RESEARCH STATEMENT
ALGEBRAIC MATROIDS: STRUCTURE AND APPLICATIONS

ZVI ROSEN

Algebraic matroids are **combinatorial** objects that can be extracted from **geometric** problems, describing the independence structure on the coordinates. In this proposal, algebraic matroids are used to analyze applied problems, and their structure is explored. We give historical background to this topic, then set forth four projects.

1. Introduction

Matroids were introduced in the early 20th century as a way of uniting disparate notions of "independence" from across mathematics. Among these notions were linear independence of vectors and graphic independence – defined by acyclicity on the subgraph corresponding to a set of edges. Algebraic independence over a field k, defined by the non-existence of polynomial relations with coefficients in k among elements of a set, is another such notion. Van der Waerden first defined algebraic matroids [vdW91], and they were studied by MacLane in an early matroid paper [ML38]. They were largely neglected after then until the 1970's. Ingleton and Main showed the existence of non-algebraic matroids in [IM75], and Bernt Lindström treated algebraic representability in several papers (e.g. [Lin83, Lin86, Lin88]), as did some others ([DL87, Gor88]).

These objects deserved to be studied in their own right: instead of thinking of elements in a field extension, let us think of coordinates of the function field of a variety. Then, any geometric problem involving special coordinates (like varieties describing physical quantitites in the real world, or varieties with combinatorially-defined ideals) boils down to an algebraic matroid. For this reason, in the past few years alone, low-rank matrix completion [KTTU12], rigidity theory [KRT13], probability distributions [KR14] and chemical reaction networks [MRHB14] have all benefited from insights given through algebraic matroids. The algorithmic tools we developed in [Ros14] make algebraic matroids readily computable and ripe for experimentation.

Exciting developments are taking place in matroid theory, and particularly at its interface with algebraic geometry. Tropical geometry (a combinatorial version of algebraic geometry) and matroid theory are continually finding deep connections (e.g. [FM12]). Huh and Katz used the homology of a particular toric variety to prove Rota's conjecture regarding log-concavity of characteristic polynomials for representable matroids [HK12]. Another conjecture of Rota, regarding the forbidden-minor criteria for representability over finite fields has also recently fallen [GGW14]. Further study of ***algebraic*** matroids can shed light on its role in this rich tapestry of research.

2. Proposed Research

My research plan breaks down into two areas – (1) delving into specific applications of algebraic matroids, and (2) further exploring the structure of algebraic matroids. I shall pursue the following four projects:

(1) Matroids for Statistical Models. [Applications]
(2) Chemical Reaction Matroids. [Applications]

(3) Algebra of Circuit Ideals and Basis Field Extensions. [Structure]
(4) Geometry of Non-Matroidal Loci. [Structure]

2.1. **Matroids for Statistical Models.** The field of algebraic statistics (reference: [DSS09]) is a dynamic area of research. In particular, graphical models study discrete events and their mutual independence structure. Statistical rankings studied in [SW12] consider parametrizations of a probability distribution on the symmetric group (or some variant). Both of these topics benefit from computation and analysis of associated algebraic matroids.

2.1.1. *Joint Probability Distributions.* An example of a simple graphical model is a pair of events that are both conditioned on r events, and are otherwise conditionally independent. This is equivalent to the matrix (p_{ij}) of joint probabilities having nonnegative rank r. A relaxed assumption is that this matrix has rank at most r; this set is defined by:

$$I = \langle (r+1) \times (r+1)\text{-minors}, \sum_{i,j} p_{ij} - 1 \rangle.$$

The circuits of the algebraic matroid allow us to rule out independence of events when only partial measurements are available. The bases allow us to reconstruct (up to finite choice) full probability distributions given the partial measurements. The fact that we are working over $\mathbb{R}$ and that we expect to find probability values in the interval $[0,1]$ complicates the simple algebraic structure with a semi-algebraic slant.

Problem 2.1. a) Characterize the bases and circuits of the matroid defined by the conditional independence model. b) Determine the defining equations and inequalities of the semi-algebraic set obtained as the projected image of the intersection with the probability simplex.

