MAXIMUM LIKELIHOOD DUALITY FOR DETERMINANTAL VARIETIES

JAN DRAISMA AND JOSE RODRIGUEZ

Abstract. In a recent paper, Hauenstein, Sturmfels, and the second author discovered a conjectural bijection between critical points of the likelihood function on the complex variety of matrices of rank r and critical points on the complex variety of matrices of co-rank $r-1$. In this paper, we prove that conjecture for rectangular matrices and for symmetric matrices, as well as a variant for skew-symmetric matrices.

1. Introduction and results

For an $m \times n$-data table $U = (u_{ij}) \in \mathbb{N}^{m\times n}$, we define the *likelihood function* $\ell_U : \mathbb{T}^{m\times n} \to \mathbb{T}$, where $\mathbb{T} = \mathbb{C}^*$ is the complex one-dimensional torus, as $\ell_U(Y) = \prod_{ij} y_{ij}^{u_{ij}}$ for $Y = (y_{ij})_{ij} \in \mathbb{T}^{m\times n}$. This terminology is motivated by the following observation. If Y is a matrix with positive real entries adding up to 1, interpreted as the joint probability distribution of two random variables taking values in $[m] := \{1, \ldots, m\}$ and $[n] := \{1, \ldots, n\}$, respectively, then up to a multinomial coefficient depending only on U, $\ell_U(Y)$ is the probability that when independently drawing $\sum_{i,j} u_{ij}$ pairs from the distribution Y, the number of pairs equal to (i, j) is u_{ij}. In other words, $\ell_U(Y)$ is the likelihood of Y, given observations recorded in the table U. A standard problem in statistics is to *maximize* $\ell_U(Y)$.

Without further constraints on Y this maximization problem is easy: it is uniquely solved by the matrix Y obtained by scaling U to lie in said probability simplex. But various meaningful statistical models require Y to lie in some *subvariety* X of $\mathbb{T}^{m\times n}$. For instance, in the model where the first and second random variable are required to be independent, one takes X equal to the intersection of the variety of matrices of rank 1 with the hyperplane $\sum_{ij} y_{ij} = 1$ supporting the probability simplex. Taking mixtures of this model, one is also led to intersect said hyperplane with the variety of rank-r matrices.

For general X, the maximum-likelihood estimate is typically much harder to find (though in the independence model it is still well-understood). One reason for this is that the restriction of ℓ_U to X may have many critical points. Under suitable assumptions, this number of critical points is finite and independent of U (for sufficiently general U), and is called the *maximum likelihood degree* or *ML-degree* of X. Finiteness and independence of U hold, for instance, for smooth closed subvarieties of tori. In that case, the ML-degree equals the signed topological Euler characteristic of X; see [Huh12] or the more general "non-compact Riemann-Roch

Date: October 2012.

The first author is supported by a Vidi grant from the Netherlands Organisation for Scientific Research (NWO).

The second author is supported by the US National Science Foundation DMS-0943745.

formula" in [FK00]. Finiteness and independence of U also hold for all varieties X studied in this paper [HRS12, HKS05] (which are smooth but not closed, and become closed but singular if one takes the closure).

We take X to be a smooth, irreducible, locally closed, complex subvariety of a torus. Doing so, we tacitly shift attention from the statistical motivation to complex geometry—in particular, we no longer worry whether the critical points counted by the ML-degree lie in the probability simplex or are even real-valued matrices. Similarly, we also no longer restrict ourselves to study integer valued tables but allow our data to vary over the complex numbers

The set of all critical points for varying data matrices U has a beautiful geometric interpretation: Given $P \in X$ and a vector V in the tangent space $T_P X$ to X at P, the derivative of ℓ_U at P in the direction V equals $\ell_U(P) \cdot \sum_{ij} \frac{v_{ij}}{p_{ij}} u_{ij}$. This vanishes if and only if U is perpendicular, in the standard symmetric bilinear form on $\mathbb{C}^{m\times n} = \mathbb{C}^{mn}$, to the entry-wise quotient $\frac{V}{P}$ of V by P. This leads us to define

$$\mathrm{Crit}(X) := \left\{ (P, U) \;\middle|\; \frac{T_P X}{P} \perp U \right\} \subseteq X \times \mathbb{C}^{m\times n},$$

which is called *the variety of critical points* of X in [Huh12], except that there U varies over projective space and the closure is taken. By construction, $\mathrm{Crit}(X)$ is smooth and irreducible, and has dimension mn; indeed, it is a vector bundle over X of rank $mn - \dim X$. The ML-degree of X is well-defined if and only if the projection $\mathrm{Crit}(X) \to \mathbb{C}^{m\times n}$ is dominant, in which case the degree of this rational map is the ML-degree of X.

In this paper, motivated by [HRS12], we consider three choices for X, all given by rank constraints: First, in the *rectangular* case, we order m, n such that $m \leq n$, fix a rank $r \in [m]$, and take X equal to

$$\mathcal{M}_r := \left\{ P \in \mathbb{T}^{m\times n} \;\middle|\; \sum_{ij} p_{ij} = 1 \text{ and } \operatorname{rk} P = r \right\}.$$

Second, in the *symmetric* case, we take $m = n$ and take X equal to

$$\mathcal{SM}_r := \left\{ P = \begin{bmatrix} 2p_{11} & p_{12} & \cdots & p_{1m} \\ p_{12} & 2p_{22} & & \\ \vdots & & \ddots & \\ p_{1m} & & & 2p_{mm} \end{bmatrix} \in \mathbb{T}^{m\times m} \;\middle|\; \begin{matrix} \sum_{i\leq j} p_{ij} = 1 \\ \text{and } \operatorname{rk}(P) = r \end{matrix} \right\}.$$

Third, in the *skew-symmetric* or *alternating* case, we take $m = n$ and, for *even* $r \in [m]$, take X equal to

$$\mathcal{AM}_r := \left\{ P = \begin{bmatrix} 0 & p_{12} & \cdots & p_{1m} \\ -p_{12} & 0 & & \\ \vdots & & \ddots & \\ -p_{1m} & & & 0 \end{bmatrix} \in \mathbb{C}^{m\times m} \;\middle|\; \begin{matrix} \sum_{i<j} p_{ij} = 1, \\ \operatorname{rk}(P) = r, \\ \text{and } \forall i<j : p_{ij} \neq 0 \end{matrix} \right\}.$$

Minor modifications of the likelihood function are needed in the latter two cases: we define as $\ell_U(P) := \prod_{i\leq j} p_{ij}^{u_{ij}}$ in the symmetric case, and as $\ell_U(P) := \prod_{i<j} p_{ij}^{u_{ij}}$ in the alternating case.

In [HRS12], using the numerical algebraic geometry software `bertini` [BHSW06], the ML-degree of $\mathcal{M}_r$ is computed for various values of r, m, n with $r \leq m \leq n$.

