{S}_2^n)$ if and only if some acyclic orientation of $G(S)$ has no alternating closed trails.*

Proof. This follows from Lemma 2.1 and Proposition 4.1 and Theorem 3.6. $\square$

Remark 4.3. The same combinatorial polynomial-time verifiable certificate as in Remark 3.5 works to confirm that a given $S \subseteq \binom{[n]}{2}$ is independent in $M(\mathcal{S}_2^n)$.

The dimension of $\mathcal{S}_2^n$ is $2n - 3$. Therefore the rank of the matroid $M(\mathcal{S}_2^n)$ is also $2n - 3$. Also note that $M(\mathcal{S}_2^k)$ is a restriction of $M(\mathcal{S}_2^n)$ whenever $k < n$. It follows that if S is a basis of $M(\mathcal{S}_2^n)$, then $G(S)$ has $2n - 3$ edges and each subgraph of size k has at most $2k - 3$ edges. It is a famous theorem of Laman [13] that graphs satisfying these constraints are exactly the minimal graphs which are generically infinitesimally rigid in the plane. These graphs are often called "Laman graphs." One might wonder whether all Laman graphs are bases of $M(\mathcal{S}_2^n)$ but this is not the case. A counterexample is $K_{3,3}$, the complete bipartite graph on two partite sets of size three (see Fig. 6). This is a Laman graph but it is dependent in $M(\mathcal{S}_2^n(\mathbb{C}))$ because the coordinates specified by $K_{3,3}$ are the entries in a 3×3 submatrix. Such a submatrix must be singular in a matrix of rank 2 and so those entries satisfy a polynomial relation. Alternatively, one could appeal to Theorem 4.2 and check that every acyclic orientation of $K_{3,3}$ induces an alternating cycle.
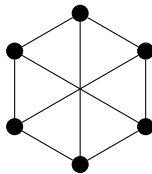
Fig. 6. $K_{3,3}$, a Laman graph that is not a basis of $M(\mathcal{S}_2^n)$.

As we will see in the proof of the following theorem, $\mathcal{M}_2^{m \times n}$ can be realized as a co-ordinate projection of $\mathcal{S}_2^k$ for any $k \geq m + n$. This implies that the matroid $M(\mathcal{M}_2^{m \times n})$ can be realized as the restriction of such a $M(\mathcal{S}_2^k)$. We can use this fact to give a characterization of the algebraic matroid of $M(\mathcal{M}_2^{m \times n})$, whose bases are also the minimally $(2,2)$-rigid graphs in [10]. The ground set of $M(\mathcal{M}_2^{m \times n})$ can be associated with the edges in the complete bipartite graph with partite vertex sets $[m]$ and $[n]$, denoted $K_{m,n}$.

Theorem 4.4. *Let S be a collection of edges of $K_{m,n}$. Then S is independent in $M(\mathcal{M}_2^{m \times n})$ if and only if some acyclic orientation of $G(S)$ has no alternating closed trails.*

Proof. Define $k := m + n$. We claim that $\mathcal{M}_2^{m \times n}$ is a coordinate projection of $\mathcal{S}_2^k$. To see this, first let $A \in \mathcal{M}_2^{m \times n}$. Then choose $u_1, u_2 \in \mathbb{C}^{m \times 1}$ and $v_1, v_2 \in \mathbb{C}^{1 \times n}$ such that $A = u_1 v_1 - u_2 v_2$. Then define the following $k \times k$ matrix

$$B := \begin{pmatrix} u_1 \\ v_2^T \end{pmatrix} \begin{pmatrix} u_2^T & v_1 \end{pmatrix} - \begin{pmatrix} u_2 \\ v_1^T \end{pmatrix} \begin{pmatrix} u_1^T & v_2 \end{pmatrix}.$$

Note that B is skew–skew symmetric of rank at most 2, and its upper-right block is equal to A. Thus the claim is proven and therefore $M(\mathcal{M}_2^{m \times n})$ is a restriction of $M(\mathcal{S}_2^k)$. The result then follows from Theorem 4.2. $\square$