When $r = 1$, the variables are independent. In joint work with K. Kubjas, we completely answered the problem; the graph notation considers matrix entries as edges of a bipartite graph with vertices = rows $\cup$ columns.

Theorem 2.2 (Kubjas-Rosen)**.** *Bases of the matroid are two-component spanning forests. The projection of all independence matrices onto an n-tree forest of entries has boundary given by coordinate hyperplanes and a single nonlinear hypersurface $\sum_{i=1}^{n} \sqrt{b_i} = 1$.*

We obtained analogous, though incomplete, results for probability tensors. This research is continuing in collaboration with T. Kahle and K. Kubjas [KKR14] with the following questions:

Problem 2.3. Find the algebraic boundary of a) projections onto the diagonal for matrices of (nonnegative) rank r, b) projections onto a set with support on $< r$ rows and columns.

2.1.2. *Statistical Rankings.* Throughout this section, our reference is [SW12]. In any context of modeling data, the circuits of the matroid are of utmost practical importance; evaluation of circuits is a criterion for model rejection using minimal data points. Many examples of statistical ranking models are toric; for instance, the Birkhoff model, the Csiszar model, and the Bradley-Terry model. The literature on toric circuits is useful there. This project will focus on non-toric statistical models, in particular the Plackett-Luce model, which has parameters θ_i and model: $p_\pi \mapsto \prod_{i=1}^{n-1} (\sum \theta_{\pi(i)})^{-1}$.

Example 2.4. The Plackett-Luce model for $n = 4$ has matroid, computed in [Ros14], with a 3-dimensional affine representation, given by a perturbed permutohedron (Figure 1).

Problem 2.5. Determine the algebraic matroids for the Plackett-Luce model, for all n.

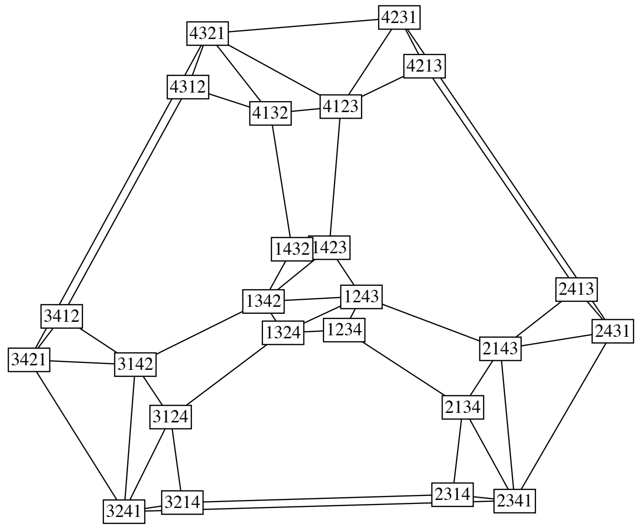

Figure 1. Permutahedron with four hexagonal faces triangulated.

2.2. **Chemical Reaction Matroids.** A chemical reaction network (CRN) can be defined as a directed graph, whose vertices are labeled with chemical complexes, and whose directed edges represent a chemical reaction that takes place. Each edge is labeled with a reaction rate, and the resulting network defines the system of ODE's governing the CRN dynamics.

Several combinatorial and geometric objects have been defined in relation to a CRN on a directed graph G. The moduli ideal M_G lives in the polynomial ring on edges of the graph – $\mathbb{Q}[\kappa_{ij} : (i,j) \in G]$ – while the ideal T_G sits in the polynomial ring whose variables correspond to chemical species – $\mathbb{Q}[c_i]$ – or vertices of the graph. Craciun et al showed in [CDSS09, Theorem 7] that the rate reactions parametrized by the variety $V(M_G)$ determine whether T_G has a positive solution. The (oriented) matroid $\mathcal{M}(G)$ associated to the directed graph has been used in various projects; for example, in [MBOS14] to detect fast flux modules. Meanwhile the algebraic matroid associated to the ideal of relations among the c_i, denoted by I_G, was used for model discrimination in [MRHB14].