	$(m,n) =$	(3,3)	(3,4)	(3,5)	(4,4)	(4,5)	(4,6)	(5,5)
$r=1$		1	1	1	1	1	1	1
$r=2$		10	26	58	191	843	3119	6776
$r=3$		1	1	1	191	843	3119	61326
$r=4$					1	1	1	6776
$r=5$								1

TABLE 1. ML-degrees of $\mathcal{M}_r$ for small values of $r \leq m \leq n$

The numbers are listed in Table 1. Observe that the numbers for rank r and rank $m-r+1$ coincide. The natural conjecture put forward in that paper is that this always holds [HRS12, Conjecture 1.2], and that there is an explicit bijection between the two sets of critical points [HRS12, Conjecture 4.2]. Moreover, similar results were conjectured for symmetric matrices. We will prove these conjectures, for which we use the term *ML-duality* suggested to us by Sturmfels.

Theorem 1 (ML-duality for rectangular matrices and for symmetric matrices)**.** *Fix a rank $r \in [m]$ and let $U \in \mathbb{N}^{m\times n}$ with $m \leq n$ ($m = n$ in the symmetric case) be a sufficiently general data matrix (symmetric in the symmetric case). Then there is an explicit involutive bijection between the critical points of ℓ_U on $\mathcal{M}_r$ (respectively, $\mathcal{SM}_r$) and the critical points of ℓ_U on $\mathcal{M}_{m-r+1}$ (respectively, $\mathcal{SM}_{m-r+1}$). In particular, the ML-degrees of $\mathcal{M}_r$ and $\mathcal{M}_{m-r+1}$ (respectively, $\mathcal{SM}_r$ and $\mathcal{SM}_{m-r+1}$) coincide. Moreover, the product $\ell_U(P)\ell_U(Q)$ is the same for all pairs consisting of a rank-r critical point P and the corresponding rank-$(m-r+1)$ point Q.*

In the alternating case, the ML-dual of $\mathcal{AM}_r$ turns out *not* to be some $\mathcal{AM}_s$ but rather an affine translate of a determinantal variety defined as follows. Let S be the skew $m \times m$-matrix

$$S := \begin{bmatrix} 0 & 1 & \cdots & 1 \\ -1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ -1 & \cdots & -1 & 0 \end{bmatrix},$$

and for even $s \in \{0, \ldots, m-1\}$ consider the variety

$$\mathcal{AM}'_s := \left\{ P \in \mathbb{C}^{m\times m} \;\middle|\; P \text{ skew}, \forall i<j : p_{ij} \neq 0, \text{ and } \operatorname{rk}(S-P) = s \right\}.$$

Note that, unlike in $\mathcal{AM}_r$, the upper triangular entries of $P \in \mathcal{AM}'_s$ are not required to add up to 1.

Theorem 2 (ML-duality for skew matrices)**.** *Fix an even rank $r \in \{2, \ldots, m\}$ and let $U \in \mathbb{N}^{m\times m}$ be a sufficiently general* symmetric *data matrix with zeroes on the diagonal. Let $s \in \{0, \ldots, m-2\}$ be the largest even integer less than or equal to $m-r$. Then there is an explicit involutive bijection between the critical points of ℓ_U on $\mathcal{AM}_r$ and the critical points of ℓ_U on $\mathcal{AM}'_s$. In particular, the ML-degrees of $\mathcal{AM}_r$ and $\mathcal{AM}'_s$ coincide. Moreover, the product $\ell_U(P)\ell_U(Q)$ is the same for all pairs consisting of a rank-r critical point P on $\mathcal{AM}_r$ and the corresponding rank-s point Q on $\mathcal{AM}'_s$.*

The proof is similar in each of the three cases. First, we determine the tangent space to X at a critical point P of ℓ_U for sufficiently general U. It turns out that this space is spanned by certain rank-one or rank-two matrices. Imposing that P be a critical point, i.e., that the derivative of ℓ_U vanishes in each of these low-rank directions leads to the conclusion that a certain matrix Q, determined from P using some involution involving the fixed matrix U, has rank at most $m-r+1$ (or s in the skew case) and is itself a critical point on the variety of matrices of its rank. Letting $k \leq m-r+1$ (respectively, $k \leq s$) be generic rank of the Qs thus obtained, we reverse the roles of P and Q to argue that k must equal $m-r+1$ (respectively, s), thus establishing the result. In the remainder of this paper we fill in the details in each of the three cases, in particular making the involution $P \to Q$ explicit.

Acknowledgments

We thank Bernd Sturmfels for his encouragement and comments on early versions of this paper.

2. Maximum likelihood duality in the rectangular case

Let $m \leq n$ be natural numbers and let $\mathcal{M}_r \subseteq \mathbb{T}^{m\times n}$ denote the variety of $m \times n$-matrices of rank r whose entries sum up to 1. Fix a sufficiently general data matrix $U = (u_{ij})_{ij} \in \mathbb{N}^{m\times n}$, which gives rise to the likelihood function $\ell_U : \mathcal{M}_r \to \mathbb{T}$, $\ell_U(P) = \prod_{i,j} p_{ij}^{u_{ij}}$. Let $P \in \mathcal{M}_r$ be a critical point for ℓ_U, which means that the derivative of ℓ_U vanishes on the tangent space $T_P\mathcal{M}_r$ to $\mathcal{M}_r$ at P. This tangent space equals

$$T_P\mathcal{M}_r = \left\{ X = (x_{ij})_{ij} \in \mathbb{C}^{m\times n} \;\middle|\; X \ker P \subseteq \operatorname{im} P \text{ and } \sum_{ij} x_{ij} = 0 \right\}. \tag{1}$$

Here the first condition ensures that X is tangent at P to the variety of rank-r matrices (see, e.g., [Har92, Example 14.6]) and the second condition ensures that X is tangent to the hyperplane where the sum of all matrix entries is 1.

Given $X \in T_P\mathcal{M}_r$, the derivative of ℓ_U in that direction equals $\ell_U(P) \cdot \sum_{ij} \frac{x_{ij}u_{ij}}{p_{ij}}$, which vanishes if and only if the second factor vanishes. We will now prove that the marginals of P are proportional to those of U (see also [HRS12, Remark 4.6]). We write $\mathbf{1}$ for the all-one vectors in both $\mathbb{C}^m$ and $\mathbb{C}^n$, and use self-explanatory notation such as $u_{i+} := \sum_j u_{ij}$ and $u_{++} := \sum_{ij} u_{ij}$.

Lemma 3. *The column vector $P\mathbf{1}$ is a non-zero scalar multiple of $U\mathbf{1}$ and the row vector $\mathbf{1}^T P$ is a non-zero scalar multiple of $\mathbf{1}^T U$.*

Proof. We prove the first statement; the second statement is proved similarly. We want to show that the 2×2-minors of the $m \times 2$-matrix $[P\mathbf{1}|U\mathbf{1}]$ vanish. We give the argument for the upper minor. Let $X = (x_{ij})$ be the $m \times n$-matrix whose first row equals p_{2+} times the first row of P, whose second row equals $-p_{1+}$ times the second row of P, and all of whose other rows are zero. Then $X \in T_P\mathcal{M}_r$, so that the derivative $\sum_{ij} x_{ij} \frac{u_{ij}}{p_{ij}}$ is zero. On the other hand, substituting X into $\sum_{ij} x_{ij} \frac{u_{ij}}{p_{ij}}$ yields $u_{1+}p_{2+} - u_{2+}p_{1+}$, hence this minor is zero as desired. The scalar multiple in both cases is $\frac{p_{++}}{u_{++}} = \frac{1}{u_{++}}$, which is non-zero. □

Define $Q = (q_{ij})_{ij}$ by $p_{ij}q_{ij} = u_{i+}u_{ij}u_{+j}$. This is going to be our dual critical point, up to a normalization factor that we determine now.