Remark 4.5. The content of Remarks 3.5 and 4.3 can be adapted to give a combinatorial polynomial-time verifiable certificate that a given $S \subseteq [m] \times [n]$ is independent in $M(\mathcal{M}_2^{m \times n})$. Namely, assuming $m > n$, we can translate S into a subset T of $\binom{[m+n]}{2}$ by mapping each $(i,j) \in S$ to the pair $\{i+m, j\}$. Then the construction given in Remark 3.5 can be applied to T.

Acknowledgments

References

[1] Louis J. Billera, Susan P. Holmes, Karen Vogtmann, Geometry of the space of phylogenetic trees, Adv. in Appl. Math. 27 (4) (2001) 733–767.

[2] Emmanuel J. Candes, Yaniv Plan, Matrix completion with noise, Proc. IEEE 98 (6) (2010) 925–936.

[3] Andreas Dress, Katharina Huber, Mike Steel, A matroid associated with a phylogenetic tree, Discrete Math. Theor. Comput. Sci. 16 (2) (2014) 41–55.

[4] Andreas W.M. Dress, Katharina T. Huber, Mike Steel, 'Lassoing' a phylogenetic tree I: basic properties, shellings, and covers, J. Math. Biol. 65 (1) (2012) 77–105.

[5] Martin Farach, Sampath Kannan, Tandy Warnow, A robust model for finding optimal evolutionary trees, Algorithmica 13 (1–2) (1995) 155–179.

[6] Alain Guénoche, Bruno Leclerc, The triangles method to build X-trees from incomplete distance matrices, RAIRO Oper. Res. 35 (2) (2001) 283–300.

[7] Alain Guénoche, Bruno Leclerc, Vladimir Makarenkov, On the extension of a partial metric to a tree metric, Discrete Math. 276 (1) (2004) 229–248.

[8] Bill Jackson, Tibor Jordán, Shin-ichi Tanigawa, Combinatorial conditions for the unique completability of low-rank matrices, SIAM J. Discrete Math. 28 (4) (2014) 1797–1819.

[9] Bill Jackson, Tibor Jordán, Shin-ichi Tanigawa, Unique low rank completability of partially filled matrices, J. Combin. Theory Ser. B 121 (2016) 432–462.

[10] Gil Kalai, Eran Nevo, Isabella Novik, Bipartite rigidity, Trans. Amer. Math. Soc. 368 (8) (2016) 5515–5545.

[11] Franz Király, Louis Theran, Ryota Tomioka, The algebraic combinatorial approach for low-rank matrix completion, J. Mach. Learn. Res. 16 (2015) 1391–1436.

[12] Franz J. Király, Zvi Rosen, Louis Theran, Algebraic matroids with graph symmetry, arXiv preprint, arXiv:1312.3777, 2013.

[13] Gerard Laman, On graphs and rigidity of plane skeletal structures, J. Engrg. Math. 4 (4) (1970) 331–340.

[14] Diane Maclagan, Bernd Sturmfels, Introduction to Tropical Geometry, vol. 161, American Mathematical Soc., 2015.

[15] Charles Semple, Mike A. Steel, Phylogenetics, vol. 24, Oxford University Press on Demand, 2003.

[16] Amit Singer, Mihai Cucuringu, Uniqueness of low-rank matrix completion by rigidity theory, SIAM J. Matrix Anal. Appl. 31 (4) (2010) 1621–1641.

[17] David Speyer, Bernd Sturmfels, The tropical grassmannian, Adv. Geom. 4 (3) (2004) 389–411.

[18] Michael S. Waterman, Temple F. Smith, On the similarity of dendrograms, J. Theoret. Biol. 73 (4) (1978) 789–800.

[19] Josephine Yu, Algebraic matroids and set-theoretic realizability of tropical varieties, arXiv preprint, arXiv:1506.01427, 2015.