Problem 2.6. Does the analogy between the ideals M_G and T_G extend to the matroids $\mathcal{M}(G)$ and $\mathcal{M}(I_G)$? In particular, do properties like connectivity in $\mathcal{M}(G)$ place constraints on $\mathcal{M}(I_G)$?

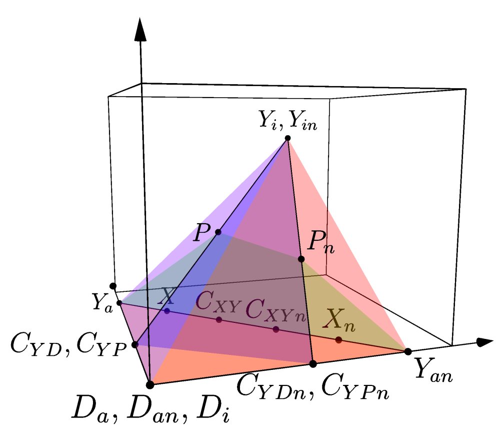

Figure 2. Matroid minor for the Wnt shuttle model, from [MRHB14].

My previous work on CRNs ([MRHB14]) had a focus on experimental design, enabling *model discrimination* and *state reconstruction* with minimal measurements. Further results in the structure section of this proposal will allow us to perform these tasks in new scenarios.

Problem 2.7. Find optimal or locally optimal measurement sets for large CRNs, where combinatorial complexity precludes computing full lists of bases and circuits.

2.3. **Algebra of Circuit Ideals and Base Degrees.** Every base of a matroid corresponds to a surjective coordinate projection with finite fibers; similarly, every circuit corresponds to a coordinate projection whose image is a hypersurface containing no coordinate subspace [KRT13].

Definition 2.8. Let $\mathcal{B}$ and $\mathcal{C}$ denote the sets of bases and circuits of an algebraic matroid $\mathcal{M}$ defined by a prime ideal I. The *circuit polynomial* of $C \in \mathcal{C}$ is the unique up to scalar polynomial relation among the elements of C. The *base degree* of $B \in \mathcal{B}$ is the degree of the field extension $[k(E) : k(x \mid x \in B)]$.

Symbolic algebra, and more recently, numerical algebraic geometry can be utilized to compute the degree of these fibers (resp. hypersurfaces), but the computation is heavy and for large cases, combinatorial complexity makes a naive approach infeasible.

Problem 2.9. Devise efficient algorithms to: a) locate bases with maximal base degree. b) locate circuits whose circuit polynomial is of maximal degree.

Computational results I obtained in [Ros14] indicate that these high-degree bases and circuits have connections to underlying symmetries in the corresponding varieties: see Figure 3 for an example. Learning a way to crawl along the basis-exchange graph towards a maximal-degree base would be of great interest.

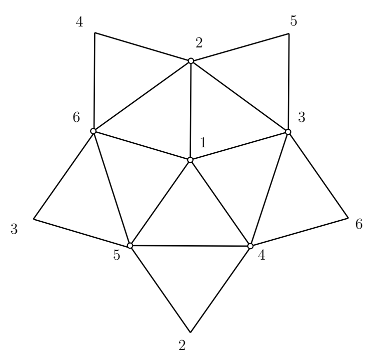

Figure 3. High-degree base of $Gr(3, 6)$, for which S_6 induces the outer automorphism.

Recent work by Aluffi [Alu14] uses Fulton-MacPherson intersection theory to explore the projections of rank loci. Properties of the Tutte group associated to an algebraic matroid, as defined by Dress and Wenzel in [DW89], indicate local relationships among base degree. We will explore both of these avenues for general algebraic matroids. We will first tackle special cases: a) toric ideals, b) ideals generated by quadratic forms, c) ideals generated by minors of special matrices. We hope to obtain results for degree bounds on circuit polynomials and base extensions. Understanding the base and circuit degrees may lead to progress on a $\geq$40-year old question:

Problem 2.10 ([Wel76])**.** Is the class of algebraic matroids closed under taking duals?