Lemma 4. *The sum* $\sum_{ij} q_{ij}$ *equals* $(u_{++})^3$.

Proof. By Lemma 3 the rank-one matrix Y defined by $y_{ij} = u_{i+}u_{+j}$ has image contained in $\operatorname{im} P$. Hence it satisfies the linear condition $Y \ker P \subseteq \operatorname{im} P$, but not the condition $\sum_{ij} y_{ij} = 0$. Similarly, P itself satisfies $P \ker P \subseteq \operatorname{im} P$, but not $\sum_{ij} p_{ij} = 0$. Hence, we can decompose Y uniquely as $cP + X$ where $c \in \mathbb{C}$ and where X satisfies $X \ker P \subseteq \operatorname{im} P$ and $\sum_{ij} x_{ij} = 0$, i.e., where $X \in T_P\mathcal{M}_r$. Then we have

$$\sum_{ij} q_{ij} = \sum_{ij} \frac{y_{ij}u_{ij}}{p_{ij}} = \sum_{ij} cu_{ij} + \sum_{ij} \frac{x_{ij}u_{ij}}{p_{ij}} = \sum_{ij} cu_{ij} + 0 = cu_{++}$$

by criticality of P. The scalar c equals

$$\frac{\sum_{ij} y_{ij}}{\sum_{ij} p_{ij}} = \frac{\sum_{ij} u_{i+}u_{+j}}{1} = (u_{++})^2,$$

which proves the lemma. □

We will use rank-one matrices in the tangent space $T_P\mathcal{M}_r$. We equip both $\mathbb{C}^m$ and $\mathbb{C}^n$ with their standard symmetric bilinear forms.

Lemma 5. *The tangent space* $T_P\mathcal{M}_r$ *at* P *is spanned by all rank-one matrices* vw^T *satisfying the following two conditions:*

- $v \in \operatorname{im} P$ *or* $w \perp \ker P$*; and*
- $v \perp \mathbf{1}$ *or* $w \perp \mathbf{1}$.

In the proof we will need that $\operatorname{im} P$ is not contained in the hyperplane $\mathbf{1}^\perp$ and that, dually, $\ker P$ does not contain $\mathbf{1}$. These conditions will be satisfied by genericity of U.

Proof. The first condition ensures that the rank-one matrices in the lemma map $\ker P$ into $\operatorname{im} P$, and the second condition ensures that the sum of all entries of those rank-one matrices is zero, so that they lie in $T_P\mathcal{M}_r$, see (1). To show that these rank-one matrices span the tangent space $T_P\mathcal{M}_r$, decompose $\mathbb{C}^m$ as $A \oplus B \oplus C$ where $A \oplus C = \mathbf{1}^\perp$ and $A \oplus B = \operatorname{im} P$. Here we use that $\operatorname{im} P$ is not contained in the hyperplane $\mathbf{1}^\perp$.

Similarly, decompose $\mathbb{C}^n = A' \oplus B' \oplus C'$ where $A' \oplus C'$ is the hyperplane $\mathbf{1}^\perp$ and $A' \oplus B' = (\ker P)^\perp$; here we use the second genericity assumption on P. These spaces have the following dimensions:

$$\begin{array}{lll} \dim A = r-1 & \dim B = 1 & \dim C = m-r \\ \dim A' = r-1 & \dim B' = 1 & \dim C' = n-r. \end{array}$$

The space spanned by the rank-one matrices in the lemma has the space $(B \otimes B') \oplus (C \otimes C')$ as a vector space complement. The dimension of this complement is $1 + (m-r)(n-r)$, which is also the codimension of $\mathcal{M}_r$. □

Let $R = \operatorname{diag}(u_{i+})_i$ and $K = \operatorname{diag}(u_{+j})_j$ be the diagonal matrices recording the row and column sums of U on their diagonals. Then, by Lemma 3, $P\mathbf{1}$ is a scalar multiple of $R\mathbf{1}$ and $\mathbf{1}^T P$ is a scalar multiple of $\mathbf{1}^T K$. This implies that, in the decompositions in the proof of Lemma 5, we may take B spanned by $U\mathbf{1} = R\mathbf{1}$ and B' spanned by $U\mathbf{1} = K\mathbf{1}$. Note that P, Q satisfy $P * Q = RUK$, where $*$ denotes the Hadamard product.

Observe also that criticality of P is equivalent to $v^T R^{-1} Q K^{-1} w = 0$ for all rank-one matrices vw^T as in Lemma 5. This criterion will be used in the proof of our duality result for $\mathcal{M}_r$.

Theorem 6 (ML-duality for rectangular matrices)**.** *Let $U \in \mathbb{N}^{m\times n}$ be a sufficiently general data matrix and let P be a critical point of ℓ_U on $\mathcal{M}_r$. Define $Q = (q_{ij})_{ij}$ by $q_{ij}p_{ij} = u_{i+}u_{ij}u_{+j}$. Then $Q/(u_{++}^3)$ is a critical point of ℓ_U on $\mathcal{M}_{m-r+1}$.*

Before proceeding with the proof, we point out that the construction of $Q' := Q/(u_{++})^3$ from P is symmetric in P and Q. As a consequence, the map $P \mapsto Q'$ from critical points of ℓ_U on $\mathcal{M}_r$ to critical points on $\mathcal{M}_{m-r+1}$ is a bijection. Moreover, it has the property that $\ell_U(P)\cdot\ell_U(Q')$ depends only on U. In particular, if one lists the critical points $P \in \mathcal{M}_r$ with positive real entries in order of decreasing log-likelihood, then the corresponding $Q' \in \mathcal{M}_{m-r+1}$ appear in order of increasing log-likelihood, since the sum $\log \ell_U(P) + \log \ell_U(Q')$ depends only on U.

Proof. Lemma 4 takes care of the normalization factor, which we therefore ignore during most of this proof. We first show that Q has rank at most $m - r + 1$. For this we take arbitrary v in the space $A = \mathbf{1}^\perp \cap \operatorname{im} \mathbf{P}$ from the proof of Lemma 5 and arbitrary $w \in \mathbb{C}^n$, so that $vw^T \in T_P\mathcal{M}_r$. From $v^T R^{-1} Q K^{-1} w = 0$ we conclude that $R^{-1} \operatorname{im} Q \subseteq A^\perp$ because v was arbitrary in A. Equivalently, since R is diagonal and hence symmetric, we conclude that $\operatorname{im} Q \subseteq (R^{-1}A)^\perp$. The latter space has dimension $m - r + 1$, which is therefore an upper bound on the rank of Q.

Similarly, for $w \in A'$ and any $v \in \mathbb{C}^m$, the matrix vw^T lies in the tangent space $T_P\mathcal{M}_r$, and we find $v^T R^{-1} Q K^{-1} w = 0$. Since v was arbitrary, this means that $QK^{-1}w = 0$, so $\ker Q$ contains $K^{-1}A'$, a space of dimension $r - 1$. If $n > m$, then by the above, $\ker Q$ *strictly* contains $K^{-1}A'$, but this will be irrelevant.

Next we prove that for any rank-one matrix xy^T such that

- $x \perp R^{-1}A$ or $y \perp K^{-1}A'$; *and*
- $x \perp \mathbf{1}$ or $y \perp \mathbf{1}$

we have $\sum_{ij} \frac{x_i u_{ij} y_j}{q_{ij}} = 0$. Note that the conclusion can be written as $x^T R^{-1} P K^{-1} y = 0$, and observe the similarity with the characterization of $T_P\mathcal{M}_r$ in Lemma 5 that will give us conditions of criticality of Q.

Given arbitrary $y \in \mathbb{C}^n$ we can write $PK^{-1}y$ as $v + cR\mathbf{1}$ with $v \in A$. Then for $x \in (R^{-1}A)^\perp$ perpendicular to $\mathbf{1}$ we find

$$x^T R^{-1} P K^{-1} y = x^T R^{-1}(v + cR\mathbf{1}) = 0 + cx^T\mathbf{1} = 0,$$

as desired. If, on the other hand, $x \in (R^{-1}A)^\perp$ is not perpendicular to $\mathbf{1}$ but $y \in \mathbb{C}^n$ is, then writing $w := K^{-1}y$ we have that the vector $v := Pw$ lies in A: indeed, $\mathbf{1}^T v$ is a scalar multiple of $\mathbf{1}^T U w$ by Lemma 3, and $\mathbf{1}^T U w = \mathbf{1}^T K w = \mathbf{1}^T y = 0$. Hence, again, $x^T R^{-1} P K^{-1} y = x^T R^{-1} v = 0$, since $x \perp R^{-1}A$. The checks for the case where $y \perp K^{-1}A'$ are completely analogous.

Now denote the rank of Q by k, so that $k \leq m-r+1$. From $\operatorname{im} Q \subseteq (R^{-1}A)^\perp$ and $(\ker Q)^\perp \subseteq (K^{-1}A')^\perp$ we conclude that the derivative of ℓ_U at Q' in the direction xy^T vanishes, in particular, when xy^T lies in the tangent space at Q' to $\mathcal{M}_k$. Hence Q' is a critical point for ℓ_U on $\mathcal{M}_k$.

Finally, we need to show that the generic rank k of Q thus obtained (from a sufficiently general U and a critical point $P \in \mathcal{M}_r$ of ℓ_U) equals $m - r + 1$, rather than being strictly smaller. For this, observe that we have constructed, for any

$r \in [m]$, a rational map of irreducible varieties

$$\psi_r : \mathrm{Crit}(\mathcal{M}_r) \dashrightarrow \mathrm{Crit}(\mathcal{M}_{f(r)}), \quad (P,U) \mapsto \left(\frac{1}{(u_{++})^3} \cdot \frac{RUK}{P} \ , \ U \right) = (Q', U)$$

where $f : [m] \to [m]$ maps r to the generic rank of the matrix Q' as (P,U) varies over $\mathrm{Crit}(\mathcal{M}_r)$. Since ψ_r commutes with the projection on the second factor, its image has dimension mn, hence ψ_r is dominant. But it is also injective—in fact, (P,U) can be recovered from (Q',U) with the exact same formula. This shows that ψ_r is birational, and that $\psi_{f(r)}$ is its inverse as a birational map. In particular we have $f(f(r)) = r$, so that f is a bijection. But the only bijection $[m] \to [m]$ with the property that $f(r) \le m - r + 1$ for all r is $r \mapsto m - r + 1$. Indeed, if r were the smallest value for which $f(r) \neq m - r + 1$, then $m - r + 1$ would not be in the image of f. This concludes the proof of the theorem. $\square$

Remark 7. It *can* happen that the rank of Q is strictly smaller than $m - r + 1$ when U is not sufficiently general. For example, in the rectangular case where $m = n = 4$, if we set

$$U := \frac{1}{40} \begin{bmatrix} 4 & 2 & 2 & 2 \\ 2 & 4 & 2 & 2 \\ 2 & 2 & 4 & 2 \\ 2 & 2 & 2 & 4 \end{bmatrix} \quad \text{and} \quad P := \frac{1}{80} \begin{bmatrix} 6+2i & 5-\sqrt{5} & 5+\sqrt{5} & 4-2i \\ 5-\sqrt{5} & 6-2i & 4+2i & 5+\sqrt{5} \\ 5+\sqrt{5} & 4+2i & 6-2i & 5-\sqrt{5} \\ 4-2i & 5+\sqrt{5} & 5-\sqrt{5} & 6+2i \end{bmatrix}$$

then (P,U) lies in $\mathrm{Crit}(\mathcal{M}_2)$. Also, the number of points in $\mathrm{Crit}(\mathcal{M}_2)$ with this choice of U is equal to the ML-degree of this model. We will find that the number of critical points in $Crit(\mathcal{M}_3)$ with this non-generic choice of U is *not* equal to the ML-degree of the model. Since u_{++}=1 we have $Q = Q'$, and

$$Q = \frac{1}{500} \begin{bmatrix} 6-2i & 5+\sqrt{5} & 5-\sqrt{5} & 4+2i \\ 5+\sqrt{5} & 6+2i & 4-2i & 5-\sqrt{5} \\ 5-\sqrt{5} & 4-2i & 6+2i & 5+\sqrt{5} \\ 4+2i & 5-\sqrt{5} & 5+\sqrt{5} & 6-2i \end{bmatrix}$$

satisfies $p_{ij}q_{ij} = \frac{u_{i+}u_{++}u_{+j}}{u_{++}^3}$. In this case, Q has rank 2 instead of rank 3. This is an important fact for numerical computations. If we were to to use the homotopy methods as in [HRS12] to find the critical points of ℓ_U on $\mathcal{M}_3$, we would track a path from a generic point of $\mathrm{Crit}(\mathcal{M}_3)$ to the point (Q,U). Since Q has rank less than 3, this will correspond to tracking a path to a singularity leading to numerical difficulties. But by determining all critical points of ℓ_U on $\mathcal{M}_2$, we can use duality to avoid these numerical difficulties. That is, to determine the points of $\mathrm{Crit}(\mathcal{M}_3)$ with U as above, we use the equation $p_{ij}q_{ij} = \frac{u_{i+}u_{++}u_{+j}}{u_{++}^3}$ given by ML-duality and determine which (q_{ij}) have rank exactly 3.