Circuit polynomials are the *support-minimal* elements of the prime ideal associated to a variety; however, they are not a generating set. For toric ideals, the following theorem characterizes the ideal generated by the circuits of the algebraic matroid:

Theorem 2.11 (Eisenbud-Sturmfels [ES96])**.**

$$Rad(I_{\mathcal{C}}) = I_{\mathcal{L}} \text{ and } Ass(I_{\mathcal{C}}) \subseteq \{I_\sigma : \sigma \text{ face of } pos(A)\}.$$

This is true for the circuits of a toric ideal, but the question has not been studied for general prime ideals. Computational tools will be helpful in building a collection of examples, which could be aided by undergraduate research. In work with Király and Theran [KRT13], we defined the multihomogenization of an ideal. The multihomogenization of an ideal I is an ideal J in a ring with twice as many variables. We proved:

$$\text{Circuits of } I \subseteq \text{Minimal Generating Set of } J$$

Problem 2.12. For an ideal I, describe the relationship among the universal Gröbner basis, the circuit set, and the minimal generating sets.

The universal Gröbner basis necessarily contains the full list of circuits, since each arises from an elimination ordering. It would be instructive and valuable to have examples of ideals that behave nicely in this respect. Robust toric ideals studied in [BR13] are ideals for which the universal Gröbner basis is also a minimal generating set. Other well-understood examples: *ideals generated by linear forms*: circuit set = universal Gröbner basis [Stu96]; *ideals of maximal minors*: minimal generating set = circuit set = universal Gröbner basis [BZ93].

2.4. **Geometry of Non-Matroidal Loci.** Every algebraic matroid over a field k of characteristic zero may be represented as a linear matroid over a field $k(T)$, where T is a finite set of transcendentals [Oxl11]. This representation is obtained by passing to the vector space of differentials. Bernt Lindstrom proved that it is possible to substitute some elements of the ground field for the set of transcendentals [Lin87].

Problem 2.13. Characterize the set of points of a variety for which substituting the coordinates for T yields the *wrong* matroid.

This question is interesting both for computing the algebraic matroid, and for understanding the way that the tangent space lines up with coordinate subspaces at different points of a variety.

Theorem 2.14 (Rosen)**.** *The set of points for which the matroid is wrong is an algebraic variety that is cut out by the intersection of principal ideals generated by nonzero minors of the Jacobian.*

This gives us a way of characterizing the *non-matroidal locus* for any particular variety. However, this doesn't give full understanding; I would like to answer the following questions: a) Given a generic variety (in the moduli sense) of dimension n and degree d, how can we describe the non-matroidal locus? b) In particular, how many irreducible components of what degree will it have?

References

[Alu14] Paolo Aluffi. Degrees of projections of rank loci. *Pre-print.*, 2014. http://arxiv.org/abs/1408.1702.

[BR13] Adam Boocher and Elina Robeva. Robust toric ideals. *Pre-print.*, 2013. http://arxiv.org/abs/1304.0603.

[BZ93] David Bernstein and Andrei Zelevinsky. Combinatorics of maximal minors. *J. Algebraic Combin.*, 2(2):111–121, 1993.

[CDSS09] Gheorghe Craciun, Alicia Dickenstein, Anne Shiu, and Bernd Sturmfels. Toric dynamical systems. *J. Symbolic Comput.*, 44(11):1551–1565, 2009.

[DL87] Andreas Dress and Lászlo Lovász. On some combinatorial properties of algebraic matroids. *Combinatorica*, 7(1):39–48, 1987.

[DSS09] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on algebraic statistics*, volume 39 of *Oberwolfach Seminars*. Birkhäuser Verlag, Basel, 2009.

[DW89] Andreas W. M. Dress and Walter Wenzel. Geometric algebra for combinatorial geometries. *Adv. Math.*, 77(1):1–36, 1989.

[ES96] David Eisenbud and Bernd Sturmfels. Binomial ideals. *Duke Math. J.*, 84(1):1–45, 1996.