3. Maximum likelihood duality in the symmetric case

Let m be a natural number and let $\mathcal{SM}_r$ denote the variety of symmetric $m \times m$-matrices of rank r whose entries sum to 2. A point P of $\mathcal{SM}_r$ and data matrix U

will be denoted by

$$P = \begin{bmatrix} 2p_{11} & p_{12} & \cdots & p_{1m} \\ p_{12} & 2p_{22} & & \\ \vdots & & \ddots & \\ p_{1m} & & & 2p_{mm} \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 2u_{11} & u_{12} & \cdots & u_{1m} \\ u_{12} & 2u_{22} & & \\ \vdots & & \ddots & \\ u_{1m} & & & 2u_{mm} \end{bmatrix}.$$

We denote the (i,j)-entries of P and U by P_{ij} and U_{ij} to distinguish them from the p_{ij} and u_{ij}, respectively. Recall that the likelihood function in the symmetric case is defined as $\ell_U(P) := \prod_{i\leq j} p_{ij}^{u_{ij}}$, which in terms of the entries of P equals $(\prod_{i<j} P_{ij}^{u_{ij}}) \cdot (\prod_i (P_{ii}/2)^{u_{ii}})$. From now on we fix a sufficiently general data matrix U and a critical point P for ℓ_U on $\mathcal{SM}_r$. The tangent space $T_P\mathcal{SM}_r$ equals

$$T_P\mathcal{SM}_r = \left\{ X \in \mathbb{C}^{m\times m} \text{ symmetric} \;\middle|\; X \ker P \subseteq \operatorname{im} P \text{ and } \sum_{ij} x_{ij} = 0 \right\}. \tag{2}$$

Given a tangent vector $X \in T_P\mathcal{SM}_r$, the derivative of ℓ_U in that direction equals

$$\sum_{i<j} \frac{X_{ij}u_{ij}}{P_{ij}} + \sum_i \frac{(X_{ii}/2)u_{ii}}{P_{ii}/2} = \sum_{i\leq j} \frac{X_{ij}u_{ij}}{P_{ij}}$$

(up to a factor irrelevant for its vanishing). We set

$$U_{i+} := \sum_j U_{ij} \text{ and } U_{++} := \sum_i \sum_j U_{ij},$$

and similarly for P. The symmetric analogue of Lemma 3 is the following.

Lemma 8. *The vector $P\mathbf{1}$ is a non-zero scalar multiple of $U\mathbf{1}$.*

Proof. We need to prove that the $m \times 2$-matrix $(P\mathbf{1}|U\mathbf{1})$ has 2×2-minors equal to zero. We prove this for the minor in the first two rows. Set $a := P_{1+}$ and $b := P_{2+}$, and define $v_1, v_2 \in \mathbb{C}^m$ by $v_1 = (b, 0, 0, \ldots, 0)^T, v_2 = (0, a, 0, \ldots, 0)$. Let w_1, w_2 be the first and second column of P, respectively. Then for each $i = 1, 2$ the matrix $X^{(i)} = v_i w_i^T + w_i v_i^T$ lies in the tangent space at P to the variety of symmetric rank-r matrices, and the difference $X := X^{(1)} - X^{(2)}$ has sum of entries equal to 0 and therefore lies in $T_P\mathcal{SM}_r$. The symmetric matrix X looks like

$$\begin{bmatrix} 2bP_{11} & (b-a)P_{12} & bP_{13} & \cdots & bP_{1m} \\ * & 2aP_{22} & -aP_{23} & \cdots & -aP_{2m} \\ * & * & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ * & * & 0 & \cdots & 0 \end{bmatrix}.$$

The derivative of ℓ_U at P in the direction X equals

$$\sum_{i\leq j} \frac{X_{ij}u_{ij}}{P_{ij}} = bU_{1+} - aU_{2+},$$

and this vanishes by criticality of P. The non-zero scalar is $\frac{P_{++}}{U_{++}} = \frac{2}{U_{++}}$. □

The analogue of R, K from the rectangular case is $R := \operatorname{diag}(U_{1+}, \ldots, U_{m+})$, which by symmetry of U equals $\operatorname{diag}(U_{+1}, \ldots, U_{+m})$. As in the rectangular case, define the symmetric matrix Q by $P * Q = RUR$, i.e., $P_{ij}Q_{ij} = U_{i+}U_{ij}U_{j+}$ for

$i, j \in [m]$. This will be our dual critical point, up to a normalizing factor to be determined now.

Lemma 9. *The sum* $\sum_{ij} Q_{ij}$ *equals* $\frac{(U_{++})^3}{2}$.

Proof. By Lemma 8 the rank-one matrix Y with entries $Y_{ij} = U_{i+}U_{j+}$ has image contained in $\operatorname{im} P$, and so does P. So we can decompose $Y = cP + X$ with $c \in \mathbb{C}$ and $X \in T_P\mathcal{SM}_r$, and we find

$$\sum_{ij} Q_{ij} = \sum_{ij} \frac{Y_{ij}U_{ij}}{P_{ij}} = \sum_{ij} cU_{ij} + \sum_{ij} \frac{X_{ij}U_{ij}}{P_{ij}} = cU_{++} + 0 = cU_{++}.$$

Moreover, the scalar c equals $\frac{Y_{++}}{P_{++}} = \frac{(U_{++})^2}{2}$, which shows that $Q_{++} = \frac{(U_{++})^3}{2}$. $\square$

As in the rectangular case, we will make use of low-rank elements in $T_P\mathcal{SM}_r$, where now "low rank" means rank two.

Lemma 10. *The tangent space* $T_P\mathcal{SM}_r$ *is spanned by all matrices of the form* $vw^T + w^Tv$ *with* $v \in im(P)$ *and* $w \in \mathbb{C}^m$*, with the additional constraint that the sum of all entries is zero, i.e., that one of* v *and* w *is perpendicular to* $\mathbf{1}$.

In the proof we will implicitly use that $\operatorname{im} P$ is not contained in $\mathbf{1}^\perp$, which is true by genericity of U.

Proof. The proof is similar to that of Lemma 5. First, the matrices in the lemma satisfy the conditions characterizing $T_P\mathcal{SM}_r$; see (2). Second, to show that they span that tangent space, split $\mathbb{C}^m$ as $A \oplus B \oplus C$ with $A \oplus B = \operatorname{im} P$ and $A \oplus C = \mathbf{1}^\perp$, so that the second symmetric power $S^2\mathbb{C}^m$ equals

$$S^2(A) \oplus S^2(B) \oplus S^2(C) \oplus (A \otimes B) \oplus (A \otimes C) \oplus (B \otimes C).$$

The matrices in the lemma span $S^2(A) + A \otimes B + (A \oplus B) \otimes C$. This space has dimension $\binom{r}{2} + (r-1) + r(n-r)$, which equals $\binom{r+1}{2} + r(n-r) - 1 = \dim \mathcal{SM}_r$. $\square$

By Lemma 10, it suffices to understand the derivative $\sum_{i \le j} \frac{X_{ij}u_{ij}}{P_{ij}}$ for X equal to $vw^T + wv^T$, in which case it equals

$$\sum_{i \le j} \frac{X_{ij}u_{ij}}{P_{ij}} = \sum_{i \le j} (v_iw_j + w_iv_j) \frac{u_{ij}}{P_{ij}} = v^T \begin{bmatrix} \frac{2u_{11}}{P_{11}} & \frac{u_{12}}{P_{12}} & \cdots & \frac{u_{1m}}{P_{1m}} \\ \frac{u_{12}}{P_{12}} & \frac{2u_{22}}{P_{22}} & & \\ \vdots & & \ddots & \\ \frac{u_{1m}}{P_{1m}} & & & \frac{2u_{mm}}{P_{mm}} \end{bmatrix} w.$$

The right-hand side can be concisely written as $v^T(\frac{U}{P})w$, where $\frac{U}{P}$ is the Hadamard (element-wise) quotient of U by P. So criticality of P is equivalent to the statement that $v^T(\frac{U}{P})w$ vanishes for all v, w as in Lemma 10. This, in turn, is equivalent to the condition that $v^TR^{-1}QR^{-1}w = 0$ for all v, w as in Lemma 10. We now state and prove our duality result in the symmetric case.

Theorem 11 (ML-duality for symmetric matrices)**.** *Let* $U \in \mathbb{N}^{m \times m}$ *be a sufficiently general symmetric data matrix, and let* P *be a critical point of* ℓ_U *on* $\mathcal{SM}_r$*. Define the matrix* Q *by* $P_{ij}Q_{ij} = U_{i+}U_{ij}U_{j+}$*. Then* $4Q/(U_{++})^3$ *is a critical point of* ℓ_U *on* $\mathcal{SM}_{m-r+1}$.

As in the rectangular case, the map $P \mapsto Q' := 4Q/(U_{++})^3$ is a bijection by virtue of the symmetry in P and Q, and the same conclusions for the cricital points with positive real entries can be drawn as in the rectangular case.

Proof. The normalizing factor was dealt with in Lemma 9 and will be largely ignored in what follows. As in the proof of Lemma 10, decompose $\mathbb{C}^m$ as $A \oplus B \oplus C$ with $A \oplus B = \operatorname{im} P$ and $A \oplus C = \mathbf{1}^\perp$. So A has dimension $r-1$, C has dimension $m-r$, and B has dimension 1. We take B to be spanned by $P\mathbf{1}$, which is a non-zero scalar multiple of $R\mathbf{1}$ by Lemma 10.

First we bound the rank of Q. To do so we prove that the image of Q is contained a space of dimension $m-r+1$. Indeed, by criticality of P we have $v^T R^{-1} Q K^{-1} w = 0$ for $w \in \mathbb{C}^m$, $v \in \operatorname{im} P$ such that $v \perp \mathbf{1}$ or $w \perp \mathbf{1}$. Taking w arbitrary and v in A, we find that $\operatorname{im} Q \subseteq (R^{-1}A)^\perp$, which has dimension $m-r+1$.

Next we show that

$$x^T R^{-1} P K^{-1} y = 0$$

for any $x \in (R^{-1}A)^\perp$ and $y \in \mathbb{C}^m$ with $x \perp \mathbf{1}$ or $y \perp \mathbf{1}$. First, suppose $x \perp \mathbf{1}$. Since $PK^{-1}y$ may be written as $a + cR\mathbf{1}$ with $a \in A$ and scalar c, we find

$$x^T R^{-1} P K^{-1} y = x^T R^{-1} a + c x^T R^{-1} R \mathbf{1} = x^T R^{-1} a + 0 = 0.$$

Otherwise, we have $y \perp \mathbf{1}$ and we may assume $x = cR\mathbf{1}$ with c a scalar. In this case, we have $x^T R^{-1} P K^{-1} y = c\mathbf{1}^T P K^{-1} y$, which by Lemma 8 equals a scalar multiple of $\mathbf{1}^T K K^{-1} y = \mathbf{1}^T y = 0$.

Let k be the rank of Q. Since $\operatorname{im} Q \subset (R^{-1}A)^\perp$ we conclude that $x^T R^{-1} P K^{-1} y = 0$ holds, in particular, for all matrices $xy^T + yx^T$ spanning the tangent space to $\mathcal{SM}_k$ at Q', so that Q' is critical. By reversing the roles of P and Q and using the involution argument at the end of the proof of Theorem 6, we conclude that for generic U the value of k equals $m-r+1$ (rather than being strictly smaller). This proves the theorem. □

4. Duality in the skew-symmetric case

The skew-symmetric case, while perhaps not of direct relevance to statistics, is of considerable algebro-geometric interest [HKS05], since the variety $\mathcal{AM}_r$, consisting of skew-symmetric matrices of *even* rank r whose upper-triangular entries are non-zero and add up to 1, is (an open subset of a hyperplane section of the affine cone over) a secant variety of the Grassmannian of 2-spaces in $\mathbb{C}^m$. Recall that we want to prove that $\mathcal{AM}_r$ (the *intersection* of a determinantal variety with an affine hyperplane) is ML-dual to the *affine translate* $\mathcal{AM}'_s$ of a determinantal variety.

A point P of $\mathcal{AM}_r$ and data matrix U will be denoted by

$$P = \begin{bmatrix} 0 & p_{12} & \cdots & p_{1m} \\ -p_{12} & 0 & & \\ \vdots & & \ddots & \\ -p_{1m} & & & 0 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 0 & u_{12} & \cdots & u_{1m} \\ u_{12} & 0 & & \\ \vdots & & \ddots & \\ u_{m1} & & & 0 \end{bmatrix}.$$

Note that U is *symmetric* rather than alternating. We fix a sufficiently general data matrix U and a critical point P for ℓ_U on $\mathcal{AM}_r$. The tangent space $T_P\mathcal{AM}_r$ equals

$$T_P\mathcal{AM}_r = \left\{ X \in \mathbb{C}^{m\times m} \text{ skew} \;\middle|\; X \ker P \subseteq \operatorname{im} P \text{ and } \sum_{i<j} x_{ij} = 0 \right\}.$$

The derivative of ℓ_U at P in the direction X equals $\sum_{i<j} \frac{x_{ij}u_{ij}}{p_{ij}}$, up to a factor irrelevant for its vanishing. The following lemma is the skew analogue of Lemmas 3 and 8.

Lemma 12. *The vector $a = \left(\sum_{j<i} p_{ji} + \sum_{j>i} p_{ij}\right)_i$ is a scalar multiple of $U\mathbf{1}$.*

Proof. We need to show that 2×2-minors of the matrix $(a|U\mathbf{1})$ are zero, and do so for the first minor. Let v_1, v_2 be the first and second column of P, respectively, and set $w_1 := (a_2, 0, \ldots, 0)$ and $w_2 := (0, -a_1, 0, \ldots, 0)$. Then each of the matrices $v_i w_i^T - w_i v_i^T$ is tangent at P to the variety of skew-symmetric rank-r matrices, and their sum

$$X = \begin{bmatrix} 0 & (a_2 - a_1)p_{12} & a_2 p_{13} & \cdots & a_2 p_{1m} \\ -(a_2 - a_1)p_{12} & 0 & -a_1 p_{23} & \cdots & -a_1 p_{2m} \\ -a_2 p_{13} & a_1 p_{23} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -a_2 p_{1m} & a_1 p_{2m} & 0 & \cdots & 0 \end{bmatrix}$$

has upper-triangular entries adding up to 0, so that X is tangent at P to $\mathcal{AM}_r$. The derivative of ℓ_U at P in the direction X, which is zero by criticality of P, equals

$$(a_2 - a_1)u_{12} + a_2 u_{13} + \ldots + a_2 u_{1m} - a_1 u_{23} - \ldots - a_1 p_{2m} = a_2 u_{1+} - a_1 u_{2+},$$

which is the minor whose vanishing was required. $\square$

Next we determine rank-two elements spanning $T_P\mathcal{AM}_r$. For this we introduce the skew bilinear form $\langle .,. \rangle$ on $\mathbb{C}^m$ defined by $\langle v, w \rangle = v^T S w = \sum_{i<j}(v_i w_j - v_j w_i)$, where S is the skew-symmetric matrix

$$S = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ -1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ -1 & \cdots & -1 & 0 \end{bmatrix}$$

from the introduction. By elementary linear algebra, this form is non-degenerate if m is even and has a one-dimensional radical spanned by $(1, -1, 1, -1, \ldots, 1) \in \mathbb{C}^m$ if m is odd.