[FM12] Alex Fink and Luca Moci. Matroids over a ring. *Pre-print.*, 2012. http://arxiv.org/abs/1209.6571.

[GGW14] Jim Geelen, Bert Gerards, and Geoff Whittle. Solving rota's conjecture. *Notices of the AMS*, 61(7):736–743, 2014.

[Gor88] Gary Gordon. Algebraic characteristic sets of matroids. *J. Combin. Theory Ser. B*, 44(1):64–74, 1988.

[HK12] June Huh and Eric Katz. Log-concavity of characteristic polynomials and the bergman fan of matroids. *Pre-print.*, 2012. http://arxiv.org/abs/1104.2519.

[IM75] Aubrey Ingleton and Roger Main. Non-algebraic matroids exist. *Bull. London Math. Soc.*, 7:144–146, 1975.

[KKR14] Thomas Kahle, Kaie Kubjas, and Zvi Rosen. Probability tensor completion. *In Preparation*, 2014.

[KR14] Kaie Kubjas and Zvi Rosen. Matrix completion for the independence model. *Pre-print.*, 2014. http://arxiv.org/abs/1407.3254.

[KRT13] Franz Király, Zvi Rosen, and Louis Theran. Algebraic matroids with graph symmetry. *ArXiv e-prints*, December 2013.

[KTTU12] Franz Király, Louis Theran, Ryota Tomioka, and Takeaki Uno. The algebraic combinatorial approach for low-rank matrix completion. *Pre-print.*, 2012. http://arxiv.org/abs/1211.4116.

[Lin83] Bernt Lindström. The non-Pappus matroid is algebraic. *Ars Combin.*, 16(B):95–96, 1983.

[Lin86] Bernt Lindström. The non-Pappus matroid is algebraic over any finite field. *Utilitas Math.*, 30:53–55, 1986.

[Lin87] Bernt Lindström. A reduction of algebraic representations of matroids. *Proc. Amer. Math. Soc.*, 100(2):388–389, 1987.

[Lin88] Bernt Lindström. Matroids, algebraic and nonalgebraic. In *Algebraic, extremal and metric combinatorics, 1986 (Montreal, PQ, 1986)*, volume 131 of *London Math. Soc. Lecture Note Ser.*, pages 166–174. Cambridge Univ. Press, Cambridge, 1988.

[MBOS14] Arne C. Müller, Frank J. Bruggeman, Brett G. Olivier, and Leen Stougie. Fast flux module detection using matroid theory. In *Research in Computational Molecular Biology - 18th Annual International Conference, RECOMB 2014, Pittsburgh, PA, USA, April 2-5, 2014, Proceedings*, pages 192–206, 2014.

[ML38] Saunders Mac Lane. A lattice formulation for transcendence degrees and p-bases. *Duke Math. J.*, 4(3), 1938.

[MRHB14] Adam MacLean, Zvi Rosen, Heather Harrington, and Helen Byrne. Parameter-free methods distinguish wnt pathway models and guide design of experiments. *Pre-print.*, 2014. http://arxiv.org/abs/1409.0269.

[Oxl11] James Oxley. *Matroid theory*, volume 21 of *Graduate Texts in Mathematics*. Oxford University Press, 2011.

[Ros14] Zvi Rosen. Computing algebraic matroids. *Pre-print.*, 2014. http://arxiv.org/abs/1403.8148.

[Stu96] Bernd Sturmfels. *Gröbner bases and convex polytopes*, volume 8 of *University Lecture Series*. American Mathematical Society, Providence, RI, 1996.

[SW12] Bernd Sturmfels and Volkmar Welker. Commutative algebra of statistical ranking. *J. Algebra*, 361:264–286, 2012.

[vdW91] Bartel Leendert van der Waerden. *Algebra. Vol. I.* Springer-Verlag, New York, 1991. Translated from the second German edition.

[Wel76] D. J. A. Welsh. *Matroid theory.* Academic Press [Harcourt Brace Jovanovich Publishers], London, 1976. L. M. S. Monographs, No. 8.