In what follows, it will be convenient to think of skew-symmetric matrices also as elements of $\bigwedge^2 \mathbb{C}^m$ or as alternating tensors.

Lemma 13. *The tangent space $T_P\mathcal{AM}_r$ is spanned by skew-symmetric matrices of the form $vw^T - wv^T$ with $v \in \operatorname{im} P$ and $\langle v, w \rangle = 0$.*

In the proof we will use that $\operatorname{im} P$ is non-degenerate with respect to $\langle .,. \rangle$. This condition will be satisfied for general U.

Proof. The proof is similar to the symmetric case and the rectangular case: a skew-symmetric matrix X lies in the tangent space if and only if $X \ker P \subseteq \operatorname{im} P$ and $\sum_{i<j} x_{ij} = 0$. The condition $v \in \operatorname{im} P$ ensures the first property and the condition that $\langle v, w \rangle = 0$ ensures the second property.

To complete the proof, decompose $\mathbb{C}^m$ as $A \oplus C$ with $A = \operatorname{im} P$ and $\langle A, C \rangle = 0$, so that $\bigwedge^2 \mathbb{C}^m$ decomposes as $\bigwedge^2 A \oplus (A \otimes C) \oplus \bigwedge^2 C$. Taking the vector w in $v^T w - wv^T$ from C we see that $A \otimes C$ is contained in the span of the matrices in the lemma. Next we argue that a codimension-one subspace of $\bigwedge^2 A$ is also contained in their span. Indeed, the (non-zero) tensors $v^T w - wv^T \in \bigwedge^2 A$ with $v, w \in A$ perpendicular with respect to $\langle ., . \rangle$ form a single orbit under the symplectic group $\mathrm{Sp}(A) = \mathrm{Sp}_r$ (recall that r is even, so that this is a reductive group), and hence their span is an $\mathrm{Sp}(A)$-submodule of $\bigwedge^2 A$. But $\bigwedge^2 A$ splits as a direct sum of only two irreducible modules under $\mathrm{Sp}(A)$: a one-dimensional trivial module corresponding to (the restriction of) $\langle ., . \rangle$ and a codimension-one module. Hence the tensors $v^T w - wv^T$ must span that codimension-one module.

Summarizing, we find that the matrices in the lemma span a space of dimension $r(n-r) + \binom{r}{2} - 1$, which equals $\dim \mathcal{AM}_r$. □

Recall that in the alternating case the likelihood function is given by $\ell_U(P) = \prod_{i<j} p_{ij}^{u_{ij}}$. The derivative of this expression in the direction of a skew-symmetric matrix X of the form $vw^T - wv^T$ equals (up to a factor irrelevant for its vanishing)

$$\sum_{i<j} x_{ij} \frac{u_{ij}}{p_{ij}} = \sum_{i<j} \frac{u_{ij}}{p_{ij}} (v_i w_j - v_j w_i) = v^T \begin{bmatrix} 0 & \frac{u_{12}}{p_{12}} & \cdots & \frac{u_{1m}}{p_{1m}} \\ -\frac{u_{12}}{p_{12}} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{u_{m-1,m}}{p_{m-1,m}} \\ -\frac{u_{1m}}{p_{1m}} & \cdots & -\frac{u_{m-1,m}}{p_{m-1,m}} & 0 \end{bmatrix} w.$$

Define the skew matrix Q by $P * Q = U$. Then criticality of P translates into $v^T Q w = 0$ for all $v \in \operatorname{im} P$ and $w \in \mathbb{C}^m$ with $\langle v, w \rangle = 0$.

Theorem 14 (ML-duality for skew matrices)**.** *Let $U = (u_{ij})_{ij}$ be a sufficiently general symmetric data matrix with zeroes on the diagonal, and let P be a critical point of ℓ_U on $\mathcal{AM}_r$, where $r \in \{2, \ldots, m\}$ is even. Let $s \in \{0, \ldots, m-2\}$ be the largest even integer less than or equal to $m - r$. Define the matrix Q by $P * Q = U$. Then the skew matrix $Q' := 2Q/U_{++}$ is a critical point of ℓ_U on the translated determinantal variety $\mathcal{AM}'_s$. Moreover, the map $P \to Q'$ is a bijection between the critical points of ℓ_U on $\mathcal{AM}_r$ and those $\mathcal{AM}'_s$.*

As in the rectangular and symmetric cases, the bijection $P \to Q'$ maps real, positive critical points to real, positive critical points in such a way that the sum of the log-likelihoods of P and Q' is constant.

Proof. By construction of Q we have $v^T Q w = 0$ for all $v \in \operatorname{im} P$ and $w \in \mathbb{C}^m$ with $v^T S w = 0$. This means that the quadratic form $(v, w) \mapsto v^T Q w$ on $\operatorname{im} P \times \mathbb{C}^m$ is a scalar multiple of the quadratic form $(v, w) \mapsto v^T S w$, denoted $\langle ., . \rangle$ earlier, on that same space. The scalar is computed by computing

$$(0, -p_{12}, \ldots, -p_{1m}) Q (1, 0, \ldots, 0)^T = U_{1+}$$

and

$$(0, -p_{12}, \ldots, -p_{1m}) S (1, 0, \ldots, 0)^T = P_{1+} = a_1,$$

where a is the vector of Lemma 12. Using that lemma and the fact that $\sum_i a_i = 2$ we find that $a_1 = 2U_{1+}/U_{++}$. We conclude that the skew bilinear form associated to $B := S - \frac{2}{U_{++}} Q$ is identically zero on $\operatorname{im} P \times \mathbb{C}^m$, hence $\ker B$ contains $\operatorname{im} P$ and $\operatorname{im} B = (\ker B)^\perp$ (where $\perp$ refers to the standard bilinear form on $\mathbb{C}^m$) is contained in $\ker P = (\operatorname{im} P)^\perp$. In particular, B has rank at most s; let $k \leq s$ denote the actual rank of B.

Next we argue that $Q' := \frac{2}{U_{++}} Q$ is critical for ℓ_U on $\mathcal{AM}'_k$. By arguments similar to (but easier than) those in Lemma 13 the tangent space $T_{Q'}\mathcal{AM}'_k$ is spanned by rank-two matrices $vw^T - wv^T$ with $v \in \operatorname{im} B$ and $w \in \mathbb{C}^m$ arbitrary. Thus proving that Q' is critical boils down to proving that $v^T P w = 0$ for all $v \in \operatorname{im} B$ and $w \in \mathbb{C}^m$. But this is immediate from $\operatorname{im} B \subseteq \ker P$. Thus Q' is critical.

Furthermore, we need to show that (for generic U) the rank k of $B = S - Q'$ is equal to s rather than strictly smaller, and that the map $P \mapsto Q'$, which is clearly injective, is also surjective on the set of critical points for ℓ_U on $\mathcal{AM}'_s$. For these purposes we reverse the arguments above: assume that Q' is a critical point on $\mathcal{AM}'_k$, where k is an even integer in the range $\{0, \dots, m-2\}$. Define $Q := \frac{U_{++}}{2} Q'$ and define P by $P * Q = U$. Also, define $B := S - Q'$. Then criticality of Q' implies that $v^T P w = 0$ for all $v \in \operatorname{im} B$ and $w \in \mathbb{C}^m$, and this implies that $\ker P \supseteq \operatorname{im} B$. Thus $l := \operatorname{rk} P$ is at most $m - k$.

Moreover, B itself lies in the tangent space $T_{Q'}\mathcal{AM}'_k$, and criticality of Q' implies that $\sum_{i<j} B_{ij} \frac{U_{ij}}{Q_{ij}} = 0$. Substituting the expression for B into this we find that

$$0 = \sum_{i<j} (1 - \frac{2}{U_{++}} Q_{ij}) \frac{U_{ij}}{Q_{ij}} = \sum_{i<j} (P_{ij} - \frac{2}{U_{++}}) = (\sum_{i<j} P_{ij}) - 1,$$

i.e., the upper-triangular entries of P add up to one. We conclude that P lies in $\mathcal{AM}_l$. Next, we argue that P is critical. Indeed, for $v \in \operatorname{im} P$ and $w \in \mathbb{C}^m$ such that $\langle v, w \rangle = (v^T S w =)0$ we find

$$v^T Q w = v^T (\frac{U_{++}}{2}(S - B)) w = \frac{U_{++}}{2}(v^T S w - v^T B w) = 0 + 0 = 0,$$

where we have used that $\operatorname{im} P \subseteq \ker B$.

Summarizing, we have found rational maps

$$\psi_r : \operatorname{Crit}(\mathcal{AM}_r) \dashrightarrow \operatorname{Crit}(\mathcal{AM}'_{f(r)}), \qquad (P, U) \mapsto (\frac{2}{U_{++}} \cdot \frac{U}{P}, U) = (Q', U) \text{ and}$$

$$\psi'_k : \operatorname{Crit}(\mathcal{AM}'_k) \dashrightarrow \operatorname{Crit}(\mathcal{AM}_{g(k)}), \qquad (Q', U) \mapsto (\frac{2}{U_{++}} \cdot \frac{U}{Q'}, U)$$

for some map f mapping even integers $r \in \{2, \dots, m\}$ to even integers $k \in \{0, \dots, m-2\}$, and some map g in the opposite direction. By the argument in the proof of Theorem 6, both ψ_r and ψ'_k are birational and $g(f(r)) = r$. Hence f is a bijection, and by the above it satisfies $f(r) \leq m - r$. The only such bijection is the one that maps r to the largest even integer less than or equal to $m - r$. This concludes the proof of the theorem. $\square$

Example 15. Now we give an explicit example illustrating dual solutions in the alternating case. For $m = 4$ the ML-degree of $\mathcal{AM}_2$ is 4 [HKS05]. Setting

$$U = \frac{1}{41} \begin{bmatrix} 0 & 2 & 3 & 5 \\ 2 & 0 & 7 & 11 \\ 3 & 7 & 0 & 13 \\ 5 & 11 & 13 & 0 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} 0 & 0.0386 & 0.0978 & 0.1075 \\ -0.0386 & 0 & 0.1563 & 0.2929 \\ -0.0978 & -0.1563 & 0 & 0.3069 \\ -0.1075 & -0.2929 & -0.3069 & 0 \end{bmatrix},$$

we have that P is a critical point of ℓ_U on $\mathcal{AM}_2$ and $U_{++} = 2$. Having Q defined as $P * Q = U$, we find that $Q(= Q')$ has full rank. But in the alternating case the ML-dual variety is an affine translate of a determinantal variety. We find that $B = S - Q$ equals

$$B = \begin{bmatrix} 0 & -0.2638 & 0.2518 & -0.1344 \\ 0.2638 & 0 & -0.0924 & 0.0841 \\ -0.2518 & 0.0924 & 0 & -0.0332 \\ 0.1344 & -0.0841 & 0.0332 & 0 \end{bmatrix},$$

and indeed B has rank $4 - 2 = 2$. We can actually compute the ML-degree of $\mathcal{AM}'_2$ symbolically to be 4 (even with the u_{ij} treated as symbols). For the data matrix U above, the minimal polynomial for q_{34} equals $434217q_{34}^4 - 1335767q_{34}^3 + 1536717q_{34}^2 - 764049q_{34} + 127426$.

5. Conclusion

We have proved that a number of natural determinantal varieties of matrices are *ML-dual* to other such varieties living in the same ambient spaces. However, we have done so without formalizing what exactly we mean by ML-duality. It would be interesting to find a satisfactory general definition, perhaps involving the condition that $(P, U) \mapsto (\frac{U}{P}, U)$, or some variant of this that takes marginals into account, is a birational map between the two varieties of critical points. Given such a definition, it would be great to discover new ML-dual pairs of varieties, for instance so-called *subspace varieties* [LW07] or varieties of consisting of *tensors* of given (border) rank. Lastly, we note that the problem of finding a formula for ML-degrees of matrix models remains wide open, though ML-duality has essentially cut this problem in half.

References

[Har92] Joe Harris. *Algebraic Geometry. A First Course.* Number 133 in Graduate Texts in Mathematics. Springer-Verlag, Berlin etc., 1992.

[HKS05] Serkan Hoşten, Amit Khetan, and Bernd Sturmfels. Solving the likelihood equations. *Found. Comput. Math.*, 5(4):389–407, 2005.

[HRS12] Jonathan Hauenstein, Jose Rodriguez, and Bernd Sturmfels. Maximum likelihood for matrices with rank constraints. 2012. Preprint, available from `http://arxiv.org/abs/1210.0198`.

[Huh12] June Huh. The maximum likelihood degree of a very affine variety. 2012. Preprint, available from `http://arxiv.org/abs/1207.0553`.

[LW07] Joseph M. Landsberg and Jerzy Weyman. On the ideals and singularities of secant varieties of Segre varieties. *Bull. Lond. Math. Soc.*, 39(4):685–697, 2007.

[BHSW06] Daniel J. Bates, Jonathan D. Hauenstein, Andrew J Sommese, and Charles W. Wampler. Bertini: Software for Numerical Algebraic Geometry. Available at `http://www.nd.edu/~sommese/bertini`.

[FK00] J. Franecki and M. Kapranov. The Gauss map and a noncompact Riemann-Roch formula for constructible sheaves on semiabelian varieties *Duke Mathematical Journal* 104:171–180, 2000.

(Jan Draisma) Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, and Centrum voor Wiskunde en Informatica, Amsterdam, The Netherlands

E-mail address: `j.draisma@tue.nl`

(Jose Rodriguez) Department of Mathematics, University of California at Berkeley, 970 Evans Hall 3840, Berkeley, CA 94720-3840 USA

E-mail address: `jo.ro@berkeley.edu`